A Subphenotype-to-Genotype Approach Reveals Disproportionate Megalencephaly Autism Risk Genes

Sierra S. Nishizaki[1,2,3,4], Natasha Ann F. Mariano[1,5], Gabriana N. La[1], José M. Uribe-Salazar[1,6], Gulhan Kaya[1], Derek Sayre Andrews[3,4], Christine Wu Nordahl[3,4], David G. Amaral[3,4], Megan Y. Dennis[1,3,4,6†]

[1]Genome Center, [2]Autism Research Training Program, [3]Department of Psychiatry and Behavioral Sciences, [4]MIND Institute, [5]Postbaccalaureate Research Education Program [6]Department of Biochemistry & Molecular Medicine, University of California Davis, CA, USA

[†]Corresponding author:
Megan Y. Dennis, Ph.D.
University of California, Davis
School of Medicine
One Shields Avenue
Genome Center, 4303 GBSF
Davis, CA 95616
Email: mydennis@ucdavis.edu

ABSTRACT

Among autistic individuals, a subphenotype with brain enlargement disproportionate to height (autism with disproportionate megalencephaly, or ASD-DM) seen at three years of age is associated with co-occurring intellectual disability and poorer prognoses later in life. However, many of the genes contributing to ASD-DM have yet to be delineated. In this study, we aim to identify additional ASD-DM associated genes to better define the genetic etiology of this subphenotype of autism. Here, we expand the previously studied sample size of ASD-DM individuals ten-fold by including probands from the Autism Phenome Project and Simons Simplex Collection, totaling 766 autistic individuals meeting the criteria for megalencephaly or macrocephaly and revealing 153 candidate ASD-DM genes harboring *de novo* protein-impacting variants. Our findings include thirteen high confidence autism genes and seven genes previously associated with DM. Five impacted genes have previously been associated with both autism and DM, including *CHD8* and *PTEN*. By performing functional network analysis, we also narrowed in on additional candidate genes, including one previously implicated in ASD-DM (*PIK3CA*) as well as 184 additional genes previously implicated in ASD or DM alone. Using zebrafish as a model, we performed CRISPR gene editing to generate knockout animals for seven of candidate genes and assessed head-size and induced seizure activity differences. From this analysis, we identified significant morphological changes in zebrafish knockout models of two genes, *ythdf2* and *ryr3*. While zebrafish knockout mutants model haploinsufficiency of assayed genes, we identified a *de novo* tandem duplication impacting *YTHDF2* in an ASD-DM proband. Therefore, we also transiently overexpressed *YTHDF2* by injection of *in vitro* transcribed human mRNA to simulate the patient-identified duplication. Following this, we observed increased head and brain size in the *YTHDF2* overexpression zebrafish matching that of the proband, providing a functional link between *YTHDF2* mutations and DM. Though discovery of additional mutations of ASD-DM candidate genes are required in order to fully elucidate the genetics associated with this severe form of autism, our study was able to provide support for *YTHDF2* as a novel putative ASD-DM gene.

INTRODUCTION

Autism is a group of neurodevelopmental traits characterized by difficulties with communication, social interactions, and behavioral challenges, prevalent in 1 out of 54 children in the United States [1,2]. Autism is highly heritable, with 50–90% of cases estimated to be driven by genetics alone [3–5]. Autism is also highly heterogeneous with recent large-scale whole exome sequencing (WES) of >63,000 autistic probands identifying 125 high confidence autism genes, with the predicted number of genes left to be discovered exceeding 1,000 [6–8]. In particular, leveraging genomic data from autism families—including parents and unaffected siblings—in the Simons Simplex Collection (SSC) [9] has identified coding *de novo* variants estimated to contribute to 30% of diagnoses [10–12]. More recently, whole genome sequencing (WGS) of SSC *de novo* noncoding mutations implicates risk in an additional 4.3% of autism cases [13,14]. Despite the combined efforts to sequence tens of thousands of genomes, known genes still only account for 10-20% of cases, and further work is required to fully elucidate genes and pathways contributing to autism etiology [15]. Combining *de novo* variation with autism sub-phenotyping has been used to address the heterogeneity of autism and identify susceptibility loci for comorbid phenotypes in an acute way [16].

Brain enlargement that is disproportionate to height, known as disproportionate megalencephaly (DM), is enriched in autistic probands with 15% of autistic boys falling under the DM subphenotype (ASD-DM) compared to 6% in typically developing (TD) boys [17]. This comorbidity is associated with more severe cognitive phenotypes, including lower IQ and language use, as well as higher rates of language regression [18–20]. This robust enrichment and distinct presentation support DM as a sub-phenotype of ASD, likely due to a shared genetic etiology between autism and DM. While a handful of genes have been associated with DM—including known autism genes impacting cell cycle and proliferation during embryonic development (*e.g.*, *CHD8* and *PTEN*)—mutations of known candidate genes make up only 3% of megalencephaly in autism probands. This leaves the genetic etiology of a majority of ASD-DM cases undiscovered [21–24]. A recent study using WES from 46 autistic families with macrocephaly (ASD-M)—defined as >2 standard deviations above the mean in head circumference for TD sex and age-matched children—successfully identified mutations in one novel and several known autism candidate genes [25], demonstrating the power of sub-phenotyping ASD-DM leading to genetic discoveries even for reduced sample size.

Zebrafish (*Danio rerio*) are an attractive model for studying neurodevelopmental traits given their rapid development, large number of progeny, transparent bodies, and that ~70% of gene orthologs are shared with humans [26,27]. Previous studies of known ASD-DM genes recapitulate macrocephaly and DM phenotypes in zebrafish knockdown and knockout experiments for *CHD8* and *KMT2E*, respectively [28,29]. This method of knocking down candidate ASD-M genes has also been used systematically to identify the contributing gene in the chromosome 16p11.2 locus in zebrafish [30]. Further, novel technologies such as the VAST BioImaging System allow for the rapid characterization of zebrafish knockout models through the generation of high resolution standardized images [31]. We recently demonstrated the utility of this approach by examining the knockdown of two genes associated with autism and microcephaly, *SLC7A5* and *SYNGAP1*, by assessing head-size phenotypes in CRISPR-generated zebrafish knockout line embryos at three and five days post fertilization (dpf) [32].

In this study, we leveraged high-coverage WGS data from two cohorts, 11 ASD-DM probands from the UC Davis MIND Institute Autism Phenome Project (APP) specifically identified using magnetic resonance

imaging (MRI) data at around three years of age, and 755 ASD-M probands with head circumference data available from the SSC cohort. Together, this represents a >10-fold increase in probands compared to the previous largest study of increased head circumference associated genes in ASD [25]. Using this sub-phenotype-to-genotype analytic strategy, we identified candidate ASD-DM and ASD-M genes harboring *de novo* loss-of-function (LoF) variants, including *CHD8* and *PTEN*, and subsequently functionally validated a subset of candidate genes using zebrafish. From this, we narrowed in on two genes impacting larval head size. One of these genes, *YTHDF2*, also lead to macrocephaly in zebrafish following embryonic microinjection of mRNA, recapitulating the phenotype seen in the ASD-DM proband harboring a partial tandem duplication of *YTHDF2*. Together our sub-phenotyping approach provides a powerful strategy to identify novel ASD-DM candidate genes and validate their role in brain development in a zebrafish model system.

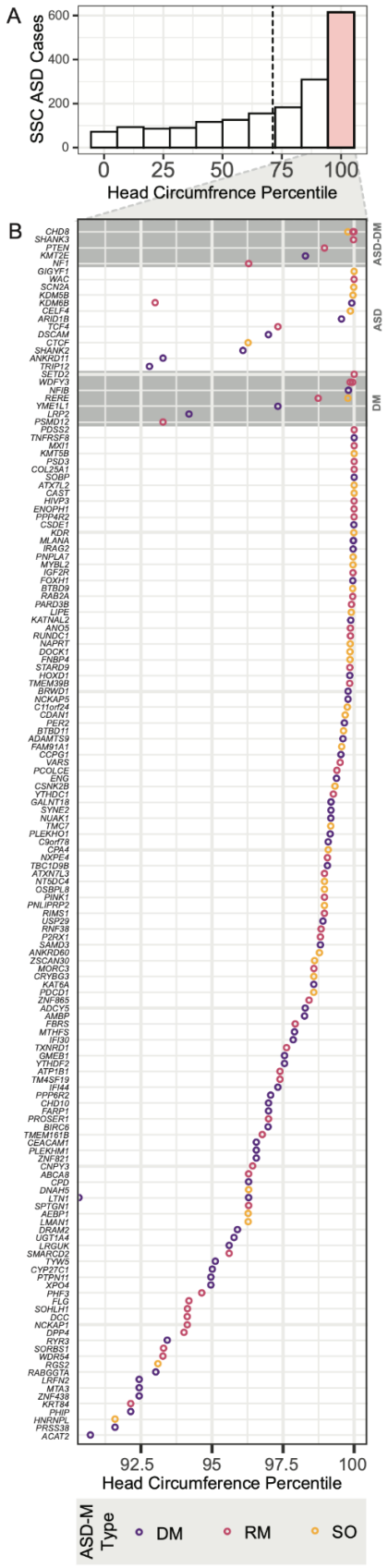
RESULTS

ASD-DM Candidate Gene Discovery

ASD-DM individuals were recruited via the UC Davis APP—a longitudinal study focused on the identification of ASD-subphenotypes [17,33,34]. Using MRI data from the study entry time point (2–3½ years of age) [33], we identified 11/89 WGS sequenced individuals in the APP cohort that met the criteria for ASD-DM, defined as a cerebral volume to height ratio >1.5 standard deviations above the mean compared to TD age-matched controls. Through a collaboration with MSSNG [35,36], WGS and variant identification/annotation was performed for the autistic probands and a subset of family members, for which we also had blood specimens, including six trios and five non-trio probands yielding over 200 thousand variants. From this, we identified two exonic, *de novo*, LoF variants from trio families, including one single-nucleotide variant (SNV) splice-site variant impacting *RYR3* and one 109-kbp duplication of *YTHDF2* and *GMEB1*. An additional ten variants were found in non-trio proband data to be exonic, LoF, and rare (not previously recorded in dbSNP) [37].

We also examined variants from the ASD-DM probands where we do not have sequenced parental genomes to determine *de novo* variant status, but may harbor rare variants associated with their ASD-DM phenotype. From these five individuals, we identified a proband harboring a chromosome 1q21.1 microduplication, a CNV previously associated with ASD-DM [38], and a single proband with variants in *CHD8* and *KMT2E* [39]. We also uncovered ten ultra-rare variants in these probands, defined as previously undescribed in dbSNP (dbSNP). Of these, three impacted genes have SFARI scores of 3S or above (*KMT2E*, *RPS6KA5*, and *TTN*), an additional three genes have known neuronal functions (*DMBT1*, *IARS2*, *FGF12*), and one gene was found recurrently carrying variants in two probands (*SPANXN4*) [40].

The SSC resource consists of genetic and phenotypic information from trios and quads of simplex autism families. SSC head circumference and age data was used to determine ASD-M status (head circumference >1.5 standard deviations above the mean for typically developing sex and age-matched children) for 755 of 1,846 SSC probands (40%) [9] (Supplementary Table 1) (Supplementary Figure 1). *De novo* variants from WGS of ASD-M trios and quads were assembled from previously published SSC *de novo* exonic variant calls [13]. In total, 157 ASD-M *de novo* variants impacted coding exons and were predicted to be LoF, overlapping a total of 150 genes (Figure 1) (Supplementary Table 2). Of note, five genes were found to



contain variants in multiple probands, including *GALNT18*, *KDM6B*, *LTN1*, *RERE*, and *WDFY3*, as well as *CHD8*, which was disrupted in three distinct probands. For SSC probands, only SNVs and indels were considered. Due to the lack of MRI data for SSC participants, for this study we used ASD-M as a proxy for ASD-DM.

In total, we identified 153 genes containing a LoF variant in the APP ASD-DM and SSC ASD-M datasets. Rates of harboring a LoF *de novo* variant in ASD-DM probands were in line with previous predictions (19.4%) and nominally enriched compared to TD SSC siblings (16.5%), though not statistically significant (chi-squared p-value = 0.1) [10,25]. Over a third of identified candidate genes (53/153) had a pLI score of > 0.9, suggesting they are intolerant to variation [41] (Supplementary Table 2). Examining *de novo* missense variants, which previously exhibit overall enrichments in affected probands versus unaffected siblings [42,43], we did not observe an enrichment in our LoF candidate genes in ASD-M probands compared to TD siblings of ASD-M probands, ASD-N probands, and siblings of ASD-N probands (Student's T-test, p-values = 0.16, 0.29, 0.76).

ASD-DM Candidate Gene Discovery Using Network Analysis

Identifying shared patterns of molecular functions and ontologies of impacted ASD-DM genes may point to additional gene candidates. Due to the highly heterogeneous nature of ASD, this type of analysis expands our ability to identify disrupted biological mechanisms and spatio-temporal expression patterns implicated in autism [24]. Here, we used as seeds 166 previously-known and identified-in-this-study ASD-DM genes to identify active interactions using the STRING

Figure 1. Macrocephaly level of candidate ASD-DM and ASD-M genes. A. A histogram representing the number of SSC probands v. head circumference percentiles shows a skew towards larger head-sizes compared to age and sex-matched typically developing children. The red bar designates those meeting the criteria for macrocephaly. The dashed line represents the distribution mean. B. ASD-DM and ASD-M genes listed by their identified proband's head circumference percentiles show genes previously associated with ASD-DM (first grey quadrant) are more likely to be associated with a higher head circumference percentiles than genes previously associated with autism (second white quadrant) and DM (third grey quadrant) alone. Color represents the macrocephaly type. DM, disproportionate macrocephaly; RM, relative macrocephaly; SO, somatic overgrowth.

database (Figure 2) [25,44]. This analysis uncovered ontology groups enriched in our dataset previously reported for ASD, including proteins involved in histone modification and chromatin organization, transcription factors, cell signaling (e.g., SMAD and E-box binding), functions key to neuronal activity (e.g., sodium and calcium ion transport and glutamate receptor binding), cell adhesion and cytoskeletal proteins, and mRNA binding (Supplementary Table 3) [45–48]. Out of our original ASD-DM candidate seed genes 41.5% (69/166) fall under one of these ontologies.

In order to identify ontologies specific to ASD-DM compared to ASD and DM alone we used the Database for Annotations, Visualization, and Integrated Discovery (DAVID) to identify unique ontologies enriched by the 166 known and candidate ASD-DM genes compared to ontologies enriched in SSC ASD-N proband LoF genes, and genes previously associated with DM [49,50]. ASD-DM is uniquely enriched for terms such as intellectual disability, synapse assembly and long term synaptic depression, histone methyltransferase activity, cytoskeletal structure (spectrin repeats) (Supplementary Table 3). Though not reaching statistical significance, ASD-DM is also weakly enriched for mRNA binding with 2/19 previously annotated m^6A-mRNA regulating genes found in ASD-DM (*YTHDC1*, *YTHDF2*) with no hits in ASD-N or in TD SSC siblings [51]. ASD-DM was also not enriched for genes in the MTOR pathway (*NF1* and *PTEN*) compared to ASD-N (*GRB10*, *PPP2R5D*, *RICTOR*, and *TSC1*) [52].

We next sought to identify putative additional candidate ASD-DM by expanding our network to include the top ten interactors for each ASD-DM seed gene (Supplementary Table 4). This list of top ten interactors included genes defined by STRING as having known protein interactions, shared homology, and co-expression patterns [44]. Of the top ten ASD-DM candidate gene interactors, 28% (518/1826) also fall under one of the ontologies found in our ASD-DM network. Interestingly, one of these genes has previously been associated with both autism and DM individually, *PIK3CA* (Supplementary Table 5). *PIK3CA* is a catalytic subunit of the mTOR pathway previously found to be associated with developmental delay and DM, including one diagnosed with autism [53].

In this top ten interactor set, twenty genes are high-confidence autism genes not previously associated with DM (*ANK2*, *ASXL3*, *CTNNB1*, *CUL3*, *DLG4*, *DYRK1A*, *GNAI1*, *GRIN2B*, *KCNMA1, KMT2A*, *NCOA1*, *NIPBL*, *NRXN1*, *PHF12*, *POGZ*, *PPP1R9B*, *SIN3A, SMARCC2*, *TBL1XR1*, *UBR1*). Eighteen additional genes from this interactor set have been implicated in DM and implicated as SFARI putative autism candidate genes (*ANK3*, *CHD2*, *CHD3*, *FRMPD4*, *HCFC1*, *HDAC4*, *HRAS*, HUWE1, *PAK1*, *PIK3R2*, *RAC1*, *SETD1A*, *SLC25A1*, *SMAD4*, *TBL1X*, *TRIO*, *USP7*, and *USP9X*), and 184 more have been implicated in DM or have a SFARI score. Especially promising are the 21 genes that contain missense variants in the SSC ASD-M probands, but do not contain *de novo* missense variants in ASD-N probands or their TD siblings, including *ABI2*, *ANK3*, *SRC*, *SRCAP*, *ATP12A*, *BAIAP2*, *CHD13*, *CH815*, *FGG*, *JUP*, *KDM2A*, *KIF20A*, *MAPK8*, *PDGFRB*, *RING1*, *SCN4A*, *SHANK1*, *SMC3*, *TCF3*, *WDR5*, *ZC3H3*. Together, network analysis and ontology point to these genes as promising ASD-DM candidate genes going forward.

Validation of Candidate Genes in a Zebrafish Model System

In order to determine if our candidate ASD-DM genes were associated with a head-size phenotype, we tested seven by CRISPR knockout in zebrafish (Supplementary Figures 2 & 3). We focused our analysis on the previously unstudied candidates from the APP cohort, including the three genes impacted by *de novo* variants, *RYR3*, *GMEB1*, and *YTHDF2*. It is known that ~20% of zebrafish genes conserved with human are also duplicated; we therefore prioritized candidate ASD-DM genes harboring rare variants in our ASD-DM non-trio dataset that have a single ortholog in zebrafish (*IARS2* and *RPS6KA*) [54]. Additionally, we included two genes from the SSC ASD-M cohort, *CHD8*, in which CRISPR F_0 mosaic knockout has previously been shown to lead to increased interorbital distance in zebrafish embryos [28], and *FAM91A1*, a gene previously not implicated in ASD. Finally, as controls, we included three genes found in SSC ASD-N probands for which we do not expect to see a head-size phenotype—*HEPACAM2*, *PAX5*, and *SCP2A*.

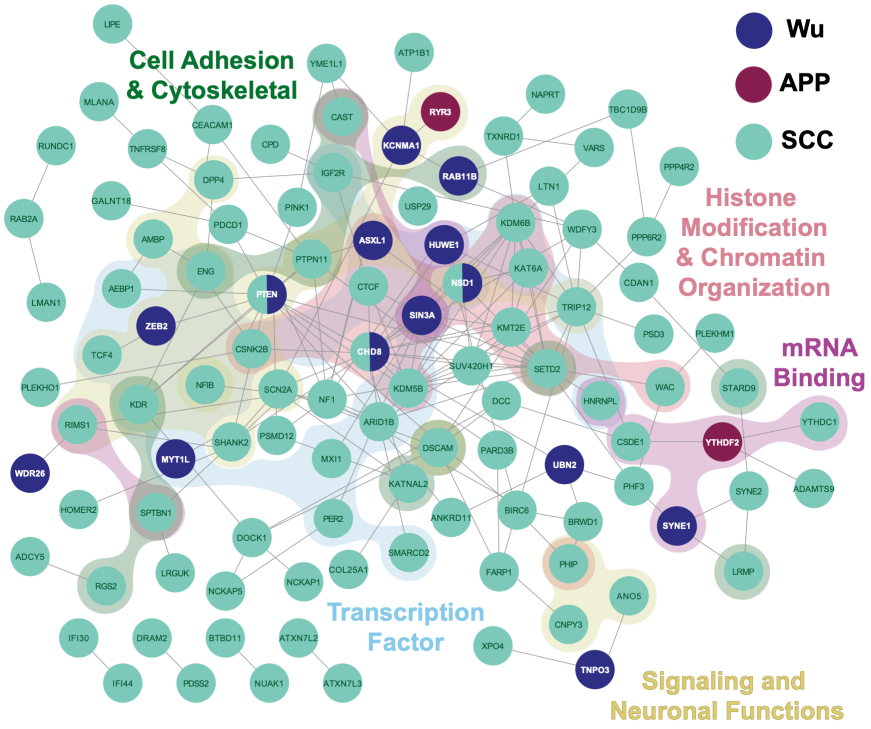


Figure 2. Network analysis and gene ontology of ASD-DM candidate genes. ASD-DM candidate genes from SSC (teal), APP (purple), and Wu (navy) probands are connected in a network via active interactions as determined by STRING [44]. Background colors represent shared gene ontology (GO) molecular functions. Disconnected gene nodes are not included.

We generated CRISPR knockout F_0 embryos by microinjection of four guide RNAs (gRNAs) targeting exonic regions for each assessed gene previously shown to result in near complete mosaic knockout of target genes with little off-target effects [25,32,55]. We then used the VAST BioImaging System paired with FishInspector software to perform automated measurements of embryo area, length, and distance between the center of the eyes (as a proxy for head size), at 3 days post fertilization (dpf). From this analysis, we identified two genes (*ryr3* and *ythdf2*) showing significant head-size differences in zebrafish knockout embryos compared to negative scrambled injection controls (Wilcoxon T-test p-values < 0.01) (Figure 3). In both cases, knockout embryos exhibited reduced head size consistent with microcephaly, with *ryr3* showing a mean reduction of 3.3% and *ythdf2* showing a mean reduction of 3.8%. *RYR3* encodes for a ryanodine receptor responsible for calcium transport in muscle and brain tissue, and *YTHDF2* encodes for part of the m^6A-mRNA degradation complex [56,57]. Though *ythdf2* also displayed a reduction of body area

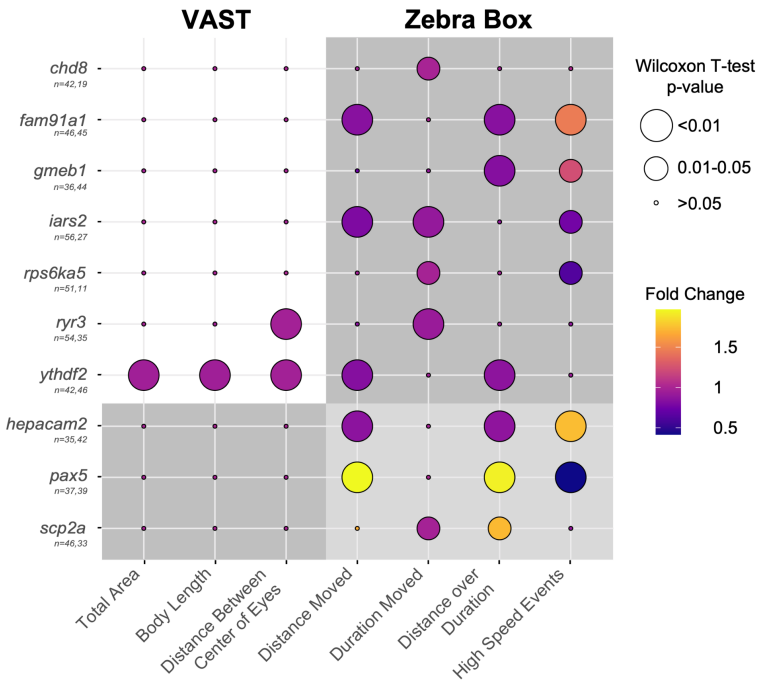


Figure 3. **Zebrafish CRISPR knockout embryos for the rapid validation of ASD-DM genes.** Phenotyping knockout zebrafish morphometric measurements using the VAST BioImaging System, we identified *ryr3* and *ythdf2* as having reduced head size compared to a negative scrambled control. Assessing for increased seizure-like activity in the presence of PTZ using ZebraBox technology we identified three genes associated with increased high speed events, and 3 associated with decreased high speed events. Circle size represents the inverse Wilcoxon T-test p-value. Circle color represents the fold change compared to a negative scrambled control.

and length, this size difference is not due to overall developmental delay measured by head trunk angle (Supplementary Figure 4). Interestingly, only genes harboring *de novo* variants from the ASD-DM APP cohort showed head-size phenotypes, though we did not observe a head-size phenotype in our *chd8* knockout embryos using our morphometric measurements, counter to previously published results [28].

As increased prevalence of seizures is highly enriched in autism [58], we assessed our knockout embryos for increased susceptibility to drug-induced seizures [59] by treating them with GABA antagonist pentylenetetrazole (PTZ; 5 mM) at 5 dpf [32,60]. High speed events, corresponding to seizure-like movements were recorded using ZebraBox [61] and quantified using published software [62] (Figure 3). This analysis revealed three gene knockout models associated with an increase in high speed events versus scrambled controls, *fam91a1*, *gmeb1*, and *hepacam2*. Of note, the human chromosome 7q21.3 microdeletion syndrome containing *HEPACAM2,* an immunoglobulin gene responsible for cell adhesion, which showed the highest fold increase in high speed events compared to controls in our zebrafish assay, have previously been associated with myoclonic seizures in humans [25,63]. Mutations in *FAM91A1*, responsible for Golgi protein trafficking, have been linked to dysregulated electrophysiological brain activity [64,65]. Though previously unassociated with seizures, *GMEB1* - a caspase activation and apoptosis inhibitor, does have a known association to schizophrenia, which shares a historical link with epilepsy [66–68]. Moving forward, we focused specifically on *YTHDF2*, found to be nearly fully duplicated in an APP ASD-DM proband and leading to reduced head size in knockout zebrafish embryos. This due to its clear mechanism of action - gain-of-function due to overexpression.

YTHDF2 in ASD-DM

Based on the findings that CRISPR loss-of-function leads to microcephaly for *ythdf2* and increased seizure activity for *gmeb1*, we sought to better characterize the *de novo* 109 kbp duplication identified in an APP proband with ASD-DM. We validated this CNV using sequence read depth (QuicK-mer2) [69,70] (Figure 4A). Split reads falling at the identified breakpoints of this copy-number variant (CNV) suggested a tandem duplication harboring the entire *GMEB1* gene and the first five of six exons of *YTHDF2* inserted at the 3' untranslated region (UTR) of the noncoding divergent transcript of *TAF12*, directly upstream of *GMEB1* (Figure 4B and C).

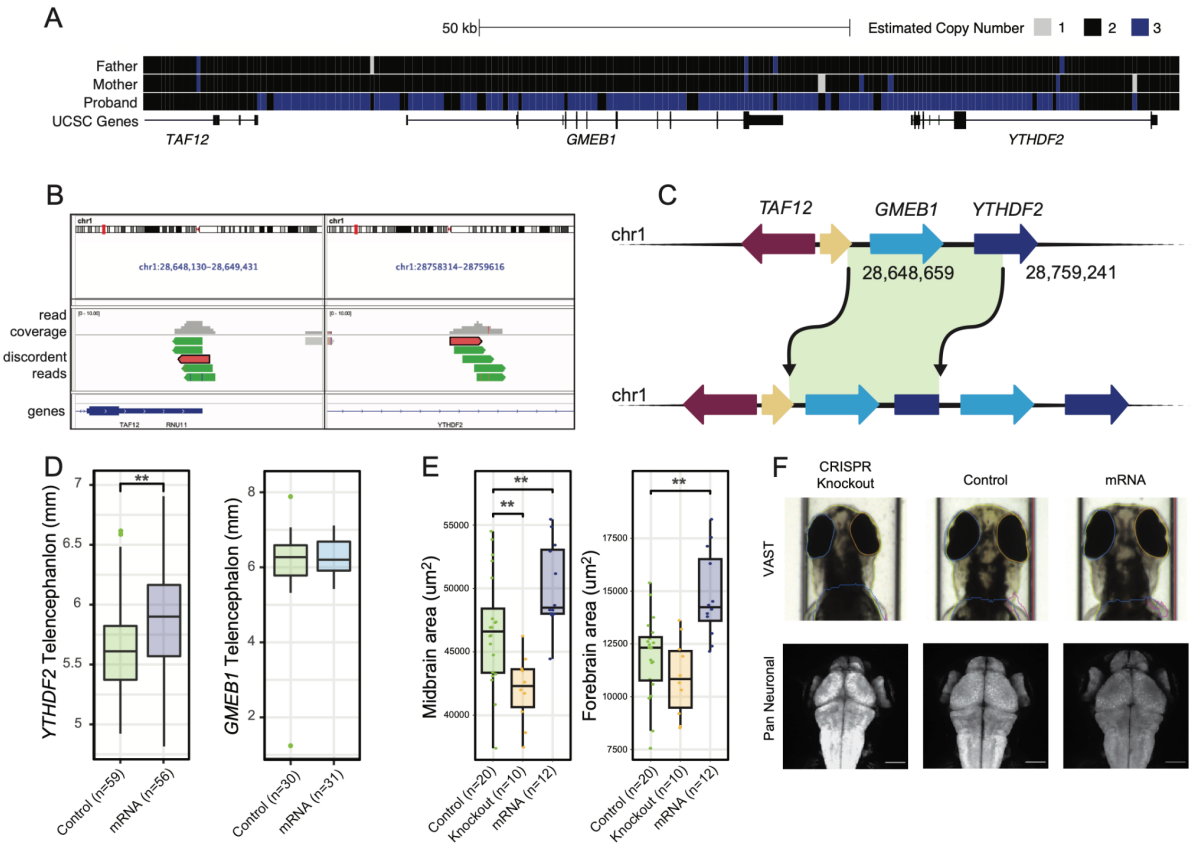


Figure 4. Disrupting *ythdf2* in zebrafish is associated with head and brain size phenotypes. A. QuickMer2 plot showing the predicted copy number of the APP proband harboring a *de novo* duplication a 110kb duplication on chromosome 1 compared to their parents harboring only 2 copies. B. IGV plot showing discordant reads in the APP proband supporting a tandem duplication. C. An illustration of the 110kb tandem duplication on chromosome 1 in an APP proband encompassing *GMEB1* and all but the last exon of *YTHDF2*. D. Head size of mRNA injected zebrafish, estimated via telencephalon measurement, are significantly increased after *YTHDF2* mRNA microinjection into zebrafish embryos (Wilcoxon T-test p-value = 0.0013), and unchanged in *GMEB1* mRNA injected embryos at 3dpf (Wilcoxon T-test p-value = 0.8). E. Knockout and mRNA injected zebrafish harboring a pan neuronal marker reveal brain size differences at 3dpf. Knockout embryos show significantly decreased midbrain volume (Wilcoxon T-test, p-value = 0.003). mRNA injected embryos show both significantly increased midbrain (Wilcoxon T-test, p-value = 0.007) and forebrain (Wilcoxon T-test, p-value = 0.0005). F. Representative control, knockout, and mRNA injected zebrafish images from wildtype embryos imaged with the VAST bioimaging system and HuC embryos harboring a pan neuronal fluorescent marker. Scale bar is 100 uM.

Using available microarray data produced from mRNA derived from whole venous blood, we found that both *GMEB1* and *YTHDF2* exhibited significantly increased expression in the APP proband harboring the duplication compared to other APP participants (Student's T-test p-value < 0.0001) [71]. *GMEB1* is an auxiliary factor in parvovirus replication known to inhibit apoptosis in neurons and previously associated with schizophrenia [67,72] and *YTHDF2* is a member of the m^6A-containing mRNA degradation complex known to be downregulated in neuronal fate determination [57], making both of these attractive potential ASD-DM candidate genes.

We hypothesized that the knockout of *ythdf2* leading to microcephaly in zebrafish larvae made its ortholog a top candidate for *YTHDF2* overexpression leading to ASD-DM. Based on their known functions, duplication of either gene could plausibly result in neurodevelopmental effects. Therefore, we modeled increased expression of *YTHDF2* and *GMEB1* by microinjecting human *in vitro* transcribed mRNA into single-cell stage zebrafish embryos. We then assessed these 'overexpression' zebrafish embryos at 3 dpf using the VAST BioImaging System (Figure 4D). For *YTHDF2*, we observed increased telencephalon length compared to a dye injected negative control consistent with increased head size but no significant differences to body area or length, matching the proband phenotype. We did not observe any difference in head size for the *GMEB1*-injected larvae.

To verify the increased head size was a result of megalencephaly, we repeated the experiment in the zebrafish transgenic line HuC-eGFP, which harbor a GFP fluorescent pan-neuronal marker (Figure 4E and F). These embryos displayed brain size differences, with knockout embryos showing significantly reduced midbrain, and mRNA injected 'overexpression' embryos exhibiting significantly increased midbrain and forebrain compared to injection controls. Together, the knockout and mRNA injected zebrafish provide evidence that gain of *YTHDF2* is associated with DM while its loss leads to microcephaly.

To better characterize *ythdf2* neurodevelopmental functions, we sought to understand expression of this gene. Previous analyses of ythdf2 expression in zebrafish via *in situ* hybridization revealed widespread robust expression [73]. This agrees with reported widespread expression in both humans and zebrafish, beginning early in development at 3 hpf in zebrafish embryos (Supplementary Figure 5) [74,75] .

Not only is *YTHDF2* highly conserved, with 95% gene conservation to mouse and 72% in zebrafish, it also has very few known annotated variants in humans, suggesting it plays a key role in development [76]. *YTHDF2* is predicted to be intolerant to mutation, with a pLI score of 1, and a LoF splice site variant found in the Gnomad Database found in a single male of Latino descent (GRCh38, chr1:28738331,C>G) [41] (Figure 5A). There are also no annotated CNVs in the 1000 Genomes Project or SSC overlapping *YTHDF2*, and no CNVs in ClinVar encompassing fewer than 16 genes (hg38, chr1:28493687-29242679; del), with some larger CNVs associated with probands diagnosed with intellectual disability and developmental delay [9,77–79]. Notably, two individuals in the SSC database were found to harbor *YTHDF* SNVs, one female proband with a *de novo* missense variant (hg38, chr1:28742997,G>A), and one unaffected parent harboring a LoF splice site mutation (hg38, chr1:28737684,T>C), consistent with incomplete penetrance.

DISCUSSION

The autism sub-phenotype of disproportionate megalencephaly (ASD-DM), which occurs in approximately 15% of autistic boys, is associated with lower language ability at age three and slower gains in IQ across early childhood resulting in a higher proportion with IQs in the range of intellectual disability by age six [17]. Others have reported similar rates of macrocephaly in autism (ASD-M) [80]. In this paper we have examined the genomes of 766 ASD-DM and ASD-M trios and quads from the APP and SSC cohorts to identify 153 ASD-DM candidate genes containing *de novo* LoF variants. We functionally tested seven of these candidate genes using a zebrafish CRISPR knockout model and discovered two of these genes to be associated with reduced head size in zebrafish: *ryr3* and *ythdf2*. In the case of the *ryr3* knockout, our findings of microcephaly are counter to those observed in the autistic proband with an identified LoF splice site variant of *RYR3* resulting in DM (chr15:33634585; G>T). Further experimentation delineating the molecular mechanisms of this proband variant will be necessary in order to determine if reduced head size in knockout zebrafish embryos is due to a potential gain-of-function mechanism or intra-species differences of *ryr3* function. For *ythdf2*, to appropriately model the identified patient *YTHDF2* duplication for which we hypothesize gene gain-of-function effects, we overexpressed *YTHDF2* in zebrafish, recapitulating the increased head-size phenotype. Together, this paper introduces two novel ASD-DM candidate genes associated with head-size phenotypes in a zebrafish model system, 141 novel unvalidated ASD-DM candidate genes, and establishes a rapid pipeline for the identification and validation of these genes. In addition to *YTHDF2* and *RYR3*, the candidate genes identified in this paper as harboring *de novo* LoF variation in ASD-DM APP and ASD-M SSC probands also make attractive targets for future phenotypic validation. This list can be further prioritized by choosing to follow-up on genes with similar functions to known ontologies in autism and ASD-DM, as well as ranking head-size phenotypes in the probands. With this expanded list of ASD-DM candidate genes, network analysis can now be leveraged to identify additional candidate genes with similar gene functions to known ASD-DM genes.

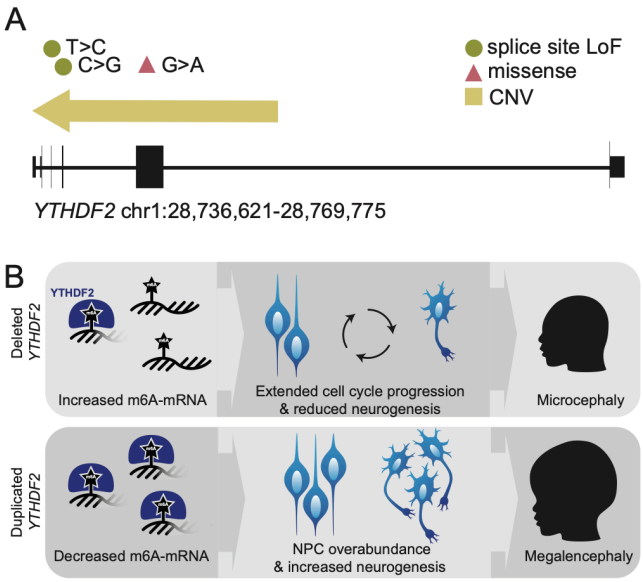


Figure 5. A potential role for *YTHDF2* in ASD-DM. A. Known *YTHDF2* variants are rare, and primarily found in autistic probands and family members. B. Proposed model of *YTHDF2* deletion and duplication phenotypes. We hypothesize *YTHDF2* deletion would lead to microcephaly due to lack of m6-mRNA degradation, extended cell cycle progression, and reduced neurogenesis. Inversely, *YTHDF2* duplication would lead to megalencephaly following increased m^6A-mRNA degradation, neural progenitor cell (NPC) overabundance, and increased neurogenesis.

Table 1. Candidate ASD-DM genes.

Category	Candidate Genes
ASD-DM (n=5)	*CHD8* [3], *KMT2E* [ASD-N], *NF1* [DP], *PTEN* [DP], *SHANK3* [DP]
ASD (n=13)	*ANKRD11* [ASD-N], *ARID1B* [ASD-N], *CELF4*, *CTCF*, *DSCAM* [DP, ASD-N], *GIGYF1* [DP, ASD-N], *KDM5B* [DP], *KDM6B* [2, DP], *SCN2A* [DP, ASD-N], *SHANK2*, *TCF4* [DP], *TRIP12*, *WAC* [DP, ASD-N]
DM (n=7)	*LRP2* [DP], *NFIB* [NO], *PSMD12*, *RERE* [2,DP], *SETD2*, *WDFY3* [2], *YME1L1* [DP]
Other (n=128)	*ABCA8* [DP], *ACAT2*, *ADAMTS9*, *ADCY5* [ASD-N], *AEBP1* [DP], *AMBP*, *ANKRD60* [NO], *ANO5* [DP], *ATP1B1* [DP, ASD-N], *ATXN7L2* [DP], *ATXN7L3*, *BIRC6*, *BRWD1*, *BTBD11* [DP], *BTBD9*, *C11orf24*, *C9orf78*, *CAST*, *CCPG1*, *CDAN1*, *CDH10* [DP], *CEACAM1*, *CNPY3*, *COL25A1*, *CPA4* [DP], *CPD* [DP], *CRYBG3* [NO], *CSDE1*, *CSNK2B*, *CYP27C1*, *DCC*, *DNAH5*, *DOCK1*, *DPP4*, *DRAM2* [DP], *ENG* [DP], *ENOPH1*, *FAM91A1*, *FARP1*, *FBRS*, *FLG* [NO], *FNBP4*, *FOXH1*, *GALNT18* [2,DP], *GMEB1*, *HIVEP3* [DP], *HNRNPL* [DP], *HOXD1*, *IFI30* [DP], *IFI44* [DP], *IGF2R*, *KAT6A*, *KATNAL2* [DP], *KDR*, *KMT5B*, *KRT84* [DP], *LIPE* [DP], *LMAN1*, *LRFN2* [DP], *LRGUK*, *LRMP*, *LTN1* [2], *MLANA* [NO], *MORC3* [DP], *MTA3*, *MTHFS* [DP], *MXI1*, *MYBL2* [DP], *NAPRT*, *NCKAP1* [ASD-N], *NCKAP5* [NO], *NT5DC4*, *NUAK1* [DP], *NXPE4* [DP], *OSBPL8*, *P2RX1*, *PARD3B* [DP], *PCOLCE* [DP], *PDCD1* [NO], *PDSS2*, *PER2*, *PHF3*, *PHIP*, *PINK1**, *PLEKHM1*, *PLEKHO1* [DP], *PNLIPRP2* [NO], *PNPLA7* [DP], *PPP4R2* [DP], *PPP6R2* [DP], *PROSER1*, *PRSS38* [NO], *PSD3*, *PTPN11** [,DP], *RAB2A*, *RABGGTA*, *RGS2*, *RIMS1* [DP, ASD-N], *RNF38*, *RUNDC1*, *RYR3*, *SAMD3* [NO], *SMARCD2*, *SOBP* [DP], *SOHLH1* [NO], *SORBS1*, *SPTBN1*, *STARD9*, *SYNE2* [DP], *TBC1D9B*, *TM4SF19* [DP], *TMC7* [NO], *TMEM161B*, *TMEM39B*, *TNFRSF8* [NO], *TXNRD1*, *TYW5*, *UGT1A4* [DP], *USP29* [DP], *VARS**, *WDR54*, *XPO4*, *YTHDC1*, *YTHDF2*, *ZNF438*, *ZNF821*, *ZNF865*, *ZSCAN30* [NO]

[2,3] Number of occurrences in cohort, if more than 1
[DP] Duplicate paralogs in zebrafish
[NO] No ortholog in zebrafish
[ASD-N] LoF variant in ASD-N SSC proband

We hypothesize that the strong enrichment between autism and DM point to a shared genetic etiology between these two traits. As 43% of *de novo* likely-gene-disrupting variants in probands have been estimated to contribute to an autism diagnosis, we expect many genes identified in this study to contribute to ASD-DM or autism alone [10]. Out of the 153 genes we identified harboring a variant in an ASD-DM or ASD-M proband, five have previously been associated with both autism and DM (Table 1). These include well-characterized ASD-DM genes *CHD8* and *PTEN*. *CHD8*, a chromatin remodeler important in early brain development for which we recovered three distinct variants, is one of the genes most strongly associated with autism and intellectual disability and [28,81]. *PTEN,* another high confidence autism gene, is a tumor suppressor gene known to function in cell proliferation [82]. To date, >26 variants of *CHD8* and >40 variants of *PTEN* have been identified in autistic probands, and both genes have been estimated to have an 85% penetrance of overgrowth and macrocephaly phenotypes [83,84]. Together, *CHD8* and *PTEN* variants have been estimated to contribute to up to 15% of all ASD-M cases [25]. Additionally, we recovered rarer high confidence autism genes also associated with macrocephaly, *KMT2E*, *SHANK3*, and *WDFY3* [6,50]. Over 31 variants have been found in *KMT2E,* which plays a role in regulating histone methylation, in cases of ASD, with an estimated 45% macrocephaly penetrance [85]. Mutations in *SHANK3*, a postsynaptic scaffolding protein, are known to be causal in the ASD-M syndromic condition Phelan–McDermid syndrome [86]. *WDFY3*, which we recovered twice in our cohorts, encodes for an autophagy scaffolding protein [87]. Interestingly, LoF variants in *WDFY3* are associated with macrocephaly, while a likely gain-of-function *WDFY3* variant has been found in an autistic individual with microcephaly.

We recovered thirteen additional high confidence autism genes not previously associated with DM [6,50] (Table 1). For 5/13 of these high confidence autism genes, including *CELF4*, *CTCF*, *KDM5B*, *KDM6B*, and *TRIP12*, we identified no LoF variants in the ASD-N probands and a variant has been previously identified in at least one ASD-DM proband, suggesting a shared etiology contributing to both autism and DM may exist for these genes but has yet to be robustly described [88–92]. Of note, *ARID1B* was identified as a candidate ASD-M gene by a previous study [25]. We identified an additional seven genes previously associated with DM, but not known to be a high confidence autism gene. Of these genes, 6/7, including *LRP2*, *NFIB*, *PSMD12*, *RERE*, *SETD2*, and *WDFY3* have a SFARI score of two or above, suggesting they have been found in a case of ASD, but do not meet the criteria for a high confidence autism gene [40]. Together, these 20 genes make attractive candidate genes, with shared gene functions that mirror high-

confidence ASD-DM genes, such as cell signaling and neuronal processes (8/19) and histone and chromatin processes (5/20). Out of the remaining 128 genes not previously associated with autism or DM, 26 have SFARI scores two and above [40]. Of these, we also identified 6 genes found to contain variants in more than 1 proband (Table 1). This includes the known ASD-DM chromatin remodeler *CHD8* - found in 3 distinct probands, the autism associated demethylase *KDM6B*, the DM-associated retinoic acid signaling *RERE* and the autophagy-linked *WDFY3*, as well as two gene previously associated with either autism or DM include *GALNT18* which functions in O-linked glycosylation and *LTN1* which functions as a ubiquitin ligase [28,87,93–96].

In addition to the work described in this paper, previously published work supports roles of *YTHDF2* in neurodevelopment. *YTHDF2* is a m^{6}A-specific RNA binding protein which designates m^{6}A labeled RNA for degradation [97]. It is known to bind over 3,000 transcripts, including targets previously associated with ASD, such as *CREBBP* [57]. Loss of m^{6}A-RNA degradation activity has been studied through the depletion of methyltransferase complex proteins, which leads to delayed neurogenesis and extended cell cycle progression of cortical neural progenitors—a mechanism associated with microcephaly [98,99]. Additionally, the downregulation of *YTHDF2* is required for neuronal fate determination and loss of *YTHDF2* is associated with delayed mitotic entry [100,101]. *YTHDF2* is also known to function inversely to the ASD-syndrome gene *FMRP*, loss of which is associated with Fragile X Syndrome, which protects its target m^{6}A-RNA from degradation by *YTHDF2* [102]. Knockout mouse models of *Ythdf2* have also shown decreased cortical thickness caused by decreased neurogenesis in early development [103]. Meanwhile, studies in pigs have associated reduced *Ythdf2* activity with inhibited porcine induced pluripotent stem cell pluripotency [104]. Together, this evidence suggests a model of *YTHDF2* duplication in which neuronal fate determination is delayed, leading to the overabundance of neuronal progenitor stem cells followed by increased neurogenesis (Figure 5B).

Interestingly, we identified another gene coding for a m^{6}A-specific RNA binding protein identified in the SSC ASD-M cohort, *YTHDC1*, an m^{6}A-RNA reader that facilitates recruitment of splicing factors [105]. Where gain of *YTHDF2* may lead to a decrease in overall m^{6}A-mRNA through increased degradation, conversely the loss of *YTHDC1* would lead to a decrease of appropriate spliced m^{6}A-mRNA [106]. Further, viral connections between m^{6}A-RNA regulators and brain size support the role of YTH domain protein family members in head size phenotypes. The m^{6}A-labeled mRNA flavivirus ZIKA is associated with severe congenital microcephaly, with *YTHDF2* found to bind and destabilize viral RNA [107,108]. These antiviral functions are controlled in part by the *METTL3* methyltransferase, which labels viral RNA for degradation, and whose knockout models are also associated with a reduced brain size in mice [109]. Our network analysis also highlighted an enrichment of mRNA binding genes in general, including genes with known splicing activity such as *HNRNPL*, a mechanism previously found to be disrupted in autism [110]. Together the m^{6}A-mRNA pathway including YTH domain proteins and m^{6}A methyltransferases and demethylases as well as mRNA binding genes will make for an interesting future area of study in regard to brain size phenotypes.

From the literature, twenty genes have been implicated as high confidence autism genes and have separately been implicated in DM [6,7,50]. Together, the candidates identified in our analysis as well as previously identified ASD-M candidate genes cover 9/20 of these genes, suggesting further work will be required to capture the full genetic heterogeneity of ASD-DM [25].

In addition to an enrichment of both megalencephaly and macrocephaly in ASD, previous studies have found both to be correlated with decreased IQ and increased levels of language regression [17,111,112]. Additionally, macrocephaly has previously been found to be highly correlated with megalencephaly, with an increased correlation in young children [113]. Unfortunately, due to the dearth of MRI evidence, not all ASD-M probands included in this study will meet the criteria for ASD-DM. Though we expect the overlap to be significant, supported by the majority of APP ASD-DM probands also meeting the criteria for macrocephaly (82%), many of the candidate ASD-DM genes identified in this study may have arisen from false positive ASD-DM probands. Additionally, these genes may harbor de novo LoF by chance, with no contribution to a head or brain size phenotype in the proband. Going forward, it is paramount to validate these candidate ASD-DM genes as related to head or brain-size phenotypes. Additionally, we expect a fraction of cases to represent false negatives due to lack of sensitivity of the instrument, molecular processes such as genetic compensation, and inefficient gene knockout which may explain our lack of phenotype seen in CRISPR knockout *chd8* fish compared to previous models [28,114]. This result highlights a disparity in the CRISPR-VAST pipeline, suggesting other methods will be necessary to uncover false-negative ASD-DM genes from this analysis.

This research represents a significant increase in the number ASD-DM and ASD-M proband genomes analyzed in search of candidate genes. The 153 candidate ASD-DM genes introduced here greatly expand our knowledge of the genetic factors specifically contributing to this severe subphenotype of ASD. Additionally, our research establishes a roadmap for the rapid functional characterization of these putative ASD-DM, proving the methodology needed to validate these candidate genes going forward.

METHODS

Megalencephaly and Macrocephaly

APP probands and determinations of megalencephaly were previously determined as part of the APP study [17]. Acquisition of MRI data for megalencephaly measurements were made during natural nocturnal sleep for children at study enrollment (time point 1), between the ages of 2 and 3.5 [115]. Blood collected during this time point was sequenced using the to 30X coverage whole genome sequencing (WGS) through a collaboration with MSSNG [35,116]. Raw data including FASTQ and VCF files can be accessed through the MSSNG access agreement: https://research.mss.ng.

Macrocephaly cases from the SSC were defined using a permissive cutoff of a head circumference >1.5 standard deviations (90%) above the mean of age matched controls [82]. Age matched TD head circumference data [117] and height data [118] from ages 4-17 were derived from publicly available standards. For males in the SSC cohort 551/1601 (34%) met the criteria for macrocephaly, for females 108/245 (45%) met this criteria.

We identified three types of macrocephaly: (1) somatic overgrowth (SO) with head circumference and height percentiles >90%; (2) disproportionate macrocephaly (DM) with height percentiles over head circumference percentile < 0.7; and (3) relative macrocephaly (RM) with height percentiles over head circumference percentile > 0.7.

Variant Annotation

De novo variants were identified in APP probands as those unique to the proband and absent from either parent via string matching (grep -Fvxf). Variant mapping and type predictions were carried out as part of the MSSNG consortium pipeline [35,116]. We considered LoF variants as those predicted to lead a frameshift, nonsense, or splice site mutation. *De novo* LoF variants were validated by IGV [119]. Rare variants identified using dbSNP as those with a minor allele frequency (MAF) < 0.2% in all 5 1000 genome super populations [77,120]. The presence of all rare and *de novo* variants identified in the APP cohort were validated by visual inspection of sequencing data via the Integrated Genomics Viewer (IGV) [70].

Network and Ontology Analyses

Network analysis of known ASD-DM genes and candidate ASD-DM genes from this study were completed using the STRING database and visualized via Cytoscape [44,121]. Gene ontology analysis was completed for known ASD-DM genes and candidate ASD-DM genes as seed genes along with their top 10 gene interactors determined using the STRING database, similar to previous studies [24]. Similar STRING database Molecular Function GO terminology were pooled. Gene ontology was completed using Database for Annotations, Visualization, and Integrated Discovery (DAVID) software [49]. Human genes were used as background for ontology analyses.

Zebrafish CRISPR and mRNA models

gRNA were chosen using CRISPRScan for gRNA with a score of 35 or higher (Supplementary Table 6) [122]. In order to achieve 90% knockdown efficiency, we designed four gRNA per target gene as in Wu et al. [123]. 2nmol crRNA were ordered from Integrated DNA Technologies (IDT). gRNA efficiency was tested via post-injection using a pooled genomic extraction of 4 embryos followed by 7.5% polyacrylamide gel visualization. Quantification of gRNA efficiency was carried out via amplicon sequencing and the CrispRvariants R package (Supplementary Figure 3, Supplementary Table 6) [124].

Injection mixes of ribonucleic protein (RNP) consisting of 4 pooled gRNA and Cas9 Nuclease, S pyogenes (New England Biolabs, M0386M) were prepared as previously described [32]. Pooled gRNA were microinjected into single cell-stage NHGRI-1 zebrafish embryos to a volume of 0.5 nL/cell as previously described using a Pneumatic MPPI-2 Pressure Injector [125]. Injection needles were made using Sutter Instrument Model P-97 Micropipette puller, with the settings Heat 680, Pull 90, Vel 90, Time 225, and thin wall glass capillaries (4m 1mm x 0.75 Fisher Scientific, 50-821-984). Scrambled injection RNP mix contained a single gRNA designed to have no target in the zebrafish genome.

Human mRNA was generated using *YTHDF2* and *GMEB1* cDNA plasmids (Horizon *YTHDF2*, MHS6278-202827242; *GMEB1*, MHS6278-202827172) [126] and prepared via the *in vitro* transcription kit mMESSAGE mMACHINE™ SP6 Transcription Kit (Thermo Fisher Scientific, AM1340). Injection mix of 100ng/uL mRNA and 0.05% phenol red were prepared as previously described and injected to a volume of 0.5 nL/cell [127].

Zebrafish Morphometric Measurements

We took dorsal and ventral images of 3dpf embryos using the Union Biometrica VAST Bioimaging System with LP Sampler using the built-in camera and manufacturer settings [31]. Zebrafish features were then identified and quantified from VAST using FishInspector software 2.0 [128]. Using FishInspector images were

assessed for total area (contourDV_regionpropsArea), embryo length (contourDV_regionpropsLengthOfCentralLine), distance between the center of the eyes (YdistanceCenter_eye1DV_eye2DV), and telencephalon distance (YdistanceEdge_eye1DV_eye2DV). Statistical analysis was done in R using the ggsignif package using the Wilcoxon test option [126,129].

Fluorescent images were taken using the Andor Dragonfly High Speed Confocal Platform with the iXon Ultra camera. Human *YTHDF2* mRNA (0.5 nL/cell, 100ng/uL human *YTHDF2* mRNA) microinjected Tg(HuC-eGFP) strain zebrafish, which harbor a GFP fluorescent pan-neuronal marker were bathed in 0.003% 1-phenyl-2-thiourea (PTU) in 10% Hank's saline between 20-24pf for 24hrs (Fisher Scientific, 5001443999). At 3dpf zebrafish embryos were embedded in 1% low melt agarose (Thermo Fisher Scientific, BP160-100) and imaged for GFP.

Zebrafish Seizure Analysis

Zebrafish movement was recorded at 30 frames per second for 1 hr with no stimulation using ViewPoint's ZebraBox Technology per manufacturer recommendations [61]. Embryos were maintained at 37C using a polystat water heating system. Movements were detected over 1 second intervals. Distance and time moved were analyzed using a custom R script and high speed events were quantified using a previously published MATLAB script [62].

Zebrafish *In Situ* Hybridization

Sense and anti-sense probes were designed to *ythdf2* and generated via the *in vitro* transcription kit mMESSAGE mMACHINE™ SP6 Transcription Kit (Thermo Fisher Scientific, AM1340) (F - GAAATTAATACGACTCACTATAGGGctcaacaaggcagtggggat; R - GAAATAATTAACCCTCACTAAAGGGAAgatccagcgcacatcgaaac). NHGRI-1 zebrafish embryos were collected and treated with 0.003% 1-phenyl-2-thiourea (PTU) in 10% Hank's saline between 20-24pf for 24hrs (Fisher Scientific, 5001443999). At 3dpf PTU treated-embryos were fixed in 4% paraformaldehyde for 12hrs at 4C (Fisher Scientific, AC416785000). Embryos were washed overnight at 70C in a solution containing 100ng of DIG-labeled probes in 250uL Hybridization Mix (Hybridization Mix: 32.5mL 65% deionized formamide, 12.5mL 5X SSC, 0.1% Tween 20, 0.1mL 50ug/mL heparin, 0.25mL 500ug/mL tRNA Type X, 0.92mL 9.2mM citric acid, 3.48mL water) (20X SSC: 87.65g NaCl, 50.26g citric acid trisodium salt dihydrate, to 500mL with nuclease-free water). Embryos were blocked for 4hrs at room temperature (Blocking solution: 0.2mL 2% sheep serum, 2mL 2mg/mL BSA, 7.8mL PBS-Tw, to 10mL with nuclease-free water) (PBS-Tw: 1xPBS containing 0.1% Tween-20) and treated with anti-DIG antibody (1:5000 in Blocking solution) overnight at 4C. Embryos were stained in the dark (staining solution: 33uL NBT and 16.5uL BCIP in 5mL AP buffer) (AP buffer: 100mL 100mM Tris, 20mL 100mM NaCl, 5mL 5mM $MgCl_2$, 2mL 0.1% Tween 20, to 1L with nuclease-free water) for 1hr at room temperature, and destained in methanol. Embryos were visualized via brightfield using a Leica M205 FA microscope.

ACKNOWLEDGEMENTS

Thank you to the APP research staff, especially Brianna Heath and Melissa Regester, as well as the undergraduate students that provide husbandry and care for our zebrafish. We are grateful to Bruce Appel from the University of Colorado for sharing the HuC-eGFP line. MYD is supported by the NIH (DP2MH119424), SSN is supported by the NIMH Autism Research Training Program T32 (MH073124)

through the UC Davis MIND Institute; NAFM is supported by the NIH as a Postbaccalaureate Research Education Program fellow (R25GM116690). A special thank you to the families who have generously shared their genetic data – you are who all of this research is for.

REFERENCES

1. Rice, C.E., Baio, J., Van Naarden Braun, K., Doernberg, N., Meaney, F.J., Kirby, R.S., and ADDM Network (2007). A public health collaboration for the surveillance of autism spectrum disorders. Paediatr. Perinat. Epidemiol. *21*, 179–190.

2. Maenner, M.J., Shaw, K.A., Baio, J., EdS1, Washington, A., Patrick, M., DiRienzo, M., Christensen, D.L., Wiggins, L.D., Pettygrove, S., et al. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill. Summ. *69*, 1–12.

3. Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C.M., and Reichenberg, A. (2014). The familial risk of autism. JAMA *311*, 1770–1777.

4. Sandin, S., Schendel, D., Magnusson, P., Hultman, C., Surén, P., Susser, E., Grønborg, T., Gissler, M., Gunnes, N., Gross, R., et al. (2016). Autism risk associated with parental age and with increasing difference in age between the parents. Mol. Psychiatry *21*, 693–700.

5. Castelbaum, L., Sylvester, C.M., Zhang, Y., Yu, Q., and Constantino, J.N. (2020). On the Nature of Monozygotic Twin Concordance and Discordance for Autistic Trait Severity: A Quantitative Analysis. Behav. Genet. *50*, 263–272.

6. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568–584.e23.

7. Fu, J.M., Satterstrom, F.K., Peng, M., Brand, H., Collins, R.L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S.P., et al. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat. Genet. *54*, 1320–1331.

8. Leblond, C.S., Le, T.-L., Malesys, S., Cliquet, F., Tabet, A.-C., Delorme, R., Rolland, T., and Bourgeron, T. (2021). Operative list of genes associated with autism and neurodevelopmental disorders based on database review. Mol. Cell. Neurosci. *113*, 103623.

9. Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron *68*, 192–195.

10. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

11. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

12. Belyeu, J.R., Brand, H., Wang, H., Zhao, X., Pedersen, B.S., Feusier, J., Gupta, M., Nicholas, T.J., Brown, J., Baird, L., et al. (2021). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. Am. J. Hum. Genet. *108*, 597–607.

13. An, J.-Y., Lin, K., Zhu, L., Werling, D.M., Dong, S., Brand, H., Wang, H.Z., Zhao, X., Schwartz, G.B., Collins, R.L., et al. (2018). Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. Science *362*, eaat6576.

14. Zhou, J., Park, C.Y., Theesfeld, C.L., Wong, A.K., Yuan, Y., Scheckel, C., Fak, J.J., Funk, J., Yao, K., Tajima, Y., et al. (2019). Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat. Genet. *51*, 973–980.

15. Rylaarsdam, L., and Guemez-Gamboa, A. (2019). Genetic Causes and Modifiers of Autism Spectrum Disorder. Front. Cell. Neurosci. *13*, 385.

16. Liu, X.-Q., Paterson, A.D., Szatmari, P., and Autism Genome Project Consortium (2008). Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. Biol. Psychiatry *64*, 561–570.

17. Amaral, D.G., Li, D., Libero, L., Solomon, M., Van de Water, J., Mastergeorge, A., Naigles, L., Rogers, S., and Nordahl, C.W. (2017). In pursuit of neurophenotypes: The consequences of having autism and a big brain. Autism Research *10*, 711–722.

18. Chawarska, K. (2011). Early Generalized Overgrowth in Boys With Autism. Archives of General Psychiatry *68*, 1021.

19. Nordahl, C.W., Lange, N., Li, D.D., Barnett, L.A., Lee, A., Buonocore, M.H., Simon, T.J., Rogers, S., Ozonoff, S., and Amaral, D.G. (2011). Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. Proc. Natl. Acad. Sci. U. S. A. *108*, 20195–20200.

20. Sacco, R., Militerni, R., Frolli, A., Bravaccio, C., Gritti, A., Elia, M., Curatolo, P., Manzi, B., Trillo, S., Lenti, C., et al. (2007). Clinical, morphological, and biochemical correlates of head circumference in autism. Biol. Psychiatry *62*, 1038–1047.

21. Hormozdiari, F., Penn, O., Borenstein, E., and Eichler, E.E. (2015). The discovery of integrated gene networks for autism and related disorders. Genome Res. *25*, 142–154.

22. Krishnan, A., Zhang, R., Yao, V., Theesfeld, C.L., Wong, A.K., Tadych, A., Volfovsky, N., Packer, A., Lash, A., and Troyanskaya, O.G. (2016). Genome-wide characterization of genetic and functional dysregulation in autism spectrum disorder (bioRxiv).

23. O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619–1622.

24. Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

25. Wu, H., Li, H., Bai, T., Han, L., Ou, J., Xun, G., Zhang, Y., Wang, Y., Duan, G., Zhao, N., et al. (2020). Phenotype-to-genotype approach reveals head-circumference-associated genes in an autism spectrum disorder cohort. Clin. Genet. *97*, 338–346.

26. Abreu, M.S. de, de Abreu, M.S., Genario, R., Giacomini, A.C.V., Demin, K.A., Lakstygal, A.M., Amstislavskaya, T.G., Fontana, B.D., Parker, M.O., and Kalueff, A.V. (2020). Zebrafish as a Model of Neurodevelopmental Disorders. Neuroscience *445*, 3–11.

27. Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498–503.

28. Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerdts, J., Baker, C., Vulto-van Silfhout, A.T., et al. (2014). Disruptive CHD8 mutations define a subtype of autism early in development. Cell *158*, 263–276.

29. Thyme, S.B., Pieper, L.M., Li, E.H., Pandey, S., Wang, Y., Morris, N.S., Sha, C., Choi, J.W., Herrera, K.J., Soucy, E.R., et al. (2019). Phenotypic Landscape of Schizophrenia-Associated Genes Defines Candidates and Their Shared Functions. Cell *177*, 478–491.e20.

30. Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature *485*, 363–367.

31. Pulak, R. (2016). Tools for automating the imaging of zebrafish larvae. Methods *96*, 118–126.

32. Colón-Rodríguez, A., Uribe-Salazar, J.M., Weyenberg, K.B., Sriram, A., Quezada, A., Kaya, G., Jao, E., Radke, B., Lein, P.J., and Dennis, M.Y. (2020). Assessment of autism zebrafish mutant models using a high-throughput larval phenotyping platform. Front. Cell Dev. Biol. *8*, 586296.

33. Nordahl, C.W., Andrews, D.S., Dwyer, P., Waizbard-Bartov, E., Restrepo, B., Lee, J.K., Heath, B., Saron, C., Rivera, S.M., Solomon, M., et al. (2021). The Autism Phenome Project: Toward Identifying Clinically Meaningful Subgroups of Autism. Front. Neurosci. *15*, 786220.

34. Ohta, H., Nordahl, C.W., Iosif, A.-M., Lee, A., Rogers, S., and Amaral, D.G. (2016). Increased Surface Area, but not Cortical Thickness, in a Subset of Young Boys With Autism Spectrum Disorder. Autism Res. *9*, 232–248.

35. C Yuen, R.K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat. Neurosci. *20*, 602–611.

36. Trost, B., Thiruvahindrapuram, B., Chan, A.J.S., Engchuan, W., Higginbotham, E.J., Howe, J.L., Loureiro, L.O., Reuter, M.S., Roshandel, D., Whitney, J., et al. (2022). Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell *185*, 4409–4427.e18.

37. Sherry, S.T., Ward, M., and Sirotkin, K. (1999). dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. *9*, 677–679.

38. Brunetti-Pierri, N., Berg, J.S., Scaglia, F., Belmont, J., Bacino, C.A., Sahoo, T., Lalani, S.R., Graham, B., Lee, B., Shinawi, M., et al. (2008). Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. Nat. Genet. *40*, 1466–1471.

39. Dolcetti, A., Silversides, C.K., Marshall, C.R., Lionel, A.C., Stavropoulos, D.J., Scherer, S.W., and Bassett, A.S. (2013). 1q21.1 Microduplication expression in adults. Genet. Med. *15*, 282–289.

40. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36.

41. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

42. Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E., MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves variant deleteriousness prediction (bioRxiv).

43. Koire, A., Katsonis, P., Kim, Y.W., Buchovecky, C., Wilson, S.J., and Lichtarge, O. (2021). A method to delineate de novo missense variants across pathways prioritizes genes linked to autism. Sci. Transl. Med. *13*,.

44. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

45. Lasalle, J.M. (2013). Autism genes keep turning up chromatin. OA Autism *1*, 14.

46. Hoffmann, A., and Spengler, D. (2021). Single-Cell Transcriptomics Supports a Role of CHD8 in Autism. International Journal of Molecular Sciences *22*, 3261.

47. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am. J. Hum. Genet. *94*, 677–694.

48. Brooks-Kayal, A. (2010). Epilepsy and autism spectrum disorders: are there common developmental mechanisms? Brain Dev. *32*, 731–738.

49. Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. *50*, W216–W221.

50. Pirozzi, F., Nelson, B., and Mirzaa, G. (2018). From microcephaly to megalencephaly: determinants of brain size. Dialogues Clin. Neurosci. *20*, 267–282.

51. Jiang, X., Liu, B., Nie, Z., Duan, L., Xiong, Q., Jin, Z., Yang, C., and Chen, Y. (2021). The role of m6A modification in the biological functions and diseases. Signal Transduct Target Ther *6*, 74.

52. Reijnders, M.R.F., Kousi, M., van Woerden, G.M., Klein, M., Bralten, J., Mancini, G.M.S., van Essen, T., Proietti-Onori, M., Smeets, E.E.J., van Gastel, M., et al. (2017). Variation in a range of mTOR-related genes associates with intracranial volume and intellectual disability. Nature Communications *8*,.

53. Yeung, K.S., Tso, W.W.Y., Ip, J.J.K., Mak, C.C.Y., Leung, G.K.C., Tsang, M.H.Y., Ying, D., Pei, S.L.C., Lee, S.L., Yang, W., et al. (2017). Identification of mutations in the PI3K-AKT-mTOR signalling pathway in patients with macrocephaly and developmental delay and/or autism. Molecular Autism *8*,.

54. Kozol, R.A., Abrams, A.J., James, D.M., Buglo, E., Yan, Q., and Dallman, J.E. (2016). Function Over Form: Modeling Groups of Inherited Neurological Conditions in Zebrafish. Front. Mol. Neurosci. *9*, 55.

55. Kroll, F., Powell, G.T., Ghosh, M., Gestri, G., Antinucci, P., Hearn, T.J., Tunbak, H., Lim, S., Dennis, H.W., Fernandez, J.M., et al. (2021). A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. Elife *10*,.

56. Nilipour, Y., Nafissi, S., Varasteh, V., Hossein-Nejad, H., Tonekaboni, S., Ravenscroft, G., Olivé, M., Laing, N., and Tajsharghi, H. (2016). Ryanodine receptor type 3 (RYR3) as a novel gene associated with nemaline myopathy and fibre type disproportion. Neuromuscular Disorders *26*, S137.

57. Sokpor, G., Xie, Y., Nguyen, H.P., and Tuoc, T. (2021). Emerging Role of m A Methylome in Brain Development: Implications for Neurological Disorders and Potential Treatment. Front Cell Dev Biol *9*, 656849.

58. Frye, R. (2013). A review of traditional and novel treatments for seizures in autism spectrum disorder: findings from a systematic review and expert panel. Frontiers in Public Health *1*,.

59. Frye, R.E., Casanova, M.F., Fatemi, S.H., Folsom, T.D., Reutiman, T.J., Brown, G.L., Edelson, S.M., Slattery, J.C., and Adams, J.B. (2016). Neuropathological Mechanisms of Seizures in Autism Spectrum Disorder. Front. Neurosci. *10*, 192.

60. Baraban, S.C., Taylor, M.R., Castro, P.A., and Baier, H. (2005). Pentylenetetrazole induced changes in zebrafish behavior, neural activity and c-fos expression. Neuroscience *131*, 759–768.

61. Lee, D.A., Oikonomou, G., and Prober, D.A. (2022). Large-scale Analysis of Sleep in Zebrafish. Bio Protoc *12*, e4313.

62. Griffin, A., Carpenter, C., Liu, J., Paterno, R., Grone, B., Hamling, K., Moog, M., Dinday, M.T., Figueroa, F., Anvar, M., et al. (2021). Phenotypic analysis of catastrophic childhood epilepsy genes. Commun Biol *4*, 680.

63. Nevado, J., Mergener, R., Palomares-Bralo, M., Souza, K.R., Vallespín, E., Mena, R., Martínez-Glez, V., Mori, M.Á., Santos, F., García-Miñaur, S., et al. (2014). New microdeletion and microduplication syndromes: A comprehensive review. Genet. Mol. Biol. *37*, 210–219.

64. Fromme, J.C., and Munson, M. (2017). Capturing endosomal vesicles at the Golgi. Nat. Cell Biol. *19*, 1384–1386.

65. Salmela, E., Renvall, H., Kujala, J., Hakosalo, O., Illman, M., Vihla, M., Leinonen, E., Salmelin, R., and Kere, J. (2016). Evidence for genetic regulation of the human parieto-occipital 10-Hz rhythmic activity. Eur. J. Neurosci. *44*, 1963–1971.

66. Kotsaris, G., Kerselidou, D., Koutsoubaris, D., Constantinou, E., Malamas, G., Garyfallos, D.A., and Hatzivassiliou, E.G. (2020). TRAF3 can interact with GMEB1 and modulate its anti-apoptotic function. J. Biol. Res. *27*, 7.

67. Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J.D., Bass, N., Bigdeli, T.B., Breen, G., Bromet, E.J., et al. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature *604*, 509–516.

68. Nadkarni, S., Arnedo, V., and Devinsky, O. (2007). Psychosis in epilepsy patients. Epilepsia *48 Suppl 9*, 17–19.

69. Shen, F., and Kidd, J.M. (2020). Rapid, Paralog-Sensitive CNV Analysis of 2457 Human Genomes Using QuicK-mer2. Genes *11*,.

70. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

71. Stamova, B.S., Tian, Y., Nordahl, C.W., Shen, M.D., Rogers, S., Amaral, D.G., and Sharp, F.R. (2013). Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. Mol. Autism *4*, 30.

72. Nakagawa, T., Tsuruma, K., Uehara, T., and Nomura, Y. (2008). GMEB1, a novel endogenous caspase inhibitor, prevents hypoxia- and oxidative stress-induced neuronal apoptosis. Neurosci. Lett. *438*, 34–37.

73. Kontur, C., Jeong, M., Cifuentes, D., and Giraldez, A.J. (2020). Ythdf mA Readers Function Redundantly during Zebrafish Development. Cell Rep. *33*, 108598.

74. Broad Institute (2021). The Genotype-Tissue Expression (GTEx) Project.

75. White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Hourlier, T., Füllgrabe, A., et al. A high-resolution mRNA expression time course of embryonic development in zebrafish.

76. Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T.I., Dahary, D., Fishilevich, S., and Lancet, D. (2021). The GeneCards Suite. Practical Guide to Life Science Databases 27–56.

77. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

78. Landrum, M.J., and Kattman, B.L. (2018). ClinVar at five years: Delivering on the promise. Hum. Mutat. *39*, 1623–1630.

79. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. The American Journal of Human Genetics *86*, 749–764.

80. Sacco, R., Gabriele, S., and Persico, A.M. (2015). Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. Psychiatry Res. *234*, 239–251.

81. Weissberg, O., and Elliott, E. (2021). The Mechanisms of CHD8 in Neurodevelopment and Autism Spectrum Disorders. Genes *12*,.

82. Klein, S., Sharifi-Hannauer, P., and Martinez-Agosto, J.A. (2013). Macrocephaly as a clinical indicator of genetic subtypes in autism. Autism Res. *6*, 51–56.

83. Ostrowski, P.J., Zachariou, A., Loveday, C., Beleza-Meireles, A., Bertoli, M., Dean, J., Douglas, A.G.L., Ellis, I., Foster, A., Graham, J.M., et al. (2019). The CHD8 overgrowth syndrome: A detailed evaluation of an emerging overgrowth phenotype in 27 patients. Am. J. Med. Genet. C Semin. Med. Genet. *181*, 557–564.

84. DeSpenza, T., Jr, Carlson, M., Panchagnula, S., Robert, S., Duy, P.Q., Mermin-Bunnell, N., Reeves, B.C., Kundishora, A., Elsamadicy, A.A., Smith, H., et al. (2021). PTEN mutations in autism spectrum disorder and congenital hydrocephalus: developmental pleiotropy and therapeutic targets. Trends Neurosci. *44*, 961–976.

85. O'Donnell-Luria, A.H., Pais, L.S., Faundes, V., Wood, J.C., Sveden, A., Luria, V., Abou Jamra, R., Accogli, A., Amburgey, K., Anderlid, B.M., et al. (2019). Heterozygous Variants in KMT2E Cause a

Spectrum of Neurodevelopmental Disorders and Epilepsy. Am. J. Hum. Genet. *104*, 1210–1222.

86. Golden, C.E.M., Wang, V.X., Harony-Nicolas, H., Hof, P.R., and Buxbaum, J.D. (2021). Reduced brain volume and white matter alterations in *Shank3* -deficient rats. Autism Research *14*, 1837–1842.

87. Le Duc, D., Giulivi, C., Hiatt, S.M., Napoli, E., Panoutsopoulos, A., Harlan De Crescenzo, A., Kotzaeridou, U., Syrbe, S., Anagnostou, E., Azage, M., et al. (2019). Pathogenic WDFY3 variants cause neurodevelopmental disorders and opposing effects on brain size. Brain *142*, 2617–2630.

88. MacPherson, M.J., Erickson, S.L., Kopp, D., Wen, P., Aghanoori, M.-R., Kedia, S., Burns, K.M.L., Vitobello, A., Tran Mau-Them, F., Thomas, Q., et al. (2021). Nucleocytoplasmic transport of the RNA-binding protein CELF2 regulates neural stem cell fates. Cell Rep. *35*, 109226.

89. Watson, L.A., Wang, X., Elbert, A., Kernohan, K.D., Galjart, N., and Bérubé, N.G. (2014). Dual effect of CTCF loss on neuroprogenitor differentiation and survival. J. Neurosci. *34*, 2860–2870.

90. Faundes, V., Newman, W.G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., Davies, S.J., Demos, M.K., Goldman, A., Gill, H., et al. (2018). Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am. J. Hum. Genet. *102*, 175–187.

91. Poot, R.A. (2020). The role of chromatin modifiers in common neurodevelopmental disorders. Stem Cell Epigenetics 279–289.

92. Rossi, M., El-Khechen, D., Black, M.H., Farwell Hagman, K.D., Tang, S., and Powis, Z. (2017). Outcomes of Diagnostic Exome Sequencing in Patients With Diagnosed or Suspected Autism Spectrum Disorders. Pediatric Neurology *70*, 34–43.e2.

93. Stolerman, E.S., Francisco, E., Stallworth, J.L., Jones, J.R., Monaghan, K.G., Keller-Ramey, J., Person, R., Wentzensen, I.M., McWalter, K., Keren, B., et al. (2019). Genetic variants in the KDM6B gene are associated with neurodevelopmental delays and dysmorphic features. Am. J. Med. Genet. A *179*, 1276–1286.

94. Fregeau, B., Kim, B.J., Hernández-García, A., Jordan, V.K., Cho, M.T., Schnur, R.E., Monaghan, K.G., Juusola, J., Rosenfeld, J.A., Bhoj, E., et al. (2016). De Novo Mutations of RERE Cause a Genetic Syndrome with Features that Overlap Those Associated with Proximal 1p36 Deletions. Am. J. Hum. Genet. *98*, 963–970.

95. Coit, P., Ortiz-Fernandez, L., Lewis, E.E., McCune, W.J., Maksimowicz-McKinnon, K., and Sawalha, A.H. (2020). A longitudinal and transancestral analysis of DNA methylation patterns and disease activity in lupus patients. JCI Insight *5*,.

96. Doamekpor, S.K., Lee, J.-W., Hepowit, N.L., Wu, C., Charenton, C., Leonard, M., Bengtson, M.H., Rajashankar, K.R., Sachs, M.S., Lima, C.D., et al. (2016). Structure and function of the yeast listerin (Ltn1) conserved N-terminal domain in binding to stalled 60S ribosomal subunits. Proc. Natl. Acad. Sci. U. S. A. *113*, E4151–E4160.

97. Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

98. Yoon, K.-J., Ringeling, F.R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., Su, Y., Kim, N.-S., Zhu, Y., Zheng, L., et al. (2017). Temporal Control of Mammalian Cortical Neurogenesis by mA Methylation. Cell *171*, 877–889.e17.

99. Gabriel, E., Wason, A., Ramani, A., Gooi, L.M., Keller, P., Pozniakovsky, A., Poser, I., Noack, F., Telugu, N.S., Calegari, F., et al. (2016). CPAP promotes timely cilium disassembly to maintain neural progenitor pool. EMBO J. *35*, 803–819.

100. Heck, A.M., Russo, J., Wilusz, J., Nishimura, E.O., and Wilusz, C.J. (2020). YTHDF2 destabilizes mA-modified neural-specific RNAs to restrain differentiation in induced pluripotent stem cells. RNA *26*, 739–755.

101. Fei, Q., Zou, Z., Roundtree, I.A., Sun, H.-L., and He, C. (2020). YTHDF2 promotes mitotic entry and is regulated by cell cycle mediators. PLoS Biol. *18*, e3000664.

102. Zhang, F., Kang, Y., Wang, M., Li, Y., Xu, T., Yang, W., Song, H., Wu, H., Shu, Q., and Jin, P. (2018). Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. Hum. Mol. Genet. *27*, 3936–3950.

103. Li, Z., Qian, P., Shao, W., Shi, H., He, X.C., Gogol, M., Yu, Z., Wang, Y., Qi, M., Zhu, Y., et al. (2018). Suppression of mA reader Ythdf2 promotes hematopoietic stem cell expansion. Cell Res. *28*, 904–917.

104. Wu, R., Liu, Y., Zhao, Y., Bi, Z., Yao, Y., Liu, Q., Wang, F., Wang, Y., and Wang, X. (2019). mA methylation controls pluripotency of porcine induced pluripotent stem cells by targeting SOCS3/JAK2/STAT3 pathway in a YTHDF1/YTHDF2-orchestrated manner. Cell Death Dis. *10*, 171.

105. Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell *61*, 507–519.

106. Gokhale, N.S., and Horner, S.M. (2017). RNA modifications go viral. PLOS Pathogens *13*, e1006188.

107. Lichinchi, G., Zhao, B.S., Wu, Y., Lu, Z., Qin, Y., He, C., and Rana, T.M. (2016). Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. Cell Host Microbe *20*, 666–673.

108. Gokhale, N.S., McIntyre, A.B.R., Mattocks, M.D., Holley, C.L., Lazear, H.M., Mason, C.E., and Horner, S.M. (2020). Altered m6A Modification of Specific Cellular Transcripts Affects Flaviviridae Infection. Molecular Cell *77*, 542–555.e8.

109. Wang, C.-X., Cui, G.-S., Liu, X., Xu, K., Wang, M., Zhang, X.-X., Jiang, L.-Y., Li, A., Yang, Y., Lai, W.-Y., et al. (2018). METTL3-mediated m6A modification is required for cerebellar development. PLoS Biol. *16*, e2004880.

110. Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. *18*, 437–451.

111. Fombonne, E., Rogé, B., Claverie, J., Courty, S., and Frémolle, J. (1999). Microcephaly and macrocephaly in autism. J. Autism Dev. Disord. *29*, 113–119.

112. Chaste, P., Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., Jeremy Willsey, A., Moreno-De-Luca, D., Yu, T.W., Fombonne, E., et al. (2013). Adjusting Head Circumference for Covariates in Autism: Clinical Correlates of a Highly Heritable Continuous Trait. Biological Psychiatry *74*, 576–584.

113. Bartholomeusz, H.H., Courchesne, E., and Karns, C.M. (2002). Relationship between head circumference and brain volume in healthy normal toddlers, children, and adults. Neuropediatrics *33*,

239–241.

114. El-Brolosy, M.A., Kontarakis, Z., Rossi, A., Kuenne, C., Günther, S., Fukuda, N., Kikhi, K., Boezio, G.L.M., Takacs, C.M., Lai, S.-L., et al. (2019). Genetic compensation triggered by mutant mRNA degradation. Nature *568*, 193–197.

115. Lee, J.K., Andrews, D.S., Ozonoff, S., Solomon, M., Rogers, S., Amaral, D.G., and Nordahl, C.W. (2021). Longitudinal Evaluation of Cerebral Growth Across Childhood in Boys and Girls With Autism Spectrum Disorder. Biol. Psychiatry *90*, 286–294.

116. Chan, A.J.S., Engchuan, W., Reuter, M.S., Wang, Z., Thiruvahindrapuram, B., Trost, B., Nalpathamkalam, T., Negrijn, C., Lamoureux, S., Pellecchia, G., et al. (2022). Genome-wide rare variant score associates with morphological subtypes of autism spectrum disorder. Nat. Commun. *13*, 6463.

117. Rollins, J.D., Collins, J.S., and Holden, K.R. (2010). United States head circumference growth reference charts: birth to 21 years. J. Pediatr. *156*, 907–913.e2.

118. Centers for Disease Control and Prevention, National Center for Health Statistics (2001). Data Table of Stature-for-age Charts.

119. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

120. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

121. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

122. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods *12*, 982–988.

123. Wu, R.S., Lam, I.I., Clay, H., Duong, D.N., Deo, R.C., and Coughlin, S.R. (2018). A rapid method for directed gene knockout for screening in G0 zebrafish. Dev. Cell *46*, 112–125.e4.

124. Lindsay, H., Burger, A., Biyong, B., Felker, A., Hess, C., Zaugg, J., Chiavacci, E., Anders, C., Jinek, M., Mosimann, C., et al. (2016). CrispRVariants charts the mutation spectrum of genome engineering experiments. Nat. Biotechnol. *34*, 701–702.

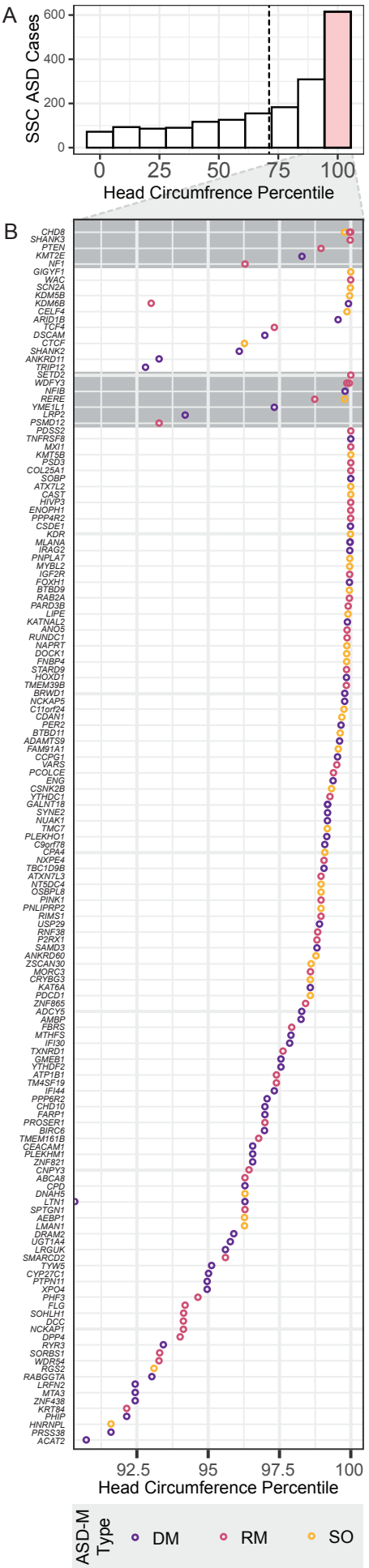
125. Jao, L.-E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc. Natl. Acad. Sci. U. S. A. *110*, 13904–13909.

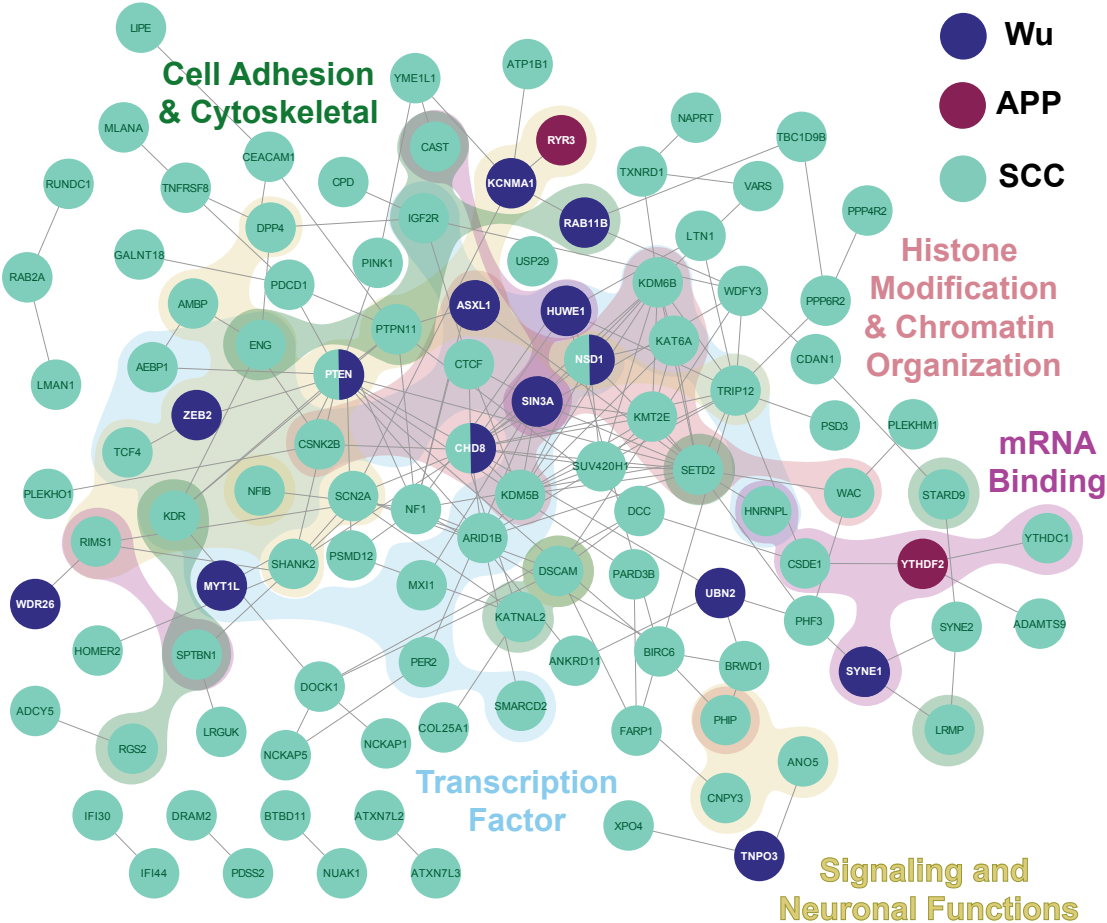
126. (2004). The Status, Quality, and Expansion of the NIH Full-Length cDNA Project: The Mammalian Gene Collection (MGC). Genome Research *14*, 2121–2127.

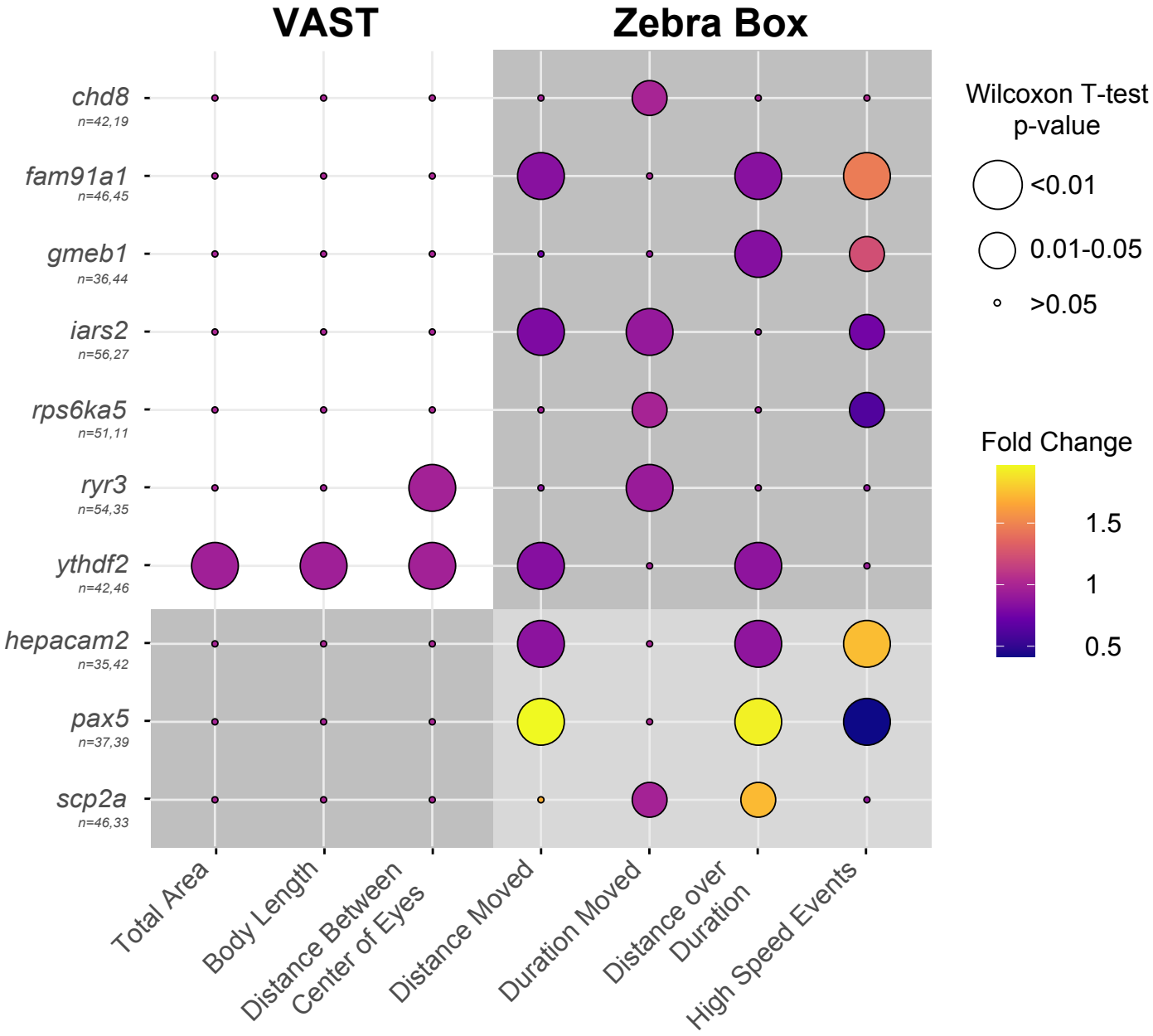
127. Yuan, S., and Sun, Z. (2009). Microinjection of mRNA and morpholino antisense oligonucleotides in zebrafish embryos. J. Vis. Exp.

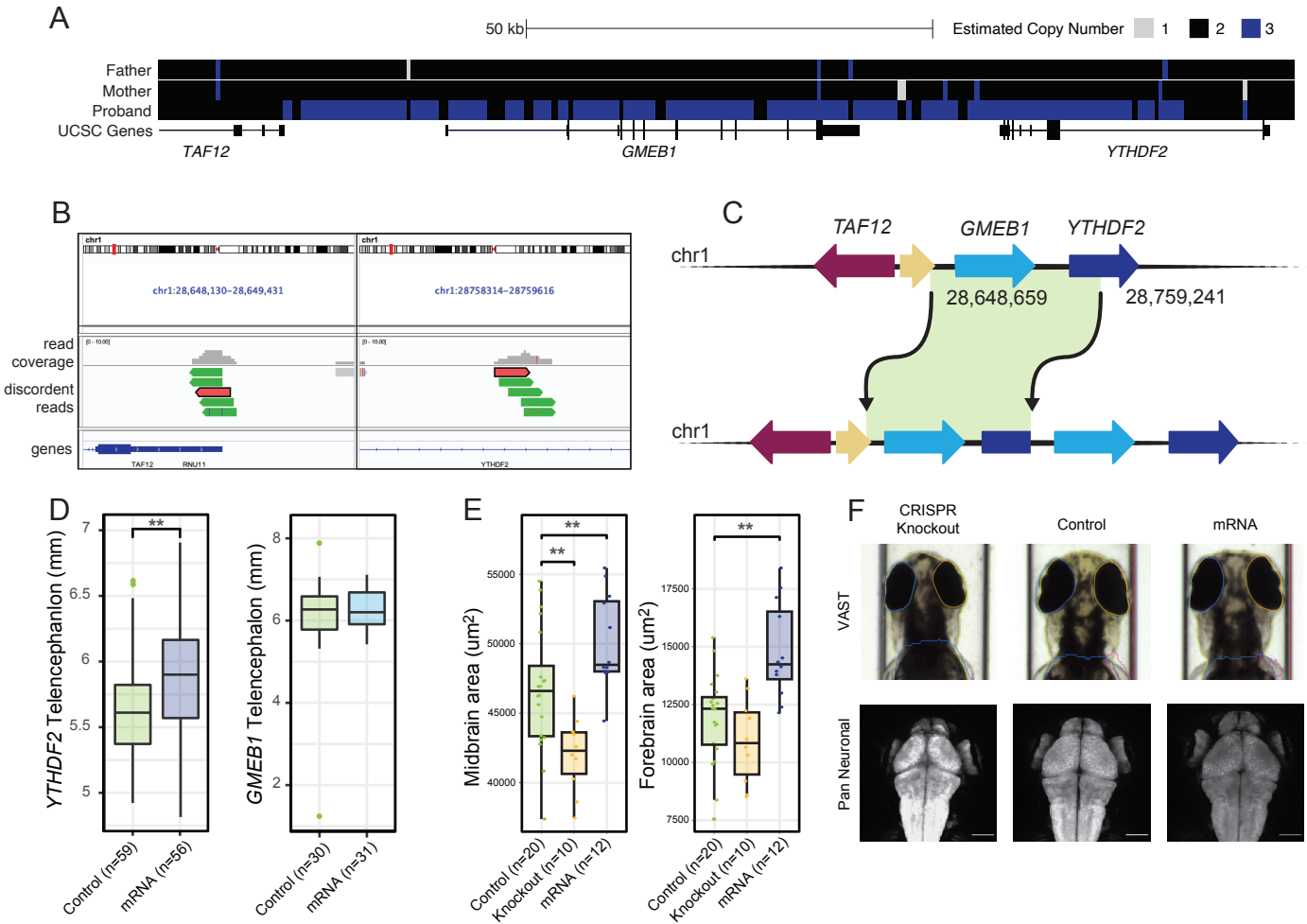
128. Teixidó, E., Kießling, T.R., Krupp, E., Quevedo, C., Muriana, A., and Scholz, S. (2019). Automated morphological feature assessment for zebrafish embryo developmental toxicity screens. Toxicol. Sci. *167*, 438–449.

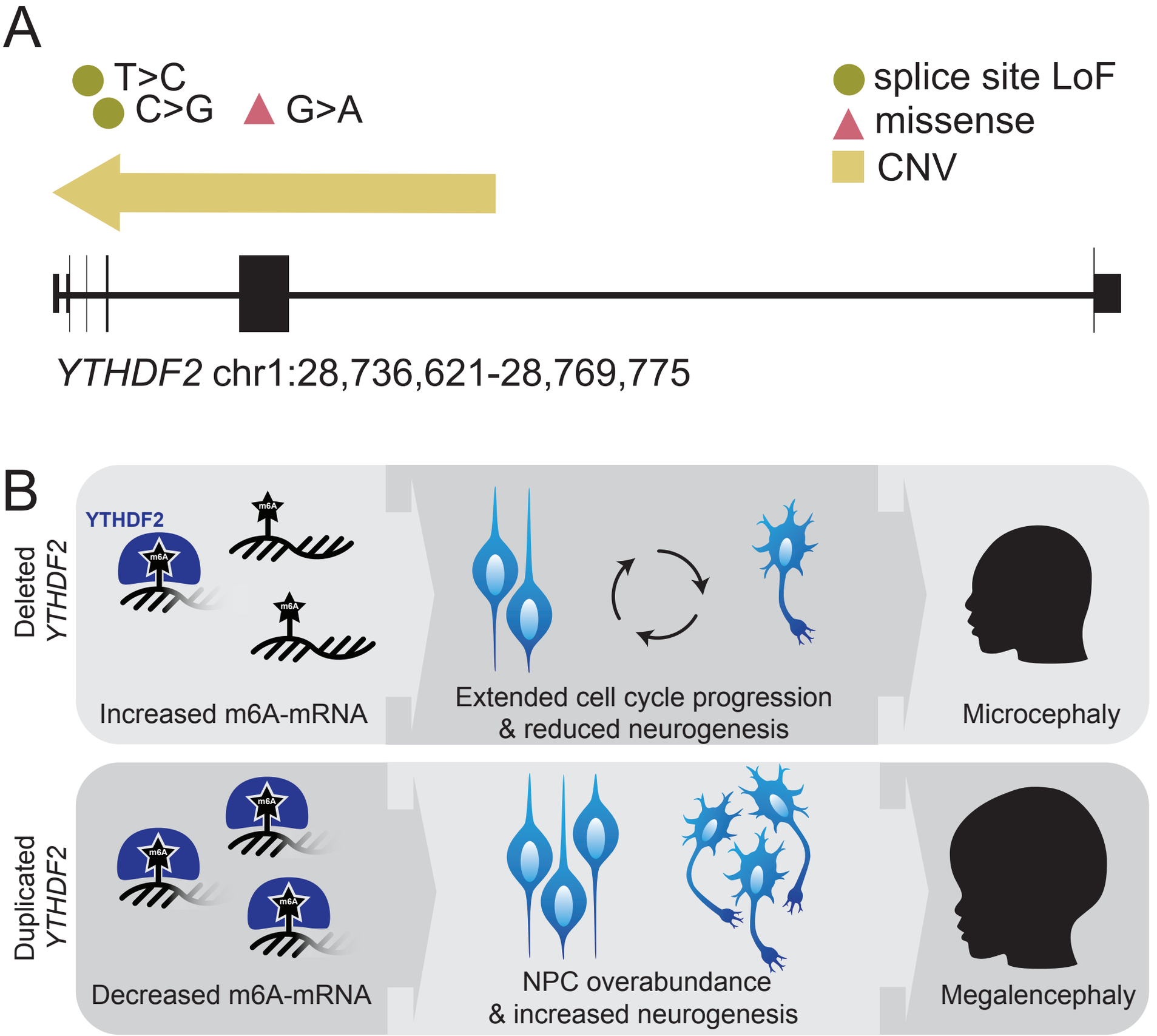
129. Ahlmann-Eltze, C., and Patil, I. ggsignif: R Package for Displaying Significance Brackets for "ggplot2."


A
SSC ASD Cases
600
400
200
0
0
25
50
75
100
Head Circumfrence Percentile
B
ASD-DM
ASD
DM
CHD8
SHANK3
PTEN
KMT2E
NF1
GIGYF1
WAC
SCN2A
KDM5B
KDM6B
CELF4
ARID1B
TCF4
DSCAM
CTCF
SHANK2
ANKRD11
TRIP12
SETD2
WDFY3
NFIB
RERE
YME1L1
LRP2
PSMD12
PDSS2
TNFRSF8
MX1
KMT5B
PSD3
COL25A1
SOBP
ATX7L2
CAST
HIVP3
ENOPH1
PPP4R2
CSDE1
KDR
MLANA
IRAG2
PNPLA7
MYBL2
IGF2R
FOXH1
BTBD9
RAB2A
PARD3B
LIPE
KATNAL2
ANO5
RUNDC1
NAPRT
DOCK1
FNIBP4
STARD9
HOXD1
TMEM39B
BRWD1
NCKAP5
C11orf24
CDAN1
PER2
BTBD11
ADAMTS9
FAM91A1
CCPG1
VARS
PCDLCE
ENG
CSNK2B
YTHDC1
GALNT18
SYNE2
NUAK1
TMC7
PLEKHO1
C9orf78
CPA4
NXPE4
TBC1D9B
ATXN7L3
NT5DC4
OSBPL8
PINK1
PNLIPRP2
RIMS1
USP29
RNF38
P2RX1
SAMD3
ANKRD60
ZSCAN30
MORC3
CRYBG3
KAT6A
PDCD1
ZNF865
ADCY5
AMBP
FBRS
MTHFS
IFI30
TXNRD1
GMEB1
YTHDF2
ATP1B1
TM4SF19
IFI44
PPP6R2
CHD10
FARP1
PROSER1
BIRC6
TMEM161B
CEACAM1
PLEKHM1
ZNF821
CNPY3
ABCA8
CPD
DNAH5
LTN1
SPTGN1
AEBP1
LMAN1
DRAM2
UGT1A4
LRGUK
SMARCD2
TYW5
CYP27C1
PTPN11
XPO4
PHF3
FLG
SOHLH1
DCC
NCKAP1
DPP4
RYR3
SORBS1
WDR54
RGS2
RABGGTA
LRFN2
MTA3
ZNF438
KRT84
PHIP
HNRNPL
PRSS38
ACAT2
92.5
95
97.5
100
Head Circumference Percentile
ASD-M Type
DM
RM
SO


Wu
APP
SCC
Cell Adhesion & Cytoskeletal
Histone Modification & Chromatin Organization
mRNA Binding
Transcription Factor
Signaling and Neuronal Functions
LIPE
YME1L1
ATP1B1
MLANA
CEACAM1
CAST
RYR3
NAPRT
TBC1D9B
RUNDC1
TNFRSF8
CPD
KCNMA1
TXNRD1
VARS
DPP4
IGF2R
RAB11B
PPP4R2
GALNT18
PINK1
USP29
LTN1
RAB2A
PDCD1
KDM6B
WDFY3
AMBP
ASXL1
HUWE1
PPP6R2
ENG
PTPN11
KAT6A
AEBP1
PTEN
CTCF
NSD1
CDAN1
LMAN1
SIN3A
TRIP12
ZEB2
KMT2E
PSD3
PLEKHM1
TCF4
CSNK2B
CHD8
SUV420H1
SETD2
PLEKHO1
NFIB
SCN2A
KDM5B
WAC
STARD9
KDR
NF1
DCC
HNRNPL
RIMS1
ARID1B
YTHDC1
SHANK2
PSMD12
DSCAM
PARD3B
CSDE1
YTHDF2
WDR26
MYT1L
MXI1
UBN2
KATNAL2
PHF3
SYNE2
ADAMTS9
HOMER2
SPTBN1
PER2
ANKRD11
BIRC6
BRWD1
SYNE1
DOCK1
SMARCD2
ADCY5
LRGUK
COL25A1
FARP1
PHIP
LRMP
RGS2
NCKAP5
NCKAP1
ANO5
CNPY3
IFI30
DRAM2
BTBD11
ATXN7L2
XPO4
TNPO3
IFI44
PDSS2
NUAK1
ATXN7L3


VAST
Zebra Box
chd8
n=42,19
fam91a1
n=46,45
gmeb1
n=36,44
iars2
n=56,27
rps6ka5
n=51,11
ryr3
n=54,35
ythdf2
n=42,46
hepacam2
n=35,42
pax5
n=37,39
scp2a
n=46,33
Total Area
Body Length
Distance Between Center of Eyes
Distance Moved
Duration Moved
Distance over Duration
High Speed Events
Wilcoxon T-test p-value
<0.01
0.01-0.05
>0.05
Fold Change
1.5
1
0.5


A
50 kb
Estimated Copy Number
1
2
3
Father
Mother
Proband
UCSC Genes
TAF12
GMEB1
YTHDF2
B
chr1:28,648,130–28,649,431
chr1:28758314–28759616
read coverage
discordent reads
genes
TAF12
RNU11
YTHDF2
C
chr1
TAF12
GMEB1
YTHDF2
28,648,659
28,759,241
chr1
D
YTHDF2 Telencephanlon (mm)
GMEB1 Telencephalon (mm)
Control (n=59)
mRNA (n=56)
Control (n=30)
mRNA (n=31)
E
Midbrain area (um2)
Forebrain area (um2)
Control (n=20)
Knockout (n=10)
mRNA (n=12)
F
CRISPR Knockout
Control
mRNA
VAST
Pan Neuronal


A
T>C
C>G
G>A
splice site LoF
missense
CNV
YTHDF2 chr1:28,736,621-28,769,775
B
Deleted YTHDF2
YTHDF2
m6A
Increased m6A-mRNA
Extended cell cycle progression
& reduced neurogenesis
Microcephaly
Duplicated YTHDF2
Decreased m6A-mRNA
NPC overabundance
& increased neurogenesis
Megalencephaly