The human pre-replication complex is an open complex

Jian Li, Jiangqing Dong, Weitao Wang, Daqi Yu, Xinyu Fan, Yan Chit Hui, Clare S.K. Lee, Wai Hei Lam, Nathan Alary, Yang Yang, et al.

The Human Pre-replication Complex is an Open Complex

Jian Li[1#], Jiangqing Dong[2#], Weitao Wang[3#], Daqi Yu[2#], Xinyu Fan[1], Yan Chit Hui[1], Clare S. K. Lee[1], Wai Hei Lam[1], Nathan Alary[3], Yang Yang[4], Yingyi Zhang[5], Qian Zhao[4], Chun-Long Chen[3]*, Bik-Kwoon Tye[6,7]*, Shangyu Dang[2,8,9]*, Yuanliang Zhai[1]*^

[1]School of Biological Sciences, The University of Hong Kong, Hong Kong

[2]Division of Life Science, Center of Systems Biological and Human Health, The Hong Kong University of Science and Technology, Hong Kong

[3]Institut Curie, PSL Research University, CNRS UMR3244, Dynamics of Genetic Information, Sorbonne Université, 75005 Paris, France

[4]State Key Laboratory of Chemical Biology and Drug Discovery, Department of Applied Biology and Chemical Technology, Hong Kong Polytechnic University, Hong Kong

[5]Biological Cryo-EM Center, The Hong Kong University of Science and Technology, Hong Kong

[6]Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong

[7]Department of Molecular Biology & Genetics, Cornell University, Ithaca, NY 14853

[8]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

[9]HKUST-Shenzhen Research Institute, Nanshan, Shenzhen 518057, China

[#]These authors contributed equally to this work

^Lead Contact: zhai@hku.hk (Y.Z.)

*Co-corresponding authors:

chunlong.chen@curie.fr (C.C.)

biktye@hku.hk (B.T.)

sdang@ust.hk (S.D.)

zhai@hku.hk (Y.Z.)

Summary:

In eukaryotes, DNA replication initiation requires assembly and activation of the minichromosome maintenance (MCM) 2-7 double hexamer (DH) to melt origin DNA strands. However, the mechanism for this initial melting is unknown. Here, we report a 2.59-Å cryo-electron microscopy structure of the human MCM-DH (hMCM-DH), also known as the pre-replication complex. In this structure, the hMCM-DH with a constricted central channel untwists and stretches the DNA strands such that almost a half turn of the bound duplex DNA is distorted with one-base pair completely separated, generating an initial open structure (IOS) at the hexamer junction. Disturbing the IOS inhibits DH formation and replication initiation. Mapping of hMCM-DH footprints indicates that IOSs are distributed across the genome in large clusters aligning well with initiation zones designed for stochastic origin firing. This work unravels an intrinsic mechanism that couples DH formation with initial DNA melting to license replication initiation in human cells.

Introduction:

Replication licensing in eukaryotes begins with the recruitment of two minichromosome maintenance (MCM) 2-7 complexes onto duplex DNA by the Origin Recognition Complex (ORC) in an ATP dependent manner[1,2]. The two hexameric MCM rings are loaded one at a time into a head-to-head double hexamer (DH) configuration[3-7] to form the pre-replication complex (pre-RC). Once licensed, the MCM-DH is transformed into two CDC45-MCM2-7-GINS (CMG) helicases, which are further assembled into replisomes that replicate DNA bidirectionally from the replication origin[2,8]. A defining step in the formation of a bidirectional fork is the melting and subsequent separation of the two origin DNA strands and the uncoupling of the two CMG helicases, which then translocate along single DNA strands crossing each other headfirst as cores of the replicating replisomes[9,10].

In the last few years, considerable efforts have been made to elucidate the structural basis of the initial melting of origin DNA bound by the MCM-DH. The first high-resolution structure determined for the yeast MCM-DH (yMCM-DH) showed that the interface of the two hexamers was twisted, tilted and offset to create a narrow central channel in the form of a kink that could cause deformation of the DNA captured at the hexamer junction[11,12]. Unfortunately, the captured DNA was unstable in this endogenously-prepared sample[4,13], consistent with the fact that the yMCM-DH is capable of sliding off DNA[7,14]. Subsequently, cryo-EM structures of yMCM-DH bound to DNA reconstituted from purified proteins showed that the duplex DNA was perfectly paired without any indication of DNA melting[5,6,15]. The structures of the yeast CMG helicase further demonstrated that the CMG helicase makes use of its amino-terminal tier of MCM ring as the front end to translocate along the leading strand template 3'-5' to unwind DNA[9,10]. These structural analyses support a model in which origin DNA bound by the MCM-DH is melted and then separated into two activated CMG helicases during helicase activation[9,10]. This model was later confirmed by biochemical and structural analyses using *in vitro* reconstituted yeast systems showing detectable DNA unwinding within the MCM chamber only after CMG assembly[9,16]. Despite these advances, it remains poorly understood when and where initial origin DNA opening begins and how the MCM-DH facilitates this process to promote replication initiation.

Licensing of replication origins by the human MCM-DH has been implicated from studies of other vertebrates[17-22]. Indeed, most of our assumed knowledge about the molecular mechanisms of replication initiation in human are inferred from studies in yeast and other

metazoans[1,2,23-25]. Other than the fact that origin selection and MCM loading are more elaborately regulated in human[23,26-30], the role of MCM in replication initiation is likely to be conserved in all eukaryotes. Recent structural studies of the human CMG helicase and replisome indicate that the architecture of the replicative helicase is conserved in yeast, drosophila, and human[31-34]. However, to advance our understanding of DNA replication-related human diseases, a detailed analysis of the DNA bound-hMCM-DH structure is required. Furthermore, a direct comparison of the independently evolved structures of the functionally conserved pre-RCs from human and yeast would provide valuable insights into the mechanism for initial strand separation. To achieve this goal, we purified the endogenous chromatin-bound hMCM-DH from HeLa G1 cells prior to DDK phosphorylation (Figures S1A-S1F) and determined its structure at 2.59 Å by single particle cryo-EM (Figures S1G-S1O, Table S1).

Results

Overall structure and domain organization of hMCM-DH bound to DNA

The overall dimensions, subunit and domain organization of the hMCM-DH bound to DNA are very similar to the reported structures of the yMCM-DH with the two hexameric rings connected at the N-terminal domains (NTDs) in a tilted and twisted fashion[4-6,15] (Figures 1 and S2). The central DNA channel runs from end-to-end of the DH with an opening of 44 Å at the C-terminal domain (CTD) ends and narrows to about 25 Å at the NTDs (Figure 1B). The channel then abruptly constricts at the hexamer interface by the offset of the two zinc-finger (ZF) rings surrounded by six β-turn loops from the oligonucleotide/oligosaccharide-binding (OB) folds of the MCM subunits, forming a narrow space of 13-Å diameter (Figures 1E-1L). This constrained DNA-binding channel ensures the capture of origin DNA by the hMCM-DH in a tight grip (Figures 1B, 1E and 1F). Notably, our 3D classification focused on the bound DNA did not identify any apo-form hMCM-DH (Figures S1P and S1Q). However, the endogenous yMCM-DH isolated from yeast G1 chromatin shows almost no DNA binding[4,13].

The quality of the density map of hMCM-DH also allowed the unambiguous assignment of the N-terminal extension (NTE) of hMCM4 (Figures S2), also known in yeast as the N-terminal serine/threonine domain (NSD), which contains multiple DDK (Dbf4-dependent kinase) phosphorylation sites and serves as the only essential substrate of DDK for DNA replication initiation[13,15,35,36]. In our structure, a stable region (residues 107-131) from the NTE of hMCM4 wraps around the NTD-A (A subdomain of NTD) of hMCM4 (Figures S2H and S2J), similar

to the conformation of the yeast NSD[13] (Figures S2I, S2K and S2L). This similarity suggests that the phosphorylation of the NTE motif of hMCM4 by DDK may be just as important for origin firing in human. Notably, the NTE of hMCM2 was previously identified as the major site for DDK phosphorylation[37] though not visualized in our structure. Further work will be required to understand the detailed roles of each NTE phosphorylation by DDK during helicase activation.

Inter-subunit interactions and ATPase centers

Resembling their counterparts in yeast, hMCM2-7 subunits oligomerize through similar inter-subunit interactions into two toroidal hexamers around dsDNA. These interactions can be categorized into three axial tiers: CTD, neck and NTD (Figure S3A-S3F). Subunit-specific extensions and insertions also help stabilize certain interfaces. For example, the CTEs from MCM2 and MCM6 engage with each other at the top of the CTD ring, contributing to a buried surface (BS) of 684 Å^2 (Figure S3A). Without this CTE dimer, the MCM2:MCM6 interface would be the smallest of all the subunit interfaces due to the presence of a unique side channel formed at the neck region (Figures S3A-S3F).

The MCM CTD ring contains six ATPase active centers, one at each inter-subunit interface, to coordinate ATP hydrolysis with conformational changes in its toroidal structure to melt origin DNA in the DH and unwind DNA in the CMG helicase[38]. In the hMCM-DH, ATP can be found at MCM2:MCM6, MCM3:MCM5, and MCM4:MCM7 dimers while a mixture of ATP/ADP at MCM7:MCM3 and ADP at MCM5:MCM2 and MCM6:MCM4 (Figures S3G-S3R). This arrangement is quite different from the ATP occupancies at the ATPase centers in the yMCM-DH which exhibits ATP binding only at the MCM2:MCM6 site[5,15] (Figures S3S and S3T). The difference in ATP occupancies in the ATPase centers between yeast and human likely reflects their very different mechanisms in regulating DH formation and their distinct functional states at origin DNA.

Initial open DNA structure at the inter-hexamer interface

By following the trajectory of both strands of the duplex DNA, at the inter-hexamer junction, half a turn of the duplex DNA is largely deformed (Figures 2A-2F). A plot of the DNA duplex width with respect to the anchor points of the relevant nucleotides and the hairpin loops in the central channel is most informative for this distorted conformation (Figure 2G). The major and minor grooves of the origin DNA are abruptly widened by 5-10 Å at the hexamer interface. As

a consequence, the duplex DNA is pried open with one base pair completely separated at the center of the DNA bound by the hMCM-DH (Figure 2). To form this initial open structure (IOS), the B-form duplex has to be untwisted and stretched (Figures 2H and 2I), probably by the coordinated untwisting motion in opposite directions of the two halves of the DH (Figure 2F; Supplementary Video 1).

This IOS is confined within the constricted space mainly formed by two staggered ZF rings at the hexamer junction (Figures 3A-3C). Among these ZFs, two pairs of ZFs from MCM5 and MCM2 make direct contacts with the melted bases as well as the associated DNA backbone (Figures 3D-3G). The ZF pair of MCM5 from opposite hexamers dimerize, with the help of the ZFs of MCM3, to align on the same plane of the opened DNA (Figures 3D, 3F and S4A-S4D). The melted bases are captured by two L209s of the MCM5 ZFs of opposite hexamers (1 and 2) through hydrophobic interactions with the symmetrically oriented aromatic rings of the separated bases (B0 and B0') (Figure 3F). This arrangement of the L209s effectively forms a barrier between the melted bases to confine them in the same constrained central chamber yet separated (Figures 3A-3F). Notably, this critical L209 residue is conserved only among metazoans (Figure S4E) although the proline of the yeast MCM5 zinc finger could also serve the function of an intercalator. In addition, the guanidinium group of R195 from each MCM5-ZF forms ionic interactions with the phosphate backbone and a cation-π interaction with the aromatic ring of the base (B-1 or B-1') (Figure 3F). Reinforcement comes from the ZF pair of MCM2 (Figures 3E and 3G). Three residues of MCM2-ZF, K331, A358, and G359, form a supporting scaffold by interacting with the sugar-phosphate backbone of the upstream nucleotides on opposite sides of the IOS (Figure 3G).

According to the polarity of the DNA strands, the ZFs of MCM2 and MCM5 from each hexamer pinch the melted strand with the NTD of the MCM ring facing 5' in the direction of translocation of the CMG helicase (Figures 3D-3G). Thus, it appears that the two opposite hexamers have each captured their prospective leading strands.

However, in the structure of the yMCM-DH[5,6,15], the bound duplex DNA is not obviously distorted or melted at the hexamer junction (Figures S4F and S4G) where the central channel is less constrained in comparison to that of hMCM-DH (Figures S4H-S4J). In addition, the ZFs of MCM2 and MCM5 from yeast engage with their respective lagging strands rather than the

leading strands of the unmelted DNA (Figures S4K and S4L). These discrepancies between yeast and human in capturing origin DNA may stem from their distinct mechanisms of DH formation or distinct functional states of the assembled DH.

The role of the IOS in replication initiation

The unique features of the IOS in the hMCM-DH invite speculations for its function. Considering the implied structural role of R195 and L209 of MCM5-ZF in establishing the IOS, we decided to mutate these two residues to test their effects on replication initiation. To do so, we tagged both alleles of the endogenous MCM5 gene with mini-Auxin-inducible degron (mAID) (MCM5-mAID) in HCT116 cells for an efficient depletion of MCM5-mAID fusion proteins (Figures S5A-S5C) while ectopically expressing the MCM5 double mutant (R195A and L209G, referred to as MCM5-RL) to a level equivalent to the endogenous MCM5 (Figures 4 and S5D-S5G). Our FACS analysis indicated that, after depleting the endogenous MCM5-mAID, the MCM5-RL mutant shows a significant delay in S phase entry in cells after release from a nocodazole block (Figures 4A and S5F). This delay is comparable to the negative control but can be rescued by the WT MCM5 (Figure 4A). To have a better understanding of this defect, we examined the chromatin association of MCM2-7, GINS2 and RPA to determine the status of replication licensing and origin firing in both WT and mutant cells. Compared to the WT control, the MCM5-RL mutant poorly supported MCM loading onto chromatin (MCM2/4 recruitment), suppressed CMG assembly (GINS2 recruitment), and prevented DNA unwinding (RPA recruitment) (Figure 4B). To evaluate the roles of individual R195 and L209, we further examined the effects of MCM5-R195A (referred to as MCM5-R) and MCM5-L209G (MCM5-L) single mutants. After MCM5-mAID depletion, obvious defects in replication initiation can also be observed in these single mutants but they are less severe when compared to the MCM5-RL double mutant (Figures S5F and S5G).

Given that the interactions between the MCM5-ZF pair and the IOS take place at the hexamer interface where the two MCM hexamers are coupled, it is possible that disrupting their interactions may interfere with stable coupling of the hexamers. To test this possibility, we sought to isolate the MCM5-RL mutant containing DHs. As the MCM5-RL mutant does not support the assembly of homo-RL/RL DHs (Figure 4B), we expressed the MCM5-RL mutant with a 3xFLAG tag (MCM5-RL-3xFLAG) in cells without depleting the endogenous WT MCM5 to isolate hetero-DHs. As expected, the MCM5-RL-3xFLAG complexes appeared to be exclusively hetero-DHs containing both WT and mutant MCM5 (referred to as RL/WT DH)

(Figure S6A). Using this sample, we determined the cryo-EM structure of the RL/WT DH (Figures S6B-S6F). We applied rigorous 3D classification focusing on the duplex DNA bound by the DH and identified four states of the RL/WT DH in which the IOS exhibits different flexibilities (Figures S6G-S6I). In contrast, the IOS from the WT/WT DHs remain largely stable (Figures S1P and S1Q). Furthermore, we found that the RL/WT mutant DHs is prone to disassembly under high salt (400 mM NaCl) treatment while the WT/WT DHs remain salt resistant (Figures S5H-S5J). Similar phenomenon was also observed with the MCM5-R and MCM5-L single mutants (Figures S5I and S5J). Together, these analyses suggest that the engagement between the MCM5-ZFs and the IOS plays a critical role in DH formation and/or maintenance.

Axial arrangement of interior hairpin loops

About 50 base pairs (bps) of relatively rigid duplex DNA span the entire length of the 3-tiered (CTD, neck, and NTD) ring of the inverted dimers of hMCM (Figures 1B and 5A). In addition to its interaction with the ZFs of MCM2 and MCM5 at the hexamer junction, the duplex DNA is also held in place by six layers of hairpin loops in the DNA channel formed by the helix-2-inserts (H2Is), presensor 1 β hairpin (PS1) and the β-turns of the OB domain (Figures 5A-5D). In the neck regions of the opposite rings, the six H2Is occupy axial positions with a helical trajectory staggering with the PS1s (Figures 5D-5F). Four H2Is of MCM6/4/7/3 are arranged in a right-handed staircase to contact the potential lagging strand for nearly half a turn while the PS1s of MCM6/4/7 contact the potential leading strand of the same minor groove (Figures 5E and 5F). Notably, the leading strand DNA at this region is situated in a position very close to the conserved MCM single-stranded binding motif (MSSB) which coordinates single-stranded DNA (ssDNA) binding for helicase translocation[33,39] (Figures 5I-5L). However, there is no contact with the duplex DNA from the H2Is of MCM2/5 and PS1s of MCM2/5/3 as they turn away from the minor groove of DNA and switch to a left-handed staircase (Figures 5E and 5F). As a result, the duplex DNA is only associated with one side of the central channel at these points on opposite hexamers with the lagging and leading strands tagged by the staggered H2Is and PS1s, respectively (Figures 1B, 5A and 5D). As for the β-turn ring at the NTD tier, the β-turn motifs of MCM3 and MCM7 interact with the lagging strand while the ZFs of MCM2 and MCM5 interact with the leading strand (Figures 5B and 5C). Again, a clear choice is made for specific motifs to tag the lagging or leading strands, as if poised to coordinate the separation of these two strands. The detailed interactions between these conserved motifs and

dsDNA are noticeably different in the yMCM-DH models (Figures 5G, 5H, S4K and S4L) either because endogenously assembled and in vitro reconstituted samples have subtle differences[5,6,15], or these features are species specific.

Global landscape of the MCM-DH footprints

As the initial melted origin DNA is tightly bound to the hMCM-DH, an examination of these bound sequences may provide valuable information about their genomic locations. We isolated DNA fragments from the endogenous DH sample for deep sequencing (Figures S7A and S7B). The size uniformity of the isolated DNA (~54-56 bp) (Figures S7B and S7D) allowed the mapping of MCM-DH footprints across the human genome at almost nucleotide resolution. Our analysis showed that the hMCM-DH sites are broadly distributed and organized into ~10,000 clusters (Figures 6A-6D), covering more than half of the mappable genome. The size of these clusters correlates with replication domains (RDs) that vary according to their replication timing[40] (Figures 6C and 6D). In general, the MCM-DH bound sites located at early RDs are relatively well-defined, clustering within a mean size of 30 kb (ranging from 5 to 200 kb) (Figure 6C). Over 70% of the early MCM-DH clusters align very well with replication initiation zones (IZs) mapped by Okazaki fragment sequencing (OK-seq)[41] (Figure 6E) and/or optical replication mapping (ORM)[42] (Figure 6F). In contrast, MCM-DH clusters associated with mid-early, mid-late and late RDs, are much more expanded in size, reaching up to several megabases in late RDs, and stronger in signals (Figures 6A-6D). The strikingly large size of the MCM-DH clusters is consistent with the reported high density of licensed initiation sites located within IZs[42,43]. Sequence composition analysis of the MCM-DH-bound DNA showed that they are highly A/T-rich, similar to the A/T content of the human genome (Figure S7E), and without an identifiable consensus motif. Interestingly, the highest A/T content is located at the center of the hMCM-DH sites, suggesting that human MCM hexamers preferentially bind easily unwinding sequences to trigger initial DNA melting before forming a stable DH.

To investigate the relationship between hORC and hMCM-DH sites, we examined the distribution of published hORC1 genomic data[44] relative to the hMCM-DH clusters in early RDs. We found that hORC1 is concentrated at the borders of the early MCM-DH clusters (Figure 6H). However, H2A.Z, which is implicated in hORC recruitment[45], is enriched at the borders and distributed throughout the early MCM-DH clusters (Figure 6I). These results support the notion that hORC together with H2A.Z play a pivotal role in the iterative loading of the hMCM-DHs within the entire IZs[45,46]. In relation to transcriptional landscapes as shown

by the available phosphorylated RNA Pol II ChIP-seq data[47] (Figure 6G) and global run-on sequencing (GRO-seq) data[48] (Figure S7I), our MCM-DH-bound sequences in early RDs are mapped in regions devoid of active transcription, which is also in agreement with previous genomic studies[49]. As well, these MCM-DH sites are largely depleted within gene bodies (Figures S7F and S7J). Instead, they associate predominantly with intergenic regions with a preference for divergent over tandem or convergent gene pairs (Figures S7K and S7L). These analyses support the notion that transcription shapes replication landscape through restricting MCM-DH formation to non-transcribed regions where a high density of replication origins is licensed for stochastic firing during helicase activation.

Discussion

In the hMCM-DH structure, half a turn of DNA is distorted and melted to form an IOS at the hexamer junction analogous to the open promoter complex observed in transcription[50]. Importantly, this IOS is required to maintain stable coupling between the two opposing MCM hexamers on DNA. Given that hORC shows little or no sequence specificity in DNA binding, the sites for replication licensing are determined mainly by local chromatin context and transcription status in human cells[23,51,52]. Consistent with previous studies[43,49,53], the hMCM-DH binding sites are found at an extremely high density within non-transcribed regions, forming broad zones widespread over the human genome. Furthermore, we found that hORC and H2A.Z show distinct patterns in relation to the MCM-DH clusters at early RDs. While both are highly enriched at the cluster borders, only H2A.Z is distributed throughout the clusters. It is likely that, with the help of H2A.Z, hORC cycles on and off chromatin to promote the loading of multiple hMCM-DHs within each cluster. The tight gripping of the duplex DNA with an IOS at the center suggests that the assembled pre-RC in human is immobile, unlike the observed sliding of MCM-DH in yeast[7,14]. A plausible explanation is that the hMCM-DH is mobile upon loading by hORC and subject to displacement by transcription machinery but once clamped down on an initiation site, it forms an immobile open complex. Formation of this tight open complex may involve rotational movement of the respective MCM rings in opposite directions. Engagement of the ZFs of MCM5 and MCM2 with the IOS may reconfigure the opposing MCM hexamers at their NTDs to form a tightly coupled hMCM-DH. Consistent with this hypothesis, our mutational analyses showed that disturbing the interactions of the IOS with any ZFs from either MCM5 or MCM2 weakened the coupling of the MCM hexamers as shown by the salt sensitivity of the mutant MCM-DHs (Figures S5I and S5J).

Importantly, this IOS was not observed in the yeast MCM-DHs reconstituted *in vitro*[5,15]. These drastically different initial structures of chromatin bound MCM-DHs observed in these studies speak to either an inherent species-specific difference between human and yeast or a manufactured difference between endogenously assembled and *in vitro* reconstituted samples.

The structural features of the hMCM-DH also provide new insights for the steps following activation of the pre-RC. To begin with, at the interface of the hMCM-DH, the ZFs of MCM2 and MCM5 from the opposing hexamers capture the melted bases and the relevant sugar-phosphate backbone of their respective leading strands (Figure 3) while the β-turns of MCM3 and MCM7 engage with the lagging strands (Figures 5B and 5C). In coordination, there seems to be a sense of strand identity in the interactions along the duplex DNA by the H2Is, PS1s and β-turns. This strand-specific arrangement by the array of MCM pore loops sets the stage for strand separation within the central channel such that further twisting of the two hexamers in opposite directions would melt additional base pairs to expand the IOS (Figures 7A-7C).

Based on the available structure of the hCMG helicase[33], it is possible to build a credible model for steps that convert the hMCM-DH into two hCMGs. The conserved MSSB elements in the hCMG helicase are arranged in a spiral staircase that contacts the backbone of the leading ssDNA for DNA unwinding[54]. However, in the hMCM-DH structure, these MSSB elements are contacting the phosphate backbone of the duplex DNA instead of ssDNA (Figures 5E and 5F). We envision that additional melting of DNA within the central channel will require remodeling of the MSSB elements and the MCM-pore loops to stably anchor the melted leading strand and keep it from reannealing (Figures 5K and 5L; Supplementary Video 2). This stabilized DNA bubble could then initiate a DNA unwinding mechanism that melts the length of the entrapped duplex DNA for lagging strand exclusion (Figures 7C and 7D). This process may involve the translocation of the still coupled hexamers, each on their designated leading strand so as to pull the duplex DNA towards its own chamber. In so doing, the duplex DNA within the central channel becomes unregistered and unwound from end to end, allowing the unwound DNA within each of the MCM chambers to merge with the IOS. The ssDNA so generated in searching for an exit may assist in the uncoupling and gate-opening of the hexamers by extruding outside of the respective overcrowding central chamber (Figures 7D and 7E). We do not know the exit gate for lagging strand exclusion. In yeast, the MCM2/5 interface is the entry gate for duplex DNA into the MCM ring during pre-RC assembly[55,56].

However, this interface is sealed by Cdc45 and GINS at the NTD tier upon CMG formation in both yeast and human, suggesting that exit of the lagging strand must go through another gate[57]. Finally, the two uncoupled CMG helicases translocate on their respective leading strands and go past each other to create the bidirectional forks that will travel long distances to replicate genomic DNA.

After the submission of this manuscript, Lewis et al using a yeast reconstituted system reported the structure of the double Cdc45-MCM-GINS-Polε (dCMGE) formed on chromatinized origin DNA[16]. Inside each MCM ring of the yeast dCMGE, the duplex DNA becomes untwisted, and base-pairing is broken, while the intervening DNA between the two CMGEs remain duplexed. These somewhat contradictory structures suggest a drastically different mechanism for the initial DNA melting step between yeast and human. Clearly, more work is needed to establish whether these different structures are species-specific, or they represent differences in the manner the samples were sourced or the different stages of the initiation process. For example, during the melting of DNA within the individual CMG chambers, the two MCM hexamers may rotate in right-handed or left-handed twists which would either overwind or underwind the intervening DNA at the interface. This may result in either reannealing of the IOS as observed in yeast[16], or expanding of the IOS into a larger bubble that would fuse with the two emerging bubbles forming inside the CMG chambers (Figures 7B and 7C). Further investigation is required to capture these intermediate states in both yeast and human for structural characterization.

Hydrolysis of ATP is the driving force of the MCM helicase. Nucleotide occupancy at the ATPase pockets provides clues of the unwinding mechanism. Interestingly, the nucleotide occupancy in the ATPase centers of the hMCM-DH is obviously different from its yeast counterpart (Figures S3S and S3T). Surprisingly, the hMCM-DH and the yeast dCMGE share a common feature in the nucleotide binding pattern around the MCM ring (Figures S3S and S3U). As DNA melting is observed in both structures, we speculate that concerted ATP hydrolysis at the six ATPase sites plays a key role in promoting DNA melting either locally at the hexamer junction, via a yet unknown mechanism, or progressively in the CTD ring of MCM complexes, via a coordinated rotary mechanism[32].

Our mapping of the hMCM-DH footprints validates the physiological relevance of the hMCM-DH sample purified from human G1 chromatin and generates a genomic map more refined than the Chromatin Endogenous Cleavage (ChEC) approach[58] (Figures S7F-S7H). We found a high density of the MCM-DHs enriched within broad clusters, and most early MCM clusters overlap well with the previously identified early IZs[41,42]. Our data support the notion that neither the size nor the signal intensity of the MCM-DH clusters correlates with origin firing efficiency arguing against the hypothesis that efficiency is determined by the number of the MCM-DHs loaded at a particular site or IZ[23,59]. This notion is further supported by the fact that the MCM-DHs are found at much higher densities and distributed more broadly at late RDs. In agreement with previous studies[49,60], hMCM-DH occupancy and active transcription are mutually exclusive. Within the gene-dense early RDs, hMCM-DHs are almost exclusively located within intergenic regions to coordinate co-directional transcription and replication. This strategy effectively avoids transcription-replication conflicts. The widespread distribution of MCM-DHs also ensures a timely completion of replication of long inter-origin segments or damaged DNA segments through initiation from redundant licensed origins.

In summary, this study presents the first high-resolution structure and footprints of the human MCM-DH. It provides important information for the initial DNA melting mechanism and valuable insights into the mechanisms of pre-RC formation, helicase activation and bidirectional replisomes assembly. Of equal importance, it provides a structural framework for the study of human diseases that have mutations mapped in the *MCM* genes and are often manifested as developmental defects such as deafness, immunodeficiency and dwarfism[61-63] and cancers[64].

Limitations of the study

Our structural and functional study indicates that pre-RC formation is coupled with the initial melting of origin DNA in human cells. However, it is unclear whether the IOS truly represents the initiation of origin DNA unwinding, or merely an anchoring structure to be reannealed upon uncoupling of the hMCM-DH. To address these questions, dynamic structures representing different frames of the process in motion are required. We anticipate that reconstituting the intermediates *in vitro* to recapitulate the many stages of the assembly and activation of the human MCM-DH will provide answers in future studies.

Acknowledgement

We thank the Biological Cryo-EM Center at the Hong Kong University of Science and Technology for the data collection of hMCM-DH samples. The center is generously supported by a donation from the Lo Kwee Seong (LKS) Foundation. This work was supported by the Research Grants Council (RGC) of Hong Kong (GRF16103918, GRF17112119, GRF17101720, GRF17119022, C7028-19GF, and C7009-20GF to Y.L.Z.; ECS26101919, GRF16103321, C7009-20GF, and C6001-21EF to S.D.), S.D. acknowledges support from Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC01-L), Guangdong Basic and Applied Basic Research Foundation (2021A1515012460) and Shenzhen Special Fund for Local Science and Technology Development Guided by Central Government (2021Szvup140). J.D. and D.Y. are supported by LKS fellowships. Work of CLC lab is supported by the YPI program of I. Curie, the ATIP-Avenir program from Centre national de la recherche scientifique (CNRS) and Plan Cancer [grant number ATIP/AVENIR: N°18CT014-00]; the Agence Nationale pour la Recherche (ANR) [grant number ReDeFINe-19-CE12-0016-02, TELOCHROM-19-CE12-0020-02]; and Institut National du Cancer (INCa) [grant number PLBIO19-076]. W.W. was supported by a COFUND IC-3i International PhD fellowship.

Author contribution

Y.L.Z. conceived the study; Y.L.Z., S.D., B.K.T. and C.L.C. supervised the project; J.L., X.Y.F, Y.C.H., Y.Y., and Q.Z. constructed cell lines, purified hMCM-DH and performed functional assays; J.D., S.D., J.L. Y.D. and Y.Y.Z prepared cryo-grids; S.D. J.D. and D.Y. collected data and processed images; J.D., S.D., D.Y. and Y.L.Z. built atomic model; J.L. and C.L. prepared DNA library for deep sequencing. C.L.C., W.W., N.A., and C.L. performed bioinformatic analysis, and Y.L.Z., B.K.T., S.D., C.L.C., J.D., J.L., D.Y., W.W., and W.H.L. prepared the figures and the manuscript.

Competing financial interest

The authors declare no competing financial interests.

Figure 1. Overall architecture of the human MCM-DH.

Side (**A-D**) and top (**E**) views of the segmented cryo-EM density map superimposed with the atomic model. **B** is same as **A** but with four MCM subunits removed to display the bound

DNA. (**F**) Cut-away top view of the density map. (**G**) Top view of the β-turn loops and ZFs. (**H-L**) Top (**H**) and side (**I-L**) views of ZF rings.

Figure 2. Conformation of the hMCM-DH-origin DNA open complex.

(**A** and **E**) Side views of the cryo-EM map with the indicated subunits visible only. (**B-D**) Zoomed-in views of the boxed regions from **A** and **E**. (**F**) A hypothetical model illustrating an ideal B-form DNA being transformed into the IOS configuration. (**G**) Plots of groove widths of the hMCM-DH-bound DNA. Groove widths of the ideal B-form DNA are shown as constant values. (**H** and **I**) The third helical turn of the melted DNA (right) is displayed side by side with an ideal B form DNA (left). The boxed regions in **H** highlight the disrupted helical structure of the melted DNA.

Figure 3. Detailed configuration of the IOS formed between the MCM2/5 ZFs and the melted DNA

(**A**) Top view of the ZF rings. (**B** and **C**) Deconstructing the ZF rings from **A** to illustrate the capturing of each prospective leading strand DNA by the corresponding ZFs. (**D** and **E**) Cartoon views of MCM2/5 ZFs engaging with the melted base pair. (**F**) Detailed interactions of R195 and L209 residues from MCM5-ZFs with the nucleotides at the melted base pair. Interatomic distances are shown in angstroms. The cation-$\boldsymbol{\pi}$ interactions between R195s and the aromatic rings of the relevant bases (B-1 and B-1') are indicated by dashed lines in red. "B" denotes base. (**G**) A358 and G359 of MCM2-ZF making contacts with the backbone of the nucleotides flanking the melted base pair.

Figure 4. The physiological function of IOS in replication initiation.

(**A** and **B**) Samples of the MCM5-RL mutant expressed in an MCM5-mAID cell line were collected at the indicated time points for FACS analysis (**A**) and chromatin binding assay (**B**). The cell lines expressing MCM5-WT or bearing an empty vector (NC) were used as positive and negative controls respectively.

Figure 5. Relationship between interior hairpin loops of hMCM single hexamer and DNA strand polarity.

(**A**) Side view of PS1, H2I, β-turn and ZF motifs from hMCM hexamer 1. (**B**) β-turn loops in surface representation around DNA. (**C**) Side views of the IOS bound by β-turn hairpin loops of MCM3/7 and the ZFs of MCM2/5. (**D**) Same as **B** but shown with PS1 and H2I loops. (**E**

and **F**) The hMCM-DH bound DNA surrounded by the PS1 and H2I loops which are arranged in right-handed (**E**) and left-handed (**F**) staircases. (**G** and **H**) Side views of the PS1 (green) and H2I (red) loops from the yMCM-DH encircling duplex DNA. (**I** and **J**) ssDNA bound to the MSSB of the human CMG structure (PDB: 6xtx). (**K**) The PS1 and H2I loops from hexamer 1 of hMCM-DH encircling its prospective leading strand DNA (magenta) superimposed with ssDNA (yellow) from hCMG (PDB:6xtx). (**L**) The PS1 and H2I loops from hCMG in contact with ssDNA superimposed with the leading strand of hMCM-DH. The MCM hexamers from hMCM-DH and hCMG were globally aligned to illustrate the relative positions of the color-coded DNA strands in **K** and **L**. Locations of the MSSBs are highlighted by the dashed line in red.

Figure 6. Global landscape of the hMCM-DH sites across the human genome.

(**A** and **B**) Genome browser screenshots illustrate the enrichment patterns of hMCM-DH sites and previously published profiles of replication timing (S50, early-replicating regions in red and late-replicating regions in blue), initiation sites identified by replication fork direction (RFD) based on OK-seq (windows with positive RFD values in red and negative values in blue), optical replication mapping (ORM), and gene expression as indicated by phosphorylated RNA Pol2-S2 (RNPII-Ⓟ). The bottom panel shows the size distribution of DNA fragments isolated from the hMCM-DHs along the genome. MCM control refers to the input DNA fragments. (**C**) Boxplot showing MCM-DH cluster size in early (E), mid-early (ME), mid-late (ML) and late (L) replicating domains (RDs). (**D)** Normalized MCM densities located in different RDs. (**E-G**) Enrichment of RFD (**E**), IZs identified by ORM (**F**) and RNPII-Ⓟ (**G**) relative to the hMCM-DH clusters at the early RDs. The upper panels show the genome-normalized relative signal around the hMCM-DHs. The lower panels show heatmaps of the same signals at each hMCM-DH cluster. (**H** and **I**) Enrichment of ORC1 (segment number of ORC1 within 1 kb bin, **H**) and H2A.Z (**I**) relative to the center and two borders (Start and End, respectively) of hMCM-DH clusters.

Figure 7. Model of DNA melting and strand separation during helicase activation.

(**A**) Origin DNA in complex with hMCM-DH forms an IOS at the hexamer interface; (**B**) upon helicase activation, additional DNA melting is induced in each CTD ring of MCM complex; (**C**) the melting bubbles further expand and merge; (**D**) followed by engagement of MSSB with the leading strand DNA as the MCM hexamers reconfigure to the translocation mode for further DNA unwinding; (**E**) complete hexamer uncoupling and lagging strand DNA exclusion

from MCM rings; (**F**) the active hCMG helicases translocating along leading strand with their NTDs headfirst in 3'-5' direction. The atomic model of hMCM-DH is shown in **A** (left panel) and hCMG (PDB: 6XTX) in **F** (right panels).

Figure S1. Sample preparation and structural determination of the hMCM-DH, related to Figures 1, 2 and 3 and STAR Methods.

(**A**) Western blotting of whole cell lysates from MCM7 and MCM7-3xFLAG tagged cell lines. (**B**) FACS analyses of MCM7-3xFLAG cells at 8 hours after release from nocodazole (Noc) synchrony. (**C**) The indicated samples were analysed by SDS-PAGE and immunoblotting. MCM2-S40-℗ represents the phosphorylation status of MCM2-S40 by DDK. The fraction at 8 hours was used for isolating hMCM-DHs. (**D**) Schematic illustration of solubilization of chromatin-bound MCM complexes from hMCM7-3xFLAG cells for hMCM-DH purification (see details in methods). (**E**) The eluted hMCM complexes were subjected to glycerol gradient sedimentation. Collected fractions were analysed by SDS-PAGE and visualized by silver staining. The fractions 8-9 were pooled and processed for EM analysis. (**F** and **G**) Representative negative-stain (**F**) and cryo-EM (**G**) image micrographs of the hMCM-DH are shown with 200 and 20 nm scale bars, respectively. (**H**) 2D class averages of the hMCM-DH particles. (**I**) Workflow of image processing. (**J**) The map is color-coded to indicate the range of the local resolutions. (**K**) Euler angle orientation distribution. (**L**) FSC curves of two independently refined maps before (red) and after (blue) post-processing. (**M**) Directional FSC from different Fourier cones. (**N**) Calculated resolution from different views. (**O**) Local density of representative regions of the final cryo-EM density map shown in blue meshes superimposed with the atomic model. (**P**) Workflow of image processing of the WT hMCM-DH particles focusing on DNA. (**Q**) EM densities of the IOS superimposed with its atomic model.

Figure S2. Domain organization and subunit-specific features of the hMCM-DH, related to Figure 1.

(A) Schematic illustration of domain organization and subunit-specific features of hMcm2-7 subunits and the archaeal MCM (Ss, *Sulfolobus solfataricus*). "-" symbols denote corresponding regions with highly disordered densities. (**B-G**) Side-by-side comparison of hMcm2-7 subunits. (**H** and **I**) Cryo-EM density maps of hMCM-DH (**H**) and yMCM-DH (**I,** EMD-3834) highlighting the conformations of MCM4-NTE. (**J** and **K**) Atomic models of the MCM4-NTDs from human (**J**) and yeast (**K**). (**L**) The MCM4-NTDs are superimposed.

Figure S3. Inter-subunit interactions and nucleotide occupancy at each ATPase activity center in the hMCM2-7 hexamer, related to Figure 1.

(**A-F**) Three-tiered inter-subunit interactions exemplified by the indicated interfaces. (**G-L**) ATPase centers at all six inter-subunit interfaces superimposed with the segmented nucleotide densities (gray mesh). Mg^{2+} is colored magenta. The map was contoured at 5σ. (**M-R**) same as **G-L** but from the opposite hexamer. (**S-U**) MCM nucleotide occupancy in the human MCM-DH (this study) (**S**), the yeast MCM-DH (PDB: 7P30) (**T**) and the yeast dCMGE (PDB: 7Z13) (**U**).

Figure S4. Detailed interactions of the ZFs of MCM3 and MCM5 around the IOS and the comparisons between the structures of origin DNA in complex with the hMCM-DH and yMCM-DH, related to Figure 3.

(**A**) Side view of the ZFs from MCM3/5 arranged around the IOS. (**B-D**) Zoomed-in views of the boxed regions in **A**. (**B**) Detailed interactions between β-turn loop of MCM3 and ZF of MCM5 from the same MCM hexamer as indicated by dashed lines. Interatomic distances are shown in angstroms. (**C**) Detailed interactions of hydrophobic residues between one of MCM3-ZFs and both MCM5-ZFs from the opposite hexamers. (**D**) Interactions between the two MCM5-ZFs. (**E**) Sequence alignments of the regions of MCM5-ZF involved in establishing the IOS in various species. (**F** and **G**) Low pass filtered EM density maps of origin DNA bound by hMCM-DH (**F**) and yMCM-DH (**G**) with magnified views of the boxed regions. (**H** and **I**) The ZF rings from opposing hexamers of hMCM-DH (**H**) and yMCM-DH (**I**). (**J**) Superimposition of the ZF rings from human (blue) and yeast (yellow) using MCM6-ZF as a reference for alignment. (**K** and **L**) The locations of the MCM2/5-ZF pair relative to DNA in hMCM-DH (**K**) and yMCM-DH (**L**).

Figure S5. **The effect of disturbing the interactions of IOS with MCM2/5-ZFs, related to Figures 3 and 4.**

(**A**) Western blotting of lysates from MCM5-WT and MCM5-mAID-3xHA tagged HCT116 cells. (**B**) Diagram illustrating a mAID system for conditional depletion of endogenous MCM5 proteins. (**C**) MCM5-mAID cell samples were collected as indicated and analyzed by SDS-PAGE and immunoblotting of the indicated proteins. (**D**) Diagram illustrating exogenous expression of MCM proteins in MCM5-mAID cell line. The indicated vectors for EXO-MCM expression were transfected using lentivirus. (**E-G**) Schematic illustration (**E**) of

nocodazole block-and-release assay for FACS analysis (**F**) and chromatin binding assay (**G**) to examine the impact of the MCM5 mutants as indicated. (**H**) The isolated MCM-DHs were treated as described in scheme. (**I**) Representative raw images and 2D class averages of the relevant DHs as indicated. (**J**) Percentages of the indicated DHs among all MCM particles. WT: WT/WT DHs; RL: MCM5-RL(R195A/L209G) mutant containing DHs; R: MCM5-R(R195A) mutant containing DHs; L: MCM5-L(L209G) mutant containing DHs; GG: MCM2-GG (A358G/P360G) mutant containing DHs.

Figure S6. Sample preparation and image processing of the hMCM-DH(RL/WT) complex, related to Figures 3 and 4 and STAR Methods.

(**A**) SDS-PAGE analysis of the glycerol gradient fractions for the hMCM-DH(RL/WT) sample. (**B**) A representative raw image micrograph of the hMCM-DH(RL/WT) sample with 20 nm scale bar and the related 2D class averages. (**C**) Workflow of image processing. (**D**) FSC curves of the final density map. (**E**) The local resolution map. (**F**) Euler angle orientation distribution. (**G**) Workflow of image processing of MCM-DH(RL/WT) particles focusing on DNA. (**H**) FSC curves of the final density map of the RL/WT-DH (State IV). (**I**) Electron densities of the origin DNA and MCM5-ZFs from States I-IV of hMCM-DH(RL/WT).

Figure S7. High-resolution mapping of hMCM-DH footprints, related to Figure 6 and STAR Methods.

(**A**) A flowchart of the procedure for extraction of origin DNA from the hMCM-DHs. (**B**) DNA fragments isolated from the hMCM-DHs were analysed by urea-PAGE. **(C)** Scatterplots showing the pairwise correlations for two biological replicates, MCM-R1 and MCM-R2 and the control sample (input DNA) in 1 kb resolution (with 10 kb smoothing and log2 transformation). **(D)** Histogram illustrating the size distribution of paired-end reads for the isolated DNA fragments. (**E**) Base composition around the centers of the MCM-DH-bound DNA fragments (blue) and the control input DNA (green). See details in methods. (**F**) Genome browser screenshots illustrating the different enrichment patterns between the hMCM footprints identified in this study (hMCM-DH, red) and those from previously published MCM2 ChEC seq data[58]. (**G** and **H**) Enrichment of MCM-DH footprints (**G**) and MCM2 ChEC-seq sites (**H**) relative to the center of early initiation zones (IZs) mapped by OKseq (RFD). (**I**) The enrichment of GRO-seq signals relative to the hMCM-DH clusters within the early RDs. The upper panel shows the relevant mean profiles around the hMCM-DHs, and the

lower panel is heatmap of the same signals at each hMCM-DH cluster. (**J**) Normalized MCM-DH density relative to the early active genes. TSS: Transcription Start Site. TTS: Transcription Termination Site. (**K**) Normalized MCM-DH densities relative to the intergenic regions between convergent (left), tandem (middle) and divergent (right) genes. S/E: left/right extremity of intergenic regions. (**L**) Percentages of intergenic regions overlapped with MCM-DH clusters for the convergent (blue), tandem (purple) and divergent (red) genes. Note that only intergenic regions with a size between 15 to 50 kb (matching the averaged IZ size: 32 kb ± 17 kb)[41] were included in this analysis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Human cell line construction
  - hMCM-DH purification
  - FACS
  - Double hexamer stability assay
  - EM data acquisition
  - Image processing
  - Model building
  - Mapping of the hMCM-DH footprints
  - MCM-DH cluster and classification
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://***.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
α-FLAG	Sigma	Cat# F1804
α-MCM2	Bethyl	Cat# A300-122A
α-MCM4	Bethyl	Cat# A300-193A
α-MCM5	[65]	Cat# A300-195A
α-MCM6	Bethyl	Cat# A300-194A
α-MCM7	Santa cruz	Cat# sc-9966

α-MCM2-S40	Abcam	Cat# ab133243
α-GINS2	Abcam	Cat# ab197123
α-RPA70	Invitrogen	Cat# PA5-17377
α-DBF4	Abcam	Cat# ab124707
α-CDC7	Santa cruz	Cat# sc-56275
α-Cyclin A2	Abcam	Cat# ab181591
α-Cyclin E1	Cell Signaling Technology	Cat# 20808
α-Vinculin	Santa cruz	Cat# sc-73614
α-β-Actin	Cell Signaling Technology	Cat# 4970
α-Histone H2B	Abcam	Cat# ab1790
Chemicals, peptides, and medium		
Nocodazole	Sigma	Cat# M1404
L-Glutamic acid potassium salt	Sigma	Cat# G1501
HEPES	Gibco	Cat# 11344-041
Magnesium acetate	Sigma	Cat# M5661
EDTA	Invitrogen	Cat# 15575-020
Sodium fluoride	Sigma	Cat# 201154
Sodium orthovanadate	Sigma	Cat# S6508
Phenylmethylsulfonyl Fluoride (PMSF)	Thermo	Cat# 36978
Triton X-100	Sigma	Cat# X100
Tergitol (NP-40)	Sigma	Cat# NP40S
Adenosine 5′-triphosphate disodium salt hydrate (ATP)	Thermo Scientific	Cat# J10585-22
Protease inhibitor, cOmplete, EDTA-free	Roche	Cat# 5056489001
Magnesium chloride	Sigma	Cat# M8266
Benzonase	7sea biotech,	Cat# RPE002
Anti-FLAG M2 affinity gel	Sigma	Cat# A2220
3x FLAG peptide	Genescript	N/A
DMEM, high glucose, GlutaMAX™ Supplement, pyruvate	Gibco	Cat# 10569044
Fetal Bovine Serum	Gibco	Cat# 10569044
Lipofectamine 2000	Invitrogen	Cat# 11668019
Geneticin	Gibco	Cat# 10131035
Hygromycin B	Gibco	Cat# 10687010
Propidium iodide	Sigma	Cat# P4170
RNase A	Sigma	Cat# R5503
Proteinase K	Thermo	Cat# 25530049
Qubit dsDNA HS Assay Kit	Invitrogen	Cat# Q32854
KAPA HyperPrep Kit	Roche	Cat# 07962347001
KAPA Library Quantification Kit	Roche	Cat# 07960298001
Palbociclib	Selleckchem	Cat# S1116
Thymidine	Sigma	Cat# T1895
Doxycycline hydrochloride	Sigma	Cat# 3447
5-Adamantyl-IAA	Tokyo Chemical Industry	Cat# A3390
Deposited data		

hMCM-DH(WT/WT) map	This study	EMD-32258
hMCM-DH(RL/WT) map	This study	EMD-33320
Atomic model for hMCM-DH(WT/WT)	This study	PDB-7W1Y
Sequencing data for mapping hMCM-DH footprints	This study	GSE202066
Pol II-pS2	[47]	https://doi.org/10.1038/s41467-020-14743-w
H2A.Z	[66]	https://www.encodeproject.org/files/ENCFF532VFI
Repli-Seq data	[67]	http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeUwRepliSeq
OK-seq	[41]	https://github.com/CL-CHEN-Lab/OK-Seq/tree/master/published_results/HeLa
ORM	[42]	https://doi.org/10.1016/j.molcel.2021.05.024
Orc1 ChIP-seq	[44]	http://genome.cshlp.org/cgi/doi/10.1101/ gr.142331.112
GRO-seq	[48]	https://www.ncbi.nlm.nih.gov/pubmed/25387874
RNA-seq	[68]	https://doi.org/10.1038/s41467-020-17858-2
MCM ChEC-seq	[58]	GSE150800
Experimental models: Organisms/strains		
HeLa S3	Lab stock	RRID:CVCL_0058
HCT116	Lab stock	N/A
HEK 293T	Lab stock	N/A
HeLa S3 MCM7-Stag-3×FLAG	This study	N/A
HeLa S3 MCM5-WT-3xFLAG	This study	N/A
HeLa S3 MCM5-R195A-L209G-3xFLAG	This study	N/A
HeLa S3 MCM5-R195A-3xFLAG	This study	N/A
HeLa S3 MCM5-L209G-3xFLAG	This study	N/A
HeLa S3 MCM2-WT-3xFLAG	This study	N/A
HeLa S3 MCM2-A358G-P360G-3xFLAG	This study	N/A
HCT116 MCM5-mAID	This study	N/A
HCT116 MCM5-mAID pLEX-EV	This study	N/A
HCT116 MCM5-mAID pLEX-MCM5-WT	This study	N/A

HCT116 MCM5-mAID pLEX-MCM5-R195A-L209G	This study	N/A
HCT116 MCM5-mAID pLEX-MCM5-R195A	This study	N/A
HCT116 MCM5-mAID pLEX-MCM5-L209G	This study	N/A
Recombinant DNA		
pX330 vector	Addgene	Cat# 42230
pMK283	Addgene	Cat# 72799
pMK284	Addgene	Cat# 72800
pMK286	Addgene	Cat# 72824
pMK287	Addgene	Cat# 72825
pMD2.G	Addgene	Cat# 12259
psPAX2	Addgene	Cat# 12260
pUC57 vector	General Biol (Anhui) Co., Ltd	N/A
pX330-sgMCM7	This study	N/A
pUC57-MCM7-3×FLAG-NeoR	This study	N/A
pUC57-MCM7-3×FLAG-HygroR	This study	N/A
pX330-sgMCM5	This study	N/A
pUC57-MCM5-mAID-NeoR	This study	N/A
pUC57-MCM5-mAID-HygroR	This study	N/A
pLVX-Tet-OsTIR1-F74A-BsdR	This study	N/A
pLEX vector	This study	N/A
pLEX-MCM5-WT	This study	N/A
pLEX-MCM5-R195A-L209G	This study	N/A
pLEX-MCM5-R195A	This study	N/A
pLEX-MCM5-L209G	This study	N/A
pLEX-MCM5-WT-3xFLAG	This study	N/A
pLEX-MCM5-R195A-L209G-3xFLAG	This study	N/A
pLEX-MCM5-R195A-3xFLAG	This study	N/A
pLEX-MCM5-L209G-3xFLAG	This study	N/A
pLEX-MCM2-WT-3xFLAG	This study	N/A
pLEX-MCM2-A358G-P360G-3xFLAG	This study	N/A
Oligonucleotides		
MCM7-sgRNA: AGACAAAAGTGATCCGTGTC	This study	Genewis
MCM5-sgRNA: CACCGAGGCGGCGCGACTCACTTG	This study	Genewis
Software and algorithms		
EPU 1.09	FEI/Thermo Fisher Scientific	N/A
MotionCor2 v1.4.0	[69]	http://msg.ucsf.edu/em/software/motioncor2.html
CTFFIND4	[70]	http://grigoriefflab.janelia.org/ctffind4
RELION3.1	[71]	https://www2.mrc-lmb.cam.ac.uk/relion/index.php?title=Main_Page

GCTF	[72]	https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/
Gautomatch	[72]	https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/
CryoSPARC	[73]	https://cryosparc.com
UCSF Chimera 1.15	[74]	https://www.cgl.ucsf.edu/chimera/
Pymol v2.0.6	Schrödinger	http://pymol.org
Coot 0.9.5	[75]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
UCSF pyem	[76]	https://github.com/asarnow/pyem
Phenix.real_space_refine	[77]	https://www.phenix-online.org/documentation/reference/real_space_refine.html
MolProbity 1.19.2	[78]	https://www.phenix-online.org/documentation/reference/molprobity_tool.html
Bowtie 2 v1.2.0	[79]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
BEDTools v2.30.0	[80]	https://bedtools.readthedocs.io/en/latest/
MarkDuplicates v2.23.4	Picard	http://broadinstitute.github.io/picard/
deepTools v3.5.1	[81]	https://deeptools.readthedocs.io/en/develop/index.html
FlowJo 10.4.0	Becton Dickinson	https://www.flowjo.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yuanliang Zhai (zhai@hku.hk).

Materials availability

Plasmids generated in this study will be distributed upon request.

Data and code availability

- Atomic coordinate and cryo-EM density map of the hMCM-DH(WT/WT) (PDB: 7W1Y, EMDB: EMD-32258) and the hMCM-DH(RL/WT) (EMDB: EMD-33320) have been deposited in the PDB and EMDB databases and are publicly available as of the data of publication. Accession codes are listed in the key resources table.
- hMCM-DH footprints-seq data have been deposited at GEO under the accession number GSE202066 and is publicly available as of the date of publication. The processed data are also available on the GitHub page: https://github.com/CL-CHEN-Lab/.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture and cell lines

Human cell lines HeLa S3, HCT116 and HEK293T were cultured in DMEM medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) and maintained at 37℃ with 5% CO_2.

MCM7-3xFLAG (HeLa S3) and MCM5-mAID (HCT116) cells were generated in this study by CRISPR/Cas system. The C-terminal tagging was confirmed by immunoblotting and DNA Sanger sequencing.

MCM5 and MCM2 expression cell lines were constructed by lentivirus infection and selected by antibiotics accordingly.

METHOD DETAILS

Human cell line construction

HeLa S3 MCM7-3xFLAG: The sgRNA (sequence: AGACAAAAGTGATCCGTGTC) targeting C-terminal *MCM7* gene was ligated into pX330 vector (Addgene, 42230). The DNA

cassettes of Stag-3×FLAG-NeomycinR and Stag-3×FLAG-HygromycinR from pMK283 (Addgene, 72799) and pMK284 (Addgene, 72799) were inserted into pUC57 flanked by the homology arms of C-terminal *MCM7* genomic sequences respectively to generate pUC57-MCM7-3×FLAG-NeoR and pUC57-MCM7-3×FLAG-HygroR. The three plasmids, pX330-sgMCM7, pUC57-MCM7-3×FLAG-NeoR and pUC57-MCM7-3×FLAG-HygroR, were co-transfected into HeLa S3 cells using Lipofectamine 2000 (Invitrogen, 11668019), according to the manufacturer's instruction. Cells were selected under treatment of Geneticin (gibco, 10131035) at 600 μg/mL and Hygromycin B (gibco, 10687010) at 250 μg/mL for about 14 days. The positive cells were then sorted into 96-well plates. The successfully engineered cell derivatives were validated by western blotting and PCR.

HCT116 MCM5-mAID: The sgRNA (sequence: CACCGAGGCGGCGCGACTCACTTG) targeting C-terminal *MCM5* gene was cloned into pX330 vector (addgene, 42230). The DNA cassettes of mAID-NeomycinR and mAID-HygromycinR from pMK286 (Addgene, 72824) and pMK287 (Addgene, 72825) were inserted into pUC57 flanked by the homology arms of C-terminal *MCM5* genomic sequences respectively to generate pUC57-MCM5-mAID-NeoR and pUC57-MCM5-mAID-HygroR. The three plasmids, pX330-sgMCM5, pUC57-MCM5-mAID-NeoR and pUC57-MCM5-mAID-HygroR, were co-transfected into HCT116 cells using Lipofectamine 2000 (Invitrogen, 11668019), according to the manufacturer's instruction. With the successfully engineered cell line, the OsTIR1-F74A was then introduced into MCM5-mAID cells using lentivirus infection.

Cell lines for ectopically expressing MCM2/5: The relevant cell lines were constructed using lentivirus infection, followed by antibiotics selection. The lentiviruses were generated by co-transfection of pLEX vector containing either WT or mutated MCM2/5, pMD2.G (Addgene, 12259) and psPAX2 (Addgene, 12260) into HEK 293T cell at a molar ration of 4:1:3. The isolated lentiviruses were used to treat HeLa S3 or mAID (HCT116) cells. Cells were selected with Puromycin at 0.5 μg/mL for further experiments.

hMCM-DH purification

To isolate the endogenous hMCM-DH(WT/WT), *MCM7*-Stag-3×FLAG HeLa S3 cells were first treated with nocodazole (Sigma, M1404) at 50 ng/ml for 16 hours. After being released into fresh medium for 8 hours, cell samples were harvested by centrifuge at 1200 rpm for 2 minutes and lysed by extraction buffer (EBX, 100 mM L-Glutamic acid potassium salt, 50 mM

HEPES/KOH pH 7.5, 10 mM Magnesium acetate, 1 mM EDTA, 0.5% Triton X-100, 3 mM ATP, 2 mM NaF, 1 mM Na_3VO_4, 1 mM phenylmethanesulfonylfluoride (PMSF), 1× protease inhibitor cocktail (Roche)) on ice for 30 minutes. Crude chromatin was collected by centrifuge at 3000 rpm for 2 minutes and treated with 1U/μL benzonase (7sea biotech, RPE002) in freshly prepared benzonase buffer (100 mM L-Glutamic acid potassium salt, 50 mM HEPES/KOH, pH 7.5, 8 mM Magnesium chloride, 1 mM EDTA, 0.02% NP-40, 3 mM ATP, 2 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF, 1× protease inhibitor cocktail (Roche)) at 37℃ for 10 minutes and subsequently on ice for 1 hour. The clarified soluble fraction was then subjected to anti-FLAG immunoprecipitation with anti-FLAG Affinity Gel (sigma, A2220) at 4℃ for 4 hours. Beads were recovered and washed extensively with benzonase buffer. For elution, 0.5 mg/mL of 3×FLAG peptide (final concentration) was added and incubated with the beads at 4℃ for 20 min. Eluates were applied on the top of 20%-40% glycerol gradient in benzonase buffer with 1× protease inhibitor. The gradients were centrifuged in a TLS-55 rotor (Beckman Optima MAX-XP Ultracentrifuge) at 105,000 g for 13 hours. The fractions containing MCM-DH were pooled and processed for electron microscopy analysis.

To prepare the samples for hMCM-DH(RL/WT) purification, pLEX-MCM5-RL-3xFLAG were transfected into HeLa S3 cells using lentivirus to generate MCM5-RL-3xFLAG cell line. Sample collection and purification of the mutant DHs were similar to the hMCM-DH(WT/WT).

FACS

Cell samples were harvested at indicated time points as shown in Figures 3G and S7F. Cells were fixed in 75% of ethanol at 4℃ overnight. After washing with 1xPBS twice, cells were suspended in 50 μg/mL of Propidium iodide (Sigma, P4170) and 100 μg/mL of RNase A (Sigma, R5503), and incubated for 30 minutes at room temperature in the dark. The stained cells were then subjected to Flow Cytometer and Cell Sorter workstation (BD FACSArialIII). Cell cycle distribution was analyzed by FlowJo.

Double hexamer stability assay

The MCM-DHs containing 3xFLAG tagged MCM5-WT, MCM5-RL, MCM5-R, MCM5-L or MCM2-GG were purified from HeLa S3 stable cell lines similar as that of the endogenous MCM7-3xFLAG cells. After glycerol gradient sedimentation, the fractions containing DHs were collected and processed. The relevant DHs were then incubated in buffers (50 mM HEPES/KOH, pH 7.5, 8 mM Magnesium chloride, 1 mM EDTA, 0.02% NP-40, 3 mM ATP,

2 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF, 1× protease inhibitor cocktail (Roche)) containing 100 mM and 400 mM of NaCl respectively, for 30 min at 30℃ in the thermomixer. Afterwards, 3 μl of the sample was applied onto copper grids and stained with 2% uranyl acetate. Images were acquired with Talos L120C microscope at 57,000x magnification. 2D classification was performed to calculate the percentage of MCM-DHs among all MCM particles as DH/(DH+SH)x100%.

EM data acquisition

For negative staining, 3 μl of protein samples were placed onto glow-discharged grids covered with a thin layer of continuous carbon film (Ted Pella) and stained with 2% (w/v) uranyl acetate. Grids were imaged on a Talos L120C microscope (Thermo Fisher Scientific) operated at 120 kV with a 4k × 4k Ceta 16M camera. A magnification of 57,000×, corresponding to a pixel size of 2.49 Å on the specimen, and a defocus around −1.5 μm were used for image recording. The information from this analysis was used to determine sample quality and estimate the relative concentration of samples used for cryo-grid preparation.

For cryo-EM, multiple-loading strategy was applied to increase particle numbers in the grid holes. Each time, 2.5 μl of concentrated hMCM-DH (WT/WT and RL/WT) samples were applied to the glow-discharged holey carbon grids with ultrathin carbon film (Quantifoil R1.2/1.3, 300 mesh Cu with 2 nm C). The grids were blotted with filter paper (Ted Pella) and then plunge frozen in liquid ethane using a Mark IV Vitrobot (Thermo Fisher Scientific) with a blotting time of 3 s at 4 ℃ and 100% humidity. Cryo-EM data collection was done with a FEI Titan Krios G3i electron microscope (Thermo Fisher Scientific) equipped with a high-brightness field emission gun operated at 300 kV. Images were collected with a K3 Summit direct electron detector (Gatan) using EPU in counting mode at a calibrated magnification of 81,000× (1.06 Å physical pixel size). The slit width of the Gatan Imaging Filter (GIF) Bio Quantum was set to 20 eV. Dose rate, total dose, and defocus range used for data collection are summarized in Table S1.

Image processing

For Cryo-EM data, drift correction was performed using MotionCor2[69]. Motion-corrected sums without dose-weighting were used for contrast transfer function (CTF) estimation with GCTF [72]. Motion-corrected sums with dose-weighting were used for all other image processing. Particles picked automatically by Gautomatch (https://www.mrc-

lmb.cam.ac.uk/kzhang/Gautomatch/) were extracted by RELION[71] and then imported into cryoSPARC[73] for 3D classification and refinement procedures. Topaz was applied for particle picking with the hMCM-DH(RL/WT) dataset[82,83]. The initial model was generated using *ab-initio* reconstruction. Particles from candidate classes by *ab-initio* reconstruction were further sorted by iterative 2D classification and heterogeneous refinement. Well-sorted particles were finally subjected to homogenous and non-uniform refinements to generate the final maps, including two half maps, an unsharpened map, and a sharpened map. Resolution was estimated using the Fourier shell correlation 0.143 criterion. Detailed information on data processing can be found in Figure S2 and S8. Local resolution map was calculated using cryoSPARC. Directional FSC (dFSC) was calculated using a similar approach as reported previously[84] with an improved script.

Classification focusing on origin DNA was also performed for both hMCM-DH(WT/WT) and hMCM-DH(RL/WT) particles. The detailed procedures were shown in Figures S1P and S6. The densities corresponding MCM-DH were subtracted from their original particles using a standard procedure[85] in cryoSPARC. The metadata of the particle information was then converted to star file format using UCSF pyem[76]. Subsequent classification with a mask on DNA densities was conducted without any alignments in relion 3.1[71]. The relevant classification information for the subtracted particles was mapped back to their original data using a home-made python script.

Model building

Ab initio model building was carried out in Coot[75] and PHENIX[77]. MCM2-7 structure in Human CMG complex (PDB: 6XTX)[33] was used as an initial model. Since the register of origin DNA engaged with endogenous hMCM-DHs is heterogenous, it is likely that there is a mix of purines and pyrimidines at each nucleotide position. For this reason we could not build the origin DNA sequence with certainty. Instead, two superposed polyA:polyT and polyT:polyA duplexes, in which all atoms is set at 50 % occupancy, were modelled to surrogate for the heterogenous sequences. For clarity, only one duplex was shown in figures. The real-space refinement of the model was carried out using Phenix.real_space_refine with Ramachandran restraint[77] and further manual adjustment in COOT. Molprobity[78] was used to evaluate the stereochemistry and geometry of the structure for manual adjustment. The program UCSF Chimera[74] and PyMOL (http://pymol.org) were used to prepare images.

Mapping of the hMCM-DH footprints

The samples for isolating DNA bound by hMCM-DHs were prepared as described in the section of hMCM-DH purification. During this process, the solubilized chromatin fraction was collected for isolating the input DNA. The protein-DNA samples were then treated with 0.5 mg/ml of proteinase K (Thermo, 25530049) and 0.25 mg/ml of RNase A (Sigma, R5503) at 65℃ for 30 min. Cold isopropanol was added to precipitate DNA overnight at 4℃. Pellet DNA by centrifuge at 21,500 g for 30 min. The pellet was washed twice with 70% ethanol and dissolved in elution buffer (5 mM Tris/HCl, pH 8.5). DNA concentration was measured by Qubit 4 Fluorometer (Invitrogen) with Qubit dsDNA HS Assay Kit (Invitorgen, Q32854) and fragment size was analyzed by Fragment Analyzer System (Agilent). 5 ng DNA was ligated to adapters at a ratio of 1:200 using KAPA HyperPrep Kit (Roche, 07962347001). Sequencing Libraries were generated with KAPA HyperPrep Kit according to the manufacturer's protocol. To recover short dsDNA fragments, a higher KAPA Pure Beads-to-sample ratio was used to purify the DNA. Quantification of adaptor-ligated DNA was performed using quantitative real-time PCR (KAPA Library Quantification Kit) before library amplification to determine the optimal PCR cycle number. The fragment size of amplified libraries was analyzed on a fragment analyzer (Agilent) before library pooling and sequencing. The libraries were sequenced on an Illumina Nextseq 500 system. Adaptor trimming was performed using Trim Galore. After trimming, reads were aligned to the human reference genome (GRCh37) with Bowtie 2 (version 1.2.0) (parameters "--phred33-quals -v 3 -m 1 -p 24 --best --strata"). PCR duplicates were removed using Picard MarkDuplicates (version 2.23.4, http://broadinstitute.github.io/picard). Only the unique mapped reads were used for the downstream analysis. DNA lengths protected by hMCM-DHs were analyzed using bamPEFragmentSize of deepTools (version 3.5.1). For further checking the size distribution of MCM reads (MCM-DH footprint size), the mapped BAM files were transformed to bed format by bedtools (version 2.30.0). All reads (paired-end read fragments) between 35 to 80 nt were then classified into 9 classes with a step size of 5 nucleotides (nt). The read depth of each class was further calculated and visualized in heatmap mode with IGV (see Figure 6B for an example).

MCM-DH cluster and classification

The bed file for raw read positions (without duplicates) were extracted from BAM files using BEDTools (version 2.30.0), and the read densities along the genome were computed using bedtools makewindow with the corresponding coverage option (e.g. 1kb or 5kb, etc). The

Pearson correlation scores between two replicates as well as the control sample were calculated (Figure S7C) using deepTools with default parameters[86] or with R (version 4.1.2 <https://www.r-project.org>). RPKM for each sample is calculated per 100bp non-overlapping window across the genome. The MCM clusters were defined with the following peak calling process: the bins with rpkm values larger than the corresponding thresholds (Rep1: > 0.6, Rep2: > 0.8) were selected and then the adjacent calling peaks within relatively small gap width (Rep1: 6kb, Rep2: 8kb) were merged to get the final clusters (with minimum size cutoff: Rep1 <= 10.5 kb, Rep2 <= 13 kb). The parameters of peak calling were chosen based on the signal noise ratio of each replicate in order to maximize the reproducibility of two replicates.

Comparative analysis of MCM-DH data with other datasets

The raw Repli-Seq data of HeLa S3 cells were downloaded from the Encode project (http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeUwRepliSeq)[67] and S50 (the fraction of S phase at which 50% of the DNA is replicated in a defined genome region) was computed[40]. The mean replication timing (S50 values) of each MCM cluster were calculated and used to classify the MCM clusters into 4 timing groups, i.e. early (S50≤0.25), mid-early (S50: 0.25~0.5), mid-late (S50: 0.5~0.75) and late (S50>0.75). The OK-seq RFD data and the corresponding initiation zones were obtained from[41]. The ORM data and the corresponding initiation zones were obtained from[42]. The data of H2AZ and RNA Pol II-pS2 (ENCSR000ECT) were downloaded from Encode project <https://www.encodeproject.org>. GRO-seq data were downloaded from a previous nascent RNA transcriptome study [48] and Orc1 ChIP-seq data were previously published [44]. Additional gene transcription (RNA-seq) and gene annotation data were retrieved from a previous study [68]. The published MCM ChEC-seq data were downloaded from GEO (GSE150800)[58]. The nucleotide composition data were downloaded from UCSC Genome Browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/gc5Base/). For analysis of base composition around the centers of the MCM-DH-bound DNA fragements (Figure S7E), only those fragments with an odd length between 51 and 59 nucleotides were used for this analysis. It should be noted that, given the fact that Benzonase favors G/C rich sequences[87], a slightly higher G/C content is enriched at the borders of both hMCM-DH-bound DNA and the input DNA. For some instances, genomic coordinates were remapped to GRCh37 using LiftOver <https://genome.ucsc.edu/cgi-bin/hgLiftOver>. The metaplots and heatmaps were generated using deepTools.

QUANTIFICATION AND STATISTICAL ANALYSIS

The resolution estimations of cryo-EM density maps are based on the 0.143 Fourier Shell Correlation (FSC) criterion. Statistical analyses were performed in deepTools or R. Statistical tests and parameter are reported in the text, figure legends and methods.

Supplemental video and Excel table titles and legends

Supplementary Video 1. Idealized melting of B-form DNA at the hexamer junction, related to Figure 2.

Supplementary Video 2. Morphing of H2I loops from lagging strand to leading strand between hMCM-DH and hCMG during MSSB formation, related to Figure 5.

Table S1 Summary of Cryo-EM Data Collection and Model Refinement, related to Figures 1, S1 and S6 and STAR Methods.

References

1 Bleichert, F., Botchan, M. R. & Berger, J. M. (2017). Mechanisms for initiating cellular DNA replication. Science 35510.1126/science.aah6317.

2 Bell, S. P. & Labib, K. (2016). Chromosome Duplication in Saccharomyces cerevisiae. Genetics 203, 1027-1067, doi:10.1534/genetics.115.186452.

3 Miller, T. C. R., Locke, J., Greiwe, J. F., Diffley, J. F. X. & Costa, A. (2019). Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. Nature 575, 704-710, doi:10.1038/s41586-019-1768-0.

4 Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B. K. & Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 A. Nature 524, 186-191, doi:10.1038/nature14685.

5 Noguchi, Y., Yuan, Z., Bai, L., Schneider, S., Zhao, G., Stillman, B., Speck, C. & Li, H. (2017). Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model. Proc Natl Acad Sci U S A 114, E9529-E9538, doi:10.1073/pnas.1712537114.

6 Abid Ali, F., Douglas, M. E., Locke, J., Pye, V. E., Nans, A., Diffley, J. F. X. & Costa, A. (2017). Cryo-EM structure of a licensed DNA replication origin. Nat Commun 8, 2241, doi:10.1038/s41467-017-02389-0.

7 Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P. & Diffley, J. F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell 139, 719-730, doi:10.1016/j.cell.2009.10.015.

8 Tanaka, S. & Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. Cold Spring Harb Perspect Biol 5, a010371, doi:10.1101/cshperspect.a010371.

9 Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. (2018). The mechanism of eukaryotic CMG helicase activation. Nature 555, 265-268, doi:10.1038/nature25787.

10 Georgescu, R., Yuan, Z., Bai, L., de Luna Almeida Santos, R., Sun, J., Zhang, D., Yurieva, O., Li, H. & O'Donnell, M. E. (2017). Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. Proc Natl Acad Sci U S A 114, E697-E706, doi:10.1073/pnas.1620500114.

11 Bochman, M. L. & Schwacha, A. (2015). DNA replication: Strand separation unravelled. Nature 524, 166-167, doi:10.1038/nature14643.

12 Zhai, Y., Li, N., Jiang, H., Huang, X., Gao, N. & Tye, B. K. (2017). Unique Roles of the Non-identical MCM Subunits in DNA Replication Licensing. Mol Cell 67, 168-179, doi:10.1016/j.molcel.2017.06.016.

13 Cheng, J., Li, N., Huo, Y., Dang, S., Tye, B. K., Gao, N. & Zhai, Y. (2022). Structural Insight into the MCM double hexamer activation by Dbf4-Cdc7 kinase. Nat Commun 13, 1396, doi:10.1038/s41467-022-29070-5.

14 Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I. & Remus, D. (2015). Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. Mol Cell 60, 797-807, doi:10.1016/j.molcel.2015.10.022.

15 Greiwe, J. F., Miller, T. C. R., Locke, J., Martino, F., Howell, S., Schreiber, A., Nans, A., Diffley, J. F. X. & Costa, A. (2022). Structural mechanism for the selective phosphorylation of DNA-loaded MCM double hexamers by the Dbf4-dependent kinase. Nat Struct Mol Biol 29, 10-20, doi:10.1038/s41594-021-00698-z.

16 Lewis, J. S., Gross, M. H., Sousa, J., Henrikus, S. S., Greiwe, J. F., Nans, A., Diffley, J. F. X. & Costa, A. (2022). Mechanism of replication origin melting nucleated by CMG helicase assembly. Nature 606, 1007-1014, doi:10.1038/s41586-022-04829-4.
17 Chong, J. P., Mahbubani, H. M., Khoo, C. Y. & Blow, J. J. (1995). Purification of an MCM-containing complex as a component of the DNA replication licensing system. Nature 375, 418-421, doi:10.1038/375418a0.
18 Thommes, P., Kubota, Y., Takisawa, H. & Blow, J. J. (1997). The RLF-M component of the replication licensing system forms complexes containing all six MCM/P1 polypeptides. EMBO J 16, 3312-3319, doi:10.1093/emboj/16.11.3312.
19 Kimura, H., Ohtomo, T., Yamaguchi, M., Ishii, A. & Sugimoto, K. (1996). Mouse MCM proteins: complex formation and transportation to the nucleus. Genes Cells 1, 977-993, doi:10.1046/j.1365-2443.1996.840284.x.
20 Walter, J. C. (2000). Evidence for sequential action of cdc7 and cdk2 protein kinases during initiation of DNA replication in Xenopus egg extracts. J Biol Chem 275, 39773-39778, doi:10.1074/jbc.M008107200.
21 Mimura, S., Kubota, Y. & Takisawa, H. (2018). MCM interference during licensing of DNA replication in Xenopus egg extracts-Possible Role of a C-terminal region of MCM3. Cell Cycle 17, 492-505, doi:10.1080/15384101.2017.1415681.
22 Gambus, A., Khoudoli, G. A., Jones, R. C. & Blow, J. J. (2011). MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. J Biol Chem 286, 11855-11864, doi:10.1074/jbc.M110.199521.
23 Hyrien, O. (2016). How MCM loading and spreading specify eukaryotic DNA replication initiation sites. F1000Res 510.12688/f1000research.9008.1.
24 Rhind, N. & Gilbert, D. M. (2013). DNA replication timing. Cold Spring Harb Perspect Biol 5, a010132, doi:10.1101/cshperspect.a010132.
25 Cvetic, C. & Walter, J. C. (2005). Eukaryotic origins of DNA replication: could you please be more specific? Semin Cell Dev Biol 16, 343-353, doi:10.1016/j.semcdb.2005.02.009.
26 Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C. & Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science 290, 2309-2312, doi:10.1126/science.290.5500.2309.
27 Parker, M. W., Botchan, M. R. & Berger, J. M. (2017). Mechanisms and regulation of DNA replication initiation in eukaryotes. Crit Rev Biochem Mol Biol 52, 107-144, doi:10.1080/10409238.2016.1274717.
28 Hayashi-Takanaka, Y., Hayashi, Y., Hirano, Y., Miyawaki-Kuwakado, A., Ohkawa, Y., Obuse, C., Kimura, H., Haraguchi, T. & Hiraoka, Y. (2021). Chromatin loading of MCM hexamers is associated with di-/tri-methylation of histone H4K20 toward S phase entry. Nucleic Acids Res 49, 12152-12166, doi:10.1093/nar/gkab1068.
29 Mei, L., Kedziora, K. M., Song, E. A., Purvis, J. E. & Cook, J. G. (2022). The consequences of differential origin licensing dynamics in distinct chromatin environments. Nucleic Acids Res10.1093/nar/gkac003.
30 Prioleau, M.-N. & MacAlpine, D. M. (2016). DNA replication origins—where do we begin? Genes & Development 30, 1683-1697, doi:10.1101/gad.285114.116.
31 Yuan, Z., Bai, L., Sun, J., Georgescu, R., Liu, J., O'Donnell, M. E. & Li, H. (2016). Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. Nat Struct Mol Biol 23, 217-224, doi:10.1038/nsmb.3170.

32 Eickhoff, P., Kose, H. B., Martino, F., Petojevic, T., Abid Ali, F., Locke, J., Tamberg, N., Nans, A., Berger, J. M., Botchan, M. R., Yardimci, H. & Costa, A. (2019). Molecular Basis for ATP-Hydrolysis-Driven DNA Translocation by the CMG Helicase of the Eukaryotic Replisome. Cell Rep 28, 2673-2688 e2678, doi:10.1016/j.celrep.2019.07.104.
33 Rzechorzek, N. J., Hardwick, S. W., Jatikusumo, V. A., Chirgadze, D. Y. & Pellegrini, L. (2020). CryoEM structures of human CMG-ATPgammaS-DNA and CMG-AND-1 complexes. Nucleic Acids Res 48, 6980-6995, doi:10.1093/nar/gkaa429.
34 Jenkyn-Bedford, M., Jones, M. L., Baris, Y., Labib, K. P. M., Cannone, G., Yeeles, J. T. P. & Deegan, T. D. (2021). A conserved mechanism for regulating replisome disassembly in eukaryotes. Nature 600, 743-747, doi:10.1038/s41586-021-04145-3.
35 Sheu, Y. J. & Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. Nature 463, 113-117, doi:10.1038/nature08647.
36 Saleh, A., Noguchi, Y., Aramayo, R., Ivanova, M. E., Stevens, K. M., Montoya, A., Sunidhi, S., Carranza, N. L., Skwark, M. J. & Speck, C. (2022). The structural basis of Cdc7-Dbf4 kinase dependent targeting and phosphorylation of the MCM2-7 double hexamer. Nat Commun 13, 2915, doi:10.1038/s41467-022-30576-1.
37 Tsuji, T., Ficarro, S. B. & Jiang, W. (2006). Essential Role of Phosphorylation of MCM2 by Cdc7/Dbf4 in the Initiation of DNA Replication in Mammalian Cells. Molecular Biology of the Cell 17, 4459-4472, doi:10.1091/mbc.e06-03-0241.
38 Bochman, M. L. & Schwacha, A. (2009). The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol Mol Biol Rev 73, 652-683, doi:10.1128/MMBR.00019-09.
39 Meagher, M., Epling, L. B. & Enemark, E. J. (2019). DNA translocation mechanism of the MCM complex and implications for replication initiation. Nat Commun 10, 3117, doi:10.1038/s41467-019-11074-3.
40 Chen, C. L., Rappailles, A., Duquenne, L., Huvet, M., Guilbaud, G., Farinelli, L., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O. & Thermes, C. (2010). Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res 20, 447-457, doi:10.1101/gr.098947.109.
41 Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C. L. & Hyrien, O. (2016). Replication landscape of the human genome. Nat Commun 7, 10208, doi:10.1038/ncomms10208.
42 Wang, W., Klein, K. N., Proesmans, K., Yang, H., Marchal, C., Zhu, X., Borrman, T., Hastie, A., Weng, Z., Bechhoefer, J., Chen, C. L., Gilbert, D. M. & Rhind, N. (2021). Genome-wide mapping of human DNA replication by optical replication mapping supports a stochastic model of eukaryotic replication. Mol Cell 81, 2975-2988 e2976, doi:10.1016/j.molcel.2021.05.024.
43 Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J. M. & Lemaitre, J. M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat Struct Mol Biol 19, 837-844, doi:10.1038/nsmb.2339.
44 Dellino, G. I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M. & Pelicci, P. G. (2013). Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate

origin selection and replication timing. Genome Research 23, 1-11, doi:10.1101/gr.142331.112.

45 Long, H., Zhang, L., Lv, M., Wen, Z., Zhang, W., Chen, X., Zhang, P., Li, T., Chang, L., Jin, C., Wu, G., Wang, X., Yang, F., Pei, J., Chen, P., Margueron, R., Deng, H., Zhu, M. & Li, G. (2020). H2A.Z facilitates licensing and activation of early replication origins. Nature 577, 576-581, doi:10.1038/s41586-019-1877-9.

46 Chou, H.-C., Bhalla, K., Demerdesh, O. E. L., Klingbeil, O., Hanington, K., Aganezov, S., Andrews, P., Alsudani, H., Chang, K., Vakoc, C. R., Schatz, M. C., McCombie, W. R. & Stillman, B. (2021). The human origin recognition complex is essential for pre-RC assembly, mitosis, and maintenance of nuclear structure. eLife 10, e61797, doi:10.7554/eLife.61797.

47 Zhang, J., Lee, D., Dhiman, V., Jiang, P., Xu, J., McGillivray, P., Yang, H., Liu, J., Meyerson, W., Clarke, D., Gu, M., Li, S., Lou, S., Xu, J., Lochovsky, L., Ung, M., Ma, L., Yu, S., Cao, Q., Harmanci, A., Yan, K.-K., Sethi, A., Gürsoy, G., Schoenberg, M. R., Rozowsky, J., Warrell, J., Emani, P., Yang, Y. T., Galeev, T., Kong, X., Liu, S., Li, X., Krishnan, J., Feng, Y., Rivera-Mulia, J. C., Adrian, J., Broach, J. R., Bolt, M., Moran, J., Fitzgerald, D., Dileep, V., Liu, T., Mei, S., Sasaki, T., Trevilla-Garcia, C., Wang, S., Wang, Y., Zang, C., Wang, D., Klein, R. J., Snyder, M., Gilbert, D. M., Yip, K., Cheng, C., Yue, F., Liu, X. S., White, K. P. & Gerstein, M. (2020). An integrative ENCODE resource for cancer genomics. Nature Communications 11, 3696, doi:10.1038/s41467-020-14743-w.

48 Andersson, R., Refsing Andersen, P., Valen, E., Core, L. J., Bornholdt, J., Boyd, M., Heick Jensen, T. & Sandelin, A. (2014). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat Commun 5, 5336, doi:10.1038/ncomms6336.

49 Kirstein, N., Buschle, A., Wu, X., Krebs, S., Blum, H., Kremmer, E., Vorberg, I. M., Hammerschmidt, W., Lacroix, L., Hyrien, O., Audit, B. & Schepers, A. (2021). Human ORC/MCM density is low in active genes and correlates with replication time but does not delimit initiation zones. eLife 10, e62161, doi:10.7554/eLife.62161.

50 Schilbach, S., Aibara, S., Dienemann, C., Grabbe, F. & Cramer, P. (2021). Structure of RNA polymerase II pre-initiation complex at 2.9A defines initial DNA opening. Cell 184, 4064-4072.e4028, doi:10.1016/j.cell.2021.05.012.

51 Lee, C. S. K., Cheung, M. F., Li, J., Zhao, Y., Lam, W. H., Ho, V., Rohs, R., Zhai, Y., Leung, D. & Tye, B. K. (2021). Humanizing the yeast origin recognition complex. Nat Commun 12, 33, doi:10.1038/s41467-020-20277-y.

52 MacAlpine, D. M. & Almouzni, G. (2013). Chromatin and DNA Replication. Cold Spring Harbor Perspectives in Biology 510.1101/cshperspect.a010207.

53 Powell, S. K., MacAlpine, H. K., Prinz, J. A., Li, Y., Belsky, J. A. & MacAlpine, D. M. (2015). Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. EMBO J 34, 531-543, doi:10.15252/embj.201488307.

54 Froelich, C. A., Kang, S., Epling, L. B., Bell, S. P. & Enemark, E. J. (2014). A conserved MCM single-stranded DNA binding element is essential for replication initiation. Elife 3, e01993, doi:10.7554/eLife.01993.

55 Zhai, Y., Cheng, E., Wu, H., Li, N., Yung, P. Y., Gao, N. & Tye, B. K. (2017). Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer. Nat Struct Mol Biol 24, 300-308, doi:10.1038/nsmb.3374.

56 Samel, S. A., Fernandez-Cid, A., Sun, J., Riera, A., Tognetti, S., Herrera, M. C., Li, H. & Speck, C. (2014). A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. Genes Dev 28, 1653-1666, doi:10.1101/gad.242404.114.
57 Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. (2019). Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. Cell 178, 600-611.e616, doi:10.1016/j.cell.2019.06.032.
58 Foss, E. J., Sripathy, S., Gatbonton-Schwager, T., Kwak, H., Thiesen, A. H., Lao, U. & Bedalov, A. (2021). Chromosomal Mcm2-7 distribution and the genome replication program in species from yeast to humans. PLoS Genet 17, e1009714, doi:10.1371/journal.pgen.1009714.
59 Das, S. P., Borrman, T., Liu, V. W. T., Yang, S. C.-H., Bechhoefer, J. & Rhind, N. (2015). Replication timing is regulated by the number of MCMs loaded at origins. Genome Research 25, 1886-1892, doi:10.1101/gr.195305.115.
60 Miotto, B., Ji, Z. & Struhl, K. (2016). Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc Natl Acad Sci U S A 113, E4810-4819, doi:10.1073/pnas.1609060113.
61 Gao, J., Wang, Q., Dong, C., Chen, S., Qi, Y. & Liu, Y. (2015). Whole Exome Sequencing Identified MCM2 as a Novel Causative Gene for Autosomal Dominant Nonsyndromic Deafness in a Chinese Family. PLoS One 10, e0133522, doi:10.1371/journal.pone.0133522.
62 Gineau, L., Cognet, C., Kara, N., Lach, F. P., Dunne, J., Veturi, U., Picard, C., Trouillet, C., Eidenschenk, C., Aoufouchi, S., Alcais, A., Smith, O., Geissmann, F., Feighery, C., Abel, L., Smogorzewska, A., Stillman, B., Vivier, E., Casanova, J. L. & Jouanguy, E. (2012). Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. J Clin Invest 122, 821-832, doi:10.1172/JCI61014.
63 Vetro, A., Savasta, S., Russo Raucci, A., Cerqua, C., Sartori, G., Limongelli, I., Forlino, A., Maruelli, S., Perucca, P., Vergani, D., Mazzini, G., Mattevi, A., Stivala, L. A., Salviati, L. & Zuffardi, O. (2017). MCM5: a new actor in the link between DNA replication and Meier-Gorlin syndrome. Eur J Hum Genet 25, 646-650, doi:10.1038/ejhg.2017.5.
64 Shima, N., Alcaraz, A., Liachko, I., Buske, T. R., Andrews, C. A., Munroe, R. J., Hartford, S. A., Tye, B. K. & Schimenti, J. C. (2007). A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. Nat Genet 39, 93-98, doi:10.1038/ng1936.
65 Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D. & Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol 8, 358-366, doi:10.1038/ncb1382.
66 Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74, doi:10.1038/nature11247.
67 Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., Dorschner, M. O., Gartler, S. M. & Stamatoyannopoulos, J. A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 107, 139-144, doi:10.1073/pnas.0912402107.

68 Promonet, A., Padioleau, I., Liu, Y., Sanz, L., Biernacka, A., Schmitz, A. L., Skrzypczak, M., Sarrazin, A., Mettling, C., Rowicka, M., Ginalski, K., Chedin, F., Chen, C. L., Lin, Y. L. & Pasero, P. (2020). Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. Nat Commun 11, 3940, doi:10.1038/s41467-020-17858-2.
69 Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y. & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331-332, doi:10.1038/nmeth.4193.
70 Rohou, A. & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216-221, doi:10.1016/j.jsb.2015.08.008.
71 Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 710.7554/eLife.42166.
72 Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J Struct Biol 193, 1-12, doi:10.1016/j.jsb.2015.11.003.
73 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290-296, doi:10.1038/nmeth.4169.
74 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612, doi:10.1002/jcc.20084.
75 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486-501, doi:10.1107/S0907444910007493.
76 Asarnow, D., Palovcak, E. & Cheng, Y. (2019). UCSF pyem v0. 5. Zenodo. Zenodo https://doi.org/10.5281/zenodo.3576630.
77 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213-221, doi:10.1107/S0907444909052925.
78 Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12-21, doi:10.1107/S0907444909042073.
79 Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359, doi:10.1038/nmeth.1923.
80 Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842, doi:10.1093/bioinformatics/btq033.
81 Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dundar, F. & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160-165, doi:10.1093/nar/gkw257.
82 Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat Methods 16, 1153-1160, doi:10.1038/s41592-019-0575-8.

83 Bepler, T., Kelley, K., Noble, A. J. & Berger, B. (2020). Topaz-Denoise: general deep denoising models for cryoEM and cryoET. Nat Commun 11, 5208, doi:10.1038/s41467-020-18952-1.

84 Dang, S., Feng, S., Tien, J., Peters, C. J., Bulkley, D., Lolicato, M., Zhao, J., Zuberbuhler, K., Ye, W., Qi, L., Chen, T., Craik, C. S., Jan, Y. N., Minor, D. L., Jr., Cheng, Y. & Jan, L. Y. (2017). Cryo-EM structures of the TMEM16A calcium-activated chloride channel. Nature 552, 426-429, doi:10.1038/nature25024.

85 Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. (2015). Sampling the conformational space of the catalytic subunit of human gamma-secretase. Elife 410.7554/eLife.11182.

86 Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research 44, W160-W165, doi:10.1093/nar/gkw257.

87 Grontved, L., Bandle, R., John, S., Baek, S., Chung, H. J., Liu, Y., Aguilera, G., Oberholtzer, C., Hager, G. L. & Levens, D. (2012). Rapid genome-scale mapping of chromatin accessibility in tissue. Epigenetics Chromatin 5, 10, doi:10.1186/1756-8935-5-10.

Figure 1

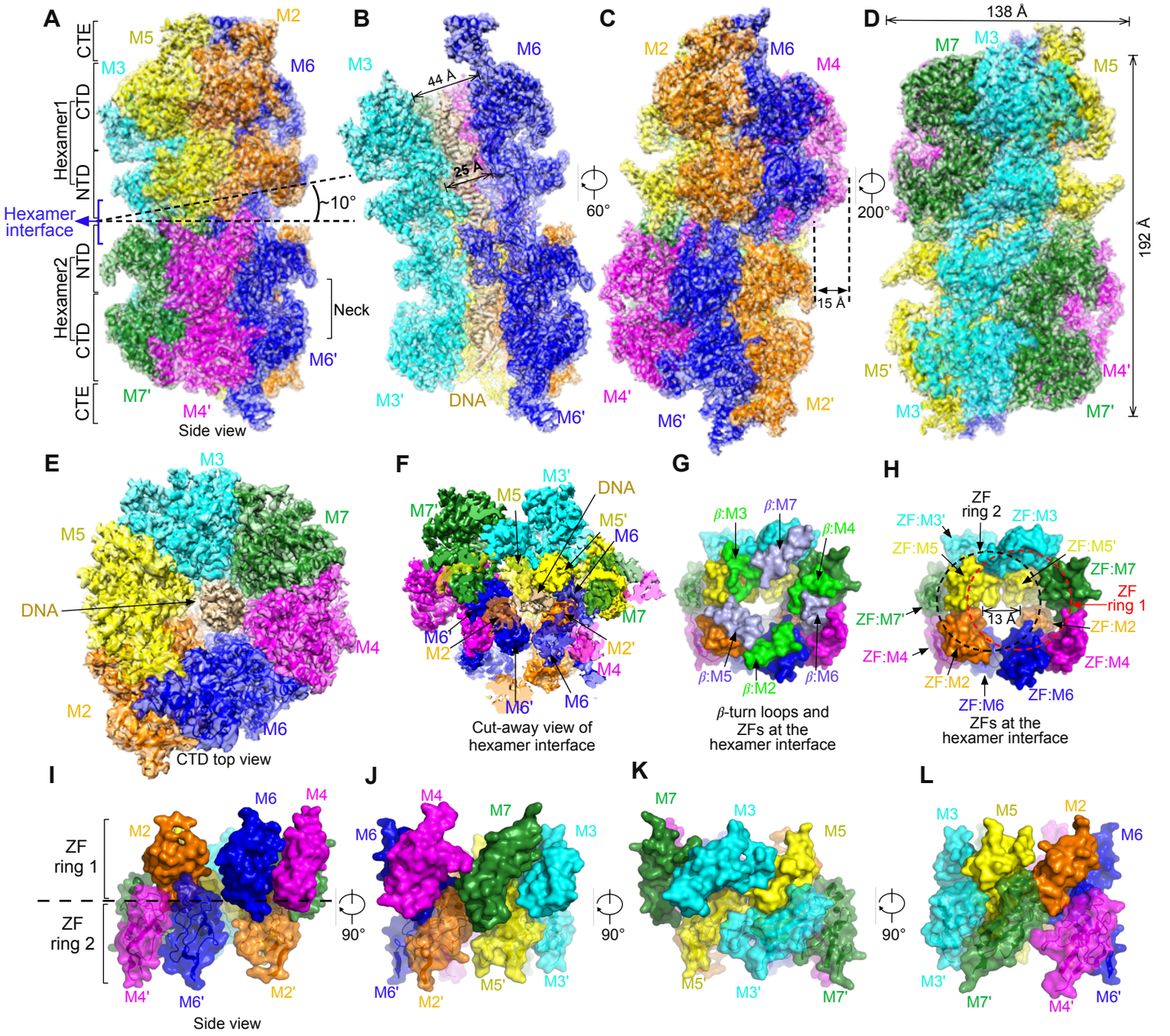

A
CTE
CTD
Hexamer1
NTD
Hexamer interface
NTD
Hexamer2
CTD
CTE
M2
M5
M3
M6
~10°
Neck
M6'
M7'
M4'
Side view
B
M3
M6
44 Å
25 Å
60°
M3'
DNA
M6'
C
M2
M6
M4
15 Å
M4'
M6'
M2'
D
138 Å
M7
M3
M5
200°
192 Å
M5'
M4'
M3'
M7'
E
M3
M7
M5
DNA
M4
M2
M6
CTD top view
F
M3'
DNA
M7'
M5
M5'
M6
M7
M6'
M2
M2'
M4
M6'
M6
Cut-away view of hexamer interface
G
β:M3
β:M7
β:M4
β:M5
β:M2
β:M6
β-turn loops and ZFs at the hexamer interface
H
ZF ring 2
ZF:M3'
ZF:M3
ZF:M5
ZF:M5'
ZF:M7
ZF ring 1
ZF:M7'
13 Å
ZF:M2
ZF:M4
ZF:M4
ZF:M2
ZF:M6
ZF:M6
ZFs at the hexamer interface
I
M6
M4
M2
ZF ring 1
ZF ring 2
90°
M4'
M6'
M2'
Side view
J
M4
M7
M6
M3
90°
M6'
M2'
M5'
M3'
K
M7
M3
M5
90°
M5'
M3'
M7'
L
M3
M5
M2
M6
M3'
M7'
M4'
M6'

Figure 2

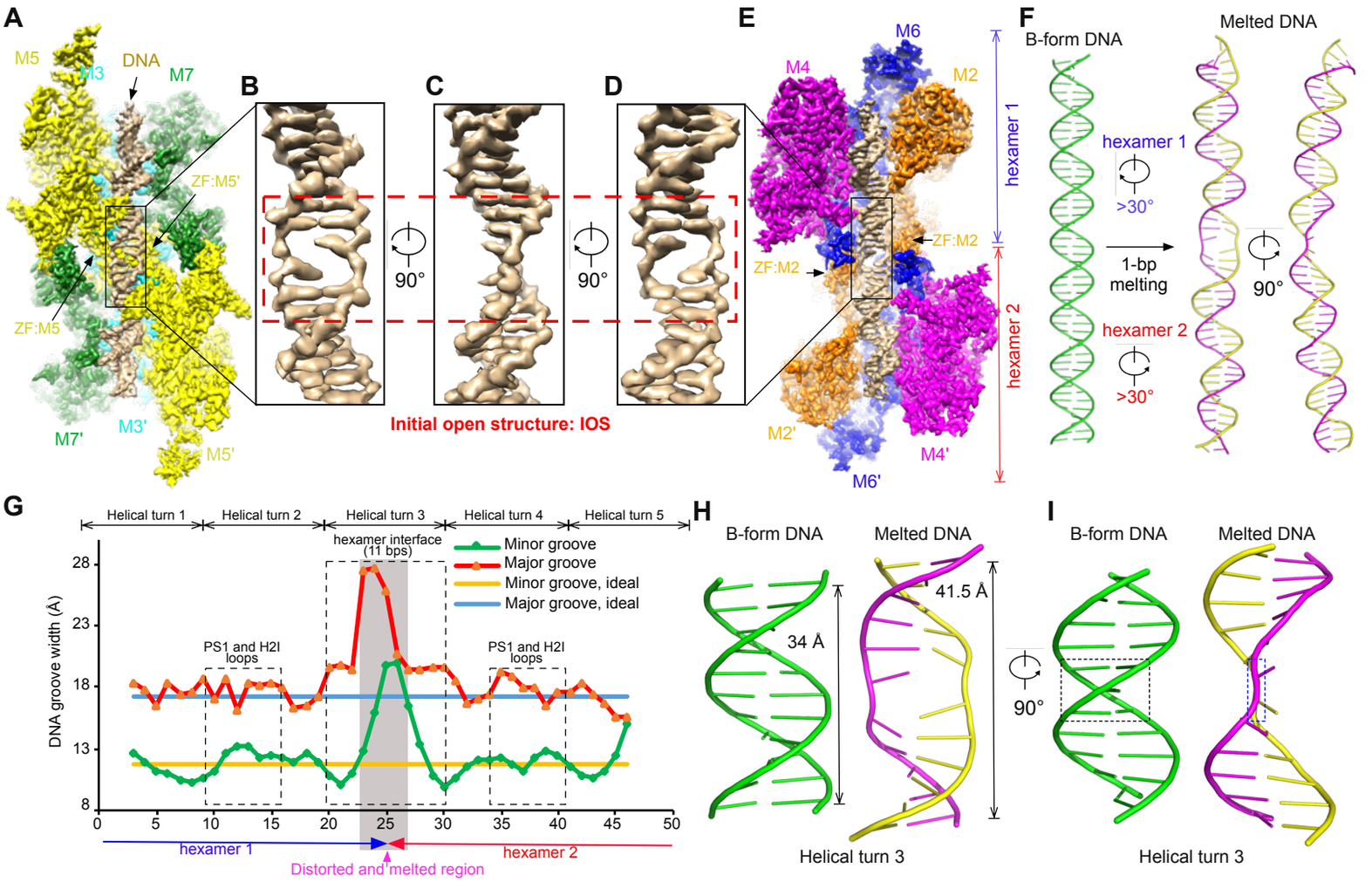

A
M5
M3
DNA
M7
ZF:M5'
ZF:M5
M7'
M3'
M5'
B
C
D
90°
90°
Initial open structure: IOS
E
M6
M4
M2
ZF:M2
ZF:M2
M2'
M4'
M6'
hexamer 1
hexamer 2
F
B-form DNA
Melted DNA
hexamer 1
>30°
1-bp
melting
hexamer 2
>30°
90°
G
Helical turn 1
Helical turn 2
Helical turn 3
Helical turn 4
Helical turn 5
hexamer interface
(11 bps)
Minor groove
Major groove
Minor groove, ideal
Major groove, ideal
DNA groove width (Å)
PS1 and H2I
loops
PS1 and H2I
loops
28
23
18
13
8
0
5
10
15
20
25
30
35
40
45
50
hexamer 1
hexamer 2
Distorted and melted region
H
B-form DNA
Melted DNA
34 Å
41.5 Å
90°
Helical turn 3
I
B-form DNA
Melted DNA
Helical turn 3

Figure 3

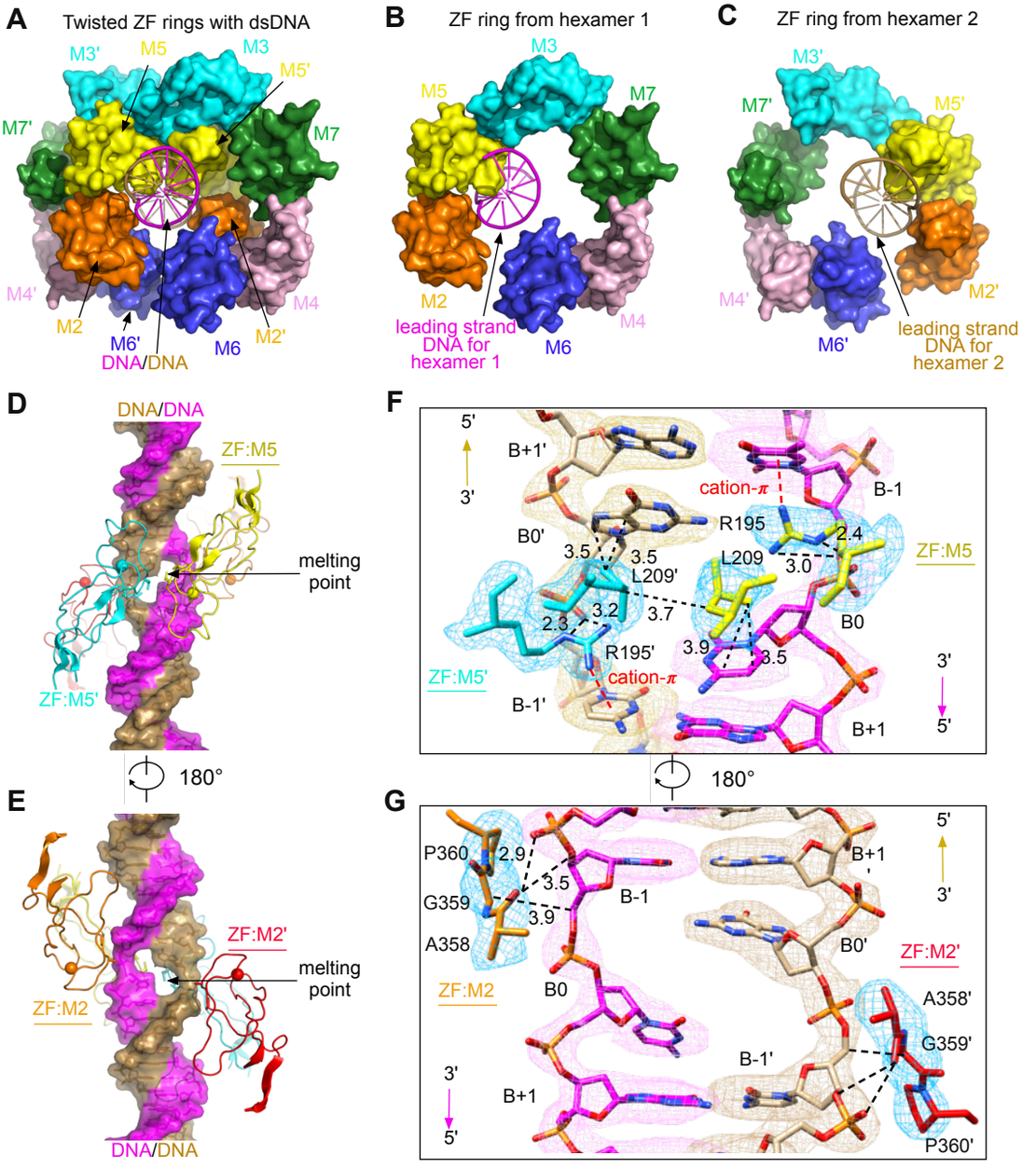

A
Twisted ZF rings with dsDNA
M3'
M5
M3
M5'
M7'
M7
M4'
M4
M2
M6'
M6
M2'
DNA/DNA
B
ZF ring from hexamer 1
M3
M5
M7
M2
M4
leading strand DNA for hexamer 1
M6
C
ZF ring from hexamer 2
M3'
M7'
M5'
M4'
M2'
M6'
leading strand DNA for hexamer 2
D
DNA/DNA
ZF:M5
melting point
ZF:M5'
180°
E
ZF:M2'
melting point
ZF:M2
DNA/DNA
F
5'
3'
B+1'
cation-π
B-1
B0'
R195
2.4
3.5
3.5
L209
3.0
ZF:M5
L209'
3.2
3.7
2.3
B0
R195'
3.9
3.5
3'
ZF:M5'
cation-π
B-1'
B+1
5'
180°
G
P360
2.9
3.5
B-1
5'
3'
G359
3.9
B+1
A358
B0'
ZF:M2'
ZF:M2
B0
A358'
G359'
B-1'
3'
B+1
5'
P360'

Figure 4

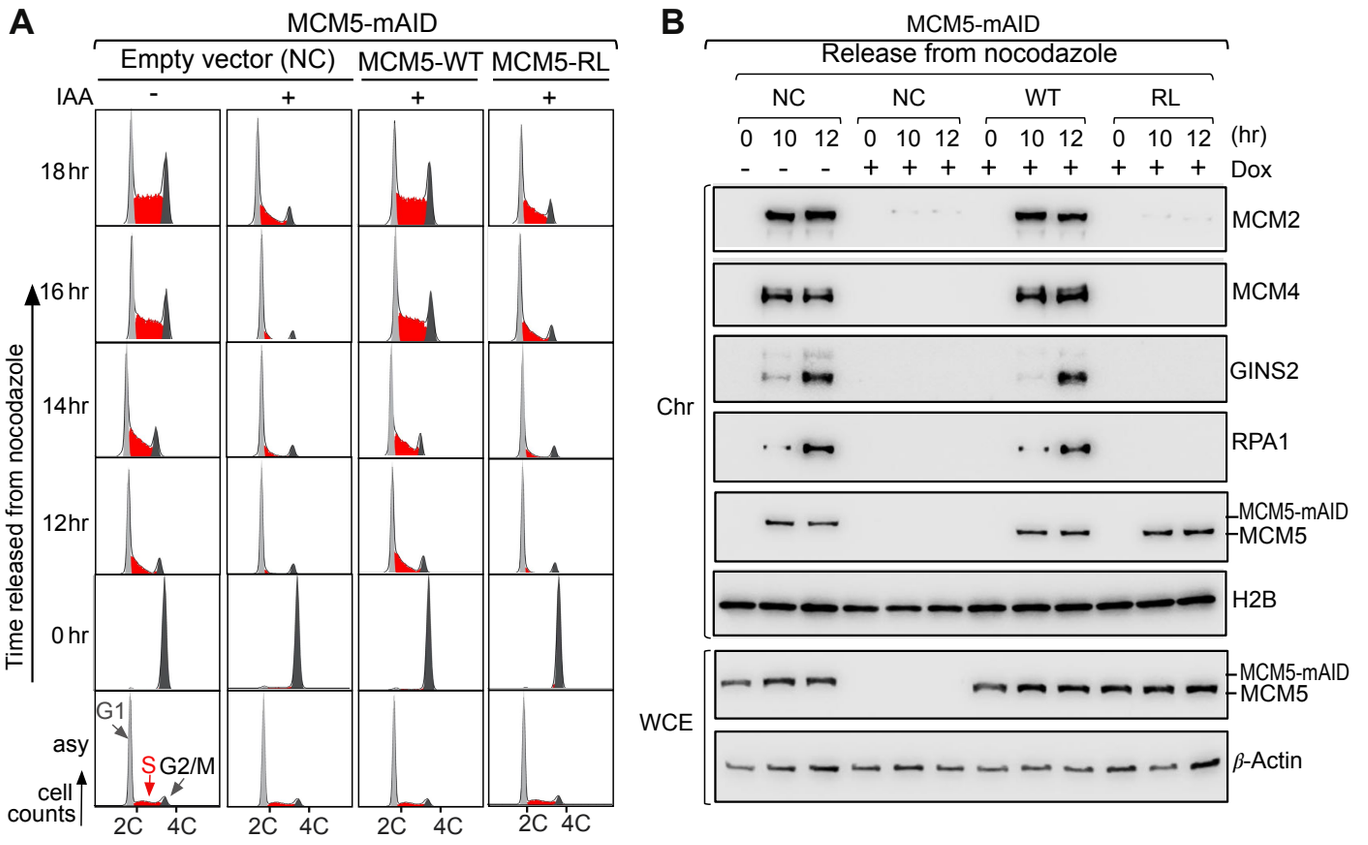

A
MCM5-mAID
Empty vector (NC)
MCM5-WT
MCM5-RL
IAA
-
+
+
+
18 hr
16 hr
14 hr
12 hr
0 hr
Time released from nocodazole
G1
S
G2/M
asy
cell counts
2C
4C
B
MCM5-mAID
Release from nocodazole
NC
NC
WT
RL
0 10 12 0 10 12 0 10 12 0 10 12 (hr)
- - - + + + + + + + + + Dox
Chr
WCE
MCM2
MCM4
GINS2
RPA1
MCM5-mAID
MCM5
H2B
MCM5-mAID
MCM5
β-Actin

Figure 5

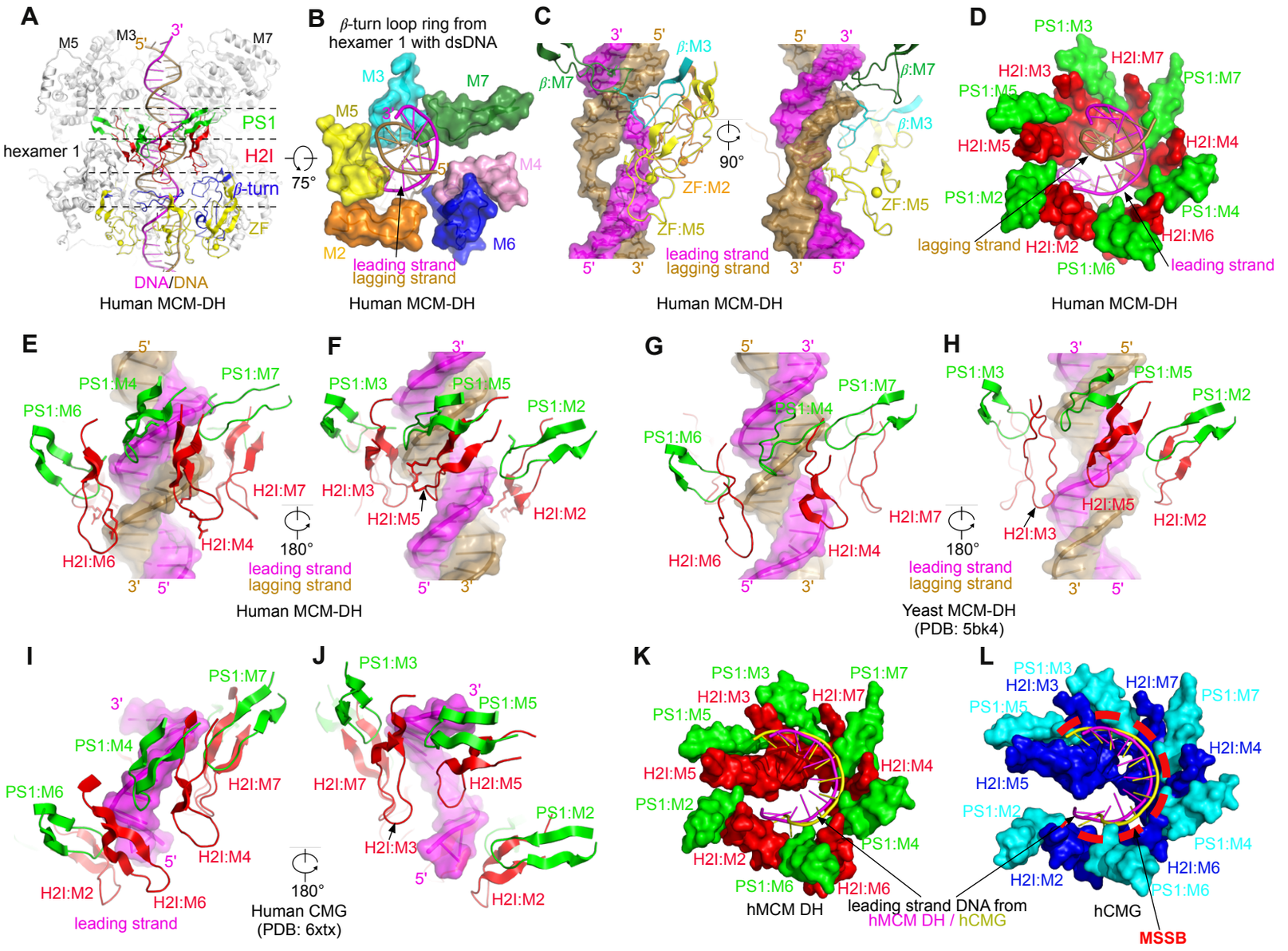

A
M5
M3
5'
3'
M7
PS1
hexamer 1
H2I
β-turn
ZF
DNA/DNA
Human MCM-DH
B
β-turn loop ring from hexamer 1 with dsDNA
M3
M7
M5
M4
M6
M2
leading strand
lagging strand
75°
Human MCM-DH
C
β:M3
β:M7
ZF:M2
ZF:M5
90°
leading strand
lagging strand
Human MCM-DH
D
PS1:M3
H2I:M7
PS1:M7
H2I:M3
PS1:M5
H2I:M5
H2I:M4
PS1:M2
PS1:M4
H2I:M6
lagging strand
H2I:M2
PS1:M6
leading strand
Human MCM-DH
E
PS1:M4
PS1:M7
PS1:M6
H2I:M7
H2I:M4
H2I:M6
180°
leading strand
lagging strand
Human MCM-DH
F
PS1:M3
PS1:M5
PS1:M2
H2I:M3
H2I:M5
H2I:M2
G
PS1:M7
PS1:M4
PS1:M6
H2I:M7
H2I:M6
H2I:M4
180°
leading strand
lagging strand
Yeast MCM-DH
(PDB: 5bk4)
H
PS1:M3
PS1:M5
PS1:M2
H2I:M5
H2I:M3
H2I:M2
I
PS1:M7
PS1:M4
PS1:M6
H2I:M7
H2I:M4
H2I:M2
H2I:M6
leading strand
180°
Human CMG
(PDB: 6xtx)
J
PS1:M3
PS1:M5
H2I:M7
H2I:M5
PS1:M2
H2I:M3
H2I:M2
K
PS1:M3
PS1:M7
H2I:M3
H2I:M7
PS1:M5
H2I:M5
H2I:M4
PS1:M2
PS1:M4
H2I:M2
PS1:M6
H2I:M6
hMCM DH
leading strand DNA from
hMCM DH / hCMG
L
PS1:M3
H2I:M3
H2I:M7
PS1:M7
PS1:M5
H2I:M4
H2I:M5
PS1:M2
PS1:M4
H2I:M6
H2I:M2
PS1:M6
hCMG
MSSB

Figure 6

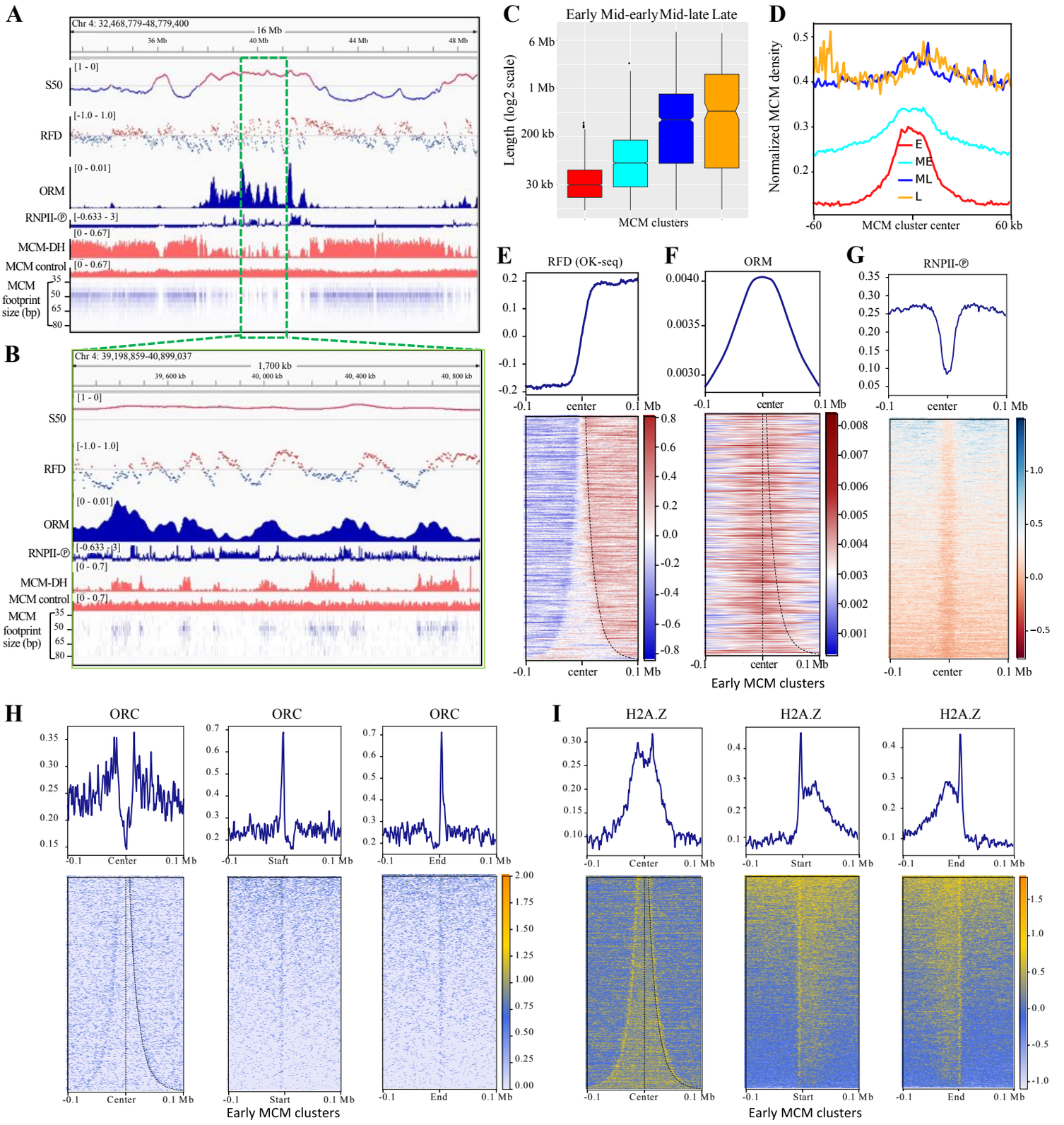

A
Chr 4: 32,468,779-48,779,400
16 Mb
36 Mb
40 Mb
44 Mb
48 Mb
S50
[1 - 0]
RFD
[-1.0 - 1.0]
ORM
[0 - 0.01]
RNPII-Ⓟ
[-0.633 - 3]
MCM-DH
[0 - 0.67]
MCM control
[0 - 0.67]
MCM footprint size (bp)
35
50
65
80
B
Chr 4: 39,198,859-40,899,037
1,700 kb
39, 600 kb
40, 000 kb
40, 400 kb
40, 800 kb
S50
[1 - 0]
RFD
[-1.0 - 1.0]
ORM
[0 - 0.01]
RNPII-Ⓟ
[-0.633 - 3]
MCM-DH
[0 - 0.7]
MCM control
[0 - 0.7]
MCM footprint size (bp)
35
50
65
80
C
Early
Mid-early
Mid-late
Late
6 Mb
1 Mb
200 kb
30 kb
Length (log2 scale)
MCM clusters
D
Normalized MCM density
0.5
0.4
0.3
0.2
E
ME
ML
L
-60
MCM cluster center
60 kb
E
RFD (OK-seq)
0.2
0.1
0.0
-0.1
-0.2
-0.1
center
0.1 Mb
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
F
ORM
0.0040
0.0035
0.0030
-0.1
center
0.1 Mb
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
G
RNPII-Ⓟ
0.35
0.30
0.25
0.20
0.15
0.10
0.05
-0.1
center
0.1 Mb
1.0
0.5
0.0
−0.5
Early MCM clusters
H
ORC
ORC
ORC
0.35
0.30
0.25
0.20
0.15
0.7
0.6
0.5
0.4
0.3
0.2
-0.1
Center
Start
End
0.1 Mb
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Early MCM clusters
I
H2A.Z
H2A.Z
H2A.Z
0.30
0.25
0.20
0.15
0.10
0.4
0.3
0.2
0.1
-0.1
Center
Start
End
0.1 Mb
1.5
1.0
0.5
0.0
-0.5
-1.0
Early MCM clusters

Figure 7

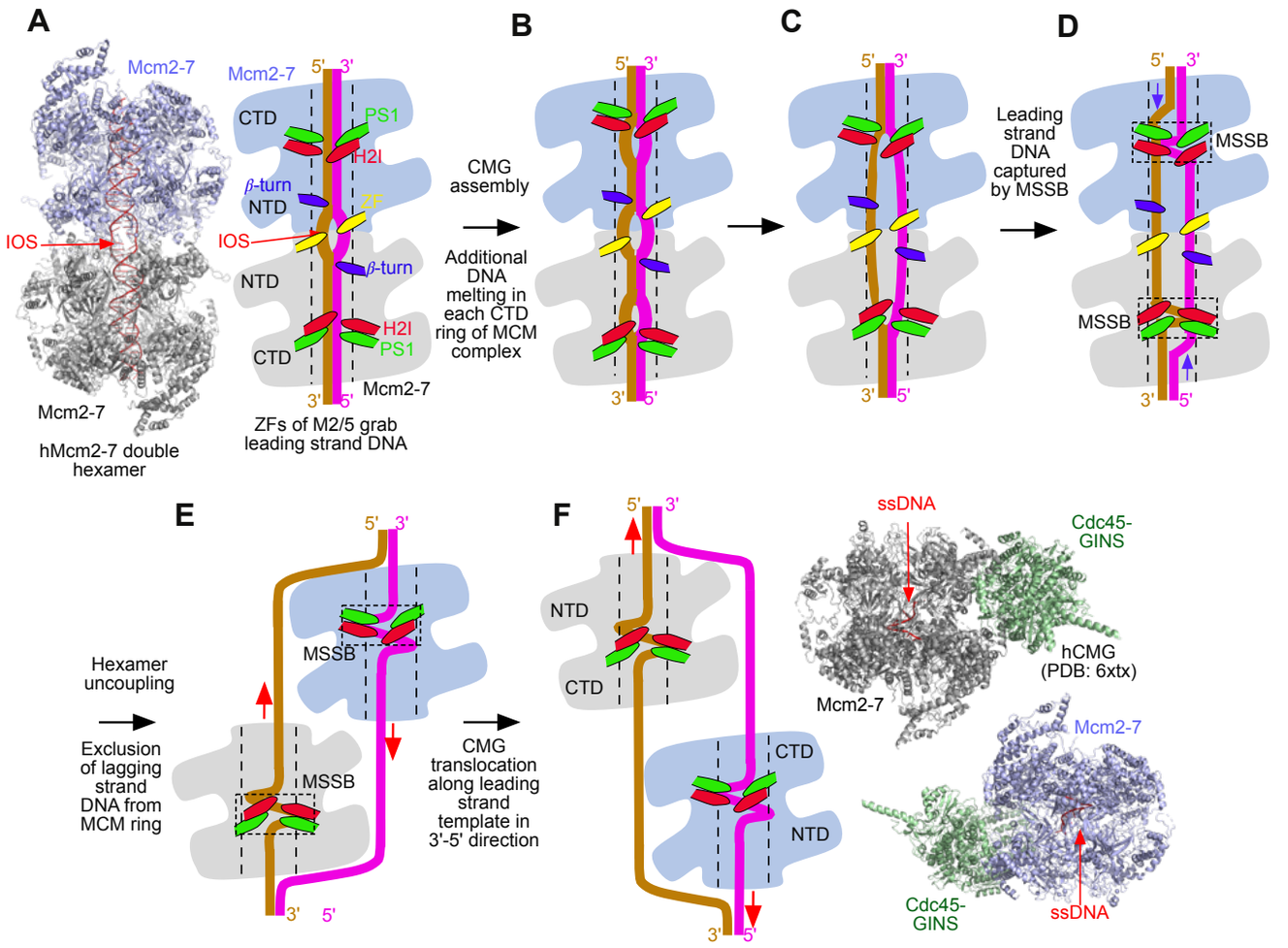

A
Mcm2-7
IOS
Mcm2-7
hMcm2-7 double hexamer
Mcm2-7
5'
3'
CTD
PS1
H2I
β-turn
NTD
ZF
IOS
NTD
β-turn
H2I
PS1
CTD
Mcm2-7
3'
5'
ZFs of M2/5 grab leading strand DNA
CMG assembly
Additional DNA melting in each CTD ring of MCM complex
B
C
Leading strand DNA captured by MSSB
D
MSSB
MSSB
E
Hexamer uncoupling
Exclusion of lagging strand DNA from MCM ring
MSSB
MSSB
CMG translocation along leading strand template in 3'-5' direction
F
NTD
CTD
CTD
NTD
ssDNA
Cdc45-GINS
hCMG (PDB: 6xtx)
Mcm2-7
Mcm2-7
Cdc45-GINS
ssDNA

Figure S1

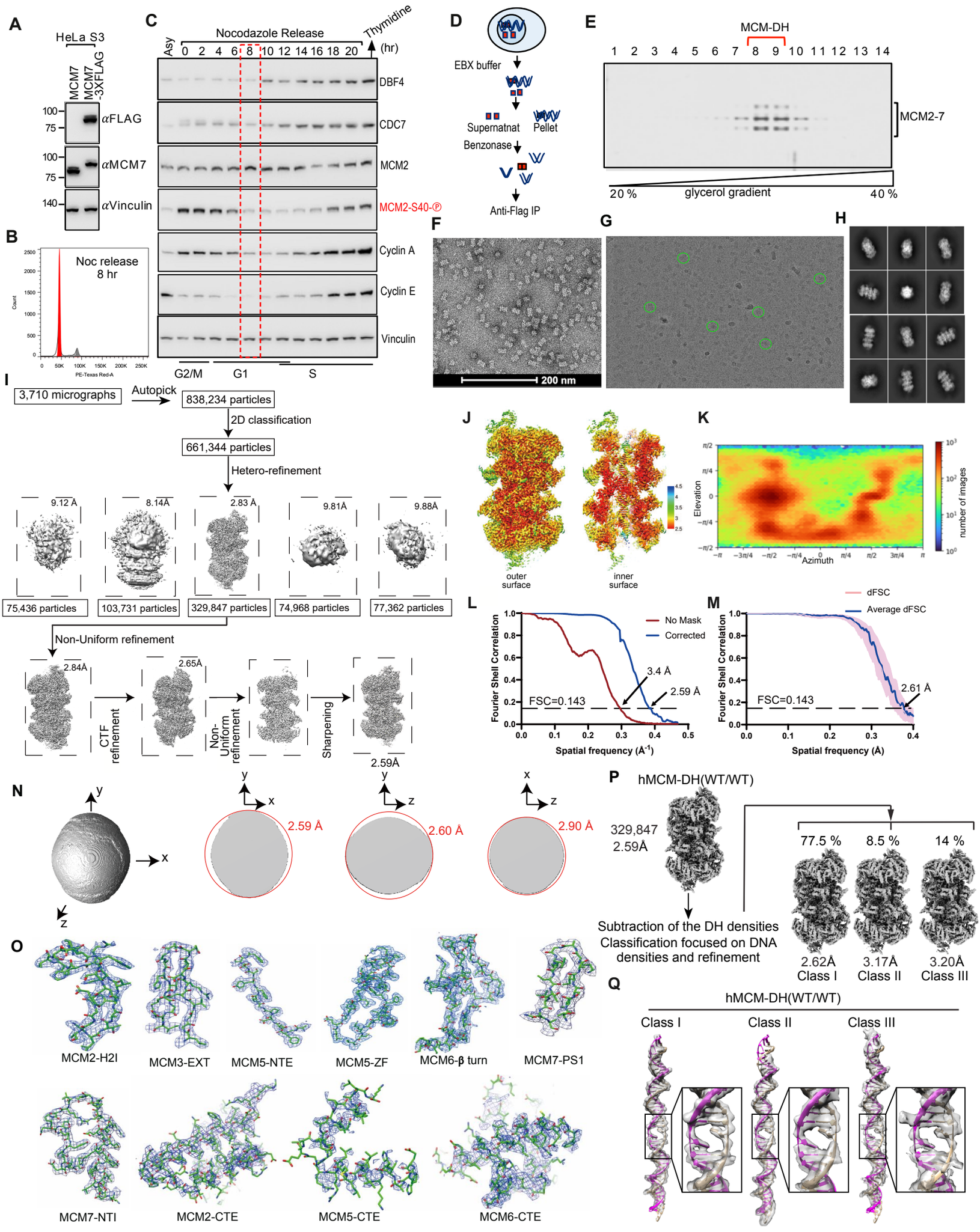

A
HeLa S3
MCM7
MCM7 -3XFLAG
αFLAG
αMCM7
αVinculin
B
Noc release 8 hr
C
Nocodazole Release
Thymidine
Asy
(hr)
DBF4
CDC7
MCM2
MCM2-S40-Ⓟ
Cyclin A
Cyclin E
Vinculin
G2/M
G1
S
D
EBX buffer
Supernatnat
Pellet
Benzonase
Anti-Flag IP
E
MCM-DH
MCM2-7
20 %
glycerol gradient
40 %
F
200 nm
G
H
I
3,710 micrographs
Autopick
838,234 particles
2D classification
661,344 particles
Hetero-refinement
75,436 particles
103,731 particles
329,847 particles
74,968 particles
77,362 particles
Non-Uniform refinement
CTF refinement
Non-Uniform refinement
Sharpening
J
outer surface
inner surface
K
Elevation
Azimuth
number of images
L
No Mask
Corrected
FSC=0.143
3.4 Å
2.59 Å
Fourier Shell Correlation
Spatial frequency (Å-1)
M
dFSC
Average dFSC
FSC=0.143
2.61 Å
Spatial frequency (Å)
N
2.59 Å
2.60 Å
2.90 Å
P
hMCM-DH(WT/WT)
329,847
2.59Å
Subtraction of the DH densities
Classification focused on DNA densities and refinement
77.5 %
8.5 %
14 %
2.62Å
Class I
3.17Å
Class II
3.20Å
Class III
O
MCM2-H2I
MCM3-EXT
MCM5-NTE
MCM5-ZF
MCM6-β turn
MCM7-PS1
MCM7-NTI
MCM2-CTE
MCM5-CTE
MCM6-CTE
Q
hMCM-DH(WT/WT)
Class I
Class II
Class III

Figure S2

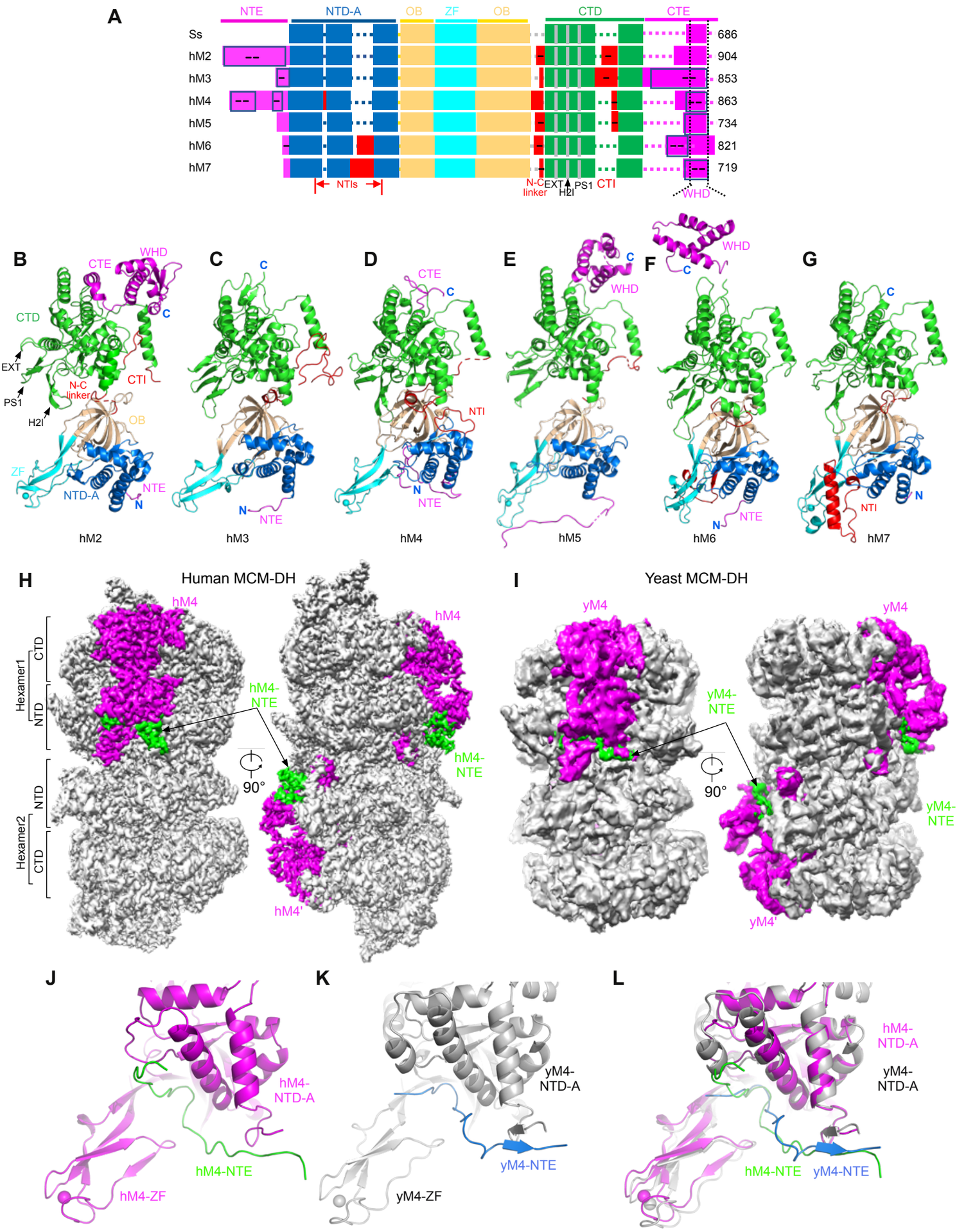
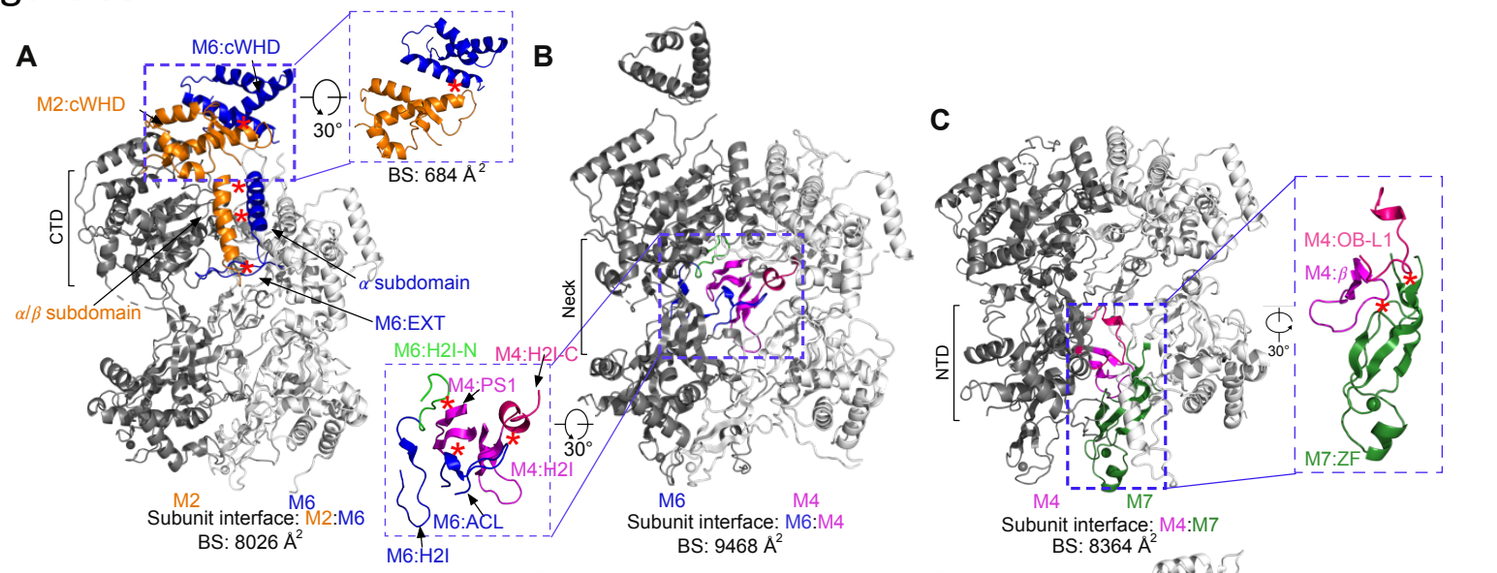
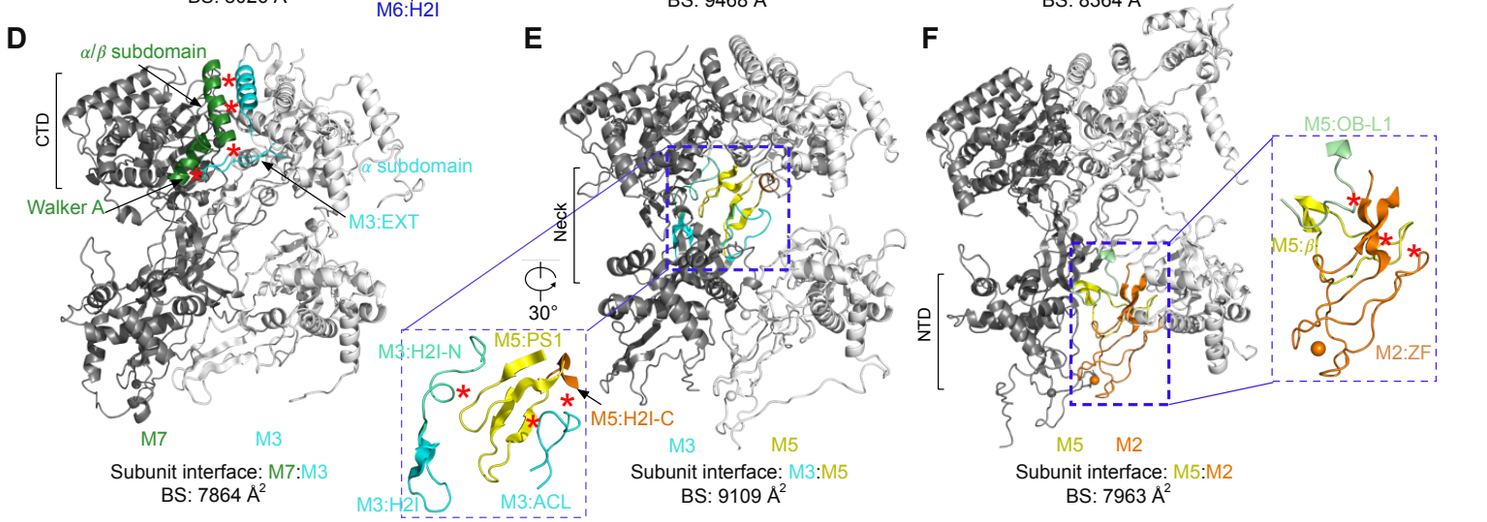
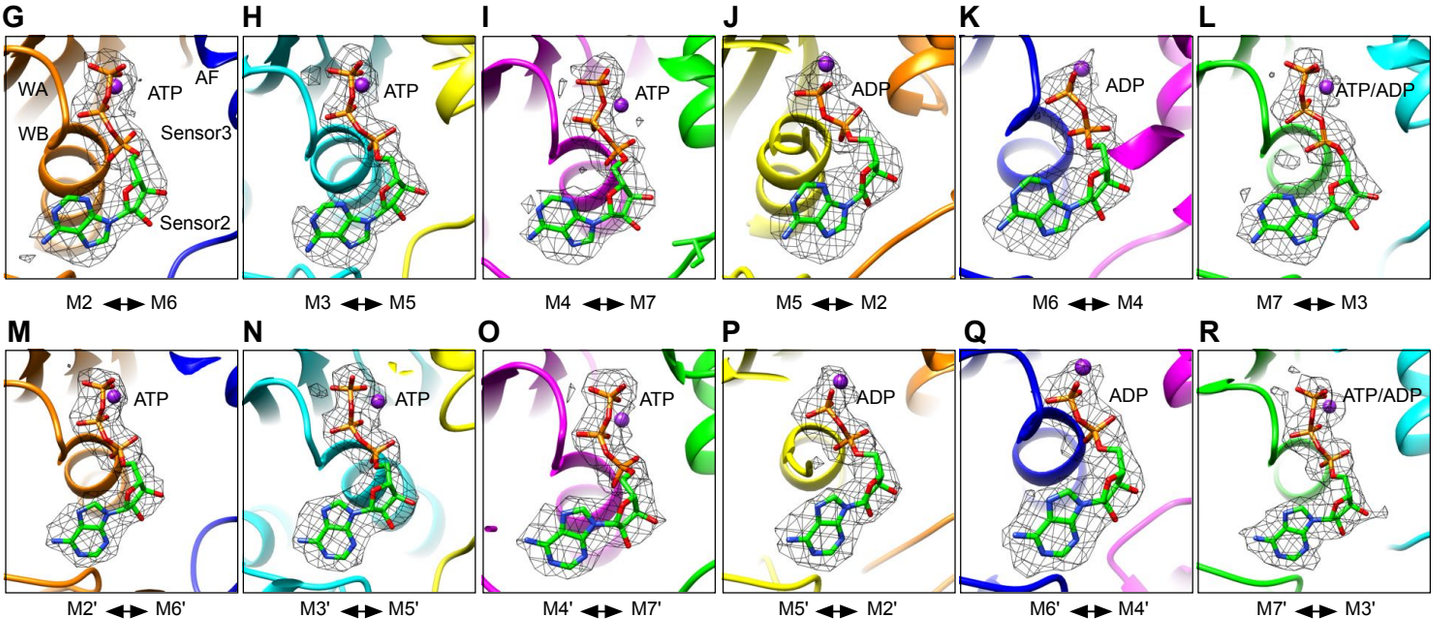

A
NTE
NTD-A
OB
ZF
OB
CTD
CTE
Ss
hM2
hM3
hM4
hM5
hM6
hM7
686
904
853
863
734
821
719
NTIs
N-C linker
EXT
H2I
PS1
CTI
WHD
B
C
D
E
F
G
hM2
hM3
hM4
hM5
hM6
hM7
H
Human MCM-DH
I
Yeast MCM-DH
Hexamer1
Hexamer2
CTD
NTD
hM4
hM4'
hM4-NTE
90°
yM4
yM4'
yM4-NTE
J
K
L
hM4-NTD-A
hM4-ZF
yM4-NTD-A
yM4-ZF
yM4-NTE

Figure S3


A
M6:cWHD
M2:cWHD
30°
BS: 684 Å²
CTD
α subdomain
α/β subdomain
M6:EXT
M6:H2I-N
M4:H2I-C
M4:PS1
M4:H2I
M6:ACL
M6:H2I
M2
M6
Subunit interface: M2:M6
BS: 8026 Å²
B
Neck
30°
M6
M4
Subunit interface: M6:M4
BS: 9468 Å²
C
NTD
30°
M4:OB-L1
M4:β
M7:ZF
M4
M7
Subunit interface: M4:M7
BS: 8364 Å²


D
α/β subdomain
CTD
α subdomain
Walker A
M3:EXT
M7
M3
Subunit interface: M7:M3
BS: 7864 Å²
E
Neck
30°
M3:H2I-N
M5:PS1
M5:H2I-C
M3:H2I
M3:ACL
M3
M5
Subunit interface: M3:M5
BS: 9109 Å²
F
NTD
M5:OB-L1
M5:β
M2:ZF
M5
M2
Subunit interface: M5:M2
BS: 7963 Å²


G
WA
ATP
AF
WB
Sensor3
Sensor2
M2 ◀▶ M6
H
ATP
M3 ◀▶ M5
I
ATP
M4 ◀▶ M7
J
ADP
M5 ◀▶ M2
K
ADP
M6 ◀▶ M4
L
ATP/ADP
M7 ◀▶ M3
M
ATP
M2' ◀▶ M6'
N
ATP
M3' ◀▶ M5'
O
ATP
M4' ◀▶ M7'
P
ADP
M5' ◀▶ M2'
Q
ADP
M6' ◀▶ M4'
R
ATP/ADP
M7' ◀▶ M3'
Map contour: 5σ

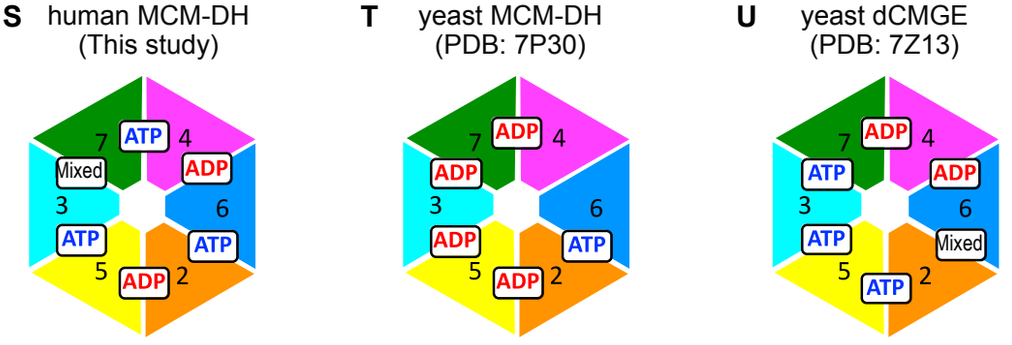

S human MCM-DH (This study)
7
ATP
4
Mixed
ADP
3
6
ATP
ATP
5
ADP
2
T yeast MCM-DH (PDB: 7P30)
7
ADP
4
ADP
3
6
ADP
ATP
5
ADP
2
U yeast dCMGE (PDB: 7Z13)
7
ADP
4
ATP
ADP
3
6
ATP
Mixed
5
ATP
2

Figure S4

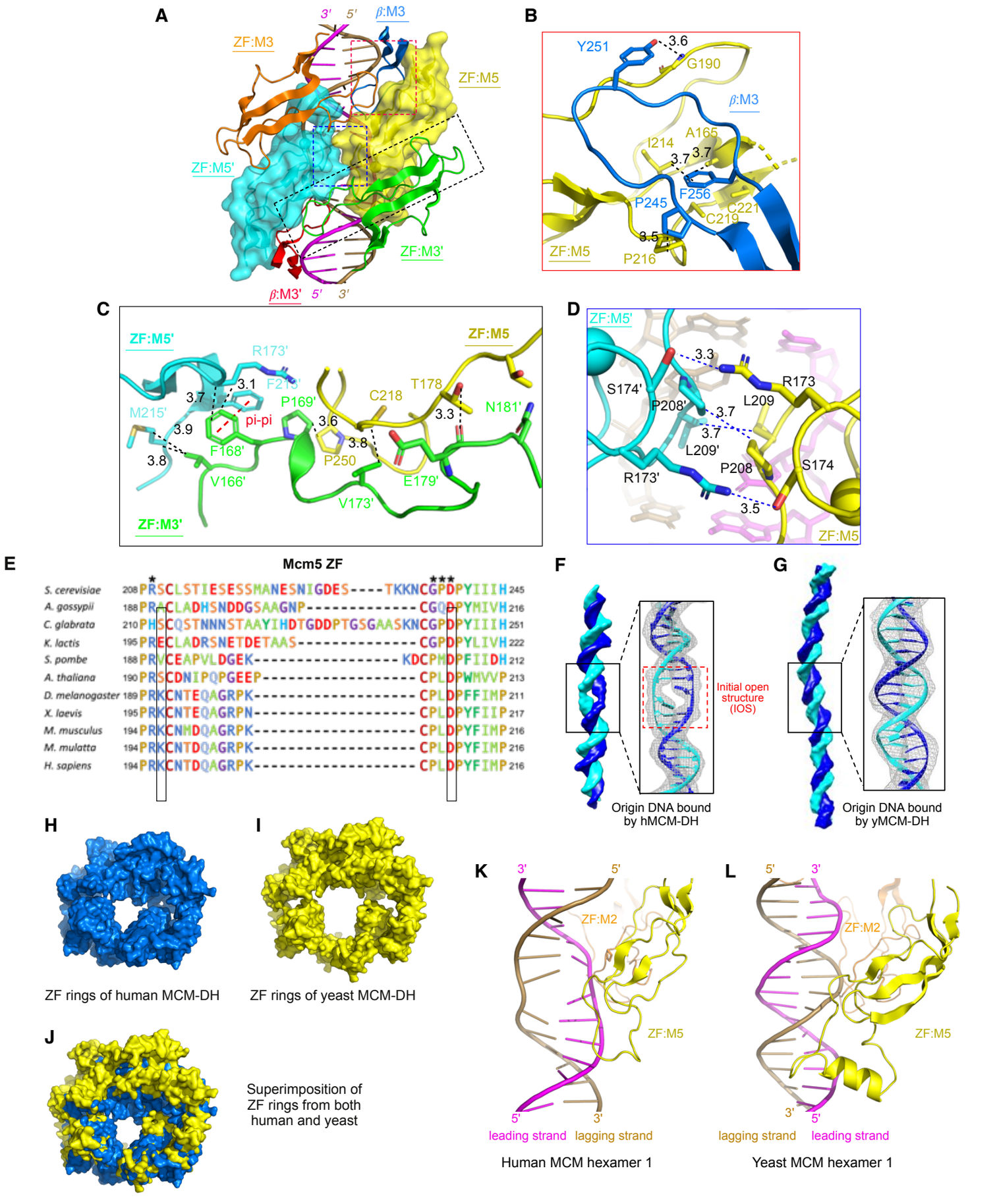

A
3′ 5′ β:M3
ZF:M3
ZF:M5
ZF:M5'
ZF:M3'
β:M3' 5′ 3′
B
Y251
3.6
G190
β:M3
A165
I214
3.7
3.7
F256
P245
C221
C219
3.5
ZF:M5
P216
C
ZF:M5'
ZF:M5
R173'
3.1
F213'
3.7
P169'
C218
T178
N181'
M215'
pi-pi
3.6
3.3
3.9
3.8
F168'
3.8
P250
V166'
E179'
V173'
ZF:M3'
D
ZF:M5'
3.3
S174'
R173
P208'
L209
3.7
3.7
L209'
P208
S174
R173'
3.5
ZF:M5
E
Mcm5 ZF
S. cerevisiae 208 PRSCLSTIESESSMANESNIGDES----TKKNCGPDPYIIIH 245
A. gossypii 188 PRACLADHSNDDGSAAGNP----------CGQDPYMIVH 216
C. glabrata 210 PHSCQSTNNNSTAAYIHDTGDDPTGSGAASKNCGPDPYIIIH 251
K. lactis 195 PRECLADRSNETDETAAS----------CGPDPYLIVH 222
S. pombe 188 PRVCEAPVLDGEK----------KDCPMDPFIIDH 212
A. thaliana 190 PRSCDNIPQPGEEP----------CPLDPWMVVP 213
D. melanogaster 189 PRKCNTEQAGRPK----------CPLDPFFIMP 211
X. laevis 195 PRKCNTEQAGRPN----------CPLDPYFIIP 217
M. musculus 194 PRKCNMDQAGRPK----------CPLDPYFIMP 216
M. mulatta 194 PRKCNTDQAGRPK----------CPLDPYFIMP 216
H. sapiens 194 PRKCNTDQAGRPK----------CPLDPYFIMP 216
F
Initial open structure (IOS)
Origin DNA bound by hMCM-DH
G
Origin DNA bound by yMCM-DH
H
ZF rings of human MCM-DH
I
ZF rings of yeast MCM-DH
J
Superimposition of ZF rings from both human and yeast
K
3′ 5′
ZF:M2
ZF:M5
5′ 3′
leading strand lagging strand
Human MCM hexamer 1
L
5′ 3′
ZF:M2
ZF:M5
3′ 5′
lagging strand leading strand
Yeast MCM hexamer 1

Figure S5

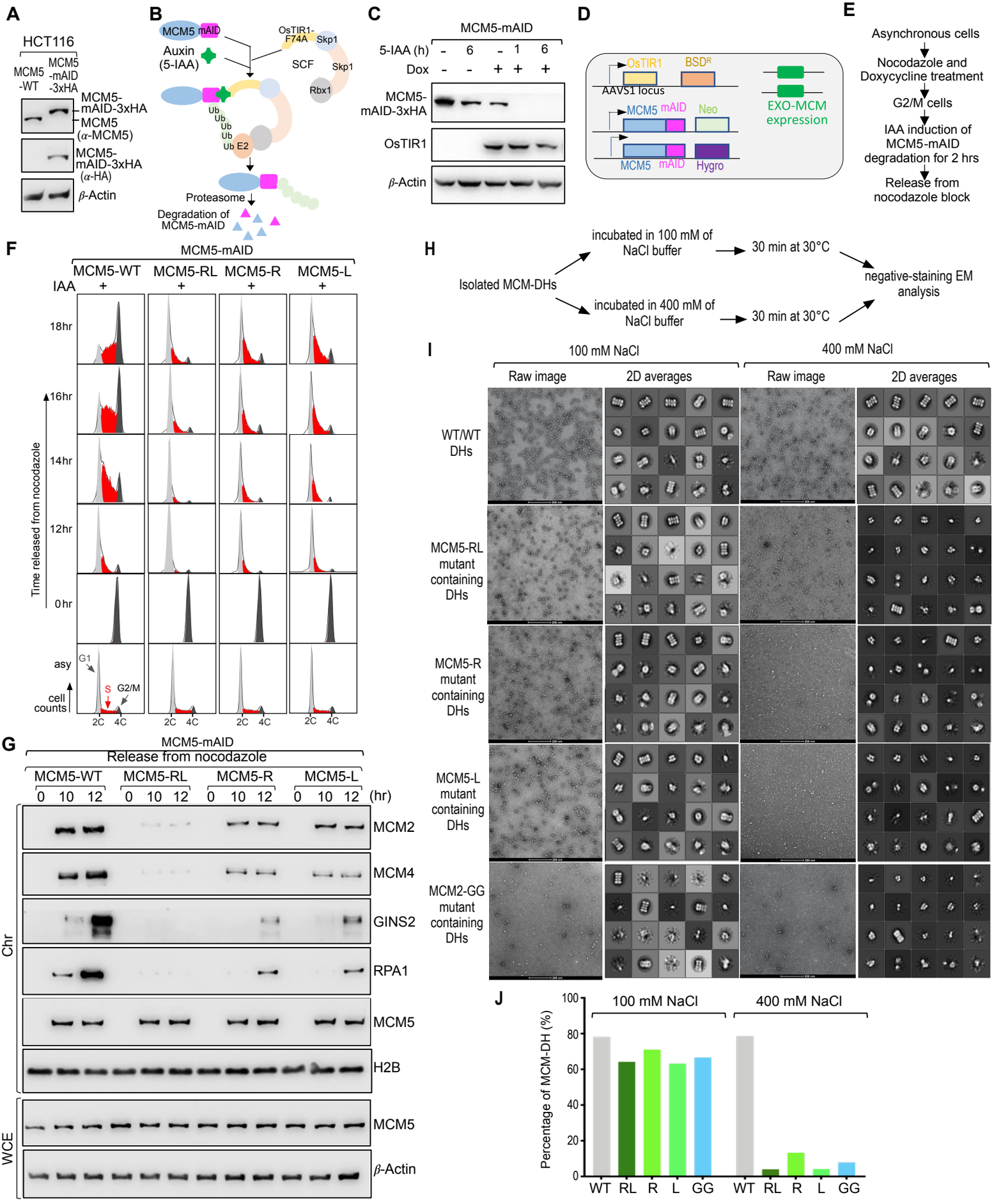

A
HCT116
MCM5 -WT
MCM5 -mAID -3xHA
MCM5-mAID-3xHA
MCM5 (α-MCM5)
MCM5-mAID-3xHA (α-HA)
β-Actin
B
MCM5 mAID
Auxin (5-IAA)
OsTIR1-F74A
Skp1
SCF
Skp1
Rbx1
Ub
Ub E2
Proteasome
Degradation of MCM5-mAID
C
MCM5-mAID
5-IAA (h) - 6 - 1 6
Dox - - + + +
MCM5-mAID-3xHA
OsTIR1
β-Actin
D
OsTIR1
BSDR
AAVS1 locus
MCM5 mAID
Neo
EXO-MCM expression
MCM5 mAID
Hygro
E
Asynchronous cells
Nocodazole and Doxycycline treatment
G2/M cells
IAA induction of MCM5-mAID degradation for 2 hrs
Release from nocodazole block
F
MCM5-mAID
MCM5-WT MCM5-RL MCM5-R MCM5-L
IAA + + + +
18hr
16hr
14hr
12hr
0 hr
Time released from nocodazole
asy
G1
S
G2/M
cell counts
2C 4C
G
MCM5-mAID
Release from nocodazole
MCM5-WT MCM5-RL MCM5-R MCM5-L
0 10 12 0 10 12 0 10 12 0 10 12 (hr)
MCM2
MCM4
GINS2
RPA1
MCM5
H2B
Chr
MCM5
β-Actin
WCE
H
Isolated MCM-DHs
incubated in 100 mM of NaCl buffer
30 min at 30°C
incubated in 400 mM of NaCl buffer
30 min at 30°C
negative-staining EM analysis
I
100 mM NaCl
400 mM NaCl
Raw image
2D averages
Raw image
2D averages
WT/WT DHs
MCM5-RL mutant containing DHs
MCM5-R mutant containing DHs
MCM5-L mutant containing DHs
MCM2-GG mutant containing DHs
J
100 mM NaCl
400 mM NaCl
Percentage of MCM-DH (%)
100
80
60
40
20
0
WT RL R L GG WT RL R L GG

Figure S6

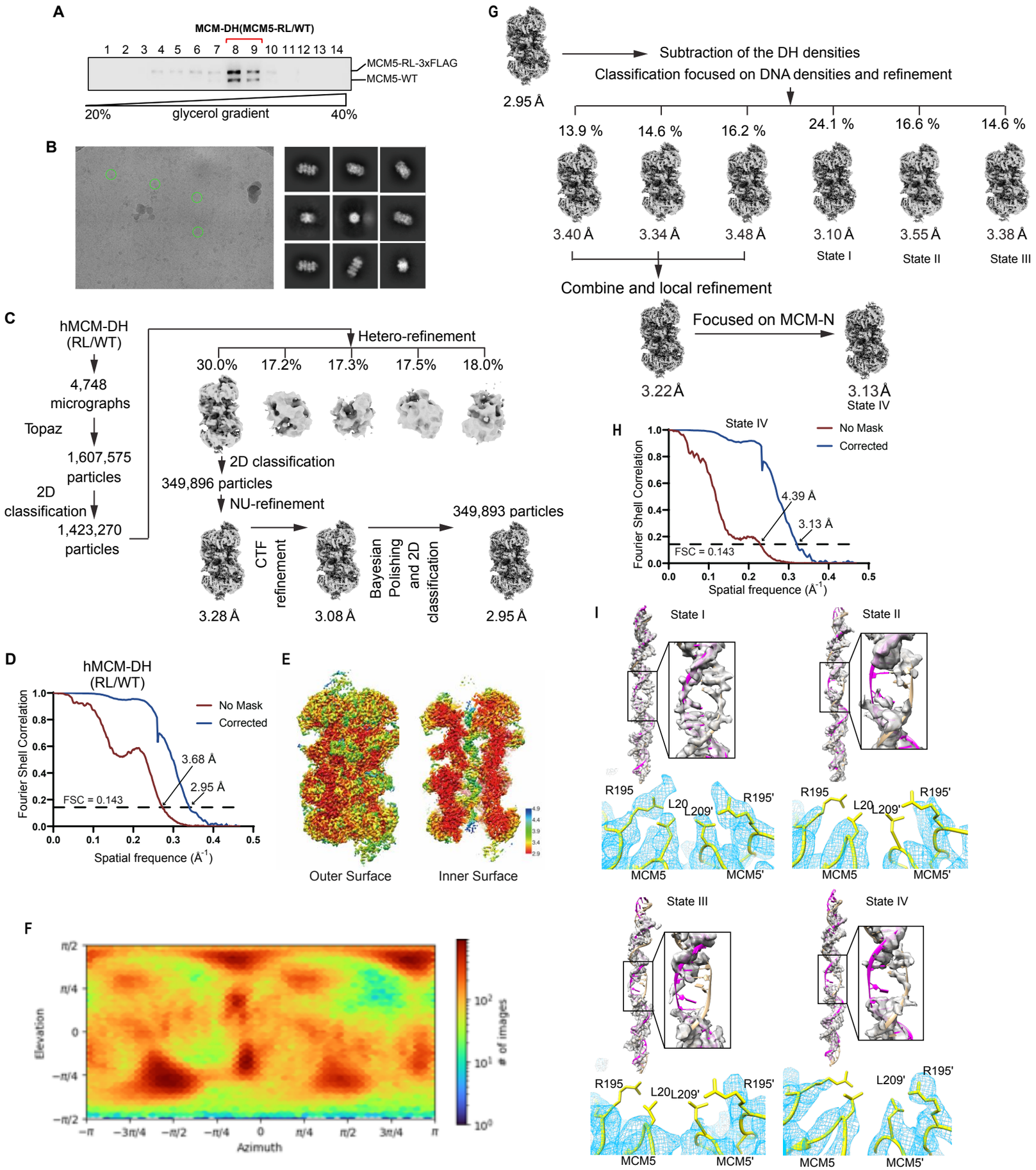
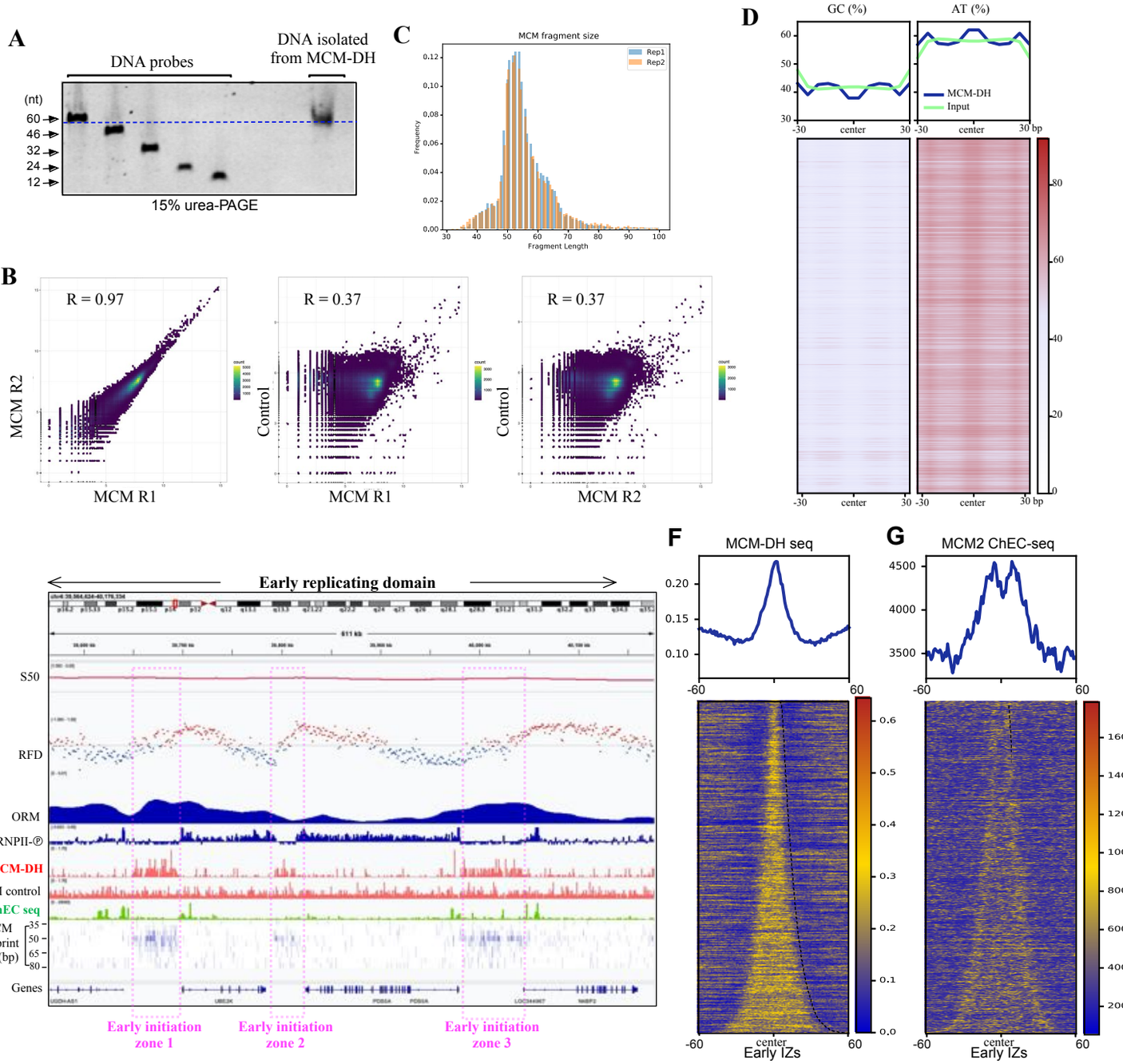

A
MCM-DH(MCM5-RL/WT)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
MCM5-RL-3xFLAG
MCM5-WT
20%
glycerol gradient
40%
B
C
hMCM-DH (RL/WT)
4,748 micrographs
Topaz
1,607,575 particles
2D classification
1,423,270 particles
Hetero-refinement
30.0%
17.2%
17.3%
17.5%
18.0%
2D classification
349,896 particles
NU-refinement
3.28 Å
CTF refinement
3.08 Å
Bayesian Polishing and 2D classification
349,893 particles
2.95 Å
D
hMCM-DH (RL/WT)
No Mask
Corrected
3.68 Å
2.95 Å
FSC = 0.143
Fourier Shell Correlation
Spatial frequence (Å-1)
E
Outer Surface
Inner Surface
F
Elevation
Azimuth
# of images
G
2.95 Å
Subtraction of the DH densities
Classification focused on DNA densities and refinement
13.9 %
14.6 %
16.2 %
24.1 %
16.6 %
14.6 %
3.40 Å
3.34 Å
3.48 Å
3.10 Å
State I
3.55 Å
State II
3.38 Å
State III
Combine and local refinement
Focused on MCM-N
3.22Å
3.13Å
State IV
H
State IV
No Mask
Corrected
4.39 Å
3.13 Å
FSC = 0.143
Fourier Shell Correlation
Spatial frequence (Å-1)
I
State I
State II
State III
State IV
R195
L20
L209'
R195'
MCM5
MCM5'

Figure S7


A
DNA probes
DNA isolated from MCM-DH
(nt)
60
46
32
24
12
15% urea-PAGE
C
MCM fragment size
Rep1
Rep2
Frequency
Fragment Length
D
GC (%)
AT (%)
MCM-DH
Input
center
30 bp
B
R = 0.97
R = 0.37
R = 0.37
MCM R2
MCM R1
Control
MCM R1
Control
MCM R2
E
Early replicating domain
S50
RFD
ORM
RNPII-Ⓟ
MCM-DH
MCM control
MCM2 ChEC seq
MCM footprint size (bp)
Genes
Early initiation zone 1
Early initiation zone 2
Early initiation zone 3
F
MCM-DH seq
G
MCM2 ChEC-seq
center
Early IZs
H
GRO-seq
center
MB
I
All early genes
S
E
80 kb
G
Convergent
Tandem
Divergent
K
Divergent
Tandem
Convergent
Percentage of MCM -DH clusters (%)