Impact of 3-dimensional genome organization, guided by cohesin and CTCF looping, on sex-biased chromatin interactions and gene expression in mouse liver

CURRENT STATUS: UNDER REVISION

Bryan J Matthews
Boston University

David J Waxman
Boston University

djw@bu.edu

ORCiD: https://orcid.org/0000-0001-7982-9206

DOI: 10.21203/rs.2.23342/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
cohesin, CTCF, intra-TAD loops, 4C-seq, liver sex differences, Nipbl, DNase hypersensitive sites
Abstract
Background: Sex differences in the transcriptome and epigenome are widespread in mouse liver and are associated with sex-bias in liver disease. Several thousand sex-differential distal enhancers have been identified; however, their links to sex-biased genes and the impact of any sex-differences in nuclear organization, DNA looping, and chromatin interactions are unknown.

Results: To address these issues, we first characterized 1,847 mouse liver genomic regions showing significant sex differential occupancy by cohesin and CTCF, two key 3D nuclear organizing factors. These sex-differential binding sites were largely distal to sex-biased genes, but rarely generated sex-differential TAD (topologically associating domain) or intra-TAD loop anchors. A substantial subset of the sex-biased cohesin-non-CTCF binding sites, but not the sex-biased cohesin-and-CTCF binding sites, overlapped sex-biased enhancers. Cohesin depletion reduced the expression of male-biased genes with distal, but not proximal, sex-biased enhancers by >10-fold, implicating cohesin in long-range enhancer interactions regulating sex-biased genes. Using circularized chromosome conformation capture-based sequencing (4C-seq), we showed that sex differences in distal sex-biased enhancer-promoter interactions are common. Sex-differential chromatin interactions involving sex-biased gene promoters, enhancers, and IncRNAs were associated with sex-biased binding of cohesin and/or CTCF. Furthermore, intra-TAD loops with sex-independent cohesin-and-CTCF anchors conferred sex specificity to chromatin interactions indirectly, by insulating sex-biased enhancer-promoter contacts and by bringing sex-biased genes into closer proximity to sex-biased enhancers.

Conclusions: These findings elucidate how 3-dimensional genome organization contributes to sex differences in gene expression in a non-reproductive tissue through both direct and indirect effects of cohesin and CTCF looping on distal enhancer interactions with sex-differentially expressed genes.

Full Text
Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the manuscript can be downloaded and accessed as a PDF.

Supplemental Information Note
Supplemental Figures are in the manuscript PDF file

Figures
Sex differences in cohesin and CTCF binding to mouse liver chromatin A. Distribution of male/female ratios for all diffReps-identified sex-differential sites that overlap a MACS2 peak
for binding of cohesin (left) and CTCF (right). The y-axis shows the number of binding sites per bin, and the x-axis shows the sex difference in binding, expressed as log2(Male/Female) fold change. Gray bars represent binding sites below the 10 read minimum count threshold, which were filtered out, and black bars represent sites that were statistically significant, but showed a |fold change| < 2 (values between -1 and 1 on the graph). Pink and blue bars respectively represent female-biased and male-biased sites above these thresholds. B. Venn diagram indicating 137 sex-differential peaks are common between cohesin and CTCF. Overlap is based on all sex-biased peaks, including male-biased and female-biased peaks on sex chromosomes (autosomal sex-biased peak numbers are shown in parenthesis, at the right). This pattern of limited overlap was also seen when the full set of unfiltered diffReps regions was examined (Fig. S1A). In total, 1,847 unique peaks exhibited significant sex bias in liver chromatin binding of CTCF and/or cohesin. C. Heat maps and aggregate plots for four sets of sex-biased cohesin (‘Coh’) or CTCF peaks. The peak set showing significant sex bias is highlighted in red at the top of each subpanel. For the heat map, read-in-peak normalized ChIP signals are shown for male and female cohesin binding (in blue) followed by male and female CTCF binding (in purple) within a 5 kb window centered around the differential peak summit. The aggregate profiles (top) represent the average signal of the heat map below for the same 5 kb window. Within each heat map, peaks are ranked based on the magnitude of sex bias from most sex-biased (top) to least sex-biased (bottom). Bimodal peaks, seen in the aggregate plots of several peak sets, is not a reflection of cohesin positioning relative to CTCF directionality, where the displacement is only ~20 nt, as compared to the >100 nt shift observed here. Rather, it may be due to differences in co-binding of transcription factors in male or female liver displacing cohesin and/or CTCF within 1 nucleosome length.
Figure 2

Cohesin and CTCF ChIP-seq binding strength and proximity to genes A. Box plots of normalized ChIP-seq signal for the peak sets indicated on the x-axis. Peaks with sex differential binding for cohesin (top graph) and CTCF (bottom graph) are shown. Each pair of
boxplots represents the male and female ChIP-seq signal for the same set of peaks, defined by their sex bias and peak type (CAC or CNC, for DCohesin peaks; and CAC or Lone CTCF, for DCTCF peaks), as indicated below the x-axis. Peak scores were calculated by average intra-peak ChIP signal, normalized by total sequence reads per million in peak (RIPM; see Methods). Female-biased peaks were, on average, stronger than male-biased peaks by M-W test: p ≤ 0.001 830 for female vs male CAC(DCoh), CAC(DCTCF), and for CNC, but not for Lone CTCF peaks. B. Distance from each indicated set of cohesin and CTCF peaks to the nearest enhancer DHS. Cumulative frequency curves indicate the fraction of each group on the y-axis, within the distance in kb to the nearest enhancer DHS indicated on the x-axis.

Enhancer DHS were defined based on their high ratio of the enhancer histone mark H3K4me1 over the promoter mark H3K4me3 at DHS [32]. Sex biased CNC peaks are closer to enhancer DHS (median distance to eDHS of 0.22 kb for male-biased CNCs and 0.12 kb for female-biased CNCs; K-S pval < 0.0001 for all comparisons) than the other CTCF and cohesin peak classes (M CAC(DCTCF): 14.98 kb; F CAC(DCTCF) 13.76 kb; M Lone DCTCF: 13.88 kb; F Lone DCTCF: 7.17 kb). Female-biased CNC peaks are significantly closer to enhancer DHS than are male-biased CNC peaks (p=0.0351; K-S t-test). Male-biased CAC(DCohesin) peaks were closer to enhancers than female-biased CAC(DCohesin) peaks (p=0.002; K-S t-test), however, the reverse was found for CAC(DCTCF) peaks (p=0.0052; K-S t-test). Distance to nearest enhancer was not significantly different between male-biased and female-biased Lone CTCF peaks (p=0.1068; K-S t-test). P values for comparisons between male-biased and female-biased peaks of the same class are shown for each plot (K-S t-test). C. Distance from each indicated set of cohesin and CTCF peaks to the nearest TSS. Cumulative frequency curves indicate the fraction of each group on the y-axis within the distance in kb to the nearest TSS indicated on the x-axis. TSS for protein coding (RefSeq) and liver IncRNA genes were considered [70]. Female-biased cohesin and CTCF peaks are closer to TSS than male-biased CTCF and cohesin peaks of the same class (significance by K-S t-test is indicated at top left of each plot). Distance to the TSS was not significantly
different for male-biased versus female-biased CNC peaks (p=0.1458; K-S t-test). D. Proximity of sex-biased cohesin and CTCF binding sites to sex-biased genes. Peak designations were as follows: Proximal, peaks < 20 kb from a sex biased gene TSS; Intra-TAD, peaks within the same intra-TAD loop as a sex-biased gene; or TAD, peaks in the same TAD as a sex-biased gene. Each of these groups is mutually exclusive. TAD loop [35] and intra-TAD loop [32] coordinates were from the indicated references. A set of 983 sex-biased biased protein-coding genes was used in this analysis (see Table S1 of [9]). E. Cumulative frequency curves show the fraction of each group (y-axis) within the distance in kb to the nearest sex-biased DHS or H3K27ac genomic region (x-axis), based on a merged list of published sex-biased DHS [15] and sex-biased H3K27ac ChIP-seq peaks [96] for male and female mouse liver. For this analysis, CAC peaks with sex-biased binding of CTCF and cohesin were combined and presented as a single group (CAC (Both)). Male-biased and female-biased CNC peaks are significantly closer to sex-biased DHS/H3K27ac than the four other peak classes (p < 0.001; K-S t-test). Female biased CNC peaks were significantly closer to sex-biased DHS/H3K27ac than male-biased CNC peaks (p=0.0094; K-S).
Proximal and distal sex-biased regulatory elements and impact of cohesin depletion. Shown are WashU Epigenome Browser screenshots of sex-biased cohesin and CTCF binding events proximal (≤ 20 kb from a sex-biased gene TSS) and distal (> 20 kb) to sex-biased genes in mouse liver. Tracks shown are (from top to bottom): TAD and TAD boundary location [indicated by a vertical shift in the TAD track, as on the left and on the right sides of panel].

Figure 3
C], H3K27ac ChIP-seq, DNase-seq, cohesin (Rad21) ChIP-seq, CTCF ChIP-seq, and Ref-seq genes. ChIP-seq and DNase-seq data in each track is shown superimposed for male (blue) and female (pink) liver after normalization to total reads per million in the union of peaks for that factor (see Methods). Below each ChIP-seq track, a horizontal bar identifies genomic regions that show significant male-bias (blue bar) or female-bias (pink bar), with darker and lighter shades indicating strict and lenient cutoffs for sex-bias, respectively (see Methods).

Green arrows indicate CTCF sex-differential CAC, and red arrows indicate CTCF sex differential CNC and Lone peaks. Male/Female stranded polyA+ RNA-seq gene expression ratios [11] are indicated above each panel. A. Slc22a29 is a female-biased gene with a distal female-biased non-anchor CAC peak overlapping a robust female-biased DHS with female-biased H3K27ac histone mark accumulation ~34 kb upstream (green arrow). The region shown spans chr19:8290981-8333632. B. Two male biased genes, Cml5 and Nat8, with proximal male-biased cohesin and CTCF peaks overlapping male-biased DHS (red, green arrows). Cml5 has a ~3 kb upstream male-biased CNC peak overlapping a strongly male-biased DHS. The Cml5 promoter shows strong male-biased H3K27ac marks and a weaker male-biased DHS. Nat8 has a weakly male-biased DHS at its promoter and a stronger male-biased DHS ~12 kb upstream that overlaps a male-biased CAC peak. The region shown spans chr6:85766132-85794443. C. Male-biased genes C8a and C8b have distal male-biased CAC and CNC peaks overlapping male biased DHS and H3K27ac marks (green and red arrows, respectively). The linear distance between the upstream male-biased CAC peaks and downstream male-biased genes is >1.5 Mb. C8a and C8b reside on opposite ends of the same TAD. Oma1 is close in linear distance to these sex-biased regulatory elements, but shows no sex differences in expression; based on TAD structure it is not be predicted to interact with the highlighted male-biased enhancers. See Fig. S4A for the full length of the TAD and a model of spatial positions. The left portion of this figure spans chr4:103027454-103167067, and the right portion spans chr4:104433514-104583344. D. Loss of cohesin binding decreases expression of C8a and C8b significantly (p ≤ 0.0001 and p=0.0020; M-W
t-test), while expression of Oma1 increases (p=0.0102; M-W t-test). Bars represent the mean expression of the group for a given gene relative to the mean of the WT group (equal to 1), and error bars show the standard deviation based on n=4 per group. E. Loss of cohesin binding has a 10-fold greater suppressive effect on male-biased genes with distal sex-biased enhancers than those with proximal sex-biased enhancers. Shown is the mean expression for cohesin-depleted versus wild-type liver, such that a value of 0.1 represents a 10-fold reduction in expression after cohesin loss. The median relative expression for DHS/H3K27ac-proximal genes is 0.69 (representing a modest suppressive effect of cohesin loss) and the corresponding median for DHS/H3K27ac-distal genes is 0.07, indicating a >10-fold greater reduction in gene expression (p 0.0087; M-W). Similar results were obtained when the definition of proximally-regulated genes was relaxed to include genes with a TSS < 20 kb from either a male biased DHS or a male biased DHS (median of 0.45 versus 0.042 for distal genes). Also see Table S2.
Figure 4

4C-Seq analysis of two female-biased and two male-biased genes A, B, C, E. Showed are WashU Epigenome Browser screenshots for four genomic regions investigated by 4C-seq.
The upper track presents 4C-seq data for four viewpoints, marked by a vertical highlight in each panel. The 4C-seq track is based on merged data from three biological replicates for each sex, calculated from the median value for a sliding window of 11 restriction fragments, in reads per million normalized 4C-seq signal per sex. These values are overlaid for visualization using the Matplotlib functionality built into the WashU genome browser with default parameters. The next four tracks show normalized DNase-seq or ChIP-seq signal for the indicated factors, and correspond to those described in Fig. 3. Sex-biased lncRNAs are shown below the Refseq gene track in pink (female-specific lncRNAs) or blue (male-specific lncRNAs). C and E also show locations of intra-TAD loops (pink arcs) below the Refseq gene track. 4C-seq data for individual biological replicates is presented in Fig. S5.

A. Distal enhancer viewpoint near A1bg and female-biased lncRNAs. The region shown includes 12 female-biased and nuclear-enriched mono-exonic lncRNAs, which fall into three clusters. The lncRNAs in each cluster are all transcribed from the same strand, as indicated by the arrow marking the TSS and direction of transcription of the most upstream lncRNA in each cluster. LncRNAs 12590-12593 show the strongest 4C-seq interactions with the viewpoint and also the most consistent female bias (Fig. S5E, Fig. S5F). The region shown spans chr15:60733512-60954051.

B. Enhancer viewpoint 8 kb upstream of Gm4794 interacts primarily with the proximal promoter (left red arrow) and a strong female-biased enhancer (right red arrow). Viewpoint enhancer and the interacting enhancer both contain female-biased CNC peaks (marked below cohesin track in dark pink). Sult3a1 shows a weak and broad pattern of interaction with this viewpoint (red bracket). The region shown spans chr10:33418446-33680888.

C. Viewpoint at the promoter of C9 interacts with a distal male-biased enhancer, bypassing the gene Dab2. This viewpoint shows a strong male-biased interaction with distal male-biased enhancer indicated by a red arrow. The nested intra-TAD loop structure (loops at bottom) may facilitate the high and male-biased expression of C9 by looping out the intervening gene Dab2. There is a weaker, but apparent male-biased interaction with the TSS of the short isoform of Dab2. Based on CAGE data and Refseq
annotation, the shorter isoform of Dab2 is predominantly expressed in liver (data not shown). The region shown spans chr15:6147917-6461799. D. Cohesin depletion significantly reduces expression of C9 and its antisense 941 lncRNA (Inc12340): 66% reduction, p=0.0056; and 82% reduction, p=0.0049, respectively. Bars represent mean expression compared to wild-type liver (WT, set equal to 1), with error bars showing standard deviation for n=4 per group. E. Nudt7 lacks male-biased DHS or H3K27ac at its promoter, but interacts with a distal male-biased enhancer within the same intra-TAD loop. This viewpoint is anchored at a male-biased DHS (red highlighting). Prominent interactions with this viewpoint (red arrows) are with a neighboring male biased enhancer and the promoter of Nudt7. While the TSS of Inc7430 and Nudt7 are close, the viewpoint enhancer interacts specifically with the TSS of Nudt7. The region shown spans chr8:116592444-116707613. F. Upon loss of chromatin-bound cohesin, expression of both Nudt7 and its bi-directionally transcribed lncRNA (Inc8430) are significantly reduced (69% decrease, p=0.0116; and 93% decrease, p=0.0113, respectively). Expression of Inc7423 was not significantly impacted.

Data presentation is as described in D.
Nox4 results with 4C-seq for two viewpoints and confirming nested intra-TAD loop structure

A. Two viewpoints for Nox4 gene region: VP1 is anchored ~11 kb upstream of the Nox4 TSS (light red highlight), and VP2 is anchored ~125 kb downstream of the TSS (green highlight). VP1 has the strongest male-biased H3K27ac mark. This region shown spans chr7:94248242-94726358. In addition to the ChIP-seq (H3K27ac, Cohesin, and CTCF) and DNase-seq tracks shown, the tracks at the top show the chromatin state of this genomic region in male liver
and female liver (bottom). Chromatin states are colored: green indicates an enhancer-like state, blue indicates a promoter-like state, and purple a transcribed-like state. Red indicates an inactive chromatin state (see Fig. S7C for further details). Regions Y1 and Y2 are in chromatin state E13 in male liver but in inactive state E2 in female liver. Y1 and Y2 both include a short enhancer state region in male liver (E11 within region Y1; E10 in region Y2). The absence of focal interactions between VP1 and VP2 supports the model of two nested and insulated intra-TAD loops shown at the bottom. All tracks were normalized and are presented as described in Fig. 4. B. Expression of Nox4, but not the neighboring gene Tyr, is cohesin-dependent. Although Nox4 is primarily regulated by proximal enhancers within the shorter intra-TAD loop, its full expression is nevertheless dependent on cohesin. This may be due to the need for the intra-TAD loop structure; however, loss of this insulation did not increase expression of Tyr. Expression of Nox4 was reduced by 62% (p<0.0001).

Data presentation is as described in Fig. 4D.
Model of the direct and indirect contribution of cohesin and CTCF 976 to sex-biased liver gene expression. A1bg exemplifies direct regulation, where female-biased CTCF binding can explain the observed female-bias in looping interactions. C9 is an example of intra-TAD loops present in male and female liver that contribute indirectly to sex-biased gene expression. Not only does the intra-TAD loop insulate the lowly-expressed sex-independent gene Dab2; it also brings the distal male-biased enhancer into close proximity. Likely additional factors, such as Mediator, YY1, or eRNAs, can contribute directly to the interactions observed in male liver.