

Article

Not peer-reviewed version

Structure-Function of the Human WAC Protein in GABAergic Neurons: Towards an Understanding of Autosomal Dominant DeSanto-Shinawi Syndrome

[Hannah C Rudolph](#)[^], [April M Stafford](#)[^], Hye-Eun Hwang, [Cheol-Hee Kim](#), [Jeremy W Prokop](#)^{*}, [Daniel Vogt](#)^{*}

Posted Date: 22 February 2023

doi: 10.20944/preprints202302.0369.v1

Keywords: cell biology; protein sorting; nuclear translocation; protein domain; WAC



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Structure-Function of the Human WAC Protein in GABAergic Neurons: Towards an Understanding of Autosomal Dominant DeSanto-Shinawi Syndrome

Hannah C Rudolph ^{1,†}, April M Stafford ^{1,†}, Hye-Eun Hwang ², Cheol-Hee Kim ²,
Jeremy W Prokop ^{1,3,*} and Daniel Vogt ^{1,4,*}

¹ Department of Pediatrics and Human Development, Michigan State University, Grand Rapids, MI 49503, USA

² Department of Biology, Chungnam National University, Daejeon 34134, Korea

³ Office of Research, Corewell Health, Grand Rapids, MI 49503, USA

⁴ Neuroscience Program, Michigan State University, East Lansing, MI 48824, USA

* Correspondence: jprokop54@gmail.com (J.W.P.); vogtdan2@msu.edu (D.V.); Tel.: 616-234-2641

† These authors contributed equally.

Simple Summary: There are several rare disrupted genes that underlie neurological dysfunction. Many have not been characterized for their role in brain function or how they work in individual brain cells. The WW domain-containing adaptor with coiled-coil, WAC, gene is one example. Dysfunction of this gene underlies a rare syndrome, DeSanto-Shinawi syndrome (DESSH), with those diagnosed having symptoms including cranio-facial changes, autism and attention deficit hyperactivity disorder. We sought to understand how the WAC protein functioned in brain cells implicated in DESSH in two ways. First, we used available technologies to predict important conserved regions in the protein that may underlie cellular function, including how it localized to distinct areas of a cell and further correlate these with reported human genetic variants in these regions. These efforts uncovered novel regions in the protein necessary and sufficient for it to localize to the nucleus. Second, we deleted/used key regions of the WAC protein to test if they were necessary/sufficient to localize WAC to distinct cell regions in brain neurons, and found that the amino-terminus of the protein fulfilled this duty. Moreover, other regions contribute to distinct biological functions of WAC and this study first highlights these aspects of this unique neurodevelopmental protein.

Abstract: Dysfunction of the WW domain-containing adaptor with coiled-coil, WAC, gene underlies a rare autosomal dominant disorder, DeSanto-Shinawi syndrome (DESSH). DESSH is associated with facial dysmorphism, hypotonia, and cognitive alterations, including attention deficit hyperactivity disorder and autism. How the WAC protein localizes and functions in neural cells is critical to understanding its role during development. To understand the genotype-phenotype role of WAC, we developed a knowledgebase of WAC expression, evolution, human genomics, and structural/motif analysis combined with human protein domain deletions to assess how conserved domains guide cellular distribution. Then assessed in a cell type implicated in DESSH, cortical GABAergic neurons. WAC contains conserved charged amino acids, phosphorylation signals, and enriched nuclear motifs, suggesting a role in cellular signaling and gene transcription. Human DESSH variants are found within these regions. We also discovered and tested a nuclear localization domain that impacts the cellular distribution of the protein. These data provide new insights into the potential roles of this critical developmental gene, establishing a platform to assess further translational studies, including the screening of missense genetic variants in WAC. Moreover, these studies are essential for understanding the role of human WAC variants in more diverse neurological phenotypes, including autism spectrum disorder.

Keywords: cell biology; protein sorting; nuclear translocation; protein domain; WAC

1. Introduction

Rare disease syndrome genes have increasingly been identified as an underlying cause of neuropsychiatric and neurological symptoms, with over one thousand genes suggested to contribute

to neurodevelopment [1]. The protein products of these genes work together at the synapse, during cellular signaling, or have a role in epigenetic and transcriptional regulation [2]. However, some gene products have not yet been functionally assessed to determine cellular roles, so understanding how they impact neurodevelopment is absent, preventing potential therapeutic interventions. Having approaches to characterize these genes and their protein products is critical for advancing sequencing into clinical utility [3]. Here we provide an example of developing screening tools for an underexplored neurodevelopment gene, elucidating novel molecular functions that could impact neural cell types.

The WW domain containing adaptor with coiled-coil (WAC) gene is associated with DeSanto-Shinawi syndrome (DESSH). Those diagnosed with DESSH have symptoms of cranial dysmorphism and hypotonia with comorbidities including attention deficit hyperactivity disorder (ADHD), autism, and seizure susceptibility [4–8]. Moreover, WAC variants rank it as a high-confidence autism spectrum disorder risk gene [9,10]. The WAC protein has a few characteristic conserved domains, yet most of the protein is structurally disordered and not yet characterized. Notably, the WAC protein is often enriched in the nucleus of cells [11,12], but no nuclear localization signal (NLS) has been defined in the protein due to no sequences matching classical NLS signals. While implicated in cellular signaling and nuclear regulation [12–15], it is unclear how regions of the WAC protein modulate cellular distribution and how human variants could alter these functions to impact critical brain cell types. Therefore, we applied a computational screen combined with cell assays to define further the domains and disorder motifs of the WAC protein, serving as an example of characterizing underexplored neurodevelopment genes.

To ascertain the structure-function of WAC in a model cell type, we examined the cellular distribution of the protein in cells recently implicated in WAC phenotypes, cortical GABAergic interneurons [16], with the rationale that cell types sensitive to WAC mutations would have the most translatable value. Thus, we chose to utilize mouse medial ganglionic eminence (MGE) cells, which give rise to the majority of GABAergic cortical interneurons [17], as a model system to examine the impact of WAC protein domain mutants on cellular distribution. This approach offers a unique ability to screen the functional regions of a protein in a relevant neural cell type, a critical need for the field of neurodevelopmental genetics. Our analyses uncovered a unique bipartite nuclear localization signal in the amino terminus with an additional C-terminal nuclear export signal. Notably, these regions are enriched for human clinical variants. This suggests that nuclear localization is critical for proper WAC function, and its dysfunction results in DESSH. Further, a high number of conserved motifs within the disordered regions of the protein, where clinical variants fall, suggest that posttranslational modifications and cellular signaling are influential aspects of WAC function.

2. Materials and Methods

2.1. Bioinformatics

The bioinformatics of WAC was assessed using a combination of previously published tools [18,19]. In short, sequences of vertebrate WAC protein were extracted from NCBI Gene Ortholog [20] during Oct 2022 and aligned with NCBI Cobalt [21]. UniProt annotations [22], gnomAD version 2 variants [23], ClinVar variants [24], and Geno2MP (<http://geno2mp.gs.washington.edu>) variants were extracted during Nov 2022. WAC protein structure was from AlphaFold prediction [25]. The human WAC protein sequence was assessed for linear motifs through the Eukaryotic Linear Motif (ELM) server [26]. WAC gene expression was annotated from GTEx human broad tissues, the Allen Brain Atlas BrainSpan [27] of different ages vs. regions of human brain, and the Allen Brain Atlas Human MTG 10X (SEA-AD) single cell dataset [28].

2.2. Animals

A colony of WT CD-1 mice were bred to generate the embryos used for primary cultures. Experimenters were blind to the parameters. All mouse procedures were performed in accordance

with NIH Guidelines for the Care and Use of Laboratory Animals and were approved by the Michigan State University Institutional Animal Care and Use Committee.

2.3. DNA vector generation

The CMV-GFP-hWAC DNA vector was generated by amplifying human WAC cDNA from a DNASU plasmid (hsCD00442491) and cloning in frame to the 3' end of GFP in a previously described GFP expressing plasmid [29]; WAC was cloned into 5' BsrGI and 3' BamHI restriction enzymes sites. We subsequently edited a 3' deletion of two adenosines in the WAC coding domain to produce the correct full-length clone. The following primers were used to generate the full length and mutants: WAC forward 5'-GAGATGTACAAGATGGTAATGTATGCGAGG-3', WAC reverse 5'-GAGAGGATCCTCACACCATGAAGGAATTC-3'. Forward primers contained a BsrGI restriction site and reverse primers a BamHI restriction site (underlined); numbers denote amino acids of WAC.

W90 Forward 5'-GAGATGTACATTGATGGTGGGACCAGTTAC-3'

W166 Forward 5'-GAGATGTACATTGAACAGAGACAAAAAGAAGC-3'

W576 Forward 5'-GAGATGTACATTGATCATGCAGAGAAGCAG-3'

W89 Reverse 5'-GAGAGGATCCTCACCTCTCTCTAACTCTGTG-3'

W165 Reverse 5'-GAGAGGATCCTCATCTTTCAAGCCACTCTTTTGG-3'

W575 Reverse 5'-GAGAGGATCCTCATGCAGGCCATCCTTGAAC-3'

The WW domain deletion construct was generated by obtaining a synthesized gene block from integrated DNA technologies (IDT) of a partial WAC coding domain that lacked the WW DNA sequence (encoding amino acids 133-165) with flanking 5' BsrGI and internal 3' NsiI restriction sites that were used to ligate in the GFP-WAC vector described above.

2.4. Immuno-fluorescence labeling and imaging

Primary neurons were washed in phospho-buffered saline containing 0.3% Triton-X100, blocked in the same solution containing 5% bovine serum albumin and then incubated in primary antibodies for 1-2 hours. They were washed 3 times and then incubated with secondary antibodies containing fluorophores for 1 hour before 3 final washes. A Leica DM2000 microscope (DFC3000G camera) captured primary cell images for quantification and Zeiss LSM800 confocal microscope used to attain representative images for figures.

2.5. MGE primary cultures

We generated MGE primary cultures as described in [30,31]. Briefly, we cultured MGE cells in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum and penicillin/streptomycin from time of seeding until one day *in vitro*. The cells were then transfected with the GFP-fusion vectors using Lipofectamine™ 2000 (ThermoFisher 11668027) on day one and after four hours, replaced with neurobasal media, supplemented with B27, glucose, glutamax and penicillin/streptomycin. Cells grew in this media for five days *in vitro*, then fixed in 4% paraformaldehyde and assessed via immuno-fluorescence labeling.

2.6. Statistics and cell assessments

Graphpad Prism 7 was used to calculate statistical significance; a p value of <0.05 was considered significant. For non-parametric data sets (proportions), we used a Chi-squared test to determine significance.

2.7. Western blotting

DNA plasmids were transfected into HEK293T cells using Lipofectamine™ 2000. After 48 hours, cells were collected and lysed in standard RIPA buffer with protease and phosphatase inhibitors and combined with Laemmli buffer (BioRad 1610737EDU) containing 2-Mercaptoethanol and incubated at 95°C for 5 minutes. Equal amounts of protein lysates were separated on 10% SDS-PAGE gels and then transferred to nitrocellulose membranes. The membranes were washed in Tris-buffered saline

containing Triton X-100 (TBST) and blocked for 1 hour in TBST containing 5% non-fat dry milk (blotto, sc-2324 SantaCruz biotechnology). Membranes were incubated with primary antibodies overnight at 4°C, washed 3 times with TBST, incubated with secondary antibodies for 1 hour at room temperature and then washed 3 more times with TBST. Membranes were incubated in ECL solution (BioRad Clarity substrate 1705061) for 5 minutes and chemiluminescent images obtained with a BioRad Chemidoc™ MP imaging system. Antibodies (all used at 1:4,000 dilution): rabbit anti-GFP (ThermoFisher A6455), rabbit anti-WAC (abCam ab109486), rabbit anti-GAPDH (Cell Signaling Technology 2118), as well as goat anti-rabbit HRP (BioRad 170-6515).

3. Results

3.1. Functional domain and motif annotations of WAC

The WAC protein contains an internal WW domain, known for protein-protein interactions with proline rich regions [32]. At the carboxy-terminus, the WAC protein has two coiled-coil domains, which can have many functions, including dimerizing with similar structures [33]. Between these regions of the WAC protein is highly disorder protein annotations with no known functions. Using 364 vertebrate WAC protein sequences, we identified regions with high conservation (Figure 1). While UniProt only annotates the WW domain and Coiled-Coil region of WAC with the majority predicted to be disordered, there is a surprising high level of conserved polar basic (blue, top panel Figure 1A), polar acidic (red, top panel Figure 1A), and Serine/Threonine (orange, top panel Figure 1A) conservation. Many of these conserved sites are found outside of the WW and coiled-coil annotations, suggesting an under annotated role of disordered motifs in WAC. Many of the conserved amino acids fall outside the annotated domains/motifs within unstructured regions of WAC (Figure 1B).

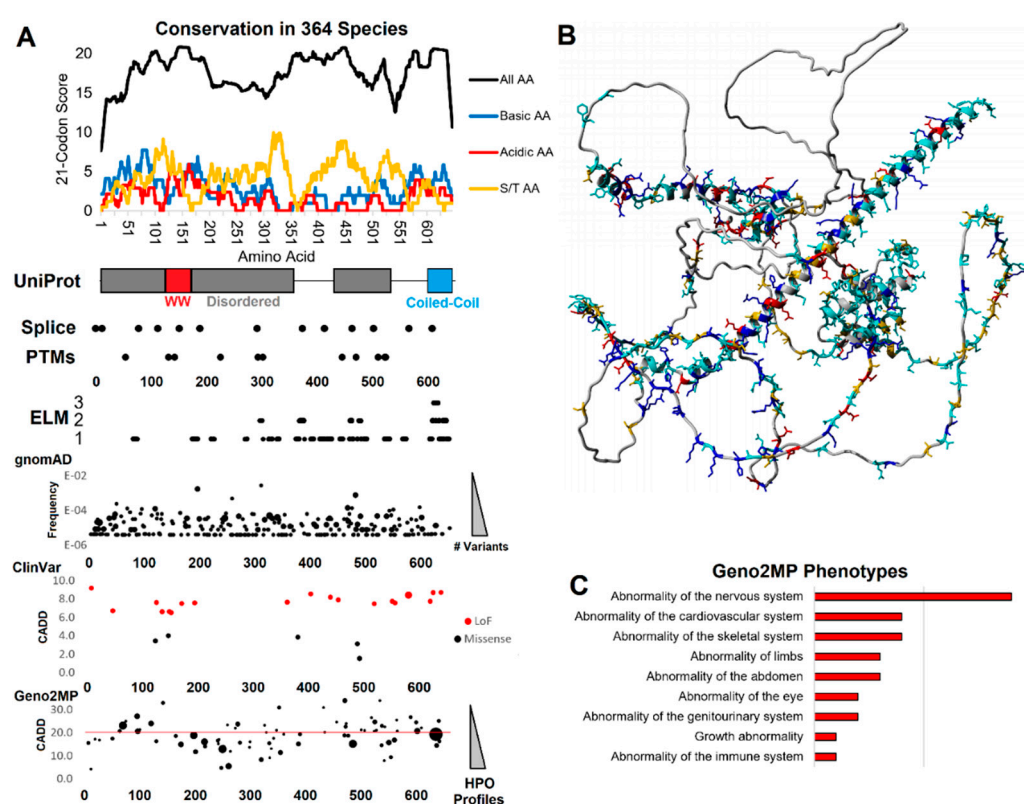


Figure 1. Genomic analysis of WAC. A) Compiled genomics insights for WAC. On top is the conservation of 364 species of WAC protein. The conservation score of all amino acids (black) is based on the amino acid % of species conserved at each amino acid (0- no conservation, 1- 100% conserved), placed on a 21 codon sliding window such that each site is added to ten upstream and ten downstream. The basic (blue, R/K/H), acidic (red, D/E), or S/T (orange) amino acid scores are

calculated using just these amino acids for the same 21 codon window. Below conservation is the amino acid locations (x-axis) for UniProt annotated domains/motifs, splice sites, posttranslational modifications (PTMs), ELM annotated motifs (y-axis lists the number of annotations at each amino acid), gnomAD (y-axis list allele frequency, bubble size is the number of unique variants), ClinVar variants (y-axis is the CADD functional score, color represents the type of variant), and Geno2MP variants (y-axis is the CADD score, bubble size the number of human phenotype profiles). B) An AlphaFold protein model of WAC showing conserved amino acids colored as blue- polar basic, red- polar acidic, orange- S/T, cyan- all other conserved amino acids. C) Geno2MP listed phenotypes for all variants with a CADD score over 20.

According to the Human Protein Atlas (<https://www.proteinatlas.org/ENSG00000095787-WAC/subcellular>), WAC is localized within the nucleus of cells but lacks an annotated nuclear localization domain. We screened the WAC sequence with multiple NLS bioinformatic screening tools, without a single tool returning a potential NLS. Surprisingly, a large portion of the linear motif predictions of WAC are also identified to be within the nuclear cell compartment, suggesting sequence convergence to nuclear function. These nuclear motifs include conserved sites for [DOC MAPK gen 1](#) (169-178, 606-614), [LIG FHA 2](#) (307-313, 618-624), [DEG SPOP SBC 1](#) (322-326), [DOC CYCLIN yCln5 NLxxxL 5](#) (357-366), [DEG APCC DBOX 1](#) (363-371), [DOC CYCLIN yCln2 LP 2](#) (366-372), [DOC PP2A B56 1](#) (383-388, 611-617), [LIG PCNA PIPBox 1](#) (400-409), [LIG EH1 1](#) (413-421), [LIG CtBP PxDSL 1](#) (437-441), [LIG 14-3-3 CanoR 1](#) (454-458, 517-526), [DOC CKS1 1](#) (455-460, 469-474), [DEG SCF FBW7 1](#) (468-475), [DOC CKS1 1](#) (469-474), [DOC PP1 RVXF 1](#) (623-629), [LIG UBA3 1](#) (626-634). It should be noted that several of these regulation motifs are linked to similar neurodevelopmental disorders including MAPK dysfunction and the [DEG_SCF_FBW7_1](#) observed modified in multiple individuals with MED13 related neurodevelopmental disorder [34]. In addition, [DEG_SPOP_SBC_1](#) suggests the protein localizes to nuclear puncta via the Cul3-RING ubiquitin ligase complex and multiple motifs suggest the protein has cell cycle and DNA damage impacts through destruction motifs and phospho signaling.

Polar basic amino acids are often critical for nuclear localization, where WAC has the highest density of conserved polar basic residues from amino acids 77-85. Polar basic amino acids 57, 60, 61, 68, 75, 79, 82, 84, 85, 87, and 89 are conserved in greater than 90% of sequences analyzed (DPSPPNKMLRRSDSPENKYSSTGHSKAKNVH~~THR~~VRERDGGTSYSPQENSHNHSALHSS). Near this region is an ELM annotated [MOD SUMO rev 2](#) site (amino acids 62-70), a SUMOylation modification site found within nuclear proteins. Additionally, a [TRG NES CRM1 1](#) nuclear export signal (NES) is found within amino acids 604-616, which is flanked by a [MOD SUMO rev 2](#) site (amino acids 599-608). This suggests that nuclear localization is potentially regulated by both the N and C-terminal regions through nonclassical motifs.

3.2. Human clinical variants annotated to functional motifs of WAC

There are multiple rare disease variants that fall within the newly identified motifs of WAC (Figure 1A). Within ClinVar, there are loss-of-function variants (LoF), suggesting that truncation of the protein results in disease. Few ClinVar missense variants have been identified to date. Yet, a large number of clinical phenotypes for rare disease patients can be observed for variants within the Geno2MP database. There is an enrichment for nervous system abnormalities (Figure 1C) of clinical variants with annotated phenotypes from Geno2MP predicted to have high pathogenic CADD scores (>20). We further analyzed three regions with clinical variants within functional motifs (Figure 2).

Within and adjacent to the possible polar basic N-terminal region of WAC that may regulate nuclear localization, five variants are found at conserved sites in Geno2MP. Two variants, S94I and Y95C, are observed multiple times in affected individuals with nervous system phenotypes. This site is a potential phosphorylation switch. The S94I variant is found homozygous in three individuals, with spastic paraplegia, abnormality of brain morphology, and intellectual disability. One individual heterozygous for the variant exhibited seizures. The Y95C variant is heterozygous in one affected individual with encephalocele. It should be noted that none of the polar basic amino acids observed in Geno2MP individuals varied.

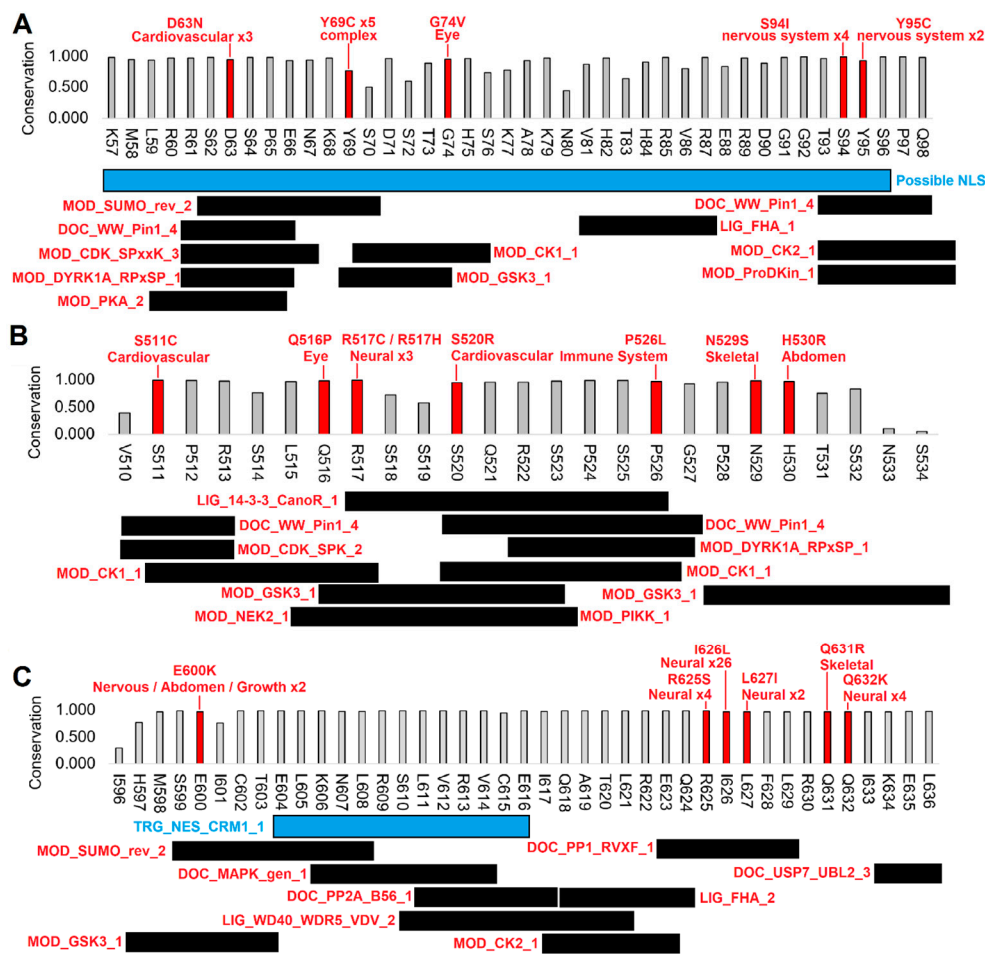


Figure 2. Highly conserved motifs within WAC containing human variants connected to medical phenotypes. Shown for the N-terminal nuclear localization motif (A), central 14-3-3 interaction motif (B), and the C-terminal nuclear export motif region of the coiled-coil domain (C). The conservation is based on the 364 amino acid sequences of vertebrate WAC (0- no conservation, 1- 100% conserved). Those amino acids in red have a known Geno2MP variant with listed phenotypes. Those with multiple individuals with overlapping phenotypes are represented with an x number of individuals. Below each region are ELM-predicted motifs.

A second conserved motif of WAC (510-534) contains multiple phosphorylation sites and a 14-3-3 interaction motif, a site that requires phosphorylation for interaction to occur. The critical amino acid R517 for 14-3-3 interaction has two variants within Geno2MP connected to neurological phenotypes. The R517C variant is found heterozygous in two affected individuals with epileptic encephalopathy and R517H is found heterozygous in an individual with neurodevelopmental abnormality. 14-3-3 signaling has been suggested to have neuroprotective function with elevation in seizure model systems [35].

The most surprising motif with functional variants connected to neurological phenotypes in Geno2MP is the C-terminal region with the nuclear export signal (WAC 596-636). This region, in addition to the nuclear export signal, has a SUMOylation motif, several phosphorylation sites, and multiple protein interaction sites. None of the Geno2MP variants are found in the nuclear export signal, rather, they flank this site. E600K is found within the SUMOylation motif and in two individuals with nervous system abnormalities. The other variants fall within a DOC_PP1_RVXF_1 motif, which is a dephosphorylation site. R625S is found homozygous in three affected individuals, one with seizures and one with abnormality of the nervous system. Four affected individuals are heterozygous for R625S with two having abnormality of brain morphology and one with intellectual disability. I626L is the most common homozygous Geno2MP variant, with 16 affected individuals annotated with abnormal brain morphology, four with spastic paraplegia, and three with intellectual

disability. The variant is absent in gnomAD and TOPmed databases, suggesting a significant enrichment in neurological disease. L627I was found heterozygous in two affected individuals with spastic paraplegia and Q632K was found heterozygous in four affected individuals with abnormality of brain morphology. Overall, this suggests that N and C-terminal regions with nuclear localization and export signals are critical for neurological development and associated with human variants of neurological phenotypes.

3.3. Expression of WAC within human brain and neural cell types

To identify cell types that WAC may impact, we performed an analysis of expression datasets (Figure 3). In the GTEx database of broad human tissue expression, several protein isoforms are observed to be expressed for WAC (Figure 3A). The longest isoform, codes for a 647 amino acid protein, with highest expression in all tissues and enriched within the brain. Smaller isoforms with spliced exons (602 and 544 amino acids) are also highly expressed in the brain and have all three of the annotated domains (Figure 3B). The BrainSpan human developmental brain atlas shows that WAC expression is relatively the same throughout post coital weeks (pcw), with higher variability later in life (Figure 3C). Multiple regions of the brain also show high expression of WAC (Figure 3D). From human single cell brain expression, both GABAergic and glutamatergic neurons express WAC, while most non-neuronal cell types have significantly lower expression (Figure 3E).

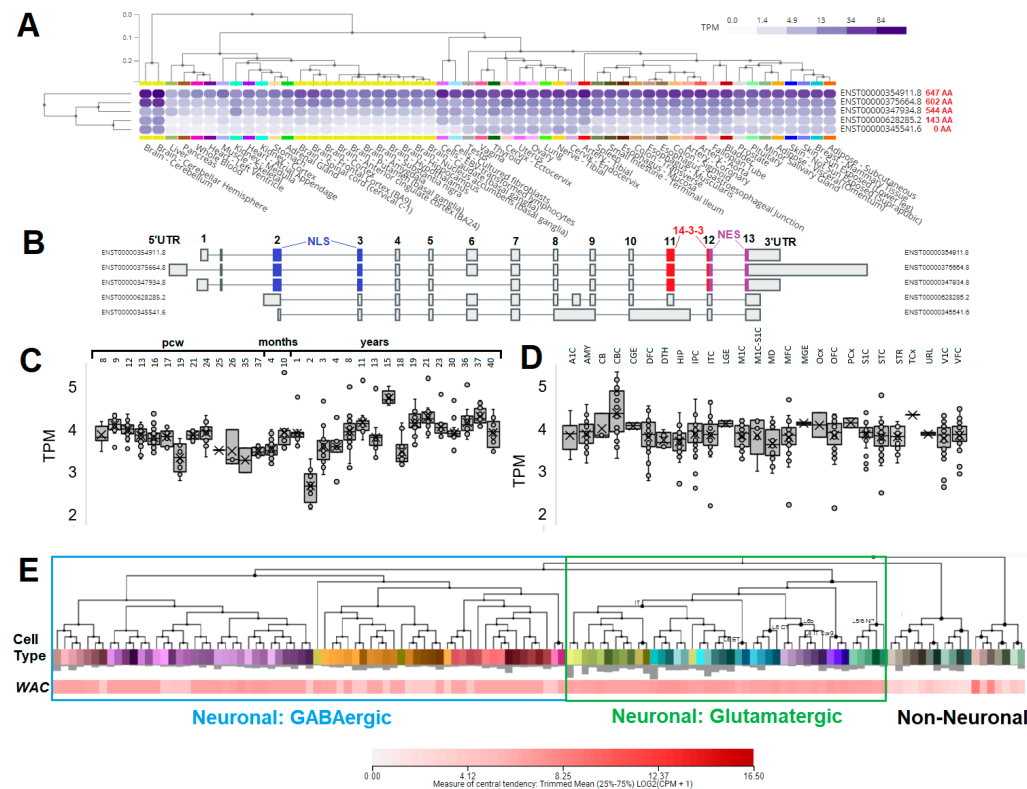


Figure 3. Analysis of WAC expression in various tissues, the brain developmental times, and within single-cell types of human brain. (A) GTEx expression of WAC isoforms in various human tissues. The darker blue color represents the higher expression. Tissues are ranked in expression. In red next to each isoform number is the amino acid size of the protein. (B) Annotated exons of each isoform as shown in panel A. In blue are exons that code for the NLS, red the 14-3-3 interaction site, and magenta the NES. (C-D) The developmental transcriptome data of BrainSpan shown as box and whisker plots of age (C) or tissue type (D). Abbreviations include: pcw- post-coital weeks, A1C- primary auditory cortex (core), AMY- amygdaloid complex, CB- cerebellum, CBC- cerebellar cortex, CGE- caudal ganglionic eminence, DFC- dorsolateral prefrontal cortex, DTH- dorsal thalamus, HIP- hippocampus (hippocampal formation), IPC- posteroventral (inferior) parietal cortex, ITC- inferolateral temporal cortex (area TEv, area 20), LGE- lateral ganglionic eminence, M1C- primary motor cortex (area M1,

area 4), M1C-S1C- primary motor-sensory cortex (samples), MD- mediodorsal nucleus of thalamus, MFC- anterior (rostral) cingulate (medial prefrontal) cortex, MGE- medial ganglionic eminence, Ocx- occipital neocortex, OFC- orbital frontal cortex, PCx- parietal neocortex, S1C- primary somatosensory cortex (area S1, areas 3,1,2), STC- posterior (caudal) superior temporal cortex (area 22c), STR- striatum, TCx- temporal neocortex, URL- upper (rostral) rhombic lip, V1C- primary visual cortex (striate cortex, area V1/17), VFC- ventrolateral prefrontal cortex. (E) Single-cell expression from the Allen Brain Atlas Human MTG – 10X (SEA-AD) data. Deeper red represents higher expression of WAC. Clustering shows grouped cells within labeled neuronal groups.

3.4. Generation and structure/function of GFP-WAC mutant fusion proteins

To assess the contribution of WAC protein domains to cellular localization and subcellular distribution, we generated a series of mutant WAC-expressing plasmids that included an amino-terminal GFP fusion; amino-terminal fusion of a different epitope was used to study WAC protein localization [11], suggesting this would be a feasible approach. The plasmids included protein domain deletions to assess whether specific regions were necessary and/or sufficient for nuclear localization and protein domain fusions to test sufficiency to localize in neurons (Schema of plasmids, Fig 4A).

We first assessed whether these GFP-WAC expression plasmids generated protein products of the predicted molecular weight via western blot. These assays assessed the GFP fusion proteins and whether the same mutant proteins were recognized by a functional commercial antibody raised against WAC. To this end, we expressed the plasmids in HEK293T cells and collected protein lysates at 48 hours for western blot analyses (Figure 4B). Notably, each of the expressed proteins exhibited GFP+ bands via western blot at the expected sizes (Figure 4C, upper panel), suggesting that these fusion proteins could be properly expressed, and the deletion mutants were stable enough to be detected via this method.

We also used a commercial (abCam ab109486) antibody raised against the WAC protein on the deletion mutants to detect where the epitope may reside in WAC (Figure 4C, lower panel). This antibody only detected protein when amino acids 166-575 were present, suggesting the antibody epitope resides within this region. The faint bands found in all lanes are likely the endogenous WAC protein isoforms expressed by the HEK293T cells. Overall, these data demonstrated that we could express the various WAC mutant proteins and determine the epitope region used to develop a reliable commercial WAC antibody.

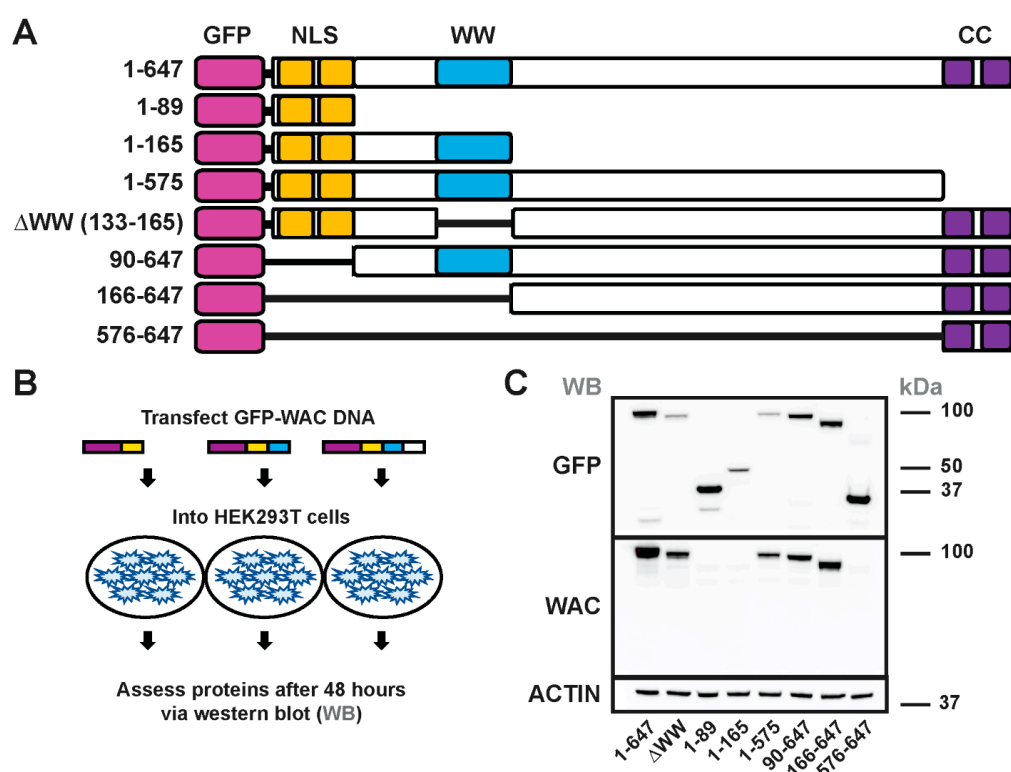


Figure 4. Assessment of GFP-WAC deletion mutant proteins. Schema showing the various GFP-WAC vectors generated to assess conserved protein domains (A). (B) GFP-WAC vectors were transfected into HEK293T cells to express proteins. (C) Western blots demonstrating expression of the various GFP-WAC proteins after 48 hours. Top panel probed with an anti-GFP antibody and bottom panel probed with a commercial anti-WAC antibody. (kDa) kilodaltons.

3.5. WAC deletion mutant proteins exhibit differential targeting and aggregation in neurons

We expressed the GFP-WAC full length gene and mutant versions in brain derived MGE cells. We collected and cultured embryonic day (E)13.5 MGE cells. The cells were transfected with the GFP fusions at 24 hours and allowed to develop for 5 days, a time when their morphology would allow us to decipher cytoplasmic vs. nuclear localization of the proteins. To this end, we found multiple phenotypes related to the necessity or sufficiency of these protein domains related to nuclear/cytoplasmic localization as well as distinct sequestration/puncta-like distribution in neurons expressing each mutant.

Consistent with a previous study [11], full length GFP-WAC preferentially localized to the nucleus in MGE cells (Figure 5A, 5A', 5I). To determine the role of distinct domains in cellular localization and distribution, we quantified whether various GFP-WAC deletion proteins localized in the nucleus and/or cytoplasm, as well as in discrete puncta within the neurons. Importantly, all mutant proteins that contained the N-terminal region localized preferentially to the nucleus (Figure 5B-5E, 5B'-5E', 5I), suggesting that the predicted NLS in this region may be functional. Notably, that region alone was sufficient to localize GFP to the nucleus (Figure 5B, 5B', 5I). In agreement with this observation, the proteins that lacked the N-terminal region were found in the cytoplasm (Figure 5F-5H, 5F'-5H', 5I), suggesting nuclear sequestration of WAC is dependent on the N-terminal basic amino acid-enriched region.

In addition to nuclear sequestration of WAC, we also noticed a puncta appearance of the protein when certain domains were present. For example, the full-length protein and WW deletion mutant both have a distinct punctate pattern in the nucleus (Figure 5A, 5A', 5E, 5E', 5J), suggesting the WW domain is dispensable for this phenotype. Consistent with the WW domain having no impact on the punctate distribution of WAC, neither the putative NLS nor the CC domains seemed to have any influence over this distribution. The region of the protein that seems important for this punctate

localization resides between amino acids 166-575 (Figure 5J), the region predicted earlier to contain several linear motifs with potential phospho targets. Thus, while we have determined the contribution of many of WAC's conserved domains on cellular localization, there is still more to be learned about this unique syndromic protein.

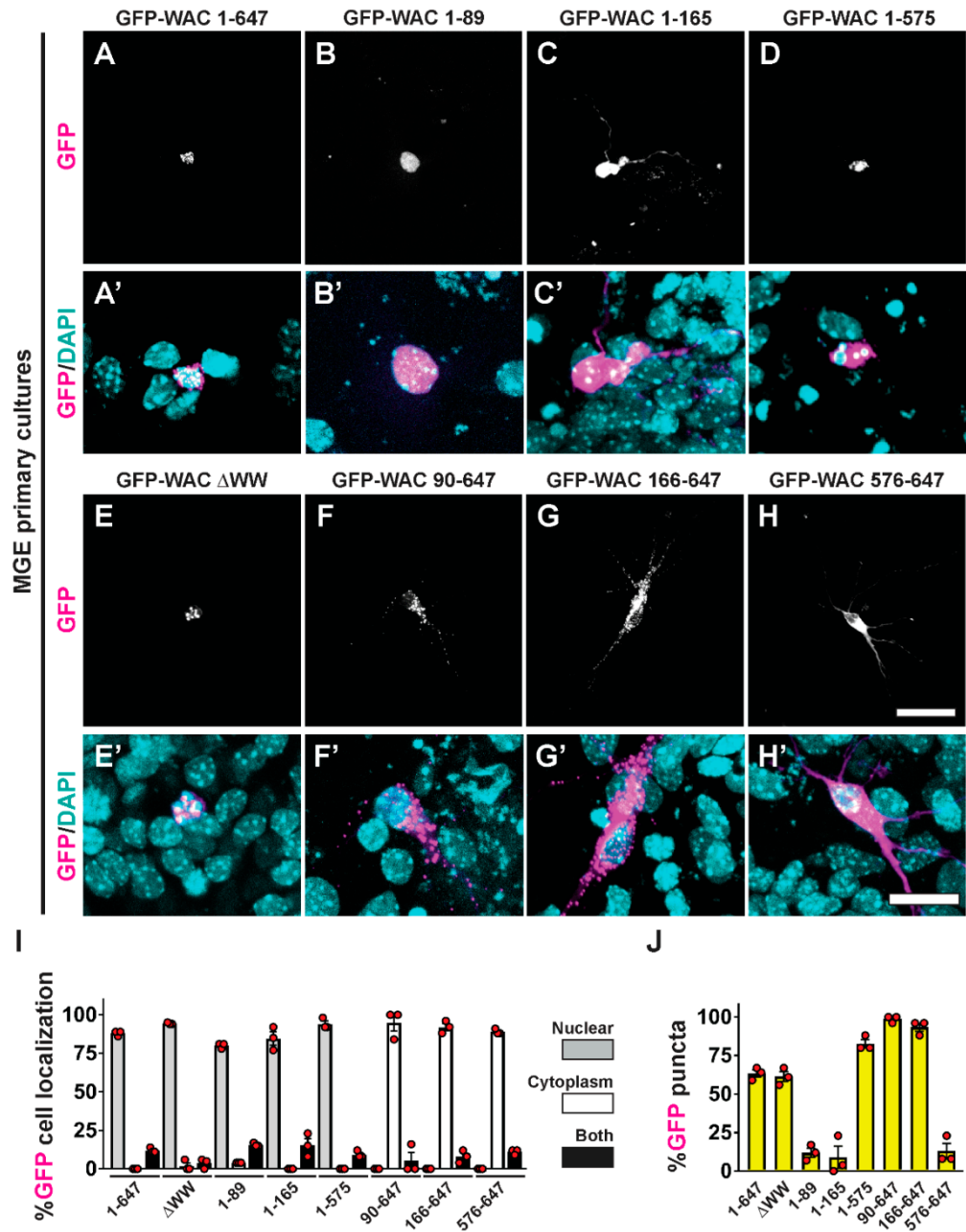


Figure 5. Distinct cellular distribution by WAC's conserved protein domains. MGE primary neurons transfected with GFP-WAC fusion proteins were assessed for GFP localization (A-H) and merged with DAPI after 5 days in vitro (A'-H'). (I) Quantification of the proportion of GFP labeled cells showing nuclear and/or cytoplasmic distribution. (J) Quantification of the proportion of GFP labeled cells with punctate localization. Scale bars in (H) = 40 μ m and (H') = 20 μ m.

4. Discussion

Dysfunction of the WAC gene underlies DESSH as well as several comorbid neurological symptoms, including autism spectrum disorder, ADHD and seizures [4,5,7,8]. This dysfunction likely arises from genetic variants within key conserved regions of the WAC gene and/or loss of function, some of these observations have elevated WAC to be considered a high-confidence autism risk gene

[9,10], conveying the importance of understanding how the gene, protein and downstream effectors are changed and potential inroads into therapeutics to remedy these alterations. Our structure-function approach begins to delve into how these regions of the WAC protein may be involved in these processes and provide a template for future work examining the role of genetic variants and other phenotypic outcomes that WAC underlies.

The WAC protein has several conserved domains and unexplored regions. The WW and CC domains have been reported in previous reports [11,14] but how WAC localizes to the nucleus and whether other regions may dictate functions have yet to be addressed. Herein, we found that the amino terminus of WAC was enriched for basic amino acids and this region was both necessary and sufficient for nuclear sequestration. Defining the region in WAC for nuclear localization will be crucial moving forward as human genetic variants in this region are assessed. In addition, the WAC protein has a punctate pattern in neurons, suggesting some type of localization to cellular compartments. This was observed in the nucleus and the cytoplasm of various deletion mutants, suggesting some part of the WAC protein can mediate this observation. While future studies are needed to assess how this punctate pattern manifests, our data suggest that amino acids 166-575 may be responsible for this cellular pattern. This region is highly conserved but lacks the conventional known conserved WW and CC domains. Instead, this region is mainly linear in nature and contains several amino acids that could act as phospho-switches. Interestingly, it adds another level of regulation to this important protein and suggests that other cellular events may be involved in WAC functions.

Our assessments herein focused on GABAergic neurons due to our previous data suggesting some of these neurons were susceptible to *Wac* loss of function [16]. We also performed an assessment of *Wac* expression in brain cell types and found a preferential enrichment for both GABAergic and glutamatergic neurons. Unfortunately, *Wac* deletion is embryonic lethal but conditional mutants may be able to uncover further functions of this gene in both GABAergic and glutamatergic neurons in future studies. The role of WAC's conserved regions reported here is a strong starting point in understanding various mechanisms by which this protein regulates each neuron subtype. We predict that while the localization of this protein may be similar between cell types, the functional partners may differ and thus phenotypes may be distinct within each brain cell type. To this end, the same WAC deletion mutants had similar localization patterns in HEK293T cells (HCR and DV, data not shown).

Wac is an important developmental gene with relatively few known functions. Herein, we developed the first assessment of this important protein's conserved regions in cellular localization. These studies will help to interpret the growing number of genetic variants in *Wac* associated with neurodevelopmental disorders, including autism [9,10]. In addition, we found unique roles for uncharacterized regions in the WAC protein that will be assessed in future studies, which may provide unique roles for how WAC operates and the cellular pathways that utilize WAC for various functions.

Author Contributions: Conceptualization, D.V. and J.W.P.; methodology, D.V. and J.W.P.; validation/analysis, H.C.R., A.M.S. and D.V.; resources, H.-E.H.; writing—original draft preparation, D.V. and J.W.P.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Spectrum Health-Michigan State University Alliance Corporation (to D.V.), Gerber Foundation and National Institutes of Health K01ES025435 (to J.W.P). C.-H.K.National Research Foundation of Korea 2018M3A9B8021980, 2020R1A5A8017671, and 2021R1A2C1008506 (to C.-H.K.).

Institutional Review Board Statement: The animal study protocol was approved by Michigan State University (protocol code PROTO202000212; date of approval July 16th, 2021).” for studies involving animals.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mitchell, K.J. The Genetics of Neurodevelopmental Disease. *Curr. Opin. Neurobiol.* **2011**, *21*, 197–203, doi:10.1016/j.conb.2010.08.009.
- Chang, J.; Gilman, S.R.; Chiang, A.H.; Sanders, S.J.; Vitkup, D. Genotype to Phenotype Relationships in Autism Spectrum Disorders. *Nat. Neurosci.* **2015**, *18*, 191–198, doi:10.1038/nn.3907.
- Bupp, C.P.; English, B.K.; Rajasekaran, S.; Prokop, J.W. Introduction to Personalized Medicine in Pediatrics. *Pediatr. Ann.* **2022**, *51*, e381–e386, doi:10.3928/19382359-20220803-03.
- DeSanto, C.; D'Aco, K.; Araujo, G.C.; Shannon, N.; DDD Study; Vernon, H.; Rahrig, A.; Monaghan, K.G.; Niu, Z.; Vitazka, P.; et al. WAC Loss-of-Function Mutations Cause a Recognisable Syndrome Characterised by Dysmorphic Features, Developmental Delay and Hypotonia and Recapitulate 10p11.23 Microdeletion Syndrome. *J. Med. Genet.* **2015**, *52*, 754–761, doi:10.1136/jmedgenet-2015-103069.
- Lugtenberg, D.; Reijnders, M.R.F.; Fenckova, M.; Bijlsma, E.K.; Bernier, R.; van Bon, B.W.M.; Smeets, E.; Vulto-van Silfhout, A.T.; Bosch, D.; Eichler, E.E.; et al. De Novo Loss-of-Function Mutations in WAC Cause a Recognizable Intellectual Disability Syndrome and Learning Deficits in Drosophila. *Eur. J. Hum. Genet. EJHG* **2016**, *24*, 1145–1153, doi:10.1038/ejhg.2015.282.
- Leonardi, E.; Bellini, M.; Aspromonte, M.C.; Polli, R.; Mercante, A.; Ciaccio, C.; Granocchio, E.; Bettella, E.; Donati, I.; Cainelli, E.; et al. A Novel WAC Loss of Function Mutation in an Individual Presenting with Encephalopathy Related to Status Epilepticus during Sleep (ESES). *Genes* **2020**, *11*, doi:10.3390/genes11030344.
- Alawadhi, A.; Morgan, A.T.; Mucha, B.E.; Scheffer, I.E.; Myers, K.A. Self-Limited Focal Epilepsy and Childhood Apraxia of Speech with WAC Pathogenic Variants. *Eur. J. Paediatr. Neurol. EJPJN Off. J. Eur. Paediatr. Neurol. Soc.* **2021**, *30*, 25–28, doi:10.1016/j.ejpn.2020.12.010.
- Morales, J.A.; Valenzuela, I.; Cuscó, I.; Cogné, B.; Isidor, B.; Matalon, D.R.; Gomez-Ospina, N. Clinical and Molecular Characterization of Five New Individuals with WAC-Related Intellectual Disability: Evidence of Pathogenicity for a Novel Splicing Variant. *Am. J. Med. Genet. A.* **2022**, doi:10.1002/ajmg.a.62648.
- Sanders, S.J.; He, X.; Willsey, A.J.; Ercan-Sencicek, A.G.; Samocha, K.E.; Cicek, A.E.; Murtha, M.T.; Bal, V.H.; Bishop, S.L.; Dong, S.; et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **2015**, *87*, 1215–1233, doi:10.1016/j.neuron.2015.09.016.
- 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. Rare Coding Variation Provides Insight into the Genetic Architecture and Phenotypic Context of Autism. *Nat. Genet.* **2022**, *54*, 1320–1331, doi:10.1038/s41588-022-01104-0.
- Xu, G.M.; Arnaout, M.A. WAC, a Novel WW Domain-Containing Adapter with a Coiled-Coil Region, Is Colocalized with Splicing Factor SC35. *Genomics* **2002**, *79*, 87–94, doi:10.1006/geno.2001.6684.
- Zhang, F.; Yu, X. WAC, a Functional Partner of RNF20/40, Regulates Histone H2B Ubiquitination and Gene Transcription. *Mol. Cell* **2011**, *41*, 384–397, doi:10.1016/j.molcel.2011.01.024.
- McKnight, N.C.; Jefferies, H.B.J.; Alemu, E.A.; Saunders, R.E.; Howell, M.; Johansen, T.; Tooze, S.A. Genome-Wide SiRNA Screen Reveals Amino Acid Starvation-Induced Autophagy Requires SCOC and WAC. *EMBO J.* **2012**, *31*, 1931–1946, doi:10.1038/emboj.2012.36.
- David-Morrison, G.; Xu, Z.; Rui, Y.-N.; Charng, W.-L.; Jaiswal, M.; Yamamoto, S.; Xiong, B.; Zhang, K.; Sandoval, H.; Duraine, L.; et al. WAC Regulates MTOR Activity by Acting as an Adaptor for the TTT and Pontin/Reptin Complexes. *Dev. Cell* **2016**, *36*, 139–151, doi:10.1016/j.devcel.2015.12.019.
- Qi, F.; Chen, Q.; Chen, H.; Yan, H.; Chen, B.; Xiang, X.; Liang, C.; Yi, Q.; Zhang, M.; Cheng, H.; et al. WAC Promotes Polo-like Kinase 1 Activation for Timely Mitotic Entry. *Cell Rep.* **2018**, *24*, 546–556, doi:10.1016/j.celrep.2018.06.087.
- Stafford, A.M.; Pacheco-Vergara, M.; Uhl, K.L.; Jager, T.E.; Li, X.; Jeong, J.; Vogt, D. A Murine Wac Model Exhibits Phenotypes Relevant to DeSanto-Shinawi Syndrome 2022, 2022.01.24.477600.
- Wonders, C.P.; Anderson, S.A. The Origin and Specification of Cortical Interneurons. *Nat. Rev. Neurosci.* **2006**, *7*, 687–696, doi:10.1038/nrn1954.
- Prokop, J.W.; Lazar, J.; Crapitto, G.; Smith, D.C.; Worthey, E.A.; Jacob, H.J. Molecular Modeling in the Age of Clinical Genomics, the Enterprise of the next Generation. *J. Mol. Model.* **2017**, *23*, 75, doi:10.1007/s00894-017-3258-3.
- Prokop, J.W.; Jdanov, V.; Savage, L.; Morris, M.; Lamb, N.; VanSickle, E.; Stenger, C.L.; Rajasekaran, S.; Bupp, C.P. Computational and Experimental Analysis of Genetic Variants. *Compr. Physiol.* **2022**, *12*, 3303–3336, doi:10.1002/cphy.c210012.

20. Maglott, D.; Ostell, J.; Pruitt, K.D.; Tatusova, T. Entrez Gene: Gene-Centered Information at NCBI. *Nucleic Acids Res.* **2005**, *33*, D54–D58, doi:10.1093/nar/gki031.
21. Papadopoulos, J.S.; Agarwala, R. COBALT: Constraint-Based Alignment Tool for Multiple Protein Sequences. *Bioinformatics* **2007**, *23*, 1073–1079, doi:10.1093/bioinformatics/btm076.
22. The UniProt Consortium UniProt: A Hub for Protein Information. *Nucleic Acids Res.* **2015**, *43*, D204–D212, doi:10.1093/nar/gku989.
23. 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. The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans. *Nature* **2020**, *581*, 434–443, doi:10.1038/s41586-020-2308-7.
24. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868, doi:10.1093/nar/gkv1222.
25. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589, doi:10.1038/s41586-021-03819-2.
26. Kumar, M.; Michael, S.; Alvarado-Valverde, J.; Mészáros, B.; Sámano-Sánchez, H.; Zeke, A.; Dobson, L.; Lazar, T.; Örd, M.; Nagpal, A.; et al. The Eukaryotic Linear Motif Resource: 2022 Release. *Nucleic Acids Res.* **2022**, *50*, D497–D508, doi:10.1093/nar/gkab975.
27. Li, M.; Santpere, G.; Imamura Kawasawa, Y.; Evgrafov, O.V.; Gulden, F.O.; Pochareddy, S.; Sunkin, S.M.; Li, Z.; Shin, Y.; Zhu, Y.; et al. Integrative Functional Genomic Analysis of Human Brain Development and Neuropsychiatric Risks. *Science* **2018**, *362*, eaat7615, doi:10.1126/science.aat7615.
28. Jones, A.R.; Overly, C.C.; Sunkin, S.M. The Allen Brain Atlas: 5 Years and Beyond. *Nat. Rev. Neurosci.* **2009**, *10*, 821–828, doi:10.1038/nrn2722.
29. Vogt, D.; Hunt, R.F.; Mandal, S.; Sandberg, M.; Silberberg, S.N.; Nagasawa, T.; Yang, Z.; Baraban, S.C.; Rubenstein, J.L.R. Lhx6 Directly Regulates Arx and CXCR7 to Determine Cortical Interneuron Fate and Laminar Position. *Neuron* **2014**, *82*, 350–364, doi:10.1016/j.neuron.2014.02.030.
30. Angara, K.; Pai, E.L.-L.; Bilinovich, S.M.; Stafford, A.M.; Nguyen, J.T.; Li, K.X.; Paul, A.; Rubenstein, J.L.; Vogt, D. Nf1 Deletion Results in Depletion of the Lhx6 Transcription Factor and a Specific Loss of Parvalbumin+ Cortical Interneurons. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, doi:10.1073/pnas.1915458117.
31. Wundrach, D.; Martinetti, L.E.; Stafford, A.M.; Bilinovich, S.M.; Angara, K.; Prokop, J.W.; Crandall, S.R.; Vogt, D. A Human TSC1 Variant Screening Platform in Gabaergic Cortical Interneurons for Genotype to Phenotype Assessments. *Front. Mol. Neurosci.* **2020**, *13*, 573409, doi:10.3389/fnmol.2020.573409.
32. Sudol, M.; Hunter, T. NeW Wrinkles for an Old Domain. *Cell* **2000**, *103*, 1001–1004, doi:10.1016/S0092-8674(00)00203-8.
33. Truebestein, L.; Leonard, T.A. Coiled-coils: The Long and Short of It. *Bioessays* **2016**, *38*, 903–916, doi:10.1002/bies.201600062.
34. Snijders Blok, L.; Hiatt, S.M.; Bowling, K.M.; Prokop, J.W.; Engel, K.L.; Cochran, J.N.; Bebin, E.M.; Bijlsma, E.K.; Ruivenkamp, C.A.L.; Terhal, P.; et al. De Novo Mutations in MED13, a Component of the Mediator Complex, Are Associated with a Novel Neurodevelopmental Disorder. *Hum. Genet.* **2018**, *137*, 375–388, doi:10.1007/s00439-018-1887-y.
35. Shimada, T.; Fournier, A.E.; Yamagata, K. Neuroprotective Function of 14-3-3 Proteins in Neurodegeneration. *BioMed Res. Int.* **2013**, *2013*, e564534, doi:10.1155/2013/564534.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.