



Published in final edited form as:

Methods. 2018 June 01; 142: 39–46. doi:10.1016/j.ymeth.2018.05.005.

5C-ID: Increased resolution Chromosome-Conformation-Capture-Carbon-Copy with in situ 3C and double alternating primer design

Ji Hun Kim^{1,*}, Katelyn R. Titus^{1,*}, Wanfeng Gong¹, Jonathan A. Beagan¹, Zhendong Cao², and Jennifer E. Phillips-Cremins^{1,2,3,**}

¹Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104 USA

²Epigenetics Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA

³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA

Abstract

Mammalian genomes are folded in a hierarchy of compartments, topologically associating domains (TADs), subTADs, and looping interactions. Currently, there is a great need to evaluate the link between chromatin topology and genome function across many biological conditions and genetic perturbations. Hi-C can generate genome-wide maps of looping interactions but is intractable for high-throughput comparison of loops across multiple conditions due to the enormous number of reads (>6 Billion) required per library. Here, we describe 5C-ID, a new version of Chromosome-Conformation-Capture-Carbon-Copy (5C) with restriction digest and ligation performed in the nucleus (*in situ* Chromosome-Conformation-Capture (3C)) and ligation-mediated amplification performed with a double alternating primer design. We demonstrate that 5C-ID produces higher-resolution 3D genome folding maps with reduced spatial noise using markedly lower cell numbers than canonical 5C. 5C-ID enables the creation of high-resolution, high-coverage maps of chromatin loops in up to a 30 Megabase subset of the genome at a fraction of the cost of Hi-C.

**Corresponding author: jcremins@seas.upenn.edu.

* Authors contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Authors' Contributions

JEPC and JHK conceived of the study. JHK, WG and JAB performed 5C. KRT implemented the computational pipeline. JEPC, JHK, KRT wrote the manuscript

Competing financial interests

The authors declare no competing financial interests.

Keywords

3D genome folding; higher-order chromatin organization; Chromosome-Conformation-Capture-Carbon-Copy; 5C; double alternating 5C primer design; *in situ* 3C; low cell number; embryonic stem cells

Introduction

Higher-order folding of chromatin in the 3D nucleus has been linked to genome function. Mammalian genomes are arranged in a nested hierarchy of territories [1], compartments [2–4], topologically associating domains [5–8] (TADs), subTADs [3, 9], and long-range looping interactions [10, 11]. Looping interactions have been linked to at least two mechanistically different modes of control over gene expression. First, enhancers can loop to distal target genes in a highly cell type-specific manner to facilitate their precise spatial-temporal regulation [12–15]. Second, long-range loops anchored by the architectural protein CTCF are often constitutive among cell types and form the structural basis for TADs/subTADs [9]. CTCF-mediated interactions connecting loop domains can create insulated neighborhoods that demarcate the search space of enhancers within the domain [16]. Specifically, CTCF anchored constitutive loops can prevent ectopic enhancer activation of genes outside of the domain or aberrant invasion of nonspecific enhancers into an inappropriate domain [16–20]. Mapping of 3D loops genome-wide across hundreds of cell types, species, and developmental lineages is an active area of intense investigation.

As genome-wide chromatin architecture maps become widely available, a critical emerging goal will be to unravel the cause and effect relationship between looping and gene expression. Indeed, there is a great need in the field to build upon descriptive mapping studies and begin to perturb the 3D genome and evaluate the link between chromatin topology and function. One major limitation preventing progress toward this goal is that Hi-C requires more than six billion reads per replicate to obtain high quality, high resolution, genome-wide looping maps [3, 13, 21]. The financial and logistical difficulties of obtaining this read depth makes it intractable to conduct studies with multiple perturbations induced by genome editing, differentiation, or drug treatment. Thus, there is a great need for a technology that creates high-resolution 3D genome folding maps at a much lower cost.

Chromosome Conformation Capture Carbon Copy (5C) is proximity ligation technology pioneered by Dekker and colleagues [22, 23]. 5C adds a hybrid capture step to the classic Chromosome Conformation Capture (3C) method to facilitate the selection of ligation products that occur only in a subset of the genome [22, 24–26]. Loop-resolution maps can be achieved at a fraction of the cost of Hi-C by only querying interactions in a 10–30 Megabase (Mb) subset of the genome [7, 9, 17, 27], allowing many samples and perturbation conditions to be screened in a high-throughput manner. The ability to query a subset of genome contacts is important because genome-editing experiments are often conducted at only one specific location in the genome. Thus, 5C enables the query of the organizing principles governing genome folding at a key subset of loops without requiring the resources to map all loops genome wide.

Despite key advantages in the original 5C technique, it also has key challenges that have held back its widespread use, including: (1) the high number of cells (>40 million) required for quality 3C template creation [2, 9, 27, 28], (2) the high amount of spatial noise caused by non-specific ligation products [29, 30], and (3) the non-comprehensive nature of the single alternating primer design [7, 22–26, 31]. Together, these limitations result in a high number of false negatives due to ligation junctions that are not queried and a high number of false positives due to spatial noise due to non-specific ligation. In the present study, we introduce two major modifications to the 5C protocol that lead to increased resolution 3D genome folding maps with reduced spatial noise using markedly lower cell numbers than canonical 5C. We conduct a comparative analysis of *in situ* [3, 32, 33] vs. canonical dilution 3C [2, 28] and a double alternating [17] vs. single alternating primer design [7, 22–26, 31] and report the downstream effect of these changes on 5C's ability to detect bona fide looping interactions.

Results

Overview

A 5C experiment starts with preparation of the 3C template (Figure 1A–B). Chromatin is fixed within a population of cells with formaldehyde. In canonical dilution 3C [2, 28], cellular and nuclear membranes are disrupted and chromatin is digested in solution with a restriction enzyme (Figure 1A). Ligation is subsequently performed under dilute conditions that promote intra-molecular ligation. By contrast, *in situ* 3C [3, 32] involves restriction enzyme digest and ligation within intact nuclei. In both methods, cross-links are reversed and DNA is isolated to create the 3C template, which represents the genome-wide library of possible hybrid ligation junctions across a population cells (Figure 1B).

The second half of the 5C protocol involves a hybrid capture step based on ligation-mediated amplification to select only a distinct subset of junctions from the genome-wide 3C library (Figure 1C–F). Canonical 5C [7, 22–26, 31] is built on an alternating primer design in which every other fragment is represented by either a Forward (FOR) primer binding to the sense strand or a Reverse (REV) primer binding to the antisense strand (Figure 1C, **left**). The single alternating design only queries approximately half of all ligation junctions in a target region because only FOR-REV primer ligation events are possible (Figure 1D–E, **left**). More recently, Dekker, Lajoie. and colleagues created a new double alternating primer design [17] which incorporates two additional “left-oriented” primers, LFOR and LREV (Figure 1C **right**). The LFOR primer orientation is designed to the antisense strand on fragments also queried by REV primers, whereas the LREV primer orientation is designed to the sense strand on fragments also queried by FOR primers. Thus, the double alternating 5C primer design, there are now two primers representing each fragment, leading to 4 possible primer ligation orientations (FOR-REV, LFOR-LREV, LFOR-REV, FOR-LREV) and the query of nearly all fragment-fragment ligation events in an a priori selected Megabase (Mb)-scale genomic region (Figure 1D–E **right**).

Double alternating primer design achieves increased loop detection sensitivity compared to single alternating design

We hypothesized that by using the double alternating design developed by Dekker and colleagues [17], we could improve canonical 5C's matrix resolution and the specificity and/or sensitivity of loop detection. To test this idea, we first started with a canonical dilution 3C template from pluripotent embryonic stem (ES) cells cultured in 2i media (detailed in Materials and Methods) and compared the quality of 5C libraries created at the same genomic region with both single alternating and double alternating primer designs. A tradeoff of the more comprehensive double alternating primer design is the possibility of artifactual 'self-circles' (*i.e.* ligation events between the 5' and 3' ends of the same restriction fragment; (5) and (6) in Figure 1D–E **right**). We counted the proportion of each possible primer ligation from the double alternating 5C experiment on a dilution 3C template from ES cells. There was an even distribution of ligation events across the four biologically informative primer-primer orientations ((1) FOR-REV: 18.60%, (2) LFOR-LREV: 17.87%, (3) LFOR-REV: 18.11%, (4) FOR-LREV: 18.41%). Importantly, self-circle ligation events ((5) LFOR-REV and (6) FOR-LREV from the same fragment) comprised only <0.1% of all primer ligations (Figure 1E), suggesting that the risk of self-ligation is very small.

We visually inspected 4 kb-binned heatmaps of 5C counts in Megabase-scale genomic regions around *Sox2* and *Zfp462* genes after matrix balancing and sequencing depth correction (detailed in Materials and Methods). We observed that the double alternating primer design results in notable improvement in specific, punctate looping signal between known long-range enhancer promoter-interactions compared to the single alternating primer design (Figure 2A–B). Double alternating 5C maps also showed less missing fragments than single alternating primer maps due to the increased complexity of ligation junctions that are queried and sequenced. In previous 5C studies, a smoothing window at least 5× greater than the bin size was required to reduce the blockiness of maps caused by missing ligation junctions [27, 34]. Here, with double alternating design, we can create heatmaps at 4 kb matrix resolution with no smoothing window and still resolve punctate loops. We provide heatmaps at a 4 kb bin resolution and a 12 kb smoothing window for ease of comparison across the technical conditions (Figure 2A–B).

To further test our qualitative observation of increased looping sensitivity with the double alternating design, we also quantified chromatin looping interactions in each 5C dataset. We modeled binned interactions as a fold-enrichment relative to a background expected model based on distance dependence and local chromatin domain architecture (detailed in Materials and Methods). As previously published [27, 30, 34], we modeled these so-called Observed/Expected values with a parameterized logistic distribution and subsequently converted p-values to interaction scores (Figure 2C; detailed in Materials and Methods). After thresholding interaction scores, we clustered adjacent looping pixels into long-range looping interaction clusters (Figure 2D; detailed in Materials and Methods). Consistent with observations in Figure 2A–B, the interaction score and loop cluster maps also highlight punctate *Sox2* and *Zfp462* gene promoter-enhancer looping clusters (Figure 2C–D). As expected, the chromatin fragments anchoring the base of detected looping interactions

contained high signal for H3K27ac, a chromatin modification known to demarcate active non-coding regulatory elements and active transcription start sites. Importantly, we identified key looping interactions between *Zfp462* and distal enhancers with the double alternating primer design that were not present with the single alternating design. The well-established *Sox2*-super enhancer interaction [5, 9, 27, 34–36] was detected by the single alternating design, but significantly more punctate and less blocky/noisy with the double alternating design. Overall, these data indicate that the double alternating primer design allows for more sensitive detection of looping interactions compared to the single alternating 5C primer design.

***In situ* 3C reduces spatial noise in 5C heatmaps compared to dilution 3C**

We next assessed the quality of the double alternating 5C experiment using *in situ* 3C and dilution 3C templates. We prepared the *in situ* and dilution 3C templates from 2 million and 40 million ES cells cultured in 2i media, respectively, as previously reported (detailed in Materials and Methods). Both dilution and *in situ* 3C led to detection of previously reported looping interactions between *Sox2* and *Zfp462* and their target enhancers (Figure 3A–D, **green arrowheads**). We then intersected cell-type specific annotations of epigenetic marks from ES cells and primary neural progenitor cells (NPCs) [34] with our identified looping clusters (Figure 3E). Looping clusters identified by *in situ* 3C are significantly enriched for ES-specific CTCF and ES-specific enhancers and depleted of NPC-specific CTCF. By contrast, looping clusters in the dilution 3C library showed minimal enrichment of the expected chromatin modifications (Figure 3E). Visual inspection of the maps revealed an extremely high degree of spatial noise and abnormal looping clusters from the dilution 3C template (Figure 3A–D). Spatial variance in dilution 3C was ~1200× and ~1380× higher than that of *in situ* 3C in the genomic regions around the *Sox2* and *Zfp462* genes, respectively (Figure 3F–G). *In situ* 3C resulted in a major improvement in spatial noise (Figure 3F–G) and led to looping interaction pixels grouped in more spherically shaped clusters with minimal background noise around the punctate looping pixels. These results indicate that *in situ* 3C is superior to dilution 3C in reducing overall spatial noise and false positive loop calls due to non-specific ligation events.

Combined implementation of a double alternating primer design and *in situ* 3C allows for the use of lower genome copies than canonical 5C

We observed that implementing a double alternating primer design and *in situ* 3C noticeably improves the quality and resolution of our 5C heatmaps by reducing background noise and allowing for more sensitive detection of chromatin looping interactions (Figure 2–3). Therefore, we hypothesized that we could lower the genome copies required for loop detection by combining the two improvements. The advantage of lowering the required number of genome copies is that lower cell number 5C could be performed in the future, opening up opportunities for conducting 5C analysis on rare cell types and human tissue samples.

We performed double alternating 5C on an *in situ* 3C template made from ES cells cultured in 2i media. Canonical 5C typically performs the ligation-mediated amplification step on 200,000 genome copies (~590 ng) of the mouse 3C template. We tried 590ng, 245 ng, 122

ng, 24ng, 12 ng, and 2.5 ng of the same *in situ* 3C library prepared from mouse ES cells in 2i media, representing 200,000, 100,000, 50,000, 10,000, 5,000 and 1,000 mouse genome copies, respectively. To ensure that the total DNA mass did not affect 5C primer binding and ligation efficiencies, we mixed 3C templates with an excess of salmon sperm DNA (to a total DNA mass of 1,500 ng).

Visual inspection of heatmaps revealed that the 3C template mass could be reduced to 50,000 genome copies and still sensitively detect all gold-standard looping interactions (Figure 4A–D, Supplementary Figure 1A–D). Notably, the quality of chromatin looping signal is drastically reduced when the number is further lowered to 1,000–10,000 genome copies (Figure 4A–D, Supplementary Figure 1A–D). Consistent with this result, quantitative chromatin enrichments were similar for 50,000–200,000 genome copies, but did not show interpretable results at 1,000–10,000 genome copies (Figure 4E). Spatial noise was generally comparable with 200,000–50,000 genome copies, but was notably higher in libraries prepared with 1,000 – 10,000 genome copies (Figure 4F). Altogether, these data demonstrate that simultaneous implementation of a double alternating primer design and *in situ* 3C allows for successful 5C using lower genome copies. The implication of these results is that 5C might be performed on smaller cell populations in future studies.

Discussion/Conclusions

The invention of the canonical 5C procedure by Dekker, Dostie and colleagues enabled the creation of high-resolution, high-coverage 3D genome folding maps from a subset of the genome (up to ~30 Mb) at a fraction of the cost of Hi-C [7, 9, 17, 27]. Due to the markedly reduced cost, 5C is poised to have high utility in addressing the significant unmet need of comprehensive inquiry of the folding of all fragments within a large genomic locus across hundreds of biological perturbation conditions. The ability to create loop resolution maps across thousands of gene editing perturbations is essential for testing the functional relationship between genome structure and function.

Despite these key advantages, canonical 5C has been limited in looping detection sensitivity and specificity due to the alternating primer design which only queries half the ligation junctions and non-specific ligations leading to a low signal to noise ratio. In the present study, we present an updated version of the classic 5C procedure, 5C *in situ* double alternating (5C-ID). 5C-ID implements a double alternating primer design [17] and *in situ* 3C [3, 32], resulting in markedly increased sensitivity for looping signal detection and reduced off-target non-specific ligation junctions. Double alternating primers comprehensively bind to all possible ligation junctions [17] missed by the single alternating design [7, 22–26, 31], leading to markedly improved loop detection sensitivity. Moreover, by conducting restriction digestion and ligation steps of 3C *in situ* in the nucleus, we dramatically reduced spatial noise caused by known non-specific ligations from classic dilution 3C [2, 3, 28–30, 32]. By combining these two changes, we were also able to maintain loop detection sensitivity with reduced genome copies. While canonical 5C was performed on 40–100 million cells [9, 15, 37], only 2 million cells were used here for 5C-ID. We observe that genome copies can be decreased to 50,000, which suggests pellets of pellets of ~25,000–50,000 cells or possibly less are possible in the future for high quality 5C maps.

Thus, 5C-ID creates high-coverage, high-quality heatmaps at loop resolution at a fraction of the cost and opens the future potential for low cell number analysis from rare cell types and human tissues.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Job Dekker and Bryan Lajoie for critical discussions, providing access to the My5C software, and assistance in designing double alternating 5C primers. J. Phillips-Cremins is a New York Stem Cell Foundation (NYSCF) Robertson Investigator and an Alfred P. Sloan Foundation Fellow. This work was funded by The New York Stem Cell Foundation (J. Phillips-Cremins), the Alfred P. Sloan Foundation (J. Phillips-Cremins), the NIH Director's New Innovator Award (1DP2MH11024701; J. Phillips-Cremins), a 4D Nucleome Common Fund grant (1U01HL12999801; J. Phillips-Cremins), and a joint NSF-NIGMS grant to support research at the interface of the biological and mathematical sciences (1562665; J. Phillips-Cremins).

References

1. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*. 2001; 2:292.
2. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragooczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. 2009; 326(5950):289–293. [PubMed: 19815776]
3. Rao, Suhas SP., Huntley, Miriam H., Durand, Neva C., Stamenova, Elena K., Bochkov, Ivan D., Robinson, James T., Sanborn, Adrian L., Machol, I., Omer, Arina D., Lander, Eric S., Aiden, Erez L. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. 2014; 159(7):1665–1680. [PubMed: 25497547]
4. Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, Barbieri M, Moore BL, Kraemer DC, Aitken S, Xie SQ, Morris KJ, Itoh M, Kawaji H, Jaeger I, Hayashizaki Y, Carninci P, Forrest AR, Semple CA, Dostie J, Pombo A, Nicodemi M. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular Systems Biology*. 2015; 11(12)
5. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012; 485(7398): 376–380. [PubMed: 22495300]
6. Hou C, Li L, Qin Zhaohui S, Corces Victor G. Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Molecular Cell*. 2012; 48(3):471–484. [PubMed: 23041285]
7. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012; 485:381. [PubMed: 22495304]
8. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell*. 2012; 148(3):458–472. [PubMed: 22265598]
9. Phillips-Cremins, Jennifer E., Sauria, Michael EG., Sanyal, A., Gerasimova, Tatiana I., Lajoie, Bryan R., Bell, Joshua SK., Ong, C-T., Hookway, Tracy A., Guo, C., Sun, Y., Bland, Michael J., Wagstaff, W., Dalton, S., McDevitt, Todd C., Sen, R., Dekker, J., Taylor, J., Corces, Victor G. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell*. 2013; 153(6):1281–1295. [PubMed: 23706625]

10. Chambeyron S, Bickmore WA. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes & Development*. 2004; 18(10):1119–1130. [PubMed: 15155579]
11. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*. 2015; 112(47):E6456–E6465.
12. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010; 467(7314):430–435. [PubMed: 20720539]
13. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013; 503(7475):290–4. [PubMed: 24141950]
14. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, Thurman RE, Cheng Y, Gulsoy G, Dennis JH, Snyder MP, Stamatoyannopoulos JA, Taylor J, Hardison RC, Kahveci T, Ren B, Gilbert DM. Topologically associating domains are stable units of replication-timing regulation. *Nature*. 2014; 515(7527):402–405. [PubMed: 25409831]
15. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012; 489:109. [PubMed: 22955621]
16. Downen, Jill M., Fan, Zi P., Hnisz, D., Ren, G., Abraham, Brian J., Zhang, Lyndon N., Weintraub, Abraham S., Schuijers, J., Lee, Tong I., Zhao, K., Young, Richard A. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell*. 2014; 159(2):374–387. [PubMed: 25303531]
17. Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, Reddy J, Borges-Rivera D, Lee TI, Jaenisch R, Porteus MH, Dekker J, Young RA. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016; 351(6280):1454–1458. [PubMed: 26940867]
18. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin David U, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang Michael Q, Ren B, Krainer Adrian R, Maniatis T, Wu Q. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. 2015; 162(4):900–910. [PubMed: 26276636]
19. Lupiáñez, Darío G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, John M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, Stefan A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., Mundlos, S. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*. 2015; 161(5):1012–1025. [PubMed: 25959774]
20. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2015; 529:110. [PubMed: 26700815]
21. Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, Brummelkamp TR, de Wit E, Rowland BD. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*. 2017; 169(4):693–707.e14. [PubMed: 28475897]
22. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*. 2006; 16(10):1299–1309. [PubMed: 16954542]
23. Dostie J, Zhan, Y., Dekker, J. *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.; 2001. Chromosome Conformation Capture Carbon Copy Technology.
24. Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols*. 2007; 2:988. [PubMed: 17446898]

25. Ferraiuolo MA, Sanyal A, Naumova N, Dekker J, Dostie J. From cells to chromatin: Capturing snapshots of genome organization with 5C technology. *Methods*. 2012; 58(3):255–267. [PubMed: 23137922]
26. van Berkum, NL., Dekker, J. Determining Spatial Chromatin Organization of Large Genomic Regions Using 5C Technology. In: Collas, P., editor. *Chromatin Immunoprecipitation Assays: Methods and Protocols*. Humana Press; Totowa, NJ: 2009. p. 189-213.
27. Beagan, Jonathan A., Gilgenast, Thomas G., Kim, J., Plona, Z., Norton, Heidi K., Hu, G., Hsu, Sarah C., Shields, Emily J., Lyu, X., Apostolou, E., Hochedlinger, K., Corces, Victor G., Dekker, J., Phillips-Cremins, Jennifer E. Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell*. 2016; 18(5):611–624. [PubMed: 27152443]
28. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing Chromosome Conformation. *Science*. 2002; 295(5558):1306–1311. [PubMed: 11847345]
29. Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature Genetics*. 2011; 43:1059. [PubMed: 22001755]
30. Gilgenast TG, Phillips-Cremins JE. Systematic evaluation of statistical methods for identifying looping interactions in 5C data. *bioRxiv*. 2017
31. Lajoie BR, van Berkum NL, Sanyal A, Dekker J. My5C: web tools for chromosome conformation capture studies. *Nature Methods*. 2009; 6:690. [PubMed: 19789528]
32. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013; 502(7469): 59–64. [PubMed: 24067610]
33. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods*. 2012; 58(3):221–30. [PubMed: 22609568]
34. Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, Lachanski CV, Gillis DR, Phillips-Cremins JE. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Research*. 2017
35. Li Y, Rivera CM, Ishii H, Jin F, Selvaraj S, Lee AY, Dixon JR, Ren B. CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells. *PLoS ONE*. 2014; 9(12):e114485. [PubMed: 25486255]
36. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-P, Tanay A, Cavalli G. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. 2017; 171(3):557–572.e24. [PubMed: 29053968]
37. Smith, Emily M., Lajoie, Bryan R., Jain, G., Dekker, J. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the *CFTR* Locus. *The American Journal of Human Genetics*. 98(1):185–201.

- We describe 5C-ID, an updated version of Chromosome-Conformation-Capture-Carbon-Copy (5C) with restriction digest and ligation performed in the nucleus (in situ Chromosome-Conformation-Capture (3C)) and ligation-mediated amplification performed with a new double alternating design.
- 5C-ID reduces spatial noise and enables higher resolution 3D genome folding maps than canonical 5C in up to a 30 Megabase subset of the genome at a fraction of the cost of Hi-C.
- 5C-ID allows for a marked improvement in sensitivity and specificity of loop detection.

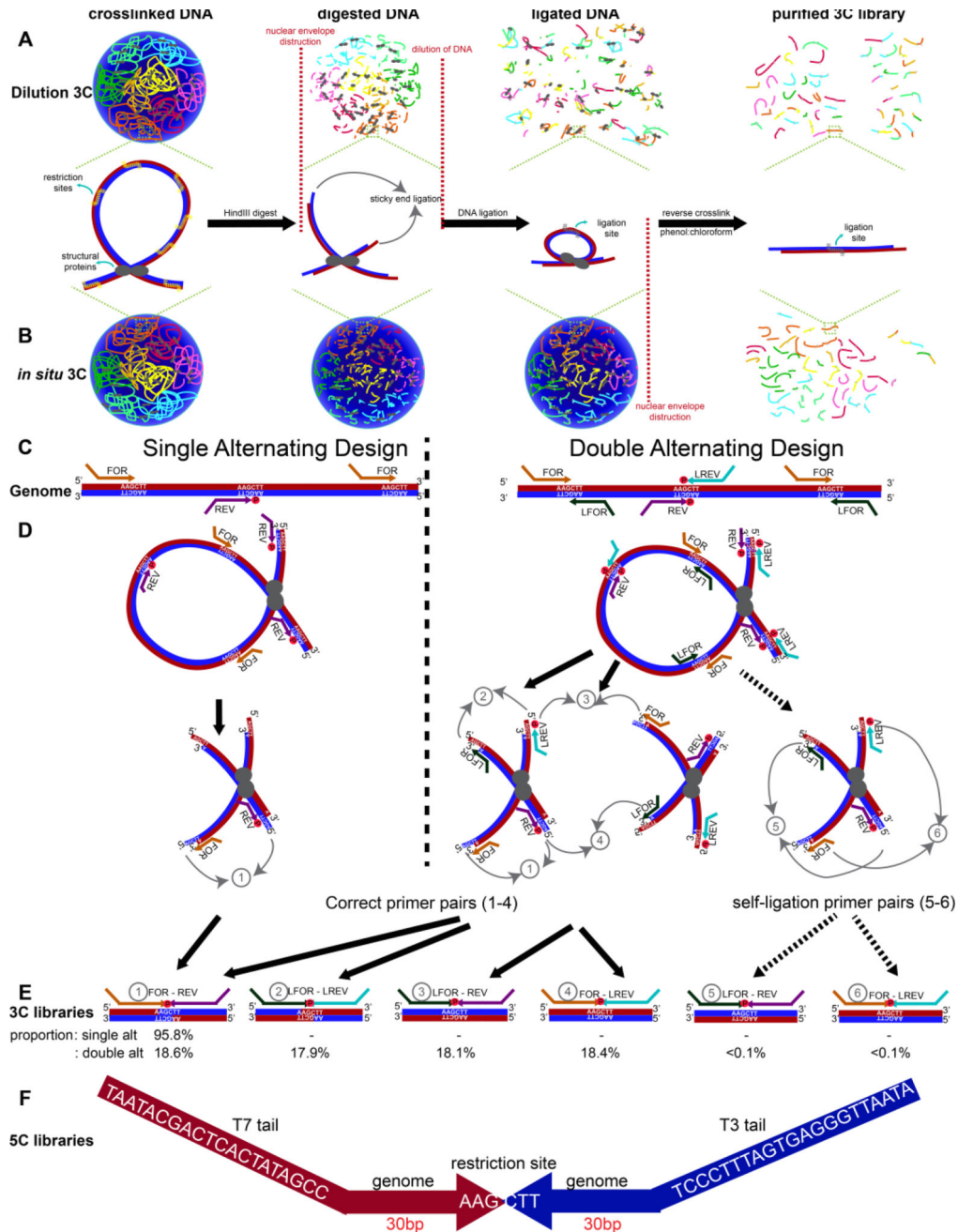


Figure 1. Overview of 5C-ID compared to canonical Chromosome-Conformation-Capture-Carbon-Copy (5C)

(A) Schematic of dilution 3C. After chromatin fixation, the cellular and nuclear membranes are disrupted and chromatin is digested with a restriction enzyme in solution. After digestion, sticky ends are subsequently ligated under dilute conditions that promote intramolecular ligation. (B) Schematic of *in situ* 3C. Unlike dilution 3C, chromatin digestion and ligation are performed *in situ* within intact nuclei. Chromatin colors illustrate DNA from independent chromosome territories. Vertical grey lines on DNA fragments illustrate ligation junctions. (C) Schematic illustrating single and double alternating 5C primer designs. Red and blue lines indicate sense and antisense DNA strands. (D–E) Schematic illustrating all

possible correct 5C primer ligations from single and double alternating designs (1–4) as well as artifactual self-circle ligations between two ends of the same fragment (5–6). Percentages of each possible primer-primer pair orientation observed from a recent 5C experiment are provided. **(F)** Schematic illustrating the 5C amplicon from ligation-mediated amplification, including the T7/T3 universal tails, the 30 base pairs that uniquely bind to the genomic DNA and the re-ligated HindIII restriction sites.

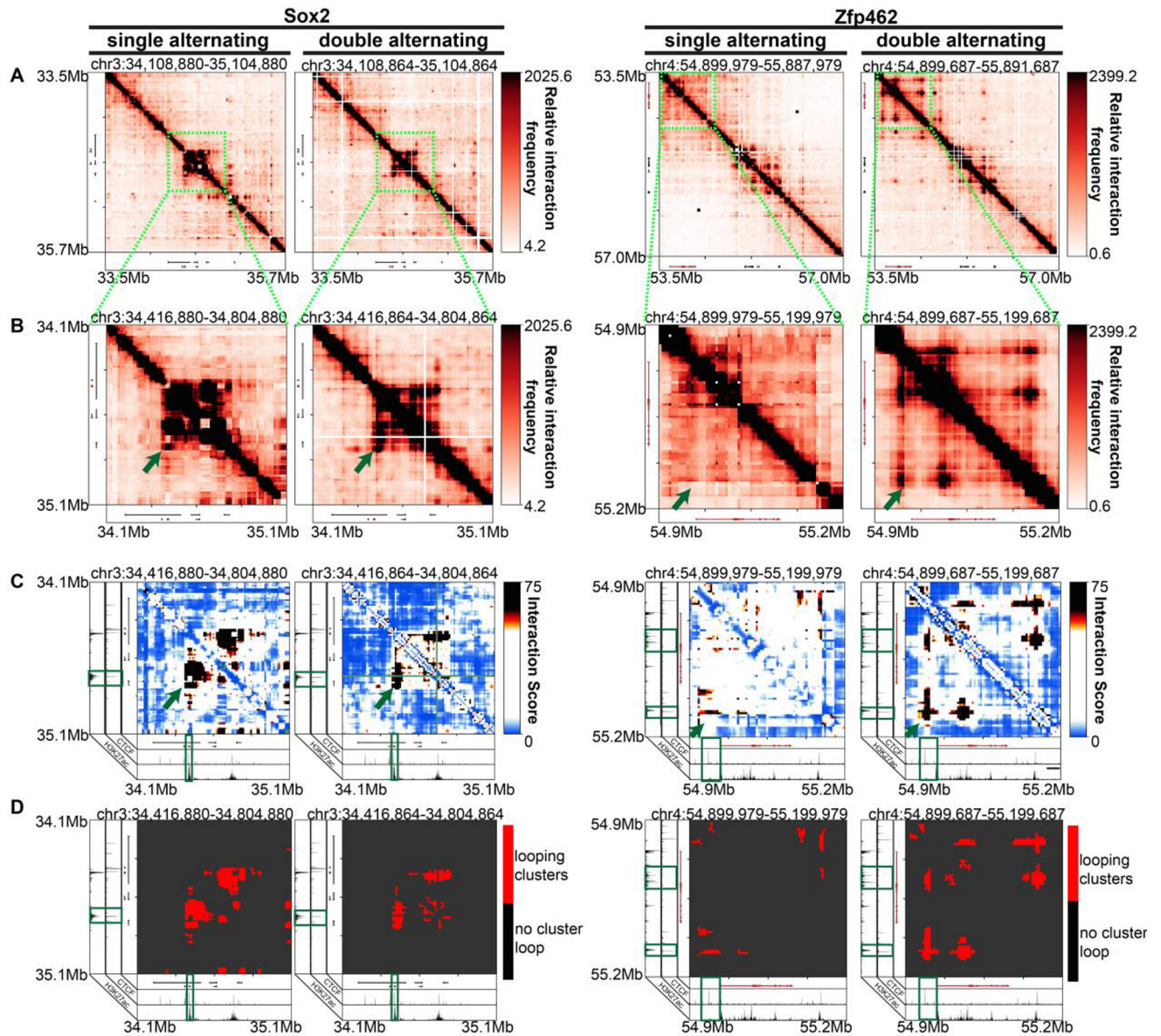


Figure 2. Improved 5C loop detection specificity with double vs. single alternating primer design
(A) Heatmaps binned at 4 kb matrix resolution showing relative chromatin interaction frequencies in 1 Mb regions surrounding *Sox2* and *Zfp462* genes across single alternating and double alternating primer designs in embryonic stem cells in 2i media. Genes of interest are highlighted in red. **(B–D)** Zoomed-in heatmaps highlighting *Sox2* and *Zfp462* interactions with published pluripotency-specific enhancers. **(B)** Relative 5C interaction frequency after sequencing depth correction, binning and matrix balancing. **(C)** Interaction scores after distance-dependence and local background expectation correction and modeling. **(D)** Long-range looping interaction clusters after thresholding on interaction scores. ChIPseq tracks for CTCF and H3K27ac from embryonic stem cells are overlaid over the maps.

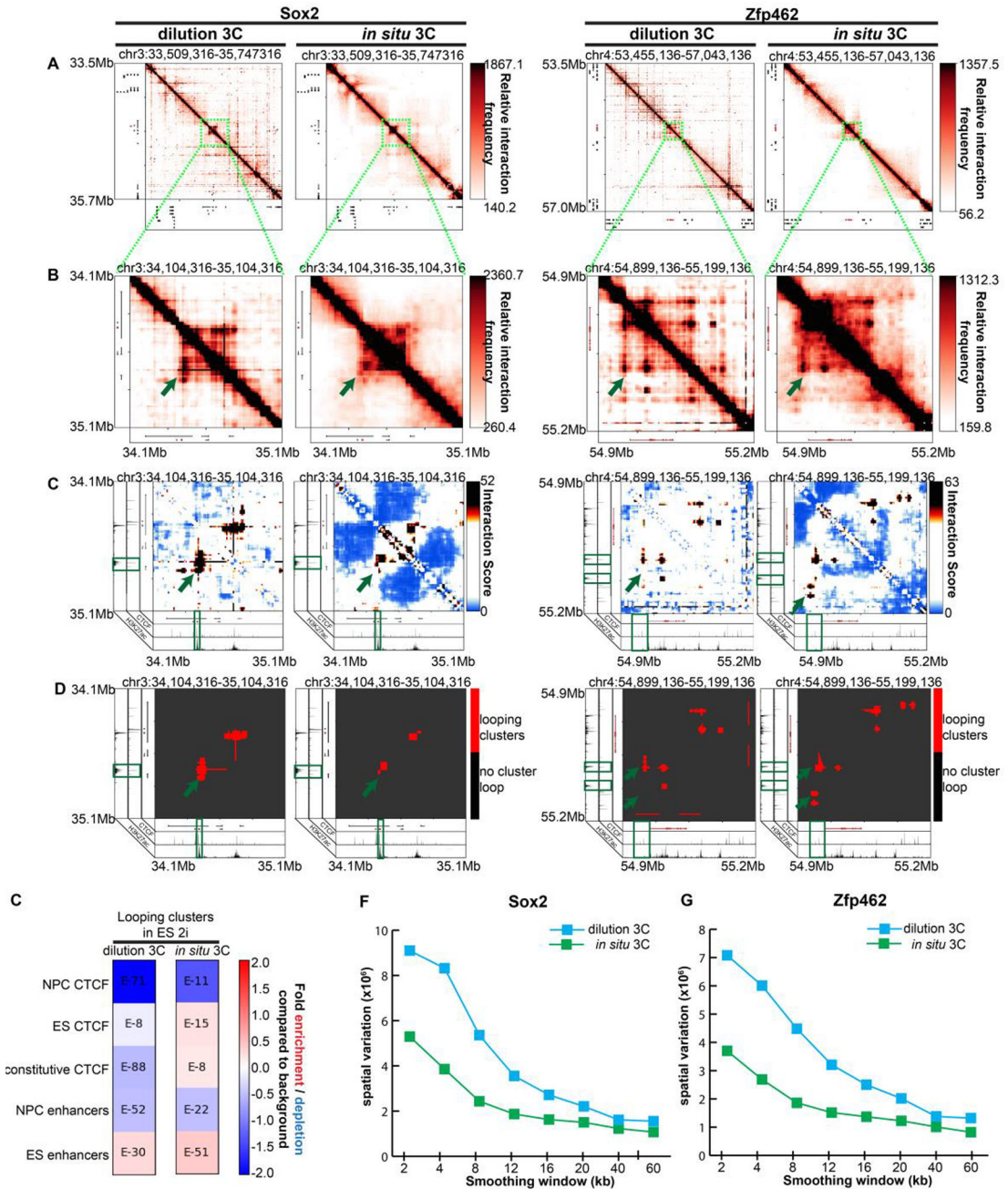


Figure 3. In situ 3C reduces spatial noise due to non-specific ligation products caused by dilution 3C

(A) Heatmaps binned at 8 kb matrix resolution showing relative chromatin interaction frequencies in 1 Mb regions surrounding *Sox2* and *Zfp462* genes. 5C libraries were created from *in situ* and dilution 3C templates from embryonic stem (ES) cells cultured in 2i media. Genes of interest are highlighted in red. (B–D) Zoomed-in heatmaps highlighting *Sox2* and *Zfp462* interactions with published pluripotency-specific enhancers. (B) Relative 5C interaction frequency after sequencing depth correction, binning and matrix balancing. (C) Interaction scores after distance-dependence and local background expectation correction

and modeling. **(D)** Long-range looping interaction clusters after thresholding on interaction scores. ChIPseq tracks for CTCF and H3K27ac from embryonic stem cells are overlaid over maps. **(E)** Enrichment of chromatin features at classified looping interactions relative to background interactions in 5C libraries using dilution vs. in situ 3C and a double alternating 5C primer design. P-values are calculated using Fisher's exact test. **(F–G)** Spatial variance of binned contact matrices around **(F)** *Sox2* and **(G)** *Zfp462* genes as a function of smoothing window size.

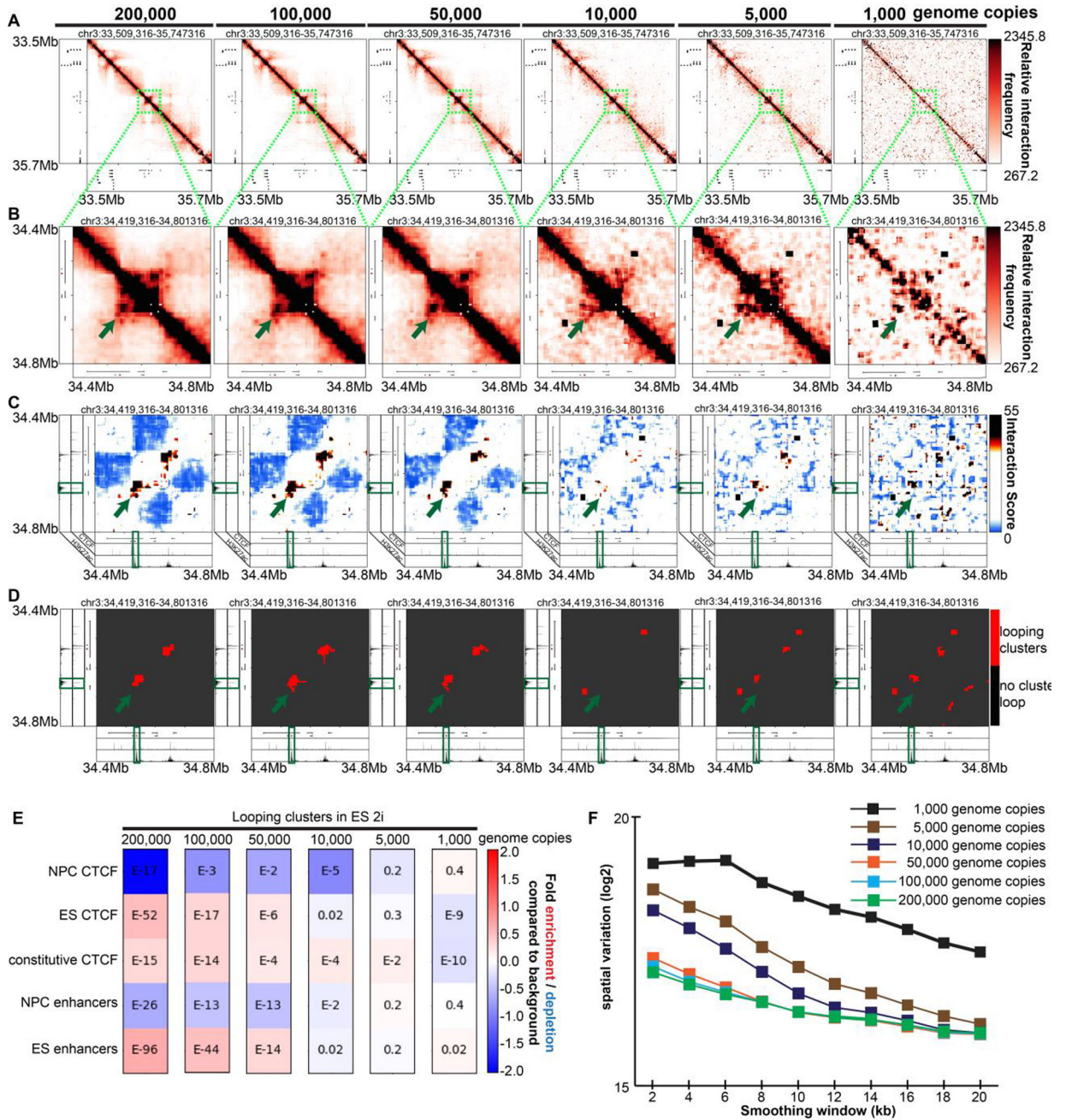


Figure 4. Combined usage of an in situ 3C template and double alternating primer design lowers the number of 3C template genome copies required to produce high quality 5C data (A) Heatmaps binned at 8 kb matrix resolution showing relative chromatin interaction frequencies in 1 Mb regions surrounding the *Sox2* gene. 5C libraries were created from a double alternating design and *in situ* 3C templates (200,000, 100,000, 50,000, 5,000, 1,000 genome copies) from embryonic stem (ES) cells cultured in 2i media. Genes of interest are highlighted in red. (B–D) Zoomed-in heatmaps highlighting *Sox2* interactions with published pluripotency-specific enhancers. (B) Relative 5C interaction frequency after sequencing depth correction, binning and matrix balancing. (C) Interaction scores after

distance-dependence and local background expectation correction and modeling. **(D)** Long-range looping interaction clusters after thresholding on interaction scores. ChIPseq tracks for CTCF and H3K27ac from embryonic stem cells are overlaid over maps. **(E)** Enrichment of chromatin features at classified looping interactions relative to background interactions in 5C libraries using in situ 3C and a double alternating 5C primer design for a range of genome copy numbers. P-values are calculated using Fisher's exact test. **(F)** Spatial variance of binned contact matrices at various genome copies as a function of smoothing window size. **(G)** Looping cluster numbers in the Sox2 region as a function of genome copy number.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript