Structure of the inner kinetochore CCAN complex assembled onto a centromeric nucleosome

Kaige Yan1,5, Jing Yang1,5, Ziguang Zhang1,5, Stephen H. McLaughlin1, Leifu Chang1,4, Domenico Fasci2, Ann E. Ehrenhofer-Murray3, Albert J. R. Heck2 & David Barford1*

In eukaryotes, accurate chromosome segregation in mitosis and meiosis maintains genome stability and prevents aneuploidy. Kinetochore complexes are large protein complexes that, by assembling onto specialized Cenp-A nucleosomes1,2, function to connect centromeric chromatin to microtubules of the mitotic spindle3,4. Whereas the centromeres of vertebrate chromosomes comprise millions of DNA base pairs and attach to multiple microtubules, the simple point centromeres of budding yeast are connected to individual microtubules5,6. All 16 budding yeast chromosomes assemble complete kinetochores using a single Cenp-A nucleosome (Cenp-ANuc), each of which is perfectly centred on its cognate centromere7–9. The inner and outer kinetochore modules are responsible for interacting with centromeric chromatin and microtubules, respectively. Here we describe the cryo-electron microscopy structure of the Saccharomyces cerevisiae inner kinetochore module, the constitutive centromere associated network (CCAN) complex, assembled onto a Cenp-A nucleosome (CCAN–Cenp-ANuc). The structure explains the interdependency of the constituent subcomplexes of CCAN and shows how the Y-shaped opening of CCAN accommodates Cenp-ANuc to enable specific CCAN subunits to contact the nucleosomal DNA and histone subunits. Interactions with the unwrapped DNA duplex at the two termini of Cenp-ANuc are mediated predominantly by a DNA-binding groove in the Cenp-L–Cenp-N subcomplex. Disruption of these interactions impairs assembly of CCAN onto Cenp-ANuc. Our data indicate a mechanism of Cenp-A nucleosome recognition by CCAN and how CCAN acts as a platform for assembly of the outer kinetochore to link centromeres to the mitotic spindle for chromosome segregation.

The 14-subunit CCAN complex assembled onto specialized Cenp-A nucleosomes (in which Cenp-A is substituted for histone H3) reconstituted using either an S. cerevisiae centromere sequence or the Widom 601 sequence, with both complexes eluting at similar volumes on size-exclusion chromatography (SEC) (Extended Data Fig. 1a–e). By contrast, CCAN did not assemble onto a canonical H3 nucleosome, indicating the specificity of the CCAN–Cenp-ANuc interaction (Extended Data Fig. 1b, f). Cryo-electron microscopy (cryo-EM) of CCAN–Cenp-ANuc (using the more stable Widom 601–Cenp-ANuc) revealed a heterogeneous population of particles that, by 3D classification, were identified as monomeric free CCAN, a monomer of CCAN in the Y-shaped structure of human CCAN12. The long N-terminal regions in CCAN resemble the negative-stain electron microscopy reconstruction of human CENP-OPQR12. The long N-terminal regions of Cenp-O and Cenp-P, which are disordered in the Klyveromyces lactis crystal structure13, are more structurally connected with Cenp-HIK and Cenp-N (Fig. 1b, c). Four subunits of Cenp-OPQU+ (Cenp-Q, Cenp-U, Nkp1 and Nkp2) form extended α-helices that associate in a parallel, interweaved fashion to create an irregular coiled-coil α-helical bundle. This shares a marked similarity to the outer kinetochore complex Mis12,14,15 (Extended Data Fig. 5c). Nkp1 and Nkp2 create an outer layer of α-helices in Cenp-OPQU+, which are probably substituted by Cenp-R in vertebrates2,2.

The Cenp-HIK module (Fig. 1c), which resembles the free Cenp-HIK complex (Extended Data Fig. 5d), is dominated by the C-terminal heat shock motif repeats of Cenp-I2,17 and the HFD domains of Cenp-T and Cenp-W. The cold-coil α-helices of Cenp-H and Cenp-K run anti-parallel to Cenp-I (Fig. 1c, Extended Data Fig. 4a–c). The base of Cenp-HIK is a four α-helical bundle comprising the N termini of Cenp-H and Cenp-K. The flexible head domain, present in free Cenp-HIK (Cenp-HIKhead), and a small population of CCAN particles (Extended Data Figs. 3c, 5b, d), matches the shape of the crystal structure of the N-terminal Cenp-I heat repeats that are associated with the C termini of both Cenp-H and Cenp-K18. (Fig. 1d). The Cenp-TW subcomplex, comprising the histone-fold domain (HFD) subunits Cenp-T and Cenp-W, is not clearly resolved in cryo-EM maps of CCAN and CCAN–Cenp-ANuc. Cenp-TW associates with Cenp-HIK in solution, consistent with previous studies11,17, and the HFD domains of Cenp-T and Cenp-W (Cenp-THFDW) interacts equally well with a complex comprising Cenp-HIKhead (Extended Data Fig. 1g–j), indicating that the HFDs of Cenp-TW interact directly with Cenp-HIKhead.

The relative organization of CCAN subunits in our cryo-EM reconstruction is in agreement with that defined from the de novo assembly

1MRC Laboratory of Molecular Biology, Cambridge, UK. 2Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, The Netherlands. 3Humboldt-Universität zu Berlin, Institut für Biologie, Berlin, Germany. 4Present address: Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. 5These authors contributed equally: Kaige Yan, Jing Yang, Ziguang Zhang. *e-mail: dbarford@mrc-lmb.cam.ac.uk

https://doi.org/10.1038/s41586-019-1609-1
of the S. cerevisiae kinetochore\(^9\) (Extended Data Fig. 1k) and consistent with a negative-stain electron microscopy reconstruction of the human CENP-HIKM–LN–OPQUR complex\(^12\). To assess the validity of our structure, we performed crosslinking mass spectrometry (XL-MS) analysis of the complexes. Numerous intra- and intersubunit crosslinks were identified (Extended Data Fig. 6a, b, Supplementary Tables 1, 2). Mapping these crosslinks onto CCAN and CCAN–Cenp-A\(^\text{Nuc}\), for which both lysines of the crosslinked pair are defined, showed that 95% of the detected crosslinks are within the expected linker-distance constraints (Extended Data Fig. 6c–f).

Kinetochore assembly onto Cenp-A\(^\text{Nuc}\) (refs \(^9,18\)), the hallmark of centromeric chromatin, with the CCAN subunits Cenp-C and Cenp-N directing this assembly\(^19,20\). In the CCAN–Cenp-A\(^\text{Nuc}\) complex, Cenp-A\(^\text{Nuc}\) is an octameric nucleosome, with DNA wrapped as a left-handed superhelix (Fig. 2, Supplementary Video 1), as previously shown for free Cenp-A\(^\text{Nuc}\) (refs \(^8,21–23\)). Consistent with these reports is that compared with canonical H3 nucleosomes, in the CCAN–Cenp-A\(^\text{Nuc}\) complex, the DNA gyre of Cenp-A\(^\text{Nuc}\) is more loosely wrapped. In CCAN–Cenp-A\(^\text{Nuc}\), only 105 bp of DNA encircle the Cenp-A-octamer, compared with 147 bp for canonical nucleosomes\(^24\) (Figs. 2, 3a, b). A total of 20 bp of DNA are unwrapped equally at each DNA terminus of Cenp-A\(^\text{Nuc}\). One of the unwrapped DNA termini, well defined in cryo-EM density, interacts with CCAN, whereas the other is disordered (Fig. 2a). We observe clearly defined α-helical density for the N-terminal segment of one Cenp-A subunit (Cenp-A\(^\text{Nuc}\)), which is inserted between the unwrapped DNA duplex and DNA gyre (Figs. 2a, 3c).

In the CCAN–Cenp-A\(^\text{Nuc}\) complex (Fig. 2, Supplementary Video 1), Cenp-A\(^\text{Nuc}\) inserts end-on into the Y-shaped opening of CCAN, with each arm of CCAN embracing opposite sides of the nucleosome. This positions the Cenp-LN module to form extensive contacts with the unwrapped DNA duplex at one of the termini of the Cenp-A\(^\text{Nuc}\) DNA gyre (Fig. 2). Cenp-LN adopts a U-shaped structure, creating an evolutionarily conserved, positively charged groove that engages the unwrapped DNA (Fig. 3b, Extended Data Fig. 7a–c). The DNA duplex runs along the Cenp-LN groove, exiting opposite to the nucleosome (Figs. 2, 3a, b). Cenp-HIK\(^\text{Head}\)–Cenp-TW also functions in Cenp-A\(^\text{Nuc}\) recognition, as indicated by the CCAN–Cenp-A\(^\text{Nuc}\) complex, in which cryo-EM density corresponding to Cenp-HIK\(^\text{Head}\)–Cenp-TW contacts the DNA gyre of Cenp-A\(^\text{Nuc}\), with Cenp-I in close proximity to Cenp-A\(^\text{Nuc}\) (Fig. 2c, Extended Data Fig. 3c, Supplementary Video 1). Compared with apo-CCAN, Cenp-HIK\(^\text{Head}\)–Cenp-TW rotates by around 90° to accommodate Cenp-A\(^\text{Nuc}\) (Extended Data Fig. 5e). Previous studies have suggested that the vertebrate Cenp-TWSX heterotetramer forms density of a 3D subclass of the overall CCAN–Cenp-A\(^\text{Nuc}\) 3D class, before application of the mask used to refine the cryo-EM map shown in a (Extended Data Fig. 3a), highlighting contacts to Cenp-A\(^\text{Nuc}\). c, The Cenp-HIK\(^\text{Head}\) module contacts Cenp-A. d, Cenp-TW contacts the DNA gyre of Cenp-A\(^\text{Nuc}\). e, The N-terminal region of Cenp-QU contacts Cenp-A and H4.
Fig. 3 | Cenp-LN interacts with the unwrapped DNA duplex of Cenp-A<sup>Nuc</sup>. a, Two orthogonal views showing the unwrapped DNA duplex of Cenp-A<sup>Nuc</sup> engaged by the DNA-binding groove of the Cenp-LN subcomplex. b, Surface of Cenp-LN showing positive electrostatic potential of the DNA-binding groove. The canonical <i>S. cerevisiae</i> H3 nucleosome (orange; Protein Data Bank (PDB) ID: 1ID3) wraps 147 bp of DNA compared with the 105 bp wrapped by the <i>S. cerevisiae</i> Cenp-A Nuc nucleosome (yellow). c, Magnified view showing insertion of the N-terminus of Cenp-A (Cenp-A<sup>α</sup>) between the unwrapped DNA duplex and DNA gyre of Cenp-A<sup>Nuc</sup>. Arg67 of the Cenp-N pyrin domain inserts into the DNA major groove.

A nucleosome-like particle to interact with DNA<sup>25</sup>. However, this is not compatible with <i>S. cerevisiae</i> Cenp-TW exactly co-localizing with centromeric Cenp-A<sup>Nuc</sup> in a Cenp-I-dependent manner<sup>24</sup>. The HDFs of Cenp-TW were assigned to cryo-EM density associated with Cenp-HIK<sup>Head</sup> contacting the DNA gyre of Cenp-A<sup>Nuc</sup>, visible in a minor 3D class of CCAN–Cenp-A<sup>Nuc</sup> (Fig. 2d, Extended Data Fig. 3c). On the opposite side of CCAN to Cenp-HIK, the N-terminal regions of Cenp-Q and Cenp-U contact the DNA gyre of Cenp-A<sup>Nuc</sup> and the N-termini of Cenp-A and H14 (Fig. 2b, right), e. This is consistent with the Cenp-Q–Cenp-U (Cenp-QU) dimer binding DNA<sup>26</sup> and recognizing the posttranslational status of the N-terminus of Cenp-A<sup>α</sup>, and further validated by our XL-MS data revealing Cenp-Q crosslinks to H2A and H2B (Extended Data Fig. 6b).

Cenp-N engages Cenp-A<sup>Nuc</sup> in the budding yeast CCAN–Cenp-A<sup>Nuc</sup> complex in a different manner to how isolated vertebrate Cenp-N subunit interacts with Cenp-A<sup>Nuc</sup> through the L1 loop of Cenp-A and the adjacent DNA gyre<sup>28,29</sup>. Because of steric clashes, the interaction of Cenp-N with Cenp-A<sup>Nuc</sup> revealed in these studies is not compatible with the position of Cenp-N in the context of the CCAN complex (Extended Data Fig. 7d). Binding of Cenp-A<sup>Nuc</sup> at this interface of CCAN, as previously proposed<sup>18</sup>, would require substantial conformational changes of CCAN. The discrepancy between our structure and that of the vertebrate system may either reflect genuine species differences in CCAN–Cenp-A<sup>Nuc</sup> architectures or result from the vertebrate Cenp-N–Cenp-A<sup>Nuc</sup> structure representing an intermediate in the CCAN–Cenp-A<sup>Nuc</sup> assembly pathway, in accordance with CCAN–Cenp-A<sup>Nuc</sup> remodelling during the cell cycle<sup>11</sup>. Cenp-C also determines kinetochore–Cenp-A<sup>Nuc</sup> interactions<sup>20</sup>, and we found that Cenp-C is required for stable assembly onto Cenp-A<sup>Cen3</sup> nucleosomes (data not shown), although not Cenp-A–Widom 601 nucleosomes (Fig. 4b). Cenp-C interacts with Cenp-A through its Cenp-C motif (Extended Data Fig. 5f), similar to vertebrates<sup>30</sup>. However, the regions of Cenp-C associated with CCAN were not visible in the cryo-EM maps. XL-MS data indicate that Cenp-C participates in multiple interactions with CCAN (Extended Data Fig. 6a, b, g, Supplementary Tables 1, 2).

To test the validity of the CCAN–Cenp-A<sup>Nuc</sup> structure, we mutated 13 Arg and Lys residues in Cenp-N that line the Cenp-LN DNA-binding groove (Fig. 4a) and tested the ability of the mutant CCAN to assemble onto Cenp-A<sup>Nuc</sup>. To avoid complications of Cenp-C interacting with Cenp-A<sup>Nuc</sup>, we used CCAN without Cenp-C (CCAN<sup>ΔCenp-C</sup>). The Cenp-N mutant (Cenp-N<sup>Mut</sup>) did not impair CCAN–Cenp-A<sup>Nuc</sup> assembly, and similar to CCAN, CCAN<sup>ΔCenp-C</sup> binds to Cenp-A–Widom 601 nucleosomes, but not H3 nucleosomes (Fig. 4b, Extended Data Figs. 8a–c, 9a, b). Cenp-N<sup>Mut</sup> disrupted CCAN–Cenp-A<sup>Nuc</sup> interactions (Fig. 4b, Extended Data Fig. 8d). By contrast, mutating the L1 loop of Cenp-A did not disrupt the binding of CCAN–Cenp-C to Cenp-A<sup>Nuc</sup> (Extended Data Figs. 8e, 9a).

We then assessed the role of the unwrapped DNA termini of Cenp-A<sup>Nuc</sup> in mediating CCAN–Cenp-A<sup>Nuc</sup> interactions. Because the αN-helix of the H3 histone stabilizes the wrapped DNA termini of canonical H3 nucleosomes<sup>22,24</sup>, to create a more closed, highly wrapped Cenp-A<sup>Nuc</sup>, we substituted the N-terminal 50 residues of H3 for the N-terminal 140 residues of Cenp-A, creating a chimeric H3<sup>N</sup>–Cenp-A (Extended Data Fig. 7e–g). The resultant H3<sup>N</sup>–Cenp-A<sup>Nuc</sup> wrapped a similar length of DNA as did H3<sup>Nuc</sup> (approximately 147 bp) (Extended Data Fig. 9c). The affinity of CCAN–Cenp-C for H3<sup>N</sup>–Cenp-A<sup>Nuc</sup> was severely disrupted, such that CCAN–Cenp-C was substantially dissociated from H3<sup>N</sup>–Cenp-A<sup>Nuc</sup> (Fig. 4b, Extended Data Fig. 8g). Binding of H3<sup>N</sup>–Cenp-A<sup>Nuc</sup> to CCAN–Cenp-C was completely disrupted with Cenp-N<sup>Mut</sup> (Fig. 4b, Extended Data Fig. 8h). The reduced affinity of CCAN for H3<sup>N</sup>–Cenp-A<sup>Nuc</sup> is not due to the lack of the Cenp-A N-terminus, because CCAN bound to full-length Cenp-A<sup>Nuc</sup> and Cenp-A<sup>Nuc</sup> in which residues 1–129 of Cenp-A are deleted (Δ<sup>N</sup>Cenp-A<sup>Nuc</sup>) equally well (Fig. 4b, Extended Data Fig. 8c, f). These biochemical studies confirm that CCAN interacts with the unwrapped DNA termini of Cenp-A<sup>Nuc</sup> and that a major role of the Cenp-LN DNA-binding groove is to engage the unwrapped DNA gyre of Cenp-A<sup>Nuc</sup>, as shown by the CCAN–Cenp-A<sup>Nuc</sup> cryo-EM structure (Fig. 3b).

Disruption of the <i>S. cerevisiae</i> Cenp-N gene (<i>CHL4</i>) causes chromosome loss and instability without affecting viability<sup>31</sup>. However, combining a <i>chl4</i> deletion with either mutation of Cenp-A (<i>CSE4</i>) or deletion of other kinetochore subunits results in synthetic growth defects and lethality<sup>9,27</sup>. Cenp-N is an essential gene in <i>S. cerevisiae</i> (ref. 27) with a role in chromatin organization<sup>27</sup>. However, the growth defect of the <i>chl4</i> mutant, Cenp-N<sup>Mut</sup> did not impair CCAN (Extended Data Fig. 7d). Binding of Cenp-A<sup>Nuc</sup> to Cenp-C was substantially disrupted with Cenp-N<sup>Mut</sup> (Fig. 4c–d). This result demonstrates a functional role for the CCAN–DNA-binding groove, and together with our biochemical data (Fig. 4b, Extended Data Fig. 8), supports the CCAN–Cenp-A<sup>Nuc</sup> architecture that we report here. In <i>S. cerevisiae</i>, Cenp-A<sup>Nuc</sup> is linked to the outer kinetochore Ndc80 complex and associated microtubules through a pathway comprising the essential proteins Cenp-C, Cenp-QU and the Mis12 complex and by a second pathway involving Cenp-TW and Cenp-N<sup>Δ</sup> (Extended Data Fig. 1k). The location of Cenp-N at the centre of CCAN is consistent with these two pathways. The unwrapped DNA termini of Cenp-A<sup>Nuc</sup> contribute to stabilizing the CCAN–Cenp-A<sup>Nuc</sup> complex through the CCAN–DNA-binding groove, augmented by contacts of both Cenp-A and the Cenp-A<sup>Nuc</sup> DNA gyre with Cenp-C (Extended Data Fig. 5f). Cenp-LN (Fig. 3b), Cenp-TW, Cenp-HIK<sup>Head</sup> and Cenp-QU<sup>7</sup> (Fig. 2c–e).

In the cryo-EM reconstruction, Cenp-A<sup>Nuc</sup> is associated with a single CCAN, whereas the expected stoichiometry is two CCANs to one Cenp-A<sup>Nuc</sup> (ref. 32). SEC with multi-angle light scattering (SEC–MALS) and analytical ultracentrifugation confirmed that the reconstituted CCAN–Cenp-A<sup>Nuc</sup> is consistent with two CCANs per Cenp-A<sup>Nuc</sup> (CCAN–Cenp-A<sup>Nuc</sup> complex) (Extended Data Fig. 10a–g). In a generated model of (CCAN)<sub>2</sub>–Cenp-A<sup>Nuc</sup>, two CCAN complexes associate through their tips of the Y, creating a slot that perfectly accommodates
Cenp-A\textsuperscript{Nuc} is inserted vertically (Fig. 4e). The two CCAN complexes cradle Cenp-A\textsuperscript{Nuc} with its unwrapped DNA duplexes stretched out, overlying the DNA-binding surface of CCAN, consistent with XL-MS crosslinks between Cenp-Q and Cenp-TW (Extended Data Fig. 6b). Extensive 2D classification of the cryo-EM data identified XL-MS crosslinks between Cenp-Q and Cenp-TW (Extended Data Fig. 8a). The DNA-binding groove functions in vivo in S. cerevisiae. Wild-type Cenp-N (CHL4\textsuperscript{WT}) rescues the growth defect of the chl4\Delta cse4-R73A mutant strain at 37°C, whereas the Cenp-N\textsuperscript{Mut} (chl4\textsuperscript{Mut}) does not. WT, wild-type strain. This experiment was performed independently ten times with similar results.

Extended Data Fig. 10h shows dynein and acetyl-CoA carboxylase. Experiments in a mutant strain (left) and loading control (right; Coomassie-blue-stained gel shows dynein and acetyl-CoA carboxylase). Experiments in d were performed independently in triplicate with similar results. c, The DNA-binding groove functions in vivo in S. cerevisiae. Wild-type Cenp-N (CHL4\textsuperscript{WT}) rescues the growth defect of the chl4\Delta cse4-R73A mutant strain at 37°C, whereas the Cenp-N\textsuperscript{Mut} (chl4\textsuperscript{Mut}) does not. WT, wild-type strain. This experiment was performed independently ten times with similar results. d, Western blot showing that Cenp-N\textsuperscript{WT} and Cenp-N\textsuperscript{Mut} are expressed at equivalent levels in the chl4\Delta cse4-R73A mutant strain (left) and loading control (right; Coomassie-blue-stained gel shows dynein and acetyl-CoA carboxylase). Experiments in d were performed independently in triplicate with similar results. e, Two views showing a representation of the (CCAN)\textsubscript{2}–Cenp-A\textsuperscript{Nuc} complex with the second CCAN protomer generated by the dyad symmetry of Cenp-A\textsuperscript{Nuc}. Sites of contact to the outer kinetochore (through Cenp-U and Cenp-T) are indicated. For gel source data, see Supplementary Fig. 1.

In S. cerevisiae, the CBF3 complex engages the CDEIII element of the approximately 125-bp centromere to direct Cenp-A nucleosome deposition. Modelling indicates that Cenp-A\textsuperscript{Nuc} can simultaneously accommodate CBF3 only when bound to a single CCAN promoter (Extended Data Fig. 9d), which suggests that CBF3 would not associate with a fully assembled kinetochore.

The (CCAN)\textsubscript{2}–Cenp-A\textsuperscript{Nuc} model suggests two possibilities for how a kinetochore-attached microtubule would segregate centromeric chromatin (Extended Data Fig. 10i, j, Supplementary Video 2). In one scenario, CCAN attaches to the microtubule through the outer kinetochore using the same face as its DNA-binding surface (Extended Data Fig. 10i). This would sandwich the DNA between CCAN and the outer kinetochore, a possibility compatible with the long flexible linkers that attach CCAN to the outer kinetochore. As the microtubule pulls on the kinetochore, CCAAN would hoist the overlying DNA. Alternatively, microtubules could attach to CCAN from the opposite face to its DNA-binding surface, so the chromosome is pulled from behind the inner kinetochore.
kinetochore (Extended Data Fig. 10). Because vertebrate Cenp-A^Nuc also wraps between 100–120 bp (of α-satellite DNA)^32 with nucleosome unwrapping enhanced by Cenp-C^33 and the human CCAN architecture^12 is similar to that of yeast, it is likely that the mechanism of recognition of the specialized Cenp-A nucleosome that we describe here for the budding yeast inner kinetochore is evolutionarily conserved.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1609-1.

Received: 22 January 2019; Accepted: 4 September 2019; Published online 2 October 2019.

1. Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmunore sera from patients with scleroderma. Chromosoma 91, 313–321 (1985).
2. Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of Saccharomyces cerevisiae. Cell 94, 607–613 (1998).
3. Cheeseman, I. M. The kinetochore. Cold Spring Harb. Perspect. Biol. 6, a015826 (2014).
4. Musacchio, A. & Desai, A. A molecular view of kinetochore assembly and function. Biolog 6, 5 (2017).
5. Clarke, L. & Carboni, J. Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature 287, 504–509 (1980).
6. Winey, M. et al. Three-dimensional ultrastructural analysis of the Saccharomyces cerevisiae mitotic spindle. J. Cell Biol. 129, 1601–1615 (1995).
7. Furuyama, S. & Biggins, S. Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc. Natl Acad. Sci. USA 104, 14706–14711 (2007).
8. Camahort, R. et al. Cse4 is part of an octameric nucleosome in budding yeast. Mol. Cell 35, 794–805 (2009).
9. Lang, J., Barber, A. & Biggins, S. An assay for de novo kinetochore assembly reveals a key role for the CENP-T pathway in budding yeast. eLife 7, e37819 (2018).
10. Hinshaw, S. M. & Harrison, S. C. The structure of the Ctf19c/CCAN from budding yeast. eLife 8, e44239 (2019).
11. McKinley, K. L. et al. The CENP-L-N complex forms a critical node in an integrated meshwork of interactions at the centromere-kinetochore interface. Mol. Cell 60, 886–898 (2015).
12. Pesenti, M. E. et al. Reconstitution of a 26-subunit human kinetochore reveals cooperative microtubule binding by CENP-Q/OPUR and NDC80. Molecular Cell 71, 923–939 (2018).
13. Schmitzberger, F. & Harrison, S. C. RWD domain: a recurring module in kinetochore architecture shown by a Ctf13–Mcm21 complex structure. EMBO Rep. 13, 216–222 (2012).
14. Dimitrova, Y. N., Jenni, S., Valverde, R., Khin, Y. & Harrison, S. C. Structure of the MIND complex defines a regulatory focus for yeast kinetochore assembly. Cell 167, 1014–1027 (2016).
15. Petrovic, A. et al. Structure of the MIS12 complex and molecular basis of its interaction with CENP-C at human kinetochores. Cell 167, 1028–1040 (2016).
16. Hu, L. et al. Structural analysis of fungal CENP-H/I/K homologs reveals a conserved assembly mechanism underlying proper chromosome alignment. Nucleic Acids Res. 47, 468–479 (2019).
17. Peķģūz Aluntkaya, G. et al. CCAN assembly configures composite binding interfaces to promote cross-linking of Ndc80 complexes at the kinetochore. Curr. Biol. 26, 2370–2378 (2016).
18. Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. In vitro centromere and kinetochore assembly on defined chromatin templates. Nature 477, 354–358 (2011).
19. Carroll, C. W., Silva, M. C., Godek, K. M., Jansen, L. E. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat. Cell Biol. 11, 896–902 (2009).
20. Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. J. Cell Biol. 189, 1143–1155 (2010).
21. Kingston, I. J., Yung, J. S. & Singleton, M. R. Biophysical characterization of the centromere-specific nucleosome from budding yeast. J. Biol. Chem. 286, 4021–4026 (2011).
22. Tachibana, H. et al. Crystal structure of the human centromeric nucleosome containing CENP-A. Nature 476, 232–235 (2011).
23. Roulland, Y. et al. The flexible ends of CENP-A nucleosome are required for mitotic fidelity. Mol. Cell 63, 674–685 (2016).
24. White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J. 20, 5207–5218 (2001).
25. Nishino, T. et al. CENP-T–W–S–X forms a unique centromeric chromatin structure with a histone-like fold. Cell 140, 487–501 (2012).
26. Hornung, P. et al. A cooperative mechanism drives budding yeast kinetochore assembly downstream of CENP-A. J. Cell Biol. 206, 509–524 (2014).
27. Anedchenko, E. A. et al. The kinetochore module Okp1^CENP-Q/Ame1^CENP-U is a reader for N-terminal modifications on the centromeric histone Cse4^CENP-A. EMBO J. 38, e98991 (2019).
28. Chittori, S. et al. Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. Science 359, 339–343 (2018).
29. Pentakota, S. et al. Decoding the centromeric nucleosome through CENP-N. eLife 6, e33442 (2017).
30. Kato, H. et al. A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. Science 340, 1110–1113 (2013).
31. Kourprina, N. et al. Identification and cloning of the CHL4 gene controlling chromosome segregation in yeast. Genetics 135, 327–341 (1993).
32. Weir, J. R. et al. Insights from biochemical reconstitution into the architecture of human kinetochores. Nature 537, 249–253 (2016).
33. Falk, S. J. et al. Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. Science 348, 699–703 (2015).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© Crown 2019
METHODS
No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cloning, expression, purification and reconstitution of recombinant CCAN–Cenp-ANuc nucleosome complex. Cloning. The genes for CTF19, OKP1, MCM21, AME1, NKP1, NKP2, CTF3, MCM16, MCM22, CNN1, WIP1, MIP2, CHL4 and IMC1 (MACM19) (see Extended Data Table 2 for vertebrate Cenp homologues) were amplified by PCR from S. cerevisiae genomic DNA and cloned into a pU1 plasmid using a modified Multibac expression system34. The intron in MCM21 was deleted by the USER method4. A double SteplI tag together with a TVE cleavage site was attached to the C termini of Ame1, Ctf3, Chl4, Mif2 and Cnn1 proteins. For expression of the Cenp-OPU+ complex (also called COMA+)—Ctf19, Okp1, Mcm21, Ame1, Nkp1 and Nkp2 gene expression cassettes in pU1 were subsequently cloned into a pF2 vector34. The gene expression cassettes for CTF3, MCM16, MCM22, CNN1 and WIP1 were cloned into pF2 to generate the Cenp-HIK–TW complex. Cenp–HIK–TW complexes. To test which regions of Cenp-H, Cenp-I and Cenp-K interact with each other and with Cenp-TW, the following fragments of Cenp-H, Cenp-I and Cenp-K were constructed: Cenp-I (residues 1–308) (Cenp–I308), Cenp-H (residues 137–182) (Cenp–H5), Cenp-H (residues 130–239) (Cenp–K7) and combinations of Cenp-H, Cenp-I and Cenp-K, together with Cenp-TW were assembled into the pU1 plasmid for Multibac expression34 for co-expression using the insect cell–baculovirus system. A double SteplI tag was added to the C terminus of Cenp-I.

To test the role of the positively charged DNA-binding groove of Cenp-N for DNA–histone nucleosome interactions, a total of 13 Arg and Lys mutations were introduced into CHL4 (Cenp–Nmut) by total gene synthesis (GeneArt/Thermo Fisher); chl42252s250b6s75kX100sK105sS108sK215sS217sK245sK249sK410sK415s. Cenp–Nmut was combined with Cenp-L to generate a Cenp–Nmut–Cenp-L co-expression baculovirus.

The baculoviruses for expression of Cenp–OPU+ , Cenp–HIK–TW, Cenp–C and Cenp–LN were prepared for expression using the insect cell–baculovirus system34.

The cDNA encoding S. cerevisiae CSE4 (S. cerevisiae CENP-A), H2A, H2B and H4 histone genes were synthesized (GeneArt/Thermo Fisher) with optimized codons for expression in Escherichia coli and were subsequently cloned into pET28a with a TEV protease cleavable N-terminal His6 tag. For the recombinant Cse4 octamer (Cenp–A octamer), four expression cassettes for CSE4, H2A, H2B and H4 histone genes were subsequently cloned into a single pET28a plasmid by USER methodology4 for E. coli expression. For S. cerevisiae H3 octamer purification, CSE4 was replaced by the H3 gene. The Cenp–A1 loop mutant (Cenp–A13, cse––A13225G4T714A327A347A348A350A351) and cse4K172S/D173A/Q174A/D175S (Cenp–Nmut) by total gene synthesis (GeneArt/Thermo Fisher): cse4(K172S/D173A/Q174A/D175S). The baculoviruses for expression of Cenp–A13, Cenp–Nmut–Cenp–L were produced with Cenp-A nucleosome followed by Cenp-HIK–TW and Cenp–OPU+. The stoichiometry of CCAN subcomplexes to Cenp–ANuc was adjusted so that CCAN subcomplexes were in excess, as judged by their separation from CCAN–Cenp–ANuc by SEC. The mixed sample was dialysed overnight in a buffer of 10 mM Heps (pH8.0), 80 mM NaCl, 1 mM EDTA and 0.5 mM TCEP at 4 °C. CCAN–Cenp–ANuc was purified by Superose 6 SEC. For cryo-EM analysis, CCAN–Cenp–ANuc was crosslinked with 5 mM BS3 (Thermo Fisher Scientific) for 1 h on ice and quenched with 50 mM Tris and then subjected to further SEC with an Agilent SEC-5 column (2.6 × 15 cm) in a buffer containing 200 mM NaCl, 10 mM Heps (pH7.5), 1 mM EDTA and 2 mM DTT. The NaCl concentration in the dialysis buffer was gradually decreased to 100 mM using an Akta pump at 1.5 ml min−1 for 16 h at 4 °C. The mixture was further dialysed against the buffer of 100 mM NaCl, 10 mM Heps (pH7.5), 1 mM EDTA, 2 mM DTT for 4 h at 4 °C.

The Cenp–A nucleosome and derivatives were stored at 4 °C.

Reconstitution of CCAN–Cenp–A nucleosome complex. The CCAN–Cenp–A nucleosome complex was reconstituted by mixing purified Cenp–C and Cenp–LN with Cenp–A nucleosome followed by Cenp–HIK–TW and Cenp–OPU+. Uncrosslinked and BS3 crosslinked CCAN–Cenp–ANuc were prepared during preparation of cryo-EM grids. To assess whether crosslinked created artefacts, we also collected a cryo-EM dataset using uncrosslinked CCAN–Cenp–ANuc.

SEC analysis of CCAN–Cenp–ANuc complexes. To analyse the formation and stability of CCAN–Cenp–ANuc complexes and mutants in CCAN and Cenp–A, all CCAN–Cenp–ANuc complexes were assembled as above (with or without Cenp–C) and then applied to an Agilent Bio SEC-5 SEC column. The eluted fractions were analysed on SDS–PAGE gels and stained with Coomassie blue and ethidium bromide to detect proteins and DNA. For assembly of the CCAN–Cenp–ANuc complexes, the concentration of Cenp–ANuc was 1.6 μM, and that for the individual CCAN subcomplexes was 1.6 μM.

SEC–MALS. SEC–MALS was performed using a Wyatt MALS system. CCAN alone, uncrosslinked and BS3 crosslinked CCAN–Cenp–ANuc complexes were injected onto an Agilent Bio SEC-5 column gel filtration column pre-equilibrated in 10 mM Heps (pH 7.5), 80 mM NaCl, 1 mM EDTA and 0.5 mM TCEP using the same buffer at 4 °C. The light scattering and protein concentration at each point across the peaks in the chromatogram were used to determine the absolute molecular mass from the intercept of the Debye plot using Zimm’s model as implemented in the ASTRA v.5.3.4.20 software (Wyatt Technologies). To determine inter-detector delay volumes, band-broadening constants and detector intensity normalization constants for the instrument, we used aldolase as a standard prior to sample measurement. Data were plotted with the program Prism v8.2.0 (GraphPad Software).

Analytical ultracentrifugation. Uncrosslinked and BS3 crosslinked CCAN–Cenp–ANuc complex at approximately 1 mg ml−1 in 10 mM Heps (pH 7.5), 80 mM NaCl, 1 mM EDTA and 0.5 mM TCEP were subjected to velocity sedimentation at 40,000 r.p.m. at 4 °C in an An50Ti rotor using an Optima XL-A analytical ultracentrifuge equipped with a dye–protein optical system and a 21.1 m optical cell, using a (c) distribution model. The partial-specific volumes (v-bar) were calculated using Sednterp (v20130813 beta) (T. Laue, University of New Hampshire). The density and viscosity of the buffer were determined with a DMA 4500D density meter.
Microccocal nuclease digestion assay. Nucleosomes were digested for 40 min with 1 unit of MNase (NEB) per microgram of DNA at room temperature (22°C). Reactions were terminated with the addition of excess EGTA. The digested nucleosome mixtures were loaded onto an agarose gel and stained to visualize the DNA.

Yeast strains and growth analysis. The S. cerevisiae strain with a chi4 deletion and cse4-R37A mutation (CHL4 Δcse4-R37A), AY6992 (MATα ade2-101 his3-11,15 trpl-1 leu2-3,112 ura3-1 can1-100 chi4-M兴kan-cxe4-R37A) and wild-type S. cerevisiae strain (W303) (MATα ade2-101 his3-11,15 trpl-1 leu2-3,112 ura3-1) have previously been described and authenticated27,28. Yeast strains do not have mycoplasma and were not tested for mycoplasma contamination. Cenp-NWT and Cenp-NStu strains were created by transferring the pYe2 plasmid incorporating either CHL4Δ or chi4Δ into strain AY6992 (MATα chi4-Mαkan-cxe4-R37A), with the native promoter of CHL4. A C-terminal double StreptII-tag on CHL4, and the URA3 selection marker. The transformants were selected on synthetic medium lacking uracil, and the plasmid-encoded CHL4 was verified by PCR using a primer pair over-spanning the CHL4 and URA3 genes. Cells were grown in drop-out uracil (SC-U) medium at 30°C and spotted in tenfold dilution steps on YPED plates. The plates were incubated at either 30°C or 37°C for three days.

Immunoprecipitation and western blotting for detecting Cenp-N expression in the chl4Δ cse4-R37A yeast. Six litres of synthetic SC-U culture were inoculated with the chl4Δ cse4-R37A yeast strain transformed with the pYe2 plasmid expressing either wild type or mutant Cenp-N with a C-terminal double StreptII-tag and collected at OD600 of approximately 0.8. Pelleted cells were lysed in buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 1 mM EDTA and 1 mM DTT) and the cleared lysate was loaded onto a 1-mL Streptactin column. Fractions were eluted with 5 mM desthiobiotin and analysed by SDS-PAGE. Western blotting was performed with a Strept-tag antibody (MCA2489P, Bio-Rad) that detected the C-terminal double StreptII-tag on Cenp-N. Total protein was analysed by Coomassie blue staining for loading controls (normalized loading).

Electron microscopy data collection. Three microlitres of the CCAN–Cenp-AΔ apo-complex at a concentration of about 1 mg mL⁻¹ was applied to glow-discharged copper 300 mesh Quantifoil R1.2/1.3 holey carbon grids (Quantifoil Micro Tools) (no carbon support). The grids were flash-frozen by being plunged into liquid ethane using an FEI Vitrobot Mark IV (waiting time, 20 s, blotting time, 2 s). Cryo-EM image stacks were collected with a Falcon III cameras in counting mode on four different FEI Titan Krios electron microscopes at a nominal magnification of 75 K (yielding pixel sizes of 1.065Å, 1.070 Å, 1.085 Å and 1.090 Å, respectively). The images were recorded at a dose rate of 0.6 electrons per pixel per second and the total exposure time was 60 s (75 frames) with the FEI automated low-dose solution for Cenp-HIK complex. The images were visualized in COOT 44 and Chimera 45. The crystal structure of Cenp-HIK was used as a reference model for building the map. The map was used for the subsequent Bayesian polishing, multi-body refinement (MBR), and the final map refinement and atomic coordinate refinement. Beam-tilt parameters of the particles were estimated based on the individual dataset in RELION 3.0. Refinements in 3D and MBRs were performed with the polished particle stacks after merging all the datasets. The dataset including all the particles generated the highest resolution reconstruction with an overall CCAN mass. The final resolution of CCAN–Cenp-AΔ apo (residues 334–450) is 3.15 Å, respectively, based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion (Extended Data Fig. 2d).

To identify (CCAN)−Cenp-AΔ apo particles, five 2D classes, with 2D averages of CCAN-Cenp-AΔ (Extended Data Fig. 2c) that showed smeared density in close proximity to Cenp-AΔ apo, were selected for further analyses. The selected particles (10,553 particles) were subject to a tandem cassette of 2D classifications, resulting in 556 particles, which showed clear C2-symmetry 2D averages. These particles were re-extracted from the micrographs with a box size of 400 pixels to accommodate the bigger symmetric particles. The re-extracted particles were then subject to further 2D classification, and classified into 20 classes, generating the representative symmetric 2D averages shown in the red box of Extended Data Fig. 2c. The reprojections of the modelled (CCAN)−Cenp-AΔ apo (filtered to 20 Å resolution) were generated with relion_project. The projections are shown in Extended Data Fig. 10. The small number of particles and highly preferred orientation on the cryo-EM grid (in the plane of the two-fold symmetry axis) precluded a 3D reconstruction.

MBR. To improve map resolution we performed MBR in RELION 3.0. Two masks were computed Cenp-LN-OPQU+ and Cenp-HK from the initial template-free particle picking. The initial mask-free particles (Cenp-HIK) and regions of Cenp-N, Cenp-U, Nkp1 and Nkp2 from small regions of Cenp-O and Cenp-P. Mask4 comprised Cenp-OP QU-N and C-terminal regions of Cenp-H, Cenp-Δ, Nkp1 and Nkp2. MBR based on mask4 showed substantially improved definition of cryo-EM densities and were used for model building (Extended Data Figs. 2d, h, i, 3b). Careful choice of the boundaries of mask2 was critical to optimizing the cryo-EM density quality for Cenp-HIK. Including specific regions of Cenp-N, N, O and P within mask2 was critical to generating maps that allowed side-chain definition of the coiled-coil regions of Cenp-H and Cenp-K (Extended Data Fig. 4a). This defined the correct assign -ment and polarity of these chains. MBR also improved definition of side chains in the base of Cenp-HIK. The subsequent MBR using mask3 and mask4 improved solution for detecting Cenp-HIK complex. Portions of the apo-CCAN density map shown in Extended Data Fig. 4. A 3D class (4% of total apo-CCAN) corresponding to dimeric apo-CCAN was determined at 9 Å resolution (Extended Data Fig. 3a).

For the uncrosslinked dataset, the same procedures were applied. A total of 123,215 particles with 1,586 micrographs were used for the final reconstruction of a map at 7.8 Å resolution for the CCAN–Cenp-AΔ apo complex (Extended Data Fig. 5a).

For the isolated Cenp-HIK complex, the same procedure was applied. A total of 374,158 particles were used for the final reconstruction of a map at 4.3 Å resolution.

Before visualization, a negative B factor determined with RELION 2.1 was applied to the density map for sharpening. The modulation transfer function of the detector was corrected in the post-processing step with RELION 3.0. The local resolution was estimated with RELION 3.0.

Model building and structure refinement. Apo-CCAN. Cryo-EM density maps were visualized in COOT and Chimera. The crystal structure of K. lactis Cenp-OPQ (PDB: 5MU3) (equivalent to S. cerevisiae Cenp-O residues 159–362, S. cerevisiae Cenp-P residues 148–361 and S. cerevisiae Cenp-Q residues 330–342) and structures of S. cerevisiae Cenp-N (residues 374–450), Cenp-L (PDB: 4JE3) and human Cenp-N N-terminal domain (NTD) (PDB: 6EQT) (equivalent to residues 12–260 of S. cerevisiae Cenp-N) were fitted into the cryo-EM density maps of apo-CCAN, with retinout and refactoring to the S. cerevisiae sequence for Cenp-NNTD, Cenp-O, Cenp-P and Cenp-Q. On the basis of the quality of the cryo-EM densities, atomic models of Nkp1, Nkp2, Cenp-U, Cenp-Q, Cenp-H (residues 7–136), Cenp-I (residues 321–728) and Cenp-K (residues 4–128) and the interdomain region of Cenp-N (residues 261 to 373) were built de novo. Only the final resolution of apo-CCAN and Cenp-U (residues 334–450) and Cenp-Q (residues 330–342) were built as polyAla (Extended Data Table 2). The secondary-structural and disordered regions of the protein sequences were analysed with PHYRE2 and PSIPred. A model for the Cenp-HIK head domain was based on the crystal structure of (Anton Paar) and an AMVn viscometer (Anton Paar). Data were plotted with the program GUSST19.
regions of the Cenp-HK assembly from Chaetomium thermophilum and Thielavia terrestris (PDB: 5Z28) corresponding to S. cerevisiae Cenp-H (residues Asp143 to Leu181), Cenp-I (residues Leu5 to Ala241) and Cenp-K (residues Ala136 to Thr236) and derived using PHYRE2. The 3.5 Å monomeric free CCAN coordinates were rigid-body-docked into the cryo-EM map. The Cenp-HK head domain was fitted to cryo-EM density of the dimeric apo-CCAN. A linker region that connects Cenp-pA-CCAN with Cenp-pA-CCAN, not present in crystal structures, was built de novo.

CCAN–Cenp-A-CCAN complex was then fit into the CCAN–Cenp-A-CCAN cryo-EM map. The nucleosome was modelled on the S. cerevisiae H3 nucleosome (PDB: 1HD3) with S. cerevisiae Cenp-A modelled on human Cenp-A (PDB: 3AN2) and mutated to the S. cerevisiae Cenp-A sequence, and the Widom 601 DNA sequence (PDB: 3LZ0). The Cenp-C model (PDB: 4X23) in the cen-

The apo-CCAN and CCAN–Cenp-A-CCAN models (excluding the Cenp-HK head domains) were optimized by several rounds of real-space refinement using PHENIX (phenix.real_space_refine)25. Standard stereochemical and secondary structural constraints were applied during the real-space refinement. The final models were evaluated with COOT44, PHENIX45 and MolProbity (http://molprobity.biochem.duke.edu/)43. Figures were prepared using ChimeraX42, Chimera45 and PyMOL (Molecular Graphics System, 2.0.3, Schrodinger). Details of the fitted and refined coordinates in Extended Data Table 2. Multiple sequence alignments were performed and displayed using JALVIEW35.

XL-MS. To assess the quality of our structure, we performed XL-MS analysis of the complexes26. Three independent crosslinking reactions were performed for each sample. The CCAN or CCAN–Cenp-A-CCAN complexes in 20 mM Hepes pH 7.5, 80 mM NaCl and 3 mM MgCl2 were crosslinked with 1 mM DSSO for 15 min at room temperature. Each reaction was quenched with Tris.HCl (pH 8.0) to 50 mM and supplemented with urea to 8 M. The samples were reduced by addition of DTT at a final concentration of 10 mM for 1 h at room temperature, and alkylated for 0.5 h at room temperature in the dark by addition of iodoacetamide to 50 mM. Protein digestion was performed with Lys-C and trypsin at an enzyme-to-protein ratio of 1:75 (w:w) at 37 °C for 16 h. The digested samples were acidified with formic acid to 1%, desalted using home-made C18 stage tips, dried and stored at −80 °C for further use.

Each sample was analysed by liquid chromatography with tandem mass spectrometry using an Agilent 1290 Infinity System (Agilent Technologies) in combination with an Orbitrap Fusion Lumos (Thermo Scientific). Reverse-phase chromatography was carried out using a 100-μm inner diameter, 2-cm trap column (packed in-house with ReproSil-Pur C18-AQ, 3 μm) coupled to a 75-μm (packed in-house with ReproSil-Pur C18-AQ, 3 μm) inner diameter, 2-cm trap column. Chromatography was carried out using a 100-μm inner diameter, 2-cm trap column (packed in-house with ReproSil-Pur C18-AQ, 3 μm) coupled to a 75-μm (packed in-house with ReproSil-Pur C18-AQ, 3 μm) inner diameter, 2-cm trap column. Reverse-phase chromatography was carried out using a 100-μm inner diameter, 2-cm trap column (packed in-house with ReproSil-Pur C18-AQ, 3 μm) coupled to a 75-μm (packed in-house with ReproSil-Pur C18-AQ, 3 μm) inner diameter, 2-cm trap column. Each sample was digested with Lys-C and trypsin at an enzyme-to-protein ratio of 1:75 (w:w) at 30 °C for 3 h, then the samples were diluted in 50 mM ammonium bicarbonate and further digested with trypsin at an enzyme-to-protein ratio of 1:75 (w:w) at 37 °C for 16 h. The digested samples were acidified with formic acid to 1%, desalted using home-made C18 tip and freeze-dried at 80 °C for further use.

Mass spectrum acquisition was performed using the MS2_MS3 strategy: the MS1 scan was recorded in Orbitrap at a resolution of 60,000, the selected precursors were fragmented in MS2 with CID and the crosslinker signature peaks recorded at a resolution of 30,000. The fragments displaying the mass difference specific for DSSO were further fragmented in a MS2 scan in the ion trap27. Each sample was analysed with Proteome Discoverer 2.3 (v2.3.0.522) with the XlinkN nodes integrated28 and searching against databases generated after bottom-up analysis of the samples. The crosslink output (Supplementary Tables 1, 2) was subsequently visualized using the xVLS35 web tool and the crosslinks mapped onto the cryo-EM structure of the nucleosome core particle at 2.8 Å resolution (Nature 389, 251–296 (1997)).

Modelling the CCAN–Cenp-A-CCAN–Cenp-A-CCAN complex. Modelling CCAN and CBF3 complexes simultaneously to the bound to Cenp-A nucleosome, we docked the free unwrapped DNA duplex of the CCAN–Cenp-A-CCAN complex onto the CBF3–Cenp-A coordinates (PDB: 6GYS)46, matching the minor and major grooves of both complexes. To avoid overlap of CBF3 and CCAN, the dyad symmetry axis of the Cenp-A nucleosome was positioned seven nucleotides upstream of the midpoint of CDEII of the Cenp-A sequence. Modelling human and S. pombe Cenp-LN complexes. To generate the human Cenp-LN complex we used residues 1–207 from PDB 6EQW47, and modelled residues 208–338 and Cenp-N by one-to-one threading in PHYRE248 using S. cerevisiae Cenp-LN as a template. S. pombe Cenp-LN was modelled with PHYRE248 using S. cerevisiae Cenp-LN as a template. The electrostatic potential of S. cerevisiae Cenp-A-CCAN and S. pombe Cenp-LN complexes were calculated and displayed in PyMOL (Molecular Graphics System, 2.0.3, Schrodinger). Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Electron microscopy maps have been deposited with the Electron Microscopy Data Bank with accession codes EMD-4580 (CCAC), EMD-4579 (CCAN–Cenp-A-CCAN), EMD-4581 (mask1) and EMD-4971 (mask2). Protein coordinates have been deposited with the PDB with accession codes 6QLE (CCAN), 6QLD (CCAN–Cenp-A-CCAN) and 6QLF (mask1). The XL-MS raw files, the associated output and databases have been deposited through the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013769. Other data are available upon reasonable request.
Acknowledgements This work was funded by MRC grant (MC_UP_1201/6) and CRUK grant (CS76/A14109) to D.B., Horizon 2020 program INFRAIA project Epic-XS (Project 823839) to A.J.R.H. and Deutsche Forschungsgemeinschaft (EH237/12-1) to A.E.E.-M. We thank the LMB, eBIC and the Universities of Cambridge and Leeds Electron Microscopy facilities for help with the electron microscopy data collection, S. Scheres for help with electron microscopy processing, members of the Barford group for useful discussions, J. Grimmett and T. Darling for computing and J. Shi for help with insect cell expression.

Author contributions Z.Z. cloned kinetochore and nucleosome constructs. J.Y. and Z.Z. purified proteins, performed the protein-complex reconstitutions and biochemical and genetic analyses. K.Y. and L.C. prepared cryo-EM grids, collected and analysed electron microscopy data and determined the 3D reconstructions of CCAN–Cenp-A\textsuperscript{Hm} and free Cenp-HIK, respectively. D.B. and K.Y. fitted coordinates and built models and J.Y. and S.H.M. performed SEC–MALS and analytical ultracentrifugation. D.F. collected and analysed XL-MS data. A.J.R.H. directed XL-MS experiments and analysis. A.E.E.-M. generated the chl4\textsuperscript{Δ} cse4-R37A and chl4\textsuperscript{Δ} yeast strains. D.B. directed the project. K.Y. and D.B. wrote the manuscript with help from all authors.

Competing interests The authors declare no competing interests.

Additional information Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1609-1.

Correspondence and requests for materials should be addressed to D.B.

Reviewer information Nature thanks Eva Nogales and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Reconstituted *S. cerevisiae* CCAN–Cenp-A\textsuperscript{Nuc} complexes. a, Size-exclusion chromatogram profiles (Agilent Bio SEC-5 column) for (i) CCAN, (ii) CCAN–Cenp-A nucleosome (with Widom 601) complex, (iii) Cenp-A nucleosome (with Widom 601), (iv) H3 nucleosome (with Widom 601) and (v) H3\textsuperscript{N}-Cenp-A\textsuperscript{Nuc} (with Widom 601). b, Comparative size-exclusion chromatogram profiles (Agilent Bio SEC-5 column) for CCAN–Cenp-A\textsuperscript{Nuc} with the Cenp-A nucleosome wrapped with either the (i) 147-bp Widom 601 positioning sequence (CCAN–Cenp-A\textsuperscript{Nuc} (Widom 601) as in a) or (ii) a 153-bp *S. cerevisiae* centromeric Cen3 sequence (CCAN–Cenp-A\textsuperscript{Nuc} (Cen3)). Both complexes eluted at the same volume. CCAN and the H3 nucleosome do not form a complex (iii). c, Coomassie-blue-stained SDS–PAGE of the 14-subunit CCAN complex. d, Coomassie-blue-stained SDS–PAGE gel of Cenp-A\textsuperscript{Nuc} (Widom 601). Lane E32, ethidium bromide-stained gel of fraction 32. e, CCAN–Cenp-A\textsuperscript{Nuc} (Widom 601) complex. Lane E13, ethidium bromide-stained gel of fraction 13. Size-exclusion chromatograms are shown in a, f, SDS–PAGE gel of CCAN and H3 nucleosome (Widom 601) SEC run shown in b, g–j, Coomassie-blue-stained SDS–PAGE gels of various Cenp-H, I and K segments co-expressed with Cenp-TW and purified with a double Strep tag on the tagged Cenp-I subunit (*). j, The HFDs of Cenp-TW (Cenp-T\textsuperscript{HFDW}) interact with the Cenp-HIK\textsuperscript{Head}. These results confirm the assignments of the Cenp-H, K and I subunits in our cryo-EM maps. k, Schematic of the organization of CCAN–Cenp-A\textsuperscript{Nuc} subunits and subcomplexes and connections to the outer kinetochore MIs12 and Ndc80 complexes. Lines indicate subcomplex connections. The two pathways connecting Cenp-A\textsuperscript{Nuc} to the Ndc80 complex and microtubules are indicated as P1 and P2 (thick lines to Ndc80). Subunits of the essential P1 pathway are labelled black and indicated with blue shading, whereas subunits of the non-essential P2 pathway are labelled white and indicated with yellow shading. The P2 pathway becomes essential when the P1 pathway is defective through defects in Dsn1 phosphorylation\textsuperscript{9}. The experiments shown in a–j were performed independently in triplicate with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Cryo-EM data of the *S. cerevisiae* CCAN–Cenp-A Nuc complex. a, A typical cryo-electron micrograph of CCAN–Cenp-A Nuc, representative of 9,002 micrographs. b, Galleries of 2D classes of CCAN, representative of 100 2D classes. c, Galleries of 2D classes of CCAN–Cenp-A Nuc, representative of 150 2D classes. The 2D class averages for the C2-symmetric CCAN–Cenp-A Nuc complex viewed in the plane of the C2-symmetry axis are outlined in red. Only a few views were observed, precluding a 3D reconstruction. Cryo-EM grids partially destabilize CCAN–Cenp-A Nuc interactions, resulting in a very low abundance of (CCAN)2–Cenp-A Nuc particles (about 0.03% of total). The two-fold symmetry axes of the (CCAN)2–Cenp-A Nuc complex are shown as dashed arrows. Experiments for data in b and c were performed independently 12 times with similar results. d, FSC curves shown for the cryo-EM reconstructions of CCAN–Cenp-A Nuc complexes: apo-CCAN, mask1 (Cenp-OPQU+ Cenp-LN), mask2 (Cenp-HIK, Cenp-LN, sub-Cenp-OP), CCAN–Cenp-A Nuc. Mask1 and mask2 used for MBR are defined in h and i and Methods. e, Angular distribution plot of CCAN–Cenp-A Nuc particles. f, Local resolution map of CCAN. g, Local resolution map of CCAN–Cenp-A Nuc. h, Local resolution map of mask1 (Cenp-OPQU+, Cenp-LN). i, Local resolution map of mask2 (Cenp-HIK, Cenp-LN, sub-Cenp-OP).
Extended Data Fig. 3 | Workflow of 3D classification of the CCAN–Cenp-A
Nuc cryo-EM dataset. **a**, After initial 2D classification, about 1.4
million particles were sorted by 3D classification into apo-CCAN (52%) and
the CCAN–Cenp-A Nuc complex (48%). For apo-CCAN, 4% existed
as dimers (black box) and 19% showed an ordered head-group (Cenp-
HIKHead) for the Cenp-HIK–TW subcomplex (blue box). A mask was
applied to the CCAN–Cenp-A Nuc cryo-EM map to exclude the structurally
variable Cenp-HIKHead domain for reconstruction of the 4.15 Å structure.
**b**, Details of the four masks used for MBR. **c**, A small 3D class of CCAN–
Cenp-A Nuc, revealing density attached to Cenp-HIKHead contacting the
DNA gyre of Cenp-A Nuc was assigned as Cenp-T HFDW.
Extended Data Fig. 4 | Cryo-EM density maps of apo-CCAN. a, Portion of cryo-EM map for the coiled coils of Cenp-H and Cenp-K. A selection of highly conserved intersubunit residues defined in b and c are labelled. These residues are well defined in the cryo-EM density, consistent with the structure. b, c, Multiple sequence alignment of the coiled-coil regions of Cenp-H (b) and Cenp-K (c). d–f, Portions of cryo-EM maps for Cenp-LN (d), Cenp-I (e) and Nkp1–Nkp2 (f). The chain assignments and polarity of Cenp-H, Cenp-I and Cenp-K of our structure agree with the cryo-EM structure of yeast Ctf3 (PDB: 6OUA)61.
Extended Data Fig. 5 | Cryo-EM densities of CCAN and CCAN–Cenp-A Nuc complexes. a, Cryo-EM reconstruction of CCAN–Cenp-A Nuc from uncrosslinked sample at 8.6 Å resolution. b, Cryo-EM map of dimeric CCAN (also Extended Data Fig. 3a, black box). Subunits are colour-coded as in Fig. 1. The 3.5 Å monomeric free CCAN coordinates were rigid-body-docked into the cryo-EM map. c, Cartoon representation of the *S. cerevisiae* MIND complex (right), showing a notable similarity to the coiled coils of Cenp-QU–Nkp1–Nkp2 of CENP-OPQU+ (left). d, View of the 4.7 Å resolution cryo-EM map of free Cenp-HIK with fitted coordinates from CCAN. e, In the context of CCAN, Cenp-HIKHead rotates to accommodate Cenp-A Nuc. The two conformations of Cenp-HIK from the apo-CCAN and CCAN–Cenp-A Nuc complexes were superimposed onto their rigid portion of Cenp-HIK (C-terminal region of Cenp-I is shown for apo-CCAN) to indicate the conformational variability of Cenp-HIKHead between the two states. Subunits of Cenp-HIKHead of CCAN–Cenp-A Nuc are coloured lighter. f, Cryo-EM density of Cenp-A Nuc showing the Cenp-C–Cenp-A binding motif.
Extended Data Fig. 6 | XL-MS analysis of the CCAN and CCAN–Cenp-A<sup>Nuc</sup> complexes. a, b, Circular plots displaying all the identified crosslinks for CCAN (a) and CCAN–Cenp-A<sup>Nuc</sup> (b). Inter- and intra-subunit crosslinks are indicated in red and blue, respectively. c, d, Histogram plots showing the C<sub>α</sub>–C<sub>α</sub> distance distribution of the crosslinks that could be mapped onto the CCAN (c) and CCAN–Cenp-A<sup>Nuc</sup> structures (d). Ninety-five per cent of the mapped crosslinks satisfy the crosslinker-imposed distance restraint of 30 Å indicated with a dashed red line. e, f, Crosslinks mapped onto the CCAN (e) and CCAN–Cenp-A<sup>Nuc</sup> complex (f). Inter and intra-subunit crosslinks are indicated in red and blue, respectively. Crosslinks exceeding the crosslinker-imposed distance restraint of 30 Å are indicated in yellow. g, Residues on CCAN shown by XL-MS that crosslink with Cenp-C are indicated on the CCAN structure. Red spheres, crosslinks in the CCAN–Cenp-A<sup>Nuc</sup> complex; yellow spheres, additional crosslinks unique to apo-CCAN. The experiments shown in a and b were performed independently in triplicate with similar results.
Extended Data Fig. 7 | The *S. cerevisiae* Cenp-A\(\text{Nuc}\) nucleosome is unwrapped. **a–c**, The positively charged electrostatic potential of the DNA-binding groove of Cenp-LN subcomplex is conserved in *S. cerevisiae*, *S. pombe* and *H. sapiens*. *S. pombe* and *H. sapiens* are represented by modelled structures. **d**, Cenp-N interacts with *S. cerevisiae* Cenp-A\(\text{Nuc}\) in the context of CCAN differently from the interaction of free human Cenp-N with Cenp-A\(\text{Nuc}\). The Cenp-N subunit of the human Cenp-N–Cenp-A\(\text{Nuc}\) structure (PDB: 6C0W\(^{29}\)) was superimposed onto Cenp-N of the *S. cerevisiae* CCAN–Cenp-A\(\text{Nuc}\) structure. In this mode of Cenp-N–Cenp-A\(\text{Nuc}\) interactions, Cenp-A\(\text{Nuc}\) would clash with Cenp-OPQU\(^{+}\) and Cenp-N of CCAN. **e, f**, Structures of *S. cerevisiae* H3\(\text{Nuc}\) (PDB: 1ID3\(^{24}\)) (e) and Cenp-A\(\text{Nuc}\) (f, this work). **g**, Sequence alignment of the N-terminal regions of *S. cerevisiae* H3 and Cenp-A (Cse4) histones. For the chimeric H3\(\text{Nuc}–\)Cenp-A\(\text{Nuc}\), residues 1–50 of *S. cerevisiae* H3 were substituted for residues 1–140 of *S. cerevisiae* Cenp-A. A similar approach was used for vertebrate Cenp-A\(\text{Nuc}\) (ref. \(^{23}\)).
Extended Data Fig. 8 | SDS–PAGE of CCANΔCenp-C–Cenp-A\textsuperscript{Nuc} complexes. Corresponding size-exclusion chromatograms are shown in Fig. 4b and Extended Data Fig. 9a. 
a, b, Mutating the Cenp-N DNA-binding groove did not impair CCAN\textsuperscript{ΔCenp-C} assembly. 
c, Wild-type CCAN\textsuperscript{ΔCenp-C} forms a complex with Cenp-A\textsuperscript{Nuc}. 
d, Mutating the Cenp-N DNA-binding groove disrupts CCAN\textsuperscript{ΔCenp-C}–Cenp-A\textsuperscript{Nuc} interactions. 
e, Mutating the L1 loop of Cenp-A did not destabilize CCAN\textsuperscript{ΔCenp-C}–Cenp-A\textsuperscript{Nuc} interactions. 
f, Deletion of the N terminus of Cenp-A (1–129) (ΔN\textsuperscript{Cenp-A\textsuperscript{Nuc}}) did not impair CCAN\textsuperscript{ΔCenp-C}–Cenp-A\textsuperscript{Nuc} interactions. 
g, Both CCAN\textsuperscript{ΔCenp-C} and CCAN\textsuperscript{ΔCenp-C}–Cenp-N\textsuperscript{Mut} bound poorly to H3\textsuperscript{N}–Cenp-A\textsuperscript{Nuc}. The experiments shown were performed independently in triplicate with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Testing of CCANΔCenp-C binding to Cenp-A Nuc.

a, Comparative SEC profiles (Agilent Bio SEC-5 column) for wild-type CCANΔCenp-C and the Cenp-N Mut of CCANΔCenp-C to Cenp-A Nuc and its modifications (Cenp-A Nuc-L1Nuc, ΔN-Cenp-A Nuc and H3N-Cenp-A Nuc) and H3Nuc. Mutating the L1 loop (Cenp-A L1Nuc) of Cenp-A or deletion of the N-terminal 129 residues (ΔNCenp-A Nuc) did not destabilize CCANΔCenp-C–Cenp-A Nuc interactions. By contrast, CCAN with the Cenp-N Mut bound less well and both CCAN and CCAN–Cenp-N Mut hardly bound to H3N–Cenp-A Nuc (CCANΔC, CCANΔCenp-C). Associated SDS–PAGE is shown in Extended Data Figs. 8, 9b. b, Coomassie-blue-stained SDS–PAGE showed that CCANΔCenp-C did not associate with H3Nuc. c, Micrococcal nuclease digestion of Cenp-A Nuc, H3Nuc and H3N–Cenp-A Nuc. Widom 601 DNA is shown as a control. The H3Nuc and H3N–Cenp-A Nuc protect a similar and longer length of DNA compared with Cenp-A Nuc. d, Model of CBF3 bound to CCAN–Cenp-A Nuc, indicating that CBF3 would not associate with a fully assembled kinetochore, consistent with proteomic data. The experiments shown in a–c were performed independently in triplicate with similar results. For gel source data, see Supplementary Fig. 1.
Cross-linked *S. cerevisiae* CCAN–Cenp-A^{Nuc}

Uncross-linked *S. cerevisiae* CCAN–Cenp-A^{Nuc}

*S. cerevisiae* CCAN alone

Extended Data Fig. 10 | See next page for caption.
Extended Data Fig. 10 | *S. cerevisiae* CCAN–Cenp-A Nuc comprises two CCAN complexes in solution. a–c, The predicted mass of (CCAN)$_2$–Cenp-A Nuc is 1.31 MDa, (CCAN)$_1$–Cenp-A Nuc is 0.77 MDa and that of a CCAN dimer 1.09 MDa (Extended Data Table 2). Representative SEC–MALS data for crosslinked *S. cerevisiae* CCAN–Cenp-A Nuc complex (a), run independently in triplicate with similar results, average molecular mass is 1.23 MDa ([CCAN]$_2$–Cenp-A Nuc); uncrosslinked *S. cerevisiae* CCAN–Cenp-A Nuc complex (b), run independently in triplicate with similar results, with average masses of 1.38 MDa ([CCAN]$_2$–Cenp-A Nuc) and 526 kDa (CCAN); and *S. cerevisiae* CCAN alone (c), run independently in duplicate with similar results, with average masses of 839 kDa for the leading edge (green) and 650 kDa for the trailing edge (magenta), suggesting a non-resolved monomer–dimer equilibrium. 
d, e, Velocity analytical ultracentrifugation of crosslinked (d) and uncrosslinked (e) *S. cerevisiae* CCAN–Cenp-A Nuc complexes with residuals to the fits shown in f and g, f, g. Fit of a $c(s)$ distribution model for the crosslinked complex (f), the major species sediments at 15.8S ($S_{w,20} = 26.1S$) with a minor species at 12.1S ($S_{w,20} = 20.0S$) that corresponds to calculated masses of 1.34 MDa ([CCAN]$_2$–Cenp-A Nuc) and 896 kDa (possibly [CCAN]$_1$–Cenp-A Nuc), respectively, with a fitted value of 1.761 for the frictional ratio. g, Fit for uncrosslinked samples, the major species is resolved into two species that sediment at 14.3S ($S_{w,20} = 22.6S$) and 15.7S ($S_{w,20} = 24.9S$) with a minor species at 12.3S ($S_{w,20} = 19.4S$), which gave masses of 1.32 MDa ([CCAN]$_2$–Cenp-A Nuc) and 1.15 MDa ([CCAN]$_2$) for the major species and 716 kDa ([CCAN]$_1$–Cenp-A Nuc) for the minor species. The experiments shown in d–g were performed independently in triplicate with similar results. h, Examples of two 2D class averages showing the (CCAN)$_2$–Cenp-A Nuc particles viewed in the plane of the C2 symmetry axis (red outline) (data from Extended Data Fig. 2c) and the 2D reprojections of a modelled (CCAN)$_2$–Cenp-A Nuc based on the CCAN–Cenp-A Nuc cryo-EM reconstruction (yellow outline) (shown in i). There is a close correspondence in shape and dimensions between the calculated reprojections and the observed 2D classes. The two-fold symmetry axes of the (CCAN)$_2$–Cenp-A Nuc complex are shown as dashed arrows. i, j, Two alternative models for how CCAN assembled onto a Cenp-A nucleosome would interact with the outer kinetochore–microtubule interface (Supplementary Video 2). i. In scenario (1), CCAN interacts with the outer kinetochrome from the same side as the DNA-binding surface. Microtubules attached to the outer kinetochrome would hoist CCAN from below the over-lying nucleosome and out-stretched DNA. j. In scenario (2), the microtubule-outer kinetochrome interface contacts CCAN from the opposite side to the CCAN DNA-binding surface. Outer-kinetochrome (outer-KT): KMN network and microtubule-attachment complexes, Dam1–DASH (budding yeast) and Ska proteins of vertebrates. The combined dimension of (CCAN)$_2$–Cenp-A Nuc (32 nm) matches that of the hub at the centre of the yeast kinetochrome$^{43}$. 


### Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

|                        | CCAN (EMDB-4580) (PDB 6QLE) | CCAN–Cenp-A<sub>Nuc</sub> (EMDB-4579) (PDB 6QLD) | Mask1 (EMDB-4581) (PDB 6QLF) | Mask2 (EMDB-4971) |
|------------------------|-----------------------------|-------------------------------------------------|-----------------------------|-----------------|
| **Data collection and processing** |                             |                                                 |                             |                 |
| Magnification          | 75,000                      | 75,000                                          | 75,000                      | 75,000          |
| Voltage (kV)           | 300                         | 300                                             | 300                         | 300             |
| Electron exposure (e-/Å²) | 32                          | 32                                              | 32                          | 32              |
| Defocus range (µm)     | 2.0-2.8                     | 2.0-2.8                                         | 2.0-2.8                     | 2.0-2.8         |
| Pixel size (Å)         | 1.09                        | 1.09                                            | 1.09                        | 1.09            |
| Symmetry imposed       | C1                          | C1                                              | C1                          | C1              |
| Initial particle images (no.) | 1,796,016                | 1,796,016                                       | 1,796,016                   | 1,796,016       |
| Final particle images (no.) | 618,459                   | 193,882                                         | 618,459                     | 618,459         |
| Map resolution (Å)     | 3.55                        | 4.15                                            | 3.45                        | 3.83            |
| FSC threshold          | 0.143                       | 0.143                                           | 0.143                       | 0.143           |
| Map resolution range (Å) | 3.0-5.5                    | 3.5-7.0                                         | 3.0-5.5                     | 3.0-5.5         |
| **Refinement**         |                             |                                                 |                             |                 |
| Initial model used (PDB code) | 5MU3, 6EQ7, 4JE3, 5W94 | 3AN2, 4X23, 5MU3, 6EQ7, 4JE3, 5W94 | 5MU3, 6EQ7, 4JE3, 5W94 | 5MU3, 6EQ7, 4JE3, 5W94 |
| Model resolution (Å)   | 3.5                         | 4.0                                             | 3.3                         | -               |
| 0.143 FSC threshold    |                             |                                                 |                             |                 |
| Model resolution range (Å) | 50 - 3.0                   | 50 - 3.6                                        | 50 - 3.0                    | -               |
| Map sharpening B factor (Å<sup>2</sup>) | -139                       | -108                                            | -135                        | -172            |
| Model composition      |                             |                                                 |                             |                 |
| Non-hydrogen atoms     | 18,058                      | 29,183                                          | 13,541                      | -               |
| Protein residues       | 2,401                       | 3,172                                           | 1,790                       | -               |
| Ligands                | 0                           | 248                                             | 0                           | -               |
| B factors (Å<sup>2</sup>) |                             |                                                 |                             |                 |
| Protein                | 78.6                        | 82.2                                            | 67.2                        | -               |
| Ligand                 | -                           | 245.8                                           | -                           | -               |
| R.m.s. deviations      |                             |                                                 |                             |                 |
| Bond lengths (Å)       | 0.004                       | 0.004                                           | 0.005                       | -               |
| Bond angles (°)        | 0.798                       | 0.793                                           | 0.828                       | -               |
| Validation             |                             |                                                 |                             |                 |
| MolProbity score       | 1.39                        | 1.57                                            | 1.45                        | -               |
| Clashscore             | 2.78                        | 4.80                                            | 2.99                        | -               |
| Poor rotamers (%)      | 0.11                        | 0.08                                            | 0.19                        | -               |
| Ramachandran plot      |                             |                                                 |                             |                 |
| Favored (%)            | 95.30                       | 94.78                                           | 94.76                       | -               |
| Allowed (%)            | 4.60                        | 5.02                                            | 5.04                        | -               |
| Disallowed (%)         | 0.10                        | 0.20                                            | 0.20                        | -               |
### Extended Data Table 2 | Table of CCAN subunits

| Subunit         | S.c. name | Length | Mol. Mass kDa | Domain/Region 1 Details | Domain/Region 2 | Domain/Region 3 | Disordered regions | Sequence built as polyA |
|-----------------|-----------|--------|---------------|--------------------------|-----------------|-----------------|-------------------|----------------------|
| Cenp-A nucleosome | Cse4      | 229    | 26.8          | α-helix and disordered 1-131 | Histone fold 132-229 PDB 3AN2 Hs Cenp-A | -               | 1-111,131-136,227-229 | 112-130               |
| H2A             |           | 132    | 14.0          | Histone fold PDB 1ID3 Sc H2A | -               | -               | -                | -                    |
| H2B             |           | 132    | 14.2          | Histone fold PDB 1ID3 Sc H2A | -               | -               | -                | -                    |
| H4              |           | 103    | 11.4          | Histone fold PDB 1ID3 Sc H2A | -               | -               | -                | -                    |
| 601 DNA         |           | 147 bp | 90.6          |                          |                 |                 |                  |                      |
| Cenp-C          | Mlf2      | 549    | 62.5          | Cenp-C motif 283-304 PDB 4X23 | Cupin fold 365-530 | -               | 1-283,306-549     | -                    |
| Cenp-H HKI–TW complex (Sctf3 complex + Cenp-TW) | | | | | | | | |
| Cenp-H          | Mcm16     | 181    | 21.1          | α-helix: De novo 4-136  | α-helix: PDB 5Z07 Cenp-I 143-181 | -               | 1-3,41-44,75-78,137-142 | -                    |
| Cenp-I          | Ctf3      | 733    | 84.3          | Heat repeats PDB 5Z07 Cenp-I 5-241 | Heat repeats: De novo | -               | 242-332,526-531,597-601,620-624,657-663,677-689 | 321-330,664-676 |
| Cenp-K          | Mcm22     | 239    | 27.6          | α-helix: De novo 7-128  | α-helix: PDB 5Z07 Cenp-I 143-236 | -               | 1-6,42-49,61-68,129-142 | -                    |
| Cenp-T          | Cnn1      | 361    | 41.3          | Histone fold           |                 | ND              | ND                | ND                   |
| Cenp-W          | Wip1      | 98     | 10.2          | Histone fold           |                 | ND              | ND                | ND                   |
| Cenp-LN complex | | | | | | | | |
| Cenp-L          | lm3       | 245    | 28.0          | α/β fold   | PDB 4JE3 Sc Cenp-L | -               | -                | -                    |
| Cenp-N          | Chl4      | 458    | 52.7          | α/β fold PDB 4JE3 Sc Cenp-L | Pyrin (1-102) Cenp-N fold (103-262) PDB 6EQT Hs Cenp-N | Cenp-N linker domain de novo (262-373) Dimerization (375-468) PDB 4JE3 Sc Cenp-N | - | - |
| Cenp-OPQ+ complex (ScCOMA+ complex) | | | | | | | | |
| Cenp-O          | Mcm21     | 368    | 43.0          | RWD PDB 5MU3 KI Ctf19 | -               | -               | 1-152,332-338     | -                    |
| Cenp-P          | Ctf19     | 369    | 42.8          | RWD PDB 5MU3 KI Ctf19 | -               | -               | 1-96,111-123,286-292,308-313 | 97-110 |
| Cenp-Q          | Okp1      | 406    | 47.4          | α-helix: De novo     | -               | -               | 1-160,220-228,304-319,392-406 | 161-219 |
| Cenp-U          | Ame1      | 324    | 37.5          | α-helix: De novo     | -               | -               | 1-130,157-165,267-276 | 131-156 |
| Nkp1            | Nkp1      | 238    | 27.0          | α-helix: De novo     | -               | -               | 1-124,135         | 24-32,217-238        |
| Nkp2            | Nkp2      | 153    | 17.9          | α-helix: De novo     | -               | -               | 1-2,25-35         | 133-153 |

Details of structured regions of CCAN subunits built into the cryo-EM density maps are indicated, including regions built as polyAla. The calculated molecular masses for CCAN and Cenp-A<sup>core</sup> complexes are (i) CCAN: 543.3 kDa, (ii) CCAN dimer: 1.09 MDa, (iii) Cenp-A<sup>core</sup>: 223 kDa, (iv) (CCAN)<sub>1</sub>–Cenp-A<sup>core</sup>: 0.766 MDa and (v) (CCAN)<sub>1</sub>–Cenp-A<sup>core</sup>: 1.31 MDa.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  Confirmed

☑️  The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☑️  A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☑️  The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☑️  A description of all covariates tested

☑️  A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☑️  A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☑️  For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

☑️  For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☑️  For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☑️  Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code.

Data collection

Commercial software: EPU from Thermo Fisher Scientific was used for automated cryo-EM data collection.

Data analysis

Cryo-EM data were analyzed using the software MotionCor2 (version 2.1), GCTF (version 0.5), RELION2.1 (version 2.1), RELION3.0 (version 3.0), SIMPLEPRIME (version 1) and RESMAP (version 1.1.4). Model building and refinement were performed using COOT (version 0.8.9.2) and Phenix (version 1.15.2) and validated in COOT (version 0.8.9.2) and MolProbity (version 4.2). Visualization was performed with COOT (version 0.8.9.2), PyMOL (version 1.8.4.1), Chimera (version 1.8.1) and ChimeraX (version 0.8). Structure figures were generated use PyMOL (v1.8.4.1) and Chimera (version 1.8.1).

Sequence alignments were performed and displayed with JALVIEW (version 1.0).

Structure prediction was performed with the PHYRE2 web tool. Protein secondary structure and disordered regions were predicted with the PHYRE2 and PSIPred web tools.

AUC data were analysed in SEDFIT v16.1 and Sednterp (version 20130813 beta). SEC-MALS data were analysed using ASTRA version 5.3.4.20 software (Wyatt Technologies) and data were plotted with the program PRISM (version 8.2.0) (GraphPad Software Inc.).

XL-MS data were analysed using Proteome Discoverer 2.3 (version 2.3.0.522) and the wXs web tool.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

EM maps are deposited with EMDB with accession codes EMD-4580 (CCAN), EMD-4579 (CCAN-Cenp-ANuc), EMD-4581 (Mask1) and EMD-4971 (Mask2). Protein coordinates are deposited with RCSB with accession codes 6QLE (CCAN), 6QLD (CCAN-Cenp-ANuc) and 6QLF (Mask1). The cross-linking mass spectrometry raw files, the associated output and databases are deposited through the ProteomeXchange Consortium 48 via the PRIDE partner repository with the dataset identifier PXD013769.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- □ Life sciences
- □ Behavioural & social sciences
- □ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reportsmg-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size We collected 9002 cryo-EM images for the CCAN-Cenp-A dataset and 910 cryo-EM images for the Cenp-HIK dataset. The total number of particles for the CCAN-Cenp-A dataset was 1,796,061 and that for the Cenp-HIK dataset was 123,215. Sample sizes were estimated on the basis of previous studies using similar methods and analyses that are widely published.

Data exclusions No data were excluded from the analysis.

Replication All attempts at replication were successful and reproducible. At least three independent biological repeats per experiment where representative data are shown. Structure determination does not require replication.

Randomization Samples were not allocated into groups. Randomization is not relevant to this study.

Blinding Blinding was not relevant to this study because there was no group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a Involved in the study       | n/a Involved in the study |
| □ □ Antibodies                  | □ ChIP-seq |
| □ □ Eukaryotic cell lines       | □ Flow cytometry |
| □ □ Palaeontology               | □ MRI-based neuroimaging |
| □ □ Animals and other organisms |         |
| □ □ Human research participants |         |
| □ □ Clinical data               |         |

Antibodies

Antibodies used One primary antibody used: MOUSE ANTI STREP-TAG CLASSIC:HRP. Anti-Strep antibody (Source: Bio-Rad, Catalogue code:MCA2489P, Batch No: 147517). Dilution 1 to 1000.

Validation Mouse anti Strep-Tag Classic antibody, clone Strep-tag II, also known as StrepMAB-Classic, recognizes Strep-tag II, a widely used tag in protein expression applications. This antibody recognizes both C- and N-terminal Strep-tag II and is especially suited to Western blot applications.

Validation: HCA182 specificity ELISA using various antigens (A: Human Serum, B: human IgG1/kappa from myeloma plasma, C: Rituximab, D: Ustekinumab, E: Infliximab, F: Adalimumab, G: Alemtuzumab and H: Bevacizumab) as coating components.
followed by Human anti Avastin®(HCA182) and HRP conjugated Mouse anti Strep-tag (MCA2489P) as detection reagent.

Literature on Bio-Rad web-site:

1. Renzi, F. et al. (2015) Glycan-Foraging Systems Reveal the Adaptation of Capnocytophaga canimorsus to the Dog Mouth. MBio. 6 (2): pii: e02507-14.
2. Gordon, C.A. et al. (2015) NUSAP1 expression is upregulated by loss of RB1 in prostate cancer cells. Prostate. 75 (5): 517-26.
3. Mavrakis, M. et al. (2016) Purification of recombinant human and Drosophila septin hexamers for TIRF assays of actin-septin filament assembly. Methods Cell Biol. 136: 199-220.
4. Oda, S. et al. (2015) Crystal Structure of Marburg Virus VP40 Reveals a Broad, Basic Patch for Matrix Assembly and a Requirement of the N-Terminal Domain for Immunosuppression. J Virol. 90 (4): 1839-48.
5. Renzi, F. et al. (2015) Glycan-foraging systems reveal the adaptation of Capnocytophaga canimorsus to the dog mouth. MBio. 6 (2): e02507.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

High-5 insect cell: Trichoplusia ni: expression system.

S. cerevisiae strains:
1. AEY4992: with a chl4 deletion and cse4-R37A mutation (chl4Δ cse4-R37A), (MATα ade2-101 lys2 his3-11,15 trp1-1 leu2-3,112 ura3-1 can1-100 chl4Δ::kanMX Cse4R37A).
2. W303 (wild type strain: MATα ade2-101 his3-11,15 trp1-1 leu2-3,112 ura3-1).

Authentication

The High-5 insect cell line was not authenticated. The S. cerevisiae strains were authenticated refs 27, 41.

Mycoplasma contamination

The High-5 insect cell line was not tested for mycoplasma contamination. Yeast strains do not have mycoplasma and were not tested for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

None