




REVIEW

Characteristics of the Kelch domain containing (KLHDC) subfamily and relationships with diseases

Courtney Pilcher^{1,2} , Paula Armina V. Bucu^{2,3} , Jia Q. Truong¹, Paul A. Ramsland^{1,4,5},
Monique F. Smeets², Carl R. Walkley^{2,3,6,7} and Jessica K. Holien^{1,2,3} 

¹ School of Science, STEM College, RMIT University, Melbourne, Australia

² St Vincent's Institute of Medical Research, Fitzroy, Australia

³ Department of Medicine, Eastern Hill Academic Centre, Melbourne Medical School, The University of Melbourne, Carlton, Australia

⁴ Department of Immunology, Monash University, Melbourne, Australia

⁵ Department of Surgery, Austin Health, The University of Melbourne, Melbourne, Australia

⁶ Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research, Melbourne, Australia

⁷ Department of Molecular and Translational Science, Monash University, Melbourne, Australia

Correspondence

J. K. Holien, School of Science, STEM College, RMIT University, Bundoora, Melbourne, VIC 3082, Australia
Tel: +61 3 9925 7256
E-mail: jessica.holien@rmit.edu.au

(Received 28 October 2024, revised 8 January 2025, accepted 13 January 2025, available online 30 January 2025)

doi:10.1002/1873-3468.15108

Edited by Hee-Jung Choi

The Kelch protein superfamily is an evolutionary conserved family containing 63 alternate protein coding members. The superfamily is split into three sub-families: Kelch like (KLHL), Kelch-repeat and bric-a-bracs (BTB) domain containing (KBTBD) and Kelch domain containing protein (KLHDC). The KLHDC subfamily is one of the smallest within the Kelch superfamily, containing 10 primary members. There is little known about the structures and functions of the subfamily; however, they are thought to be involved in several cellular and molecular processes. Recently, there have been significant structural and biochemical advances for KLHDC2, which has aided our understanding of other KLHDC family members. Furthermore, small molecules directly targeting KLHDC2 have been identified, which act as tools for targeted protein degradation. This review utilises this information, in conjunction with a thorough exploration of the structural aspects and potential biological functions to summarise the relationship between KLHDCs and human disease.

Keywords: Kelch family; KLHDC; protein degradation; protein function

The Kelch protein superfamily

The Kelch domain was first described as a repeat element in the sequence of the *Drosophila* ORF1 protein in 1993 [1]. Since then, multiple Kelch-repeat proteins

have been identified through sequencing a diverse range of organisms including viruses, plants, fungi and mammals [2]. The Kelch domain consists of several

Abbreviations

ASK1, apoptosis signal-regulating kinase 1; BACK, BTB and C-terminal Kelch; BTB, bric-a-bracs; CCT, chaperonin-containing TCP1; CRISPR, clustered regularly interspaced short palindromic repeats; Cryo-EM, Cryo-electron microscopy; CTLH, alpha-helical segment, C-terminal to LiSH; CUL, Cullin; DesCEND, destruction *via* C-end degrons; DNA, deoxyribonucleic acid; ELO, elongin; ER, endoplasmic reticulum; GST, glutathione S-transferase; HGNC, HUGO Gene Nomenclature Committee; KBTBD, Kelch-repeat and bric-a-bracs; KLHDC, Kelch domain containing protein; KLHL, Kelch like; LiSH, Lissencephaly-1; MKLN1, Muskeliin 1; NMR, nuclear magnetic resonance; NSCLC, non-small-cell lung cancer; PDP2, pyruvate dehydrogenase; POZ, poxvirus and zinc finger domain; PP5, protein phosphatase 5; PPP1R15A, protein phosphatase 1 regulatory subunit 15A; PROTAC, proteolysis targeting chimeras; PTOV, prostate tumour-overexpressed protein; RABEPK, Rab9 effector protein with Kelch motifs; RBX, ring-box protein; RBX1, ring-box protein 1; RING, really interesting new gene; RNA, ribonucleic acid; Sec, selenocysteine; Sel, selenoprotein; SEPHS2, selenophosphate synthetase 2; SRR, substrate recognition receptor; TCAP, telethonin; TSPYL, testis-specific Y-encoded-like; Ub, ubiquitin; UPS, ubiquitin proteasomal system; USP, ubiquitin-specific peptidase; UTR, untranslated region.

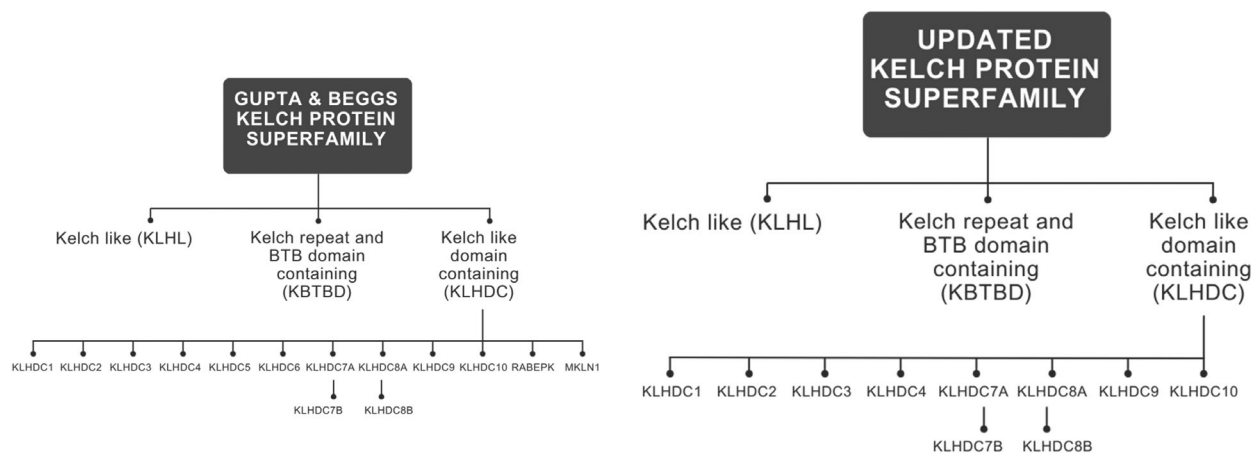


Fig. 1. Kelch protein superfamily. There are three subfamilies – Kelch like (KLHL), Kelch-repeat and bric-a-bracs domain containing (KBTBD) and Kelch domain containing protein (KLHDC). Initial phylogenetic analysis suggested 12 members of the KLHDC subfamily (left); however, recent evidence has suggested that there are only 10 subfamily members [95].

Kelch repeats and is one of the largest evolutionary conserved protein folds [1,2]. Each Kelch-repeat structure, termed a ‘blade’, consists of three to four twisted antiparallel β -strands connected by intrablade loops. Five to seven Kelch repeats form the distinct ‘ β -propeller’ of the Kelch domain [3]. Kelch domains function as scaffolds for protein–protein interactions, primarily binding substrates for ubiquitination and subsequent degradation.

The complement of human Kelch proteins was initially defined with 57 Kelch-repeat proteins. The list was subsequently expanded to include 63 alternate genes for Kelch-repeat domain containing proteins [1–4], of which three are non-protein-coding genes [3]. These were initially grouped into five subgroups [1,5] defined by the position of the β -propeller domain and the presence of other protein domains within the primary sequences. However, emerging details surrounding the discovery of BACK domain [6] and review of the KLHL (Kelch-like) literature [5,6] ultimately led to the update of this group, as proposed by Gupta and Beggs [3]. This redefined the five subgroups into three – KLHL, KBTBD (Kelch repeat and broad complex, Tram track domain containing) and KLHDC (Kelch domain containing) – based on the number of Kelch repeats in addition to the presence of alternative domains (Fig. 1). Additionally, a phylogenetic analysis of the human Kelch family proteins was also completed by Gupta and Beggs [3], which showed that most members in each of the subfamilies (KLHL, KBTBD and KLHDC) largely clustered together with the KLHDC subfamily, suggesting that KLHDC subfamily diverged first with a gain or loss of Kelch family domains occurring at later stages of evolution.

Since this review focuses on the KLHDC family, it will not describe the other domains in detail, but these have been well documented by others [2–7].

The KLHL subfamily contains all atypical Kelch protein domains, that is, an N-terminal BTB (bric-a-brac, tram track, broad complex)/POZ (poxvirus and zinc finger) domain, a BACK (for BTB and C-terminal Kelch) [6] domain and between two and eight Kelch repeats. KLHL is known to form a complex with Cullin3 and RBX1 [8–10], where the BTB/POZ domain is responsible for binding to Cullin3, while the Kelch domain determines the substrate specificity, and the BACK domain links both BTB and Kelch domains [11].

The KBTBD subfamily members typically lack the BACK domain [6], highlighting that these likely diverged in the later stages of evolution [3]. The KLHDC subfamily does not contain either BTB or BACK domains [5,6]; however, some have alternate domains such as the transmembrane and glycine rich domains [3], which are known to be functionally important.

Kelch-like domain containing protein (KLHDC) subfamily

In 2014, Gupta and Beggs [3] proposed a KLHDC subfamily that contained 12 primary members (Fig. 1), with some members having 1–3 additional isoforms (denoted with *) – KLHDC1, KLHDC2*, KLHDC3, KLHDC4*, KLHDC7A and KLHDC7B, KLHDC8A and KLHDC8B, KLHDC9*, KLHDC10, RABEPK and MKLN1, based on a phylogenetic analysis using each of their Kelch domain sequences. Additionally,

two more proteins (KLHDC5 and KLHDC6) have also been classified as KLHDC subfamily members previously. However, we propose a revised number of members from 12 to 10 (Fig. 1), with the following justifications. In the case of KLHDC5, experimental data obtained by Cummings *et al.* [12] initially found that KLHDC5 contained a BTB domain; additionally, it was Dhanoa *et al.* [5] who also indicated that based on the structure (presence of BTB, BACK and 5–6 Kelch repeats), KLHDC5 should be reannotated from KLHDC subfamily (KLHDC5) to the KLHL subfamily (KLHL42). Furthermore, these structural findings are also supported by the ALPHAFOLD prediction server [13], which predicts KLHL42 (Uniprot Entry [14]: Q9P2K6, ALPHAFOLD [13] entry: AF-Q9P2K6-F1) to have both a BTB and BACK domain with ‘very high’ model confidence. Consequently, nomenclature and classification systems, such as the HUGO Gene Nomenclature Committee (HGNC) in the human genome [15] and UniProt [14], have also recognised these updates and have since reclassified KLHDC5 to KLHL42. KLHDC6 is relatively less researched; however, complete sequencing of full length human cDNAs [16], the use of domain prediction software such as INTERPRO ([14,17]: Q3ZCT8) and some initial alignment and sequence searches revealed KLHDC6 to contain a BTB domain [18]; hence, KLHDC6 has now been reclassified to KBTBD12. Like in the case of KLHL42, the structural findings are also supported by ALPHAFOLD prediction server [13], which predicts that KBTBD12 (Uniprot entry [14]: Q3ZCT8, ALPHAFOLD [13]: AF-Q3ZCT8-F1) contains a BTB domain with ‘very high’ model confidence. This reclassification is also acknowledged by HGNC [15] and UniProt [14], who have also reflected this change within their online servers.

Finally, while Muskulin 1 (MKLN1) and Rab9 effector protein with Kelch motifs (RABEPK) show initial characteristics that could warrant their inclusion in the KLHDC subfamily – such as the presence of similar structural domains (Kelch-repeat motifs) – further evidence, such as an up to date and detailed phylogenetic and structural analysis to confidently classify them within the KLHDC subfamily, is severely lacking. MKLN1 contains both a LiSH (Lissencephaly-1 homology – an alpha helical region of approximately 60 amino acids at the N terminus) motif and a CTLH (alpha-helical segment, C-terminal to LiSH) motif [19]. Additionally, experimental data [19] identified an N-terminal region as a predicted discoidin-like domain, making it the only Kelch-repeat protein identified in the human genome to possess a discoidin domain [2]. The presence of these motifs and domains

distinguishes MKLN1 from other Kelch-repeat proteins [19]. Furthermore, through sequence identity comparison of MKLN1 and the rest of the KLHDC subfamily, MKLN1 has the lowest percent identity (average ~ 16%) (Table 1), and thus, we believe that MKLN1 cannot be classified at this time specifically within the KLHDC subfamily. Similarly, RABEPK also shares a low percent identity (average ~ 19%) (Table 1), which is likely reflective of the presence of the Kelch-repeat motifs; however, this suggests weak evolutionary conservation. Additionally, the functional role of RABEPK diverges significantly to that of the KLHDC subfamily, with it primarily acting as an effector protein for Rab9, a small GTPase involved in vesicle transport [20,21]. This has little overlap with the functions typically associated with the KLHDC subfamily; for example, it requires binding of GTP for activity [21].

In comparison to the KLHL and KBTBD subfamilies, the KLHDC subfamily has a distinct lack of literature specifically pertaining to their physiological roles, structural domains and function. KLHDC2 is currently the only member of the KLHDC subfamily to have an X-ray crystal structure, as well as extensive literature elucidating its function [22–26]. Despite the lack of data, what is known about KLHDC2, and other proteins containing related motifs, can be used to predict and understand the functions of other KLHDC members.

As described above, KLHDC family proteins lack both the BTB and BACK domains found in the other Kelch subfamilies. However, multiple KLHDC members contain alternate domains, many with known function (Fig. 2). The variety of domains within the KLHDC subfamily is likely to contribute to functional differences [3] and suggests that there is extensive diversity in the biological functions of KLHDC members [4,27].

Despite being of the same subfamily, there is low sequence identity (< 26%) among the KLHDC members (Table 1, Fig. 3), with some exceptions. For example, KLHDC1 and KLHDC2 share the highest sequence identity (44.24%), which aligns with the phylogenetic prediction that they are closely related. Similarly, KLHDC8A and KLHDC8B also share one of the highest levels of sequence identity (46%), alongside KLHDC7A and KLHDC7B, both of which are also phylogenetically predicted to be closely related. However, KLHDC7A, KLHDC7B, KLHDC8A and KLHDC8B also have the lowest sequence identity with the remaining members of the family (< 21%).

Unsurprisingly, the region sharing the most identity between the subfamily members is within the Kelch β -

Table 1. Sequence identity comparison of all KLHDC subfamilies. Sequences obtained and aligned in UniProt [14]. Where isoforms are known to exist, only the first isoform was used. Coloured based on % identity with high identity coloured green and low identity coloured red.

Uniprot code	Protein name	Q8N7A1	Q9Y2U9	Q9BQ90	Q8TBB5	Q5VTJ3	Q96G42	Q81YD2	Q81XV7	Q8NEP7	Q6PID8	Q7Z6M1	Q9UL63
% identity	KLHDC1	100											
	KLHDC2	44.24	100										
	KLHDC3	25.29	25.08	100									
	KLHDC4	20.29	19.88	20.06	100								
	KLHDC7A	16.67	15.07	16.85	15.43	100							
	KLHDC7B	19.17	16.25	21.32	17.42	38.27	100						
	KLHDC8A	16.67	14.69	18.64	13.73	24.91	21.16	100					
	KLHDC8B	17.27	14.29	15.6	17.07	20.77	20.95	46	100				
	KLHDC9	22.18	22.39	18.93	20.83	14.18	15.64	16.36	16.22	100			
	KLHDC10	25.38	23.93	24.93	21.04	17.17	21.16	19.86	21.16	18.49	100		
	RABEPK	23.12	25.3	20.66	18.32	12.94	18.62	18.45	16.54	18.55	23.95	100	
	MKLN1	20.61	17.96	16.67	17.77	11.38	14.08	16.72	13.64	18.67	16.49	16.76	100

propeller domain. Bioinformatic analysis of all known Kelch-repeat proteins [1] highlighted eight key conserved residues within each Kelch repeat, which allows for the formation of the β -propeller [28] (Fig. 3). Specifically, these include a set of four hydrophobic residues, followed by a double glycine element separated from two aromatic residues, most often a tyrosine and tryptophan which was found to be conserved in 90% of the sequences [1,28] (Fig. 3). For the KLHDC subfamily specifically, most members contain six Kelch repeats forming the β -propeller Kelch domain, except for KLHDC8A and KLHDC8B which contain seven (Figs 2 and 4). Most of the members conserve the characteristic gly-gly doublet within each Kelch repeat. However, there are varying degrees of conservation of the other conserved Kelch residues. Only one di-glycine element is conserved across every member of the KLHDC subfamily, while six other conserved residues are observed at varying degrees across all members; except for members KLHDC4 and KLHDC7A which only have three and two di-gly conserved elements, respectively, that align with the remaining family members (Fig. 3). It is also important to note that there is variable spacing between the glycine pair and conserved aromatic residues within each Kelch repeat, which does not always appear to form a consistent pattern [3].

KLHDC structure

Only KLHDC2 has an experimentally determined structure (X-ray crystal, Cryo-EM, NMR) available [22,23,26]. There are currently 10 experimental structures of KLHDC2 in the protein data bank (Oct 2024) [29]. Eight of these are X-ray crystal structures (PDB IDs: 8SGF, 8SGE, 8PIF, 8EBM, 8EBL, 6DO5, 6DO4, 6DO3), containing only the Kelch domain in complex with bound substrates. In 2023, an additional X-ray crystal structure was determined, containing the full KLHDC2 structure in complex with bound adaptor proteins ELOB and ELOC (PDB: 8EBN). This structure revealed that the functional complex was likely to be a tetrameric structure, containing four KLHDC2 monomers with their adaptor proteins [22]. More recently, this tetrameric assembly was also supported by the determination of a Cryo-EM structure (PDB: 8SH2) and adjacent biochemical studies which highlighted that this tetrameric structure can be modulated by substrate engagement [24]. This structure also depicts a flexible 'lasso' type region formed by amino acids 363–406, in which the end of this lasso (amino acids 400–406) can bind into the substrate binding pocket of another KLHDC2 within the same complex,

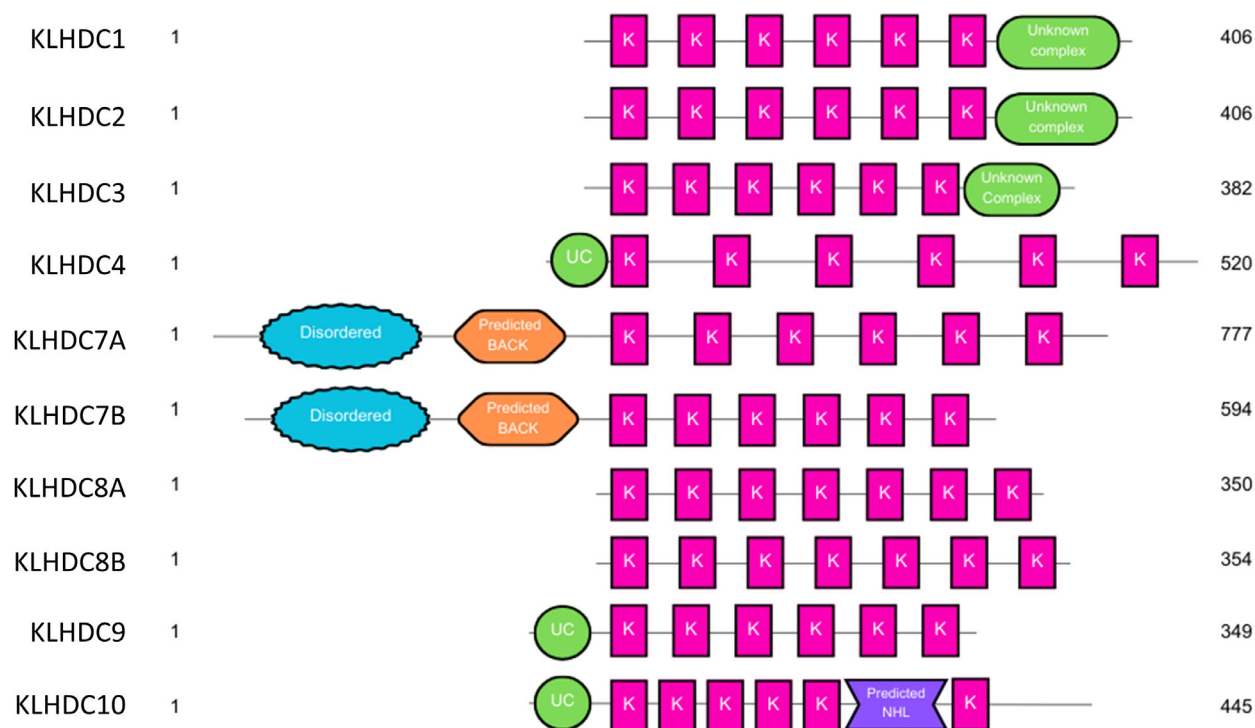


Fig. 2. Schematic of structural differences between each KLHDC sub-family member. In order top to bottom: KLHDC1, KLHDC2, KLHDC3, KLHDC4, KLHDC7A, KLHDC7B, KLHDC8A, KLHDC8B, KLHDC9, KLHDC10. Pink box: single Kelch repeat; green circle/oval: unknown complex; orange hexagon: predicted BACK complex; blue bubble: disordered region; purple inverted hexagon: predicted NHL repeat complex. Analysis was completed using ALPHAFOLD structures; these structures, along with an X-ray crystal structure of KLHDC2 (PDB: 8EBN, with other complexes removed), were used to analyse and compare the structural domains of the different family members.

and thus, the C-terminal region is able to form its own substrate.

For the purposes of this review and to analyse the structural aspects of all other KLHDC members, ALPHAFOLD2 [13] structures were obtained (Fig. 4). These structures, along with an X-ray crystal structure of KLHDC2 (PDB: 8EBN [22], with other complexes removed), were used to analyse and compare the structural domains of the different family members. Additionally, INTERPRO [17], a bioinformatics resource that provides a functional analysis of protein sequences by classifying them into families and predicting their domains and important sites, was also used to determine possible functions of the additional structural domains. These two sets of analyses are described below.

Kelch domain

The intrablade loops which connect each β -strand together are highly variable between each of the different KLHDC subfamily proteins. Additionally, the loops that connect each repeat to each other are also

highly variable, with ranges in length predicted to be between 1 residue and ~ 35 residues. Since the conserved β -propeller shape without these loops would be a flat, conserved surface, these differences in loop lengths lead to differences in the 3D shape of the proteins [1]. The centre of the Kelch domain is where the substrates are shown to bind [24,30], and thus, these loops likely alter what other proteins each family member interacts with, leading to alternate functions [1,26].

Additional structural domains

Another feature of some KLHDC subfamily members is the presence of additional (non-Kelch) structural domains, which may explain the wide variety of functions of KLHDC subfamily members.

Despite originally being defined as a KLHDC subfamily members, both KHLDC5 and KLHDC6 were predicted to contain BTB/POZ or BACK domains. Specifically, KLHDC5 was predicted to have a BTB/POZ domain between residues 5–72 and a BACK domain between residues 128–237, whereas KLHDC6

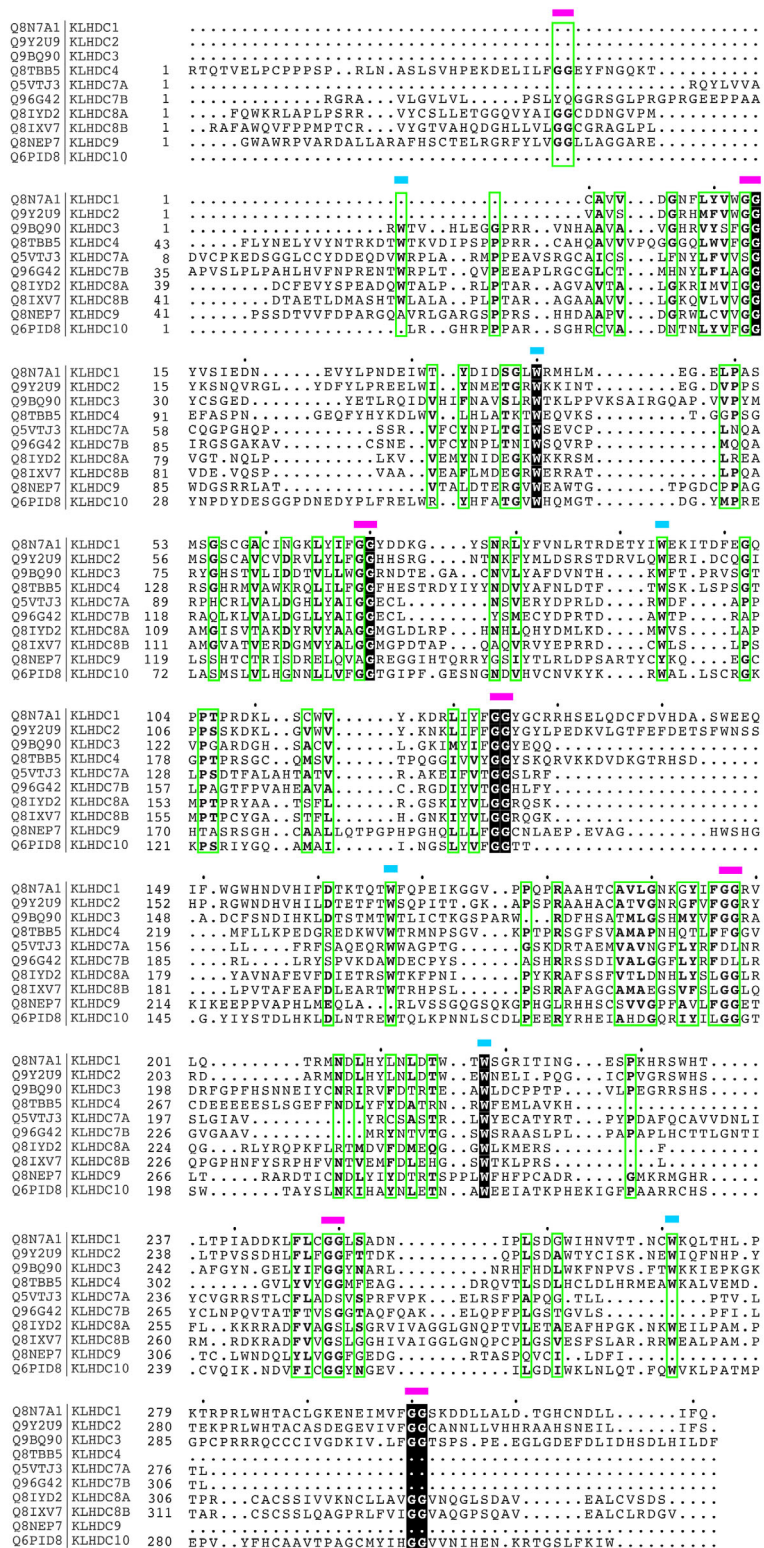


Fig. 3. Alignment of Kelch β -propeller of each of the KLHDC subfamily members. Protein sequences of human KLHDC subfamily members obtained from UniProt (codes within figure). Dots indicate every tenth amino acids of the protein sequence. Magenta bar: gly-gly doublet; cyan bar: aromatic residue Tryptophan. White letters on black background: identical amino acids; bold letters in green box: highly similar amino acids according to physico-chemical properties; standard letters: different amino acids. Alignment was performed using CLUSTAL OMEGA [96].

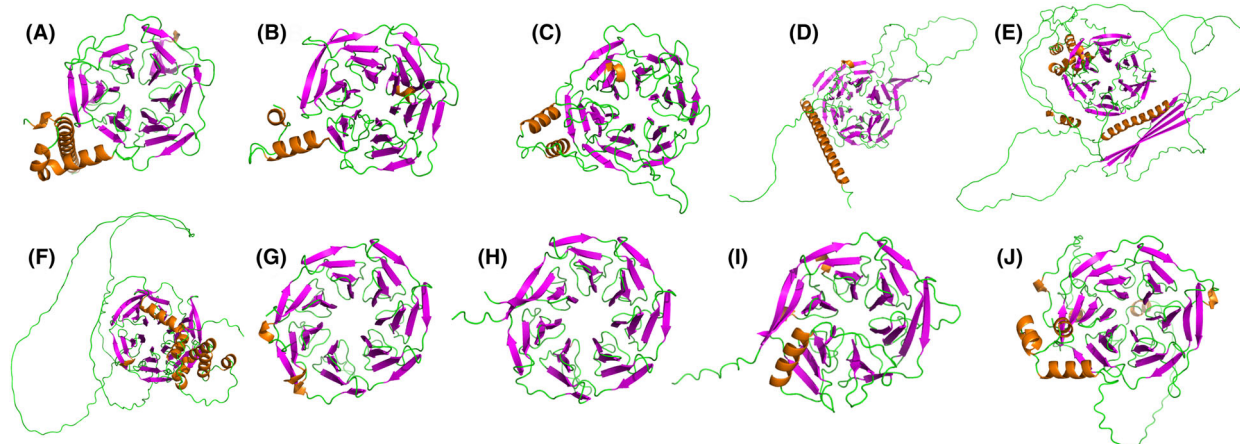


Fig. 4. Predicted structures of KLHDC sub-family. (A) Predicted crystal structure of KLHDC1 (AF_Q8N7A1-F1), (B) experimental crystal structure of KLHDC2 from Scott *et al.* [22] (PDB: 8EBN), (C) predicted crystal structure of KLHDC3 (AF-Q9BQ90-F1), (D) predicted crystal structure of KLHDC4 (AF-Q8TBB5-F1), (E) predicted crystal structure of KLHDC7A (AF-Q5VTJ3-F1), (F) predicted crystal structure of KLHDC7B (AF_Q96G42-F1), (G) predicted crystal structure of KLHDC8A (AF-Q8IYD2-F1), (H) predicted crystal structure of KLHDC8B (AF_Q8IXV7-F1), (I) predicted crystal structure of KLHDC9 (AF-Q8NEP7-F1) and (J) predicted crystal structure of KLHDC10 (AF-Q6PID8-F1). All structures obtained from ALPHAFOLD2 [13] and visualised in PYMOL (THE PYMOL MOLECULAR GRAPHICS SYSTEM, VERSION 3.0 SCHRÖDINGER, LLC.).

was predicted to contain a BTB/POZ domain at residues 28–128 and a BACK domain at residues 133–235 (Fig. 2). This prediction is supported by KLHDC5 also shown to interact with Cullin 3 (CUL3), which binds uniquely to BTB domains [12,31]. KLHDC7A was also predicted to have a BACK domain between residues 430–505. It is also predicted to be extracellular, at residues 39–777, although there is no literature to support this (Fig. 2). KLHDC10 is predicted to have an NHL repeat at residues 281–382, a highly conserved six-bladed β -propeller, present in a wide variety of eukaryotic and prokaryotic proteins (Fig. 2). Like a Kelch repeat, an NHL repeat is involved in protein binding; however, it is also suggested to be important for RNA recognition [32].

The remaining subfamily members – KLHDC1, KLHDC2, KLHDC3, KLHDC4, KLHDC7B and KLHDC9 – had no predictions as to their additional structural motifs, despite having primary sequences which are predicted to fold outside of the Kelch domain (Figs 2 and 3). Meanwhile, KLHDC8A and KLHDC8B do not appear to have any distinct structural elements in addition to the Kelch β -propeller domain (Figs 2 and 4). With so little known about their biology, it is unclear why these two members differ structurally and if they should be considered as a separate subgroup. Further research will have to elucidate this. While there are no predictions of the structural motifs of KLHDC2 or KLHDC3, research shows that the additional structural element – outside of the Kelch domain – is involved in interactions with

the CRL2 complex, comprising of CUL2, RBX1, Elongin B (ELOB) and Elongin C (ELOC), in order to facilitate substrate binding and trigger protein degradation (Fig. 2) [22,24,30,33,34].

Physiological roles of KLHDC family members

Localisation and expression

The KLHDC subfamily is distributed throughout the cell with some members shown to localise to the nucleoplasm (KLHDC2, KLHDC3, KLHDC9, KLHDC10) [35,36], nuclear speckles (KLHDC7A) [35], cytoplasm (KLHDC3) [7,33,37] and nucleoli (KLHDC4) [35]. This is suggestive of a role related to DNA and RNA biological processes. Some members are shown to reside in the cytosol such as KLHDC1 and KLHDC8B [35]. KLHDC1 was shown to be cytoplasmic, despite being fused to strong nuclear localisation signals, leading to the suggestion that it must be sequestered to the cytosol *via* unknown partner protein(s) [38]. KLHDC7B is located within the plasma membrane [35], which supports the prediction of this protein having an extracellular domain. The specific cell localisation of KLHDC8A remains unknown. Finally, in HeLa cells, KLHDC8B was shown to only be expressed in mitotic cells, concentrating in the region where the cells separate, suggesting a role in cytokinesis [39].

Where the tissue localisation and expression are known, most KLHDC subtypes are widely expressed

Table 2. The localisation, expression and physiological roles of KLHDCs. Individual references are within the table.

Gene name	Localisation	Expression	Function
KLHDC1	Intracellular – cytosol [36,38]	Widely expressed – high levels in skeletal muscle, pancreas and liver [38]	Enables ubiquitin ligase-substrate adaptor activity [60] Destabilises truncated selenoproteins [60] Involved in ubiquitin-dependent protein catabolic process <i>via</i> the C-end degron rule pathway [60]
KLHDC2	Intracellular – nucleoplasm, nuclear membrane, nuclear bodies [36,38]	Widely expressed – high levels in liver, testis [36], skeletal muscle, heart [38]	Substrate-recognition component of a Cul2-RING (CRL2) E3 ubiquitin-protein ligase complex of the DesCEND (destruction <i>via</i> C-end degrons) pathway [26,43,44] It may also act as an indirect repressor of CREB3-mediated transcription by interfering with CREB3-DNA-binding [76] Can engage the substrate binding domains of another protomer using its C-terminal Gly-Ser motif to mimic a degron [22]
KLHDC3	Intracellular – nucleoplasm [36]	Widely expressed – high levels in testis and skeletal muscle [7,36]	Involved in the activation of the V(D)J recombination [7] In mice, this gene is found to be expressed specifically in testis [7] CRL2 adaptor part of the ubiquitin-proteasome system and is known as a Cullin2-RING E3 ubiquitin ligase (CRL2) complex adaptor that recognises specific protein substrates and targets these for degradation by the CRL2 ubiquitin proteasome system [44,53] Targets proteins containing glycine at their C-terminal residues [26,44] Mediates ubiquitination and degradation of truncated SELENOV and SEPHS2 selenoproteins produced by failed UGA/Sec decoding, which end with a glycine [26,43]
KLHDC4	Intracellular – nucleoli and kinetochore [36]	Unknown	Unknown
KLHDC7	Intracellular – specific location unknown [36]	Unknown – however KLHDC7A is thought to express mainly in the kidney [36]	Believed to be responsible for protein coding [77]
KLHDC7A	Intracellular – nuclear speckles [36]		
KLHDC7B	Intracellular – plasma membrane [36]		
KLHDC8	Intracellular – specifics unknown [36]	Unknown – however KLHDC8A is thought to express in the ovary [36]	KLHDC8B is understood to be required for mitotic integrity and maintenance of chromosomal stability. Additionally, it is involved in pinching off the separated nuclei at the cleavage furrow and in cytokinesis I [41] and is thought to participate in cytokinesis, where reduced expression leads to formation of binucleated cells [39]
KLHDC8A	Intracellular – specifics unknown [36]		
KLHDC8B	Intracellular – cytosol [36]		
KLHDC9 (previously known as KARCA1)	Intracellular – nucleoplasm, nucleoli, nucleoli rim, mitotic chromosome [36]	Widely expressed – mainly within testis [36,78]	Believed to enable cyclin binding [78]
KLHDC10	Intracellular – nucleoplasm [36]	Widely expressed – high levels in testis and brain	Participates in oxidative stress-induced cell death through MAP3K5 activation [71] Inhibits PPP5C phosphatase activity on MAP3K5 [71] Specifically recognise proteins with a proline-glycine at the C terminus just as KLHDC3 [44]

[35] (Table 2). However, some members have been shown to be expressed at high levels in tissues involved in reproduction, specifically the testis (KLHDC1,

KLHDC3, KLHDC8B and KLHDC10) [40], epididymis (KLHDC8B) [40], seminal vesicle (KLHDC8B) [40] and endometrium and breast (KLHDC4) [40],

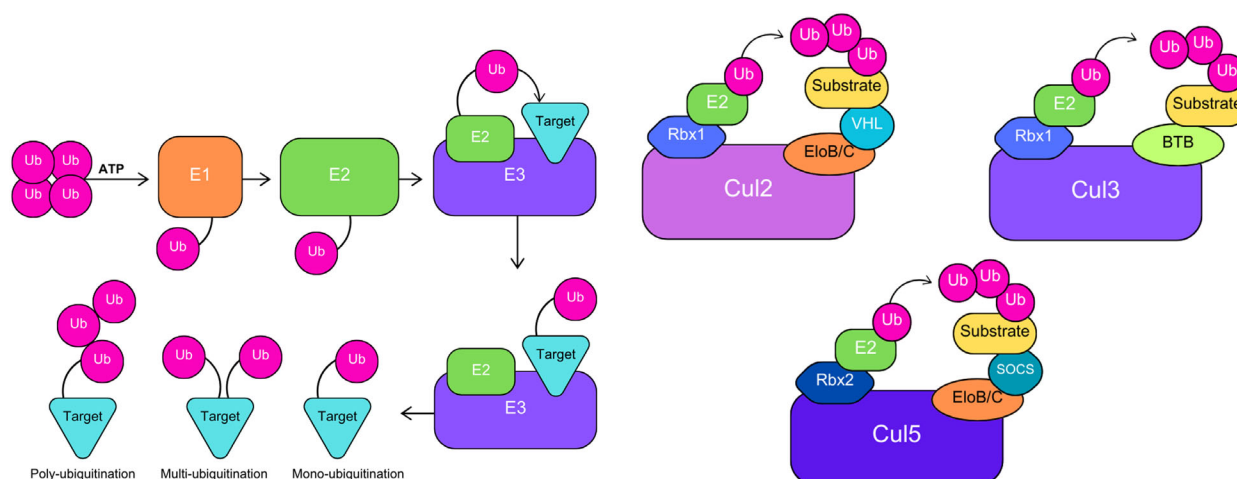


Fig. 5. Schematic of KLHDC role in ubiquitination. Left panel: Simplified schematic of the cascade process for ubiquitination to occur. E1 activates Ub through a thioester bond in an ATP-dependent process. Then, the activated Ub is 'transferred' to an E2 enzyme's active-site cysteine. Finally, an E3 ligase recognises the E2 complex and transfers Ub from E2 to the target substrate. Based on how Ub binds to the substrate, ubiquitination can be mono-, multi-, or poly-ubiquitination. Right panel: Schematic of three E3 CRL complexes specific to this review; Cul2, Cul3 and Cul5 recruit specific adaptor proteins (Elongin B/C for Cul2 and Cul5, BTB protein for Cul3) and receptor proteins (VHL-box for Cul2, SOCS for Cul5) to form CRL E3 ubiquitin ligases with the RING protein (Rbx1/2). These ligases transfer ubiquitin from Rbx1/2-bound E2 to substrate proteins. Figures adapted from [45].

highlighting their potential role in cell division, specifically meiotic recombination [41]. Additionally, KLHDC subfamily members may also play a potential role in regenerative cells since there is also a high expression of KLHDC1, KLHDC2 and KLHDC3 in skeletal muscle [7,38] and liver cells [40] (Table 2).

The KLHDC subfamily role in the ubiquitination pathway

All members of the KLHDC family are thought to be substrate receptors of E3 ligase complexes [3] responsible for targeting and tagging misfolded or otherwise damaged proteins for degradation [24]. The degradation of these proteins is generally happening through two methods, the ubiquitin proteasomal system and/or autophagy [24,42], which are both essential for the maintenance of homeostasis. As the name suggests, the ubiquitin proteasomal system (UPS) is the term for the cascade of events leading to the ubiquitination of specific proteins, that is, where ubiquitin is 'transferred' onto proteins by covalent conjugation. This process requires a set of three enzymes: E1s (ubiquitin activating enzymes), E2s (ubiquitin conjugating enzymes) and E3 ligases which selectively attach ubiquitin to lysine, serine, threonine or cysteine residues to the specific substrates in need of degradation [42] (Fig. 5). Specifically, KLHDC proteins are specifically essential for the destruction *via* C-end degrons

(DesCEND) mechanism of ubiquitination. This mechanism specifically targets substrates with specific C-terminal motifs for degradation [26,43,44].

E3 ligases are generally categorised according to their structure and function. These form complexes consisting of a scaffold subunit, a catalytic subunit and an adaptor protein. The complex then binds to a group of proteins that are responsible for substrate recognition or the substrate recognition receptor (SRR) [42,45]. It is thought that Kelch proteins act as SRRs in the cullin-RING (Really Interesting New Gene) ubiquitin ligase (CRL) network [46,47]. CRLs comprised four major components: Cullin proteins (CUL1, CUL2, CUL3, CUL4A/4B, CUL5 and CUL7) that act as the central scaffold which binds to a RING-box protein (in most cases RBX1 or RBX2) which then binds to a E2 ubiquitin conjugating enzyme at the N-terminal and an adaptor protein-substrate receptor complex which recognises the target protein and connects the SRRs back to the Cullin at the C-terminal [42,45]. Each CRL subfamily has a distinct set of adaptors or substrate recruiters. Of specific interest to this review are CUL2 and CUL5, which both interact with Elongin B and Elongin C (ELOB/C), and CUL3, which uses BTB/BACK/POZ (broad complex, tramtrack and bric-à-brac/poxvirus and zinc finger)-domain proteins as both adaptor and substrate receptor [45] (Fig. 5). It is these ELOB/C and BTB/BACK/POZ complexes

which recruit BC-box proteins to act as substrate receptors [34,48].

Since the KLHDC subfamily members lack BTB-BACK domains, they are required to recruit BC-Box proteins to enable efficient ubiquitination. The BC-box has a consensus sequence of (S,T,P)LXXX(C,S,A)XXXΦ, where X represents any amino acid, and Φ represents the hydrophobic amino acids [49]. It is then through these BC-box proteins that the selectivity of substrates occurs through the recognition of specific short peptide motifs termed degrons [50,51]. Emerging research has indicated that some KLHDC subfamily members such as KLHDC2 and KLHDC3 utilise the CRL2 complex to facilitate these previously described steps of ubiquitination [26,30,43,44,52–54].

KLHDC2

The most well-studied member is KLHDC2, which has been shown to be a substrate-recognition component of the Cul2-RING E3 ubiquitin-protein ligase complex which functions in the DesCEND pathway. Thus, the C-terminal region of CUL2 binds to RING-box protein 1 (RBX1) and the N-terminal region binds to Elongin B/C, followed by the BC-box proteins – VHL (Von Hippel–Lindau) being the prototype – to form this complex [43,48]. KLHDC2 recognises specific sequences located near the C-terminal regions of target proteins, known as C-degrons; the recognition of these C-degrons then leads to their ubiquitination and degradation [26,43,44]. These C-degrons bind within the substrate-binding pockets, located at the top of the β-propeller of the KLHDC proteins. Specifically of note is the di-glycine motif at the C-terminal [22,26], referred to as a di-Glycine degron, which is present in most substrates. This has been highlighted by crystal structures of KLHDC2 in complex with an early-terminated selenoprotein – SelK and SelKS – C-end degron peptide (PDB code: 6DO3 [26]). This selenoprotein SelK incorporates a selenocysteine (Sec), which is encoded by the stop codon UGA [55]. This codon is interpreted as Sec due to a Sec insertion sequence element in the 3′ untranslated region (UTR) of SelK mRNA, facilitating the synthesis of the complete protein. When selenium is deficient, the absence of Sec-transfer RNA leads to premature translation termination [56]. Recent research indicates that the truncated SelK protein produced under these conditions possesses a C-terminal di-Glycine degron, marking it for CRL2-KLHDC2-mediated proteasomal degradation [26]. Several additional X-ray structures such as a peptide derived from EPHB2 (PDB code: 8EBL [22]) have revealed that

KLHDC2 is also able to bind with other C-degron containing proteins.

The 3D structures, along with extensive biochemical studies, have shown that a specific Gly-Gly motif is essential at the C-terminal. However, in attempting to resolve how the KLHDC2-ELOB/C substrate receptor is regulated [22], these studies were also able to reveal that KLHDC2 (which terminates at the C-terminal end with a Gly-Ser [24]) can mimic a di-Glycine degron with its C-terminal 401-NNTSGS-406 C-degron mimic. This degron mimic has been shown to bind at lower affinity than the substrates, potentially due to the additional atoms of Ser406, causing Trp270 into a slightly less energetically favourable orientation [24]. Interestingly, structural data suggested that KLHDC2's activity could be self-regulated [22]. Using this degron mimic, KLHDC2 forms a complex where one KLHDC2 C terminus binds to another KLHDC2 protein, thus self-regulating its activity [22,24].

KLHDC2 has recently had ligands developed against it which provided pharmacological evidence that KLHDC2 could be a good PROteolysis TArgeting Chimeras (PROTAC) target for specific protein degradation [23–25]. PROTACs are designed as bifunctional molecules that consist of two main components. One part of the molecule binds to the target protein you want to degrade, and the other part recruits an E3 ubiquitin ligase, in this case KLHDC2, which tags the target protein for destruction. This approach allows for targeted degradation of proteins that are often difficult to inhibit with traditional small molecules, offering a new strategy for drug development, especially in cancer and other diseases.

KLHDC3 and KLHDC10 are also classified as CUL2-RING E3 ligases

KLHDC3 has been often studied concurrently with KLHDC2 [22]. Like KLHDC2, it is a substrate-recognition component of a CRL2 E3 ubiquitin-protein ligase [44,53]; however unlike KLHDC2 substrates which contain the diGlycine degron as discussed above, a study by Koren *et al.* [44] discovered that only the C-terminal glycine residue is absolutely critical for degradation; KLHDC3 recognises substrates through interaction with other exposed C-terminal degrons such as -RG, -KG or -QG motif. KLHDC3 was originally identified by its ability to distinguish defective selenoproteins by their C-terminal residues. Thus, it can mediate ubiquitination and degradation of truncated SelS and SEPHS2 selenoproteins produced by failed UGA/Sec decoding, which also end with a glycine [26,43]. Moreover, like KLHDC2,

KLHDC3 is able to form complexes with both ELOC and ELOB [22]. Notably, like KLHDC2's GlySer autoinhibitory motif, KLHDC3 contains a C-terminal HisGly motif and can form a similar tetrameric structure as KLHDC2 [22] although it is unclear whether it has the same autoinhibitory function. On the other hand, while KLHDC10 can form a complex with ELOB/C akin to its other subfamily members, KLHDC2 and KLHDC3, it was not able to form a tetrameric structure, instead remaining monomeric [22]. Since KLHDC10 has a C-terminal ending in Leu-Lys, this finding aligns with the fact that the C-terminal sequence has diverged from the cognate degron [22], potentially preventing the engagement of its substrate-binding pocket [24]. However, ALPHA-FOLD2 [13] models suggest the N terminus of KLHDC10 may fold back on itself and bind to the central substrate pocket of KLHDC10. The N terminus has Alanine in positions 3 and 4, and a modelled Ala-tail of at least 5 consecutive Alanine's with a short GFP-derived sequence was found to consistently bind to KLHDC10 [57]. Although these interactions are yet to be further validated and understood, they could suggest a similarity with KLHDC2's autoinhibition and a similar mechanism for this protein to regulate its activity.

KLHDC1 binds Cullin-5 for ubiquitin activity

Like CUL2, CUL5 also interacts with ELOB/C. However, unlike CUL2 which tends to form a complex with VHL box (BC box and CUL2 box)-containing proteins (named after VHL the prototype substrate receptor of CUL2), CUL5 usually forms a complex with SOCS box (BC box and CUL5 box) containing proteins of which SOCS2 is the prototype [45]. The interaction of CUL2 vs. CUL5 with ELOB/C is subtle and purely due to the difference in the CUL2 or CUL5 boxes of the proteins they each recruit [45,58,59]. KLHDC1 interacts with CUL5, forming a CUL5-RING ubiquitin ligase complex, which is thought to target proteins for ubiquitination and degradation in a manner similar to CUL2-KLHDC2 [48]. Additionally, KLHDC1 has recently been shown to target the same di-Gly degron and destabilise truncated selenoproteins [60].

KLHDC4, KLHDC7A, KLHDC7B, KLHDC8A, KLHDC8B, KLHDC9

Limited research exists for KLHDC4, which suggests that it may not be able to interact with CUL2 or CUL3 [61]. Structural analysis performed by Lian

et al. [61] suggested that KLHDC4 lacks the consensus modules required to interact with Cullin proteins, which supports the concept that KLHDC4 may exert its function through other means. No current information regarding their function within the ubiquitin system exists for KLHDC7A, KLHDC7B, KLHDC8A, KLHDC8B and KLHDC9.

Non-ubiquitin-related KLHDC functions are mostly related to protein coding and cell cycle regulation

Despite their large role in the ubiquitin pathway, there are some literature studies which support alternate roles for KLHDC family members. There is some evidence that KLHDC3 and KLHDC8B may be involved in cell cycle regulation [7,41]. KLHDC3 is structurally akin to recombination activating gene 2, a protein involved in the activation of the V(D)J recombination, thus has a proposed role in meiotic division [7], though this was based solely on the β -propeller of RAG2 which at the time of this publication was the most closely related structure. However, since KLHDC3 primarily localises to pachytene spermatocyte cytoplasm and meiotic chromatin [7], this could suggest that KLHDC3 plays a role in meiotic division. KLHDC8B is involved in pinching off the separated nuclei at the cleavage furrow and in cytokinesis 1, ultimately thought to protect cells against mitotic errors, centrosomal amplifications, micronucleus formation and aneuploidy [41].

Known protein–protein interactions (PPI) and substrates of KLHDCs

Although there is limited literature on the KLHDC subfamily and their physiological roles, as well as a lack of experimentally determined crystal structures, predicted PPIs and substrate binding can provide insights into the potential physiological functions of different KLHDC proteins. The functions of most KLHDC subfamily members have been derived from research not specific to KLHDC subfamily, and therefore, we have collated these data (Table 3) along with known/predicted protein interactions to better understand the roles of the subfamily.

Protein–protein interactions between KLHDC2, KLHDC3 and KLHDC10 and CUL2/ELOC/ELOB are supported by experimental research

PPIs of KLHDC2, KLHDC3 and KLHDC10 are among the highest predictions of the family. Discussed

previously, KLHDC2, KLHDC3 and KLHDC10 have shown to be SSRs of CUL2, as well as ELOB and ELOC, and thus, the prediction of interactions between these KLHDC subfamily members and CUL2 is unsurprising. However, there is also a predicted interaction between CUL5 and KLHDC2 (Table 3). While there is currently a lack of experimental data to support this interaction [34], we know that KLHDC1 and KLHDC2 share some sequence identity (17.8%), and like KLHDC1, KLHDC2 is able to recognise the diGly degron at the C terminus of truncated SELENOS, as previously described [60]. Thus, although it is possible that KLHDC2 interacts with CUL5, experimental validation would be required to confirm this prediction.

KLHDC-substrate relationships

The information gleaned from KLHDC2 can be used to understand the role of other KLHDC subfamily members. As previously discussed, KLHDC1, KLHDC2 and KLHDC3 have been shown to recognise specific degrons (diGly and single Gly at the C terminus, respectively), while KLHDC3 recognises substrates through interaction with other exposed C-terminal degrons, such as -RG, -KG or -QG motif [44].

As aforementioned, the early termination of both SelK and SelS proteins rendered their C-end degron to be -Gly-Gly, allowing KLHDC2 to target these proteins. The C-terminal peptide of these truncated selenoproteins binds to the centre of KLHDC2 between its propeller blades which form a binding pocket [26]. KLHDC1 has also been determined to bind to truncated SelS [60]. In addition, from a global protein stability assay, KLHDC3 was found to target both UGA-terminated selenoproteins SelS and selenophosphate synthetase 2 (SEPHS2) [62].

Only a handful of full-length proteins have been identified and validated as targets of KLHDCs. These include PTOV1, PDP2 and MIC19 for KLHDC2 and PPP1R15A, USP49, TCAP, p14ARF and TSPYL1 for KLHDC3 [43,44,54]. A study investigating the C-degron pathway provides a comprehensive list with 20 KLHDC2-binding full-length proteins identified through GST pull-down and mass spectrometry, most of which are diGly-ending, although not all have been verified [54].

In human cells, KLHDC3 binds p14ARF and targets it for degradation in a CRL2 complex-dependent manner through the recognition of p14ARF's C-terminal degron [33,63]. This degron is, however, not conserved in mice. It has also been shown that

KLHDC3 interacts with and contributes to the degradation of c-Myc in tandem with Fbxw7, functioning in the regulation of cell growth and tumour formation [37].

Recent studies on KLHDCs have emphasised that the conformation of the substrate is the determining factor when it comes to their binding tendencies. KLHDC3 recognises the C-degron sequence (-RGRG) of the herpesvirus UL49.5 protein which acts in ER-associated degradation pathways. A mutation leading to the protein having a similar conformation preserves crucial interactions, whereas a mutation that can significantly alter the C-end's conformation results in the loss of the substrate's ability to adopt in the KLHDC3 binding cleft [64]. A related study suggested the importance of an arginine residue being the fourth amino acid from the C terminus and a positively charged residue located directly upstream of the C-terminal Gly (at the penultimate position); however, more evidence will be necessary to support this [65]. This is also supported by others who established the crucial role of either a charged or polar residue, of which positively charged residues fit, as the penultimate amino acid for a degron to be recognised [54].

There is limited information on the rest of the KLHDC family's interacting substrates. KLHDC8A binds Chaperonin-Containing TCP1 (CCT) and is speculated to act as an adapter that facilitates CCT- α -tubulin interaction which ultimately contributes to the regulation of tubulin biogenesis [66]. KLHDC9, also known as KARCA1, interacts with the cyclin A1-CDK2 complex [67]. Finally, aside from the previously discussed potential of KLHDC10 to have interactions that are reminiscent of KLHDC2's C-degron mimicry, this member of the KLHDC family generally recognises substrates enriched with -WG, -PG and -AG endings [44]. KLHDC10 binds to the C-terminal phosphatase domain of protein phosphatase 5 (PP5). This in turn suppresses its phosphatase activity, although this function is not dependent on its role as a substrate receptor of the CRL2 complex [68]. KLHDC10 also binds to homopolymeric Alanine tails (with at least four Ala residues) and Alanine-rich C-degrons, mediating their degradation in the ribosome-associated quality control pathway [57,69].

Diseases related to the KLHDC family

Cancer

Unsurprising based upon their role in the ubiquitination pathway, some KLHDC subfamily members have been reported to play a role in the development of several

Table 3. Known Protein–protein interactions of KLHDC family members. Protein–protein interaction (PPI) data collated from various databases (BioGrid [79], IntAct [80] and STRING [81]) using default parameters. Each data set was analysed, and the top five known PPIs based on their highest amount of evidence surrounding their interactions listed. Shown are the eight C-terminal amino acids of each protein, anticipated to be the C-degron each KLHDC protein member interacts with.

Gene name	Number of protein–protein interactions per database			Top 5 known PP interactions (in no order)	C-terminal sequence
	BioGrid	IntAct	STRING		
KLHDC1	3	2	33	CUL5 [82] TCEB2 [82] SPPL3 [83] ELOB [84]	INTFIYMA SANEQAVQ SSSRFLEV SANEQAVQ
KLHDC2	133	64	61	CUL2 [54,85] CUL5 [86] APPBP2 [54] RBX1 [84] KLHDC3 [54,60]	ADEYSYVA INTFIYMA QNVEGPSC WEFQKYGH RPIVSSHG
KLHDC3	172	56	38	CUL2 [60] ELOB [54] RBX [60] APPBP2 [43] KLHDC2 [60]	ADEYSYVA SANEQAVQ WEFQKYGH QNVEGPSC GSNNTSGS
KLHDC4	41	20	26	ARMC6 [87] CROT [85,86] AHR [85] CES2 [85] ZCCHC14 [88]	GORGNLAP QLMNSTHL DLTSSGFL PEERHTEL AESLDSTD
KLHDC7A	2	–	5	KLHDC10 [3] ORF9B IGSF21 [89] SLC25A45 [90] PPM1J [90]	OGLIERLK EFVVVTVK LTVILELT EYLLRWWG LGGPGSYS
KLHDC7B	27	72	3	CLM2 [91] OR4F21 [92] ZNF487 [92] EPHB6 [93] A1CF [94]	PQWAPPGR LVIYKKIS VDFTQEEW RQQGSVEV RGDGYGTF
KLHDC8A	20	18	10	PEX7 [85] TXNDC9 [86] CCT2 [86] CCT3 [86] CCT4 [86]	PACLTIPA DSDSDDD RVPDHHPC GAPDAGQE IDDVVNTR
KLHDC8B	12	4	10	CCT2 [85] CCT3 [85] CCT5 [85] CCT7 [85,86] IMPDH1 [85,86]	RVPDHHPC GAPDAGQE RKPGESEE RGRGRPH HSYEKRLY
KLHDC9	37	5	10	CA3 [85] CAV3 [85] CCDC134 [85] CEACAM8 [85] CKB [85]	RVVRASFK KVVLKKEV ISRSQSEL VLARVALI DDLMPAQK
KLHDC10	74	20	10	CUL2 [44] ELOB [44] APPBP2 [44] KLHDC3 [44] KLHDC2 [44]	ADEYSYVA SANEQAVQ QNVEGPSC RPIVSSHG GSNNTSGS

tumours (Table 2). However, their precise relationship remains largely unknown. In non-small-cell lung cancer (NSCLC), high KLHDC3 expression is associated with poorer survival rates [53]. A CRISPR/Cas9 knockout of KLHDC3 significantly reduced lung cancer cell growth and suppressed tumour growth *in-vivo* [53]. The study also found that KLHDC3 interacts with p14ARF through its β -propellor, recruiting p14ARF to the CRL2 complex and promoting its N-terminal ubiquitylation [33]. This process leads to p14ARF degradation *via* the UPS, thereby promoting NSCLC proliferation [33,53]. KLHDC3 may also play roles in other types of tumours such as testicular and ovarian tumours [33,53]. Inhibiting KLHDC3 with peptides or small molecule inhibitors could offer a promising therapeutic strategy for cancers overexpressing KLHDC3, although no inhibitors currently exist. Similarly, KLHDC4 has been implicated in nasopharyngeal cancer by suppressing apoptosis, and CRISPR/Cas9 knockout of KLHDC4 has shown reduced tumour growth and migration [61], making it another potential therapeutic target, despite the lack of available inhibitors. KLHDC7A, KLHDC7B and KLHDC9 have been associated with breast and ovarian cancer, where their expression correlates with cancer cell proliferation, but these findings are based on transcriptome-wide studies and require further experimental validation.

Both KLHDC8A and KLHDC8B have been implicated in glioma cell apoptosis and classical Hodgkin lymphoma, respectively. KLHDC8A is significantly overexpressed in high-grade gliomas compared to low-grade and normal brain tissues. This expression is correlated with poor prognosis in patients [70]. The study showed that upregulation of KLHDC8A promotes glioma cell proliferation, migration and invasion while inhibiting apoptosis [70]. Additionally, the introduction of lactate induces KLHDC8A, linking metabolic changes to tumour progression [70]. Mutations of KLHDC8B were shown to be associated with an increased risk of developing classical Hodgkin lymphoma [70]. The study specifically highlights that these mutations could disrupt normal cell division processes, which lead to chromosomal instability. In addition to this finding, the study also found that KLHDC8B regulates proteins including Bcl2/BAX and p21/CDK2, all thought to indirectly influence the mitotic checkpoint, further highlighting the importance of maintaining genomic stability [70]. Finally, KLHDC10 has been implicated in the regulation of oxidative stress-induced ASK1 activation by suppressing PP5 [71]. It is thought that KLHDC10 possesses two functions, one as a SRR as described above and another as a signal regulator through its interaction with PP5 [71];

however, associations with cancer are not well documented.

Like KLHDC3 and KLHDC4, the research conducted on the other family members indicates that targeting this subfamily could be a potential therapeutic strategy for various KLHDC-overexpressing cancer types; however, potential inhibitors are not well documented. Conversely, targeted protein degradation using proteolysis targeting chimeras (PROTACs) is currently being explored as a therapeutic strategy for targeting cancer-related oncoproteins that cannot be treated with classical small molecule inhibitors. As such, KLHDC2 is emerging as a novel powerful degrader, and the hope is that through the development of potent small molecule ligands that recruit KLHDC2 to the target protein, it will be able to specifically degrade proteins involved in cancer [72].

Neurological diseases

The KLHDC subfamily has also been implicated in neurological functions and disorders [46,73–75] (Table 2). KLHDC1 and KLHDC3 may be involved in neural differentiation and have been associated with psychiatric disorders, particularly schizophrenia [46,74], suggesting a role in the development and function of the nervous system; however, this is not well defined. As discussed, KLHDC8A is linked to glioma [70], indicating a possible role in neurological malignancies. KLHDC10, known for its role in oxidative stress response [71], may be relevant to neurodegenerative diseases, but further research is required to establish its direct implications. However, for the remainder of the KLHDC subfamily – KLHDC1, KLHDC3, KLHDC4, KLHDC7A and 7B, KLHDC8B and KLHDC9 – direct associations with neurological disorders are not well documented, and comprehensive research is still required to fully understand their roles, if any.

Conclusion

The KLHDC subfamily has gained significant interest in recent years, especially KLHDC2 which has been highlighted as an excellent E3 ligase substrate adaptor suitable for use as a PROTAC drug. The new knowledge gathered from these recent structural and biochemical advances for KLHDC2, along with predictive software such as ALPHAFOLD and INTERPRO, has allowed for a greater understanding of the other KLHDC family members. This review consolidates all the known structure and functional roles of these family members. There has been some debate in the field over which proteins should be included in the KLHDC family. The

structural analysis conducted in this review in combination with the supporting literature suggested that there should only be 10 family members: KLHDC1, KLHDC2, KLHDC3, KLHDC4, KLHDC7A and 7B, KLHDC8A and 8B, KLHDC9 and KLHDC10. With the emergence of PROTACs in the drug discovery field, a greater understanding of the KLHDC family members may lead to the design of novel potent PROTAC reagents and allow diverse targeting of various pathology-related proteins.

Acknowledgements

This work was supported by the Australian Government Medical Research Future Fund (MRFF) – Emerging Priorities and Consumer Driven Research Initiative – 2020 Childhood Cancer Research Grant (MRFF2007435 to CRW, JKH and MFS). JKH was supported by an RMIT Vice Chancellors Fellowship. CRW was supported by a National Health and Medical Research Council Australia (NHMRC) Investigator Award (GNT2018098). CP was supported by an RMIT Research Stipend Scholarship. The work was supported, in part, by the Victorian State Government Operational Infrastructure Support Scheme to St Vincent's Institute. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Open access publishing facilitated by RMIT University, as part of the Wiley - RMIT University agreement via the Council of Australian University Librarians.

Author contributions

CP created the models and the figures and wrote the first draft. PAVB wrote the section on substrates. JQT conducted the protein domain analysis. JKH conceptualised this manuscript and conducted extensive editing. In conjunction with all authors, PAR, MFS and CRW edited and reviewed the manuscript.

References

- Adams J, Kelso R and Cooley L (2000) The Kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol* **10**, 17–24.
- Prag S and Adams JC (2003) Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. *BMC Bioinformatics* **4**, 42.
- Gupta VA and Beggs AH (2014) Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skelet Muscle* **4**, 11.
- Shi X, Xiang S, Cao J, Zhu H, Yang B, He Q and Ying M (2019) Kelch-like proteins: physiological functions and relationships with diseases. *Pharmacol Res* **148**, 104404.
- Dhanoa BS, Cogliati T, Satish AG, Bruford EA and Friedman JS (2013) Update on the Kelch-like (KLHL) gene family. *Hum Genomics* **7**, 13.
- Stogios PJ and Prive GG (2004) The BACK domain in BTB-kelch proteins. *Trends Biochem Sci* **29**, 634–637.
- Ohinata Y, Sutou S and Mitsui Y (2003) A novel testis-specific RAG2-like protein, peas: its expression in pachytene spermatocyte cytoplasm and meiotic chromatin. *FEBS Lett* **537**, 1–5.
- Teng F, Wang Y, Liu M, Tian S, Stjepanovic G and Su M-Y (2023) Cryo-EM structure of the KLHL22 E3 ligase bound to an oligomeric metabolic enzyme. *Structure* **31**, 1431–1440.e5.
- Enchev RI, Schulman BA and Peter M (2015) Protein neddylation: beyond cullin–RING ligases. *Nat Rev Mol Cell Biol* **16**, 30–44.
- Geyer R, Wee S, Anderson S, Yates J and Wolf DA (2003) BTB/POZ domain proteins are putative substrate adaptors for Cullin 3 ubiquitin ligases. *Mol Cell* **12**, 783–790.
- Ye G, Wang J, Yang W, Li J, Ye M and Jin X (2022) The roles of KLHL family members in human cancers. *Am J Cancer Res* **12**, 5105–5139.
- Cummings CM, Bentley CA, Perdue SA, Baas PW and Singer JD (2009) The Cul3/Klhdc5 E3 ligase regulates p60/katanin and is required for Normal mitosis in mammalian cells. *J Biol Chem* **284**, 11663–11675.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
- UniProt Consortium (2022) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* **51**, D523–D531.
- Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L and Bruford EA (2022) Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res* **51**, D1003–D1009.
- Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K *et al.* (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**, 40–45.
- Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L *et al.* (2022) InterPro in 2022. *Nucleic Acids Res* **51**, D418–D427.
- Stogios PJ and Privé GG (2004) The BACK domain in BTB-kelch proteins. *Trends Biochem Sci* **29**, 634–637.
- Prag S, Collett GD and Adams JC (2004) Molecular analysis of muskellin identifies a conserved discoidin-like

- domain that contributes to protein self-association. *Biochem J* **381**, 547–559.
- 20 Díaz E, Schimmöller F and Pfeffer SR (1997) A novel Rab9 effector required for endosome-to-TGN transport. *J Cell Biol* **138**, 283–290.
- 21 Ikononov OC, Sbrissa D, Mlak K, Deeb R, Fligger J, Soans A, Finley RL Jr and Shisheva A (2003) Active PIKfyve associates with and promotes the membrane attachment of the late endosome-to-trans-Golgi network transport factor Rab9 effector p40. *J Biol Chem* **278**, 50863–50871.
- 22 Scott DC, King MT, Baek K, Gee CT, Kalathur R, Li J, Purser N, Nourse A, Chai SC, Vaithiyalingam S *et al.* (2023) E3 ligase autoinhibition by C-degron mimicry maintains C-degron substrate fidelity. *Mol Cell* **83**, 770–786.e9.
- 23 Scott DC, Dharuman S, Griffith E, Chai SC, Ronnebaum J, King MT, Tangallapally R, Lee C, Gee CT, Yang L *et al.* (2024) Principles of paralog-specific targeted protein degradation engaging the C-degron E3 KLHDC2. *Nat Commun* **15**, 8829.
- 24 Hickey CM, Digianantonio KM, Zimmermann K, Harbin A, Quinn C, Patel A, Gareiss P, Chapman A, Tiberi B, Dobrodziej J *et al.* (2024) Co-opting the E3 ligase KLHDC2 for targeted protein degradation by small molecules. *Nat Struct Mol Biol* **31**, 311–322.
- 25 Kim Y, Seo P, Jeon E, You I, Hwang K, Kim N, Tse J, Bae J, Choi HS, Hinshaw SM *et al.* (2023) Targeted kinase degradation via the KLHDC2 ubiquitin E3 ligase. *Cell Chem Biol* **30**, 1414–1420.e5.
- 26 Rusnac DV, Lin HC, Canzani D, Tien KX, Hinds TR, Tsue AF, Bush MF, Yen HS and Zheng N (2018) Recognition of the diglycine C-end degron by CRL2 (KLHDC2) ubiquitin ligase. *Mol Cell* **72**, 813–822.e4.
- 27 Ehrlich KC, Baribault C and Ehrlich M (2020) Epigenetics of muscle- and brain-specific expression of KLHL family genes. *Int J Mol Sci* **21**, 8394.
- 28 Hudson AM and Cooley L (2008) Phylogenetic, structural and functional relationships between WD- and Kelch-repeat proteins. *Subcell Biochem* **48**, 6–19.
- 29 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The protein data Bank. *Nucleic Acids Res* **28**, 235–242.
- 30 Timms RT, Mena EL, Leng Y, Li MZ, Tchasovnikarova IA, Koren I and Elledge SJ (2023) Defining E3 ligase–substrate relationships through multiplex CRISPR screening. *Nat Cell Biol* **25**, 1535–1545.
- 31 Canning P, Cooper CDO, Krojer T, Murray JW, Pike ACW, Chaikuad A, Keates T, Thangaratnarajah C, Hojzan V, Marsden BD *et al.* (2013) Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases. *J Biol Chem* **288**, 7803–7814.
- 32 Kumari P, Aeschmann F, Gaidatzis D, Keusch JJ, Ghosh P, Neagu A, Pachulska-Wieczorek K, Bujnicki JM, Gut H, Grosshans H *et al.* (2018) Evolutionary plasticity of the NHL domain underlies distinct solutions to RNA recognition. *Nat Commun* **9**, 1549.
- 33 Zhang P, Gao K, Zhang L, Sun H, Zhao X, Liu Y, Lv Z, Shi Q, Chen Y, Jiao D *et al.* (2022) CRL2-KLHDC3 E3 ubiquitin ligase complex suppresses ferroptosis through promoting p14(ARF) degradation. *Cell Death Differ* **29**, 758–771.
- 34 Mahrour N, Redwine WB, Florens L, Swanson SK, Martin-Brown S, Bradford WD, Staehling-Hampton K, Washburn MP, Conaway RC and Conaway JW (2008) Characterization of Cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to Elongin BC-based ubiquitin ligases. *J Biol Chem* **283**, 8005–8013.
- 35 Pontén F, Gry M, Fagerberg L, Lundberg E, Asplund A, Berglund L, Oksvold P, Björling E, Hober S, Kampf C *et al.* (2009) A global view of protein expression in human cells, tissues, and organs. *Mol Syst Biol* **5**, 337.
- 36 Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM *et al.* (2017) A subcellular map of the human proteome. *Science* **356**, eaal3321.
- 37 Motomura S, Yumimoto K, Tomonaga T and Nakayama KI (2024) CRL2(KLHDC3) and CRL1 (Fbxw7) cooperatively mediate c-Myc degradation. *Oncogene* **43**, 1917–1929.
- 38 Chin KT, Xu HT, Ching YP and Jin DY (2007) Differential subcellular localization and activity of kelch repeat proteins KLHDC1 and KLHDC2. *Mol Cell Biochem* **296**, 109–119.
- 39 Salipante SJ, Mealiffe ME, Wechsler J, Krem MM, Liu Y, Namkoong S, Bhagat G, Kirchhoff T, Offit K, Lynch H *et al.* (2009) Mutations in a gene encoding a midbody kelch protein in familial and sporadic classical Hodgkin lymphoma lead to binucleated cells. *Proc Natl Acad Sci U S A* **106**, 14920–14925.
- 40 Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419.
- 41 Timms AE and Horwitz MS (2010) KLHDC8B in Hodgkin lymphoma and possibly twinning. *Commun Integr Biol* **3**, 154–158.
- 42 Yang Q, Zhao J, Chen D and Wang Y (2021) E3 ubiquitin ligases: styles, structures and functions. *Mol Biomed* **2**, 23.
- 43 Lin HC, Yeh CW, Chen YF, Lee TT, Hsieh PY, Rusnac DV, Lin SY, Elledge SJ, Zheng N and Yen HS (2018) C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol Cell* **70**, 602–613.e3.
- 44 Koren I, Timms RT, Kula T, Xu Q, Li MZ and Elledge SJ (2018) The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* **173**, 1622–1635.e14.

- 45 Jeong Y, Oh A-R, Jung YH, Gi H, Kim YU and Kim K (2023) Targeting E3 ubiquitin ligases and their adaptors as a therapeutic strategy for metabolic diseases. *Exp Mol Med* **55**, 2097–2104.
- 46 Niculescu AB, Levey DF, Phalen PL, Le-Niculescu H, Dainton HD, Jain N, Belanger E, James A, George S, Weber H *et al.* (2015) Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. *Mol Psychiatry* **20**, 1266–1285.
- 47 Bennett EJ, Rush J, Gygi SP and Harper JW (2010) Dynamics of Cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics. *Cell* **143**, 951–965.
- 48 Kamura T, Maenaka K, Kotoshiba S, Matsumoto M, Kohda D, Conaway RC, Conaway JW and Nakayama KI (2004) VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev* **18**, 3055–3065.
- 49 Okumura F, Matsuzaki M, Nakatsukasa K and Kamura T (2012) The role of Elongin BC-containing ubiquitin ligases. *Front Oncol* **2**, 10.
- 50 Lucas X and Ciulli A (2017) Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. *Curr Opin Struct Biol* **44**, 101–110.
- 51 Varshavsky A (1991) Naming a targeting signal. *Cell* **64**, 13–15.
- 52 Zhang P, Gao K, Zhang L, Sun H, Zhao X, Liu Y, Lv Z, Shi Q, Chen Y, Jiao D *et al.* (2021) Crl2-klhdc3 E3 ubiquitin ligase complex suppresses ferroptosis through promoting p14ARF degradation. *Cell Death Differ* **29**, 758–771.
- 53 Liu Y, Luo Y, Yan S, Lian Y-F, Wu S, Xu M, Feng L, Zhang X, Li R, Zhang X *et al.* (2022) CRL2KLHDC3 mediates p14ARF N-terminal ubiquitylation degradation to promote non-small cell lung carcinoma progression. *Oncogene* **41**, 3104–3117.
- 54 Yeh CW, Huang WC, Hsu PH, Yeh KH, Wang LC, Hsu PW, Lin HC, Chen YN, Chen SC, Yeang CH *et al.* (2021) The C-degron pathway eliminates mislocalized proteins and products of deubiquitinating enzymes. *EMBO J* **40**, e105846.
- 55 Ambrogelly A, Palioura S and Söll D (2007) Natural expansion of the genetic code. *Nat Chem Biol* **3**, 29–35.
- 56 Driscoll DM and Copeland PR (2003) Mechanism and regulation of selenoprotein synthesis. *Annu Rev Nutr* **23**, 17–40.
- 57 Patil PR, Burroughs AM, Misra M, Cerullo F, Costas-Insua C, Hung HC, Dikic I, Aravind L and Joazeiro CAP (2023) Mechanism and evolutionary origins of alanine-tail C-degron recognition by E3 ligases Pirh2 and CRL2-KLHDC10. *Cell Rep* **42**, 113100.
- 58 Cardote TAF, Gadd MS and Ciulli A (2017) Crystal structure of the Cul2-Rbx1-EloBC-VHL ubiquitin ligase complex. *Structure* **25**, 901–911.e3.
- 59 Kim YK, Kwak MJ, Ku B, Suh HY, Joo K, Lee J, Jung JU and Oh BH (2013) Structural basis of intersubunit recognition in elongin BC-cullin 5-SOCS box ubiquitin-protein ligase complexes. *Acta Crystallogr D Biol Crystallogr* **69**, 1587–1597.
- 60 Okumura F, Fujiki Y, Oki N, Osaki K, Nishikimi A, Fukui Y, Nakatsukasa K and Kamura T (2020) Cul5-type ubiquitin ligase KLHDC1 contributes to the elimination of truncated SELENOS produced by failed UGA/sec decoding. *iScience* **23**, 100970.
- 61 Lian YF, Yuan J, Cui Q, Feng QS, Xu M, Bei JX, Zeng YX and Feng L (2016) Upregulation of KLHDC4 predicts a poor prognosis in human nasopharyngeal carcinoma. *PLoS One* **11**, e0152820.
- 62 Lin HC, Ho SC, Chen YY, Khoo KH, Hsu PH and Yen HC (2015) SELENOPROTEINS. CRL2 aids elimination of truncated selenoproteins produced by failed UGA/sec decoding. *Science* **349**, 91–95.
- 63 Liu Y, Luo Y, Yan S, Lian YF, Wu S, Xu M, Feng L, Zhang X, Li R, Zhang X *et al.* (2022) CRL2 (KLHDC3) mediates p14ARF N-terminal ubiquitylation degradation to promote non-small cell lung carcinoma progression. *Oncogene* **41**, 3104–3117.
- 64 Slusarz MJ (2024) How kelch domain-containing protein 3 distinguishes between the C-end degron of herpesviral protein UL49.5 and its mutants – insights from molecular dynamics. *Bioorg Med Chem* **109**, 117795.
- 65 Wahalska M, Riepe C, Slusarz MJ, Graul M, Borowski LS, Qiao W, Foltynska M, Carette JE, Bienkowska-Szewczyk K, Szczesny RJ *et al.* (2023) The herpesvirus UL49.5 protein hijacks a cellular C-degron pathway to drive TAP transporter degradation. *bioRxiv*. doi: [10.1101/2023.09.27.559663](https://doi.org/10.1101/2023.09.27.559663)
- 66 Lee D, Gimple RC, Wu X, Prager BC, Qiu Z, Wu Q, Daggubati V, Mariappan A, Gopalakrishnan J, Sarkisian MR *et al.* (2023) Superenhancer activation of KLHDC8A drives glioma ciliation and hedgehog signaling. *J Clin Invest* **133**, e163592.
- 67 Diederichs S, Baumer N, Ji P, Metzelder SK, Idos GE, Cauvet T, Wang W, Moller M, Pierschalski S, Gromoll J *et al.* (2004) Identification of interaction partners and substrates of the cyclin A1-CDK2 complex. *J Biol Chem* **279**, 33727–33741.
- 68 Fan Q, Huang LZ, Zhu XJ, Zhang KK, Ye HF, Luo Y, Sun XH, Zhou P and Lu Y (2014) Identification of proteins that interact with alpha A-crystallin using a human proteome microarray. *Mol Vis* **20**, 117–124.
- 69 Thrun A, Garzia A, Kigoshi-Tansho Y, Patil PR, Umbaugh CS, Dallinger T, Liu J, Kreger S, Patrizi A, Cox GA *et al.* (2021) Convergence of mammalian RQC and C-end rule proteolytic pathways via alanine tailing. *Mol Cell* **81**, 2112–2122.e7.
- 70 Zhu X, Chen T, Yang H and Lv K (2020) Lactate induced up-regulation of KLHDC8A (Kelch domain-

- containing 8A) contributes to the proliferation, migration and apoptosis of human glioma cells. *J Cell Mol Med* **24**, 11691–11702.
- 71 Sekine Y, Hatanaka R, Watanabe T, Sono N, Iemura S-i, Natsume T, Kuranaga E, Miura M, Takeda K and Ichijo H (2012) The Kelch repeat protein KLHDC10 regulates oxidative stress-induced ASK1 activation by suppressing PP5. *Mol Cell* **48**, 692–704.
- 72 Röth S, Kocaturk NM, Sathyamurthi PS, Carton B, Watt M, Macartney TJ, Chan K-H, Isidro-Llobet A, Konopacka A, Queisser MA *et al.* (2023) Identification of KLHDC2 as an efficient proximity-induced degrader of K-RAS, STK33, β -catenin, and FoxP3. *Cell Chem Biol* **30**, 1261–1276.e7.
- 73 Mohammadnejad A, Soerensen M, Baumbach J, Mengel-From J, Li W, Lund J, Li S, Christiansen L, Christensen K, Hjelmberg JVB *et al.* (2021) Novel DNA methylation marker discovery by assumption-free genome-wide association analysis of cognitive function in twins. *Aging Cell* **20**, e13293.
- 74 Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ, Luo S, Zhang L, van Velkinburgh JC, Farmer AD *et al.* (2008) Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS One* **3**, e3625.
- 75 Chai K, Liang J, Zhang X, Cao P, Chen S, Gu H, Ye W, Liu R, Hu W, Peng C *et al.* (2021) Application of machine learning and weighted gene co-expression network algorithm to explore the hub genes in the aging brain. *Front Aging Neurosci* **13**, 707165.
- 76 Zhou HJ, Wong CM, Chen JH, Qiang BQ, Yuan JG and Jin DY (2001) Inhibition of LZIP-mediated transcription through direct interaction with a novel host cell factor-like protein. *J Biol Chem* **276**, 28933–28938.
- 77 Martin-Pardillos A and Cajal SRY (2019) Characterization of Kelch domain-containing protein 7B in breast tumours and breast cancer cell lines. *Oncol Lett* **18**, 2853–2860.
- 78 Diederichs S, Bäumer N, Ji P, Metzelder SK, Idos GE, Cauvet T, Wang W, Möller M, Pierschalski S, Gromoll J *et al.* (2004) Identification of interaction partners and substrates of the cyclin A1-CDK2 complex. *J Biol Chem* **279**, 33727–33741.
- 79 Stark C (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535–D539.
- 80 Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N *et al.* (2013) The MIntAct project—Intact as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358–D363.
- 81 Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al.* (2019) String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607–D613.
- 82 Hüttenhain R, Xu J, Burton LA, Gordon DE, Hultquist JF, Johnson JR, Satkamp L, Hiatt J, Rhee DY, Baek K *et al.* (2019) ARIH2 is a Vif-dependent regulator of CUL5-mediated APOBEC3G degradation in HIV infection. *Cell Host Microbe* **26**, 86–99.e7.
- 83 Yang J, Guo F, Chin HS, Chen GB, Ang CH, Lin Q, Hong W and Fu NY (2023) Sequential genome-wide CRISPR-Cas9 screens identify genes regulating cell-surface expression of tetraspanins. *Cell Rep* **42**, 112065.
- 84 Blondelle J, Biju A and Lange S (2020) The role of Cullin-RING ligases in striated muscle development, function, and disease. *Int J Mol Sci* **21**, 7936.
- 85 Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, Gygi MP, Thornock A, Zarraga G, Tam S *et al.* (2021) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022–3040.e28.
- 86 Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H *et al.* (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509.
- 87 Zeissig S, Petersen BS, Tomczak M, Melum E, Huc-Claustre E, Dougan SK, Laerdahl JK, Stade B, Forster M, Schreiber S *et al.* (2015) Early-onset Crohn's disease and autoimmunity associated with a variant in CTLA-4. *Gut* **64**, 1889–1897.
- 88 Seeley AH, Durham MA, Micale MA, Wesolowski J, Foerster BR and Martin DM (2014) Macrocerebellum, epilepsy, intellectual disability, and gut malrotation in a child with a 16q24.1-q24.2 contiguous gene deletion. *Am J Med Genet A* **164a**, 2062–2068.
- 89 Lin X, Wang J, Yun L, Jiang S, Li L, Chen X, Li Z, Lu Q, Zhang Y and Ma X (2016) Association between LEKR1-CCNL1 and IGSF21-KLHDC7A gene polymorphisms and diabetic retinopathy of type 2 diabetes mellitus in the Chinese Han population. *J Gene Med* **18**, 282–287.
- 90 Kelly DM and Rothwell PM (2022) Disentangling the relationship between chronic kidney disease and cognitive disorders. *Front Neurol* **13**, 830064.
- 91 Kalra G, Milon B, Casella AM, Herb BR, Humphries E, Song Y, Rose KP, Hertzano R and Ament SA (2020) Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty. *PLoS Genet* **16**, e1009025.
- 92 Luo S, Wang L, Li S, Wang H, Huang S, Zhang Z, Wang R, Guan H and Huang Y (2021) Identification of key molecules and lncRNA-miRNA-mRNA ceRNA network in preeclampsia. *Int J Gen Med* **14**, 7579–7590.

- 93 Salokas K, Liu X, Öhman T, Chowdhury I, Gawryski L, Keskitalo S and Varjosalo M (2022) Physical and functional interactome atlas of human receptor tyrosine kinases. *EMBO Rep* **23**, e54041.
- 94 Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloreaux B *et al.* (2020) A reference map of the human binary protein interactome. *Nature* **580**, 402–408.
- 95 Capalbo L, Bassi ZI, Geymonat M, Todesca S, Copoiu L, Enright AJ, Callaini G, Riparbelli MG, Yu L, Choudhary JS *et al.* (2019) The midbody interactome reveals unexpected roles for PP1 phosphatases in cytokinesis. *Nat Commun* **10**, 4513.
- 96 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, and Higgins DG. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**. <https://doi.org/10.1038/msb.2011.75>