

SCIENTIFIC REPORTS



OPEN

Crystal structure of the N-terminal domain of human CDC73 and its implications for the hyperparathyroidism-jaw tumor (HPT-JT) syndrome

Wei Sun¹, Xiao-Lin Kuang^{1,2}, Yan-Ping Liu¹, Li-Fei Tian¹, Xiao-Xue Yan¹ & Wenqing Xu^{1,3}

CDC73/Parafibromin is a critical component of the Paf1 complex (PAF1C), which is involved in transcriptional elongation and histone modifications. Mutations of the human *CDC73/HRPT2* gene are associated with hyperparathyroidism-jaw tumor (HPT-JT) syndrome, an autosomal dominant disorder. CDC73/parafibromin was initially recognized as a tumor suppressor by inhibiting cell proliferation via repression of *cyclin D1* and *c-myc* genes. In recent years, it has also shown oncogenic features by activating the canonical Wnt/ β -catenin signal pathway. Here, through limited proteolysis analysis, we demonstrate that the evolutionarily conserved human CDC73 N-terminal 111 residues form a globularly folded domain (hCDC73-NTD). We have determined a crystal structure of hCDC73-NTD at 1.02 Å resolution, which reveals a novel protein fold. CDC73-NTD contains an extended hydrophobic groove on its surface that may be important for its function. Most pathogenic *CDC73* missense mutations associated with the HPT-JT syndrome are located in the region encoding CDC73-NTD. Our crystal and biochemical data indicate that most CDC73 missense mutations disrupt the folding of the hydrophobic core of hCDC73-NTD, while others such as the K34Q mutant reduce its thermostability. Overall, our results provide a solid structural basis for understanding the structure and function of CDC73 and its association with the HPT-JT syndrome and other diseases.

Human CDC73 (hCDC73), also known as parafibromin, is a human nuclear protein encoded by the cell division cycle 73 (*CDC73*) gene which is located on chromosome 1q31.2^{1,2}. Mutations of human *CDC73* are associated with hyperparathyroidism-jaw tumor (HPT-JT) syndrome, which is an autosomal dominant disorder characterized by parathyroid tumors, fibro-osseous jaw tumors, cystic kidney lesions and uterine tumors^{1,3-5}. *CDC73* mutations are also associated with parathyroid carcinomas and renal tumors⁶⁻¹⁰ (Supplementary Table S1). CDC73 is a core subunit of the polymerase-associated factor 1 complex (PAF1C) that binds to RNA polymerase II¹¹⁻¹³. PAF1C participates in recruitment of histone modification factors and promotion of transcriptional elongation¹⁴⁻¹⁶.

Human CDC73 was initially proposed as a tumor suppressor which can repress expression of cyclin D1^{17,18} and *c-myc* proto-oncogene¹⁹. For example, CDC73 can recruit the histone methyltransferase (HMTase) SUV39H1 to the *cyclin D1* gene to inhibit its expression by inducing methylation of H3 K9¹⁸. Loss-of-function mutations of hCDC73 can thus lead to overproduction of cell cycle control proteins and promote cell proliferations. However, in some other occasions, CDC73 possesses oncogenic features. In particular, it can serve as a transcriptional coactivator of the Wnt/ β -catenin signaling pathway by directly interacting with β -catenin in the nucleus²⁰. This interaction can be stabilized by SHP2-mediated dephosphorylation of CDC73 residues Y290, Y293 and Y315 in the nucleus²¹. Recent studies also show that CDC73 can act as a scaffold protein to coordinate Wnt/Hedgehog/

¹National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, P. R. China. ²College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China. ³Department of Biological Structure, University of Washington, Seattle, Washington, 98195, USA. Correspondence and requests for materials should be addressed to X.-X.Y. (email: snow@ibp.ac.cn) or W.X. (email: wuxu@u.washington.edu)

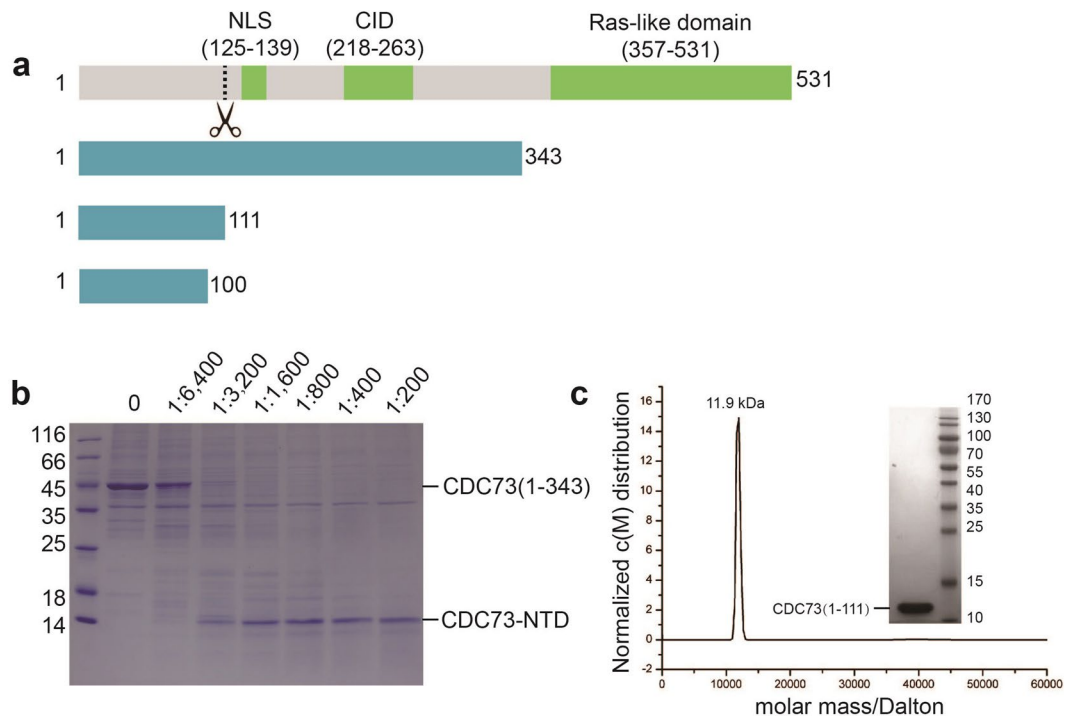


Figure 1. Biochemical characterization of the N-terminal domain of CDC73 (CDC73-NTD). **(a)** Schematic representation of human CDC73 and CDC73 fragments used in this study. Full-length human CDC73 consists of 531 residues containing a nuclear localization signal (NLS) and a β -catenin interaction domain (CID). The dash line indicates the boundary of a protease-resistant fragment. **(b)** One trypsin-resistant fragment can be obtained through limited proteolysis of CDC73(1–343), which contains the first 111 residues of hCDC73 (CDC73-NTD). The marked ratios are the molar ratios of trypsin over hCDC73(1–343) used in limited proteolysis. The SDS-PAGE gel was stained with Coomassie Brilliant Blue. **(c)** Analytical ultracentrifugation (AUC) analysis demonstrates that CDC73(1–111) has a molecular weight of 11.9 kDa, which is close to the nominal MW of monomeric hCDC73(1–111) (12.7 kDa). The inset shows the SDS-PAGE of our hCDC73(1–111) sample used for crystallographic and AUC analysis, stained with Coomassie Brilliant Blue.

Notch pathway and plays important roles during embryogenesis and organogenesis^{22,23}. Both tumor-suppressing and oncogenic effects of CDC73 rely on the N-terminal portion of hCDC73^{18,20}.

The C-terminus of hCDC73 (residues 357–531) bears a conserved Ras-like domain homologous to yeast CDC73, which promotes the interaction between the PAF1 complex and chromatin²⁴. Crystal structures of the C-terminal domain of yeast CDC73 have been reported^{24,25}, but there is no 3D structural information of the hCDC73 N-terminal region to date. In addition to the β -catenin interaction domain (CID, residues 218–263), the hCDC73 N-terminal region also contains a functional nuclear localization signal (residues 125–139)²⁶, as well as three proposed nucleolar localization signals^{27,28}. Most of the documented CDC73 mutations are nonsense truncation mutations which result in physiologically inactive proteins³ (Supplementary Table S1). Although a number of missense variants of hCDC73 have been found in HPT-JT and other diseases, only a few of them, including K34Q, I60N, L64P and R91P, have been shown to be pathogenic (Supplementary Table S1). Strikingly, all of them are located in the N-terminal region of CDC73, underlying the functional importance of this region^{17,29,30}. For example, I60N shows decreased protein stability and a loss of its ability to down-regulate c-myc expression, and is associated with ossifying fibroma³⁰. The L64P mutation found in HPT-JT patients was found to be able to retain nuclear expression and form PAF1C, but showed reduced interaction with the Set1 histone methyltransferase complex and decreased its HMTase activity^{17,31}.

To understand how the CDC73 N-terminal domain contributes to CDC73 functions and how CDC73 pathogenic mutations may lead to HPT-JT and other diseases, we have characterized the structural features of the CDC73 N-terminal regions and revealed a previously unidentified N-terminal globular domain. Furthermore, we have determined a crystal structure of CDC73-NTD at ultrahigh 1.02 Å resolutions and characterized CDC73 pathogenic missense mutants using biochemical assays. Our study provides novel insights into CDC73 function and molecular mechanisms of HPT-JT mutants.

Results

N-terminal half of hCDC73 contains a conserved and protease-resistant domain. To understand how parafibromin is involved in β -catenin-mediated Wnt signal activation, we tested the overexpression and purification of numerous hCDC73 fragments. We found that human CDC73(1–343), which contains β -catenin interaction domain of CDC73 (CID) and the nuclear localization signal (NLS, Fig. 1a), can be purified in decent quantities. However, during purification, hCDC73(1–343) was susceptible to degradation and robustly led to a

Data collection	CDC73(1–100) - native	CDC73(1–100) - derivatives	CDC73(1–111) - native
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
Cell dimensions			
A, b, c (Å)	28.59, 53.50, 65.73	28.55, 53.52, 66.01	38.09, 43.68, 67.63
α, β, γ (°)	90.00, 95.66, 90.00	90.00, 95.73, 90.00	90.00, 90.00, 90.00
Wavelength (Å)	0.9782	0.9782	0.9782
Resolution (Å)	50–1.40 (1.42–1.40)	50–1.40 (1.42–1.40)	50–1.02 (1.05–1.02)
R_{merge} (%)	4.3 (18.1)	11.5 (57.4)	12.2 (35.9)
$I/\sigma I$	41.92 (10.0)	72.38 (6.5)	38.61 (4.9)
CC1/2	0.996 (0.986)	0.989 (0.964)	0.964 (0.891)
Completeness (%)	98.9 (99.1)	99.0 (99.5)	99.0 (79.8)
Redundancy	6.6 (6.3)	25.3 (18.8)	6.3 (5.8)
Wilson B factors (Å ²)	11.0	15.9	9.2
Refinement			
Resolution (Å)	19.5–1.40		19.4–1.02
No. Reflections	38451		57501
$R_{\text{work}}/R_{\text{free}}$ (%)	13.52/17.42		11.94/13.94
No. atoms	1661		1740
No. of water	348		284
Average B factors (Å ²)			
Protein	15.37		13.18
H ₂ O	33.25		32.27
R.m.s deviations			
Bond lengths (Å)	0.012		0.011
Bond angles (°)	1.361		1.516
Ramachandran plot (%)			
Most Favorable	100.0		100.0
Allowed	0.0		0.0
Outliers	0.0		0.0

Table 1. Summary of crystallographic statistics.

stable ~12kD fragment. Thus we performed limited proteolysis of hCDC73(1–343) with trypsin and consistently observed one stable fragment (Fig. 1b). LC-MS/MS analysis revealed that this protease-resistant fragment contains at least the first N-terminal 100 residues and likely the first 111 residues of hCDC73. We thus named this protease-resistant region as hCDC73-NTD. Based on sequence conservation and the size of trypsin-resistant fragment, we generated new plasmids and purified both hCDC73(1–100) and hCDC73(1–111). SEC and analytical ultracentrifugation (AUC) analysis clearly show that both fragments are monomeric in solution (Fig. 1c).

Crystal structures of human CDC73-NTD. Since hCDC73-NTD is evolutionarily conserved and harbors most of the documented pathogenic CDC73 missense mutations, we determined crystal structures of hCDC73-NTD. First, we determined the crystal structure of CDC73(1–100) using the SAD method (with a soaked platinum derivative) at 1.4 Å resolution, and then the crystal structure of CDC73(1–111) at 1.02 Å resolution (Table 1; Fig. 2a,b). There are one and two molecules in each asymmetric unit for CDC73(1–111) and CDC73(1–100) crystals, respectively. The model corresponds well to the electron density map (Supplementary Fig. S1). The overall structure of CDC73(1–111) is almost identical with CDC73(1–100), with a r.m.s.d. of α positions being 0.894 Å. Both CDC73(1–100) and CDC73(1–111) fold into a globular domain. The extra 11 residues of CDC73(1–111) pack tightly on surface of the CDC73(1–100) core structure (Fig. 2c). Analysis using the DALI server shows that CDC73(1–100) has a novel fold (Supplementary Table S2). This domain structure is composed of four α helices and one anti-parallel β -sheet consisting of three β -strands (Fig. 2b).

Analysis of pathogenic mutations in human CDC73-NTD. Most pathogenic mutations of human CDC73 are nonsense or frame-shift mutations (Supplementary Table S1). However, at least four missense mutations have been characterized (Fig. 3a). They are K34Q, I60N, L64P, R91P^{6,29–31}. Mutations L64P and R91P, occurring at residues located in the helices H2 and H4, respectively, are helix-disruptive and may thus destroy the CDC73-NTD folding (Fig. 3b). Ile60 forms part of the hydrophobic core of CDC73-NTD. The I60N mutation would cause steric clashes in the CDC73-NTD hydrophobic core and thus disrupt its structure (Fig. 3b). To examine the biochemical properties of these mutants, we tested overexpression and purification of several human CDC73(1–111) missense mutants in *E. coli*, including K34Q, I60N, L64P and R91P. Almost all of them, including I60N, L64P, and R91P, are essentially insoluble in *E. coli*, consistent with our structure-based prediction that these missense mutations disrupt the structure of CDC73-NTD.

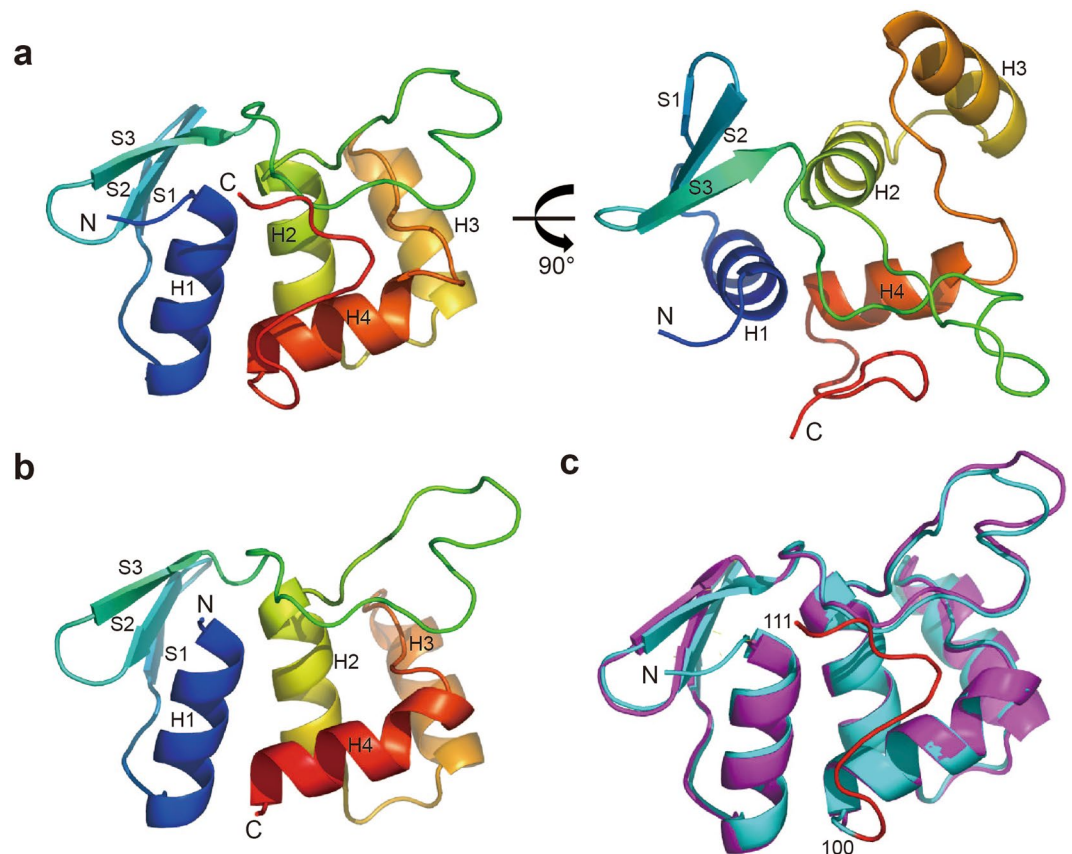


Figure 2. Overall structure of CDC73-NTD. **(a)** Two orthogonal views of the crystal structure of CDC73(1–111). **(b)** Overall structure of CDC73(1–100). **(c)** Superposition of CDC73(1–111) and CDC73(1–100) crystal structures. CDC73(1–111) and CDC73(1–100) are shown in cyan and magenta, respectively. The red loop represents the extra 11 C-terminal residues of CDC73(1–111).

Characterization of the K34Q mutant and the role of the hCDC73(1–111) tail. Among CDC73 missense mutants we tested, K34Q was the only one which we could produce its soluble form in decent amount for further characterization. To test whether the K34Q mutation changed the secondary structure of CDC73-NTD, the CD spectra of wild-type (WT) and K34Q mutant forms of CDC73(1–111) were measured. Not surprisingly, the overall secondary structure compositions are essentially identical (Fig. 4a). In contrast, the del(5–10) mutant of hCDC73(1–111), with deletion of residues 5–10 in helix 1 (H1), could be purified with a very low yield and demonstrated a CD spectrum of a largely unfolded protein (Fig. 4a,d). To examine whether the K34Q mutation changes its thermostability, we measured the melting temperature of WT and K34Q using a ThermoFluor assay. Apparently, the K34Q mutant ($T_m = 46^\circ\text{C}$) is thermodynamically less stable than the WT human CDC73(1–111) ($T_m = 56^\circ\text{C}$) (Fig. 4b). In our crystal structure, the Lys34 sidechain can adopt two conformations, both forming salt bridges with adjacent Asp23 and Asp58. The K34Q mutant may thus destabilize the CDC73-NTD structure by disrupting these salt bridges (Fig. 4c).

This C-terminal 11 residues of CDC73(1–111) interact extensively with the CDC73(1–100) core, which may stabilize the overall structure (Fig. 4e). Indeed, our ThermoFluor assay showed that hCDC73(1–100) and hCDC73(1–111) have T_m values of 46°C and 56°C , respectively (Fig. 4b).

A hydrophobic groove on the CDC73-NTD surface and the NoLS. Human CDC73-NTD has a hydrophobic groove on the surface. It is formed by CDC73-NTD β -strand S1 and helices H1, H2 and H4 (Fig. 5a and Supplementary Fig. S2). Hydrophobic residues involved in this groove, L8, I18, V20, V25, F27, I60, L61, L63, L64, L67, L95 and L98, are mostly conserved among CDC73 proteins from yeast to human, suggesting a conserved structural feature that may have functional implications (Fig. 3c). Interestingly, pathogenic mutations I60N and L64P are located inside of this groove, while the other two pathogenic mutants K34Q and R91P are located on two ends of this groove. Although the function of this groove remains unclear, it is likely that the deformation of this groove is associated with CDC73 malfunction.

A separate issue regarding CDC73-NTD is a proposed nucleolar localization signal (NoLS), which is among the three proposed NoLSs that mediate the nucleolar localization of hCDC73²⁷. In our crystal structure, this NoLS (residues 76–92) spans helices H3, H4 and the H3–H4 loop, and forms an exposed surface of CDC73-NTD that is adjacent but distinct from the hydrophobic groove (Supplementary Figs S2 and S3). This proposed NoLS appears conserved among species and may be important for CDC73 functions, since three HPT-JT associated

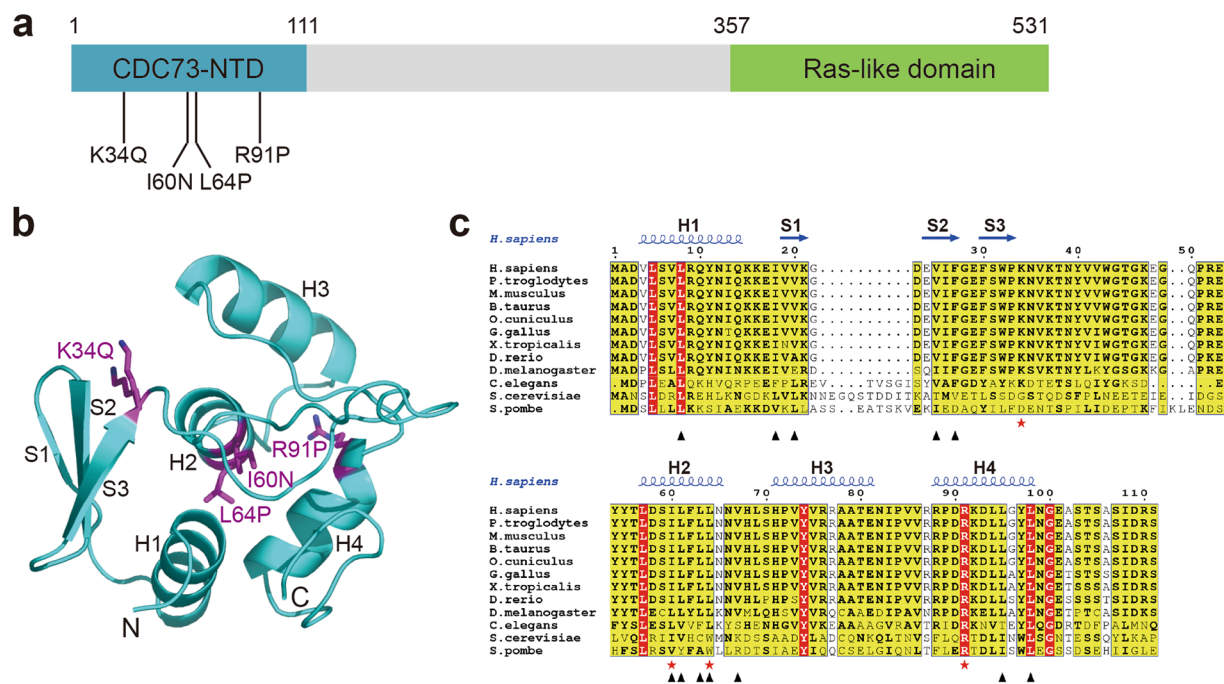


Figure 3. CDC73-NTD harbors all reported pathogenic missense mutations. **(a)** Domain structure of CDC73 shows that four characterized pathogenic mutations are clustered in CDC73-NTD. **(b)** Positions of those four pathogenic missense mutations are labeled in the structure of CDC73-NTD. **(c)** Sequence alignment of CDC73(1–111) from ten species shows that CDC73-NTD is evolutionarily conserved from yeast to human. The ★ indicate positions of CDC73 pathogenic missense mutations. The ▲ mark residues forming the hydrophobic groove of CDC73-NTD.

mutants, R77P, del(NIP) and del(VV) have been identified and characterized within this NoLS²⁸. There is no general mechanism how NoLS promotes nucleolar localization^{32,33}, and it remains unclear how this proposed CDC73 NoLS works.

Discussion

HPT-JT, parathyroid carcinomas and renal tumors are clinically severe cases and threat many human lives. Discovery of pathogenic CDC73 mutants, especially pathogenic missense mutations observed in CDC73-NTD, strongly demonstrate the biological and medical significance of the CDC73 N-terminal domain. It remains unclear how CDC73 exerts its biomedical effects. Recently, a mouse model for HPT-JT with *Cdc73* depletion, showing parathyroid tumors and uterine neoplasms, has been established and will be helpful for future research on these tumors³⁴.

Here, we demonstrate that CDC73-NTD can form a well-folded globular domain, and is conserved from human to yeast. Most of the reported CDC73 missense pathogenic mutations are clustered in this region. Our limited proteolysis data suggests that the region between the well folded CDC73-NTD (residues 1–111) and the Ras-like C-terminal domain (residues 357–531) is largely unfolded, but is involved in interactions with a number of CDC73 binding partners including β -catenin^{18,20,23,35}. Previous works showed that most *HRPT2* mutations (Supplementary Table 1) would result in truncation of the CDC73 protein³. Lack of the C-terminal Ras-like domain may cause deficiency in histone ubiquitination and methylation^{36,37}. Some truncations lacking the region interacting with other Paf1 complex subunits would impair the formation of this complex¹⁸. CDC73 is predominantly a nuclear protein, and is rich in nucleoli^{26,27,38}. Although it remains unclear how the proposed NoLS in CDC73-NTD works, the association of CDC73 mutations within this NoLS with HPT-JT suggests biological significance of this NoLS.

The pathogenic CDC73 K34Q mutant was identified from patient with renal tumors. The K34Q mutant is localized to the nucleus and can effectively participate in PAFIC formation⁶. Unlike other mutants, the CDC73(1–111) K34Q mutant is as soluble as the WT protein. Thermofluor assay shows that the melting temperature of K34Q is 10 °C lower than that of WT. This suggests that the K34Q mutant may lose the ability to suppress expression of *cyclin D1* due to lower thermal stability. Alternatively, since Lys34 is located at one end of the hydrophobic groove and may be involved in interaction with other CDC73 binding partner(s), the K34Q mutation may lead to functional loss by abolishing these CDC73-partner interaction(s).

In summary, we reveal a well-folded and conserved N-terminal domain of CDC73. Through crystallographic analysis, we provide solid structural basis for understanding CDC73 pathogenic mutants within this domain. We suggest that most, if not all, CDC73 pathogenic mutations lead to loss of CDC73 functions by disrupting the folding or thermostability of CDC73-NTD. The observation of a hydrophobic groove on CDC73 surface will also help with understanding the function of this conserved domain.

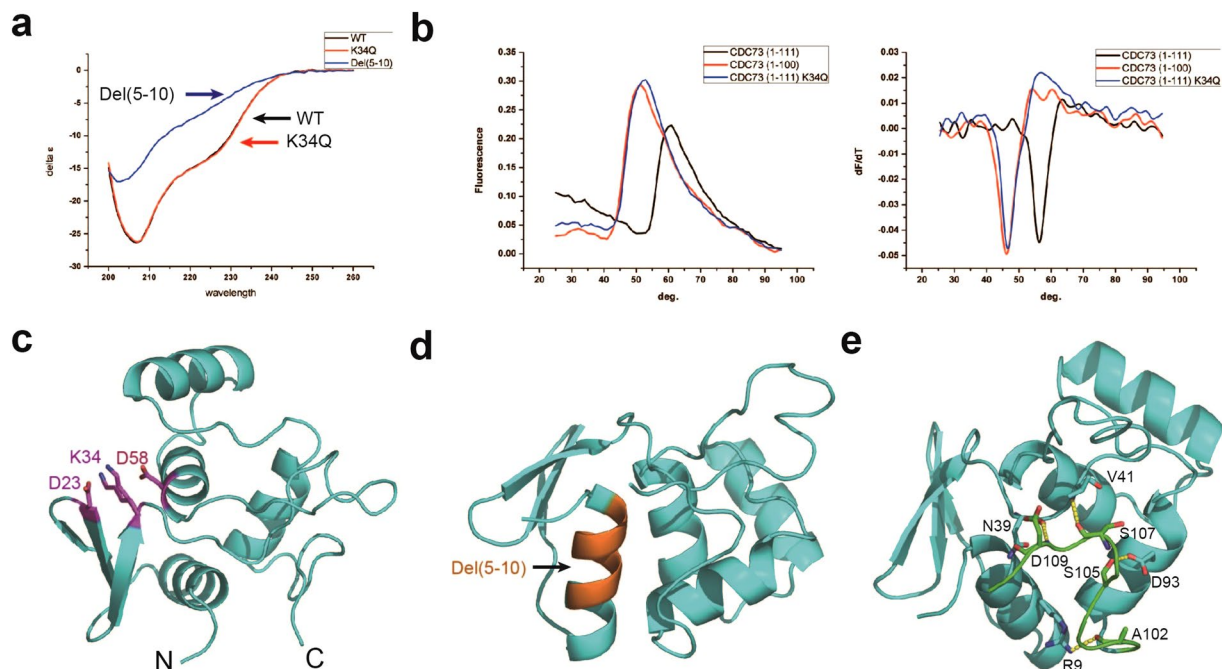


Figure 4. Characterization of the CDC73 K34Q mutant and analysis of the differences between CDC73(1–111) and CDC73(1–100). **(a)** CD spectra of the CDC73(1–111) K34Q mutant, with wild-type (WT) and the del(5–10) mutant as positive and negative controls, respectively. **(b)** Thermofluor assays reveal a T_m difference of 10 °C for both WT vs K34Q CDC73(1–111) and CDC73(1–111) vs CDC73(1–100). The left panel shows the fluorescence change corresponding to temperature. The right panel is the calculation of melting temperatures. **(c)** In our crystal structure, the K34 sidechain (with dual conformations observed) forms salt bridges with D23 and D58. **(d)** Positions of residues 5–10 in the CDC73 H1 helix, shown in orange in the structure. Internal deletion of these 6 residues disrupts the overall fold of hCDC73-NTD. **(e)** The CDC73-NTD C-terminal tail (residues 101–111) interacts with the rest of CDC73-NTD via hydrogen bonds and van der Waals forces. CDC73(1–100) and the C-terminal tail are shown in cyan and green, respectively. The hydrogen bonds between them are labeled with yellow dash.

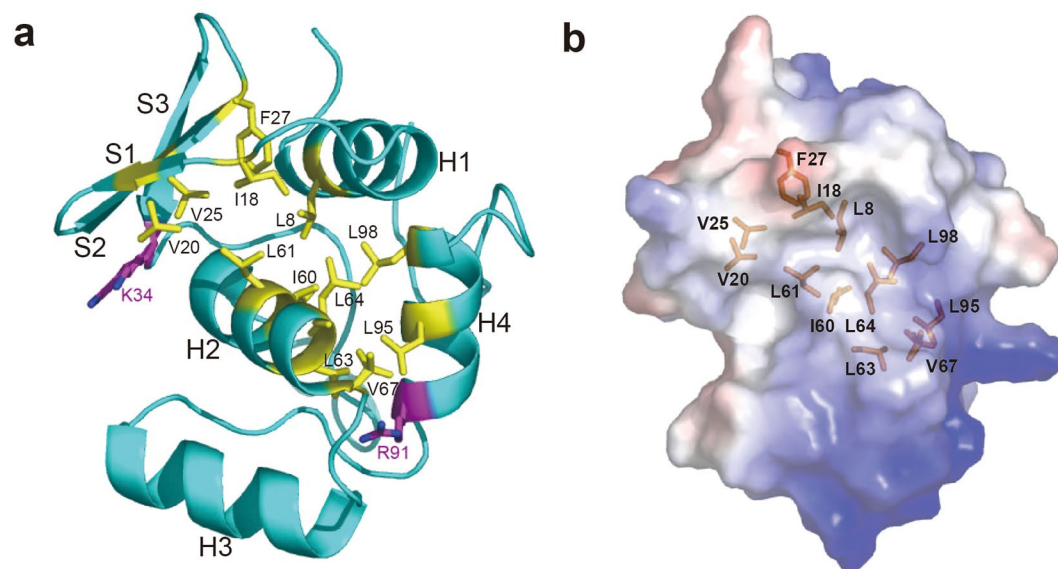


Figure 5. A hydrophobic groove on the surface of CDC73-NTD. **(a)** Residues composing the hydrophobic groove are shown as yellow sticks. Two residues associated with CDC73 pathogenic mutations, K34 and R91, are located at two ends of the hydrophobic groove. **(b)** Electrostatic potential surface map was shown with the same orientation as the structure in panel a. Residues forming the hydrophobic groove are shown in sticks.

Materials and Methods

Cloning, purification and crystallization. The full-length human CDC73 gene was cloned from a cDNA library, and constructed into pET28a vector. Distinct fragments were also inserted into a pET28a vector. Mutant plasmids of hCDC73 were constructed with the Stratagene Quickchange Kit.

Constructed plasmids were transformed into BL21(DE3) for protein expression. Cell cultures were induced with 0.2 mM IPTG for 20 h at 18 °C when OD₆₀₀ reached roughly 0.8. Cell pellets were disrupted by sonication. Afterwards cell pellets were spun down at a speed of 30,700 g. CDC73 proteins were purified by a sequential procedure: nickel column, his-tag cleavage by TEV protease, second nickel column, ion-exchange column and size exclusion column. Purified proteins were stored in a buffer containing 20 mM Tris pH7.5, 400 mM NaCl and 4 mM DTT.

After purification, proteins were concentrated to 10 mg/ml and used for crystallization screens. When an initial hit was observed, optimization was conducted employing the hanging-drop method. The growth condition of CDC73(1–100) crystal is 0.2 M MgCl₂, 0.1 M Tris pH8.5, 25% PEG3350, and CDC73(1–111) crystals were obtained in 0.1 M Tris pH8.5, 25% PEG3350.

Data collection and structure determination. The native X-ray diffraction data sets of CDC73(1–100) and CDC73(1–111), as well as the potassium tetrapermanganate-derivative data set of CDC73(1–100), were collected at the BL19U beam line in Shanghai Synchrotron Radiation Facility (SSRF). All data sets were processed with HKL3000³⁹. The phase problem of CDC73(1–100) was solved by single anomalous dispersion (SAD) method using Autosol in PHENIX⁴⁰, and autobuilding was done with ARP/wARP in CCP4 suite⁴¹. Model building was performed with COOT⁴² and iterative cycles of refinement were done with Refmac5 in CCP4 suite⁴¹ and PHENIX⁴⁰. Thereafter, the structure of CDC73(1–111) was solved by molecular replacement using the Phaser in PHENIX, with our CDC73(1–100) crystal structure as the template. Due to the ultrahigh resolution of CDC73(1–111) crystals, we added alternative conformations and refined occupancy for each conformation during refinement. Anisotropic B factor refinement was used. Finally, hydrogen atoms were added to the structure. All structural pictures were generated using PyMOL⁴³. Crystallographic statistics are shown in Table 1.

Limited proteolysis and mass spectrometry. His-CDC73(1–343) was digested with trypsin at different molar ratios of trypsin:CDC73 (1:6,400, 1:3,200, 1:1,600, 1:800, 1:400 and 1:200). The incubation buffer is 20 mM Tris pH7.5, 400 mM NaCl. Samples were incubated at room temperature for 1 h. After that protein loading buffer was added and boiled at 100 °C for 10 min for SDS-PAGE. The degraded fragment after limited proteolysis was cut from the SDS-PAGE gel and sent for LC-MS/MS analysis in the Laboratory of Proteomics, Institute of Biophysics, CAS.

Analytical ultracentrifugation, CD spectrum and thermofluor assay. Sedimentation velocity experiment was performed with a Beckman ProteomeLab XL-I analytical ultracentrifuge at 20 °C. Proteins were prepared in a buffer containing 20 mM Tris pH7.5, 400 mM NaCl, with A₂₈₀ absorbance of ~0.7. Data collection was performed at 60krpm at a wavelength of 280 nm. Finally, interference sedimentation coefficient distribution (c(M)) was calculated using software SEDFIT.

Circular dichroism was performed using a ChirascanPlus CD spectrometer. The protein buffer was changed to a phosphate buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄) before the experiment, and proteins were concentrated to 0.2 mg/ml.

Melting temperatures of the wild-type, K34Q and del(5–10) mutants of CDC73(1–111) were measured using quantitative PCR machine Rotor-Gene 6600. Before experiment, proteins were concentrated to 4 mg/ml, and then 18 μl PBS was mixed with 1 μl protein and 1 μl SYPRO™ ORANGE dye. The temperature gradient was set to a range from 25 °C to 95 °C with 1 °C/min increment.

Data availability. Atomic coordinates and structure factors of hCDC73(1–100) and hCDC73(1–111) have been deposited in the Protein Data Bank under accession numbers of 5YDF and 5YDE, respectively.

References

1. Carpten, J. D. *et al.* HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nature genetics* **32**, 676–680, <https://doi.org/10.1038/ng1048> (2002).
2. Newey, P. J., Bowl, M. R. & Thakker, R. V. Parafibromin - functional insights. *Journal of Internal Medicine* **266**, 84–98, <https://doi.org/10.1111/j.1365-2796.2009.02107.x> (2009).
3. Newey, P. J., Bowl, M. R., Cranston, T. & Thakker, R. V. Cell division cycle protein 73 homolog (CDC73) mutations in the hyperparathyroidism-jaw tumor syndrome (HPT-JT) and parathyroid tumors. *Human mutation* **31**, 295–307, <https://doi.org/10.1002/humu.21188> (2010).
4. Chen, Y. *et al.* CDC73 gene mutations in sporadic ossifying fibroma of the jaws. *Diagnostic pathology* **11**, 91, <https://doi.org/10.1186/s13000-016-0532-0> (2016).
5. Gill, A. J. Understanding the genetic basis of parathyroid carcinoma. *Endocrine pathology* **25**, 30–34, <https://doi.org/10.1007/s12022-013-9294-3> (2014).
6. Zhao, J. *et al.* Sporadic human renal tumors display frequent allelic imbalances and novel mutations of the HRPT2 gene. *Oncogene* **26**, 3440–3449, <https://doi.org/10.1038/sj.onc.1210131> (2007).
7. Teh, B. T. *et al.* Autosomal dominant primary hyperparathyroidism and jaw tumor syndrome associated with renal hamartomas and cystic kidney disease: linkage to 1q21–q32 and loss of the wild type allele in renal hamartomas. *The Journal of clinical endocrinology and metabolism* **81**, 4204–4211, <https://doi.org/10.1210/jcem.81.12.8954016> (1996).
8. Szabo, J. *et al.* Hereditary hyperparathyroidism-jaw tumor syndrome: the endocrine tumor gene HRPT2 maps to chromosome 1q21–q31. *American journal of human genetics* **56**, 944–950 (1995).
9. Bricaire, L. *et al.* Frequent large germline HRPT2 deletions in a French National cohort of patients with primary hyperparathyroidism. *The Journal of clinical endocrinology and metabolism* **98**, E403–408, <https://doi.org/10.1210/jc.2012-2789> (2013).

10. Cetani, F. *et al.* CDC73 mutational status and loss of parafibromin in the outcome of parathyroid cancer. *Endocrine connections* **2**, 186–195, <https://doi.org/10.1530/ec-13-0046> (2013).
11. Wang, P. F. *et al.* Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice. *Molecular and Cellular Biology* **28**, 2930–2940, <https://doi.org/10.1128/mcb.00654-07> (2008).
12. Xu, Y. *et al.* Architecture of the RNA polymerase II–Paf1C–TFIIS transcription elongation complex. *Nature communications* **8**, 15741, <https://doi.org/10.1038/ncomms15741> (2017).
13. Shi, X. *et al.* Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription. *Mol Cell Biol* **16**, 669–676 (1996).
14. Kim, J., Guermah, M. & Roeder, R. G. The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. *Cell* **140**, 491–503, <https://doi.org/10.1016/j.cell.2009.12.050> (2010).
15. Chu, X. *et al.* Structural insights into Paf1 complex assembly and histone binding. *Nucleic Acids Res* **41**, 10619–10629, <https://doi.org/10.1093/nar/gkt819> (2013).
16. Tomson, B. N. & Arndt, K. M. The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochimica Et Biophysica Acta-Genes Regulatory Mechanisms* **1829**, 116–126, <https://doi.org/10.1016/j.bbagr.2012.08.011> (2013).
17. Woodard, G. E. *et al.* Parafibromin, product of the hyperparathyroidism-jaw tumor syndrome gene HRPT2, regulates cyclin D1/PRAD1 expression. *Oncogene* **24**, 1272–1276, <https://doi.org/10.1038/sj.onc.1208274> (2005).
18. Yang, Y.-J., Han, J.-W., Youn, H.-D. & Cho, E.-J. The tumor suppressor, parafibromin, mediates histone H3 K9 methylation for cyclin D1 repression. *Nucleic Acids Research* **38**, 382–390, <https://doi.org/10.1093/nar/gkp991> (2010).
19. Lin, L., Zhang, J. H., Panicker, L. M. & Simonds, W. F. The parafibromin tumor suppressor protein inhibits cell proliferation by repression of the c-myc proto-oncogene. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17420–17425, <https://doi.org/10.1073/pnas.0710725105> (2008).
20. Mosimann, C., Hausmann, G. & Basler, K. Parafibromin/Hyrax activates Wnt/Wg target gene transcription by direct association with beta-catenin/Armadillo. *Cell* **125**, 327–341, <https://doi.org/10.1016/j.cell.2006.01.053> (2006).
21. Takahashi, A. *et al.* SHP2 Tyrosine Phosphatase Converts Parafibromin/Cdc73 from a Tumor Suppressor to an Oncogenic Driver. *Molecular Cell* **43**, 45–56, <https://doi.org/10.1016/j.molcel.2011.05.014> (2011).
22. Kikuchi, I. *et al.* Dephosphorylated parafibromin is a transcriptional coactivator of the Wnt/Hedgehog/Notch pathways. *Nature communications* **7**, 12887, <https://doi.org/10.1038/ncomms12887> (2016).
23. Mosimann, C., Hausmann, G. & Basler, K. The role of Parafibromin/Hyrax as a nuclear Gli/Ci-interacting protein in Hedgehog target gene control. *Mechanisms of development* **126**, 394–405, <https://doi.org/10.1016/j.mod.2009.02.002> (2009).
24. Amrich, C. G. *et al.* Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin. *The Journal of biological chemistry* **287**, 10863–10875, <https://doi.org/10.1074/jbc.M111.325647> (2012).
25. Chen, H. *et al.* Crystallographic analysis of the conserved C-terminal domain of transcription factor Cdc73 from *Saccharomyces cerevisiae* reveals a GTPase-like fold. *Acta crystallographica. Section D, Biological crystallography* **68**, 953–959, <https://doi.org/10.1107/s0907444912017325> (2012).
26. Hahn, M. A. & Marsh, D. J. Identification of a functional bipartite nuclear localization signal in the tumor suppressor parafibromin. *Oncogene* **24**, 6241–6248, <https://doi.org/10.1038/sj.onc.1208778> (2005).
27. Hahn, M. A. & Marsh, D. J. Nucleolar localization of parafibromin is mediated by three nucleolar localization signals. *FEBS letters* **581**, 5070–5074, <https://doi.org/10.1016/j.febslet.2007.09.050> (2007).
28. Paziienza, V. *et al.* Identification and Functional Characterization of Three NoLS (Nucleolar Localisation Signals) Mutations of the CDC73 Gene. *PLoS one* **8**, doi:<https://doi.org/10.1371/journal.pone.0082292> (2013).
29. Cetani, F. *et al.* Different somatic alterations of the HRPT2 gene in a patient with recurrent sporadic primary hyperparathyroidism carrying an HRPT2 germline mutation. *Endocr Relat Cancer* **14**, 493–499, <https://doi.org/10.1677/erc-06-0092> (2007).
30. Masi, G. *et al.* Characterization of a new CDC73 missense mutation that impairs Parafibromin expression and nucleolar localization. *PLoS one* **9**, e97994, <https://doi.org/10.1371/journal.pone.0097994> (2014).
31. Rozenblatt-Rosen, O. *et al.* The parafibromin tumor suppressor protein is part of a human Paf1 complex. *Molecular and Cellular Biology* **25**, 612–620, <https://doi.org/10.1128/mcb.25.2.612-620.2005> (2005).
32. Emmott, E. & Hiscox, J. A. Nucleolar targeting: the hub of the matter. *Embo Reports* **10**, 231–238, <https://doi.org/10.1038/embor.2009.14> (2009).
33. Pederson, T. & Tsai, R. Y. L. In search of nonribosomal nucleolar protein function and regulation. *Journal of Cell Biology* **184**, 771–776 (2009).
34. Walls, G. V. *et al.* Mice deleted for cell division cycle 73 gene develop parathyroid and uterine tumours: model for the hyperparathyroidism-jaw tumour syndrome. *Oncogene*, doi:<https://doi.org/10.1038/ncr.2017.43> (2017).
35. Iwata, T., Mizusawa, N., Taketani, Y., Itakura, M. & Yoshimoto, K. Parafibromin tumor suppressor enhances cell growth in the cells expressing SV40 large T antigen. *Oncogene* **26**, 6176–6183, <https://doi.org/10.1038/sj.onc.1210445> (2007).
36. Kim, J. & Roeder, R. G. Direct Bre1–Paf1 complex interactions and RING finger-independent Bre1–Rad6 interactions mediate histone H2B ubiquitylation in yeast. *The Journal of biological chemistry* **284**, 20582–20592, <https://doi.org/10.1074/jbc.M109.017442> (2009).
37. Wood, A., Schneider, J., Dover, J., Johnston, M. & Shilatifard, A. The Paf1 complex is essential for histone monoubiquitination by the Rad6–Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *The Journal of biological chemistry* **278**, 34739–34742, <https://doi.org/10.1074/jbc.C300269200> (2003).
38. Bradley, K. J. *et al.* Parafibromin is a nuclear protein with a functional monopartite nuclear localization signal. *Oncogene* **26**, 1213–1221, <https://doi.org/10.1038/sj.onc.1209893> (2007).
39. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallographica Section D-Biological Crystallography* **62**, 859–866, <https://doi.org/10.1107/s0907444906019949> (2006).
40. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213–221, <https://doi.org/10.1107/s0907444909052925> (2010).
41. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **50**, 760–763, doi:<https://doi.org/10.1107/s0907444994003112> (1994).
42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501, <https://doi.org/10.1107/s0907444910007493> (2010).
43. DeLano, W. L. & Brunger, A. T. Helix packing in proteins: prediction and energetic analysis of dimeric, trimeric, and tetrameric GCN4 coiled coil structures. *Proteins* **20**, 105–123, <https://doi.org/10.1002/prot.340200202> (1994).

Acknowledgements

We appreciate the support and guidance of Prof. Dong-Cai Liang. We thank assistance in data collection from the staff of beamline BL19U1/BL17U at the Shanghai Synchrotron Radiation Facility, China. We also thank Xiao-Xia Yu and Jianhui Li for help with AUC and CD analyses at the Institute of Biophysics, Chinese Academy of

Sciences. This work was supported in part by National Natural Science Foundation of China grants 31500610, 31570794 and 31629002, and Chinese Academy of Sciences Pilot Strategic Science and Technology Projects B grant XDB08010303. This work was also funded by Chinese Postdoctoral Science Foundation.

Author Contributions

W.S., X.L.K., Y.P.L. and L.F.T. built constructs, performed protein purification, conducted the biochemical and biophysical assays. W.S. and X.X. Y. obtained crystals of the CDC73-NTD crystals, performed the X-ray diffraction measurements and determined the crystal structure. W.S., X.X.Y. and W.X. wrote the paper. All authors provided editorial comments.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-15715-9>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017