# UC Davis
## UC Davis Previously Published Works

**Title**

DNA Topoisomerase I differentially modulates R-loops across the human genome

**Permalink**

**Journal**

**ISSN**

**Authors**

Manzo, Stefano G
Hartono, Stella R
Sanz, Lionel A
et al.

**Publication Date**

**DOI**

**Copyright Information**

Peer reviewed

Genome Biology

**RESEARCH**

**Open Access**

CrossMark

# DNA Topoisomerase I differentially modulates R-loops across the human genome

Stefano G. Manzo[1,4†], Stella R. Hartono[2†], Lionel A. Sanz[2], Jessica Marinello[1], Sara De Biasi[3], Andrea Cossarizza[3], Giovanni Capranico[1*] and Frederic Chedin[2*]

## Abstract

**Background:** Co-transcriptional R-loops are abundant non-B DNA structures in mammalian genomes. DNA Topoisomerase I (Top1) is often thought to regulate R-loop formation owing to its ability to resolve both positive and negative supercoils. How Top1 regulates R-loop structures at a global level is unknown.

**Results:** Here, we perform high-resolution strand-specific R-loop mapping in human cells depleted for Top1 and find that Top1 depletion results in both R-loop gains and losses at thousands of transcribed loci, delineating two distinct gene classes. R-loop gains are characteristic for long, highly transcribed, genes located in gene-poor regions anchored to Lamin B1 domains and in proximity to H3K9me3-marked heterochromatic patches. R-loop losses, by contrast, occur in gene-rich regions overlapping H3K27me3-marked active replication initiation regions. Interestingly, Top1 depletion coincides with a block of the cell cycle in G0/G1 phase and a trend towards replication delay.

**Conclusions:** Our findings reveal new properties of Top1 in regulating R-loop homeostasis in a context-dependent manner and suggest a potential role for Top1 in modulating the replication process via R-loop formation.

## Background

Biological processes such as transcription and replication generate torsional stress on the DNA double helix that, if not properly dealt with, can lead to genome instability [1]. R-loop structures, a prevalent non-B DNA structure in mammalian genomes, have been particularly linked to genomic instability by causing interference between the replication and transcription machineries [2, 3]. R-loops are formed during transcription upon reannealing of the nascent transcript to the DNA template strand, forming an RNA:DNA hybrid and forcing the non-template strand to loop out. Mapping data indicate that these non-B DNA structures are prevalent in mammalian genomes, where they form dynamically over conserved regions [4, 5]. Negative supercoiling generated behind the elongating RNA polymerase [6] is thought to facilitate R-loop formation by inducing

an underwound DNA state favorable to the re-annealing of the nascent transcript [7]. DNA Topoisomerase I (Top1) is one main cellular factor controlling topological homeostasis [8, 9]. Top1 activity can relax negative supercoils by cutting one of the DNA strands, creating a transient Top1-DNA cleavage complex (Top1cc), and performing a controlled rotation of the cut strand around the uncut strand [10, 11]. The relaxation activity on negative supercoils is thought to reduce co-transcriptional R-loop formation which in turns prevents replication / transcription interference and favors genome stability. Indeed, deletion of the bacterial *topA* gene, an enzyme that only relaxes negative supercoils, creates R-loop-prone hypernegatively supercoiled DNA and causes a growth defect that can be suppressed by over-expression of Ribonuclease H (RNase H), an enzyme that degrades RNA strands in RNA:DNA hybrids [7, 12, 13]. Furthermore, persistent depletion of Top1 in mammalian cells leads to replicative stress and replication-transcription conflicts that can be rescued by overexpression of RNase H [14]. Finally, stabilization of Top1cc by Top1 inhibitors such as camptothecin and its derivatives [15] leads to R-loop stabilization in human cells upon short treatment [16, 17]

* Correspondence: giovanni.capranico@unibo.it; flchedin@ucdavis.edu
†Stefano G. Manzo and Stella R. Hartono contributed equally to this work.
[1]Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy
[2]Department of Molecular and Cellular Biology and Genome Center, University of California, Davis, USA
Full list of author information is available at the end of the article

Manzo *et al. Genome Biology* (2018) 19:100

Page 2 of 18

and to transcription-dependent DNA breakage that can be partially suppressed by RNase H expression [18].

Thus, while it is clear that Top1 regulates R-loops and prevents R-loop-induced genomic instability, the range of loci that are sensitive to R-loop modulation by Top1 is not known. Addressing this gap in knowledge is important given rising evidence that R-loops are abundant in mammalian genomes and also participate in important biological processes [19–21]. For instance, R-loops are involved in regulating chromatin states [4, 5, 22], in mediating transcription termination [23], and in immunoglobulin class switch recombination [24]. Studies also suggest a role for R-loops in priming DNA replication in prokaryotic systems and yeast [25–28]. How R-loop formation is dynamically regulated to permit the physiological roles of R-loops while minimizing the negative impacts of excessive R-loops on genome stability is not clear. In this study, we used the DRIPc-seq technique [4] to map R-loop structures genome-wide in human cells experiencing an acute but transient depletion of Top1. Our work reveals that Top1 modulates R-loop structures differently according to genomic context and provide new evidence that R-loops may play a role in the replication process.

## Results

### Top1 depletion causes subtle R-loop gains and losses

To investigate how global R-loop patterns change upon Top1 depletion, we used siRNA transfection to efficiently silence Top1 in human HEK293 cells (Fig. 1a) and quantified global R-loop levels using dot blots, taking advantage of the anti-DNA:RNA hybrid S9.6 antibody [29, 30]. This approach reproducibly showed a 1.5–2-fold increase in overall R-loop loads in Top1-depleted cells five days post-transfection (Fig. 1b and Additional file 1: Figure S1A). As a control, we pre-treated the genomic DNA with RNase H, which completely abolished the R-loop signal in both control and Top1 knockdown samples, demonstrating specificity of these methods (Additional file 1: Figure S1B). A time course experiment showed that this global R-loop increase was detectable 5 and 6 days post initial transfection but could not yet be detected 4 days post-transfection (Additional file 1: Figure S1C). To capture the early R-loop response to Top1 depletion all subsequent experiments were performed 5 days after initial transfection.

We next quantified R-loop formation genome wide by employing DRIPc-seq, a technique that allows high-resolution, strand-specific genomic mapping of R-loop structures [4]. R-loop structures were observed over 69,066 peaks using a standard peak calling algorithm, (Additional file 2: Table S1), covering ∼ 200 megabases (Mb) of genomic space, which is in close agreement with previous data [4]. As expected, R-loop formation was predominantly genic, with promoters and terminators representing hotspots of signal (Additional file 1: Figure S1D).

Detection of R-loop signal changes [4] indicated that only a small subset of R-loop peaks (4.07%) showed significant changes upon Top1 depletion. However, inspection of signal in Top1-depleted samples revealed numerous instances of signal spreading from existing R-loop peaks (Additional file 1: Figure S1E). To properly account for these events, we optimized a high-sensitivity version of our peak-calling algorithm and applied it to all samples (see Methods). This method identified a total of 399,953 peaks and significant, reproducible R-loop signal gains and losses upon Top1 depletion occurred at 15,112 and 12,977 peaks, respectively (7.02% of total peaks) (Additional file 1: Figure S1F, see Additional file 2: Table S1 for a comparison between standard and high-sensitivity methods). These changes were independently validated using DRIP-qPCR at representative test loci (Additional file 1: Figure S1G). Similar results were obtained when we induced Top1 depletion with an additional siRNA and similar trends towards R-loop gains and losses were also detected at earlier Top1 depletion time points (Additional file 1: Figure S1H and I). Finally, similar trends towards R-loop gains and losses were observed with the S9.6-independent method DRIVE-qPCR [5], in which a catalytically inactive RNASEH1 protein is used to capture R-loops (Additional file 1: Figure S1J).

Quantitative analysis of DRIPc-seq signal over high-sensitivity peaks of R-loop gain (RLG) and R-loop loss (RLL) was consistent with the increased S9.6 signal detected by dot blot: RLG peaks occupied ∼ 100 Mb of sequence space whereas RLL peaks occupied 36 Mb (Fig. 1c). Furthermore, the intensity of R-loop signals measured as the total number of reads over all RLG and RLL peaks also showed an overall net increase in Top1-depleted cells (Fig. 1d). Since ribosomal DNA arrays harbor a major source of cellular R-loops [19] and Top1 depletion in yeast was shown to cause R-loop gains over the 5′ETS region [31], we reasoned that ribosomal R-loops could also contribute to the increased R-loop signal observed by dot blots. To address this, we visualized R-loop loads over the ribosomal DNA region, which revealed that R-loops increased over the 5′ ETS, but not the transcribed 28S region (Fig. 1e). This was further validated by DRIP-qPCR and similar results were obtained with a second siRNA against Top1 (Additional file 1: Figure S1K and L). Therefore overall, results from genomic profiling and dot blot analysis were consistent and point to the fact that Top1 depletion results in a net increase in cellular R-loop loads, although only a minority of R-loop peaks are directly affected.

### R-loop gains and losses in Top1-depleted cells define distinct gene categories

Peaks of RLG and RLL in Top1-depleted cells appeared to have distinct genomic distributions. RLG peaks overwhelmingly (81.2%) overlapped with gene body (Fig. 1c),
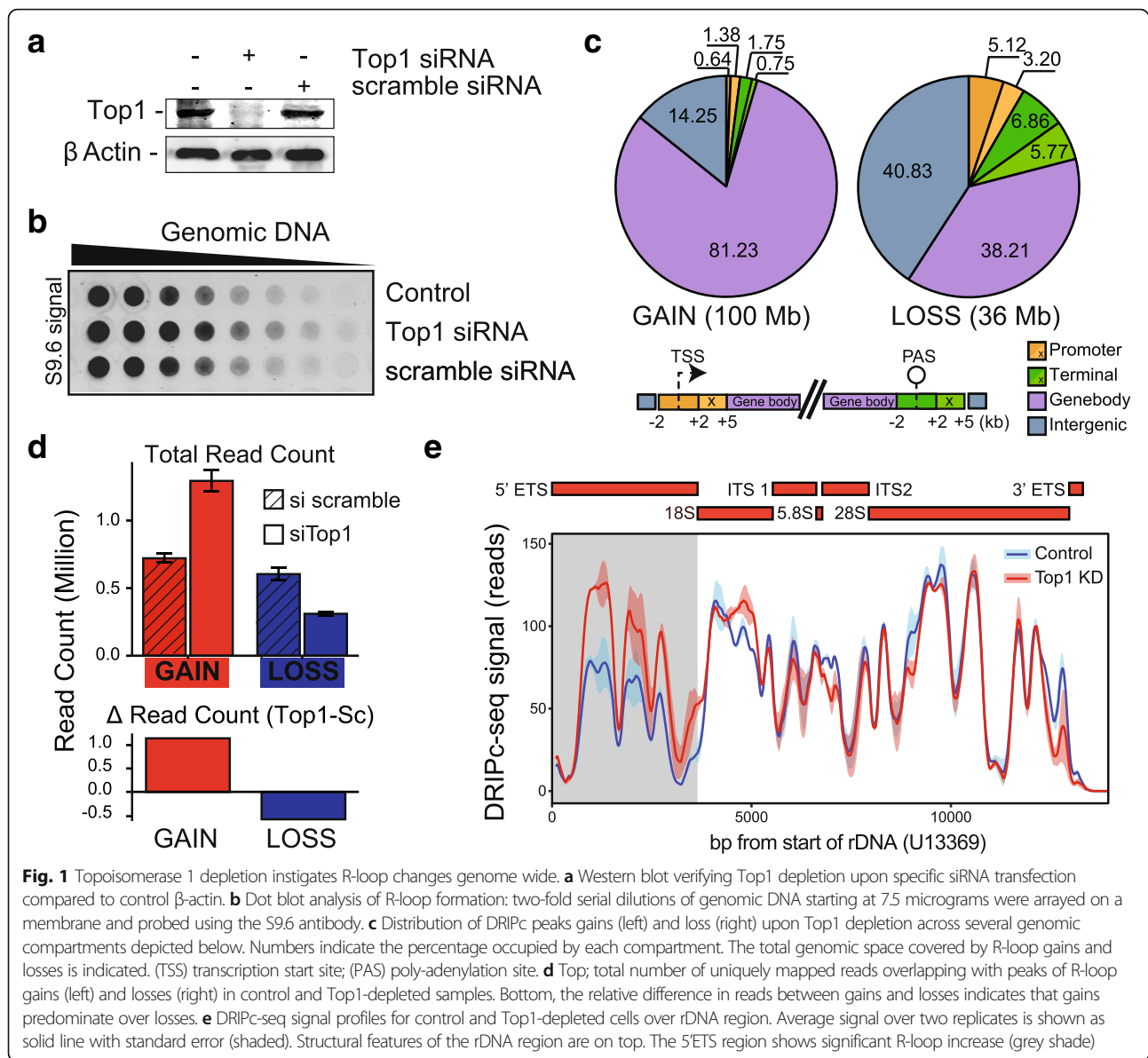
Manzo *et al. Genome Biology* (2018) 19:100

Page 3 of 18



**Fig. 1** Topoisomerase 1 depletion instigates R-loop changes genome wide. **a** Western blot verifying Top1 depletion upon specific siRNA transfection compared to control β-actin. **b** Dot blot analysis of R-loop formation: two-fold serial dilutions of genomic DNA starting at 7.5 micrograms were arrayed on a membrane and probed using the S9.6 antibody. **c** Distribution of DRIPc peaks gains (left) and loss (right) upon Top1 depletion across several genomic compartments depicted below. Numbers indicate the percentage occupied by each compartment. The total genomic space covered by R-loop gains and losses is indicated. (TSS) transcription start site; (PAS) poly-adenylation site. **d** Top; total number of uniquely mapped reads overlapping with peaks of R-loop gains (left) and losses (right) in control and Top1-depleted samples. Bottom, the relative difference in reads between gains and losses indicates that gains predominate over losses. **e** DRIPc-seq signal profiles for control and Top1-depleted cells over rDNA region. Average signal over two replicates is shown as solid line with standard error (shaded). Structural features of the rDNA region are on top. The 5'ETS region shows significant R-loop increase (grey shade)

indicating that they associate with transcription elongation. RLL peaks, by contrast, overlapped with promoter and terminator regions, with only a minority (38.2%) mapping to gene bodies. Moreover, RLL peaks most often showed an "intergenic" distribution that often corresponded to a terminator-downstream location immediately outside of the gene boundaries used here for classification (Fig. 1c). Mapping RLG and RLL peaks onto genes allowed us to identify three classes of genes: RLG genes that predominantly gained R-loops, RLL genes that mostly lost R-loop signal, and a large class of genes that showed both gains and losses (Additional file 1: Figure S2A, B and Additional file 2: Table S1). To simplify the analysis, we selected RLG genes with a minimal 5:1 ratio of peak gains to loss (and vice-versa for RLL genes). This

delineated two clear groups of RLG ($n = 959$) and RLL genes ($n = 2046$), respectively (Fig. 2a). A large fraction of genes underwent mixed changes (gain and loss; $n = 9375$), while a small portion did not undergo detectable changes ($n = 2464$) (Fig. 2a). Altogether, Top1 depletion triggered subtle shifts in genic R-loop distribution marked by both gains and losses of R-loops over a large portion of genes. Overall, a significant subset of genes (11.8% of total) was marked by nearly exclusive patterns of R-loop gains or losses upon Top1 depletion.

To understand the nature of the differential response to Top1 depletion, we focused on RLG and RLL genes and first investigated their lengths. RLG genes were significantly longer (2.7 fold on average) compared to RLL genes or genes with no or mixed R-loop change (Fig. 2b).
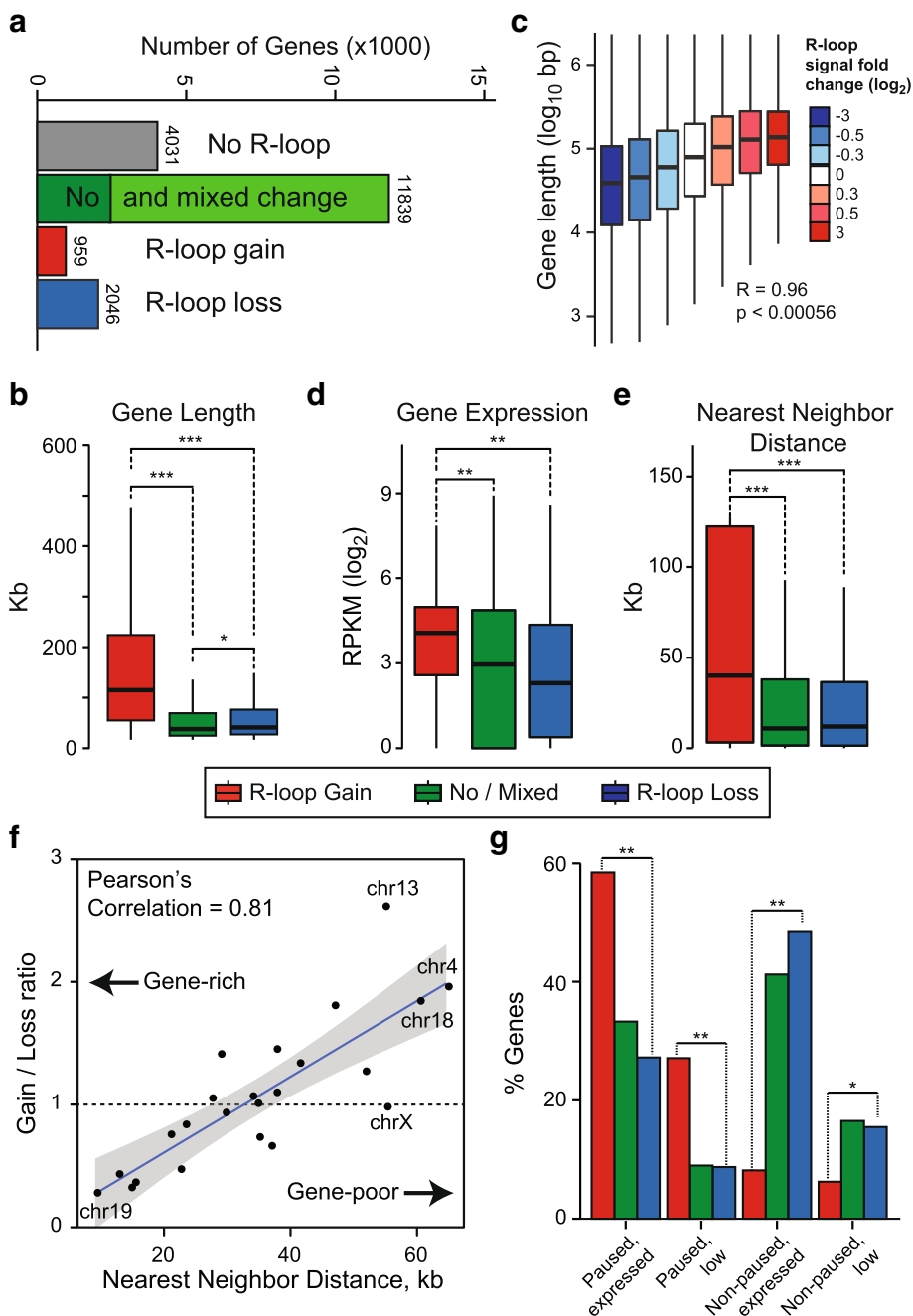
Manzo *et al. Genome Biology* (2018) 19:100

Page 4 of 18



**Fig. 2** R-loop gains and losses occur on genes with distinct categories. **a** Breakdown of genes according to their R-loop status upon Top1 depletion; numbers indicate gene numbers in each category. **b** Quartile plot depicting distribution of lengths for genes undergoing R-loop gains, losses, or no/mixed change (color code is indicated below the plot). Stars (*, **, and ***) indicate *p*-value less than $10^{-10}$, $10^{-25}$, $10^{-40}$, respectively (Wilcoxon Mann-Whitney). **c** Gene length is plotted as a function of the fold R-loop signal change. Bins were chosen so they contain similar number of genes to avoid biased sampling. Median Pearson correlation coefficient and associated p-value are indicated. **d**, **e** Same as (**b**) except gene expression and gene distance are plotted. **f** XY plot between gene density measured on each individual chromosome (represented by a dot) and the ratio of gene numbers undergoing R-loop gains and R-loop losses upon Top1 depletion. The regression line along with 95% confidence interval and Pearson correlation coefficient are indicated. **g** Distribution of genes undergoing R-loop gains and losses according to the RNAP stalling and expression status of each gene. Color code is as in (**b**)

Manzo *et al. Genome Biology* (2018) 19:100

Page 5 of 18

By contrast, RLL genes were only slightly longer than control genes with no or mixed changes in R-loops. This suggests that longer genes are more prone to RLG upon Top1 depletion. To confirm this, we clustered all R-loop peaks by their ratio of R-loop signal change from control to Top1-depleted conditions regardless of whether they belonged to RLL or RLG genes. We then measured gene lengths as a function of R-loop signal fold change in each cluster. We observed a strong positive correlation between the relative intensity of R-loop signal changes and each cluster's length (Fig. 2c). This, together with the marked distribution of RLG peaks to gene bodies, suggests that transcription elongation through long genes is more prone to R-loop stabilization in absence of Top1. To determine if expression levels could also distinguish RLG and RLL genes, we performed total RNA-seq on both control and Top1-depleted cells. This analysis revealed that RLG genes were significantly more expressed (1.5–3.5 fold on average) than RLL genes, which themselves were not significantly different than control genes (Fig. 2d). Thus, RLG genes tend to be long and highly expressed. Since long genes often reside in gene-poor areas of the genome, we measured the distance between RLG and RLL genes and their nearest neighbor. RLG genes were located significantly further away from potential neighbors than control or RLL genes (Fig. 2e). This indicates that RLG genes tend to occupy gene-poor neighborhoods. Consistent with this, there was a strong inverse correlation between chromosomal gene density and R-loop loss/gain ratio: gene-rich chromosomes predominantly showed loss of R-loop signal loss following Top1 knockdown, while gene-poor chromosomes favored R-loop gain events (Fig. 2f).

Finally, given that Top1 can regulate RNAP release from promoter-proximal pausing [32–34], we asked whether RLG and RLL genes differ in how frequently they undergo pause-release. For this, we performed RNA Polymerase II ChIP-seq in control and Top1-depleted cells and determined the pausing index of each gene according to well-defined categories [35]. RLG genes were strongly enriched in genes that undergo promoter-proximal pausing (Fig. 2g). By contrast, RLL genes mostly corresponded to genes that do not undergo pausing and their distribution was not significantly different from that of control genes. More broadly, our analysis confirmed that Top1 depletion causes RNAP accumulation in the vicinity of the TSS, particularly for paused genes [32] (Additional file 1: Figure S2C). Altogether, this shows that the differential response to Top1 depletion defines two broadly distinct classes of genes.

### Top1 depletion favors co-transcriptional R-loop gains
The preferential localization of RLG peaks to gene bodies suggests that R-loop gains occur during transcription elongation. Gene metaplots confirmed that RLG peaks

were typically circumscribed to the transcribed portion of long genes, with no apparent gradient towards the 5′ or 3′-ends (Fig. 3a). By contrast, RLL peaks were more prominent for shorter genes and showed preferential distribution around the promoter and terminator regions. To determine if the increased R-loops over RLG genes was caused through a co-transcriptional mechanism, we asked if the strandedness of RLG peaks was concordant with the strandedness of their respective genes. Gains of R-loop signal were indeed only observed on the template strand (Fig. 3b). By contrast, the patterns of R-loop signal loss over RLL genes were more complex, even in control cells (Fig. 3c). This complexity most likely reflects the fact that RLL genes reside in gene-rich areas and therefore often possess immediate neighbors with divergent promoters and convergent terminators. As a result, R-loop formation appears on both template and non-template strands in metagene plots. Under conditions of Top1 depletion, both template and non-template signals were reduced. In agreement with their lower transcription levels (Fig. 2d), RLL genes showed a lower median R-loop signal compared to RLG genes (compare Fig. 3b and c). Moreover, gene expression analysis showed that the large majority (93%) of RLG and RLL genes did not undergo significant up or downregulation, arguing that the changes described here are not simply a result of altered transcriptional states (Additional file 1: Figure S2D). Overall, the data are consistent with R-loop changes triggered by Top1 depletion being mainly co-transcriptional.

### R-loop gains upon Top1 depletion preferentially associate with heterochromatin and nuclear lamina
To understand the cause(s) driving RLG in the absence of Top1, we asked if RLG peaks possess specific predictive chromatin features. To do this, we compared the chromatin states of RLG peaks to that of other R-loop-forming loci matched for position, expression, and length that showed no or mixed changes in response to Top1 depletion (thereafter referred to as "matched"). This stringent comparison approach was necessary to remove known confounding variables that may affect chromatin states and is similar to the approach previously used to identify chromatin features of R-loop forming genes [4]. Matching on GC skew and overall R-loop levels was also performed and did not change the conclusions (data not shown). At first, we analyzed chromatin association by the extent of peak overlap, keeping promoters, gene bodies, and terminal regions separate. Out of a wide array of available datasets including ChromHMM states [36] and ChIP-seq information for nearly a hundred types of histone modifications and chromatin factors [37], RLG peaks showed modest but significant overlap enrichment only for a few chromatin marks,
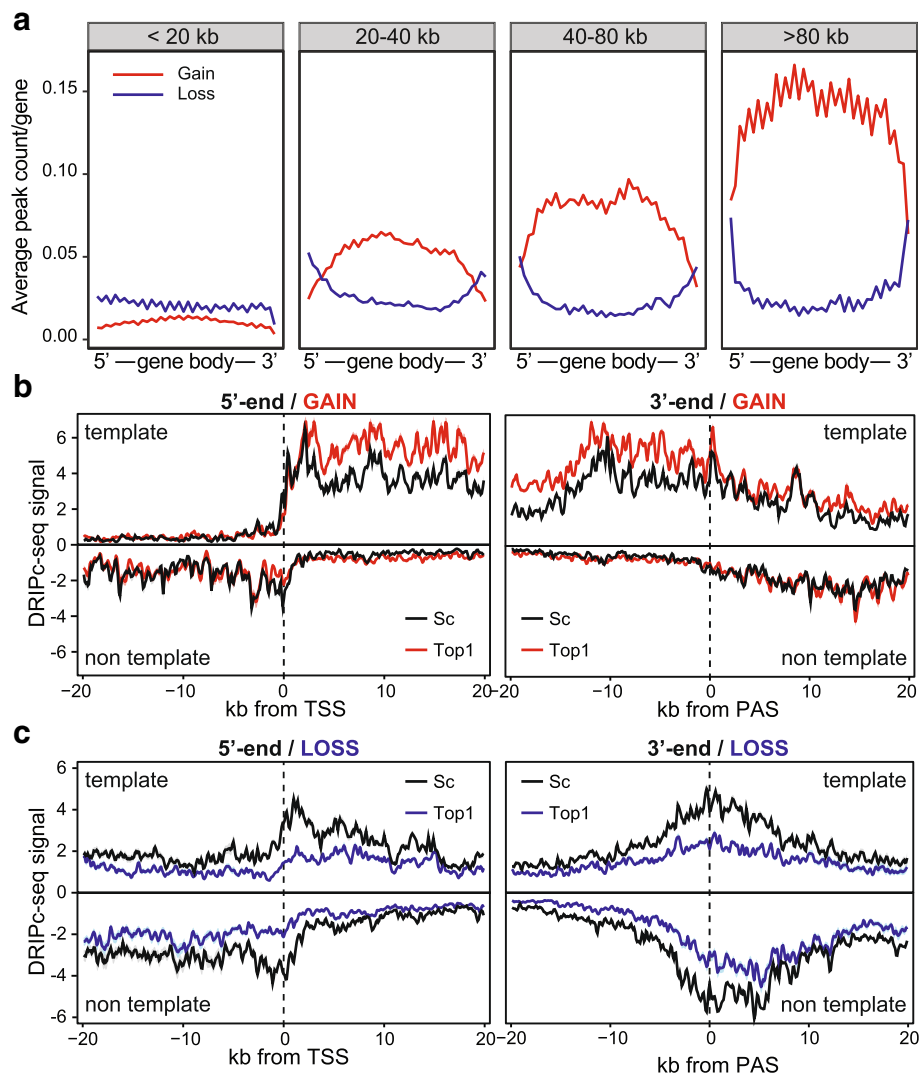
Manzo *et al. Genome Biology*  (2018) 19:100

Page 6 of 18



**Fig. 3** R-loop signal at RLG genes are co-transcriptional. **a** Distribution of all peaks of R-loop gain and loss (p-adjusted < 0.1) along a gene metaplot, normalized by number of genes in each length category. Genes are broken down by length, as indicated at top. Genes were binned in 40 bins and peak counts reported by bin. **b**–**c** Metaplots of DRIPc-seq signal over RLG genes (**b**) and RLL genes (**c**) along a 20 kilobase window centered on their TSS at left, or PAS at right. Values are median and shown with standard deviation (shaded). Samples are color-coded as indicated

highlighting their significance. H3K36me3 and H3K79me2, two marks associated with transcription elongation, and the ChromHMM transcription elongation state itself, were enriched over RLG peaks (Fig. 4a). RLL peaks, by contrast, were significantly depleted for these marks compared to matched sets. In addition, RLG peaks showed significantly higher overlap with the heterochromatin mark H3K9me3 and with lamina-associated domains (LADs) [38] while RLL genes showed the opposite trend (Fig. 4a). These observations are consistent with RLG genes being long and highly expressed and occupying gene-poor neighborhoods (Fig. 2) that are more likely to include heterochromatin-like features such as H3K9me3 and nuclear lamina binding.

Since lamin and H3K9me3 association might represent physical constraints to the dissipation of

transcription-induced DNA supercoils in the absence of Top1, we therefore investigated the relationship between RLG and these parameters further. We calculated the distance between RLG and RLL peaks and the nearest annotated H3K9me3 or LAD peak. RLG peaks were significantly closer to H3K9me3 peaks compared to matched R-loop forming peaks (Fig. 4b). This proximity was true regardless of the genic location of the RLG peak (promoter, gene body and terminator) and was striking given that the median value for distance was close to zero. By contrast, RLL peaks were located further away from H3K9me3 peaks (median distance of 8 kb) than matched control peaks. RLG peaks therefore tend to reside in immediate proximity to H3K9me3 peaks. We next asked if the
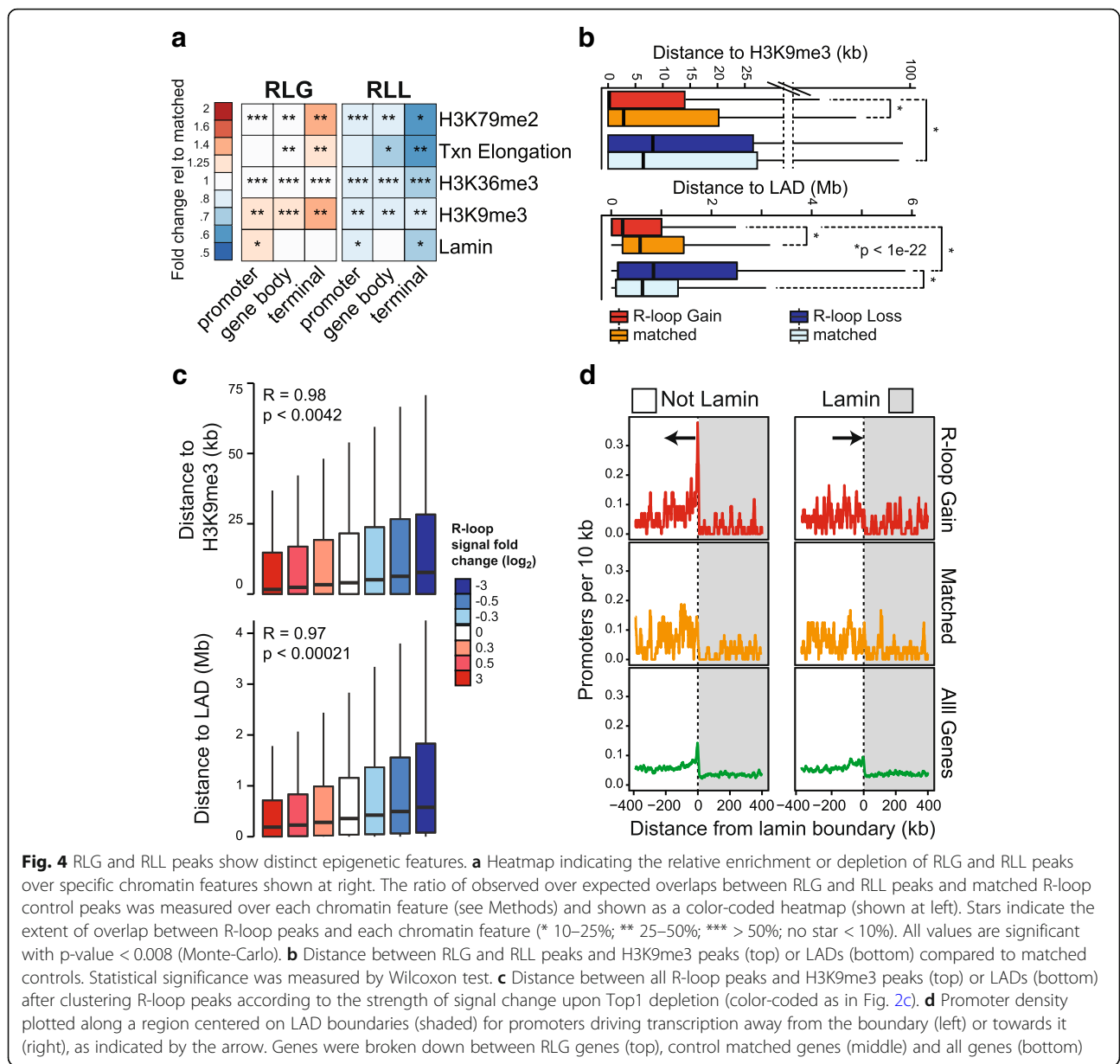
Manzo *et al. Genome Biology* (2018) 19:100

Page 7 of 18



**Fig. 4** RLG and RLL peaks show distinct epigenetic features. **a** Heatmap indicating the relative enrichment or depletion of RLG and RLL peaks over specific chromatin features shown at right. The ratio of observed over expected overlaps between RLG and RLL peaks and matched R-loop control peaks was measured over each chromatin feature (see Methods) and shown as a color-coded heatmap (shown at left). Stars indicate the extent of overlap between R-loop peaks and each chromatin feature (* 10–25%; ** 25–50%; *** > 50%; no star < 10%). All values are significant with p-value < 0.008 (Monte-Carlo). **b** Distance between RLG and RLL peaks and H3K9me3 peaks (top) or LADs (bottom) compared to matched controls. Statistical significance was measured by Wilcoxon test. **c** Distance between all R-loop peaks and H3K9me3 peaks (top) or LADs (bottom) after clustering R-loop peaks according to the strength of signal change upon Top1 depletion (color-coded as in Fig. 2c). **d** Promoter density plotted along a region centered on LAD boundaries (shaded) for promoters driving transcription away from the boundary (left) or towards it (right), as indicated by the arrow. Genes were broken down between RLG genes (top), control matched genes (middle) and all genes (bottom)

intensity of R-loop signal gains was correlated to the distance to H3K9me3 peaks. For this, we clustered all R-loop peaks by their R-loop change ratios and measured the distance between these loci and the nearest H3K9me3 peak in each cluster. A strong correlation between the two parameters was observed such that the R-loop peaks with the strongest gains were located closest to H3K9me3 peaks. Vice-versa, R-loop peaks with the strongest losses were located furthest away from H3K9me3 peaks (Fig. 4c). Furthermore in 70% of cases, the H3K9me3 peaks were located upstream of the RLG peaks relative to gene transcription (not shown). Similar observations were made with LADs: RLG peaks were located significantly

closer to LADs than RLL genes or control matched genes (Fig. 4b). In addition, we observed a strong correlation between distance to LADs and strength of R-loop change (Fig. 4c). While the median distance between RLG and annotated LADs was comparatively large (~ 200 kb; Fig. 4b), we note that top-ranked RLG genes in terms of signal gains were often closely juxtaposed to LADs (Additional file 1: Figure S3A). Similarly, promoters of RLG genes were characterized by a strong upstream Lamin B1 signal and marked lamin B1 depletion around the TSS region (Additional file 1: Figure S3B). To further characterize the arrangement of RLG genes relative to LAD boundaries, we calculated the density of promoters around LAD boundaries as a function

Manzo *et al. Genome Biology* (2018) 19:100

Page 8 of 18

of genic orientation. RLG genes transcribing away from LADs showed a sharp promoter density peak at or near the LAD border (Fig. 4d). By contrast, no peak was observed for genes transcribing towards the LAD. A corresponding promoter peak was not observed for matched control genes and only weakly for all genes. Altogether, this data shows that long, highly transcribed genes accumulate R-loops likely due to the proximity to H3K9me3 peaks and/or LADs, which may impede the dissipation of DNA supercoils along the chromatin template and therefore allow the local accumulation of R-loop-favorable negative DNA supercoils.

## R-loop losses upon Top1 depletion preferentially associate with early, active, replication origins

To gain insights into the functional significance of R-loop loss in the absence of Top1, we also analyzed the chromatin states of RLL peaks. Compared to matched R-loop-forming loci, RLL peaks showed significantly higher intersect with chromatin features characteristic of regions repressed by the Polycomb group complex. These features include the "repressed" ChromHMM state, the Polycomb mark H3K27me3, and several subunits of the PRC1 and PRC2 complexes (Fig. 5a). In all cases, a significant decreased overlap was observed for RLG peaks, suggesting that these chromatin features are specific. Additional overlap enrichment was observed for terms related to the chromHMM "insulator" state and for the CTCF and RAD21 factors that often colocalize at chromatin loops (Additional file 1: Figure S4A).

Since Top1 activity has been mapped to a well-established, conserved human replication origin [39], we also included replication initiation sites mapped by Short Nascent Strand sequencing (SNS-seq) [40] to our analysis. To our surprise, RLL peaks showed strong overlap enrichment with such loci compared to matched controls while RLG peaks showed clear depletion (Fig. 5a). To investigate the relationship between RLL peaks and origins further, we measured the distance between RLL and RLG loci to the nearest annotated SNS-seq loci. RLL peaks were located significantly closer to origins than matched controls and dramatically closer than RLG peaks which themselves were located further away from origins than matched controls (Fig. 5b). Furthermore, we observed a strong correlation between the intensity of the R-loop signal gain/loss upon Top1 depletion measured over RLL and RLG peaks and the distance from these peaks to the nearest SNS-seq origin (Fig. 5c). Strikingly, peaks with the strongest loss tended to directly match onto SNS-seq origins (median distance of zero). Increased overlap of RLL peaks over CpG island loci, which are often replication origins [40], was also observed, while RLG peaks showed the opposite trend

(Additional file 1: Figure S4A). Thus, topoisomerase I depletion associates with a loss of R-loop signal at peaks that are proximal to replication origins.

To understand whether the association of RLL peaks with replication origins may have functional significance, we next asked whether RLL peaks also overlapped with SNS-seq signal, which reflects the frequency of replication initiation events. SNS-seq signal was significantly higher for RLL peaks over both promoters and terminators compared to the signal observed over matched controls (Fig. 5d). Similarly, SNS-seq signal over RLG peaks was lower than that of matched controls. Overall, a strong correlation was observed between replication signal and the strength of R-loop gains and losses (Fig. 5e). This indicates that RLL peaks delineate regions with high replication initiation activity while RLG peaks match further away from sites of replication initiation onto regions with poor replication initiation potential. Analysis of replication timing data (Repli-seq, [41]) confirmed that RLL peaks replicate as early as late G1 (G1b) and S1 phases (Fig. 5f). Compared to matched loci, RLL peaks were significantly more likely to replicate in late G1 and less likely to replicate in later phases. By contrast, RLG peaks replicated predominantly in later phases of the cell cycle (S1 and S2) and showed a significant tendency towards later replication compared to matched loci.

To ensure that the association between RLL peaks and replication origins is robust, we analyzed independent datasets where origins were mapped through Okazaki fragment sequencing (OK-seq) [42]. While SNS-seq and OK-seq datasets produce distinct replication initiation maps, both methods nonetheless show significant overlap in particular around gene bodies and terminal genic regions (data not shown). RLL peaks showed increased overlap with OK-seq-derived initiation peaks (AS peaks, [42]) while RLG peaks showed decreased overlap (Fig. 5a). Likewise RLG peaks were located further away than expected from matched control genes, while RLL peaks were distributed as expected from, or closer than, control peaks (Additional file 1: Figure S4B). Finally, RLG peaks showed reduced densities of AS peaks compared to matched controls while RLL peaks showed the opposite trend (Additional file 1: Figure S4C).

While the analysis above was restricted to genic regions so we could ensure stringent matching procedures, we also investigated intergenic RLL peaks and found determined that intergenic RLL peaks also showed a 4–5 times higher overlap with replication initiation regions (SNS-seq or OK-seq) than expected at random (data not shown). Thus, altogether the present genomic analyses show a robust association between peaks of R-loop loss in response to Top1 depletion and active, early replication origins, suggesting a role for Top1 in modulating the replication process.
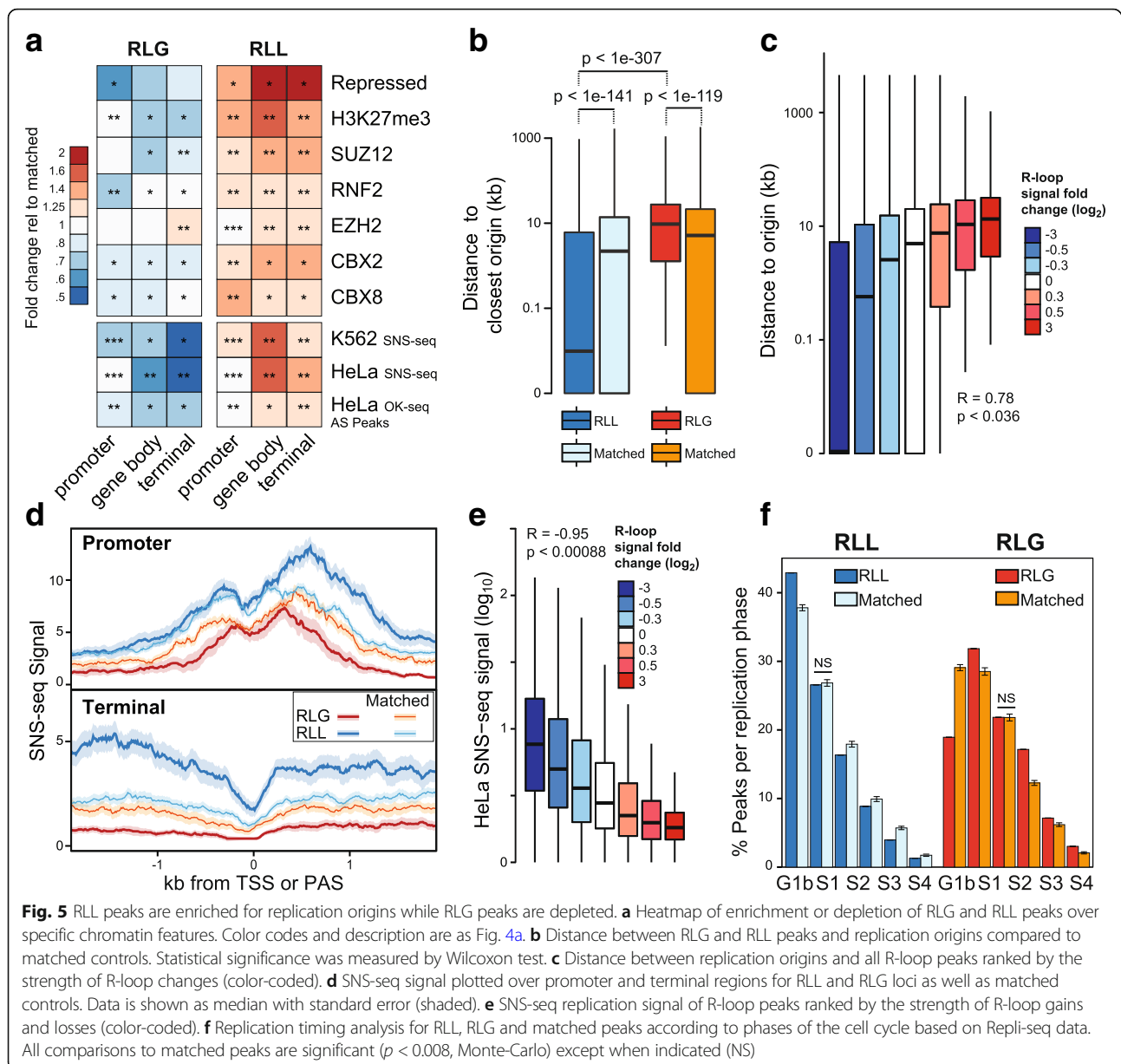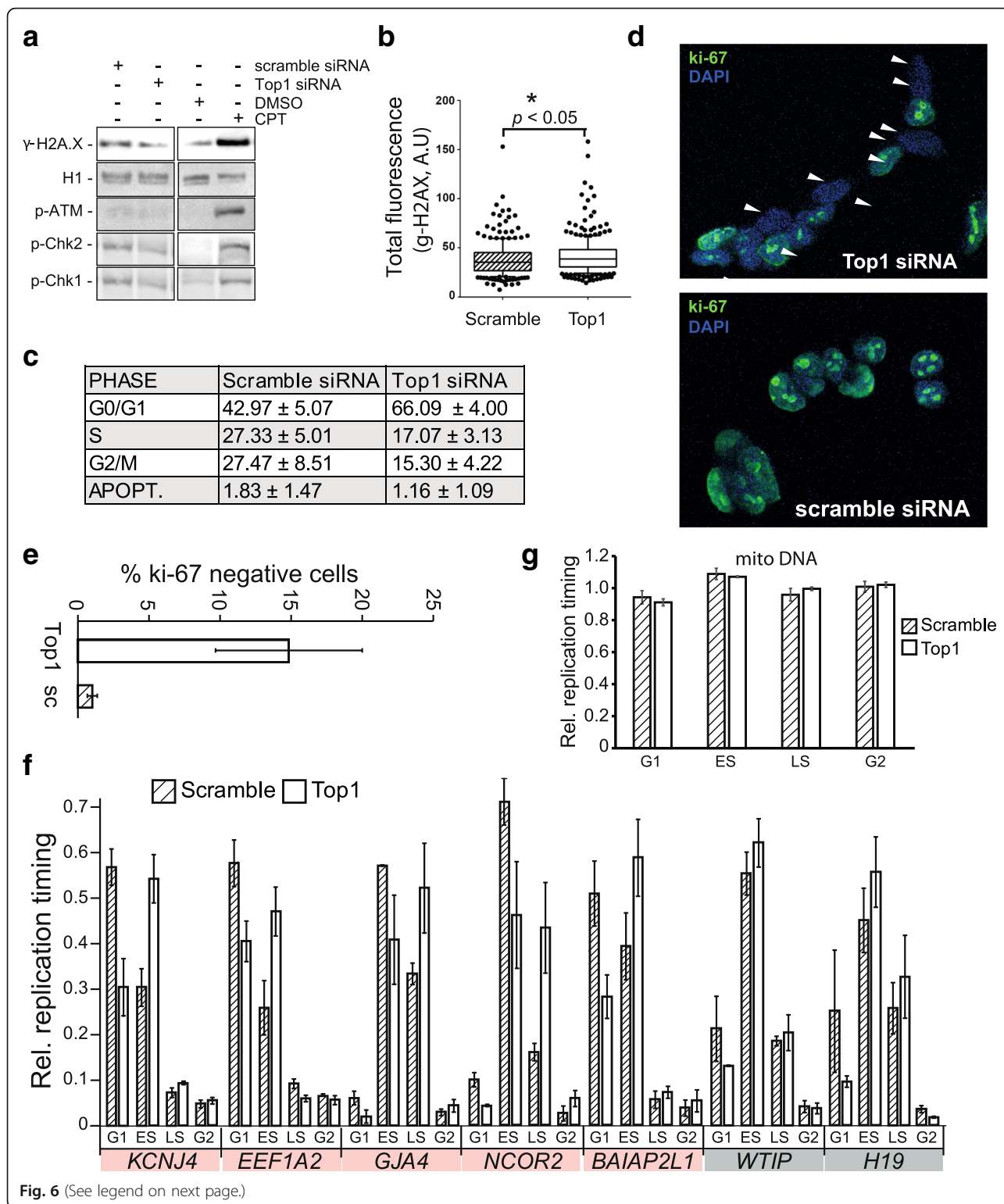
Manzo *et al. Genome Biology* (2018) 19:100

Page 9 of 18



**Fig. 5** RLL peaks are enriched for replication origins while RLG peaks are depleted. **a** Heatmap of enrichment or depletion of RLG and RLL peaks over specific chromatin features. Color codes and description are as Fig. 4a. **b** Distance between RLG and RLL peaks and replication origins compared to matched controls. Statistical significance was measured by Wilcoxon test. **c** Distance between replication origins and all R-loop peaks ranked by the strength of R-loop changes (color-coded). **d** SNS-seq signal plotted over promoter and terminal regions for RLL and RLG loci as well as matched controls. Data is shown as median with standard error (shaded). **e** SNS-seq replication signal of R-loop peaks ranked by the strength of R-loop gains and losses (color-coded). **f** Replication timing analysis for RLL, RLG and matched peaks according to phases of the cell cycle based on Repli-seq data. All comparisons to matched peaks are significant ($p < 0.008$, Monte-Carlo) except when indicated (NS)

## Top1 depletion triggers G0/G1 block and replication timing delays with minimal DNA damage

Since R-loop accumulation, as seen here over long genes in Top1-depleted cells, often associates with increased genomic instability, we tested if Top1 depletion was associated with induction of the DNA damage response. Surprisingly, Western blots did not indicate significant hyper-phosphorylation of the histone variant H2AX, a marker of DNA breaks and replicative stress (Fig. 6a). Camptothecin treatment, which leads to covalent Top1-DNA complexes, strongly induced this modification (Fig. 6a). A more sensitive examination of γH2AX levels by immunofluorescence microscopy revealed a small but significant increase in Top1-depleted cells

(Fig. 6b). However, no significant increased phosphorylation of ATM, CHK1, or CHK2 could be measured by Western blots, suggesting that ATM and ATR DNA damage sensing pathways are not broadly activated upon acute Top1 depletion in human HEK293 cells (Fig. 6a). Since DNA damage induced by Top1 depletion is dependent on S-phase [14], we next asked if Top1 depletion caused any cell cycle changes by performing cyto-fluorimetric analyses. Surprisingly, Top1 depletion induced a consistent block in G0/G1 phase, with marked reduction of the S and G2/M phases (Fig. 6c), which was not due to checkpoint activation (Fig. 6a). Immunofluorescence assays using the Ki-67 proliferation marker confirmed that a fraction of Top1-depleted cells exit out of

Manzo et al. Genome Biology (2018) 19:100

Page 10 of 18



**Fig. 6** (See legend on next page.)

Manzo *et al. Genome Biology* (2018) 19:100

Page 11 of 18

the cell cycle and rest in G0 (Fig. 6d and e). Similar results were obtained using an independent Top1 siRNA (Additional file 2: Figure S5). Thus, a pronounced, short term (5 days), depletion of Top1 impairs the G1 to S phase transition. This observation may account for the modest amount of DNA damage observed under these conditions since R-loop mediated genomic instability is thought to be caused by replication-transcription conflicts. Indeed, prolonged Top1 depletion under selective conditions that force cells to undergo division is associated with R-loop-induced DNA damage [14].

Given the accumulation of Top1-depleted cells in G0/G1, we wondered if the R-loop gains and losses we observed could be due to preferential R-loop formation by RLG and RLL genes within and outside of G1, respectively. To test this, we synchronized cells in G2, released them, and monitored R-loop formation in G1 and mid-S by DRIP-qPCR at a range of loci. Other than an expected drop of R-loop formation in G2/M, we did not observe any specific trend across RLG and RLL loci analyzed here (Additional file 2: Figure S6A). To further determine if the pattern of R-loop gains and losses could be explained by an accumulation of cells in G0/G1, we took advantage of a recent R-loop dataset obtained from breast cancer cells (MCF7) in the presence or absence of estradiol [43]. Without estradiol, 85% of MCF7 cells accumulate in G0/G1. Upon stimulation with estradiol, cells rapidly re-enter the cell cycle. We reasoned that if R-loop gains and losses are caused by the G0/G1 arrest, then a similar, perhaps even amplified, pattern should be observed in MCF7 cells. To test this, we determined the overlap between MCF7 R-loop peaks in unstimulated or estradiol-stimulated conditions (2 and 24 h) and both RLG genes and RLL genes. The proportion of RLG and RLL genes that intersected with R-loops (i.e. at least on R-loop peak in the gene) was relatively constant through time (Additional file 2: Figure S6B). Thus we did not observe that the proportion of R-loop-positive RLG genes decreased with cell cycle re-entry as would be expected if R-loop formation at RLG genes was a property of the G0/G1 state. Similarly, we did not observe a significant increase in the proportion of R-loop positive RLL genes with cell cycle re-entry, as would be expected if R-loop formation was only allowed outside the G0/G1 phase in

RLL genes. Similar results were obtained when we measured the total R-loop loads (i.e. length occupied by R-loop peaks) of RLL and RLG genes (Additional file 2: Figure S6B). Overall, we conclude that the patterns of R-loop losses and gains observed upon Top1 depletion are unlikely to simply reflect shifts in cell cycle patterns and are more likely to result from the response to Top1 depletion.

The observation that Top1-depleted cells undergo a G1/S transition block, combined with the close association of RLL peaks with early, highly active, replication origins, suggested that Top1 depletion may affect the replication program. To test this, we analyzed the replication timing of multiple early-replicating RLL regions using BrdU incorporation to mark newly replicated strands, followed by immunoprecipitation and qPCR after cell sorting into G0/G1, early S, late S and G2/M phases. Top1-depleted cells showed delayed replication timing with a consistent switch from G1 to early S or from early S to late S phase for 5 out of 7 RLL loci (Fig. 6f). It should be noted however, that genes with no or mixed changes in R-loop distribution also showed a similar trend for 7 out of 8 loci tested (Additional file 2: Figure S6B). A minority of RLG genes (3 out of 10) also showed a modest tendency towards replication delay (Additional file 2: Figure S6C). This effect was specific for the nuclear genome, as mitochondrial DNA replication timing was not affected by Top1 knockdown (Fig. 6g). Thus, when cells are able to overcome the G1/G0 block and initiate S phase, they nonetheless show a trend towards replication delay in a way that appears influenced by, but not strictly dictated by, R-loop gains or losses.

## Discussion

Multiple studies have identified Top1 as a factor that prevents R-loop formation since the enzyme relaxes negative supercoils during transcription [44, 45] thereby preventing an R-loop favorable underwound DNA state [7, 12]. To understand how Top1 modulates R-loop formation *in vivo*, we profiled these structures globally in Top1-depleted human cells. Surprisingly, Top1 depletion caused both increases and decreases of R-loop levels depending on the genomic context (Figs. 1 and 2).

Manzo *et al. Genome Biology* (2018) 19:100

Page 12 of 18

Consistent with the expectation that Top1 prevents R-loops, we identified a clear class of genes that respond to Top1 depletion by gaining R-loops. These genes were long, highly transcribed, and located in gene-poor areas of the genome. These observations are consistent with prior studies and allow us to refine a model for Top1 activity during transcription elongation. Long genes were shown to be more sensitive to Top1 poisoning by Camptothecin [46] or to Top1 depletion in mouse and human neurons [47]. These studies are in agreement with our observations that RLG peaks arise co-transcriptionally on long and highly expressed genes, where they principally match to gene bodies (Figs. 1, 2 and 3). Our work also shows that RLG genes preferentially undergo promoter stalling (Fig. 2) which is in agreement with a recent study showing that Top1 becomes physically associated with the RNAP complex and catalytically activated upon release of the transcription machinery into elongation from a promoter-proximal paused state [32]. Similarly, these observations are in agreement with prior observations that Top1 inhibition with camptothecin has an immediate effect on the RNAP II pause/release cycle at active promoters including at the long human HIF-1α gene [34]. Our present work supports the view that Top1 facilitates transcription elongation and precisely defines the class of genes that are most dependent on this enzyme: Top1 efficiently prevents co-transcriptional R-loops specifically for long gene units with high transcription levels.

Interestingly, the generation of topological stress during transcription requires that the DNA fiber is placed under some physical constraints so as to prevent spontaneous dissipation of supercoils [1]. Our genome-wide data reveals that in the case of R-loop stabilization through RLG genes, proximity to H3K9me3-marked chromatin and lamin-associated domains may represent the main source of such physical constraints. A subset of RLG genes were located in close proximity to LADs and faced away from the LAD boundary, suggesting that LADs might physically trap supercoils, causing an increase in negative supercoil density behind the transcribing RNAP in the absence of Top1. Indeed, the strength of R-loop gains clearly correlated with the proximity to LADs (Fig. 4). We therefore suggest that the association of genes to the nuclear envelope sensitizes them to topological disruptions. In addition to LADs, we identify heterochromatic H3K9me3-marked patches as a second important distinguishing feature of RLG genes. These patches often were in close proximity to RLG peaks and the strength of RLGs was inversely correlated to their distance from H3K9me3 peaks. We suggest that H3K9me3-marked heterochromatic patches might prevent dissipation of torsional tension because of their closed chromatin nature. Additionally, H3K9me3 was shown to mediate perinuclear anchoring, which could

further prevent supercoil dissipation [48]. Altogether, our data reveals that long, highly expressed genes in proximity to LADs or H3K9me3 patches are important reservoirs of R-loops that require proper topological control by Top1. Top1 depletion may be less critical for genes without such topological constraints where activity of Top2 isoforms may be sufficient to substitute for Top1 absence.

Given the association between R-loops and RNAP pausing [4] as well as DNA breakage [3], we speculate that R-loop suppression by Top1 plays an important role in ensuring proper gene expression and genome stability. We note that, in contrast to other studies [14, 49], we did not detect telltale signs of genomic instability (Fig. 6). Importantly, these studies used cell lines in which Top1 was stably knocked down and that were forced to undergo replication by passaging and selection. By contrast, our study only involved transient Top1 knockdown and caused a strong G0/G1 cell cycle block (Fig. 6; see below). Given that passage through S phase is required for R-loop-induced DNA breakage and instability phenotypes [2, 14], the reduction of cells in S phase likely counteracted the accumulation of DNA damage in our cell model. We speculate, however, that RLG genes may represent a source of genomic instability once cells are able to replicate. We also note that Top1 depletion in our system did not result in major changes in gene expression (Additional file 1: Figure S2D) or notable accumulation of RNAP at sites of RLG (data not shown). This indicates that, while transient Top1 depletion caused R-loop accumulation in RLG genes, gene expression still proceeded mostly unchanged. It is possible that the loss of Top1 activity, particularly in removing positive supercoils that might hinder RNAP progression, was compensated by the redundant activity of Top2. Recent evidence indeed shows that genes with high transcriptional outputs require Top2 activity to properly handle the resulting torsional stress [44]. Thus, unlike widely held views, Top1 depletion does not result in a global R-loop increase but rather affects a specific subset of genes. This study identifies RLG genes as uniquely Top1-responsive and reveals the molecular features that render these genes dependent on Top1 for R-loop control.

Unexpectedly, Top1 depletion also led to R-loop losses over a class of genes entirely distinct from RLG genes. RLL genes were of average length, resided in gene-rich neighborhoods, and were moderately expressed (Figs. 2, 4). The most striking feature of RLL peaks was their tendency to co-localize with replication initiation regions as defined either by SNS-seq or OK-seq (Fig. 5). This co-localization was underscored by the fact that RLL loci showed higher SNS-seq signal than matched or RLG loci. Initiation regions highlighted by their RLL overlap replicated early (predominantly G1), earlier than other Top1-invariant R-loop forming loci matched for gene expression and gene

densities. Studies of replication origins in mammalian systems indicate that early origins are characterized by marks of open chromatin and by transcription [50–52]. Fittingly, co-transcriptional R-loops preferentially associate with increased DNase accessibility, histone H3 acetylation, and histone H3 lysine 4 methylation [4, 22]. Top1-responsive RLL peaks further include a significant association with the H3K27me3 Polycomb mark and components of the PRC complexes (Fig. 5a). Interestingly, a subset of early, highly efficient replication origins was previously associated with a very similar chromatin pattern [40, 51]. Thus RLL peaks correspond to Top1-responsive R-loop forming loci that are enriched over a subset of early active replication origins and preferentially carry chromatin marks previously defined for these loci.

Interestingly, a primary cellular response to Top1 depletion is the accumulation of cells in G1/G0 and a delay in replication timing at certain genomic loci. One possible mechanism to account for this observation is if Top1 and R-loops participate in origin function. Top1 is known to bind to replication origin sequences [39, 53] as part of the replication progression complex (RPC) which comprises the MCM and GINS proteins [54]. In the SV40 system, almost all RPC components are individually dispensable for activation of SV40 origin in crude extracts, except for Top1 and its interaction with the T antigen for the priming of viral replication [55, 56]. Top1 is therefore a part of the basal complex responsible for origin activation and nascent fork formation. Furthermore, Top1 DNA cleavage sites have been mapped at the lamin B2 origin and Top1 inhibition by low camptothecin concentrations abolished origin firing, suggesting that Top1 and DNA topology play a key role in this process [53]. A plausible hypothesis is that the catalytic activity of Top1 is necessary at replication origins to remove the positive, but not negative, supercoils generated by the unwinding of DNA mediated by MCM helicases [57], leaving the DNA template more negatively supercoiled and thus favoring DNA strand separation. If so, the absence of Top1 will cause the inefficient removal of positive supercoils which in turns will disfavor R-loop formation and cause the appearance of RLL loci. Our work therefore highlights that Top1 may play an important role in modulating replication origin function in human cells at a subset of early origins. It nonetheless remains possible that Top1 depletion may affect replication timing and cell cycle progression in an indirect and more complex manner; further investigations will be necessary to fully define the molecular mechanisms linking Top1 and replication origin activity.

In addition, our work also raises the possibility that R-loop formation may be linked to replication origin function in human cells. The notion that R-loops may contribute to origin function is supported by a wide array of observations. As mentioned above, R-loop forming regions associate with chromatin signatures that are typical of replication origins. R-loops and origins both show hotspots of distribution at gene ends [42, 58]. CpG island promoters in particular, are R-loop and origin hotspots [5, 52, 59–63], and associate with conserved patterns of GC skew [58, 64], a sequence characteristic that intrinsically favors the formation of G-rich signatures often referred to as origin G-rich repeated elements [65]. Such G-rich motifs have the potential to form G quadruplex structures that have been implicated as determinants of origin positioning and efficiency [65, 66]. While it is unclear if G quadruplex can spontaneously nucleate in the context of double-stranded DNA, it is reasonable to propose that R-loop structures can favor G4 formation on the looped out single-strand [67]. Interestingly, the ORC1 subunit of the origin Recognition Complex was shown to bind G4-preferrable ssDNA [68], thereby suggesting that R-loop formation may favor origin licensing. Several historical precedents further underscore the connections between R-loops and origins. In the T4 bacteriophage and in ColEI-replicons in *E. coli*, R-loops function as replication origins [25–27, 69]. In *E. coli*, recombination-mediated R-loops in RNase H-deficient strains support an OriC-independent mode of replication [70, 71]. Increased R-loop formation in RNase H-deficient yeast strains subjected to Top1 inhibition led to origin-independent DNA replication initiation in the rDNA [28]. Finally, the mitochondrial genome is thought to initiate DNA replication priming through R-loop intermediates [72–74] and a recent study showed that replication origins are specified in an R-loop dependent manner at murine class switch immunoglobulin regions [75]. Thus, as judged from location overlaps, chromatin features, and functional associations, our work is consistent with an intimate connection between R-loop formation and replication origin specification [76]. Future work will be necessary to delineate the detailed mechanistic connections that link transcription, R-loop formation, topoisomerase activity, and replication initiation.

## Conclusions

Altogether, our work establishes that Top1 regulates R-loop formation in a context-dependent manner. Long, highly transcribed genes for which supercoil dissipation is not possible due to physical anchoring were particularly susceptible to Top1 depletion and responded by gaining R-loops. By contrast, a class of loci overlapping with efficient early replication origins showed an unexpected loss of R-loops upon Top1 depletion. Many genes in addition, showed a mixed response including R-loop gains and losses. This shows that unlike what was previously believed, Top1 exerts subtle effects on genomic R-loop formation, and highlights the importance of using precise R-loop genomic mapping technologies to

Manzo *et al. Genome Biology* (2018) 19:100

Page 14 of 18

determine the effect of a given factor on R-loop metabolism.

## Methods

### Cell Lines and Drugs
HEK293 cells (ATCC) were maintained in DMEM (Thermofisher) supplemented with 10% FBS in a humidified incubator at 5% of $CO_2$. Camptothecin (Sigma Aldrich) was dissolved in DMSO at 10 mM concentration, stored in aliquots at $-20°$, and used as a 1,000× stock during 1 h treatments.

### Top1 Knockdown
HEK293 cells were counted and seeded at 150,000 cells per 35 mm dish. 24 h after seeding, cells were reverse transfected using RNAimax transfection reagent (Thermofisher) and with 10 nM of Top1-specific validated siRNA (ThermoFisher) targeting exon 16 (siRNA #1; Cat: S14304) and exon 15 (siRNA #2; Cat: S14306) of the nuclear Top1 transcript, or with a negative control RNA (scramble; Cat: AM4613). 48 h after the first transfection, one fifth of the cells were transfected again in a similar manner. Cells were harvested 72 h after the second transfection for all subsequent analysis. Knockdown was verified by Western blot (Fig. 1) and at the RNA level by RNA-seq. The mitochondrial Top1 enzyme (TOP1MT) was not affected by the knockdown as measured by RNA-seq.

### Dot Blot Analysis
Genomic DNA was extracted according to DRIP protocol and digested with restriction enzyme cocktail mix. Two-fold serial dilutions starting from 7.5 micrograms of DNA were spotted on a nitrocellulose membrane and crosslinked with UV light (120 mJ/cm$^2$). Membrane was blocked with PBS-Tween (0.1%) and 3% BSA for 30 min and then incubated with S9.6 antibody diluted to 1 µg/ml in PBS-Tween (0.1%), 3% BSA. After washing, membrane was incubated with HRP-conjugated or Alexa-fluor 488 anti-mouse secondary antibodies, further washed and developed with ECL techniques or directly in fluorescence scanning. In case of treatment with RNase H genomic DNA was pre-incubated with 10 U of enzyme for two hours at 37 °C. To ensure equal loading, we systematically withdrew an aliquot of DNA prior to application on the membrane and loaded it on an agarose gel. Densitometry was used to confirm that all samples were equally digested and of equal intensities (Additional file 1: Figure S1A).

### Western Blot
Western blot analysis was performed according to standard procedures. Membranes were incubated with the following antibodies: anti Top1 (c15, sc5342), anti beta-actin (I-19, sc1616), anti p-ATM (10H11.E12, sc47739), anti

histone H1 (AE-4, sc8030) from Santa Cruz Biotechnology. Anti Phospho-H2AX antibody (ser139, JBW301) was from Millipore. Anti Phospho-ChK1 (Ser345, 133D3) and anti Phospho-Chk2 (Thr68, C13C1) were from Cell Signaling.

### DRIP and DRIPc-seq
DRIPc-seq was performed as previously described [4]. Briefly, DRIP immunoprecipitates obtained from 40 micrograms of digested genomic DNA were collected and treated with DNase I (Fermentas). The resulting RNA strands were purified and reverse-transcribed using the iScript kit (Bio-Rad). Second strand synthesis was performed using dUTP instead of dTTP. Ligation of Illumina Truseq adapters was performed according to manufacturer's instructions and a UDG glycosylase treatment was introduced before library amplification to permit strand-specific R-loop detection. In case of treatment with RNase H or RNase A, digested genomic DNA was pre-treated with 10 units of RNase H or 10 µg/ml of RNase A for two hours at 37 °C before DRIP.

### RNA Pol II ChIP-seq and total RNA-seq
RNA Pol II ChIP was performed as previously described [77]. Immunoprecipitated DNA was purified and used to construct Illumina NGS libraries according to manufacturer procedures. Total RNA-seq was performed after ribosomal RNA depletion using an Illumina Truseq RNA-seq kit according to the manufacturer's instructions.

### DRIPc-seq, RNA-seq, and RNA Pol II ChIP-seq Mapping and Peak Calling
Sequenced single-end reads were subjected to standard quality control pipeline using fastq-mcf software and mapped using Tophat2 for RNA-seq and Bowtie2 for the rest with default parameters. Sequencing read depths were normalized by number of mapped reads between samples, and only uniquely mapped reads were considered. High copy-number or contamination-prone regions such as rDNA, mitochondria, centromere, and ENCODE blacklisted regions were excluded. DRIPc-seq peak calling was performed using a previously developed Hidden Markov Model [4] modified to enable higher sensitivity in particular when dealing with lower and trailing signal (see https://github.com/srhartono/highsenshmm). This method was about 2.5-fold more sensitive, generating about 200,000 peaks of signal covering about 500 MB of genomic space in each replicate. For analysis, all DRIPc-seq peaks present in at least one sample were considered and regions showing significant differences in signal between Top1-depleted and control cells were identified using DESeq2 using significance thresholds of an adjusted *p*-value < 0.1 and signal fold-change higher than 1.25× or lower than 0.8× (using a more stringent adjusted p-value < 0.05 did not affect our

Manzo *et al. Genome Biology* (2018) 19:100

Page 15 of 18

conclusions; data not shown). Genes shorter than 5 kb were eliminated from this analysis.

## Overlap Analysis with Other Datasets

Datasets for lamin, chromatin marks, ChromHMM states, SNS-seq and OK-seq replication origins or zones were downloaded from published sources. The RNAP pausing state of each gene was categorized as in [78] using RNA Pol II ChIP-seq datasets generated from control HEK293 cells. The enrichment or depletion of RLL and RLG peaks over chromatin features of interest was first measured in terms of peak overlap. For this, we determined the overlap of RLL and RLG peaks over chromatin peaks of interest and then calculated the peak overlap for control peaks. These control peaks were stringently selected following an earlier strategy [4]. In brief, these peaks belonged to expression- and length-matched R-loop forming genes that were not affected by Top1 depletion (no and mixed changes in Fig. 2a). In all cases, these peaks were matched to a similar-sized R-loop peak on the matched gene. In the case of promoters and terminators, the precise position of the initial and shuffled peaks was maintained. Each initial RLL or RLG peak was independently matched multiple times to avoid outliers. We next determined the ratio of overlaps between RLL or RLG peaks and control peaks and expressed this ratio as a heatmap. The absolute overlap of chromatin features with RLL and RLG peaks is indicated by stars, as shown in Figs. 4 and 5. SNS-seq origin peaks were from [40] for human K562 and HeLa cells. OK-seq data was downloaded as RFD values from [42] for human HeLa cells. The RFD signal was processed using an HMM model configured as described by Petryk et al. [42] to call replication initiation zones. Overlap between SNS-seq origins and OK-seq initiation zones was measured relative to stringent matched controls. Given that R-loop mapping was performed in HEK293 cells, it is likely that the overlap with replication initiation regions was under-estimated.

## Immunofluorescence

48 h after second round of transfection with siRNA oligonucleotides, cells were detached and seeded at 300,000 cells per 35 mm dish on a glass coverslip pre-treated with gelatin. 24 h after seeding, cells were methanol fixed and treated with acetone. Blocking and ki-67 (Abcam, ab15580) or γH2AX antibody incubation were performed in 4× SSC and 3% BSA at 20 °C for 30 min and 2 h, respectively. Secondary antibody was anti-rabbit or anti-mouse Alexa-fluor 488. Nuclei were counter-stained with DAPI.

## Cell Cycle Analysis and Replication Timing

Cell cycle analysis and replication timing were performed as described previously [79]. Briefly, cells were pulse-labeled with BrdU (50 μM) for two hours. Cells were then harvested, fixed in 70% ethanol, and stored at − 20 °C. Before cell cycle analysis and sorting, cells were labeled with Propidium Iodide (50 μg/ml) and treated with RNase A (250 μg/ml). Cells were analyzed and sorted with a Biorad S3e cell sorter. After sorting, cells were lysed and genomic DNA was extracted. DNA was immunoprecipitated with anti BrdU antibody (B44, BD Biosciences, 347,580), purified, and used as template in qPCR. To assess R-loop formation across the cell cycle, cells were first synchronized in G2 after thymidine block (24 h) and released into nocodazole-containing media (12 h). Cells were then allowed to cycle in fresh medium and harvested in G1 and mid-S of the following cycle for DRIP-qPCR analysis. Cytofluorometric analysis after propidium iodide staining confirmed that > 85% of the cells were in the correct cell cycle phases.

## Additional files

**Additional file 1: Figure S1. A**. Quantitation of R loop signal from dot blot shown in Fig. 1. Error bars are SE of three independent experiments. A representative DNA loading control is shown at right after agarose gel electrophoresis. **B**. Dot blot and relative quantitation to measure R loop loads in genomic DNA from control or Top1 depleted cells, with and without RNase H pre-treatment. **C**. Dot blot depicting change of global R-loop signal 4, 5, and 6 days after initial Top1 depletion. Samples are indicated at left. **D**. Distribution of DRIPc signal (right) across genomic compartments compared to expected distribution in the genome (left). Color codes are as described over Fig. 1c. **E**. Genome browser screenshot illustrating signal "spreading" in the Top1-depleted samples. The R-loop signal for the (+) strand is shown in red. The high-sensitivity peak calls developed to capture signal spreading are found below each track. **F**. XY plot of average log(10) count-normalized DRIPc-seq signals for control samples (scramble − x axis) and top1-depleted samples (y-axis). **G**. Validation by DRIP-qPCR of R-loop gain and loss loci identified by DRIPc-seq. Error bars are SE of three independent experiments. **H**. Validation by DRIP-qPCR R-loop gain (left) and loss (right) upon depletion of Top1 by a second, independent siRNA. The inset above shows a Western blot verifying Top1 depletion. **I**. Validation by DRIPqPCR of RLL and RLG loci identified by DRIPc-seq 5 and 6 days after Top1 depletion. Error bars represent SE of 2 independent experiments. **J**. Validation of R-loop loss and gain loci identified by DRIPc-seq using S9.6-independent DRIVE-pPCR method. The average and standard deviation of two independent replicates is shown. The *RPL13A* locus represents an invariant control. **K**. DRIPqPCR analysis of R-loop formation over the 5′ ETS and 28S rDNA regions. Results are average of 3 independent experiments shown with standard deviation. RNase A and RNase H pre-treatments are indicated below. **L**. DRIPqPCR analysis of R-loop formation over the 5′ ETS and 28S rDNA regions with a second siRNA against Top1. Results are average of 2 independent experiments shown with SEM. **Figure S2.** Examples of DRIPc profiles for control and Top1-depleted cells for genes showing gain and loss (**A**) or mixed R-loop changes (**B**) after Top1 knockdown. Stars indicate statistically significant differences. **C**. Ratio plots of the RNA polymerase II ChIP-seq signal between Top1-depleted and control cells around the TSS of specific gene categories according to expression and pausing status. Top1 depletion causes an increase in RNAPII levels around the TSS of paused genes. Peak shape differences between our and previous studies 30 are likely due to the use of different antibodies (we used a pan-RNAPII Ab whereas others used an anti-S5P RNAPII Ab). This allowed us to observe progressive RNAPII accumulation downstream of the TSS, consistent with RNAPII encountering difficulty during elongation in the absence of Top1. **D**. Venn diagrams depicting the overlap between RLG and RLL genes and genes undergoing up or down regulation in Top1-depleted cells. Differentially expressed genes were identified with a 1.5-fold up or down minimal threshold (and adjusted *p*-value< 0.05). **Figure S3. A**. Distance between the top 100 RLG and RLL peaks (as measured

Manzo *et al. Genome Biology* (2018) 19:100

Page 16 of 18

to signal gains and losses) and LADs compared to matched controls. Statistical significance was measured by Wilcoxon test. **B**. Metaplots of Lamin B1 signal over the promoters and terminals of RLG and RLL genes measured over a +/− 50 kb window. Genes are aligned at their TSS or polyadenylation sites (PAS). Values are median and shown with standard deviation (shaded). **Figure S4. A**. Heatmap of enrichment or depletion of RLG and RLL peaks over specific chromatin features as measured by peak overlap (CTCF, Rad21, Insulator, CpG Islands, and ORC1). Color codes and description is as Fig. 4a. Stars indicate the extent of overlap between R-loop peaks and each chromatin feature (* 10–25%; ** 25–50%; *** > 50%; no star < 10%). All values are significant with p-value < 0.08 (Monte-Carlo) **B**. Heatmap of distance of RLG and RLL peaks from OK-seq, SNS-seq, and CpG island peaks. The numbers indicate the median distance (in kb) from RLG and RLL peaks and each feature. **C**. Average density of OK-seq replication initiation zones (AS) were plotted around promoter and terminal regions for RLL, RLG and corresponding matched invariant genes. Red arrows indicate loss of origin density for RLG genes compared to matched genes. Blue arrows indicate gain of origin densities compared to matched genes. (PDF 3624 kb)

**Additional file 2: Figure S5.** Validation of cell cycle block using an independent Top1 siRNA. **A**. Representative images of ki-67 staining for Top1-depleted and control cells. Cells were counter-stained with DAPI. **B**. Quantification of ki-67 staining (350 cells for each sample from one representative experiment). **C**. Cell cycle analysis for Top1-depleted cells and controls. Results are from one representative experiment. **D**. Cytofluorimetric profile for control and Top1-depleted cells. X-axis is propidium iodide signal, y-axis is cell count. **Figure S6.** (**A**) R-loop levels measured by DRIP-qPCR at nine different loci across the G2/M, G1 and S phases of the cell cycle are showed here normalized to R-loop levels in asynchronous cells. Results are shown as average and standard deviation from two experiments. (**B**) Re-analysis of R-loop mapping data from MCF7 cells before and 2 and 24 h after stimulation with estradiol (data from GSE81851). Top: overlap between R-loop peaks and RLG and RLL genes. Bottom: Total length covered by R-loop peaks in RLG and RLL peaks. (**C** and **D**) Analysis of replication timing at a range of Mixed/no change loci (**C**) and RLG loci (**D**) in Top1-depleted cells and controls. Recovery of BrdU-labeled immuno-precipitated DNA for each phase was normalized based on total signal for each sample. Error bars are SE of two replicates. Red and grey shading indicate genes with significant and non-significant replication timing delays, respectively. **Table S1.** Number of R-loop peaks and genes after peak calling with the standard method [4] or the high sensitivity method implemented here. The numbers are broken down between loci that undergo R-loop gains, Rloop losses, no change, or both. is propidium iodide signal, y-axis is cell count. (PDF 17536 kb)

**Additional file 3:** Review history. (DOCX 320 kb)

## Data access

All high-throughput sequencing data have been deposited into NCBI GEO entry GSE102474. Publicly available datasets used are as follows: Lamin-associated domain [38], ChromHMM [36], ENCODE histones [37], SNS-seq [40], and OK-seq [42] coordinates.

## Review history

The review history is available as Additional file 3.

## Authors' contributions

SGM, FC, and GC designed research. SGM and LS performed experiments. SRH performed bioinformatics analysis. SDB and AC performed cell sorting analysis. SGM, SRH, FC, GC analyzed the data and wrote the paper. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

These are not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

[1]Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy. [2]Department of Molecular and Cellular Biology and Genome Center, University of California, Davis, USA. [3]Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, Modena, Italy. [4]Present address: Division of Gene Regulation, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

## References

1. Bermejo R, Lai MS, Foiani M. Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. Mol Cell. 2012;45:710–8.
2. Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. Mol Cell. 2012;46:115–24.
3. Hamperl S, Cimprich KA. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. DNA Repair (Amst). 2014;19:84–94.
4. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, Chedin F. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell. 2016;63:167–78.
5. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012;45:814–25.
6. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A. 1987;84:7024–7.
7. Drolet M, Bi X, Liu LF. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. J Biol Chem. 1994;269:2068–74.
8. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol. 2002;3:430–40.
9. Baranello L, Kouzine F, Levens D. DNA topoisomerases beyond the standard role. Transcription. 2013;4:232–7.
10. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. Annu Rev Biochem. 2001;70:369–413.
11. Capranico G, Marinello J, Chillemi G. Type I DNA Topoisomerases. J Med Chem. 2017;60:2169–92.
12. Masse E, Phoenix P, Drolet M. DNA topoisomerases regulate R-loop formation during transcription of the rrnB operon in Escherichia coli. J Biol Chem. 1997;272:12816–23.
13. Drolet M, Phoenix P, Menzel R, Masse E, Liu LF, Crouch RJ. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc Natl Acad Sci U S A. 1995;92:3526–30.
14. Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol. 2009;11:1315–24.
15. Pommier Y. Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer. 2006;6:789–802.
16. Marinello J, Bertoncini S, Aloisi I, Cristini A, Malagoli Tagliazucchi G, Forcato M, Sordet O, Capranico G. Dynamic Effects of Topoisomerase I Inhibition on R-Loops and Short Transcripts at Active Promoters. PLoS One. 2016;11:e0147053.
17. Marinello J, Chillemi G, Bueno S, Manzo SG, Capranico G. Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. Nucleic Acids Res. 2013;41:10110–23.
18. Sordet O, Redon CE, Guirouilh-Barbat J, Smith S, Solier S, Douarre C, Conti C, Nakamura AJ, Das BB, Nicolas E, et al. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. EMBO Rep. 2009;10:887–93.

Manzo *et al. Genome Biology* (2018) 19:100

Page 17 of 18

19. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. Nat Rev Genet. 2015;

20. Costantino L, Koshland D. The Yin and Yang of R-loop biology. Curr Opin Cell Biol. 2015;34:39–45.

21. Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 2014;28:1384–96.

22. Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat Struct Mol Biol. 2015;22:999–1007.

23. Proudfoot NJ. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science. 2016;352:aad9926.

24. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol. 2003;4:442–51.

25. Carles-Kinch K, Kreuzer KN. RNA-DNA hybrid formation at a bacteriophage T4 replication origin. J Mol Biol. 1997;266:915–26.

26. Masukata H, Tomizawa J. A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. Cell. 1990;62:331–8.

27. Itoh T, Tomizawa J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc Natl Acad Sci U S A. 1980;77:2450–4.

28. Stuckey R, Garcia-Rodriguez N, Aguilera A, Wellinger RE. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. Proc Natl Acad Sci U S A. 2015;112:5779–84.

29. Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods. 1986;89:123–30.

30. Phillips DD, Garboczi DN, Singh K, Hu Z, Leppla SH, Leysath CE. The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. J Mol Recognit. 2013;26:376–81.

31. El Hage A, French SL, Beyer AL, Tollervey D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. 2010;24:1546–58.

32. Baranello L, Wojtowicz D, Cui K, Devaiah BN, Chung HJ, Chan-Salis KY, Guha R, Wilson K, Zhang X, Zhang H, et al. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. Cell. 2016;165:357–71.

33. Khobta A, Ferri F, Lotito L, Montecucco A, Rossi R, Capranico G. Early effects of topoisomerase I inhibition on RNA polymerase II along transcribed genes in human cells. J Mol Biol. 2006;357:127–38.

34. Baranello L, Bertozzi D, Fogli MV, Pommier Y, Capranico G. DNA topoisomerase I inhibition by camptothecin induces escape of RNA polymerase II from promoter-proximal pause site, antisense transcription and histone acetylation at the human HIF-1alpha gene locus. Nucleic Acids Res. 2010;38:159–71.

35. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008;322:1845–8.

36. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9:215–6. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHmm/wgEncodeBroadHmmK562HMM.bed.gz

37. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74. https://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeBroadHistone

38. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453:948–51. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/laminB1Lads.txt.gz

39. Falaschi A. Binding of DNA topoisomerases I and II to replication origins. Methods Mol Biol. 2009;582:131–43.

40. Picard F, Cadoret JC, Audit B, Arneodo A, Alberti A, Battail C, Duret L, Prioleau MN. The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet. 2014;10:e1004282. http://pbil.univ-lyon1.fr/members/fpicard/oriseq/

41. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A. 2010;107:139–44.

42. Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, Silvain M, Thermes C, Chen CL, Hyrien O. Replication landscape of the human

genome. Nat Commun. 2016;7:10208. http://157.136.54.88/cgi-bin/gbrowse/gbrowse/okazaki_ref/

43. Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chedin F, Swigut T, Cimprich KA. Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife. 2016;5:e17548.

44. Kouzine F, Gupta A, Baranello L, Wojtowicz D, Ben-Aissa K, Liu J, Przytycka TM, Levens D. Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat Struct Mol Biol. 2013; 20:396–403.

45. Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, Eijk PP, Cockroft SL, Bradley M, Ylstra B, Gilbert N. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat Struct Mol Biol. 2013;20:387–95.

46. Solier S, Ryan MC, Martin SE, Varma S, Kohn KW, Liu H, Zeeberg BR, Pommier Y. Transcription poisoning by Topoisomerase I is controlled by gene length, splice sites, and miR-142-3p. Cancer Res. 2013;73:4830–9.

47. King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, et al. Topoisomerases facilitate transcription of long genes linked to autism. Nature. 2013;501:58–62.

48. Gonzalez-Sandoval A, Towbin BD, Kalck V, Cabianca DS, Gaidatzis D, Hauer MH, Geng L, Wang L, Yang T, Wang X, et al. Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos. Cell. 2015;163:1333–47.

49. Miao ZH, Player A, Shankavaram U, Wang YH, Zimonjic DB, Lorenzi PL, Liao ZY, Liu H, Shimura T, Zhang HL, et al. Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses. Cancer Res. 2007;67:8752–61.

50. Fragkos M, Ganier O, Coulombe P, Mechali M. DNA replication origin activation in space and time. Nat Rev Mol Cell Biol. 2015;16:360–74.

51. Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Mechali M. The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res. 2015;25:1873–85.

52. Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M. Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet. 2009;5:e1000446.

53. Abdurashidova G, Radulescu S, Sandoval O, Zahariev S, Danailov MB, Demidovich A, Santamaria L, Biamonti G, Riva S, Falaschi A. Functional interactions of DNA topoisomerases with a human replication origin. EMBO J. 2007;26:998–1009.

54. Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol. 2006;8:358–66.

55. Simmons DT, Melendy T, Usher D, Stillman B. Simian virus 40 large T antigen binds to topoisomerase I. Virology. 1996;222:365–74.

56. Wold MS, Weinberg DH, Virshup DM, Li JJ, Kelly TJ. Identification of cellular proteins required for simian virus 40 DNA replication. J Biol Chem. 1989;264:2801–9.

57. Falaschi A, Abdurashidova G. Molecular mechanics and DNA replication regulation. HFSP J. 2007;1:215–9.

58. Ginno PA, Lim YW, Lott PL, Korf I, Chedin F. GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res. 2013;23:1590–600.

59. Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, Quesneville H, Prioleau MN. Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A. 2008;105:15837–42.

60. Necsulea A, Guillet C, Cadoret JC, Prioleau MN, Duret L. The relationship between DNA replication and human genome organization. Mol Biol Evol. 2009;26:729–41.

61. Delgado S, Gomez M, Bird A, Antequera F. Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. 1998;17:2426–35.

62. Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Mechali M. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. 2011;21:1438–49.

63. Karnani N, Taylor CM, Malhotra A, Dutta A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol Biol Cell. 2010;21:393–404.

Manzo *et al. Genome Biology* (2018) 19:100

Page 18 of 18

64. Hartono SR, Korf IF, Chedin F. GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. Nucleic Acids Res. 2015;43(20):9729-41.

65. Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, Mechali M. New insights into replication origin characteristics in metazoans. Cell Cycle. 2012;11:658–67.

66. Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintome C, Riou JF, Prioleau MN. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. 2014;33:732–46.

67. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. 2004;18:1618–29.

68. Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, Nakatsuka A, Kunichika T, Obuse C, Waga S. Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J Biol Chem. 2013;288:30161–71.

69. Kreuzer KN, Brister JR. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol J. 2010;7:358.

70. Kogoma T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol Mol Biol Rev. 1997;61: 212–38.

71. Maduike NZ, Tehranchi AK, Wang JD, Kreuzer KN. Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. Mol Microbiol. 2014;91:39–56.

72. Akman G, Desai R, Bailey LJ, Yasukawa T, Dalla Rosa I, Durigon R, Holmes JB, Moss CF, Mennuni M, Houlden H, et al. Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria. Proc Natl Acad Sci U S A. 2016;113:E4276–85.

73. Lee DY, Clayton DA. Initiation of mitochondrial DNA replication by transcription and R-loop processing. J Biol Chem. 1998;273:30614–21.

74. Xu B, Clayton DA. A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. Mol Cell Biol. 1995;15:580–9.

75. Wiedemann EM, Peycheva M, Pavri R. DNA Replication Origins in Immunoglobulin Switch Regions Regulate Class Switch Recombination in an R-Loop-Dependent Manner. Cell Rep. 2016;17:2927–42.

76. Lombrana R, Almeida R, Alvarez A, Gomez M. R-loops and initiation of DNA replication in human cells: a missing link? Front Genet. 2015;6:158.

77. Manzo SG, Zhou ZL, Wang YQ, Marinello J, He JX, Li YC, Ding J, Capranico G, Miao ZH. Natural product triptolide mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II. Cancer Res. 2012;72:5363–73.

78. Core LJ, Waterfall JJ, Gilchrist DA, Fargo DC, Kwak H, Adelman K, Lis JT. Defining the status of RNA polymerase at promoters. Cell Rep. 2012;2:1025–35.

79. Ryba T, Battaglia D, Pope BD, Hiratani I, Gilbert DM. Genome-scale analysis of replication timing: from bench to bioinformatics. Nat Protoc. 2011;6:870–95.