

Disruptions in protein-protein interactions between HTT, PRPF40B, and MECP2 are involved in Lopes-Maciel-Rodan syndrome

Netrang Desai¹, Devon Stork²

¹ Juanita High School, Kirkland, Washington

² Harvard University, Cambridge, Massachusetts

SUMMARY

Lopes-Maciel-Rodan syndrome (LOMARS) is a rare and serious neurodevelopmental disorder caused by two compound heterozygous missense mutations in the Huntingtin gene (*HTT*). LOMARS is related to Rett syndrome, which is primarily caused by mutations in X-linked Methyl-CpG-Binding Protein 2 (*MECP2*), since it results in Rett-like neurological phenotypes and manifestations. While *HTT* and *MECP2* proteins are not known to directly bind to each other, it is plausible that other protein(s) enable and enhance this protein-protein interaction (PPI). This study reports the involvement of pre-mRNA Processing Factor 40 Homolog B (*PRPF40B*) in mediating the PPI between *HTT* and *MECP2*. By using coevolution data, we found that *PRPF40B* interacts with both *HTT* and *MECP2* through its group II WW domain, a modular protein domain that mediates protein-protein interactions. We also revealed that multiple missense and splice region variants in *PRPF40B* result in LOMARS and Rett-like phenotypes, suggesting that weakened interactions between mutant *PRPF40B* and wild-type *MECP2* are likely to be associated with the phenotypes. Furthermore, upon docking wild-type and mutant models of *HTT* with *PRPF40B*, we observed that the LOMARS-associated mutations significantly weaken *HTT*-*PRPF40B* interactions. We also performed similar docking experiments on *SIN3A* and *PRPF40A*, which is a protein that is highly related in structure and function to *PRPF40B* and observed that the mutations also weaken their interactions with *HTT*. Overall, this study demonstrates the significant role of *HTT*-*PRPF40B*-*MECP2* interactions in the development and progression of LOMARS and suggests the involvement of similar PPI disruption in Rett syndrome.

INTRODUCTION

Lopes-Maciel-Rodan Syndrome (LOMARS) is a rare autosomal recessive neurodevelopmental disease that develops during infancy (1, 2). Compound heterozygous missense mutations (p.P703L and p.T1260M) located in the Huntingtin gene (*HTT*) are associated with the onset of LOMARS (1, 2). A total of four patients have so far met the diagnostic criteria for LOMARS. An 18-year-old girl exhibited developmental regression at six months of age and developed seizures at eight months (1). Other phenotypes experienced

by the patient include severe intellectual disability, poor or absent speech, teeth grinding, spinal curvature, small feet/cold feet, gait instability, cerebral and cerebellar atrophy, and difficulty swallowing (1). Three siblings with age ranging from newborn to 12 years exhibited similar phenotypes along with myopia, delayed psychomotor development, spasticity, and abnormal muscle contraction (2).

LOMARS is reminiscent of Rett syndrome, which is a progressive neurodevelopmental disease primarily caused by mutations in the X-linked Methyl-CpG Binding Protein 2 (*MECP2*) (3–8). *MECP2* is a chromatin-associated protein that regulates transcription of methylated DNA by repressing gene expression (6, 7, 9). *MECP2* is composed of five major domains: methyl-CpG binding domain, N-terminal domain, inter-domain, transcription repression domain, and C-terminal domain (4). One way that *MECP2* regulates gene expression is through DNA methylation, which plays a key role in long-term gene silencing. (5). Through the N-terminal domain, *MECP2* recognizes methyl CpGs, which are sites in DNA where a phosphate group separates cytosine and guanine, in gene promoters and binds to non-methylated DNA (5). Thereafter, *MECP2* methylates DNA and represses transcription of targeted genes (5). Since the clinical manifestation of LOMARS is closely related to that of Rett syndrome (3–8), it is possible that *MECP2* is also involved in the pathology of LOMARS.

In order to initiate DNA methylation, *MECP2* recruits and binds to other transcriptional regulators and chromatin remodeling enzymes to silence gene expression (6). Out of multiple proteins associated with transcriptional regulation, the Huntingtin (*HTT*) protein is known to interact with *MECP2* in mouse and cellular models (6, 7). *HTT* is a ubiquitously expressed nuclear protein that binds to various transcription factors to regulate transcription (10). A particular mutation of a trinucleotide repeat (CAG) in *HTT* results in a polyglutamine (polyQ) expansion in the N-terminal domain of *HTT* (6, 10, 11). As a result, mutant *HTT* leads to the development of Huntington's disease (HD), a severe neurodegenerative disease. Recently, a study conducted by McFarland et al. reported the polyQ expansion to significantly increase the binding between *Htt* and *Mecp2* in the striatum of mouse models with mutant *Htt* in comparison to those with wild-type (WT) *Htt* (6). Moreover, in the striatal neurons of the mouse models, the polyQ expansion increased *Htt*-*Mecp2* interactions both in the nucleus and cytoplasm. As a result,

the mutant Htt increased the binding of Mecp2 to methylated DNA, which led to increased repression by Mecp2 and the underexpression of genes that are essential for the maintenance of neuronal health and brain development (6).

Given that the interaction between Htt and Mecp2 is significantly increased in the presence of the polyQ repeat, we suspect that the binding affinity between Htt and Mecp2 may substantially decrease in the absence of the polyQ repeat. McFarland et al. observed the interaction between the two proteins to be weaker in the striatum of the mouse models with only WT Htt (6). Likewise, they observed the interactions to be weaker in the cytoplasm of the striatal neurons with WT Htt (6). Moreover, an *in vivo* study by Roux et al. reported WT Htt levels to be significantly decreased in Mecp2 deficient mouse brains (7). Roux et al. found reduced Htt levels in the hippocampus, hypothalamus, and the cortex (7). Additionally, they found abnormal transcription rates and disruption of cellular process, such as protein trafficking, to be associated with decreased Htt levels in Mecp2 deficient brains (7). Together, this suggests that HTT and MECP2 are likely to interact *in vivo* without the polyQ expansion and in the absence of MECP2.

When the polyQ repeat is absent, we hypothesize that other proteins might mediate and enhance the PPI between WT HTT and MECP2. While several proteins are known to interact with HTT and MECP2, prior experiments have shown pre-mRNA processing factor homolog B (PRPF40B) to be a common interactor between the two proteins (5, 11). PRPF40B consists of two WW domains, including group I and II WW domains, and six FF domains. WW domains, which consist of two tryptophan residues that are 20–22 amino acids apart, are small protein domains that interact with proline-rich regions by forming three antiparallel β -sheets (5, 11). An FF domain, which is ~50–60 amino acids in length and composed of three α -helices, is known to be involved in transcriptional regulation as it is located in multiple splicing factors (29). One particular domain of PRPF40B, the group II WW domain (p.L133–p.D166), is known to bind to both HTT and MECP2 (5, 11). Existing studies have shown that weakened binding of Prpf40b with mutant Htt (polyQ expansion) and Mecp2 results in phenotypes that are closely related to that of LOMARS and Rett syndrome, making it an important protein to investigate. One study has identified several mutations on Mecp2, including C-terminal truncations and frameshift mutations, that abolish Mecp2's binding with the group II WW domain of prpf40b in mouse models (5). Likewise, another study has found the polyQ expansion in Htt to influence the binding between prpf40b and Htt (11). PRPF40B is also involved in pre-mRNA splicing and vesicle transport, which are some of the major processes that are likely to be disrupted in LOMARS and Rett syndrome (11). Given recent experimental evidence and PRPF40B's functional role, it seems possible that PRPF40B could be involved in LOMARS.

Here, we attempt to investigate the role of PRPF40B in mediating the PPI between HTT and MECP2. We

hypothesize that the mutations on HTT can abrogate binding between the group II WW domain of PRPF40B and HTT resulting in an overall disruption between HTT, PRPF40B, and MECP2 interactions. We first establish the interaction of PRPF40B with HTT and MECP2 by examining coevolution and the genome wide association data (GWAS), which identifies several mutations in PRPF40B that result in LOMARS and Rett syndrome phenotypes. Importantly, in order to evaluate our hypothesis and investigate the effects of the LOMARS-associated mutations on the HTT-PRPF40B interaction, we perform molecular docking between WT and mutant models of HTT and the group II WW domains of PRPF40B using the HADDOCK2.4 software. We found the interactions between HTT and the group II WW domains to be significantly weakened in mutant complexes. Overall, the results from our protein docking analysis reinforce our prediction that the LOMARS-associated mutations disrupt binding between HTT and PRPF40B, leading to a complete HTT-PRPF40B-MECP2 PPI disruption. Furthermore, similar docking experiments were performed to find potential effects of the LOMARS-associated mutations on the interactions of SIN3A and PRPF40A, a mammalian ortholog of the yeast pre-mRNA splicing factor Prp40, with HTT and MECP2. Our analysis revealed that the mutations weaken the interaction of PRPF40A and SIN3A with HTT, suggesting that they may also be involved in the pathology of LOMARS and Rett syndrome.

Missense Variants	
Genomic Position	Amino Acid Substitution
g.49631906C>T	p.A92V
g.49633074G>A	p.D137N
g.49634020A>G	p.D247A
g.49635149C>T	p.A351V
g.49635209C>T	p.T371I
g.49635395G>T	p.E399D
g.49641923G>C	p.V595M
g.49642281A>C	p.E644A
g.49643364C>T	p.R783W
g.49643695T>A	p.H795Q
Splice Region Variants	
NM_001031698.3:c.580+3G>A	
NM_001031698.2:c.1166+5G>C	
NM_001031698.2:c.580+7T>A	

Table 1: Ten missense and three splice region variants located in PRPF40B. gDNA and amino acid substitutions for each missense variant. We found one mutation (g.49633074G>A), shaded in green, in the group II WW domain encoding region of PRPF40B to result in an amino acid substitution of p.D137N. We also show the DNA substitutions for each splice region variant. Amino acid substitutions were absent for the splice region variants. All variants were mapped to GRCh38.13

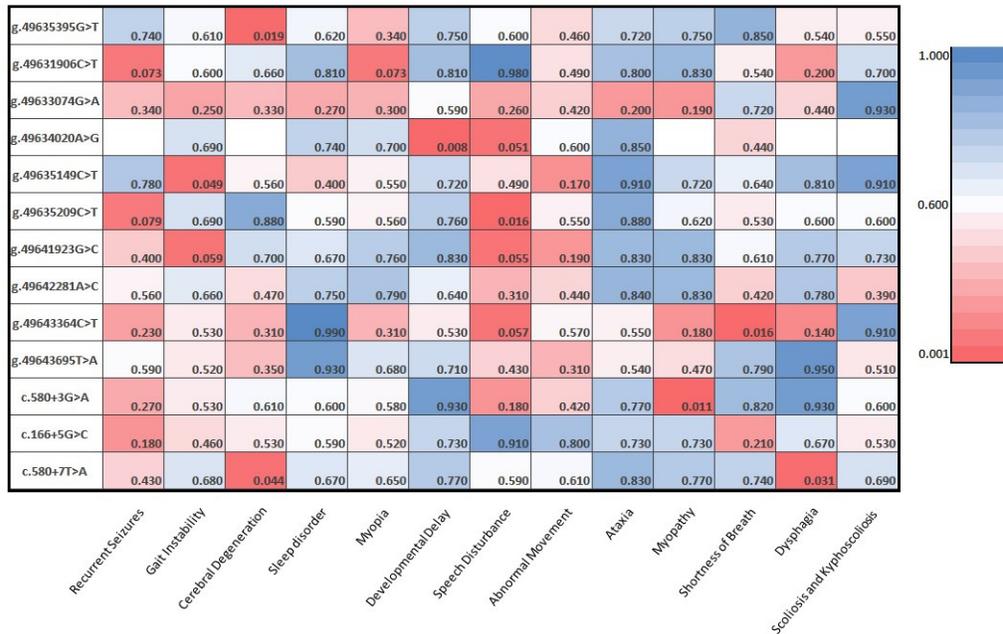


Figure 1: Mutations in PRPF40B are associated with LOMARS and Rett-like phenotypes. The heatmap illustrates the association between 10 missense and 3 splice region variants in *PRPF40B* and 13 phenotypes that are closely related to LOMARS and Rett syndrome. The *p* values, which are rounded to three decimal places, represent the genotype-phenotype association. The deep red color indicates a stronger association while the deep blue color indicates a weaker association between variants and phenotypes. Empty boxes indicate absence of phenotype in patients with a specific variant.

RESULTS

Differentiation between the compound heterozygous missense mutations

Compound heterozygous missense mutations located on *HTT* are likely responsible for the onset of LOMARS (1, 2). We obtained mutational data from gnomAD v2.1.1 and ClinVar to examine the two mutations at an individual level (16). The first mutation, g.3133374C>T, which has a total population frequency of 0.00008055%, results in a p.P703L missense mutation in *HTT* and is classified as pathogenic in ClinVar (Accession ID: VCV000417742.1, Variation ID: 417742). The second mutation, g.3162034C>T, which has a population frequency of 0.008306%, results in a p.T1260M missense mutation in *HTT* and is classified as having conflicting interpretations of pathogenicity in ClinVar (Accession ID: VCV000417743.2, Variation ID: 417743). Since the p.T1260M mutation is found at a higher level in a “healthy” population, it is less likely to be deleterious than the p.P703L mutation. However, the effects of p.T1260M may increase when present *in trans* with p.P703L (2). This combinational scenario can increasingly impair the WT function of *HTT* and potentially impact other *HTT* interacting proteins, leading to LOMARS and Rett-like phenotypes (2).

Co-evolution analysis between *HTT* and *MECP2* indicates *PRPF40B* is a common binding partner between the two proteins

HTT and *MECP2* exhibit weak interactions without the presence of polyQ expansion (6). Therefore, it is possible

for other proteins to mediate interactions between *HTT* and *MECP2*. Through the MatrixMatchMaker database of co-evolving proteins (MMM-D), we were able to find 78 proteins that are predicted to interact with *HTT*, and 27 proteins that are predicted to interact with *MECP2* (see materials and methods). The database uses three different criteria for predicting coevolution between proteins: known interaction, unknown interaction, and MMM score. The typical cutoff value for MMM scores is ≥ 5 (12). When the MMM score is high (≥ 5), the coevolution between proteins is stronger, which indicates a high likelihood of PPI (12). For known protein interactions, scores below five indicate weak interactions between two or more proteins. While the low scores do not necessarily indicate an absence of interactions between proteins, they suggest that the strength of the physical interactions (residue-residue) at the binding interfaces is weaker (12).

We found 31 *HTT* interacting proteins to have a MMM score greater than or equal to 6, while the remaining 47 *HTT* interacting proteins had a score in the range of 1–5. In comparison, we found all 27 *MECP2* interacting proteins to have a score at or below 5. In order to find out whether *MECP2* and *HTT* share common interacting proteins or not, we compared the MMM score of 105 proteins (see materials and methods). Consequently, we located three known proteins that interact with both *MECP2* and *HTT*: *SIN3A*, *PRPF40B*, and *SP1*. The *SIN3A* (MMM = 5) and *PRPF40B* (MMM = 6) proteins were more likely to coevolve with *HTT* than *SP1* (MMM = 5). On the other hand, in the coevolution dataset of *MECP2*, all the three proteins had a MMM score of

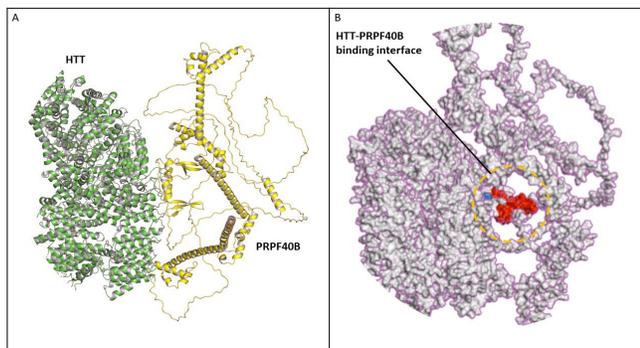


Figure 2: WT HTT and PRPF40B interaction. (A) The green protein is HTT and the yellow protein is PRPF40B. This panel depicts the overall WT HTT-PRPF40B interaction. (B) This panel depicts the binding interface between WT HTT and PRPF40B (orange circle). The blue color represents the defined active residue of HTT (p.P703), and the red color represents the active residues of PRPF40B (p.L133–p.D166). Protein structures created using PyMOL.

5. Since the MMM score of SP1 was lower in the coevolution dataset of HTT, it suggests that the interaction between SP1 and HTT might be significantly weaker than that of SIN3A and PRPF40B. Therefore, SIN3A and PRPF40B are more likely to bind and interact with both HTT and MECP2.

Co-evolutionary connections between protein families also highlight functional relationships between coevolving proteins (12). Both SIN3A and PRPF40B participate in transcriptional regulation, which may indicate a close functional relationship with MECP2 and HTT. SIN3A inhibits chromatin binding, participates in transcriptional coactivation and corepressor, and enables protein binding (17, 18). Likewise, the PRPF40B protein is known to mediate protein-protein binding and regulate pre-mRNA splicing (11, 17). Existing experimental evidence has established Prpf40b as an important binding partner of Htt and Mecp2 in mouse models. A study found that a C-terminal truncation of Mecp2[1–471] by 15 amino acids, large frameshift deletions, and an internal deletion result in developmental regressive phenotypes, including moderate mental retardation, by abrogating binding with Prpf40b (5). Additionally, disruptive interactions between mutant Htt (polyQ repeat) and Prpf40b result in subtle differences in their binding and are possibly associated in HD pathology (11). Altogether, both the MMM scores and existing experimental evidence suggest that the HTT and MECP2 proteins bind and interact with PRPF40B.

Mutations in PRPF40B are associated with LOMARS and Rett like-phenotypes

An *in vivo* experiment conducted by Buschdorf et al. has identified a large cluster of Rett causing frameshift mutations in Mecp2 that abrogates binding between the group II WW domain of Prpf40b and Mecp2 (5). These mutations are located in the proline-rich region of mutant Mecp2[1–400] (5). Mutant Mecp2[1–438], which has a Rett syndrome causing

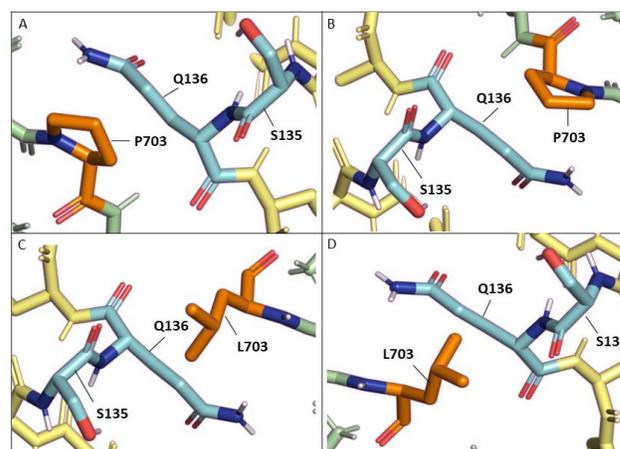


Figure 3: LOMARS-causing mutations disrupt interaction between HTT and PRPF40B. The dark blue color represents nitrogen atoms, the red color represents oxygen atoms, and the white color represents hydrogen atoms. The green protein is HTT, and the yellow protein is PRPF40B. Within the active residues of PRPF40B, p.S135 and p.Q136 (light blue) are specifically interacting with residue 703 (orange) of HTT. (A, B) The WT HTT (p.P703) interacts with p.S135 and p.Q136 by forming close bonds. (C, D) However, the mutant HTT (p.L703) weakens the normal bonding by creating several clashes and repelling the residue-residue interactions. Protein structures created using PyMOL.

frameshift mutation at residue 436, restricts Mecp2 from binding to the group II WW domain *in vitro* (5). Additionally, the study reports a C-terminal truncation, along with two mutations in Mecp2[325–486], and large in-frame deletion to result in mental retardation (5). The results from this experiment indicate that mutations in Mecp2 disrupt Mecp2's binding with the group II WW domain of Prpf40b and result in Rett-like phenotypes.

As Mecp2 mutations abrogate binding to Prpf40b, we reasoned that the mutations in PRPF40B could also be associated with Rett-like phenotypes. In order to evaluate our prediction, we used the UKBiobank TOPMed-imputed PheWeb database to locate mutations in PRPF40B. We found ten missense (Table 1) and three splice region (Table 1) mutations to be located at several regions in PRPF40B (19). The mutations were associated with 13 phenotypes (Figure 1). Out of those 13 phenotypes, 8 of them were previously reported in patients diagnosed with LOMARS, including recurrent seizures, gait instability, sleep disorder, myopia, speech disturbance, shortness of breath, dysphagia, scoliosis, and kyphoscoliosis. The remaining five phenotypes, which are cerebral degeneration, developmental delay, abnormal movement, ataxia, and myopathy, are closely related to the clinical synopsis and pathology of LOMARS and Rett syndrome. Furthermore, we found one specific missense mutation (g.49633074G>A) to be located in the group II WW domain encoding region of PRPF40B. Further examination of the mutation revealed its association with two additional phenotypes: abnormalities of the jaw ($p = 0.014$) and walking

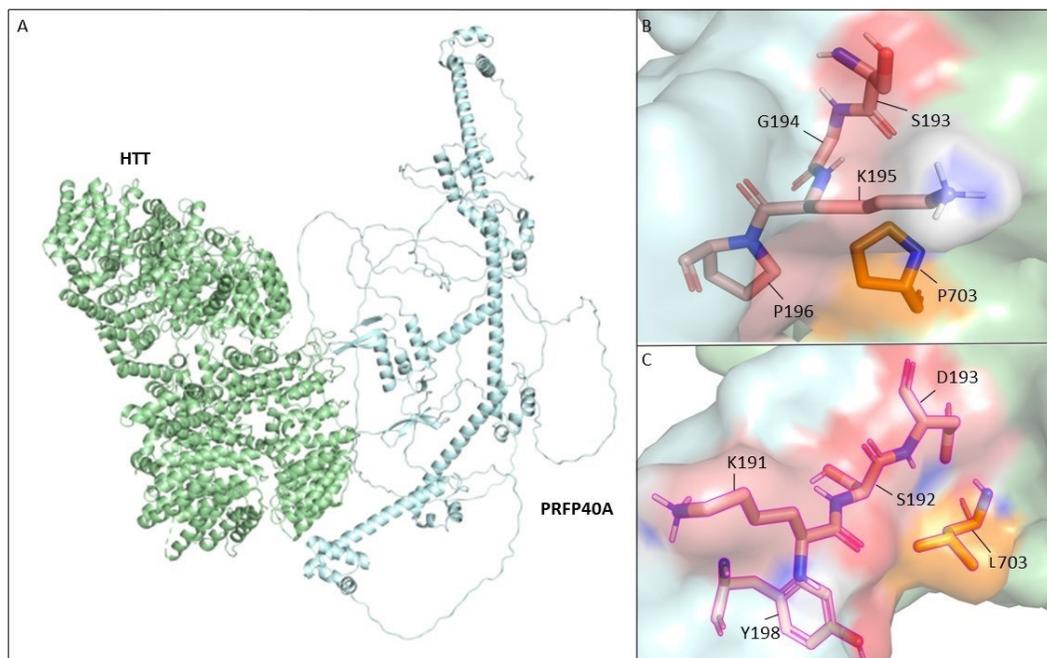


Figure 4: LOMARS-causing mutations disrupt interaction between HTT and PRPF40A. The dark blue color represents nitrogen atoms, the red color represents oxygen atoms, and the white color represents hydrogen atoms. (A) WT HTT (green) interacting with PRPF40A (light blue). (B) In the WT complex, the p.S193, p.G194, p.K195, and p.P196 active residues (pink) of PRPF40A are interacting with the residue p.P703 (orange) of WT HTT. (C) In the double-mutant complex, the p.K191, p.S192, p.D193, and p.Y198 active residues of PRPF40A are interacting with the residue p.L703 of mutant HTT. The surface and stick view depicts the interaction interfaces in panels B and C to eliminate unnecessary surrounding residues. Protein structures created using PyMOL.

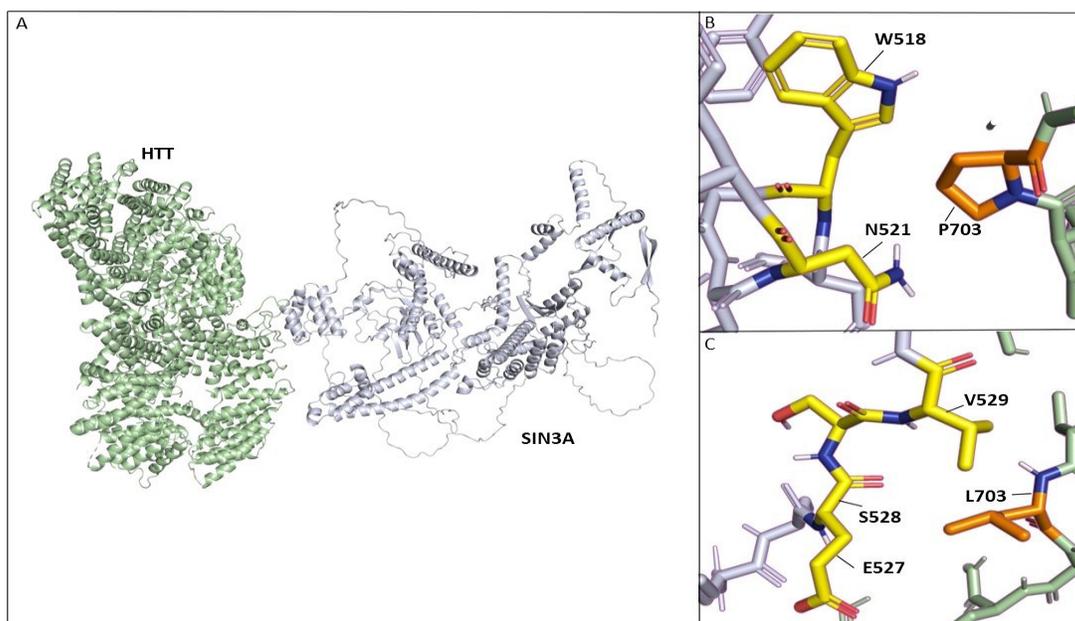


Figure 5: LOMARS-causing mutations disrupt interaction between HTT and SIN3A. The dark blue color represents nitrogen atoms, the red color represents oxygen atoms, and the white color represents hydrogen atoms. (A) WT HTT (green) interacting with SIN3A (light blue). (B) In the WT complex, the residues p.W518 and p.N521 (yellow) of SIN3A are interacting with the residue p.P703 of WT HTT. (C) In the double-mutant complex, the residues p.E527, p.S528, and p.V529 of SIN3A are interacting with p.L703 of mutant HTT. Protein structures created using PyMOL.

difficulties ($p = 0.041$). While the p values (see materials and methods) for majority of the phenotypes are greater than 0.05, it is suggestive that mutations in PRPF40B cause LOMARS and Rett-like phenotypes by weakening or abrogating binding with MECP2. While mutations abrogating the interaction between Mecp2 and Prpf40b result in Rett syndrome, it is possible that mutations affecting the interaction between HTT and PRPF40B may also result in similar phenotypes.

Structural analysis of protein-protein interaction between PRPF40B and HTT

Since HTT is important in the pathology of LOMARS, understanding its interaction with the group II WW domain of PRPF40B could give insights into the key binding areas between the two proteins. In order to analyze the interactions between HTT and PRPF40B, we used the High Ambiguity Driven protein-protein Docking software 2.4 (HADDOCK 2.4). HADDOCK uses experimental data from Crystallography and NMR system (CNS) as restraints and then applies energetics and shape complementarity to dock biomolecules together (13, 14). We obtained the full-length structure of HTT from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (PDB ID: 6RMH) and the predicted full-length structure of PRPF40B (UniProt ID: Q6NWY9) from AlphaFold (20, 21). We first docked WT HTT with the group II WW domains of PRPF40B (**Figure 2A**). As a general docking protocol, HADDOCK requires users to define the Ambiguous Interaction Restraints (AIRs), which are classified as active and passive residues, before docking. Although p.T1260 is the site of one of the LOMARS-associated mutations, we suspect that p.T1260 will be absent at or near the HTT-PRPF40B binding interface due to its low level of pathogenicity. Therefore, we defined p.P703 as the active residue for HTT and p.L133–p.D166 (group II WW domain) as the active residues of PRPF40B (**Figure 2B**). For post-docking summary and results, HADDOCK generated the top 10 predicted “clusters”, each consisting of 4 predicted structures of protein complexes of the HTT-PRPF40B interaction. The clusters were ranked based on their HADDOCK and z scores. A cluster is deemed reliable and accurate if it has a larger negative HADDOCK and z score. We selected the first, or top ranking, WT cluster based on its low HADDOCK (~ -35.5) and z score (-1.9). Within the WT cluster, we observed the four predicted structures to be highly similar (RMSD < 0.2). Further visual analysis of the WT cluster revealed that p.T1260 was, indeed, not located at or near the active sites of HTT and PRPF40B, suggesting that it may not solely disrupt HTT-PRPF40B interaction.

In order to observe differences in binding in the presence of LOMARS-associated mutations in HTT, we docked three separate models of mutant HTT with PRPF40B: i) HTT p.P703L, ii) HTT p.T1260M, and iii) HTT p.P703L and p.T1260M. We defined the active residues for model i as p.L703, model ii as p.M1260, and model iii as both p.L703 and p.M1260. Here, we defined p.M1260 as the active residue to observe its impact on binding between HTT and PRPF40B. We selected

the first cluster of mutant HTT and PRPF40B based on the low HADDOCK (~ -30.5) and z scores (~ -1.8) of the cluster. Within each cluster, we observed the four predicted protein complexes to be also similar (RMSD < 0.26). Despite defining p.M1260 as the active residue in models ii and iii, we observed the residue was not located near the binding interface, which suggests that p.M1260 may not be directly interacting with the group II WW domain of PRPF40B.

We used the predicted structural complexes of HTT and PRPF40B from HADDOCK to measure binding affinity in Protein Binding Energy Prediction Software (PRODIGY) (22,23). PRODIGY uses a robust linear regression algorithm of residue-residue contacts (interfacial contacts) and 3D structures of protein complexes to predict binding affinity (23). The Gibbs free energy (ΔG), measured in kcal mol⁻¹, and the equilibrium dissociation constant (K_d), measured in molar units (M), describe the binding affinity. A high ΔG and K_d represent a stronger binding affinity between proteins in a complex. We observed the WT cluster to possess stronger binding affinity ($\Delta G = -11.3$ kcal mol⁻¹, $K_d = 5.4E-09$ M) (**Figure 3A–B**) when compared to other mutant structures of the HTT-PRPF40B complex. The HTT-PRPF40B complex with HTT p.P703L had slightly weaker binding affinity ($\Delta G = -9.9$ kcal mol⁻¹, $K_d = 5.7E-08$ M) (**Figure 3C–D**), while the complex with HTT p.P703L and p.T1260M had a much weaker binding affinity ($\Delta G = -7.4$ kcal mol⁻¹, $K_d = 4.0E-06$ M). In contrast, the binding affinity of HTT p.T1260M ($\Delta G = -10.2$ kcal mol⁻¹, $K_d = 3.1E-08$ M) was nearly similar to the binding affinity of the WT cluster. These findings indicate that HTT p.P703L has a greater effect on the ability of HTT to bind to PRPF40B than HTT p.T1260M. Interestingly, we observed the binding affinity between HTT and PRPF40B to be significantly weaker under the presence of both LOMARS-associated mutations (double mutants), suggesting that a combinational effect is far more disruptive. While individual mutations are likely to reduce the binding affinity to some degree, a combination of p.P703L and p.T1260M may completely abrogate the binding between HTT and PRPF40B.

Structural analysis of protein-protein interactions between PRPF40A/SIN3A and HTT

In order to assess whether the docking results are specific to PRPF40B, we performed similar experiments on PRPF40A and SIN3A proteins which are both known to interact with HTT. We obtained the predicted full-length structures of PRPF40A (UniProt ID: Q9R1C7) and SIN3A (UniProt ID: Q60520) from AlphaFold (21). While PRPF40A also contains six domains (group I and II WW and six FF domains), its group II WW domain (p.L181–p.E214) binds specifically to HTT (11). In order to observe the effects of the LOMARS-associated mutations on HTT-PRPF40A interaction, we docked two separate HTT models with the group II WW domain of PRPF40A: WT HTT and the double mutant HTT. Considering that a combinational effect of mutations has the most damaging effect on protein interactions, we chose to focus only on the double mutant

HTT model. In HADDOCK, we defined p.L181–p.E214 as the active residues of PRPF40A. For the WT interaction, we chose the top ranked cluster with a low HADDOCK (-48.9 +/- 1.7) and z score (-2.3). Similarly, for the interaction with double mutant HTT, we chose the top ranked cluster with a low HADDOCK (-33.6 +/- 6.2) and z score (-1.7). Further analysis of each of the clusters in PRODIGY revealed that the WT complex ($\Delta G = -9.8$ kcal mol⁻¹, $K_d = 6.4E-08$ M) (Figure 4A–B) possesses stronger binding affinity than the complex with double mutant HTT ($\Delta G = -8.6$ kcal mol⁻¹, $K_d = 5.3E-07$ M) (Figure 4C). This suggests that under the presence of both the LOMARS-associated mutations, the interaction between the group II WW domain of PRPF40A and HTT is weakened.

Furthermore, we repeated the same protein docking experiment with SIN3A. Since this protein also has an MMM score of five and coevolves with HTT, we predicted that LOMARS-associated mutations would have similar effects on its interaction with HTT. Little is known about the interaction between SIN3A and HTT, so we defined the surface residues of SIN3A as passive. This allowed HADDOCK to take all of the surface residues into consideration while disregarding the ones that are not likely to belong to the interaction interface. Upon performing docking, we chose the top ranked WT cluster with a low HADDOCK (-61.6 +/- 0.6) and z score (-0.9), and also the top ranked cluster of double mutant HTT with a low HADDOCK (-53.6 +/- 6.3) and z score (-1.0). We found the complex with double mutant HTT ($\Delta G = -6.6$ kcal mol⁻¹, $K_d = 1.5E-05$ M) (Figure 5C) to weaken the binding between SIN3A and HTT when compared to the WT complex ($\Delta G = -7.4$ kcal mol⁻¹, $K_d = 3.5E-06$ M) (Figure 5A–B). Together, the results from PRODIGY suggest that the double LOMARS-associated mutations on HTT are likely to disrupt the interaction between SIN3A and HTT.

DISCUSSION

In this study, we investigated the relationships of protein-protein interactions between HTT, PRPF40B, and MECP2. We initially found the interaction between PRPF40B and HTT/MECP2 through the MMM-D database of co-evolving proteins and existing experimental evidence (5, 11). The high MMM scores of PRPF40B suggest that PRPF40B coevolves with HTT and MECP2 and, thus, interacts with them. The high MMM score can be explained by two concepts: common phylogenetic diversity and correlated rates of evolution. One explanation is that the interacting proteins (PRPF40B, HTT, and MECP2) are likely to follow a main phylogenetic signal and remain conserved in eukaryotes, resulting in high rates of coevolution and strong interactions (12). Another possibility of the high MMM score is that the proteins are likely to possess correlated rates of evolution over a larger evolutionary period (12). Together, these concepts could explain why the interactions between PRPF40B, HTT, and MECP2 are strong. Additionally, the GWAS data from UK Biobank TOPMed-Imputed PheWeb supports our initial

coevolution findings that suggest PRPF40B interacts with MECP2. We found the mutations in PRPF40B to result in phenotypes associated with LOMARS and Rett syndrome. Our GWAS findings suggest that mutations at several regions in PRPF40B, including one missense mutation in the group II WW domain, can perhaps impair PRPF40B's binding with MECP2 and lead to similar deteriorating phenotypes. There is a favorable likelihood of correlation between mutations and reduced interactions between PRPF40B and MECP2. More importantly, our docking and binding affinity results from HADDOCK and PRODIGY reveal details about the binding interface between HTT and the group II WW domain of PRPF40B. While previous studies have used mouse models to observe interactions between HTT and PRPF40B, we utilized the predicted full-length structures of human PRPF40B and HTT. We found the mutant HTT-PRPF40B complexes to possess weaker binding affinity than the WT complex. This supports our initial hypothesis that mutations in HTT lead to the development of LOMARS by disrupting HTT's interaction with PRPF40B and MECP2.

Furthermore, our results from PRODIGY show that all HTT-PRPF40B complexes with mutant structures of HTT are responsible for weakening the interaction to some extent. Consistent with our initial exploration of the mutations from gnomAD database, p.M1260 is possibly less pathogenic when compared to p.L703 as p.M1260 is far away from the binding interface. We also found a synergistic effect of p.L703 and p.M1260 to reduce the binding affinity between HTT and PRPF40B substantially. While p.L703 can weaken the binding at the interaction interface, p.M1260 can lead to conformational changes and alter the structure of HTT slightly. Together, these two mutations can significantly damage HTT and disrupt its interaction with PRPF40B. As a result, these disruptive mutations likely impair PRPF40B's function to regulate and mediate PPI between HTT and MECP2, stimulating the onset and development of LOMARS. PRPF40B's function can be disrupted by two possible mechanisms. First, the mutations can gradually deteriorate the binding between the group II WW domain of PRPF40B and HTT over time and make PRPF40B dysfunctional, which in turn could restrict MECP2 from interacting with HTT. Second, the mutations could disrupt the activity of either HTT or MECP2 and lead to toxic consequences, including transcriptional dysregulation. These aberrant interactions could also potentially lead to unwanted consequences such as the formation of protein aggregates.

The weakened interactions between HTT, PRPF40B, and MECP2 could potentially lead to harmful consequences, including dysregulation of genes that are directly controlled by MECP2 and involved in brain development and maintenance of neuronal health. Brain-derived neurotrophic factor gene (*BDNF*), which is under the control of MECP2, is dysregulated in Rett syndrome. BDNF is known to support neuronal survival and engages in axonal and dendritic differentiation, maturation, and synaptic plasticity in CNS. (6–8). In Rett

syndrome patients, loss of function of MECP2 has been known to lead to downregulation of *BDNF* and results in neurological dysfunction (6–7). Other genes controlled by MECP2 include glutamate receptor delta-1 subunit (*GRID1*), Ubiquitin-protein ligase E3A (*UBE3A*), Membrane Palmitoylated Protein 1 (*MPP1*), and Guanidinoacetate N-methyltransferase (*GAMT*) (8), which are also known to be dysregulated in Rett syndrome due to a loss of function of MECP2 (8). Therefore, it is possible that these genes might also be dysregulated in LOMARS due to the weakened interactions between HTT, PRPF40B, and MECP2.

Besides the importance of HTT, PRPF40B, and MECP2 interactions, interactions involving PRPF40A and SIN3A are also significant. Unlike other signaling and PPI mediating proteins, PRPF40A is unique in that its structure and function are closely related to that of PRPF40B. Along with being involved in pre-mRNA splicing, PRPF40A plays a vital role in neuronal trafficking through cytoskeletal organization (24). While PRPF40A's interaction with HTT is necessary for normal functioning of neurons, the LOMARS-associated mutations might significantly disrupt the interaction, as shown by our docking results. This disruption could likely negatively affect cytoskeletal-based transport and organelle trafficking in certain domains of neurons. Similarly, we observed the interaction between SIN3A and HTT to be weakened due to the LOMARS-associated mutations. SIN3A is an important transcriptional regulator as it participates in mediating PPI. SIN3A's function could also be disrupted by HTT mutations leading to transcriptional dysregulation which could, in turn, dysregulate the levels of proteins that are necessary for proper neuronal function. Overall, the possible interaction of PRPF40A and SIN3A with HTT and MECP2 could be disrupted by the LOMARS-associated mutations, leading eventually to deteriorating phenotypes.

Existing research on LOMARS has been focused on the genotype-phenotype effects of the two compound heterozygous mutations on *HTT* (1, 2). However, this study attempts to provide novel insights into the role of protein-protein interactions in the pathology of LOMARS. The findings in this study could not only advance current understanding of LOMARS but could also provide insights into the importance of similar protein-protein interactions in Rett syndrome. We believe our protein docking results could be improved by utilizing more realistic models of PRPF40B, PRPF40A, and SIN3A proteins. While AlphaFold provides accurate predictions of the structures of proteins, it currently only predicts the structure of a single protein chain with WT sequence (21). Future changes to AlphaFold could result in the modelling of proteins that are more realistic in nature. Additionally, we can refine our docking experiments by obtaining more information on the binding interfaces of proteins, particularly for the SIN3A-HTT interaction. Along with additional valid experimental data, intermolecular NOE and RDCs could also strengthen our protein docking experiments (25). Moreover, HADDOCK does not allow users to submit

parameter files for modified residues or bases, since it utilizes topallhdg format convention for CNS amino acid parameters (13). Future modifications to HADDOCK could allow users to submit all modifications made to a protein structure at the residue level, including topologies and parameters of protein molecules (13).

While this study shows that the group II WW domain of PRPF40B is essential for binding with HTT and MECP2, further research shall be directed towards understanding the importance of other domains in PRPF40B, including group I WW domains (p.G92–p.V125) and six FF domains (p.R276–p.V682) (17). Previous studies have found that the group I WW domains interact with HTT (11). Indeed, through our GWAS analysis, we found that one missense variant was located in the group I WW domain encoding region of *PRPF40B* and five other missense mutations located in the FF domain encoding region of *PRPF40B*. The variants in these domains are known to result in LOMARS and Rett-like phenotypes. Thus, determining if group I WW domains and FF domains are responsible for mediating interactions between HTT and MECP2 would help therapeutics to be targeted towards those domains. It would be also beneficial to investigate the role of other WW domain containing proteins in LOMARS. Few WW domain containing proteins that interact with HTT include FBP21 (26, 28), SETD2 (27), and FBP30 (26, 28). All of these proteins are known to consist of group I and II WW domains and mediate PPI between proteins. Additional studies will be required to determine whether these proteins interact with MECP2 or not and if they are involved in the pathology of LOMARS and Rett syndrome.

MATERIALS AND METHODS

MatrixMatchMaker database

We performed the coevolution analysis for HTT and MECP2 by utilizing the MatrixMatchMaker database (MMM-D) of co-evolving proteins. MMM-D uses coevolution data to predict interactions between proteins, which can reveal functional relationships and interaction networks important for cellular function (12). We used the MMM score to determine the interaction of PRPF40B, PRPF40A, and SIN3A with HTT and MECP2. The database calculates the MMM score by finding the largest common submatrix between distance matrices of two protein families (12). The database measures coevolution between families of distinct size, including both paralogs and orthologs, and predicts multiple coevolving partners. The database includes around 60,000 known protein interactions and 5,000,000 predicted interactions (12).

UKBiobank TOPMed-imputed PheWeb

The UKBiobank TOPMed-imputed PheWeb is a GWAS/PheWas database developed through the consortium of UKBiobank and TOPMed. The PheWeb consists of genotype-phenotype data for about 57 million variants (19). From a group of 77,465 cases and 328,796 controls, 10 missense and 3 splice region variants in *PRPF40B* were selected based on

their association with LOMARS and Rett-like phenotypes and predicted loss-of-function nature.

HADDOCK2.4

We used the HADDOCK2.4 protein docking software to dock WT and mutant HTT structures with PRPF40B, PRPF40A, and SIN3A. HADDOCK is a unique software since it uses bioinformatical and computational methods, in combination with data from NMR and X-ray crystallography, to accurately predict the structures of PPI complexes (13–15). HADDOCK, unlike existing molecular docking servers such as ClusPro, GRAMM-X, PatchDock/FireDock, Hex, and RosettaDock, allows full structural flexibility to proteins when forming a complex (13).

AIRs is classified as active or passive residues. Active residues are more likely to be involved in the interacting interface of proteins and show a significant chemical shift perturbation upon complex formation. Passive residues can be involved in protein interactions but are not important and show a less significant chemical shift perturbation (13, 14).

The HADDOCK score of each cluster is determined by calculating the average score of the four predicted complexes. The software calculates the HADDOCK score by using the following equation:

$$\text{HADDOCK}_{\text{score}} = 1.0 * E_{\text{vdw}} + 0.2 * E_{\text{elec}} + 1.0 * E_{\text{desol}} + 0.1 * E_{\text{air}}$$

where E_{vdw} represents the intermolecular van der Waals energy, E_{elec} represents the intermolecular electrostatic energy, E_{desol} represents an empirical desolvation energy term, and E_{air} represents the Ambiguous Interaction Restraint (AIR) energy (13, 14).

The z score is the number of standard deviations away from the mean of a given HADDOCK score of protein clusters (13,14).

PRODIGY

We used PRODIGY to measure binding affinity between WT and three mutant models of HTT with PRPF40B. Likewise, we used PRODIGY to measure binding affinity between WT and mutant models of HTT with PRPF40A and SIN3A. PRODIGY calculates the Gibbs free energy (ΔG) by using the following equation:

$$\Delta G = -0.09459 \text{ ICs}_{\text{charged/charged}} - 0.10007 \text{ ICs}_{\text{charged/apolar}} + 0.19577 \text{ ICs}_{\text{polar/polar}} - 0.22671 \text{ ICs}_{\text{polar/apolar}} + 0.18681 \% \text{NIS}_{\text{apolar}} + 0.3810 \% \text{NIS}_{\text{charged}} - 15.9433$$

where $\text{ICs}_{\text{xxx/yyyy}}$ represents the number of interfacial contacts found on the binding interface of HTT and PRPF40B (22,23).

PRODIGY calculates the dissociation constant by using the following equation:

$$\Delta G_{\text{fe}} = RT \ln K_d$$

where ΔG_{fe} is the predicted free energy, R represents the ideal gas constant (kcal K⁻¹ mol⁻¹), and T represents the temperature (K) (22, 23). We set the temperature to 298.15K (25°C) for binding affinity analysis.

ACKNOWLEDGMENTS

We would like to acknowledge gnomAD, MMM-D, ClinVar, UK Biobank TOPMed-Imputed PheWeb, HADDOCK2.4, and PRODIGY for using their databases and software. The FP7 WeNMR (project# 261572), H2020 West-Life (project# 675858), the EOSC-hub (project# 777536) and the EGI-ACE (project# 101017567) European e-Infrastructure projects are acknowledged for the use of their web portals, which make use of the EGI infrastructure with the dedicated support of CESNET-MCC, INFN-PADOVA-STACK, INFN-LNL-2, NCG-INGRID-PT, TW-NCHC, CESA, IFCA-LCG2, UA-BITP, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan and the US Open Science Grid.

Additionally, we thank Deyri Garcia for reading our paper and providing constructive feedback.

Received: August 26, 2021

Accepted: October 22, 2021

Published: July 11, 2022

REFERENCES

1. Rodan, Lance H., *et al.* "A Novel Neurodevelopmental Disorder Associated with Compound Heterozygous Variants in the Huntingtin Gene." *European Journal of Human Genetics*, vol. 24, no. 12, 2016, pp. 1826–27. Crossref, doi:10.1038/ejhg.2016.74.
2. Lopes, Fátima, *et al.* "Identification of Novel Genetic Causes of Rett Syndrome Like phenotypes." *Journal of Medical Genetics*, vol. 53, no. 3, 2016, pp. 190–99. Crossref, doi:10.1136/jmedgenet-2015-103568.
3. Amir, Ruthie E., *et al.* "Rett Syndrome Is Caused by Mutations in X-Linked MECP2, Encoding Methyl-CpG-Binding Protein 2." *Nature Genetics*, vol. 23, no. 2, 1999, pp. 185–88. Crossref, doi:10.1038/13810.
4. Liyanage, Vichithra R. B., and Mojgan Rastegar. "Rett Syndrome and MeCP2." *NeuroMolecular Medicine*, vol. 16, no. 2, 2014, pp. 231–64. Crossref, doi:10.1007/s12017-014-8295-9.
5. Buschdorf, Jan P., and Wolf H. Strtling. "A WW Domain Binding Region in Methyl-CpG-Binding Protein MeCP2: Impact on Rett Syndrome." *Journal of Molecular Medicine*, vol. 82, no. 2, 2004, pp. 135–43. Crossref, doi:10.1007/s00109-003-0497-9.
6. McFarland, K. N., *et al.* "MeCP2: A Novel Huntingtin

- Interactor." *Human Molecular Genetics*, vol. 23, no. 4, 2013, pp. 1036–44. Crossref, doi:10.1093/hmg/ddt499.
7. Roux, Jean-Christophe, *et al.* "Modification of Mecp2 Dosage Alters Axonal Transport through the Huntingtin/Hap1 Pathway." *Neurobiology of Disease*, vol. 45, no. 2, 2012, pp. 786–95. Crossref, doi:10.1016/j.nbd.2011.11.002.
 8. Ehrhart, Friederike, *et al.* "Rett Syndrome – Biological Pathways Leading from MECP2 to Disorder Phenotypes." *Orphanet Journal of Rare Diseases*, vol. 11, no. 1, 2016. Crossref, doi:10.1186/s13023-016-0545-5.
 9. Swanberg, Susan E., *et al.* "Reciprocal Co-Regulation of EGR2 and MECP2 Is Disrupted in Rett Syndrome and Autism." *Human Molecular Genetics*, vol. 18, no. 3, 2008, pp. 525–34. Crossref, doi:10.1093/hmg/ddn380.
 10. Futter, M., *et al.* "Wild-Type but Not Mutant Huntingtin Modulates the Transcriptional Activity of Liver X Receptors." *Journal of Medical Genetics*, vol. 46, no. 7, 2009, pp. 438–46. Crossref, doi:10.1136/jmg.2009.066399.
 11. Faber, P. "Huntingtin Interacts with a Family of WW Domain Proteins." *Human Molecular Genetics*, vol. 7, no. 9, 1998, pp. 1463–74. Crossref, doi:10.1093/hmg/7.9.1463.
 12. Tillier, E. R. M., and R. L. Charlebois. "The Human Protein Coevolution Network." *Genome Research*, vol. 19, no. 10, 2009, pp. 1861–71. Crossref, doi:10.1101/gr.092452.109.
 13. ries, Sjoerd J. de, *et al.* "The HADDOCK Web Server for Data-Driven Biomolecular Docking." *Nature Protocols*, vol. 5, no. 5, 2010, pp. 883–97. Crossref, doi:10.1038/nprot.2010.32.
 14. Zundert, G. C. P. van, *et al.* "The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes." *Journal of Molecular Biology*, vol. 428, no. 4, 2016, pp. 720–25. Crossref, doi:10.1016/j.jmb.2015.09.014.
 15. Honorato, Rodrigo V., *et al.* "Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem." *Frontiers in Molecular Biosciences*, vol. 8, 2021. Crossref, doi:10.3389/fmolb.2021.729513.
 16. Karczewski, Konrad J., *et al.* "The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans." *Nature*, vol. 581, no. 7809, 2020, pp. 434–43. Crossref, doi:10.1038/s41586-020-2308-7.
 17. Bateman, Alex, *et al.* "UniProt: The Universal Protein Knowledgebase in 2021." *Nucleic Acids Research*, vol. 49, no. D1, 2020, pp. D480–89. Crossref, doi:10.1093/nar/gkaa1100.
 18. Yang, Xiaoyong, *et al.* "Recruitment of O-GlcNAc Transferase to Promoters by Corepressor MSin3A." *Cell*, vol. 110, no. 1, 2002, pp. 69–80. Crossref, doi:10.1016/s0092-8674(02)00810-3.
 19. UK BioBank. "PheWeb." UKBiobank TOPMed-Imputed PheWeb, UK BioBank, pheweb.org/UKB-TOPMed/ about. Accessed 25 Aug. 2021.
 20. Jung, Taeyang, *et al.* "The Polyglutamine Expansion at the N-Terminal of Huntingtin Protein Modulates the Dynamic Configuration and Phosphorylation of the C-Terminal HEAT Domain." *Structure*, vol. 28, no. 9, 2020, pp. 1035–1050.e8. Crossref, doi:10.1016/j.str.2020.06.008.
 21. Jumper, John, *et al.* "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature*, 2021. Crossref, doi:10.1038/s41586-021-03819-2.
 22. Vangone, Anna, and Alexandre MJJ Bonvin. "Contacts-Based Prediction of Binding Affinity in Protein–Protein Complexes." *ELife*, vol. 4, 2015. Crossref, doi:10.7554/elife.07454.
 23. Xue, Li C., *et al.* "PRODIGY: A Web Server for Predicting the Binding Affinity of Protein–Protein Complexes." *Bioinformatics*, 2016, p. btw514. Crossref, doi:10.1093/bioinformatics/btw514.
 24. Bai, Siau Wei, *et al.* "Identification and Characterization of a Set of Conserved and New Regulators of Cytoskeletal Organization, Cell Morphology and Migration." *BMC Biology*, vol. 9, no. 1, 2011. Crossref, doi:10.1186/1741-7007-9-54.
 25. Dominguez, Cyril, *et al.* "HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information." *Journal of the American Chemical Society*, vol. 125, no. 7, 2003, pp. 1731–37. Crossref, doi:10.1021/ja026939x.
 26. Salah, Zaidoun. "WW Domain-Containing Proteins: Retrospectives and the Future." *Frontiers in Bioscience*, vol. 17, no. 1, 2012, p. 331. Crossref, doi:10.2741/3930.
 27. Gao, Yong-Guang, *et al.* "Autoinhibitory Structure of the WW Domain of HYPB/SETD2 Regulates Its Interaction with the Proline-Rich Region of Huntingtin." *Structure*, vol. 22, no. 3, 2014, pp. 378–86. Crossref, doi:10.1016/j.str.2013.12.005.
 28. Ilsley, Jane L., *et al.* "The WW Domain: Linking Cell Signalling to the Membrane Cytoskeleton." *Cellular Signalling*, vol. 14, no. 3, 2002, pp. 183–89. Crossref, doi:10.1016/s0898-6568(01)00236-4.
 29. Allen, Mark *et al.* "The structure of an FF domain from human HYPA/FPB11." *Journal of molecular biology* vol. 323,3 (2002): 411-6. doi:10.1016/s0022-2836(02)00968-3
- Copyright:** ©2022 Desai, Stork. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.