

## Research Article

# Redefining the catalytic HECT domain boundaries for the HECT E3 ubiquitin ligase family

Emma I. Kane, Steven A. Beasley, Johanna M. Schafer, Justine E. Bohl, Young Sun Lee, Kayla J. Rich, Elizabeth F. Bosia and  Donald E. Spratt

Gustaf H. Carlson School of Chemistry and Biochemistry, Clark University, 950 Main Street, Worcester, MA 01610, U.S.A.

**Correspondence:** Donald E. Spratt (dspratt@clarku.edu)



There are 28 unique human members of the homologous to E6AP C-terminus (HECT) E3 ubiquitin ligase family. Each member of the HECT E3 ubiquitin ligases contains a conserved bilobal HECT domain of approximately 350 residues found near their C-termini that is responsible for their respective ubiquitylation activities. Recent studies have begun to elucidate specific roles that each HECT E3 ubiquitin ligase has in various cancers, age-induced neurodegeneration, and neurological disorders. New structural models have been recently released for some of the HECT E3 ubiquitin ligases, but many HECT domain structures have yet to be examined due to chronic insolubility and/or protein folding issues. Building on these recently published structural studies coupled with our in-house experiments discussed in the present study, we suggest that the addition of ~50 conserved residues preceding the N-terminal to the current UniProt defined boundaries of the HECT domain are required for isolating soluble, stable, and active HECT domains. We show using *in silico* bioinformatic analyses coupled with secondary structural prediction software that this predicted N-terminal  $\alpha$ -helix found in all 28 human HECT E3 ubiquitin ligases forms an obligate amphipathic  $\alpha$ -helix that binds to a hydrophobic pocket found within the HECT N-terminal lobe. The present study brings forth the proposal to redefine the residue boundaries of the HECT domain to include this N-terminal extension that will likely be critical for future biochemical, structural, and therapeutic studies on the HECT E3 ubiquitin ligase family.

## Introduction

In 1995, a novel class of enzymes containing an approximate 350 residue conserved domain at their respective C-terminus were found to be homologous to the C-terminal domain of E6-associated protein (E6AP) and had the capabilities of forming a thioester bond with ubiquitin [1]. The most unique aspect of this study revealed that a highly conserved catalytic cysteine residue within this C-terminal domain was responsible for polyubiquitin chain formation. This discovery led to the classification of the homologous to E6AP C-terminus (HECT) E3 ubiquitin ligase family, with currently 28 human members that differ in size and the presence of various N-terminal protein-protein interaction domains involved in substrate recruitment [2]. Further investigation through structural analysis revealed this 350 residue domain was bilobal with numerous conserved residues within the hinge region found between the N- and C-terminal lobes that provide flexibility to support HECT-dependent ubiquitylation activity [3]. This cleft region is suggested to also be critical to its HECT function as loss-of-function mutations in E6AP within this region have been linked to Angelman syndrome [4]. Both lobes were also identified to have unique characteristics involved in ubiquitin handling by the HECT E3 ubiquitin ligases – where the N-terminal lobe is responsible for E2 ubiquitin conjugating enzyme recruitment and binding, while the C-terminal lobe contains the highly conserved catalytic cysteine residue that facilitated ubiquitin transfer onto a specific substrate [5,6]. Conservation within the hinge between the N- and C-terminal lobes have been suggested to allow for the flexible rotation to bring the catalytic cysteine residue in the HECT C-terminal lobe into close proximity

Received: 09 May 2022  
Revised: 07 September 2022  
Accepted: 08 September 2022

Accepted Manuscript online:  
16 September 2022  
Version of Record published:  
04 October 2022

with the E2-ubiquitin complex bound to the HECT N-terminal lobe [6]. Studies have also suggested a competing proximal addition model where the HECT domain can form higher oligomeric structures to facilitate polyubiquitin chain formation [7–11].

Members of the HECT E3 ubiquitin ligase family have been broadly categorized into three subfamilies based upon their domain architecture in their variable N-terminal protein–protein interaction domains and their functional similarities. These subfamilies include the neuronal precursor cell-expressed developmentally down-regulated 4 subfamily (NEDD4, 9 members), the HECT and RLD domain subfamily (HERC, 6 members), and the unclassified ‘other’ subfamily (13 members) [12]. The NEDD4 E3 ubiquitin ligases possess a Ca<sup>2+</sup>-binding C2 domain for phospholipid recognition [13] and two-to-four WW domains required for binding specifically to proline-rich motif-containing substrates [14]. Members of the HERC subfamily are divided based upon size, with only the large HERCs (HERC1/HERC2) containing more than one RCC1-like domain (RLD) for multiple substrate recognition and binding opportunities [15]. The remaining 13 HECT E3 ubiquitin ligases have diverse protein–protein interacting domains within their N-terminus, with some homology existing such as ankyrin repeat and armadillo repeat domains [2]. Thorough phylogenetic analyses of the HECT family suggest that it may be more appropriate to have 16 separate HECT subfamilies, including the notable division of small and large members of the HERC subfamily [16,17].

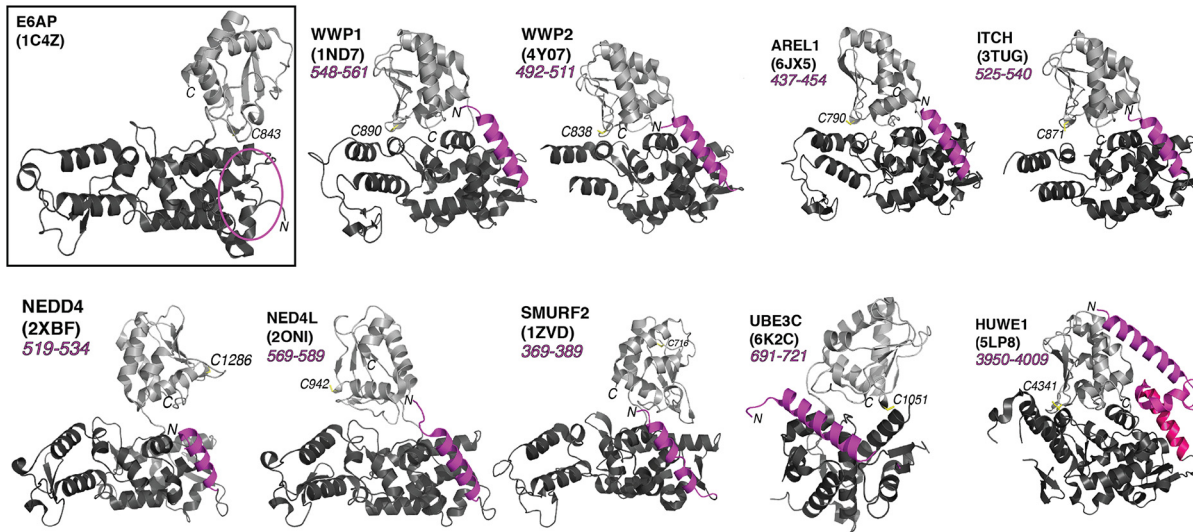
HECT E3 ubiquitin ligases are involved in a myriad of biological processes and cellular pathways including protein degradation, DNA damage repair, apoptosis, subcellular localization, and immunological response [2,18–20]. Numerous studies have implicated the dysfunction of members from the HECT E3 ubiquitin ligase family in disease development such as various cancers, immunological disorders, neurodevelopmental disorders, and neurodegenerative disease [18,21–24]. Since the HECT domain possesses the most prominent functional regions of this protein family, it is logical that a research focus would be directed towards structural and functional characterization of this catalytic region of the HECT E3 ubiquitin ligases. New structural insights into the HECT domain have shown that the introduction of extending the N-terminal sequence of the HECT domain is important for folding. While many of these crystal structures exist for HECT domains with an additional  $\alpha$ -helix [5,25–27], the HUWE1 structure contains up to three  $\alpha$ -helices [27].

Here, we show through bioinformatic analysis and biochemical characterization that up to four N-terminal  $\alpha$ -helices beyond the ~350 residue UniProt defined boundary are predicted to be present within all 28 HECT E3 ubiquitin ligases and that including these N-terminal extensions for the HECT domain improves solubility and increases ubiquitylation activity. Our bioinformatic evaluation of the various members of the HECT E3 ubiquitin ligases through multiple sequence alignment (MSA) with Jalview [28] coupled with secondary structure prediction and evaluating available PDB structures using PyMol [29] suggests there exists at least one amphipathic  $\alpha$ -helix – with some predicted to have two to four  $\alpha$ -helices – prior to the N-terminal lobe of the HECT domain. Our assertion is further supported by the recent crystal structures of the HECT domains of apoptosis-resistant E3 ubiquitin protein ligase-1 (AREL1, PDB: 6JX5) and ubiquitin-protein ligase E3C (UBE3C; PDB: 6K2C) HECT domains that included an additional 50 amino acids preceding the N-terminal lobe to the isolated HECT domains [25,30]. Our studies coupled with recent structural reports on different HECT domains indicate that the currently annotated UniProt HECT domain residue boundaries need to be revisited and revised.

## Results

In the years since the original E6AP HECT domain structure was published in 1999, every solved HECT structure has included at least one  $\alpha$ -helical extension immediately upstream from the defined N-terminal lobe (Figure 1) [5,6,25–27,30–33]. The additional  $\alpha$ -helix (often referred to in the literature as  $\alpha 1'$ ) sits astride the HECT N-terminal lobe on the opposite face from the E2 enzyme binding site. The ClustalW and T-Coffee MSA of the 28 human HECT paralogs resulted in an alignment that includes a JPred predicted  $\alpha$ -helix (Figure 2).

The predicted  $\alpha$ -helix shows a periodicity for the highly conserved hydrophobic residues resulting in their collective orientation on the same face of the  $\alpha$ -helix. Further analysis of the HECT domains with known structures reveals these residues form a hydrophobic core along with the hydrophobic residues from three distinct  $\alpha$ -helices found in the HECT N-terminal lobe that are also conserved (Figure 3A). For example, in the AREL1 HECT domain structure, these comprise F439, V443, F446, L450, and V453 on the  $\alpha$ -helical extension interacting with residues L563, Y564, L691, L692, I694, F695, and L703 on the HECT N-terminal lobe creating a hydrophobic pocket (Figure 3B,C). Further structural analysis revealed that removing this protective helix exposes a contiguous hydrophobic patch for all the HECT proteins analyzed. For example, AREL1 has a hydrophobic patch measured to be 627 Å and WWP1 has a hydrophobic patch of 698 Å (Figure 3C). This exposed pocket may have contributed to the insolubility of the shorter



**Figure 1. Structural analysis of the solved HECT domains models of various HECT E3 ubiquitin ligases**

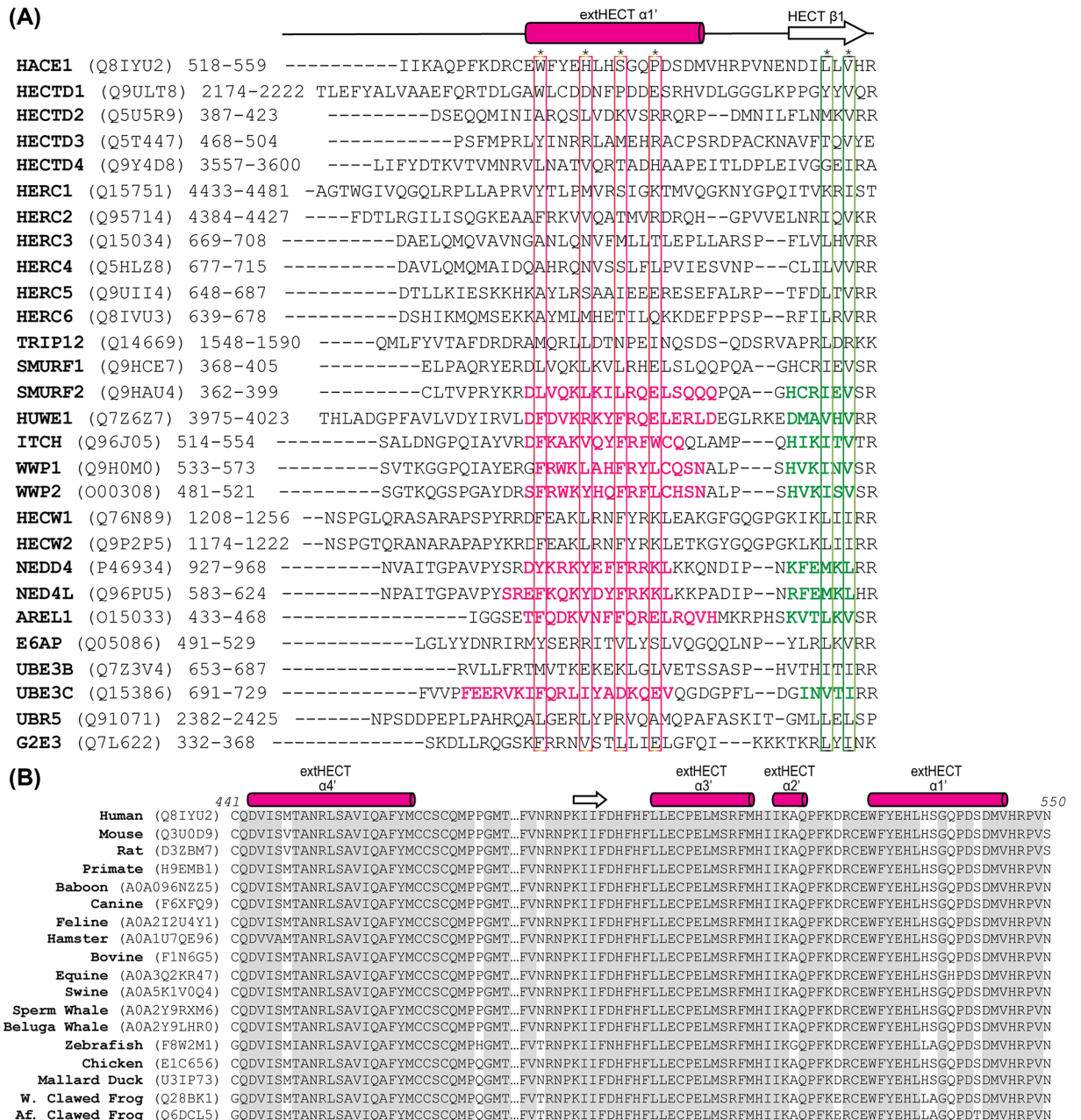
Representative extended HECT domain models within the HECT E3 ubiquitin ligase family. With the amphipathic  $\alpha$ -helix included, all models show a similar tertiary structure to its namesake E6AP (PDB: 1C4Z). The bilobal structure is shown with the C-lobe containing the catalytic cysteine (light gray), the N-lobe with the E2~ubiquitin binding interface and sites required for oligomerization (dark gray), and the N-terminal  $\alpha$ -helical extension (magenta). These crystal models are also missing some atoms including loops and the C-terminal tail of the HECT domain.

constructs. It is interesting to note that the equivalent hydrophobic patch residues of the originally published E6AP structure missing the N-terminal  $\alpha$ -helix (PDB: 1C4Z) are found in the core of the trimeric crystal interface.

The protection of this hydrophobic pocket probably provides structural integrity to the HECT domain, which is supported by the marked increase in solubility we observed for our extended HECT domain constructs (Figure 4A). Each extension of the HECT domain we incorporated into our extended HECT (extHECT) expression constructs was based off the location of predicted boundaries of  $\alpha$ -helices, with the inclusion of each  $\alpha$ -helix showing improved solubility and HECT-dependent activity. Taken together, our MSA results support and further demonstrate that identical or similar residues within these corresponding positions within the predicted  $\alpha$ -helix extension are also found in all 28 human members of the HECT E3 ubiquitin ligase family.

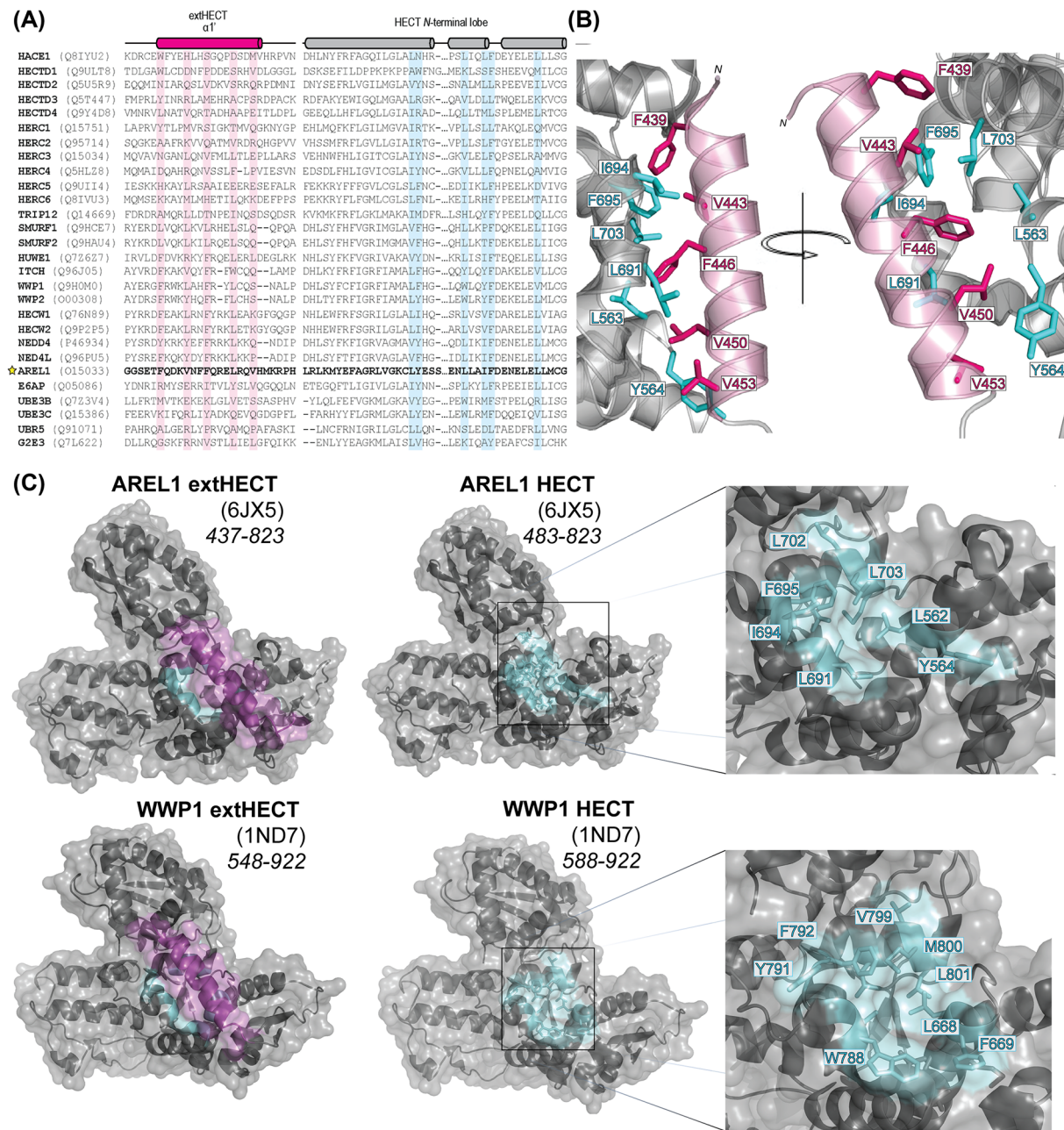
The HECT E3 ubiquitin ligases we assessed in this study (HECW2, HERC4, G2E3, and HACE1) showed significantly increased solubility upon induction with IPTG and higher yields when comparing our extHECT constructs to the current UniProt HECT domain boundaries (Figure 4). Our results demonstrate that the susceptibility of the exposed hydrophobic cleft at the N-terminus of the HECT domain can be mitigated by the addition of these additional  $\alpha$ -helices. Each hydrophobic residue within the extended  $\alpha$ -helix/helices suggests that this conserved region is required to support normal HECT-dependent ubiquitylation. The solubility of each HECT E3 ubiquitin ligase appears to be dependent on the addition of at least one of these predicted  $\alpha$ -helical extensions. Both paralog and ortholog secondary structure prediction indicate that the extended region would likely have  $\alpha$ -helical content (Figures 2 and 3).

The overexpression and purification of heterologous proteins in *Escherichia coli* greatly depends on the overall stability and native fold, while mitigating solvent accessibility of hydrophobic regions of the expressed protein [34–36]. Despite our repeated attempts to modify buffers, reagents, and purification procedures, we were unable to improve stability, activity, and/or the yield of any of our original HECT constructs using the UniProt defined domain boundaries. Our experimental optimization included (i) varying temperatures for expression following IPTG induction, (ii) altering IPTG concentrations, (iii) scaling down production culture volumes, and (iv) subcloning the HECT domain construct to be expressed with other purification fusion tags. While we were able to successfully purify some of these UniProt defined HECT constructs, the overall yield we obtained for our new extHECT constructs resulted in approximately a 50-fold increase in protein yield (Figure 4).

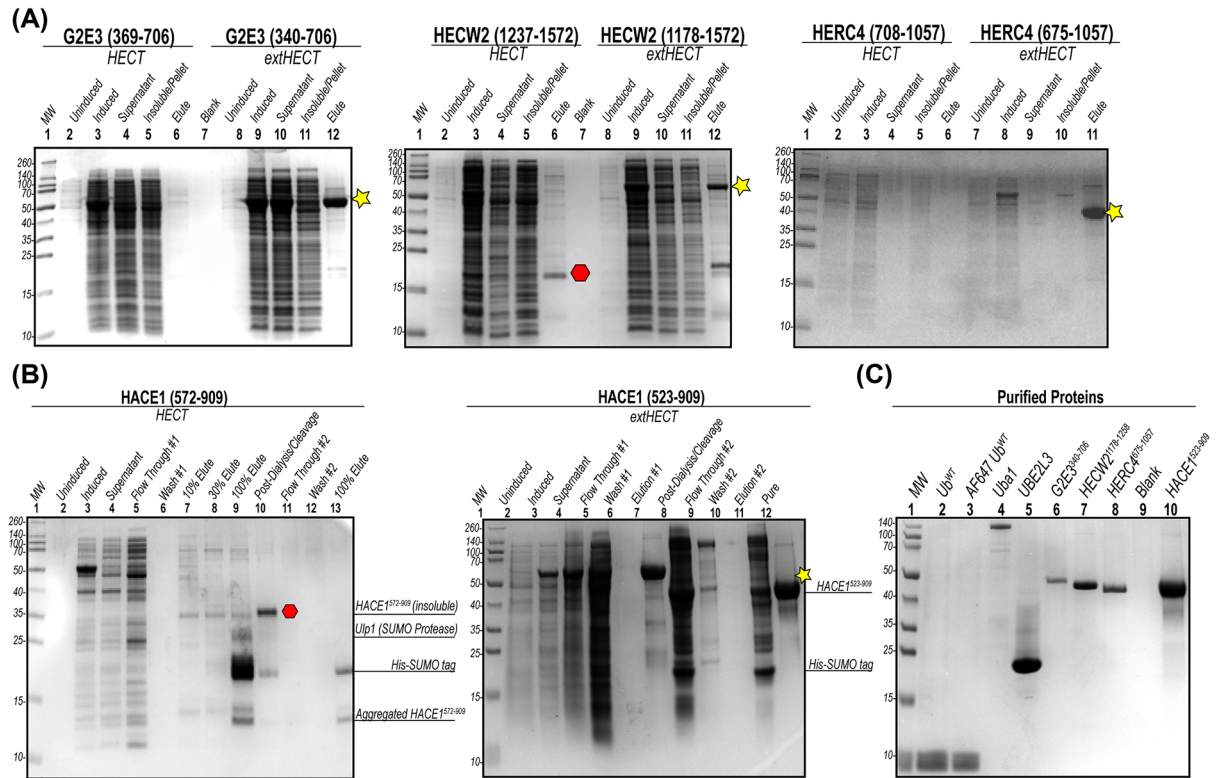


**Figure 2. Paralog and ortholog MSA of the HECT E3 ligase family with secondary structure prediction through Jalview and JPred**

(A) Paralog MSA of the junction with the extended HECT  $\alpha$ -helix ( $\alpha 1'$ ) and the initial  $\beta$ -sheet ( $\beta 1$ ) of the HECT domain. Sequences and residue numbers were obtained through UniProt for all human HECT E3 ubiquitin ligases. Structural elements from solved crystal models with the  $\alpha$ -helical extension (magenta) and N-lobe  $\beta$ -sheet (green) are indicated. Residues facing towards the hydrophobic core are designated with an asterisk. The alignment was produced using ClustalW and T-Coffee and manually curated by Jalview. (B) Ortholog results revealed high conservation for the HECT E3 ubiquitin ligase HACE1 where all four  $\alpha$ -helices are predicted to fold preceding the N-lobe of the HECT domain. This finding complements the paralog analysis shown in panel (A). Residues with absolute conservation are shown in grey.



**Figure 3. Secondary structure prediction reveals a hydrophobic pocket protected by the N-terminal  $\alpha$ -helical extension**  
 (A) MSA analysis coupled with secondary structure prediction for all HECT E3 ubiquitin ligases indicates a conserved amphipathic  $\alpha$ -helix (shown in magenta) with hydrophobic residues suggested to interact with residues found in the hydrophobic cleft of the HECT N-terminal lobe (shown in cyan). This alignment takes into consideration the conservation and secondary structure prediction with a focus on the 1'  $\alpha$ -helix preceding the HECT domain. (B) The AREL1 HECT (PDB: 6JX5) with conserved hydrophobic residues shown within the predicted  $\alpha$ -helical extension (magenta) are predicted to interact with the conserved N-lobe residues (cyan). This alignment also confirms the role that extending the known HECT domain boundaries are important for domain stability and solubility. (C) Surface maps of the extended (right) and UniProt defined (left) WWP1 and AREL1 HECT domains. The conserved sidechain HECT N-lobe residues highlighted in panel (B) reside within the hydrophobic pocket (cyan). The  $\alpha$ -helical extension (magenta) reduces the surface area exposed from the hydrophobic patch, as observed through the decrease in color. The inset (right) shows a closer look at the conserved hydrophobic residues in the hydrophobic patch.



**Figure 4. Solubility and purification of UniProt defined HECT domain against extended HECT domains**

**(A)** The solubility of each HECT and extHECT construct were analyzed by SDS-PAGE. Solubility following induction showed that the extHECT constructs of the HECT E3 ubiquitin ligases evaluated in this paper (shown with yellow stars) were more soluble than the shorter HECT domains using the domain boundaries annotated on UniProt. **(B)** Example SDS-PAGE of the purification steps for HACE1 HECT (residues 574–909 based on UniProt defined domain boundaries) and extHECT (residues 523–909) constructs. The HECT construct was deemed to be insoluble (shown with red octagon) while the extHECT HACE1 was significantly more soluble and with a higher yield (shown with yellow star). **(C)** Analysis of protein purity for all proteins used in this study including the isolated extHECT constructs. All samples were loaded at 5 mg of protein, except for UBE2L3 and HACE1<sup>523-909</sup> which were loaded at 10 mg.

We also assessed the ubiquitylation activity of our HECT and extHECT proteins using an *in vitro* fluorescent ubiquitylation assay that has been optimized to demonstrate E2~ubiquitin charging/discharging and improved extHECT~ubiquitin charging. We observed a marked increase in ubiquitylation activity by our extHECT constructs, whereas the HECT domains showed little to no activity (Figure 5). All extended HECT domains used for these assays demonstrated that the HECT domain could be charged with ubiquitin. To demonstrate optimization of UBE2L3~Ub charging, UBE2L3 discharging and HECT/extHECT~Ub charging, we performed an additional assessment to find the optimal concentration of UBE2L3 to ubiquitin to limit the excess free ubiquitin found at the bottom of the gel. Once determining the ideal concentration of ubiquitin (5  $\mu$ M) and UBE2L3 (10  $\mu$ M), we performed an optimization assessment of the E3 at a 1:1 ratio of UBE2L3 to reproducibly show E3~ubiquitin charging and UBE2L3 discharge of ubiquitin. We found that using a HECT or extHECT concentration at 10  $\mu$ M at a 1:1 ratio showed the robust activity of our extended HECT domain constructs. With the difficulty in isolating the shorter HECT proteins, our data shows that the inclusion of additional N-terminal  $\alpha$ -helices provide overall greater structural integrity coupled with increased solubility that aid in HECT-dependent ubiquitylation activity.

## Discussion

Building on previous studies, the present study proposes and reinforces the need to redefine the HECT domain boundaries currently annotated on UniProt and the major databases. The HECT domain that was originally annotated using the BLAST search engine of the NCBI databases yielded only 14 hits [1]. The NCBI GenBank databases at the end of 1994 contained less than 238,000 total sequences as compared with nearly 250,000,000 sequences currently



in the databases as of the writing of this study. The original boundaries for the HECT domain were defined from the sequences found in Pfam entry PF00643 and InterPro entry IPR000569. Subsequent HECT domain boundaries and analyses have used this definition as the basis for phylogenetic studies such as the identification of HECT proteins in soybeans [37].

Identifying the exact domain boundaries within a protein often requires access to solved structural models; however, bioinformatic approaches using MSA and secondary structure and disorder prediction have proved to be successful approaches to predict domain boundaries. For example, this bioinformatic MSA approach was used to successfully predict the boundaries for the novel Parkin RING0 domain, and the disordered regions found between domains such as the ubiquitin-like (UblD) and RING-BRcat-Rcat (RBR) domains of Parkin [38] before 3D structures of full-length parkin were released in 2013 [39–42]. Although the PONDR results concerning the human HECT domains can show some differences, as it often predicts disorder in known structures at the N-terminal, the PONDR server reproducibly predicted that there was disorder before our new extended HECT domain boundaries. Previous bioinformatic analysis performed on the HECT family have identified conserved residues that regulate autoinhibition upon modification. As with NEDD4 and E6AP, there are lysine and threonine residues that reside within the first  $\alpha$ -helix predicted to fold preceding the N-lobe that are subject to methylation and phosphorylation [43,44].

While it was found that this N-terminal extension also induces HECT-dependent oligomerization, the exact structural mechanisms used by many of the members in the HECT E3 ubiquitin ligase family that employ this modification remain only partially understood. Previous studies have also pointed out that the inclusion of the extended  $\alpha$ -helices in-front of the HECT domain are necessary to perform structural analysis of the HECT domain that coincidentally also include highly conserved, essential residues that mediate HECT-dependent oligomerization [3,8,11,31,43,45]. Future studies are essential to examine the oligomerization and/or autoinhibition potential for the currently illusive members of the HECT E3 ubiquitin ligases.

Within the Human Gene Mutation Database (HGMD), numerous missense mutations have been annotated in the HECT E3 ubiquitin ligase family that are in proximity to the HECT domain (Table 1) [46]. Of these, some are relatively conserved and correlate to the onset of similar diseases and/or disorders. For example, an N-terminal missense mutation of an arginine to stop codon, irrespective of their subfamily classification, is linked to intellectual disability. Specifically assessing the HECW2 (R1330W) and HACE1 (R585W) mutants, this arginine residue is relatively conserved in the N-lobe and linked to the onset of neurological disorders [47,48]. Interestingly, the HECTD3 mutant (R478C), which has been linked to Tourette syndrome, lies i-3 from the conserved phenylalanine of the  $\alpha 1'$ -helical extension that protects the hydrophobic pocket. Also, the HECW2 F1193V missense mutation identified in neurodevelopmental delay and hypotonia is the conserved  $\alpha 1'$  residue that interacts with the N-terminal  $\alpha$ -helix to protect the hydrophobic pocket [49]. This demonstrates that some of these novel missense mutations within the HECT E3 ubiquitin ligase family likely are involved, whether directly or indirectly, with protection of this hydrophobic pocket to maintain its structure and ubiquitylation function.

HECT E3 ubiquitin ligases have been implicated in numerous diseases and disorders in relation to their altered expression and functional capabilities [2]. For example, the altered expression levels of HACE1 leads toward the development of Wilms' tumor [77]. Interestingly, overexpression of HUWE1 has been correlated to breast, brain, and prostate cancer progression, but also is down-regulated in both colorectal and lung cancer [78]. The effect that these residue substitutions listed in Table 1 have on the structure and function of the HECT E3 ubiquitin ligase family members require further investigation to determine if these residue substitutions affect protein stability, autoregulation, or HECT domain concentration-dependent oligomerization.

While this manuscript was in preparation, the structure prediction AlphaFold database was publicly released [79]. The domain boundaries from our study and the predicted structures from AlphaFold are summarized in Table 2. The four largest human HECT E3 ubiquitin ligases are absent from the AlphaFold database as they exceed their 2700 amino acid length criterion (i.e. HERC1, HERC2, HECTD4, and UBR5). In general, there was consensus between the extended HECT and AlphaFold database domain boundary predictions with an obligate  $\alpha$ -helix present to stabilize the HECT N-terminal lobe (Table 2).

The significant difference was the size of the loop between the canonical HECT and the helical extension, and this was most pronounced in the HERC subfamily (i.e. HERC1, HERC2, HERC3, HERC4, HERC5, and HERC6) as well as HECTD1, HECTD2, and HECTD3. It should also be noted that the extended  $\alpha$ -helix is predicted to lie in the hydrophobic groove with its N-terminus proximal to the C-lobe, with the only exception being TRIP12 where the N-terminal helix is predicted to lie in the opposite orientation (Figure 6). Regardless, these new computational models further support our suggestion that the HECT domain boundaries need to be expanded to include the N-terminal helix.



**Table 1** HECT domain mutations annotated in the HGMD database linked to disease

HECT	Residue substitution	Location in the HECT domain	Clinical manifestation
NED4L	Y679C; Q974H	N-lobe	Periventricular neuronal heterotopia [50]
	E893K; R897Q	C-lobe	
SMURF2	T641A	C-lobe	Neurodevelopmental disorder [51]
HECW2	R1191Q; F1193V	ext- $\alpha^2$ <sup>a</sup>	Neurodevelopmental delay and hypotonia [49]
	R1330W	N-lobe	Developmental delay, absent speech, epilepsy, encephalopathy, hypotonia, and macrocephaly [47]
	D1442G		Autism spectrum disorder [52]
	E1445G		Neurodevelopmental delay and hypotonia [49]
HERC1	L4154X <sup>b</sup>	ext- $\alpha^3$ <sup>a</sup>	Autism spectrum disorder [53]
	G4520E	N-lobe	Overgrowth, intellectual disability, and facial dysmorphism [54]
HERC2	D4267E	ext- $\alpha^3$ <sup>a</sup>	Neurological disease [55]
TRIP12	R1595Q	N-lobe	Intellectual disability [56]
	S1840L		
	Q1916X <sup>b</sup>	C-lobe	Intellectual disability [57]
HUWE1	R3267H	ext- $\alpha^4$ <sup>a</sup>	Craniosynostosis [58]
	R4013W	ext- $\alpha^1$ <sup>a</sup>	Mental retardation, X-linked [59]
	R4023C;		Intellectual disability [60]
	N4075K; Y4106C; L4157V; E4244D		
	R4063Q	N-lobe	Intellectual disability, microcephaly, and postnatal growth failure [61]
	R4130Q		Developmental delay, Turner-type [62]
	R4187H		Intellectual disability, X-linked [63]
	R4187C		Mental retardation, X-linked [59]
	G4229D		Multiple congenital anomalies [64]
	K4295N	C-lobe	Intellectual disability [60]
	G4310R		Intellectual disability, microcephaly, and postnatal growth failure [61]
HACE1	R585W	N-lobe	Autism spectrum disorder [48]
	Gln618fs <sup>c</sup> ; P674fs <sup>c</sup> ; R748X <sup>b</sup>		Spastic paraplegia and psychomotor retardation with or without seizures [65,66]
	L832X <sup>b</sup>	C-lobe	
	A861P		Hereditary spastic paraplegia [67]
HECTD3	R478C		Tourette syndrome [68]
AREL1	P779L	C-lobe	Pulmonary inflammation [69]
UBE3B	Q700X; G779R	N-lobe	Kaufman oculocerebrofacial syndrome [70]
	Q727P		Blepharophimosis-ptosis-intellectual-disability syndrome [71]
	R922C		Autism spectrum disorder [72]
	R997P	C-lobe	Kaufman oculocerebrofacial syndrome [70]
	Q1005P		Kaufman oculocerebrofacial syndrome [73]
UBE3C	S845F	N-lobe	Autism spectrum disorder [74]
	F996C	C-lobe	
E6AP	I827K; G870D	C-lobe	Angelman syndrome [75,76]

<sup>a</sup>ext- $\alpha$  = predicted  $\alpha$ -helix N-terminal to current UniProt defined boundary

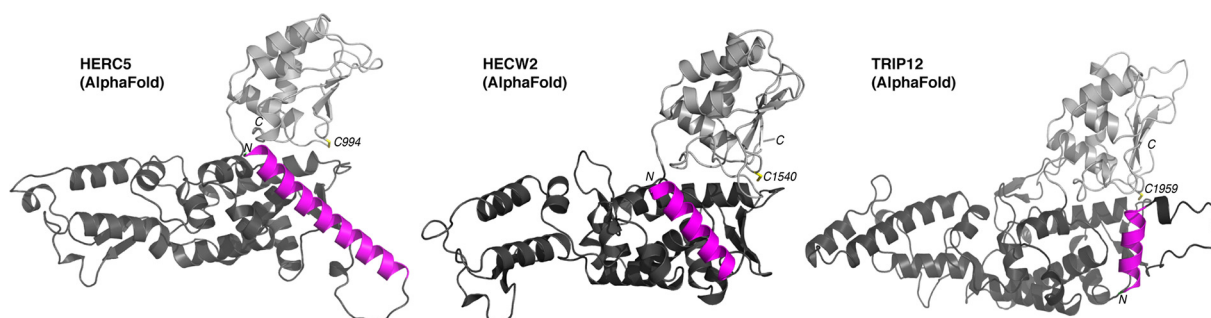
<sup>b</sup>X = stop codon

<sup>c</sup>fs = codon frameshift

In summary, our study demonstrates that our extHECT constructs have increased solubility, improved stability, higher yields, and robust ubiquitylation activity. While some HECT E3 ubiquitin ligases have solved HECT domain models [6,25–27,30,32,33], many others remain unknown. To date, most of these HECT domain structures have been solved for members of the NEDD4 subfamily, with only a few members in the ‘other’ subfamily having a HECT domain structure available on the Protein Data Bank. The present study clearly shows that extending the N-terminal

**Table 2** Uniprot PROSITE and predicted AlphaFold boundaries of the HECT E3 ligase family

Protein ID	UniProt	MSA/Jalview	AlphaFold	Notes	
HERC Family	HERC1	4501-4848	4451-4861	N/A	-
	HERC2	4457-4794	4400-4834	N/A	-
	HERC3	951-1050	681-1050	659-1050	Longer predicted loop and helix
	HERC4	730-1057	689-1057	667-1057	Longer predicted loop and helix
	HERC5	702-1024	659-1024	638-1024	Longer predicted loop and helix
	HERC6	693-1017	650-1022	629-1022	Longer predicted loop and helix
NEDD Family	NEDD4	984-1318	940-1319	941-1319	PDB 2XBF; [33]
	NED4L	640-974	594-975	595-975	PDB 2ONI (X-ray: SGC)
	ITCH	569-903	527-903	528-903	PDB 3TUG (X-ray: SGC)
	WWP1	588-922	547-922	547-922	PDB 1ND7; X-ray [6]
	WWP2	536-870	494-870	494-870	PDB 4Y07; X-ray [32]
	SMURF1	420-757	377-757	369-757	Similar prediction
	SMURF2	414-748	371-748	367-748	PDB 1ZVD; X-ray [26]
	HECW1	1271-1606	1226-1606	1223-1606	Similar prediction
	HECW2	1237-1572	1192-1572	1189-1572	Similar prediction
Other Family	TRIP12	1885-1992	1561-1992	1558-1992	unique helix orientation PDB 3G1N, 3H1D, 5LP8;
	HUWE1	4038-4374	3993-4374	N/A	X-ray [27,31] PDB 7MWE; Cryo-EM [80]
	HACE1	574-909	529-909	524-909	Similar prediction
	HECTD1	2151-2610	2192-2610	2091-2610	Longer predicted loop and helix
	HECTD2	437-776	395-776	377-776	Longer predicted loop and helix
	HECTD3	512-857	474-861	451-861	Longer predicted loop and helix
	HECTD4	3627-3996	3570-3996	N/A	-
	UBR5	2462-2799	2395-2799	N/A	-
	AREL1	483-823	439-823	439-823	PDB 6JX5; X-ray [25]
	G2E3	371-698	341-706	345-706	Similar prediction
	UBE3B	702-1068	659-1068	650-1068	Longer predicted loop and helix
	UBE3C	744-1083	695-1083	695-1083	PDB 6K2C; X-ray [30]
E6AP	776-875	500-875	486-875	Longer predicted loop and helix	



**Figure 6.** AlphaFold predicts an  $\alpha$ -helix preceding the HECT domain

Predicted models of the extended HECT domain of HECW2, HERC5, and TRIP12 from the AlphaFold database (<https://alphafold.ebi.ac.uk/>) showing different  $\alpha$ -helical extension configurations. The HECT N-terminal lobe (dark gray), C-terminal lobe (light gray),  $\alpha$ -helical extension that stabilizes the N-terminal lobe (magenta), and the catalytic cysteine (yellow) shown that AlphaFold models predicts that different orientations for the N-terminal extension in the extHECT construct may be possible.

lobe and redefining the HECT domain boundaries for all the human HECT E3 ubiquitin ligases is warranted. This knowledge will be an important consideration for future biochemical, structural, and therapeutic studies on the HECT E3 ubiquitin ligase family.

## Methods

### Bioinformatic analysis of the HECT E3 ubiquitin ligases

The sequences for the 28 human HECT paralogs were acquired from the UniProt database ([www.uniprot.org](http://www.uniprot.org)) using the canonical HECT E3 ubiquitin ligase isoform and parsing out splice variants. The sequences were aligned using ClustalW [81] and T-Coffee [82] followed by manual curation in Jalview [28]. Secondary sequence prediction by JPred [83] was used on the aligned sequences within Jalview to assist in determining domain boundaries. Furthermore, the Protein Data Bank (PDB; [www.rcsb.org](http://www.rcsb.org)) database was scoured to identify all the deposited PDB structures for isolated HECT domains for PyMol analysis [29,84] (Figures 1 and 2). Jalview was also used to assess the ortholog family of HACE1, showing both high conservation and  $\alpha$ -helical prediction location immediately N-terminal to the annotated HECT domain (Figure 2B). Ortholog MSA analysis of both protein families were downloaded from Ensembl ([ensembl.org](http://ensembl.org)) and analyzed in Jalview [28,85]. Sequences that were obvious fragments (lacking an intact HECT domain that included the catalytic cysteine), contained stretches of unknown residues, or isoforms from the same organism were removed from the analyses. The organisms were chosen to span disparate eukaryotic clades. PONDR VSL1 analysis was also used to search the HECT E3 ubiquitin ligases for predicted disordered regions to assist in domain boundary predictions ([www.pondr.com](http://www.pondr.com)). ExPasy ProtParam was used to predict the physicochemical properties of the predicted extended HECT domain constructs ([web.expasy.org/protparam](http://web.expasy.org/protparam)) [86,87].

The proteins focused on in this study (HECW2, HERC4, G2E3, and HACE1) do not have any deposited PDB structures, and our initial studies with our constructs based upon UniProt HECT domain boundaries had proven difficult to express and purify for downstream biochemical and biophysical analysis. To determine the size of the hydrophobic patch, the PDB files for the known HECT structures that include the helical extension were analyzed using PyMol. The helical extension was removed from the original PDB file as the basis for the shorter HECT domain analysis. Only the A chain was considered, all waters and other heteroatoms removed, and hydrogens were added. Hydrophobic residues (A,G,V,I,L,E,M,Y,W) sidechain atoms only were selected and the PyMol command 'get\_area' was used to quantitatively determine the surface area coverage.

### HECT domain expression plasmid design and construction

Expression plasmids containing the full-length open reading frame of the corresponding HECT E3 ubiquitin ligases were designed with *E. coli* codon optimization and synthesized by Synbio Technologies, Inc. (Monmouth Junction NJ, U.S.A.). The plasmids contained the essential components desired for protein overexpression in *E. coli* – a T7 promoter induced through the addition of  $\beta$ -D-1-thiogalactopyranoside (IPTG), a lac operon, origin of replication and terminator sequence, antibiotic resistant marker for ampicillin (amp) for selection, an N-terminal His<sub>6</sub>-SUMO tag for increased protein solubility and downstream Ni<sup>2+</sup>-immobilized metal affinity chromatography (Ni<sup>2+</sup>-IMAC) purification, and flanking restriction sites for *Bam*HI (5') and *Xho*I (3') in the multiple cloning site.

Primers were designed to isolate both the UniProt defined HECT (HECT) and extended HECT (extHECT) domains for each protein from the full-length sequence plasmids by deleting residues from the full-length vector using single primers in parallel (SPRINP) [88] and quick-change mutagenesis (QCM) [89] strategies (Table 3). The boundaries for each extHECT construct were chosen based on bioinformatic predictions including the MSA and reference to the recent AREL1 and UBE3B HECT domain PDB structures [25,30]. These included the NEDD4 family member HECW2 (HECT: 1237-1572; extHECT: 1190-1572), HERC protein HERC4 (HECT: 708-1057; extHECT: 675-1057), and 'other' HECT subfamily members G2E3 (HECT: 369-706; extHECT: 340-706) and HACE1 (HECT: 574-909; extHECT: 523-909). All primer sequences were synthesized by Thermo-Fisher. SPRINP and QCM were performed using a T100 Thermal Cycler (Bio-Rad). The HECW2 and HERC4 extHECT expression constructs were made through PCR amplification of the desired ORF, restriction digestion with flanking *Bam*HI (5') and *Xho*I (3'), and ligation with T4 DNA ligase into the His<sub>6</sub>-SUMO plasmid. Colonies were screened by colony PCR and restriction digestion, and positive transformants were verified by Sanger DNA sequencing (Psomagen, Cambridge MA, U.S.A.).

### Overexpression and purification of HECT and extHECT domains

The His<sub>6</sub>SUMO-HECT and His<sub>6</sub>SUMO-extHECT expression plasmids for HECW2, HERC4, G2E3, and HACE1 were transformed into either BL21(DE3) or LOBSTR(DE3) *E. coli* cell lines and grown at 37°C at 200 rpm in Luria-Bertani media supplemented with 100  $\mu$ g/ml ampicillin. When the cultures reached an OD<sub>600</sub> of approximately 0.6–0.8,

**Table 3 Oligonucleotide primers for recombinant HECT and extHECT plasmids**

Primers	Orientation	Nucleotide sequence
HACE1		
572-909	Sense	5'-GGAGGTGGATCCAAGGCTAATTGTGCTAAG-3'
	Antisense	3'-CTTAGCACAATTAGCCTTGGATCCACCTCC-5'
523-909	Sense	5'-GGAGGTGGATCCCAGCCGTTCAAGGATCGC-3'
	Antisense	3'-GCGATCCTTGAACGGCTGGGATCCACCTCC-5'
G2E3		
369-706	Sense	5'-GATTGGAGGTGGATCCACGAAGCGTCTTTAC-3'
	Antisense	3'-CTAACCTCCACCTAGGTGCTTCGCAGAAATG-5'
340-706	Sense	5'-GATTGGAGGTGGATCCTCAAATTCGGCGC-3'
	Antisense	3'-CTAACCTCCACCTAGGAGTTTTAAGGCCGC-5'
HECW2		
1237-1572	Sense	5'-GGAGGTGGATCCAGCAGGAAGGAC-3'
	Antisense	3'-CCTCCACCTAGGTGCTCCTCCTG-5'
1190-1572	Sense	5'-GGAGGTGGATCCAACCGCATTTCCG-3'
	Antisense	3'-CCTCCACCTAGGTTTGCCTAAAGC-5'
HERC4		
708-1057	Sense	5'-CAGATTGGAGGTGGATCCTGCCTAATCTTGTT-3'
	Antisense	3'-GTCTAACCTCCACCTAGGACGGATTAGAACCAA-5'
683-1057	Sense	5'-CAGATTGGAGGTGGATCCCAGATGGCGATTGAC-3'
	Antisense	3'-GTCTAACCTCCACCTAGGGTCTACCGCTAACTG-5'

the cultures were induced with 0.5 mM IPTG and incubated at 16°C at 200 rpm for 16–20 h. The cells were harvested by centrifugation at 10,000 rpm for 10 min at 4°C using a Sorvall LYNX 4000 superspeed centrifuge with a Fiberlite F10-4x1000 LEX Carbon Fiber rotor (Thermo-Fisher). Cell pellets were resuspended with either a Tris or phosphate-based Wash Buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> or Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0) supplemented with ProBlock™ Protease Inhibitor Cocktail (GoldBio, St. Louis MO, U.S.A.; 100 µl/ml), lysed using an EmulsiFlex-C5 homogenizer (Avestin, Ottawa ON, Canada), and clarified by ultracentrifugation using a Optima L-80 XP ultracentrifuge with a 70.1 Ti rotor (Beckman-Coulter) at 41,000 rpm for 40 min. The supernatants were then passed through a 0.45 µm syringe sterile filter (Thermo-Fisher) prior to being applied to 10 ml of Ni<sup>2+</sup>-IMAC resin (GoldBio) pre-equilibrated with wash buffer. After the resin was extensively washed with 25 column volumes of wash buffer, the HECT or extHECT protein was eluted with 30–50 ml of Elution Buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> or Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 8.0). Fractions containing eluted protein were pooled and incubated with recombinant SUMO protease (Ulp1; 40 µg/ml) to cleave the His<sub>6</sub>SUMO tag, then dialysed against Wash Buffer in 3.5 kDa MWCO Spectra/Por 3 Dialysis Tubing (Thermo-Fisher) at 4°C with stirring overnight. The Ulp1 cleaved protein was then reappplied to 10 ml of equilibrated Ni<sup>2+</sup>-IMAC resin in Wash Buffer and the flowthrough containing the desired HECT or extHECT protein was collected and pooled. The protein was then concentrated using an Amicon 15 mL centrifugal filter with a 10 kDa MWCO (Millipore) and loaded on to a HiLoad 16/60 Superdex 75 or 200 column (Cytiva Life Sciences) equilibrated with Gel Filtration Buffer (20 mM Tris-HCl, 100 mM NaCl, 1 mM TCEP, 1 mM EDTA) at a flow rate of 1 ml/min using an ÄKTA Pure 25L FPLC or ÄKTA Start FPLC system (Cytiva Life Sciences). Fractions containing pure HECT or extHECT protein, as assessed by 15% SDS-PAGE and visualized with an iBright FL1000 gel imager (Thermo-Fisher), were pooled, aliquoted, flash frozen in N<sub>2</sub>(l), and stored at –80°C.

### ***In vitro* fluorescent ubiquitylation assay**

All soluble HECT and extHECT proteins were assessed for their ability to catalyze ubiquitin transfer between all cascading enzymes using a fluorescent auto-ubiquitylation assay. Specific attention was given to observing if the inclusion of the N-terminal extension impacted HECT-dependent ubiquitylation activity. Once HECT/extHECT~ubiquitin charging was confirmed through assay optimization, ubiquitylation assays were conducted with 5 µM Alexa Fluor 647 N-terminally tagged ubiquitin (0.1 µM E1 activating enzyme UBE1 (Uba1), 10 µM E2 conjugating enzyme UBE2L3 (UbcH7) or UBE2D3 (UbcH5c), and 10 µM HECT or extHECT, 0.5 mM DTT, 4 µM ATP, 10 µM MgCl<sub>2</sub> in 50 mM HEPES pH 7.5, 100 mM NaCl. Each reaction was incubated at 37 °C at 200 rpm for 1 to 2 h. To determine the presence of ubiquitin~thioester intermediates, appropriate samples were supplemented with 4 mM DTT. Reactions were terminated by adding SDS-PAGE loading dye and heating at 95 °C in a dry bath for one minute. The samples were

then loaded on to a 12% Bis-Tris gel at pH 6.4 and run for 1 h at 130 V. The gels were removed from the apparatus and immediately visualized on an iBright FL1000 imaging system (Thermo-Fisher) using the setting for Alexa Fluor 647 fluorescence gel imaging.

## Data Availability

All materials for this project could be made available upon request.

## Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

This research was supported by the National Institutes of Health [grant number R15GM126432 (to D.E.S.)].

## CRedit Author Contribution

**Emma I. Kane:** Conceptualization, Writing—original draft, Writing—review & editing. **Steven A. Beasley:** Conceptualization, Writing—original draft, Writing—review & editing. **Johanna M. Schafer:** Conceptualization. **Justine E. Bohl:** Conceptualization. **Young Sun Lee:** Conceptualization. **Kayla J. Rich:** Conceptualization. **Elizabeth F. Bosia:** Conceptualization. **Donald E. Spratt:** Conceptualization, Supervision, Writing—original draft, Writing—review & editing.

## Abbreviations

AREL1, apoptosis-resistant E3 ubiquitin protein ligase-1; E6AP, E6-associated protein; HECT, homologous to E6AP C-terminus; HGMD, Human Gene Mutation Database; MSA, multiple sequence alignment; RLD, RCC1-like domain.

## References

- Huibregtse, J.M., Scheffner, M., Beaudenon, S. and Howley, P.M. (1995) A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2563–2567, <https://doi.org/10.1073/pnas.92.7.2563>
- Wang, Y., Argiles-Castillo, D., Kane, E.I., Zhou, A. and Spratt, D.E. (2020) HECT E3 ubiquitin ligases - emerging insights into their biological roles and disease relevance. *J. Cell Sci.* **133**, jcs228072, <https://doi.org/10.1242/jcs.228072>
- Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P.M., Huibregtse, J.M. et al. (1999) Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321–1326, <https://doi.org/10.1126/science.286.5443.1321>
- Kishino, T., Lalande, M. and Wagstaff, J. (1997) UBE3A/E6-AP mutations cause Angelman syndrome. *Nat. Genet.* **15**, 70–73, <https://doi.org/10.1038/ng0197-70>
- Kamadurai, H.B., Souphron, J., Scott, D.C., Duda, D.M., Miller, D.J., Stringer, D. et al. (2009) Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol. Cell.* **36**, 1095–1102, <https://doi.org/10.1016/j.molcel.2009.11.010>
- Verdecia, M.A., Joazeiro, C.A., Wells, N.J., Ferrer, J.L., Bowman, M.E., Hunter, T. et al. (2003) Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol. Cell.* **11**, 249–259, [https://doi.org/10.1016/S1097-2765\(02\)00774-8](https://doi.org/10.1016/S1097-2765(02)00774-8)
- Ronchi, V.P., Kim, E.D., Summa, C.M., Klein, J.M. and Haas, A.L. (2017) In silico modeling of the cryptic E2 approximately ubiquitin-binding site of E6-associated protein (E6AP)/UBE3A reveals the mechanism of polyubiquitin chain assembly. *J. Biol. Chem.* **292**, 18006–18023, <https://doi.org/10.1074/jbc.M117.813477>
- Ronchi, V.P., Klein, J.M., Edwards, D.J. and Haas, A.L. (2014) The active form of E6-associated protein (E6AP)/UBE3A ubiquitin ligase is an oligomer. *J. Biol. Chem.* **289**, 1033–1048, <https://doi.org/10.1074/jbc.M113.517805>
- Ronchi, V.P., Klein, J.M. and Haas, A.L. (2013) E6AP/UBE3A ubiquitin ligase harbors two E2~ubiquitin binding sites. *J. Biol. Chem.* **288**, 10349–10360, <https://doi.org/10.1074/jbc.M113.458059>
- Todaro, D.R., Augustus-Wallace, A.C., Klein, J.M. and Haas, A.L. (2017) The mechanism of neural precursor cell expressed developmentally down-regulated 4-2 (Nedd4-2)/NEDD4L-catalyzed polyubiquitin chain assembly. *J. Biol. Chem.* **292**, 19521–19536, <https://doi.org/10.1074/jbc.M117.817882>
- Todaro, D.R., Augustus-Wallace, A.C., Klein, J.M. and Haas, A.L. (2018) Oligomerization of the HECT ubiquitin ligase NEDD4-2/NEDD4L is essential for polyubiquitin chain assembly. *J. Biol. Chem.* **293**, 18192–18206, <https://doi.org/10.1074/jbc.RA118.003716>
- Rotin, D. and Kumar, S. (2009) Physiological functions of the HECT family of ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **10**, 398–409, <https://doi.org/10.1038/nrm2690>
- Plant, P.J., Yeger, H., Staub, O., Howard, P. and Rotin, D. (1997) The C2 domain of the ubiquitin protein ligase Nedd4 mediates Ca<sup>2+</sup>-dependent plasma membrane localization. *J. Biol. Chem.* **272**, 32329–32336, <https://doi.org/10.1074/jbc.272.51.32329>
- Staub, O., Dho, S., Henry, P., Correa, J., Ishikawa, T., McGlade, J. et al. (1996) WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na<sup>+</sup> channel deleted in Liddle's syndrome. *EMBO J.* **15**, 2371–2380, <https://doi.org/10.1002/j.1460-2075.1996.tb00593.x>
- Sanchez-Tena, S., Cubillos-Rojas, M., Schneider, T. and Rosa, J.L. (2016) Functional and pathological relevance of HERC family proteins: a decade later. *Cell. Mol. Life Sci.* **73**, 1955–1968, <https://doi.org/10.1007/s00018-016-2139-8>

- 16 Marin, I. (2010) Animal HECT ubiquitin ligases: evolution and functional implications. *BMC Evol. Biol.* **10**, 56, <https://doi.org/10.1186/1471-2148-10-56>
- 17 Grau-Bove, X., Sebe-Pedros, A. and Ruiz-Trillo, I. (2013) A genomic survey of HECT ubiquitin ligases in eukaryotes reveals independent expansions of the HECT system in several lineages. *Genome Biol. Evol.* **5**, 833–847, <https://doi.org/10.1093/gbe/evt052>
- 18 Bernassola, F., Chillemi, G. and Melino, G. (2019) HECT-Type E3 ubiquitin ligases in cancer. *Trends Biochem. Sci.* **44**, 1057–1075, <https://doi.org/10.1016/j.tibs.2019.08.004>
- 19 Zheng, N. and Shabek, N. (2017) Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157, <https://doi.org/10.1146/annurev-biochem-060815-014922>
- 20 Mathieu, N.A., Levin, R.H. and Spratt, D.E. (2021) Exploring the roles of HERC2 and the NEDD4L HECT E3 ubiquitin ligase subfamily in p53 signaling and the DNA damage response. *Front. Oncol.* **11**, 659049, <https://doi.org/10.3389/fonc.2021.659049>
- 21 Scheffner, M. and Kumar, S. (2014) Mammalian HECT ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochim. Biophys. Acta* **1843**, 61–74, <https://doi.org/10.1016/j.bbamcr.2013.03.024>
- 22 Kane, E.I., Waters, K.L. and Spratt, D.E. (2021) Intersection of redox chemistry and ubiquitylation: post-translational modifications required for maintaining cellular homeostasis and neuroprotection. *Cells* **10**, 2121
- 23 Melino, G., Cecconi, F., Pelicci, P.G., Mak, T.W. and Bernassola, F. (2019) Emerging roles of HECT-type E3 ubiquitin ligases in autophagy regulation. *Mol. Oncol.* **13**, 2033–2048, <https://doi.org/10.1002/1878-0261.12567>
- 24 Goto, J., Otaki, Y., Watanabe, T. and Watanabe, M. (2021) The role of HECT-Type E3 ligase in the development of cardiac disease. *Int. J. Mol. Sci.* **22**, 6065, <https://doi.org/10.3390/ijms22116065>
- 25 Singh, S., Ng, J., Nayak, D. and Sivaraman, J. (2019) Structural insights into a HECT-type E3 ligase AREL1 and its ubiquitination activities in vitro. *J. Biol. Chem.* **294**, 19934–19949, <https://doi.org/10.1074/jbc.RA119.010327>
- 26 Ogunjimi, A.A., Briant, D.J., Pece-Barbara, N., Le Roy, C., Di Guglielmo, G.M., Kavsak, P. et al. (2005) Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. *Mol. Cell.* **19**, 297–308, <https://doi.org/10.1016/j.molcel.2005.06.028>
- 27 Sander, B., Xu, W., Eilers, M., Popov, N. and Lorenz, S. (2017) A conformational switch regulates the ubiquitin ligase HUWE1. *Elife* **6**, e21036, <https://doi.org/10.7554/eLife.21036>
- 28 Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191, <https://doi.org/10.1093/bioinformatics/btp033>
- 29 Maples-Keller, J.L., Norrholm, S.D., Burton, M., Reiff, C., Coghlan, C., Jovanovic, T. et al. (2022) A randomized controlled trial of 3,4-methylenedioxymethamphetamine (MDMA) and fear extinction retention in healthy adults. *J. Psychopharmacol.* **36**, 368–377, <https://doi.org/10.1177/02698811211069124>
- 30 Singh, S. and Sivaraman, J. (2020) Crystal structure of HECT domain of UBE3C E3 ligase and its ubiquitination activity. *Biochem. J.* **477**, 905–923, <https://doi.org/10.1042/BCJ20200027>
- 31 Pandya, R.K., Partridge, J.R., Love, K.R., Schwartz, T.U. and Ploegh, H.L. (2010) A structural element within the HUWE1 HECT domain modulates self-ubiquitination and substrate ubiquitination activities. *J. Biol. Chem.* **285**, 5664–5673, <https://doi.org/10.1074/jbc.M109.051805>
- 32 Gong, W., Zhang, X., Zhang, W., Li, J. and Li, Z. (2015) Structure of the HECT domain of human WWP2. *Acta Crystallogr F. Struct. Biol. Commun.* **71**, 1251–1257, <https://doi.org/10.1107/S2053230X1501554X>
- 33 Maspero, E., Mari, S., Valentini, E., Musacchio, A., Fish, A., Pasqualato, S. et al. (2011) Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation. *EMBO Rep.* **12**, 342–349, <https://doi.org/10.1038/embo.2011.21>
- 34 Rosano, G.L. and Ceccarelli, E.A. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, 172, <https://doi.org/10.3389/fmicb.2014.00172>
- 35 Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411–421, [https://doi.org/10.1016/S0958-1669\(99\)00003-8](https://doi.org/10.1016/S0958-1669(99)00003-8)
- 36 Gopal, G.J. and Kumar, A. (2013) Strategies for the production of recombinant protein in *Escherichia coli*. *Protein J.* **32**, 419–425, <https://doi.org/10.1007/s10930-013-9502-5>
- 37 Meng, X., Wang, C., Rahman, S.U., Wang, Y., Wang, A. and Tao, S. (2015) Genome-wide identification and evolution of HECT genes in soybean. *Int. J. Mol. Sci.* **16**, 8517–8535, <https://doi.org/10.3390/ijms16048517>
- 38 Hristova, V.A., Beasley, S.A., Rylett, R.J. and Shaw, G.S. (2009) Identification of a novel Zn<sup>2+</sup>-binding domain in the autosomal recessive juvenile Parkinson-related E3 ligase parkin. *J. Biol. Chem.* **284**, 14978–14986, <https://doi.org/10.1074/jbc.M808700200>
- 39 Spratt, D.E., Martinez-Torres, R.J., Noh, Y.J., Mercier, P., Manczyk, N., Barber, K.R. et al. (2013) A molecular explanation for the recessive nature of parkin-linked Parkinson's disease. *Nat. Commun.* **4**, 1983, <https://doi.org/10.1038/ncomms2983>
- 40 Wauer, T. and Komander, D. (2013) Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J.* **32**, 2099–2112, <https://doi.org/10.1038/emboj.2013.125>
- 41 Riley, B.E., Lougheed, J.C., Callaway, K., Velasquez, M., Brecht, E., Nguyen, L. et al. (2013) Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nat. Commun.* **4**, 1982, <https://doi.org/10.1038/ncomms2982>
- 42 Trempe, J.F., Sauve, V., Grenier, K., Seirafi, M., Tang, M.Y., Menade, M. et al. (2013) Structure of parkin reveals mechanisms for ubiquitin ligase activation. *Science* **340**, 1451–1455, <https://doi.org/10.1126/science.1237908>
- 43 Attali, I., Tobelaim, W.S., Persaud, A., Motamedchaboki, K., Simpson-Lavy, K.J., Mashahreh, B. et al. (2017) Ubiquitylation-dependent oligomerization regulates activity of Nedd4 ligases. *EMBO J.* **36**, 425–440, <https://doi.org/10.15252/embj.201694314>
- 44 Yi, J.J., Berrios, J., Newbern, J.M., Snider, W.D., Philpot, B.D., Hahn, K.M. et al. (2015) An autism-linked mutation disables phosphorylation control of UBE3A. *Cell* **162**, 795–807, <https://doi.org/10.1016/j.cell.2015.06.045>

- 45 Ronchi, V.P. and Haas, A.L. (2012) Measuring rates of ubiquitin chain formation as a functional readout of ligase activity. *Methods Mol. Biol.* **832**, 197–218, [https://doi.org/10.1007/978-1-61779-474-2\\_14](https://doi.org/10.1007/978-1-61779-474-2_14)
- 46 Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S. et al. (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577–581, <https://doi.org/10.1002/humu.10212>
- 47 Halvardson, J., Zhao, J.J., Zaghlool, A., Wentzel, C., Georgii-Hemming, P., Mansson, E. et al. (2016) Mutations in HECW2 are associated with intellectual disability and epilepsy. *J. Med. Genet.* **53**, 697–704, <https://doi.org/10.1136/jmedgenet-2016-103814>
- 48 Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D. et al. (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221, <https://doi.org/10.1038/nature13908>
- 49 Berko, E.R., Cho, M.T., Eng, C., Shao, Y., Sweetser, D.A., Waxler, J. et al. (2017) De novo missense variants in HECW2 are associated with neurodevelopmental delay and hypotonia. *J. Med. Genet.* **54**, 84–86, <https://doi.org/10.1136/jmedgenet-2016-103943>
- 50 Broix, L., Jagline, H., Ivanova, E., Schmucker, S., Drouot, N., Clayton-Smith, J. et al. (2016) Mutations in the HECT domain of NEDD4L lead to AKT-mTOR pathway deregulation and cause periventricular nodular heterotopia. *Nat. Genet.* **48**, 1349–1358, <https://doi.org/10.1038/ng.3676>
- 51 Reuter, M.S., Tawamie, H., Buchert, R., Hosny Gebрил, O., Froukh, T., Thiel, C. et al. (2017) Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *JAMA Psychiatry* **74**, 293–299, <https://doi.org/10.1001/jamapsychiatry.2016.3798>
- 52 Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K. et al. (2015) Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588, <https://doi.org/10.1038/ng.3303>
- 53 Hashimoto, R., Nakazawa, T., Tsurusaki, Y., Yasuda, Y., Nagayasu, K., Matsumura, K. et al. (2016) Whole-exome sequencing and neurite outgrowth analysis in autism spectrum disorder. *J. Hum. Genet.* **61**, 199–206, <https://doi.org/10.1038/jhg.2015.141>
- 54 Ortega-Recalde, O., Beltran, O.I., Galvez, J.M., Palma-Montero, A., Restrepo, C.M., Mateus, H.E. et al. (2015) Biallelic HERC1 mutations in a syndromic form of overgrowth and intellectual disability. *Clin. Genet.* **88**, e1–e3, <https://doi.org/10.1111/cge.12634>
- 55 Charng, W.L., Karaca, E., Coban Akdemir, Z., Gambin, T., Atik, M.M., Gu, S. et al. (2016) Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Med. Genomics* **9**, 42, <https://doi.org/10.1186/s12920-016-0208-3>
- 56 Bramswig, N.C., Ludecke, H.J., Pettersson, M., Albrecht, B., Bernier, R.A., Cremer, K. et al. (2017) Identification of new TRIP12 variants and detailed clinical evaluation of individuals with non-syndromic intellectual disability with or without autism. *Hum. Genet.* **136**, 179–192, <https://doi.org/10.1007/s00439-016-1743-x>
- 57 Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P. et al. (2016) Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196, <https://doi.org/10.1038/nn.4352>
- 58 Clarke, C.M., Fok, V.T., Gustafson, J.A., Smyth, M.D., Timms, A.E., Frazer, C.D. et al. (2018) Single suture craniosynostosis: Identification of rare variants in genes associated with syndromic forms. *Am. J. Med. Genet. A* **176**, 290–300, <https://doi.org/10.1002/ajmg.a.38540>
- 59 Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C. et al. (2008) Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am. J. Hum. Genet.* **82**, 432–443, <https://doi.org/10.1016/j.ajhg.2007.11.002>
- 60 Moortgat, S., Berland, S., Aukrust, I., Maystadt, I., Baker, L., Benoit, V. et al. (2018) HUWE1 variants cause dominant X-linked intellectual disability: a clinical study of 21 patients. *Eur. J. Hum. Genet.* **26**, 64–74, <https://doi.org/10.1038/s41431-017-0038-6>
- 61 Friez, M.J., Brooks, S.S., Stevenson, R.E., Field, M., Basehore, M.J., Ades, L.C. et al. (2016) HUWE1 mutations in Juberg-Marsidi and Brooks syndromes: the results of an X-chromosome exome sequencing study. *BMJ Open* **6**, e009537, <https://doi.org/10.1136/bmjopen-2015-009537>
- 62 Strauss, K.A., Gonzaga-Jauregui, C., Brigatti, K.W., Williams, K.B., King, A.K., Van Hout, C. et al. (2018) Genomic diagnostics within a medically underserved population: efficacy and implications. *Genet. Med.* **20**, 31–41, <https://doi.org/10.1038/gim.2017.76>
- 63 Niranjana, T.S., Skinner, C., May, M., Turner, T., Rose, R., Stevenson, R. et al. (2015) Affected kindred analysis of human X chromosome exomes to identify novel X-linked intellectual disability genes. *PLoS ONE* **10**, e0116454, <https://doi.org/10.1371/journal.pone.0116454>
- 64 Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F. et al. (2016) Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* **18**, 696–704, <https://doi.org/10.1038/gim.2015.148>
- 65 Nagy, V., Hollstein, R., Pai, T.P., Herde, M.K., Buphamalai, P., Moeseneder, P. et al. (2019) HACE1 deficiency leads to structural and functional neurodevelopmental defects. *Neurol. Genet.* **5**, e330, <https://doi.org/10.1212/NXG.0000000000000330>
- 66 Deng, H.X. (2019) HACE1, RAC1, and what else in the pathogenesis of SPPRS? *Neurol. Genet.* **5**, e326, <https://doi.org/10.1212/NXG.0000000000000326>
- 67 Sager, G., Turkiyilmaz, A., Ates, E.A. and Kutlubay, B. (2021) HACE1, GLRX5, and ELP2 gene variant cause spastic paraplegias. *Acta Neurol. Belg.* **391–399**, <https://doi.org/10.1007/s13760-021-01685-3>
- 68 Willsey, A.J., Fernandez, T.V., Yu, D., King, R.A., Dietrich, A., Xing, J. et al. (2017) De novo coding variants are strongly associated with Tourette disorder. *Neuron* **94**, 486e489–499e489, <https://doi.org/10.1016/j.neuron.2017.04.024>
- 69 Lear, T.B., McKelvey, A.C., Evankovich, J.W., Rajbhandari, S., Coon, T.A., Dunn, S.R. et al. (2019) KIAA0317 regulates pulmonary inflammation through SOCS2 degradation. *JCI Insight* **4**, e129110, <https://doi.org/10.1172/jci.insight.129110>
- 70 Basel-Vanagaite, L., Yilmaz, R., Tang, S., Reuter, M.S., Rahner, N., Grange, D.K. et al. (2014) Expanding the clinical and mutational spectrum of Kaufman oculocerebrofacial syndrome with biallelic UBE3B mutations. *Hum. Genet.* **133**, 939–949, <https://doi.org/10.1007/s00439-014-1436-2>
- 71 Basel-Vanagaite, L., Dallapiccola, B., Ramirez-Solis, R., Segref, A., Thiele, H., Edwards, A. et al. (2012) Deficiency for the ubiquitin ligase UBE3B in a blepharophimosis-ptosis-intellectual-disability syndrome. *Am. J. Hum. Genet.* **91**, 998–1010, <https://doi.org/10.1016/j.ajhg.2012.10.011>
- 72 Chahrour, M.H., Yu, T.W., Lim, E.T., Ataman, B., Coulter, M.E., Hill, R.S. et al. (2012) Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet.* **8**, e1002635, <https://doi.org/10.1371/journal.pgen.1002635>

- 73 Yilmaz, R., Szakszon, K., Altmann, A., Altunoglu, U., Senturk, L., McGuire, M. et al. (2018) Kaufman oculocerebrofacial syndrome: Novel UBE3B mutations and clinical features in four unrelated patients. *Am. J. Med. Genet. A* **176**, 187–193, <https://doi.org/10.1002/ajmg.a.38538>
- 74 O’Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P. et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250, <https://doi.org/10.1038/nature10989>
- 75 Beasley, S.A., Kellum, C.E., Orlomski, R.J., Idrizi, F. and Spratt, D.E. (2020) An Angelman syndrome substitution in the HECT E3 ubiquitin ligase C-terminal lobe of E6AP affects protein stability and activity. *PLoS ONE* **15**, e0235925, <https://doi.org/10.1371/journal.pone.0235925>
- 76 Martinez, F., Caro-Llopis, A., Rosello, M., Oltra, S., Mayo, S., Monfort, S. et al. (2017) High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J. Med. Genet.* **54**, 87–92, <https://doi.org/10.1136/jmedgenet-2016-103964>
- 77 Anglesio, M.S., Evdokimova, V., Melnyk, N., Zhang, L., Fernandez, C.V., Grundy, P.E. et al. (2004) Differential expression of a novel ankyrin containing E3 ubiquitin-protein ligase, Hace1, in sporadic Wilms’ tumor versus normal kidney. *Hum. Mol. Genet.* **13**, 2061–2074, <https://doi.org/10.1093/hmg/ddh215>
- 78 Su, C., Wang, T., Zhao, J., Cheng, J. and Hou, J. (2019) Meta-analysis of gene expression alterations and clinical significance of the HECT domain-containing ubiquitin ligase HUWE1 in cancer. *Oncol. Lett.* **18**, 2292–2303, <https://doi.org/10.3892/ol.2019.10579>
- 79 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589, <https://doi.org/10.1038/s41586-021-03819-2>
- 80 Hunkeler, M., Jin, C.Y., Ma, M.W., Monda, J.K., Overwijn, D., Bennett, E.J. et al. (2021) Solenoid architecture of HUWE1 contributes to ligase activity and substrate recognition. *Mol. Cell.* **81**, 3468e3467–3480e3467, <https://doi.org/10.1016/j.molcel.2021.06.032>
- 81 Li, K.B. (2003) ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* **19**, 1585–1586, <https://doi.org/10.1093/bioinformatics/btg192>
- 82 Procter, J.B., Carstairs, G.M., Soares, B., Mourao, K., Ofoegbu, T.C., Barton, D. et al. (2021) Alignment of Biological Sequences with Jalview. *Methods Mol. Biol.* **2231**, 203–224, [https://doi.org/10.1007/978-1-0716-1036-7\\_13](https://doi.org/10.1007/978-1-0716-1036-7_13)
- 83 Cole, C., Barber, J.D. and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201, <https://doi.org/10.1093/nar/gkn238>
- 84 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242, <https://doi.org/10.1093/nar/28.1.235>
- 85 Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R. et al. (2021) Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891, <https://doi.org/10.1093/nar/gkaa942>
- 86 Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788, <https://doi.org/10.1093/nar/gkg563>
- 87 Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. and Uversky, V.N. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **1804**, 996–1010, <https://doi.org/10.1016/j.bbapap.2010.01.011>
- 88 Edelheit, O., Hanukoglu, A. and Hanukoglu, I. (2009) Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol.* **9**, 61, <https://doi.org/10.1186/1472-6750-9-61>
- 89 Liu, H. and Naismith, J.H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91, <https://doi.org/10.1186/1472-6750-8-91>