CTCF and Cohesin link sex-biased distal regulatory elements to sex-biased gene expression in mouse liver

Bryan J. Matthews and David J. Waxman

Department of Biology and Bioinformatics Program, Boston University, Boston, MA 02215

Running title: CTCF, cohesin and sex-biased liver gene expression

FUNDING - This work was supported by National Institutes of Health grants [DK33765, ES024421 to D.J.W]; and by the National Science Foundation predoctoral fellowship [DGE-1247312 to B.J.M].

Corresponding author:
David J. Waxman
Dept. of Biology, Boston University
5 Cummington Mall
Boston, MA 02215
Email: djw@bu.edu

ABSTRACT

More than 1,000 genes are differentially expressed between male and female mouse liver. However, ~90% of 4,000 sex-differential open chromatin regions harboring regulatory elements implicated in liver sex bias are distal enhancers. To elucidate the role of 3D nuclear organization and address the challenge of linking distal regulatory regions to sex-biased gene targets, we performed ChIP-seq for CTCF and cohesin, known mediators of 3D nuclear organization, and identified ~1,000 binding sites showing significant differential occupancy by these factors in male and female mouse liver. Distal genomic regions showing sex-differential CTCF and/or cohesin binding were evaluated as potential factors in sex-dependent looping controlling sex-biased proximal effects on gene expression. Results were integrated with a study of cohesin depletion in mouse liver to evaluate the impact of cohesin loss on the expression of putative gene targets. Further, circularized chromosome conformation capture with sequencing (4C-seq) was used to discover striking sex-differences in 3D looping, linking distal sex-biased enhancers to proximal sex-biased gene expression for both male-biased (*C9*, *Nudt7*, *Nox4*) and female-biased genes (*A1bg, Sult3a2*). Our findings also illustrate the importance of a nested intra-TAD loop for insulation of sex-specific enhancer–promoter contacts for *Nox4*. These studies reveal how chromatin interactions and 3-dimensional genome organization guided both directly and indirectly by CTCF and cohesin looping contribute to liver sex bias.

INTRODUCTION

Human sex differences are observed in several non-reproductive tissues, notably brain [1], immune system [2], kidney [3], and liver [4]. The health impacts of sex differences in liver include a higher incidence of aggressive liver cancer in males [5], increased susceptibility to autoimmune hepatitis in females [6], and sex differences in metabolism of some pharmaceuticals and environmental chemicals [7]. Transcriptomic and epigenetic sex-bias is best characterized in mouse liver, where ~1,000 genes [8], including many lncRNA genes [9, 10] and miRNAs [11], exhibit sex-biased expression that may be regulated by thousands of sex-specific DNase I hypersensitive sites (DHS) [12, 13]. This prior work has shown that the majority of sex-biased genes do not have gene-proximal sex-biased chromatin marks or DHS, suggesting most epigenetic sex bias occurs at distal elements such as enhancers [12, 13]. Although not directly tested to date, three-dimensional looping is one mechanism that could link the ~4,000 mostly distal sex-biased enhancers to the ~1,000 sex-biased genes.

Two major transcription factor (TF) regulators of 3D genomic architecture are CCCTC-binding factor (CTCF) and the multi-protein complex cohesin. CTCF has primarily been studied for its role in DNA looping and insulation [14, 15]. Cohesin, named for its role in sister chromatid cohesion [16], is the molecular motor powering DNA looping via a "loop extrusion" mechanism [17, 18]. Loss of CTCF or cohesin is embryonic lethal in mice [19, 20], however, in adult mouse liver the induced degradation of the cohesin loading factor Nipbl results in a dose-dependent loss of both cohesin binding and virtually all focal DNA looping [21]. In other systems, a similar loss of looping has been observed as a result of direct depletion of either cohesin [22] or CTCF [23]. Thus, both CTCF and cohesin are required for DNA looping.

Cohesin is loaded and unloaded from chromatin by specific protein complexes but lacks sequence-specific DNA binding [24], thus its role is largely dependent on its binding partners. When associated with mediator and tissue-specific transcriptional regulators, cohesin mediates shorter-range looping between enhancers and promoters—often referred to as enhancer-promoter loops or gene loops [25-27]. Regions bound by cohesin but not CTCF (cohesin-non-CTCF sites; CNC) tend to be highly tissue specific but are frequently weaker than sites where cohesin is bound together with CTCF [25, 28, 29]. When cohesin and CTCF are both bound, at cohesin-and-CTCF (CAC) sites, cohesin forms insulating loops, which have been variously characterized as insulated neighborhoods [26], loop domains [30], Topologically Associating Domains (TADs) [31], and intra-TAD loops [32]. The majority of insulating CAC-mediated insulated loops are conserved across multiple tissues or cell types [30-33], and are often even conserved across mammalian species [34]. A distinct subset of CAC sites are important for enhancer-promoter interactions [35, 36], and may specifically help structure super-enhancers around key constituents enhancers, known as hubs [37]. The role of CTCF binding in the absence of cohesin (lone CTCF sites) is less clear. Unlike cohesin, CTCF binds specific DNA motifs via its 11 zinc fingers [38], yet the function of lone CTCF sites is likely also dependent on additional interacting proteins, such as the TFs YY1 [39], BORIS [40], and STAT5 [41].

CAC-mediated insulating loops are typically anchored by a pair of highly-bound CTCF sites oriented by the non-palindromic CTCF motif [30]. Supporting this finding, inversion or deletion of a CAC-bound anchor results in a loss of looping [17, 42]. This stereotypical pattern of convergently-oriented CTCF binding enables the accurate prediction of CAC-mediated loops from CTCF/cohesin binding activity and motif orientation alone [17, 32, 43]. Computational prediction of loops linking enhancers and promoters, i.e., gene loops, is a more elusive goal, although some recent progress has been made [44, 45]. To directly measure looping at this scale, experimental methods based on chromatin conformation capture (3C) technology use formaldehyde crosslinking and restriction enzyme digestion, followed by proximity ligation to probe the interaction frequency between regions in the genome [46]. Circularized chromatin conformation capture with sequencing (4C-seq) is a variant of 3C that assesses all interactions between a single region of interest and the rest of the genome ("one-versus-all") [47].

Here we took a multi-pronged approach to characterize the role of architectural proteins and 3D genome organization in the maintenance of mouse liver sex differences. First, we performed ChIP-seq for CTCF and cohesin using chromatin from male and female mouse livers and show that the majority of sex-biased binding of these factors occurs at distal intergenic sites. Only a minority of these binding sites are *cis* to sex-biased genes, which may *cis*-proximal (within 20 kb) or more *cis*-distal (> 20 kb, but within the same TAD) to a sex biased gene. Specific examples of *cis* proximal and distal CTCF/cohesin binding are shown for male-biased genes (proximal: *Cml5*, *Nat8*; distal, *C8a*, *C8b*) and female-biased genes (proximal: *Fam84b*; distal: *Slc22a29*). Using a publicly available cohesin depletion dataset for male liver [21], we find that distally-regulated male-biased genes are considerably more sensitive to cohesin loss than those with proximal sex bias. To directly measure differences in chromatin interactions, we use 4C-seq for six viewpoints at or nearby five genes showing liver sex bias in expression, ranging from 2-fold to 700-fold (female-biased: *A1bg* and *Sult3a1*; male-biased: *C9*, *Nudt7*, and *Nox4*). Finally, we show the importance of a nested intra-TAD loop for insulation of enhancer-promoter contacts for *Nox4*. Overall, we describe how 3-dimensional genome organization contributes to liver sex bias in both direct and indirect ways.

RESULTS

Global differences in CTCF/Cohesin binding between male and female mouse liver - To understand the differences between male and female 3D genome organization, we determined the distribution of CTCF and cohesin binding genome wide by ChIP-seq. Using chromatin prepared from mouse liver, we performed ChIP-seq using polyclonal antibodies to CTCF (n=4 livers per sex) and the cleavable cohesin complex protein Rad21 (n=3 livers per sex). We identified 975 CTCF peaks (750 on autosomal chromosomes; Table S1A) and 1,011 cohesin peaks (841 autosomal; Table S1B) showing >2-fold sex-difference in factor binding (Figure 1A). Peaks with sex differential binding for both CTCF and cohesin are represent only 13.4% of differential cohesin peaks and 16.3% of differential CTCF peaks (Figure 1B). A similar low percentage overlap of sex-differential CTCF binding with sex-differential cohesin binding was found using several alternative filters or approaches to define sex-differential binding (Figure S1A). Aggregate plots and heatmaps show that, on average, peaks with sex-biased CTCF binding show the same trend of sex bias in cohesin binding (and vice versa), although the magnitude is generally less than 2-fold (Figure 1C).

Location of sex-biased binding sites relative to genes and regulatory elements – Given the strong dependence of cohesin and CTCF function on the presence of their binding partners, we subdivided the 1,845 sex differential CTCF and/or cohesin sites into 4 groups: CAC sites with sex-differential binding of cohesin (ΔCoh) or CTCF (ΔCTCF), CNC sites with sex-differential cohesin binding, and Lone CTCF with sex-differential CTCF binding (see Methods). Sex-differential CAC sites comprise 45-66% of all sex-differential sites (Figure S1B). The sex-differential CAC sites are stronger peaks than CNC or lone CTCF sites, with the female-biased sites generally found at stronger peaks than the male-biased sites (Figure 2A, Figure S2A). This difference is reflected in the percent of peaks with a CTCF motif and the motif quality (Figures S2B,C). A higher fraction of male-biased Lone CTCF sites contain a CTCF motif than similar female-biased sites (Figure S2B), but there is no significant difference in normalized ChIP signal or motif score (Figure 2A, Figure S2C). As CTCF binding is inhibited by DNA methylation [48, 49], the increased fraction of male-biased "Lone CTCF" peaks could be due to the reported hypomethylation of DNA in male compared to female [50], which could open up additional binding sites (see Discussion).

Female-biased CTCF/cohesin binding sites are significantly closer to genes and enhancer DHS (eDHS) than male-biased sites (Figure 2B,C). As expected, sex-biased CNCs are more closely associated with eDHS than sex-biased CAC sites. Specifically, the majority of sex-biased CNCs are within ~200 bp of an enhancer (median distance: 202 bp in males, 119 bp in females), compared to a median distance of 8.8 kb for male-biased CAC and 17.5 kb for female-biased CAC sites (Figure 2C, *left*). This supports the proposal that, although sex-biased CNCs show weaker binding than CACs, they are often present at active enhancers and likely functional.

We identified 137 CAC peaks with sex differences in both CTCF and cohesin binding, 84 of which are on sex chromosomes (Table S1C). Of these, only 17 overlap TAD or intra-TAD loop anchors in liver [32], and again a majority of these are on sex chromosomes (9 of 17; Table S1C). Consistent with this, ~90% of intra-TAD loops and anchors predicted in male liver are also predicted in female liver (Figure S2D). Thus, CAC-mediated intra-TAD loops in mouse liver are largely conserved between the sexes, and therefore do not directly regulate sex-biased gene expression. Rather, TAD and intra-TAD scale insulating loops are likely to indirectly guide enhancer-promoter looping in a sex-biased manner, as is described below.

We compared the binding of sex differential sites to CTCF binding and gene expression across 15 mouse tissues using data for male liver from the ENCODE consortium [51, 52]. Strikingly, female-biased CAC sites (ΔCoh or ΔCTCF) showed greater overlap with CTCF binding sites from other mouse tissues (i.e. are less liver specific; Figure S3A), even though the ENCODE data used came from male liver. By comparison, female-biased Lone CTCF sites are more liver-specific than male-biased sites, while there was no difference between the highly liver-specific male- and female-biased CNC sites (Figure S2C). The higher tissue conservation for female-biased CAC peaks is most likely a direct reflection of their binding strength, as stronger peaks show greater conservation for both sex-differential and all CTCF peaks (Figure S3B). When comparing the tissue specificity in expression of the nearest liver-expressed gene (FPKM > 1; TSS within 20 kb of peak), female-biased CNCs are nearby liver-specific genes relative to all expressed genes or female-biased CACs (p=0.0013 and p=0.0071, MW test; Figure S3C). While the closest genes to male-biased CNCs are more tissue specific than liver-expressed genes overall, they are not significantly more tissue specific than genes neighboring male-biased CACs (p=0.0247 and p=0.1050, MW test; Figure S3C). Thus, it appears that sex-biased CNC peaks also participate in tissue-specific transcriptional regulation, as was described for CNC peaks generally [25, 28, 29].

Sex-biased CTCF/cohesin sites tend to be distal from sex-biased genes and sex-biased DHS or H3K27ac-marked regions (Figures 2D,E). Fewer than 20% of sex-biased CTCF/cohesin binding sites are within 20 kb of a sex-biased gene, while at most 53% of male-biased CNCs could be considered potential *cis* regulators of sex-biased genes (within the same intra-TAD or TAD loop as a sex-biased gene; Figure 2D). Sex-biased CNCs are significantly closer to sex-biased *cis* regulatory elements than CACs or Lone CTCF sites, however the majority of all sites remain quite distal (Figure 2E).

Proximal and distal sex-biased CTCF/Cohesin peaks and impact of cohesin depletion - Examples of sex-biased genes with proximal sex-biased CTCF/cohesin binding are shown in Figures 3A and 3B. These include the male-biased genes *Cml5* and *Nat8* (M/F expression ratios of 20.2 and 4.2, respectively; Figure 3A) as well as the female biased gene *Fam84b* (F/M expression ratio of 2.8; Figure 3B). Approximately 3 kb upstream of the TSS of *Cml5* a male biased CNC overlaps a male-biased DHS (red arrow in Figure 3A), while a male biased CAC overlapping a male-biased DHS is positioned ~12 kb upstream of the TSS of *Nat8*. Presumably, cohesin and/or CTCF play a role in looping these proximal male-biased enhancers to these neighboring genes to enhance the male-biased in gene expression. Even more proximal, the female-biased CNC shown in Figure 3B overlaps the TSS of *Fam84b* and female-biased H3K27ac accumulation at the promoter.

A significant fraction of sex-biased CTCF/cohesin occurs within the same TAD as a sex-biased gene (Figure 2D). Examples of sex-biased CTCF/cohesin binding sites that are distal yet within the same TAD as similarly sex-biased genes are shown in Figures 3C and 3D. Figure 3D shows the female-biased gene *Slc22a29* (F/M ratio, 8.7). 33 kb upstream of the *Slc22a29* TSS is a highly female-biased enhancer (F-biased DHS and H3K27ac mark accumulation) which also overlaps a female-biased CAC. *Slc22a29* is the closest gene and TSS to this female-biased enhancer. Figure 3C shows two male biased genes, the complement C8 genes *C8a* (M/F = 3.3) and *C8b* (M/F = 2.8), which are linearly distal (~1.5 Mb) from a cluster of strongly male-biased enhancers. Based on known TAD structure, we would predict they are more proximal spatially than the adjacent gene, *Oma1* (also see screenshot in Figure S4A).

A recent study of cohesin function in male mouse liver revealed that one consequence of loss of cohesin binding (achieved by depleting its loading factor, Nipbl) is the loss of distal enhancer-promoter contacts in favor of local ectopic contacts—activating proximal genes [21]. Figure 3E shows the relative expression change for wild type and cohesin-depleted male mouse liver for the 3 genes shown in Figure 3C: *Oma1*, *C8a*, and *C8b*. Expression of *C8a* and *C8b* is dramatically reduced (98.2% and 82.3% reduction, respectively) upon loss of chromatin-bound cohesin, while expression of *Oma1* increases modestly (+22%), perhaps by an enhancer hijacking mechanism. By comparison, there is no significant difference in expression between WT and cohesin-depleted mice for the more proximally-regulated male-biased genes Cml5 and Nat8 (Figures 3A, S4B).

Beyond specific examples of distally-regulated (*C8a*, *C8b*) or proximally-regulated (*Cml5*, *Nat8*) male-biased genes, we also sought to characterize the requirement of cohesin for proper expression of male-biased genes. As a general rule, male-biased genes with proximal sex-biased cis regulatory elements are less sensitive to loss of cohesin than male-biased genes dependent on distal sex-biased enhancers (Figure 3F). Conceivably, alternative transcription factors (e.g. Mediator[25] or YY1[53]) are sufficient to maintain enhancer-promoter loops over short genomic distances, while cohesin is required for the upkeep of distal interactions likely via a direct or indirect looping mechanism.

Using 4C-seq to directly assess sex-biased chromatin interactions in mouse liver - To more directly assess the role of enhancer promoter loops in epigenetic sex bias, we performed 4C-seq for 6 viewpoints at or near 5 sex-biased genes. The viewpoints in Figure 4 show merged results from 3 biological replicates (n=3 per sex), each library prepared from a separate CD-1 mouse liver (see Methods) and example libraries shown in Figure S5A,B.

Figure 4A shows a viewpoint anchored at a female-biased enhancer near the strongly female-biased gene A1bg (F/M expression ratio of 155) as well as 12 female-biased, mono-exonic lncRNAs [10]. In the female samples, robust interactions are observed between the viewpoint enhancer and three regions, indicated by red arrows. While the viewpoint region interacts with a stronger female-biased enhancer (*right* arrow) and the promoter of *A1bg* (*middle*), the strongest interaction is actually with a cluster of lncRNAs (lncRNAs 12590-93; *left*). Of the 12 lncRNAs in this screenshot, this interacting cluster represents more highly expressed (Figure S5C) and more consistently female-biased genes across various RNA-seq datasets (Figure S5D). Traces for 4C-seq biological replicates are shown in Figure S6A. The precise link between the female-biased expression of these lncRNAs and the female-bias in 3D interactions with a distal enhancer is not known, however, it does point to specific candidates for further characterization. It is possible that the 3D interactions are mediated by these nuclear-enriched lncRNAs (as described for Xist [54] and Firre [55]), or alternatively, their interaction is primarily that of a regulatory enhancer driving expression of several genes—including *A1bg* and multiple lncRNAs. The presence of female-biased CTCF binding at both the strong female-biased enhancer and the lncRNA cluster (*left* and *right* arrows) provides some mechanistic support for the latter interpretation. As CTCF is known to functionally interact with lncRNAs with high affinity [56], these two mechanisms are not necessarily mutually exclusive; *cis*-acting lncRNAs (expressed only in female liver) could help selectively guide CTCF binding and interactions unique to female liver.

Figure 4B shows 4C-seq results for an enhancer proximal to the highly female-biased gene *Gm4794* (F/M ratio = 704; also known as *Sult3a2*) and the mono-exonic female-biased lncRNAs 8820 and 8821 (F/M = 34.2 and 90.3). This enhancer is also distal to two other female-biased protein coding genes, *Sult3a1* (F/M = 288) and *Rsph4a* (F/M = 33.6). Female-biased interactions are observed with the proximal promoter of Gm4794 (*left* red arrow) and the strong female-biased enhancers overlapping lncRNA 8820. Unlike the enhancer neighboring A1bg, these interactions may be mediated by female-biased CNC peaks present at both the viewpoint enhancer and downstream interacting enhancer. Despite the absence of insulator elements (i.e. no intra-TAD loops present), all focal interactions are local within ~35 kb of the viewpoint. In addition to the enhancer-promoter interactions indicated by the red arrows in Figure 4B, the viewpoint enhancer also shows weak interactions from the female-biased promoter upstream *Sult3a1* to the end of the gene body (red bracket). Despite robust female-biased DHS and H3K27ac mark accumulation, this region lacks cohesin binding, which

may account for the lack of direct interactions with the viewpoint enhancer. It is possible that *Gm4794* and *Sult3a1* both interact with the enhancer that overlaps with lncRNA 8820, but the indirect nature of this weak association cannot be captured by proximity ligation under standard 4C-seq conditions. Traces for biological replicates are shown in Figure S6B.

In the absence of sex-biased CTCF or cohesin binding, intra-TAD loops may indirectly coordinate enhancer promoter contacts preferential to one sex. To test this hypothesis, we performed 4C-seq for the promoter of the component complement gene *C9* (M/F = 3.5; Figure 4C). This gene also overlaps an antisense lncRNA that is also male-biased (lncRNA 12340; M/F = 25.5). A strongly male-biased enhancer region lies ~230 kb upstream of the TSS of *C9*, and the former shows male-biased DHS and H3K27ac mark accumulation, while the latter only has male-biased DHS. Both of these regions fall just outside a nested intra-TAD loop (< 10 kb upstream for enhancers and downstream for *C9* promoter) insulating the lowly expressed gene Dab2 (FPKM 0.7 in M and 1.2 in F). In both male and female liver we observe that the promoter of *C9* is interacting with the cluster of distal enhancers (Figure 4C; red arrow), with a stronger interaction observed in males. In general, only very weak, mostly nonfocal interactions are observed between the promoter of *C9* and within the intra-TAD loops. This insulation allows a strong male biased enhancer to bypass a more proximal gene, *Dab2*, to drive expression of *C9*. Beyond male bias in expression, this is a clear mechanism allowing high expression of *C9* (61.3 FPKM in M) compared to *Dab2* (0.7 FPKM in M). However, the shorter isoform of *Dab2* does show weak, male-specific interaction despite a lack of any sex-bias in gene expression. The movement of cohesin is at least in part linked to transcriptional activity [18, 57], thus higher levels of transcription in male liver could lead to less cohesin pausing at loop anchors and elevated interactions crossing these loop boundaries in males. Traces for individual replicates are shown in Figure S6C.

Given the interaction with a strong distal enhancer element, we hypothesized that the expression of *C9* and lncRNA 12340 would be sensitive to loss cohesin binding. In fact, both genes show substantial reduction in expression in cohesin-depleted mouse liver (~3-5-fold; $p < 0.01$ for both), while the insulated gene *Dab2* increases in expression but not to a significant extent ($p=0.219$; Figure 4D).

Nudt7 is an example of a highly expressed, male-biased gene with distal male-biased enhancers within the same intra-TAD loop (M/F = 2.5). Although there are some weak male-biased DHS along the gene body of *Nudt7* (Figure 4E), there is no apparent sex bias at the shared promoter with lncRNA 7430 (M/F = 4.0; FPKM 2.9). ~22 and 34 kb upstream are two male-biased enhancers with male-biased DHS and H3K27ac mark accumulation, the latter also overlaps the TSS of the male-biased lncRNA 7423 (M/F 7.0; FPKM 2.75 in M). This cluster of enhancers also is one of 503 super-enhancers found in male and female liver and maintained despite epigenetic variation associated with circadian rhythms [32]. In male liver interactions between the viewpoint enhancer and both a neighboring male-biased enhancer as well as the promoter of *Nudt7* (red arrows; Figure 4E). By comparison, no focal interactions (in either sex) are observed between the viewpoint enhancer and a sex independent enhancer 15.7 kb upstream of Nudt7. Traces for individual replicates are shown in Figure S6D. While the interaction with the promoter region is primarily with the TSS of *Nudt7* and not lncRNA7430, both genes are downregulated upon cohesin depletion (Figure 4F). Expression of lncRNA 7423 is not significantly reduced, potentially due to the proximity of strong male-biased enhancers.

Sex-independent, nested intra-TAD loops restrict *Nox4* to proximal enhancer-promoter interactions - We next sought to characterize chromatin interactions for two male-biased enhancers neighboring the male-biased gene NADPH Oxidase 4 (*Nox4*; M/F expression ratio 7.7). In humans, *Nox4* is substantially upregulated across many cancer types, including hepatocellular carcinoma (Figure S7A). In mice that spontaneously-develop liver tumors, *Nox4* is substantially upregulated in tumors over adjacent control tissue (Figure S7B; [58]). Figure 5A shows both viewpoints, with the "VP Enh1" track showing interactions for an enhancer ~11kb upstream of the TSS of *Nox4* (*above* in red) and the "VP Enh2" track showing interactions for an enhancer ~125 kb downstream of the TSS of *Nox4* (*below* in green). Enh2 does not appear to interact with Enh1 or with the promoter of *Nox4* in either sex. In fact, Enh2 does not appear to have any clear male-biased interactions.

Comparatively, Enh1 shows interactions associated with "transcribed-like" chromatin states that are specific to male liver (yellow highlighting in Figure 5A). The upstream highlighted region does not correspond to any annotated gene, however the chromatin state in male, but not female, indicates a narrow enhancer (green) and transcribed-like chromatin state (purple; full chromatin state key in Figure S7C) [13]. Downstream, the red arrows indicate the position of the 3' anchor of the shorter intra-TAD loop. Immediately upstream of this anchor we can see interactions in both male and female liver indicative of insulation within the intra-TAD loop. Just downstream of the anchor however we see interactions only for males (yellow highlighting; *right*), which is likely due to the increased movement of cohesin beyond this loop anchor most probably as a result of increased transcription of Nox4 in male liver [18, 57]. Both highlighted regions are in a transcribed chromatin state only in male liver. Traces for individual replicates are shown in Figure S7D,E.

These data support the predicted model of two nested intra-TAD loops, with the smaller enclosed loop insulated from the larger enclosing loop. Domain predictions from other mouse tissues support the conclusion that Enh1 and Enh2 are in separate domains and thus only Enh1 would be predicted to interact with the *Nox4* promoter (Figure S7F; [26, 51]). Generally, the cis regulatory elements relevant for the regulation of *Nox4* appear to be contained within the smaller intra-TAD loop. It is less clear what regulatory function Enh2 plays, as it does not interact with either *Nox4* or the downstream inactive gene *Tyr* (FPKM < 0.01 in both sexes).

DISCUSSION

This work represents the first study of sex differences in autosomal 3D genome organization. We show that more than 1,000 CTCF or cohesin peaks are occupied at different levels between male and female mouse liver. While cohesin-non-CTCF (CNC) sites are closer to enhancers than other peak classes, only a minority of any sex-biased CTCF/cohesin peaks are proximal to sex-biased genes or enhancers. We present several examples of sex differential CTCF/cohesin binding putatively linked to sex-biased genes that are proximal (male-biased: *Cml5*, *Nat8*; female-biased: *Fam84b*) or distal (male-biased: *C8a*, *C8b*; female-biased *Slc22a29*). Using publicly available data, we show that these distally regulated male-biased genes are more sensitive to cohesin depletion, providing additional evidence that cohesin is directly or indirectly involved in linking distal sex-biased enhancers to proximal gene targets. To directly test for differences in chromatin interactions between male and female mouse liver, we performed circularized chromosome conformation capture with sequencing (4C-seq) for 6 regions neighboring 5 highly sex-biased genes. These highlight examples of the direct contribution of female-biased CTCF binding towards female-biased interactions for *A1bg* (F/M=155) as well as female-biased cohesin binding facilitating proximal enhancer-promoter interactions near *Gm4794* (F/M= 704). The indirect contribution of CTCF and cohesin to liver sex differences is shown for the male-biased gene *C9* (M/F=3.5), which interacts more strongly with a distal male-biased enhancer, bypassing a lowly expressed gene insulated by an intra-TAD loop. Intra-TAD loops also insulate the male-biased gene *Nudt7* and an upstream super-enhancer and limit *Nox4* to proximal enhancer regulation by restricting access to an intronic enhancer. Taken together, chromatin interactions link distal sex-biased enhancers to genes which is guided both directly and indirectly by CTCF and cohesin looping.

Despite identifying ~1,000 sex-biased binding sites for CTCF or cohesin, the majority are distal from enhancers or genes with known sex bias. Although these distal sites lack any obvious link to sex-biased gene expression in wild type liver, they could have a priming effect and contribute to sex-specific responses to hepatic stressors, such as high fat diet [59] or xenobiotic exposure [60]. Just as short-term feeding of a high fat diet can leave a lasting "epigenetic memory" in the form of epigenetic modifications [61], sex-biased CTCF/cohesin binding sites may prime each sex for differential looping patterns in response to subsequent chemical exposure or dietary stressors. It should also be noted that our datasets likely capture only a subset of sex-biased CTCF/cohesin binding sites for adult mice at 8 weeks of age. Liver sex differences in DNA methylation increase with age [50, 62], and the binding of CTCF is highly sensitive to DNA methylation within its motif [48, 49, 63]. The hypomethylation of enhancers in male liver (lower methylation in males than females; [50]) could open up additional CTCF sites, an effect that would presumably be more pronounced in older mice. At 8 weeks of age,

we found that the majority of male-biased CTCF binding is at Lone CTCF sites, which tend to be closer to both enhancers and TSS than CAC sites (Figure 2B,C).

There is substantial evidence that the majority of insulated, CAC-mediated loops are invariant across most tissues and cell types studied [30]. A significant minority of these loops are tissue-specific, typically on the order of 20-30% [32]. Using a computational method for loop prediction that we recently described [32], we do not find substantial evidence for the presence of sex-specific insulated intra-TAD loops. A cell-type-specific intra-TAD loop was described in erythroid cells, despite no obvious differences in CTCF or cohesin binding from embryonic cells [64]. If this phenomenon is widespread, then some, and perhaps many sex-biased intra-TAD loops could be formed by sex-independent CTCF/cohesin anchors via an unknown mechanism. In the absence of global assessment of all sex-biased looping (i.e. using Hi-C), we conclude that intra-TAD loops are primarily sex-independent features that contribute indirectly to sex-biased gene expression.

Cohesin-and-CTCF sites may directly link enhancers and promoters, in addition to their well-characterized role in domain-level organization [35, 42]. This finding suggests that in addition to the largely invariant structural features of TADs and insulated intra-TAD loops, such "non-anchor" CAC sites may contribute to tissue-specific enhancer–promoter or enhancer–enhancer interactions. As the majority of sex-differential CTCF/cohesin binding occurs at non-anchor CAC sites, long distance CAC-mediated interactions may link sex-biased enhancers to promoters of sex-biased genes. Specific examples include the enhancer–enhancer contacts observed for *A1bg* (in the context of female-biased CTCF binding; Figure 4A), and may also characterize the distal female-biased enhancer neighboring *Slc22a29* (female-biased CTCF and cohesin binding at a robust female-biased enhancer; Figure 3D).

The interacts of the promoter of *C9* with a male-biased enhancer >200 kb upstream is an interesting example of CAC-looping indirectly guiding enhancer–promoter contact. An intervening intra-TAD loop insulates the sex independent (and weakly-expressed) gene *Dab2* gene, while bringing *C9* and a cluster of male-biased enhancers in closer proximity in both male and female liver. In the absence of any sex-biased CTCF or cohesin binding to explain the increased contact in male liver, it is probable that other known looping factors augment direct enhancer–promoter interactions specifically in male liver. Mechanistically, this could be driven by protein (such as YY1[53] or mediator[25]) or non-protein factors such as enhancer RNAs (eRNAs) [65, 66]. Although data is only available for male mouse liver, 3 of the 4 distal upstream enhancers are actively transcribed to produce bidirectional eRNA transcripts [67] and 4 of 4 are bound by the protein YY1 [68].

Sex differences in mouse and human liver are linked to the sex-specific pattern of growth hormone secretion by the pituitary gland [69-71]. While we have demonstrated the presence of proximal and distal sex-biased looping in mouse liver, it is not clear how dynamic these loops are in response to factors that dysregulate sex-biased gene expression following alterations in hormone patterns [72] or xenobiotic exposure [60]. In an extreme example of dynamic looping, complete loss of cohesin followed by its reintroduction (achieved by auxin-induced targeted degradation and withdrawal) establishes that virtually all loops are lost after 40 min then regained within an hour of auxin withdrawal [22]. During macrophage differentiation, extensive changes are observed to gene expression, the enhancer landscape, and chromatin interactions linking them [73]; however IMR90 cells show extensive changes in enhancer activity and gene expression in response to TNFα treatment yet virtually no differences in DNA looping are observed [74]. Studies on circadian effects in mouse liver have shown selective interactions between circadian oscillating genes and oscillating enhancers, but did not directly show a related rhythmicity of 3D interactions (either no observed differences [75], or the comparison was not made [76]). Based on the sensitivity of distally-regulated genes to cohesin depletion, we predict that looping from distal sex-biased enhancers would be more responsive to alterations of the plasma growth hormone pattern. Alternatively, the majority of loops might not be directly responsive to changes in growth hormone stimulation, as in treated IMR90 cells [74], or that sex differences in 3D interactions are a result sex hormones or other factors.

Chromosome conformation capture technologies have been extended beyond the interaction analysis to study genomic rearrangements and copy number variation [77]. Here we also show that 3C-based assays may also be a useful tool to determine candidate lncRNAs for further analysis. The highly female-biased gene *A1bg* is nearby many lncRNAs, however, we observed that a subset of interacting lncRNAs was consistently female-biased across different studies, experiment types, and mouse strains (Figure S5C,D) — suggestive they may play a functional role. Currently, enhancer-associated lncRNAs are defined as those transcribed from enhancers and show interesting properties such as increased enhancer activity, tissue-specificity, and association with expression of nearby protein coding genes [78, 79]. Our 4C-seq results for the *A1bg* enhancer suggest that this model could be extended to enhancer-interacting lncRNAs as an approach to characterize or screen for candidate functional lncRNAs. This would likely best be accomplished by implementation of a more global interaction-based assay, such as Hi-C.

The vast majority of sex-biased CTCF and/or cohesin binding is intergenic and distal to TSS and sex-biased genes (Figure 2C,F). While most sex-biased CNC peaks overlap enhancers (median distance ~0.2 kb; Figure 2D), only ~25% of sex-biased CNC peaks are within 1 kb of a sex-biased enhancer (Figure 2G). Overall, we estimate that less than half of sex-biased CTCF/cohesin peaks contribute to sex-biased gene expression, as they are not at least within the same TAD as a sex-biased gene (Figure 2F). However, this does not mean that 3D organization in general does not contribute to the extensive transcriptomic and epigenetic sex bias in mouse liver. Previously described intra-TAD loops may indirectly facilitate male-biased chromatin interactions for genes like *C9* and insulate the highly active and super-enhancer regulated *Nudt7*, all despite lacking sex-biased CTCF or cohesin binding. Similarly, the nested intra-TAD loops insulating *Nox4* prevent distal enhancer-promoter interaction with an intronic enhancer in favor of more proximal regulation by a cluster of upstream enhancers. Highly female-biased genes, such as *A1bg* and *Gm4794,* show female-biased CTCF and cohesin binding, respectively, as well as female-biased chromatin interactions at distances ranging from 10 to 100 kb. Thus, CTCF and cohesin can contribute in both direct and indirect ways to guide sex-biased enhancer–promoter contacts in mouse liver.

MATERIALS AND METHODS

Animal Work and Tissue Processing - Adult CD-1 mice were purchased from Charles River Laboratories [cat#:Crl:CD1(ICR)] and sexes were housed separately in the Boston University Laboratory Animal Care Facility. Animals were treated and housed until 8 weeks of age using protocols specifically reviewed for ethics and approved by Boston University's Institutional Animal Care and Use Committee (IACUC). Mice were then euthanized via cervical dislocation and their livers removed and rinsed in ice cold Phosphate Buffered Saline (PBS). Livers were homogenized in a Potter-Elvehjem homogenizer in high sucrose homogenization buffer (10 mM HEPES (pH 7.5), 25 mM KCl, 1 mM EDTA, 2 M sucrose, 10% glycerol, 0.05 mM DTT, 1 mM PMSF, 0.15 mM spermine, 0.2% (v/v) spermidine, 1 mM Na orthovanadate, 10 mM NaF, and 1X Roche Complete Protease Inhibitor Cocktail) to preserve nuclei structure. This suspension was layered on top of a 3 ml cushion of homogenization buffer and ultracentrifuged at 25,000 RPM for 30 min at 4°C in an SW41 Ti rotor to obtain clean nuclei in a loose pellet. The supernatant was removed, and the pelleted nuclei were resuspended in 1 ml of crosslinking buffer (10 mM HEPES buffer (pH 7.6), 25 mM KCl, 0.15 mM 2-mercaptoethanol, 0.34 M sucrose, 2 mM MgCl2). This nuclear suspension was transferred to a 1.5 ml Eppendorf tube at 30°C, formaldehyde was added to a final concentration of 0.8% (v/v), and incubated with occasional mixing for 9 min at 30°C. Crosslinking was halted by the addition of 110 µl of 1 M glycine, followed by a 5 min incubation at room temperature. The crosslinked suspension was then layered atop homogenization buffer and pelleted, as above, prior to proceeding with sonication.

Pelleted samples were resuspended in 1 ml of 4°C 1X Radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% IPEGAL, 0.5% deoxycholic acid, and 1X Roche Complete Protease Inhibitor Cocktail) containing 0.5% SDS by gentle pipetting. This material was then transferred to a 15 ml polystyrene tubes (BD Falcon, cat#:352095) and sonicated according to the manufacturer's instructions with a Bioruptor

Twin instrument (UCD-400). This consisted of 75 cycles at high intensity setting in a 4°C water bath, where 1 cycle is ON for 30 s and OFF for 30 s. The sonicated chromatin was spun at 18,000 x g for 10 min at 4°C to clear any debris, and the supernatant was snap frozen in liquid nitrogen. The bulk of this material was stored at -80°C prior to ChIP, while a small aliquot was reverse crosslinked to quantify the amount of sonicated chromatin and the fragment size distribution. This aliquot was adjusted to a final concentration of 0.2M NaCl and crosslinks were reversed for 6 h at 65°C in a thermocycler with heated lid. This sample was dilute 3-fold in water, then digested with RNase A (Novagen, cat#:70856) for 30 min at 37°C followed by proteinase K (Bioline, cat#: BIO-37084) for 2 h at 56°C. Fragment size distribution for the resulting DNA was assessed by 1% agarose gel electrophoresis to ensure the bulk of material was < 500 bp in size. Chromatin quantification was performed using a Quanti-iT PicoGreen assay (Invitrogen, cat#:Q33130), back calculating to determine the undiluted concentration of the bulk chromatinized DNA stored for downstream ChIP experiments.

Chromatin Immunoprecipitation and Sequencing (ChIP-seq) Experimental Protocol - All male ChIP samples in this study are from a previous work [32], deposited at GSE102997. Male and female biological replicates are from individual mouse livers with 8 total CTCF ChIP-seq samples per sex (n=4 males and n=4 females) and 6 total Rad21 ChIP-seq samples (n=3 males and n=3 females). All samples were processed using the same protocol and conditions throughout, as follows.

Immunoprecipitation of chromatin from mouse liver was performed as described previously [13]. Specifically, 5 µl of antibody to either CTCF (Millipore, cat#:07-729) or to the cohesin subunit Rad21 (Abcam, cat#:ab992) were mixed with 30 µl of Protein A Dynabeads (Invitrogen, cat#:1002D) and incubated in blocking solution (0.5% bovine serum albumin in PBS) for 3 h at 4°C. Beads were then washed with blocking solution, followed by incubation overnight with 70 µg of liver chromatin in 1X RIPA (containing 0.1% SDS). After washing with 1X RIPA (containing 0.1% SDS), formaldehyde crosslinks were reversed by incubation for 6 h at 65°C followed by RNase A digestion (Novagen, cat#:70856) at 37°C for 30 min then proteinase K digestion (Bioline, cat#:37084) at 56 °C for 2 h. The resulting crude DNA extract was purified using the QIAquick Gel Extraction Kit (Qiagen, cat#:28706) and quantified with the Qubit HS DNA kit (Invitrogen, cat#:Q32854).

Circularized Chromatin Conformation Capture with Sequencing (4C-seq) Experimental Protocol - Nuclei isolation and crosslinking were performed as above in tissue processing, through pelleting of crosslinked nuclei. Digestion, proximity ligation, and inverse PCR were carried out as described previously [32]. Pelleted nuclei were resuspended in Buffer A (15 mM Tris-HCl pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.5 mM spermidine, 0.3 mM spermine) and quantified using the Countess Automated Cell Counter (Invitrogen, cat#:C10227). Nuclei were aliquoted so that 10 million nuclei were used per individual 4C experiment, pelleted at 1,000 x g for 5 min at 4°C, and resuspended in 1X NEBuffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9; NEB, cat#:B7003S). Primary restriction enzyme digestion of intact nuclei was conducted with 50,000 U of DpnII (NEB: #R0543) overnight at 37°C with agitation at 900 RPM. DpnII was inactivated with a final concentration of 2% SDS, then diluted 5-fold in 1X ligation buffer (Enzymatics, cat#:B6030). Proximity ligation was performed overnight at 16°C with 200U of T4 DNA ligase (Enzymatics, cat#:L6030). At this point, DNA was reverse crosslinked and purified using a standard phenol/chloroform cleanup according to the manufacturer's protocol (VWR, cat#:0883). Secondary digestion of purified DNA was performed with 50 U of Csp6I (Thermo Scientific, cat#:ER0211) overnight at 37°C in 500 µl of 1X Buffer B (Fermentas, cat#BB5; 10 mM Tris-HCl (pH 7.5), 10 mM MgCl2, 0.1 mg/ml BSA). To halt enzymatic activity, Csp6I was heat inactivated for 30 min at 65°C. Samples were diluted 10-fold and secondary ligation 10-fold and secondary ligation was carried out as above, overnight at 16°C. Primary digestion, proximity ligation, secondary digestion, and secondary ligation were each independently verified by reverse crosslinking and gel electrophoresis as described in the sonication section of "Animal Work and Tissue Processing," above. An example gel image is shown in Figure S5A. The final PCR template was purified by phenol/chloroform clean up, followed by QiaPrep 2.0 column cleanup (Qiagen, cat#:27115). This resulted in a standard, circularized 4C inverse PCR template amplified by specific viewpoint primers, below.

PCR reactions were performed using primer pair sequences for valid 4C-seq viewpoints (VPs) from the 4CSeqpipe primer database (http://www.wisdom.weizmann.ac.il/~atanay/4cseq_pipe/) with the addition of 5' dangling truncated Illumina adapters (Table S3A). Candidate viewpoints were selected based on the following criteria. First, the VP was required to be within the same TAD as a sex-biased protein-coding or lncRNA gene with >3-fold M/F bias in expression. Second, the VP must either fall within 1 kb of the TSS sex biased gene or else overlap a sex-biased enhancer (minimum 2-fold M/F bias in DHS opening or K27ac mark accumulation). Third, the nonreading primer was required to uniquely map to the genome, while the reading primer was more stringently required to have > 89% unique sequence identity (i.e., no 18-mer within the primer mapped elsewhere in the genome). Amplification of 4C templates was performed with Platinum Taq DNA polymerase (Invitrogen, cat#:10966026) under the following conditions: 94°C for 2 min, 25 cycles at (94°C 30 s, 55°C 30 s, 72°C 3 min), then 4°C hold. To limit the effect of PCR artifacts, a total of 6 identical inverse PCR reactions per sample were processed in parallel and then pooled prior to library preparation. For Nox4 M1, M2, F1, and F2 samples, libraries were prepared independently for singular inverse PCR reactions which were then pooled at the raw read level (equivalent of a pool of 2 identical PCR reactions). Example pooled 4C libraries are shown in Figure S5B.

4C-seq samples were amplified using NEB barcoded primers such that each biological replicate had a unique barcode (NEB, cat#:E7335). To prevent overamplification, each sample was amplified with 5 additional cycles of PCR to fill in the full Illumina adapter sequence necessary for sequencing. Due to the low complexity in the first 20 bases of the sequence read, 4C libraries were multiplexed and sequenced with additional high complexity libraries (i.e. RNA or ChIP-seq). In practice, 4C libraries constituted no more than 15% of the total library pool, by molarity. Samples were sequenced on an Illumina HiSeq 2500 with either 50, 125, or 150 bp paired end read lengths at the New York Genome Center (50 and 125 bp reads) and Novogene (150 bp reads).

Computational Analysis of ChIP-seq Datasets - Sequence reads were split by barcode and mapped to the mm9 genome assembly with Bowtie2 (v2.2.9). All reads not uniquely mapped to the genome were removed for downstream analyses. Peak calling was performed using MACS2 (v2.1.1) with default parameters, then peaks overlapping blacklisted genomic regions (www.sites.google.com/site/anshulkundaje/projects/blacklists) were filtered out. Additionally, we removed spurious peaks that contain only PCR duplicated reads, defined as 5 or more identical sequence reads that do not overlap any other reads. All BigWig tracks for visualization in a genome browser were normalized for sequencing depth and sample quality, expressed as reads in peaks per million mapped reads (RIPM). In practice, the y-axis shows the read count from a given sample divided by the total number of reads in peaks (reads that overlap a called peak in any sample, or the union peak list), per million. Normalization was performed separately for CTCF and cohesin datasets. This approach is functionally similar to the common quality control metric known as Fraction of Reads in Peaks (FRiP) used by the ENCODE consortium and others [80]. All samples used in this study were of good quality with a minimum FRiP of 0.1034 (male Rad21 ChIP-seq; G133_M6) with a full listing in Table S3B.

Differential peaks were identified as described previously with some modifications [REF_andy]. DiffReps (v1.55.4) was used to identify in-peak differential sites with default parameters and a window size of 200 bp. This output was then filtered to require that differential sites overlap a peak from the per-factor union peak list, contain a minimum of 10 sequence reads, show a minimum 2-fold sex difference, and have an FDR < 0.05. Distance to features and overlap analysis were conducted using Bedtools (v2.26.0). TAD and Intra-TAD coordinates for mm9 were downloaded from reference [32], where TAD definitions are based on prior experimental Hi-C analysis in male mouse liver [34]. For browser visualization, a secondary "lenient" sex-biased peak list was also defined as follows. Biological replicates were merged by sex and factor (M CTCF, M cohesin, F CTCF, and F cohesin), then the merged samples with a higher FRiP (compared to the same factor, but opposite sex) was downsampled so that the mapped read files for each sex contained the same number of reads in peaks. In practice, this meant that the mapped reads from merged female cohesin and merged male CTCF were downsampled by a factor of 0.979869 and 0.745955, respectively. MAnorm was then used to compare single replicates from each sex to define the "lenient" sex-differential binding for a given factor

(minimum 2-fold sex bias; p adj < 0.01; read count > 15 for the upregulated peak; [81]). The strict and lenient lists are mutually exclusive and do not overlap.

Differential peaks were then further subdivided into 4 groups [CAC(ΔCoh), CAC(ΔCTCF), CNC(ΔCoh), or Lone CTCF(ΔCTCF)] based on the following criteria, separately for male and females. Differential cohesin peaks are considered CAC if they overlap peaks in 3 or 4 of the CTCF ChIP-seq biological replicates (out of a total of 4). Alternatively, they are considered CNC if they overlap peaks in 0 or 1 of the CTCF-seq biological replicates (out of a total of 4; those overlapping 2 CTCF replicates were removed). Similarly, differential CTCF peaks are considered CAC if they overlap peaks in 2 or 3 of the cohesin ChIP-seq biological replicates (out of a total of 3). Differential CTCF peaks are considered "Lone CTCF" if they overlap peaks in 0 or 1 of the cohesin ChIP-seq biological replicates (out of a total of 3). Control "sex-independent peaks" for CTCF and cohesin were defined by ranking each peak according to the ratio of the RIPM normalized ChIP signal for the merged male over the RIPM normalized ChIP signal for the merged female sample (separately for CTCF and cohesin). The 1,000 peaks whose ratio was closest to 1 were defined as "sex-independent" CTCF or cohesin peaks.

Enhancer DHS were defined by a high ratio of H3K4me1 to H3K4me3 ChIP signal as defined previously [32]. CTCF motif discovery was performed using the FIMO option from MEME Suite (v4.10.0), and presence of a motif is defined as a motif score > 10.

Intra-TAD loops for female samples were identified exactly as described previously for male liver [32]. The pipeline was run with default parameters for all CAC sites for an initial loop count of 20,000, again identical to previous parameters for male liver [32]. Overall, we identified 9,724 loops in female (compared to 9,052 in male) with 10,273 loop anchors (compared to 9,052 in male). Considering the lack of sex-differential CAC at TAD or intra-TAD anchors we did considered the majority of these to be the result of technical noise. Of the ~10% of loops that were unique to males or females they tended to be weaker than shared loops, narrowly meeting the cutoff in one sex but not the other. Reciprocal overlap between loops was performed using bedtools (bedtools intersect –wa –u –r –f 0.8) as described previously [32].

Computational Analysis of 4C-seq Datasets - Biological replicates were first demultiplexed by index read barcode. As fastq files for each biological replicate contained multiple viewpoints, reads were further split based on matches to the reading primer for each viewpoint (Table S3A). Then, 20 nts from the 5' end of reads were removed (representing the reading primer) using the FASTX-Toolkit (v0.0.14) prior to mapping. For consistency, 25 nts were also trimmed from the 3' end of 150 nt reads so that they were identical in length to the 125 nt libraries (after 5' and 3' trimming, both were 105 bases in length). Reads were then iteratively mapped to a reduced mouse genome using Bowtie2 (v2.2.2). Rather than mapping to the full genome, this reduced genome contained only 105 bp upstream and 105 bp downstream from any DpnII cut site in the genome (GATC). An iterative mapping strategy was used because some reads contain multiple ligation junctions, rendering the full-length read unmappable as described in some Hi-C pipelines [82]. Any non-uniquely mapped reads were trimmed by ~10% of their total length (starting from the 3' end of the read), and mapping was reattempted as above. This process was repeated to a minimum read length of 20 nt, with a stepsize of 2 bp (for shorter reads; 50 nt) or 10 bp (for longer reads; 125 or 150 nt) per iteration. The net effect of trimming and remapping increased the overall percent of uniquely mapping reads by 2-6% for shorter read lengths and 8-22% for longer read length libraries.

After mapping bedGraph files were smoothed using a sliding window of 11 restriction fragments taking the median value in the window. Smoothed BigWig tracks were normalized by total reads per million to account for differences in sequencing depth. Merged replicates shown in the main Figure panels were generated by taking the median signal at a given restriction fragment per viewpoint and sex. The window overlapping the viewpoint fragment was removed prior to BigWig generation for merged replicates. Tracks were visualized in the WashU genome browser with replicates overlaid using the Matplot feature. Table S3C shows the sequences used for demultiplexing 4C libraries as well as basic statistics for read mapping and quality control.

Analysis of Publicly Available Datasets - Plots showing the impact of cohesin depletion on expression of protein coding and lncRNA genes were based on data from [21], using RNA-seq datasets from n=4 (each) male mouse livers for WT and Nipbl-depleted replicates. Using RPKM normalized data, reads are expressed as expression relative to the mean of the WT group (set as 1). This was accomplished on a per gene basis by dividing the expression value for each individual replicate by the mean of the WT group plus a small psuedocount to avoid dividing by zero (1e-6). This data is presented as mean relative expression ± SD for all plots. The underlying values for all genes are shown in Table S2A.

The following datasets did not originate from this study but were used to generate screenshots: H3K27ac ChIP-seq data (Lau Corona *et. al., in Review*), Dnase-seq [12], male CTCF/cohesin [32], sex-biased lncRNAs [83], and M/F expression ratio [72]. Unless otherwise noted, FPKM and M/F expression ratios for protein coding genes are based on Ribo-depleted total RNA while lncRNA values are based on Ribo-depleted nuclear RNA. Bigwigs were RIPM normalized separately per experiment type as described above ("Computational Analysis of ChIP-seq Datasets"). Below DHS and H3K27ac BigWig signal tracks the regions with significant sex bias are indicated. Sex-biased DHS and H3K27ac-marked regions were split into groups on the principal of defining "strict" and "lenient" subclasses of sex-biased binding. In practice, DHS are shaded from dark to light for high, standard, and low stringency male- (blue) and female-biased (pink) DHS, as defined previously [12]. The low and high stringency bear the greatest resemblance to the "lenient" and "strict" definitions used here. For the H3K27ac sex-biased tracks, four total groups were defined based on the magnitude of the FC difference between male and female samples (M/F or F/M > 2.5 defined as "strict"; 1.5 < M/F or F/M < 2.5 defined as "lenient"). Both "strict" and "lenient" sex-biased H3K27ac peaks were defined using MAnorm (p adj < 0.05; read count > 15 for the upregulated peak; [81]). These cutoffs resulted in 1,583 female-biased H3K27ac peaks (380 strict, 1203 lenient) as well as 2,241 male-biased H3K27ac peaks (604 strict, 1637 lenient).

Pan-cancer human expression data was downloaded for Nox4 using the web software TIMER [84]. Only tissues with matched "Normal" and "Tumor" expression values are shown with otherwise default parameters. Nox4 expression for male C3H mouse liver and neoplasms was downloaded as processed, normalized RNA-seq counts from a previous study (E-MTAB-6972; [58]).

Basic Statistical Tests and Approaches - Boxplots, cumulative distribution plots, and statistical analyses were performed using GraphPad Prism 7. Unless otherwise indicated, all pairwise comparisons used two-tailed, nonparametric t tests. Comparisons of distributions were performed using a Kolmogorov–Smirnov (KS) test, while comparisons of values used a Mann-Whitney (MW) test. The results are annotated where **** indicates $p \leq 0.0001$, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$, and ns $p > 0.05$.

ACKNOWLEDGEMENTS - We would like to thank Aram Shin for optimizing Rad21 and CTCF ChIP-seq experiments in this lab. Also, thanks to Tisha Melia and Andy Rampersaud for developing the pipelines used in ChIP-seq and RNA-seq analysis. Further we'd like to thank Nicholas O'Neill for useful comments and improvements to the 4C-seq pipeline and mapping strategy.

CONFLICT OF INTEREST - The authors declare that they have no competing interests.

REFERENCES

1. Mayne BT, Bianco-Miotto T, Buckberry S, Breen J, Clifton V, Shoubridge C, Roberts CT: **Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans.** *Frontiers in Genetics* 2016, **7:**183.
2. Qu K, Zaba LC, Giresi PG, Li R, Longmire M, Kim YH, Greenleaf WJ, Chang HY: **Individuality and variation of personal regulomes in primary human T cells.** *Cell Systems* 2015, **1:**51-61.
3. Rinn JL, Rozowsky JS, Laurenzi IJ, Petersen PH, Zou K, Zhong W, Gerstein M, Snyder M: **Major molecular differences between mammalian sexes are involved in drug metabolism and renal function.** *Developmental Cell* 2004, **6:**791-800.
4. Zhang Y, Klein K, Sugathan A, Nassery N, Dombkowski A, Zanger UM, Waxman DJ: **Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease.** *PLoS One* 2011, **6:**e23506.
5. Ruggieri A, Barbati C, Malorni W: **Cellular and molecular mechanisms involved in hepatocellular carcinoma gender disparity.** *Int J Cancer* 2010, **127:**499-504.
6. Guy J, Peters MG: **Liver disease in women: the influence of gender on epidemiology, natural history, and patient outcomes.** *Gastroenterology Hepatology* 2013, **9:**633.
7. Buzzetti E, Parikh PM, Gerussi A, Tsochatzis E: **Gender differences in liver disease and the drug-dose gender gap.** *Pharmacol Res* 2017, **120:**97-108.
8. Clodfelter KH, Holloway MG, Hodor P, Park SH, Ray WJ, Waxman DJ: **Sex-dependent liver gene expression is extensive and largely dependent upon signal transducer and activator of transcription 5b (STAT5b): STAT5b-dependent activation of male genes and repression of female genes revealed by microarray analysis.** *Molecular Endocrinology* 2006, **20:**1333-1351.
9. Melia T, Hao P, Yilmaz F, Waxman DJ: **Hepatic Long Intergenic Noncoding RNAs: High Promoter Conservation and Dynamic, Sex-Dependent Transcriptional Regulation by Growth Hormone.** *Mol Cell Biol* 2016, **36:**50-69.
10. Melia T, Waxman DJ: **Sex-biased lncRNAs inversely correlate with sex-opposite gene co-expression networks in Diversity Outbred mouse liver.** *Endocrinology* 2019.
11. Hao P, Waxman DJ: **Functional Roles of Sex-Biased, Growth Hormone-Regulated MicroRNAs miR-1948 and miR-802 in Young Adult Mouse Liver.** *Endocrinology* 2018, **159:**1377-1392.
12. Ling G, Sugathan A, Mazor T, Fraenkel E, Waxman DJ: **Unbiased, genome-wide in vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression.** *Molecular and Cellular Biology* 2010, **30:**5531-5544.
13. Sugathan A, Waxman DJ: **Genome-Wide Analysis of Chromatin States Reveals Distinct Mechanisms of Sex-Dependent Gene Regulation in Male and Female Mouse Liver.** *Molecular and Cellular Biology* 2013, **33:**3594-3610.
14. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137:**1194-1211.
15. Ghirlando R, Felsenfeld G: **CTCF: making the right connections.** *Genes Development* 2016, **30:**881-891.
16. Mehta GD, Rizvi SMA, Ghosh SK: **Cohesin: a guardian of genome integrity.** *Molecular Cell Research* 2012, **1823:**1324-1342.
17. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al: **Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes.** *Proceedings of the National Academy of Sciences* 2015, **112:**E6456-6465.
18. Davidson IF, Goetz D, Zaczek MP, Molodtsov MI, Huis In 't Veld PJ, Weissmann F, Litos G, Cisneros DA, Ocampo-Hafalla M, Ladurner R, et al: **Rapid movement and transcriptional re-localization of human cohesin on DNA.** *EMBO J* 2016, **35:**2671-2685.
19. Fedoriw AM, Stein P, Svoboda P, Schultz RM, Bartolomei MS: **Transgenic RNAi Reveals Essential Function for CTCF in H19 Gene Imprinting.** *Science* 2004, **303:**238.

20. Xu H, Balakrishnan K, Malaterre J, Beasley M, Yan Y, Essers J, Appeldoorn E, Tomaszewski JM, Vazquez M, Verschoor S, et al: **Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice.** *PLoS One* 2010, **5:**e12112.
21. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L: **Two independent modes of chromatin organization revealed by cohesin removal.** *Nature* 2017, **551:**51.
22. Rao SS, Huang S-C, St Hilaire BG, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID: **Cohesin loss eliminates all loop domains.** *Cell* 2017, **171:**305-320. e324.
23. Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG: **Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization.** *Cell* 2017, **169:**930-944. e922.
24. Ciosk R, Shirayama M, Shevchenko A, Tanaka T, Toth A, Shevchenko A, Nasmyth K: **Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins.** *Molecular Cell* 2000, **5:**243-254.
25. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467:**430-435.
26. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, Young RA: **Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes.** *Cell* 2014, **159:**374-387.
27. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, et al: **Architectural protein subclasses shape 3D organization of genomes during lineage commitment.** *Cell* 2013, **153:**1281-1295.
28. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT: **A CTCF-independent role for cohesin in tissue-specific transcription.** *Genome Res* 2010, **20:**578-588.
29. Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, Xu H, Ramsay RG, Odom DT, Flicek P: **Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules.** *Genome Research* 2012, **22:**2163-2175.
30. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell* 2014, **159:**1665-1680.
31. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: **Topological domains in mammalian genomes identified by analysis of chromatin interactions.** *Nature* 2012, **485:**376-380.
32. Matthews BJ, Waxman DJ: **Computational prediction of CTCF/cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver.** *Elife* 2018, **7: 10.7554/eLife.34077**.
33. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al: **Chromatin architecture reorganization during stem cell differentiation.** *Nature* 2015, **518:**331-336.
34. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S: **Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture.** *Cell Reports* 2015, **10:**1297-1309.
35. Ren G, Jin W, Cui K, Rodrigez J, Hu G, Zhang Z, Larson DR, Zhao K: **CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression.** *Molecular Cell* 2017, **67:**1049-1058. e1046.
36. Splinter E, Heath H, Kooren J, Palstra R-J, Klous P, Grosveld F, Galjart N, de Laat W: **CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus.** *Genes and Development* 2006, **20:**2349-2354.
37. Huang J, Li K, Cai W, Liu X, Zhang Y, Orkin SH, Xu J, Yuan G-C: **Dissecting super-enhancer hierarchy based on chromatin interactions.** *Nature Communications* 2018, **9:**943.
38. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137:**1194-1211.
39. Donohoe ME, Zhang L-F, Xu N, Shi Y, Lee JT: **Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch.** *Molecular Cell* 2007, **25:**43-56.

40. Pugacheva EM, Rivero-Hinojosa S, Espinoza CA, Méndez-Catalá CF, Kang S, Suzuki T, Kosaka-Suzuki N, Robinson S, Nagarajan V, Ye Z: **Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions.** *Genome Biology* 2015, **16:**161.
41. Park HJ, Li J, Hannah R, Biddie S, Leal-Cervantes AI, Kirschner K, Flores Santa Cruz D, Sexl V, Gottgens B, Green AR: **Cytokine-induced megakaryocytic differentiation is regulated by genome-wide loss of a uSTAT transcriptional program.** *EMBO Journal* 2016, **35:**580-594.
42. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al: **CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function.** *Cell* 2015, **162:**900-910.
43. Oti M, Falck J, Huynen MA, Zhou H: **CTCF-mediated chromatin loops enclose inducible gene regulatory domains.** *BMC Genomics* 2016, **17:**252.
44. Whalen S, Truty RM, Pollard KS: **Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin.** *Nature Genetics* 2016, **48:**488.
45. Yang Y, Zhang R, Singh S, Ma J: **Exploiting sequence-based features for predicting enhancer–promoter interactions.** *Bioinformatics* 2017, **33:**i252-i260.
46. Dekker J, Rippe K, Dekker M, Kleckner NJs: **Capturing chromosome conformation.** 2002, **295:**1306-1311.
47. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, et al: **Robust 4C-seq data analysis to screen for regulatory DNA interactions.** *Nature Methods* 2012, **9:**969-972.
48. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM: **CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.** *Nature* 2000, **405:**486.
49. Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV: **Critical dna binding interactions of the insulator protein ctcf a small number of zinc fingers mediate strong binding, and a single finger-dna interaction controls binding at imprinted loci.** *Journal of Biological Chemistry* 2007, **282:**33336-33345.
50. Reizel Y, Spiro A, Sabag O, Skversky Y, Hecht M, Keshet I, Berman BP, Cedar H: **Gender-specific postnatal demethylation and establishment of epigenetic memory.** *Genes and Development* 2015, **29:**923-933.
51. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B: **A map of the cis-regulatory sequences in the mouse genome.** *Nature* 2012, **488:**116-120.
52. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC: **Comparison of the transcriptional landscapes between human and mouse tissues.** *Proceedings of the National Academy of Sciences* 2014, **111:**17224-17229.
53. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL: **YY1 is a structural regulator of enhancer-promoter loops.** *Cell* 2017, **171:**1573-1588. e1528.
54. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES: **The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome.** *Science* 2013, **341:**1237973.
55. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR: **Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.** *Nature Structural Molecular Biology* 2014, **21:**198.
56. Kung JT, Kesner B, An JY, Ahn JY, Cifuentes-Rojas C, Colognori D, Jeon Y, Szanto A, del Rosario BC, Pinter SF, et al: **Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF.** *Molecular Cell* 2015, **57:**361-375.
57. Borrie MS, Campor JS, Joshi H, Gartenberg MR: **Binding, sliding, and function of cohesin during transcriptional activation.** *Proceedings of the National Academy of Sciences* 2017, **114:**E1062-E1071.
58. Aitken SJ, Ibarra-Soria X, Kentepozidou E, Flicek P, Feig C, Marioni JC, Odom DT: **CTCF maintains regulatory homeostasis of cancer pathways.** *Genome Biology* 2018, **19:**106.
59. Keleher MR, Zaidi R, Hicks L, Shah S, Xing X, Li D, Wang T, Cheverud JM: **A high-fat diet alters genome-wide DNA methylation and gene expression in SM/J mice.** *BMC genomics* 2018, **19:**888.

60. Lodato NJ, Melia T, Rampersaud A, Waxman DJ: **Sex-Differential Responses of Tumor Promotion-Associated Genes and Dysregulation of Novel Long Noncoding RNAs in Constitutive Androstane Receptor-Activated Mouse Liver.** *Toxicological Sciences* 2017, **159:**25-41.
61. Leung A, Trac C, Du J, Natarajan R, Schones DE: **Persistent chromatin modifications induced by high fat diet.** *Journal of Biological Chemistry* 2016, **291:**10446-10455.
62. Takasugi M, Hayakawa K, Arai D, Shiota K: **Age- and sex-dependent DNA hypomethylation controlled by growth hormone in mouse liver.** *Mechanisms of Ageing and Development* 2013, **134:**331-337.
63. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R: **Widespread plasticity in CTCF occupancy linked to DNA methylation.** *Genome Research* 2012, **22:**1680-1688.
64. Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, Downes DJ, Gosden M, Sharpe JA, Sloane-Stanley JA, Hughes JR, et al: **Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo.** *Nature Cell Biology* 2017, **19:**952-961.
65. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X: **Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.** *Nature* 2013, **498:**516.
66. Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RAJS: **Transcription factor trapping by RNA in gene regulatory elements.** 2015, **350:**978-981.
67. Fang B, Everett LJ, Jager J, Briggs E, Armour SM, Feng D, Roy A, Gerhart-Hines Z, Sun Z, Lazar MA: **Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo.** *Cell* 2014, **159:**1140-1152.
68. Schwalie PC, Ward MC, Cain CE, Faure AJ, Gilad Y, Odom DT, Flicek P: **Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes.** *Genome Biology* 2013, **14:**R148.
69. Jansson J-O, Edén S, Isaksson O: **Sexual Dimorphism in the Control of Growth Hormone Secretion*.** *Endocrine Reviews* 1985, **6:**128-150.
70. Wauthier V, Sugathan A, Meyer RD, Dombkowski AA, Waxman DJ: **Intrinsic sex differences in the early growth hormone responsiveness of sex-specific genes in mouse liver.** *Molecular Endocrinology* 2010, **24:**667-678.
71. Waxman DJ, Holloway MG: **Sex differences in the expression of hepatic drug metabolizing enzymes.** *Mol Pharmacol* 2009, **76:**215-228.
72. Lau-Corona D, Suvorov A, Waxman DJ: **Feminization of male mouse liver by persistent growth hormone stimulation: Activation of sex-biased transcriptional networks and dynamic changes in chromatin states.** *Molecular and Cellular Biology* 2017.
73. Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, Love MI, Aiden EL, Bassik MC, Snyder MP: **Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development.** *Molecular Cell* 2017, **67:**1037-1048. e1036.
74. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B: **A high-resolution map of the three-dimensional chromatin interactome in human cells.** *Nature* 2013, **503:**290-294.
75. Xu Y, Guo W, Li P, Zhang Y, Zhao M, Fan Z, Zhao Z, Yan J: **Long-range chromosome interactions mediated by cohesin shape circadian gene expression.** *PLoS Genetics* 2016, **12:**e1005992.
76. Yeung J, Mermet J, Jouffe C, Marquis J, Charpagne A, Gachon F, Naef F: **Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs.** *Genome Research* 2018, **28:**182-191.
77. Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, Collins VP, Fraser P: **Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours.** *Genome Biology* 2017, **18:**125.
78. Marques AC, Hughes J, Graham B, Kowalczyk MS, Higgs DR, Ponting CP: **Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs.** *Genome Biology* 2013, **14:**R131.
79. Gil N, Ulitsky I: **Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity.** *Cell Systems* 2018, **7:**537-547.e533.

80. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P: **ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.** *Genome Research* 2012, **22:**1813-1831.
81. Shao Z, Zhang Y, Yuan G-C, Orkin SH, Waxman DJ: **MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets.** *Genome Biology* 2012, **13:**R16.
82. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E: **HiC-Pro: an optimized and flexible pipeline for Hi-C data processing.** *Genome Biology* 2015, **16:**259.
83. Melia T, Hao P, Yilmaz F, Waxman DJ: **Hepatic Long Intergenic Noncoding RNAs: High Promoter Conservation and Dynamic, Sex-Dependent Transcriptional Regulation by Growth Hormone.** *Molecular and Cellular Biology* 2016, **36:**50-69.
84. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS: **TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells.** *Cancer Research* 2017, **77:**e108-e110.

FIGURE LEGENDS

Figure 1: Analysis of male and female CTCF/cohesin ChIP-seq

A. Distribution of M/F ratios for in-peak differential sites for cohesin (*left*) and CTCF (*right*). Only the diffreps sites overlapping a MACS2 peak from any individual or merged are shown (see Methods). The y-axis shows the number of sites per bin, while the x-axis shows the log2(Male/Female) fold change (FC). Gray bars represent sites filtered out because they fell below the 10 read minimum count threshold, while black bars represent sites with a |FC| < 2 (or between -1 and 1 on the graph). Pink and blue bars represent female- and male-biased sites (respectively) above these thresholds.
B. There is limited overlap of sex differential cohesin and CTCF peaks. Differential cohesin peaks are shown in blue while differential CTCF sites are shown in blue. There are 136 differential cohesin peaks that overlap differential CTCF peaks. There are 137 differential CTCF peaks that overlap a differential cohesin peak. This overlap is shown for all sex-biased peaks (1,011 and 975 differential peaks for cohesin and CTCF, respectively); this includes both male- and female-biased peaks on autosomal and sex chromosomes. As seen in Figure S1A, this limited overlap is reproducible for unfiltered DiffReps as well.
C. Heatmaps and aggregate plots for sex-biased peaks. The group with significant sex bias is highlighted in red per subpanel. Clockwise from the top left, the groups with significant sex bias are: male-biased cohesin, female-biased cohesin, female-biased CTCF, and male-biased CTCF peaks. For the heatmap, read-in-peak normalized ChIP signal is shown for male and female cohesin (in blue) followed by male and female CTCF (in purple) within a 5 kb window around the differential peak summit. The aggregate profiles (*top*) represent the average signal of the heatmap below for the same 5 kb window.

Figure 2: Comparison of CTCF/cohesin ChIP-seq binding strength and proximity to relevant genes

A. With the exception of lone CTCF differential peaks, female-biased differential peaks tend to be stronger than male-biased differential peaks. The y-axis shows normalized ChIP signal for the indicated groups on the x-axis. Peaks with sex differential cohesin signal (*Top*) and CTCF signal (*Bottom*) are shown separately. Each pair of boxplots represents the male and female ChIP signal for the same set of peaks, defined by their sex bias and peak type (CAC or CNC for ΔCohesin peaks and CAC or Lone CTCF for ΔCTCF peaks). Both are indicated below the x-axis. Peak scores are calculated by average intra-peak ChIP signal, normalized by the total reads per million in peak (RIPM; see Methods).
B. In most cases, female-biased CTCF/cohesin peaks tend to be significantly closer to TSS than male-biased CTCF/cohesin peaks of the same class (CAC-ΔCohesin, CAC-ΔCTCF, and "Lone CTCF"). There is no significant difference in distance to TSS for male- versus female-biased CNC peaks (p=0.1458; KS t-

test). Cumulative frequency curves show the percent of each group on the y-axis within the distance to the nearest TSS indicated on the x-axis (in kilobases). TSS for protein coding (RefSeq) and liver lncRNA genes were considered [60]. P values for comparisons between male- and female-biased peaks of the same class are shown in the upper left of each plot (KS t-test).

C. Male- and female-biased CNC peaks are closer to enhancers than other classes of CTCF and cohesin peaks, and female-biased CNC peaks are significantly closer the enhancer DHS (eDHS) than male-biased CNC peaks (p=0.0351; KS t-test). Male-biased CAC-ΔCohesin peaks tend to be closer to enhancers than female-biased CAC-ΔCohesin peaks (p=0.002; KS), however the reverse is true for CAC-ΔCTCF peaks (p=0.0052; KS). There is no significant difference in distance to nearest enhancer between male- and female-biased Lone CTCF peaks (p=0.1068; KS t-test). Cumulative frequency curves show the percent of each group on the y-axis within the distance to the nearest eDHS indicated on the x-axis (in kilobases). Enhancers were defined previously based on a high ratio of the enhancer histone mark H3K4me1 over the promoter mark H3K4me3 at DHS [32]. P values for comparisons between male- and female-biased peaks of the same class are shown on the right of each plot (KS t-test).

D. Regardless of CTCF/cohesin peak class, most sex-biased sites are not proximal to sex-biased genes. Male-biased CNC peaks are the most proximal to sex-biased genes, while female-biased CAC-ΔCohesin are the least proximal. Peaks within each group were considered "proximal" if they were within 20 kb of a sex-biased gene, "In intra-TAD" if they fell within the same intra-TAD loop as a sex-biased gene, or "In TAD" if they overlapped the same TAD as a sex-biased gene. Each of these groups are mutually exclusive. Coordinates for Intra-TAD loops [32] and TADs [34] are from publicly available datasets.

E. Male- and female-biased CNC peaks are closer to sex-biased enhancers than any other class of sex-biased CTCF/cohesin ($p < 0.001$ for all; KS). Female-biased CNC peaks are also closer to sex-biased DHS or K27ac peaks than male-biased CNC peaks, although this difference is more modest in magnitude (p=0.0094; KS). Cumulative frequency curves show the percent of each group on the y-axis within the distance to the nearest sex-biased genomic feature indicated on the x-axis (in kilobases). The nearest sex-biased genomic features were a merged list of previously published sex-biased DHS [12], and unpublished sex-biased K27ac ChIP-seq peaks (Lau-Corona *et al.*). For visualization purposes in this Figure, CAC peaks with sex biased CTCF (CAC-ΔCTCF) and cohesin (CAC-ΔCohesin) binding are combined and present as a single group (CAC-ΔBoth). There are no significant differences between male- or female-biased CAC-ΔCTCF and/or CAC-ΔCohesin peaks in distance to nearest sex-biased enhancer (data not shown).

Figure 3: Examples of proximal and distal regulation of sex-biased gene expression

A-B. UCSC browser screenshots showing proximal sex-biased CTCF and/or cohesin binding events relevant to sex-biased gene expression in mouse liver. As in Figure 2E, "proximal" is defined as ≤20 kb from a sex-biased gene. From top to bottom the tracks are: TAD position, male and female H3K27ac ChIP-seq, male and female DNase-seq, male and female cohesin (Rad21) ChIP-seq, male and female CTCF ChIP-seq, female CTCF ChIP-seq, and Ref-seq genes. Each male (in blue) and female (in pink) pair of ChIP- or DNase-seq tracks are normalized to reflect the total reads per million in the union of peaks for that factor (see Methods). Below the signal track the regions with male- or female-bias are indicated with blue or pink, with a darker shade indicating a stricter cutoff for sex-bias and a lighter shade indicating a more lenient definition (differs by factor, see Methods). Arrows are used to further highlight the regions in question, with green arrows indicating CAC and red arrows indicating CNC or "Lone CTCF" sex-differential peaks.

A. Two male biased genes, Cml5 and Nat8, with proximal male-biased CTCF/cohesin peaks overlapping male-biased DHS (indicated by arrows). Cml5 has a male-biased CNC peak overlapping a strongly male-biased DHS approximately 3kb from the TSS. The promoter of Cml5 has strong male-biased H3K27ac accumulation and weaker male-biased DHS. Nat8 on the other hand only has a weakly male-biased DHS at

its promoter, but a stronger male-biased DHS ~12 kb away overlaps a male-biased CAC. In this case only the CTCF is differential across all replicates, however the cohesin signal is also differential using a more lenient comparison of single merged male versus merged female cohesin replicates (see Methods). This screenshot is for the region spanning chr6:85766132-85794443. The previously published ratio of male over female gene expression is indicated above the screenshot: 26.5- and 5.8-fold male bias for Cml5 and Nat8, respectively [83].

B. Fam84b is an example of a female-biased gene with a proximal female-biased CNC peak overlapping the TSS and female-biased H3K27ac accumulation (indicated by the red arrow). There is also a sex-independent CAC overlapping the second exon of Fam84b that does not overlap the female-biased H3K27ac peak at the promoter. This screenshot is for the region chr15:60647974-60662508. The previously published ratio of female over male gene expression (3.3-fold) is indicated above the screenshot [83].

C-D. UCSC browser screenshots showing distal sex-biased CTCF and/or cohesin binding events relevant to sex-biased gene expression in mouse liver. As in Figure 2E, "distal" is defined as >20 kb from a sex-biased gene, but within the same TAD. The tracks, annotations, and normalization are identical to Figures 3A and 3B.

C. Two male biased genes, C8a and C8b, with distal male-biased CAC and CNC peaks overlapping male-biased DHS and H3K27ac mark accumulation (indicated by the green and red arrows, respectively). While the linear distance between the upstream male-biased CAC peaks and downstream male-biased genes is more than 1.5 Mb, they reside on opposite ends of the same TAD (see model in Figure S3A). Despite being close in linear distance to Oma1, this gene shows no sex differences in gene expression and based on TAD structure would <u>not</u> be predicted to interact with the highlighted male-biased enhancers. A screenshot covering the full length of the TAD and a potential model of their spatial position is shown in Figure S3A. The left portion of the panel is spans the region from chr4:103027454-103167067, while the right portion spans the region from chr4:104433514-104583344. The previously published ratio of male over female gene expression is indicated above the screenshot: 3.3- and 3.0-fold male bias for C8a and C8b, respectively [83].

D. Slc22a29 is an example of a female-biased gene with a distal female-biased CAC peak overlapping a robust female-biased DHS with female-biased H3K27ac histone mark accumulation (indicated by the green arrow). With a distance of ~34 kb, Slc22a29 is the closest gene to this female-biased, non-anchor CAC. The screenshot spans the region from chr19:8290981-8333632. The previously published ratio of female over male gene expression for Slc22a29 (9.9) is indicated above the screenshot [83].

E. Upon loss of cohesin binding, expression of C8a and C8b is significantly reduced ($p \leq 0.0001$ and $p=0.0020$; MW t-test), while expression of Oma1 is increased ($p=0.0102$; MW t-test). This fits with general model that loss of cohesin results in proximal over distal regulatory interactions, including enhancer-hijacking by linearly proximal genes. Bars represent the mean expression of the group for a given gene relative to the mean of the WT group (equal to 1), with error bars showing the standard deviation for n=4 per group.

F. Loss of cohesin binding has a more drastic impact on male-biased genes dependent on distal enhancers than those with proximal sex bias. Expressed, strongly male-biased genes (FPKM > 1 and M/F > 3) were divided into two groups based on distance from their TSS to the nearest male-biased DHS or K27ac-marked region. DHS-proximal male-biased genes (n=48) were defined as having TSS < 20 kb from the nearest male-biased DHS, while DHS-distal male-biased genes (n=21) were those with TSS > 20 kb from such regions. Plotted is the average expression for cohesin-depleted liver divided by the average expression in WT mice, such that a value of 0.1 represents a 10-fold reduction in expression after loss of cohesin binding. The average relative expression for DHS-proximal genes is 1.05 (representing no change between WT and cohesin depleted), while the average for DHS-distal is 0.5, representing an average reduction of 2-fold reduction in gene expression upon cohesin depletion ($p = 0.0087$; MW).

Figure 4: 4C-seq results for two female and two male-biased genes

A. A distal enhancer near with A1bg and select female-biased lncRNAs. The screenshotted region contains 12 female-biased lncRNAs (and 0 male-biased lncRNAs), all of which are monoexonic. Each labelled cluster of lncRNAs are on the same strand and only the TSS of the most upstream (with respect to direction of transcription) is shown. This cluster of lncRNAs is more consistently female-biased across multiple RNA-seq variants and two mice strains (Figure S5C,D). The linearly proximal lncRNAs 12600-12603 show the weakest interactions and least consistency in female bias. A full listing of relevant expression data for these genes is presented in Figure S5C,D. Panels 4A,B,C,E show the median reads per million normalized 4C-seq signal per sex (*Top* track). The tracks below are identical to Figure 3, showing normalized DNase- or ChIP-seq signal for the indicated factor. If present, intra-TAD loops or sex-biased lncRNAs are shown below the Refseq gene track in blue. This screenshot spans the region from chr15:60733512-60954051.
B. An enhancer 8 kb upstream from Gm4794 interacts primarily with the proximal promoter (*left* red arrow) and a strong female biased enhancer (*right* red arrow). Both the viewpoint enhancer and the interacting enhancer contain female-biased CNC peaks (indicated below cohesin track in dark pink). Sult3a1 only shows a weak and broad pattern of interaction with this viewpoint enhancer (red bracket). This screenshot spans the region from chr10:33418446-33680888.
C. The promoter of C9 interacts with a distal male biased enhancer, bypassing the gene Dab2. This viewpoint is anchored at the promoter of the male-biased gene C9, and predominantly shows a strong male biased interaction with the distal male-biased enhancer indicated by the red arrow. The nested intra-TAD loop structure may facilitate the high and male-biased expression of C9 by "looping out" the intervening gene Dab2. There is also 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). This screenshot spans the region from chr15:6147917-6461799.
D. Upon loss of chromatin-bound cohesin, expression of both C9 and its antisense lncRNA (12340) are significantly reduced (66% reduction, p=0.0056 and 82% reduction, p=0.0049; respectively). This is likely a reflection of the dependence of these genes on distal regulatory elements for normal expression. Dab2 expression is increased however this difference is not significant (p=0.219). Bars represent the mean expression of the group for a given gene relative to the mean of the WT group (equal to 1), with error bars showing the 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 indicated by the red box. The two more prominent interactions between this viewpoint is with a neighboring male-biased enhancer and the promoter of Nudt7. While TSS of lncRNA 7430 and Nudt7 are very close, the viewpoint enhancer appears to interact more directly with the TSS of Nudt7 specifically. This screenshot spans the region from chr8:116592444-116707613.
F. Upon loss of chromatin-bound cohesin, expression of both Nudt7 and its bidirectionally transcribed lncRNA (8430) are significantly reduced (69% reduction, p=0.0116 and 93% reduction, p=0.0113; respectively). Expression of lncRNA 7423 is not significantly impacted (p=0.166). Bars represent the mean expression of the group for a given gene relative to the mean of the WT group (equal to 1), with error bars showing the standard deviation for n=4 per group.

Figure 5: Nox4 results, combining luciferase assay with 4C for multiple viewpoints

A. No observable interactions occur between two male-biased enhancers neighboring the male-biased gene Nox4. Two viewpoints are shown in this Figure: "VP Enh1" is anchored at the red highlighted region ~11 kb upstream from the Nox4 TSS, while "VP Enh2" is anchored at the green highlighted region ~120 kb downstream from the TSS of Nox4. Enh1 is a region with the strongest male-biased H3K27ac mark This screenshot spans the region from chr7:94248242-94726358. In addition to the ChIP-seq (H3K27ac, Cohesin, and CTCF) and DNase-seq, two tracks indicating the chromatin state in male (top) and female liver

(bottom) are shown above. Green indicates enhancer-like, blue promoter-like, and purple transcribed-like (Figure S6C). While no focal interactions between "Enh1" and "Enh2" these two viewpoints support the model of two nested and insulated intra-TAD loops as shown below the screenshot. All tracks were normalized and as described in Figure 4.

B. Expression of Nox4, but not the neighboring gene Tyr, is cohesin-dependent. While the results in Figure 5A indicate Nox4 is primarily regulated by the proximal enhancers within the shorter intra-TAD loop, proper expression of Nox4 is still dependent on the presence of cohesin. This may be due to the necessity of the intra-TAD loop structure; however, loss of this insulation does not appear to increase expression of Tyr as might have been expected (i.e. "enhancer hijacking"). Expression of Nox4 is reduced 62% ($p<0.0001$). Bars represent the mean expression of the group for a given gene relative to the mean of the WT group (equal to 1), with error bars showing the standard deviation for n=4 per group.

C. An overall model of the direct and indirect contribution of CTCF/cohesin contributing to sex-biased gene expression in mouse liver. A1bg is a good example of direct regulation, where female-biased CTCF binding could explain the observed female-bias in looping interactions. C9 is an example of intra-TAD loops present in male and female liver contribute indirectly to sex-biased gene expression. Not only does the intra-TAD loop insulated the lowly-expressed sex-independent gene Dab2, it also brings the distal male-biased enhancer into close proximity. Likely some additional factors like mediator, YY1, or eRNAs directly contribute to the direct interactions observed in male liver.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: Additional features of sex-biased differential CTCF/cohesin peaks

A. The limited overlap of sex-biased CTCF and sex-biased cohesin binding is not an artifact of filtering. Similar to the results presented in Figure 1B, the overlap of sex-biased CTCF and cohesin binding sites is limited. From top to bottom, overlap of sex-biased cohesin (*Left*; blue) and CTCF (*Right*; purple) for the following groups (top to bottom): (1) All sex differential sites from diffReps without filtering for overlap of a MACS2 peak, (2) All sex differential sites from diffReps that overlap a MACS2 peak for a given factor, (3) All sex differential hotspots identified by diffReps, which is an alternate method for identifying differential sites. Overlap is defined as $\geq$1 bp using Bedtools.

B. A substantially larger fraction of male-biased ΔCTCF peaks fall into the "Lone CTCF" peak class compared to female-biased ΔCTCF peaks. Otherwise, the majority of sex-biased CTCF/cohesin peaks are at CAC. Four pie charts are shown for (from top to bottom): male-biased ΔCohesin peaks, female-biased ΔCohesin peaks, male-biased ΔCTCF peaks, and female-biased ΔCTCF peaks. For each of these four groups, the fraction of peaks at CAC sites is shown in purple while the fraction of peaks at either CNC (for ΔCohesin) or "Lone CTCF" (for ΔCTCF) is shown in blue. The total number of differential peaks in each group is also indicated below each chart.

Figure S2: Comparison of sex-biased CTCF/cohesin peaks

A. Related to Figure 2A, female-biased CTCF/cohesin peaks tend to be stronger than male-biased peaks. While Figure 2A only shows normalized ChIP signal for the factor with differential signal (i.e. male and female cohesin ChIP-seq signal for ΔCohesin peaks), here we show ChIP signal for CTCF and cohesin for both ΔCohesin and ΔCTCF peaks. In aggregate, CAC peaks with significant sex-biased cohesin binding show the same directionality of sex-bias for CTCF (and vice versa), albeit at a reduced magnitude (see also Figure 1C). The y-axis shows normalized ChIP signal for the indicated groups along the x-axis. Peaks with male-and

female-biased cohesin binding (*Left*) and male- and female-biased CTCF binding (*Right*) are shown separately. Each group of 4 boxplots represents the male and female ChIP signal for cohesin followed by CTCF the same set of peaks. Each plot represents all differential peaks for a given sex (male or female) and factor (CTCF or cohesin). These four sets are further subdivided by peak type (CAC or CNC for ΔCohesin peaks and CAC or Lone CTCF for ΔCTCF peaks), indicated below the x-axis. Peak scores are calculated by average intra-peak ChIP signal, normalized by the total reads per million in peak (RIPM; see Methods).

B. In addition to being stronger than male-biased peaks, a larger fraction of female-biased CAC peaks contain CTCF binding motifs. Despite no significant difference in peak strength between male- and female- biased "Lone CTCF" peaks, a larger fraction of male-biased Lone CTCF peaks contain CTCF motifs. A larger fraction of female-biased CNC peaks contain CTCF motifs, however the vast majority do not contain CTCF motifs, as expected (< 20% for all groups). In all cases the percent for each group is comparable to a matched set of sex-independent peaks. The y-axis shows the percent of peaks in each group (separately for male-biased, female-biased, and sex-independent) considered to have a CTCF motif present within the peak. Presence of motif is defined as the presence of the canonical core CTCF binding motif (MA0139.1) as identified by FIMO with a minimum score > 10. In addition to this binary consideration of motif presence/absence, the related plot in Figure S2B shows the motif strength for each group.

C. Female-biased CAC peaks contain higher quality CTCF motifs than male-biased CAC peaks [p=0.0212 for CAC(ΔCoh) and p=0.0023 for CAC(ΔCTCF) peaks; MW t-test]. In addition to a larger proportion of female-biased CAC peaks containing motifs (Figure 2B), the overall quality of the motif matches present is higher as reflected by FIMO motif score. This log-likelihood ratio score is reflection of how closely the best intra-peak motif matches the provided canonical CTCF motif MA0139.1. There is no significant difference between motif scores for male- and female-biased "Lone CTCF" or male- and female-biased CNC peaks (p=0.7671 and p=0.1329; MW t-test). The dashed line at a FIMO score of 10 reflects the cutoff used to define the presence or absence of a motif in Figure 2B.

D. Male and female intra-TAD loops and loop anchors are mostly shared. Using a computational intra-TAD loop prediction algorithm, 9,543 and 9,724 loops are predicted in male and female liver, respectively [32]. 87.9% of intra-TAD loops in male are also identified in female liver (*left)*) and 93.4% of male intra-TAD loop anchors are also predicted to be loop anchors in female liver. Given the limited number of autosomal CAC peaks with sex differences in CTCF and cohesin binding (53 total), this result is not surprising.

Figure S3: Tissue conservation features of sex-differential peaks

A. Female-biased CAC peaks are less tissue specific than male-biased CAC peaks, while male-biased "Lone CTCF" peaks are less liver specific than female-biased peaks of the same class. The tissue ubiquity of CTCF binding for female-biased CAC peaks could be due to the fact that female-biased peaks are stronger and contain higher quality CTCF motifs, however male-biased "Lone CTCF" peaks are not significantly stronger, nor do they contain higher quality motifs than female-biased peaks of the same class. The apparent difference could be due to the lack of female CTCF ChIP-seq data in most mouse tissues. CTCF ChIP-seq data from non-liver, non-reproductive tissues was generated by the ENDCOE consortium from male mice [51]. For male- and female-biased CNC peaks, over 80% are not bound by CTCF in other mouse tissues. This provides additional evidence that CNCs are found at liver-specific cis regulatory elements, and that these sites do not appear to act as insulators in other non-liver tissues (i.e. CTCF binding is lost in liver or gained in some other tissue). The x-axis indicates the number of ENCODE tissues (out of 15) that also have CTCF bound, where a value of "15" indicates tissue-ubiquitous CTCF binding and a value of "0" indicates liver-specific CTCF binding. The y-axis shows the proportion of male- (blue) or female-biased (red) peaks that falls into a given bin. P values comparing the distribution of tissue-specificity values for CTCF binding is indicated in the upper left corner of each plot (KS t-test).

B. There is a strong association between peak strength and tissue conservation, which likely explains the modestly higher tissue conservation of female-biased CTCF/cohesin peaks. The reads-in-peaks normalized

ChIP-seq data is shown for all CTCF peaks, male differential CTCF peaks, and female differential peaks. These are grouped according to the number of non-liver ENCODE tissues with a CTCF peak, where 0 means a peak is liver-specific and 15 means all mouse tissues with ENCODE datasets have CTCF bound at that position.

C. The tissue specificity of neighboring genes varies significantly by the class of CTCF/cohesin binding site. Shown are the Tau values, where a value close to 1 indicates the pattern of expression across mouse ENCODE RNA-seq datasets is highly tissue specific, while lower values (< ~0.3) are considered housekeeping genes. Nearest genes were defined on distance to TSS and only liver expressed genes were considered (FPKM > 1). Tau values are calculated based on the average of two replicates from all tissues except testis generated by ENCODE and analyzed previously. Both female and male CNCs are near genes that are more tissue specific than liver expressed genes, however only genes near female-biased CNCs are significantly different than genes near similarly sex-biased CACs (p=0.007; MW). This difference is not a reflection of the male or female CAC group being compared against, as genes near female CNCs are significantly more liver specific than genes near male CACs (p=0.0171; MW) while the opposite comparison for male CNCs vs female CACs is still not significant (p=0.07; MW).

Figure S4: Additional C8a/C8b screenshot and cohesin depletion effects

A. A full screenshot potential model linking the distal male-biased enhancers and component complement genes C8a and C8b within a single TAD on chromosome 4. This screenshot spans chr4:102960671-104603975. The tracks, normalization, and annotations are identical to Figure 3.
B. For the proximally-regulated male-biased genes shown in Figure 3A (Nat8 and Cml5) loss of cohesin does not significantly impact their expression.

Figure S5: Quality control of 4C Samples

A. Agarose gel quality control for ligated, digested, and re-ligated 4C samples. Lane (i) indicates material after proximity ligation, lane (ii) shows material after Csp6i digestion, and lane (iii) shows material after self-circularization ligation. Lane (iii) represents the final material used as input for inverse PCR with viewpoint-specific primers (Table S3A). The indicated sizes (in kb) are shown on the left of the gel in red.
B. Agarose gel quality control for post inverse PCR final 4C-seq libraries. A diverse mixture of PCR products (presenting as a "smear" on the gel), primarily below ~1kb indicates a high quality library and allows for efficient sequencing. The indicated sizes (in kb) are shown on the right of the gel in red.
C. Relevant to Figure 4A, the expression of 12 monoexonic lncRNAs and A1bg across several 6 RNA-seq datasets from CD-1 mouse liver and 1 RNA-seq dataset from C57/Bl6 mouse liver. From left to right, the columns indicate the mean expression level of a given gene in female liver (FPKM) for: Total PolyA+ unstranded RNA-seq (G83; [72]), Total PolyA+ unstranded RNA-seq
D. Relevant to Figure 4A, the significance and magnitude of female biased expression for 12 monoexonic lncRNAs and A1bg across several 6 RNA-seq datasets from CD-1 mouse liver and 1 RNA-seq dataset from C57/Bl6 mouse liver.

Figure S6: Additional 4C-seq examples

Figure S5 shows the 6 replicates per viewpoint. Main figures show the merged replicate signal by sex, taking the median value for a given smoothed window of 11 restriction fragments. The gene tracks and sex-biased sites are identical to those shown in Figure 4. These bed tracks from top to bottom: sex-biased H3K27ac peaks,

sex-biased DHS, sex-biased cohesin peaks, and sex-biased CTCF peaks. Finally, protein coding genes, sex-biased lncRNA genes, and intra-TAD loops are shown (if present).

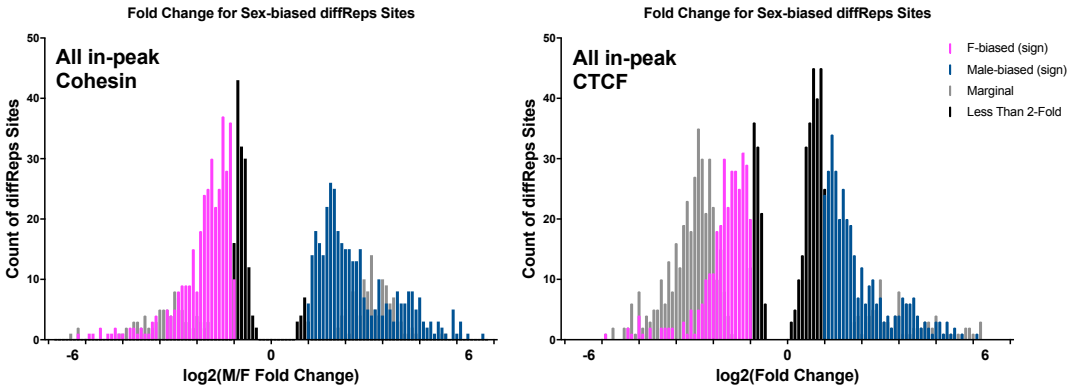
A. All replicates for the A1bg enhancer viewpoint (chr15:60733512-60954051).
B. All replicates for the Gm4794 enhancer viewpoint (chr10:33418446-33680888).
C. All replicates for the C9 promoter viewpoint (chr15:6147917-6461799).
D. All replicates for the Nudt7 enhancer viewpoint (chr8:116592444-116707613).

Figure S7: High expression of Nox4 in HCC (both human and mouse model)

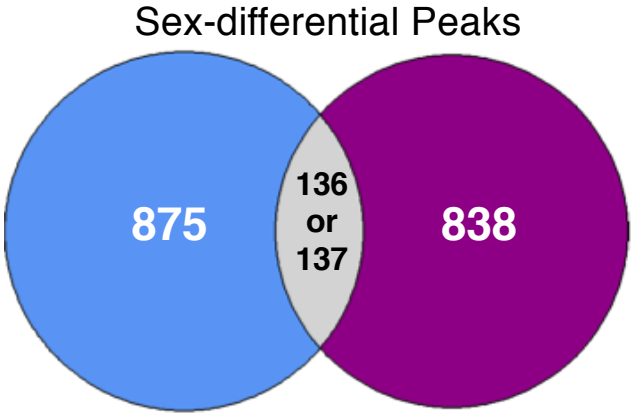
A. In human patient samples, Nox 4 is consistently upregulated in tumor tissue relative to normal controls, including for hepatocellular carcinoma (red brackets). Only data for cancer types with matched primary tumor and normal tissue controls are shown, with Nox4 showing significant upregulation in tumors for 14 of 18 matched pairs. Expression datasets are from The Cancer Genome Atlas (TCGA) and were analyzed by the Tumor IMmune Estimation Resource (TIMER) webtool with default parameters [84]. The significance of comparisons between Normal and Tumor tissue was calculated by Wilcoxon test and is indicated on the chart as: 0 ≤ *** < 0.001 ≤ ** < 0.01 ≤ * < 0.05. For example, the difference between liver tumor versus normal tissue (p = 1.35E-25) is indicated as ***.
B. In mice that spontaneously develop tumors (strain C3H; [58]), Nox4 is substantially upregulated in tumor relative to healthy tissue (p=0.0006; MW).
C. Chromatin state key for relevant to Figure 6A, based on a 14 state model of chromatin states in male and female liver [13].
D. All replicates for the Nox4 enhancer 1 viewpoint (chr7:94248242-94726358). Tracks are otherwise identical to Figure S5.
E. All replicates for the Nox4 enhancer 2 viewpoint (chr7:94248242-94726358). Tracks are otherwise identical to Figure S5.
F. Three separate models of regulatory domain prediction support the model that Enh1 and Enh2 are not likely to interact. The position of the Enh1 viewpoint is indicated by the red arrow and the position of the Enh2 viewpoint is indicated by the green arrow. From top to bottom: Enhancer Promoter Units (EPUs) are domains based on the pan-tissue correlation of ChIP-seq signal at enhancer and promoters [51]; intra-TAD loops are computationally predicted based on CTCF and cohesin ChIP-seq data from mouse liver [32]; cohesin ChIA-PET loops show cohesin-anchored interactions from mouse embryonic stem cells [26].

Fig. 1

A.



B.



Cohesin (Rad21):

1,011 (841 autosomal)

- 475 (395) M-biased
- 536 (441) F-biased

CTCF:

975 (750 autosomal)

- 395 (328) M-biased
- 580 (421) F-biased

C.

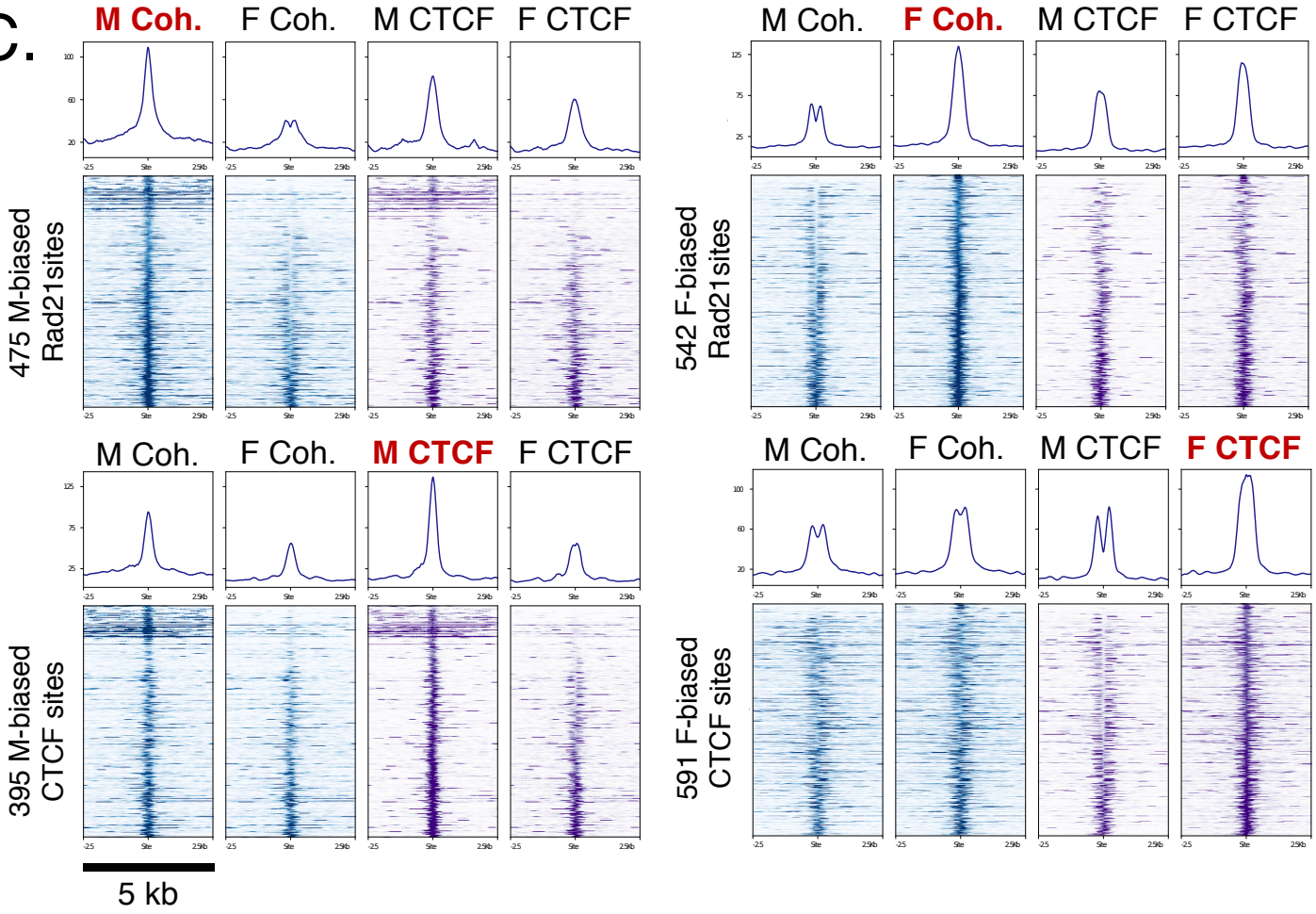
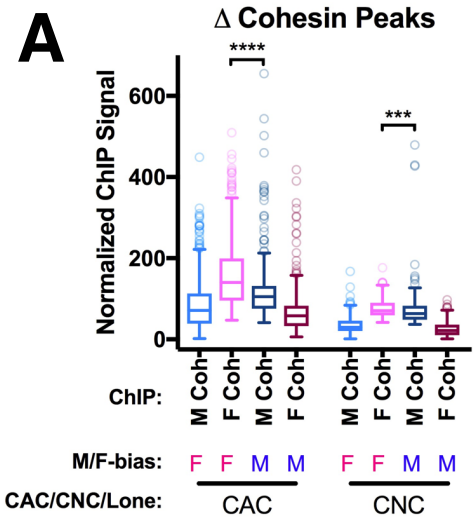
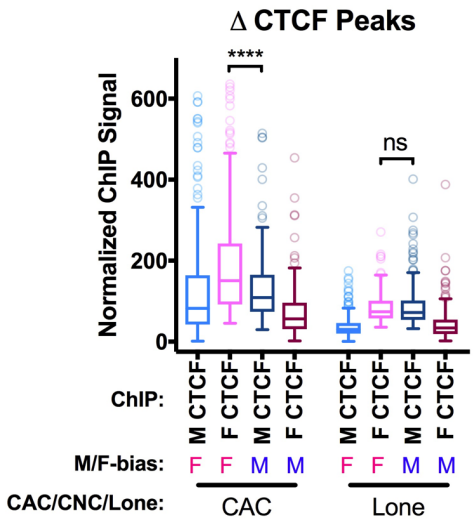
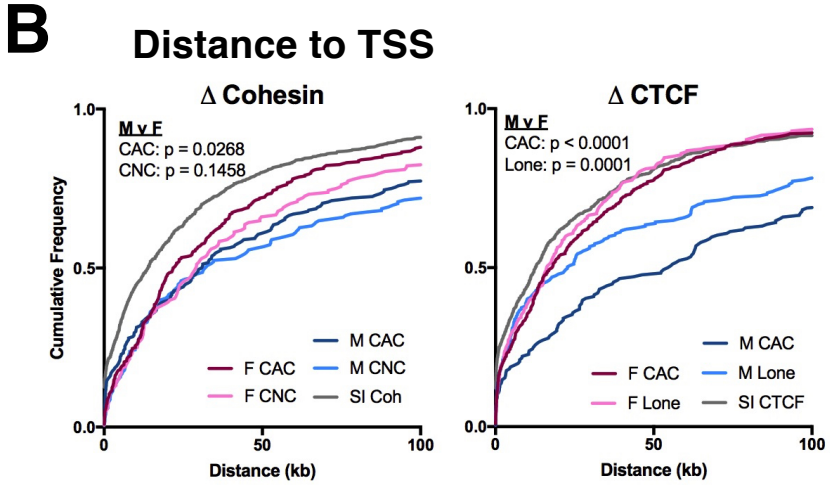
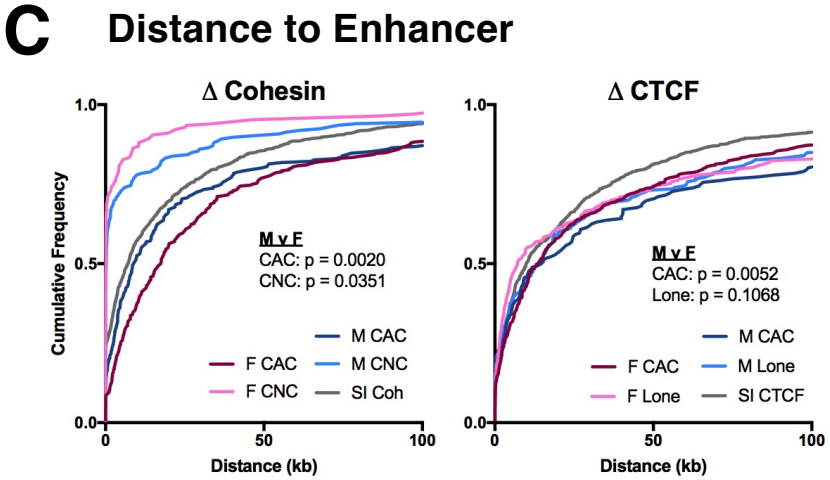


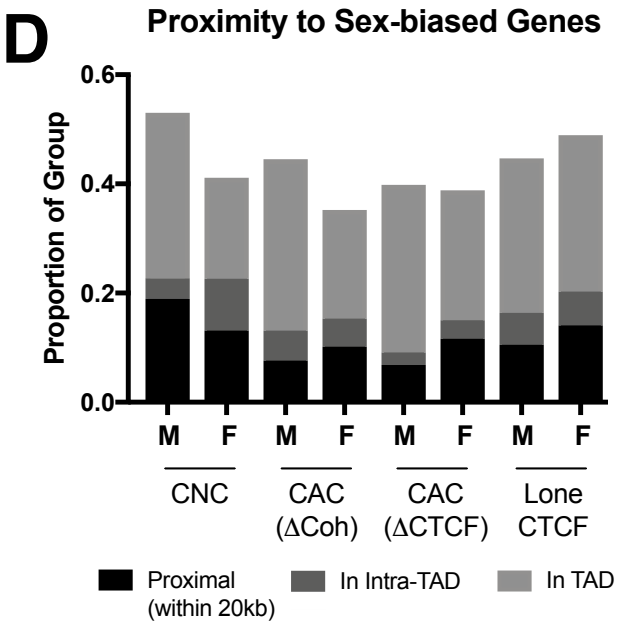
Fig. 2


A
Δ Cohesin Peaks
Normalized ChIP Signal
****
***
ChIP:
M Coh
F Coh
M Coh
F Coh
M Coh
F Coh
M Coh
F Coh
M/F-bias:
F F M M F F M M
CAC/CNC/Lone:
CAC
CNC


Δ CTCF Peaks
Normalized ChIP Signal
****
ns
ChIP:
M CTCF
F CTCF
M CTCF
F CTCF
M CTCF
F CTCF
M CTCF
F CTCF
M/F-bias:
F F M M F F M M
CAC/CNC/Lone:
CAC
Lone


B
Distance to TSS
Δ Cohesin
M v F
CAC: p = 0.0268
CNC: p = 0.1458
Cumulative Frequency
Distance (kb)
M CAC
F CAC
M CNC
F CNC
Sl Coh
Δ CTCF
M v F
CAC: p < 0.0001
Lone: p = 0.0001
M CAC
F CAC
M Lone
F Lone
Sl CTCF


C
Distance to Enhancer
Δ Cohesin
M v F
CAC: p = 0.0020
CNC: p = 0.0351
Cumulative Frequency
Distance (kb)
M CAC
F CAC
M CNC
F CNC
Sl Coh
Δ CTCF
M v F
CAC: p = 0.0052
Lone: p = 0.1068
M CAC
F CAC
M Lone
F Lone
Sl CTCF


D
Proximity to Sex-biased Genes
Proportion of Group
M F M F M F M F
CNC
CAC (ΔCoh)
CAC (ΔCTCF)
Lone CTCF
Proximal (within 20kb)
In Intra-TAD
In TAD

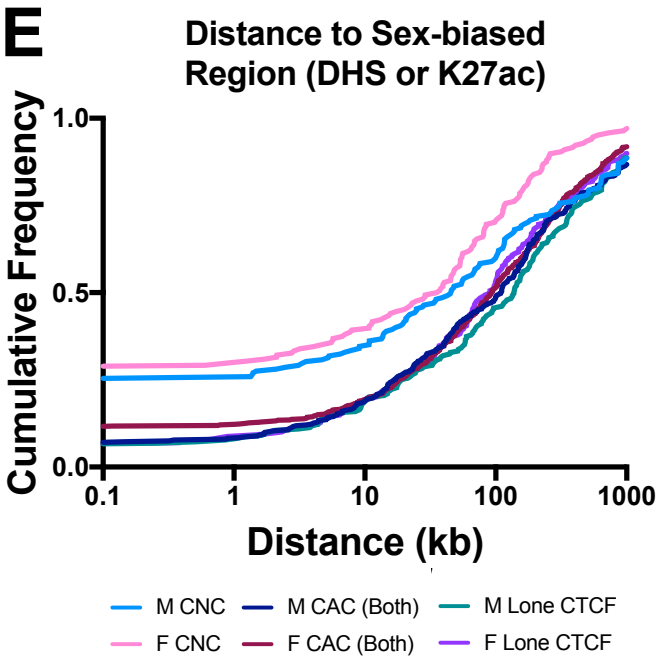

E
Distance to Sex-biased Region (DHS or K27ac)
Cumulative Frequency
Distance (kb)
M CNC
F CNC
M CAC (Both)
F CAC (Both)
M Lone CTCF
F Lone CTCF

Fig. 3

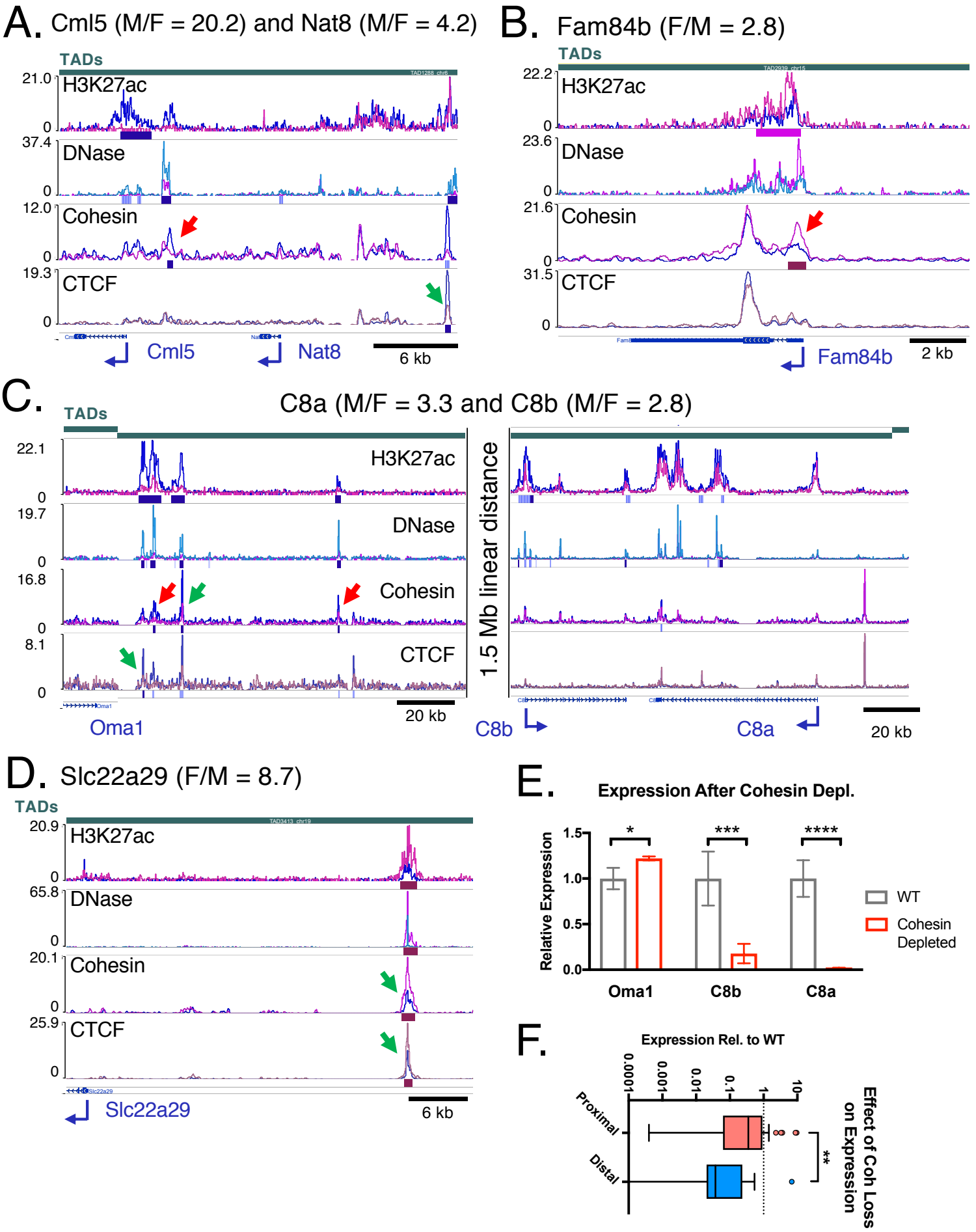

A. Cml5 (M/F = 20.2) and Nat8 (M/F = 4.2)
TADs
H3K27ac
DNase
Cohesin
CTCF
Cml5
Nat8
6 kb
B. Fam84b (F/M = 2.8)
TADs
H3K27ac
DNase
Cohesin
CTCF
Fam84b
2 kb
C. C8a (M/F = 3.3 and C8b (M/F = 2.8)
TADs
H3K27ac
DNase
Cohesin
CTCF
1.5 Mb linear distance
Oma1
20 kb
C8b
C8a
20 kb
D. Slc22a29 (F/M = 8.7)
TADs
H3K27ac
DNase
Cohesin
CTCF
Slc22a29
6 kb
E. Expression After Cohesin Depl.
Relative Expression
WT
Cohesin Depleted
Oma1
C8b
C8a
F. Expression Rel. to WT
Proximal
Distal
Effect of Coh Loss on Expression

Fig. 4

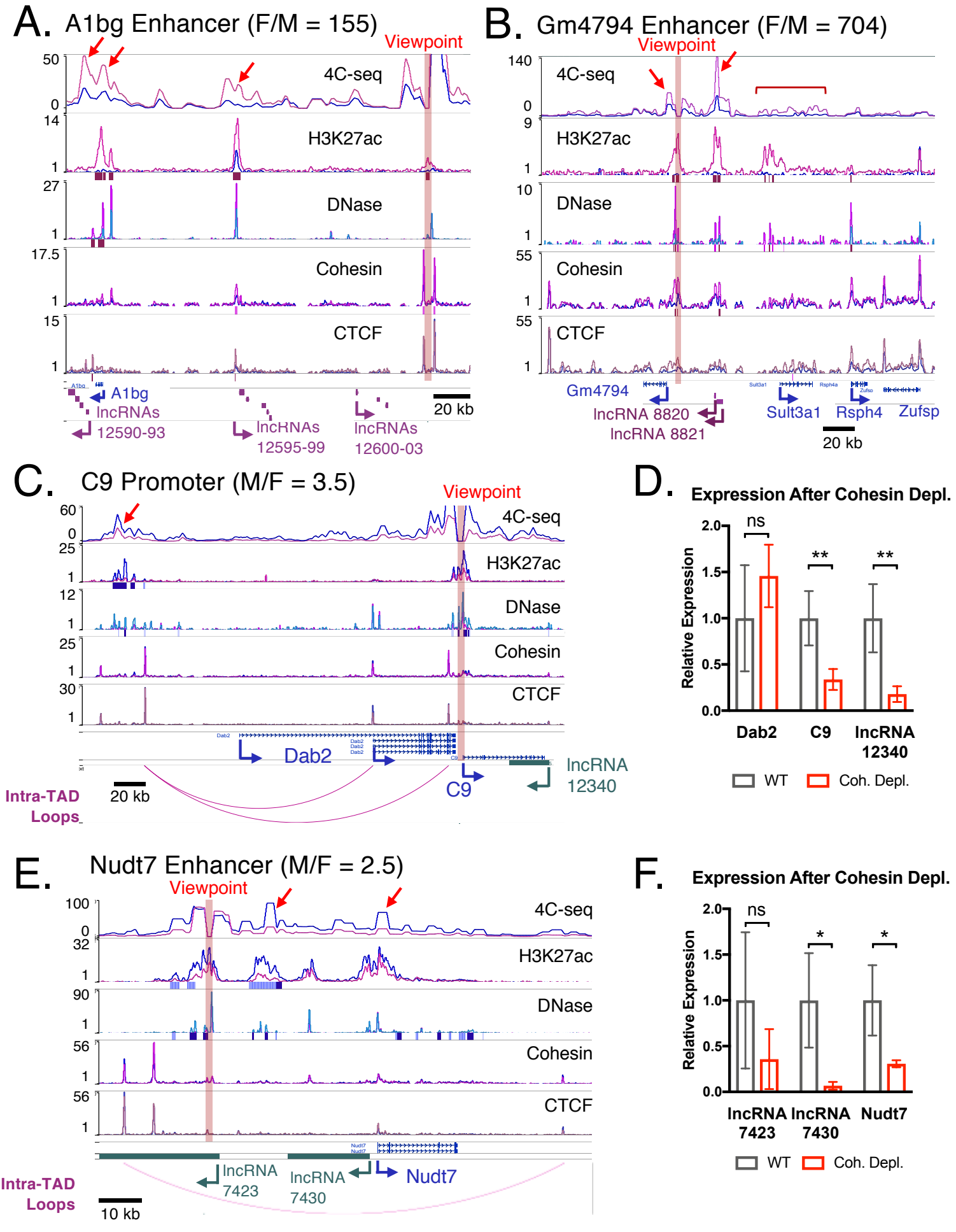
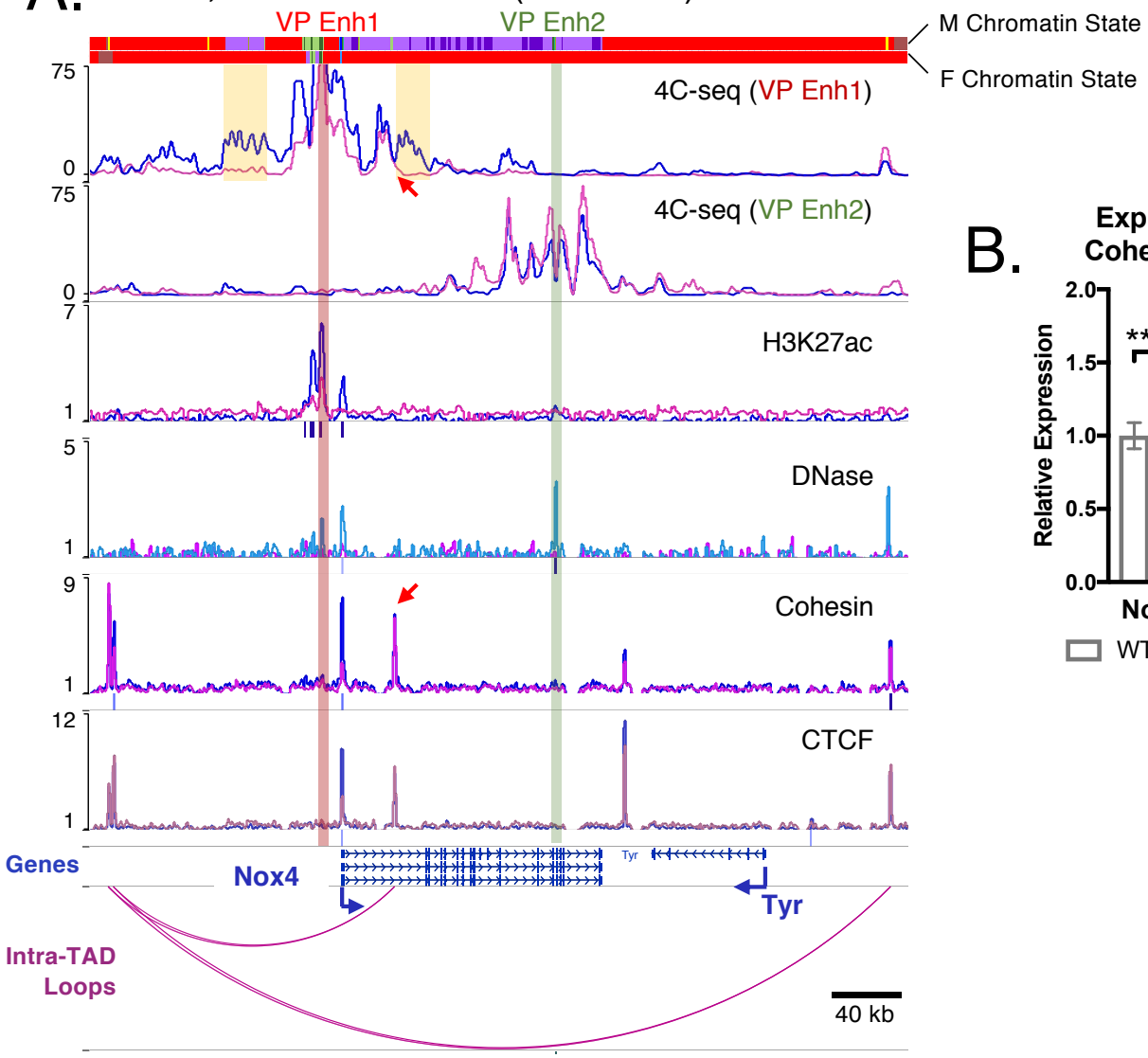
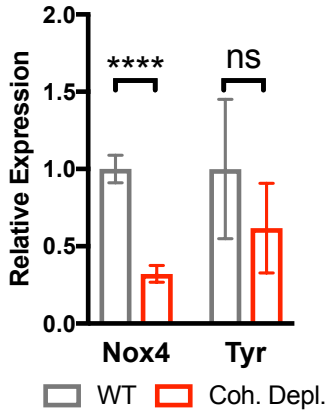

A. A1bg Enhancer (F/M = 155)
Viewpoint
4C-seq
H3K27ac
DNase
Cohesin
CTCF
A1bg
lncRNAs 12590-93
lncRNAs 12595-99
lncRNAs 12600-03
20 kb
B. Gm4794 Enhancer (F/M = 704)
Viewpoint
4C-seq
H3K27ac
DNase
Cohesin
CTCF
Gm4794
lncRNA 8820
lncRNA 8821
Sult3a1
Rsph4
Zufsp
20 kb
C. C9 Promoter (M/F = 3.5)
Viewpoint
4C-seq
H3K27ac
DNase
Cohesin
CTCF
Dab2
C9
lncRNA 12340
Intra-TAD Loops
20 kb
D. Expression After Cohesin Depl.
Relative Expression
ns
**
**
Dab2
C9
lncRNA 12340
WT
Coh. Depl.
E. Nudt7 Enhancer (M/F = 2.5)
Viewpoint
4C-seq
H3K27ac
DNase
Cohesin
CTCF
lncRNA 7423
lncRNA 7430
Nudt7
Intra-TAD Loops
10 kb
F. Expression After Cohesin Depl.
Relative Expression
ns
*
*
lncRNA 7423
lncRNA 7430
Nudt7
WT
Coh. Depl.

Fig.5

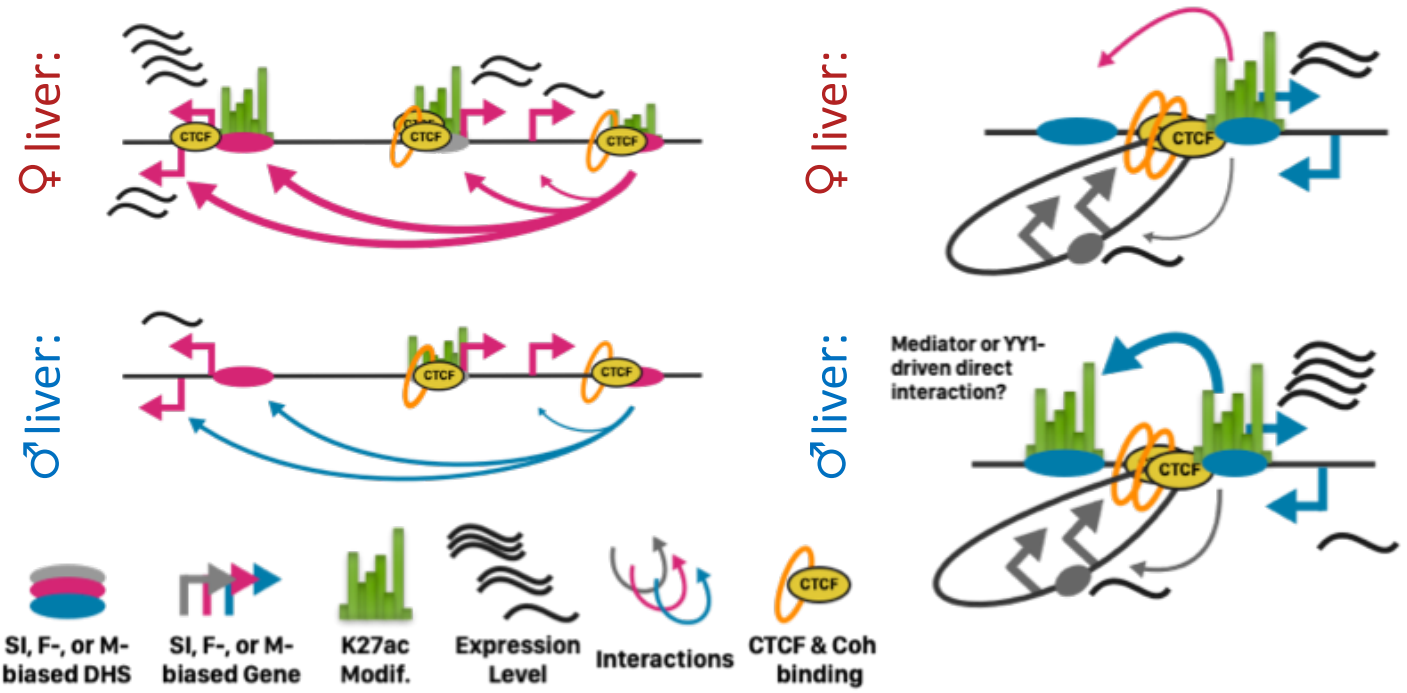
A. Nox4, 2 Enhancer VPs (M/F = 7.7)



B.



C.



Supplemental Figures

Figure S1: Additional features of sex-biased differential CTCF/cohesin peaks

Figure S2: Comparison of sex-biased CTCF/cohesin peaks

Figure S3: Tissue conservation features of sex-differential peaks

Figure S4: Additional C8a/C8b screenshot and cohesin depletion effects

Figure S5: Quality control of 4C-seq Samples and A1bg lncRNA analysis

Figure S6: Additional 4C-seq screenshots with replicates

Figure S7: High expression of Nox4 in HCC (both human and mouse model) and other Nox4 screenshots

Fig. S1

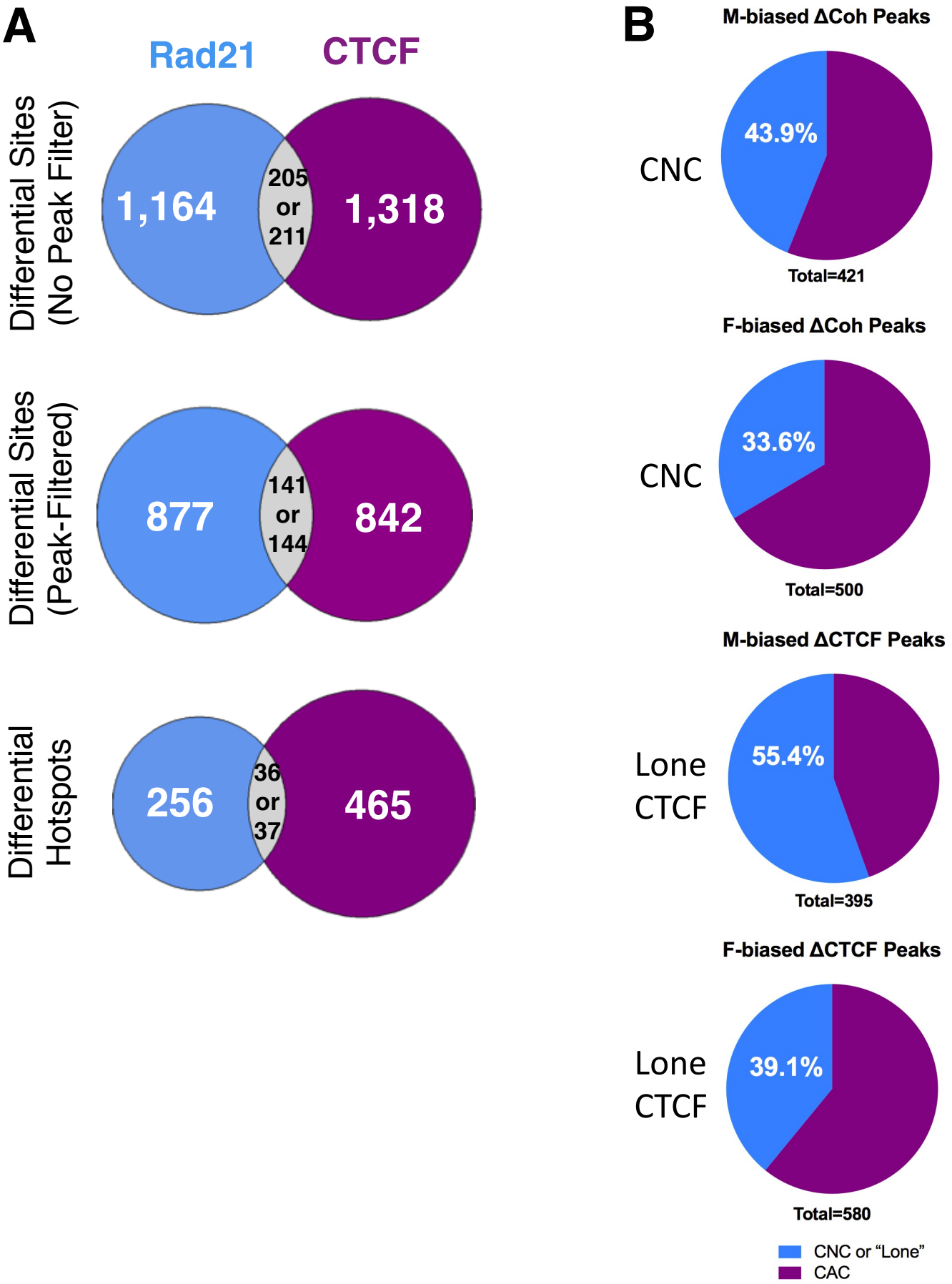
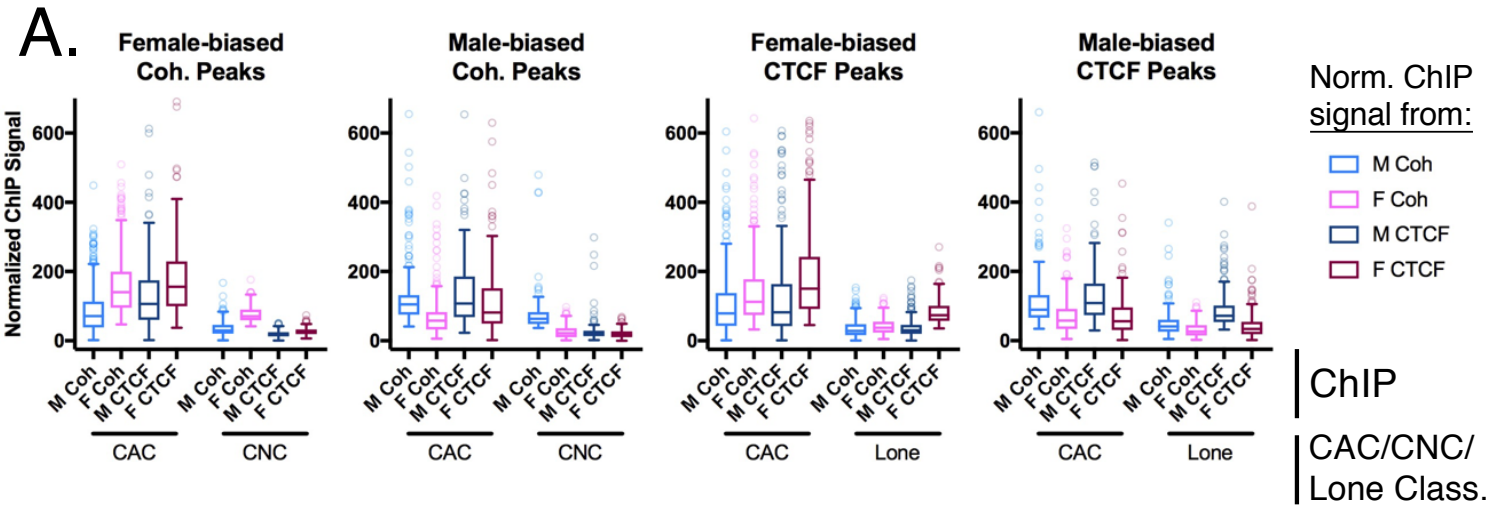
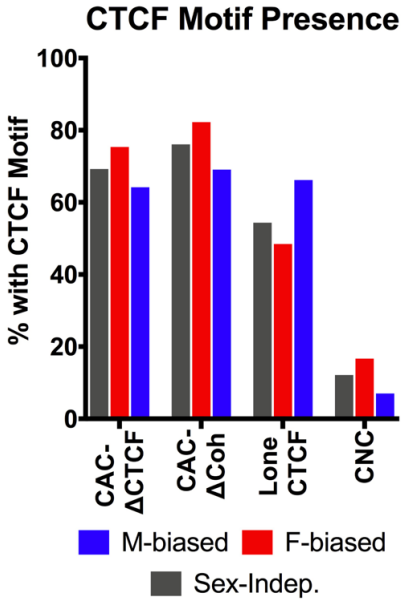
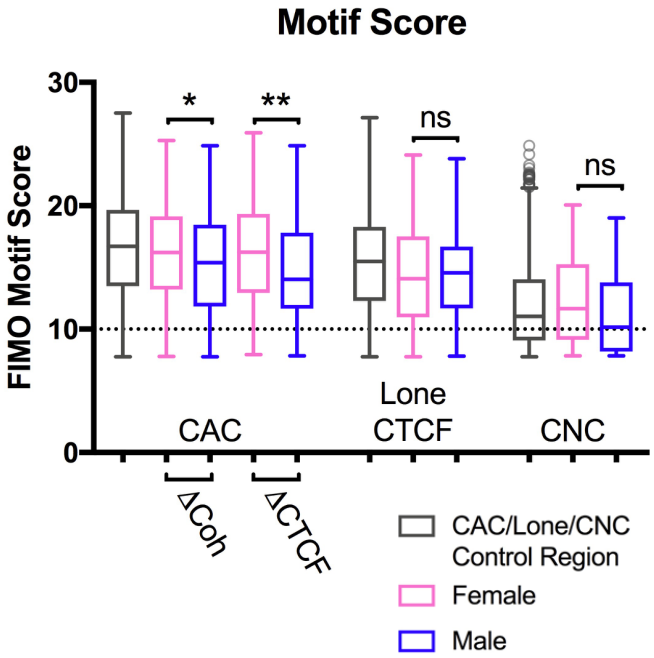

A
Rad21
CTCF
Differential Sites (No Peak Filter)
1,164
205 or 211
1,318
Differential Sites (Peak-Filtered)
877
141 or 144
842
Differential Hotspots
256
36 or 37
465
B
M-biased ΔCoh Peaks
CNC
43.9%
Total=421
F-biased ΔCoh Peaks
CNC
33.6%
Total=500
M-biased ΔCTCF Peaks
Lone CTCF
55.4%
Total=395
F-biased ΔCTCF Peaks
Lone CTCF
39.1%
Total=580
CNC or "Lone"
CAC

Fig. S2


A.
Female-biased Coh. Peaks
Male-biased Coh. Peaks
Female-biased CTCF Peaks
Male-biased CTCF Peaks
Normalized ChIP Signal
0
200
400
600
M Coh
F Coh
M CTCF
F CTCF
CAC
CNC
Lone
Norm. ChIP signal from:
M Coh
F Coh
M CTCF
F CTCF
ChIP
CAC/CNC/ Lone Class.


B.
CTCF Motif Presence
% with CTCF Motif
0
20
40
60
80
100
CAC-ΔCTCF
CAC-ΔCoh
Lone CTCF
CNC
M-biased
F-biased
Sex-Indep.


C.
Motif Score
FIMO Motif Score
0
10
20
30
*
**
ns
ns
CAC
Lone CTCF
CNC
ΔCoh
ΔCTCF
CAC/Lone/CNC Control Region
Female
Male

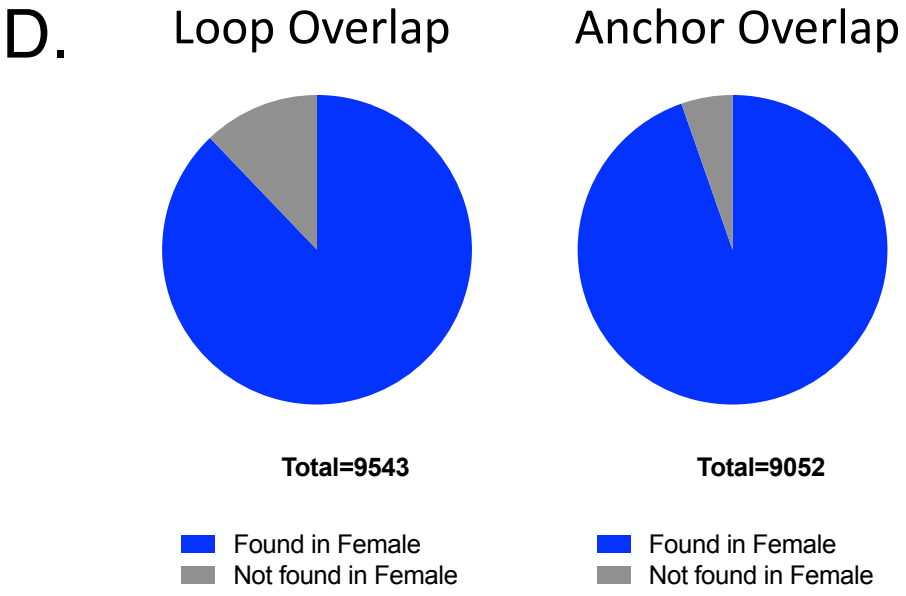
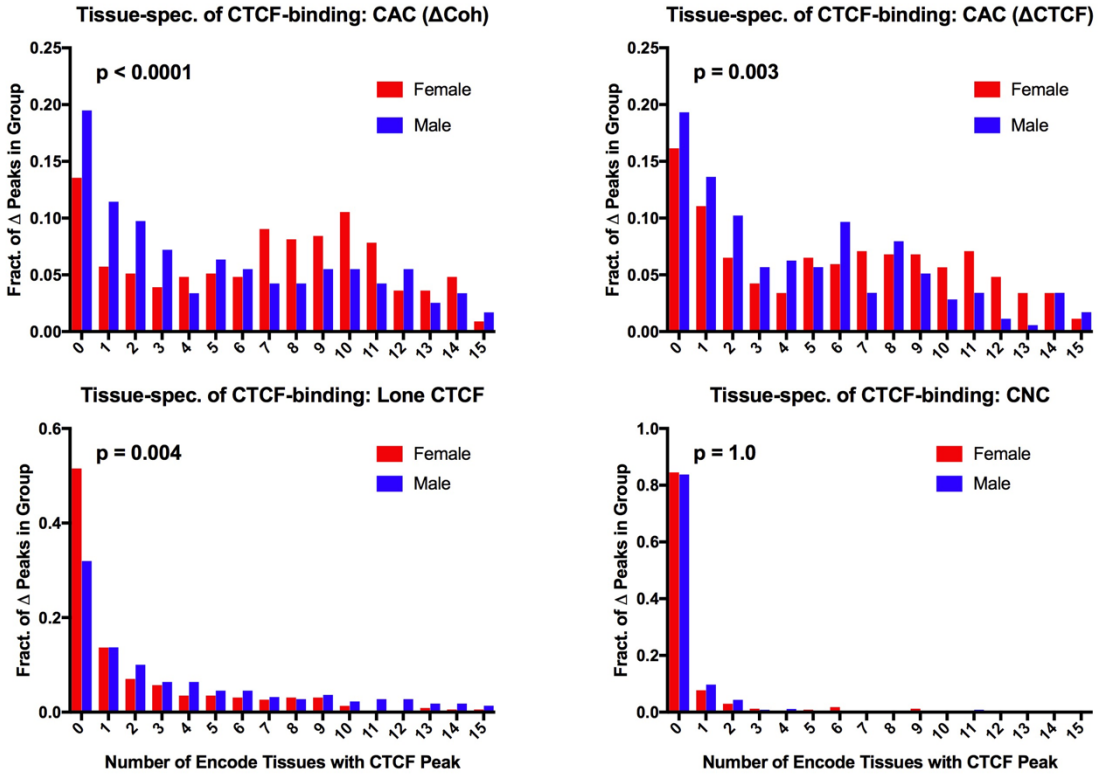
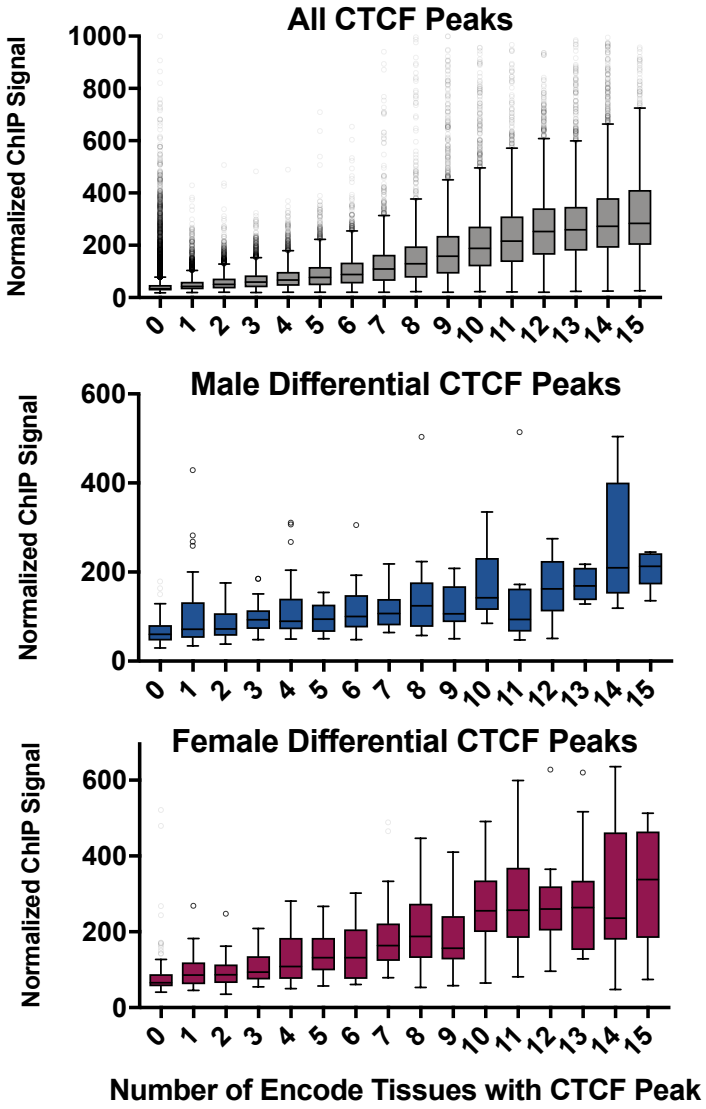

D.
Loop Overlap
Anchor Overlap
Total=9543
Total=9052
Found in Female
Not found in Female
Found in Female
Not found in Female

Fig. S3

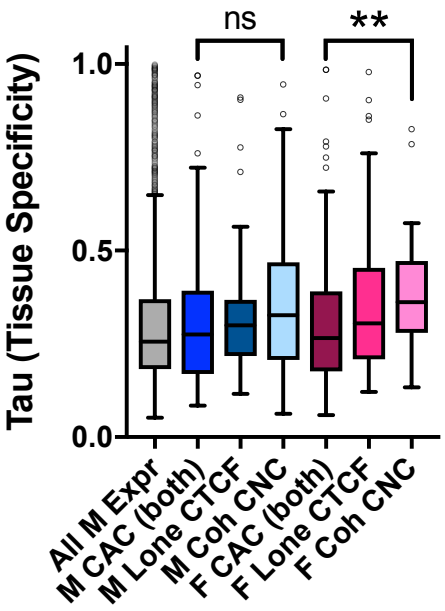
A.



B.



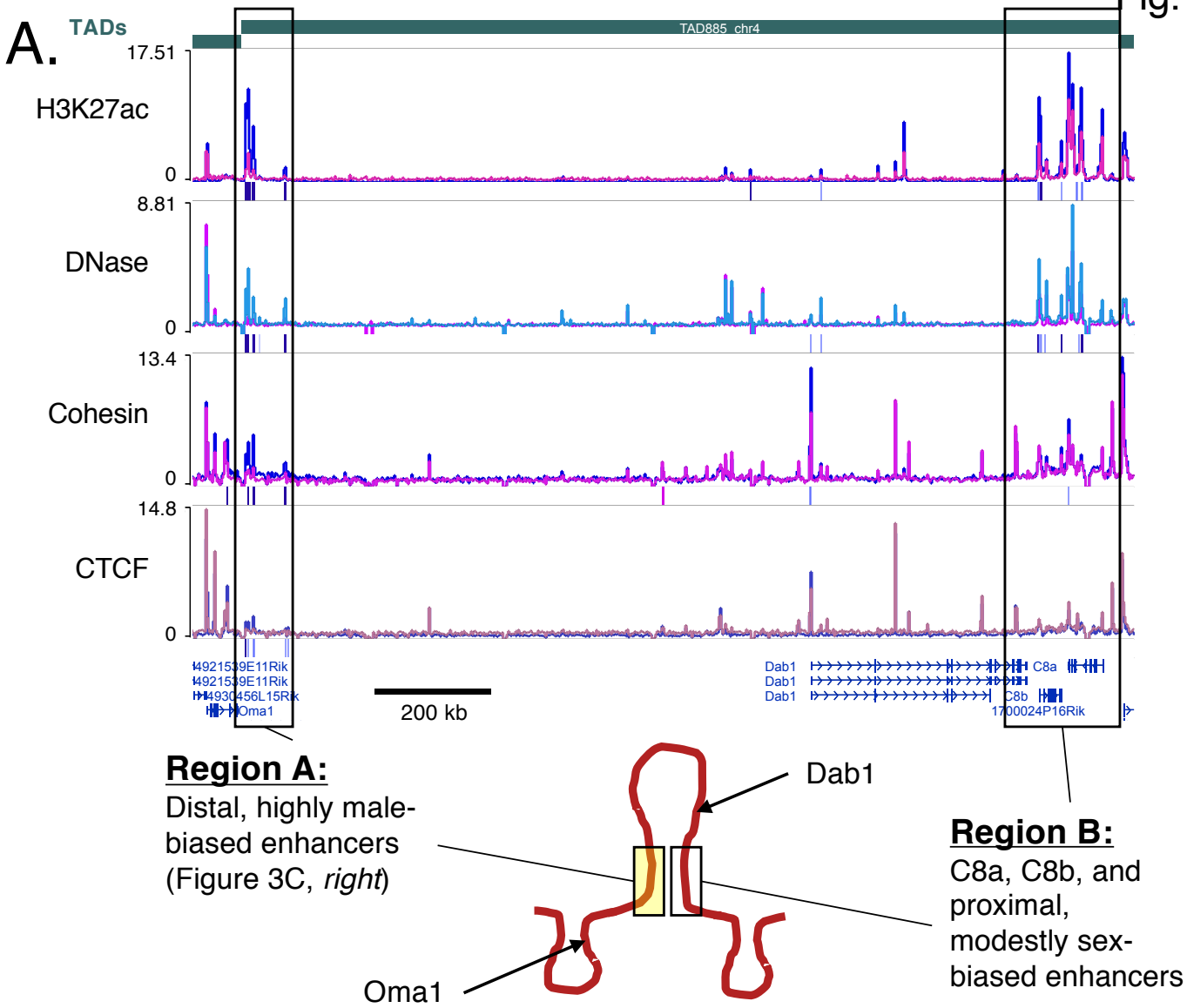
C.



Higher value = more tissue specific

Fig. S4

A.


TADs
TAD885_chr4
17.51
H3K27ac
0
8.81
DNase
0
13.4
Cohesin
0
14.8
CTCF
0
4921539E11Rik
4921539E11Rik
4930456L15Rik
Oma1
200 kb
Dab1
Dab1
Dab1
C8a
C8b
1700024P16Rik
Region A:
Distal, highly male-biased enhancers (Figure 3C, right)
Dab1
Oma1
Region B:
C8a, C8b, and proximal, modestly sex-biased enhancers

B.

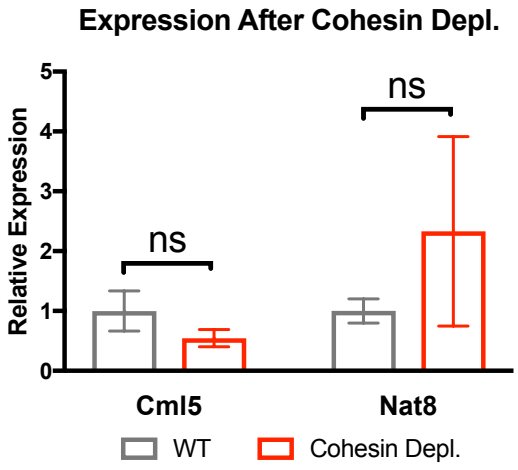
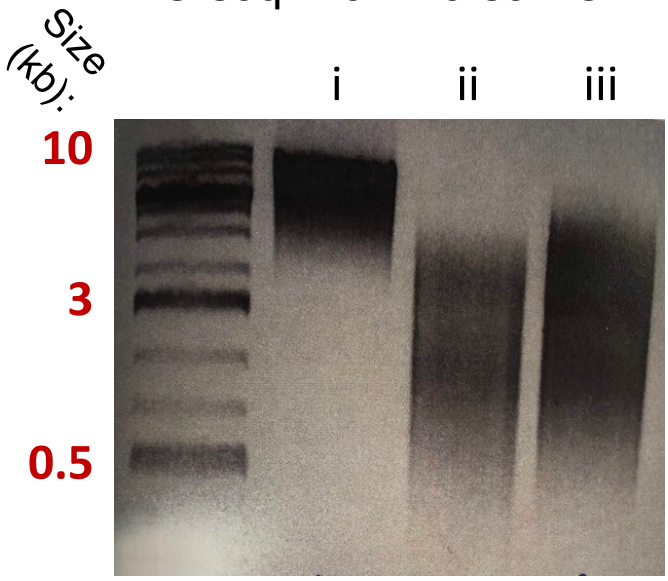
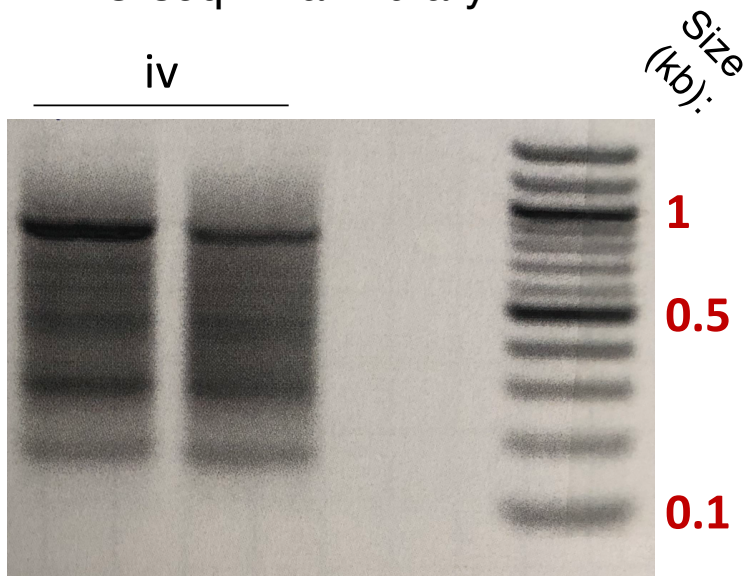

Expression After Cohesin Depl.
Relative Expression
5
4
3
2
1
0
ns
ns
Cml5
Nat8
WT
Cohesin Depl.

Fig. S5

A.



B.



C. **Expression Across Datasets (F FPKM)**

										Nuclear Enrichment	
	Max FPKM Male	Max FPKM Female	Total PolyA Unstr F (G83)	PolyA Unstr F (G85)	Total PolyA Strnd F	Nuclear PolyA Strnd F	Total RiboM Strnd F	Nuclear RiboM Strnd F	B6 Total RiboM Strnd F	Nuclear Enrichment (PolyA)	Nuclear Enrichment (RiboM)
ncRNA_inter_chr15_12590	0.01	2.71	0.62	1.30	0.54	2.57	0.67	2.71	0.13	4.8	4.1
ncRNA_inter_chr15_12591	0.01	2.11	0.35	0.97	0.50	1.15	0.53	2.11	0.18	2.3	4.0
ncRNA_inter_chr15_12592	0	0.99	0.25	0.35	0.06	0.22	0.29	0.99	0.07	3.6	3.5
ncRNA_inter_chr15_12593	0	1.80	0.06	0.19	0.09	1.06	0.31	1.80	0.12	11.4	5.9
ncRNA_inter_chr15_12595	0.01	1.13	0.01	0.01	0.01	0.33	0.23	1.13	0.03	38.8	5.0
ncRNA_inter_chr15_12596	0.02	1.77	0.02	0	0.03	0.24	0.15	1.77	0.07	9.3	11.8
ncRNA_inter_chr15_12597	0	0.91	0.01	0.03	0.02	0.34	0.09	0.91	0.04	15.0	9.9
ncRNA_inter_chr15_12598	0	0.87	0.06	0.00	0.07	0.68	0.11	0.87	0.02	9.2	8.1
ncRNA_inter_chr15_12599	0.02	0.81	0.02	0.04	0.03	0.81	0.09	0.73	0.01	27.9	7.8
ncRNA_inter_chr15_12600	0	0.37	0	0	0.02	0	0.03	0.37	0	0.0	11.8
ncRNA_inter_chr15_12601	0.01	0.31	0	0	0	0.06	0.00	0.31	0	inf	inf
ncRNA_inter_chr15_12602	0	0.34	0	0	0.01	0.03	0.02	0.34	0	2.6	15.1
A1bg	4.35	1,422.30	846.40	1,422.30	1,238.83	382.97	250.99	37.52	522.16	0.31	0.15

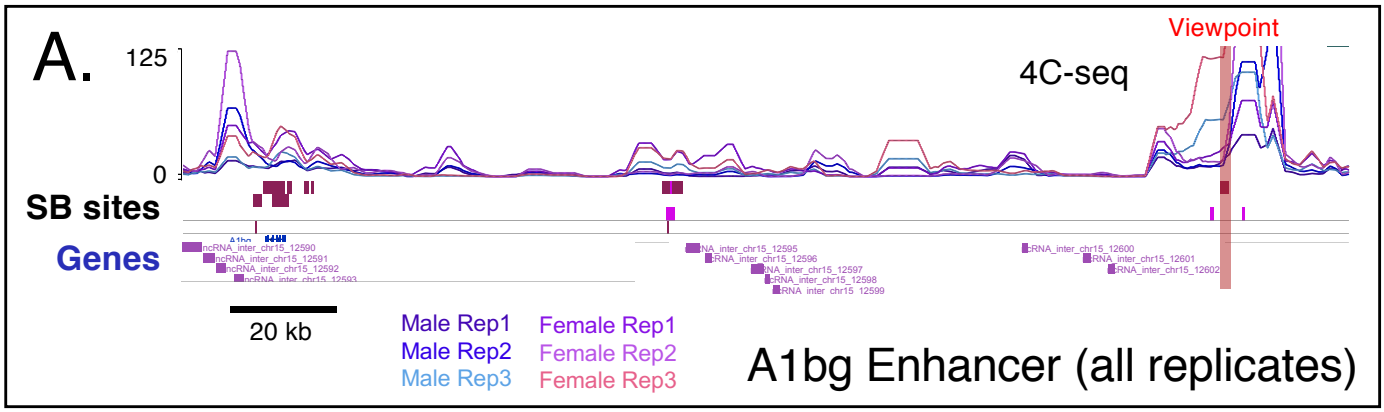
Highlighting = relative within column for lncRNAs only (excluding A1bg)
NE = (Female Nuclear FPKM / Female Total FPKM)

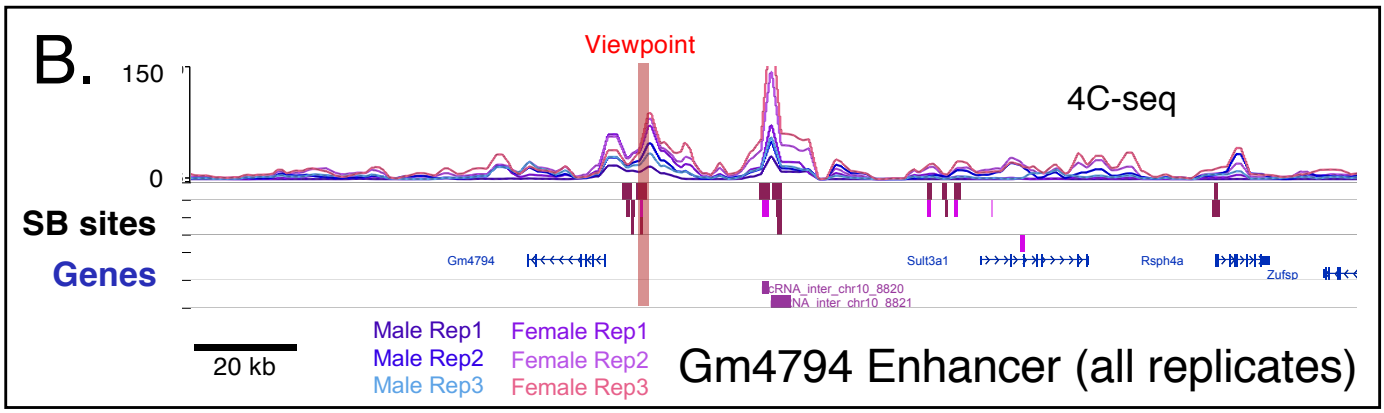
D. **FC of Significant Sex Bias (log2; M/F)**

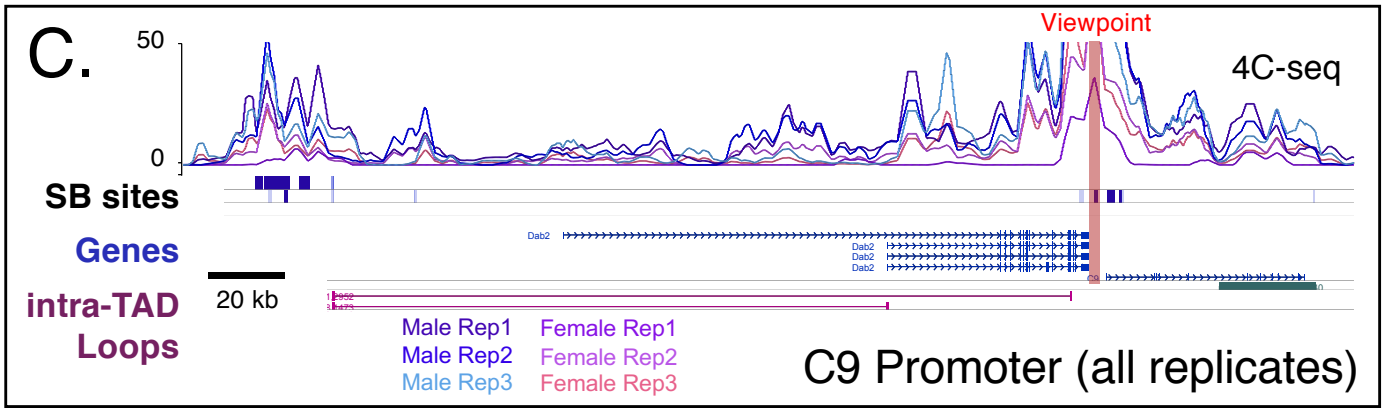
	Log2 FC (M/F) from EdgeR							
	PolyA Unstr F (G83)	PolyA Unstr F (G85)	Total PolyA Strnd F	Nuclear PolyA Strnd F	Total RiboM Strnd F	Nuclear RiboM Strnd F	B6 Total RiboM Strnd F	# of datasets sign. in
ncRNA_inter_chr15_12590	-6.519	-6.870	-8.203	-10.232	-7.557	-9.850	-4.453	7
ncRNA_inter_chr15_12591	-4.937	-8.111	-5.436	-8.313	-6.344	-8.715	-4.078	7
ncRNA_inter_chr15_12592	-6.358	-6.359	ns	-5.682	-5.331	-7.359	ns	5
ncRNA_inter_chr15_12593	ns	ns	ns	-7.893	-5.347	-8.189	-3.330	4
ncRNA_inter_chr15_12595	ns	ns	ns	-6.804	-5.586	-6.170	ns	3
ncRNA_inter_chr15_12596	ns	ns	ns	-5.284	ns	-7.703	ns	2
ncRNA_inter_chr15_12597	ns	ns	ns	-6.741	ns	-7.682	ns	2
ncRNA_inter_chr15_12598	ns	ns	ns	-6.147	ns	-6.023	ns	2
ncRNA_inter_chr15_12599	ns	ns	ns	-7.073	ns	-6.459	ns	2
ncRNA_inter_chr15_12600	ns	ns	ns	ns	ns	-5.199	ns	1
ncRNA_inter_chr15_12601	ns	ns	ns	ns	ns	-5.464	ns	1
ncRNA_inter_chr15_12602	ns	ns	ns	ns	ns	-5.388	ns	1
A1bg	-7.620	-10.348	-8.914	-9.734	-7.278	-7.964	-12.38	7

Highlighting = relative within table for sex bias and relative within column for "# of datasets sign. In"
"# of datasets..." = the number of RNA-seq datasets (left) in which significant sex bias is observed (max of 7)

Fig. S6


A.
125
0
Viewpoint
4C-seq
SB sites
Genes
20 kb
Male Rep1
Male Rep2
Male Rep3
Female Rep1
Female Rep2
Female Rep3
A1bg Enhancer (all replicates)


B.
150
0
Viewpoint
4C-seq
SB sites
Genes
Gm4794
Sult3a1
Rsph4a
Zufsp
20 kb
Male Rep1
Male Rep2
Male Rep3
Female Rep1
Female Rep2
Female Rep3
Gm4794 Enhancer (all replicates)


C.
50
0
Viewpoint
4C-seq
SB sites
Genes
intra-TAD
Loops
Dab2
20 kb
Male Rep1
Male Rep2
Male Rep3
Female Rep1
Female Rep2
Female Rep3
C9 Promoter (all replicates)

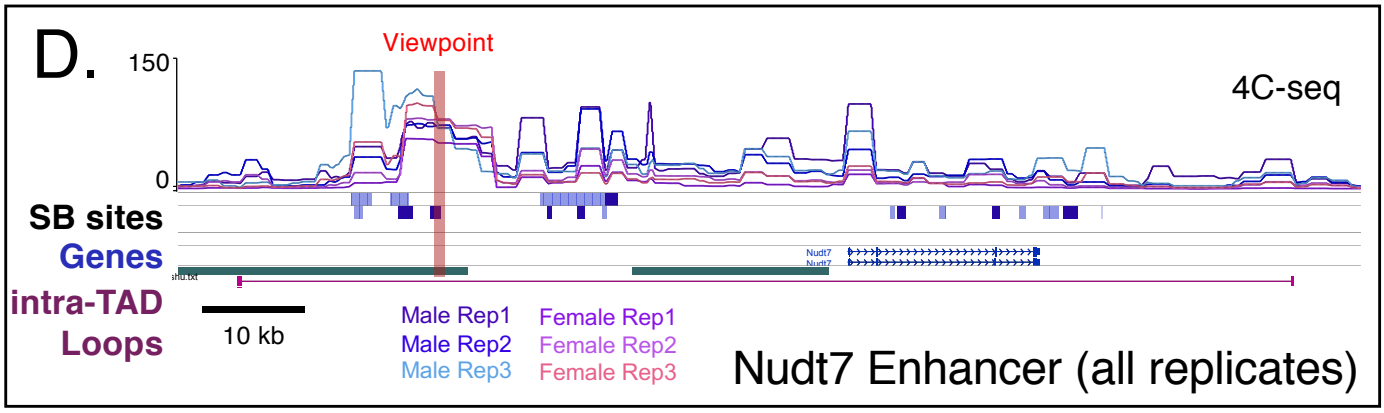
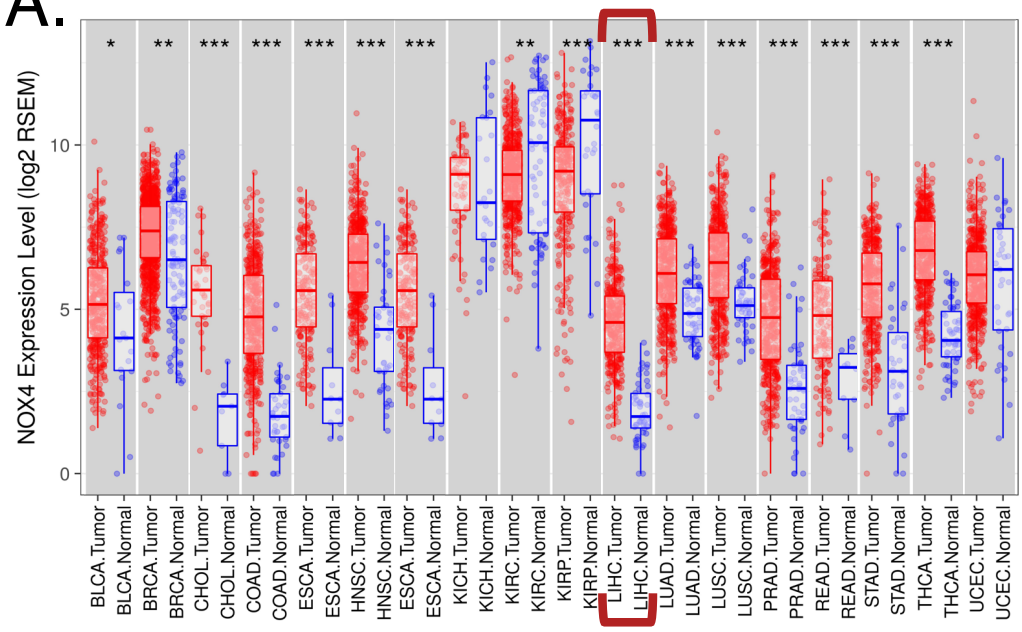
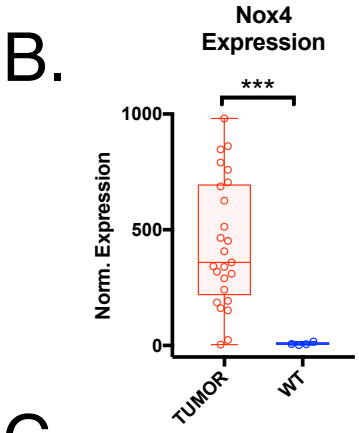
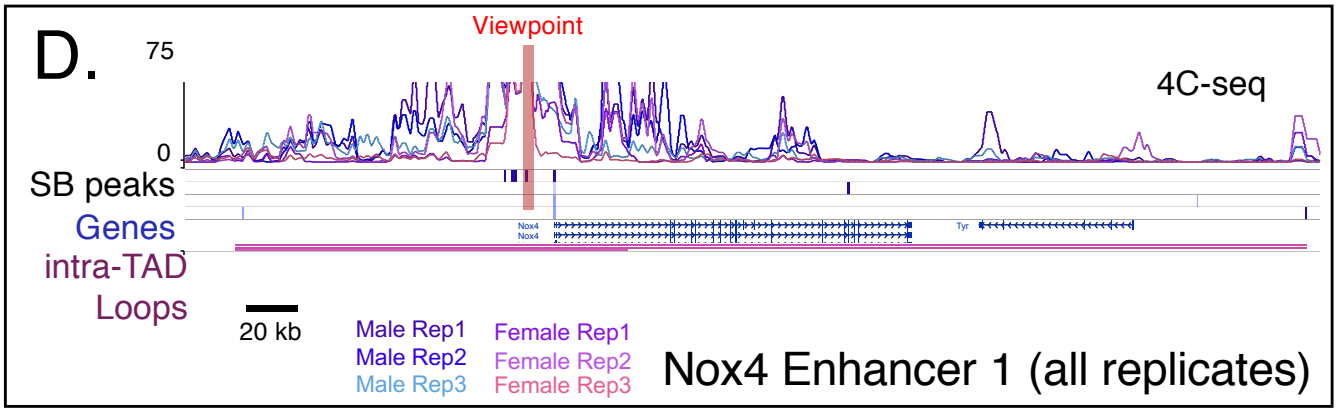

D.
150
0
Viewpoint
4C-seq
SB sites
Genes
intra-TAD
Loops
Nudt7
10 kb
Male Rep1
Male Rep2
Male Rep3
Female Rep1
Female Rep2
Female Rep3
Nudt7 Enhancer (all replicates)

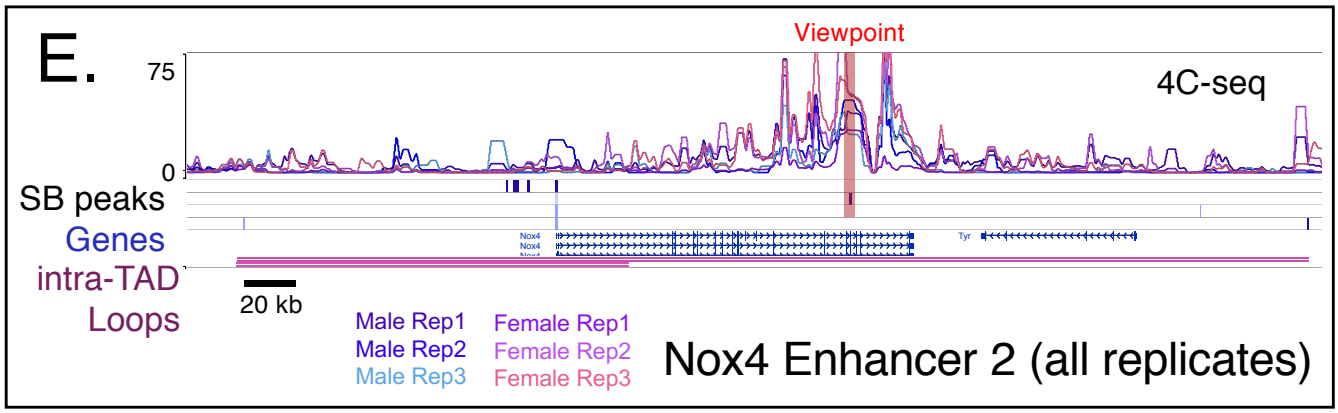
Fig. S7

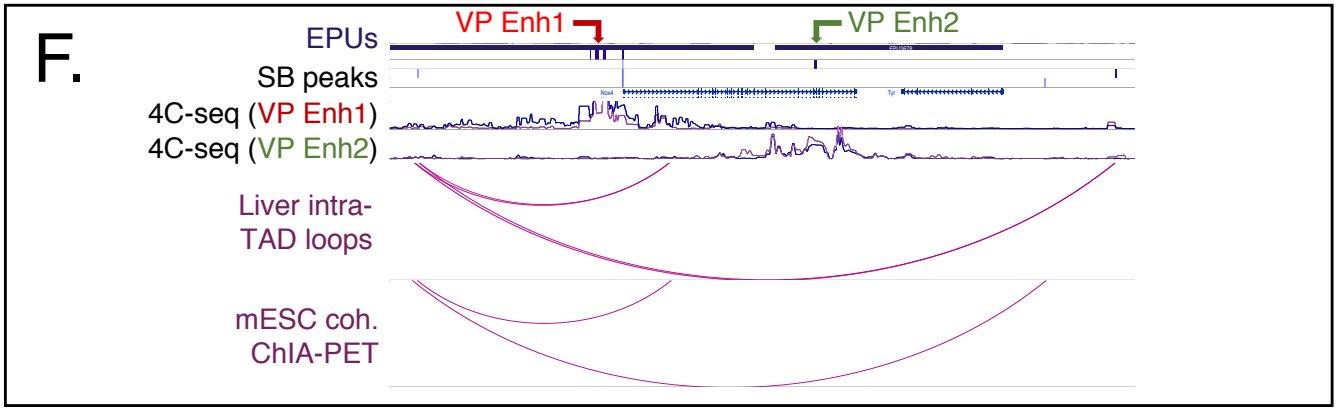

A.
NOX4 Expression Level (log2 RSEM)
BLCA.Tumor
BLCA.Normal
BRCA.Tumor
BRCA.Normal
CHOL.Tumor
CHOL.Normal
COAD.Tumor
COAD.Normal
ESCA.Tumor
ESCA.Normal
HNSC.Tumor
HNSC.Normal
ESCA.Tumor
ESCA.Normal
KICH.Tumor
KICH.Normal
KIRC.Tumor
KIRC.Normal
KIRP.Tumor
KIRP.Normal
LIHC.Tumor
LIHC.Normal
LUAD.Tumor
LUAD.Normal
LUSC.Tumor
LUSC.Normal
PRAD.Tumor
PRAD.Normal
READ.Tumor
READ.Normal
STAD.Tumor
STAD.Normal
THCA.Tumor
THCA.Normal
UCEC.Tumor
UCEC.Normal


B.
Nox4 Expression
Norm. Expression
TUMOR
WT

C.
No marks
Inactive
Bivalent
Enhancer-like
Promoter-like
Transcribed-like


D.
Viewpoint
4C-seq
SB peaks
Genes
intra-TAD
Loops
20 kb
Male Rep1
Male Rep2
Male Rep3
Female Rep1
Female Rep2
Female Rep3
Nox4 Enhancer 1 (all replicates)


E.
Viewpoint
4C-seq
SB peaks
Genes
intra-TAD
Loops
20 kb
Male Rep1
Male Rep2
Male Rep3
Female Rep1
Female Rep2
Female Rep3
Nox4 Enhancer 2 (all replicates)


F.
EPUs
VP Enh1
VP Enh2
SB peaks
4C-seq (VP Enh1)
4C-seq (VP Enh2)
Liver intra-TAD loops
mESC coh. ChIA-PET