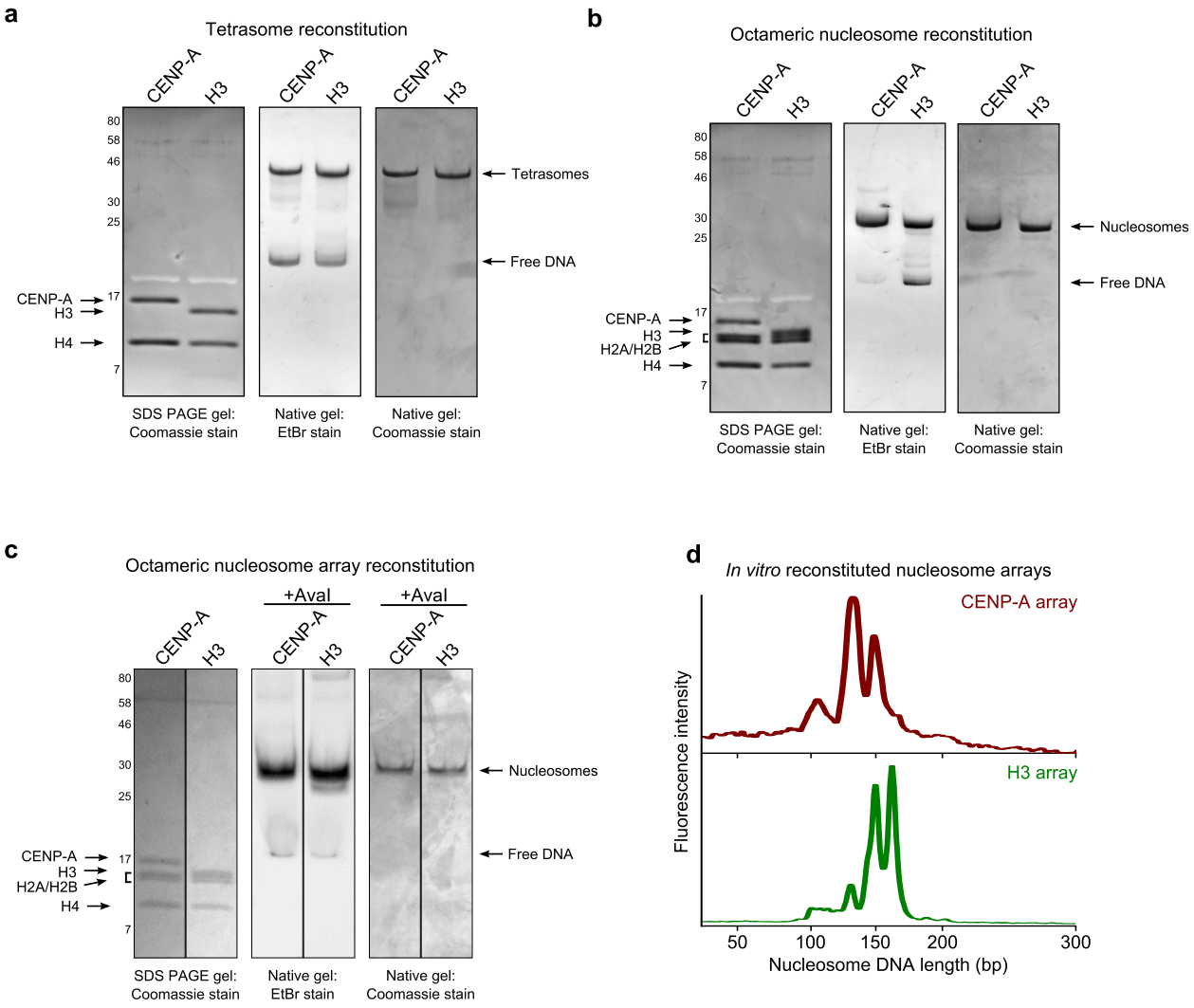
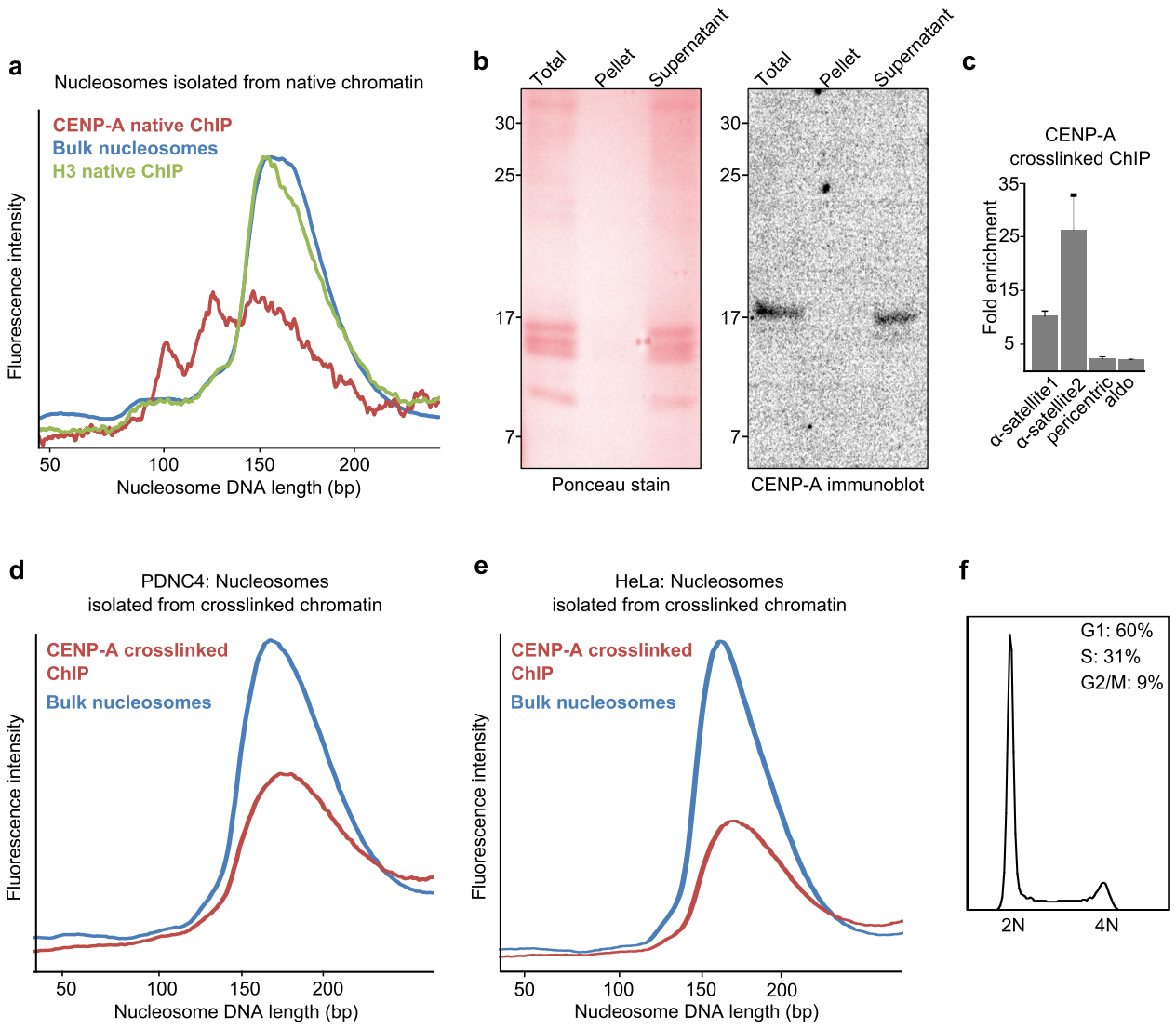
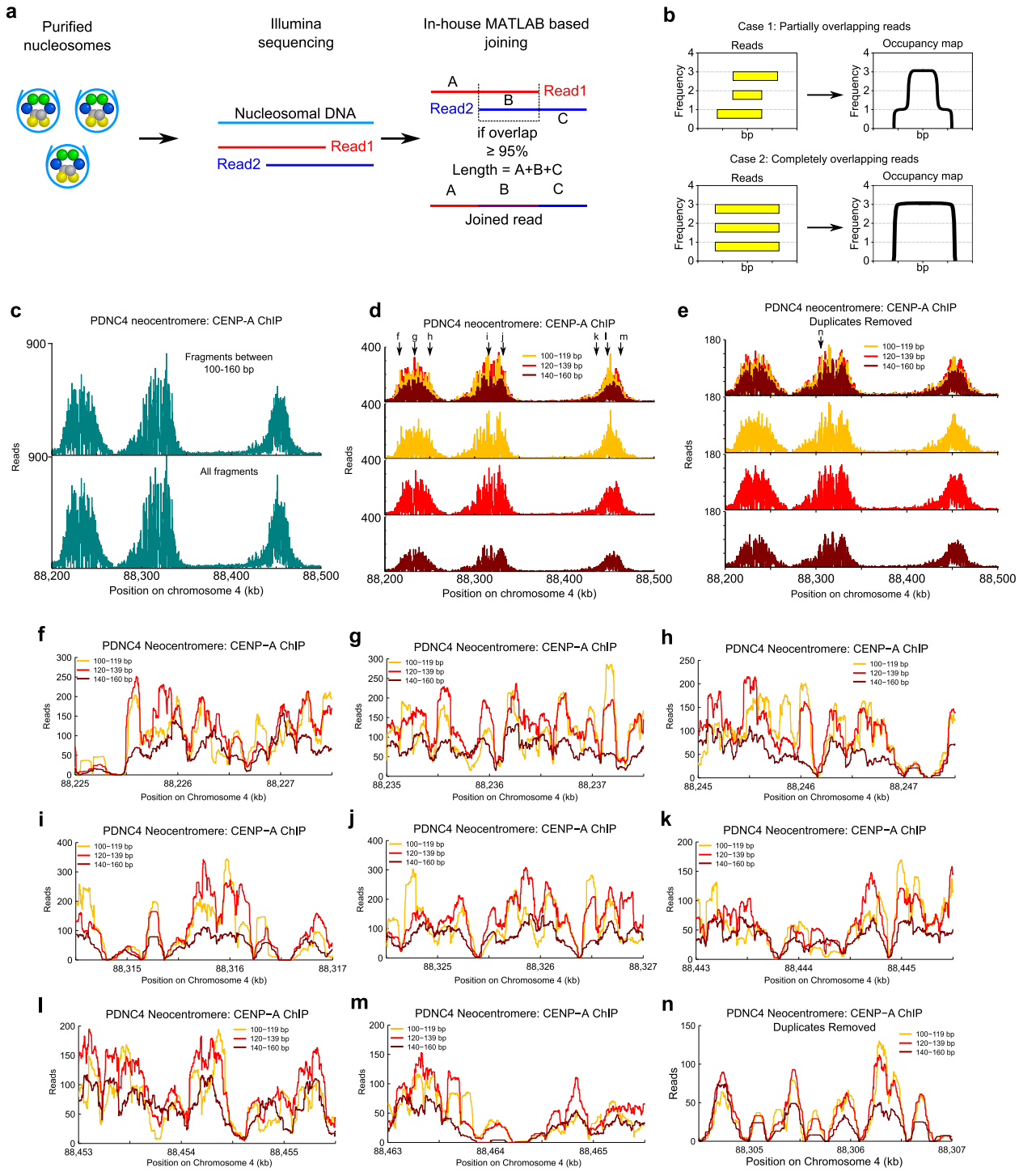


Supplementary Figure 1 | Assembly and MNase digestion of CENP-A- and H3-containing tetrasomes, octameric nucleosomes, and nucleosomal arrays. (**a,b**) Composition of CENP-A- or H3-containing tetrasomes (**a**) or octameric mononucleosomes (**b**) reconstituted on '601' 1x 200 bp DNA. All tetrasomes and octamers appear to have stoichiometric amounts of all histones (SDS PAGE gel on left). CENP-A-containing complexes migrate similarly to the H3-containing complexes on native PAGE gels (middle & right). This indicates that CENP-A assembles into tetrasomes as $(CENP\text{-}A\text{–}H4)_2$ and into octamers as $(CENP\text{-}A\text{–}H4\text{–}H2A\text{–}H2B)_2$ just as H3 assembles as $(H3\text{–}H4)_2$ and $(H3\text{–}H4\text{–}H2A\text{–}H2B)_2$, respectively. (**c**) Composition of CENP-A- or H3-containing nucleosomal arrays reconstituted on '601' 12x 200 bp DNA template. Composition of nucleosomal arrays is determined after cleavage with AvaI between individual 601 monomers to produce mononucleosomes. Nucleosomal arrays appear to have stoichiometric amounts of all histones (SDS PAGE gel on left). CENP-A-containing complexes migrate similarly to the H3-containing complexes on native PAGE gels (middle & right). This indicates that CENP-A assembles into octameric nucleosomal arrays just as H3 does. (**d**) MNase digestion profiles of CENP-A- or H3-containing nucleosomal arrays. Nucleosomal arrays represent stretches of nucleosomes found on chromosomes and produce a nearly identical MNase digestion profile as do octameric mononucleosomes (Fig. 1c).

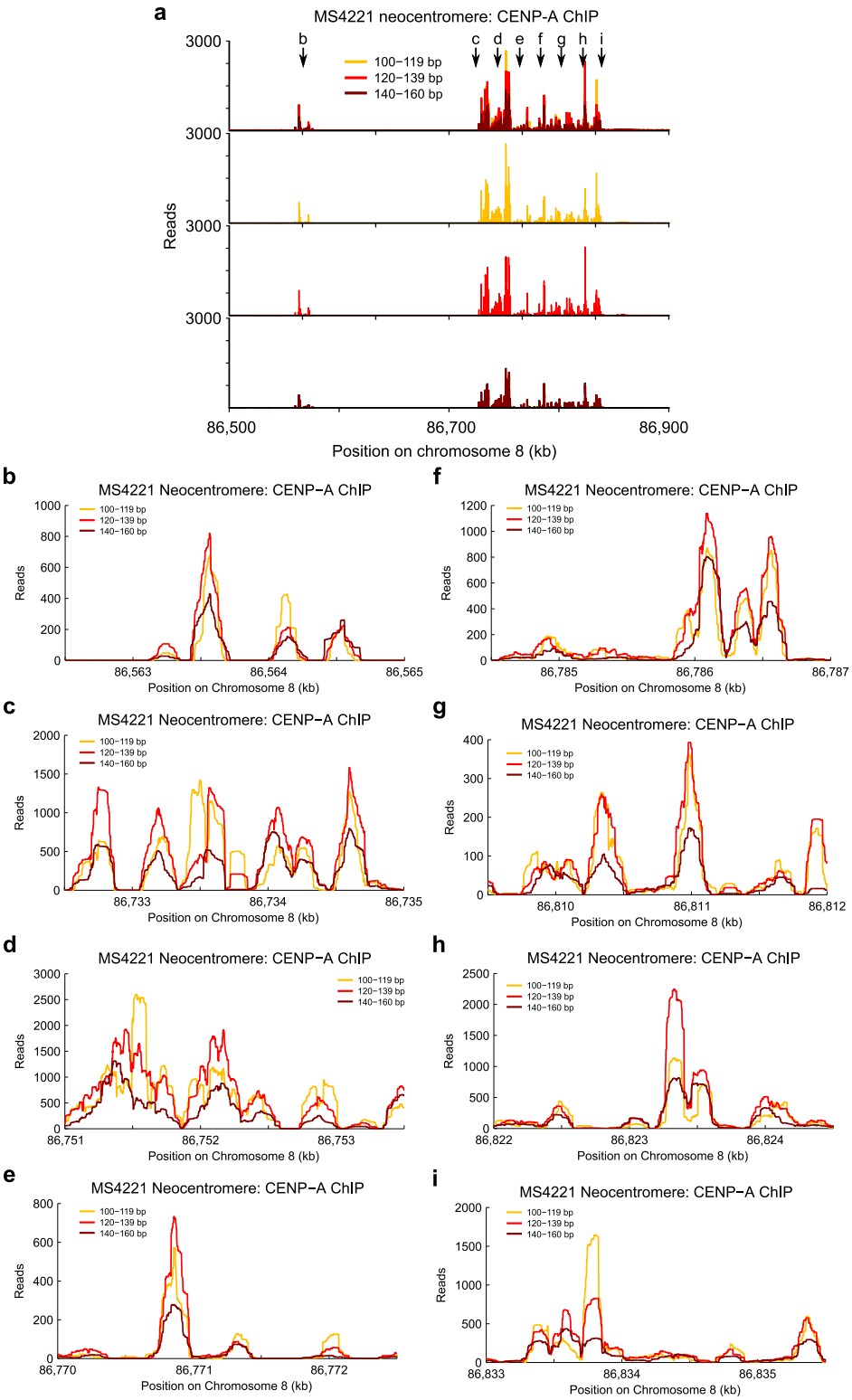


Supplementary Figure 2 | Characterization of native and crosslinked CENP-A ChIP. (**a**) DNA length distributions of MNase digested CENP-A native ChIP, H3 native ChIP and bulk nucleosomes from the same preparation. H3 native ChIP using H3K9me3 antibody yields nearly identical size distribution as bulk nucleosomes (input to the ChIP) suggesting that CENP-A size distribution is not an artifact of immunoprecipitation. Both α-H3K9me3 and α-H3.3 native ChIP produce patterns nearly identical to the bulk (input) nucleosomes. (**b**) Ponceau-stained blot (left) showing that all nucleosomes are released into the supernatant after MNase digestion of crosslinked cells and α-CENP-A blot (right) showing that all of the CENP-A protein is present in the supernatant from which CENP-A is immunoprecipitated. (**c**) qPCR analysis comparing enrichment of CENP-A crosslinked ChIP DNA relative to bulk nucleosome DNA. CENP-A ChIP sequences are specifically enriched for α-satellite regions (α-satellite1, α-satellite2), but not at pericentric or promoter (aldo) regions, as expected. Error bars represent s.e.m. from three independent replicates. (**d**) DNA length distributions of MNase digested CENP-A crosslinked ChIP and bulk nucleosomes from the same preparation from PDNC4 cells. Transiently unwrapped termini are predictably stabilized by standard chromatin crosslinking, yielding CENP-A nucleosomes that protect DNA from MNase to a similarly sized broad range as bulk mononucleosomes from the input material to the CENP-A ChIP. (**e**) DNA length distributions of MNase digested CENP-A crosslinked ChIP and bulk nucleosomes from the same preparation from HeLa cells. (**f**) FACS analysis of propidium iodide stained cycling HeLa cells.

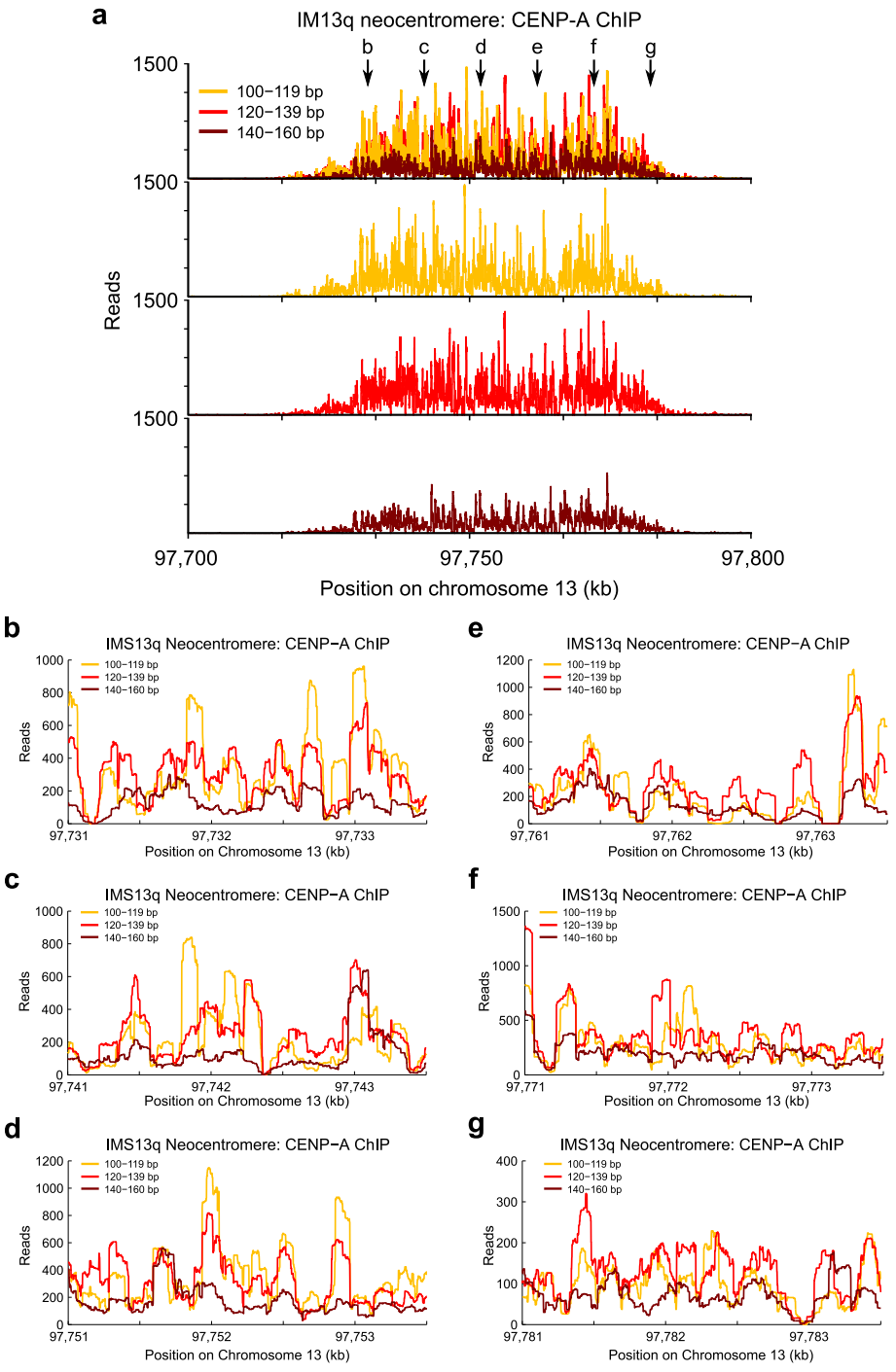


Supplementary Figure 3 | All three size classes of CENP-A nucleosomes map to the same locations at PDNC4 neocentromere. (**a**) Scheme for joining paired-end reads. DNA from native ChIP purification of nucleosomes is extracted and subjected to 100 bp paired-end Illumina sequencing. The first 100 bp from either end of the sequence are sequenced (Read1, Read2). Since mononucleosomes protect less than 200 bp DNA (100 bp + 100 bp), Read1 and Read2 will contain identical overlapping sequence. We used the MATLAB local alignment function to determine the overlapping region (requiring ≥ 95% overlap identity) that was common to both sequence and created the resulting joined sequence. (**b**) Scheme for plotting nucleosome

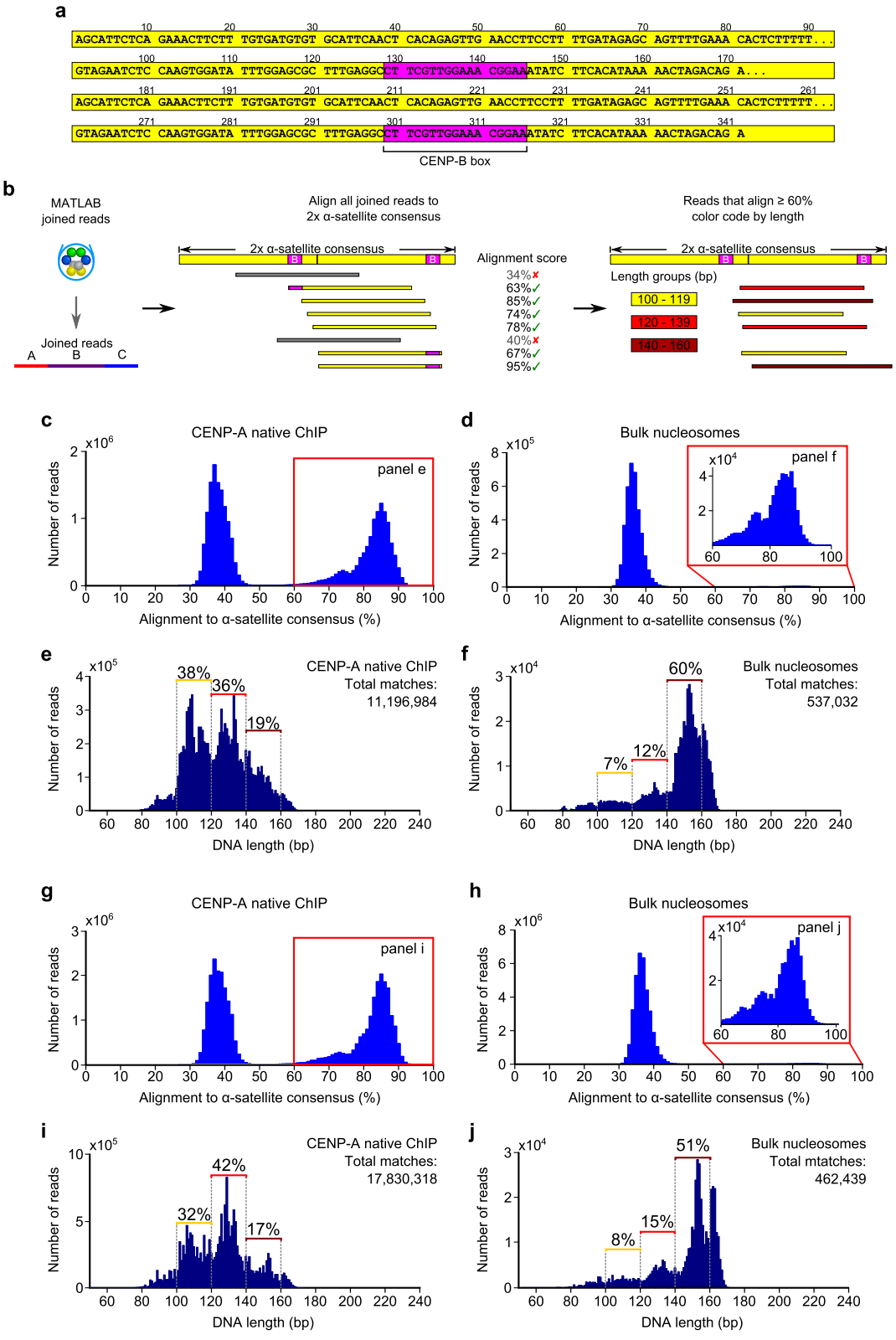
occupancy maps. The number of reads that overlap at each particular base pair along the coordinate are summed and plotted to yield the nucleosome occupancy map (Fig. 3d-i and Supplementary Figs. 3-5). (**c**) Occupancy maps demonstrating that the sum of the three size classes of CENP-A nucleosomes chosen for subsequent analysis represent the vast majority of DNA fragments. (**d**) Occupancy maps for the three different size classes of CENP-A nucleosomes along the length of the PDNC4 neocentromere and at various subsections (2500 bp windows) in (**f-m**). The 2500 bp subsections were chosen at random by choosing 3 regions (each 10 kb apart) within each of the three ~60 kb CENP-A domains. In all cases, p-values < 0.0001 (Supplementary Table 2). (**e**) Occupancy maps for the three different size classes of CENP-A nucleosomes after duplicate reads (reads that are identical in length and sequence) were removed along the length of the entire PDNC4 neocentromere and in a 2500 bp window (**n**) (see Fig. 3e for comparison). Nucleosome occupancy maps retain very similar patterns before and after duplicate removal and p-values remain < 0.0001 (Supplementary Table 2).



Supplementary Figure 4 | All three size classes of CENP-A nucleosomes map to the same locations at MS4221 neocentromere. (**a-i**) Occupancy maps for the three different size classes of CENP-A nucleosomes along the length of the MS4221 neocentromere (**a**) and at various subsections (2500 bp windows) in (**b-i**). Location of the 2500 bp windows are indicated with black arrows in (**a**). The 2500 bp subsections were chosen at random (one region within each prominent CENP-A peak from (**a**)). In all cases, p-values < 0.0001 (Supplementary Table 2).

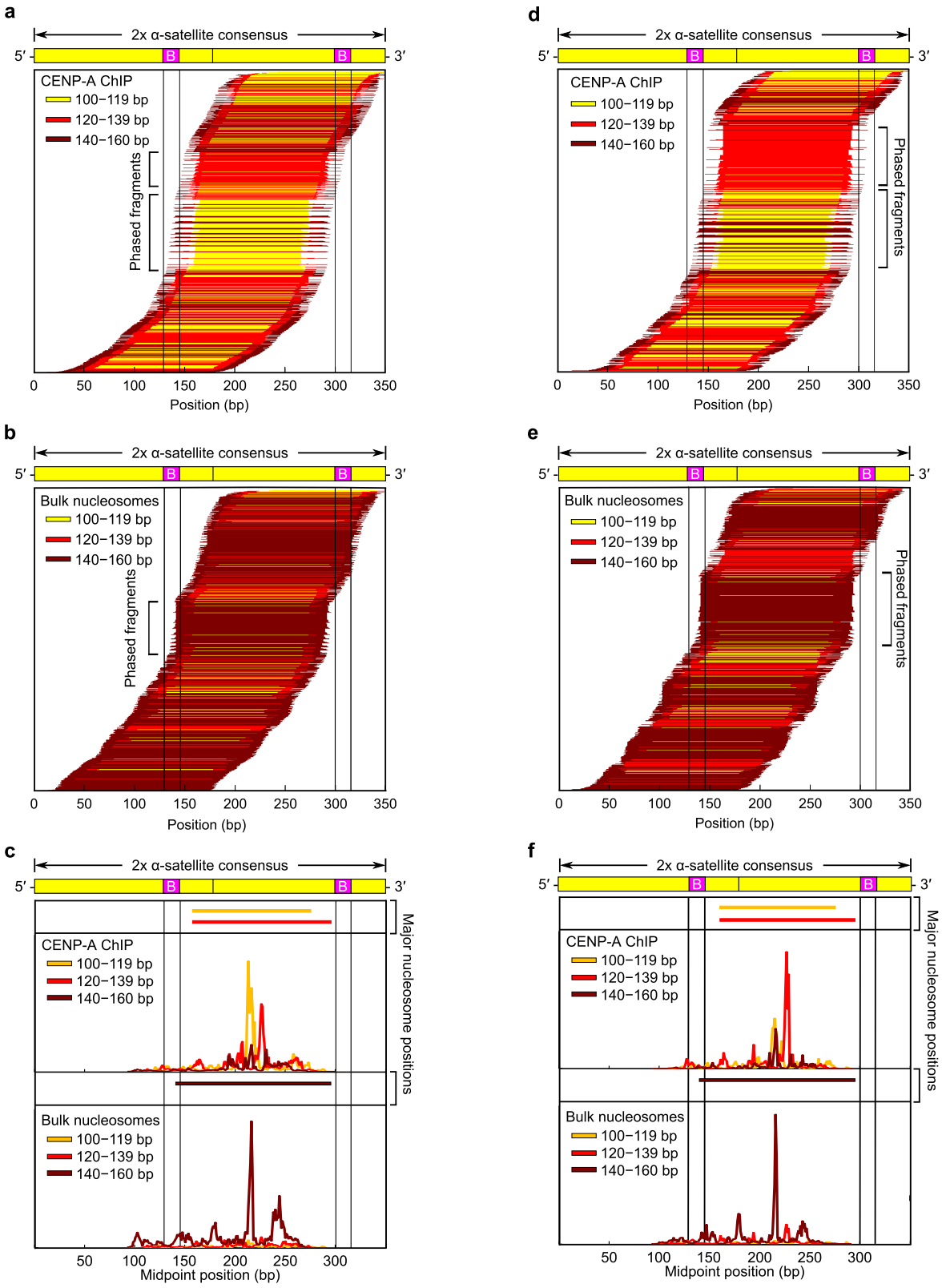


Supplementary Figure 5 | All three size classes of CENP-A nucleosomes map to the same locations at IMS13q neocentromere. (**a-g**) Occupancy maps for the three different size classes of CENP-A nucleosomes along the length of the entire IMS13q neocentromere (**a**) and at various subsections (2500 bp windows) in (**b-g**). Location of the 2500 bp windows are indicated with black arrows in (**a**). The 2500 bp subsections were chosen at random (each 10 kb apart). In all cases, p-values < 0.0001 (Supplementary Table 2).



Supplementary Figure 6 | CENP-A nucleosomes at normal centromeres in MS4221 and IMS13q cells have a tripartite distribution of sizes. (**a**) Sequence of the dimer α-satellite consensus is highlighted in yellow with the location of the CENP-B boxes highlighted in magenta. (**b**) All joined reads (as described in

Supplementary Fig. 3a) are aligned along the dimer α-satellite consensus sequence and only those sequences which align with ≥ 60% identity to the consensus are color-coded based on length and are plotted along the length of the consensus sequence. (**c,d**) Alignment of MS4221 CENP-A (**c**) or bulk nucleosome (**d**) fragments to the α-satellite consensus sequence. (**e,f**) DNA length distribution of all CENP-A (**e**) or bulk nucleosome (**f**) fragments aligning to the α-satellite consensus sequence with ≥ 60% identity. (**g,h**) Alignment of IMS13q CENP-A (**g**) or bulk nucleosome (**h**) fragments to the α-satellite consensus sequence. (**i,j**) DNA length distribution of all CENP-A (**i**) or bulk nucleosome (**j**) fragments aligning to the α-satellite consensus sequence with ≥ 60% identity.



Supplementary Figure 7 | Terminally unwrapped CENP-A nucleosomes and their conventional counterparts with wrapped termini are similarly phased at normal centromeres in MS4221 and IMS13q cells. (**a**,**b**) The position of each individual MS4221 derived CENP-A (**a**) or bulk nucleosome (**b**) fragment

along the dimer α-satellite consensus sequence. (**c**) Frequency of midpoint position along the dimer α-satellite consensus sequence for CENP-A or bulk nucleosome fragments derived from MS4221 cells. (**d,e**) The position of each individual IMS13q derived CENP-A (**d**) or bulk nucleosome (**e**) fragment along the dimer α-satellite consensus sequence. (**f**) Frequency of midpoint position along the dimer α-satellite consensus sequence for CENP-A or bulk nucleosome fragments derived from IMS13q cells. Each horizontal line represents a single nucleosomal fragment that is color-coded based on length in (**a**), (**b**), (**d**), and (**e**). The most prominent nucleosome positions are displayed with horizontal lines in (**c**) and (**f**). Solid vertical lines indicate the location of the 17 bp CENP-B box in (**a-f**).

Supplementary Table 1. CENP-A joined reads are specifically enriched at neocentromeres

	Number of joined reads at the neocentromere		
	PDNC4	*MS4221*	*IMS13q*
CENP-A native ChIP	326,764	340,371	338,724
Bulk nucleosomes	4407	6187	1216
H3 native ChIP	2732	5427	1501

Supplementary Table 2. Pearson correlation coefficients for nucleosome occupancy maps at neocentromeres demonstrate that all size classes colocalize.

Genomic Region (bp)	Correlation between 100-119 bp 120-139 bp (p-value)	Correlation between 100-119 bp 140-160 bp (p-value)	Correlation between 120-139 bp 140-160 bp (p-value)	Correlation between 100-119bp Random (p-value)	Correlation between 120-139 bp Random (p-value)	Correlation between 140-160bp Random (p-value)	Figure Reference
PDNC4 (chr4)							
88200000-88500000	0.8757 (<0.0001)	0.7881 (<0.0001)	0.9115 (<0.0001)	-0.0018 (0.3271)	-0.0014 (0.4526)	-0.0009 (0.6392)	Fig. 3d; Fig. S3d
88225000-88227500	0.8151 (<0.0001)	0.6315 (<0.0001)	0.764 (<0.0001)	-0.0021 (0.9175)	-0.001 (0.9614)	0.0059 (0.7667)	Fig. S3f
88235000-88237500	0.7297 (<0.0001)	0.3497 (<0.0001)	0.7068 (<0.0001)	-0.0119 (0.5523)	-0.0018(0.9283)	0.0312 (0.1187)	Fig. S3g
88245000-88247500	0.7587 (<0.0001)	0.5566 (<0.0001)	0.8261 (<0.0001)	0.0151 (0.4506)	0.0209 (0.2969)	0.0243 (0.2238)	Fig. S3h
88304500-88307000	0.8373 (<0.0001)	0.6790 (<0.0001)	0.9068 (<0.0001)	-0.0045 (0.8203)	0.0041 (0.8373)	-0.0032 (0.8743)	Fig. 3e
88314500-88317000	0.7924 (<0.0001)	0.8054 (<0.0001)	0.8802 (<0.0001)	-0.0191 (0.3406)	0.0036 (0.8589)	0.0034 (0.8632)	Fig. S3i
88324500-88327000	0.5328 (<0.0001)	0.2900 (<0.0001)	0.7275 (<0.0001)	0.0192 (0.3381)	0.0013 (0.9500)	-0.0056 (0.7799)	Fig. S3j
88443000-88445500	0.7070 (<0.0001)	0.6132 (<0.0001)	0.7852 (<0.0001)	-0.0228 (0.2537)	-0.0234 (0.2412)	-0.0182 (0.3632)	Fig. S3k
88453000-88453500	0.7669 (<0.0001)	0.4449 (<0.0001)	0.8301 (<0.0001)	-0.0025 (0.8989)	0.0083 (0.6771)	0.0169 (0.3969)	Fig. S3l
88463000-88463500	0.8209 (<0.0001)	0.6905 (<0.0001)	0.8540 (<0.0001)	0.0097 (0.6279)	0.0022 (0.9139)	0.0023 (0.9094)	Fig. S3m
IMS13q (chr13)							
97700000-97800000	0.7812 (<0.0001)	0.6625 (<0.0001)	0.8017 (<0.0001)	0.0009 (0.6723)	0.0007 (0.6572)	0.0002 (0.8215)	Fig. 3h, Fig. S5a
97731000-97733500	0.7390 (<0.0001)	0.3456 (<0.0001)	0.5265 (<0.0001)	0.0194 (0.3326)	0.0142 (0.4765)	-0.0038 (0.8509)	Fig. S5b
97741000-97743500	0.4935 (<0.0001)	0.2246 (<0.0001)	0.6353 (<0.0001)	-0.0106 (0.5959)	-0.0143 (0.4755)	-0.0249 (0.2132)	Fig. S5c
97751000-97753500	0.7513 (<0.0001)	0.3607 (<0.0001)	0.4924 (<0.0001)	-0.0044 (0.8275)	-0.0028 (0.8870)	0.0029 (0.8842)	Fig. S5d
97761000-97763500	0.7866 (<0.0001)	0.6493 (<0.0001)	0.6497 (<0.0001)	0.0089 (0.6576)	0.0100 (0.6178)	0.0050 (0.8026)	Fig. 3i; Fig. S5e
97771000-97773500	0.6070 (<0.0001)	0.6007 (<0.0001)	0.7132 (<0.0001)	0.0137 (0.4945)	0.0104 (0.6030)	-0.0050 (0.8039)	Fig. S5f
97781000-97783500	0.6956 (<0.0001)	0.1633 (<0.0001)	0.3336 (<0.0001)	-0.0134 (0.5042)	-0.0133 (0.5061)	0.0281 (0.1599)	Fig. S5g
MS4221 (chr8)							
86500000-86900000	0.8661 (<0.0001)	0.8009 (<0.0001)	0.9342 (<0.0001)	0.0006 (0.7210)	0.0009 (0.5896)	0.0009 (0.5489)	Fig. 3f, Fig. S4a
86562500-86565000	0.8856 (<0.0001)	0.8272 (<0.0001)	0.9237 (<0.0001)	-0.0267 (0.1825)	-0.0219 (0.2747)	-0.0296 (0.1384)	Fig. S4b
86732500-86735000	0.7114 (<0.0001)	0.5724 (<0.0001)	0.9179 (<0.0001)	-0.0082 (0.6830)	-0.0069 (0.7290)	-0.0138 (0.4907)	Fig. S4c
86751000-86753500	0.7259 (<0.0001)	0.5494 (<0.0001)	0.8916 (<0.0001)	-0.0035 (0.8615)	-0.0188 (0.3742)	-0.0275 (0.1690)	Fig. S4d
86770000-86772500	0.9456 (<0.0001)	0.9481 (<0.0001)	0.9726 (<0.0001)	0.0257 (0.1982)	0.0290 (0.1474)	0.0290 (0.1471)	Fig. S4e
86784500-86787000	0.9596 (<0.0001)	0.8767 (<0.0001)	0.9532 (<0.0001)	-0.0162 (0.4180)	-0.0192 (0.3362)	-0.0077 (0.7010)	Fig. S4f
86809500-86812000	0.9416 (<0.0001)	0.8276 (<0.0001)	0.8535 (<0.0001)	-0.0207 (0.3010)	-0.0176 (0.3802)	-0.0018 (0.9287)	Fig. 3g; Fig. S4g
86822000-86824500	0.9158 (<0.0001)	0.8387 (<0.0001)	0.8975 (<0.0001)	-0.0175 (0.3812)	-0.0171 (0.3917)	-0.0250 (0.2121)	Fig. S4h
86833000-86835500	0.8543 (<0.0001)	0.6433 (<0.0001)	0.9194 (<0.0001)	-0.0116 (0.5612)	-0.0026 (0.8983)	-0.0039 (0.8464)	Fig. S4i
PDNC4 (chr4) (Duplicates removed)							
88200000-88500000	0.9178 (<0.0001)	0.8448 (<0.0001)	0.9410 (<0.0001)	0.0015 (0.4261)	0.0013 (0.4820)	0.0011 (0.5366)	Fig. S3e
88304500-88307000	0.9149 (<0.0001)	0.7111 (<0.0001)	0.8877 (<0.0001)	0.0172 (0.3898)	0.0047 (0.8142)	-0.0104 (0.6020)	Fig. S3n

NOTE: See supplementary figure legends and methods for information about how sub-regions were selected for each cell line.