Architecture of the CBF3-centromere complex of
the budding yeast kinetochore

Kaige Yan1,2, Ziguo Zhang1,2, Jing Yang1, Stephen H. McLaughlin1,1 and David Barford1,1*

Kinetochore are multicomponent complexes responsible for coordinating the attachment of centromeric DNA to mitotic-spindle microtubules and are associated with the centromere-specific nucleosome Cse4. Deposition of Cse4 at CEN loci is dependent on the CBF3 complex that engages CDEIII to direct Cse4 nucleosomes to CDEII. To understand how CBF3 recognizes CDEIII and positions Cse4, we determined a cryo-EM structure of a CBF3–CEN complex. CBF3 interacts with CEN DNA as a head-to-head dimer that includes the whole of CDEIII and immediate 3' regions. Specific CEN-binding of CBF3 is mediated by a Cep3 subunit of one of the CBF3 protomers that forms major groove interactions with the conserved and essential CCG and TGT motifs of CDEIII. We propose a model for a CBF3–Cse4–CEN complex with implications for understanding CBF3-directed deposition of the Cse4 nucleosome at CEN loci.

Centromeres mediate the attachment of chromosomes to the mitotic spindle by coordinating the assembly of kinetochores, large proteinaceous complexes that bind microtubules. In mitosis, kinetochores couple chromosome motion to the energy released by microtubule depolymerization, which powers chromosome segregation. Incorrectly attached kinetochores trigger a mitotic checkpoint pathway that inhibits the anaphase-promoting complex to prevent chromosome mis-segregation. The regional centromeres of higher eukaryotes comprise megabases of chromosomal DNA and engage multiple microtubules. In contrast, the point centromeres of budding yeast chromosomes are defined by conserved ~125 nucleotide segments that are sufficient to direct CENP-A nucleosome deposition centered onto the adjacent CDEII, creating a structure with pseudo-3' lobe (Fig. 1). The short CDEI and CDEIII segments are highly conserved across all 16 Saccharomyces cerevisiae chromosomes, whereas CDEII elements, although lacking sequence similarity, share an AT-rich region of 80–90 nucleotides. Disruption of the invariant CCG motif of CDEIII inactivates the centromere.

Centromeres of all species are marked by centromere-specific nucleosomes in which the canonical histone H3 is replaced by a specific histone (CENP-A), CENP-A in animals and Cse4 in S. cerevisiae, CENP-A nucleosomes coordinate kinetochore assembly by recruiting inner kinetochore scaffolding subunits such as CENP-C (Mif2) and CENP-N (Chl4). CENP-A nucleosome deposition at human centromeres is epigenetically specified by preexisting CENP-A nucleosomes. In S. cerevisiae, the multiprotein CBF3 complex assembles on CDEIII of CEN loci to direct Cse4 nucleosome deposition centered onto the adjacent CDEII, creating a nucleosome-resistant region of ~200 base pairs (bp) that is essential for yeast viability with Cep3, Ctf13 and Skp1 forming CBF3core. The Cep3 homodimer incorporates two DNA-binding Gal4-like Zn2Cys6 clusters, one of which provides sequence specificity by engaging the essential CCG motif of CDEIII. The Cep3 dimer, Ctf13, and Skp1 without DNA. However, other particles had larger sizes, with one notable class featuring an extended structure with two prominent lobes joined by a smaller central lobe (Fig. 2). We also observed monomeric CBF3 complexes and used focused 3D refinement in RELION, segmented rigid domains of CBF3–CEN3 were determined at higher resolution (Fig. 1 and Supplementary Fig. 2d–g,i). We also observed monomeric CBF3–CEN3 complexes (Supplementary Fig. 2d and Supplementary Table 1). To obtain a higher-resolution reconstruction of CBF3 for atomic-model building, we masked the EM density corresponding to a single CBF3m complex with 4.4 Å resolution. Using multibody refinement in RELION, segmented rigid domains of CBF3–CEN3 were determined at higher resolution. The resultant CBF3m reconstruction was determined to 3 Å, sufficient for de novo modeling with confidence.

Results

Overall architecture of CBF3. To generate a CBF3–CEN complex, all four CBF3 proteins were coexpressed, and the resulting complex was purified and then assembled onto the 147-bp CEN3 DNA duplex (Supplementary Fig. 1a). Cryo-electron micrographs of the CBF3–CEN3 complex revealed a heterogeneous mixture of particles (Supplementary Fig. 2a–d, Table 1 and Supplementary Table 1). Approximately one-third of particles consisted of CBF3–CEN3 complexes, bridged by a central Ndc10 dimer, are bound to a single CEN3 DNA duplex. This creates a structure with pseudo dyad symmetry. The entire CBF3–CEN3 complex was reconstructed to an overall resolution of 4.4 Å. Using multibody refinement in RELION, the resultant CBF3m reconstruction was determined to 3 Å, sufficient for de novo modeling with confidence (Table 1 and Supplementary Fig. 2d,h,i).

In a related complex (CBF3m–Ndc10–CEN3), we replaced full-length Ndc10 with Ndc10DBD (residues 1–540) that includes only its DNA-binding domain (DBD) and lacks the C-terminal...
residues (Ndc10C) shown to mediate Ndc10 dimerization \(^3\)\(^{1,2}\) (Supplementary Fig. 1b). EM analysis of CBF3\(^{core}\)–Ndc10\(^{DBD}\)–CEN3 showed mainly a mixture of DNA-free CBF3\(^{core}\) and CBF3\(^{core}\)–Ndc10\(^{DBD}\) particles suggesting that Ndc10C enhances CBF3–CEN3 DNA binding (Supplementary Fig. 3a). We determined 3D reconstructions of DNA-free CBF3\(^{core}\) and CBF3\(^{core}\)–Ndc10\(^{DBD}\) at 3.9 Å and 3.6 Å, respectively (Supplementary Fig. 3, Table 1 and Supplementary Table 1). The high quality of the 3 Å EM density map of CBF3\(^{msk}\) derived from the CBF3–CEN3 cryo-EM data allowed atomic modeling and refinement of its complete structure (Supplementary Fig. 2h,j,k and Table 1). This was then used to interpret the 3.6 Å map of CBF3\(^{core}\)–Ndc10\(^{DBD}\) (without DNA) (Supplementary Fig. 3f and Table 1) and the CEN3-bound CBF3 dimer (Fig. 1, Supplementary Fig. 3e–g and Table 1). CBF3\(^{core}\) forms a rigid assembly through extensive contacts involving its component subunits: a Cep3 homodimer and Ctf13–Skp1 heterodimer (Fig. 1b,d). Our CBF3\(^{core}\) structure resembles that of the recently published CBF3\(^{core}\) (ref. \(^1\)), although at higher resolution, allowing the building of Ctf13. The Cep3B subunit of the Cep3 homodimer contacts both subunits of the Ctf13–Skp1 heterodimer with its N-terminal Zn\(_2\)Cys\(_6\) (ZnF) cluster becoming ordered by a composite interface formed of Ctf13 (including the F-box) and Skp1 (Fig. 2a,b). Skp1 interacts with Cep3B and cullin subunits through the same interface, explaining how Cep3 protects Ctf13 from ubiquitinit-dependent degradation\(^14\). The Zn\(_2\)Cys\(_6\) cluster of the Cep3A subunit of CBF3\(^{core}\)–Ndc10\(^{DBD}\) (without DNA) is disordered, similar to that in crystal structures of free Cep3\(^{35,36}\). However, in both Cep3 subunits, Met1 and Phe2 are anchored within a hydrophobic pocket formed by the αA and αV helices. For Cep3A, this would restrict the conformational variability of its Zn\(_2\)Cys\(_6\) cluster. When bound to CEN3, the Cep3A Zn\(_2\)Cys\(_6\) cluster engages the CCG motif of CDEIII\(^2\) (Figs. 1a and 3a,g,i). In contrast to the large interfaces shared between subunits of CBF3\(^{core}\), Ndc10 associates only with Ctf13 and this is through a relatively small contact surface area (1,354 Å\(^2\)) situated opposite the Ctf13–Cep3B interface (Figs. 1a and 2c). Due to this small interface, Ndc10 appears flexible in the DNA-free complex (Supplementary Fig. 4b).

In the 3 Å map of CBF3\(^{msk}\), well-defined EM density for Ctf13 allowed complete tracing of the protein sequence (Table 1). As proposed from biochemical studies\(^2\), Ctf13 is at the center of the complex (Figs. 1a and 2a). The N-terminal F-box of Ctf13 (residues

---

**Fig. 1 | Overall structure of the dimeric CBF3–CEN3 complex.** a, c, Two orthogonal views of the EM density map of the complex. b, d, Corresponding cartoon representations. The TGT and CCG motifs of CDEIII are colored blue in d. e, g, Electrostatic surface representation of the dimeric CBF3–CEN3 complex. The color code is displayed below. f, h, Surface conservation of the CBF3–CEN3 dimer. The conservation score is indicated below. i, Region of CEN3 sequence and 3’ flanking region recognized by the dimeric CBF3–CEN3 complex. Sequences indicated with diamonds and circles are the CCG and TAA motifs recognized by the Zn\(_2\)Cys\(_6\) cluster of CBF3\(^{A}\) and CBF3\(^{B}\), respectively. The TGT motif contacted by the α-MN helix of Cep3A is indicated with arrowheads.
1–28) forms hydrophobic interactions with both Skp1 and Cep3B (Fig. 2a,b). Relative to other Skp1-F-box interactions, Ctf13 forms extensive contacts to Skp1 involving its mainly α-helical N-terminal domain (Ctf13NTD). This domain also contacts the Zn 2Cys6 cluster and Ndc10DBD (Fig. 2a,b). Interactions of the conserved Leu12, Pro13 and Leu24 residues of the F-box with Skp1 are consistent with their biochemically defined function of stabilizing Ctf13−Skp1 interactions28 (Fig. 2b). The C-terminal region of the LRR domain of Ctf13 (Ctf13LRR) contacts both Skp1 and Cep3B, the latter contact augmented by the α-helical insert of Ctf13LRR.

Ndc10 interacts with CBF3core through an N-terminal α-helical hairpin that inserts into a crevice between Ctf13NTD and Ctf13LRR (Fig. 2c). Helix αA of the hairpin, not visible in free Ndc1031,32, becomes ordered through interactions with Ctf13. The C-terminal region of the LRR domain of Ctf13 (Ctf13LRR) contacts both Skp1 and Cep3B, the latter contact augmented by the α-helical insert of Ctf13LRR.

| Table 1 | Cryo-EM data collection, refinement and validation statistics |
|---------------------------------|---------------------------------|---------------------------------|
| **CBF3<sub>core−Ndc10<sup>DBD</sup></sub> (EMD-0095, PDB 6GYP)** | **CBF3−CEN3 (EMD-0096, PDB 6GYS)** | **CBF3<sub>msk</sub> (EMD-0097, PDB 6GYU)** |
| **Data collection and processing** | | |
| Magnification | 132,075 | 132,075 | 132,075 |
| Voltage (kV) | 300 | 300 | 300 |
| Electron exposure (e/Å<sup>2</sup>) | -28 | -28 | -28 |
| Defocus range (μm) | 2.4–3.5 | 2.4–3.5 | 2.4–3.5 |
| Pixel size (Å) | 1.06 | 1.06 | 1.06 |
| Symmetry imposed | C1 | C1 | C1 |
| Initial particle images (no.) | 757,433 | 987,484 | 987,484 |
| Final particle images (no.) | 73,894 | 22,668 | 198,010 |
| Map resolution (Å) | 3.6 | 4.4 | 3.0 |
| FSC threshold | 0.143 | 0.143 | 0.143 |
| Map resolution range (Å) | 3.3–4.5 | 3.7–8.3 | 2.5–3.7 |
| **Refinement** | | |
| Initial model used (PDB code) | 2QUQ,4ACO,1FQV,2HAP | 2QUQ,4ACO,1FQV,2HAP | 2QUQ,1FQV,2HAP |
| Model resolution (Å) | 8.0–3.3 | 8.0–3.7 | 8.0–2.5 |
| FSC threshold | 0.143 | 0.143 | 0.143 |
| Map sharpening B factor (Å<sup>2</sup>) | −120 | −110 | −120 |
| **Model composition** | | |
| Non-hydrogen atoms | 17,811 | 38,882 | 18,157 |
| Protein residues | 2,800 | 5,688 | 2,800 |
| Ligands | 0 | 8 | 2 |
| B factors (Å<sup>2</sup>) | | |
| Protein | 174.4 | 304.1 | 148.9 |
| Ligand | - | 395.4 | 181.3 |
| R.m.s. deviations | 0.006 | 0.007 | 0.005 |
| Bond lengths (Å) | 0.895 | 1.018 | 0.872 |
| Bond angles (°) | 0.693 | 0.811 | 0.780 |
| **Validation** | | |
| MolProbity score | 1.71 | 1.77 | 1.72 |
| Clashscore | 4.86 | 6.07 | 5.00 |
| Poor rotamers (%) | 0.50 | 0.29 | 0.74 |
| Ramachandran plot | | |
| Favored (%) | 92.83 | 93.36 | 93.58 |
| Allowed (%) | 7.08 | 6.48 | 6.28 |
| Disallowed (%) | 0.09 | 0.16 | 0.14 |

**CBF3 interacts with CEN3 DNA as a dimer.** We fitted the refined CBF3<sub>msk</sub> coordinates and a 47-bp DNA duplex into the EM density map of the CBF3−CEN3 dimer (Fig. 1a–d). Two CBF3 protomers (CBF3<sup>a</sup> and CBF3<sup>b</sup>), bound to a gently bent DNA duplex, associate through the central dimerized Ndc10 (Figs. 1a,b and 3a and 3b), respectively. Replacing Ile76 and Tyr79 with either Ala or Arg abolished Ctf13 binding to Ndc10 (Supplementary Fig. 1c), validating the importance of this interaction in stabilizing CBF3. A mitotic stability assay, assessing the efficiency of mini-chromosome segregation, confirmed the importance of this interface in vivo (Supplementary Fig. 5). In yeast strains harboring the Ctf13 mutants, the efficiency of mini-chromosome segregation was severely reduced, equivalent to that caused by disruption of the essential CCG motif of CDEIII. Combining the Ctf13 and CCG mutants did not further impair mini-chromosome segregation, showing that the Ctf13 and CCG mutations disrupt an interdependent function.
Supplementary Video 1). EM density for Ndc10 subunits in CBF3−CEN3 is similar to that in CBF3−Ndc10DBD (without DNA), with no visible density for Ndc10C in the majority of defined 3D classes (Supplementary Fig. 2d). This indicates notable flexibility of Ndc10C. In the low-pass filtered maps, we visualized the fragmented densities of two Ndc10DBD C termini situated in close proximity, suggesting the possibility that two Ndc10 subunits, self-associated through Ndc10DBD on DNA, may also self-associate through their Ndc10C (Fig. 3c and Supplementary Fig. 6a). Ndc10C-dimerization would probably stabilize the DNA-bound complex, an idea supported by the very small proportion of dimeric CBF3−CEN3 complexes in the CBF3core−Ndc10DBD−CEN3 cryo-EM dataset.

At 4.3 Å, the resolution of the CEN3 EM density map did not allow definitive identification of DNA bases (Fig. 1a,c and Supplementary Fig. 2g). Thus, to determine the orientation and register of CEN3 DNA, we prepared CEN3 DNA biotinylated at the CDEIII end for avidin labeling. The labeled DNA maintained affinity for CBF3. A cryo-EM reconstruction of the CBF3−CEN3−avidin complex at 12.5 Å revealed clear EM density for the avidin label only at one DNA end (Fig. 3e). We then fitted a DNA duplex that is consistent with CBF3 CEN-DNA crosslinking data27, and that positions the Cep3A Zn2Cys6 cluster at the CDEIII CCG motif28.

The CBF3 dimer binds to CEN3 DNA in a roughly two-fold symmetric manner, although because the protein complex extends beyond the 3′ end of CEN3 DNA, the DNA duplex is positioned asymmetrically with respect to the CBF3 dimer (Fig. 1b,d). CBF3A contacts CDEII and 6 bp of CDEII. Strikingly, this corresponds to the 5′ end of the 56-bp region protected from DNA cleavage in the CBF3−CEN3 DNA complex17 (Fig. 1i). CBF3B contacts the entire DNA duplex (14 bp) 3′ of CDEIII (Fig. 1i). Only 12 bp of CDEII are visible, with CEN3 DNA 5′ of this segment being disordered. Both CBF3A and CBF3B engage the DNA duplex through a channel created by the Cep3 dimer and Ctf13 (Figs. 1a and 3a). The 3′ end of the CEN3 DNA exiting the CBF3A DNA-binding channel then contacts Ndc10 of CBF3A some 10 bp 3′ of the CCG motif (Fig. 3a–d). Ndc10-DNA interactions in the CBF3−CEN3 complex involve non-specific DNA sugar–phosphate interactions and are identical to those of the Kluyveromyces lactis Ndc10−DNA complex31, and consistent with a mutagenesis study of S. cerevisiae Ndc1032 (Fig. 3b–d). In the CBF3−CEN3 complex, however, a dimer of

Fig. 2 | CBF3 inter-subunit interfaces. a, Interactions of the Cep3B Zn2Cys6 cluster at the Skp1–Ctf13 interface. b, Contacts of the Ctf13 F-box with Skp1 and Cep3B. L12, P13 and L24 are essential for Ctf13-Skp1 interactions. c, The N terminus of Ndc10 contacts Ctf13. Mutating I76 and Y79 of Ctf13 disrupt Ctf13-Ndc10 interactions (Supplementary Fig. 1c).
Ndc10DBD is arranged with dyad symmetry on the DNA duplex. The two Ndc10DBD self-associate through a small interface of 451 Å² surface area that involves the α-helix responsible for mediating Ndc10-DNA interactions. As Ndc10DBD is not known to dimerize without DNA, this protein–protein interface is presumably of low intrinsic affinity.

CBF3B extends beyond the DNA duplex present in the CBF3–CEN3 complex. We therefore modeled an extended DNA duplex downstream of CBF3B guided by the DNA contacts with CBF3A (Fig. 4a). This indicated that the DNA in contact with the CBF3 dimer comprises CDEIII together with 6 bp of CDEII and 24 bp to the right of CDEIII (4CDEIII) (Figs. 1i and 4a). The modeled complex exactly matches the 56-bp DNase resistant core protected by CBF3 (Fig. 4f).

Cep3A of CBF3A forms two important sequence-specific contacts to CDEIII. The Zn2Cys6 cluster engages the major groove of DNA contacting the essential CCG motif (Fig. 3f–i). Modeling of this interaction was based on the Hap1–DNA structure. The second sequence-specific contact is created by residues 318–328 of Cep3A, which, in the presence of CEN3, fold into an α-MN helix (α-MN) that inserts into the major groove of DNA 1.25 turns 5' to the CCG motif. The α-MN helix contacts the conserved TGT motif that is essential for optimal CBF3 binding (Figs. 3g,h and Supplementary Fig. 6b).

Cep3B of CBF3B forms two important sequence-specific contacts to CDEIII. The Zn2Cys6 cluster engages the major groove of DNA contacting the essential CCG motif (Fig. 3f–i). Modeling of this interaction was based on the Hap1–DNA structure. The second sequence-specific contact is created by residues 318–328 of Cep3A, which, in the presence of CEN3, fold into an α-MN helix (α-MN) that inserts into the major groove of DNA 1.25 turns 5' to the CCG motif. The α-MN helix contacts the conserved TGT motif that is essential for optimal CBF3 binding (Figs. 3g,h and Supplementary Fig. 6b). In between these two Cep3A sequence-specific contact sites, the Ctf13 loop (comprising residues 290–310) inserts into the minor groove and interacts with the DNA phosphate backbone (Fig. 3g). In the CBF3B promoter of the CBF3 dimer, the Zn2Cys6 cluster of Cep3A inserts into the major groove of 4CDEIII some
to surround the histone octamer\textsuperscript{15}. Budding yeast Cse4 nucleosomes deposited on CEN DNA are centered on CDEII and protect a micrococal nuclease-resistant kernel of 125–135 bp\textsuperscript{14}. In our CBF3–CEN cryo-EM structure the curvature of the 12 bp of CDEII matches that of the CENP-A nucleosome DNA\textsuperscript{16}. We modeled a CBF3–Cse4–CEN3 complex, assuming that the Cse4 nucleosome is an octamer and wraps 115–120 bp of DNA\textsuperscript{26,41}. We placed the entry point of the nucleosome 4 bp 5' of the CDEII–CDEIII junction (Figs. 1i and 4b,c). The modeled CBF3–Cse4–CEN3 complex is devoid of steric clashes, superimposes the major and minor grooves of the CBF3-bound and nucleosome DNA, and allows Ndc10 to interact with CDEII and possibly with Cbf1 engaged at CDEI, in agreement with a previous study\textsuperscript{42}. The N terminus of Cse4 approaches Ctf13 (Fig. 4c), suggesting a mechanism for specific recognition of Cse4 nucleosomes by CEN-bound CBF3\textsuperscript{43}. Our model is compatible with both Cse4–H4–H2A–H2B hemisomes\textsuperscript{28} and octameric nucleosomes\textsuperscript{31}. Whereas a left-handed supercoiled DNA can be readily accommodated, consistent with previous reports\textsuperscript{23,41}, our modeling suggests that a nucleosome with right-handed chirality would collide with CBF3\textsuperscript{43} (Supplementary Fig. 6c,d). The 120 bp wrapped by the Cse4 nucleosome together with 56 bp of the CBF3 dimer would generate a protein–DNA interface of ~180 bp. This matches well with the single-base pair resolution mapping of Cse4 nucleosomes\textsuperscript{46,47} and the size of nucleosome-resistant centromere cores\textsuperscript{30,31}. Due to the symmetry of the CBF3 dimer bound to the CEN3 DNA, an interesting feature of this model is the possibility that the CBF3 dimer interacts with two Cse4 nucleosomes. Budding yeast centromeres containing more than one Cse4 nucleosome have been proposed\textsuperscript{40}.

In the CBF3–CEN3 structure, Ndc10\textsuperscript{108} is dimerized on DNA (Fig. 3a–c). Using size exclusion chromatography–multi-angle light scattering (SEC–MALS) we analyzed the molecular mass of CBF3 in the presence and absence of CEN3 DNA. This indicated that CEN3 DNA enhanced the dimerization of a CBF3\textsuperscript{108} dimer to generate a CBF3–CEN3 complex incorporating two Ndc10 dimers (Supplementary Fig. 7), consistent with Ndc10 forming higher order multimers\textsuperscript{32}. Extended 89-bp DNA segments (33 nucleotides to the right of the core 56-bp segment) bind an additional CBF3 complex that is stimulated by Ndc10\textsuperscript{108}. One possibility is that a dimer of Ndc10 dimers mediates bridging of two DNA-bound CBF3 dimers facilitating either DNA loop or pairing of sister chromatids (Fig. 4d). Intriguingly, AFM images of CBF3–CEN DNA complexes revealed paired complexes from which project three to four DNA arms\textsuperscript{30}.

**Discussion**

The structure of a CBF3–CEN3 complex provides the first insights into the architecture of point centromeres, revealing a head-to-head CBF3 dimer that explains the CBF3 nucleosome-protected region of CEN DNA\textsuperscript{17} and CBF3–CEN3 crosslinking data\textsuperscript{27}. Three of the CBF3 subunits (Cep3, Ctf13 and Ndc10) interact with DNA. Two zinc fingers of the CBF3 dimer, one from each Cep3 dimer of the two CBF3 protomers, directly engage the major groove of the DNA. The DNA-binding zinc finger of Cep3A of CBF3\textsuperscript{3} interacts with the conserved CCG of the CDEIII motif of CEN3, explaining the crucial role and conservation of the CCG motif\textsuperscript{13–15}. The second sequence-specific contact within CDEIII involves the α-MN helix of Cep3B of CBF3\textsuperscript{3} that contacts the conserved TGT motif that is essential for optimal CBF3 binding\textsuperscript{25}. Both the zinc finger of Cep3A and the α-MN helix of Cep3B assume order on binding CDEIII. We present a model for the CBF3–Cse4–CEN3 complex that supports the idea that CENP-A nucleosomes wrap DNA as a left-handed helix. The close proximity of the CENP-A-specific Cse4 protein to CBF3 suggests that CENP-A nucleosomes wrap DNA in a manner that allows CBF3 to directly engage the major groove of the DNA.

**Implications for CBF3–Cse4 nucleosome interactions.** In human CENP-A nucleosomes, 120 bp of α-satellite DNA generate 1.4 gyres that extend two DNA arms subtended by ~55°. Similar to our structure, these AFM images revealed a bilobal globular domain from which DNA emerges (Fig. 4). The DNA-free monomeric CBF3 core–Ndc10DBD structure\textsuperscript{11} (PDB 3AN2). The two CENP-A subunits (labeled as Cse4A and Cse4B) are highlighted. In our model, CBF3 and Cse4 are considered as a single-domain complex. This is consistent with previous reports\textsuperscript{41} and Ndc10 dimers mediate (Fig. 4e). Intriguingly, AFM images of CBF3–CEN DNA complexes revealed paired complexes from which project three to four DNA arms\textsuperscript{30}.
framework for understanding the structure of the centromere–inner kinetochore interface.

**Online content**
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41594-018-0154-1.

Received: 22 June 2018; Accepted: 10 October 2018; Published online: 26 November 2018

**References**
1. Bloom, K. S. et al. Chromatin conformation of yeast centromeres. *Proc. Natl Acad. Sci. USA* 92, 12026–12030 (1995).
2. Espelin, C. W., Kaplan, K. B. & Sorger, P. K. Probing the architecture of a simple kinetochore using DNA-protein crosslinking. *J. Cell Biol.* 139, 1383–1396 (1997).
3. Russell, I. D., Grancell, A. S. & Sorger, P. K. The unstable F-box protein p58-Cih13 forms the structural core of the CBF3 kinetochore complex. *J. Cell Biol.* 145, 933–950 (1999).
4. Dechassa, M. L. et al. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat. Commun.* 2, 313 (2011).
5. Nakane, T., Kinniani, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *elife* 7, e36861 (2018).
6. Cho, U. S. & Harrison, S. C. Ndc10 is a platform for inner kinetochore assembly in budding yeast. *Nat. Struct. Mol. Biol.* 19, 48–55 (2011).
7. Kaplan, K. B., Hyman, A. A. & Sorger, P. K. Regulating the yeast kinetochore by ubiquitin-dependent degradation and Skp1-mediated phosphorylation. *Cell* 91, 491–500 (1997).

**Acknowledgements**
This work was funded by a MRC grant (no. MC_UP_1201/6) and a CR-UK grant (no. CR-UK-M016). D.B. thanks S. Chen, G. Cannone and G. McMullan for help with EM data collection, J. Grimmett and T. Darling for computing and J. Shi for help with insect cell expression.

**Author contributions**
D.B. directed the project. Z.Z. cloned all constructs, I.Y. and Z.Z. purified complexes, Z.Z. performed the miototic stability assay, I.Y. and S.M. performed SEC-MALS experiments. K.Y. prepared EM grids, collected and analyzed EM data. K.Y. and D.B. wrote the manuscript.
Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41594-018-0154-1.

Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.B.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018
Methods

Cloning, expression and purification of recombinant CBF3–CEN3 complex. Coding fragments of Cep3, Ndc10, Ctf13 and Skp1 were amplified by PCR from \textit{S. cerevisiae} genomic DNA and cloned into a pUC plasmid using a modified Multilinker expression system, as described. A double StreptII tag together with a Tobacco Etch Virus (TEV) cleavage site was attached to the C terminus of Cep3. All four genes were further assembled into a pUC plasmid for CBF3 complex expression. We define CBF3 as Cep3, Ctf13, Skp1 and Ndc10, and CBF3DBD as Cep3, Ctf13 and Skp1. CBF3 is equivalent to CBF3DBD–Ndc10. Ndc10 DBD (Ndc10 DBD) was combined with Cep3, Ctf13 and Skp1 into pUCl2 for CBF3–Ndc10 DBD expression. The C termini of Ndc10 and Ndc10 DBD were modified with a TEV His-tag for Ndc10 and Ndc10 DBD purification. The Ctf13 DBD-T7 and Ctf13 DBD-T7S mutations were made and combined with Cep3, Ndc10 and Skp1 in pUCl2 for the mutant CBF3 complexes. All complexes were expressed in the insect cell–baculovirus system using Hi5 cells as described.

To prepare CEN3 DNA, 24 copies of the 147 bp of CEN3 \textsuperscript{0} (sequence: ATC AAA TAG TAC AAA TAA [GTCA ACA TGA TGA TTA ATT TTA TTA TTA TTA AAA AAG TAA AAA ATA AAA AGT AGT TTA TTA TTA AAA AAA ATT AAT ATT AGT GTA TTA GAT TTC CAA AAG TAA AAA AAG AAA TAG TAA GAA GTA TTT GAT TTC CGA AAG TTA AAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT AAA TAG TAC AAA TAA [GTC ACA TGA TGA TAT TTG ATT TTA TTA TAT]

Video frames were first aligned using MotionCor2 \textsuperscript{51}. Contrast transfer function parameters were estimated with Gctf \textsuperscript{52}. The initial template-free 3D classification was applied to the image processing and a low-pass filtered structure was fitted into the monomeric CBF3DBD as a reference in the 3D classification. A total of 145,747 particles were assigned to CBF3DBD–Ndc10 DBD, whereas 184,356 were assigned to CBF3DBD. After multireference-based 3D classifications, 87,156 particles of CBF3DBD–Ndc10 DBD were selected for the final refinement. For CBF3DBD, different conformations of the zinc-finger motifs were characterized after a multireference 3D classification pipeline including the particles generated by the highest resolution reconstruction. The final resolutions for CBF3DBD–Ndc10 DBD and CBF3DBD are 3.58 Å and 3.91 Å, respectively, based on the gold-standard Fourier shell correlation FSC=0.143 criterion \textsuperscript{53} (Supplementary Fig. 3g).

For the CBF3–CEN3 dataset (Supplementary Fig. 2 and Table 1), a similar workflow was applied to the image processing and a low-pass filtered structure of CBF3–CEN3 DBD was used as a reference in the 3D classification. A dimeric CBF3–CEN3 complex and a monomeric CBF3–CEN3 complex were characterized. A total of 99,089 particles were assigned to the dimeric CBF3–CEN3 complex, and 22,688 (class 5) of these were seen to be longer than a DNA density and were class-based refined with an initial 4.4 Å map (Supplementary Fig. 2h). 197,322 particles were assigned to the monomeric CBF3–CEN3 complex. We applied multireference ambiguity in RELION \textsuperscript{17} to segment three domains of the dimeric CBF3–CEN3 complex to improve their local resolutions: (1) the upper CBF3 DBD; (2) the CEN3 DNA with associated Ndc10 DBD; and (3) the lower CBF3 DBD, resulting in reconstructions at 3.89 Å, 4.33 Å and 4.16 Å, respectively (Supplementary Table 1 and Supplementary Fig. 2h). The flexible orientation between the CBF3 DBD and Ndc10 DBD is stabilized by the CEN3 DNA (Supplementary Fig. 4). A local mask encompassing the upper CBF3 complex only (CBF3 \textsuperscript{+}) with the Zn cluster of C3p3B was employed for both 3D classification and subsequent density improvement in the refinement. The particles assigned to either the dimeric CBF3–CEN3 or the monomeric CBF3–CEN3 complexes were subject to 3D classification and the particles in the best-quality classes were saved. A total of 198,010 particles were used for the final reconstruction, and a 3.05 Å resolution map was obtained for CBF3+ (Supplementary Fig. 2h, j,k).

Before visualization, a negative B factor determined with RELION was applied to the density map for sharpening. The model fitting function of the detector was corrected in the postprocessing step with RELION. The local resolution was estimated with ResMap 	extsuperscript{17} integrated in RELION 2.1.

Model building and structure refinement. Three maps are defined as: (1) 3.6 Å map of CBF3DBD–Ndc10 DBD (no CEN3 DNA) (Supplementary Fig. 3f); (2) 4.4 Å map of CBF3–CEN3 DNA (Supplementary Fig. 2d) and (3) 3.0 Å map of CBF3DBD (Supplementary Fig. 2h; Table 1 and Supplementary Table 1).

The crystal structures of the \textit{S. cerevisiae} Cep3 dimer \textsuperscript{18} (PDB:2QQU) and Ndc10 DBD (PDB: 4ACO) (residues 44–337 and the model of the \textit{S. cerevisiae} Skp1 was fitted into monomeric CBF3DBD–Ndc10 DBD–CEN3. The \textit{S. cerevisiae} Skp1 model was based on human Skp1 (PDB:1QF7) with the PDBsum server (http://zhanglab.ccmb.med.umich.edu/I-TASSER). Based on the excellent quality of the EM density map, an atomic model of the Ctf13 was build de novo using COOT \textsuperscript{19}. The secondary structure and disordered regions of the Ctf13 sequence were analyzed with PHyre2 \textsuperscript{21} and PSIPred \textsuperscript{22}. Amino acids were assigned based mainly on the predicted secondary structures and clearly defined bulky residue densities. The final model of Ctf13 lacks the N-terminal two residues and two unstructured loop regions (50–54 and 205–252). The zinc-finger motif (1–53) in the Cep3 N-terminal was built guided by the Hap1 structure (PDB: 2HAJ). The extreme N terminus of Ndc10, except the missing first 26 amino acids, shows clear side chain densities that facilitated the model building. This CBF3 model was fitted into the cryo-EM density map of the dimeric CBF3–CEN3 complex following multiple refinement steps. Based on the position of avidin in the CBF3–CEN3–Ctf13 crosslinking data \textsuperscript{24}, we built the CEN3 DNA model. The model was optimized with several rounds of real-space refinement in PHENIX (phenix.real_space_refine) \textsuperscript{25}. Standard stereochemical and secondary-structural constraints were applied during the real-space refinement. The final models were evaluated with MolProbity (http://molsprobity.biochem.duke.edu/)

Modeling of the CBF3–Cep3–Cenp-A nucleosome. To model the CBF3–Cep3–CEN3 complex we made the following assumptions:

(1) CBF3: The position of CBF3 on CEN3 DNA remains unchanged and that CBF3 undergoes no conformational changes. Relative to its DNA-free state, CBF3 does not undergo major conformational change in the CBF3–CEN3 complex.

(2) Csf1: The Csf1 nucleosome was modeled based on the CENP-A crystal structure \textsuperscript{26}. In this structure the CENP-A nucleosome is formed from a 147-bp palindromic DNA derived from a human α-satellite DNA sequence; 120 bp of
Articles

A fragment of ARS1::TRP1::CEN3 flankcd by two EcoRI sites was cloned in pUC18 plasmid. Based on this plasmid, a CEN3<sup>Δctf13</sup>-<sup>ΔcdeIII</sup> mutation was made. Mini-chromosomes (1442 bp) of CEN3<sup>Δctf13</sup> and CEN3<sup>Δctf13</sup>-<sup>ΔcdeIII</sup> were made by cycling plasmid ligation fragment ARS1::TRP1::CEN3 released from the pUC18 plasmids by EcoRI. The mini-chromosomes were transformed into yeast strains harboring Ctf13 wild type or its mutants, selected on synthetic media lacking leucine, tryptophan and surplus G418.

Single colonies of yeast strains carrying wild-type and mutant Ctf13, in combination with a mini-chromosome with either wild-type or mutant CEN3 from synthetic complete (Sc)-Leu-Tip + G418 selective plates, were selected to inoculate 3 ml of Sc-Leu + G418 medium, a nonselective growth medium for mini-chromosomes. After 48 h growth at 30°C, cells were plated for single colonies onto Sc-Leu + G418 plates. After 3 days' incubation at 30°C, colonies were plated onto Sc-Leu-Tip + G418 selective plates. These plates were incubated for 3 days at 30°C, and the percentage of growth population bearing mini-chromosomes was determined.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** EM maps have been deposited with EMDB (EMD-0095, EMD-0096, EMB-0097). Protein coordinates have been deposited with RCSB (6GYS, 6GYS, 6GYY). Other data are available upon reasonable request.

**References.**

49. Fitzgerald, D. J. et al. Multiprotein expression strategy for structural biology of eukaryotic complexes. *Structure* 15, 275–279 (2007).
50. Zhang, Z., Yang, J. & Barford, D. Recombinant expression and reconstitution of multiprotein complexes by the USER cloning method in the insect cell-baculovirus expression system. *Methods* 95, 13–25 (2016).
51. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332 (2017).
52. Zhang, K. Geometric real-time CTF determination and correction. *J. Struct. Biol.* 193, 1–12 (2016).
53. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D Struct. Biol.* 73, 496–502 (2017).
54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501 (2010).
55. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858 (2015).
56. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* 41, W349–W357 (2013).
57. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221 (2010).
58. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21 (2010).
59. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 399, 251–260 (1999).
60. Chiti, S. et al. Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. *Science* 359, 339–343 (2018).
61. Pentakota, S. et al. Decoding the centromeric nucleosome through CENP-N. *eLife* 6, e33442 (2017).
62. Bao, Y., White, C. L. & Luger, K. Nucleosome core particle containing a poly(dA·dT) sequence element exhibit a locally distorted DNA structure. *J. Mol. Biol.* 361, 617–624 (2006).
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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
- [ ] An indication of 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 statistics 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
- [ ] Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

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, GCTF, RELION2.1, SIMPLEPRIME and REFSMAP. Model building and refinement were performed using COOT and Phenix. Visualization was performed with COOT, Pymol and Chimera. ConSurf was used for analyzing structural conservation and PISA for analyzing protein interface surface areas. |

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 upon request. 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

The three dimensional cryo-EM density maps have been submitted to EM-DIR under accession numbers 0095, 0096, 0097. Coordinates have been deposited with RCSB under accession numbers 6GYP, 6GY5, 6GYU. All other data are available from the corresponding author on reasonable request.

Field-specific reporting

Please select 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/authors/policies/reportingSummary-flat.pdf

Life sciences study design

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

Sample size
For the mitotic stability assay, a sample size of 100 was used (yeast colonies) per yeast variant. Data were analyzed using paired t-tests. The experimental data are the result of five replicates.

Data exclusions
No data were excluded.

Replication
All attempts at replication were successful and reproducible.

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.

Behavioural & social sciences study design

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

Study description
Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample
State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.

Sampling strategy
Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size or if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.

Data collection
Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.

Timing
Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.

Data exclusions
If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Non-participation
State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

Randomization
If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.
Ecological, evolutionary & environmental sciences study design

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

**Study description**
Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

**Research sample**
Describe the research sample (e.g. a group of tagged Passer domesticus, all Sterocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

**Sampling strategy**
Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

**Data collection**
Describe the data collection procedure, including who recorded the data and how.

**Timing and spatial scale**
Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken.

**Data exclusions**
If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

**Reproducibility**
Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

**Randomization**
Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

**Blinding**
Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

**Did the study involve field work?**  ○ Yes  ○ No

---

**Field work, collection and transport**

**Field conditions**
Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

**Location**
State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

**Access and import/export**
Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

**Disturbance**
Describe any disturbance caused by the study and how it was minimized.

---

**Reporting for specific materials, systems and methods**

---

**Materials & experimental systems**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | Unique biological materials |
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |

---

**Methods**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |
Unique biological materials
Policy information about availability of materials

Obtaining unique materials NA

Antibodies

Antibodies used NA
Validation NA

Eukaryotic cell lines
Policy information about cell lines

Cell line source(s) Insect cells lines for baculovirus expression: Sf9 and High five.
Authentication Insect cells from Invitrogen were not further authenticated.
Mycoplasma contamination Protein expression cell lines were not tested.
Commonly misidentified lines (See: ICGC register) None

Palaeontology

Specimen provenance NA
Specimen deposition Indicate where the specimens have been deposited to permit free access by other researchers.
Dating methods If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms
Policy information about studies involving animals, ARRIVE guidelines recommended for reporting animal research

Laboratory animals NA
Wild animals Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method, if released, say where and when) OR state that the study did not involve wild animals.
Field-collected samples For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Human research participants
Policy information about studies involving human research participants

Population characteristics Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."
Recruitment Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

ChIP-seq

Data deposition
☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.
☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.
Data access links
May remain private before publication.

Files in database submission
Provide a list of all files available in the database submission.

Genome browser session
(e.g. UCSC)
Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates
Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth
Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies
Describe the antibodies used for the ChiP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters
Specify the command line program and parameters used for read mapping and peak calling, including the ChiP, control and index files used.

Data quality
Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software
Describe the software used to collect and analyze the ChiP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
☐ All plots are contour plots with outliers or pseudocolor plots.
☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument
Identify the instrument used for data collection, specifying make and model number.

Software
Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance
Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy
Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary information.

Magnetic resonance imaging

Experimental design

Design type
Indicate task or resting state; event-related or block design.

Design specifications
Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures
State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).
### Acquisition

**Imaging type(s)**
Specify: functional, structural, diffusion, perfusion.

**Field strength**
Specify in Tesla

**Sequence & imaging parameters**
Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

**Area of acquisition**
State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

**Diffusion MRI**
- [ ] Used
- [ ] Not used

### Preprocessing

**Preprocessing software**
Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

**Normalization**
If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

**Normalization template**
Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

**Noise and artifact removal**
Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

**Volume censoring**
Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

### Statistical modeling & inference

**Model type and settings**
Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or aut-correlation).

**Effect(s) tested**
Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

**Specify type of analysis:**
- [ ] Whole brain
- [ ] ROI-based
- [ ] Both

**Statistic type for inference**
(See Eklund et al., 2016)
Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

**Correction**
Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

### Models & analysis

**n/a Entered in the study**
- [ ] Functional and/or effective connectivity
- [ ] Graph analysis
- [ ] Multivariate modeling or predictive analysis

**Functional and/or effective connectivity**
Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

**Graph analysis**
Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

**Multivariate modeling and predictive analysis**
Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.