Modeling clinical phenotypes with protein function to explain the disease spectrum in patients with CHIP mutations

Sabrina C. Madrigal[*,1], **Zipporah McNeil**[*,1], **Chang-he Shi**[†,2], **Cam Patterson**[‡,2], **Kenneth Matthew Scaglione**[§,2] **and Jonathan C. Schisler**[*,**,3]

[*]McAllister Heart Institute, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, [†]Department of Neurology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou, Henan, China, [‡]University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA, [§]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC 27710, USA, [**]Department of Pharmacology and Department of Pathology and Lab Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

ABSTRACT Monogenetic disorders that cause cerebellar ataxia are characterized by defects in gait and atrophy of the cerebellum, however, patients often suffer from a spectrum of disease, complicating treatment options. Spinocerebellar autosomal recessive 16 (SCAR16) is caused by coding mutations in *STUB1*, a gene that encodes the multi-functional enyzme CHIP (C-terminus of HSC70-interacting protein). The spectrum of disease found in SCAR16 patients includes a wide range in the age of disease onset, cognitive dysfunction, increased tendon reflex, and hypogonadism. Although SCAR16 mutations span the multiple functional domains of CHIP, it is unclear if the location of the mutation contributes to the clinical spectrum of SCAR16 or with changes in the biochemical properties of CHIP. In this study, we examined the associations and relationships between the clinical phenotypes of SCAR16 patients and how they relate to changes in the biophysical, biochemical, and functional properties of the corresponding mutated protein. We found that the severity of ataxia did not correlate with age of onset; however, cognitive dysfunction, increased tendon reflex, and ancestry were able to predict 54% of the variation in ataxia severity. We further identified domain-specific relationships between biochemical changes in CHIP and clinical phenotypes, and specific biochemical activities that associate selectively to either increased tendon reflex or cognitive dysfunction, suggesting that specific changes to CHIP-HSC70 dynamics contributes to the clinical spectrum of SCAR16. Finally, using linear models of SCAR16 and Monte Carlo simulations, our data support the hypothesis that further inhibiting mutant CHIP activity lessens the severity of SCAR16.

KEYWORDS ataxia; ubiquitin; protein quality control; modeling disease

Ataxia is a general term used to describe a loss of coordination. Ataxia can be caused by a variety of diseases, including metabolic disorders, vitamin deficiencies, peripheral neuropathy, cancer, or brain injuries. In addition to deterioration in movement and balance, ataxia can be accompanied by a spectrum of secondary disorders, including impairments in speech, vision, and cognitive ability. Ataxia is most often caused by the progressive deterioration of the cerebellum, known as cerebellar ataxia (CA), of which there are several causes: hyperthyroidism, alcoholism, stroke, multiple sclerosis, and traumatic injury. Additionally, there are known genetic mutations that are thought to cause CA, and these forms of CA are classified by their inheritance patterns Bird (1993). CA mutations are inherited most commonly in an autosomal recessive manner (estimated prevalence is 7 per 100,000). CA can also manifest as an autosomal dominant disorder (estimated prevalence is 3 per 100,000), in addition to less prevalent X-linked or mitochondrial form of inheritance. Most forms of autosomal dominant CAs are

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Thursday 25th April, 2019
[1]These authors contributed equally to this work.
[2]These authors contributed equally to this work.
[3]Corresponding author: Please insert the affiliation correspondence address and email for the corresponding author. The corresponding author should be marked with the relevant number in the author list, as shown in the example.

caused by polyglutamine expansions within a protein coding region, in contrast to autosomal recessive CAs that are caused by conventional mutations within the coding region Bird (1993). The age of onset, prognosis, accompanying symptoms vary both among and within the genetic forms of CA, and importantly, there are currently no front-line medications for CA Ronnebaum *et al.* (2014).

Spinocerebellar ataxia autosomal recessive 16 (SCAR16, MIM 615768) is a recessive form of cerebellar ataxia with a wide ranging disease spectrum, that can also include hypogonadism, cognitive dysfunction, dysarthria, and increased tendon reflex Ronnebaum *et al.* (2014). Using whole exome sequencing, we identified a mutation in *STUB1* in two patients initially diagnosed with ataxia and hypogonadism Shi *et al.* (2014). Subsequently, numerous clinical reports identified *STUB1* mutations in patients with ataxia, confirming our initial identification of a new disease Shi *et al.* (2014, 2013); Bettencourt *et al.* (2015); Cordoba *et al.* (2014); Depondt *et al.* (2014); Synofzik *et al.* (2014); Hayer *et al.* (2017); Heimdal *et al.* (2014). Remarkably, *STUB1* mutations were found in nearly 2% of Caucasian patients with degenerative ataxia, and these mutations appeared to be specific to the ataxia phenotype and not rare ubiquitous polymorphisms Synofzik *et al.* (2014). *STUB1* encodes the multi-functional enzyme CHIP (C-terminus of HSC70-interacting protein), recognized as an active member of the cellular protein quality control machinery and has multiple functions as both a chaperone Rosser *et al.* (2007); Schisler *et al.* (2013), co-chaperone Ballinger *et al.* (1999); Dai *et al.* (2003), and ubiquitin ligase enzyme Jiang *et al.* (2001); Meacham *et al.* (2001). As a chaperone, CHIP can cause structural changes to proteins to either maintain solubility or increase specific activity. As a co-chaperone, CHIP directly interacts with heat shock proteins (HSP) and can aid in the stabilization and refolding of HSP-bound substrates. Conversely, as a ubiquitin ligase, CHIP ubiquitinates terminally-defective proteins and targets them for degradation by the ubiquitin proteasome system.

SCAR16 mutations span the three functional domains of CHIP (Figure 1A): the N-terminal TPR domain that binds HSPs, the coiled-coiled domain that is important for dimerization, and the C-terminal Ubox domain that is responsible for the ubiquitin ligase function Ronnebaum *et al.* (2014). Currently, it is not known if the location of these mutations mediate specific aspects of the SCAR16 spectrum. Equally so, it is not known how changes in CHIP properties caused by substitution mutations relate to clinical phenotypes. In this report, we combined clinical data provided by numerous reports and a recent report that characterized the biochemical repercussions of several of these SCAR16 disease mutations (Figure 1A) Kanack *et al.* (2018). Our approach allowed us to identify the specific biochemical changes to CHIP that are coupled to SCAR16 clinical characteristics. We developed linear models and used simulations to identify which properties of mutant CHIP proteins may impact disease severity. Defining the relationship between changes in specific features of CHIP and disease phenotypes may reveal new clues to both the spectrum of this disease, and ultimately, guide precision medicine-based strategies to treat SCAR16.

Materials and Methods

SCAR16 patient data

Clinical data were obtained from published reports (Table S1) Shi *et al.* (2014, 2013); Bettencourt *et al.* (2015); Cordoba *et al.* (2014); Depondt *et al.* (2014); Synofzik *et al.* (2014); Hayer *et al.* (2017); Heimdal *et al.* (2014). One measure of disease severity is the score from the Scale for the Assessment and Rating of Ataxia (SARA). When SARA scores were not implicitly stated, SARA was imputed based on the clinical report as indicated in Table S1 Schmitz-Hübsch *et al.* (2006).

CHIP mutation data

All biophysical and biochemical properties of CHIP proteins with disease-associated substitution mutations were obtained from published data (Table S2) Kanack *et al.* (2018). HSP70 ubiquitination at 37 °C or 25 °C was measured by densitometry analysis and represented by the total amount of HSP70 that was modified by ubiquitination; wild-type (WT) CHIP ubiquitinated 73% or 81% of the HSP70 total in the reaction, respectively.

Statistical Analysis

All analyses were performed using JMP Pro (v14.2.0). Continuous and categorical clinical variable distributions were analyzed using the Shapiro-Wilk W test or chi-squared test, respectively. Bivariate analysis was performed using either t test, one-way ANOVA, linear regression, or contingency analysis using Fisher's exact test, depending on the two variables being compared. A false discovery rate of <10% (Benjamini-Hochberg) was used to control for multiple test correction for each dependent variable (Table S3), with raw P values reported. Post hoc tests, when applicable, are described in the figure legends. Partial least squares regression was initially performed using all variables, using leave-one-out cross validation and then refined by retaining all variables with VIP values > 0.8 Garthwaite (1994); Cox and Gaudard (2013). The model coefficients for the original data are provided in the corresponding data tables.

Monte Carlo simulations were performed by modeling SARA and age of onset using a multiple Y partial least squares regression equation using the indicated variables Garthwaite (1994); Wold *et al.* (2001); Eriksson *et al.* (2013); Cox and Gaudard (2013). Targets for SARA and age of onset improvement were set at one standard deviation from the mean. Simulations were run using all X variables as random with a normal distribution. 5000 baseline simulations were used to fit the original data. The simulation was adjusted to maximize desirability via an additional 5000 simulations to identify parameters that maximize improvement of both Y variables Derringer and Suich (1980).

Data Availability

Table S1 contains clinical variables. Table S2 contains the biochemical data. Table S3 contains the adjusted P values. All supplementary files are available at UNC digital repository, DOI: 10.17615/8dqf-e678.

Results and Discussion

Clinical variable analyses from SCAR16 patients

We collated the phenotypic data from the various clinical reports encompassing 24 patients (Table S1) Shi *et al.* (2014, 2013); Bettencourt *et al.* (2015); Cordoba *et al.* (2014); Depondt *et al.* (2014); Synofzik *et al.* (2014); Hayer *et al.* (2017); Heimdal *et al.* (2014). Several variables had skewed distributions (Table 1), most notably, hypogonadism was found in only four patients, whereas over 70% of the patients suffered from increased tendon reflex and/or cognitive dysfunction . There was nearly an equal number of males and females, as well as homozygous and hemizygous patients. The median clinical score for ataxia (SARA) was 18.5, a value associated with moderate dependence for daily

activities Schmitz-Hübsch *et al.* (2006), and the median age of onset was 17 years of age.

Bivariate analysis: We explored the co-penetrance of three reported SCAR16 clinical phenotypes, cognitive dysfunction, increased tendon reflex, and hypogonadism. There was no association between cognitive dysfunction and increased tendon reflex (Fisher's exact test, $P = 1.000$, suggesting that these two phenotypes are distinct from each other. Hypogonadism was only reported in four patients, limiting the power of the contingency analysis, however, all four patients with hypogonadism also had cognitive dysfunction (Fisher's exact test, $P < 0.0001$).

We next examined the relationship between two measures of disease severity, the age of onset and SARA, a reliable and valid clinical measure of ataxia Schmitz-Hübsch *et al.* (2006). We hypothesized that more severe ataxia would correlate with an earlier age of onset. Counter to our hypothesis, there was no correlation between age of onset and ataxia (Figure 1B), indicating that these two variables were independent of each other. We further explored the relationships of age of onset and SARA with the other clinical variables reported in SCAR16 patients (Table 1). Patients with homozygous mutations had a reported disease onset 12 years earlier than patients with compound heterozygous mutations (Figure 1C). However, unlike other genetic forms of ataxia caused by trinucleotide repeats, we did not find any additional associations between either homozygozity or age of onset with either SARA or any of the remaining clinical variables (Figure 1D, Table S2, S3). In contrast, when we measured the associations between SARA and the remaining clinical variables (Table 2) we observed that SCAR16 patients with cognitive dysfunction scored 10 points higher on the SARA assessment (Figure 1E). We also found that patients with European ancestry had the highest average SARA (Figure 1F, median = 32) whereas there was no difference between those with Han Chinese or Middle Eastern/North African ancestry (median = 15 and 14.5, respectively). These data demonstrate that SCAR16 patients with cognitive dysfunction had more severe deficits in motor function and that other genetic factors may potentially influence the effect of CHIP mutations on the severity of ataxia.

Multivariate analysis: Given the covariance structure of the patient variables and the mixture of continuous and nominal variables, we employed partial least squares regression to model ataxia severity (using SARA) as a function of the clinical characteristics. This modeling technique permits the usage of correlated explanatory phenotype variables in describing the distribution of SARA. Using leave-one-out cross-validation we identified three factors: ancestry, cognitive dysfunction, and tendon reflex (Figure 2A) that could explain 54% of the variation in SARA (Figure 2B). We found that cognitive dysfunction had the largest estimate in the regression analysis (Equation 1, CD = cognitive dysfunction, TR = tendon reflex, Y = yes, N = no, EAS = East Asian ancestry, EUR = European ancestry). Also, European decent and increased tendon reflex contributed to higher SARA. Our model suggests that genetic factors associated with ancestry may influence the severity of SCAR16, even to a greater extent than homozygosity, age of onset, and sex, as these later variables did not contribute to the predictive power of the modeling (Table 2).

$$\begin{aligned} SARA_{adj} = 17.2 + [5.6(CD_Y), -5.6(CD_N)] + [4.7(EUR), \\ 0.9(SAS), -1.4(MENA), -1.7(AMR), -3.4(EAS)] \\ +[2.7(TR_Y), -2.7(TR_N)] \end{aligned} \quad (1)$$

Changes in CHIP biochemistry caused by SCAR16 mutations

A recent study analyzed the effect of SCAR16 mutations on the properties of CHIP at the protein level Kanack *et al.* (2018). We used these data to look for relationships between various biophysical and biochemical properties to better understand the effect of these mutations on CHIP function. All continuous data were normally distributed except for K_D, which did not meet normalcy testing even after various transformations, therefore, we performed all analyses using non-transformed data. As expected, there was a strong positive correlation between ubiquitin chain formation (a readout of the alignment between CHIP and the E2 enzyme) Kanack *et al.* (2018) and the extent of HSP70 ubiquitination ($\rho = 0.64$). Also, HSP70 ubiquitination was inversely correlated with the K_D between CHIP and a peptide containing the HSP70 binding motif (ρ = -0.76). We observed a positive correlation between $B_{\max}$ and K_D regarding CHIP interactions with the HSP70 peptide ($\rho = 0.66$). This unexpected positive correlation suggests that some mutant CHIP proteins, including CHIP-T246M, may have an increased binding capacity towards chaperones, consistent with our previous studies Shi *et al.* (2014, 2018). Throughout the initial correlation analyses (Figure 3A) we observed clustering of proteins based on the domain that harbors the mutation. Therefore, we performed hierarchical clustering of mutant CHIP proteins and their corresponding biochemical parameters. Mutant proteins clustered primarily by the domain harboring the mutation (Figure 3B), consistent with the premise that the domain affected by the mutation may have differential actions on CHIP activities.

The impact of the location of SCAR16 mutations on CHIP biochemistry: We hypothesized that the location of the mutation would differentially impact the multiple functions of CHIP to either bind to chaperones and/or mediate poly-ubiquitin chain formation. To test this hypothesis, we associated the variability in biochemical properties of CHIP mutations with the domain harboring the mutation. This analysis revealed three associations at a FDR < 10% (Table 3). First, binding studies between CHIP and the HSP70 tail peptide revealed that Ubox mutations had increased binding capacity towards HSP70 (Figure 3C); however, there was no difference in binding affinity across domains (Figure 3D). Further, as shown by Kanack *et al.* (2018), all mutant CHIP proteins maintained some capacity to polyubiquitinate HSP70, except for two Ubox mutants, M240T and T246M (Figure 3E). Moreover, every Ubox mutant had a reduced capacity to form polyubiquitin chains, perhaps due to altered interactions with E2 enzymes, whereas TPR and CC mutations were not defective in these same conditions (Figure 3F). These data suggest that Ubox mutations, in particular, affect functions that span both the co-chaperone and ubiquitin ligase functions of CHIP.

Determining the relationship between patient phenotypes and the altered biochemical properties of mutant CHIP proteins

Finally, we analyzed the clinical phenotypic data from SCAR16 patients and the corresponding biochemical characteristics of mutant CHIP proteins to determine if clinical variables were related to biochemical changes caused by the coding mutations. This approach has several limitations. First, only substitution mutations were analyzed; therefore, we do not capture the phenotype-functional associations from indel mutations. Second, given the recessive and sometimes hemizygous nature of SCAR16 we performed these analyses on a per allele basis. Thus, in the case of hemizygous patients, our analysis cannot distinguish if one mutation has a more functional role than the other

regarding the clinical manifestations of SCAR16. However, the majority of hemizygous genotypes include one allele that results in a pre-terminal stop codon, predicted to be degraded by non-sense mediated RNA decay (Table S1).

Connecting the location of CHIP mutations with SCAR16 clinical variables: We used oneway ANOVA to measure the relationship between the domain location harboring the mutation and the continuous clinical variables. We did not identify any variation in either age of onset, SARA, or $SARA_{adj}$ that associated with the domain location of the mutation (P = 0.51, 0.86, or 0.14, respectively). However, we observed a highly skewed distribution between the mutation location and cognitive dysfunction (Figure 4). Most notably, 94% of alleles with Ubox mutations associated with cognitive dysfunction, with no difference in the frequency of cognitive dysfunction between TRP or CC allele mutations (59% of TPR or CC alleles associated with cognitive dysfunction). In constrast, hypogonadism and increased tendon reflex were equally distributed in regards to the location of CHIP mutations (Fisher's exact test = 0.15 and 1.00, respectively). These data suggest that domain-specific changes in CHIP function may contribute to the clinical spectrum of SCAR16, particularly Ubox mutations that are linked with cognitive dysfunction.

Bivariate analysis of the biochemical changes caused by CHIP mutations with cognitive dysfunction and tendon reflex: Changes in the biochemical properties of mutant CHIP may influence the clinical spectrum of SCAR16. As discussed above, there is no link between cognitive dysfunction and tendon reflex phenotypes in SCAR16 patients. As such, we tested the relationships between the corresponding biochemical properties of mutant CHIP proteins encoded by disease alleles and these two independent phenotypes. We found that T_m and steady-state expression levels of mutant CHIP proteins did not associate with either cognitive dysfunction (P = 0.57 and 0.58) or increased tendon reflex (P = 0.28 or 0.62). However, cognitive dysfunction associated with 35% lower activity in CHIP/E2-mediated ubiquitin chain formation (Figure 5A), suggesting this activity is essential to maintain cerebellar cognition. CHIP functions primarily as a dimer, and seven of the 13 characterized mutations maintain oligomeric distributions similar to CHIP-WT Kanack *et al.* (2018). Interestingly, dimeric forms of mutant CHIP associated with increased tendon reflex (Figure 5B). Therefore, loss of ubiquitin ligase function may be a prominent contributor to cognitive dysfunction, whereas CHIP that still forms dimers, and perhaps some degree of altered CHIP function, may contribute to increased tendon reflex. Accordingly, aspects of the CHIP-HSC70 interaction changed reciprocally when compared to either cognitive dysfunction or tendon reflex. Alleles from SCAR16 patients with cognitive dysfunction had increased HSP70 binding capacity (delta = 14.3 μmol per min, (Figure 5C) with no change in K_D (Figure 5D). In contrast, alleles from SCAR16 patients with increased tendon reflex corresponded to mutant CHIP proteins with similar binding capacity (Figure 5E) and affinity to HSC70 (Figure 5F) as wild-type CHIP. Although only reaching marginal significance, we again observed opposite effects with cognitive dysfunction and tendon reflex regarding the ability of CHIP to ubiquitinate HSP70 (Figure 5G, 5H), effects that are consistent with the patterns in ubiquitin chain formation (Figure 5A) and HSP70 binding (Figure 5C - 5F). Differential biochemical activities of CHIP linked with these clinical phenotypes support the concept that altered CHIP-HSC70 dynamics, caused by disease mutations, contribute to the clinical spectrum of SCAR16.

Multivariate modeling of disease onset and SARA as a function of the biochemical properties of mutant CHIP: We predicted that the biochemical changes to CHIP via disease mutations would affect both the age of onset and disease severity. We first performed bivariate analysis of age of onset and SARA with each of the biochemical properties of the mutant alleles (Table 4). Alleles that encode mutant CHIP proteins with lower activity and higher $B_{\max}$ associated with an *earlier* age of onset. Additionally, alleles encoding mutant CHIP proteins forming higher-order oligomers and reduced binding affinities associated with *lower* SARA scores. These data suggest that non-functional forms of CHIP (higher-order oligomers and decreased HSC70 binding affinity) results in less severe disease as opposed to mutant CHIP proteins that still maintain normal tertiary structure and binding activities towards chaperones. We then developed multivariate linear models using partial least squares for both age of onset (Figure 6A, Equation 2) and SARA (Figure 6B, Equation 3) as functions of the biochemical properties of mutant CHIP proteins. T_m and %expression were the only variables that did not contribute to the modeling of either onset or SARA (VIP > 0.8). The remaining five variables accounted for 70% of the variance in the age of onset; however, SARA modeling was less powerful, accounting for only 20% of the variance (Table 4).

$$\begin{aligned} Onset_{adj} = 22.5 + [5.8(oligomer), -5.8(dimer)] \\ -0.3[(\%HSP70ub - 43.7)/15.5] \\ +3.8[(\%chainformation - 69.9)/39.0] \\ +0.5[(K_D - 7.8)/6.0] \\ -4.5[(B_{\max} - 212.0)/16.6] \end{aligned} \quad (2)$$

$$\begin{aligned} SARA_{adj2} = 18.2 + [1.7(oligomer), -1.7(dimer)] \\ +0.7[(\%HSP70ub - 43.7)/15.5] \\ +0.8[(\%chainformation - 69.9)/39.0] \\ +0.9[(K_D - 7.8)/6.0] \\ -0.7[(B_{\max} - 212.0)/16.6] \end{aligned} \quad (3)$$

Effects of CHIP stability on disease onset and severity : One approach in designing therapies for SCAR16 would be to restore CHIP activity. Kanack et al. hypothesized that mutant CHIP proteins could be stabilized by decreasing the reaction temperature below the T_m of the mutant proteins Kanack *et al.* (2018). It was subsequently demonstrated that decreasing the reaction temperature to 25 °C, partially recovered CHIP activities Kanack *et al.* (2018). Statistical analysis of these data demonstrate that decreasing the reaction temperature increased HSP70 ubiquitination (Figure 7A) and partially recovered some of the in vitro binding alterations seen in mutant CHIP proteins (Figure 7B, 7C). We used the data from experiments performed at 25 °C, including %HSP70ub, K_D $B_{\max}$, in the biochemical models of onset and SARA to examine the impact on these two phenotypes of SCAR16. Increasing mutant CHIP activity led to an earlier predicted age of onset (Figure 8A) with the mean of the model predictions equaling a six year lower age of onset (Figure 8B). We found that mutant CHIP parameters at 25 °C had a less dramatic impact on SARA (Figure 8C) and predicted a small increase in SARA (Figure 8D). Considering the effects on both SARA and age of onset, the effect of increasing mutant CHIP activities, by

lowering the reaction temperature, predict more severe disease phenotypes.

Ultimately, we were interested in identifying changes in CHIP properties and activities that may result in delaying onset and decreasing severity (SARA). We ran Monte Carlo simulations with a combined partial least squares regression model of both age of onset and SARA, incorporating the four continuous biochemical variables identified in Table 3. The simulation that fit the original SCAR16 patient data indicate decreases in both ubiquitination-related activities and a modest reduction in binding affinity compared to wild-type CHIP (Table 5). Next, we set the improvement targets at one σ from the mean, equating to a 10.7 year increase in age of onset and a 10 point decrease in SARA. The results of the simulation were consistent towards both Y variables, with the overall trend of further inhibiting mutant CHIP activity towards HSP70, by decreasing E3 activity and/or reducing the binding affinity to HSP70, would both delay onset and severity of SCAR16 (Figure 9). Our simulations successfully delayed the age of onset by 7.2 years and decreased SARA by 8.2 points by decreasing the affinity of mutant CHIP proteins to HSP70 to 17 μM and decreasing the ubiquitin ligase activity to 7.2%, relative to wild-type CHIP (Table 5). Not surprisingly, there was a strong negative correlation between the amount of HSP70 ubiquitination and K_D of the CHIP-HSP70 interaction (Figure 3A, ρ = -0.76), suggesting that blocking the interaction mutant CHIP with HSP70 represents a therapeutic opportunity that can be explored in future studies.

What remains unclear is why Ubox mutations, with disrupted ubiquitin-related activities, strongly associate with cognitive dysfunction. Ubox mutations also have higher Bmax, and we previously observed that the Ubox mutant CHIP-T246M pulled down more HSC70 and HSP70 compared to CHIP-WT in multiple cell models and in our engineered mouse line that expresses CHIP-T246M from the endogenous locus Shi *et al.* (2014, 2018). One possibility is that Ubox mutants still bind E2 enzymes, but the inability to transfer ubiquitin disrupts E2 function, and perhaps activity of E2 enzymes towards other E3 ligases. Alternatively, the propensity to form higher-order oligomers Kanack *et al.* (2018); Shi *et al.* (2018) and changes in solubility Shi *et al.* (2018) could also affect the function of proteins that still interact with Ubox mutants.

Additional SCAR16 patients with new, uncharacterized mutations continue to be reported Gazulla *et al.* (2018); Turkgenc *et al.* (2018); Hayer *et al.* (2017); García *et al.* (2017); Kawarai *et al.* (2016), in addition to a possible variant that functions in a dominant manner Genis *et al.* (2018). With additional clinical data and with the advent of new molecular, cellular, and preclincal models to study CHIP function, it is likely that precision-based approaches could be developed based on the specific mutation or the specific loss of function. For example, patients not affected cognitively could benefit from blocking the interaction between mutant CHIP and chaperones, however, SCAR16 patients with cognitive dysfunction may benefit from the use of molecular chaperones to prevent the oligmerization of these mutant CHIP proteins. CHIP impacts several cellular pathways, and identifying the CHIP-dependent pathways drive the specific pathologies in SCAR16, such as necroptosis Seo *et al.* (2016), IGF1 Tawo *et al.* (2017), mitophagy Lizama *et al.* (2018), autophagy Ravi *et al.* (2018); Guo *et al.* (2015), or water balance Wu *et al.* (2017), may also uncover therapeutically relevant targets. Additionally, gene therapy approaches may also be applicable to SCAR16. The obvious solution is to use gene editing approaches to correct these mutations, however, given the recessive nature of the disease, delivery of a functional copy of CHIP may also be beneficial Chandler *et al.* (2017); Cabral-Miranda *et al.* (2017); alternatively, antisense oligonucleotide therapy that can downregulate mutant CHIP protein levels may also prove to be an an effective approach Scoles *et al.* (2017). Finally, by looking more broadly at the various autosomal recessive ataxias, additional themes and targets that could be effective across multiple ataxia diseases may also come to light in the years to come Synofzik *et al.* (2019).

Acknowledgements

We thank members of the Schisler Laboratory, including Rebekah Sanchez-Hodge for a critical review of the manuscript, Kalleen Kelley, and the McAllister Heart Institute administration team. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

Literature Cited

Ballinger, C. A., P. Connell, Y. Wu, Z. Hu, L. J. Thompson, *et al.*, 1999 Identification of CHIP, a Novel Tetratricopeptide Repeat-Containing Protein That Interacts with Heat Shock Proteins and Negatively Regulates Chaperone Functions. Molecular and Cellular Biology **19**: 4535–4545, 00784.

Bettencourt, C., J. G. de Yébenes, J. L. López-Sendón, O. Shomroni, X. Zhang, *et al.*, 2015 Clinical and Neuropathological Features of Spastic Ataxia in a Spanish Family with Novel Compound Heterozygous Mutations in STUB1. Cerebellum (London, England) **14**: 378–381, 00003.

Bird, T. D., 1993 Hereditary Ataxia Overview. In *GeneReviews®*, edited by M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. Bean, K. Stephens, and A. Amemiya, University of Washington, Seattle, Seattle (WA).

Cabral-Miranda, F., E. Nicoloso-Simões, J. Adão-Novaes, V. Chiodo, W. W. Hauswirth, *et al.*, 2017 rAAV8-733-Mediated Gene Transfer of CHIP/Stub-1 Prevents Hippocampal Neuronal Death in Experimental Brain Ischemia. Molecular Therapy **25**: 392–400.

Chandler, R. J., I. M. Williams, A. L. Gibson, C. D. Davidson, A. A. Incao, *et al.*, 2017 Systemic AAV9 gene therapy improves the lifespan of mice with Niemann-Pick disease, type C1. Human Molecular Genetics **26**: 52–64.

Cordoba, M., S. Rodriguez-Quiroga, E. M. Gatto, A. Alurralde, and M. A. Kauffman, 2014 Ataxia plus myoclonus in a 23-year-old patient due to STUB1 mutations. Neurology **83**: 287–288, 00014.

Cox, I. and M. Gaudard, 2013 *Discovering partial least squares with JMP*. SAS Institute.

Dai, Q., C. Zhang, Y. Wu, H. McDonough, R. A. Whaley, *et al.*, 2003 CHIP activates HSF1 and confers protection against apoptosis and cellular stress. The EMBO Journal **22**: 5446–5458, 00259.

Depondt, C., S. Donatello, N. Simonis, M. Rai, R. van Heurck, *et al.*, 2014 Autosomal recessive cerebellar ataxia of adult onset due to STUB1 mutations. Neurology **82**: 1749–1750, 00014.

Derringer, G. and R. Suich, 1980 Simultaneous optimization of several response variables. Journal of quality technology **12**: 214–219.

Eriksson, L., T. Byrne, E. Johansson, J. Trygg, and C. Vikström, 2013 *Multi- and Megavariate Data Analysis Basic Principles and Applications*. Umetrics Academy, Google-Books-ID: 58qLBQAAQBAJ.

García, A. M., S. Abrevaya, G. Kozono, I. G. Cordero, M. Córdoba, *et al.*, 2017 The cerebellum and embodied semantics: evidence from a case of genetic ataxia due to STUB1 mutations. Journal of Medical Genetics **54**: 114–124.

Garthwaite, P. H., 1994 An interpretation of partial least squares. Journal of the American Statistical Association **89**: 122–127.

Gazulla, J., S. Izquierdo-Alvarez, E. Sierra-Martínez, M. E. Marta-Moreno, and S. Alvarez, 2018 Inaugural cognitive decline, late disease onset and novel STUB1 variants in SCAR16. Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology **39**: 2231–2233.

Genis, D., S. Ortega-Cubero, H. San Nicolás, J. Corral, J. Gardenyes, *et al.*, 2018 Heterozygous STUB1 mutation causes familial ataxia with cognitive affective syndrome (SCA48). Neurology **91**: e1988–e1998.

Guo, D., Z. Ying, H. Wang, D. Chen, F. Gao, *et al.*, 2015 Regulation of autophagic flux by CHIP. Neuroscience Bulletin **31**: 469–479.

Hayer, S. N., T. Deconinck, B. Bender, K. Smets, S. Züchner, *et al.*, 2017 STUB1/CHIP mutations cause Gordon Holmes syndrome as part of a widespread multisystemic neurodegeneration: evidence from four novel mutations. Orphanet Journal of Rare Diseases **12**.

Heimdal, K., M. Sanchez-Guixé, I. Aukrust, J. Bollerslev, O. Bruland, *et al.*, 2014 STUB1 mutations in autosomal recessive ataxias – evidence for mutation-specific clinical heterogeneity. Orphanet Journal of Rare Diseases **9**, 00013.

Jiang, J., C. A. Ballinger, Y. Wu, Q. Dai, D. M. Cyr, *et al.*, 2001 CHIP Is a U-box-dependent E3 Ubiquitin Ligase IDENTIFICATION OF Hsc70 AS A TARGET FOR UBIQUITYLATION. Journal of Biological Chemistry **276**: 42938–42944, 00507.

Kanack, A. J., O. J. Newsom, and K. M. Scaglione, 2018 Most mutations that cause spinocerebellar ataxia autosomal recessive type 16 (SCAR16) destabilize the protein quality-control E3 ligase CHIP. The Journal of Biological Chemistry **293**: 2735–2743.

Kawarai, T., R. Miyamoto, Y. Shimatani, A. Orlacchio, and R. Kaji, 2016 Choreoathetosis, Dystonia, and Myoclonus in 3 Siblings With Autosomal Recessive Spinocerebellar Ataxia Type 16. JAMA neurology **73**: 888–890.

Lizama, B. N., A. M. Palubinsky, V. A. Raveendran, A. M. Moore, J. D. Federspiel, *et al.*, 2018 Neuronal Preconditioning Requires the Mitophagic Activity of C-terminus of HSC70-Interacting Protein. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience .

Meacham, G. C., C. Patterson, W. Zhang, J. M. Younger, and D. M. Cyr, 2001 The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. Nature Cell Biology **3**: 100–105, 00752.

Ravi, S., T. L. Parry, M. S. Willis, P. Lockyer, C. Patterson, *et al.*, 2018 Adverse Effects of Fenofibrate in Mice Deficient in the Protein Quality Control Regulator, CHIP. Journal of Cardiovascular Development and Disease **5**.

Ronnebaum, S. M., C. Patterson, and J. C. Schisler, 2014 Emerging evidence of coding mutations in the ubiquitin–proteasome system associated with cerebellar ataxias. Human Genome Variation **1**: 14018, 00005.

Rosser, M. F. N., E. Washburn, P. J. Muchowski, C. Patterson, and D. M. Cyr, 2007 Chaperone Functions of the E3 Ubiquitin Ligase CHIP. Journal of Biological Chemistry **282**: 22267–22277, 00116.

Schisler, J. C., C. E. Rubel, C. Zhang, P. Lockyer, D. M. Cyr, *et al.*, 2013 CHIP protects against cardiac pressure overload through regulation of AMPK. The Journal of Clinical Investigation **123**: 3588–3599, 00024.

Schmitz-Hübsch, T., S. T. du Montcel, L. Baliko, J. Berciano, S. Boesch, *et al.*, 2006 Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology **66**: 1717–1720.

Scoles, D. R., P. Meera, M. D. Schneider, S. Paul, W. Dansithong, *et al.*, 2017 Antisense oligonucleotide therapy for spinocerebellar ataxia type 2. Nature **544**: 362–366.

Seo, J., E.-W. Lee, H. Sung, D. Seong, Y. Dondelinger, *et al.*, 2016 CHIP controls necroptosis through ubiquitylation- and lysosome-dependent degradation of RIPK3. Nature Cell Biology **18**: 291–302.

Shi, C.-h., C. Rubel, S. E. Soss, R. Sanchez-Hodge, S. Zhang, *et al.*, 2018 Disrupted structure and aberrant function of CHIP mediates the loss of motor and cognitive function in preclinical models of SCAR16. PLoS Genetics **14**.

Shi, C.-H., J. C. Schisler, C. E. Rubel, S. Tan, B. Song, *et al.*, 2014 Ataxia and hypogonadism caused by the loss of ubiquitin ligase activity of the U box protein CHIP. Human Molecular Genetics **23**: 1013–1024, 00047.

Shi, Y., J. Wang, J.-D. Li, H. Ren, W. Guan, *et al.*, 2013 Identification of CHIP as a Novel Causative Gene for Autosomal Recessive Cerebellar Ataxia. PLoS ONE **8**, 00025.

Synofzik, M., H. Puccio, F. Mochel, and L. Schöls, 2019 Autosomal Recessive Cerebellar Ataxias: Paving the Way toward Targeted Molecular Therapies. Neuron **101**: 560–583.

Synofzik, M., R. Schüle, M. Schulze, J. Gburek-Augustat, R. Schweizer, *et al.*, 2014 Phenotype and frequency of STUB1 mutations: next-generation screenings in Caucasian ataxia and spastic paraplegia cohorts. Orphanet Journal of Rare Diseases **9**: 57, 00016.

Tawo, R., W. Pokrzywa, Kevei, M. E. Akyuz, V. Balaji, *et al.*, 2017 The Ubiquitin Ligase CHIP Integrates Proteostasis and Aging by Regulation of Insulin Receptor Turnover. Cell **169**: 470–482.e13.

Turkgenc, B., B. Sanlidag, A. Eker, A. Giray, O. Kutuk, *et al.*, 2018 STUB1 polyadenylation signal variant AACAAA does not affect polyadenylation but decreases STUB1 translation causing SCAR16. Human Mutation **39**: 1344–1348.

Wold, S., M. Sjöström, and L. Eriksson, 2001 PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems **58**: 109–130.

Wu, Q., H. B. Moeller, D. A. Stevens, R. Sanchez-Hodge, G. Childers, *et al.*, 2017 CHIP Regulates Aquaporin-2 Quality Control and Body Water Homeostasis. Journal of the American Society of Nephrology: JASN .

Table 1 Patient variables

Trait	Abundance (N = 24)	*P*
Hypogonadism (yes)	17%	0.0006
Increased tendon reflex (yes)	75%	0.0122
Ancestry (AMR, SAS, MENA, EUR, EAS)	4%, 4%, 25%, 33%, 33%	0.0126
Cognitive dysfunction (yes)	71%	0.0382
Onset age (year)	17 (10)	0.0674
SARA (score)	18.5 (16.4)	0.1413
Homozygosity (yes)	46%	0.6829
Sex (male)	54%	0.6829

Abundance for each trait is represented by either the percentage or median (and interquartile range) for categorical or continuous variables, respectively. Super populations for ancestry: AMR = Ad Mixed American, SAS = South Asian, MENA = Middle Eastern North African, EUR = European, EAS = East Asian. SARA = Scale for the Assessment and Rating of Ataxia.

Table 2 Analysis and modeling of SARA with clinical variables

Variable	Statistic	*P*	Model	Coefficient
Cognitive dysfunction (yes)	13.8	0.002	Y	4.9303
Ancestry (AMR, SAS, MENA, EUR, EAS)	4.13[a]	0.032	Y	-1.4603, 0.7768, -1.2308, 4.1984, -3.0372
Increased tendon reflex (yes)	6.56	0.199	Y	2.3424
Homozygosity (yes)	-4.38	0.328	N	-1.5642
Onset age (year)	-0.08[b]	0.721	N	-0.2934
Sex (male)	1.19	0.792	N	0.4248
Hypogonadism (yes)	1.00	0.869	N	0.3573

Student's t tests were used to calculate the difference (statistic) and *P* value between SARA and each variable, unless otherwise indicated. Variables that were included in the regression modeling are indicated as well as the beta coefficients from the initial fit. Scale for the Assessment and Rating of Ataxia. Super populations for ancestry: AMR = Ad Mixed American, SAS = South Asian, MENA = Middle Eastern North African, EUR = European, EAS = East Asian.

[a] ANOVA was used and the F ratio is provided as the statistic
[b] Linear regression was used and the beta is provided as the statistic

Table 3 Analysis of SCAR16 mutation location with biochemical properties of the encoded mutant CHIP protein

Biochemical property	F ratio	*P*
B_{max}	10.7	0.0055
%chain formation	9.3	0.0053
%HSP70 ubiquitination	5.2	0.0281
K_D	1.5	0.2871
T_m	0.6	0.5430
%expression	0.1	0.8658

The F ratio and *P* value from ANOVA testing the three domains of CHIP with the variance of biochemical properties. B_{max} = amount of binding, K_D = equilibrium dissociation constant. T_m = thermal melting temperature.

Table 4 **Analysis and modeling age of onset and SARA with biochemical properties of mutant CHIP proteins**

	Onset (y)				SARA			
Variable	Statistic	*P*	Model	Coefficient	Statistic	*P*	Model	Coefficient
%chain formation	0.182	0.0001	Y	3.80	0.046	0.2865	Y	0.78
B_{max}	-0.286	0.0044	Y	-4.53	-0.190	0.1014	Y	-0.67
K_D	-0.148	0.6203	Y	0.46	-0.715	0.0228	Y	-0.91
Oligomeric (yes)	4.066[a]	0.3372	Y	5.77	-9.083[a]	0.0070	Y	-1.73
%HSP70 ubiquitination	0.071	0.5862	Y	-0.29	0.204	0.0559	Y	0.70
T_m	-0.503	0.3170	N	N/A	0.665	0.1081	N	N/A
%expression	0.020	0.7916	N	N/A	-0.056	0.3643	N	N/A

Linear regression was used to calculate the beta (statistic) and *P* value between SARA or Onset with the indicated variables, unless otherwise indicated. Variables that were included in the regression modeling are indicated as well as the beta coefficients from the initial fit. SARA = Scale for the Assessment and Rating of Ataxia, B_{max} = amount of HSP70 binding, K_D = equilibrium dissociation constant. T_m = thermal melting temperature.

[a] T test was used and the difference reported as the statistic

Table 5 **Monte Carlo simulations of onset and SARA in SCAR16 using biochemical properties of mutant CHIP protein**

	%HSP70 ub	%chain	K_D	B_{max}	Onset (y)	SARA
wild-type CHIP	73.1%	100%	2.9	204	N/A	N/A
Simulation						
SCAR16	44%	70%	7.8	212	20.6	18.8
$SCAR16_{opt}$	7.2%	15%	17	217	27.8	10.6

The values of %HSP70 ubiquitination (ub), %ubiquitin chain formation, equilibrium dissociation constant (K_D), and amount of HSP70 binding (B_{max}) for wild-type CHIP is provided. The results of modeling age of onset and SARA using the experimental biochemical data of mutant CHIP proteins (SCAR16) compared to the model optimized to reduce disease severity ($SCAR16_{opt}$) are indicated.

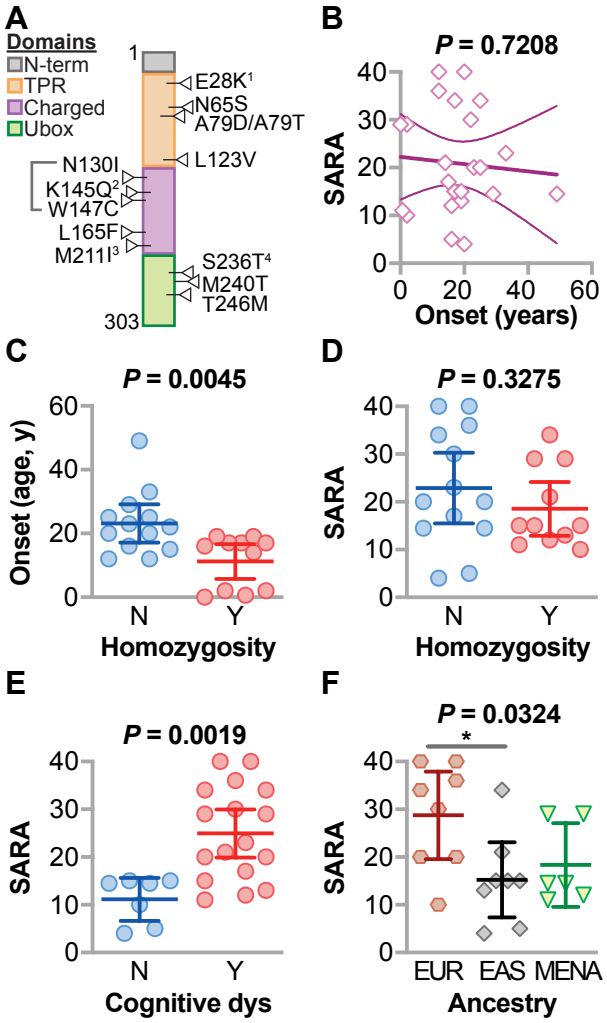

Figure 1 Relationship between SCAR16 clinical variables. (A) Locations of the biochemically characterized SCAR16 mutations across the various functional domains of CHIP. Compound mutations with a second pre-terminal stop codon allele are indicated: [1]K144*, [2]Y230Cfx*8, [3]E238*, [4]Y207*. (B) Regression analysis of age of onset and SARA summarized by the best fit line and the 95% confidence interval. (C) Age of onset and (D) SARA score of SCAR16 patients stratified by homozygosity as well as (E) SARA score of SCAR16 patients stratified by cognitive dysfunction (dys); data are represented by dot plot and summarized by the mean ± 95% confidence interval, analyzed by t test. (F) SARA score of SCAR16 patients stratified by ancestry; data are represented by dot plot and summarized by the mean ± 95% confidence interval, analyzed by ANOVA: * $P < 0.05$ via Tukey post hoc test.

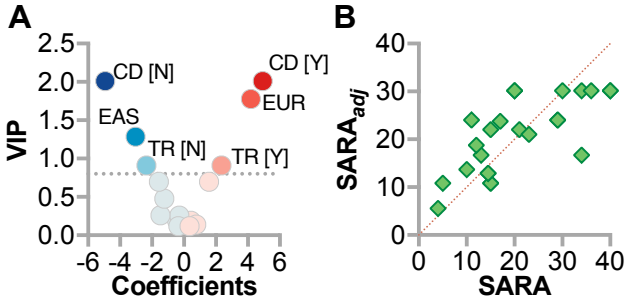

Figure 2 Multivariate regression model of SARA in SCAR16 patients. (A) Model coefficients and their variable importance in projection values from PLS regression analysis represented by scatter plot: CD = cognitive dysfunction, TR = tendon reflex, Y = yes, N = no, EAS = East Asian ancestry, EUR = European ancestry. The dotted line represents the VIP cutoff for inclusion into the final model. (B) Actual SARA by predicted SARA (SARA*adj*) plot of the PLS model of each SCAR16 patient

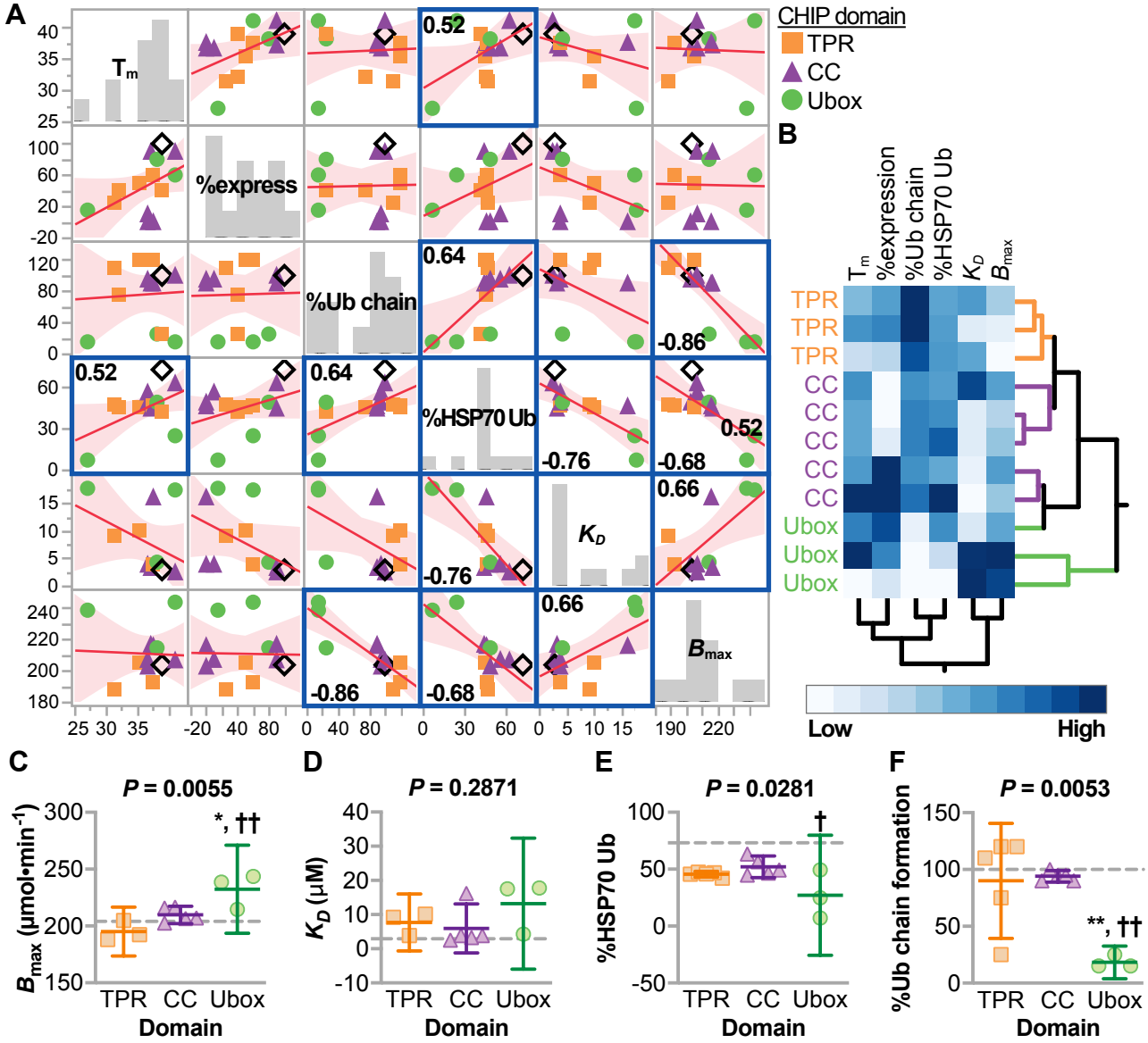

Figure 3 Domain-specific changes in the biochemistry of mutant CHIP proteins. (A) Multivariate correlations of the biochemical properties of mutant CHIP proteins represented by scatter plot and summarized by the best fit line and the 95% confidence interval; highlighted correlations indicate the Pearson correlation coefficients (ρ) < 10% FDR. (B) Unsupervised hierarchical clustering of the biochemical variables from mutant CHIP proteins identified by the domains harboring the mutation. (C - F) Biochemical variables stratified by the location of the mutation; data are represented by dot plot and summarized by the mean ± 95% confidence interval, analyzed by ANOVA: *,** indicate $P < 0.05$, or 0.001 compared to the CC domain, †, †† indicate $P < 0.05$, or 0.001 compared to the TPR domain via Tukey post hoc test. The dashed lines indicate levels measured with CHIP-WT levels.

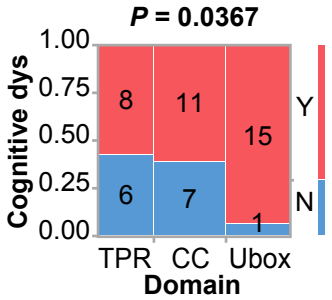

Figure 4 Ubox mutations are linked to a higher incidence of cognitive dysfunction. Contingency analysis of the location of disease alleles and the presence of cognitive dysfunction in SCAR16 patients, analyzed using Fisher's exact test.

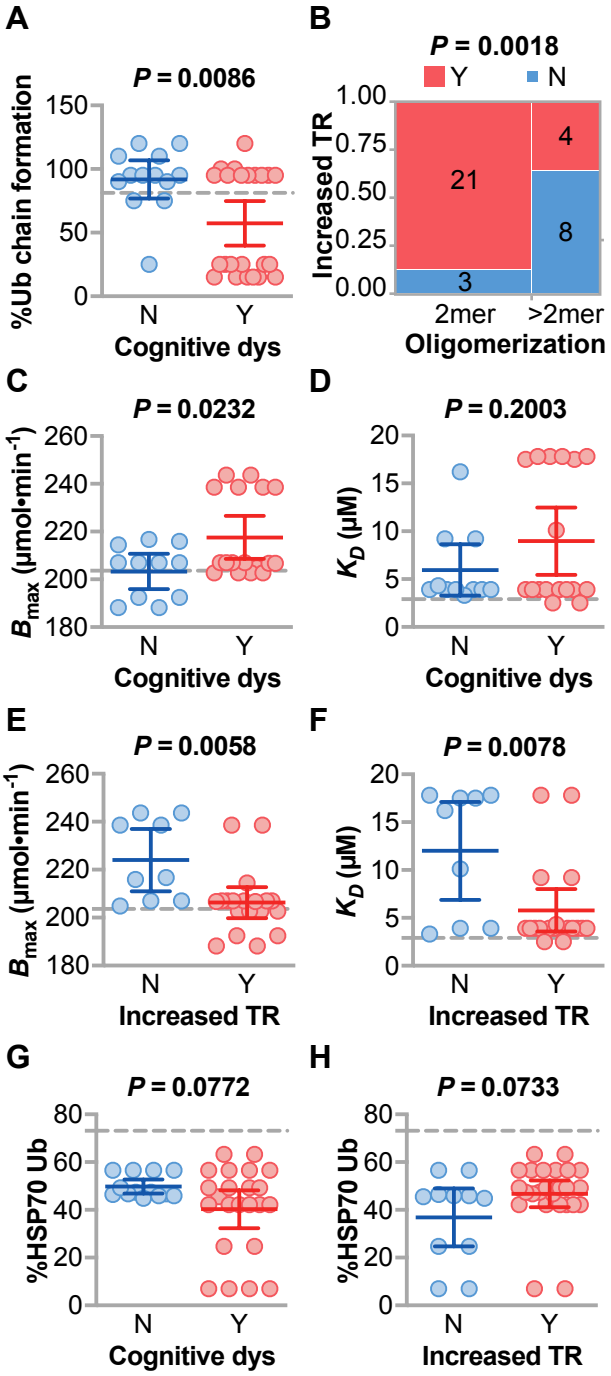

Figure 5 Changes in the biochemical properties of mutant CHIP proteins associate with clinical features of SCAR16. (A) Ubiquitin chain formation of mutant CHIP proteins stratified by the presence of cognitive dysfunction (dys) found in SCAR16 patients; data are represented by dot plot and summarized by the mean ± 95% confidence interval, analyzed by t test. The dashed line indicates the activity in wild-type CHIP. (B) Contingency analysis of the oligomerization pattern of mutant CHIP proteins and the presence of tendon reflex in SCAR16 patients, analyzed using Fisher's exact test. (C - H) the indicated biochemical variables of mutant CHIP proteins stratified by either the presence of cognitive dysfunction (dys) or increased tendon reflex (TR), found in SCAR16 patients; data are represented by dot plot and summarized by the mean ± 95% confidence interval, analyzed by t test. The dashed line indicates the activity of wild-type CHIP.

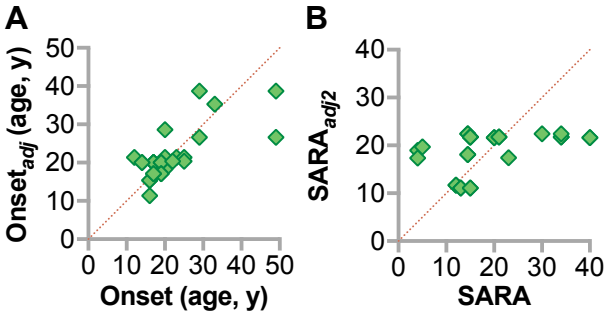

Figure 6 Regression model performance of age of onset and SARA as a function of the biochemical properties of mutant CHIP proteins. (A) Scatter plot of actual age of onset by predicted onset ($Onset_{adj}$) via the PLS model. (B) Scatter plot of actual SARA by predicted SARA ($SARA_{adj2}$) plot of the PLS model.

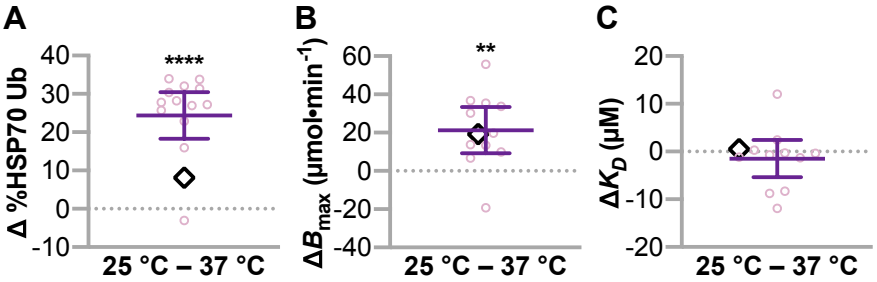

Figure 7 Temperature-dependent changes in the biochemical properties of mutant CHIP proteins. The absolute change (Δ) in (A) HSP70 ubiquitination, (B) B_{max}, and (C) K_D at 25 °C compared to 37 °C. Each mutant protein data are represented by dot plot and summarized by the mean ± 95% confidence interval: **, **** correspond to $P < 0.01$, 0.0001 via one sample t test versus the hypothetical value of 0 represented by the dotted line.

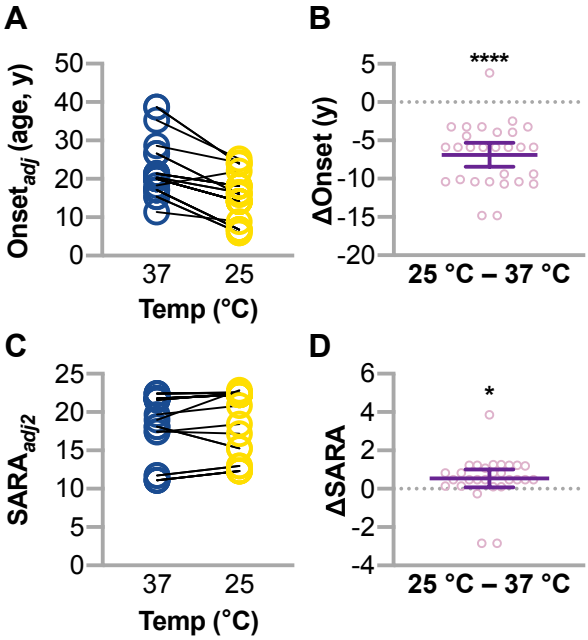

Figure 8 Predicted age of onset and SARA with temperature-dependent changes in mutant CHIP proteins. Model results of (A) ($Onset_{adj}$) or (C) $SARA_{adj2}$ via the PLS model at 37 °C and 25 °C with each allele represented by a before-after plot. The absolute change (Δ) in (B) $Onset_{adj}$ or $SARA_{adj2}$ at 25 °C compared to 37 °C. Each allele is represented by dot plot and summarized by the mean ± 95% confidence interval: *, **** correspond to $P < 0.05$, 0.0001 via one sample t test versus the hypothetical value of 0 represented by the dotted line.

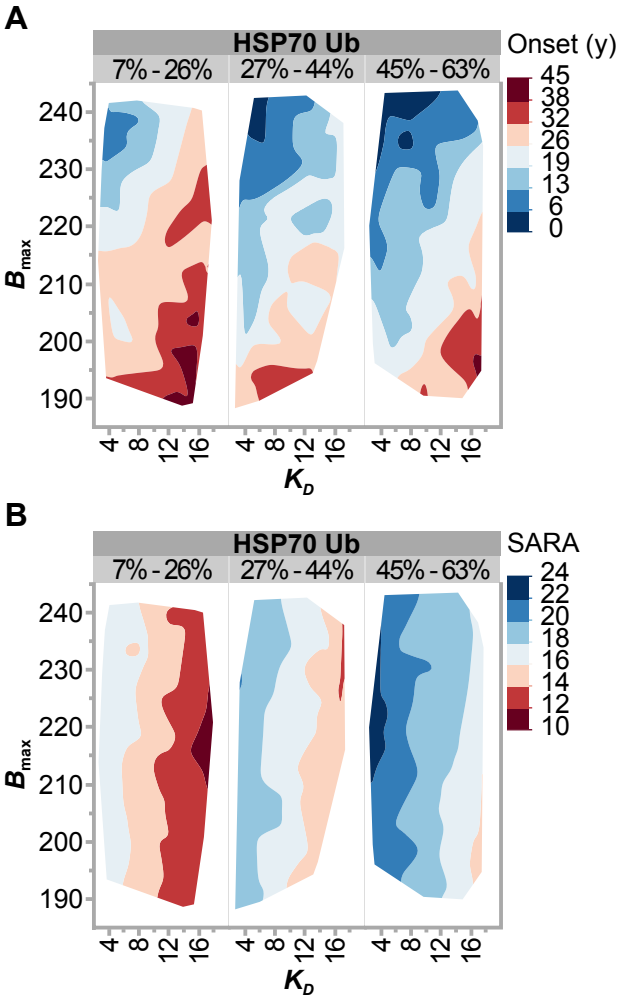

Figure 9 Simulation results of CHIP biochemical properties that improve disease phenotypes. Simulations of (A) age of onset and (B) SARA, represented as contour plots of B_{max} as a function of and K_D grouped by the percentage of HSP70 ubiquitination relative to wild-type CHIP.