# DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome-wide

SUPPLEMENTARY INFORMATION

Nicolas Altemose[1,2*Π], Annie Maslan[1,2,3Π], Owen K. Smith[4,5Π], Kousik Sundararajan[4Π], Rachel R. Brown[4], Reet Mishra[1], Angela M. Detweiler[6], Norma Neff[6], Karen H. Miga[7,8], Aaron F. Straight[4Σ], Aaron Streets[1,2,3,6Σ]

[1] Department of Bioengineering, University of California, Berkeley, CA 94720
[2] UC Berkeley-UCSF Graduate Program in Bioengineering, University of California, Berkeley, Berkeley, CA 94720
[3] Center for Computational Biology, University of California, Berkeley, CA 94720
[4] Department of Biochemistry, Stanford University, Stanford, CA 94305
[5] Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305
[6] Chan Zuckerberg Biohub, San Francisco, CA 94158
[7] Department of Molecular & Cell Biology, University of California, Santa Cruz, CA 95064
[8] UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA 95064
*current address: Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720

[Π]These authors contributed equally, listed alphabetically
[Σ]These authors co-supervised the study; to whom correspondence should be addressed: astreets@berkeley.edu, astraigh@stanford.edu

# Contents


**Supplementary Table 1** **3**

**Supplementary Table 2** **5**

**Supplementary Table 3** **6**

**Supplementary Figure 1** **7**

**Supplementary Figure 2** **8**

**Supplementary Note 1** **9**

**Supplementary Note 2** **10**

**Supplementary Note 3** **12**

**Supplementary Note 4** **13**

**Supplementary Note 5** **14**

**Supplementary Note 6** **15**

**Supplementary Note 7** **16**

**Supplementary Note 8** **17**

**Supplementary Note 9** **18**

**Supplementary Note 10** **20**

**Supplementary Note 11** **21**

**Supplementary Note 12** **24**

**Supplementary Note 13** **25**

**Supplementary Note 14** **26**

**Supplementary Note 15** **28**

**References** **29**

# Supplementary Table 1

| ID | Batch | BC | Cell Line | Ab | Ab dil. | Other / Notes | pA/G | Link. len. (aa) | MTase | [MTase] (nM) | Ab2 | pA/G bind temp | Act. buf. | Act. time (min) | Act. [SAM] (uM) | Read number | Total bases sequenced | Mean read len. | ON:OFF | ON-target prop. mA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 6 | HEK293T | LMNB1 | 500 | SRE XL | pAG | 29 | EcoGII | 50 | N | 4C | A | 30 | 500 | 86,481 | 1,976,058,348 | 22,850 | 1.67 | 4.46E-05 |
| 2 | 2 | 9 | HEK293T | LMNB1 | 500 | SRE XL, Ab2 bind at 4C | pAG | 29 | EcoGII | 150 | Y | 4C | A | 30 | 500 | 38,539 | 847,065,308 | 21,979 | 2.93 | 1.57E-04 |
| 3 | 2 | 10 | HEK293T | LMNB1 | 500 | SRE XL | pAG | 29 | EcoGII | 150 | N | 4C | A | 30 | 500 | 36,921 | 856,905,359 | 23,209 | 2.50 | 1.37E-04 |
| 4 | 2 | 14 | HEK293T | LMNB1 | 500 | NP40 0.5%,SRE XL,Ab2@4C | pAG | 29 | EcoGII | 150 | Y | 4C | A | 30 | 500 | 60,602 | 1,490,024,882 | 24,587 | 1.21 | 1.18E-04 |
| 5 | 2 | 15 | HEK293T | LMNB1 | 500 | NP40 0.5%, SRE XL | pAG | 29 | EcoGII | 150 | N | 4C | A | 30 | 500 | 38,772 | 943,586,111 | 24,337 | 1.35 | 7.23E-05 |
| 6 | 3 | 16 | HEK293T | LMNB1 | 500 | | pAG | 29 | EcoGII | 527 | N | RT | A | 30 | 500 | 162,252 | 1,752,177,765 | 10,799 | 1.38 | 5.32E-05 |
| 7 | 3 | 17 | HEK293T | LMNB1 | 500 | | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 198,264 | 2,102,108,611 | 10,603 | 3.03 | 9.52E-05 |
| 8 | 3 | 18 | HEK293T | LMNB1 | 500 | | pAG | 29 | EcoGII | 150 | N | RT | A | 30 | 500 | 106,892 | 1,227,114,851 | 11,480 | 2.05 | 2.99E-05 |
| 9 | 3 | 19 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | RT | A | 30 | 500 | 238,437 | 2,415,676,861 | 10,131 | 4.80 | 1.74E-04 |
| 10 | 4 | 20 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | Y | 4C | A | 30 | 500 | 121,419 | 1,261,478,986 | 10,389 | 4.73 | 4.01E-04 |
| 11 | 4 | 21 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | 4C | A | 30 | 500 | 123,908 | 1,276,684,561 | 10,303 | 3.98 | 2.99E-04 |
| 12 | 4 | 23 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | 4C | A | 60 | 500 | 112,198 | 1,162,823,359 | 10,364 | 3.73 | 3.23E-04 |
| 13 | 4 | 24 | HEK293T | LMNB1 | 100 | SAM replenished | pAG | 29 | EcoGII | 527 | N | 4C | A | 60 | 500*2 | 119,153 | 1,207,239,004 | 10,132 | 3.05 | 3.13E-04 |
| 14 | 4 | 1 | HEK293T | LMNB1 | 100 | SAM replenished | pAG | 29 | EcoGII | 527 | N | 4C | A | 60 | 500*2 | 96,385 | 1,076,704,141 | 11,171 | 2.65 | 2.87E-04 |
| 15 | 5 | 4 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 80,135 | 839,079,448 | 10,471 | 3.56 | 4.22E-04 |
| 16 | 5 | 5 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | Y | RT | A | 15 | 500 | 101,911 | 1,067,611,582 | 10,476 | 3.87 | 3.25E-04 |
| 17 | 5 | 6 | HEK293T | LMNB1 | 100 | 30C act | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 68,779 | 651,361,202 | 9,470 | 7.54 | 4.95E-04 |
| 18 | 5 | 7 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 83,487 | 902,381,502 | 10,809 | 4.85 | 4.40E-04 |
| 19 | 5 | 8 | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 527 | Y | RT | A | 30 | 500 | 26,351 | 227,185,814 | 8,622 | 5.64 | 7.04E-05 |
| 20 | 5 | 9 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | Y | RT | A | 30 | 500 | 19,112 | 169,551,133 | 8,871 | 7.42 | 8.39E-05 |
| 21 | 5 | 10 | HEK293T | LMNB1 | 100 | | pA | mix | Hia5 | 527 | Y | RT | A | 30 | 500 | 15,778 | 151,798,963 | 9,621 | 9.84 | 7.88E-05 |
| 22 | 5 | 11 | HEK293T | LMNB1 | 100 | pAG-EcoGII+pA-Hia5-both | mix | mix | Both | 527 | Y | RT | A | 30 | 500 | 15,975 | 149,309,499 | 9,346 | 11.82 | 2.65E-04 |
| 23 | 5 | 12 | HEK293T | LMNB1 | 100 | NP40 0.1% | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 86,033 | 860,416,335 | 10,001 | 2.48 | 2.50E-04 |
| 24 | 5 | 14 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | RT | A | 30 | 500 | 101,662 | 993,731,709 | 9,775 | 2.51 | 2.97E-04 |
| 25 | 5 | 15 | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 527 | N | RT | A | 30 | 500 | 106,830 | 985,205,512 | 9,222 | 12.43 | 7.89E-04 |
| 26 | 5 | 16 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | N | RT | A | 30 | 500 | 74,873 | 793,137,043 | 10,593 | 11.72 | 1.06E-03 |
| 27 | 5 | 17 | HEK293T | LMNB1 | 100 | | pA | mix | Hia5 | 527 | N | RT | A | 30 | 500 | 93,269 | 988,096,097 | 10,594 | 17.68 | 8.36E-04 |
| 28 | 5 | 18 | HEK293T | LMNB1 | 100 | pAG-EcoGII+pA-Hia5-both | mix | mix | Both | 527 | N | RT | A | 30 | 500 | 92,725 | 945,637,416 | 10,198 | 6.09 | 7.71E-04 |
| 29 | 5 | 19 | HEK293T | LMNB1 | 100 | **light fixation** | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 91,520 | 686,940,638 | 7,506 | 5.37 | 4.31E-04 |
| 30 | 6 | 21 | HEK293T | LMNB1 | 100 | Triton 0.1% | pA | 26 | Hia5 | 527 | N | RT | A | 30 | 500 | 100,177 | 1,126,513,813 | 11,245 | 4.66 | 9.10E-05 |
| 31 | 6 | 22 | HEK293T | LMNB1 | 100 | Triton 0.1% | pA | 26 | Hia5 | 527 | N | RT | B | 30 | 500 | 114,462 | 1,105,844,974 | 9,661 | 21.18 | 3.68E-04 |
| 32 | 6 | 23 | HEK293T | LMNB1 | 100 | NP40 0.1% | pA | 26 | Hia5 | 527 | N | RT | A | 30 | 500 | 105,547 | 934,810,760 | 8,857 | 3.68 | 6.12E-05 |
| 33 | 6 | 24 | HEK293T | LMNB1 | 100 | NP40 0.1% | pAG | 29 | EcoGII | 527 | N | RT | A | 30 | 500 | 84,310 | 868,314,441 | 10,299 | 4.31 | 1.11E-04 |
| 34 | 6 | 1 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | RT | A | 30 | 500 | 91,978 | 1,189,823,234 | 12,936 | 3.41 | 1.49E-04 |
| 35 | 6 | 2 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | N | RT | A | 30 | 500 | 105,110 | 1,264,716,882 | 12,032 | 24.21 | 5.33E-04 |
| 36 | 6 | 3 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | N | RT | B | 30 | 500 | 103,978 | 1,311,352,305 | 12,612 | 38.05 | 2.06E-03 |
| 37 | 7 | 4 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | N | RT | B* | 30 | 500 | 90,785 | 752,583,312 | 8,290 | 26.47 | 4.08E-03 |
| 38 | 7 | 5 | HEK293T | LMNB1 | 100 | | pAG | 7 | Hia5 | 527 | N | RT | B* | 30 | 500 | 48,122 | 459,680,632 | 9,552 | 29.83 | 2.23E-03 |
| 39 | 7 | 6 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | RT | B* | 30 | 500 | 81,695 | 727,342,635 | 8,903 | 7.48 | 1.21E-03 |
| 40 | 7 | 7 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | N | RT | B | 30 | 500 | 65,117 | 583,081,677 | 8,954 | 23.49 | 3.05E-03 |
| 41 | 7 | 8 | HEK293T | LMNB1 | 100 | | pAG | 7 | Hia5 | 527 | N | RT | B | 30 | 500 | 56,317 | 540,388,209 | 9,595 | 29.75 | 2.08E-03 |
| 42 | 7 | 9 | HEK293T | LMNB1 | 100 | | pAG | 29 | EcoGII | 527 | N | RT | B | 30 | 500 | 47,684 | 449,643,915 | 9,430 | 8.22 | 1.15E-03 |
| 43 | 7 | 10 | HEK293T | LMNB1 | 100 | 30C act | pA | 26 | Hia5 | 527 | N | RT | B | 30 | 500 | 58,489 | 523,464,949 | 8,950 | 29.98 | 2.24E-03 |
| 44 | 7 | 11 | HEK293T | LMNB1 | 100 | 30C act | pAG | 7 | Hia5 | 527 | N | RT | B | 30 | 500 | 57,609 | 511,675,484 | 8,882 | 17.42 | 1.26E-03 |
| 45 | 7 | 12 | HEK293T | LMNB1 | 100 | 30C act | pAG | 29 | EcoGII | 527 | N | RT | B | 30 | 500 | 61,562 | 551,698,966 | 8,962 | 4.40 | 5.33E-04 |
| 46 | 7 | 13 | HEK293T | LMNB1 | 100 | 30C act | pAG | 29 | EcoGII | 527 | N | RT | A | 30 | 500 | 71,900 | 628,811,974 | 8,746 | 8.18 | 3.02E-04 |
| 47 | 7 | 14 | HEK293T | LMNB1 | 100 | Ab2 (GP) | pA | 26 | Hia5 | 527 | Y | RT | B | 30 | 500 | 99,602 | 843,096,447 | 8,465 | 21.37 | 2.33E-03 |
| 48 | 7 | 15 | HEK293T | LMNB1 | 100 | Ab2 (GP) | pA | 26 | Hia5 | 527 | Y | RT | A* | 30 | 500 | 54,677 | 484,736,060 | 8,865 | 8.61 | 4.85E-04 |
| 49 | 7 | 16 | HEK293T | LMNB1 | 100 | Ab2 (GP) | pAG | 7 | Hia5 | 527 | Y | RT | B | 30 | 500 | 63,400 | 571,575,224 | 9,015 | 13.88 | 1.43E-03 |
| 50 | 7 | 17 | HEK293T | LMNB1 | 100 | Ab2 (GP), 30C act | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 60,290 | 517,001,320 | 8,575 | 7.27 | 3.43E-04 |
| 51 | 7 | 18 | HEK293T | LMNB1 | 100 | pA-Hia5 short+long | pA | mix | Hia5 | 527 | N | RT | B | 30 | 500 | 47,923 | 468,996,420 | 9,786 | 21.99 | 3.19E-03 |
| 52 | 7 | 19 | HEK293T | LMNB1 | 100 | pA-Hia5 200 nM, Ab2 (GP) | pA | 26 | Hia5 | 200 | Y | RT | B | 30 | 500 | 52,514 | 513,948,158 | 9,787 | 19.59 | 2.82E-03 |
| 53 | 7 | 20 | HEK293T | LMNB1 | 100 | pA-Hia5 50 nM, Ab2 (GP) | pA | 26 | Hia5 | 50 | Y | RT | B | 30 | 500 | 48,723 | 441,526,404 | 9,062 | 20.30 | 2.55E-03 |
| 54 | 7 | 22 | HEK293T | LMNB1 | 100 | **frozen** | pA | 26 | Hia5 | 527 | N | RT | B | 30 | 500 | 71,276 | 488,718,798 | 6,857 | 28.06 | 2.98E-03 |
| 55 | 7 | 23 | HEK293T | LMNB1 | 100 | Ab2 (Goat), 30C act | pAG | 29 | EcoGII | 527 | Y | RT | A | 30 | 500 | 61,444 | 554,745,469 | 9,028 | 8.71 | 5.32E-04 |
| 56 | 8 | 1 | HEK293T | LMNB1 | 100 | poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 45,362 | 345,941,926 | 7,626 | 7.68 | 2.36E-03 |
| 57 | 8 | 2 | HEK293T | LMNB1 | 100 | poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 40,187 | 353,173,477 | 8,788 | 6.86 | 2.38E-03 |
| 58 | 8 | 3 | HEK293T | LMNB1 | 50 | poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 33,387 | 278,128,346 | 8,330 | 9.62 | 3.97E-03 |
| 59 | 8 | 7 | GM12878 | LMNB1 | 100 | poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 55,380 | 462,838,466 | 8,358 | 8.39 | 2.75E-03 |
| 60 | 8 | 8 | HG002 | LMNB1 | 100 | poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 59,508 | 439,509,594 | 7,386 | 9.34 | 3.13E-03 |
| 61 | 8 | 9 | Hap1 | LMNB1 | 100 | poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 59,648 | 424,749,575 | 7,121 | 7.47 | 3.43E-03 |
| 62 | 8 | 15 | HEK293T | LMNB1 | 100 | **light fixation**, poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 52,734 | 315,838,720 | 5,989 | 9.74 | 3.58E-03 |
| 63 | 8 | 17 | Hap1 | LMNB1 | 100 | primary at RT, poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 54,120 | 354,171,388 | 6,544 | 8.12 | 3.29E-03 |
| 64 | 8 | 19 | HEK293T | LMNB1 | 100 | **conA beads**, poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 61,614 | 494,542,424 | 8,026 | 8.56 | 3.29E-03 |
| 65 | 8 | 20 | GM12878 | LMNB1 | 100 | **conA beads**, poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 64,420 | 461,182,886 | 7,159 | 5.15 | 3.08E-03 |
| 66 | 8 | 24 | Hap1 | LMNB1 | 100 | **conA beads**, poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 54,354 | 404,350,726 | 7,439 | 4.45 | 3.01E-03 |
| 67 | 9 | 1 | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 45,503 | 360,723,378 | 7,927 | 10.65 | 1.84E-03 |
| 68 | 9 | 4 | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 527 | N | RT | B* | 30 | 500 | 62,799 | 407,120,248 | 6,483 | 16.01 | 3.81E-03 |

| ID | Batch | BC | Cell Line | Ab | Ab dil. | Other / Notes | pA/G | Link. len. (aa) | MTase | [MTase] (nM) | Ab2 | pA/G bind temp | Act. buf. | Act. time (min) | Act. [SAM] (uM) | Read number | Total bases sequenced | Mean read len. | ON:OFF | ON-target prop. mA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **69** | *9* | *5* | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 33,562 | 297,569,878 | 8,866 | 18.09 | 3.28E-03 | |
| **70** | *9* | *6* | HEK293T | LMNB1 | 100 | | pA | 26 | Hia5 | 200 | N | RT | B* | 30 | 500 | 27,725 | 284,031,573 | 10,245 | 19.49 | 3.78E-03 | |
| **71** | *9* | *7* | HEK293T | LMNB1 | 100 | | pAG | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 31,665 | 307,111,011 | 9,699 | 11.34 | 1.75E-03 | |
| **72** | *9* | *8* | HEK293T | LMNB1 | 50 | | pAG | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 40,276 | 349,442,835 | 8,676 | 17.40 | 1.94E-03 | |
| **73** | *10* | *14* | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 202,447 | 1,902,202,299 | 9,396 | 19.71 | 2.52E-03 | |
| **74** | *10* | *15* | HEK293T | LMNB1 | 100 | RNAse | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 376,220 | 2,775,298,200 | 7,377 | 13.63 | 1.64E-03 | |
| **75** | *10* | *17* | HEK293T | LMNB1 | 100 | 300mM NaCl final wash | pA | 26 | Hia5 | 200 | N | RT | B* | 30 | 500 | 226,875 | 1,978,617,616 | 8,721 | 17.77 | 2.73E-03 | |
| **76** | *11* | *20* | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 129,920 | 1,273,519,587 | 9,802 | 24.33 | 2.85E-03 | |
| **77** | *11* | *21* | HEK293T | LMNB1 | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 138,407 | 1,318,014,140 | 9,523 | 34.31 | 2.90E-03 | |
| **78** | *12* | *7* | HEK293T | LMNB1 | 50 | 2M cells | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 105,855 | 977,322,733 | 9,233 | 15.73 | 3.50E-03 | v1 |
| **79** | *12* | *8* | HEK293T | LMNB1 | 50 | 5M cells | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 80,057 | 694,123,573 | 8,670 | 15.07 | 3.97E-03 | |
| **80** | *12* | *9* | HEK293T | LMNB1 | 50 | 10M cells | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 101,673 | 861,972,532 | 8,478 | 13.25 | 2.40E-03 | |
| **98** | *13* | *13* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 97,515 | 860,318,284 | 8,822 | 8.71 | 2.23E-03 | |
| **99** | *13* | *14* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 62,174 | 588,488,369 | 9,465 | 10.45 | 2.70E-03 | |
| **100** | *13* | *15* | HEK293T | LMNB1 | 50 | No spermidine at activation | pA | 7 | Hia5 | 200 | N | RT | B*-- | 30 | 800 | 49,639 | 506,842,274 | 10,211 | 13.31 | 4.15E-03 | |
| **101** | *13* | *16* | HEK293T | LMNB1 | 50 | 0.05 mM spermidine at activation | pA | 7 | Hia5 | 200 | N | RT | B*- | 30 | 800 | 40,995 | 459,789,338 | 11,216 | 14.99 | 4.21E-03 | |
| **102** | *13* | *17* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 60 | 800 | 44,601 | 547,230,128 | 12,269 | 11.99 | 4.12E-03 | |
| **103** | *13* | *18* | HEK293T | LMNB1 | 50 | pipetted mid-way through act. | pA | 7 | Hia5 | 200 | N | RT | B* | 120 | 800 | 58,127 | 594,768,652 | 10,232 | 13.05 | 5.22E-03 | |
| **104** | *13* | *19* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 66,138 | 651,244,669 | 9,847 | 9.81 | 2.05E-03 | |
| **105** | *13* | *20* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 60 | 800 | 54,069 | 564,188,122 | 10,435 | 14.23 | 4.14E-03 | |
| **106** | *13* | *21* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 120 | 800 | 49,174 | 538,723,774 | 10,955 | 13.86 | 5.06E-03 | |
| **107** | *13* | *22* | HEK293T | LMNB1 | 50 | No spermidine at all steps | pA | 7 | Hia5 | 200 | N | RT | B*-- | 30 | 800 | 62,941 | 613,561,043 | 9,748 | 6.02 | 1.28E-03 | |
| **108** | *13* | *23* | HEK293T | LMNB1 | 50 | 0.05 mM spermidine at all steps | pA | 7 | Hia5 | 200 | N | RT | B*- | 30 | 800 | 49,696 | 519,922,031 | 10,462 | 11.12 | 2.57E-03 | |
| **109** | *13* | *24* | HEK293T | LMNB1 | 50 | Ca + Mg instead of spermidine, but removed at act. | pA | 7 | Hia5 | 200 | N | RT | B*-- | 30 | 800 | 49,574 | 553,202,469 | 11,159 | 14.72 | 3.09E-03 | |
| **110** | *14* | *7* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800 | 34,594 | 406,707,546 | 11,757 | 20.61 | 3.00E-03 | |
| **111** | *14* | *8* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 30,919 | 368,525,479 | 11,919 | 21.17 | 5.15E-03 | |
| **112** | *14* | *9* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 21,318 | 285,761,191 | 13,405 | 18.69 | 5.65E-03 | |
| **113** | *14* | *10* | HEK293T | LMNB1 | 50 | no spin post act. | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 19,530 | 240,481,368 | 12,313 | 18.55 | 3.03E-03 | |
| **114** | *14* | *11* | HEK293T | LMNB1 | 50 | no spin post act. | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 24,613 | 305,984,775 | 12,432 | 21.42 | 5.44E-03 | |
| **115** | *14* | *12* | HEK293T | LMNB1 | 50 | spin max speed post act. | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 20,713 | 267,789,859 | 12,929 | 26.70 | 5.77E-03 | |
| **116** | *14* | *13* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 27,158 | 346,224,365 | 12,749 | 29.76 | 1.96E-03 | |
| **117** | *14* | *14* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 500 | N | RT | B*- | 120 | 800*2 | 26,971 | 331,687,491 | 12,298 | 20.87 | 5.82E-03 | |
| **118** | *14* | *15* | HEK293T | LMNB1 | 50 | 2h pA binding | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 23,237 | 288,234,035 | 12,404 | 21.17 | 4.81E-03 | |
| **119** | *14* | *16* | HEK293T | LMNB1 | 50 | | pA | 7 | Hia5 | 200 | N | 4C | B*- | 120 | 800*2 | 21,467 | 259,041,131 | 12,067 | 22.20 | 5.20E-03 | |
| **120** | *14* | *17* | HEK293T | LMNB1 | 50 | 2h pA binding | pA | 7 | Hia5 | 200 | N | 4C | B*- | 120 | 800*2 | 25,022 | 307,073,669 | 12,272 | 20.53 | 6.21E-03 | v2 |
| **121** | *14* | *18* | HEK293T | LMNB1 | 50 | Ca + Mg instead of spermidine, removed at act., no spin | pA | 7 | Hia5 | 200 | N | RT | B*-- | 120 | 800*2 | 15,375 | 189,086,758 | 12,298 | 21.14 | 3.23E-03 | |
| **122** | *14* | *24* | HEK293T | LMNB1 | 50 | No tween throughout | pA | 7 | Hia5 | 200 | N | RT | B*- | 120 | 800*2 | 33,527 | 365,439,027 | 10,900 | 20.07 | 4.71E-03 | |
| **123** | *15* | *2* | HEK293T w/ EcoGII-LMNB1 | - | - | Stably transduced with inducible EcoGII-LMNB1 | - | 19 | EcoGII | - | - | - | - | - | - | 61,294 | 445,922,529 | 7,275 | 10.46 | 3.97E-03 | |

**Supplementary Table 1. All LMNB1-directed conditions tested.** Conditions are specified on the left side of the table, and outputs are summarized on the right hand side. For each protocol parameter (columns), the option that was selected for the final protocol is highlighted in green. ON:OFF represents the ratio of the proportion of adenines (q>=10) methylated (p>=0.9) in cLADs (on-target regions) to the proportion of adenines methylated in ciLADs (off-target regions), and cells are colored by the magnitude of this ratio (white=low, magenta=high). The neighboring column is shaded white to purple to correspond to the proportion of adenines methylated in cLADs. Each batch was run on a separate day. A list of abbreviations follows. Ab: Primary Antibody, Ab2: Secondary Antibody (either Goat if not specified or GP for guinea pig), Ab. Dil.: the dilution factor of the antibody (50 = 1:50), BC: barcode, RT: Room Temperature, Buffer A: same as wash buffer in final protocol, Buffer A*: same as A but with 75 mM NaCl, Buffer B: activation buffer (15 mM Tris, pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, pH 8.0, 0.5 mM EGTA, pH 8.0, 0.5 mM Spermidine, 800 µM SAM) without BSA, Buffer B*: Buffer B with 0.1% BSA, Buffer B*-: Buffer B* with 0.05 mM spermidine, Buffer B*--: Buffer B* with no spermidine, conA: concanavalin A beads, MTase: methyltransferase, RNAse: the nuclei were treated with RNAse prior to antibody binding, SAM: S-adenosylmethionine (methyl donor), SRE XL: the Circulomics Short Read Eliminator XL kit was used to select longer fragments prior to sequencing. The red arrows indicate instances of the v1 and v2 protocols (conditions 78 and 120).

# Supplementary Table 2

| ID | Batch | BC | Cell Line | Ab | Ab dil. | Other / Notes | pA/G | Link. len. (aa) | MTase | [MTase] (nM) | Ab2 | pA/G bind temp | Act. buf. | Act. time (min) | Act. [SAM] (uM) | Read number | Total bases sequenced | Mean read len. | ON:OFF | ON-target prop. mA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **81** | *4* | *22* | HEK293T | LMNB1 | 100 | **no SAM** | pAG | 29 | EcoGII | 527 | N | 4C | A | 30 | **0** | 144,251 | 1,297,558,669 | 8,995 | 1.22 | 3.83E-05 |
| **82** | *8* | *14* | HEK293T | LMNB1 | 100 | **no SAM**, poor batch | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | **0** | 37,978 | 301,486,227 | 7,938 | 2.58 | 5.13E-04 |
| **83** | *1* | *7* | HEK293T | **IgG** | 500 | SRE XL | pAG | 29 | EcoGII | 50 | N | 4C | A | 30 | 500 | 39,167 | 872,653,096 | 22,280 | 1.11 | 2.22E-05 |
| **84** | *2* | *13* | HEK293T | **IgG** | 500 | SRE XL | pAG | 29 | EcoGII | 150 | N | 4C | A | 30 | 500 | 66,878 | 1,500,527,829 | 22,437 | 0.98 | 4.86E-05 |
| **85** | *4* | *2* | HEK293T | **IgG** | 100 | | pAG | 29 | EcoGII | 527 | Y | 4C | A | 30 | 500 | 165,186 | 1,608,748,957 | 9,739 | 0.82 | 6.05E-05 |
| **86** | *4* | *3* | HEK293T | **IgG** | 100 | | pAG | 29 | EcoGII | 527 | N | 4C | A | 30 | 500 | 95,567 | 982,777,929 | 10,284 | 0.84 | 7.71E-05 |
| **87** | *8* | *10* | GM12878 | **IgG** | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 35,649 | 304,927,237 | 8,554 | 1.77 | 4.57E-04 |
| **88** | *8* | *11* | HG002 | **IgG** | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 58,023 | 418,761,436 | 7,217 | 1.45 | 4.44E-04 |
| **89** | *8* | *12* | Hap1 | **IgG** | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 73,437 | 453,120,139 | 6,170 | 1.86 | 4.77E-04 |
| **90** | *8* | *13* | HEK293T | **IgG** | 100 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 500 | 50,184 | 395,339,186 | 7,878 | 1.56 | 4.15E-04 |
| **91** | *12* | *16* | GM12878 | **IgG** | 50 | | pA | 7 | Hia5 | 200 | N | RT | B* | 30 | 800 | 169,811 | 1,344,179,912 | 7,916 | 1.47 | 1.50E-04 |
| **92** | *5* | *13* | HEK293T | - | - | **free floating EcoGII** | pAG | 29 | EcoGII | 527 | - | - | A | 30 | 500 | 102,650 | 983,478,388 | 9,581 | 0.57 | 2.80E-04 |
| **93** | *5* | *20* | HEK293T | - | - | **light fix., free floating EcoGII** | pAG | 29 | EcoGII | 527 | - | - | A | 30 | 500 | 118,802 | 260,574,348 | 2,193 | 0.57 | 8.96E-05 |
| **94** | *8* | *16* | HEK293T | - | - | **free floating Hia5** | - | - | Hia5 | 200 | - | | B* | 30 | 500 | 49,149 | 363,924,351 | 7,405 | 1.10 | 7.55E-03 |
| **95** | *10* | *21* | HEK293T | - | - | **free floating Hia5** | - | - | Hia5 | 200 | - | | B* | 30 | 500 | 224,739 | 1,954,375,136 | 8,696 | 1.17 | 6.34E-03 |
| **96** | *10* | *22* | HEK293T | - | - | **free floating Hia5** w/ RNAse | - | - | Hia5 | 200 | - | | B* | 30 | 500 | 192,453 | 1,755,187,759 | 9,120 | 1.23 | 6.96E-03 |
| **97** | *12* | *17* | GM12878 | - | - | **free floating Hia5** | - | - | Hia5 | 200 | - | | B* | 30 | 800 | 94,126 | 831,589,618 | 8,835 | 1.12 | 7.98E-03 |

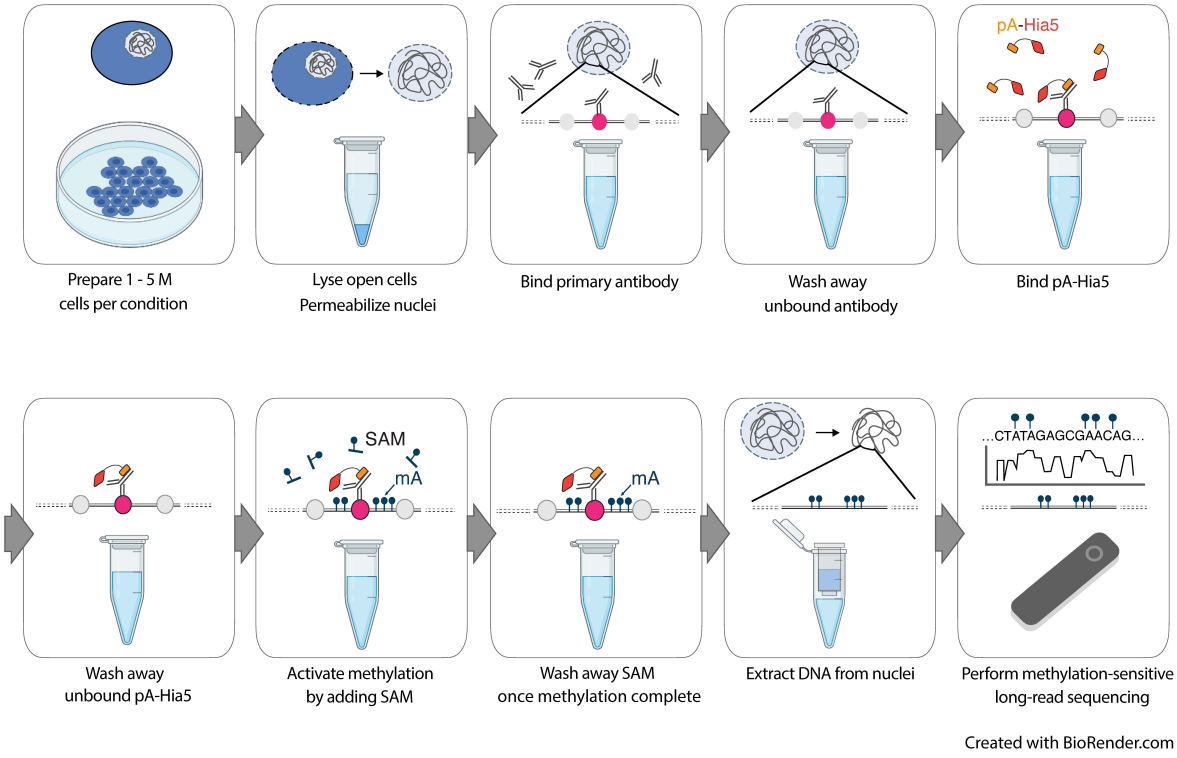
**Supplementary Table 2. Control conditions tested.** Same as Supplementary Table 1 but for 3 different negative control conditions (not expected to show methylation, or not expected to show enrichment in cLADs): no SAM added at activation/methylation step, nonspecific IgG isotype control antibody used, or free-floating enzyme was added at the activation/methylation step to methylate all accessible DNA.

# Supplementary Table 3

| Sample number | Target / description | Cell line | Reads | Bases | Mean read length | Protocol version |
|---|---|---|---|---|---|---|
| 1 | CTCF | GM12878 | 729097 | 8824849107 | 12104 | v1 |
| 2 | CTCF | GM12878 | 209699 | 2864997263 | 13662 | v1 |
| 3 | CTCF | GM12878 | 215664 | 7052382103 | 32701 | v1 |
| 4 | CTCF | GM12878 | 161144 | 1961203749 | 12170 | v1 |
| 5 | CTCF | GM12878 | 194417 | 2202030238 | 11326 | v2 - opt1 |
| 6 | CTCF | GM12878 | 231715 | 2395286461 | 10337 | v2 - opt2 |
| 7 | CTCF | GM12878 | 201517 | 1719224973 | 8531 | v2 - opt3 |
| 8 | CTCF | GM12878 | 191839 | 2218327728 | 11563 | v2 |
| 9 | CTCF | GM12878 | 1781121 | 23009057167 | 12918 | v2 |
| 10 | CTCF | GM12878 | 1820026 | 22882150932 | 12572 | v2 |
| 11 | free pA-Hia5 | GM12878 | 970282 | 10371517658 | 10689 | v1 |
| 12 | IgG | GM12878 | 1421526 | 11531597797 | 8112 | v1 |
| 13 | H3K9me3 | HG002 | 896511 | 8057506035 | 8988 | v1 |
| 14 | H3K9me3 | HG002 | 155656 | 1847831908 | 11871 | v1 |
| 15 | H3K9me3 | HG002 | 233920 | 5798858000 | 24790 | v1 |
| 16 | free pA-Hia5 | HG002 | 71713 | 1433794520 | 19994 | v1 |
| 17 | free pA-Hia5 | HG002 | 204504 | 3771902818 | 18444 | v1 |
| 18 | free pA-Hia5 | HG002 | 64758 | 1219166120 | 18826 | v1 |
| 19 | free pA-Hia5 | HG002 | 145380 | 2878390589 | 19799 | v1 |
| 20 | IgG | HG002 | 306270 | 7857066233 | 25654 | v1 |
| 21 | IgG | HG002 | 132045 | 2685488328 | 20338 | v1 |
| 22 | in vitro methylated genomic DNA | GM12878 | 330573 | 2785142705 | 8425 | v1 |
| 23 | unmethylated genomic DNA | GM12878 | 437135 | 4040533340 | 9243 | v1 |
| 24 | CENP-A | HG002 | 258948 | 2278497590 | 8799 | v1 |
| 25 | CENP-A | HG002 | 325860 | 2490299385 | 7642 | v1 |
| 26 | CENP-A | HG002 | 65205 | 454714846 | 6974 | v1 |
| 27 | CENP-A | HG002 | 320201 | 6719916952 | 20986 | v1 |
| 28 | CENP-A | HG002 | 1669667 | 5013914326 | 3003 | v1 |
| 29 | free pA-Hia5 | HG002 | 230216 | 2018755168 | 8769 | v1 |
| 30 | free pA-Hia5 | HG002 | 114364 | 848652995 | 7421 | v1 |
| 31 | free pA-Hia5 | HG002 | 257095 | 1740051056 | 6768 | v1 |
| 32 | IgG | HG002 | 252694 | 2349991082 | 9300 | v1 |
| 33 | IgG | HG002 | 235264 | 2068374069 | 8792 | v1 |
| 34 | IgG | HG002 | 84856 | 624068929 | 7354 | v1 |
| 35 | untreated | HG002 | 260016 | 2167665966 | 8337 | v1 |
| 36 | untreated | HG002 | 436472 | 3262286345 | 7474 | v1 |
| 37 | untreated | HG002 | 86613 | 573212999 | 6618 | v1 |

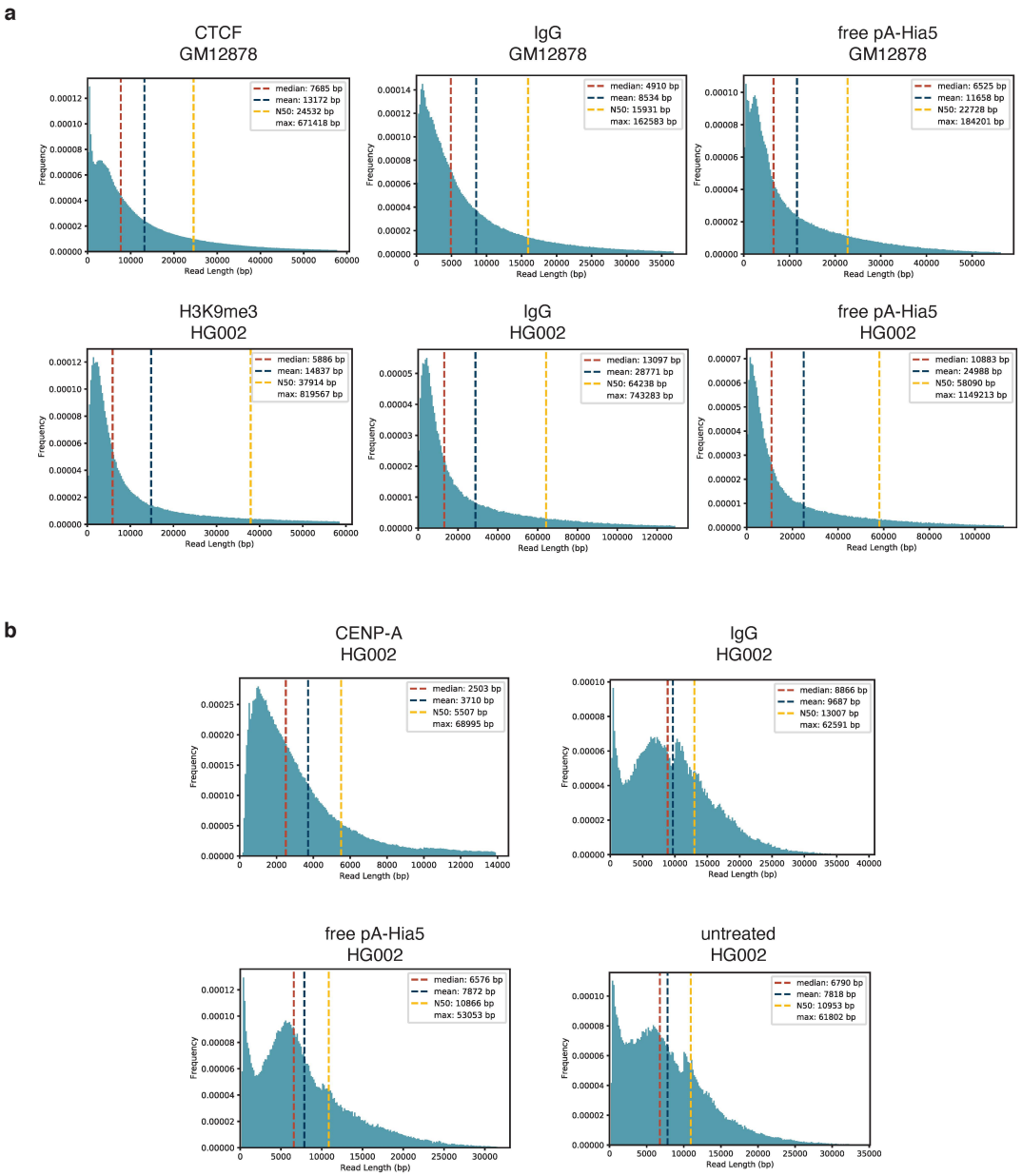
**Supplementary Table 3. Sequencing summary metrics.** The number of reads, bases, and mean read length are indicated for CTCF-, H3K9me3-, and CENP-A-directed DiMeLo-seq, along with accompanying controls. Protocol version indicates whether the standard protocol (v1) or the protocol for optimized methylation efficiency (v2) was used (Methods). For CTCF samples used in optimization for v2 protocol development, the optimization conditions are: opt1: 2 hour activation, 0.05 mM spermidine at activation, replenish SAM; opt2: 2 hour activation, 0.05 mM spermidine at activation, replenish SAM, 500 nM pA-Hia5; opt3: 2 hour activation, no spermidine, 1 mM Ca++ and 0.5 mM Mg++ buffer.

# Supplementary Figure 1




**Supplementary Figure 1. Workflow of DiMeLo-seq *in situ* methylation, DNA extraction, and sequencing.** Schematic of the DiMeLo-seq *in situ* methylation protocol, which involves a series of binding steps and washes followed by DNA extraction and sequencing.

# Supplementary Figure 2




**Supplementary Figure 2. Read length distribution histograms.** Read length distributions for mapped reads passing quality filters with median, mean, N50, and max read length indicated. Outliers above Q3 + 3*IQR are not shown. Distributions contain reads merged across all experiments for a given target and cell line described in Supplementary Table 3. **a,** Histograms from samples prepared with standard DiMeLo-seq with variable library preparation cleanup methods (Methods). Larger fragments are maintained in HG002 samples where we pelleted and resuspended the DNA between library preparation steps rather than performing traditional bead-based cleanups. **b,** Histograms from samples prepared with DiMeLo-seq with AlphaHOR-RES.

# Supplementary Note 1

***In vitro* DNA methylation assay**

Hia5, pA-pHia5, and pAG-pHia5 concentrations were estimated using the extinction coefficients. Serial dilutions (100, 10, 1, 0.1 nM for Hia5 and pA-Hia5 comparison, 30, 3, 0.3 nM for Hia5 and pAG-Hia5 comparison) were made using Buffer A (15 mM Tris-HCl, pH 8.0; 15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA, and 0.5 mM spermidine). Proteins were then mixed with Buffer A supplemented with S-adenosyl-methionine (NEB B9003S) containing 1 ug of either naked unmethylated DNA (Plasmid ASP3552, 2x601, prepared from GM2163 dam- *E. coli* strain) or methylated DNA (Plasmid ASP3552, 2x601, prepared from DH5a *E. coli* strain). Reactions were incubated for 1 hr at 37°C. PCR purification was performed to extract DNA which was then digested with DpnI for 1.5 hours at 37°C and run on an agarose gel to assess the degree of methylation.

# Supplementary Note 2

**Reconstituted chromatin experiments**

1x601 DNA containing plasmid was obtained from Addgene (pGEM-3z/601 Plasmid #26656). A 730 bp region containing 1x601 sequence in the middle was amplified (Forward primer - CAGGTTTCCCGACTGGAAAG, Reverse primer with NheI and AscI sites- GATCGCTAGCGGCGCGCCATTCAGGCTGCGCAAC) and digested with NheI. After digestion, the DNA was was then biotinylated by filling in NheI 5'overhang with dGTP (NEB), α-thio-dCTP (Chemcyte), α-thio-dTTP (Chemcyte), and biotin-14-dATP (Thermo Fisher Scientific) using Large Klenow fragment 3'-5' exo- (NEB).

18x601 DNA array was obtained as previously described[1]. To summarize, puC18 vector with 18 repeats of the "601" nucleosome positioning sequence[2] (ASP 696) was transformed into competent dam- *E. coli* strain, GM2163, and purified using a QIAGEN Gigaprep kit. The unmethylated 18x601 plasmid was digested with EcoRI, XbaI, DraI, and HaeII. Array DNA used in directed methylation experiments was then biotinylated by filling in EcoRI and XbaI 5'overhangs with dGTP (NEB), α-thio-dCTP (Chemcyte), α-thio-dTTP (Chemcyte), and biotin-14-dATP (Thermo Fisher Scientific) using Large Klenow fragment 3'-5' exo- (NEB).

Histones for chromatin assembly (CENP-A, H3, H4, H2A, and H2B) were purified as previously described[1,3]. Chromatin was reconstituted using salt dialysis as described previously[1]. 1x601 or 18x601 biotinylated DNA, H2A/H2B histone dimer, and tetramer (H3/H4 or CENP-A/H4 histone tetramer) were added to high salt buffer (10 mM Tris-HCl, pH 7.5; 0.25 mM EDTA; 2 M NaCl). The mixture was gradually dialyzed over the course of ~67 hours at a rate of 0.5 mL/min from high salt buffer into low salt buffer (10 mM Tris-HCl, pH 7.5; 0.25 mM EDTA; 2.5 mM NaCl). CENP-A/H4 or H3/H4 tetramer concentrations were titrated to obtain chromatin of varying saturation. Nucleosome assembly on 1x601 DNA was verified using overnight digestion at room temperature with BsiWI (restriction site at the center of 601 sequence) followed by native acrylamide gel shift analyses and agarose gels. After digestion, intact nucleosome occupied 1x601 chromatin (uncut by BsiWI, 730bp) was separated from digested nucleosome unoccupied 1x601 DNA (cut by BsiWI, 360bp and 370bp) using glycerol gradient ultracentrifugation and fractionation. 50 ul of overnight BsiWI digested chromatin was pipetted on top of a 5 mL 5-30% (w/v) linear glycerol gradient in buffer containing 20 mM Tris pH 7.5, 0.25 mM EDTA, 0.1% Igepal/NP40, 10 mM KCl, and spun for 16 hours at 35000 rpm at 4 ℃ in a SW-55 Ti swinging-bucket rotor (Beckman Coulter). After the spin, 100 ul fractions were collected from the top of glycerol gradient. Fractions containing high migrating nucleosome bands (> 730 bp) were collected and concentrated for 1x601 experiments.

Chromatin assembly on 18x601 array was verified using a native acrylamide gel shift analysis after overnight restriction digestion using AvaI at room temperature (18x601 array DNA contains engineered AvaI recognition sites between adjacent 601 positions) (Extended Data Fig. 2a,c)[1].

# Supplementary Note 3

### *In vitro* chromatin methylation

In experiments involving free pA-Hia5 (non-targeted) methylation on chromatin, reconstituted chromatin was incubated in activation buffer (15 mM Tris-Cl pH 8.0, 15 mM NaCl, 60 mM KCl, 0.1% w/v Bovine Serum Albumin (BSA)) containing 0.8 mM SAM and 25 nM pA-Hia5 or pAG-Hia5 for 30 minutes at 37 ℃. In antibody-directed methylation experiments, chromatin reconstituted on biotinylated 18x601 array DNA was used. In DNA LoBind Eppendorf tubes, M-280 Streptavidin-coated Dynabeads (Invitrogen) were washed in bead buffer (50 mM Tris, pH 7.4, 75 mM NaCl, 0.25 mM EDTA, 0.05% Triton X-100, and 2.5% polyvinyl alcohol (30kDa - 70 kDa)) and incubated with biotinylated CENP-A or H3 containing chromatin (at 12.5 nM 601 concentration or 0.7 nM 18x601 concentration) for 1 hour at room temperature with constant agitation. Chromatin-coated beads were then magnetically separated and washed twice with Chromatin wash buffer (CWB)-75 (20 mM HEPES pH 7.5, 75 mM KCl, 0.05% Triton X-100, 0.1% BSA) and then incubated in CWB-75 containing 1 ug/mL of rabbit anti-CENP-A antibody[4], mouse anti-H3 antibody (MABI 0301, Active Motif), or rabbit or mouse IgG (Jackson Immuno Research) control for 30 minutes with agitation at room temperature. Beads were then washed twice with CWB-75 and incubated in CWB-75 for 30 minutes with agitation, followed an additional wash in CWB-75 before incubation in CWB-75 containing 25 nM pA-Hia5 (in CENP-A-directed methylation experiments) or pAG-Hia5 (in H3-directed methylation experiments). After incubation with pA-Hia5 or pAG-Hia5, beads were washed twice with CWB-100 (20 mM HEPES pH 7.5, 100 mM KCl, 0.05% Triton X-100, 0.1% BSA) to remove unbound pA-Hia5 and then resuspended in activation buffer (15 mM Tris-Cl pH 8.0, 15 mM NaCl, 60 mM KCl, 0.1% w/v BSA) containing 0.8 mM SAM for 30 minutes at 37℃. Beads were then split into two tubes and processed separately for immunostaining (with anti-mA antibody) and library preparation (for long-read sequencing). For library preparation, chromatin was released from beads using BamHI and KpnI digestion (cuts near biotinylated ends of 18x601 array DNA), or using AscI digestion (cuts near biotinylated end of 1x601), DNA was extracted, and processed using Oxford Nanopore Technology native barcoding (PCR-free) kit (EXP-NBD104 or EXP-NBD114) with the Ligation Sequencing Kit (SQK-LSK109).

# Supplementary Note 4

***In vitro* chromatin DiMeLo-seq analyses**

Reads from *in vitro* experiments were initially basecalled with Guppy (4.4.2) using the fast basecalling model (dna_r9.4.1_450bps_fast.cfg). After initial basecalling reads were demultiplexed and split by barcode using the guppy_barcoder and fast5_subset from ont_fast5_api. Fast5s for each barcode were then aligned and modification basecalled with Megalodon (2.2.9) using the rerio all-context basecalling model (res_dna_r941_min_modbases-all-context_v001.cfg) with --guppy_params "trim_barcodes" and --mod_min_prob 0. In experiments involving 1x601 array, reads less than 700 bp, representing BsiWI digestion products of unoccupied 1x601, were removed from downstream analyses. In experiments involving 18x601 array, reads less than 3.6 kb (i.e. 18x 200 bp repeats of 601), representing partial arrays, were removed from downstream analyses. Modification basecalled reads were smoothed by calculating rolling average over a 50 bp window in a NaN-sensitive manner (averaging only over adenine bases). Following smoothing, adenine bases with methylation probability score > 0.6 were assigned as methylated (mA). The threshold of 0.6 was empirically determined by comparing pA-Hia5 treated and untreated naked 1x601 DNA, false detection rate (FDR) < 5%, (Extended Data Fig. 1f). FDR was estimated using binned probability scores for reads from naked DNA methylated with free pA-Hia5 or untreated, corresponding to True Positive or True Negative respectively. For a given cutoff, a read is classified as methylated if the percentage of methylation (i.e. % mA/A) on that read is greater than the cutoff. FDR was calculated as (FPR/(FPR+TPR)). For classifying 1x601 reads as methylated, we empirically determined the minimum percentage of each read to be methylated above a given threshold (0.6) from Receiver Operator Characteristic curves comparing binned methylation on reads from 1x601 CENP-A chromatin after CENP-A-directed methylation (as TPR) to IgG-directed methylation (Extended Data Fig. 1g) or no treatment (Extended Data Fig. 1h) (as FPR). In experiments estimating extent of methylation on 1x601 reads (Fig. 2d), we classify a portion of the read centered at the 601 dyad as methylated if 20% of its length is methylated above the threshold of 0.6 (Purple dot in Extended Data Fig. 1g, h, Dotted line on Extended Fig. 1k).

For clustering and visualizing methylation on individual 18x601 reads, we first classified each 601 position as with or without nucleosome. A region spanning 400 bp centered at each theoretical 601 dyad position was classified as containing a nucleosome if > 10% and < 60% of that region was methylated. (< 60% is used to filter out regions that do not show nucleosome protection). Reads were then clustered by performing hierarchical clustering of jaccard distances of inferred nucleosome positions on either 18x601 or 4x601 region.

# Supplementary Note 5

**Chromatin-coated beads immunostaining and imaging**

Following incubation in activation buffer, chromatin coated beads were incubated in CWB-2M (20 mM HEPES pH 7.5, 2M NaCl, 0.05% Triton X-100, 0.1% BSA) for 1 hour at 55 ℃ to denature protein. Beads were then washed twice in CWB-2M to remove denatured protein while retaining biotinylated DNA on beads. (Anti-CENP-A antibody and anti-methyladenine antibody are both derived from rabbit, therefore, to avoid cross-reactivity with anti-rabbit conjugated secondary antibody used for immunofluorescence, chromatin coated beads were washed with CWB-2M as mentioned above to remove CENP-A antibody prior to staining with anti-N6-methyladenosine antibody.) Beads were then washed twice with Antibody dilution buffer or AbDil (20 mM Tris-HCl, pH 7.4, 150 mM NaCl with 0.1% Triton X-100, and 2% BSA) and dropped onto poly-L-lysine-coated coverslips and allowed to attach for 30 minutes. Coverslips were incubated with AbDil containing 1 ug/ml rabbit anti-N6-methyladenosine antibody (Millipore Sigma ABE572) for 30 minutes, washed twice with AbDil, and incubated with AbDil containing 2 ug/ml Alexa 647 fluorophore conjugated goat anti-rabbit secondary antibody (Molecular Probes) for 30 minutes. Coverslips were then washed twice with AbDil, incubated with AbDil containing 1 ug/ml propidium iodide (Sigma) for 10 minutes, washed twice with AbDil and phosphate buffered saline (PBS), blotted gently, mounted in 90% glycerol, 10 mM Tris-Cl pH 8.8, and 0.5% *p*-phenylenediamine, and sealed using clear nail polish.

Imaging was performed using IX70 (Olympus) microscope with a DeltaVision core system (Applied precision) with a Sedat quad-pass filter set (Semrock) and monochromatic solid-state illuminators, controlled via softWoRx 4.1.0 software (Applied Precision). Images were acquired using a 100x 1.4 NA Plan Apochromatic oil immersion objective (Olympus) and captured using a CoolSnap HQ CCD camera (Photometrics). Z-stacks were acquired at 0.2 uM intervals over a total 3 uM total axial distance. Bead images were analyzed using custom ruby software[3]. At least 50 beads were analyzed for each condition per experiment.

# Supplementary Note 6

**Modification calling thresholds**

Basecalling was performed using Oxford Nanopore Technologies's Guppy software (v4.5.4) and Megalodon software (v2.3.1) with the Rerio res_dna_r941_min_modbases-all-context_v001.cfg basecalling model. To estimate false positive rates (FPR) at each mA probability score threshold, we counted the fraction of As called as mA on untreated GM12878 genomic DNA, which should lack any methyladenines (Extended Data Fig. 3). To provide a lower bound on the true positive rate (TPR), we counted the fraction of As called as mA on purified GM12878 genomic DNA treated with pA-Hia5 *in vitro* (Extended Data Fig. 3). Using these values, we could estimate a lower bound on the FDR (FPR/(FPR+TPR)). For Guppy modified base calls, we used a modification probability threshold of 0.6 (basecalling 0.0009 FDR, 0.000245 FPR, 0.281 TPR lower bound). For Megalodon's modified base calls, we used a modification probability threshold of 0.75 (basecalling 0.0008 FDR, 0.000159 FPR, 0.203 TPR lower bound). For some analyses higher thresholds were used; for example, a stringent Guppy threshold of 0.9 was used for LMNB1 analyses (Extended Data Fig. 4). We note that the predominant source of background noise in DiMeLo-seq stems from off-target methylation, as opposed to false-positive methylation calls. Using higher mA score thresholds effectively serves as a threshold on higher mA density, to distinguish on-target methylation from off-target methylation.

# Supplementary Note 7

**LMNB1 data analysis**

All sequencing was performed on ONT MinION v9.4 flow cells. Basecalling and modification calling were performed on Amazon Web Services g4dn.metal instances, which have 8 NVIDIA T4 GPUs, 96 CPUs, 384 Gb memory, and 2x900 Gb local solid-state storage; this configuration allows for efficient parallelization and high basecalling speed. Basecalling was first performed using Oxford Nanopore Technologies's Guppy software (v4.5.4), using a Rerio res_dna_r941_min_modbases-all-context_v001.cfg basecalling model, and demultiplexing when appropriate. Modification calls were extracted from fast5 output files using ont-pyguppy-client-api. Basecalled reads were aligned to the T2T-CHM13v1.0 reference sequence using Winnowmap (v2.03), which is adapted to perform better than other long-read aligners in repetitive regions[5]. Fast5 files were split by barcode using fast5_subset then re-basecalled using ONT's Megalodon software (v2.3.1), using the same reference and model file. Custom code was used to parse output files and isavailable on Github. To evaluate performance, cLAD and ciLAD coordinates[6] were lifted over from hg38 to the T2T-CHM13v1.0 reference[7]. Single-cell Dam-LMNB1 data were re-mapped to T2T-CHM13v1.0 and processed as described in Altemose et al.[6]. Browser plots were made using the WashU Epigenome Browser[8].

Figure 3e: In single-cell DamID, each 100 kb bin of the genome is given a binary classification indicating whether it was in contact with the nuclear lamina or not in that particular cell during an ~18 hour incubation period when Dam-LMNB1 is expressed *in vivo*[9]. Across a sample of 32 single cells, we used these binary classifications to estimate a scDamID-based LMNB1 interaction frequency for each bin of the genome across the sample of cells[6,9]. We then performed a similar binary classification of individual LMNB1-targeted DiMeLo-seq reads based on each read's proportion of methylated adenines, determining a lamina interaction threshold (mA/A > 0.001 with a stringent mA calling threshold of 0.9) to identify reads from cLADs with 59% sensitivity and 94% specificity (Extended Data Fig. 4). The DiMeLo-seq-based interaction frequency for each bin was then computed as the proportion of overlapping reads with mA/A above the lamina interaction threshold. A read was determined to overlap a bin if it aligned to it with more than 50% of its length, and any mA calls on that read were assigned to that bin for browser plotting. 100 kb genomic bins were filtered to those with at least 60 overlapping DiMeLo-seq reads, and with a single-cell combined mean-squared-error estimate <0.004, to select for regions with higher-confidence interaction frequency estimates.

# Supplementary Note 8

**LMNB1 optimization**

We found that we could reliably estimate protocol performance parameters (on-target methylation and on-target:off-target methylation) using only ~0.2X genome-wide coverage per sample, allowing us to multiplex several conditions on the same MinION flow cell and achieve sufficient coverage after only 24 hours of sequencing. Using the v2 protocol and applying a stringent methylation score threshold of 0.9 (Extended Data Fig. 3a,b-4, Supplementary Note 6), we regularly achieve on-target methylation of 0.3-0.6% of adenines in cLADs, with an on-target:off-target ratio in the range of 15-30 (Supplementary Table 1). These performance metrics depend on the choice of mA score threshold (Extended Data Fig. 4c), which was chosen to balance sensitivity and specificity in distinguishing regions with on-target and off-target methylation. We note that this threshold does not primarily serve to reduce false-positive mA calls, which occur at an extremely low rate (Extended Data Fig. 3a,b; see full discussion of threshold evaluation in Supplementary Note 6). Unlike other protein-DNA mapping methods, which use sequencing coverage as a readout of interaction frequency, DiMeLo-seq sequences the entire genome without enrichment for interacting regions. Thus, as further validation we can plot DiMeLo-seq's coverage and methylation frequency as separate tracks in a browser representing the T2T-CHM13 complete reference sequence, and we can compare these to the results obtained for the same protein target in the same cells by conventional bulk DamID (Fig. 3c).

Surprisingly, we found no improvement in on-target methylation when using a secondary antibody to recruit more methyltransferase molecules to each site, perhaps due to steric effects, and we saw no improvement when increasing the linker length between pA and Hia5 (Extended Data Fig. 3 and Supplementary Table 1). We saw a slight drop in performance when using pAG-Hia5 compared to pA-Hia5, also potentially due to steric effects. We also found that cell permeabilization with NP40 or Triton X-100 (vs. standard digitonin) actively reduces methylation downstream (Supplementary Table 1). While optimization was carried out in HEK293T cells, we also validated that the protocol worked in other human cell lines: Hap1, GM12878, and HG002.

# Supplementary Note 9

**DiMeLo-seq with Concanavalin A coated magnetic beads & input considerations**

Concanavalin A coated magnetic beads (Bangs Laboratories BP531) were tested as an alternative to centrifugation for cell pelleting throughout the protocol, adapted from the CUT&RUN protocol[10,11]. To equilibrate beads, conA bead slurry was resuspended by gentle vortexing and 10 µL of bead slurry per sample was added to 1.5 ml conA binding buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM CaCl2, 1 mM MnCl2), then placed on a magnet. Supernatant was removed and beads were resuspended in 1.5 ml conA binding buffer again, then cleared on a magnet again. Washed beads were resuspended in 20 ul conA binding buffer per sample. For the experiments numbered 64-66 in our Supplementary Table 1, we used conA beads with 500k, 430k, and 500k cells each for HEK293T (~triploid), GM12878 (diploid), and Hap1 (haploid) cells, respectively. Prior to the permeabilization step, cells were first washed with PBS 3x by centrifugation, then resuspended in 1 ml wash buffer (20 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1 Roche Complete tablet -EDTA per 50 ml buffer). 10 µl of equilibrated bead slurry was added while gently vortexing the cell suspension, and the beads + cells were incubated for 10 minutes at room temperature on a rotator. Bead-bound cells were pelleted on a magnet and resuspended in Dig-Wash buffer to begin the permeabilization step, and the remainder of the protocol was carried out as described above, substituting magnetic separation for centrifugation. Note: conA beads may interfere with the ability to perform quality control by IF. At the DNA extraction step, cells were lysed on beads, and beads were separated from lysate on a magnet prior to proceeding with DNA precipitation. The final DNA yield was 20% for HEK293T, 39% for GM12878, and 75% for Hap1, relative to what one would theoretically expect from the input number of cells, after accounting for ploidy (estimated as 3*ploidy pg per cell). This variance in DNA recovery may have to do with the propensity for each cell type to bind conA beads and resist nuclear envelope rupture, or possibly to do with relative cell sizes and the binding capacity of the conA beads.

The input requirements for DiMeLo-seq ultimately depend on a multitude of factors: the desired coverage, the desired fragment length distribution, the genome size, the ploidy of the cell type, and the efficiency of the DNA extraction and library prep protocols being used. For the conA bead experiment with 430k GM12878 cells, we yielded 500 ng of DNA after extraction, and ~200 ng after library prep. If prepared with an lsk-110 kit, this would be enough to load a minION flowcell twice while maintaining high pore occupancy (100 ng per loading). Each loading of a flowcell yields ~9 Gb on average, so this amount of DNA would provide 6x coverage of the human genome. Thus, based on our empirical results from this replicate, we can estimate that around 200k diploid cells are needed for 3x human genome coverage. For a line/protocol with higher recovery efficiency, this could fall closer to 100k cells per 3x human

genome coverage. These input requirements may continue to decrease with flow cell designs and new library prep chemistries.

# Supplementary Note 10

**Creation and induction of stable cell lines for *in vivo* DiMeLo-seq**

Stable HEK293T cell lines were created by retroviral transduction followed by drug selection. Retroviral plasmids containing DDdegron-EcoGII-V5linker-LMNB1 were obtained from Addgene (#122083; Sobecki et al.[12]). Retroviruses were produced in the Phoenix Ampho packaging cell line (obtained from the UC Berkeley cell culture facility). Phoenix cells were seeded in standard growth medium (DMEM with 10% FBS and 1X P/S) in a T75 flask 24 hours before transfection, aiming for 70% confluence at the time of transfection. 25 µg of plasmid DNA was combined with 75 µl FUGENE-HD transfection reagent in 1200 µl optiMEM and incubated for 10 minutes, then added to the media. After 12 hours, the media was replaced with fresh media, and the cells were incubated at 32 °C with 5% $CO_2$ and 100% humidity to help preserve viral particles. 36 hours later, the virus-containing media was harvested and centrifuged at 1800 rpm for 5 minutes to remove any Phoenix cells. The media was supplemented with 10 µl/ml of 1 M HEPES and 4 µg/ml of polybrene. For HEK293T cells, 2.5 ml of this media was added to each well of a 6-well plate containing adhered cells at 40-50% confluence. Plates were spinoculated in a centrifuge with a swinging-bucket plate rotor at 1300x*g* for 1 hour at room temperature, then incubated at 37 °C overnight. The media was replaced the next morning. After 24 hours, puromycin was added to the media at a concentration of 1 µg/ml and the media was replenished every 48 hours for 10 days. Surviving cells were expanded and frozen for later use. 15 hours prior to harvesting, 1 µM Aqua-Shield-1 reagent (AOBIOUS AOB6677, made to 0.5 mM stock) was added to the media to stabilize protein expression. DNA was harvested using an NEB Monarch Genomic DNA Purification Kit (T3010S), sheared to a target of 8 kb using a Covaris g-tube (Covaris 520079), and purified with a Circulomics SRE XS kit (SS-100-121-01), then barcoded and library prepped with method 1 described below.

The higher methylation observed in the *in vivo* sample likely owes to the effectively longer incubation time during which methyl groups can be deposited on adenines *in vivo* (15 h) compared to *in situ* (2 h), as well as to chromatin dynamics *in vivo* that may make a greater fraction of the genome accessible to the methyltransferase[13]. However, compared to *in situ* DiMeLo-seq with pA-Hia5, the *in vivo* EcoGII-LMNB1 approach produced 36% less on-target methylation and 25% more off-target methylation.

# Supplementary Note 11

**CTCF data analysis**

For GM12878 samples (CTCF-targeted, IgG control, free pA-Hia5, in vitro methylated genomic DNA, and untreated genomic DNA), Megalodon modified basecalls were used for analysis. Reference GM12878 ChIP-seq peaks (ENCFF797SDL, ENCODE Project Consortium[14]) were lifted over from hg38 to T2T chm13v1.0. These peaks were intersected with known CTCF motifs that were also lifted over to T2T chm13v1.0[15]. Reference GM12878 ATAC-seq peaks (ENCFF748UZH, ENCODE Project Consortium[14]) were also lifted over from hg38 to T2T chm13v1.0. Enrichment in CTCF ChIP-seq peaks and ATAC-seq peaks was calculated using bedtools (v2.28.0). The anti-CTCF antibody (ab188408) was confirmed by personal correspondence to bind a peptide in the first 600 C-terminal amino acids of the protein.

For analysis of CTCF-targeted DiMeLo-seq data, modified basecalls for reads spanning CTCF ChIP-seq motifs were extracted within -1 kb to 1 kb of the motif center. To extract single molecules spanning peaks, pysam (v0.15.3) was used with code adapted from De Coster et al.[16]. If the motif was on the - strand, the positions of bases relative to the motif center were flipped. Only non-overlapping CTCF sites within the -1 kb to 1 kb display were considered, and only mA called with probability $\geq$ 0.75 were plotted. Aggregate profiles were plotted with a moving average of 50 bp. For peak and read counts considered, see Extended Data Figure 5j. For joint analysis of mA and mCpG on the same molecules, only molecules spanning motifs in the top decile of ChIP-seq peaks that have at least one mA called with probability $\geq$ 0.75 and one mCpG called with probability $\geq$ 0.75 were considered, resulting in 23,147 reads considered.

To test that C-terminal and N-terminal CTCF targeting produce significantly different methylation enrichment patterns, we performed a Fisher's exact test comparing the fraction of methylated adenines 3' to the motif center (-300 bp to 0 bp) to the fraction of methylated adenines 5' to the motif center (0 bp to +300 bp) for the two samples (p-value of $5.9 \times 10^{-8}$). We used the fisher_exact function from scipy (v1.4.1) with the alternative hypothesis "greater" for this comparison of C-terminal to N-terminal targeting. The significance still held when narrowing the test region to include only the central peak (-100 bp to 100 bp), and in this region, the p-value was 0.00010.

We detected peaks in our DiMeLo-seq data in aggregate to create Extended Data Figure 5f. First, we took the average probability of methylation reported across all reads for a given base in the reference. We then computed the mean methylation probability in a 200 bp sliding window with a 20 bp step size. Next, for various average methylation cutoff thresholds from 0 to 50, we classified 200 bp bins as true positive (above threshold in DiMeLo-seq, overlapping ChIP-seq peak), false negative (below threshold in DiMeLo-seq, overlapping ChIP-seq peak), false

positive (above threshold in DiMeLo-seq, not overlapping ChIP-seq peak), or true negative (below threshold in DiMeLo-seq, not overlapping ChIP-seq peak). We then created the ROC curves having performed this peak calling method with 25X, 20X, 15X, 10X, and 5X coverage for our CTCF-targeted sample. The area under the curve was calculated for the various sequencing depths using sklearn.metrics.auc function (v.0.24.2).

To call CTCF peaks on single molecules, all molecules spanning top decile ChIP-seq peaks that had at least one mA detected with probability ≥ 0.9 were considered (25,122 reads). We used a more stringent probability threshold to select reads that contained confident methylation because the goal was to determine where the peak center is detected on reads that call a peak. Reads were also filtered to require they span the CTCF motif with at least 100 bp covered on each side of the motif. With a sliding window of 20 A's, the probability that at least one A was methylated within the bin was computed by calculating 1-exp(sum(log(1-p))) for each mA with probability > 0.5. We used a lower probability cutoff for calculating binned probabilities to detect peaks on single molecules because on single molecules, we wanted to capture any mA calls, even lower confidence calls, to increase our sensitivity at the cost of specificity. Reads that had at least one binned probability ≥ 0.8 have a called peak, and the peak center was calculated as the midpoint of the longest stretch of mA with binned probability ≥ 0.8 (1.2% FDR). The FDR was calculated as the fraction of adenines methylated in the unmethylated control divided by the fraction of adenines methylated in the CTCF-targeted sample using these same filtering criteria.

To estimate the single-molecule sensitivity for detecting CTCF binding events, we performed a binary classification of CTCF-targeted DiMeLo-seq reads based on the proportion of adenines methylated. We quantified this proportion in both on-target peak regions, defined as +/- 150 bp of the CTCF binding motif center in top decile ChIP-seq peaks and in off-target regions, defined as -2000 to -1850 and +1850 to +2000 bp of the top decile motif center, once baseline background *in situ* methylation levels have been reached. Using this approach, we calculated TPR and FPR as a function of the number of mA in these 300 bp regions required to consider a CTCF binding event detected, and approximated the single-molecule sensitivity to be 54% (FPR 5.7%) when requiring 6 mA with probability ≥ 0.75 in 300 bp for a CTCF binding event detection.

For analysis of single molecules spanning two CTCF sites, peak pairs that were 2 to 10 kb apart were selected from all CTCF ChIP-seq peaks). As in peak calling, binned qualities in bins of 20 A's were computed. Here, if a binned probability > 0.9 fell within 100 bp on either side of at least one of the two CTCF binding sites, the read was considered to have a called peak and the molecule was included in Figure 4c. A total of 1959 peak pairs were considered with a total of 3036 reads spanning these peaks with a peak detected at at least one of the two sites (4207 total reads spanned these pairs of sites). Reads were clustered using k-means clustering (scikit-learn v0.24.2) with 3 clusters.

A vcf file containing high-quality phased heterozygous polymorphisms in GM12878 were obtained from https://hgdownload.soe.ucsc.edu/gbdb/hg38/platinumGenomes/hg38.hybrid.vcf.gz, which combines variant calls from the Platinum Genomes and Genome in a Bottle projects[17,18]. This vcf was lifted over from hg38 to CHM13v1.0 using VCF-liftover (https://github.com/hmgu-itg/VCF-liftover) with a chain file from http://t2t.gi.ucsc.edu/chm13/hub/t2t-chm13-v1.0/hg38Lastz/hg38.t2t-chm13-v1.0.over.chain.gz. DiMeLo-seq alignments were phased using NanoMethPhase v1.0[19] with parameters --mapping_quality 10 --min_SNV 1 --average_base_quality 10. Because NanoMethPhase requires base quality values, the input bam files for CTCF phasing were obtained by merging guppy output bam file information with alignment position information from winnowmap using custom in-house code available on github. IGV v2.11.4[20] was used for initial data exploration (note: for megalodon mod_mappings bam files, "C+Z" was replaced with "C+m", and "A+Y" was replaced with "A+a" in each line of the bam file for proper visualization). Final single-molecule plots were made with custom in-house code available on github. CTCF site coordinates within the H19/IGF2 Imprinting Control Region were obtained from Ulaner et al.[21].

# Supplementary Note 12

**PacBio data analysis**

Starting with the hifi_reads.bam file output from the sequencer, we used SMRTLink (v10) command-line tools to process the data. First we used ccs-kinetics-bystrandify to create a bam file with forward and reverse strands as separate reads. We aligned this bam file using pbmm2 align to the T2T-CHM13v1.0 reference and extracted reads that overlapped the top decile of CTCF ChIP-seq peaks using bedtools (v2.28.0) intersect. We then ran a custom script provided by PacBio to compute an IPD ratio for each base. We aligned this output using pbmm2 align to the T2T-CHM13v1.0 reference. We then used custom scripts to extract single base IPD ratios for comparison to nanopore for Extended Data Fig. 8. In particular, methylated base calls +/- 100 bp around the CTCF motif center for top decile CTCF ChIP-seq peaks were extracted. For PacBio, we plotted the fraction of adenines methylated in this peak region as a function of IPD ratio and number of passes. For Nanopore, we plotted the peak methylation as a function of mA probability. For both, we compared to the methylation detected in the untreated control in this same peak region. We then selected a constant peak methylation level of 10% of adenines methylated and compared the profiles for PacBio and CTCF with thresholds corresponding to a peak methylation rate of 10%.

# Supplementary Note 13

**H3K9me3 data analysis**

For all HG002 samples (H3K9me3-targeted, IgG control, and free pA-Hia5) a merged bam file was created with samtools (v1.8) from the Guppy bam and winnowmap outputs aligned to a special male reference genome (CHM13+HG002X+hg38Y: autosomes from the T2T chm13v1.0 genome combined with a T2T assembly of HG002 chromosome X[7] and the chrY sequence from hg38), and a mapping quality threshold of 10 was applied. To compare to CUT&RUN, broad peaks were called using macs2 (v2.1.1) on a H3K9me3 CUT&RUN bam file from HG002[22]. Regions outside H3K9me3 CUT&RUN regions for Figure 5a were defined as regions of the genome outside of the called broad H3K9me3 peaks with 10 kb buffer on each side of the called peaks. Centromere and HOR boundaries were defined from the T2T centromere annotation[22]. Enrichment in CUT&RUN peaks, centromeres, and active HOR arrays was computed using bedtools (v2.28.0). For analyzing mA signal at HOR boundaries, the mean mA/A in a 100 kb rolling window from -300 kb within the HOR to 300 kb outside of the HOR was computed. A total of 2,359 reads spanned this region. HOR boundaries considered were those that transition quickly into non-repetitive sequences: 1p, 2pq, 6p, 9p, 13q, 14q, 15q, 16p, 17pq, 18pq, 20p, 21q, 22q. For single molecule browser visualization, modified bases were extracted as in CTCF analysis using custom python scripts, and modified bases with probability $\geq 0.6$ were displayed. Single-molecule browser plots were generated using plotly (v4.5.2) with code adapted from De Coster et al.[16].

# Supplementary Note 14

**CENP-A data analysis**

Basecalling for centromere enriched samples was performed twice both times using Guppy (5.0.7). The first basecalling used the "super accuracy" basecalling model (dna_r9.4.1_450bps_sup.cfg), followed by alignment to the CHM13+HG002X+hg38Y reference genome using Winnowmap (v2.03). These alignments were then filtered for only primary alignments and mapq score greater than 10 using samtools view -F 2308 -q10. A second round of basecalling was then performed again using Guppy (5.0.7) but now with the rerio all-context basecalling model (res_dna_r941_min_modbases-all-context_v001.cfg) with --bam_out and --bam_methylation_threshold 0.0. Modified basecalls were then merged by read id with winnowmap alignments to generate bam files with high confidence alignments combined with modification calls for downstream processing. For CENP-A-directed experiments four independent biological replicates were used, and for controls (IgG-directed, free-floating pA-Hia5, and untreated), two independent biological replicates were used. For all samples the first replicate was sequenced on two separate flow cells and all sequencing runs were merged for the final analysis.

To calculate centromere enrichment samtools bedcov was used to calculate the total bases that mapped to alpha satellite active HORs in free-floating Hia5 treated samples treated with and without centromere enrichment[23]. Reported coverage at each centromeric region is relative to the length of that region. Chromosomes with more than one active HOR had the mean value of length-normalized coverage reported. deepTools2 bamCoverage (v3.3.1)[24] was used to generate bigWigs with 10 kb bin size, that were plotted on HG002 chromosome X using pygenometracks (v3.6)[25] to compare chromosome-wide coverage between centromere enriched and unenriched samples.

A *k*-mer counting pipeline was used to identify CENP-A enriched *k*-mers from chm13 Native ChIP-seq experiment[26,27]. After separating DiMeLo-seq reads into those that did and did not have a CENP-A enriched *k*-mer, methylation frequency for each subset was calculated, as well as the fold enrichment for percentage mA/A of reads containing CENP-A enriched *k*-mers over those that did not.

For single molecule browser visualization, modified bases were extracted as in CTCF analysis using custom python scripts, and modified bases with mA probability (from Guppy) > 0.6 and mCpG probability (from Guppy) > 0.6 were displayed.

Average fraction of mA or mCpG methylation for aggregate views were calculated as the fraction of reads at each adenine or CpG that have a probability score (from Guppy) greater than

0.6. Representative plots show average fraction of reads at each adenine or CpG with methylation probability score above threshold binned by smoothing over a rolling window of 250 bp for better visualization. Coverage plots indicate the number of reads that are aligned within the region.

For estimating the density of CENP-A-containing nucleosomes per read, a 5 kb window was slid across each read (step size 1 bp), and within that 5 kb window, the proportion of all 200 bp windows (step size 1 bp) containing at least 3 mAs was computed (using a Guppy mA probability threshold of 0.6). On average within alpha satellite, this threshold of 3 mAs corresponds to 5% of all A bases within a 200 bp window. These values were then averaged across all 5 kb read windows overlapping each 5 kb reference window to produce the density plot in Fig. 6g. The thresholds for bin size, minimum percentage methylation, and probability score were empirically determined to produce a 5% FDR, using IgG DiMeLo-seq reads as a true negative control set.

# Supplementary Note 15

**Protein purification**

Histones for chromatin assembly (CENP-A, H3, H4, H2A, and H2B) were purified as previously described[1,3]. pA-Hia5, pAG-Hia5, or Hia5 purification were purified according to Stergachis et al.[28] with a few modifications. Plasmids were transformed into T7 Express lysY competent *E. coli* cells (NEB #C3010I) for recombinant protein expression. 200 mL starter culture was grown in LB broth at 37°C with 50 ug/mL kanamycin and 34 ug/mL chloramphenicol to an $OD_{600}$ of 0.6. Starter cultures were then diluted to 2L culture in LB broth at 37°C with 50 ug/mL kanamycin to an $OD_{600}$ of 0.8 - 1.0. Protein expression was induced using a final concentration of 1mM IPTG (Isopropyl beta-D-1-thiogalactopyranoside) for 4 hours at 20°C with shaking. Cells were then pelleted at 5000 x g for 15 minutes at 4°C. Pelleted cells were resuspended in 35 mL lysis buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 10% glycerol; 0.5% Triton X-100). Resuspended cell pellets were flash frozen in liquid nitrogen and stored at -80°C until purification. After thawing frozen cell pellets, EDTA-free protease inhibitor tablets (Roche 11873580001) and 10 mM β-mercaptoethanol were added. Cells were lysed by probe sonication (6 pulses, 30s on, 1 min off at 200W). Lysed cells were centrifuged for 1 hour at 40,000 x g (4°C) in 50 mL Oakridge tubes. Ni-NTA agarose was prepared with 2x washes of 30 mL equilibration buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 20 mM imidazole) per 5 mL of slurry. Cell lysate was incubated with Ni-NTA agarose and rotated for 1 hour at 4°C. Mixture was poured onto a gravity column, then washed with 40 mL buffer 1 (50 mM HEPES, pH 7.5; 300 mM NaCl; 50 mM imidazole), 30 mL of buffer 2 (50 mM HEPES, pH 7.5; 300 mM NaCl; 70 mM imidazole), and eluted with 30 mL of elution buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 250 mM imidazole). Eluted protein was filtered with a 0.2 μm filter, then loaded onto a HiPrep 26/10 Desalting column (Cytiva) to buffer exchange eluate into FPLC buffer A (50 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1mM DTT). Following buffer exchange, the sample was applied in tandem onto HiTrap Q HP and HiTrap SP HP (Cytiva) columns. Both columns were washed with 5x combined column volumes of FPLC buffer A. The HiTrap Q HP (Cytiva) column was removed and protein was eluted from the SP column using a linear gradient of 20 column volumes with increasing linear gradient of FPLC buffer B (50 mM Tris-HCl, pH 8.0; 1 M NaCl; 1 mM DTT). Fractions were collected and quantified using A280 absorbance. Elution peak fractions were concentrated using a 10K Amicon Ultra-15 tube to final protein concentration > 5 uM. The final concentrated protein was supplemented with 10% glycerol final concentration, aliquoted, and stored at -80°C.

# References


1. Guse, A., Fuller, C. J. & Straight, A. F. A cell-free system for functional centromere and kinetochore assembly. *Nat. Protoc.* **7**, 1847–1869 (2012).
2. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology* vol. 276 19–42 (1998).
3. Westhorpe, F. G., Fuller, C. J. & Straight, A. F. A cell-free CENP-A assembly system defines the chromatin requirements for centromere maintenance. *J. Cell Biol.* **209**, 789–801 (2015).
4. Carroll, C. W., Silva, M. C. C., Godek, K. M., Jansen, L. E. T. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat. Cell Biol.* **11**, 896–902 (2009).
5. Jain, C. *et al.* Weighted minimizer sampling improves long read mapping. *Bioinformatics* **36**, i111–i118 (2020).
6. Altemose, N. *et al.* µDamID: A Microfluidic Approach for Joint Imaging and Sequencing of Protein-DNA Interactions in Single Cells. *Cell Syst* **11**, 354–366.e9 (2020).
7. Nurk, S. *et al.* The complete sequence of a human genome. *bioRxiv* 2021.05.26.445798 (2021) doi:10.1101/2021.05.26.445798.
8. Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU Epigenome Browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).
9. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).
10. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with

high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).

11. Janssens, D. & Henikoff, S. CUT&RUN: Targeted in situ genome-wide profiling with high efficiency for low cell numbers v3. (2019) doi:10.17504/protocols.io.zcpf2vn.
12. Sobecki, M. *et al.* MadID, a Versatile Approach to Map Protein-DNA Interactions, Highlights Telomere-Nuclear Envelope Contact Sites in Human Cells. *Cell Rep.* **25**, 2891–2903.e5 (2018).
13. van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P. & van Steensel, B. Cell cycle dynamics of lamina-associated DNA. *EMBO Rep.* **21**, e50636 (2020).
14. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
15. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).
16. De Coster, W., Stovner, E. B. & Strazisar, M. Methplotlib: analysis of modified nucleotides from nanopore sequencing. *Bioinformatics* **36**, 3236–3238 (2020).
17. Eberle, M. A. *et al.* A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. *Genome Res.* **27**, 157–164 (2017).
18. Zook, J. M. *et al.* An open resource for accurately benchmarking small variant and reference calls. *Nat. Biotechnol.* **37**, 561–566 (2019).
19. Akbari, V. *et al.* Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. *Genome Biol.* **22**, 68 (2021).
20. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
21. Ulaner, G. A. *et al.* CTCF binding at the insulin-like growth factor-II (IGF2)/H19

imprinting control region is insufficient to regulate IGF2/H19 expression in human tissues. *Endocrinology* **144**, 4420–4426 (2003).

22. Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science* **375**, (2022).
23. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
24. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).
25. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
26. Logsdon, G. A. *et al.* The structure, function and evolution of a complete human chromosome 8. *Nature* **593**, 101–107 (2021).
27. Smith, O. K. *et al.* Identification and characterization of centromeric sequences in Xenopus laevis. *Genome Res.* **31**, 958–967 (2021).
28. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).