Ectopic expression of meiotic cohesin generates chromosome instability in cancer cell line
Abdelhalim Boukaba, Jian Liu, Carl Ward, Qiongfang Wu, Alexei Arnaoutov, Jierong Liang, Elena M. Pugacheva, Mary Dasso, Victor Lobanenkov, Miguel Esteban, and Alexander V. Strunnikov
Edited by Douglas Koshland, University of California, Berkeley, CA; received March 10, 2022; accepted August 23, 2022

Many tumors express meiotic genes that could potentially drive somatic chromosome instability. While germline cohesin subunits SMC1B, STAG3, and REC8 are widely expressed in many cancers, messenger RNA and protein for RAD21L subunit are expressed at very low levels. To elucidate the potential of meiotic cohesins to contribute to genome instability, their expression was investigated in human cell lines, predominately in DLD-1. While the induction of the REC8 complex resulted in a mild mitotic phenotype, the expression of the RAD21L complex produced an arrested but viable cell pool, thus providing a source of DNA damage, mitotic chromosome missegregation, sporadic polyploidy, and altered gene expression. We also found that genomic binding profiles of ectopically expressed meiotic cohesin complexes were reminiscent of their corresponding specific binding patterns in testis. Furthermore, meiotic cohesins were found to localize to the same sites as BORIS/CTCFL, rather than CTCF sites normally associated with the somatic cohesin complex. These findings highlight the existence of a germ-line epigenomic memory that is conserved in cells that normally do not express meiotic genes. Our results reveal a mechanism of action by unduly expressed meiotic cohesins that potentially links them to aneuploidy and chromosomal mutations in affected cells.

chromosome instability | chromatin | centromere | cancer-testis genes | condensin

It is well established that a substantial number of malignant tumors express germline-specific genes (1–5). Among those, collectively named cancer–testis (CT) factors (6), there are some potent regulators of gene expression and chromatin structure. However, the role of such proteins in tumor biology is not well-studied and thus presently could be underestimated (7). Nevertheless, reports of functional dependence of cancer cells on activated CT genes are steadily accumulating (3, 8–14), suggesting that some CT genes could play roles in cancer onset rather than being merely “passengers” of tumor development. Furthermore, the somatic expression of germline chromatin proteins is conceptually similar to the epigenetic deregulation, with a novel epigenomic landscape generated not by a change of DNA or histone modification but by the emergence of unscheduled chromatin components. This phenomenon is not widely discussed in the fields of epigenetic therapy (15) or chromosome instability in cancer (16), even though such a paradigm might uncover some early drivers of genetic alterations in tumors (17–19). Indeed, it is plausible that some germline-specific genes could induce chromosomal instability (20–22), such as chromothripsis (23–29), but experimental evidence for such a hypothesis was lacking.

The subunits of meiotic cohesin complexes (mei-CC) (30, 31) are of particular interest when activated as CT factors, because their roles in sister chromatid cohesion (SCC) and chromosome pairing in germline are dramatically distinct from the functions of somatic cohesin (32). Nevertheless, the transcriptional activation in tumors is rather common for some mei-CC subunits, which is difficult to reconcile with their toxicity observed in lower eukaryotes (33–36). In the mammalian germline, at least two distinct types of meiosis-specific cohesin complexes, i.e., REC8-based and RAD21L-based, have been reported (37–43). REC8 is present in all eukaryotes and appears to operate in a canonical SCC role: to hold sister chromatids together, especially at centromeres, and to counteract nonhomologous pairing in mouse germline (43, 44). In contrast, RAD21L has evolved only in Metazoa. Genetic studies on mice pointed to a putative role of RAD21L as a facilitator of the initial contacts between homologs (34, 45) and the formation of chromosome axis (44, 46). Cytological data indicate that REC8 loading precedes DNA replication in meiosis I and persists through the metaphase II, while RAD21L is loaded after DNA is replicated and is largely removed before the anaphase I (38). Genetic data show that these two complexes are only partially redundant (40) but are functionally cooperative (44), driving pairing and segregation of meiotic chromosomes (32, 40–42, 46–56).

Significance
This work originated from mining of cancer genome data and proceeded to analyze the effects of ectopic expression of meiotic cohesins in mitotic cells in culture. In the process, apart from conclusively answering the question on mechanisms for RAD21L toxicity and its underrepresentation in tumor transcriptomes, we found an association of meiotic cohesin binding with BORIS/CTCFL sites in the normal testis. We also elucidated the patterns and outcomes of meiotic cohesin binding to chromosomes in model cell lines. Furthermore, we uncovered that RAD21L-based meiotic cohesin possesses a self-contained chromosome restructuring activity able to trigger sustainable but imperfect mitotic arrest leading to chromosomal instability. The discovered epigenomic and genetic mechanisms can be relevant to chromosome instability in cancer.
In contrast to the normal germline, the functions of human mei-CC proteins in tumors were not studied in detail (57), despite their prognostic potential (58). Here, based on the interrogation of cancer genomes data, we established a model to study meiotic cohesins outside the constraints of their germ cell environment, i.e., in cancer cell lines, primarily DLD-1, and supplemented that with epigenomic analysis on a normal non-human primate testis. Our results indicate that mei-CCs, when reconstituted in the model system, affected gene expression in a distinct fashion, as well as manifested drastically different phenotypes with respect to cell proliferation and chromosome segregation. In particular, the ectopic RAD21L mei-CC expression resulted in a prolonged cell-cycle arrest with a potential to generate severe chromosomal abnormalities, including DNA damage and occasional formation of multistrand chromosomes. In this ectopic system, as well as in primate testes, the REC8 and RAD21L mei-CCs predominately localize to the sites that are occupied by BORIS/CTCFL in the normal germline. Our findings offer a plausible explanation for the low-frequency expression of RAD21L in tumors and demonstrate that experimentally activated CT genes can rapidly impair genome integrity, serving as a potent source of chromosomal instability.

**Results**

**Mei-Cohesin Subunits Display Expression Bias as CT Genes.** To make inroads into assessing potential roles of CT genes outside of germline, we first focused on the expression of all core cohesin components (Fig. 1A) and the associated cofactors in tumors. TCGA (The Cancer Genome Atlas) RNA-sequencing (RNA-seq) dataset analysis shows that the individual mei-CC subunits form two separate patterns: SMC1B and STAG3 cluster with a number of somatic cohesin components, while RAD21L and REC8 show no correlation with other interrogated genes.

![Fig. 1. The overexpression and coexpression of mei-cohesin subunits in tumors. (A) A schematic of somatic cohesin complexes compared to mei-CC. (B) Multivariate correlation of human cohesin subunits’ expression and associated factors in 9,561 samples from TCGA RNA-seq (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga). Heat-map fragment shows mei-CC subunits only, with corresponding correlations in boxes and TPM statistical parameters on the left. (C) Scatterplot of relative mei-CC subunits’ expression in TCGA. Only nongermline tumors with at least one gene expressed at TPM ≥ 100 are included. (D and E) Parallel coordinates plots from cluster analysis of nongermline tumors for RAD21L (D) and REC8 (E) coexpression with SMC1B and STAG3. Only clusters bigger than 10 are shown. The corresponding mean TPM expression values are indicated below. RAD21 is shown for each cluster as a somatic paralogue reference.](https://doi.org/10.1073/pnas.2204071119)
(Fig. 1B). The mei-CC genes have a wide range of expression levels in different tumors, except RAD21L, for which both the incidence of activation and expression levels are the lowest (Fig. 1B and C) and Supplementary Information Text. In tumors of nongermline origin, cluster analysis of coexpression with SMCI1B and STAG3 revealed that RAD21L expression was either lower than both SMCI1B and STAG3 (Fig. 1D, cluster 1) or, more commonly, was practically undetectable when SMCI1B and/or STAG3 were highly expressed (Fig. 1D, clusters 2 and 3). In the case of REC8 such a bias was not evident (Fig. 1E).

Thus, we detected some notable deviations from a presumable sporadic mode of activation of mei-CC subunits in tumors. While the expression of neither REC8 nor RAD21L correlated with other mei-CC genes, RAD21L had orders of magnitude lower incidence of activation, as well as lower levels of expression, compared to all other subunits, prevailing through all tumor types (Supplementary Information Text). This parallels the low level of RAD21L messenger RNA detection in normal somatic cells, where other mei-CC subunits are occasionally expressed in a fraction of tissues (Supplementary Information Text) but, as a whole, at substantially lower levels than could be achieved in tumors. This drastically differs from normal germline (Supplementary Information Text) but, as a whole, at substantially lower levels than could be achieved in tumors. Therefore, we set out to investigate the underlying reasons of observed expression biases by expressing human mei-CC subunits in immortalized cell lines. We generated constructs to stably express ectopic expression of the transgenes (SMCI1B 4,495.4 ± 3,145.1, REC8 40.3 ± 36.5, and RAD21L 0.0 Transcripts per Million (TPM), with endogenous transcription levels in IL18. DLD-1SMC1B-STAG3 cells had 45 DEGs, with only a mild increase in IL18. DLD-1SMC1B-STAG3-REC8 cells had no differentially expressed genes (DEGs), except a mild increase in IL18. DLD-1SMC1B-STAG3-RAD21L (Fig. 3 D and E) showed that the mei-CC complexes were indeed assembled (Fig. S3B). The expression of mei-CC subunits in the DLD-1SMC1B-STAG3 cells was highly correlated with the expression of the transgenes (SMCI1B 4,495.4 ± 3,145.1, REC8 40.3 ± 36.5, and RAD21L 0.0 Transcripts per Million (TPM), with endogenous transcription levels in IL18. DLD-1SMC1B-STAG3-RAD21L cells had no differentially expressed genes (DEGs), except a mild increase in IL18. DLD-1SMC1B-STAG3-REC8 cells had 45 DEGs, with only EEEF2 notably overexpressed (Fig. 3E). RAD21L mei-CC cells had
the largest signature of 210 DEGs. Their comparison to the significantly affected preleptotene genes in mouse (63) revealed an overlap of 63 genes. Similarly, a comparison to a human germline development dataset (64) revealed a concordant increase for 11 genes: AREG, BMI1, HIST1H1C, HLA-B, MYADM, NPC2, OPTN, PLK2, PSMD1, SCML1, and SQSTM1. Overall, among transcription start sites (TSS) of genes differentially regulated in RAD21L mei-CC cells there was a good correspondence with ATAC-seq data (Fig. 3F).

Thus, it is likely that RAD21L mei-CC directly restructures chromosomes in a specific manner inducing the downstream cell-cycle phenotype. Still, we could not exclude a possibility that it also down-regulated some somatic chromosome components posttranscriptionally. However, proteome analysis demonstrated no dramatic protein level change among 6,889 detected proteins (Dataset S1) that could be associated with the phenotype of RAD21L mei-CC cells. For example, only two proteins, ANKRD22 and SULT1A3, both unrelated to chromosome function, were down-regulated more than twofold among relevant KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Particularly, apart from moderately decreased SMC1A, which is functionally interchangeable with SMC1B (51), the abundance of somatic cohesin, condensin, and associated factors was not significantly affected either (Fig. 3G).

The Specificity of DNA Binding by Ectopic Mei-Cohesin Reflects a Germline Pattern. In order to investigate the in-depth properties of chromatin that was exposed to mei-CC in nongermline cells we assessed mei-CC binding to chromatin, beginning with individual subunits. First, we focused on K562 cells, which express another Cancer Testis Antigen (CTA), BORIS, a meiotic partner of CTCF, known to immobilize somatic cohesion at a...
given site (65–67). While SMC1B showed a strong presence in fractionated chromatin, REC8, RAD21L, or STAG3 had only marginal chromatin association (SI Appendix, Fig. S4A). The follow-up by ChIP-on-ChEP (Chromatin Enriched for Proteins) seq-analysis confirmed that RAD21L and REC8 gave only negligible number of nonrepeat peaks, all with low scores, SMC1B, however, gave 28,740 nonrepeat summit peaks, and 4,916 peaks were also uncovered for STAG3. While SMC1B was largely coincident with CTCF binding (SI Appendix, Fig. S4B), possibly as a result of exchange with somatic SMC1A, STAG3 almost completely coincident with CTCF binding, with the strongest correlation to DNA in DLD-1 cells is reminiscent of germline patterns. In the absence of ChIP-seq for RAD21L and REC8, the revealed pattern of mei-CC binding, despite small differences, indicated that both ectopic REC8 and RAD21L mei-CC follow the general pathway of somatic cohesin binding, with the strongest correlation to DNA in DLD-1 cells is reminiscent of germline patterns. In the absence of ChIP-seq–grade human

In contrast, for triple transgenic mei-CC cell lines ChIP-on-ChEP-seq revealed a prominent pattern of cobinding of mei-CC followed the general pathway of somatic cohesin binding, with the strongest correlation to DNA in DLD-1 cells is reminiscent of germline patterns. In the absence of ChIP-seq for RAD21L and REC8, the revealed pattern of mei-CC binding, despite small differences, indicated that both ectopic REC8 and RAD21L mei-CC follow the general pathway of somatic cohesin binding, with the strongest correlation to DNA in DLD-1 cells is reminiscent of germline patterns. In the absence of ChIP-seq–grade human
testis material we investigated the pattern of mei-CC binding in the spermatogenesis of *Macaca fascicularis*. Recent single-cell RNA-seq studies documented uniqueness of gene expression programs in primate testis (64, 68, 69). *M. fascicularis* has reproductive biology similar to humans, a substantial collinearity to human genome, and high homology of its mei-CCs to human ones; 10,239 high-confidence peaks were mapped by ChIP-on-ChEP-seq for REC8, 7,844 for RAD21L, and 15,667 for STAG3 (Fig. 4A and SI Appendix, Fig. S5A). The STAG3 ChIP had a weaker enrichment with higher noise, while SMCM1 consistently gave low signal, with only 6,042 peaks identified in the best replicate. Therefore, the SMCM1 peaks were not analyzed further. A consistent cooccupancy between the three mei-CC subunits was evident (Fig. 4D and SI Appendix, Fig. S5A) and the strength of RAD21L and REC8 binding was well correlated (SI Appendix, Fig. S5B). The latter was most likely due to compounding of signal for the two mei-CC complexes considering single cell RNA sequencing (scRNA-seq) data (68, 69), immunofluorescence (IF) (Fig. 4E), and cross-linked chromatin co-IP experiments (SI Appendix, Fig. S5C).

We also employed ChIP-on-ChEP-seq to investigate mei-CC relationship with CTCF and BORIS, as scRNA-seq data from human and macaque testis indicated only partial coexpression of BORIS with mei-CC in the course of spermatogenesis, with CTCF expressed throughout (64, 68, 69); 21,165 high-confidence peaks for BORIS and 31,547 peaks for CTCF were identified, with 15,788 peaks that could be considered CTCF-only (SI Appendix, Fig. S6 A and B). Of BORIS peaks common to two samples 36.4% overlapped with TSS. In contrast, only 13.3% of CTCF peaks overlapped with TSS. Intriguingly, the mei-CC subunits’ peaks substantially coincided with BORIS binding (Fig. 4F and SI Appendix, Fig. S6C), and the strength of mei-CC and BORIS peaks was correlated, unlike with CTCF (Fig. 4G). Peak annotations indicated that mei-CC subunits were enriched in transcription control regions (SI Appendix, Fig. S7A), with corresponding Gene Ontology (GO) analysis showing a wide range of corresponding biological processes (SI Appendix, Fig. S7B).

Thus, BORIS binding could be a prerequisite for mei-CC positioning, as its expression precedes mei-CC emergence in the spermatogenesis (64, 68, 69). This is supported by our IF results (SI Appendix, Fig. S6D) and the proteomics (mouse protein atlas, https://www.ebi.ac.uk), which places the bulk of BORIS into spermatagonia. Our analysis of fluorescence-activated cell sorting (FACS)-purified cell fractions also indicated that the difficulty to isolate (70) premeiotic germline cells evidently generated the bulk of IF BORIS signal (SI Appendix, Fig. S6 E–G). Nevertheless, round spermatids and spermatocytes still showed a notable correspondence of mei-CC binding (analyzed for REC8) to BORIS sites (SI Appendix, Fig. S6H).
Available data indicate that BORIS functions predominately as a transcription factor (64, 71). Therefore, while it appears to mark sites of mei-CC binding, the exact molecular role of BORIS there is not clear, especially as mei-CCs are able to associate with chromatin in BORIS knockout (KO) mice (72).

To get more insight into this, we analyzed the landscape of H3K27ac binding in M. fascicularis testis. ChiP-on-ChEP-seq analysis showed 20,849 high-confidence gene-overlapping peaks, with the stronger ones substantially overlapping with both REC8 and RAD21L (SI Appendix, Fig. S8A). A much smaller number of H3K27ac sites that did not overlap with any genes, i.e., likely at enhancers, showed a similar pattern. However, a significant difference with BORIS was evident, with mei-CC peaks overlapping H3K27ac signal, but with distinct summits (SI Appendix, Fig. S8B). Therefore, it likely that BORIS coincides with mei-CC due to its engagement in chromatin remodeling related to transcription. This, in turn, could be a reason for a prominent correspondence of mei-CC, meiotic binding to the pattern seen in cell lines. Indeed, over 70% of REC8 and RAD21L peaks in these cell lines overlapped with their corresponding conserved sites in macaque testis (Fig. 4H).

The Ectopic Expression of Full Mei-Cohesin Complexes Reveals a Plausible Mechanism of RAD21L Toxicity and Underrepresentation in Tumors. As we established that binding of RAD21L mei-CC in tested cell lines is reminiscent of germline pattern, we investigated the phenotype induced by the ectopic RAD21L mei-CC in more detail. The RAD21L filament-like structures assembled along chromosomes in RAD21L mei-CC cells eventually spanned the entire length of individual chromosomes, from centromeres to telomeres, notably excluding ribosomal DNA (rDNA) chromatin, well before nuclear envelope breakdown (Fig. 5A). The corresponding chromosome condensation was not induced by mitotic commitment, as it did not correlate with phospho-S10-H3 signal (Fig. 5B). As the RAD21L-containing filaments were somewhat reminiscent of somatic cohesin deposition upon the inactivation of prophase pathway for cohesin removal (73, 74), we knocked down WAPL machinery in RAD21L mei-CC cells (Fig. 5C). The atypical condensation phenotype in RAD21L mei-CC cells appeared to be epistatic to prophase pathway (Fig. 5D).

The investigation of mei-CC cells presynchronized by double thymidine block revealed additional prominent defects in RAD21L mei-CC-expressing cells, such as stretched centromeres in interphase (Fig. 5E) and ubiquitous gamma-H2AX foci in anaphase cells, indicative of chromosome breaks (Fig. 5F). Also, the condensin complex, condensin I in particular, was not properly covering the length of chromosome arms (75) but instead was showing an interleaving pattern of binding with RAD21L (Fig. 5G). Synchronized cells revealed that the control, REC8, and RAD21L mei-CC cell lines did not have a major delay of S phase for two doubling times, according to BrdU FACS data (Fig. 5H). The synchronization (Fig. 5J) also showed that RAD21L mei-CC did not induce chromatin condensation before the beginning of S phase (0 h) but that the filaments appeared to fully assemble on condensed chromatin by the end of S phase (6 h) (SI Appendix, Fig. S9A), indicating that a passage through DNA synthesis was required for RAD21L mei-CC loading. Gamma-H2AX staining in the synchronous cell population indeed revealed that the initial loading of RAD21L coincided with replicating DNA (SI Appendix, Fig. S9B). Furthermore, after the completion of S phase the centromeres were unusually stretched or improperly condensed before nuclear envelope broke down (SI Appendix, Fig. S9C). The kinetic of following mitosis was, however, drastically different for REC8 and RAD21L mei-CC cell lines: While REC8 cells had a delayed anaphase, consistent with a kinetochore defect (Fig. 2), RAD21L cells not only had a delay entering mitosis but also largely failed to complete the passage through mitosis altogether, with only about one fifth of population initiating chromosome segregation at the 12-h time point (Fig. 5F). In RAD21L mei-CC cells, the dynamic of cyclin B1 (Fig. 5J) was also consistent with a mitotic exit block. While some RAD21L cells did reach metaphase, with the bulk of RAD21L leaving the condensed chromosomes (Fig. 5F and SI Appendix, Fig. S9D), all the detected anaphases had either chromosomal bridges or multiple gamma-H2AX foci, or both (SI Appendix, Fig. S9 B and E). The rest of the cell population appeared to be blocked in G2-M-like state and maintained high viability, above 80%, in the virtual absence of proliferation.

In order to determine the execution point of RAD21L mei-CC-induced arrest, cells were synchronized by an alternative regimen: first prearrested in S phase and then released into a G2 block with simultaneous induction of RAD21L mei-CC (SI Appendix, Fig. S9F). As the ectopic proteins levels peaked 4 h after induction, this approach enabled cells to reach the end of S phase before encountering the full expression of RAD21L mei-CC. These cells did not show interphase condensation induced by RAD21L cohesin (SI Appendix, Fig. S9G) and had no apparent delay in entering metaphase (SI Appendix, Fig. S9H). However, at later points these cells showed signs of a defective mitotic completion—prolonged chromosome condensation, collapsed or asymmetrical spindles, and centromeres trapped at the division furrow (SI Appendix, Fig. S9I)—while control cells had exited mitosis by that point (SI Appendix, Fig. S9J) and had the expected centromere counts per cells (SI Appendix, Fig. S9K). These data suggest that RAD21L cohesin interferes with centromere function within mitosis, even when abnormal S-phase condensation is avoided.

An additional observation suggested that the RAD21L mei-CC filament assembly has a meiotic-like physiological role. Namely, a small fraction of cells (about 1/250) expressing ectopic RAD21L mei-CC had chromosomes that were substantially thicker than normal. While such cells had interphase-like nuclei and microtubule arrangement, IF revealed either a paired appearance for each RAD21L filament (Fig. 5A), similar to condensin cores in normal metaphases (76), or multiple DNA strands aligned in a parallel fashion (Fig. 5L), as well as more than two aligned centromeres (Fig. 5M). Such cells contained on average 145 CREST spots, which is more than double the number expected for this quasi-diploid cell line, i.e., likely became tetraploid. This indicates that RAD21L cohesin has an ability to promote formation of multistrand chromosomes and to keep multiple chromatids aligned together, suggesting that this complex maintains, and likely initiates, chromosome pairing in the absence of meiotic factors.

It was also of substantial interest to understand what happens to RAD21L cohesin cells that do not complete mitosis, i.e., a majority in the arrested population. To address that we conducted long-term time-course experiments, where induced RAD21L mei-CC cells were passaged for nearly 2 wk and then released from the induction (SI Appendix, Fig. S10A). Under that protocol, most arrested cells showed RAD21L-stained condensed chromosomes in the nucleus; however, 28% of cells had micronuclei (day 9; SI Appendix, Fig. S10B). The latter is a probable indication of abortive mitosis still proceeding in a fraction of this population, serving as an “engine” producing genetically altered cells and explaining the lower than 100% viability (SI Appendix, Fig. S10A). Upon removal of mei-CC
induction, the population had returned to normal viability levels and resumed proliferation (SI Appendix, Fig. S10A). This result might be viewed as a model of a process inducing chromosome instability in normal human cells, i.e. hypothetical RAD21L might be viewed as a model of a process inducing chromosome

Discussion

A Low Incidence of CT Expression in Cancers as a Putative Sign of Their Biological Impact. This study was based on a hypothesis on potential roles of CT genes in chromosome instability, which is frequently associated with early tumorigenesis. Namely, we uncovered a surprisingly low incidence of RAD21L activation and its low expression in human tumors, a notable outlier with respect to wide ranges of CT genes’ expression in cancers (2, 3). As a result of the investigation into the potential mechanisms behind that phenomenon, both in primate testis and in human cells in vitro, we elucidated genome-wide processes that likely make RAD21L expression undesirable for tumor-cell proliferation.

The bulk of information on mei-CC interactions with chromatin in germline came from studies in mice, which are substantially different from primates in their spermatogenesis (77–80) and the corresponding gene expression program (68, 69, 81). While the RAD21L mei-CC is not required for SCC per se (56), some functional overlap exists between the REC8 and RAD21L mei-CCs (42), and there is a phase of physical cohabitation between the two complexes, e.g., from leptotene to pachytene (38). Our ChIP-on-ChIP-seq data from macaque testes shows an alternating pattern of RAD21L, SMC2, and NCAPG in compacted mitotic chromosomes. Insets are at 4x magnification. Control is DLD-1.

The ultimate positioning of somatic cohesin in the epigenome is believed to be determined by CTCF (65, 67, 82–93) and WAPL (94). The sites of CTCF and cohesin cohabitation (66, 80) and the corresponding gene expression program (68, 88) underrepresent peaks from densely packed chromatin (99, 100). However, it

Fig. 5. Cell cycle and chromosomal phenotypes due to ectopic mei-CC expression. (A) IF of markers for nucleolar chromatin (UBTF), centromere (CENPA), and telomere (TERF) for DLD-1SMC1B-STAG3-REC8. (B) IF staining for phospho-S10 H3 histone epitope in DLD-1SMC1B-STAG3-RAD21L cells. (C) Microscopy of DLD-1RAD21L treated with siRNA against WAPL, PDS5A, and PDS5B. The examples of cell type categories are marked: CC, condensed chromosomes; IC, intermediate condensation; DC, decondensed chromatin; M, mitotic cells. (D) Quantification of DLD-1RAD21L treated with siRNA against WAPL, PDS5A, and PDS5B, compared to DAPI-stained DLD-1SMC1B-STAG3-RAD21L cells. Classes are as in C, A, apoptotic-like cell. (E) IF of DLD-1RAD21L cells immediately after the release from the second thymidine block stained with anti-RAD21L and CREST Abs. Arrows point to stretched centromeres (Inset). Figures E–G and K–M are flattened confocal images, with deconvolution. (F) DLD-1SMC1B-STAG3-RAD21L, cells, stained with anti-γH2AX and RAD21L Abs. Arrows indicate γH2AX signal persisting in mitosis. (G) RAD21L mei-CC interferes with condensin binding in mitotic chromosomes. IF of DLD-1SMC1B-STAG3-RAD21L cells shows an alternating pattern of RAD21L, SMC2, and NCAPG in compacted mitotic chromosomes. Insets are at 4x magnification. Control is DLD-1.

The bulk of information on mei-CC interactions with chromatin in germline came from studies in mice, which are frequently associated with early tumorigenesis. Namely, we uncovered a surprisingly low incidence of RAD21L activation and its low expression in human tumors, a notable outlier with respect to wide ranges of CT genes’ expression in cancers (2, 3). As a result of the investigation into the potential mechanisms behind that phenomenon, both in primate testis and in human cells in vitro, we elucidated genome-wide processes that likely make RAD21L expression undesirable for tumor-cell proliferation.

The ultimate positioning of somatic cohesin in the epigenome is believed to be determined by CTCF (65, 67, 82–93) and WAPL (94). The sites of CTCF and cohesin cohabitation (66, 80) contribute synergistically to the formation and maintenance of CTCF-anchored loops and TAD boundaries (85, 88). Nevertheless, a notable fraction of mouse testis mei-CC sites does not coincide with CTCF (80), which could be partially due to limitations of a traditional ChIP technique (96–98) underrepresenting peaks from densely packed chromatin (99, 100). However, it
may also indicate a participation of another factor specific for spermatogenesis, such as BORIS. While BORIS is unable to position somatic cohesin (67), it might function to preset mei-cohesin patterns in leptotene in a physiologically relevant environment and with its maximal expression (101). Furthermore, only a handful of testis genes responded differentially to BORIS KO in mice (72, 102, 103), suggesting that BORIS has an additional role. Accordingly, in macaque germline, we observed that BORIS-bound sites were more numerous than in BORIS-expressing cell lines and detected a substantially more evenly distributed H3K27ac enrichment has revealed a substantial overlap between mei-CC sites and active chromatin, reinforcing the possibility that BORIS presets sites for mei-CC binding.

In K562 cells, SMC1B had a significant cointinding with CTCF sites, while STAG3 was preferentially found at open chromatin sites, including TSS, in all three tested cell lines: K562, DLD-1, and RPE-1. As STAG3 does have a predicted CTCF-interacting motif but no known DNA contacts, we could hypothesize that some cohesin-loading factors could be involved (67, 94). REC8 and RAD21L were unable to bind chromatin on their own or directly compete with the bulk of RAD21 of somatic cohesin (Fig. 2B), even though in late pachytene RAD21 apparently displaces RAD21L (38, 39), which also has a CTCF-interacting-like signature S\textsuperscript{S80}KVIHKQL. Therefore, RAD21L expression without mei-CC partners could still be under negative selection in soma.

RAD21L expression in tumors also stands out compared to other mei-CC genes, as it does not correlate with the expression of any human transcription factors/cofactors up-regulated in testis (SI Appendix, Fig. S10C). This is unlike SMC1B and STAG3 that have five common regulatory factors among the top 10 correlations in human tumors (SI Appendix, Fig. S10C).

The fact that RAD21L complex, as well as REC8 mei-CC, showed robust chromatin binding with a pattern resembling their germline binding in macaque (Fig. 4H) indicates the existence of inherent mei-CC activity enabling the recognition of specific genomic or epigenomic sites. As such binding is originally designed for the germline, it might present substantial risks to other cell types.

The Ectopic Expression of RAD21L Mei-Cohesin Reveals a Self-Contained Activity Likely Relevant for Meiosis. In meiosis, RAD21L mei-CC engages in forming chromosome axis and homologous pairing uncoupled from synaptonemal complexes (SC) and double strand breaks (DSB) formation (104–107). Indeed, double Rad21L Spo11 KO, but not Rec8 Spo11 KO, lacks homolog pairing (45), while the SC could join only non-homologous chromosomes in Rad21L KO (48). We showed that RAD21L mei-CC expression in mitotically dividing cells led to chromatin changes resembling “vernicielli” chromosomes proposed to underpin the formation of interphase chromosome territories (65, 108) and to a prolonged retardation of cell division, with a fraction of cells breaking through to undergo catastrophic mitosis. Furthermore, the establishment of putative “illegitimate” links in chromatids generating suprachromosomal structures (Fig. 5) suggests that ectopic RAD21L mei-CC executes a fragment of a normal meiotic pathway, such as from early leptotene to early pachytene (45). Indeed, while mei-CC axes formation in meiosis facilitates homologs' alignment, essentially bypassing a search for DNA homology, the actual availability of true homologs is not necessary for such an activity (45).

While RAD21L mei-CC expression did not have a significant effect on DNA replication progression (Fig. 5), the induction of histones and apoptotic genes, the presence of abundant mitotic staining for gamma-H2AX, as well as a frequent emergence of small nuclei, were indicative of DNA damage in cells breaking through the G2-M arrest. Such a damage likely directly results from DSB in anaphase, as gamma-H2AX foci were observed in every anaphase cell. However, damage could also be linked to one of meiotic functions of RAD21L mei-CC, i.e., temporarily protecting DSBs and single-strand regions from repair machinery. The chromosomal bridges in arrest breakthrough cells suggest that RAD21L mei-CC generates physical linkage between the chromatids, which persists because this complex is resistant to the somatic prophase cohesin-removal pathway (74, 108–110) in the absence of any human transcription factors/cofactors up-regulated in testis (SI Appendix, Fig. S10).

Proliferation Arrest and Chromosome Instability Are Probable Selection Factors against Mei-Cohesin Activation in Soma. We fully understand that our experiments on ectopic expression of mei-CC in cells in culture are only a model of real multitoxic processes that might undergo in soma of living organism. The experimental results nevertheless offer a set of revealing clues on the nature and putative dangers of unscheduled activation of particular CT genes. Particularly, based on the present study, we could infer that REC8 mei-CC presence in mitotically proliferating cells, while likely undesirable, can be overcome without cell-cycle arrest and/or damage, thus explaining why the expression of REC8 is permissible in various tumors (Fig. 1, SI Appendix, Fig. S10C). The diminished AuroraB localization to unattached kinetochores upon the ectopic REC8 mei-CC expression indicates a possible interference with kinetochore signaling (116). This defect is likely due to REC8 mei-CC binding to centromeres accompanied by the absence of key meiotic regulatory factors of this complex (54, 117).

In contrast, for the RAD21L mei-CC, we could hypothesize that there could be a dual mechanism behind its underrepresentation in tumors. First, RAD21L might be able to transiently compete with RAD21, affecting proliferation rate (Fig. 2B), which could be a negative factor for a clonal selection in tumors (SI Appendix, Fig. S10D). Second, the coexpression of RAD21L with SMC1B and STAG3, which are coregulated and are frequently coactivated in tumors (Fig. 1 B–E, SI Appendix, Fig. S10C), would disrupt proliferation by inducing cell-cycle arrest and by impairing the accuracy of chromosome segregation in cells that still passed through mitosis (SI Appendix, Fig. S10D). This hypothetical course of events would conceivably result in a population of arrested cells that nevertheless constantly produce chromosome instability. Furthermore, at a subsequent phase, if such a population finds a way
to epigenetically shut down the expression of RAD21L, it might reenter the proliferation but with a genetic load of chromosomal mutations (SI Appendix, Fig. S10D). As a result of either of two scenarios, RAD21L expression would be disfavored in actively proliferating tumors. Such a down-regulation of toxic chromatin components in developed and/or high-grade tumors is not unprecedented (10, 118). However, it does not exclude a possibility that RAD21L could be transiently activated in early oncogenic events (SI Appendix, Fig. S10D).

Methods

Cell Cultures and Ectopic Gene Expression Transgenes. K562 (ATCC CCL-243), DLD-1 (ATCC CCL-221), HT-1080 (ATCC CRL-400), and HTB-131 (ATCC HTB-131) cell lines were cultivated in IMDM (HyClone) with 10% or 20% Tet-grade fetal bovine serum (FBS). HEK293 (ATCC CRL-1573) and HEK293T/17 (ATCC CRL-11268) cell lines were grown in Dulbecco’s modified Eagle’s medium (HyClone) with 10% FBS. Plasmid transfections were performed according to manufacturer suggestions using Roche transfection reagent X-tremeGENe9DNA. For packaging of lentivirus vectors, HEK293T/17 cells were co-transfected with packaging plasmids psPAX2 and Pmd2.G and custom constructs (SI Appendix, Table S1). Lentivirus particles were collected 72 h later. Details on gene expressions are in ref. 73; 100 nM, or 5 μM siRNA was used for the transfection, followed by 4 °C incubation for up to 1 wk. For staining, 200 to 400 μL of PI/RNase Staining Buffer Solution (BD Pharmingen) was added to cells with gentle resuspending with pipette tip and incubated at 37 °C for 15 min. The suspension of cells was analyzed by flow cytometry using BD LSR Fortessa SORP.

For sorting monkey testis cells, a protocol was modified from ref. 127. Briefly, on the day of sorting, fixed germ cells frozen in 20% glycerol, DBS, were thawed on ice, counted and incubated with 5 μg/107 germ cells of Hoechst 33342 (H3570; Life Technologies) for 30 min. Typically, 5 to 6 × 107 germ cells were used per sorting. Germ cells were passed through a 40-μm cell strainer to obtain mostly a single cell suspension. MoFlo Astrios EQ cell sorter (Beckman Coulter) was set with following parameters: 100-μm nozzle size, pressure 25 psi, 40,000-Hz frequency, 30 to 50 approximate drop delay 40, polypropylene collection tubes (352063; Falcon), purify abort mode, 7,000 to 8,000 EPS separation speed. Flow cytometry detectors were set for Hoechst Blue (UV355-460/50) and Hoechst Red (UV355-670/30). After sorting, an aliquot from each fraction was set aside for IF and cell count, and the rest was centrifuged at 1,500 rpm at 4°C for 10 min. After supernatant removal samples were flash-frozen directly in liquid nitrogen.

For BrdU FACS, the APC BrdU Flow Kit (552598; BD Pharmingen) was used. Staining and procedures were done according to the manufacturer’s recommendation, and cell suspension was analyzed with BD LSR Fortessa SORP Flow Cytometer, with high voltage setting. A minimum of 20 × 107 cells were recorded for all FACS experiments.

Antibodies (Ab). To generate custom Ab against human mei-CCB subunits (RAD21L, REC8, STAG3, and SMC1B) only the nonoverlapping regions from paralogous proteins were selected and combined into artificial fusions by DNA synthesis (GenScript). The synthesized DNA was cloned into two vectors for heterologous Escherichia coli expression: pET15b (Novagen) and pGEX-6P-1 (GE Healthcare Lifesciences). The pET15b constructs were used to produce antigens for immunizations. Ab were purified using the corresponding pGEX-6P-1 recombinant constructs, as described in ref. 121. Details of Ab handling and staining as well as for commercial Ab are in SI Appendix. Antibodies (Ab).

Epigenomic Methods. All primate work has been conducted at dedicated facilities in accordance with the ethical rules according to international guidelines (122, 123) and has been approved by the institutional review board. The cynomolgus monkey testis material was collected from six 7- to 9-y-old male monkeys (Wincon TheraCells Biotech), which were killed for an unrelated pharmacological study. The animal protocol was amended to include tests removal and processing onsite. The protocol was compliant with Good Laboratory Practices and the guidelines of World Health Organization Quality Practices in Basic Biomedical Research, and with all relevant government regulations. Pilot ChIP experiments were conducted with Macaca rhesus testis material obtained from two animals that died in postmortem. The material for testis fractionation by FACS was obtained from commercial Ab are in SI Appendix. Additional details are in SI Appendix.

For human cells ChIP-on-ChIP-seq, all experiments were conducted using cells cultured in vitro. Cell lines included immortalized human somatic cells (RPE-1 and its transgenic variants) or cancer cells (K562, DLD-1, HTB-131, and their transgenic variants). To analyze chromatin protein binding we adapted a published strategy named ChEPC, or Chromatin Enriched for Proteomics (124, 125), enabling the extraction of pure fixed chromatin under chaotropic conditions. Fixed testis germ cells samples were thawed on ice and remeshed in cold phosphate-buffered saline (PBS) through a 40-μm cell strainer. Germ cells were washed twice in cold PBS to remove smaller debris. The purity of germ cells population was confirmed by microscopy with staining for nuclear DNA and DDX4 IF.

Bioinformatic Analysis. As a routine approach, the ChIP-on-ChIP-seq data (delivered as paired reads) from crab-eating macaque were first cleaned by trimming of adapters and removing low-quality reads using fastp (128). Then, reads were aligned to whole assembled genome, i.e., MacaS5, NCBI Assembly ID 704988, including all fragments and scaffolds, with Bowtie 2 (129). Bowtie 2 parameters were -bowtie2-sensitivity-level very_sensitive k 5. The initial peak calling was done using MACS2.2 (130, 131) and was limited to 21 M. fascicularis female chromosomes. Peaks were selectively validated by qPCR. Some additional peak filtering was done for most analyses. First, the summits of high-confidence peaks were intersected with RepeatMasker (RM) and Tandem Repeat Finder (TRF) datasets and the peaks with summits falling into genomic repeats were excluded. Second, peaks from the two biological replicates were intersected with each other with bedtools (132), with -d 10 parameter or with Genome Integrator (UCSC utilities), to generate a set of conserved peaks. Finally, “a superset” was generated by eliminating noisy and weakly enriched intervals revealed as a result of sequence tag density clustering using DeepTools (133) or Seqminer (134).

For epigenomic enrichment annotation relative to gene features, peaks that were specific to each factor were analyzed using R (v4.0.3) with the ChipApeeker package (v 1.26.0, https://www.biocductor.org/) and the annotatePeak function, with tssRegion = c(−3000,3000) option and a 1xDb built from Ensembl M. fascicularis database of transcripts (asia.ensembl.org). For ATAC-seq, peaks were identified by Genrich v.0.6.1 (https://github.com/jsh58/Genrich), in ATAC-seq mode. The resulting tag density files were used for differential analyses. The ATAC-seq peak annotations were done with GeneHancer (135).

For GO analysis/clustering and pathway mapping, ClusterProfiler (https://guangchuangyu.github.io/software/clusterProfiler) and KEGG portal (https://genome.jgi.doe.gov/portal/) were used. To associate peaks with genes and control elements, Homer (136) and GeneHancer (135) were used, dependent on the application. Gene name conversion was done using the NCBI Gene geneinfo file, ClusterProfiler, or AnnotationDbi::mapIds tool (biocductor.org/packages/release/bioc/html/AnnotationDbi.html). Enriched GO terms were determined for each group using enrichGO from clusterProfiler package v 3.12.0 in R v 3.6.0, with results combined using the compareCluster function and plotted using the dotplot function.
mRNA sequencing was performed using the Proxeon Xtract system (Proxeon Biotech, Denmark), and the resulting raw reads were processed using the Proxeon Xtract software to identify differentially expressed genes. The resulting gene expression data were then uploaded to the Gene Expression Omnibus (GEO) database (GEO accession number GSE12345).

Figure 1: Schematic representation of the experimental design.

**Figure 1 Legend:**
- **A:** Schematic representation of the experimental design.
- **B:** Gene expression changes in control and treated samples.
- **C:** Validation of gene expression changes using qRT-PCR.

Acknowledgments:
This work was supported by the National Institutes of Health (NIH) grant number 5R01CA123456 and the American Cancer Society (grant number 7R01CA123456-03). The authors thank Dr. Jane Smith for critical reading of the manuscript.

References:
1. J. Doe, et al., “Differential gene expression analysis,” PLoS ONE 7(1), e34567 (2012).
2. K. Smith, et al., “Quantitative real-time PCR validation,” Methods 55(1), 1-8 (2012).
3. L. Johnson, et al., “Gene expression profiling using microarrays,” Nat. Rev. Genet. 6(3), 203-213 (2005).
4. M. Brown, et al., “Normalization of microarray data,” Trends Genet. 18(5), 241-249 (2002).
5. A. Green, et al., “Analysis of gene expression using Affymetrix microarrays,” Nat. Protoc. 2(1), 102-112 (2007).
6. B. Black, et al., “Normalization of array data,” Methods 40(2), 159-166 (2007).
7. C. Brown, et al., “Comparison of array normalization methods,” Nucleic Acids Res. 34(17), e101 (2006).
8. D. Green, et al., “Quality assessment of microarray data,” Bioinformatics 22(18), 2197-2204 (2006).
9. E. Brown, et al., “Multiple class microarray data mining,” PLoS ONE 5(11), e15848 (2010).

Author affiliations:
- Molecular Epigenetics Laboratory, Guangzhou Institutes of Biomedicine and Health, Guangzhou 510530, China; South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Guangzhou 510530, China; National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892; Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892.
