ARTICLE

doi:10.1038/nature26003

Integrative structure and functional anatomy of a nuclear pore complex

Seung Joong Kim[1]*, Javier Fernandez-Martinez[2]*, Ilona Nudelman[2]*, Yi Shi[3]†*, Wenzhu Zhang[3]*, Barak Raveh[1], Thurston Herricks[4], Brian D. Slaughter[5], Joanna A. Hogan[6], Paula Upla[7], Ilan E. Chemmama[1], Riccardo Pellarin[1]†, Ignacia Echeverria[1], Manjunatha Shivaraju[5], Azraa S. Chaudhury[2], Junjie Wang[3], Rosemary Williams[2], Jay R. Unruh[5], Charles H. Greenberg[1], Erica Y. Jacobs[3], Zhiheng Yu[8], M. Jason de la Cruz[8]†, Roxana Mironska[2], David L. Stokes[7], John D. Aitchison[4,9], Martin F. Jarrold[6], Jennifer L. Gerton[5], Steven J. Ludtke[10], Christopher W. Akey[11], Brian T. Chait[3], Andrej Sali[1] & Michael P. Rout[2]

Nuclear pore complexes play central roles as gatekeepers of RNA and protein transport between the cytoplasm and nucleoplasm. However, their large size and dynamic nature have impeded a full structural and functional elucidation. Here we determined the structure of the entire 552-protein nuclear pore complex of the yeast *Saccharomyces cerevisiae* at sub-nanometre precision by satisfying a wide range of data relating to the molecular arrangement of its constituents. The nuclear pore complex incorporates sturdy diagonal columns and connector cables attached to these columns, imbuing the structure with strength and flexibility. These cables also tie together all other elements of the nuclear pore complex, including membrane-interacting regions, outer rings and RNA-processing platforms. Inwardly directed anchors create a high density of transport factor-docking Phe-Gly repeats in the central channel, organized into distinct functional units. This integrative structure enables us to rationalize the architecture, transport mechanism and evolutionary origins of the nuclear pore complex.

Nuclear pore complexes (NPCs) are large proteinaceous assemblies studded through the nuclear envelope, the double-membraned barrier that surrounds the nucleus; NPCs are the sole mediators of macromolecular transport between the nucleus and the cytoplasm, and carry key regulatory platforms for numerous nuclear processes[1]. NPCs are also major targets for viral manipulation and defects in this transport machine are directly linked to human diseases, including cancers[2]. Each NPC is an eight-fold symmetric, cylindrical assembly consisting of approximately 550 copies of about 30 different proteins of the nucleoporin family (Nups). These Nups assemble into sub-complexes that form higher-order structures called spokes. Eight spokes assemble into even larger modules: coaxial outer and inner rings form a symmetric core scaffold, which is connected to a membrane ring, a nuclear basket and cytoplasmic RNA export complexes[3]. The scaffold surrounds a central channel that is formed in part by multiple intrinsically disordered Phe-Gly (FG) repeat motifs that extend from nucleoporins termed FG Nups. These FG motifs mediate selective nucleocytoplasmic transport through specific interactions with nuclear transport factors (NTFs), which carry their cognate macromolecular cargoes[4]. It has also previously been suggested that the central channel contains a feature called the central transporter[5]. Although partial structures have previously been described[3,6,7], a complete, high-resolution structure for the entire NPC in any organism has hitherto been lacking, leaving open key questions as to how the NPC is organized and functions, and how it evolved. To address these questions, we have determined an integrative structure of the yeast NPC at sub-nanometre precision.

Solving the structure of the *S. cerevisiae* NPC

We developed a method to rapidly and gently isolate native yeast NPCs, enabling us to determine the type and amount of each Nup in the NPC, the proximities between Nups resolved to the amino acid residue level, and the mass and detailed morphology of the entire NPC. These data were then used to solve the structure using an integrative modelling approach[8,9] (Extended Data Fig. 1, Supplementary Results and Discussion, and Methods).

We determined the mass of the entire NPC and a definitive stoichiometry for every Nup and associated molecules using mass spectrometric and *in vivo* imaging methods. The native NPC has a mass of 52 MDa, or about 87 MDa when including the membrane, cargo and NTFs (Fig. 1 and Extended Data Figs 2, 3c). To inform the proximities, orientations and conformations of the Nups, isolated NPCs were subjected to cross-linking with mass spectrometric readout[9,10]. This approach identified 3,077 unique cross-linked pairs of residues, and provided the distance restraints between them, both within and between Nups (Fig. 2, Supplementary Table 1 and Methods). The morphology of the NPC was determined using cryo-electron tomography (cryo-ET) and sub-tomogram averaging[11] (Methods). This approach provided a final 3D map at approximately 28 Å resolution, with a local resolution of 20–25 Å for the inner ring, which has approximate C_2 symmetry (Fig. 3 and Extended Data Figs 4–6). The NPCs retained a considerable amount of nuclear envelope membrane, which forms a continuous belt around the midline of the structure (Fig. 3a, b, d, e). We found that a membrane protein ring interconnects adjacent spokes

[1]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California 94158, USA. [2]Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, New York 10065, USA. [3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York 10065, USA. [4]Institute for Systems Biology, 401 Terry Ave. N., Seattle, Washington 98109, USA. [5]Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. [6]Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA. [7]Skirball Institute and Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA. [8]Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20148, USA. [9]Center for Infectious Disease Research, Seattle, Washington 98109, USA. [10]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA. [11]Department of Physiology and Biophysics, Boston University School of Medicine, 700 Albany Street, Boston, Massachusetts 02118, USA. †Present addresses: Structural Bioinformatics Unit, Institut Pasteur, CNRS UMR 3528, Paris, France (R.P.); Department of Cell Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, USA (Y.S.); Structural Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA (M.J.d.l.C.).
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

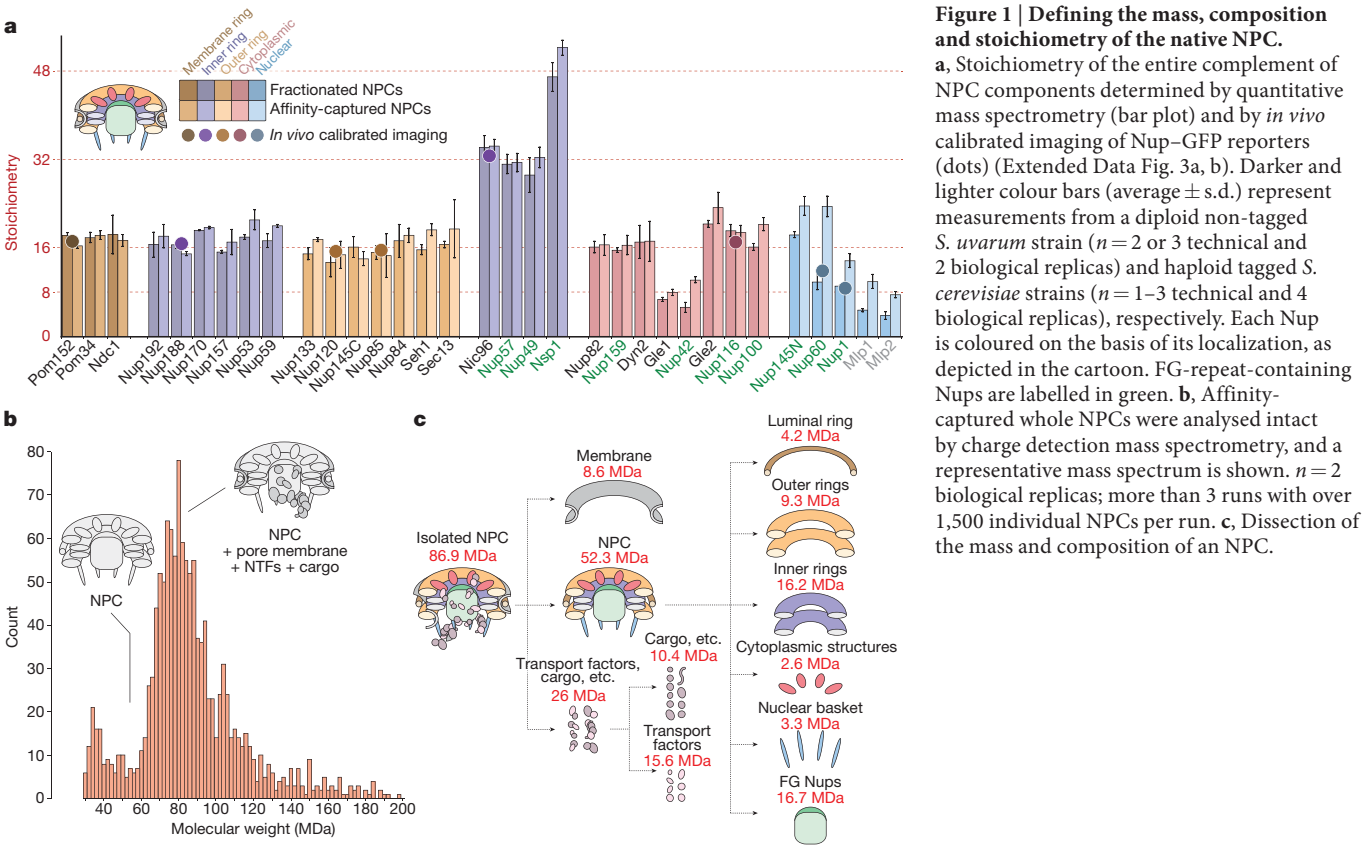


Figure 1 | Defining the mass, composition and stoichiometry of the native NPC. **a**, Stoichiometry of the entire complement of NPC components determined by quantitative mass spectrometry (bar plot) and by *in vivo* calibrated imaging of Nup–GFP reporters (dots) (Extended Data Fig. 3a, b). Darker and lighter colour bars (average ± s.d.) represent measurements from a diploid non-tagged *S. uvarum* strain ($n = 2$ or 3 technical and 2 biological replicas) and haploid tagged *S. cerevisiae* strains ($n = 1$–3 technical and 4 biological replicas), respectively. Each Nup is coloured on the basis of its localization, as depicted in the cartoon. FG-repeat-containing Nups are labelled in green. **b**, Affinity-captured whole NPCs were analysed intact by charge detection mass spectrometry, and a representative mass spectrum is shown. $n = 2$ biological replicas; more than 3 runs with over 1,500 individual NPCs per run. **c**, Dissection of the mass and composition of an NPC.

within the nuclear envelope lumen (Fig. 3a, e), a feature largely absent from recent electron microscopy maps. A cylindrically averaged, bi-lobed density fills the central channel (the 'central transporter', Fig. 3a, b). Individual Nups and their domains, as well as the subcomplexes of the NPC, were represented on the basis of published crystallographic structures, integrative structures and comparative models[9,10,12] (Supplementary Table 2 and Methods), and validated by small-angle X-ray scattering profiles for 18 Nups (147 constructs; Supplementary Table 6 and Methods). An ensemble of structural solutions for the NPC that sufficiently satisfied all experimental data was calculated by extensive configurational sampling[8,9] (Supplementary Table 3 and Methods). Variability among these solutions defines the precision of our structure, as quantified by the average root-mean-square deviation between solutions in the final ensemble[9]. Our final structure defines the positions of 552 Nups (Fig. 4 and Supplementary Videos 1–3), with an overall precision of about 9 Å (Extended Data Fig. 1e, f). The centroid solution is used as the representative structure. The structure was validated by numerous independent tests (Extended Data Figs 1, 7, 8, Supplementary Tables 3, 4 and Methods).

The multiple functionalities and enormous size of the NPC present unique and substantial structural challenges: it must form a stable passageway with a fixed inner diameter; it must be anchored to the nuclear envelope and stabilize the pore membrane within which it resides, with a height appropriate for the thickness of the nuclear envelope; it must correctly position the transport machinery; and it must resist stresses that might lead to disassembly or malfunction. Our structure suggests how each of these challenges is met and—by comparison with the vertebrate scaffold[6]—how different organisms may meet these challenges (see the section 'Evolutionary origin and diversity of the NPC').

Forming a stable and defined passageway

The fitness defects of strains containing Nup truncations provide an estimate of the structural importance of the truncated regions[9,13]. We quantified the fitness defect of strains containing systematic truncations of every major symmetric Nup using ODELAY[14] (an automated phenotypic analysis platform; Extended Data Fig. 9). Results were heat mapped onto the NPC structure to reveal critical elements of NPC stability (Fig. 5a). The inner ring of the NPC contains crucial stabilizing elements, including Nic96, which forms the heart of a diagonally oriented column within each spoke (Fig. 5b) and interacts with every other protein in the inner ring (Fig. 4d–f). This high connectivity explains why Nic96 is an essential keystone, holding in place much of the scaffold of the NPC. The remainder of each diagonal column is made of Nup157 and Nup170, which flank Nic96 (Fig. 5b); Nup157 and Nup170 are functionally redundant, but are synthetically lethal[15] and together form another essential element of the diagonal column. Inter-spoke connections represent a second crucial stabilizing element. Nup192 probably serves as a cross-brace between adjacent spokes (Fig. 5a, c). The N termini of Nup170 and disordered regions of Nup53 and Nup59 also form key connections between adjacent spokes[16] (Figs 4d, 5c). The inter-spoke connections are established largely through small, hinge-like contacts that may confer flexibility to the interface between adjacent spokes; the diagonal arrangement of the central columns may also enable rotation or local flexing (Fig. 5b), by accommodating compression and expansion forces from nuclear envelope distortions and from the central transporter and the transit of cargoes. Nup188 and Nup192 act as radial separators between the Nic96 column and the triple coiled-coil domains of Nsp1, Nup57 and Nup49, which form a discontinuous ring that defines the narrowest part of the passageway and may allow some dilation of the NPC (Fig. 4d–f). This architecture sets a soft upper limit of about 40 nm for the size of cargoes that can transit the NPC[4].

How the NPC shapes the nuclear envelope

The pore membrane, where the inner and outer membranes of the nuclear envelope join, defines the inner surface of a torus and therefore has both concave and convex curvatures (Fig. 5d). The inner ring is anchored to the pore membrane through membrane-binding motifs (MBMs) on the β-propellers at the N termini of Nup157 and Nup170, and on the C termini of Nup53 and Nup59 (refs. 17–19). These proteins

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

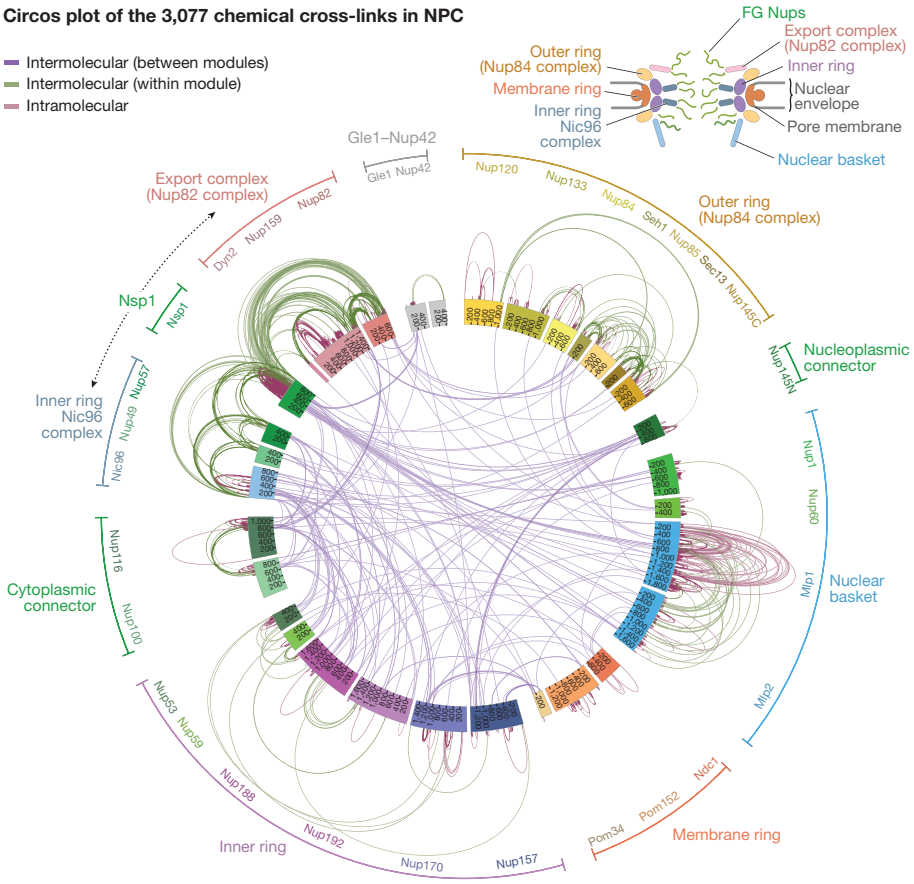


Figure 2 | Chemical cross-linking and mass spectrometry reveals nucleoporin connectivity in the NPC. Circular plot showing the distribution of chemical cross-links (Supplementary Table 1), mapped to each nucleoporin represented as a coloured segment and with the amino acid residues indicated. The identity of each module and Nup is shown in the periphery of the plot. Key in top left gives the types of cross-links; diagram in top right illustrates the relative positions of modules in the NPC.

also interact with the scaffold-facing regions of the transmembrane domain (TMD) carrying Nups, such as Pom152, Ndc1 and Pom34 (Fig. 4d). Together, these MBMs and TMDs form an NPC-anchoring girdle of membrane-associated motifs around the scaffold equator, defining the concave curve of the pore membrane. The convex curvature is defined by both outer and inner rings (Figs 4a, e, 5d). Each outer ring is formed by eight Y-shaped Nup84 complexes arranged head-to-tail and joined by an interaction between the N termini of Nup120 and Nup133 (ref. 20), creating another hinged spoke-to-spoke interface and a minor fitness hotspot (Fig. 5a, c). The outer rings also help define the overall height of the NPC such that it is appropriate for the width of the nuclear envelope. Each Nup84 complex is anchored to the pore membrane by MBMs situated within the N-terminal β-propellers of Nup133 and Nup120[10,12,21] (Fig. 5d). The convex curvature of the pore membrane is thus defined and stabilized by a ring of MBMs underneath the outer rings and by the thick girdle of MBMs and TMDs around the NPC equator. On the nuclear side, MBMs from Nup1 and Nup60 help anchor the basket to the nuclear envelope[22] (Fig. 5d, e).

In the membrane ring, the luminal domain of Pom152 is composed of nine immunoglobulin-like fold repeats[23] that oligomerize in an anti-parallel fashion to form eight circumferential arches within the nuclear envelope lumen, producing additional connections between adjacent spokes. Pom152 appears to be pre-stressed by assembly into these arches (Fig. 4e); the resulting tension may minimize elliptical distortion of the NPC[23]. Each arch also delimits a channel (300 × 120 Å wide) between itself and the underlying pore membrane (Figs 3e, 4e). The outer rings form a series of circumferential arches that align with the Pom152 luminal arches (Fig. 4b, e). These arches align with hinges in the inner ring (Fig. 5c) that could flex to form lateral openings between spokes. This juxtaposition of arches and transient openings may delineate conduits for nucleocytoplasmic transport of transmembrane proteins[24], potentially resolving the issue of how membrane proteins transit the NPC[25].

Positioning the RNA processing platforms

Whereas the core scaffold is symmetric about the plane of the nuclear envelope, two machineries associated with RNA processing and transport—the basket and export platform— are located at the nuclear and cytoplasmic faces of the NPC, respectively (Fig. 4d–f). At the core of the export platform is the Nup82 complex, which has a coiled-coil bundle that is attached to the Nup85–Seh1 arm and hub region of the Nup84 complex in the cytoplasmic outer ring (Fig. 4f). Together, they form a lateral platform that faces the central channel. An α-helical rod that extends from the Nup82 complex holds Gle1, the RNA helicase Dbp5 and the FG-repeat-carrying Nup42 over the middle of the central channel[9,25,26]. As a result, numerous transport-factor-docking sites and ATP-dependent RNA remodelling proteins are aligned above the cytoplasmic exit of the NPC to efficiently receive exporting RNAs, to remodel and then release them into the cytoplasm. Likewise, Mlp1 and Mlp2 in the nuclear basket are anchored to the core scaffold mainly by the Nup85–Seh1 arm, similar to the Nup82 complex (Figs 4d, e, 5e). The nuclear basket serves as a platform for the first stages of RNA processing and export[27], and the export platform organizes the last stages of export[26]. Similarities between the export platform and basket suggest that these structures are ancient homologues (Extended Data Fig. 10); their asymmetric localization directs unidirectional export of transcripts out of the nucleus.

Flexible connectors tie the NPC together

Certain disordered connectors have recently been shown to be important for holding parts of the scaffold together[6,7,16]; the complete structure presented here highlights the extent to which such connectors are critical to NPC integrity. Remarkably, flexible connectors run the entire length of each spoke, tying together every major element in the NPC (Fig. 5c, e). They link the periphery and outer rings to the inner rings, both inner rings to the pore membrane and adjacent spokes to one another. We identified two types of connectors

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

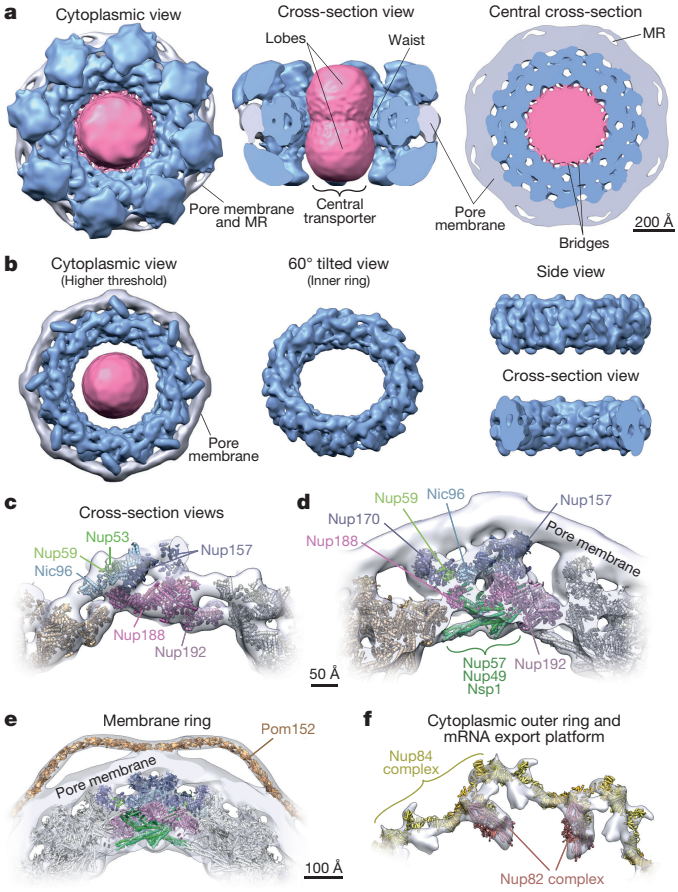


Figure 3 | Morphology of the NPC. a, **b**, Cryo-ET map of the NPC: core scaffold, blue; membrane region, grey; central transporter, pink. MR: membrane ring. In **a**, cytoplasmic top view (left); cross-section side view (middle); and central cross-section top view (right). **b**, Cryo-ET map from **a** presented at a higher threshold. Top view (left); 60°-tilted view of the inner ring (middle); and side (right, top) and cross-section views (right, bottom) of the inner ring. Scale bar, 200 Å. **c**–**f**, Cross-section views show a representative structure embedded within the cryo-ET density (grey), presented with different filtering and thresholding to show the good fit to the cryo-ET map in the inner ring (**c**, **d**), the membrane ring (**e**) and the cytoplasmic outer ring and mRNA export platform (**f**). Nups indicated as in Fig. 4. Scale bars, 50 Å (**c**, **d**) and 100 Å (**e**, **f**).

(Supplementary Results and Discussion). First, there are vertical connections, aligned parallel to the cylindrical axis of the NPC and constituting the main anchor points between the export platform and the inner ring. On the nuclear side, similar connections are present between the nuclear basket and the inner ring, with an additional connection between the basket and outer ring (Figs 4d, 5e). Second, there are horizontal flexible connectors that link the central channel to the pore membrane between adjacent spokes (Fig. 5e). Collectively, these flexible connectors may serve to allow limited movement of the more rigid modules with respect to one another, thereby providing the NPC with another degree of flexibility in response to deformation[28].

Organization of the transport machinery

Despite its critical function, the central gating machinery has largely been excluded from recent NPC maps[6,7,29] and its properties have remained controversial[4]. Here we confirm the existence of a large central transporter with two high-density 'lobes' connected by a narrower 'waist' of lower density[5] (Figs 3, 6 and Extended Data Figs 4–6). This central transporter comprises multiple FG repeats that account for about 9 MDa, together with approximately 26 MDa of NTFs and their cargoes caught in transit (though they may be somewhat averaged out in our map) (Fig. 1b, c and Extended Data Fig. 3c). Indeed, even after isolation, each NPC carried 10–80 copies of each of the observed NTFs[30], reflecting the huge and varied transport flux through NPCs.

The localization of FG-repeat anchor points reveals three patterns. First, a vertical path is formed along each spoke by a continuous array of FG repeats (Fig. 6a, c and Extended Data Fig. 11b–d). By binding to these repeats, NTFs may follow these paths across the entire NPC. Second, the FG anchor points of Nsp1–Nup57–Nup49 form a central ring on the equator of the NPC (Fig. 6b). Thin bridges in our cryo-ET map coincide with the location of these FG anchor points, which indicates that these bridges comprise the FG repeats themselves, emanating from their anchor points (Fig. 6b). Third, the structured regions of the NPC largely direct the FG-repeat regions inwards toward the axis of the central channel (Fig. 6a), instead of projecting from the NPC towards the cytoplasm and nucleoplasm as they are often represented[9]. This geometry generates a highly concentrated (25–150 mM) and dynamic FG-repeat phase through which cargo-carrying NTFs readily pass, facilitated by their specific FG interactions, whereas nonspecific macromolecular diffusion is hindered by this same dense phase[31].

It has previously been suggested that the two main types of FG repeat ('Phe-X-Phe-Gly/Phe-Gly' (FXFG/FG) and 'Gly-Leu-Phe-Gly' (GLFG)) are segregated in the NPC to define functionally distinct zones of the gating machinery[32]. Consistent with this, and with the known role of FXFG/FG-type repeats in docking RNAs during export[33], we find that FXFG/FG-type repeats are enriched in the nuclear and cytoplasmic peripheries of the NPC, where the RNA-associating export platform and basket reside (Fig. 6d and Extended Data Fig. 11c). By contrast, the GLFG-type repeats are enriched in regions adjacent to the inner ring and near the cytoplasmic entrance to the central channel. This cytoplasmic localization coincides with the position of FG repeats that are most important for limiting the passage of nonspecific macromolecules (Fig. 6e and Extended Data Fig. 11d), and is consistent with the known role of GLFG-type repeats in maintaining the passive permeability barrier[34,35].

Evolutionary origin and diversity of the NPC

NPCs share architectural features with vesicle coating complexes (Extended Data Fig. 10), which led us to hypothesize that they share a common evolutionary ancestor, the 'protocoatomer'[36]. Two major families of coating complexes exist: COPI/clathrin and COPII, each of which have discrete vesicle recognition and trafficking roles[37,38]. We find both COPI/clathrin-like and COPII-like features in the NPC, suggesting that ancestral COPI and COPII coating families evolved first and were followed by the NPC, which may have evolved through a partnership of COPI and COPII coats. This hypothesis implies that the nucleus was a later addition in the evolutionary path of the first eukaryotes (Supplementary Results and Discussion).

Despite substantial conservation of some elements of NPC architecture, other elements can vary widely between species. Generally, the inner ring appears most conserved[39,40], as is seen in a comparison of our yeast structure with that of the human scaffold[6], although the latter is more expanded (Extended Data Fig. 12). By contrast, peripheral elements exhibit considerable lineage-specific losses and duplications[40,41]. In yeast, each outer ring is formed by 8 copies of the Nup84 complex (Figs 3f, 4), whereas in vertebrates each outer ring contains 16 copies of the equivalent Y-shaped complex arranged in two interlocked rings[17,42]. Moreover, we see neither an additional copy of Nup157 or Nup170 connecting the outer and inner rings nor Nup188 or Nup192 in the outer rings, as indicated in humans[6] (Fig. 4d–f and Extended Data Fig. 12a). Another previous model assumed that fungal and human core scaffolds have essentially identical structures[7]. Our data invalidate this assumption (Fig. 1, Supplementary Table 2a and Supplementary Results and Discussion), as well as an earlier 'fencepost' model[43].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

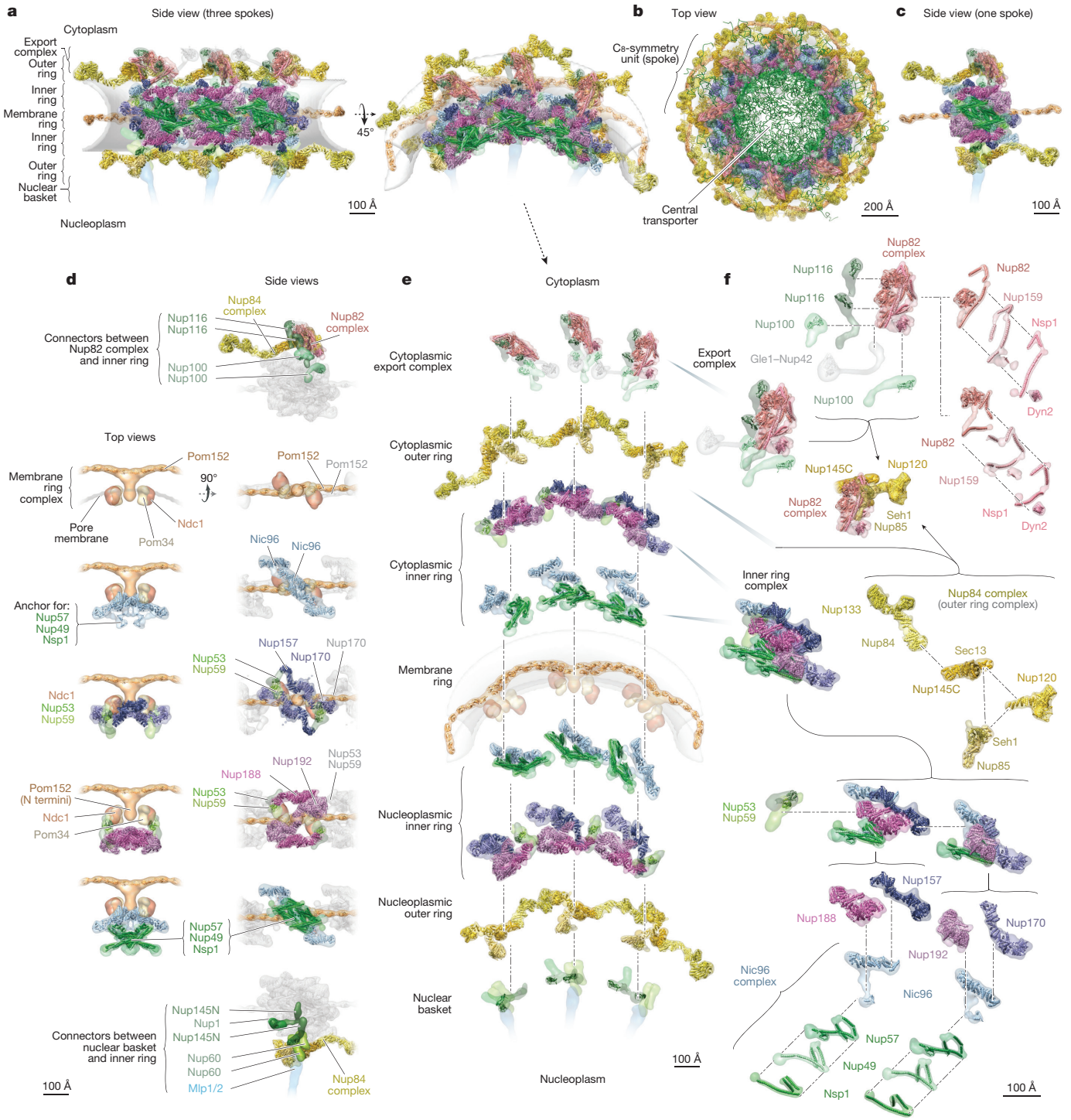


Figure 4 | Structural dissection of the NPC. Complete structure of the NPC and its components shown in different orientations, with a model of the pore membrane region shown in grey (Supplementary Videos 1–3). For each Nup, the localization probability density of the ensemble of structures is shown with a representative structure from the ensemble embedded within it (Supplementary Table 2). **a**, Two views of three consecutive NPC spokes (C_8-symmetry units), showing how the coaxial outer, inner and membrane rings run continuously between spokes. **b**, Cytoplasmic top view of the complete NPC structure with modelled FG-repeat regions (green). **c**, Side view of a single NPC spoke. **d**, Relative position of major NPC components and connections both within and between spokes. Top (left) and side (right) views are shown. The membrane ring (beige) is included for reference. Flexible connectors between outer and inner rings are shown in the top and bottom panels, with the inner and membrane rings shown as faded grey densities. **e**, Exploded view of three consecutive spokes, spanning from the cytoplasmic face (top) to the nuclear face (bottom), with dashed lines connecting neighbouring rings. **f**, Cytoplasmic mRNA export complex (top), the Nup84 complex (centre) and the inner ring complex, including the Nic96 complex (bottom), from a single spoke. The complexes are shown as an exploded diagram, with dashed lines connecting neighbouring components.

In summary, there is no single universal NPC structure; instead, similar structural elements are used in somewhat different arrangements to generate many lineage-specific adaptations.

Conclusions

We have described the structure of the entire yeast NPC at sub-nanometre precision. At the heart of the inner ring, rigid diagonal columns

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

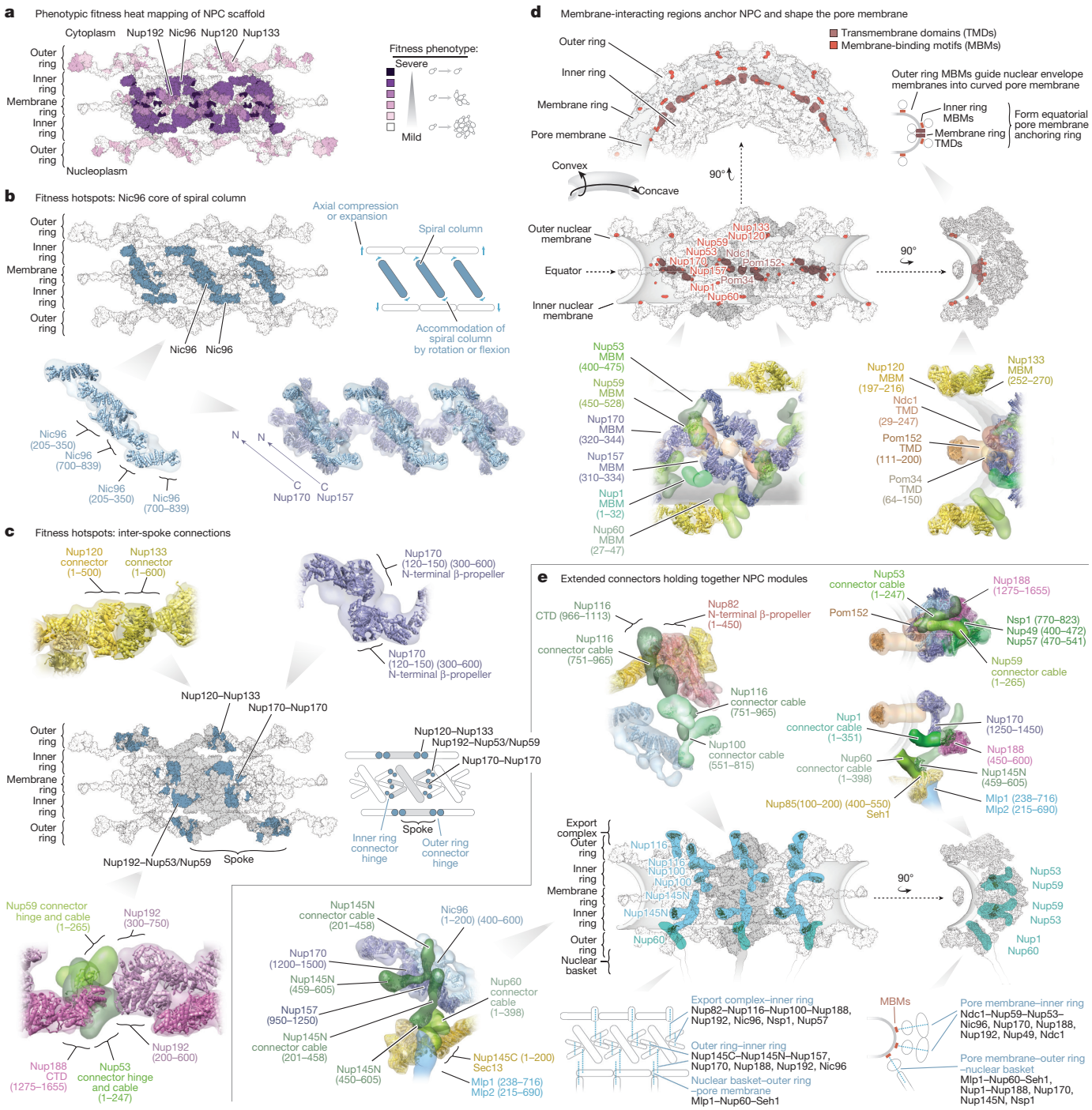


Figure 5 | Key NPC architectural features and principles. a, Severity of fitness defects, indicated in increasing intensity of purple shades for specific truncations of nucleoporins (Extended Data Fig. 9), mapped onto three spokes of the NPC. **b**, **c**, Structures corresponding to the position of the most severe defects (dark blue). In top panels in **b**, diagonally oriented columns reinforcing the core scaffold may accommodate NPC compression and expansion (diagram to right). Bottom, molecular details of Nup arrangement (relevant residue numbers indicated). In middle panels in **c**, the position of hotspots coincides with spoke-to-spoke connections. Central spoke, grey; flanking spokes, white (schematic, middle right). Top and bottom, molecular details of spoke-to-spoke connector hinges. **d**, Top left and centre left, three spokes shown as top and front views; centre right, one spoke in side view. Schematic indicates convex and concave pore membrane curvatures. Positions of TMDs and MBMs are depicted and their proteins are labelled in brown and orange, respectively. Top right, diagrammatic side view showing how the MBMs and TMDs curve the pore membrane. Bottom, molecular details of the Nups containing the TMDs and MBMs. **e**, Second row left, three spokes in front view, showing how vertical connector Nups (cyan) spanning from the cytoplasmic to nuclear sides of the NPC connect the rings. Second row right, one spoke in side view, showing how horizontal connector Nups (aquamarine) connect modules spanning from the pore membrane to the central channel. First row and bottom row left show molecular details of the connectors within the NPC. Bottom row centre and right, diagrammatic views of the connectors depicted as blue dotted lines; modules connected labelled in blue; major Nups being contacted by connectors listed in grey.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

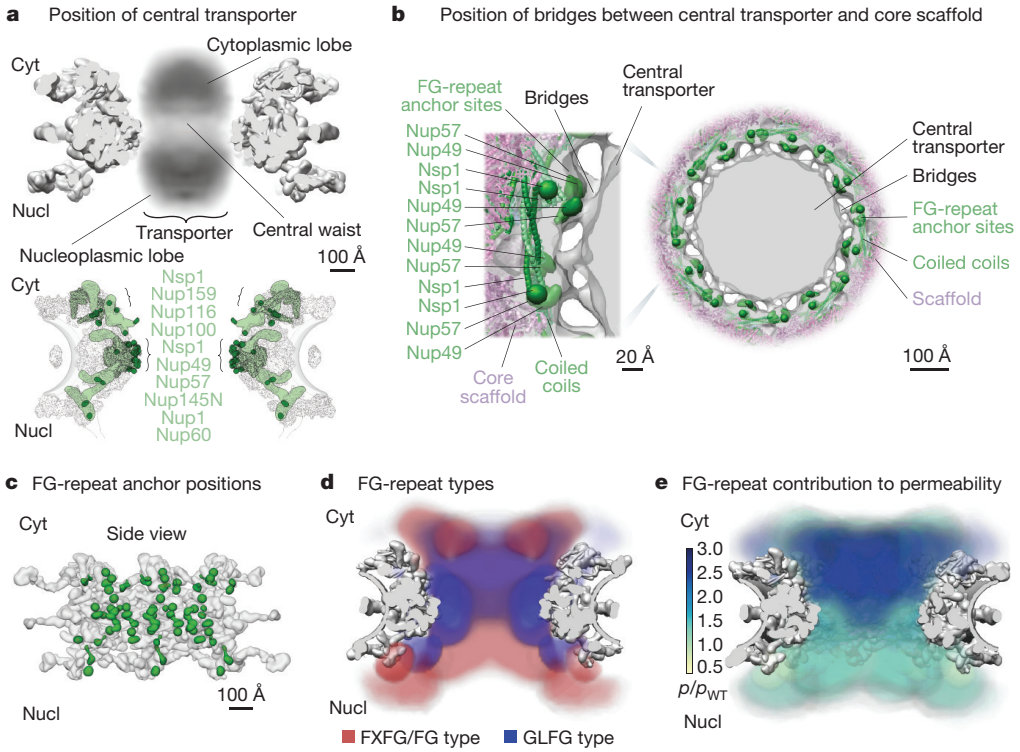


Figure 6 | The distribution of FG repeats informs the NPC transport gating mechanism. a, Central transporter density from the cryo-ET map (Fig. 3) is shown within the structure of the NPC scaffold (grey) (top). Features of the central transporter are indicated. Anchors (light green) in FG Nups largely direct the FG-repeat emanating points (dark green) towards the central channel (bottom). Scale bar, 100 Å. **b**, Central cross-section of the cryo-ET map (grey) with embedded representative NPC structure (Fig. 4), showing the central transporter and the bridges connecting it to the core scaffold in top view (scale bar, 100 Å), with a magnified view of one spoke on the left (scale bar, 20 Å). The anchor points for the FG repeats of Nup49, Nup57 and Nsp1 are depicted as green densities. **c**, Position of FG-repeat anchor points (green) within a side view of three spokes of the scaffold (grey). Scale bar, 100 Å. **d**, Heat mapping of repeats of FXFG/FG type (red) and GLFG type (blue), from Brownian dynamics simulations (Methods), showing partitioning to different regions of the central channel. Scale bar, 100 Å. **e**, Heat mapping of the effect of FG-repeat region truncations on NPC permeability; the severity of the permeability defect[34] (measured as permeability relative to permeability in wild type, (p/p_{WT})) is indicated in increasing shades from minor (light green) to severe (dark blue). Cyt, cytoplasm; nucl, nucleoplasm. Scale bar, 100 Å.

reinforce the structural integrity of the NPC. Membrane-binding and transmembrane Nups are strategically placed throughout the core scaffold to stabilize pore membrane curvature and clamp the NPC to the nuclear envelope. Connectors run the length of each spoke, flexibly tying together all the major modules in the NPC. The architecture of the NPC is reminiscent of a suspension bridge, in which rigid supporting columns are firmly anchored to a substrate and flexible suspension cables connect the columns and roadway to provide a strong and resilient structure. We show that most FG Nup anchor points face inwards toward the NPC central channel to generate a highly concentrated milieu of FG repeats: FXFG/FG repeats form mRNA docking 'traps' at the entrance and exit of the channel, and GLFG repeats help form a cytoplasmically biased permeability barrier.

Despite differences, yeast and human NPCs retain a notable degree of structural conservation (Extended Data Fig. 12a, b). As a result, many of the conclusions drawn here should be applicable to the human NPC. To illustrate this point, we mapped the positions of yeast homologues of the oncogenic hotspot human Nup214, Nup98 and Tpr (ref. 2) (Extended Data Fig. 12c). Rather than being randomly scattered, these positions coincide with RNA-binding platforms on the cytoplasmic and nucleoplasmic faces of the NPC as well as with several critical connectors and associated FG regions. This conservation suggests that alterations in RNA export, and changes in NPC architecture induced by defective connectors, may underlie the altered behaviour of NPCs in cancer cells. Thus, our yeast structure provides a roadmap with the potential to advance our understanding of NPC physiology and nuclear transport in general.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 June 2017; accepted 6 February 2018.
Published online 14 March 2018.

1. Ptak, C., Aitchison, J. D. & Wozniak, R. W. The multifunctional nuclear pore complex: a platform for controlling gene expression. *Curr. Opin. Cell Biol.* **28,** 46–53 (2014).
2. Nofrini, V., Di Giacomo, D. & Mecucci, C. Nucleoporin genes in human diseases. *Eur. J. Hum. Genet.* **24,** 1388–1395 (2016).
3. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450,** 695–701 (2007).
4. Stanley, G. J., Fassati, A. & Hoogenboom, B. W. Biomechanics of the transport barrier in the nuclear pore complex. *Semin. Cell Dev. Biol.* **68,** 42–51 (2017).
5. Akey, C. W. & Goldfarb, D. S. Protein import through the nuclear pore complex is a multistep process. *J. Cell Biol.* **109,** 971–982 (1989).
6. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352,** 363–365 (2016).
7. Lin, D. H. *et al.* Architecture of the symmetric core of the nuclear pore. *Science* **352,** aaf1015 (2016).
8. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450,** 683–694 (2007).
9. Fernandez-Martinez, J. *et al.* Structure and function of the nuclear pore complex cytoplasmic mRNA export platform. *Cell* **167,** 1215–1228.e25 (2016).
10. Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics* **13,** 2927–2943 (2014).
11. Briggs, J. A. Structural biology *in situ*—the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23,** 261–267 (2013).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

12. Kim, S. J. *et al.* Integrative structure–function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. *Mol. Cell. Proteomics* **13,** 2911–2926 (2014).
13. Fernandez-Martinez, J. *et al.* Structure-function mapping of a heptameric module in the nuclear pore complex. *J. Cell Biol.* **196,** 419–434 (2012).
14. Herricks, T. *et al.* One-cell doubling evaluation by living arrays of yeast, ODELAY! *G3* **7,** 279–288 (2017).
15. Aitchison, J. D., Rout, M. P., Marelli, M., Blobel, G. & Wozniak, R. W. Two novel related yeast nucleoporins Nup170p and Nup157p: complementation with the vertebrate homologue Nup155p and functional interactions with the yeast nuclear pore-membrane protein Pom152p. *J. Cell Biol.* **131,** 1133–1148 (1995).
16. Fischer, J., Teimer, R., Amlacher, S., Kunze, R. & Hurt, E. Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol.* **22,** 774–781 (2015).
17. von Appen, A. *et al. In situ* structural analysis of the human nuclear pore complex. *Nature* **526,** 140–143 (2015).
18. Marelli, M., Lusk, C. P., Chan, H., Aitchison, J. D. & Wozniak, R. W. A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope. *J. Cell Biol.* **153,** 709–724 (2001).
19. Vollmer, B. *et al.* Dimerization and direct membrane interaction of Nup53 contribute to nuclear pore complex assembly. *EMBO J.* **31,** 4072–4084 (2012).
20. Seo, H. S. *et al.* Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc. Natl Acad. Sci. USA* **106,** 14281–14286 (2009).
21. Drin, G. *et al.* A general amphipathic α-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14,** 138–146 (2007).
22. Mészáros, N. *et al.* Nuclear pore basket proteins are tethered to the nuclear envelope and can regulate membrane curvature. *Dev. Cell* **33,** 285–298 (2015).
23. Upla, P. *et al.* Molecular architecture of the major membrane ring component of the nuclear pore complex. *Structure* **25,** 434–445 (2017).
24. Meinema, A. C. *et al.* Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. *Science* **333,** 90–93 (2011).
25. Knockenhauer, K. E. & Schwartz, T. U. The nuclear pore complex as a flexible and dynamic gate. *Cell* **164,** 1162–1171 (2016).
26. Folkmann, A. W., Noble, K. N., Cole, C. N. & Wente, S. R. Dbp5, Gle1–IP6 and Nup159: a working model for mRNP export. *Nucleus* **2,** 540–548 (2011).
27. Saroufim, M. A. *et al.* The nuclear basket mediates perinuclear mRNA scanning in budding yeast. *J. Cell Biol.* **211,** 1131–1140 (2015).
28. Meseroll, R. A. & Cohen-Fix, O. The malleable nature of the budding yeast nuclear envelope: flares, fusion, and fenestrations. *J. Cell. Physiol.* **231,** 2353–2360 (2016).
29. Eibauer, M. *et al.* Structure and gating of the nuclear pore complex. *Nat. Commun.* **6,** 7532 (2015).
30. Paradise, A., Levin, M. K., Korza, G. & Carson, J. H. Significant proportions of nuclear transport proteins with reduced intracellular mobilities resolved by fluorescence correlation spectroscopy. *J. Mol. Biol.* **365,** 50–65 (2007).
31. Adams, R. L. & Wente, S. R. Uncovering nuclear pore complexity with innovation. *Cell* **152,** 1218–1221 (2013).
32. Patel, S. S., Belmont, B. J., Sante, J. M. & Rexach, M. F. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* **129,** 83–96 (2007).
33. Adams, R. L., Terry, L. J. & Wente, S. R. Nucleoporin FG domains facilitate mRNP remodeling at the cytoplasmic face of the nuclear pore complex. *Genetics* **197,** 1213–1224 (2014).
34. Timney, B. L. *et al.* Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* **215,** 57–76 (2016).
35. Yamada, J. *et al.* A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics* **9,** 2205–2224 (2010).
36. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2,** e380 (2004).
37. Faini, M., Beck, R., Wieland, F. T. & Briggs, J. A. Vesicle coats: structure, function, and general principles of assembly. *Trends Cell Biol.* **23,** 279–288 (2013).
38. Rout, M. P. & Field, M. C. The evolution of organellar coat complexes and organization of the eukaryotic cell. *Annu. Rev. Biochem.* **86,** 637–657 (2017).
39. Obado, S. O. *et al.* Interactome mapping reveals the evolutionary history of the nuclear pore complex. *PLoS Biol.* **14,** e1002365 (2016).
40. Iwamoto, M. *et al.* Compositionally distinct nuclear pore complexes of functionally distinct dimorphic nuclei in the ciliate *Tetrahymena*. *J. Cell Sci.* **130,** 1822–1834 (2017).
41. Obado, S. O., Field, M. C. & Rout, M. P. Comparative interactomics provides evidence for functional specialization of the nuclear pore complex. *Nucleus* **8,** 340–352 (2017) .
42. Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155,** 1233–1243 (2013).
43. Debler, E. W. *et al.* A fence-like coat for the nuclear pore membrane. *Mol. Cell* **32,** 815–826 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Webb (UCSF) for help with the Integrative Modelling Platform, the Rockefeller University Outreach Program for support for A.S.C., the NYULMC OCS Microscopy Core, K. Uryu and the EMRC Resource Center (Rockefeller University) for assistance with negative-stain electron microscopy, F. Alber, M. C. Field, N. Ketaren, S. Obado, R. Hayama and D. Simon for feedback and critical reading of the manuscript, and L. Herlands for support and encouragement. The work was supported by a NSF GRF 1650113 (I.E.C.), a NSF grant CHE-1531823 (M.F.J.), the SIMR (J.L.G.), NIH grants R01 GM080477 (J.L.G.), U54 GM103511 (B.T.C., A.S., J.D.A. and M.P.R.), R01 GM112108 (M.P.R. and J.D.A.), P41 GM109824 (M.P.R., A.S., J.D.A. and B.T.C.), P50 GM076547 (J.D.A.), R01 GM063834 (C.W.A.), R01 GM080139 (S.J.L.), P41 GM103314 (B.T.C.), R01 GM083960 (A.S.) and U54 DK107981 (M.P.R. and J.D.A.). We are grateful for the support provided by G. Blobel, who inspired the work presented here.

Author Contributions The order of first co-authors was determined through a random selection process. I.N., J.F.-M., A.S.C., R.W. and M.P.R. performed the affinity purifications; W.Z., J.F.-M., R.W., R.M., E.Y.J., M.P.R. and B.T.C. performed the quantitative mass spectrometry; M.S., B.D.S., J.R.U. and J.L.G. performed the calibrated imaging; J.A.H.. B.T.C. and M.F.J. performed the charge detection mass spectrometry; Y.S., J.F.-M., R.W., I.N., J.W. and B.T.C. performed the chemical crosslinking with mass spectrometry; C.W.A., S.J.L., I.N., Z.Y. and M.J.d.l.C. performed the cryo-ET; S.J.K. performed the small-angle X-ray scattering; T.H., J.F.-M. and J.D.A. performed the phenotypic profiling; P.U. and D.L.S. performed the negative-stain electron microscopy; S.J.K., B.R., I.E.C., R.P., I.E., C.H.G. and A.S. performed the integrative structure computations; S.J.L., C.W.A., B.T.C., A.S. and M.P.R. supervised the project; S.J.K., J.F.-M., I.N., Y.S., W.Z., B.R., S.J.L., C.W.A., B.T.C., A.S. and M.P.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.P.R. (rout@rockefeller.edu), B.T.C. (chait@rockefeller.edu), A.S. (sali@salilab.org), C.W.A. (cakey@bu.edu) or S.J.L. (sludtke@bcm.edu).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

1. Yeast strains and materials. All *S. cerevisiae* strains used in this study are listed in Supplementary Table 5, with the exception of the Nup84 complex truncation mutants[13] and the Pom152 truncation mutants[23]. Unless otherwise stated, strains were grown at 30 °C in YPD medium (1% yeast extract, 2% bactopeptone and 2% glucose). The diploid *S. uvarum* strain (ATCC 9080) was grown and processed for nuclear envelope purification as previously described[44].

The following materials were used in this study: Dynabeads M-270 Epoxy (143.02D; Invitrogen); rabbit IgG (55944; MP Biomedicals); protease inhibitor cocktail (P-8340; Sigma-Aldrich); and Solution P (2 mg Pepstatin A, 90 mg PMSF, and 5 ml of absolute ethanol).

2. Immuno-purification of the endogenous *S. cerevisiae* NPC. An immuno-purification protocol for the isolation of endogenous whole NPCs from *S. cerevisiae* was developed using previously published methodology[3,45–49]. *S. cerevisiae* Mlp1-, Nup84- or Nup82-encoding genes were genomically tagged with PrA preceded by the human rhinovirus 3C protease (PPX) target sequence (GLEVLFQGPS). Cells were grown in YPD medium at 30 °C until early log phase ($\sim 2 \times 10^7$ cells/ml), collected, frozen in liquid nitrogen and cryogenically lysed in a planetary ball mill PM 100 (Retsch) (http://lab.rockefeller.edu/rout/protocols). Frozen cell powder was resuspended in 9 volumes of resuspension buffer (20 mM HEPES–KOH pH 7.4, 50 mM potassium acetate, 20 mM NaCl, 2 mM $MgCl_2$, 0.5% (w/v) Triton X-100, 0.1% (w/v) Tween-20, 1 mM DTT, 10% (v/v) glycerol, 1/500 (v/v) protease inhibitor cocktail (Sigma)). Cell lysate was clarified by centrifugation at 2,500 RCF for 5 min followed by filtration through 1.6-μm filters (Whatman glass microfibre syringe filters). Magnetic beads (Invitrogen) conjugated to rabbit IgG antibodies (http://lab.rockefeller.edu/rout/protocols) were added to the clarified cell lysate at a concentration of 50 μl slurry per 1 g of frozen cell powder and incubated for 30 min at 4 °C. Beads were washed once with 1 ml of elution buffer without protease inhibitors (20 mM HEPES–KOH pH 7.4, 50 mM potassium acetate, 20 mM NaCl, 2 mM $MgCl_2$, 0.1% (w/v) Tween-20, 1 mM DTT, 10% (v/v) glycerol). For native elution of the complex, the desired volume of elution buffer with PreScission protease (GE Healthcare) (1/15 (v/v)) was added to the beads and incubated for 45 min at 4 °C. A magnet was used to remove the beads and collect the supernatant. Beads were subsequently washed with the desired volume of elution buffer containing 1/500 (v/v) protease inhibitor cocktail (Sigma). The total elution volume was centrifuged at 20,000*g* for 5 min to remove the residual magnetic beads. Typical yield of the immuno-purification is ~4 μg of isolated NPCs per 1 g frozen cell powder (see Extended Data Fig. 2b for SDS-PAGE analysis; for gel source data, see Supplementary Fig. 1).

3. Mass and stoichiometry of the native *S. cerevisiae* NPC. Quantification of the absolute stoichiometry of each nucleoporin in the native NPCs was performed using a strategy that combined several orthologous methods (Extended Data Fig. 2a): (1) use of synthetic concatemers of tryptic peptides or QconCATs[50] to define the relative stoichiometry of each component by quantitative mass spectrometry in affinity-captured NPCs; (2) *in vivo* calibrated imaging analysis of GFP-tagged Nups[51], to quantify the absolute copy number per NPC of Nups selected to represent each major module of the NPC; and (3) charge detection mass spectrometry to measure the total mass of affinity-captured NPCs[52]. For the calculation of the integrative NPC structure, the final copy numbers were rounded to fit the known NPC C_8-symmetry and these values are indicated in Supplementary Table 2a.

3.1 NPC QconCAT design and purification. Mass spectrometry quantification of the relative amounts of each nucleoprotein in the purified NPC complex was performed using two specifically designed, heavy-labelled synthetic internal standards or QconCATs[50,53] (Extended Data Fig. 2d, e) formed by concatenated quantotypic nucleoporin peptides. To minimize the potential effect of having different residues flanking the trypsin cleavage site on the cleavage efficiency, we included the native three-residue flanking sequences framing the trypsin cleavage site for each peptide[54]. For QconCAT-A (Extended Data Fig. 2d), two peptides for each of the nucleoporins and one peptide for *Staphylococcus aureus* protein A and *Aequorea victoria* GFP proteins were selected (Supplementary Table 7) on the basis of their favourable signal responses in liquid chromatography–mass spectrometry analyses of NPC samples and by fitting to the following criteria (when possible): (1) the native three-residue flanking sequences at both sides of the trypsin cleavage sequence do not contain Lys or Arg; (2) avoid the presence of Cys or Met residues within the peptide; (3) avoid the presence of potential internal trypsin cleavage sites (Lys or Arg residues); (4) peptides should be less than 3,000 Da (small size); and (5) avoid peptides showing obvious interferences from co-eluting peptides during the liquid chromatography separation for mass spectrometry analyses. QconCAT-B included two quantotypic peptides for Nup159, Mlp2, Nup192, Nup84, Nup85, Nup120, Nup49, Nup57, Pom152 and Nic96, and the same GFP peptide as in QconCAT-A (Supplementary Table 7). As an internal control, both QconCAT-A and -B included the same peptides for Nic96, Pom152 and GFP. Each synthetic gene was designed by concatenation of the sequences encoding the selected peptides and addition of a 6× His C-terminal tag (Extended Data Fig. 2d). A 3×FLAG peptide was also included at the N terminus of QconCAT-A, resulting in a protein of 148.2 kDa. The *Escherichia coli* codon optimized sequences were cloned into: (1) plasmid pET15-b (as a NcoI–XhoI fragment) in the case of QconCAT-A; and (2) pGEX6p-1 (as a BamHI–XhoI fragment) in the case of QconCAT-B, resulting in the expression of a 68.1 kDa protein with a N-terminal GST tag that was mainly used as a sacrificial peptide[55]. The QconCAT proteins were expressed by growing 300 ml of BL21 *E. coli* cells at 37 °C to $OD_{600} = 0.6$ in minimal M9 medium without ammonium chloride[50,53] supplemented with light amino acids and 0.5 mg/ml of heavy arginine and lysine (L-arginine:HCl $^{13}C_6$; L-lysine:2HCl $^{13}C_6$, Cambridge Isotope Laboratories). IPTG (1 mM) was used to induce expression of the constructs for 3 h at 37 °C. Collected cells were processed using BugBuster Extraction Reagent (Novagen), as indicated by the manufacturer, to isolate the inclusion bodies where the QconCAT protein is accumulated. The full-length QconCAT-A was then purified by resuspending the inclusion bodies pellet in binding buffer (20 mM sodium phosphate pH 7.4, 45 mM imidazole, 500 mM NaCl, 6 M guanidinium chloride, 10 mM TCEP (0.5 M Bond-Breaker TCEP solution, Thermo Fisher Scientific), 1/500 protease inhibitor cocktail (Roche)) and passed through an equilibrated His-Trap HP (GE Healthcare) at room temperature. The retained NPC QconCAT-A was then eluted in 20 mM sodium phosphate pH 7.4, 500 mM imidazole, 500 mM NaCl, 6 M guanidinium hydrochloride, 1 mM TCEP, 1/500 protease inhibitor cocktail. One-hundred-microlitre aliquots of the resulting elution were precipitated to eliminate the guanidinium hydrochloride by adding ice-cold ethanol to a final concentration of 90% and incubating the samples at −20 °C for 2 h. Samples were then centrifuged for 10 min at 14,000 r.p.m. and 4 °C to pellet the precipitated protein. The resulting pellet was washed with ice-cold 90% ethanol and allowed to air-dry until most of the liquid was evaporated, leaving a wet pellet. These pellets were solubilized with 5% SDS, 500 mM Tris–HCl pH 8.0, 5 mM TCEP buffer, by incubating for 5 min at room temperature and 5 min at 72 °C and centrifuged for 10 min at 14,000 r.p.m. at room temperature. The supernatants were recovered and two of them combined and injected into a TSKgel G4000SWxl size-exclusion column (TOSOH Bioscience) coupled to a TSKgel SWxl guard column (TOSOH Bioscience), pre-equilibrated in running buffer (40 mM HEPES–KOH pH 7.0, 150 mM NaCl, 0.1% SDS, 5 mM TCEP, 1 mM EDTA). Two-hundred-microlitre fractions were collected and analysed by SDS-PAGE to detect the presence of the QconCAT-A peak. Fractions containing the full-length pure protein were supplemented with a final 20% glycerol (v/v), aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C for further use. In the case of QconCAT-B, the protein was purified using His-Trap HP and the elution precipitated and prepared as described for QconCAT-A. The resulting sample was injected into a TSKgel Super SW3000 size-exclusion column (TOSOH Bioscience) pre-equilibrated in running buffer (40 mM HEPES–KOH pH 7.0, 150 mM NaCl, 0.1% SDS, 5 mM TCEP, 1 mM EDTA). One-hundred-microlitre fractions were collected and analysed by SDS-PAGE to detect the presence of the QconCAT-B peak. Fractions containing the full-length pure protein were stored as indicated for QconCAT-A.

For the quantitative mass spectrometry analysis, the native NPCs from *S. cerevisiae* PPX–PrA-tagged haploid strains were affinity captured as described above, or purified as enriched NPCs from a diploid *S. uvarum* strain using a subfractionation method previously described in detail[44,56–58] (http://lab.rockefeller.edu/rout/protocols), using 0.035 mg heparin per mg of fraction protein. For affinity-captured NPCs, the natively eluted NPCs (5 μg) were concentrated by pelleting at 40,000 r.p.m. for 20 min at 4 °C in a TLA 55 rotor (Beckman). In the case of subfractionation-enriched NPCs, a volume of the 1.45 M:1.85 M sucrose gradient fraction that contained an estimated 5 μg of NPCs was diluted 1/5 (v/v) in bt-DMSO buffer (10 mM bis-Tris-HCl pH 6.5, 0.1 mM $MgCl_2$, 20% DMSO) and pelleted at 15,000 r.p.m. for 450 min at 4 °C in a TLA 55 rotor (Beckman). For in-solution mass spectrometry analysis of subfractionation-enriched NPCs, 0.1 μg of QconCAT-A were immobilized on Dynabeads His-Tag Isolation and Pulldown resin (Thermo Fisher Scientific) pre-equilibrated in binding buffer (20 mM HEPES, 150 mM NaCl, 8 M urea, 5 mM TCEP). The purified protein sample was incubated with the resin for 20 min at room temperature, and washed with binding buffer 5 × 200 μl to eliminate residual SDS; in-solution and in-gel analyses showed consistent results (not shown), so most of the further analyses were performed in-gel to improve consistency, speed and throughput. For the solid-state in-gel mass spectrometry analyses, pelleted NPCs were solubilized in 10 μl of 0.5 M Tris-HCl pH 8.0, 5% SDS by incubating at 72 °C for 5 min and then diluted 1:1 with 20% glycerol, 50 mM TCEP, 0.5 mM EDTA, 0.05% (w/v) bromophenol blue. Approximately equimolar amounts of 0.1 μg of purified QconCAT-A or 0.045 μg of purified QconCAT-B were added to each 5-μg NPC sample. Samples were then incubated at 72 °C for 10 min,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

cooled to room temperature and treated with a final 30 mM of iodoacetamide (Sigma), at room temperature in the dark for 30 min. Samples was then loaded into a 4% (37.5:1) stacking acrylamide SDS-PAGE gel prepared in-house. The resulting bands, containing a mixture of whole NPCs and QconCAT protein (labelled with a stable isotope), were excised and processed for quantitative mass spectrometry analyses.

3.2 Mass spectrometry characterization of QconCAT labelled with a stable isotope. The mass of purified intact QconCAT-A protein, labelled with a stable isotope, was analysed by matrix-assisted laser desorption/ionization (MALDI) (Extended Data Fig. 2e) on a JEOL JMS-S3000 SpiralTOF mass spectrometer using the ultra-thin-layer sample preparation method[59,60] in which α-cyano-4-hydroxycinnamic acid (Sigma) was used as the matrix. The mass of QconCAT-A was internally calibrated with horse myoglobin. Mass calibration and background subtraction were carried out with the JEOL msTornado control software, and additional analyses were carried out with the MoverZ software[61]. The QconCAT-A protein was also characterized by peptide mapping, in which tryptic peptides from in-gel digestion were loaded onto a PicoFrit column (New Objective) with an integrated emitter tip (360-mm O.D., 50-mm I.D., 10-mm tip) self-packed with 6 cm of reverse-phase C18 material (ReproSil-Pur C18-AQ, 3-mm beads, Dr. Maisch GmbH), and analysed with a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific), with a Agilent 1200 series HPLC system (Agilent) and a micro electrospray source built in-house. The purified QconCAT-B was characterized by peptide mapping on a Thermo Orbitrap Fusion mass spectrometer, with a Thermo Easy-nLC 1000 HPLC and a Thermo Easy-Spray electrospray source.

3.3 Stoichiometry quantification of NPC using QconCAT and by mass spectrometry. Mixtures of yeast NPC proteins and QconCAT, labelled with a stable isotope, were enzymatically digested either in solution in the presence of urea or inside a SDS-PAGE gel matrix. For in-solution digestion, a mixture of the NPCs and immobilized QconCATs on His-Dynabeads were sequentially digested at room temperature by Endoproteinase Lys-C in 8 M urea for 66 h and by trypsin in 2 M urea for 3 h. For in-gel digestion, proteins in the gel matrix were digested in 100 mM Tris-HCl at room temperature either sequentially by 0.25–2 μg Endoproteinase Lys-C for 66 h and by 3–25 μg trypsin for 3 h, or—in later experiments—by 25 μg trypsin alone for 3 h. The resulting peptides were analysed in duplicate by liquid chromatography–mass spectrometry using a Thermo Fusion or a Thermo Q Exactive Plus mass spectrometer, with a Thermo Easy-nLC 1000 HPLC and a Thermo Easy-Spray electrospray source. The ratios of light nucleoporin (L) to heavy QconCAT proteins (H) for standard peptides were obtained using MaxQuant[62], complemented with manual determination.

We incorporated two standard peptides from each nucleoporin into the QconCAT standard to enable us to check for internal consistency of the measured L/H ratios for each nucleoporin. Our check required that the relative standard deviations of L/H ratios for two standard peptides from two duplicate liquid chromatography–mass spectrometry runs—that is, for a total of four measurements per nucleoporin—be ≤ 25%. When deriving relative stoichiometry for any given preparation of NPCs analysed in different replication experiments, we corrected for variations in the mixing ratio of light nucleoporins and heavy QconCAT proteins by scaling the measured L/H ratios to minimize the sum of the relative standard deviations of the L/H ratios over all nucleoporins. The resulting scaled L/H ratios from different experiments were used to derive the average L/H ratios and standard deviations. To assay for potential nucleoporin stoichiometry bias arising from capture through particular affinity handles, we used stable isotope labelling with amino acids in cell culture followed by mass spectrometry (SILAC–MS) analysis of these preparations versus the nuclear envelope preparation. We performed $n = 2$ or 3 technical and 2 biological replicas for NPCs purified by subfractionation procedures from a diploid, non-tagged *S. uvarum* strain, and $n = 1$–3 technical and 4 biological replicas for the nuclear-envelope-corrected affinity-captured NPCs from haploid, tagged *S. cerevisiae* strains (Fig. 1a).

The absolute stoichiometry (Fig. 1 and Supplementary Table 2a) was then determined by normalizing the summed copies of Nup188, Nup120 and Nic96 per NPC to 64 copies (that is, 16 for Nup188 and Nup120, and 32 for Nic96).

3.4 SILAC-MS analyses of the NPC stoichiometry. A preparation of yeast nuclear envelopes obtained by a previously established subfractionation method[44] does not involve disruption of the nuclear envelope membrane by detergents and generates sheets of nuclear envelope studded with intact NPCs. To assess the degree to which the affinity-captured NPCs were intact, we used SILAC-MS to compare the levels of each Nup in the affinity-captured preparation relative to those in the nuclear envelope preparation. To do this, the nuclear envelope sample labelled with light isotopes was mixed with Mlp1-PPX-PrA affinity-captured NPC sample labelled with a heavy isotope (L-lysine:2HCL $^{13}C_6$) in a SILAC experiment. Mixtures of nuclear envelope proteins and NPCs labelled with stable isotopes, purified using the Mlp1–PPX-PrA handle, were digested sequentially in gel matrix by Endoproteinase LysC and by trypsin. Resulting peptides were analysed by liquid chromatography–mass spectrometry on a Thermo Q Exactive Plus mass spectrometer, with a Thermo Easy-nLC 1000 HPLC and a Thermo Easy-Spray electrospray source. H/L ratios for all peptides were obtained using MaxQuant[62], complemented with manual examinations. The H/L ratios of peptides were used to derive the H/L ratios of nucleoporins and standard deviations (data not shown). The result showed that the affinity-capture process does not affect the overall ratios of the major Nups and NPC modules relative to the nuclear envelope samples (data not shown), indicating that the affinity-capture procedure generates intact NPCs. We also used this comparison relative to nuclear envelopes to correct for the slight increases observed in the ratios of Nups closely associated with the Mlp1 handle in the affinity-captured NPCs (Fig. 1a).

3.5 In vivo *calibrated imaging analysis of GFP-tagged Nups.* Calibrated imaging data were acquired as previously described[51]. Using the avalanche photodiode imaging module of a Zeiss confocor 3, confocal *z*-stacks of live yeast were acquired with a 40× 1.2 NA Plan-Apochromat water objective. The 488-nm laser line was used to excite GFP, with a 405/488/561 dichroic. Emission was reflected with a LP580 emission dichroic and collected through a BP 505–540-nm emission filter. The pinhole was set to 1 Airy unit. The zoom was set to maintain a pixel size of 55 nm, and a *z*-step size of 400 nm was used. After acquisition, images were binned in XY by 2, resulting in an effective pixel size of 110 nm, and anaphase cells were analysed for diffraction-limited Nup spots along the anaphase bridge. These spots, when present, were fit to a 2D Gaussian curve to obtain the amplitude of the signal. The *z*-slice with the maximum signal intensity of the spot was analysed. Fluorescence correlation spectroscopy was used to convert the amplitude of the Gaussian fit of the Nup spot number of molecules of GFP. In brief, using a strain expressing only cytosolic GFP, fluorescence correlation spectroscopy determined the average number of molecules in the focal volume, as previously described[51]. Then, the amplitude of the signal of the Nup spot was compared to the intensity of cytosolic GFP, taken with the same imaging setup. For all measurements, number 1.5 coverslips were measured for uniformity, and the correction collar of the water objective was optimized for this thickness using signal intensity of Alexa Fluor 488 in solution. For each day data were acquired, the calibration using cytosolic monomer GFP was obtained.

3.6 Phospholipid analysis. These analyses were performed by Avanti Polar Lipids using their standard protocols.

3.7 Label-free mass spectrometry quantification of the NPC and associated proteins. Raw mass spectrometry files from QconCAT Mlp1-PPX-PrA immuno-isolation experiments were analysed using the MaxQuant iBAQ method[63]. Only peptides that were not isotopically labelled (that is, not QconCAT) were considered. Proteins were filtered to require more than three unique peptides per protein, and stoichiometries normalized to the absolute minimum value of the difference between label-free and the QconCAT stoichiometry for all the Nups (Extended Data Fig. 3c and Supplementary Table 8). Stoichiometries were multiplied by molecular weight to obtain mass per NPC complex and the results summed to obtain total mass of the NPC (Fig. 1c and Extended Data Fig. 3c).

3.8 Living mass of the NPC with charge detection mass spectrometry. The charge detection mass spectrometry instrument has previously been described[52,64]. In brief, the measurements are made by trapping single ions in a linear electrostatic ion trap. As the ions oscillate back and forth in the trap, they pass through a cylindrical electrode. The charge induced on the electrode is detected by a charge sensitive preamplifier. The resulting signal is amplified and digitized, and then analysed using fast Fourier transforms. The fundamental frequency provides the m/z and the magnitude is proportional to the charge. The mass of each ion is then obtained by multiplying the charge and m/z. Each NPC sample was characterized by measuring the masses of several thousand ions individually and then binning the masses to yield a true mass spectrum (Fig. 1b, c).

4. Chemical cross-linking and mass spectrometry analysis of the cross-linked NPC. NPCs were immuno-purified from Mlp1-PPX-PrA, Nup82-PPX-PrA and Nup84-PPX-PrA *S. cerevisiae* strains. After native elution, 0.5 or 1.0 mM disuccinimidyl suberate (DSS) was added and sample incubated at room temperature for 30 min with gentle shaking (~1,000 r.p.m.). The reaction was quenched with 50 mM ammonium bicarbonate or SDS-PAGE buffer containing 100 mM Tris-HCl. The sample was then precipitated using 90% methanol at −80 °C or concentrated in a speed vacuum before separation by SDS electrophoresis.

The sample was reduced by 10 mM tris-(2-carboxyethyl)-phosphine (Invitrogen) at 80 °C for 15–20 min, cooled to room temperature and alkylated by 50 mM iodoacetamide for 20 min in the dark to block the formation of disulfide bonds. After reduction and alkylation, the cross-linked complexes were separated by 3–8% SDS-PAGE (NuPAGE Tris-Acetate Fisher) to reduce the complexity of the sample. For in-gel digestion, the high-molecular-weight-region gel bands (>460 kDa, estimated by the high-molecular-weight protein markers, Invitrogen)

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

corresponding to the cross-linked NPC proteins were sliced and proteolysed by trypsin as previously described[10,65]. In brief, gel plugs were crushed into small pieces, ~5–10 μg of sequencing-grade trypsin (Promega) per ~100 μg protein was added with subsequent 6–8 h incubation. This proteolysis was repeated once more to ensure optimal results. Peptides were extracted by formic acid and acetonitrile, desalted on C18 cartridges (Waters) and snap-frozen before fractionation.

To reduce the complexity of the sample, proteolysed mixtures were separated by an orthogonal two-step fractionation strategy. First, size exclusion chromatography[66] was used for size-based separation of peptides into 2–4 fractions (~2–10 kDa). Then, a secondary fractionation using a self-packed basic (at pH 10) C18 resins (Dr. Maisch GmbH) resulted in 10–12 peptide fractions, which were subsequently analysed by liquid chromatography–mass spectrometry.

Each peptide fraction was dissolved in the sample loading buffer (5% MeOH and 0.2% FA) and analysed either by an Orbitrap Q Exactive Plus mass spectrometer or a LTQ Velos Orbitrap Pro mass spectrometer (Thermo Fisher Scientific). The Q Exactive Plus instrument was directly coupled to an easy-nLC system (Thermo Fisher Scientific) for electrospray. The cross-linked peptides were loaded onto the Easy-Spray columns (15-cm prepacked columns that are filled with C18 reverse-phase material of 2 or 3 μm particle size, 200 Å pore size and 50 μm inner diameter, Thermo Fisher Scientific) that were heated to 35 °C. Mobile phase A consisted of 0.1% formic acid and mobile phase B of 100% ACN with 0.1% formic acid. Peptides were eluted in liquid chromatography gradients of 120 min (for example, a liquid chromatography gradient of 3–7% B, 0-6 min; 7–28% B, 6-101 min; 28–100%B, 101–113 min; followed by equilibration with 100% A until 120 min). Flow rates were set at ~250–275 nl/min. Other instrumental parameters for chemical cross-linking and mass spectrometry analyses include: capillary temperature: 250–275 °C; target mass resolutions (at 200 Th): 70,000 for mass spectrometry and 17,500 for tandem mass spectrometry; AGC targets: $1–3 \times 10^6$ (full mass) and 2×10^5 (tandem mass spectrometry); mass spectrometry range of 300–1,700 Th; isolation window: 1.3–1.7 Th; higher-energy collisional dissociation normalized energy: 24–29; dynamic exclusion allowed once per 75–90 s. The top 8 most abundant ions (with charge stage of 3–7 and intensity thresholds of 3,000–7,500 ions) were selected for fragmentation by higher-energy collisional dissociation. The maximum injection times were set at 200 ms (for mass spectrometry) and 500–800 ms (for tandem mass spectrometry). For samples that were analysed by Orbitrap Velos, the cross-linked peptide mixtures were pressure-loaded onto a self-packed PicoFrit column with integrated electrospray ionization emitter tip (360 O.D, 75 I.D with 15 μm tip, New Objective). The column was packed with 10–15 cm reverse-phase C18 material (3 μm porous silica, 200 Å pore size, Dr. Maisch GmbH). Mobile phase A consisted of 0.5% acetic acid and mobile phase B of 70% ACN with 0.5% acetic acid. The peptides were eluted in a 120- or a 140-min liquid chromatography gradient (8% B to 50% B, 0–93 min, followed by 50% B to 100% B, 93–110 min and equilibrated with 100% A until 120 or 150 min) using a HPLC system (Agilent), and analysed with a LTQ Velos Orbitrap Pro mass spectrometer using similar parameters to the Q Exactive Plus instrument.

The raw data were searched by pLink[67] using a FASTA database containing 34 NPC protein sequences. An initial MS1 search window of 5 Da was allowed to cover all isotopic peaks of the cross-linked peptides. The data were automatically filtered using a mass accuracy of MS1 $\leq$ 10 p.p.m. (parts per million) and MS2 $\leq$ 20 p.p.m. of the theoretical monoisotopic (A0) and other isotopic masses (A + 1, A + 2, A + 3 and A + 4) as specified in the software. Other search parameters included cysteine carbamidomethyl as a fixed modification and methionine oxidation as a variable modification. A maximum of two trypsin missed-cleavage sites was allowed. The initial search results were obtained using a default 5% false discovery rate expected by target–decoy search strategy. All spectra were manually verified as previously described[9,10,65,68,69]. The cross-linking data were analysed and plotted by an online software tool, CX-Circos (http://cx-circos.net; W.J. *et al.*, manuscript in preparation) (Fig. 2).

5. Cryo-electron tomography of whole NPCs. We used cryo-ET and sub-tomogram averaging (Extended Data Fig. 4a) to obtain a final map with a global resolution of 28 Å; the inner ring was solved at 20–25 Å (Extended Data Fig. 5 and Supplementary Table 9). To create this map, NPCs were immuno-purified from Mlp1-PPX-PrA *S. cerevisiae* strain, in a final buffer of 20 mM HEPES (at pH 7.5), 50 mM Potassium Acetate, 20 mM NaCl, 2 mM $MgCl_2$, 0.1% Tween 20 and 1 mM DTT (see '2. Immuno-purification of the endogenous *S. cerevisiae* NPC' for details). The concentration was estimated by SDS-PAGE to be ~0.3–0.4 mg/ml. Freshly cleaned Quantafoil 300 mesh copper grids with 2-μm holes in the support film were prepared with a continuous carbon support film that spanned the holes. Before use, the grids were glow discharged in air, floated on 5 μl sample drops for 45 min and then washed by serial transfer on 4×20 μl drops of elution buffer without glycerol. Each grid was mounted on forceps in a Mark III Vitrobot (FEI) at room temperature and 100% relative humidity. Buffer on the grid was removed by blotting from the bottom with a tool that held a filter paper wedge, using access through the left-hand port. Then, 2 μl of freezing buffer was added to the grid from the right-hand port and the grid was plunge-frozen in liquid ethane after blotting.

Cryo-ET data collection was done on a Titan-Krios electron microscope operating at 300 kV, equipped with an X-FEG, a post-column energy filter set to 20 eV and a spherical aberration (Cs) corrector (Supplementary Table 9). Images were recorded with a Gatan K2 Summit direct electron detector in integration mode, with single frames taken at each tilt with UCSF Tomo[70], at a nominal spacing of 5.6 Å per pixel. A total of 253 tilt series were collected in steps between −60°, 0° and 60° in increments of 2.5–4° for different tilt series. Although the full tilt range was used for tomogram reconstruction, in the final sub-tomogram averaging step only data up to ± 45° tilt from each sub-tomogram were included in the final average. The dose target for each tilt series was 90–100 electrons per $Å^2$ and followed a cosine α dose curve with a flux of 20 electrons per pixel per second, and a dose of 3.5 electrons per $Å^2$ for the zero tilt image. Extended Data Fig. 4a presents the strategy we used to reconstruct the 3D map of the whole yeast NPC. The 3dmod viewer in IMOD[71] was used to visually screen tilts for defects with the Fourier transform, to gauge image motion. In total, 120 tilt series with a defocus range between −4.6 and −7.5 μm were kept for further processing (Supplementary Power Point Presentation slides 1, 2). After interactive test runs with etomo[71], we processed the tilt series in an automated fashion with batchtomo using 7×7 patch tracking to create aligned tilt series and calculated tomograms with back-projection and the simultaneous iterative reconstruction technique (SIRT), which were contrast transfer function (CTF)-corrected by phase-flipping each image in the tilt series (Supplementary Power Point Presentation slides 3–6). The final SIRT tomograms were binned 3× and used for interactive sub-tomogram 'particle' picking with e2spt_boxer.py in the EMAN2 single particle tomography package[72,73] with a low pass filter of 100 Å. In total, 6,416 fully sampled unfiltered sub-tomograms were extracted from the back-projection tomograms in $300 \times 300 \times 300$ voxel volumes.

In the alignment and averaging process, new algorithms for high-speed 3D alignment with automatic missing wedge compensation and averaging[73] were used throughout, and were critical for processing such large sub-tomograms. An initial reference was prepared by averaging a small subset of sub-tomograms to produce a low-resolution reference using the C_8-symmetry of the complex[74]. Owing to the large size and distinct shape of the particles, alignments were unambiguous. The alignment and averaging strategy for the final map was adapted from previously described procedures[72], and applied iteratively. The observed flexibility of the NPC ring initially limited the overall resolution to ~38 Å with 5,245 sub-tomograms (data not shown); the sub-tomograms discarded at this stage were those with the worst quality when compared to the average, generally owing to higher noise levels but in some cases due to particle damage or false positives during sub-tomogram picking. We realized that observed flexibility of the NPCs limited our resolution to ~38 Å and therefore used a tactic to locally align all individual spokes (C_8-symmetry units) to the reference rather than aligning whole rings, which could contain long-range deviations from a perfect toroid. It is important to note that these deviations are not large; across the entire NPC, they are on the order of the 38 Å resolution achieved without local alignment. All previous NPC cryo-ET maps have been produced using the approach of dividing the NPC into subunits[6,29,75,76]. In brief, two reference volumes were prepared; (1) the entire NPC and (2) a masked volume in which one-fourth of the ring had been retained, centred roughly on the mass of a single subunit. Each NPC was rotationally and translationally aligned to the reference ring. Using this initial alignment, each NPC was replicated into its eight pseudo-symmetric orientations, then a translation-only alignment against the masked reference was performed. This had the effect of bringing one asymmetric unit per replicated ring into register with the reference at a consistent radius. Although small per-subunit rotations might have occurred, this possibility was not included in the local alignment. The average from the eight subunits was then used to construct a symmetric ring by applying an azimuthal linear ramp mask centred on the mask used for alignment, which fell to zero at an angle of 45° in both directions, and then imposing the C_8-symmetry. This interpolates smoothly from one side of the subunit to the other symmetry-related side, to produce a complete symmetrized ring. This processing dramatically reduces the blurring caused by local fluctuations in subunit position and the resulting 3D volume was used as a reference in the next cycle of iterative refinement, which was repeated until no further improvement was observed.

At this stage we realized that the preferred orientation of the particles within our tomograms was leading to anisotropic resolution in the final structure, with 2/3 of the NPC rings oriented within 30° of the C_8-symmetry axis (as clearly observed in our raw data). Producing an isotropic average required balancing the various ring orientations by discarding the lowest-contrast rings in the over-populated

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

orientations to more evenly balance the orientation distribution[77–79]. This normalizing of orientations, and not 3D classification, was what led to discarding a fraction of our data (Extended Data Fig. 4a). When doing this, we elected to discard noisier sub-tomograms. This was achieved by comparing the agreement of each sub-tomogram with the overall average. In each angular range, we then retained roughly the same number of sub-tomograms, keeping those with the best quality. The discarded sub-tomograms were nearly as good, meaning we could equally well have used the next best subset of the sub-tomograms with virtually no effect on the final structure. Indeed, in the less common orientations, we were forced to use virtually all of the sub-tomograms irrespective of quality. Thus, in our work, a large fraction of the sub-tomograms were discarded not owing to their poor quality or conformational variability, but rather because of the preferred orientation of the particles within the tomogram. The final reconstruction used 1,864 (of the 6,416 initial) sub-tomograms. These sub-tomograms were further divided randomly into two groups for resolution assessment. 'Gold standard' refinement was used for resolution testing and to ensure self-consistency (Supplementary Power Point Presentation slides 7, 8). Both global and local resolution assessments were done using a set of tiled Gaussian masks to estimate the local resolution and reproducibility of the structure, which is one of the standard methods for local resolution assessment. In brief, a 3D Gaussian shape is generated with a 'full width at half maximum' parameter of at least 2× the anticipated resolution, and generally even larger. This Gaussian shape is then applied as a mask to both maps and a Fourier shell correlation (FSC) curve is computed. This process is repeated in a tiled pattern throughout the volume. This provides a resolution for each sampled location in the volume. This procedure involves a trade-off: with smaller Gaussians, FSC curves are less precise, but with larger Gaussians the resolution estimate is less localized. The Gaussians overlap to provide better sampling of the volume. The resulting resolution map is similar to those produced by ResMap[80], but it measures FSC, which is filter-independent, unlike ResMap[80] that requires unfiltered volumes and would not work well on sub-tomogram averages. The global resolution of our cryo-ET map at the standard $FSC_{0.143}$ cutoff is ~28 Å and the local resolution distribution ranges from 20 to 38 Å, with the inner ring being in the 20–25 Å range (Extended Data Fig. 5a–d). This local resolution estimate was used to locally filter the density map, to produce a map with the appropriate level of detail in each area (Extended Data Fig. 5). The size of the Gaussian window was 140 Å, indicating the smallest region over which the resolution is considered to vary. Although this may seem large, it is small compared to the size of the overall NPC (Extended Data Fig. 5). CTF phase-flipping was applied during tomographic reconstruction and a final approximate amplitude correction was applied to the averaged NPC ring. Therefore, theoretical CTF curves for the mean defocus values present in the tomograms were averaged assuming 10% amplitude contrast. The reciprocal of this curve was then applied as a filter to the final uncorrected map. The cryo-ET density map was refined at 5.3 Å per pixel on the basis of a recalibration of the map with known structures.

In parallel, and as an additional validation of our final map, we also carried out a tomographic analysis of the yeast NPC dataset (the same 6,416 sub-tomograms) using RELION 1.4, and incorporated a CTF model[81,82]. In brief, we calculated back-projection tomograms without phase-flipping corrections for the individual tilted images, and binned the output sub-tomograms twofold to 10.6 Å per pixel. The datasets underwent sequential rounds of 2D classification using *Z*-projections of the sub-tomograms to eliminate poor particles. A subsequent 2D classification identified near-top, tilted and side views; the latter provided an independent estimate of NPC thickness perpendicular to the nuclear envelope (620–640 Å). A 3D reconstruction using the best sub-tomograms with RELION produced a map at ~35 Å resolution (data not shown) with similar features to those obtained with the e2spt in the EMAN2 single particle tomography package[72,73], including distinct connections between each spoke and the transporter, further validating these features in our cryo-ET map (Fig. 3). Finally, *Z*-projections of original sub-tomograms that were roughly aligned along the C_8-symmetry axis were used for an additional unsupervised 2D classification, which produced classes with central transporters without using the C_8-symmetry restraint (Extended Data Fig. 6a, b). Differences in the apparent resolution of the class averages in Extended Data Fig. 6a, b reflect different particle numbers in the classes. As mentioned above, the dataset of particles has a strong orientational bias, in which the NPCs tend to bind to the carbon support film with a range of 0–30 degrees of tilt. The class averages are based on 2D projections along the *z* axis of the original sub-tomograms, to avoid issues with the missing wedge, and there is therefore a disparity in particle numbers in the classes. Tilting in the tomographic data collection helped to fill in the missing data, but we took great care to ensure an equal coverage of Fourier space in the calculation of our final map, to avoid distortions, and also took a number of other steps to ensure that radiation damage and loss of data quality in later tilts was minimized by using only information in Fourier space at ± 45° from each particle sub-tomogram when they were combined to form the final map. The RELION map serves as a strong validation of our final map, because if our map was flawed, a reconstruction with RELION would have resulted in a different map (Extended Data Fig. 5e). Additionally, the fact that a RELION reconstruction resulted in a ~35 Å resolution map—virtually the same resolution as obtained in our 'intermediate' map described above (~38 Å)—validates our methodology and the quality of our final map (Extended Data Fig. 5e). An additional point that provides *prima facie* evidence that our cryo-ET map was calculated correctly is that local two-fold symmetry (C_2-symmetry), which was expected in the inner ring of the NPC, emerges without any enforcement, whereas the overall map shows a clear asymmetry with large and distinctly different features on the nuclear and cytoplasmic face of the yeast NPC (which were also observed in the RELION map) and a slightly tapered appearance, as is shown in Fig. 3a and Extended Data Figs 4d, 5b.

6. Small angle X-ray scattering. Small angle X-ray scattering (SAXS) measurements for 147 constructs of 18 Nups[9,12,23,83–86] (Supplementary Table 6; S.J.K. *et al.*, manuscript in preparation; and Source Data) were carried out both at the Stanford Synchrotron Radiation Lightsource Beamline 4-2 in the SLAC National Accelerator Laboratory and at the SIBYLS Beamline 12.3.1 of the Advanced Light Source in the Lawrence Berkeley National Laboratory. SAXS data were collected at concentrations ranging from 0.5 to 5.0 (or higher, depending on the sample) mg/ml, using the previously defined standard protocol[12,23,83]; approximately 20 one-second exposures were used for each sample and buffers at 15 °C. Further details of the SAXS experiments have previously been published[9,12,23,83–86].

7. Phenotypic analysis by one-cell doubling evaluation by living arrays of yeast. Yeast growth phenotypes were quantified using the one-cell doubling evaluation by living arrays of yeast (ODELAY) assay, as previously described[14]. In brief, yeast was cultured in YPD medium in 96-well plates overnight. Cultures were diluted to an OD_{600} of 0.09 and allowed to grow for 6 h at 30 °C. The cultures were then diluted again to an OD_{600} of 0.02 and spotted onto YPD agarose medium. The resulting cultures were then observed using time-lapse microscopy for 48 h with 30 min intervals between images. All images were collected on Leica DMI6000 microscopes with a 10× 0.3NA lens using bright field microscopy. MATLAB scripts using the Micro-Manager interface controlled the image collection process[87]. Six independent experiments were performed. The population growth rates were scored against each other using the following equation:

$$Z_{\text{mean}} = \frac{1}{n}\sum_{i}^{n}\frac{d_i - \mu_i}{\sigma_i}$$

in which d_i is the *i*th decile of doubling time of the query population, μ_i is the mean of the *i*th decile of the doubling time of the parent strain and σ_i is the standard deviation of the *i*th decile of the doubling time of the parent strain. The mean and standard deviation deciles were calculated from at least 4 separate populations containing at least 200–300 individuals. All calculations were performed using MATLAB scripts. Following *Z*-scoring of the populations, an additional weight was added to the scoring for truncation strains that occurred in haploid versus diploid strains of yeast.

8. Negative-stain electron microscopy of the native Nic96 complex. An affinity-captured and natively eluted sample of the endogenous Nic96 complex (composed of Nic96, Nsp1, Nup49 and Nup57) was applied to a glow-discharged grid and stained with 1% uranyl formate. Images were collected on a Philips CM200 transmission electron microscope (FEI) operating at 200 kV at 50,000× magnification and a defocus of ~1.5 μm (2.03 pixels per Å). Images were recorded on a Gatan UltraScan 1000 2k × 2k CCD camera (Gatan). Particles were selected using Boxer from EMAN[88], normalized and then phase-flipped using ctfit from EMAN. In total, 34 class averages (selected classes shown in Extended Data Fig. 7g) were generated through ISAC[89] that classified ~86% of the original set of 5,458 particles.

9. Integrative structure determination of the *S. cerevisiae* NPC. The structure of the *S. cerevisiae* NPC, including the scaffold, membrane rings, cytoplasmic export platform and nuclear baskets in the context of the pore membrane—but excluding the flexible FG regions—was solved by integrative structure determination (see '9.1 Integrative structure determination of the *S. cerevisiae* NPC scaffold, membrane rings, cytoplasmic export platform and nuclear basket'). Moreover, the distributions of the FG regions and the cargo-bound NTFs, comprising the central transporter, were computed by Brownian dynamics simulation (see '9.2 Brownian dynamics simulation of FG repeats and NTFs').

9.1 Integrative structure determination of the S. cerevisiae *NPC scaffold, membrane rings, cytoplasmic export platform and nuclear basket.* Integrative structure determination of the *S. cerevisiae* NPC proceeded through four stages[8,90–92] (Extended Data Fig. 1, Supplementary Table 3 and Supplementary Videos 1–3): (1) gathering data, (2) representing subunits and translating data

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

into spatial restraints, (3) configurational sampling to produce an ensemble of structures that satisfies the restraints and (4) analysing and validating the ensemble structures and data (Extended Data Figs 1, 7, 8 and Supplementary Tables 2–4). The integrative structure modelling protocol (stages 2, 3 and 4) was scripted using the Python modelling interface (PMI) package version 4d97507, which is a library for modelling macromolecular complexes based on our open-source integrative modelling platform (IMP) package[90] version 2.6 (https://integrativemodeling.org). The current procedure is an updated version of previously described protocols[9,10,12,93–96].

9.1.1 Stage 1: gathering data. The stoichiometry of Nups in the NPC was determined using native mass spectrometry and biochemical quantification of the purified NPC complex (Fig. 1 and Extended Data Figs 2, 3). In total, 3,077 intra- and intermolecular DSS and EDC unique cross-links were identified using mass spectrometry (Fig. 2 and Supplementary Table 1), which informed the spatial proximities among the 32 Nups and their conformations. The density map of the entire NPC was determined by cryo-ET at an average resolution of 28 Å, with the local resolution as high as ~20 Å for the inner ring, which informed the shape of the NPC (Fig. 3 and Extended Data Figs 4–6). Re-interpreted immuno-electron microscopy data[3,8] informed the positions of 29 Nups. Predictions of the transmembrane domains obtained from the *Saccharomyces* Genome Database[97] (http://yeastgenome.org) and predictions of MBMs from the HeliQuest webserver[12,98] informed about their respective proximities to the pore membrane. Previous immuno-electron microscopy measurements[99] informed the end-to-end distance for Mlp1 and Mlp2. Low-resolution electron microscopy images of the NPC[44] informed the diameter of the distal basket ring formed by Mlp1 and Mlp2.

Representations of individual Nups and some of their sub-complexes (Supplementary Table 2 and references therein) relied on (1) atomic structures of 21 yeast Nup domains and 3 sub-complexes determined by X-ray crystallography or nuclear magnetic resonance spectroscopy; (2) our previously determined structures of Nup116, Nup133, Nup145N, Nup192 and Pom152, as well as the Nup82 and Nup84 sub-complexes solved by integrative structure determination[9,10,12,23,83–86]; (3) 29 comparative models built with MODELLER 9.13[100] on the basis of known structure(s) detected by HHpred[101,102]; (4) SAXS profiles for 147 constructs of 18 Nups[9,12,23,83–86] (Supplementary Table 6; S.J.K. *et al.*, manuscript in preparation); (5) secondary structure, disordered regions, and domain boundaries predicted by PSIPRED[103,104], DISOPRED[105], and DomPred[106], respectively; (6) coiled-coil regions of Nup82, Nup159, Nsp1, Nup49, Nup57, Mlp1, and Mlp2 predicted by COILS/PCOILS[107] and Multicoil2[108]; (7) an atomic structure of the Nup53$^{229–365}$ RRM domain from *S. cerevisiae* determined by X-ray crystallography (P. Sampathkumar *et al.*, manuscript in preparation); and (8) negative-stain electron micropscopy density maps of full-length Nup192 (EMD-5556[86]) and Pom152 (EMD-8543[23]). See Supplementary Table 2 and references therein for all above (1) to (8).

Our previously published topological map of the NPC[3] and the 82 composites determined by affinity purification and overlay assay[8] were not used for computing the current NPC structure, but were used for validating the current NPC structure.

9.1.2 Stage 2: representing subunits and translating data into spatial restraints. Information about the modelled system (see '9.1.1 Stage 1: gathering data') can in general be used for defining the system's representation, defining the scoring function that guides sampling of alternative structural models, limiting sampling, filtering of good-scoring structures obtained by sampling and final validation of the structures. Here the NPC representation relies primarily on stoichiometry as well as atomic structures, integrative structures, comparative models and SAXS profiles of Nups and their sub-complexes (Supplementary Tables 2 and 6, and references therein); the scoring function relies on chemical cross-links, the cryo-ET density map, immuno-electron microscopy localizations, excluded volume, sequence connectivity, the shape of the pore membrane and four types of sequence-based localization relative to the membrane (below); the sampling benefits from symmetry constraints (below); and the validation of the final structure relies in part on the SAXS profiles (Supplementary Table 6) and composites determined by affinity purification and overlay assays[8] (below).

To improve computational efficiency and avoid a representation that was too coarse, we represented the NPC in a multi-scale fashion. A rigid-body consisting of multiple beads was defined for each X-ray structure, NMR structure, comparative model and integrative structure of the NPC components (Supplementary Table 2). The remainders of the Nup sequences not in rigid-bodies (36.8% of residues, excluding FG repeats) were represented as flexible strings of beads. In a rigid-body, the beads have their relative distances constrained during configurational sampling, whereas in a flexible string the beads are restrained by the sequence connectivity, excluded volume and potentially additional restraints, such as chemical cross-links, as exemplified in previous publications[9,10,23,93,109].

Rigid-bodies (63.2% of residues, excluding FG repeats) were coarse-grained using two resolutions, in which beads represented either individual residues or segments of up to ten residues. The coordinates of a 1-residue bead were those of the corresponding C_α atom. The coordinates of a 10-residue bead were the centre of mass of the ten constituent 1-residue beads. Finally, the remaining regions without an atomic representation (that is, the predicted transmembrane and disordered regions) were represented by a flexible string of beads encompassing 25 to 100 residues each; the low-resolution representation of these regions is justified because their conformations are likely to be 'decoupled' from the structure of the rest of the NPC[3,110].

We used the SAXS data to confirm the rigid-body representations of eight Nups with X-ray structures, comparative models and previously published atomic integrative structures[9,12,23,83–86] (Extended Data Fig. 7f and Supplementary Tables 2, 6). The rigid-body representation of a Nup construct was validated by a χ-value that quantifies the difference between the computed (from an atomic rigid-body representation using FoXS[111]) and experimental SAXS profiles, except for several constructs of Nup133 and Nup192 that were flexible during integrative modelling and were thus evaluated as previously described[12,86,112]. The χ-value validation assumes that each Nup construct, corresponding in most cases to a single domain (not the whole protein), has the same conformation in solution and in complex; this assumption is consistent with other data (for example, the chemical cross-links and cryo-ET map). The SAXS validation is necessarily limited to *S. cerevisiae* constructs of Nups that exist as a rigid monomer in solution and do not contain FG repeats; rigid-body representations of the constructs from other species, constructs that oligomerize in solution and constructs that include FG repeats cannot be easily used for validation, because of the sensitivity of a computed SAXS profile to the differences in the sequence and stoichiometry, as well as to potential errors in comparative modelling (especially of insertion and deletion).

After producing this validated representation, we next encoded the spatial restraints on the basis of information gathered in Stage 1, according to the following steps (Supplementary Table 4; for the definition of the scoring function consisting of these restraints, see '9.1.3.2 Scoring function'):

(1) Cross-link restraints: 1,643 of the 3,077 unique cross-links (Fig. 2 and Supplementary Table 1a) were used to restrain the distances spanned by the cross-linked residues, relying on a Bayesian scoring function[10]. The evaluation takes into account the ambiguity due to multiple copies of identical subunits and, for cross-links involving the same protein type, due to the lack of knowledge of whether they are intra- or intermolecular[9,93,109]; the ambiguous cross-link restraint considers all intra- and intermolecular assignments in multiple copies of identical subunits, with only the least violated distance contributing to the score. The remaining 1,434 DSS and EDC cross-links (Fig. 2 and Supplementary Table 1b–f) were already used as restraints to build the integrative structures of the Nup84[10] and Nup82 sub-complexes[9], represented here as rigid-bodies. The two homo-dimer DSS cross-links between two copies of residue 62 of Pom152[23] and two copies of residue 151 in Nup60[22] were transformed into harmonic upper-distance bounds, enforcing the homo-dimer configuration.

(2) Cryo-ET density restraint: the cryo-ET density restraint was applied, which corresponded to the cross-correlation between the Gaussian mixture model (GMM) representation of most Nups and the GMM representation of the cryo-ET density map[95,113–115] (Fig. 3 and Extended Data Figs 4–6); we used a GMM representation for the sake of computational efficiency, necessitated by the large size of the NPC. An assessment of a given structure against a density map is much faster when both are represented with a mixture model (because the number of components in a mixture model is much smaller than the number of grid points covering the maps). However, these two scores are very strongly correlated. Thus, the structures obtained with a grid representation, if we had sufficient computational power, would certainly be indistinguishable from the current NPC structures[115].

A 90° arc of the cryo-ET density map was approximated by the GMM, which contained 1,750 components computed using the expectation-maximization algorithm implemented in scikit-learn (http://scikit-learn.org); the cryo-ET GMM appeared to be sufficient to reproduce the complete features of the density map (excluding the central transporter region). To use a comparable number of GMM components for Nups, a Nup was approximated by a GMM component for each of its 100 to 500 residues. The cross-correlation quantified the degree of overlap between the Nup GMM components and the cryo-ET GMM components.

(3) Immuno-electron microscopy localization restraints: the immuno-electron microscopy localization restraint was used to localize the C-terminal beads of 29 of the 32 Nups, on the basis of previous immuno-electron microscopy data[3,8,116]. This goal was achieved by imposing upper and lower harmonic bounds on the axial and radial coordinates of the restrained bead, reflecting the uncertainty in the immuno-electron microscopy data[8]. The three remaining Nups (Nsp1, Sec13

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

and Seh1) were not restrained by the immuno-electron microscopy data because of their high uncertainty, presumably due to the positional heterogeneity of the tagged Nup in the multiple superposed electron microscopy images of the NPC. This heterogeneity is more likely to occur for Nups with multiple copies per C_2-symmetry unit, which are unlikely to share the same radial and axial coordinates.

(4) Excluded volume restraints: the protein excluded volume restraints were applied to each 10-residue bead, using the statistical relationship between the volume and the number of residues that it covered[8–10,117].

(5) Sequence connectivity restraints: we applied sequence connectivity restraints, using a harmonic upper bound on the distance between consecutive beads in a subunit, with a threshold distance equal to three times the sum of the radii of the two connected beads. The bead radius was calculated from the excluded volume of the corresponding bead, assuming standard protein density[8–10,117].

(6) Membrane exclusion restraints: the membrane exclusion restraints were applied to beads in the non-membrane-spanning Nups or to their segments to prevent these beads from penetrating the pore membrane. A lower harmonic bound at 0 Å was applied to the distance between a bead and the closest point on the pore-side membrane surface[3,8] (modelled as a half-torus with the large and small radii of 390 and 150 Å, respectively; Supplementary Table 3), for all coarse beads (10 residues or more per bead) in all Nups but Pom152, Ndc1 and Pom34; the restraint was also applied to all non-membrane coarse beads of Pom152$^{1–110}$, Ndc1$^{1–28}$, Ndc1$^{248–655}$, Pom34$^{1–63}$ and Pom34$^{151–299}$.

(7) Transmembrane domain restraints: the transmembrane domain restraint was used to localize the coarse beads in the predicted transmembrane domains (Pom152$^{111–200}$, Ndc1$^{29–247}$ and Pom34$^{64–150}$; Supplementary Table 2 and references therein) within the pore membrane, which is 45 Å thick[3,8]. This aim was achieved by imposing an upper harmonic bound at 45 Å and a lower harmonic bound at 0 Å on the distance between the bead and the closest point on the pore-side membrane surface.

(8) Membrane surface binding restraints: the membrane surface binding restraint was used to localize the coarse beads in the predicted MBMs (Nup1$^{1–32}$, Nup60$^{27–47}$, Nup120$^{135–152}$, Nup120$^{197–216}$, Nup133$^{252–270}$, Nup157$^{310–334}$, Nup170$^{320–344}$, Nup53^{475} and Nup59^{528}; Supplementary Table 2 and references therein), within the pore membrane up to 12 Å from the pore-side membrane surface[118]. This aim was achieved by imposing an upper harmonic bound at 12 Å within the pore membrane and a lower harmonic bound at 0 Å on the distance between the bead and the closest point on the pore-side membrane surface. For Nup120, only the best satisfied of the Nup120$^{135–152}$ and Nup120$^{197–216}$ restraints was used[10,12] (conditional restraint).

(9) Pom152 perinuclear volume restraint: only the C-terminal region of Pom152 (residues 201–1337) was restrained to the perinuclear lumen of the pore membrane[23]. This aim was achieved by imposing a lower harmonic bound at 0 Å on the distance between the Pom152 beads and the closest point on the perinuclear side of the membrane surface.

(10) Distal basket ring restraints: the conformations of Mlp1 and Mlp2 were restrained by an upper harmonic bound at 350 Å and a lower harmonic bound at 230 Å on the distance between the N-terminal and C-terminal beads, on the basis of immuno-electron microscopy measurements[99]. In addition, the radius of the distal basket ring was restrained by an upper harmonic bound at 170 Å and a lower harmonic bound at 130 Å on the radial coordinates of the C-terminal beads of Mlp1 and Mlp2, on the basis of low-resolution electron microscopy images of the NPC[44]. The nuclear basket was also informed by cross-linking restraints and the C_8-symmetry constraint (see '9.1.3.1 Sampling space with symmetry constraints').

9.1.3 Stage 3: Configurational sampling. We used the configurational sampling to produce an ensemble of structures that satisfies the restraints, as described below

9.1.3.1 Sampling space with symmetry constraints. We aimed to maximize the efficiency of the configurational sampling: more specifically, we aimed to maximize the precision at which the sampling of good-scoring solutions was exhaustive (see '9.1.4 Stage 4: analysing and validating the ensemble structures and data'). Therefore, we reduced the number of independently moving parts in the NPC structure by explicitly considering the C_8- and C_2-symmetries of the NPC, as follows. The entire NPC consists of 8 clones of the C_8-symmetry unit, related by multiples of a 45° rotation around the z axis (Fig. 3 and Extended Data Figs 4–6). The C_8-symmetry unit was further broken into two C_2-symmetry units and non-C_2-symmetric Nups (Supplementary Table 2a); the C_2-symmetry unit contains Nups that occur equally on both the cytoplasmic and nucleoplasmic sides[3,8,116]. For computational efficiency, we defined the coordinate system such that the C_2-symmetry is imposed simply by cloning a bead in the C_2-symmetry unit at (x, y, z) to $(x, -y, -z)$ (equivalent to a rotation of 180° around the x axis). This aim was achieved by fitting both copies of Pom152[23] into the cryo-ET density map, followed by moving the centre of the map to the origin of the coordinate system and orienting the map such that the $x, -y, -z$ transformation applies to Pom152.

With these symmetries in hand, we sampled only the positions of rigid-bodies and beads corresponding to the Nups in the C_2-symmetry unit and non-C_2-symmetric Nups. There are no Nups that occur on both the cytoplasmic and nuclear sides and are not related by the C_2-symmetry; there are no Nups that occur with a different stoichiometry on both sides. In addition, the luminal domain of Pom152 was considered already well-positioned given its fit into the cryo-ET density map (Fig. 3e) and peripheral location in the NPC[23], and was not sampled further.

9.1.3.2 Scoring function. The scoring function included restraints on the sampled Nups and the Pom152 luminal domain as well as restraints across the interfaces with neighbouring symmetry units: (1) the cryo-ET density restraint and distal basket ring restraint applied to the Nups in the sampled C_8-symmetry unit; (2) sequence connectivity, immuno-electron microscopy localization, and the four types of sequence-based localizations relative to the membrane applied to the Nups in the sampled C_2-symmetry unit and non-C_2-symmetric Nups; and (3) cross-link and excluded volume restraints applied to the pairs of beads for Nups within the sampled C_8-symmetry unit and across the interfaces with neighbouring symmetry units.

9.1.3.3 Sampling algorithm. The search for good-scoring structures relied on replica exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm[9,10] (Supplementary Table 3). The Monte Carlo moves included random translation and rotation of rigid-bodies (up to 4 Å and 0.04 radians, respectively) and random translation of individual beads in the flexible segments (up to 4 Å). As indicated above, these operations were applied only to the sampled rigid-bodies and beads. The remaining, symmetry-constrained rigid-bodies and units were moved in lockstep to maintain the exact C_8- and C_2-symmetries at each sampling step, as described above. Up to 64 replicas were used, with a 1.0–5.0 temperature range. Forty-two independent sampling calculations were performed, each one starting with a random initial configuration. The coordinates were saved every 10 Gibbs sampling steps, each consisting of a cycle of Monte Carlo steps that moved every rigid-body and flexible bead once.

To further increase the efficiency of sampling, we first applied the above Monte Carlo algorithm separately to the following four subsets of Nups, which are co-localized on the basis of previous characterizations[3,8,9,23] and the current cryo-ET density map: (1) Nup82 and Nup84 sub-complexes, (2) Pom152, (3) inner-ring Nups (Nup157, Nup170, Nup188, Nup192, Pom34, Ndc1, Nup53, Nup59 and Nic96$^{205–839}$), and (4) Mlp1 and Mlp2. Next, the best-scoring solutions from sampling each of the first three subsets were combined; they were already in the same reference frame, because they were all obtained by fitting the same cryo-ET density map and immuno-electron microscopy data. The rest of the Nups and the Mlp1–Mlp2 heterodimer were then added in random positions and orientations, followed by another application of the above Monte Carlo algorithm to all sampled Nups. This sampling produced a total of 100,453 modelled structures in 42 independent runs (the score ranges from 88,545.0 to 103,589.5, with the mean and standard deviation of 88,831.5 and 187.4, respectively), requiring ~10 weeks on a cluster of ~2,500 CPU cores. For the most detailed specification of the sampling procedure, see the IMP modelling script (https://salilab.org/npc2018).

We considered for further analysis only the 5,529 modelled structures with the scores better than 88,644.1 (1 standard deviation below the mean value); this threshold implies satisfaction of the input datasets within their uncertainties (Supplementary Table 4; see '9.1.4.3 Fit to input information'). These structures are already superposed because they were fit into the same Cryo-ET map and sampling did not move the luminal domain of Pom152 (see '9.1.3.1 Sampling space with symmetry constraints').

9.1.4 Stage 4, analysing and validating the ensemble structures and data. Input information and output structures need to be analysed to estimate structure precision and accuracy, detect inconsistent and missing information, and to suggest more informative future experiments. We used the previously published analysis and validation protocol[8,9]. Assessment began with a test of the thoroughness of structural sampling, followed by structural clustering of the modelled structures and estimating their precision based on the variability in the ensemble of good-scoring structures, quantification of the structure fit to the input information and structure assessment by data not used to compute it; structure assessment by cross-validation was not performed in this case, because it takes ~10 weeks on approximately 2,500 CPU cores to compute an ensemble of structures for a single set of input datasets. These validations are based on the nascent wwPDB effort[92] toward archiving, validating and disseminating integrative structures. We now discuss each one of these validations in turn.

9.1.4.1 Thoroughness of the configurational sampling. We must first estimate the precision at which sampling found the most good-scoring solutions (sampling precision); the sampling precision must be at least as high as the precision of the

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

structure ensemble that is consistent with the input data (structure precision). As a proxy for testing the thoroughness of sampling, we performed four tests of sampling convergence[119], as follows.

The first convergence test confirmed that the scores of refined structures do not continue to improve as more structures are computed, essentially independently of each other (Extended Data Fig. 1c).

The second convergence test confirmed that the good-scoring structures in independent sampling runs 1–21 (structure sample 1; $n^{\text{sample1}} = 2{,}359$ structures) and 22–42 (structure sample 2; $n^{\text{sample2}} = 3{,}170$ structures) satisfied the data equally well. The non-parametric Kolmogorov–Smirnov two-sample test[120,121] (two-sided) indicates that the difference between the two score distributions is insignificant (P value (1.0) > 0.05). In addition, the magnitude of the difference is small, as demonstrated by the Kolmogorov–Smirnov two-sample test statistic, D, of 0.045 (Extended Data Fig. 1d). Thus, the two score distributions are effectively equal.

Next, we considered the 5,529 good-scoring structures themselves (not their scores as in the two tests described above). For stochastic sampling methods, thoroughness of sampling can be assessed by showing that multiple independent runs (for example, using random starting configurations and different random number generator seeds, as is the case for structure samples 1 and 2) do not result in noticeably different structures[8–10,13]. We tested the similarity between structure samples 1 and 2 in the following two ways.

The third convergence test[119] relied on the χ^2-test (one-sided) for homogeneity of proportions[122] between structure samples 1 and 2 (Extended Data Fig. 1e, f). The test involves clustering structures from both samples, followed by comparing the proportions of structures from each sample in each cluster. No adjustment was made for multiple comparisons. A comparison of two NPC structures considered only the beads representing Nups with a single copy per C_2-symmetry unit and the Nic96 complex (including all Nups in the inner, outer and membrane rings, but excluding Nup100, Nup116, Nup145N, Nup1, Nup60, Gle1, Nup42, Mlp1 and Mlp2), to avoid the combinatorial explosion in identification of topologically equivalent Nup copies. The sampling precision is defined as the largest root-mean-square deviation (r.m.s.d.) between a pair of NPC structures within any cluster, in the finest clustering for which each sample contributes structures proportionally to its size (considering both the significance and magnitude of the difference) and for which a sufficient proportion of all structures occur in sufficiently large clusters. The sampling precision for our NPC structure is 9 Å (Extended Data Fig. 1e).

Threshold-based clustering[123] results in a single dominant cluster containing 80.3% of the good-scoring structures (Extended Data Fig. 1e, f) with a root-mean-square fluctuation (r.m.s.f.) of 9 Å (cluster precision). The remaining 19.7% of the structures are similar to those in the dominant cluster; the largest r.m.s.d. value from a structure in the dominant cluster is 17 Å (the mean and standard deviation of the r.m.s.d. values are 13.3 and 1.3 Å, respectively). Therefore, there is effectively a single good-scoring solution, at the structure precision of 9 Å (equal to the cluster precision). The sampling precision of 9 Å (r.m.s.d.) is sufficiently high for computing a structure at 9 Å precision (r.m.s.f.; r.m.s.d. is approximately $\sqrt{2} \times$ r.m.s.f.[124]). For the remainder of our analysis, we use only the structures in the dominant cluster.

The fourth convergence test relied on a comparison of two localization probability density maps for each Nup, obtained for dominant cluster structures in samples 1 and 2. A localization probability density map defines the probability of any voxel (here, $6 \times 6 \times 6$ Å^3) being occupied by a specific protein in a set of structure densities, which in turn are obtained by convolving superposed structures with a Gaussian kernel (here, with a standard deviation of 5.4 Å). The average cross-correlation coefficient between the two maps for each Nup is 0.90, indicating that the positions of most Nups in the two samples are nearly identical at the structure precision of 9 Å.

In conclusion, all four sampling tests indicate that the sampling was exhaustive at 9 Å precision (Supplementary Table 3). The caveat is that passing these tests is necessary but not sufficient evidence of thorough sampling; a positive outcome of the tests may be misleading if, for example, the landscape contains only a narrow—and thus difficult to find—pathway to the pronounced minimum corresponding to the correct structure. Moreover, our sampling was not completely stochastic because it proceeded in two steps, the first of which prepared the starting configuration for the second step. As a result, the actual structure precision might be worse[125–128] than the estimated 9 Å.

9.1.4.2 Clustering and structure precision. An ensemble of good-scoring structures needs to be analysed in terms of the precision of its structural features[3,8,9]. The precision of a component position can be quantified by its variation in an ensemble of superposed good-scoring structures. It can also be visualized by the localization probability density for each of the components of the NPC structure.

As described above, integrative structure determination of the NPC resulted in effectively a single good-scoring solution, at the precision of ~9 Å. This precision is sufficiently high to pinpoint the locations and orientations of the constituent Nups, demonstrating the quality of the input data, including the chemical cross-links (Fig. 2 and Extended Data Fig. 7a–c) and the cryo-ET density map (Fig. 3 and Extended Data Fig. 8).

9.1.4.3 Fit to input information. An accurate structure needs to satisfy the input information used to compute it. Because the sampling was exhaustive at ~9 Å precision, overfitting is not a problem at this precision; all structures at this precision that are consistent with the data are provided in the ensemble.

The dominant cluster satisfies 90% of the DSS cross-links (Extended Data Fig. 7a–c and Supplementary Tables 1, 4); a cross-link restraint is satisfied by a cluster of structures if the corresponding C_α–C_α distance in any of the structures in the cluster (considering restraint ambiguity) is <35 Å (Extended Data Fig. 7a–c; shown in blue). Therefore, the dominant cluster essentially satisfies the cross-linking data within its uncertainty (the false detection rate is approximately 5% to 10%[129,130]). Most of the cross-link violations are small, and can be rationalized by local structural fluctuations, coarse-grained representations of some Nup domains, and/or finite structural sampling, as shown in Extended Data Fig. 7a (a histogram presenting the distribution of the cross-linked C_α–C_α distances).

The localization probability densities for the dominant cluster overlap well with the cryo-ET density map, with the cross-correlation coefficient of 0.92 (Fig. 3, Extended Data Fig. 8 and Supplementary Table 4). Additional density is present in the cryo-ET map for the Nup82 complex (cytoplasm) and basket attachment sites (nucleoplasm). This density may arise from local flexibility of these modules or may be due to the presence of cargo or transport factors associated with the NPC (Fig. 1b, c and Extended Data Fig. 3c). For visualization, the localization probability densities are typically smoothed and contoured at the threshold that results in approximately twice the protein volume estimated from its sequence (Fig. 4).

The remainder of the restraints are harmonic, with a specified standard deviation. The dominant cluster generally satisfied at least 95% of restraints of each type (Supplementary Table 4); a restraint is satisfied by a cluster of structures if the restrained distance in any structure in the cluster (considering restraint ambiguity) is violated by less than 3 standard deviations, specified for the restraint. Most of the violations are small, and can be rationalized by local structural fluctuations, coarse-grained representations of some Nup domains and/or finite structural sampling.

9.1.4.4 Satisfaction of data and considerations that were not used to compute structures. The most direct test of a modelled structure is by comparing it to the data that were not used to compute it (a generalization of cross-validation).

First, our current NPC structure is consistent with our previously published data and topological map[3,8] (Extended Data Fig. 7d, e). Our current structure satisfies all 82 composites determined by affinity purification and overlay assays[3,8], even though these were not used in this calculation. For example, Pom152, Pom34, Ndc1, Nup157 and Nup170 are connected with each other (left panel in Extended Data Fig. 7e), consistent with the composites determined in a previous publication using the affinity purification data[3,8] (right panel in Extended Data Fig. 7e). Moreover, the position of each Nup in the current structure is generally similar to that in the previous topological map[3,8], although the current structure is determined at a precision that is an order of magnitude higher than in the previous map (Extended Data Fig. 7d).

Second, the atomic structures of eight Nups are consistent with the corresponding SAXS profiles for their constructs (Extended Data Fig. 7f and Supplementary Tables 2, 6), as discussed in '9.1.2 Stage 2: representing subunits and translating data into spatial restraints'. For example, the SAXS profile calculated from the atomic structure of Pom152$^{718–1148}$ (red curve in Extended Data Fig. 7f) using FoXS[111] is well matched ($\chi = 1.48$) to the corresponding experimental SAXS profile[23] (black dots in Extended Data Fig. 7f; $n = 20$ exposures). For visualization purposes, the Pom152$^{718–1148}$ structure (represented as a ribbon) is shown along with the best fit of the *ab initio* shape (represented as a transparent envelope) computed from the experimental SAXS profile, in Extended Data Fig. 7f.

Third, the structures of the Nic96 complex (composed of Nic96, Nsp1, Nup49 and Nup57) in the dominant cluster can be projected well on most of the 2D class averages obtained for the natively isolated complex (Extended Data Fig. 7g; see '8. Negative-stain electron microscopy of the native Nic96 complex'). More specifically, the electron microscopy 2D validation fits the structure of the Nic96 complex in the whole NPC context to the electron microscopy class averages of the Nic96 complex, and computes a score that quantifies the match. The computation proceeds in three stages: (1) generation of alternative model projections, (2) alignment of the class average and each model projection, and (3) calculation of the fitting score for each projection, as follows. First,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

1,000 uniformly distributed projections of the low-pass-filtered structure of the NPC on the sphere (stage 1) were generated. Second, each projection was optimally aligned to each of the class averages in Fourier space (stage 2). Finally, a score corresponding to the cross-correlation coefficient was computed (stage 3). For example, the experimental class averages were satisfied by the structure with cross-correlation coefficients of 0.85 and 0.80, respectively (Extended Data Fig. 7g).

Fourth, the structure was also validated by its comparison to the core scaffold maps of the *Homo sapiens* NPC, which are based primarily on electron microscopy density maps[6,7,17,29] (Extended Data Fig. 12). Overall, the inner-ring architecture is similar in both yeast and vertebrates, consistent with it being the most conserved part of the NPC[39].

Finally, the structure allows us to rationalize the functional fitness (Fig. 5), the transport through the NPC (Fig. 6 and Extended Data Fig. 11) and the evolution (Extended Data Fig. 10), therefore increasing our confidence in the structure compared to not being able to rationalize these aspects of the NPC[8,131,132].

9.2 Brownian dynamics simulation of FG repeats and NTFs. Distributions of the FG repeats and NTFs were computed by Brownian dynamics simulation[133], using our previously published protocol[34] implemented in IMP[90] version 2.6. The simulated system included the static NPC ring determined in this study, the pore membrane, disordered and flexible FG-repeat domains as well as freely diffusing NTFs and inert macromolecules, all enclosed within a bounding box of $2{,}000 \times 2{,}000 \times 2{,}000$ Å^3.

The pore membrane was represented as a 250 Å slab with a cylindrical pore of radius 375 Å that contains the static NPC ring (this pore membrane representation is simplified compared to the toroidal pore used for solving the structure of the static NPC ring, for reasons of computational efficiency). Each of the FG-repeat domains was represented as a flexible string of beads; a bead had a radius of 6 Å and encompassed 20 residues to achieve a compromise between computational efficiency and accuracy[34,110,134–136]. Consecutive beads were restrained by a bond with an equilibrium length of 18 Å and a constant force of 1.0 kcal per mol per Å, approximating the spring-like nature of flexible polymers[137] in general and FG-repeat domains[110,134–136,138–140] in particular. The freely diffusing molecules included 1,600 NTFs and 1,600 inert macromolecules (0.33 mM each), each consisting of 8 subgroups of 200 macromolecules, ranging in radius from 4 to 28 Å in 2 Å increments (10 to 75 kDa, assuming constant protein density of 1.38 g/cm^2). Excluded volume interactions were applied to pairs of overlapping beads and to beads penetrating the pore membrane or the bounding box, using a constant repulsive force of 10.0 kcal per mol per Å. The potential binding energy between a binding site on an FG motif and a binding site on an NTR was modelled by an anisotropic harmonic potential dependent on the distance and orientation between the two sites that reproduces measured apparent dissociation constants in our simulations (B.R. *et al.*, manuscript in preparation).

The Brownian dynamics of the entire system were simulated at 297.15 K for 40 microseconds with a time step of 1,047 femtoseconds, independently 400 times; the first 10 microseconds of each trajectory were considered equilibration time and ignored in subsequent analysis. The distributions of the FG Nup and NTR positions were then computed from the total of 12 milliseconds of simulations over a cubic grid with a voxel size of $10 \times 10 \times 10$ Å^3, at time intervals of 9.5 picoseconds, from all 400 trajectories; these distributions were averaged by relying on the C_8-symmetry of the NPC.

Code availability. Files containing integrative structure modelling scripts, as well as the input data and output results are available at http://salilab.org/npc2018. The source code for calibrated imaging is available at https://github.com/jayunruh/Jay_Plugins.

Data availability. Source Data for Fig. 1a are provided as an excel file. Original data underlying the image calibration data (source data for Fig. 1 and Extended Data Fig. 3) can be accessed from the Stowers Original Data Repository at http://www.stowers.org/publications/LIBPB-1267.

Raw data for the chemical cross-links (source data for Fig. 2 and Supplementary Table 1) are available via the Zenodo data repository, at http://doi.org/10.5281/zenodo.1149746.

The cryo-ET density map (source data for Fig. 3) is deposited in the Electron Microscopy Data Bank (EMDB) with the accession code EMD-7321.

The cryo-ET raw data (120 tilt series; source data for Extended Data Figs 4–6) are deposited in the Electron Microscopy Public Image Archive (EMPIAR) (https://www.ebi.ac.uk/pdbe/emdb/empiar/) with the accession code EMPIAR-10155 (see Supplementary Power Point Presentation slides 1, 2 as examples).

The integrative NPC structure (source data for Fig. 4) is deposited in the nascent public Protein Data Bank (PDB) repository, PDB-dev (https://pdb-dev.wwpdb.org/), under the accession codes PDBDEV_00000010, PDBDEV_00000011 and PDBDEV_00000012.

Source data for Extended Data Fig. 2b are provided in the Supplementary Information (Supplementary Fig. 1).

SAXS data (source data for Extended Data Fig. 7f) are deposited in the Small Angle Scattering Biological Data Bank (SASBDB; https://www.sasbdb.org/), under the accession codes SASDBV9, SASDBW9, SASDBX9, SASDBY9 and SASDBZ9. In addition, all SAXS data (Supplementary Table 6) are provided as source data with the article.

Raw data for the negative-stain electron microscopy of the native Nic96 complex (source data for Extended Data Fig. 7g) are deposited in the Electron Microscopy Public Image Archive (EMPIAR) (https://www.ebi.ac.uk/pdbe/emdb/empiar/) with the accession code EMPIAR-10162.

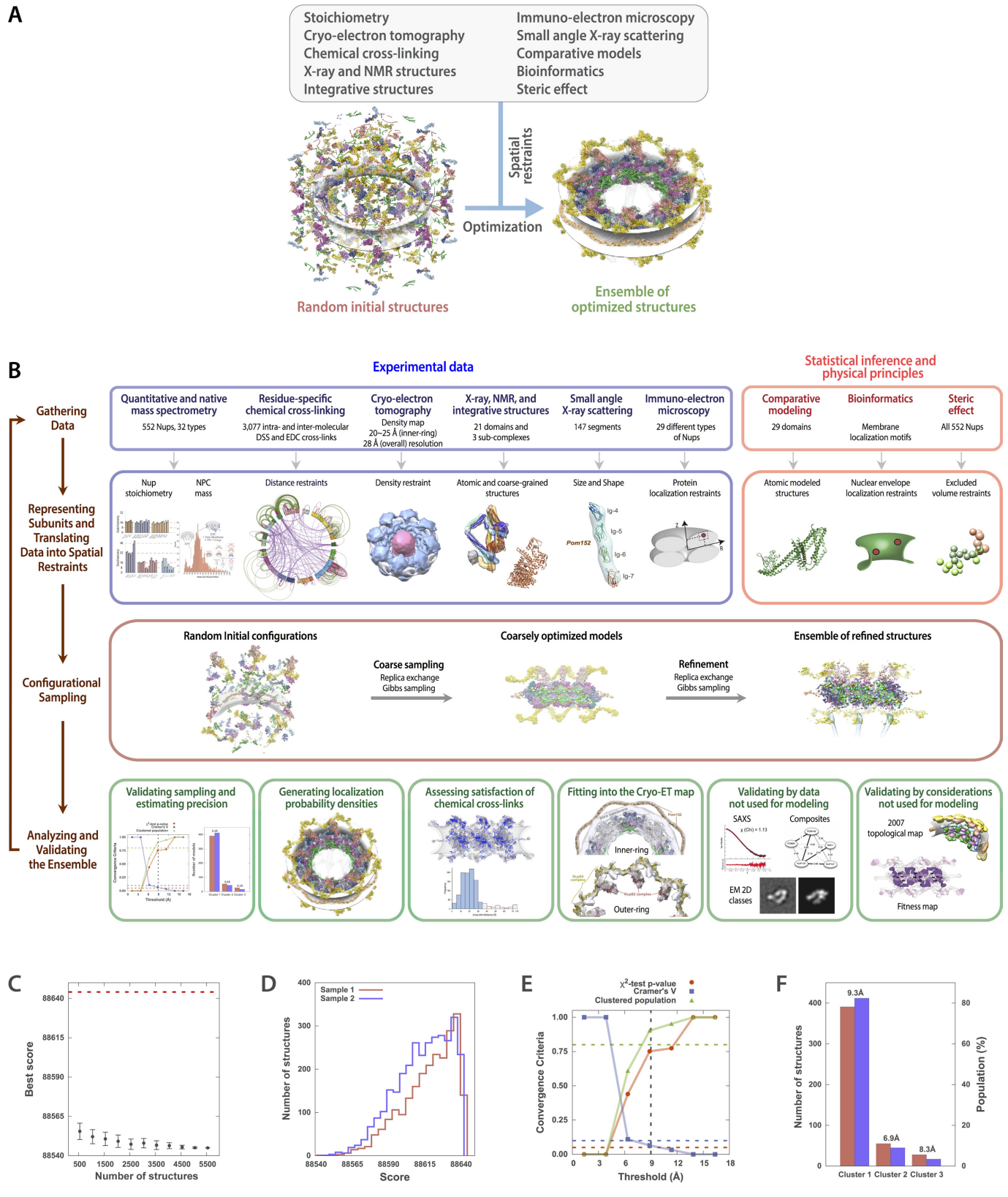
All other data are available from the corresponding author upon reasonable request.

44. Rout, M. P. & Blobel, G. Isolation of the yeast nuclear pore complex. *J. Cell Biol.* **123,** 771–783 (1993).
45. LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z. & Rout, M. P. Protein complex purification by affinity capture. *Cold Spring Harb. Protoc.* http://doi.org/10.1101/pdb.top077545 (2016).
46. LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z. & Rout, M. P. Optimized affinity capture of yeast protein complexes. *Cold Spring Harb. Protoc.* http://doi.org/10.1101/pdb.prot087932 (2016).
47. LaCava, J., Fernandez-Martinez, J. & Rout, M. P. Native elution of yeast protein complexes obtained by affinity capture. *Cold Spring Harb. Protoc.* http://doi.org/10.1101/pdb.prot087940 (2016).
48. Oeffinger, M. *et al.* Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods* **4,** 951–956 (2007).
49. Hakhverdyan, Z. *et al.* Rapid, optimized interactomic screening. *Nat. Methods* **12,** 553–560 (2015).
50. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2,** 587–589 (2005).
51. Shivaraju, M. *et al.* Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. *Cell* **150,** 304–316 (2012).
52. Keifer, D. Z., Motwani, T., Teschke, C. M. & Jarrold, M. F. Measurement of the accurate mass of a 50 MDa infectious virus. *Rapid Commun. Mass Spectrom.* **30,** 1957–1962 (2016).
53. Pratt, J. M. *et al.* Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protocols* **1,** 1029–1043 (2006).
54. Kito, K., Ota, K., Fujita, T. & Ito, T. A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.* **6,** 792–800 (2007).
55. Ding, C. *et al.* Quantitative analysis of cohesin complex stoichiometry and SMC3 modification-dependent protein interactions. *J. Proteome Res.* **10,** 3652–3659 (2011).
56. Rout, M. P. & Kilmartin, J. V. Components of the yeast spindle and spindle pole body. *J. Cell Biol.* **111,** 1913–1927 (1990).
57. Strambio-de-Castillia, C., Blobel, G. & Rout, M. P. Isolation and characterization of nuclear envelopes from the yeast *Saccharomyces*. *J. Cell Biol.* **131,** 19–31 (1995).
58. Rout, M. P. & Strambio-de-Castillia, C. in *Cell Biology: A Laboratory Handbook 2* (ed. Celis, J. E.) 143–151 (Academic, 1998).
59. Cadene, M. & Chait, B. T. A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.* **72,** 5655–5658 (2000).
60. Fenyo, D. *et al.* MALDI sample preparation: the ultra thin layer method. *J. Vis. Exp.* **192,** 192 (2007).
61. Field, H. I., Fenyö, D. & Beavis, R. C. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics* **2,** 36–47 (2002).
62. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).
63. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
64. Contino, N. C., Pierson, E. E., Keifer, D. Z. & Jarrold, M. F. Charge detection mass spectrometry with resolved charge states. *J. Am. Soc. Mass Spectrom.* **24,** 101–108 (2013).
65. Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular assemblies. *Nat. Methods* **12,** 1135–1138 (2015).
66. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11,** M111.014126 (2012).
67. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9,** 904–906 (2012).
68. Cevher, M. A. *et al.* Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat. Struct. Mol. Biol.* **21,** 1028–1034 (2014).
69. Sun, J. *et al.* The architecture of a eukaryotic replisome. *Nat. Struct. Mol. Biol.* **22,** 976–982 (2015).
70. Zheng, S. Q. *et al.* UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J. Struct. Biol.* **157,** 138–147 (2007).
71. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116,** 71–76 (1996).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

72. Galaz-Montoya, J. G., Flanagan, J., Schmid, M. F. & Ludtke, S. J. Single particle tomography in EMAN2. *J. Struct. Biol.* **190,** 279–290 (2015).
73. Galaz-Montoya, J. G. *et al.* Alignment algorithms and per-particle CTF correction for single particle cryo-electron tomography. *J. Struct. Biol.* **194,** 383–394 (2016).
74. Yang, Q., Rout, M. P. & Akey, C. W. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* **1,** 223–234 (1998).
75. Beck, M. *et al.* Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306,** 1387–1390 (2004).
76. von Appen, A. & Beck, M. Structure determination of the nuclear pore complex with three-dimensional cryo electron microscopy. *J. Mol. Biol.* **428,** 2001–2010 (2016).
77. Ludtke, S. J. Single-particle refinement and variability analysis in EMAN2.1. *Methods Enzymol.* **579,** 159–189 (2016).
78. Iwanczyk, J. *et al.* Structure of the Blm10–20 S proteasome complex by cryo-electron microscopy. Insights into the mechanism of activation of mature yeast proteasomes. *J. Mol. Biol.* **363,** 648–659 (2006).
79. Elad, N. *et al.* The dynamic conformational landscape of γ-secretase. *J. Cell Sci.* **128,** 589–598 (2015).
80. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).
81. Bharat, T. A. & Scheres, S. H. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protocols* **11,** 2054–2065 (2016).
82. Bharat, T. A., Russo, C. J., Löwe, J., Passmore, L. A. & Scheres, S. H. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23,** 1743–1753 (2015).
83. Sampathkumar, P. *et al.* Atomic structure of the nuclear pore complex targeting domain of a Nup116 homologue from the yeast, *Candida glabrata*. *Proteins* **80,** 2110–2116 (2012). 10.1002/prot.24102
84. Sampathkumar, P. *et al.* Structure of the C-terminal domain of *Saccharomyces cerevisiae* Nup133, a component of the nuclear pore complex. *Proteins* **79,** 1672–1677 (2011).
85. Sampathkumar, P. *et al.* Structures of the autoproteolytic domain from the *Saccharomyces cerevisiae* nuclear pore complex component, Nup145. *Proteins* **78,** 1992–1998 (2010). 10.1002/prot.22707
86. Sampathkumar, P. *et al.* Structure, dynamics, evolution, and function of a major scaffold component in the nuclear pore complex. *Structure* **21,** 560–571 (2013).
87. Edelstein, A. D. *et al.* Advanced methods of microscope control using μManager software. *J. Biol. Methods* **1,** e10 (2014).
88. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
89. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20,** 237–247 (2012).
90. Russel, D. *et al.* Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10,** e1001244 (2012).
91. Schneidman-Duhovny, D., Pellarin, R. & Sali, A. Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* **28,** 96–104 (2014).
92. Sali, A. *et al.* Outcome of the first wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **23,** 1156–1167 (2015).
93. LoPiccolo, J. *et al.* Assembly and molecular architecture of the phosphoinositide 3-kinase p85α homodimer. *J. Biol. Chem.* **290,** 30390–30405 (2015).
94. Luo, J. *et al.* Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Mol. Cell* **59,** 794–806 (2015).
95. Robinson, P. J. *et al.* Molecular architecture of the yeast Mediator complex. *eLife* **4,** e08719 (2015).
96. Webb, B. *et al.* Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol. Biol.* **1091,** 277–295 (2014).
97. Cherry, J. M. *et al. Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40,** D700–D705 (2012).
98. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific α-helical properties. *Bioinformatics* **24,** 2101–2102 (2008).
99. Niepel, M. *et al.* The nuclear basket proteins Mlp1p and Mlp2p are part of a dynamic interactome including Esc1p and the proteasome. *Mol. Biol. Cell* **24,** 3920–3938 (2013).
100. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
101. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33,** W244–W248 (2005).
102. Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21,** 951–960 (2005).
103. 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).
104. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195–202 (1999).
105. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20,** 2138–2139 (2004).
106. Marsden, R. L., McGuffin, L. J. & Jones, D. T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci.* **11,** 2814–2824 (2002).
107. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252,** 1162–1164 (1991).
108. Trigg, J., Gutwin, K., Keating, A. E. & Berger, B. Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. *PLoS ONE* **6,** e23519 (2011).
109. Algret, R. *et al.* Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. *Mol. Cell. Proteomics* **13,** 2855–2870 (2014).
110. Raveh, B. *et al.* Slide-and-exchange mechanism for rapid and selective transport through the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **113,** E2489–E2497 (2016).
111. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38,** W540–W544 (2010).
112. Schneidman-Duhovny, D., Kim, S. J. & Sali, A. Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct. Biol.* **12,** 17 (2012).
113. Kawabata, T. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a Gaussian mixture model. *Biophys. J.* **95,** 4643–4658 (2008).
114. Jonić, S. *et al.* Denoising of high-resolution single-particle electron-microscopy density maps by their approximation using three-dimensional Gaussian functions. *J. Struct. Biol.* **194,** 423–433 (2016).
115. Hanot, S. *et al.* Multi-scale Bayesian modeling of cryo-electron microscopy density maps. *Preprint at* https://www.biorxiv.org/content/early/2018/02/09/113951 (2017).
116. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148,** 635–651 (2000).
117. Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15,** 2507–2524 (2006).
118. Campelo, F. & Kozlov, M. M. Sensing membrane stresses by protein insertions. *PLOS Comput. Biol.* **10,** e1003556 (2014).
119. Viswanath, S., Chemmama, I. E., Cimermancic, P. & Sali, A. Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophys. J.* **113,** 2344–2353 (2017).
120. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, 1956).
121. McCarroll, D. *Simple Statistical Tests for Geography* (Chapman and Hall/CRC, 2016)
122. McDonald, J. H. *Handbook of Biological Statistics 3rd edn* (Sparky House, 2014).
123. Daura, X. *et al.* Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed. Engl.* **38,** 236–240 (1999).
124. Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* **98,** 861–871 (2010).
125. Read, R. J. *et al.* A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* **19,** 1395–1412 (2011).
126. Montelione, G. T. *et al.* Recommendations of the wwPDB NMR Validation Task Force. *Structure* **21,** 1563–1570 (2013).
127. Henderson, R. *et al.* Outcome of the first Electron Microscopy Validation Task Force meeting. *Structure* **20,** 205–214 (2012).
128. Trewhella, J. *et al.* Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. *Structure* **21,** 875–881 (2013).
129. Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl Acad. Sci. USA* **111,** 9455–9460 (2014).
130. Erzberger, J. P. *et al.* Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex. *Cell* **158,** 1123–1135 (2014).
131. Alber, F., Förster, F., Korkin, D., Topf, M. & Sali, A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77,** 443–477 (2008).
132. Alber, F., Chait, B. T., Rout, M. P. & Sali, A. in *Protein–Protein Interactions and Networks: Identification, Characterization and Prediction* (eds Panchenko, A. & Przytycka, T.) 99–114 (Springer, 2008).
133. Ermak, D. L. & Mccammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **69,** 1352–1360 (1978).
134. Hough, L. E. *et al.* The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *eLife* **4,** e10027 (2015).
135. Milles, S. *et al.* Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163,** 734–745 (2015).
136. Sakiyama, Y., Mazur, A., Kapinos, L. E. & Lim, R. Y. Spatiotemporal dynamics of the nuclear pore complex transport barrier resolved by high-speed atomic force microscopy. *Nat. Nanotechnol.* **11,** 719–723 (2016).
137. van der Maarel, J. R. C. *Introduction to Biopolymer Physics* (World Scientific, 2008).
138. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl Acad. Sci. USA* **100,** 2450–2455 (2003).
139. Lemke, E. A. The multiple faces of disordered nucleoporins. *J. Mol. Biol.* **428,** 2011–2024 (2016).
140. Lim, R. Y. *et al.* Flexible phenylalanine–glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA* **103,** 9512–9517 (2006).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

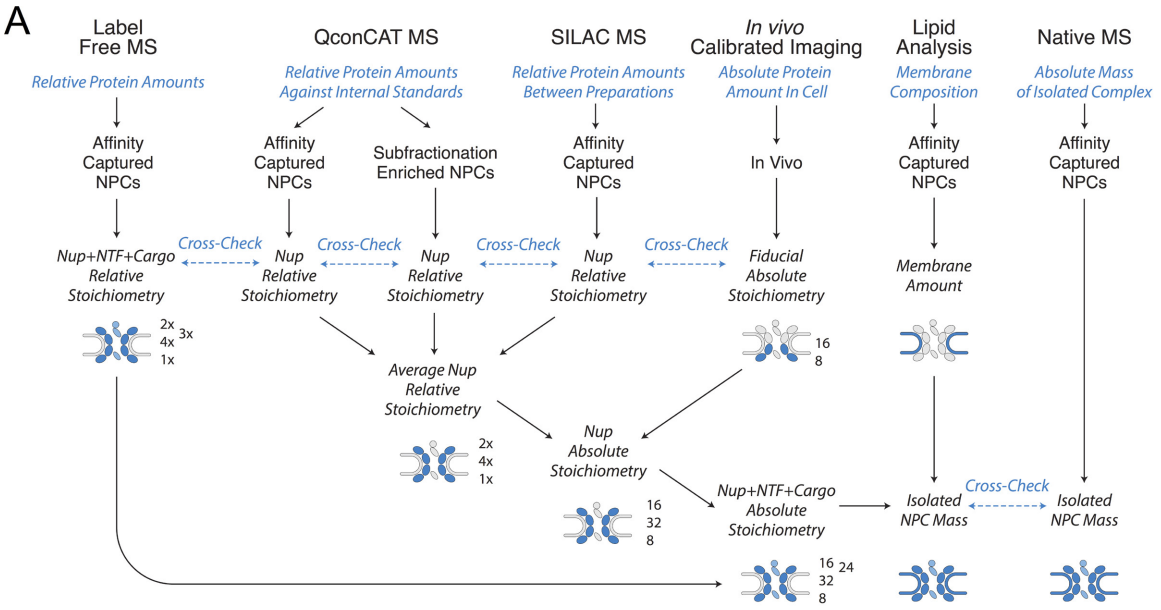


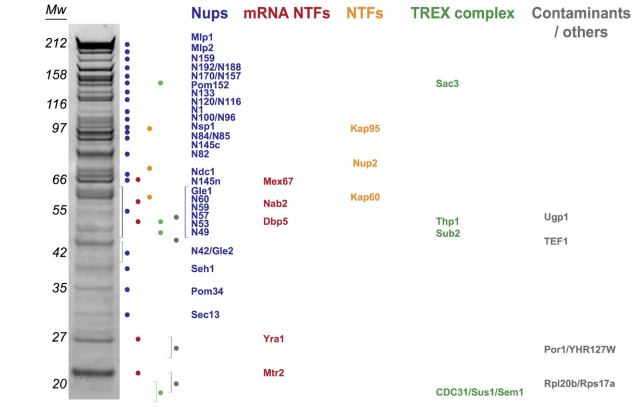
Extended Data Figure 1 | See next page for caption.

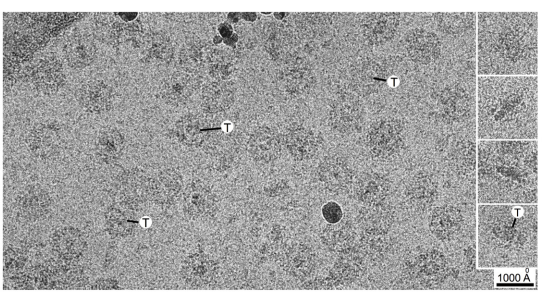
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

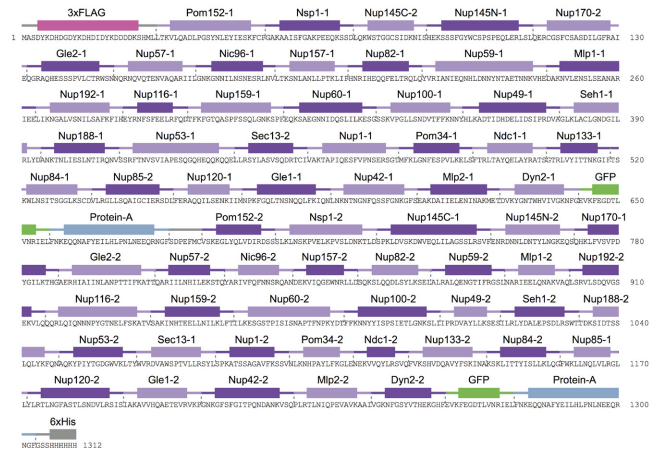
Extended Data Figure 1 | Integrative structure determination of the *S. cerevisiae* NPC at 9 Å precision. a, Schematic of integrative structure determination of the *S. cerevisiae* NPC. Random initial structures of the Nups and their sub-complexes were optimized by satisfying spatial restraints implied by the input information. **b**, The full description of integrative structure determination of the *S. cerevisiae* NPC, proceeded through four stages[8,90–92] (Supplementary Table 3): (1) gathering data, (2) representing subunits and translating data into spatial restraints, (3) configurational sampling to produce an ensemble of structures that satisfies the restraints and (4) analysing and validating the ensemble structures and data (Extended Data Figs 7, 8 Supplementary Tables 2–4 and Methods). The integrative structure modelling protocol (stages 2, 3 and 4) was scripted using the Python modelling interface package version 4d97507, which is a library for modelling macromolecular complexes based on our open-source IMP package[90] version 2.6 (https://integrativemodeling.org). **c**, Convergence of the structure score for the 5,529 good-scoring NPC structures; the scores do not continue to improve as more structures are computed, essentially independently of each other. The error bar represents the standard deviations of the best scores, estimated by repeating sampling of NPC structures ten times ($n = 10$, mean score values plotted). The red dotted line indicates the total score threshold (88,644.1) that defines the good-scoring NPC structures (Methods). **d**, Distribution of scores for structure samples 1 (red) and 2 (blue), comprising the 5,529 good-scoring NPC structures ($n^{\text{sample1}} = 2{,}359$ and $n^{\text{sample2}} = 3{,}170$ structures). The non-parametric Kolmogorov–Smirnov two-sample test[120,121] (two-sided) indicates that the difference between the two score distributions is insignificant (P value $(1.0) > 0.05$). In addition, the magnitude of the difference is small, as demonstrated by the Kolmogorov–Smirnov two-sample test statistic, D, of 0.045. Thus, the two score distributions are effectively equal. **e**, Three criteria for determining the sampling precision (y axis), evaluated as a function of the r.m.s.d. clustering threshold[123] (x axis) ($n = 5{,}529$ structures). First, the P value is computed using the χ^2-test (one-sided) for homogeneity of proportions[122] (red dots). Second, an effect size for the χ^2-test is quantified by the Cramer's V value (blue squares). Third, the population of structures in sufficiently large clusters (containing at least ten structures from each sample) is shown as green triangles. The vertical dotted grey line indicates the r.m.s.d. clustering threshold at which three conditions are satisfied (χ^2-test P value $(0.75) > 0.05$ (red, horizontal dotted line), Cramer's V $(0.065) < 0.10$ (blue, horizontal dotted line) and the population of clustered structures $(0.90) > 0.80$ (green, horizontal dotted line)), thus defining the sampling precision of 9 Å. The three solid curves (in red, blue and green) were drawn through the points to help visualize the results. **f**, Population of sample 1 and 2 structures in the three clusters obtained by threshold-based clustering[123] using an r.m.s.d. threshold of 12 Å. The dominant cluster (cluster 1) contains 80.3% of the structures. Cluster precision is shown for each cluster. The precision of the dominant cluster defines the structure precision.

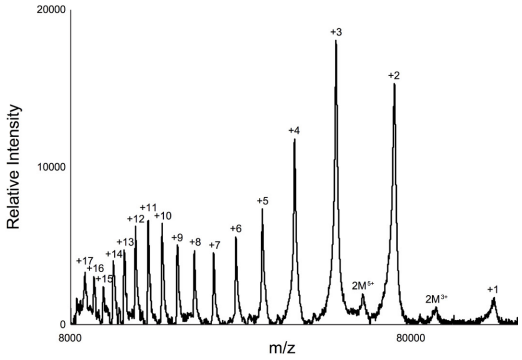
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.









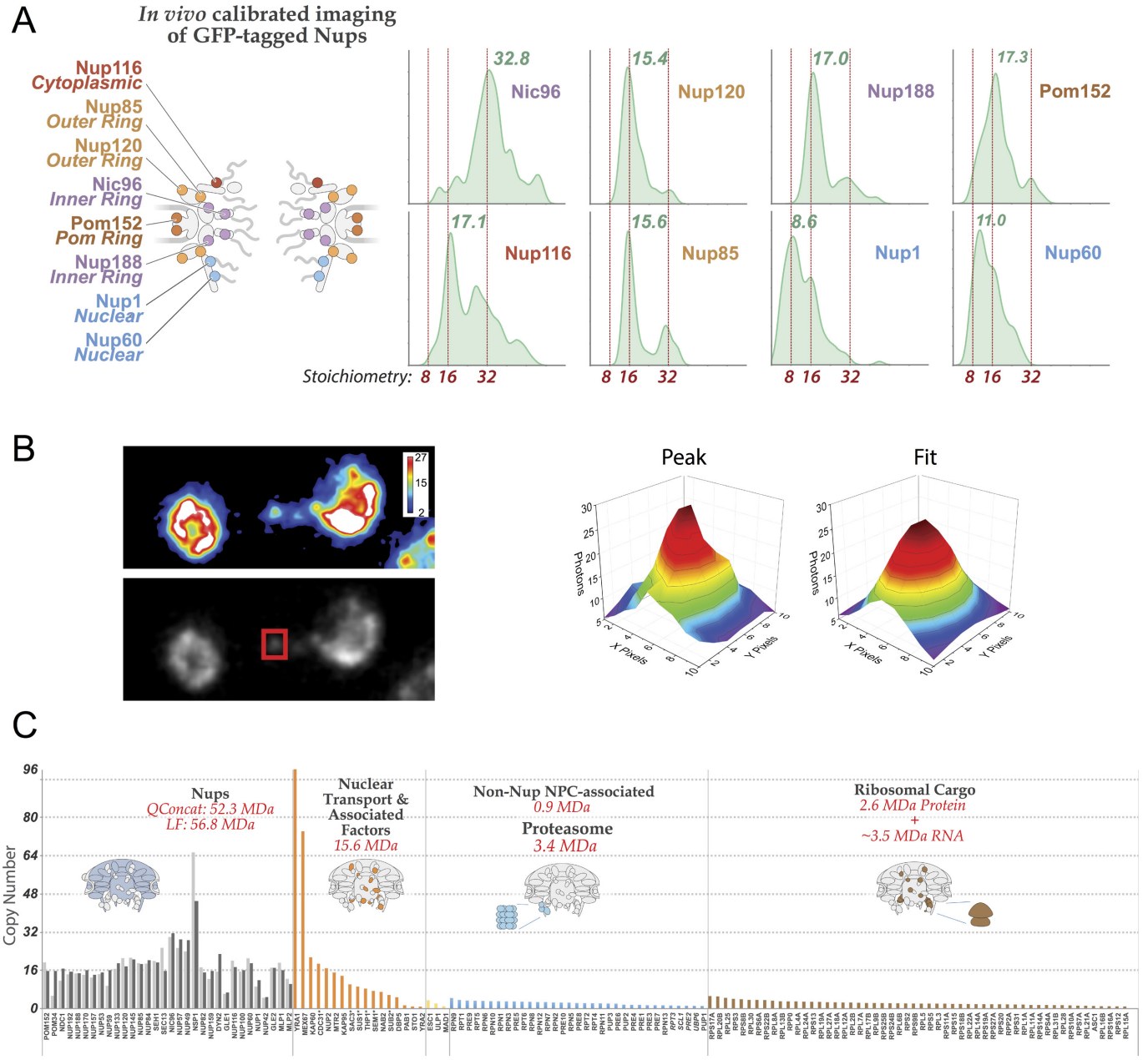


Extended Data Figure 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

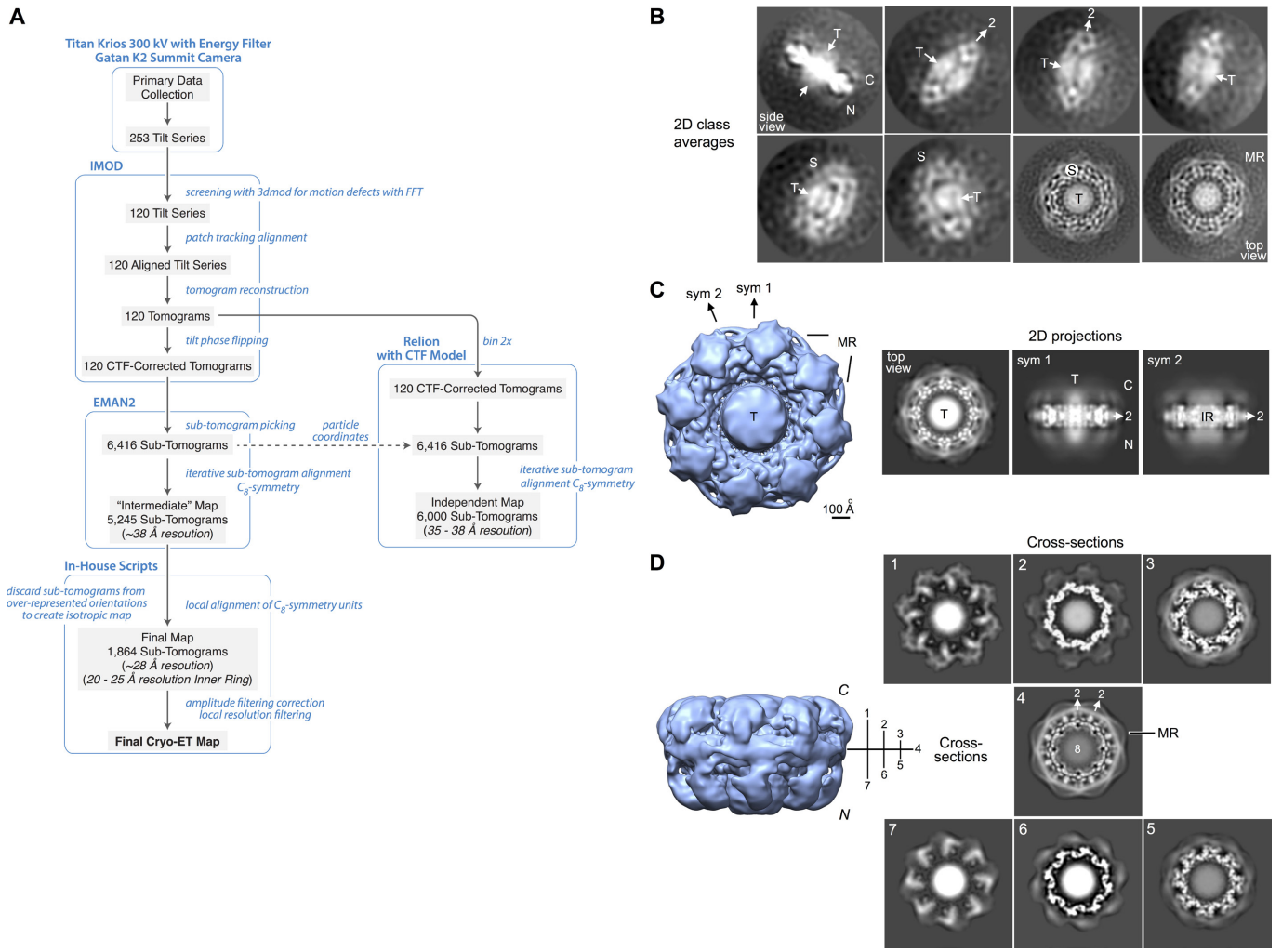
Extended Data Figure 2 | Quantitative analysis of the mass and stoichiometry of the endogenous NPC (part 1). a, A multipronged approach to accurately define the mass, stoichiometry and composition of native macromolecular assemblies. Schematic is shown of the multiple orthologous methods that are integrated within our strategy for the analysis of native assemblies. The main experimental methods are listed on top, followed by the characteristic that they help to quantify (in blue) and the type of sample to which they were applied. The final outcome of each method is indicated (black arrows); the steps of each method that are compared to serve as a cross-check control are indicated (blue dashed lines). At bottom, the integration of the different data points into a final comprehensive description of the endogenous assembly is depicted. Small cartoon insets of the NPC are used to illustrate the analysis. **b**, SDS-PAGE analysis of the affinity-captured *S. cerevisiae* NPCs isolated from an MLP1–PPX-PrA tagged strain ($n > 20$ independent experiments). Molecular weight marker values (Mw) are indicated to the left of the gel lane. Dots signify the main protein components of the isolated NPCs as identified by liquid chromatography–mass spectrometry (Extended Data Fig. 3c). Proteins are grouped and coloured by functional categories or membership of discrete macromolecular assemblies. Nups, blue; mRNA transport factors (TFs), red; transport factors, orange; transcription-export (TREX), green; contaminants and/or others, grey. For gel source data, see Supplementary Fig. 1. **c**, Cryo-electron microscopy analysis of the affinity-captured NPCs ($n > 20$ independent experiments). The particles have a clear preferred orientation (Methods). Some side views are presented in the inset. The central transporter is present in every NPC (indicated by 'T'). Scale bar, 1,000 Å. **d**, Schematic showing the primary amino acid sequence of the 148.2 kDa synthetic QconCAT-A. It includes two peptides for each Nup (thick bars), arranged in the indicated order. The native three amino acid residues flanking regions (thin bars) of each peptide were included to preserve the native trypsin target sequence. A N-terminal 3×FLAG tag was included, as well as a C-terminal 6×His tag for purification under denaturing conditions. The stringent criteria used for the selection of the QconCAT peptides are described in the Methods. **e**, MALDI mass spectrometry spectrum of intact, purified full-length QconCAT-A labelled with stable isotope, showing that a single species was detected. Numbers above peaks denote the QconCAT-A protein species with n positive charges and the 2M QconCAT-A protein dimer. The measured molecular weight of the QconCAT-A labelled with stable isotope was 149,049 ± 38 Da, consistent with its calculated molecular weight of 148,200 Da (Methods).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



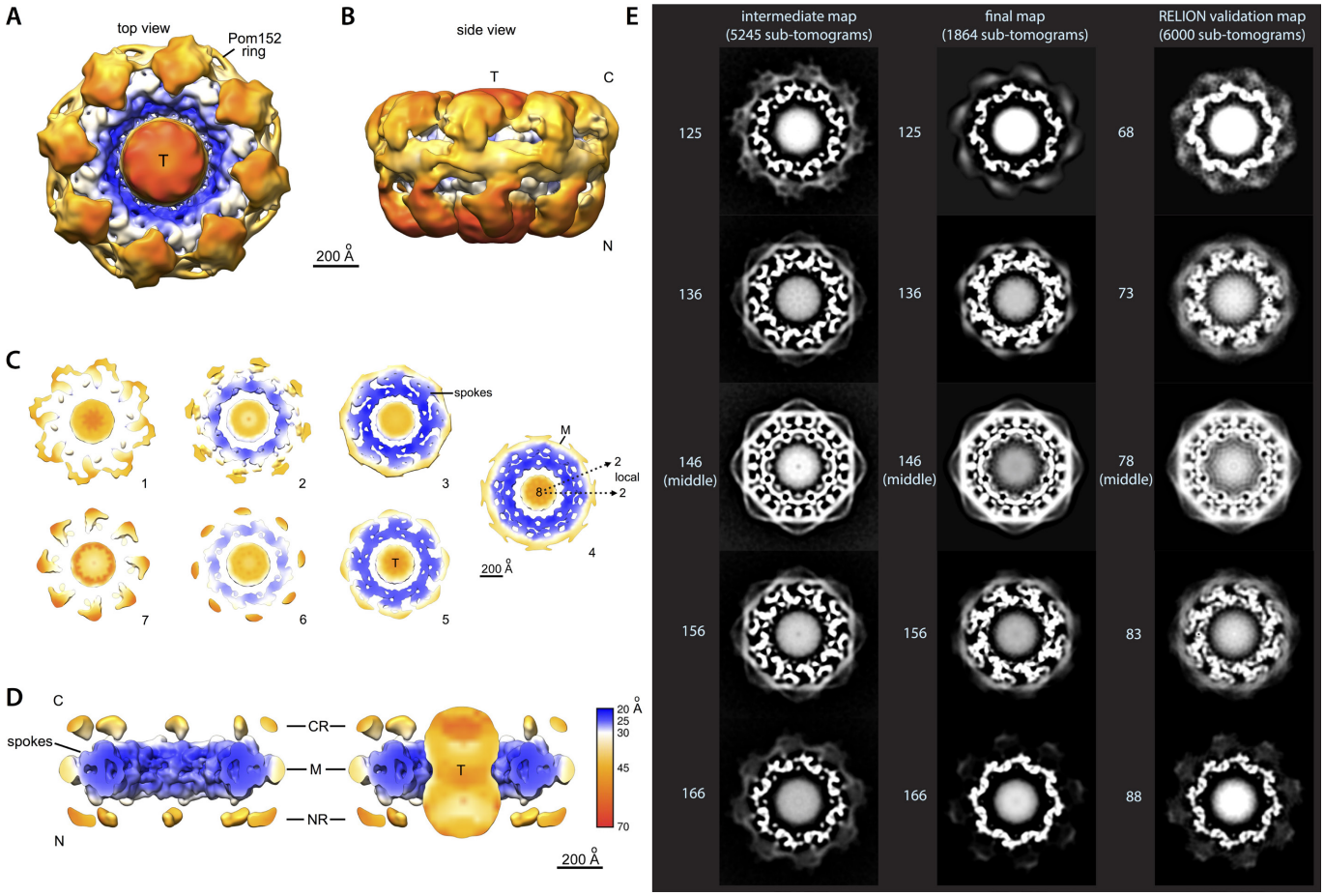
Extended Data Figure 3 | Quantitative analysis of the mass and stoichiometry of the endogenous NPC (part 2). a, Left, schematic localization of the Nup–GFP reporters selected for the *in vivo* calibrated imaging stoichiometry analyses. Nups were selected to represent every major NPC module and to provide comprehensive coverage of the assembly. Right, Kernel density estimation of distributions of GFP proteins per Nup were calculated from the calibrated imaging data. $n = 48–178$. **b**, Heat map of a yeast cell expressing Nup85–GFP. Image (left) was acquired as described in Methods. In addition, for illustration purposes, a maximum projection along the *z* axis was performed, and the image was smoothed with a Gaussian blur of radius 1. A heat map was used to illustrate intensity units in raw photon counts. For the area outlined in a red rectangle, a 2D distribution of photon counts and the corresponding Gaussian fit are shown (right). **c**, Stoichiometries of main components associated with the affinity-captured NPCs, as determined by label-free mass spectrometry quantification (at least three peptides per protein). Proteins are grouped and coloured by functional categories or membership of discrete macromolecular assemblies. The TREX complex components are included in the 'Transport & Associated Factors' category and labelled with an asterisk. QconCAT-derived stoichiometries for all the Nups (dark grey bars) are shown for comparison.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



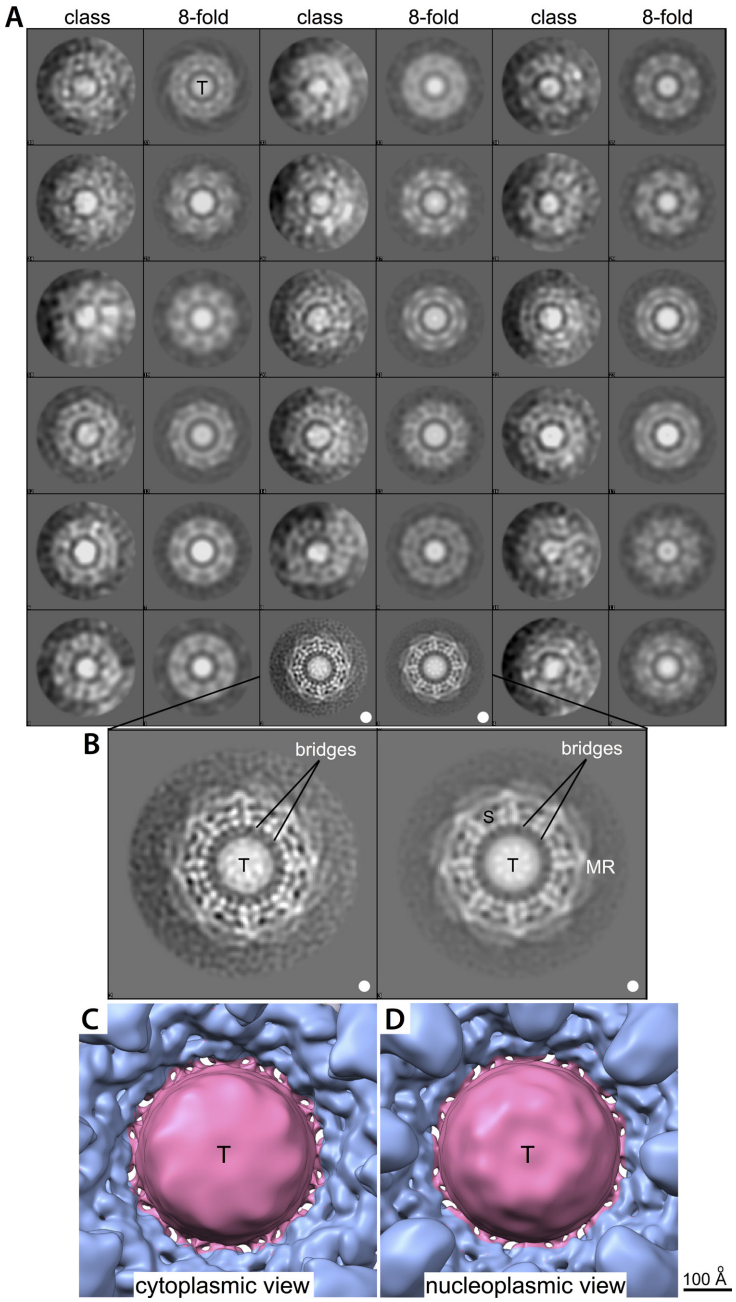
Extended Data Figure 4 | Cryo-ET strategy and the resulting 3D cryo-ET map of the NPC exhibiting non-enforced local C_2-symmetry axes in the inner ring. a, Diagram describing the methodology used to obtain the whole *S. cerevisiae* NPC cryo-ET map (Methods). **b**, 2D class averages are shown (protein in white), which were calculated using the original unaligned sub-tomograms projected along the *z* axis. The overall thickness of the *S. cerevisiae* NPC is apparent in a side-view class, and the local C_2-symmetry axes in the inner ring are also apparent (indicated with '2'). The transporter density is present in every class. **c**, Left, top view of the cryo-ET map with the two local C_2-symmetry axes indicated by arrows and labels (sym 1 and sym 2). They are 22.5° apart, owing to the C_8-symmetry axis. Right, 2D projections of the top view, and two side views along the two local C_2-symmetry axes (side 1 and side 2 projected along axes sym 1 and sym 2, respectively). **d**, Seven cross-sections of the cryo-ET map are shown on the right (labelled 1–7) with their positions in the 3D map indicated in the side view on the left. The local C_2-symmetry of the inner ring is apparent in cross-sections 2–6, mirrored about the central section in panel 4. Labels throughout: C, cytoplasm; N, nucleoplasm; T, central transporter; S, core scaffold; MR, membrane ring; IR, inner ring. Scale bar, 100 Å.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



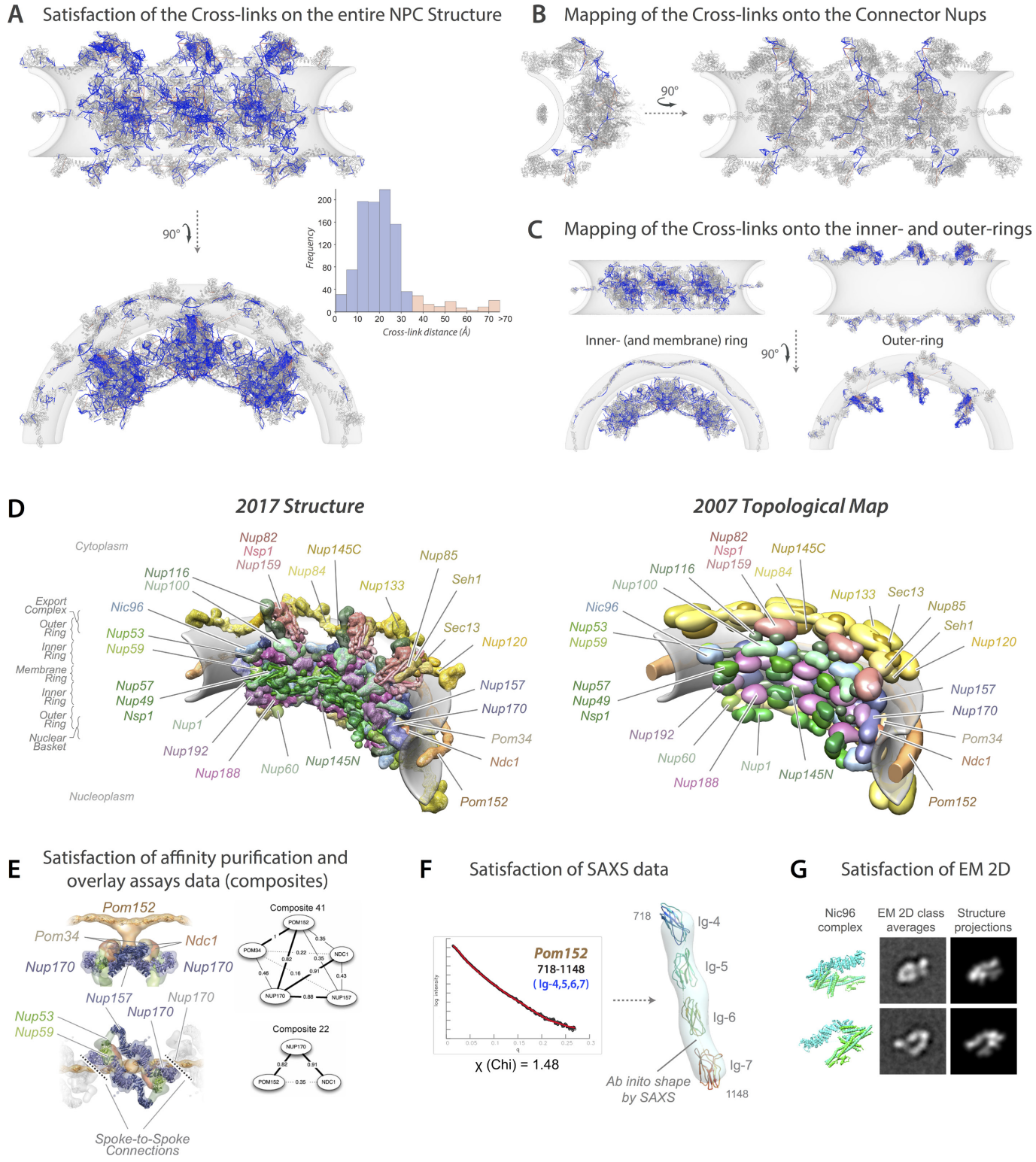
Extended Data Figure 5 | Resolution estimates for the cryo-ET map of the NPC and comparison of cross-sections between the intermediate, final and RELION cryo-ET maps. a, **b**, Top (**a**) and side (**b**) views of the cryo-ET map are colour-coded according to local resolution estimates (colour bar), and are shown at a low threshold to reveal weaker density features at the periphery of the NPC that are more flexible. **c**, Cross-sections are shown at a reduced scale, colour coded according to local resolution estimates (colour bar). A remnant of the pore membrane (M) is present, encircling the entire mid-line of the NPC. Sections 3–5 are shown at a higher threshold. In section 3, the inner-ring region is indicated by 'spokes'. In section 4, local C_2-symmetry axes are indicated by dashed arrows. **d**, Thick sections of the inner ring are shown at a higher threshold, as viewed along the membrane plane. Note that the inner ring (indicated) is almost entirely in the 20–25 Å resolution range. C, cytoplasmic side; N, nuclear side; CR, cytoplasmic ring; NR, nuclear ring; M, pore membrane; T, central transporter. **e**, Comparison of five cross-sections (cross-section number on left) in the inner-ring region of the NPC between cryo-ET maps in different stages of the reconstruction process (Extended Data Fig. 4a): intermediate map (left column), final map (middle column) and an independent validation map, reconstructed with RELION at a twice-reduced Å per pixel size (right column). Details on the reconstruction of maps are provided in Methods.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | 2D classification of projections from 1,864 original NPC sub-tomograms aligned with their C_8-symmetry axis nearly along the *z* axis. a, In total, 18 good class averages are shown after maximum likelihood classification (using RELION 1.4[81,82]) without symmetry imposed. Each class average (on the left) is paired with a C_8-symmetry enforced image of itself (on the right). Central transporter densities are present in each of the class averages (both with and without imposed C_8-symmetry), indicating that the central transporter is generally present in these particles. **b**, An expanded view of a large class from **a** shows bridges (indicated) between the core scaffold and the central transporter, both before (left) and after averaging (right) using the C_8-symmetry of the NPC. S, core scaffold; MR, membrane ring; T, central transporter. Matching panels in **a** and **b** are marked with white dots. **c**, **d**, The cryo-ET 3D map is presented as in Fig. 3 and is zoomed in to show the meshwork of bridges between the scaffold and the central transporter, as viewed from the cytoplasm and the nucleoplasm, respectively.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

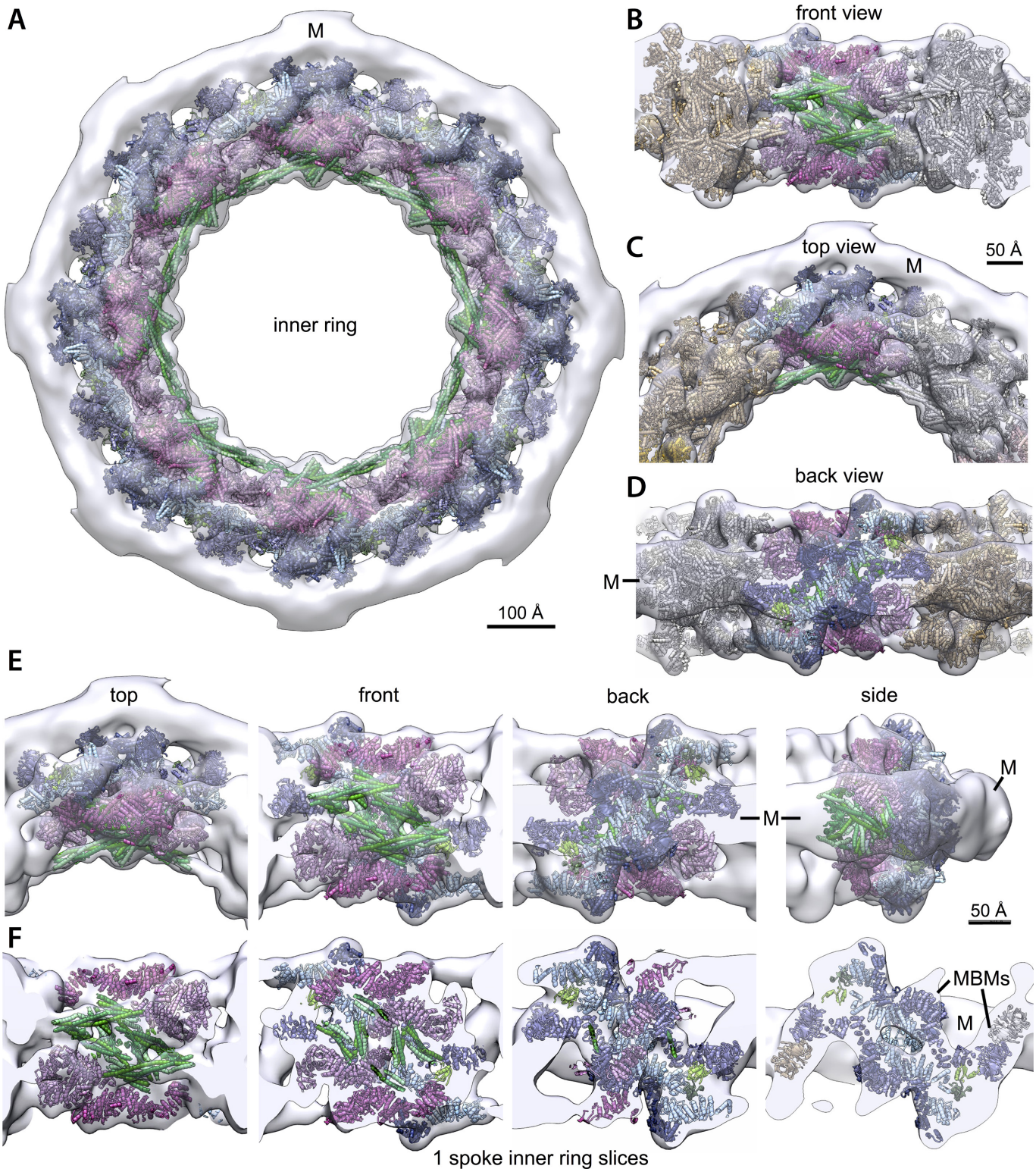


Extended Data Figure 7 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

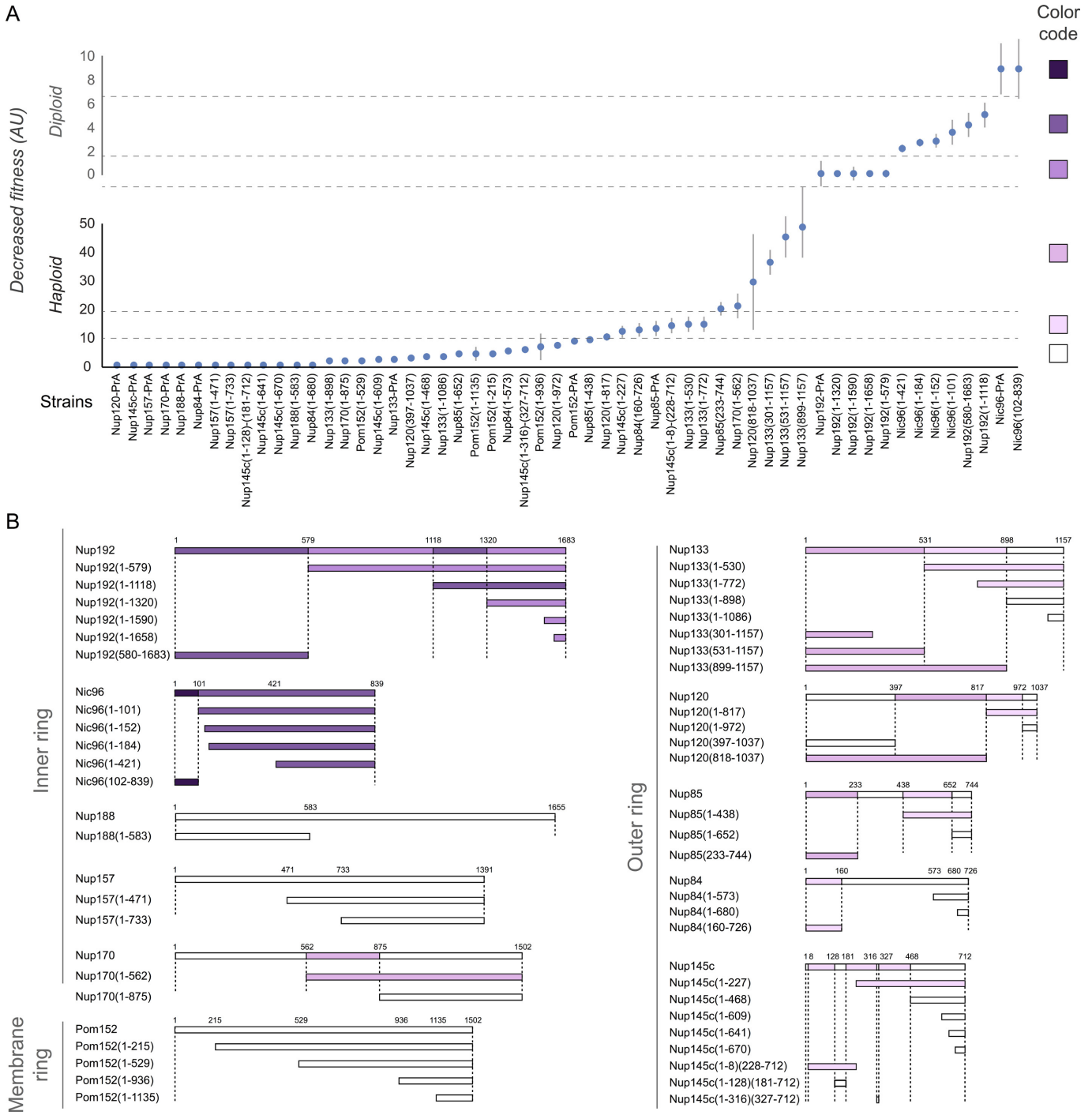
Extended Data Figure 7 | Validation of the NPC structure (part 1). **a**–**c**, Satisfaction of the chemical cross-links. **a**, Identified chemical cross-links were mapped onto the integrative structure of the entire NPC, as shown in front (upper) and top (lower) views. Satisfied cross-links, with C_α–C_α distances that fall within the distance threshold of 35 Å in at least one good-scoring NPC structure, are shown in blue. Violated cross-links, with C_α–C_α distances that are larger than 35 Å, are shown in orange. The histogram on the right shows the distribution of the cross-linked C_α–C_α distances, validating the NPC structure. **b**, Mapping of the cross-links onto the cytoplasmic and nucleoplasmic connector Nups (Nup116, Nup100, Nup145N, Nup1 and Nup60). Front (right) and side (left) views show how the NPC outer rings are connected to the inner ring through a network of connector Nups across the length of the spoke. **c**, Mapping of the cross-links onto the inner (and membrane) and outer rings, in front (upper) and top (lower) views. **d**–**g**, Satisfaction of data and considerations that were not used to compute the structure. **d**, Our integrative structure of the NPC (left) was compared a previously published topological map[3,8] (right). The two structures are consistent with each other, though our integrative structure is defined at an order of magnitude higher precision. **e**, Satisfaction of affinity purification and overlay assays data (composites); our current structure satisfies all 82 composites determined by affinity purification and overlay assays[3,8], even though these data were not used in its determination. For example, Pom152, Pom34, Ndc1, Nup157 and Nup170 are connected with each other (left), consistent with the previously published composites determined using the affinity purification data[3,8] (right). **f**, Satisfaction of SAXS data; the atomic structures of eight Nups are consistent with the corresponding SAXS profiles for their constructs[9,12,23,83–86] (Supplementary Tables 2, 6 and Methods). For example, the SAXS profile calculated from the atomic structure of Pom152$^{718–1148}$ (red curve) using FoXS[111] is well-matched ($\chi = 1.48$) to the corresponding experimental SAXS profile[23] (black dots; $n = 20$ exposures). For visualization purposes, the Pom152$^{718–1148}$ structure (represented as a ribbon) is shown along with the best fit of the *ab initio* shape (represented as a transparent envelope) computed from the experimental SAXS profile. **g**, Satisfaction of the negative-stain electron microscopy 2D class averages for the Nic96 complex; the structures of the Nic96 complex (composed of Nic96, Nsp1, Nup49 and Nup57) in the dominant cluster can be projected well on 2D class averages obtained for the natively isolated complex ($n = 5{,}458$ particles; Methods). The experimental class averages were satisfied by the structure with cross-correlation coefficients of 0.85 and 0.80, respectively (Methods).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



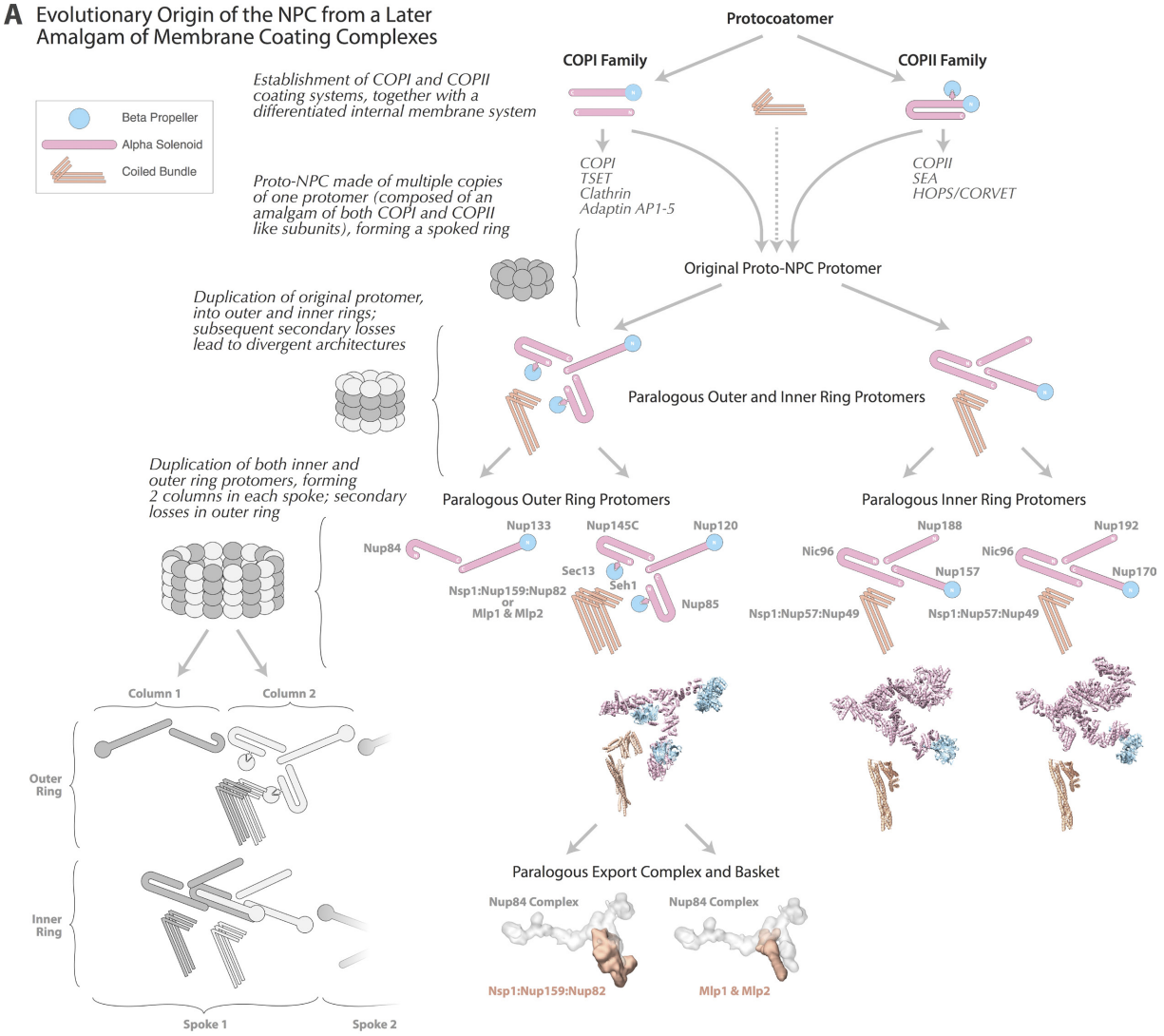
Extended Data Figure 8 | Validation of the NPC structure (part 2), showing consistency between the NPC structure and the cryo-ET density map. The cryo-ET density map is shown at a high-density threshold (grey) to reveal details of the inner ring. A representative structure of the inner ring is shown docked into the density, showing the excellent fit. All Nups are coloured as in Fig. 4. The pore membrane is indicated by M. **a**, Full eight-spoke inner ring (scale bar, 100 Å). **b**–**d**, front (**b**), top (**c**) and back (**d**) views of three spokes with neighbours coloured brown and grey (scale bar, 50 Å). **e**, Different views of a single spoke (scale bar, 50 Å) are shown within the density map. **f**, Thick cross-sections are shown through a single spoke in the inner ring, as viewed from the central C_8-symmetry axis (scale bar, 50 Å); MBMs (see Fig. 5) are indicated.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Functional analysis of the fitness of nucleoporin mutants using ODELAY. a, The fitness defect phenotype was quantified and plotted (mean *Z*-score; $n = 6$ experiments, containing at least 200–300 individuals per point; see Methods for details) for each nucleoporin truncation or C-terminal protein-A tagged mutant in order of decreasing fitness (increasing number of units), as observed by ODELAY assay[14] (Methods). Strains for which truncations in a haploid background were found to lead to lethality after tetrad dissection (Nic96 and Nup192) were assigned the maximum level of defect and plotted on top of the rest (diploid), on the basis of the fitness phenotype observed for the indicated Nic96 and Nup192 mutants in a diploid background (in which a wild-type copy of the nucleoporin is also present and expressed). Six divisions were assigned based on decreasing levels of fitness[9,13]; white (wild type) to dark purple (severe defect). AU, arbitrary unit; error bar = standard deviation. **b**, Mapping of the colour code described in **a** into the NPC components. Horizontal lines represent the amino acid residue length of each protein and truncated version; amino acid residue positions are shown on top of the lines.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

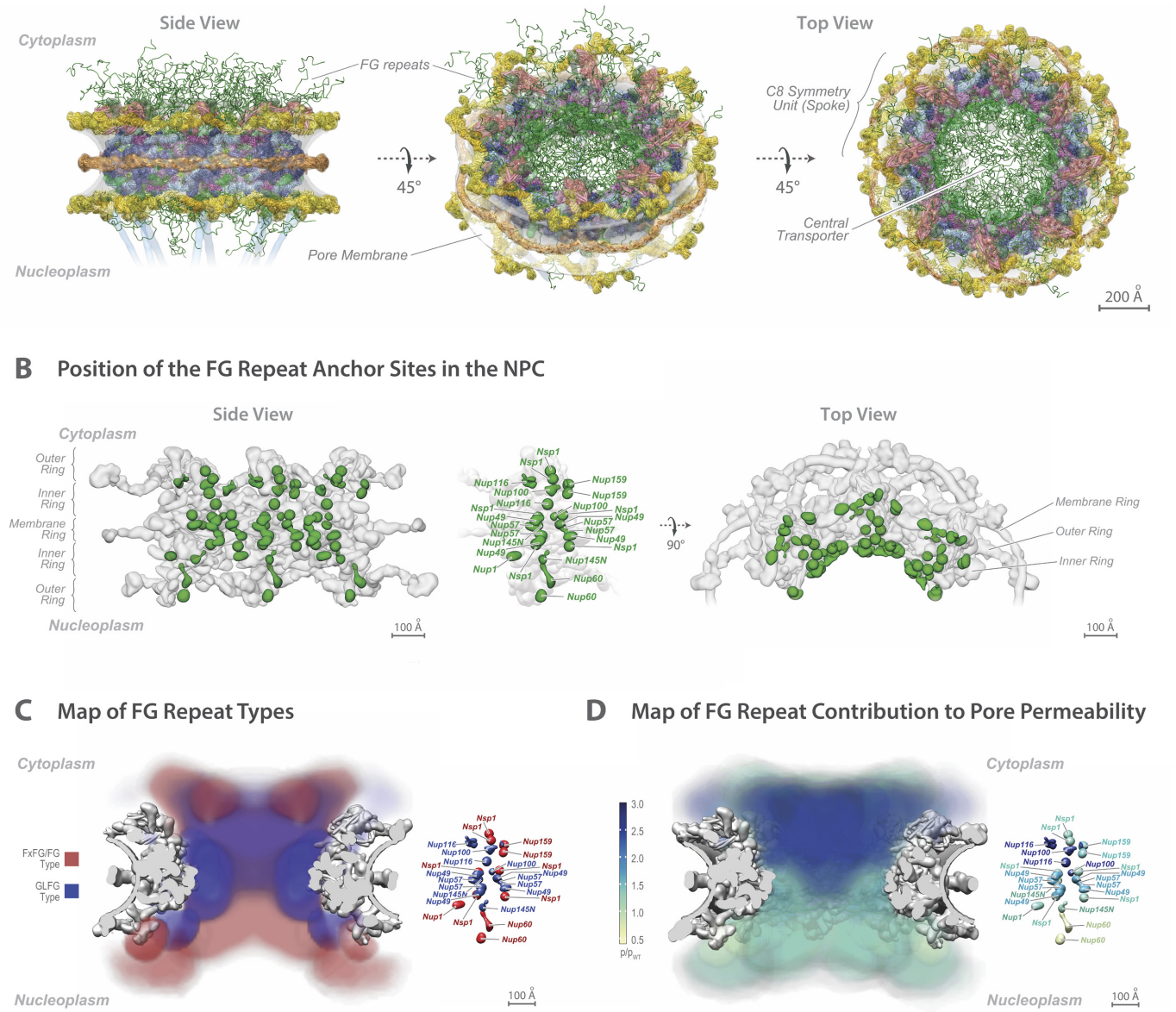


Extended Data Figure 10 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

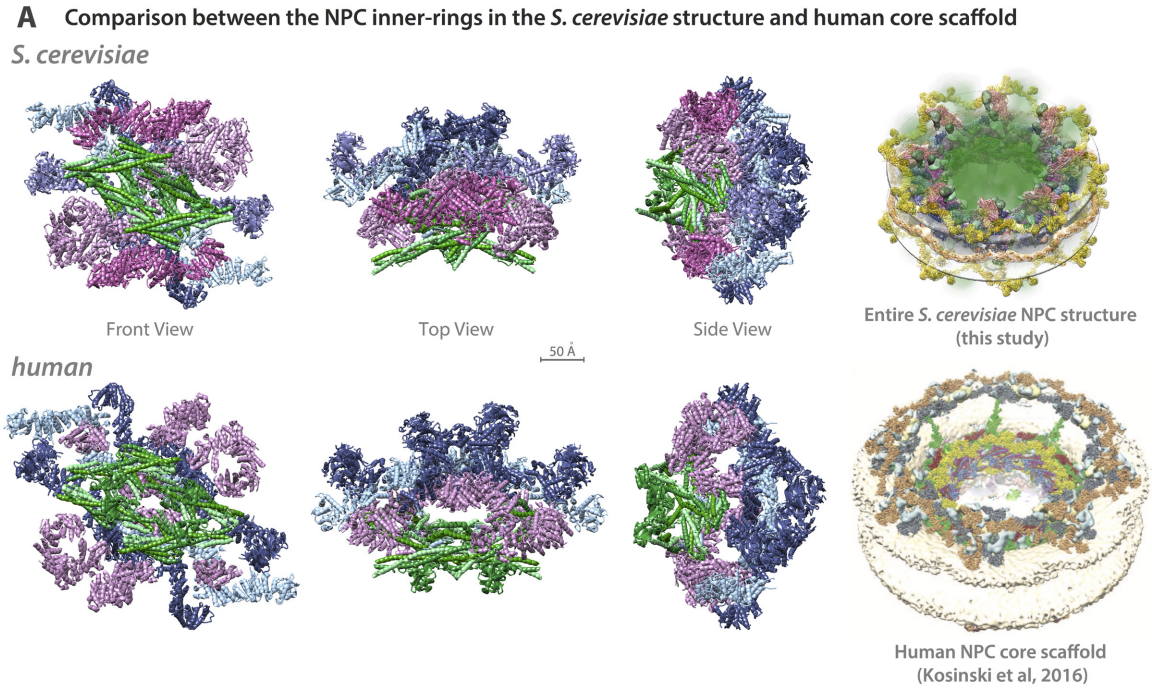
Extended Data Figure 10 | Proposed evolutionary origin of the NPC from a later amalgam of membrane coating complexes. a, Diagram depicting how the NPC may have originated from an ancestral coatomer module through a series of duplications, divergence and secondary loss events. Top, the origin of an ancestral proto-NPC coatomer module from an amalgamation of COPI-like and COPII-like complexes. Middle, the initial duplication leading to the origin of the inner and outer rings, and their associated coiled bundles. Presumed secondary losses removed the additional COPII-like subunit of the inner-ring protomer; loss of the adaptin-like subunit from the outer ring may have occurred here, or later in only certain lineages. Bottom, another duplication and divergence within each spoke may then have generated two parallel and laterally-offset paralogous columns; in the outer ring, a COPII-like subunit was then lost from one of the duplicates. The coiled bundles of the outer rings gave rise to the cytoplasmic export complex and nuclear basket by subsequent duplication; the export complex itself is a duplicate with a dimer of trimeric coiled bundles in its core. Outer-ring duplications are not shown. Relevant nucleoporin domains are depicted as follows: β-propellers (cyan circles), α-solenoids (pink bars) and coiled-coil domains (orange sticks). Left, diagrams (grey) exemplify the path of duplications within the whole NPC. Examples of ribbon representations for each module are presented. The anchoring points of the coiled-coil cytoplasmic Nup82 complex and the nuclear basket (orange densities) into an equivalent region of the outer-ring Nup84 complex (grey density) are shown. **b**, Conserved structural motifs connecting spokes in the outer and inner rings. Diagram showing how the spoke-to-spoke connection is established through similar head-to-head connections of heterodimers containing one COPI-like and one COPII-like subunits in both the outer (left) and the inner (right) NPC rings. Top, nucleoporin domains coloured as in **a**; bottom, COPI-like Nups in red, COPII-like Nups in blue.

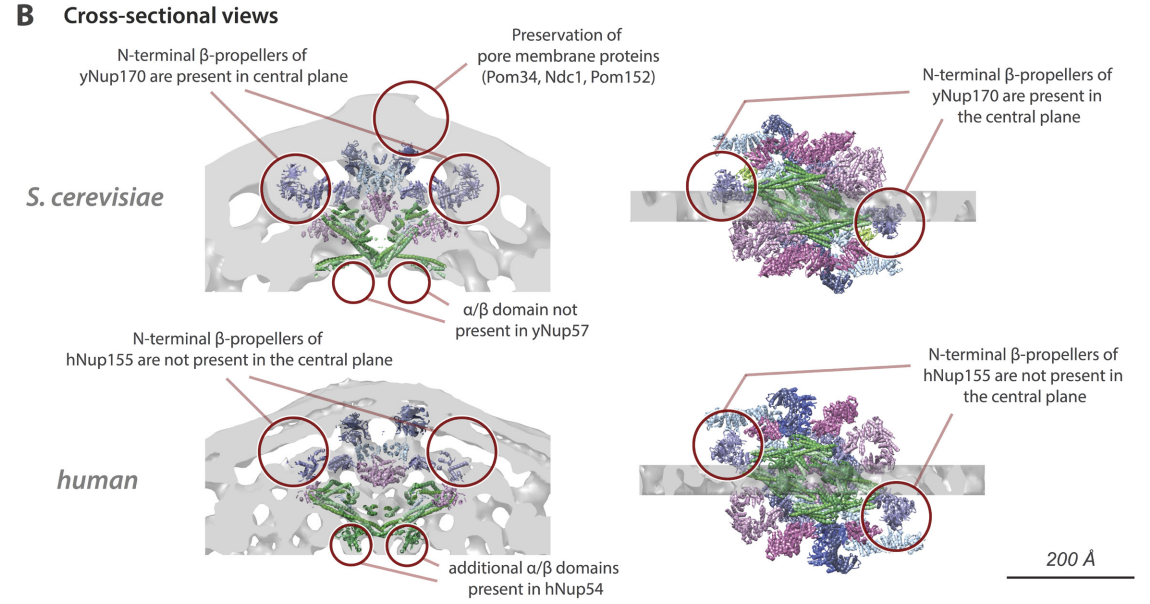
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

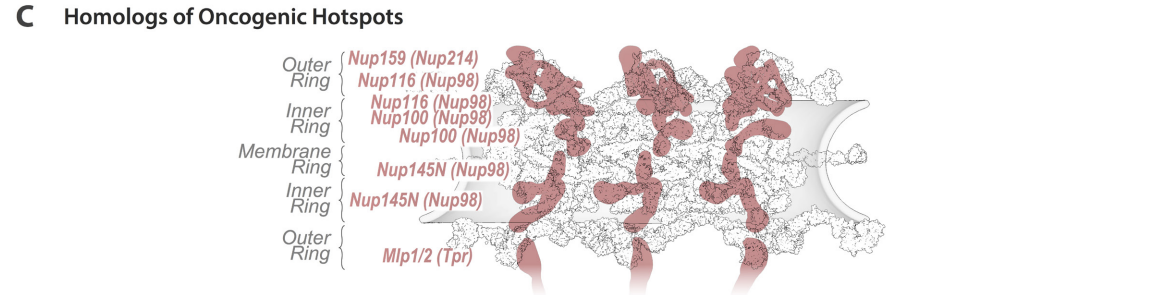


Extended Data Figure 11 | Position of the FG-repeat anchor points and heat mapping of the FG repeats. a, Three views of the complete structure of the NPC are shown with major structural features (coloured as in Fig. 4 and Supplementary Table 2) and a snapshot of modelled FG-repeat regions (indicated in green). For each Nup, the localization probability density of the ensemble of structures is shown with a representative structure from the ensemble embedded within it. See also Supplementary Videos 1–3. Scale bar, 200 Å. **b**, Positions of FG-repeat anchor points within the ensemble of solutions are depicted as green surfaces; the Nups to which they belong are labelled in the centre image. Left, side view of three spokes; centre, side view of one spoke; right, top view of three spokes. Scale bar, 100 Å. **c**, Heat mapping of the type of FG-repeat region of each FG Nup (FXFG/FG type, red; GLFG type, blue), showing partitioning of the FG types to different regions of the central transporter. Identity of mapped Nups is shown in the diagram on the right. Scale bar.100 Å. **d**, Heat mapping of the effect on NPC permeability of the truncation of an FG repeat in each FG Nup, relative to the wild-type strain (p/p_{WT}); the severity of the permeability defect is indicated in increasing intensity of shades of blue from minor defect (light green) to severe defect (dark blue), thereby defining the FG repeats that are most important in maintaining the passive permeability barrier. Identities of mapped Nups are shown in the diagram on the right. Scale bar, 100 Å.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.







Extended Data Figure 12 | Comparison of the *S. cerevisiae* NPC structure and human NPC core scaffold. a, Comparison between the inner rings in the structure of the *S. cerevisiae* NPC (first row) and the core scaffold of the human NPC (second row) (Protein Data Bank code: 5IJN[6]). Yeast Nups are coloured as in Fig. 4; human Nup homologues are coloured as their yeast counterparts. All copies of human Nup155 are coloured as yeast Nup157, and all copies of human Nup205/Nup188 are coloured as yeast Nup192. Only homologue Nups present in both yeast and human are shown. The human NPC core scaffold includes two additional copies of Nup155 that are absent in yeast (owing to the different stoichiometry between organisms). Yeast Nup53 and Nup59 are not shown because their counterparts are not present in the human NPC core scaffold. **b**, Major differences in the inner ring between the *S. cerevisiae* and human NPCs are highlighted, in the cross-sectional view near the equator. **c**, Positions of yeast Nups homologous to oncogenic human Nups (in parentheses) are shown in red, mapped onto three spokes of the NPC.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Michael P. Rout

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

The sample size for integrative structure determination was determined by counting 100,453 modeled structures and 5,529 good-scoring structures. For all other experiments, we counted and reported accordingly how many times we performed each experiment.

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses

3. Replication

Describe whether the experimental findings were reliably reproduced.

Yes, the experimental findings were reliably reproduced.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization was required for the experiments performed in this study

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No blinding was required for the experiments performed in this study

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Integrative Modeling Platform (IMP 2.6), MODELLER 9.3, HHPred, PSIPRED, DISOPRED, DomPred, COILS/PCOILS, Multicoil2, UCSF Chimera, CX-Circos, matplotLib, GNUPLOT 5.0, IMOD, EMAN2, RELION, pLink

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restriction, all materials readily available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No cell lines were used in this study
b. Describe the method of cell line authentication used.	No cell lines were used in this study
c. Report whether the cell lines were tested for mycoplasma contamination.	No cell lines were used in this study
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No cell lines were used in this study

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	This study did not involve human research participants.