*Bioinformatics*, YYYY, 0–0

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Manuscript Category

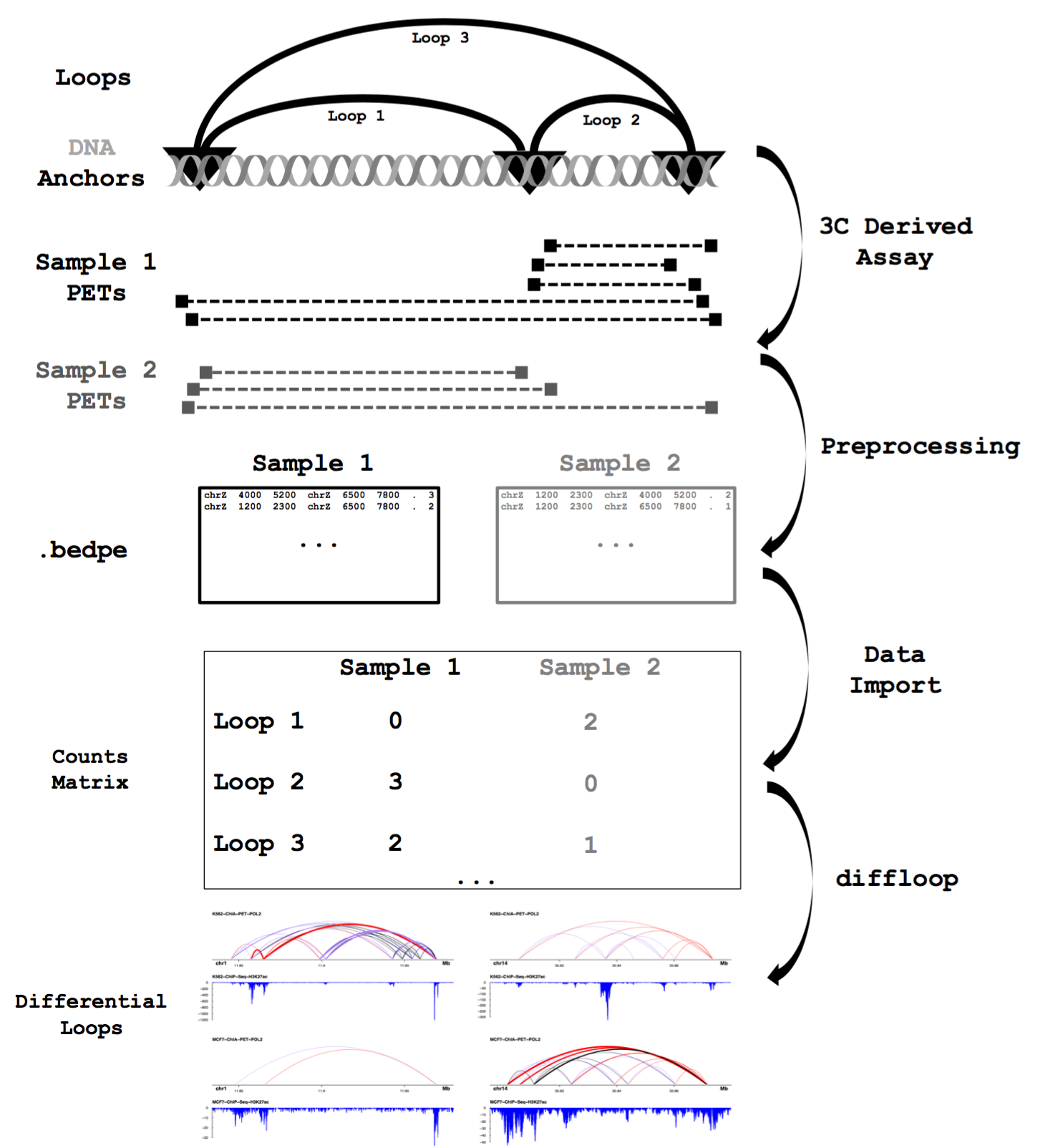
|  |
| --- |
| Genome Analysis  diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data  Caleb A. Lareau1,2,3 and Martin J. Aryee1,2,3,\*  1Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Ave. Boston, MA 02115. USA., 2Department of Pathology, Massachusetts General Hospital, 55 Fruit St. Boston, MA 02114. USA. 3Cell Circuits and Epigenomics Program, Broad Institute of MIT and Harvard, 415 Main St, Cambridge, MA 02142. USA.  \*To whom correspondence should be addressed.  Associate Editor: XXXXXXX  Received on XXXXX; revised on XXXXX; accepted on XXXXX  Abstract  **Motivation:** The three-dimensional architecture of DNA within the nucleus is a key determinant of interactions between genes, regulatory elements, and transcriptional machinery. As a result, differences in DNA looping structure are associated with variation in gene expression and cell state.  **Results:** To assess changes in DNA looping architecture between samples, we introduce *diffloop*, an R/Bioconductor package that provides a suite of functions for the quality control, statistical testing, annotation, and visualization of DNA loops. We demonstrate this functionality by detecting differences in DNA loops between ENCODE ChIA-PET datasets and relate looping to variance in epigenetic states and gene expression.  **Availability:** https://bioconductor.org/packages/release/bioc/html/diffloop.html.  **Contact:** aryee.martin@mgh.harvard.edu  **Supplementary information:** Supplementary data are available at *Bioinformatics* online. |

# Introduction

The organization of DNA within the nucleus into hierarchical three-dimensional (3D) structures plays a key role in regulating gene expression by determining the accessibility of genes to the transcriptional machinery as well as the proximity of genes to their distal regulatory elements. Differences in 3D architecture, such as the presence or absence of “loops” between specific enhancers and their target genes, are associated with transcriptional variation in both normal and disease states (Matharu and Ahituv, 2015). Intriguingly, several recent studies have implicated pathogenic alterations in genome topology with a diverse set of phenotypes including cancers and autoimmune diseases (Flavahan, et al., 2016; Hnisz, et al., 2016; Wang, et al., 2013).

Recent experimental techniques that couple chromatin conformation capture (3C) (Dekker, et al., 2002) with high-throughput sequencing have made the genome-wide identification of 3D interactions feasible. One such technique, Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET), uses chromatin immunoprecipitation to capture interactions mediated by specific structural, regulatory, or transcriptional machinery proteins and is especially suited for high-resolution identification of 3D interactions. In addition, the high-throughput chromosome conformation capture (Hi-C) assay, which yields a potentially complete map of DNA proximity, can also be used to identify DNA loops (Durand, et al., 2016; Rao, et al., 2014).

In order to fully explore the role that 3D genome organization plays in determining normal and pathogenic cell states, statistical tools are needed to identify differences in topology in a similar manner to which differential expression analysis is applied to transcriptional data. Additionally, the systematic integration of biological prior knowledge, such as the location of active enhancer regions, into topology analyses can provide annotation and insight into the regulatory role of a loop. Examples of loop annotation categories include enhancer-promoter (Matharu and Ahituv, 2015) and disrupted neighborhoods (Hnisz, et al., 2016). Moreover, a computational framework that facilitates integration of other genomic data (e.g. DNA variation, RNA transcription, DNA and histone modification, etc.) in 3D genome analyses will help resolve the functional implication of DNA looping and relate these topological features to phenotypic variation.



To address these needs, we have developed *diffloop*, an R/Bioconductor (Huber, et al., 2015) package that implements statistical testing for differential DNA looping between samples from ChIA-PET and other 3D assays. *diffloop* additionally provides functionality for quality control and visualization of differential DNA topological features and facilitates a platform for integrative analysis with other types of genomic data. To demonstrate the utility of *diffloop*, we compare differences in DNA topology between three cancer cell lines using RNA Polymerase II ChIA-PET data from the ENCODE project including four samples representing the MCF-7 breast cancer cell line (N=2), the K562 leukemia cell line (N=2) (Consortium, 2012).

# Methods

## Preprocessing

Assays such as ChIA-PET enrich for chromatin interactions where interacting loci, termed anchors, are bound by a protein of interest. Anchors linked by paired-end tags/reads (PETs) represent distal regions of DNA that co-localize in three-dimensional space. The preprocessed data input to *diffloop* consists of interaction counts between putatively interacting anchors, as represented in **Figure 1**. Raw data preprocessing pipelines that produce the standard bed paired-end (.bedpe) input data format. A number of software solutions, including the Mango and ChIA-PET2 pipelines, exist to process raw sequencing reads from a ChIA-PET assay into putative loops in the bedpe format that is imported into diffloop (Li, et al., 2016; Li, et al., 2010; Paulsen, et al., 2014; Phanstiel, et al., 2015). These tools typically consider interactions for a single sample at a time when identifying statistically significant loops. For the purposes of differential analysis, it is important to make the determination of valid 3D interactions using all samples collectively on order to define a common set of loops. We therefore recommend the use of inclusive preprocessing parameters that do not filter out interactions, retaining all data for import and subsequent filtering in diffloop. In Mango, retaining is achieved by setting the reportallpairs flag to TRUE. *diffloop* collates these read counts and assembles a list of anchors and a counts matrix as part of a larger S4 data structure to be utilized in differential association as shown in the **Figure 1**.

When loops data is imported into *diffloop*, the precise anchor loci may very slightly from sample to sample. As defining a common feature space is necessary for differential testing, users have the option to pad loop anchors when using the **loopsMakes()** function. This function first loops over each file to determine the union of loop anchors, merging and padding anchors as specified by the user to define a common feature or loop space. One this set of common loops is established, the PETs that support each loop in each sample are imported and the counts matrix is populated. In effect, the counts matrix provides the level of evidence supporting each putative loop (row) for each sample (column). Under this construct, familiar techniques such as differential testing, normalization, and principal component analyses can be applied to the counts matrix.

**Fig. 1.** **Overview of DNA looping data and its representation.** The ChIA-PET and other 3C-derived assays generate sequencing reads that represent chromatin interactions associated with a protein of interest. Preprocessing involves identifying the interacting loci (loop anchors) and counting the number of reads supporting each interaction. These data are typically represented in .bedpe or a closely related format. The core functionality of *diffloop* imports these preprocessed loop data and assembles a counts matrix as part of a larger structure with metadata that can be used to assess differential looping.

## diffloop data structure

## The core functionality of *diffloop* is designed around loops objects, a novel S4 class implemented in the package. Each loops object contains five slots that provide an efficient storage of three-dimensional data in the R environment. Specifically, the anchors slot contains a GRanges (Lawrence, et al., 2013) object specifying the genomic coordinates of the DNA anchors identified through peak calling; the interactions slot represents each loop (row) by a pair of indices (two columns) specifying its two anchors; the counts matrix summarizes the number of supporting PETs for each loop (row) per sample (column); the colData slot contains per-sample information, such as group labels and normalizing constants; and the rowData slot, which provides per-loop annotation, such as loop width, loop type, and statistical significance measures.

## Differential Loop Calling

To determine differential topological features or loops, *diffloop* implements a version of the negative binomal regression model originally reported in the edgeR package to model PET counts from the loops object counts matrix (Robinson, et al., 2010). To account for variation in sequencing depth, a per-sample size factor is included as an offset in the model, as similarly implemented in DESeq2 (Love, et al., 2014). Loop-wise dispersion estimates are stabilized using an empirical Bayes shrinkage procedure. In effect, the implementation of differential loop calling in *diffloop* applies the RNA-Seq gene read count test from edgeR to loop PET counts. We’ve additionally implemented a modified version of the limma-voom framework for identifying differential loops using precision weights though the negative binomial regression is presently the default in *diffloop*. (Law, et al., 2014)

**2.4 Loop Annotation**

The **annotateLoops()** function classifies each loop as enhancer-promoter, promoter-promoter, enhancer-enhancer, or no special annotation for POL2 loops. Additional annotation for loops mediated by CTCF or cohesin subunits is also currently supported. Moreover, loops that connect distal regulatory elements and gene promoters can be selected using the **keepEPloops()** function. For the analyses described in this manuscript, enhancer regions were defined by a 1kb radius around an H3K27ac peak for either cell type. Other epigenetic markings may also be suitable for nuanced logic in defining promoter and regulatory regions. For our analyses, peaks were downloaded from GEO using the accession numbers provided in **Supplementary Table 1**. Promoter regions were defined as being within a 1kb radius of a RefSeq transcription start site.

**2.5 Integration of 1D Epigenetic and RNA data**

To better characterize the epigenetic and transcriptional correlates associated with differential looping, we’ve implemented several functions in *diffloop* that enable a spectrum of –omics analyses in the context of differential looping. Details of the methods and data used to complement the ENCODE ChIA-PET analysis are described in the Supplement.

# Results

## 3.1 Analysis of ENCODE ChIA-PET Samples

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample Name** | **SRR .fastq**  **Files** | **# unique**  **loops** | **# useful PETs** | **# raw**  **PETs** |
| K562\_r1 | SRR372747 | 31,721 | 95,210 | 25,778,343 |
| K562\_r2 | SRR372748 | 50,072 | 161,053 | 37,685,853 |
| MCF7\_r1 | SRR372741 | 23,329 | 50,216 | 29,098,917 |
| MCF7\_r2 | SRR372742 | 36,199 | 85,640 | 40,474,778 |

All four POL2 ChIA-PET samples were preprocessed from raw reads to loops individually using Mango preprocessing pipeline (Phanstiel et al., 2015). Across the union of our four samples considered for our analyses, we observed 87,456 anchor pairs (loops) involving 24,576 autosomal loci. Individual-level summary statistics are shown in **Table 1.** As we wanted to exclude bias related to copy number variation (CNV) in these cancer genomes, we removed all interactions that were associated with anchors in known CNV regions in either of the MCF-7 or K562 cell lines. CNVs are known to affect chromatin interaction counts (Wu and Michor, 2016) and likely to bias differential looping analysis, especially in cancer genomes. In order to exclude such bias, we only considered interactions where neither anchor overlapped one of the nearly 500 CNV regions for either the MCF-7 or K562 cell lines (Consortium, 2012). Across all samples, we observed that over 9,000 of anchor regions overlapping a CNV region by at least 1 base and removed all loops involving at least one of these anchors, leading to 57,744 CNV-free loops in the K562 and MCF-7 samples.

**Table 1.** Summary Statistics for ChIA-PET from GSE33664

We next filtered the observed interactions to retain only those that support valid loops suitable for differential analysis through a two-step process. First, we identified statistically significant loops by retaining only those anchor pairs whose interaction counts were significantly higher than that expected based on the background chromatin interaction frequency. *diffloop* combines counts across samples and assigns statistical significance to each putative loop using the method developed by Phanstiel *et al*. for the Mango ChIA-PET preprocessing pipeline (Phanstiel, et al., 2015). We retained 23,935 loops in our combined K562 and MCF-7 samples that were significant at an FDR of 0.01 from the Mango model. Finally, we excluded anchor pairs not supported by at least two samples with two PETs each as proposed in other loop analyses (Ji, et al., 2016), retaining 9,320 loops. Retaining loops with sufficiently high PET counts is analogous to filtering for genes that have a reasonable expression level in expression analyses (Love, et al., 2014). After defining a quality-controlled set, we could identify differential loops.

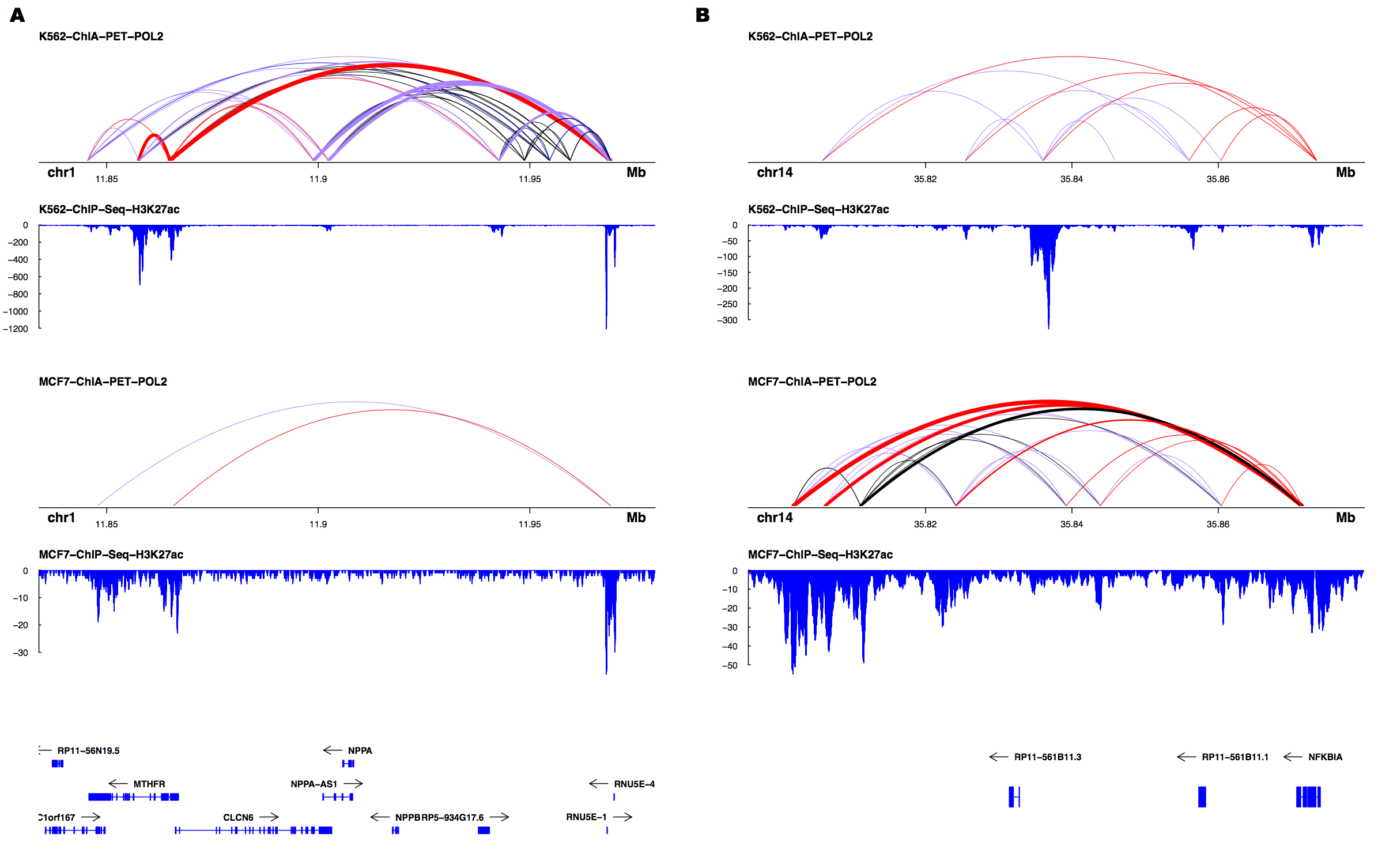
## 3.2 Differential loops between samples

At an FDR of 1%, we identified 2,633 differential loops between the cell types, including 1,974 loops that were annotated as enhancer-promoter loops. **Supplemental Table 2** summarizes the top 5 differential enhancer-promoter loops unique to each cell line. To characterize the topological differences more systematically, we identified pathways enriched for genes involved in differential enhancer-promoter looping. We assessed MsigDB hallmark gene sets (Subramanian, et al., 2005) using a Wilcoxon rank sum test, which yieldednine pathways enriched (FDR q < 0.01) for genes with differential enhancer-promoter loops have existing evidence of relevance to the two cell types and cancers (See Supplement). Genes related to estrogen response such as GREB1 and XBP1 (Sengupta, et al., 2010; Sun, et al., 2007), for example, are linked by several strong loops to unique enhancers in the MCF-7 breast cancer cell line. Conversely, targets associated with c-MYC transcription factor, which plays a well-documented role in leukemia (Delgado and Leon, 2010), and heme metabolism were enriched in K562. These results suggest that differential topology analyses can systematically uncover known and novel regulatory loops related to disease and other phenotypes of interest.

**Figure 2** shows two examples of regions with differential looping with corresponding tracks for the active enhancer mark, H3K27 acetylation (H3K27ac). Panel **(A)** shows a dynamic 3D landscape with many differential loops present more strongly in the K562 leukemia cell line than the MCF-7 breast cancer cell line. Most prominently, several unique enhancer-promoter loops (red) link the *MTHFR* promoter with nearby enhancers. Several of the enhancers are also linked by enhancer-enhancer loops (purple). Consistent with the increased looping the *MTHFR* gene is expressed at a higher level in the K562 (leukemia) cell line relative to MCF-7 (breast cancer) cell line (FDR = 2.95 x 10-11). Notably, variants near this gene have been associated with increased risk for a variety of leukemias.(Moon, et al., 2007; Rossen, et al., 1985) Conversely, Panel **(B)** shows a more active three-dimensional regulatory system in the MCF-7 cell line near the *NFKBIA* gene, which has been linked to breast cancer (Curran, et al., 2002) and is overexpressed in MCF-7 compared to K562 (FDR = 0.0211).

## 3.3 Epigenetic correlates of differential looping

To assess the relationship between differential looping and chromatin state, we compared the MCF-7/K562 log fold-change in PET count to corresponding ratios for DNase hypersensitivity (open chromatin), DNA methylation from the Illumina 450k array, and ChIP-Seq data for RAD21 (a cohesin subunit) and H3K27ac. As it has been suggested that the alteration of one loop anchor is sufficient to disrupt the loop(Ji, et al., 2016), we used the anchor with the greatest difference in epigenetic mark



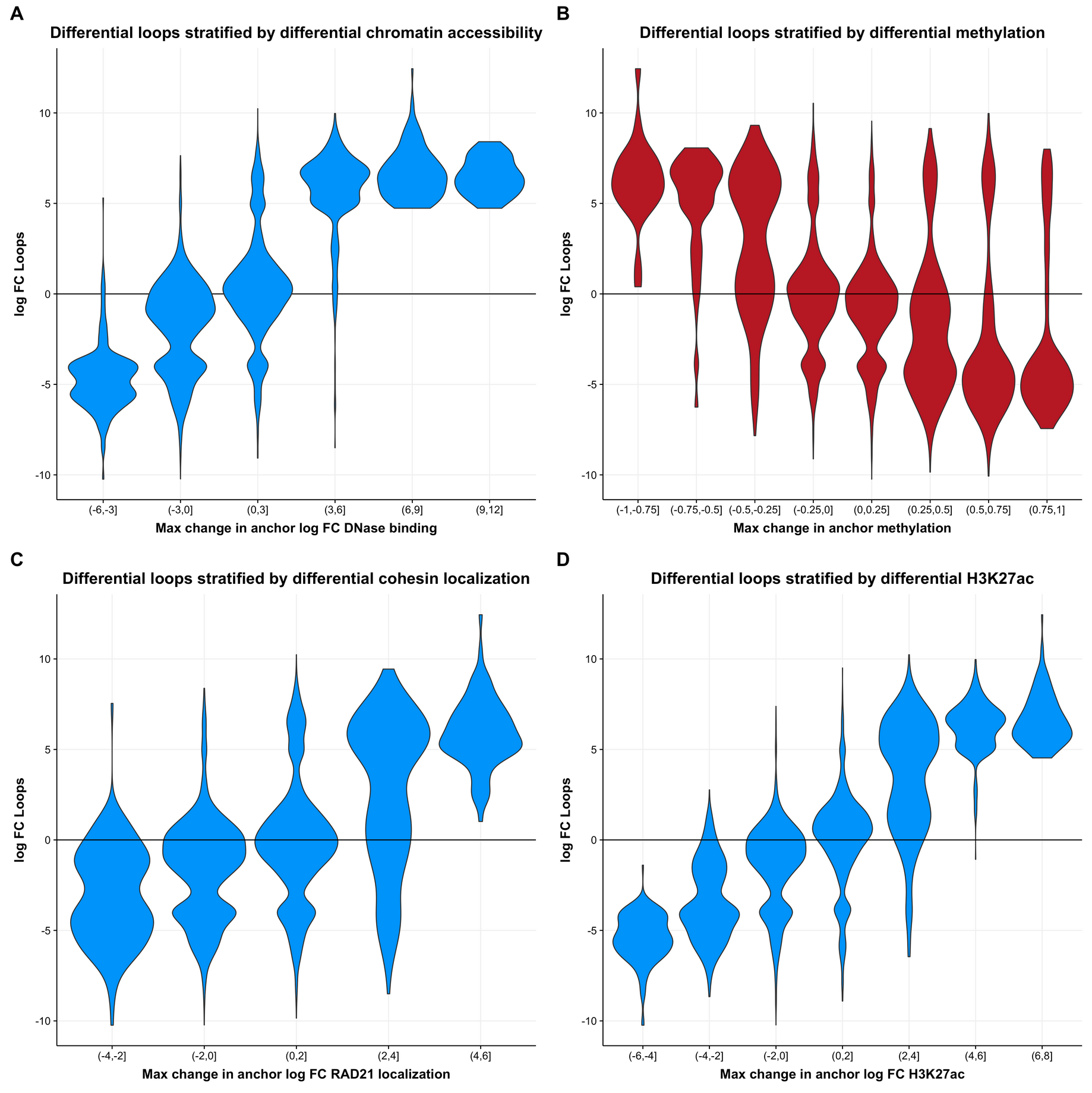
**Fig. 2. Sample visualizations of differential looping**. Each panel shows the combined POL2 ChIA-PET replicates for the K562 and MCF-7 cell lines as well as the cell type-specific H3K27ac ChIP-Seq track. Line widths are indicative of the number of PETs supporting a loop while colors represent biological annotation (red: enhancer-promoter; purple: enhancer-enhancer; black: no special annotation). *MTHFR* (A) and*NFKBIA* (B) have variable topological and epigenetic landscapes between the two cell lines and have been implicated in leukemia and breast cancer, respectively.

in each case. For the sequencing data (ChIP-Seq, DNase), we subtracted the mean of the changes to account for variable total sequencing depth. **Figure 3** summarizes these results, which indicate that increasing loop strength in a cell type is positively correlated with **(A)** open chromatin, **(C)** cohesin localization, and **(D)** active enhancer markings. On the other hand, **(B)** DNA methylation was negatively correlated with loop strength, suggesting that hypermethylation of a genomic locus may inhibit loop formation.

## 3.4 Differential looping and gene expression

Having ascertained that loop strength is tightly correlated with epigenetic state, we next sought to characterize the relationship between looping and transcription with a focus on the role of enhancer-promoter interactions. **Figure 4** **(A)** summarizes differential expression in MCF-7 vs. K562 as a function of loop strength showing that differences in enhancer-promoter loop strength between the cell types is strongly positively correlated with differences in gene expression level. Of the pairs of enhancer-promoter loops/transcripts where both the loop and gene expression were significantly differential (FDR < 0.01; N = 230), 96.1% agreed in direction. We propose then that in the context of enhancer-promoter looping, identifying differential loops between cell states can be a useful means for uncovering correlates of transcriptional change observed in RNA-Seq analyses.

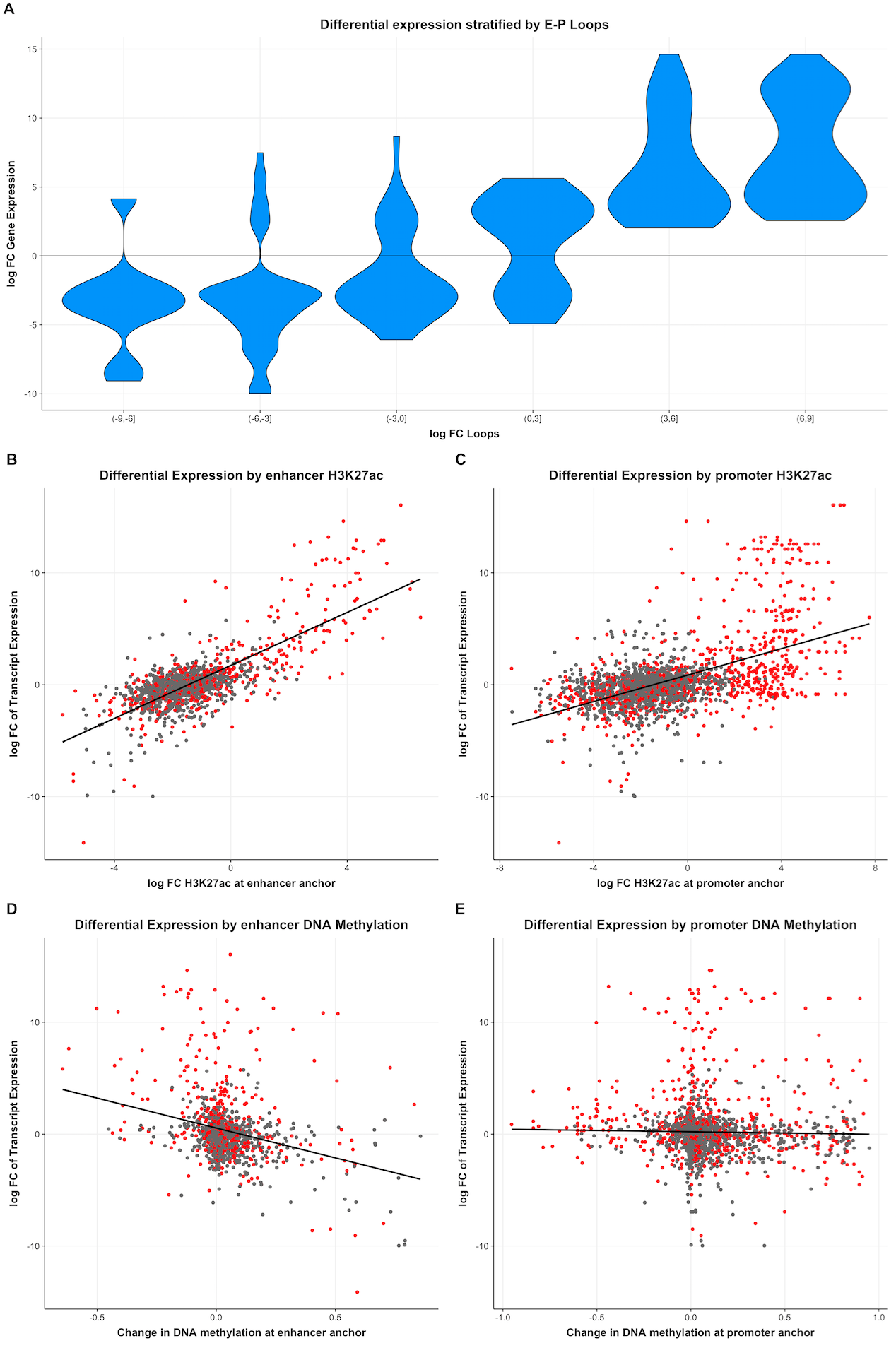
As both H3K27 acetylation and DNA methylation are known to correlate with transcription(Bell, et al., 2011), we sought to investigate whether these marks might play different roles at gene promoters vs. distal enhancers. **Figure 4** shows that, as expected, bothdistal **(B)** and promoter **(C)** changes in H3K27ac are positively correlated with expression

****

**Fig. 3. Differential loop strength correlated with various epigenetic factors.** In each panel, the log2 fold change computed by *diffloop* for K562/MCF-7 is plotted against the log2 fold change of the epigenetic feature averaged over the region defined by the loop anchor. The anchor with the largest deviation from zero was chosen for each loop. Higher loop strength is associated with higher levels of (A) open chromatin, (C) cohesin binding, and (D) H3K27ac and lower levels of DNA methylation (B).

whereas both distal **(D)** and promoter **(E)** changes in DNA methylation were negatively correlated with expression. Interestingly, the effect of epigenetic alterations was far more pronounced at the distal enhancer

region compared to the promoter region. In particular, we note that the variation in enhancer methylation had a higher correlation with gene expression than that of the promoter, emphasizing the role of enhancer-promoter contacts available through looping data. Given that identifying the enhancers associated with a gene is non-trivial, and that over 40% of enhancers skip the closest gene when interacting with their target promoters (Li, et al., 2012), this finding highlights the utility of a genome topology map for integrative analyses of transcription and epigenetics. Differential loops (indicated in red) did not appear to deviate from the overall trend of epigenetic values at enhancers/promoters correlated with the transcriptional outcomes.

****

**Fig. 4**. **Effects of differential looping on transcription**. (A) All loops that uniquely map to a single promoter region of differentially expressed transcript (FDR < 0.01). Increasing enhancer-promoter loop strength is positively correlated with increased transcription. We linked enhancer-promoter loops to their corresponding genes and assessed the effects of H3K27ac (B) distally and (C) proximally to the transcription start site. Similarly, the (D) distal and (E) proximal effects of DNA methylation were plotted against the expression fold-difference of the corresponding gene. Differential loops (FDR < 0.01) are highlighted in red in panels (B) - (E).

**4 Discussion**

A number of recent studies have implicated genome topology alterations in pathogenesis (Flavahan, et al., 2016; Hnisz, et al., 2016; Matharu and Ahituv, 2015; Wang, et al., 2013). As the precise causes and effects of variability in DNA looping between cellular states is not well characterized, computational tools are needed to identify topological differences with statistical rigor and integrate putative regions of differential looping with the wealth of existing -omics data. To address this need, we have presented *diffloop*, an R/Bioconductor package that borrows much of its statistical methodology from differential expression analysis methods for RNA-Seq count data. Our package provides a full suite of functions to identify, annotate and contextualize DNA loops that vary between samples. The implementation of *diffloop* readily integrates with other R/Bioconductor packages and workflows and provides straightforward functionality for integrating genetic, epigenetic, and transcriptional data in the context of variability in the three-dimensional genome.

The base functionality of *diffloop* is designed to import processed data in a form that resembles .bedpe files (as shown in Figure 1) and perform differential loop calling two or more conditions that each have two or more samples. We note that integrating ChIA-PET data from experiments that target different factors is not advised. Moreover, while conditions without replicates can be analyzed, *diffloop* will not assign statistical confidence for differences in this setting. Thus, we focused our analysis on ChIA-PET data from the K562 and MCF-7 cell lines that had replicates of ChIA-PET data against RNA POL2. Other analyses may focus on loops derived from ChIA-PET protocols against other commonly studied factors such as CTCF, SMC1, or RAD21 as well as loops derived from Hi-C and similar chromosome conformation capture protocols. Moreover, *diffloop* can identify topological differences and provide biological annotation of DNA loops for arbitrarily complex experimental designs though we focus on two-group pairwise comparisons for simplicity in this paper.

We first compared the aggregate of the topological loops from ChIA-PET between three types of cancers and determined that cellular phenotypes indeed cluster based on their distinctive topology. While our comparison is close to minimal for statistical inference (2 groups by 2 replicates each), *diffloop* can handle arbitrarily complex designs, providing unique avenues to access the topological changes associated with cell lineages, such as differentiating stem cells (Ji, et al., 2016) or pre-cancerous, oncogenic, and metastatic cell states.

Since chromatin interaction counts are known to be strongly distance dependent, we additionally investigated whether there could be between-sample variation in this relationship. This would suggest making the size factor dependent on loop width or loop PET count, similar to an approach used for Hi-C (Lun and Smyth, 2015). In this dataset, we did not find evidence for such a loop width dependency (Figure S3) and hence decided that a single scale factor correction value for each sample was sufficient for normalization of the examined datasets. While the default method of association in *diffloop* stems from the negative binomial regression model in edgeR, multiple different models, including those mentioned in the supplement are available in the current implementation of the package.

Though the application of the edgeR framework to our counts data was intuitive, we note the mean-variance trend in Figure S1 that fails to show overdispersion beyond that in the Poisson model at the current sequencing depths. As this pattern is also observable at low counts for many RNA-Seq datasets, (Anders and Huber, 2010) we expect that topology libraries with better sequencing depths will likely also demonstrate greater overdispersion and benefit from application of the Negative Binomial regression model. Newer high-throughput techniques (Fang, et al., 2016; Mumbach, et al., 2016) that yield higher loop counts from chromatin conformation assays can also be seamlessly analyzed with *diffloop* and are expected to be over-dispersed in association settings.

When assessing the correlation between variability in epigenetic features (open chromatin, DNA methylation, histone acetylation, and cohesin localization) and differential loops (**Figure 4**), we determined that strong changes in these one-dimensional epigenetic data at either anchor were highly correlated with the presence of a differential loop. It is possible that significant alterations in DNA methylation or open chromatin may affect the ability of loop-mediating proteins to bind to the genome, and that resulting loss of a DNA loop prevents the transcriptional machinery from activating target genes. We therefore expect that the topological structure of the genome is a vital link in understanding the effect of epigenetic variability on gene expression. Similarly, much of the effect of distal genetic variation on transcription as uncovered through expression quantitative trait loci analyses may be mediated through the 3D architecture of the genome.

One key finding was that variation in distal DNA methylation and distal enhancer markings (indicated by H3K27ac) are more highly correlated with differential expression than the epigenetic signature at proximal promoter regions. Our finding corroborates previous findings that enhancer DNA methylation more highly correlated with gene expression (Aran, et al., 2013) and provides a framework to use topological data to conveniently link distal enhancer regions to their specific transcription start sites. While the distal regulatory elements accounted for a significant proportion of the variability in gene expression, the complexities associated with multiple enhancer-promoter and enhancer-enhancer loops per transcription start site convolutes the direct effect of epigenetic variation on gene expression. We anticipate that subsequent iterations of *diffloop* will include functionality to synthesize the full spectrum of connections when linking epigenetic data to transcriptional variation as mediated through the 3D genome.

An estimated nearly 40% of enhancer regions affect genes that are not the closest (in one-dimension) (Li, et al., 2012) implying analysis of the topology of the genome is vital to understanding what epigenetic modifications affect transcriptional processes. Moreover, as significant variability in transcription leads to distinct cellular states, characterizing the variable topology of the genome is vital for mechanistically bridging epigenetic changes like variable open chromatin and DNA methylation to their transcriptional effects. While 3C assays have provided a means to localize distal regulatory regions in the topological landscape to specific genes, statistically rigorous and computationally flexible tools are needed to fully characterize the 3D genome. Our analyses of the topological variability between cancers from the ENCODE project suggest that *diffloop* may be a useful resource for integrating genome topology data into existing workflows.

**Availability and Accession Numbers**

*diffloop* is available from R/Bioconductor: https://bioconductor.org/packages/release/bioc/html/diffloop.html. This package requires R > = 3.3.0 and depends on several R/Bioconductor packages including GenomicRanges, rtracklayer, and ggplot2. Version 1.3.2 of *diffloop* was used for all analyses described in this manuscript. All data used in this study are made freely available through GEO at the accession numbers specified in the Materials and Methods section. Code to reproduce analyses, figures, and tables are available through Github: **https://github.com/aryeelab/diffloop\_paper/**

**Supplementary Data**

**Supplementary PDF.** Supplementary information regarding additional diagnostics and analyses described in the differential loop calling implementation are available online

Acknowledgements

We thank D.Day, D. Hnisz, and members of the R. Young Lab for their useful discussion and insight.

Funding

This work was supported by the National Science Foundation [Graduate Research Fellowship #DGE1144152 to C.A.L.]. M.J.A. is supported by a Massachusetts General Hospital Department of Pathology startup fund. Funding for open access charge: Massachusetts General Hospital.

*Conflict of Interest:* none declared.

References

Anders, S. and Huber, W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11(10):R106.

Aran, D., Sabato, S. and Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol* 2013;14(3):R21.

Bell, J.T.*, et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 2011;12(1):R10.

Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489(7414):57-74.

Curran, J.E., Weinstein, S.R. and Griffiths, L.R. Polymorphic variants of NFKB1 and its inhibitory protein NFKBIA, and their involvement in sporadic breast cancer. *Cancer Lett* 2002;188(1-2):103-107.

Dekker, J.*, et al.* Capturing chromosome conformation. *Science* 2002;295(5558):1306-1311.

Delgado, M.D. and Leon, J. Myc roles in hematopoiesis and leukemia. *Genes Cancer* 2010;1(6):605-616.

Durand, N.C.*, et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* 2016;3(1):99-101.

Fang, R.*, et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* 2016;26(12):1345-1348.

Flavahan, W.A.*, et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 2016;529(7584):110-114.

Hnisz, D.*, et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 2016;351(6280):1454-1458.

Huber, W.*, et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 2015;12(2):115-121.

Ji, X.*, et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* 2016;18(2):262-275.

Law, C.W.*, et al.* voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15(2):R29.

Lawrence, M.*, et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;9(8):e1003118.

Li, G.*, et al.* ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res* 2016.

Li, G.*, et al.* ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol* 2010;11(2):R22.

Li, G.*, et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 2012;148(1-2):84-98.

Love, M.I., Huber, W. and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550.

Lun, A.T. and Smyth, G.K. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics* 2015;16:258.

Matharu, N. and Ahituv, N. Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genet* 2015;11(12):e1005640.

Moon, H.W.*, et al.* MTHFR 677CC/1298CC genotypes are highly associated with chronic myelogenous leukemia: a case-control study in Korea. *Leuk Res* 2007;31(9):1213-1217.

Mumbach, M.R.*, et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 2016;13(11):919-922.

Paulsen, J.*, et al.* A statistical model of ChIA-PET data for accurate detection of chromatin 3D interactions. *Nucleic Acids Res* 2014;42(18):e143.

Phanstiel, D.H.*, et al.* Mango: a bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* 2015;31(19):3092-3098.

Rao, S.S.*, et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;159(7):1665-1680.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139-140.

Rossen, R.D.*, et al.* Human peripheral blood monocytes release a 30,000 dalton factor (30 KD MF) that stimulates immunoglobulin production by activated B cells. *J Immunol* 1985;135(5):3289-3297.

Sengupta, S., Sharma, C.G. and Jordan, V.C. Estrogen regulation of X-box binding protein-1 and its role in estrogen induced growth of breast and endometrial cancer cells. *Horm Mol Biol Clin Investig* 2010;2(2):235-243.

Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545-15550.

Sun, J., Nawaz, Z. and Slingerland, J.M. Long-range activation of GREB1 by estrogen receptor via three distal consensus estrogen-responsive elements in breast cancer cells. *Mol Endocrinol* 2007;21(11):2651-2662.

Wang, S.*, et al.* An enhancer element harboring variants associated with systemic lupus erythematosus engages the TNFAIP3 promoter to influence A20 expression. *PLoS Genet* 2013;9(9):e1003750.

Wu, H.J. and Michor, F. A computational strategy to adjust for copy number in tumor Hi-C data. *Bioinformatics* 2016.