

OPEN ACCESS

Citation: Moreno-Ramos OA, Olivares AM, Haider NB, de Autismo LC, Lattig MC (2015) Whole-Exome Sequencing in a South American Cohort Links ALDH1A3, FOXN1 and Retinoic Acid Regulation Pathways to Autism Spectrum Disorders. PLoS ONE 10(9): e0135927. doi:10.1371/journal.pone.0135927

Editor: Michael Schubert, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, FRANCE

Received: February 20, 2015

Accepted: July 28, 2015

Published: September 9, 2015

Copyright: © 2015 Moreno-Ramos et al. This is an open access article distributed under the terms of the <u>Creative Commons Attribution License</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Support for sequencing was provided by proyecto semilla 2013-1, Facultad de ciencias, Universidad de los Andes (OAM) and Vicerrectoria de investigaciones, Universidad de los Andes (MCL). Support for molecular and animal studies was provided by the Chase Fund (NBH) and Edward N. & Della L. Thome Memorial Foundation, Bank Of America, N.A., Trustee (NBH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

RESEARCH ARTICLE

Whole-Exome Sequencing in a South American Cohort Links ALDH1A3, FOXN1 and Retinoic Acid Regulation Pathways to Autism Spectrum Disorders

Oscar A. Moreno-Ramos¹, Ana María Olivares², Neena B. Haider², Liga Colombiana de Autismo³¹, María Claudia Lattig¹*

1 Departamento de Ciencias Biológicas, Facultad de Ciencias, Universidad de los Andes, Bogotá D.C., Colombia, 2 Schepens Eye Research Institute, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, MA, United States of America, 3 Liga Colombiana de Autismo (LICA), Bogotá D.C., Colombia

¹ Membership of the Liga Colombiana de Autismo is listed in the Acknowledgments. * mlattig@uniandes.edu.co

Abstract

Autism spectrum disorders (ASDs) are a range of complex neurodevelopmental conditions principally characterized by dysfunctions linked to mental development. Previous studies have shown that there are more than 1000 genes likely involved in ASD, expressed mainly in brain and highly interconnected among them. We applied whole exome sequencing in Colombian-South American trios. Two missense novel SNVs were found in the same child: ALDH1A3 (RefSeq NM_000693: c.1514T>C (p.I505T)) and FOXN1 (RefSeq NM_003593: c.146C>T (p.S49L)). Gene expression studies reveal that Aldh1a3 and Foxn1 are expressed in ~E13.5 mouse embryonic brain, as well as in adult piriform cortex (PC; ~P30). Conserved Retinoic Acid Response Elements (RAREs) upstream of human ALDH1A3 and FOXN1 and in mouse Aldh1a3 and Foxn1 genes were revealed using bioinformatic approximation. Chromatin immunoprecipitation (ChIP) assay using Retinoid Acid Receptor B (Rarb) as the immunoprecipitation target suggests RA regulation of Aldh1a3 and Foxn1 in mice. Our results frame a possible link of RA regulation in brain to ASD etiology, and a feasible non-additive effect of two apparently unrelated variants in ALDH1A3 and FOXN1 recognizing that every result given by next generation sequencing should be cautiously analyzed, as it might be an incidental finding.

Introduction

Autism spectrum disorders (ASD) are a range of complex neurodevelopmental conditions characterized by dysfunctions in mental development [1]. At the same time, it has been hypothesized that ASD are a group of continuous disorders instead of a single discrete disorder [2], explaining why ASD patient symptomatology is so variable. Although prevalence of ASD has increased worldwide [3], in Colombia there are no official reports regarding its prevalence.



Competing Interests: The authors declare funding from a commercial source: Bank of America. This does not alter their adherence to all the PLOS ONE policies on sharing data and materials.

A recent study has shown that ASD heritability is calculated closely to 50% [4] (lower than the ~90% previously estimated [2,5]) leading to higher environmental factor's impact on the occurrence of the disorder. In fact, no clear environmental factors that have been unfolded since the recurrence of ASD does not change statistically between dizygotic twins and complete siblings [4].

Most of our understanding of the molecular basis of ASD comes from reports of rare copy number variants (CNVs) [$\underline{6}-\underline{8}$] and the discovery of *de novo* single nucleotide variants (SNVs) using Whole Exome Sequencing (WES) [$\underline{9}-\underline{13}$] or Whole Genome Sequencing (WGS) [$\underline{14}$]. More than a 1000 genes are thought to be involved in ASD [$\underline{15}$], most of them located in genes of known function and involved in pathways related to cell and neuronal development, projection, motility and proliferation; also in genes involved in spine and dendrite plasticity regulation, and gene regulation [$\underline{8}, \underline{15}, \underline{16}$]. Additionally, variants affecting recurrent genes have been associated to ASD by their single impact ignoring a possible synergic impact between two or more de novo events in the same individual. Nonetheless, all this information has been derived mostly from Caucasian individuals. Thus, novel genes and even novel pathways involved in ASD pathogenesis might be found in understudied populations such as Colombian.

WES was applied in search of *de novo* variants that might be causative of ASD in four family trios from Colombia. Two *de novo* non-synonymous mutations affecting *ALDH1A3* (RefSeq NM_000693, MIM:600463) and *FOXN1* (RefSeq NM_003593.2, MIM:600838) genes were uncovered in the same child (FAM07). Recently, retinoic acid (RA) pathways mediated by Retinoid Orphan Receptor Alpha (RORA) have been implicated in ASD [17]. Furthermore, homozygous *ALDH1A3* missense and nonsense mutations in humans have been linked to anophthalmia and microphthalmia (MIM: 615113; A/M) with some affected individuals also exhibiting autistic traits [18–23]. Though no clear evidence supports a link between variants in ALDH1A3 with autism, mouse studies show that lack of Aldh1a3 (Gene NC_000073.6) results in abnormal GABAergic neuronal differentiation in the forebrain basal ganglia [24]. Malfunctions in this inhibitory system have been found to be associated to mental disorders like ASD, schizophrenia and bipolar disorder [25,26]. Moreover, failures of this enzyme in the striatum are related to elimination of dopamine receptor D2 in the nucleus accumbens [27].

Although little is known about *FOXN1*, it is essential for proper immune system function, especially in thymus development and maintenance during adulthood [28–30]. Foxn1 (Gene NC_000077.6) knockout mice (commonly known as NUDE mouse) overexpress pro-inflammatory cytokines (Specially T_{H1}) [31]. This overexpression has also been reported in ASD children [32]. Besides T-cell immunodeficiency, mutations in FOXN1 have also been observed together with congenital alopecia and nail dystrophy (MIM: 601705). A homozygote mutation in FOXN1 was reported in a 15-week-old fetus with anencephaly and severe neural tube defect (MIM: 601705) [33].

As ALDH1A3 is an enzyme responsible for Retinoic Acid (RA) synthesis, and FOXN1 has been found to regulated by RORA (gene involved in RA cascade) [17], together with RA's major role in gene expression regulation during brain development [34], it was questioned if RA might possibly regulate ALDH1A3 and FOXN1 through RA receptors (RARs). Predicted RA Response Elements (RAREs) in the promoter regions of both genes where evaluated to be recognized by RARs, using C57BL/6J (B6) mice, in two developmental stages. Rarb (Gene NC_000080.6) was selected as the chromatin association target protein since it is known to regulate dorsal and ventral telencephalon development, hippocampal plasticity and brain barrier development [35–38].

Results indicate that Rarb regulates Aldh1a3 and Foxn1 (NC_000077.6) in adult (~P30) mice PC and in embryo (~E13.5) whole brain, as demonstrated by ChIP essay. We suspect that both variants in genes ALDH1A3 and FOXN1 interact in a non-additive manner, where an

epistatic interaction effect can be expected, possibly explaining the etiology in the affected child.

Methods

Ethics statement

The study presented here involves human participation and informed consent and assent. The study and consent procedure was approved by the ethical committee of Universidad de los Andes (Acta 11, 2011). Informed consent and assent was given and explained to all participants. All participants provided a written informed consent to participate in this study. Each consent was marked with a family and individual unique identifier, names and identifying information are kept securely stored. Only the principal investigator is able to recognize personal data with unique identifiers.

Animal procedures were approved by Schepens Eye Research Institute Animal Care and Use Committee (Permit number: S-309–0714) in compliance with the Animal Welfare Act Regulations.

Cohort selection

Four family trios (family ID: Fam02, Fam07, Fam09 and Fam10) referred by Liga Colombiana de Autismo (LICA) were selected to participate in the study. All individuals are from Bogotá D.C.—Colombia. Affected individuals were engaged in a series of tests-tasks to evaluate patterns and maturity of the child, interaction with parents and family patterns. Inclusion in the present study required meeting criteria for autism on the ADOS [39] and ADI-R [40] and subject were confirmed to meet DSM-V (APA, 2012) criteria for autism by trained professionals who observed and interacted with them over several visits. After psychological examination, a genetic counselor physician discarded any comorbidity with other ASD related syndrome.

DNA sample preparation, exome sequencing and analysis

DNA extraction was carried out using FlexiGene DNA kit (Qiagen, Gaithersburg, MD, USA). The extraction was carried out parting from 300uL of leucocytes following manufacturer instructions. A mean of 8.7ug of DNA per individual was sent to Otogenetics Corporation in Atlanta GA, USA. We applied WES using Agilent Sureselect All Exon V4 platform and PE100 Illumina HiSeq2000 platform to the four family trios selected (all family trios DNA samples met the quality criteria for WES by Otogenetics Corporation, Atlanta, GA-USA). Sequence reads were cleaned using the latest version of FASTX-Toolkit (V 0.0.13.2) (http://hannonlab. cshl.edu/fastx_toolkit/index.html). Sequence mapping and variant calling was carried first by mapping the short sequences using Burrows-Wheeler aligner (BWA) (from http://bio-bwa. sourceforge.net) [41] on the latest version of the human genome available at UCSC Genome Browser (HG19). SAM tools (from http://samtools.sourceforge.net) [42] and PicardTools (from http://picard.sourceforge.net) were used to manipulate and mark sequence duplicates. SNPs and small indel calling was achieved using HaplotypeCaller walker of Genome Analysis Toolkit (GATK, from http://www.broadinstitute.org/gatk/) [43,44]. Best practices of GATK were followed in order to guarantee an efficient variant calling. Annotation was performed using snpEff (Human genome reference 37.5) [45]. Putative de novo variants were identified using AWK language.

Variant prioritization and validation

After identifying all the non-inherited variants, minor allele frequency (MAF) was revised from dbSNP. Those variants that presented a MAF > 5% were discarded. Different sets of primers were manually designed to flank the variants position obtained during calling. PCR amplification protocol described in Perea, C.S. et al., 2012 was used [46]. Sanger sequencing at Macrogen, Seul—Korea; and at Universidad de los Andes' Sanger sequencing facility (3500 Genetic Analyzer; Applied Biosystems, USA) was used to corroborate the presence of each of the variants unveiled.

After validation of each of the variants was performed, depending on the predicted impact, different bioinformatic tools were used to predict their impact. The possible effect of the synonymous variant on mRNA stability was analyzed using SilVA [47], and ESEfinder for possible *de novo* splicing regions [48,49]. Non-synonymous variants were analyzed using SIFT [50], PolyPhen [51] and PROVEAN [52].

Retinoic acid regulation pattern discovery

Searching for the relaxed Retinoic Acid Response Element [RARE; 5'-(A/G)G(G/T)(G/T)(G/C)A-3' motif separated by one, two or five nucleotides, which interacts with heterodimers of RARs and retinoid X receptors (RXRs)] [34,53,54], was achieved using the online interface of Patser (online site: http://stormo.wustl.edu/consensus/html/Html/main.html) [55]. Genomic region spanning 30000bp upstream to 1000bp downstream from the initial transcription site for humans and mice (Human genome build NCBI 37.5 and Mouse genome build GRCm38. p2) in both negative and positive strands was analyzed.

Animals

Pregnant females and adults were obtained under standard conditions from Jackson Laboratory and housed in vivariums at Schepens Eye Research Institute. The study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institute of Health. Tissues were collected from C57BL/6J (B6) postnatal (~P30) and from B6 embryonic (~E13.5).

Aldh1a3 and Foxn1 expression study

Transcription was observed after total RNA extraction from piriform cortex (PC) from both lobules of adult C57BL/6J (B6) brain mice (~P30), and B6 embryonic whole brain (~E13.5) using Trizol. PC was chosen since Aldh1a3 expression has been reported in this region in postnatal stages [56]. A total of 2µg of total RNA was reversed using random decamers, in independent biological triplicates using Retroscript (Ambion). PCR, using primers flanking the conserved splice regions of both genes (S2 Table), was used to check Foxn1 and Aldh1a3 transcription. PCR products were run on agarose (2%) ethidium bromide stained gel.

Chromatin immunoprecipitation

Chromatin immunoprecipitacion assay (ChIP) against RARs-RARE interaction was used to validate the bioinformatic predictions. ChIP, followed by PCR against the interaction of Rarb with the predicted RAREs, was performed as described in Haider, N.B. et al., 2009 with modifications [57]. Briefly, three adult (~P30) mice PC and three embryo (~E13.5) mouse whole brain were collected. Each sample was sonicated using three cycles. First cycle: 1s pulse*1s wait between pulses; total of 20 pulses. Second cycle: 19 steps of ten pulses, 1s pulse*1s wait between pulses, 30 seconds wait between steps; total of 190 pulses. Third cycle: 10s pulse*20s wait



	Cono nomo	Chr		Def Allele		A A Change	Conner Validation
Family ID	Gene name	Chr.	Chr. Pos.	Rel. Allele	Altern. Allele	A.A. Change	Sanger validation
Fam07	ALDH1A3	15	101454953	Т	С	I505T	+
	FOXN1	17	26851543	С	Т	S49L	+
Fam10	NR4A2	2	157186342	С	А	S119	+
	GAS8-AS1	16	90095617	G	GCTGCGGGGCAGC		+

Table 1. Non-inherited variants found in families Fam07 y Fam10.

doi:10.1371/journal.pone.0135927.t001

between pulses; total of three pulses. Average DNA fragment size obtained was ~300bp. Aliquots for Input (Positive control), Rarb antibody (SC-552, Santa Cruz Biotechnology, Inc.) immunoprecipitation and rabbit IgG (2027, Santa Cruz Biotechnology, Inc.) (Negative control) were taken and incubated overnight. PCR was performed using primers flanking each RARE sequence found in mice (<u>S3 Table</u>), to amplify approximately 200 base pair amplicon. PCR conditions used are previously described in Haider, N.B. et al., 2001 where an annealing temperature of 58°C and 35 cycles were applied, and a dilution of 1:10 for the Input sample was used [<u>58</u>]. PCR products were visualized by ethidium bromide staining.

Results

Exome sequencing, technical outcomes, variant calling and *de novo* variant discovery

After applying WES, a ~69.1X coverage was yielded. In average, ~228,116.1 variants per family trio were obtained. A total of three SNVs and one Indel were validated by PCR amplification followed by Sanger sequencing (Table 1). We observed a transvertion-transition ratio of 1:2 and non-coding SNV or Indel were found in two of the families studied. None of the genes found in this study have been reported in the 982 WES studies [9–13], or in any of the ten WGS family trios [14] performed to date. All other NGS data is stored at the Universidad de los Andes bioinformatic facility and it is freely available upon request.

Variant prioritization and analysis

One synonymous *de novo* SNV, located in *NR4A2* (<u>Table 1</u>, RefSeq NM_006186.3, MIM: 601828; c.779G>T [p.S119]) and one *de novo* insertion in the ORF GAS8-AS1 (<u>Table 1</u>, RefSeq NC_000016.10, MIM:605179; g.90095617_90095618insCTGCGGGGCAGC) were found in the proband of Fam10. The possible effect of the synonymous variant on mRNA stability was analyzed using SilVA [<u>47</u>], as in O'Roak, B.J. et al., 2012 another synonymous variant was reported in this gene [<u>10</u>], and ESEfinder for possible *de novo* splicing regions [<u>48,49</u>], but no effects on mRNA stability or possible new splice sites were found (<u>Table 2</u>). The insertion in

Family	Variant class	Gene	Software					
			SilVA	ESEfinder	PROVEAN (Cutoff<- 2.5)	SIFT (Cutoff<0.05)	PolyPhen (Cutoff<0.878)	
FAM07	Non- Synonymous	ALDH1A3	NA	NA	Deleterious (-3.534)	Tolerated (0.142)	Possibly Damaging (0.878)	
		FOXN1	NA	NA	Neutral (-0.241)	Harmful (0.006)	Benign (0.137)	
FAM10	Synonymous	NR4A2	Likely benign	No splice site found	NA	NA	NA	
	Insertion	GAS8-AS1	NA	No splice site found	NA	NA	NA	

doi:10.1371/journal.pone.0135927.t002





Fig 1. Pedigrees of family FAM07 showing chromatograms where the de novo Novel mutation event occurred for genes ALDH1A3 (c.1618T>C [p. Ile505Thr]) and FOXN1 (c.175C>T [p.Ser49Leu]).

doi:10.1371/journal.pone.0135927.g001

GAS8-AS1 (<u>Table 1</u>) was analyzed using ESEfinder to determine if it might alter *GAS8* gene (MIM:605178) splice since it is located in intron 2. ESEfinder did not show any new probable splicing site for the Indel in GAS8-AS1 or even *GAS8* (<u>Table 2</u>).

Two non-synonymous *de novo* variants were uncovered in the affected child of family Fam07 within *ADLH1A3* (RefSeq NM_000693: c.1514T>C (p.I505T)) and *FOXN1* (RefSeq NM_003593: c.146C>T (p.S49L)) (Fig_1). SIFT [50] predictions indicate that the *FOXN1* alteration is harmful; while PolyPhen [51] and PROVEAN [52] predict that the SNV located in *ADLH1A3* is deleterious (Table 2). The other two families: Fam02 and Fam09 did not reveal any de novo mutations on the affected probands. The study further focused on studying both *ALDH1A3* and *FOXN1* since the variants found have a potential negative impact.

RAREs prediction

After looking for the RAREs patter using Patser in the positive and negative strands of the promoter region of genes ALDH1A3 and FOXN1 in humans, as in Aldh1a3 and Foxn1 in mice [55], four possible RAREs for *ALDH1A3* and 18 for *FOXN1* in humans, and ten possible RAREs for Aldh1a3 and 11 for Foxn1 in mice were found (<u>S1 Table</u>).

Aldh1a3 and Foxn1 expression in mice brain tissues and regulation by Rarb

According to the expression analysis performed on adult (~P30) PC and in embryo (~E13.5) whole brain, electrophoretic run indicates that both genes are transcribed in ~P30 mouse PC and in ~E13.5 mouse whole brain Fig 2A.





doi:10.1371/journal.pone.0135927.g002

LOS ONE

Chromatin Immunoprecipitation results demonstrate that Rarb binds to one predicted RARE upstream of Aldh1a3 and to two predicted RAREs upstream of Foxn1 in the adult PC (Fig 2B, Table 3) in mice. In the embryo's stage, we observed that Rarb binds to one predicted RARE for Aldh1a3 and two for Foxn1 (Fig 2B, Table 3) in mice. These results indicate that both RAREs are probably recognized independently of the developmental stage.

Discussion

Most ASD studies highlight a strong genetic heterogeneity where both *de novo* germline and rare inherited variants are distributed across numerous genes yet, interconnected in similar

Table 3. RARE sequences and their relative positions to the initial transcription site, found by ChIP for genes Foxn1 and Aldh1a3 in mice in adult (~P30) piriform cortex and embryo (~E13.5) whole brain.

Gene	RARE Sequence	Relative Position	Strand
Aldh1a3	GGGGGAGTGGGGGA	-16982	-
Found	AGGTGACAATGGGGTGA	-7422	+
FOXIT	GGTTCATCAGTTCA	-17785	+

doi:10.1371/journal.pone.0135927.t003

biological processes such as chromatin remodeling and transcription regulation [8,10-13,16] or target of fragile X mental retardation protein (FMRP) [59]. Furthermore, previous WES studies associate genes to ASD only if two or more disrupting variants are observed in the same gene in different individuals. Variant and gene association depends uniquely on the effect the affected gene has [6,10-13,15,59], though it is also important to examine how two or more de novo variants in the same individual might play a synergic role on ASD phenotype. Moreover, every result given by WES or WGS trio studies should be carefully analyzed, as it might be an incidental finding unless function studies are performed.

The findings of RAREs sequences in the promoter regions, in humans as in mice, made us consider that RARs family might regulate FOXN1 and ALDH1A3 genes. Thus, we aimed to determine if any of those RAREs predicted might be functional using mice as models. We first confirm expression of Aldh1a3 in PC during in adult (~P30) mouse brain as previously shown [56] and evaluated co-expression of Foxn1 in the same region and stage. Moreover, we demonstrate that both genes are also expressed in embryo (~E13.5) mice. ChIP-PCR results suggest that Foxn1 and Aldh1a3 expression is promoted by Rarb binding to different RARES in both tissues (adult PC and embryo whole brain; Fig 2B, Table 3). Since ChIP was only carried out to analyze Rarb-RARE interaction, we could not assure that any other members of the Rar family or other transcription factors might contribute to Foxn1 or Aldh1a3 regulation during the stages analyzed. Therefore, it is possible that gene RA regulation is exerted by other proteins in both tissues at both stages. Based on known functions of RARs/RXRs, these results predict a possible positive regulation by Rarb for both genes. The RA regulation model predicts that in RA presence, RARs/RXRs release co-repressor proteins and recruit co-activation complexes [34]. In fact, it has been demonstrated that RA promotes Rars binding to putative RAREs in mouse embryonic stem cells working as transcription enhancers [60]. As Rarb and the other RAR family members regulate many genes at different stages and tissues at time-specific periods, lack of RA might contribute to a transcriptional deregulation.

Recently, non functional *ALDH1A3* alleles have been reported as responsible for ophthalmologic diseases as A/M in which four of the six reports mention that some affected individuals also have ASD like features or intellectual disability in different degrees, even though these findings are generally considered as incidental [18,19,21,23]. However, it is not possible to discard the idea that mutations in *ALDH1A3* might alter brain development since it has been described that RA is essential for brain cortex development facilitating transcription of essential genetic markers [61]. Additionally, FAM07 child carries a non-synonymous variant, predicted to be pathogenic in *ALDH1A3* (Table 2). One would thus predict that RA synthesis and concentration in basal ganglia, nucleus accumbens, striatum, and PC would be lower than normal levels.

Since RA mediates transcription of different genes, variants in ALDH1A3 in conjunction with other inherited rare variants or *de novo* variants potentially act synergistically causing ASD. This might also be the case in the A/M phenotype where some of the affected individuals also have ASD or intellectual disability [19]. Yet, if another pathogenic variant is present on a gene that is regulated by retinoic acid, the altered concentrations levels of RA together with a deficient function of a second gene might end up in a phenotype such as ASD.

As no clear data of FOXN1 regulation of brain development is clear, we cannot predict a direct effect of the non-synonymous variant observed in this child. Since this nuclear receptor is apparently regulated by RA via Rarb, the lack of this metabolite due to the mutation in *ALDH1A3* might cause a lower transcription of *FOXN1*, leading to a dysregulation of genes and pathways not yet described. Moreover, nude affected humans by mutations or truncations in the FOXN1 gene do not present gross central nervous system alterations [62], but they however present changes in the corpus callosum, neural tube and choroid plexus [33,62]. This

suggest that *FOXN1* has similar functions as other member of the FOX family like *FOXP1* and *FOXP2* [63,64], which have been already associated to ASD [6].

In conclusion, the results indicate that next generation sequencing from under-studied cohorts provide new genetic data that might conduct to new perspectives and hypothesis to explain genetic basis of ASD. Our results are in concordance with previous studies demonstrating that transcription factors play significant roles in ASD [8,10-13,16]. A previous study of our group reported that genes involved in transcription regulation are essential to serotoninergic pathway stability in depression [65]. Moreover, even though ASD is highly heritable, a multitude of environmental influences play a fundamental role on ASD incidence [4], which might interact with the proband's genetic variants causing the disorder. Though carrying a variant in one allele of ALDH1A3 might not cause a defined phenotype such as A/M, this can act as a risk allele to ASD since lower levels of RA might contribute to a poor transcription of several genes, one of them being FOXN1. Additionally, the second gene variant present in FOXN1, might affect this transcription factor's function by not properly regulating yet uncovered pathways in brain. Finally, when attributing causation to *de novo* variants prudence needs to be exercised.

Supporting Information

S1 Table. Retinoic Acid Response Elements (RAREs) found 30000bp upstream and 1000bp downstream from the initial transcription site in both DNA strands for ALDH1A3 and FOXN1 genes in humans, and Aldh1a3 and Foxn1 genes in mice. Relative positions are given from the initial transcription site of each gene. (PDF)

S2 Table. Primer sequences flanking conserved regions of Aldh1a3 and Foxn1 transcripts in mice. Sequences are given $5' \rightarrow 3'$ direction. (PDF)

S3 Table. Primer sequences used to amplify each flanking RARE bioinformatically predicted sequences for Aldh1a3 and Foxn1 genes in mice. Sequences are given $5' \rightarrow 3'$ direction. (PDF)

Acknowledgments

We thank the children, their families for deciding to be part of this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Liga Colombiana de Autismo participant's member list is as follows: Edith B. Roncancio, Camilo O. Vargas, Francisco J. Paredes, Sandra M. Grisales, Anyela M. Peña, Miguel A. Robayo, María E. Rodriguez and María E. Reyes.

Author Contributions

Conceived and designed the experiments: MCL OAM NBH. Performed the experiments: OAM AMO. Analyzed the data: OAM MCL. Contributed reagents/materials/analysis tools: NBH MCL. Wrote the paper: OAM MCL NBH. Patient referral, clinical and psychological evaluation: Liga Colombiana de Autismo.

References

- Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet. 2008; 9:341–55. doi: <u>10.1038/nrg2346</u> PMID: <u>18414403</u>
- 2. Constantino JN, Todd RD. Autistic traits in the general population: a twin study. Arch Gen Psychiatry. American Medical Association; 2003 May 1; 60(5):524–30. PMID: <u>12742874</u>

- Zaroff CM, Uhm SY. Prevalence of autism spectrum disorders and influence of country of measurement and ethnicity. Soc Psychiatry Psychiatr Epidemiol. 2012 Mar; 47(3):395–8. doi: <u>10.1007/s00127-011-</u> <u>0350-3</u> PMID: <u>21293843</u>
- Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA. American Medical Association; 2014 May 7; 311(17):1770–7. doi: <u>10.1001/jama.2014.</u> 4144 PMID: 24794370
- Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics. 2011 Sep; 128(3): e488–95. doi: <u>10.1542/peds.2010-2825</u> PMID: <u>21844053</u>
- Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. Elsevier; 2011 Jun 9; 70(5):863–85. doi: <u>10.1016/j.neuron.2011.05.002</u> PMID: <u>21658581</u>
- Vorstman JAS, van Daalen E, Jalali GR, Schmidt ERE, Pasterkamp RJ, de Jonge M, et al. A double hit implicates DIAPH3 as an autism risk gene. Mol Psychiatry. Macmillan Publishers Limited; 2011 Apr; 16 (4):442–51. doi: <u>10.1038/mp.2010.26</u> PMID: <u>20308993</u>
- Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am J Hum Genet. 2014 Apr 1; 94(5):677–94. doi: 10.1016/j.ajhg.2014.03.018 PMID: 24768552
- O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011 Jun; 43(6):585–9. doi: <u>10.1038/</u> <u>ng.835</u> PMID: <u>21572417</u>
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012 May 10; 485(7397):246–50. doi: <u>10.</u> <u>1038/nature10989</u> PMID: <u>22495309</u>
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012 May 10; 485(7397):242–5. doi: <u>10.1038/nature11011</u> PMID: <u>22495311</u>
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012 May 10; 485(7397):237– 41. doi: 10.1038/nature10945 PMID: 22495306
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. Elsevier; 2012 Apr 26; 74(2):285–99. doi: <u>10.1016/j.neuron.</u> 2012.04.009 PMID: 22542183
- Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. Elsevier; 2012 Dec 21; 151(7):1431–42. doi: 10.1016/j.cell.2012.11.019 PMID: 23260136
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014 Oct 29; 515(7526):209–15. doi: <u>10.</u> <u>1038/nature13772</u> PMID: <u>25363760</u>
- Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. Neuron. 2012 Dec 20; 76 (6):1052–6. doi: <u>10.1016/j.neuron.2012.12.008</u> PMID: <u>23259942</u>
- Sarachana T, Hu VW. Genome-wide identification of transcriptional targets of RORA reveals direct regulation of multiple genes associated with autism spectrum disorder. Mol Autism. 2013 Jan; 4(1):14. doi: <u>10.1186/2040-2392-4-14</u> PMID: <u>23697635</u>
- Fares-Taie L, Gerber S, Chassaing N, Clayton-Smith J, Hanein S, Silva E, et al. ALDH1A3 mutations cause recessive anophthalmia and microphthalmia. Am J Hum Genet. 2013 Feb 7; 92(2):265–70. doi: 10.1016/j.ajhg.2012.12.003 PMID: 23312594
- Roos L, Fang M, Dali C, Jensen H, Christoffersen N, Wu B, et al. A homozygous mutation in a consanguineous family consolidates the role of ALDH1A3 in autosomal recessive microphthalmia. Clin Genet. 2013 Sep 11;
- Yahyavi M, Abouzeid H, Gawdat G, de Preux A- S, Xiao T, Bardakjian T, et al. ALDH1A3 loss of function causes bilateral anophthalmia/microphthalmia and hypoplasia of the optic nerve and optic chiasm. Hum Mol Genet. 2013 Aug 15; 22(16):3250–8. doi: <u>10.1093/hmg/ddt179</u> PMID: <u>23591992</u>

- Aldahmesh MA, Khan AO, Hijazi H, Alkuraya FS. Mutations in ALDH1A3 cause microphthalmia. Clin Genet. 2013 Aug; 84(2):128–31. doi: 10.1111/cge.12184 PMID: 23646827
- Mory A, Ruiz FX, Dagan E, Yakovtseva EA, Kurolap A, Parés X, et al. A missense mutation in ALDH1A3 causes isolated microphthalmia/anophthalmia in nine individuals from an inbred Muslim kindred. Eur J Hum Genet. Macmillan Publishers Limited; 2014 Mar; 22(3):419–22. doi: <u>10.1038/ejhg.</u> <u>2013.157</u> PMID: <u>23881059</u>
- Semerci CN, Kalay E, Yıldırım C, Dinçer T, Olmez A, Toraman B, et al. Novel splice-site and missense mutations in the ALDH1A3 gene underlying autosomal recessive anophthalmia/microphthalmia. Br J Ophthalmol. 2014 Jun; 98(6):832–40. doi: <u>10.1136/bjophthalmol-2013-304058</u> PMID: <u>24568872</u>
- Chatzi C, Brade T, Duester G. Retinoic acid functions as a key GABAergic differentiation signal in the basal ganglia. PLoS Biol. 2011 Apr; 9(4):e1000609. doi: <u>10.1371/journal.pbio.1000609</u> PMID: <u>21532733</u>
- Benes FM, Berretta S. GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. Neuropsychopharmacology. 2001 Jul; 25(1):1–27. PMID: <u>11377916</u>
- Rubenstein JLR, Merzenich MM. Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes Brain Behav. 2003 Oct; 2(5):255–67. PMID: <u>14606691</u>
- Molotkova N, Molotkov A, Duester G. Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone. Dev Biol. 2007 Mar 15; 303 (2):601–10. PMID: <u>17207476</u>
- Bredenkamp N, Nowell CS, Blackburn CC. Regeneration of the aged thymus by a single transcription factor. Development. 2014 Apr; 141(8):1627–37. doi: <u>10.1242/dev.103614</u> PMID: <u>24715454</u>
- Chen L, Xiao S, Manley NR. Foxn1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner. Blood. 2009 Jan 15; 113(3):567–74. doi: <u>10.1182/blood-2008-05-156265</u> PMID: <u>18978204</u>
- Nowell CS, Bredenkamp N, Tetélin S, Jin X, Tischner C, Vaidya H, et al. Foxn1 regulates lineage progression in cortical and medullary thymic epithelial cells but is dispensable for medullary sublineage divergence. PLoS Genet. 2011 Nov; 7(11):e1002348. doi: <u>10.1371/journal.pgen.1002348</u> PMID: <u>22072979</u>
- Suckow MA, Rosen ED, Wolter WR, Sailes V, Jeffrey R, Tenniswood M. Prevention of human PC-346C prostate cancer growth in mice by a xenogeneic tissue vaccine. Cancer Immunol Immunother. 2007 Aug; 56(8):1275–83. PMID: <u>17242926</u>
- **32.** Onore C, Careaga M, Ashwood P. The role of immune dysfunction in the pathophysiology of autism. Brain Behav Immun. 2012 Mar; 26(3):383–92. doi: <u>10.1016/j.bbi.2011.08.007</u> PMID: <u>21906670</u>
- 33. Amorosi S, D'Armiento M, Calcagno G, Russo I, Adriani M, Christiano AM, et al. FOXN1 homozygous mutation associated with anencephaly and severe neural tube defect in human athymic Nude/SCID fetus. Clin Genet. 2008 Apr; 73(4):380–4. doi: 10.1111/j.1399-0004.2008.00977.x PMID: 18339010
- Rhinn M, Dollé P. Retinoic acid signalling during development. Development. 2012 Mar; 139(5):843– 58. doi: <u>10.1242/dev.065938</u> PMID: <u>22318625</u>
- Liao W-L, Wang H-F, Tsai H-C, Chambon P, Wagner M, Kakizuka A, et al. Retinoid signaling competence and RARbeta-mediated gene regulation in the developing mammalian telencephalon. Dev Dyn. 2005 Apr; 232(4):887–900. PMID: <u>15736225</u>
- 36. Liao W-L, Liu F-C. RARbeta isoform-specific regulation of DARPP-32 gene expression: an ectopic expression study in the developing rat telencephalon. Eur J Neurosci. 2005 Jun; 21(12):3262–8. PMID: <u>16026464</u>
- Bonhomme D, Pallet V, Dominguez G, Servant L, Henkous N, Lafenêtre P, et al. Retinoic acid modulates intrahippocampal levels of corticosterone in middle-aged mice: consequences on hippocampal plasticity and contextual memory. Front Aging Neurosci. 2014 Jan; 6:6. doi: <u>10.3389/fnagi.2014.00006</u> PMID: 24570662
- Mizee MR, Wooldrik D, Lakeman KAM, van het Hof B, Drexhage JAR, Geerts D, et al. Retinoic acid induces blood-brain barrier development. J Neurosci. 2013 Jan 23; 33(4):1660–71. doi: <u>10.1523/</u> JNEUROSCI.1338-12.2013 PMID: 23345238
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000 Jun; 30(3):205–23. PMID: <u>11055457</u>
- Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994 Oct; 24(5):659–85. PMID: <u>7814313</u>
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15; 25(14):1754–60. doi: 10.1093/bioinformatics/btp324 PMID: 19451168

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15; 25(16):2078–9. doi: <u>10.1093/bioinformatics/btp352</u> PMID: <u>19505943</u>
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep; 20(9):1297–303. doi: 10.1101/gr.107524.110 PMID: 20644199
- DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May; 43 (5):491–8. doi: <u>10.1038/ng.806</u> PMID: <u>21478889</u>
- 45. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6(2):80–92.
- Perea CS, Paternina AC, Gomez Y, Lattig MC. Negative affectivity moderated by BDNF and stress response. J Affect Disord. 2012 Feb; 136(3):767–74. doi: <u>10.1016/j.jad.2011.09.043</u> PMID: <u>22044630</u>
- Buske OJ, Manickaraj A, Mital S, Ray PN, Brudno M. Identification of deleterious synonymous variants in human genomes. Bioinformatics. 2013 Aug 1; 29(15):1843–50. doi: <u>10.1093/bioinformatics/btt308</u> PMID: 23736532
- Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet. 2006 Aug 15; 15 (16):2490–508. PMID: <u>16825284</u>
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. 2003 Jul 1; 31(13):3568–71. PMID: <u>12824367</u>
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009 Jan; 4(7):1073–81. doi: <u>10.1038/nprot.2009.86</u> PMID: <u>19561590</u>
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. Nature Publishing Group; 2010 Apr; 7 (4):248–9. doi: <u>10.1038/nmeth0410-248</u> PMID: <u>20354512</u>
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012 Jan; 7(10):e46688. doi: <u>10.1371/journal.pone.0046688</u> PMID: 23056405
- Balmer JE, Blomhoff R. Gene expression regulation by retinoic acid. J Lipid Res. 2002 Nov; 43 (11):1773–808. PMID: <u>12401878</u>
- Balmer JE, Blomhoff R. A robust characterization of retinoic acid response elements based on a comparison of sites in three species. J Steroid Biochem Mol Biol. 2005 Sep; 96(5):347–54. PMID: <u>16081280</u>
- Hertz GZ, Stormo GD. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics. 1999; 15(7–8):563–77. PMID: <u>10487864</u>
- Wagner E, Luo T, Dräger UC. Retinoic acid synthesis in the postnatal mouse brain marks distinct developmental stages and functional systems. Cereb Cortex. 2002 Dec; 12(12):1244–53. PMID: <u>12427676</u>
- 57. Haider NB, Mollema N, Gaule M, Yuan Y, Sachs AJ, Nystuen AM, et al. Nr2e3-directed transcriptional regulation of genes involved in photoreceptor development and cell-type specific phototransduction. Exp Eye Res. 2009 Sep; 89(3):365–72. doi: <u>10.1016/j.exer.2009.04.006</u> PMID: <u>19379737</u>
- Haider NB, Naggert JK, Nishina PM. Excess cone cell proliferation due to lack of a functional NR2E3 causes retinal dysplasia and degeneration in rd7/rd7 mice. Hum Mol Genet. 2001 Aug 1; 10(16):1619– 26. PMID: <u>11487564</u>
- Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014 Oct 29; 515(7526):216–21. doi: <u>10.1038/</u> <u>nature13908</u> PMID: <u>25363768</u>
- Mahony S, Mazzoni EO, McCuine S, Young RA, Wichterle H, Gifford DK. Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. Genome Biol. 2011 Jan; 12(1):R2. doi: <u>10.</u> <u>1186/gb-2011-12-1-r2</u> PMID: <u>21232103</u>
- Siegenthaler JA, Ashique AM, Zarbalis K, Patterson KP, Hecht JH, Kane MA, et al. Retinoic acid from the meninges regulates cortical neuron generation. Cell. 2009 Oct 30; 139(3):597–609. doi: <u>10.1016/j.</u> cell.2009.10.004 PMID: 19879845
- Amorosi S, Vigliano I, Del Giudice E, Panico L, Maruotti GM, Fusco A, et al. Brain alteration in a Nude/ SCID fetus carrying FOXN1 homozygous mutation. J Neurol Sci. 2010 Nov 15; 298(1–2):121–3. doi: 10.1016/j.jns.2010.08.066 PMID: 20864124

- **63.** Rousso DL, Gaber ZB, Wellik D, Morrisey EE, Novitch BG. Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. Neuron. 2008 Jul 31; 59 (2):226–40. doi: <u>10.1016/j.neuron.2008.06.025</u> PMID: <u>18667151</u>
- 64. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001 Oct 4; 413(6855):519–23. PMID: <u>11586359</u>
- 65. Moreno-Ramos OA, Lattig MC, González Barrios AF. Modeling of the hypothalamic-pituitary-adrenal axis-mediated interaction between the serotonin regulation pathway and the stress response using a Boolean approximation: a novel study of depression. Theor Biol Med Model. 2013 Jan; 10:59. doi: <u>10.</u> 1186/1742-4682-10-59 PMID: 24093582