Lipid and sterol gene sequence variation in autism and correlates with neurodevelopmental status: A pilot study

Trevor A. Hall *, Robert D. Steiner 1, Hollis Wright, Beth Wilmot, Jean-Baptiste Roullet 2, Meaghan Peters, Michael Harris

Oregon Health & Science University, 707 SW Gaines Street, Portland, OR 97239-3098, United States

ABSTRACT

Objective: Research has uncovered potential links between lipid and sterol metabolism and autism spectrum disorder (ASD). We worked to characterize genetic sequence variants in lipid/sterol related genes in children affected with ASD to investigate the association between lipid/sterol gene sequence variation and neurodevelopmental phenotype that could identify new etiologies for ASD and eventually aid to focus intervention strategies.

Design and methods: Children with confirmed ASD were recruited from a regional academic health center. Participants included 24 children (20 male and 4 female) between the ages of 40 to 81 months (M=60 months). Several neurodevelopmental measures were administered which provided an assessment of neurodevelopmental functional status. We applied our exome sequencing workflow to perform alignment to the Human Reference Sequence (build 37) base calling for every base pair in the reads that align to the reference sequence QC evaluation of the annotation of all genetic variants different from the reference sequence using dbSNP and the 1000 Genomes databases. We investigated whether novel variants identified were related to neurodevelopmental functioning.

Results and conclusions: Variants occurred in 355 total genomic positions, 53 of which were not previously annotated as variant positions in either dbSNP or the 1000 Genomes Project’s variant annotation. Of these 355 variants, 169 were nonsynonymous (31 were novel). The total number of variants observed in the exons of captured regions of an individual participant ranged from 88 to 117; novel variants ranged from four to 10 per participant, while nonsynonymous variants ranged from 36 and 51 per participant. The total number of nonsynonymous variants per subject was significantly associated with neurodevelopmental function. Further, several genes involved in sterol and lipid metabolism including NPC1, DHCR24 and others that when mutated cause diseases with ASD characteristics, were associated with ASD in Network analysis. Altogether, the findings suggest that nonsynonymous variants in lipid/sterol related genes may be a biological marker of neurodevelopment status in ASD. Results support an association between lipid and sterol metabolism and ASD and suggest the need for further research attempting to elucidate the mechanisms behind the association and the etiology and neurodevelopmental effects of ASD.

Focal points:

• Bedside
  Understanding the association between genetics and metabolism and ASD will contribute to the scientific understanding of complex neurodevelopmental disorders. Continued study into the association between lipid and sterol genes and ASD may lead to new novel and effective treatment options for ASD.

• Benchside
  A large number of genetic sequence variations are associated with ASD. Additional research is still needed to determine which associations contribute to various subtypes of ASD presentations, and to elucidate the mechanisms behind the observed association.

*This work was completed when all authors were at Oregon Health & Science University, Portland, OR, United States.
* Corresponding author.
E-mail address: halltr@ohsu.edu (T.A. Hall).
  1 Current address: Marshfield Clinic Research Foundation and University of Wisconsin, School of Medicine and Public Health, 1000 North Oak Avenue, Marshfield, WI 54449, United States.
  2 Current address: Washington State University, College of Pharmacy, 412 E. Spokane Falls, Blvd., Spokane, WA 99210-1495, United States.

http://dx.doi.org/10.1016/j.nhtm.2015.09.001
2307-5023/© 2015 European Society for Translational Medicine. Published by Elsevier Ltd. All rights reserved.
1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder defined by impairments in the areas of social communication, as well as patterns of restricted or repetitive interests and behaviors [3]. In the United States the prevalence of ASD has steadily increased from an estimated 1 in 150 individuals [9,15,24,38] to the current estimate of 1 in 68 individuals [8]. With the increasing number of children identified with ASD, it has become important to promote research that seeks to better understand the etiology of the disorder so that novel medical treatments and refined neurobehavioral intervention strategies may be developed, tested, applied, and then monitored longitudinally.

One avenue for research that has begun to receive attention is investigation of metabolic derangements in ASD. A growing number of genetic disorders are associated with ASD, including but not limited to 1q21 duplication, 3p deletion/duplication, 7q11 duplication, 15q duplication, 15q13 deletion, 16p11 deletion, 22q11 deletion, 22q13 deletion, 15q11–13 (Angelman syndrome), 11p15 (Beckwith-Weidemann syndrome), 7q35-36 (Cortical dysplasia focal epilepsy), 10q23 (Cowden-Bannayan Riely–Ruvalcaba syndrome), trisomy 21 (Down syndrome), 15p11 (Potocki–Lupski Duplication), 15q11–13 (Prader–Willi syndrome), Xq26 (Rett syndrome), 12p13 (Timothy syndrome), and 9q34 & 16p13 (Tuberous sclerosis) [1,10,16,21,23,25,35,43,49,51,52]. Moreover, ASD has also been associated with a few metabolic diseases including mucopolysaccharidosis type 3, untreated phenylketonuria, adenylosuccinate lyase deficiency, and Smith–Lemli–Opitz syndrome [10,41,46].

Smith–Lemli–Opitz syndrome (SLOS) is a multiple congenital anomalies/intellectual disability syndrome with microcephaly, dysmorphism, growth impairment, limb anomalies, and often eye, heart, genitourinary and other anomalies as key features [22,34,42]. SLOS is a disorder of cholesterol synthesis, with resulting cholesterol deficiency. Core ASD symptoms (i.e., problems with communication and socialization) are almost universal in individuals with SLOS [41] and with the growing associations between SLOS and ASD and cholesterol deficiency and ASD, [5,46] the role of cholesterol and more broadly sterol and lipid metabolism in the etiology of ASD is important to understand. The major roles for cholesterol in myelination, synaptogenesis, and neurosteroid synthesis together with the demonstration of ASD in SLOS and serotonin abnormalities in a mouse model for SLOS [18,31,32,33] provide a strong rationale for the study of cholesterol and sterol related genes in ASD.

It is important to investigate why the sterol disorder SLOS and cholesterol deficiency are associated with ASD, and whether genes involved in lipid and sterol metabolism may play a role in the etiology of ASD and/or in its neurodevelopmental correlates. Because a defect in cholesterol synthesis is the cause of SLOS, it is questioned whether a subgroup of children with ASD who do not have SLOS might have a subtle variation in their lipid or sterol related genome. Despite its behavioral similarities, the extent to which SLOS may share neurobiological mechanisms with non-SLOS autism remains unclear. Although we did not expect undiagnosed SLOS to be present in significant numbers of individuals with ASD selected from a general ASD clinic, we did hypothesize that subtle perturbations identifiable in lipid and sterol related genes would be found in some individuals with an ASD who do not have SLOS.

We investigated lipid and sterol related genes in children with ASD in an effort to determine the association between lipid and sterol gene sequence variation and ASD neurodevelopmental phenotypes that could eventually help elucidate etiologies of ASD and its neurodevelopmental correlates and aid to focus treatment intervention strategies.

2. Material and methods

2.1. Participants and procedures

Children with ASD were recruited from the Child Development and Rehabilitation Center’s (CDRC) Autism Program at Oregon Health & Science University (OHSU). Eligibility criteria included: 1) children aged 24 months to 84 months; 2) a medical diagnosis of Autistic Disorder [2] which was clinically rendered at the CDRC by a multidisciplinary diagnostic team of professionals; 3) scores on the Autism Diagnostic Observation Scale–Generic (ADOS-G; [36]) had to meet or exceed cutoffs for categorization of “Autism” (all staff who administered the ADOS-G had gained research-reliable status with a certified reliability trainer); and 4) the parents of the children needed to speak fluent English and give informed consent. Children signed assent forms when applicable. This study was carried out under an Institutional Review Board approved protocol.

A medical record review yielded 129 potentially interested parents/caregivers of eligible children. Following a phone conversation, 68 children were scheduled for neurodevelopmental assessment and a blood draw; however, 18 of these participants voluntarily withdrew because of the required blood draw. Therefore, a total of 50 children completed the neurodevelopmental and blood draw portions of the study. After the collection of biological samples, only 24 of those samples were deemed usable for exome sequencing. As such, participants included 24 children (20 male and 4 female) between the ages of 40 to 81 months (M=60 months). Upon successful completion of the initial eligibility requirements, the child and his/her family were scheduled for a second study visit to complete the neurodevelopmental assessment. From there, the children participated in a study visit at the...
Oregon Clinical and Translational Research Institute where approximately 14 milliliters of blood was drawn from the arm (most commonly the median cubital vein) in each participant. Blood draws were completed by a phlebotomist who had experience working with young children and individuals with neurodevelopmental disabilities.

2.2. Measures

As previously indicated, the ADOS-G was used to establish initial eligibility. The ADOS-G is a semi-structured, standardized, play-based assessment measure and consists of a set of activities that assess social interaction, communication, play skills, and repetitive and stereotyped behaviors. Cutoff scores in the domains of Communication, Social Interaction, and Combined (Communication + Social Interaction), allow an individual to be placed in a (n) Autism, ASD, or non-spectrum category in each of these three areas [28].

The Merrill–Palmer Scale-Revised (MP-R) [37] is the revision to the 1931 Merrill–Palmer Scale. The MP-R is comprised of a Cognitive Battery which includes general cognitive, memory, speed of cognition, receptive language, visual motor, and fine motor domains; the MP-R also has a Gross Motor section. Moreover, the MP-R includes co-normed parent report measures of social-emotional development, expressive language, and self-help/adaptive behaviors. This revision of the Merrill–Palmer follows the natural developmental progression of activities for children from birth to age 78 months. The MP-R provides developmental growth scores as well as criterion-referenced standard scores. Also provided is a Developmental Index for all assessment scales related to neurodevelopmental status. The scales which psychometrically load to the Developmental Index include Cognitive, Fine Motor, Receptive Language, Memory, Speed, and Visual Motor. The MP-R has been field tested extensively over a five-year period for both content and construct validity as well as for fairness of assessment [19,37].

Vineland Adaptive Behavior Scales, Second Edition: Parent Report Edition (VABS-II) [44]. The VABS-II is a widely used standardized caregiver interview designed to measure adaptive behavior in children from birth to 9 years, 11 months. It consists of several items falling in four general domains of functioning: communication, daily living skills, socialization, and motor.

PDD Behavior Inventory (PDDBI) [13]. The PDDBI is a caregiver-report rating scale designed to assess both adaptive and maladaptive behaviors in children with ASD. The parent version of the inventory consists of 176 items organized into 10 subscales. The Maladaptive subscales are sensory/perceptual approach behaviors, specific fears, arousal problems, aggressiveness, social pragmatic problems, and semantic/pragmatic problems. The Adaptive scales are social approach behaviors, learning/memory/receptive language, phonological skills, and semantic/pragmatic ability. Evidence for convergent validity with the ADI-R, Vineland and other measures has been found [11,12].

2.3. Sequencing methods and statistical analysis

Centrifugation occurred at +4 °C and plasma aliquots were stored at −80 °C until analysis. DNA was isolated from a freshly drawn 3 ml EDTA tube using Puregene DNA isolation kits (Qiagen, Valencia, CA 91355). All remaining blood and DNA obtained from the sample were banked indefinitely for future research. DNA samples were sequenced on an Illumina HiSeq2000 using the NimbleGen SeqCap EZ exome v2.0 capture library. We applied our exome sequencing workflow to perform alignment to the Human Reference Sequence (build 37) base calling for every base pair in the reads that align to the reference sequence QC evaluation of the

annotation of all genetic variants different from the reference sequence using dbSNP and the 1000 Genomes databases (Fig. 1).

In terms of alignment and case calling, reads were aligned to the hg19 build of the human genome using the Bowtie 0.12.7 aligner. Only reads that aligned uniquely to the reference sequence were retained. Aligned reads were then processed using samtools 0.1.15 into consensus genotypes.

Quality assessment of the coverage of the aligned reads and the quality of genotype calls was performed using the Bioconductor package ShortRead 1.10.4 in R 2.13.0. For quality filtering, we defined the minimum consensus quality and coverage criteria for the assembly of the exome data at a given genomic position for a confident genotype call to be made to be a Phred-scaled consensus quality score of > 20 (giving a probability of .01 or less of a miscall at the position) and coverage > 10 reads at a given base. We refer to such bases as high-quality.

Only high-quality bases were considered for further annotation. After quality filtering, high-quality genotypes that were variant with regard to the reference sequence were annotated for functional effects using the Annovar package (www.openbioinformatics.org/annovar), version 06-18-2011). Functional annotation included coding and protein alterations relative to the RefSeq transcript (hg19) for the gene or genes affected. Variations not present in either dbSNP (Build 132) or the Phase 1 release of the 1000 Genomes Project data were considered to be novel and are indicated in the provided results tables as “unannotated” (i.e., these variants were not present in the current versions of dbSNP and 1000 genomes databases).

We determined whether any of the novel variants were common to multiple participants in the study. Additionally, we investigated whether variants were related to neurodevelopmental function in ASD by using a two-sample Wilcoxon test. For each novel variant, we compared neurodevelopmental function between carriers of the variant and those participants in whom the variant was not present. We examined whether there was a correlation between overall number of variants or novel and/or nonsynonymous variants in the sequenced genes and neurodevelopmental severity using a Kendall ranked correlation test.

2.4. Genetic, physical and pathway analysis of lipid and sterol genes with nonsynonymous variants

We mapped our sterol and lipid-related gene set into the GeneMANIA [50] gene network creation software, using the GeneMANIA plugin for the Cytoscape [40] network visualization software, version 3.1.1. GeneMANIA creates interaction networks between genes in an input gene set based on evidence aggregated from publications in the literature and interaction databases. A gene interaction network was created based on all genetic
interaction, physical interaction and pathway data sets included in the database, using GeneMANIA’s automatic weighting method. After inference, the network’s properties such as node degree and connectivity were analyzed using Cytoscape’s Network Analyzer function. We then performed a Poisson regression predicting to the number of nonsynonymous variants (novel and known) in each gene based the node degree of that gene in the inferred network for all genes with at least one edge.

2.5. Weighted gene co-expression network analysis clustering and regression of neurodevelopmental function of per-gene nonsynonymous variant load in steroid and lipid genes

We applied Weighted Gene Coexpression Network Analysis (WGCNA, [53]) to cluster the 40 members of our lipid and sterol gene set possessing nonsynonymous variants based on the per-individual correlations of numbers of nonsynonymous variants in each gene. Variant load refers to the total number of non-reference bases resulting in a nonsynonymous substitution located in a gene or group of genes of interest in a given individual. The number of variant alleles at a given position was not incorporated into variant load; only the presence of at least one variant allele was considered in this analysis. While WGCNA is normally applied to discover co-expression clusters or “modules” of genes whose messenger RNA levels are tightly correlated in gene expression data sets, we reasoned that modules of genes whose nonsynonymous variant loads were correlated might represent gene sets with shared functions or predictive power relevant to ASD neurodevelopmental functioning. Kendall’s rank correlations were used since absolute differences in total number of variants between genes were small and calculated between each pair of genes to produce a correlation matrix. After module inference using a power value of 1 and a minimum module size of 2, 6 modules were produced. We then performed a Kendall ranked correlation test of correlation between nonsynonymous variant loads of any individual gene or subset of genes with ASD severity, as well as each pair and each trio of genes.

2.6. Overlap of sterol gene set with a variant-enriched gene set in a parent-sibling study of ASD

A recent whole genome sequencing study by Yuen et al. [48] of ASD siblings and parents found that the presence of higher numbers of sequence variants associated with a subset of genes known as PhHs_MindFun_All was correlated with stronger severity of ASD symptoms in their subjects. We compared the overlap of our sterol and lipid gene list with the PhHs_MindFun_All gene set as compared to all genes in the HUGO-annotated human gene set using a hypergeometric test.

3. Results

We set out to characterize genetic variants in lipid and sterol related genes (Table 1) in 24 participants affected with ASD. The sample is predominantly male, reflecting the demographics of ASD. The majority of the children are white (87%) and non-Hispanic (4% identified as being of Hispanic ethnicity and 8% identified as being of Other ethnicity). One third (33%) of the sample has an MP-R Developmental Index greater than 70. Table 2 shows the neurodevelopmental characteristics of the 24 children. In addition, the MP-R Developmental Index was significantly associated with VABS-II Communication Total and the VABS-II Motor Skills Total. The VABS-II Communication, Socialization, and Motor Skills Totals were associated with PDDBI Total. Because higher scores on the PDDBI indicate more problems, the results indicate that lower neurodevelopmental skills are related to more ASD related problems as measured by the PDDBI (see Table 3). We assessed the data for the amount of targeted exome capture regions covered by high-quality sequence. Median high-quality coverage of a given target region was 100% for all lanes, while the total percentage of target regions with at least one base of high-quality coverage was 99.7% for all lanes. No batch or lane effects were apparent.

Variants occurred in 355 total genomic positions, 53 of which were not previously annotated as variant positions in either dbSNP or the 1000 Genomes Project’s variant annotation (Table 4). Of

---

Table 1

| Sterol/lipid related genes which were sequenced/probed. |
|--------------------------------------------------------|
| Gene (Location) | Gene (Location) | Gene (Location) |
|----------------|----------------|----------------|
| SC5DL (11q23.3) | SREBF2 (22q13) | EBP (Xp11.23-p11.22) |
| SCP2 (1p32) | SREBF1 (17p12) | SCAMOL (4q32-q34) |
| EBLP (13q12-q13) | STARD4 (5q21.1) | CYP8B1 (3p22-p23.1) |
| LOC120498 (11q23.3) | LOC404297 (16q21) | CEL (9q43.3) |
| SCAP (3p21.3) | SOAT1 (1q25) | CYP51A1 (7q22.2-q23.1) |
| C20orf79 (20p11.13) | SOAT2 (12q13.13) | LPR (10q22.3-22q3.3) |
| CYP7A1 (2p31-qter) | LOC19902 (11q13.1) | LOC119901 (11q13.1) |
| LOC563793 (22q13.2) | LOC403548 (Xp21) | LOC219548 (13q12.2) |
| LOC1504367 (22q13.2) | LOC145035 (16q21) | LOC139151 (Xp11.3) |
| LMAN1 (12q13) | LOC72917 (16q21) | LOC145099 (13q12.3) |
| HSD17B4 (5q21) | TRHDIE (14q5-q21) | SCAMP5 (Xp21.1) |
| NPC1 (18q11-q12) | CH2SH (10q23) | STOML1 (15q24-q25) |
| HSD1L2 (19q22) | C14orf1 (14q24.3) | SQLE (8q24.1) |
| ARCA1 (9q11.1) | STARDS3 (17q11-q12) | STARDS5 (15q25) |
| ARCG5 (2p21) | CSorF4 (5q31-q32) | SDR5A2 (2p23) |
| ARG11F10 (1p36.13) | ARCG8 (2p21) | NPC2 (14q24.3) |
| STARD6 (18q21.2) | PCSK9 (1p32.2) | INSG1 (7q36) |
| OSBP (11q2-q13) | ARV1 (1q42) | FF16237 (2p11) |
| OSTalpha (3q29) | ACS52 (2q11.22) | EGF (4q25) |
| SCABR1 (12q24.31) | NSDD1 (Xp28) | OSP1L2 (20q13.3) |
| MBTS1 (16q24) | STAR (8p11.22) | DHC17A2 (p33-p31) |
| INS (11p15.5) | TWIST2 (2q37.3) | MVK (12q24) |
| ACAA2 (18q11.1) | CYP7B1 (8q21.3) | FADS2 (11q2-q13.1) |
| NR0B2 (1p31.1) | NRAS2 (4q21) | BTVN2A (6p23.1) |
| ID3 (1p36.13-p36.12) | FOXO3 (9p21) | CEB1 (Xp11.24) |
| OSTAbeta (15q22.1) | FBXW7 (14q13.1) | PAF1 (20q13.33) |
| INSIG2 (2q14.1) | BTN2A1 (6p22.1) | TMT5F2 (11q13) |
| HNF4A (20q12-q13.1) | FASN (17q25) | CEP5 (16q21) |

Table 2

| Summary of developmental data. |
|-------------------------------|
| MP-R area | Mean score | Range |
|----------------|------------|-------|
| Cognitive | 54.71 (28.17) | 15–97 |
| Fine Motor | 60.43 (34.01) | 12–105 |
| Receptive Language | 57.81 (34.45) | 10–113 |
| Memory | 52.58 (29.12) | 10–96 |
| Speed | 71.23 (30.49) | 12–114 |
| Visual Motor | 54.32 (29.80) | 10–95 |
| Developmental Index | 50.50 (32.46) | 10–93 |
| Gross Motor | 67.96 (20.36) | 30–96 |
| Overall Expressive Language Score | 50.58 (15.74) | 35–93 |
| Overall Language Score | 56.10 (21.99) | 20–92 |
| Social-Emotional Development | 69.32 (21.29) | 30–115 |
| Self-Help/Adaptive | 73.33 (22.59) | 31–109 |
these 355 variants, 169 were nonsynonymous (31 of which were novel), while one novel variant resulted in a stop codon. The total number of variants observed in the exons of captured regions of an individual subject ranged from 88 to 117; novel variants ranged from four to 10 per subject, while nonsynonymous variants ranged from 36 to 51 per participant.

One novel variant introduced a stop codon (in exon 17 of the FASN gene) in two unrelated participants. No sex-exclusive variation was detected (e.g., no variant was common to all females but not found in males or vice versa). Seven novel variants are found in 6 participants and there were 138 novel variants observed overall (Table 5). 31 variant loci involved nonsynonymous substitutions, while 21 were synonymous.

We did not identify any novel variants associated with differences in neurodevelopmental function (overall MP-R Developmental Index growth score indicator, ranked ordered) in this study. The total number of novel and nonsynonymous variants was not significantly correlated with neurodevelopmental severity using Kendall’s ranked correlation test ($p = 0.27$), while total novel variants was suggestively correlated ($p = 0.068$, $\tau = -0.29$). However, the total number of nonsynonymous variants (annotated and novel) per subject was very significantly correlated with neurodevelopmental severity ($\tau = -0.4, p = 0.007$; Fig. 2). We therefore carried forward all annotated and all novel nonsynonymous variants for further analysis.

We constructed a gene interaction network based on interactions from the literature contained in the GeneMANIA database (visualized in Fig. 3). Using a Poisson regression, we found that the degree of the node in the network was a significant negative predictor of the number of nonsynonymous variants observed in the gene ($\beta = -0.07$, $p = 0.04085$). In addition, we found that overall variant load among the genes with degree equal or less than the median degree of 3 edges for the network was more strongly correlated (Kendall’s rank correlation, $\tau = -0.465$, $p = 0.002$) with neurodevelopmental severity than variants in genes with connectivity equal to the median or greater ($\tau = -0.004$, $p = 0.97$). Analysis using the DAVID functional annotation tool [17] indicates that genes with connectivity higher than the median in this gene interaction network are enriched for functional annotation terms related to transcriptional control and DNA binding, while genes with connectivity equal or lower to the median are enriched in terms related to direct lipid processing, such as lipid transport and sterol biosynthesis.

We then used the WGCNA method to cluster genes with correlated nonsynonymous variant loads to find gene sets with strongly related levels of nonsynonymous variation. Of the

### Table 3

Correlation matrix for full neurodevelopmental battery.

|        | 1         | 2         | 3         | 4         | 5         | 6         | 7         |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1. MP-R Developmental Index | 1.00      | –.221     | .448\*    | .340      | .275      | .459\*    | –.269     |
| 2. ADOS Total | 1.00      | –.213     | .306      | .256      | –.101     | –.303     | .085      |
| 3. VABS-II Communication Total | 1.00      |          |           | .669**    | .752**    | .687**    | –.522**   |
| 4. VABS-II Daily Living Skills Total | 1.00      |          |           |          | .788**    | .798**    | –.273     |
| 5. VABS-II Socialization Total |          |          |           |          |          |          | 1.00      |
| 6. VABS-II Motor Skills Total |          |          |           |          |          |          | –.460**   |
| 7. PDDBI Total |          |          |           |          |          |          | 1.00      |

Correlations that are bolded are statistically significant.

* $p < 0.05$ level.

** $p < 0.01$ level.

### Table 4

Summary of variants observed in study participants.

|              | Synonymous | Nonsynonymous | Stopgain |
|--------------|------------|---------------|----------|
| Annotated    | 164        | 138           | 0        |
| Unannotated  | 21         | 31            | 1        |

### Table 5

Frequency of novel variants.

| Chromosome | Position | Gene            | Frequency |
|------------|----------|-----------------|-----------|
| 1          | 18014130 | ARHGEF10L       | 1         |
| 1          | 18023575 | ARHGEF10L       | 1         |
| 2          | 18023654 | ARHGEF10L       | 1         |
| 4          | 11080638 | EFGR            | 1         |
| 6          | 110901983| EGFR            | 1         |
| 10         | 26388425 | BNP232A         | 2         |
| 12         | 10691759 | TUBB            | 1         |
| 13         | 10019286 | ACAT2           | 10        |
| 16         | 11636729 | CAV1            | 2         |
| 18         | 38003911 | STAR            | 1         |
| 20         | 35005783 | STAR            | 1         |
| 752         | 56711088 | CYFPP7B         | 2         |
| 64          | 126030917| SQLE            | 1         |
| 69          | 107553232| ABCA1           | 1         |
| 27          | 135947017| CEL             | 1         |
| 27          | 135947023| CEL             | 1         |
| 2           | 90796395 | LIPA            | 1         |
| 2           | 61566004 | FADS3           | 1         |
| 37          | 50371360 | LAMA1           | 1         |
| 37          | 110011272| MMAB            | 1         |
| 18          | 101034320| MKV             | 2         |
| 12          | 88722058 | MVD             | 12        |
| 16          | 90001236 | TUBB3           | 1         |
| 17          | 76098024 | RAII            | 1         |
| 17          | 17700321 | RAII            | 1         |
| 17          | 17700557 | RAII            | 1         |
| 17          | 17700788 | RAII            | 2         |
| 17          | 17716129 | SREBF1          | 1         |
| 17          | 17717585 | SREBF1          | 13        |
| 17          | 73942784 | ACXO1           | 1         |
| 17          | 73947570 | ACXO1           | 1         |
| 17          | 800043018| FASN            | 1         |
| 17          | 800043372| FASN            | 2         |
| 17          | 800045849| FASN            | 1         |
| 17          | 800046104| FASN            | 2         |
| 17          | 800049486| FASN            | 1         |
| 17          | 800053197| FASN            | 1         |
| 19          | 11241983 | LDLR            | 1         |
| 19          | 18794823 | C20orf79        | 1         |
| 19          | 18794852 | C20orf79        | 1         |
| 22         | 42276902 | SREBF2          | 21        |
| X           | 48382197 | EBP             | 1         |

We then used the WGCNA method to cluster genes with correlated nonsynonymous variant loads to find gene sets with strongly related levels of nonsynonymous variation. Of the
6 modules produced in this analysis, only the sum of variant load of the largest module of 10 genes was significantly associated with ranked autism severity ($p=0.0009$, Kendall rank correlation test). We then performed a standard linear regression of variant load on ranked severity using each gene in this module as a predictor in a single linear model to determine if variant load at any specific genes appeared to be associated with neurodevelopmental function. Only variant load at NPC1 was a significant predictor in this model ($p=0.009$), though variant load at DHCR24 was suggestively significant ($p=0.0152$). A model using only variant load at DHCR24 and NPC1 showed variant loads at these genes were significant predictors of neurodevelopmental severity ($p=0.0001$ and 0.0009, respectively). The sum of nonsynonymous variant loads in NPC1 and DHCR24 genes is also significantly associated with a more severe neurodevelopmental rank in our data set using Kendall’s ranked correlation test ($\tau=-0.581$, $p=0.0003$). It is also the second strongest two-gene variant load association among all possible pairs of sterol and lipid-related genes with variations in this study, second only to the combination of BTN2A1 and NPC1 ($\tau=-0.61$, $p=0.00014$). Variant loads at BTN2A1, DHCR24 and NPC1 together are significant predictors in a linear model of neurodevelopmental function ($p=0.047, 0.0252$ and 0.0004, respectively). Combined, the cumulative variation load of all three genes had the strongest correlation of all three-gene combinations with ASD severity ($\tau=-0.6584903$, $p<0.00003$). This association remains significant after Benjamini–Hochberg adjustment for testing all trios of genes in the largest module at an adjusted alpha of 0.05 (adjusted $p=0.003$), and remains suggestively significant even after adjustment for testing all possible trios of genes (adjusted $p=0.0767$).

A hierarchical clustering of neurodevelopmental severity (rows) and variant load (columns) among genes in module 1 is illustrated in Fig. 4. This clustering largely recapitulates the results of our regression and correlation analysis: for example, it is easy to see that no subject with a relatively high variant load at both NPC1 and DHCR24 has a ranked neurodevelopmental severity less than 13, placing all but one of those subjects in the lower half of the sample’s severity distribution. The association of variant load at BTN2A1 with variants at NPC1 and DHCR24 and of all three genotypes with severity is also apparent from the clustering.

We then examined the subnetwork of the genes in the largest WGCNA module and their first-neighbors in the GeneMANIA interaction network of lipid and sterol genes (visualized in Fig. 5). Three of the 10 genes in the largest WGCNA module (DHCR24, ABCG5 and SCAP) interact directly with members of the highly-connected complex of FBXW7, SREBF1 and SREBF2 either through genetic interaction (DHRC24, ABCG5) or through direct physical interaction of protein products (SCAP), while NPC1 is connected...
4. Discussion

ASD is a neurodevelopmental disorder which impacts a person’s ability to utilize social communication skills and it causes patterns of restricted or repetitive interests and behaviors. In addition, tissue growth and elaboration of neuronal architecture and connectivity is theorized to occur prematurely in autism which may contribute directly to the expression of ASD behaviors [14]. The overgrowth produces defects in neural patterning, wiring, and cortical interactions between brain regions [14]. Further, young children with ASD show 18% more white matter and 12% more gray matter than their typically developing peers with the temporal lobes having the largest white matter increase [6]. This is compelling because within the central nervous system (CNS), cholesterol has evolved as the essential component of myelin in white matter. As a consequence, the concentration of cholesterol in the CNS is much higher than in any other organ or tissue. During late prenatal and early postnatal development, the rate of myelination is directly correlated with the rate of cholesterol synthesis in the CNS and to the average cholesterol concentration in the brain [18]. In addition, cholesterol has been identified as the signal for synaptogenesis and is critical in glial cells and neuron proliferation [18]. For example, it has been documented that by artificially changing the cholesterol level within glial cells, the frequency of synaptic events are either increased or decreased [32]. Cholesterol is synthesized locally in all regions of the CNS [18]. As humans mature, the need for cholesterol in the CNS decreases, and a pathway for removing cholesterol from the CNS to the plasma is established (i.e., the formation and excretion of 24S-OH-cholesterol (24S) into the blood). Unfortunately, little is known about how CNS cholesterol synthesis affects human brain function. However, work with patients suffering from Alzheimer’s disease has found higher levels of 24S in their plasma when compared to healthy control participants, suggesting a link between cholesterol turnover and brain pathology [29].

Cholesterol synthesis, both in the body as well as the brain, is critical for the production of steroids. All steroid hormones have cholesterol as precursor [39]. In addition, it has been well documented that the brain is capable of producing its own steroids, called neurosteroids [31]. Increasing evidence suggests that these neurosteroids play a vital role in many neuronal processes, including neuroendocrine functions and behavior. Neurosteroids also contribute to neurotransmitter functioning [33]. Neurosteroid deficiency has been documented in ASD [30,45]. Together, these data have led to the hypothesis that ASD may occur as the result of impaired brain sterol homeostasis secondarily causing decreased neurosteroid synthesis, aberrant synaptic plasticity and abnormal sonic hedgehog patterning [26].

Understanding the relationship between neural information processing and genetic components (e.g., lipid/sterol-regulating genes) in ASD may be useful to better inform targets for therapy as highlighted in a 2004 report by Belmonte et al.[5]. To that end, the main finding from our study suggests that the number of nonsynonymous variants present in a young child with ASD is correlated with his or her overall neurodevelopmental severity. This possibility is compelling for numerous reasons. Firstly, nonsynonymous variants in lipid and sterol comprise a group of single nucleotide polymorphisms (SNPs) that, together with SNPs in regulatory regions, are believed to have the highest impact on phenotype [36]. We also find a significant overlap of the sterol genes we assayed with a set of genes with high mutation rates in ASD-affected siblings and their parents that was recently reported by Yuen et al. Among the seven overlapping genes (CYP7B1, DHCR24, EBP, HSD17B4, INS, MVK, and NSHDL), we found that variant load in DHCR24 in particular was found to be a strong predictor of ASD severity in our study. Interestingly, mutations in

Finally, we examined the overlap of our sterol and lipid gene set with the PhHS_MindFun_All gene set reported as containing a high level of variation in ASD siblings and their parents by Yuen et al. Seven of the sterol genes investigated here (CYP7B1, DHCR24, EBP, HSD17B4, INS, MVK, and NSHDL) are included in PhHS_MindFun_All, a statistically significant overlap when compared to the HUGO-annotated human genome as a whole ($p=.0117$, hypergeometric test).

Fig. 4. Hierarchical clustering and heatmap of nonsynonymous variant load and ASD severity for genes in the largest WGCNA module. Heatmap values indicate variant load, with bright green indicating the lowest observed variant loads and bright red the highest. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Sub-network of genes in the largest WGCNA module and first neighbors in the GeneMANIA interaction network of sterol and lipid genes. Genes in the largest WGCNA module are indicated as diamonds.
DHCR24 are causative in Smith–Lemli–Opitz Syndrome; children affected by this syndrome are often initially diagnosed as affected by ASD [7]. In addition, Yuen et al. found that siblings with similar patterns of variance in ASD-related genes tended to display similar types and severity of ASD symptoms as compared to siblings with different variants. Though those authors did not directly correlate clinical measures of ASD severity with mutation load or specific mutation patterns in their analysis, their findings provide additional support for the idea that an increased number of variants in autism-related genes may predispose persons to develop ASD, and that similar patterns of mutation may lead to similar symptoms and severity of the disorder. Overall, these studies suggest that the current findings support what is known about the functional impact of nonsynonymous variants. Interestingly, to our knowledge, this is the first time that a relationship has been suggested between genetic sequence variants in lipid and sterol genes and ASD.

We also observe a significant relationship between high numbers of nonsynonymous variants and low connectivity in our GeneMANIA network analysis. While a literature bias in the GeneMANIA or DAVID databases cannot be ruled out, these poorly-connected genes are significantly enriched for lipid transport and synthesis annotations, suggesting that a substantial portion of variation in neurodevelopmental severity in ASD may be affected by nonsynonymous variants in genes directly involved in sterol transport and synthesis, rather than in the transcription factors controlling the process. Some of this effect may be due to a higher evolutionary permissibility for functional variation in the lipid processing and transport genes assessed here as compared to transcription factors; interestingly, only one novel and nonsynonymous mutation is observed in the SREBF1 transcription factor in our cohort.

Additional support for the potential impact of cumulative nonsynonymous variation in sterol processing and signaling genes in ASD comes from the result of our analysis of correlated nonsynonymous variant loads using WGCNA. The sum of nonsynonymous variant loads in DHCR24, NPC1 and BTN2A1 strongly predict neurodevelopmental severity of ASD in our cohort; two of these genes are involved directly in lipid metabolism through either transport (NPC1) or oxioreductase activity (DHCR24), while all three genes are involved in signaling activity and BTN2A1 has been associated with ASD in copy-number variation studies [27].

NPC1 and DHCR24 are also closely connected to the central transcriptional control cluster of SREBF1, SREBF2 and FXBW7 in our network analysis, along with the cholesterol transporter ACG5 and the SREBF chaperone protein SCAP (Brown and Goldstein 1998) that are grouped with NPC1 and DHCR24 in our WGCNA analysis. SREBF factors are well-characterized as important to the transcriptional control of sterol biosynthesis, while FXBW7 has been implicated in regulating the activity of SREBF1/2 by targeting them for degradation via ubiquitination. This suggests that one unifying hypothesis to explain the observed association between variant load in NPC1, DHCR24, SCAP and ACG5 and neurodevelopmental severity in ASD is through direct or downstream effects of their gene products on sterol biosynthesis pathways mediated by SREBF1/2 and FXBW7 due to aberrations in lipid metabolism, such as the alterations of sterol transport observed in patients with ACG5 mutations involved in sitosterolemia [20], or the abnormal sterol accumulations present in persons with NPC1 mutations that lead to Niemann–Pick Type C disease. Some children with Niemann–Pick Type C disease show sensory integration deficits similar to those observed in ASD children [4]: researchers have also recently reported administration of cyclodestrin to cats with the feline version of this disorder substantially ameliorates its symptoms and progression [47], raising intriguing possibilities if NPC1-mediated lipid processing deficits influence ASD.

Abnormalities in the signaling behaviors of mutated versions of NPC1, DHCR24 and BTN2A1 might also influence the neurodevelopmental severity of ASD, as could changes in the functionality of SCAP as a SREBF chaperone.

There are some notable limitations to this study. The human reference sequence is a dynamic resource and updates to the database often change the position of a variant relative to a gene. Also, the dbsNP and 1000 Genomes data are dynamic and are continually updated. Therefore, variants which are novel at this time may be found in other samples as more data is incorporated into these databases. In terms of sample size and power, it is unknown whether these variants could be found in a similar size control cohort by chance.

In terms of future directions, it would likely be informative to genotype all novel variants rather than limiting to lipid and sterol genes in a cohort of ASD participants as well as controls to confirm their association to ASD. Comparing these results with other whole exome data may also be beneficial.

Conflicts of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property. We further confirm that any aspect of the work covered in this manuscript that has involved either experimental animals or human patients has been conducted with the ethical approval of the appropriate ethical committees and has been conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The study was approved and monitored by the Oregon Health & Science University Institutional Review Board (IRB; approval number 4031).

Ethical approval

The study submitted for publication conforms to the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, issued by the International Committee for Medical Journal Editors (ICMJE), and to the Committee on Publication Ethics (COPE) code of conduct for editors. The study was conducted to a high ethical standard and it adheres to local regulations and standards for gaining scrutiny and approval.

The work described in the article was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The study was approved and monitored by the Oregon Health & Science University Institutional Review Board (IRB; approval number 4031).
Funding sources

Research reported in this publication was supported by the Northwest Health Foundation under the Mark O. Hatfield Research Fellowship on Autism. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Northwest Health Foundation.

Research reported in this publication was also supported by National Center for Advancing Translational Sciences of the National Institutes of Health under award number ULTR000128. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments

The authors thank Clive Woffendin, Bob Searles, Shannon McWeeney, Cheryl Maslen, Karen Grant, and Celeste Flachsbart for their valuable contributions to the project. The authors also thank the OHSU Autism Clinical Program for its support of this research, as well as the children and families who participated.

References

[1] B. Abrahams, Genetics: Many roads to the autism spectrum disorders, in: D. Fein (Ed.), The Neuropsychology of Autism, Oxford University Press, Oxford New York, 2011, pp. 35–47.
[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th edition, APA, Washington DC, 2000 (Text Revision, DSM-IV-TR).
[3] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th ed., American Psychiatric Association, Washington, DC, 2013.
[4] G.N. Andrade, S. Molholm, J.S. Butler, A.B. Brandwein, S.U. Walkley, J.J. Foxe, Arterial multisensory integration in Niemann-Pick type C disease-towards potential biomarkers, Orphanet J. Rare Dis. 9 (149) (2014).
[5] M. Belmonte, E. Cook, J. Rubenstein, W. Greenough, A. Beckel-Mitchener, E. Courchesne, L. Boulanger, S. Powell, P. Levitt, E. Perry, Y. Jiang, T. DeLorey, E. Tierney, Autism as a disorder of neural information processing: directions for research and targets for therapy, Mol. Psychiatry 9 (7) (2004) 646–663.
[6] F. Dinstein, A. Hardan, S. Ucelli di Nemi, J. Perez, J. Soares, F. Barale, Brain anatomy and development in autism: review of structural MRI studies, Brain Res. Bull. 61 (2003) 557–569.
[7] I. Bukels, F.D. Porter, A.W. Zimmerman, E. Tierney, Smith–Lemli–Opitz syndrome and autism spectrum disorder, Am. J. Psychiatry 164 (11) (2007) 1661.
[8] I. Bukels, F.D. Porter, A.W. Zimmerman, E. Tierney, Smith–Lemli–Opitz syndrome and autism spectrum disorder, Am. J. Psychiatry 164 (11) (2007) 1661.
[9] J. Blancato, Hypercholesterolemia in children with Smith-Magenis syndrome: place of birth on the risk of autism: a nationwide register-based study, J. Child Neurol. Psychiatry 16 (5) (2004) 938–943.
[10] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and applications, Nucleic Acids Res. 30 (17) (2002) 3894–3900.
[11] M. Jorde, S. Hasstedt, E. Ritvo, A. Mason-Brothers, B. Freeman, C. Pingree, No evidence for ethnic and racial bias in the Tryout of the Merrill–Palmer Scale-Revised, Psychol. Rep. 94 (1) (2004) 217–220.
[12] J.A. Hubacek, K.E. Berge, J.C. Cohen, H.H. Hobbs, Mutations in ATP-cassette binding proteins G5 (ABCC5) and G8 (ABCC8) causing sitosterolemia, Hum. Mutat. 18 (4) (2001) 359–360.
[13] G. Dennis Jr, B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, R. Lempicki, DAVID: database for annotation, visualization, and integrated discovery, Genome Biol. 4 (5) (2003) 11–16.
[14] J.M. Dietchy, S.D. Turley, Cholesterol metabolism in the central nervous system during early development and in the mature mammal, J. Lipid Res. 45 (8) (2004) 1375–1397.
[15] R. Floyd, K. Gathercoal, G. Roid, No evidence for ethnic and racial bias in the Tryout Edition of the Merrill–Palmer Scale-Revised, Psychol. Rep. 94 (1) (2004) 217–220.
[16] J. Blancato, Hypercholesterolemia in children with Smith-Magenis syndrome: place of birth on the risk of autism: a nationwide register-based study, J. Child Neurol. Psychiatry 16 (5) (2004) 938–943.
[17] R. Lee, E. Tierney, Hypothesis: The role of sterols in autism spectrum disorder, Autism Res. Treat. 2011 (2011).
[18] R. Santos, S. Smith, L. Leib, D. Smith, L. Lemli, J.E. Opitz, A newly recognized syndrome of multiple congenital anomalies, J. Pediatr. 139 (1) (2001) 1.
cyclodextrin prevents cerebellar dysfunction and Purkinje cell death in feline Niemann-Pick type C1 disease, Sci. Transl. Med. 276 (2015).

[48] R. Yuen, B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Malopathamkalam, G. Pellecchia, Y. Liu, M. Gazzellone, L. D’Abate, E. Deneault, J. Howe, R. Liu, A. Thompson, M. Zarrei, M. Uddin, C. Marshall, R. Ring, L. Zwagenbaum, P. Ray, R. Wexberg, M. Carter, B. Fernandez, Whole-genome sequencing of quartet families with autism spectrum disorder, Nat. Med. 21 (2015) 185–191.

[49] K. Wang, H. Zhang, D. Ma, M. Bucan, J. Glessner, B. Abrahams, D. Salyakina, M. Imielinski, J. Bradford, P. Sleiman, C. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Takahashi, T. Sakurai, E. Rappaport, C. Lajonchere, J. Munson, A. Estes, O. Kovatska, J. Piven, L. Sonnenblick, A. Retuerto, E. Herman, H. Dong, T. Hutmen, M. Signum, S. Ozono, A. Klin, T. Owley, J. Sweeney, C. Brune, R. Cantor, R. Bernier, J. Gilbert, M. Cuccaro, W. McMahon, J. Miller, M. State, T. Wassink, H. Coon, S. Levy, R. Schultz, J. Nurnberger, J. Haines, J. Sutcliffe, E. Cook, N. Minshew, J. Busbaum, G. Dawson, S. Grant, D. Geschwind, M. Pericak-Vance, G. Schellenberg, H. Hakonarson, Common genetic variants on 5p14.1 associate with autism spectrum disorders, Nature 459 (2009) 528–533.

[50] D. Warde-Farley, S. Donaldson, O. Comes, K. Zuberi, R. Badawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. Bader, O. Morris, The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function, Nucleic Acids Res. 38 (2010) 214–220.

[51] L. Weiss, D. Arking, The Gene Discovery Project of Johns Hopkins & the Autism Consortium, A genome-wide linkage and association scan reveals novel loci for autism, Nature 461 (2009) 802–808.

[52] L. Weiss, Y. Shen, J. Korn, D. Arking, D. Miller, R. Fosdal, E. Saeumunds, H. Stefansson, M. Ferreira, T. Green, O. Platt, D. Ruderfer, C. Walsh, D. Altschuler, A. Chakraverti, R. Tanzi, K. Stefansson, S. Santangelo, J. Gusella, P. Sklar, B. Wu, M. Daly, Association between Microdeletion and Microduplication at 16p11.2 and Autism, N. Engl. J. Med. 358 (2008) 667–675.

[53] W. Zhao, P. Langfelder, T. Fuller, J. Dong, A. Li, S. Hoverth, Weighted gene coexpression network analysis: State of the art, J. Biopharm. Stat. 20 (2010) 281–300.