bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NSD2 overexpression drives clustered chromatin and transcriptional changes in insulated domains

Priscillia Lhoumaud[1], Sana Badri[1*], Javier Rodriguez Hernaez[1*], Gunjan Sethia[1], Andreas Kloetgen[1], MacIntosh Cornwell[1], Sourya Bhattacharyya[3,4], Ferhat Ay[3,4], Richard Bonneau[2], Aristotelis Tsirigos[1,5] & Jane A. Skok[1#]

[1]Dept. of Pathology, New York University Langone Health, New York, NY 10016, USA; [2]Dept. of Biology, Center for Genomics and Systems Biology, NYU, New York, NY 10003, USA; [3]Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA 92037, USA; [4]School of Medicine, University of California, San Diego, La Jolla, CA 92093, USA; [5]Applied Bioinformatics Laboratories, NYU School of Medicine, New York, NY 10016, USA.

* Equal contribution.

[#] Corresponding author

Correspondence: Jane.Skok@nyulangone.org

Running title: NSD2 overexpression drives 3D genome reorganization

Key words: NSD2; H3K36me2; H3K27ac; CTCF; insulated domains; 3D chromosome organization; compartments; TADs; loops; chromatin

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bullet Points (4 of max 75 words):

- H3K36me2 provides a chromatin landscape favorable for H3K27ac and CTCF enrichment which drives 3D genome reorganization
- CTCF, H3K27ac and gene expression changes cluster within a subset of insulated domains implicating 3D chromosome organization as a key factor in the NSD2-mediated phenotype
- NSD2 High cells co-opt global alterations in chromatin modifications to drive oncogenic transcriptional programs
- Alterations in chromatin modifiers in disease states enables dissection of the functional interplay between chromatin and 3D chromosome structure

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

CTCF and cohesin play a key role in organizing chromatin into TAD structures. Disruption of a single CTCF binding site is sufficient to change chromosomal interactions leading to spreading of active chromatin marks and altered gene regulation. However, the extent to which alterations in chromatin modifications can disrupt 3D chromosome organization leading to changes in gene regulation is not known**.** In multiple myeloma a 4;14 translocation induces overexpression of the histone methyltransferase, NSD2 resulting in expansion of H3K36me2 and shrinkage of antagonistic H3K27me3 domains. Using isogenic cell lines producing high and low levels of NSD2, we demonstrate significant enrichment in H3K27ac and CTCF binding in expanded H3K36me2 domains. Alterations in H3K27ac and CTCF cluster together with gene expression changes in a subset of insulated domains, implicating 3D chromosome organization in driving the altered transcriptional program of NSD2 overexpressing cells. These findings have implications for other diseases involving altered chromatin modifiers.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The 3D organization of chromosomes enables cells to balance the biophysical constraints of the crowded nucleus with the functional dynamics of gene regulation. Chromosomes are divided into large domains that physically separate into two nuclear compartments (A and B) of active and inactive chromatin, respectively (Lieberman-Aiden et al., 2009). Loci within each compartment interact more frequently with other loci from the same compartment irrespective of whether these regions are proximal on the linear chromosome. Compartments can be further subdivided into 1MB sized highly self-interacting 'topologically associated domains' (TADs), which are separated from each other by insulating boundaries enriched for the architectural proteins, CTCF and cohesin. (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Since TADs are largely conserved between cell types and across different species they are considered the basic organizational unit of eukaryotic genomes. They play an important role in gene expression by limiting the influence of enhancers to genes located within the same TAD (Dowen et al., 2014; Ji et al., 2016; Sun et al., 2019) .

CTCF and cohesin are major contributors in shaping architecture within TADs and depletion of either factor leads to loss of TAD structure (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). Depletion of cohesin also leads to strengthening of compartments, suggesting that compartments and TADs are formed by two different mechanisms and that TAD structures antagonize compartmentalization by mixing regions of disparate chromatin status (Rao et al., 2017; Schwarzer et al., 2017). The best-accepted model to explain TAD formation and maintenance, involves a loop-extrusion mechanism whereby cohesin rings create loops by actively extruding DNA until the complex finds two CTCF binding sites in convergent orientation (Fudenberg et al., 2016). In this way cohesin forms chromatin loops as a result of its ability to hold together two double-strand DNA helices via its ring structure (Fudenberg et al., 2016; Nasmyth, 2001; Nuebler et al., 2018; Sanborn et al., 2015; Tang et al., 2015). Indeed, genome-wide

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analyses revealed that loops are preferentially formed between convergently orientated CTCF binding sites (Rao et al., 2014; Vietri Rudan et al., 2015) and divergent sites delineate boundary regions (Gomez-Marin et al., 2015). These sites are thought to serve as a block to the movement of cohesin on chromatin.

A number of labs have demonstrated that disruption of a single CTCF binding site is sufficient to alter chromosomal interactions leading to the spreading of active chromatin and altered gene regulation (Guo et al., 2011; Narendra et al., 2015; Sanborn et al., 2015; Xiang et al., 2011). However, it is not known if the reverse is true and whether alterations in chromatin modifications themselves can impact chromosome organization at the level of A and B compartments, TAD structure, CTCF binding and enhancer-promoter contacts in a manner that corresponds to changes in gene regulation.

An alteration in the balance of the antagonistic marks H3K36me2 and H3K27me3 is a hallmark of many different cancers. Translocation between the immunoglobulin heavy chain locus, *IGH* on chromosome 14 with the *NSD2* locus (also known as MMSET or WHSC1) on chromosome 4, leads to NSD2 overexpression in 15-20% of multiple myeloma (MM) patients that have a poor survival rate and do not respond well to cytotoxic chemotherapy (Huang et al., 2013; Keats et al., 2006; Kuo et al., 2011; Lauring et al., 2008). NSD2 is a histone methyl transferase (HMT) that is responsible for deposition of the H3K36 mono- and di-methyl mark. In a wild-type setting, H3K36me2 accumulates on active gene bodies and acts as a signature of transcriptional activity. However, when NSD2 is overexpressed as a result of the 4;14 translocation, H3K36me2 spreads outside of active gene bodies into intergenic regions. Expansion of H3K36me2 domains results in contraction of H3K27me3 domains, altering gene expression programs in the absence of driver mutations in a manner that is poorly understood (Popovic et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Of note, similar changes in chromatin are detected in other cancers such as B and T acute lymphoblastic leukemia (B- and T-ALL) and a number of advanced stage solid tumors, including prostate, colon and skin cancers (Hudlebusch et al., 2011a; Hudlebusch et al., 2011b). In some cases, increased H3K36me2 results from an E1099K mutation in NSD2 that affects the catalytic domain of this enzyme (Oyer et al., 2014). In addition, a mutation in H3.3 in which the lysine at position 27 is mutated to a methionine (H3K27M) results in a similar H3K36me2 versus H3K27me3 imbalance by impacting the action of EZH2 in two pediatric brain cancers, diffuse intrinsic pontine glioblastoma (DIPG) and supratentorial glioblastoma multiforme (GBMs) (Stafford et al., 2018; Sturm et al., 2012; Wu et al., 2012). DIPG and GBM are the most aggressive primary malignant brain tumors in adults and children and the median survival of this group of patients is approximately 1 year. Given the poor prognosis of patients suffering from these different cancers, it is important to gain a better understanding of the mechanisms underlying changes in gene expression in diseases with an H3K36me2 versus H3K27me3 imbalance.

In multiple myeloma, alterations in gene expression have been shown to be dependent on the histone methyl-transferase activity of NSD2 (Martinez-Garcia et al., 2011). Although the impact of NSD2 overexpression on chromatin modifications has been well documented, there has been no in-depth analysis into the mechanisms underlying the changes in gene expression that occur downstream of the expansion and reduction of active H3K36me2 and repressive H3K27me3 domains. Using isogenically matched multiple myeloma patient derived cell lines that differ only in the levels of NSD2 they express, we demonstrate that spreading of H3K36me2 from active gene bodies into intergenic regions is accompanied by changes in H3K27ac marks (a feature of regulatory elements) as well as an alteration in the genome wide profile of CTCF binding. Both changes are linked to significant alterations in gene expression and oncogene activation. Expansion of H3K36me2 domains also drives compartment switching and alterations in intra-TAD interactions, while altered boundary insulation scores overlap differential CTCF and Rad21

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

binding. Importantly, analysis of Hi-C and CTCF HiChIP data demonstrates that changes in CTCF, H3K27ac together with alterations in transcriptional output, cluster within a subset of insulated regions in NSD2 high expressing cells. These results reveal a bidirectional relationship between 2D and 3D chromatin organization in gene regulation and demonstrate that cells can co-opt altered chromatin domains to drive oncogenic transcriptional programs that are regulated within insulated boundaries.

Results

NSD2 overexpression leads to alterations in H3K27ac

NSD2 overexpression leads to spreading of H3K36me2 from active gene bodies into intergenic regions leading to a more open chromatin conformation (Kuo et al., 2011). Alterations in gene expression have previously been shown to be dependent on the histone methyl-transferase activity of NSD2, highlighting the importance of deposition and expansion of H3K36me2 domains in the activation of oncogenic transcriptional pathways (Martinez-Garcia et al., 2011). However, it is not known how intergenic spreading of H3K36me2 alters gene regulation and whether 3D organization of the genome plays a role.

In order to address this question, we used two isogenic cell lines generated from a patient derived KMS11 t(4;14) multiple myeloma cell line: NTKO (non-translocated knockout) and TKO (translocated knockout) cells, which have the endogenous allele or the translocated *NSD2* allele, respectively inactivated (**Figure 1A**) (Lauring et al., 2008). Importantly, NTKO and TKO cells differ solely in the level of NSD2 they express and henceforth are referred to as NSD2 High and Low cells. The paired cell lines provide us with an opportunity to analyze the impact of NSD2 overexpression independent of other confounding genetic or epigenetic alterations. In this respect they are more useful than patient-based samples which do not have appropriate controls. Using RNA-seq we observed that NSD2 High expression leads to the deregulation of many genes (1650

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

up and 303 downregulated genes) (**Figure 1B**) and principal component analysis (PCA) revealed that NSD2 High and Low replicates separated into distinct clusters (**Figure S1A**), a finding that is consistent with previous studies (Kuo et al., 2011; Popovic et al., 2014). Genes associated with multiple myeloma and KRAS pathways were enriched as shown by Gene Set Enrichment Analysis (GSEA, (Subramanian et al., 2005) **Figure S1B**).

Since the vast majority of regulatory elements are located within intergenic regions, we asked whether spreading of H3K36me2 into these locations could impact their activation status. To test this idea, we analyzed H3K27ac, a feature of both enhancers and promoters. H3K27ac ChIP-seq revealed that a total of 2597 peaks were significantly affected by NSD2 overexpression and PCA revealed that NSD2 High and Low replicates clustered separately (**Figure S2A**). Specifically, we identified 1896 increased and 701 decreased peaks in NSD2 High versus low cells (**Figure 1C and 1D**). Increased and decreased peaks were predominantly located in intergenic and intronic regions suggesting that enhancer activity could be affected (**Figure 1E**). This pattern was much more pronounced for increased peaks.

Previous studies have reported changes in the activity of super-enhancers in multiple myeloma (Fulciniti et al., 2018; Loven et al., 2013). The term super enhancer is used to describe clusters of enhancers with high levels of activating histone marks including H3K27ac (Whyte et al., 2013). Super-enhancers are occupied by lineage specific master regulator transcription factors (TFs) known to control cell identity (Hnisz et al., 2013; Whyte et al., 2013). Furthermore, super-enhancers are generated at oncogenes and other loci that are important in tumorigenesis (Hnisz et al., 2013; Whyte et al., 2013). To separate super-enhancers from typical enhancers we used 'ROSE' (rank ordering of super-enhancers (Whyte et al., 2013)). With this approach, we identified 260 and 278 super-enhancers in NSD2 low and high cells, respectively (Hnisz et al., 2013; Loven et al., 2013) (**Figure S2B**). Differential super-enhancers were identified from the merged group

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

based on alterations in H3K27ac enrichment using an FDR cutoff of 0.1. With this strategy, we identified about a third of super-enhancers (91) with significantly altered H3K27ac signal in NSD2 High cells, 51 with increased and 40 with decreased signal (**Figure 1F**).

Changes in H3K27ac are associated with changes in gene regulation

To determine which transcription factors (TFs) could bind gained and lost super-enhancers in NSD2 High versus Low cells, we analyzed motifs using ATAC-seq. We identified 119 ATAC-seq peaks in the 51 gained super-enhancers and 102 ATAC-seq peaks in the 40 super-enhancers that were lost (**Figure 1F**). TRAP (transcription factor affinity prediction) was used to identify which motifs were present in the ATAC-seq peaks (Thomas-Chollier et al., 2011). Some of the motifs were shared between gained and lost super-enhancers (CTCF, RUNX1 etc), while others were identified in only one category. The expression of some of the genes that encode these TFs goes up or down as denoted respectively by a red or blue asterisk in **Figure 1G.** Most of the motifs in super-enhancers were also found in individual gained and lost H3K27ac peaks, with the exception of the CTCF motif (**Figure S2C**). This is consistent with recent studies showing that super-enhancers are enriched with CTCF more frequently than typical enhancers (Gong et al., 2018; Huang et al., 2018).

To examine the connections between alterations in H3K27ac and gene regulation we separated H3K27ac peaks into those located at promoters, non-promoter sites (intragenic and intergenic combined) and super-enhancers. GREAT (Genomic Regions Enrichment of Annotations Tool; (McLean et al., 2010)) was used to associate peaks at distal sites and super-enhancers with gene expression changes, using an arbitrary cutoff of 250kb, based on findings from a previous study showing that the mean distance from enhancer to promoter is around 196kb (Jin et al., 2013). H3K27ac peaks at all three sites were significantly correlated with gene expression changes (**Figure 1H**), with the strongest effect seen at promoters, possibly because the latter are easiest

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to connect to target genes. Together these data indicate that overexpression of NSD2 leads to significant changes in H3K27ac linked to deregulation of gene expression.

Changes in CTCF binding profile are linked to alterations in gene expression

Nuclear architecture is a powerful regulator of gene expression. In particular, precise transcriptional control is exerted by restricting the influence of enhancers to target genes within TADs whose boundaries are enriched for the insulating protein, CTCF. Given that marks associated with regulatory elements are significantly altered in a manner that corresponds to changes in gene expression in NSD2 overexpressing cells, we next asked whether alterations in histone modifications could lead to changes in chromosome organization. For this analysis, we started with a CTCF ChIP-seq and identified a surprising difference in the CTCF binding profile: 2197 and 295 CTCF peaks were increased and decreased, respectively in the NSD2 High cells (**Figure 2A and B**). These separated into distinct clusters by PCA analysis (**Figure S2D**). Increased CTCF peaks were predominantly located in intergenic and intragenic regions, as compared with decreased and stable CTCF peaks (**Figure 2C**). Moreover, CTCF peaks that were enriched or depleted at promoters and distal sites were significantly correlated with gene expression changes (**Figure 2D**), with the strongest effect seen at promoters. This finding is consistent with previous studies showing that loss of CTCF at promoters leads to downregulation of genes (Nora et al., 2017).

New CTCF and H3K27ac peaks are located within expanded H3K36me2 domains

To better understand the links between NSD2 overexpression and the altered genomic profiles of H3K27ac and CTCF, we analyzed the chromatin landscape surrounding the 1000 most enriched new H3K27ac and CTCF peaks in the NSD2 High cells. Importantly, new H3K27ac and CTCF peaks were found to be located in newly enriched H3K36me2 regions while H3K27me3 remained unchanged at these locations (**Figure 2E**). Thus, increases in H3K27ac peaks were found to be

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

independent of H3K27me3 changes. New H3K27ac and CTCF peaks were also associated with a gain in overlapping ATAC-seq peaks. This data suggests that spreading of H3K36me2 provides a more accessible chromatin landscape favorable for CTCF and H3K27ac enrichment. As shown in **Figure 2F** these changes are linked to the activation of oncogenes involved in multiple myeloma including *SYK* (Koerber et al., 2015; Lin et al., 2017; Liu and Mamorska-Dyga, 2017; Lorenz et al., 2016; Qin et al., 2017)*, MET* (Baljevic et al., 2017; Phillip et al., 2009; Zaman et al., 2015) and *SH3GL3* (Chen et al., 2016) in NSD2 overexpressing cells (**Figure 2F**). Of note, few changes were observed for the 1000 most stable and decreased H3K27ac and CTCF peaks (**Figure S3**).

The vast majority of chromosomal loops involve cohesin. These can be separated into two main categories: loops that are CTCF dependent and CTCF independent cell type specific, dynamic loops that form between enhancers and promoters and involve cell type specific TFs (Snetkova and Skok, 2018). To investigate whether new H3K27ac and CTCF peaks had the potential to be involved in looping in the NSD2 High cells we performed ChIP-seq for Rad21 (a component of the cohesin complex). Interestingly, increased Rad21 signal was associated with both newly enriched H3K27ac and CTCF sites in the NSD2 overexpressing cells. However, gained CTCF and H3K27ac peaks did not overlap with each other (**Figure 2E**), suggesting that loops involving new regulatory elements associated with enriched H3K27ac were generated in a CTCF independent manner.

NSD2 overexpression drives A/B compartment switching

Given the finding that there are significant changes in CTCF or H3K27ac peaks overlapping Rad21 alterations in NSD2 High cells, we next asked whether 3D organization was altered. To determine this, we performed Hi-C in duplicate using the Arima Kit and processed the data using Hi-C bench (Lazaris et al., 2017). The number of valid read pairs was consistent between replicates (~120 million to ~180 million (**Figure S4A**)) and the PCA showed that NSD2 High

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

versus Low replicates separated as expected (**Figure S4B**). A/B compartment status was analyzed in merged NSD2 Low and High replicates using the Eigen vector (principal component 1, PC1, see STAR Method for details) (Lieberman-Aiden et al., 2009). We observed significant compartment alterations in NSD2 High cells, as exemplified by the changes visualized on chromosome 7 in **Figure 3A**. Overall, the number of A and B compartments was lower in NSD2 High versus Low cells and we detected 324 regions that had switched from A to B and 491 regions that switched from B to A (**Figure 3B, C and D**). Although there were more regions switching from B to A, these regions were smaller and the portion of the genome was comparable to the regions switching from A to B (**Figure S4C, D and E**). Compartment switching from B to A overlapped with expanded H3K36me2 domains and a reduction in H3K27me3, while the reverse was true for switching from compartment A to B (**Figure 3D and E**).

Changes in TAD boundaries and intra-TAD interactions

Both Hi-C replicates from each condition harbored similar number of TADs with similar sizes (**Figure S4F and G**) and they were merged for TAD analysis to determine the impact of NSD2 overexpression on 3D genome organization. Using Hi-C bench, we assigned a mean boundary score to each replicate using a resolution of 40kb (Lazaris et al., 2017) to determine if there were changes in TAD boundaries (see STAR Method for details). NSD2 overexpression was associated with a significant (FDR < 0.05) increase in insulation at TAD boundaries (red, 61) and much fewer boundaries that were weakened (blue, 5) **(Figure 4A)**. These changes were synonymous with changes in CTCF and Rad21 binding (**Figure 4B**) and consistent with the finding that the majority of CTCF alterations were associated with increased rather than decreased binding (**Figure 2A and B**). In addition, we found that NSD2 High cells were predominantly associated with significant gains (red, 229) and fewer decreases in intra-TAD interactions (blue, 30) (**Figure 4C**). Increased intra-TAD interactions were linked to expanded

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

H3K36me2 domains and a reduction in H3K27me3, while decreases in intra-TAD interactions were linked to a reduction in H3K36me2 (**Figure 4D**).

As mentioned above, NSD2 overexpression was associated with a significant strengthening of TAD boundaries concordant with increases in CTCF and Rad21 binding (**Figure 4B**). These boundary changes were also associated with transcriptional activation. As an example we highlight activation of *FZD8* (log2 fold-change = 3.16, FDR = 8.18E-34), a gene that is involved in the WNT signaling pathway (Spaan et al., 2018) (**Figure 4E**). *FZD8* is located at a CTCF enriched boundary that makes a new contact with a downstream region that gains a CTCF/Rad21 site (blue stripe in **Figure 4E**) in NSD2 High cells. This is demonstrated by an increase in boundary score (arrows in **Figure 4E and F**) and the formation of a new Hi-C loop (circle in **Figure 4F**). Consistent with the data in **Figure 2D**, the whole domain overlaps a region that is enriched for H3K36me2 and within the new loop there are increases in H3K27ac marks.

Previous studies have proposed that *FGF13* expression enables cells to cope more effectively with RAS-mediated stress and oncogene-mediated excessive protein synthesis, which is known to occur in multiple myeloma in the form of antibody production (Bublik et al., 2017). In **Figure S5A and B** we show that activation of *FGF13* occurs at a region with an altered TAD boundary and increased intra-TAD interactions. This region harbors newly formed H3K27ac peaks, enriched chromatin accessibility and overlaps a domain that is enriched for H3K36me2 that undergoes B to A compartment switching. Increased CTCF and Rad21 binding are at the edges of increased Hi-C loops, suggesting that newly formed insulated neighborhoods could facilitate the interaction between *FGF13* and regulatory elements.

We also found preexisting CTCF independent contacts between H3K27ac enriched regions that fall within a loop that has enriched CTCF binding. As an example, we highlight the *KRAS*

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

oncogene (yellow strip **Figure S5C and D**) that plays an important role in driving the multiple myeloma phenotype (Walker et al., 2012; Xu et al., 2017). As shown in **Figure S5B**, upregulation of *KRAS* (log2 fold change = 0.86, FDR=1.38E-06) coincides with a preexisting loop between *KRAS* and a super enhancer whose activity is enriched in NSD2 high cells (see arrow and blue stripe **Figure S5C** and arrows and circle in **S5D**). The figure also highlights an increase in intra-TAD activity. The above examples indicate that significant changes in H3K27ac, CTCF and transcriptional output are frequently clustered together.

NSD2 overexpression drives clustered chromatin and transcriptional changes in insulated domains

To investigate the connections between alterations in gene expression, CTCF and H3K27ac from a global perspective, we focused our analysis on TADs and CTCF-mediated loops. To identify CTCF-mediated loops, we performed CTCF HiChIP in NSD2 High and Low cells (Mumbach et al., 2016). Statistical analysis of HiChIP data was carried out using FitHiChIP ((Bhattacharyya et al., 2018), see STAR Method for details). CTCF-mediated interactions were considered significant using an FDR <0.01.

To determine whether significant changes in CTCF, H3K27Ac and gene expression in NSD2 High were localized together, we focused on common TADs and common CTCF-mediated loops (see details in STAR Methods). This allowed us to reliably compare between the NSD2 high versus Low condition. The number of significant changes in CTCF, H3K27Ac and gene expression in these common domains is shown in (**Figure S6A and B**). In general, we found that TADs and CTCF loops had a median size of 600kb and 210kb, respectively (**Figure S6C**).

As shown in the scheme in **Figure 5A**, we selectively analyzed TADs and CTCF-mediated loops containing at least one CTCF and H3K27Ac peak as well as at least one transcriptionally active

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gene in either the NSD2 Low or High condition. Without any filtering we found that up or down regulation of the three variables was positively correlated in both TADs and CTCF-mediated loops as shown by the red and blue dots, where each dot represents a single TAD or CTCF-mediated loop (**Figure S7A and B**). Positively correlated features are shown by pairwise and three-way log2 fold change comparisons in the 2D and 3D plots, respectively. We also included in our analysis a pairwise and three-way comparison of intra-TAD interaction and PC1 changes (see methods for details) to examine compartment differences within insulated regions that contained alterations in CTCF, H3K27Ac and transcriptional output (**Figure S7A**). Again, we found that alterations in intra-TAD interactions and compartment switching were concordant within TADs with alterations in CTCF, H3K27Ac and transcriptional output.

We next redid the pairwise and three-way comparisons restricting our analyses to only the significant changes in CTCF, H3K27Ac and transcriptional output within TADs and CTCF-mediated loops and found an improved correlation (**Figure 5B and C**). We also confirmed that high PC1 mean difference values are associated to regions switching from B to A using the Homer algorithm. This analysis showed that 83% of TADs with B to A switching regions (orange dots) overlap with active features in the 3D graph in common TADs with differential changes in CTCF, H3K27ac as well as RNA (25% of the total positively correlated active TADs) (**Figure 5B**). Together these analyses indicate that changes in CTCF, H3K27ac and transcriptional output are positively correlated and cluster within insulated domains.

Concordant and discordant transcriptional and chromatin changes cluster within a subset of common TADs and CTCF HiChIP loops

To further determine the relationship between significant gene expression changes, H3K27ac and CTCF, we separated common TADs and CTCF-mediated loops into concordant and discordant changes (i.e. positively correlated and negatively correlated, respectively) (**Figure 6A**). As shown

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in this figure, the vast majority of changes were concordant. We define concordant and discordant TADs and CTCF loops according to the average log2fc of CTCF, H3K27ac and transcription. As a result, a discordant loop with a positive change in CTCF and expression and overall negative change in H3K27ac could include enhancers that are strongly downregulated as well as those that are upregulated. Thus, even though the overall H3K27ac change is negative we cannot exclude the possibility that there will not be a newly active enhancer present in the domain. Furthermore, enriched binding of CTCF could have an insulating effect by interfering with the contact between an enhancer and its target gene resulting in transcriptional repression. For these reasons we included both concordant and discordant changes into our analysis.

Significant changes in gene expression in the concordant and discordant TADs and CTCF-mediated loops were further subdivided into two groups: (i) those that overlapped with significant changes in CTCF *and* H3K27ac and (ii) those that overlapped with significant changes in *either* CTCF or H3K27ac (**Figure 6B**). A total of 819 differentially regulated genes were captured in this analysis, which amounted to 70% of the 1179 differentially expressed genes found in common TADs or CTCF-mediated loops, which is more than expected by random chance (19%). Thus, significant concordant and discordant transcriptional changes cluster with CTCF and/or H3K27ac changes within a subset of common TADs (49%) and CTCF HiChIP loops (22% as shown by a hypergeometric test (p-value = 9.5 e-395).

Within the TADs and loops that harbor significant changes in CTCF, H3K27Ac and transcriptional output we detected several newly activated oncogenes associated with multiple myeloma related pathways. These include the protein tyrosine phosphatase gene, *PTPN13* (Sardina et al., 2014; Zhou et al., 2009), *FGF13* (Mahtouk et al., 2010) and *ETV5* (Chang-Yew Leow et al., 2013; Croonquist et al., 2003). The *PTPN13* gene, which regulates cell growth (Zhuang et al., 2017), differentiation (Sardina et al., 2014), mitotic cycle, and oncogenic transformation is located within

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

a CTCF-mediated loop which harbors significant changes in CTCF and H3K27Ac (**Figure 6C**), and it is found within a region that undergoes B to A compartment switching. *ETV5*, *a* gene involved in the KRAS pathway (Hollenhorst et al., 2011; Zhang et al., 2017), is another example of gene activation that occurs in a CTCF-mediated loop and a TAD that harbors significant changes in CTCF, H3K27Ac (**Figure 6C**). As with *PTPN13,* the insulated region contains new H3K27ac and CTCF peaks, overlaps a region that has switched from compartment B to A, and in addition has increased intra-TAD interactions (**Figure 6C**).

Activation of *SYK* involves increased promoter contacts within a CTCF-mediated loop

SYK is another example of an oncogene that is activated within a CTCF-mediated loop that has significant increases in H3K27ac and CTCF peaks. To investigate whether upregulation of this gene in NSD2 High cells involves alterations in promoter contacts we performed high resolution 4C-seq from a viewpoint (bait) located on a CTCF site 27Kb downstream of the *SYK* promoter. We identified significantly increased interactions with upstream and downstream regions (FDR < 0.05; marked by the small ovals on the 4C plot in **Figure 7A**). All of these sites correspond to enriched CTCF or H3K27ac peaks that overlap with enriched cohesin peaks in the NSD2 High cells (**Figure 7B**). The interactions identified by 4C-seq can also be visualized by Hi-C (arrows and circles in **Figure 7C**) or by HiChIP (**Figure 7B**, loop in “CTCF HiChIP interactions” track). Interestingly, the change in interactions seen in **Figure 7A** overlap with a region that changes from compartment B to A and has increased intra-TAD interactions. As in other examples, these changes all occur in a domain where H3K36me2 is enriched in the NSD2 High cells. Taken together our data indicate that gene expression changes and activation of oncogenes in NSD2 overexpressing cells are found in a subset of insulated regions with accompanying alterations in H3K27ac and CTCF peaks. These data are consistent with the widely accepted idea that the vast majority of enhancers exert their effects on target genes within the same insulated neighborhood. Consistent with this model, we find that alterations in H3K27ac (which reflect the activation status

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of enhancers) and/or CTCF (which mediate changes in interactions and boundaries), occur together with changes in the transcriptional output of genes in the same insulated domain.

Discussion

Many labs have shown that deletion of individual CTCF sites can disrupt chromosome interactions allowing the spreading of active chromatin and altering gene regulation (Guo et al., 2011; Narendra et al., 2015; Xiang et al., 2011). However, it is not known whether alterations in chromatin can affect CTCF binding, 3D genome organization and gene expression. To address this question we examined the effect of NSD2 overexpression in t(4;14) multiple myeloma. The activation of oncogenic transcriptional pathways in this context has been shown to rely on the histone methyl transferase activity of NSD2 and deposition of H3K36me2 (Martinez-Garcia et al., 2011). Here we show that NSD2 overexpressing cells can co-opt global changes in chromatin modifications to drive a disease specific transcriptional program that is regulated within insulated boundaries.

Expansion of H3K36me2 outside of active gene bodies provides a favorable environment for enrichment of H3K27ac and binding of CTCF. Hi-C and HiChIP analyses reveal that changes in H3K27ac, CTCF and transcriptional activity occur in a predominantly concordant manner in common TADs (72%) and CTCF loops (82%). Furthermore, intra-TAD interactions and B to A compartment changes are positively correlated with changes in H3K27ac, CTCF and transcriptional activity in TADs. Importantly, we find that 70% of the deregulated genes cluster together in a subset of TADs or CTCF-mediated loops that harbor significant changes in CTCF and/or H3K27ac binding. Alterations in gene regulation within this subset of insulated domains occurs at a frequency that is significantly above that expected by random chance. These findings suggest that the presence of alterations in either CTCF and/or H3K27ac in the same insulated region has an impact on gene expression. As highlighted in the examples in this manuscript, these

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

changes provide an explanation for the activation of many oncogenes associated with multiple myeloma. Our findings are consistent with a model in which gene regulation is constrained at the 3D level by insulated boundaries that restrict the influence of enhancers.

Since the NSD2 phenotype relies on the histone methyl transferase activity of the protein, the upstream causal event is the expansion of H3K36me2 domains. Beyond this change it is not clear what order the other events occur in, or whether they are reliant on each other. It is unlikely that alterations in H3K27ac enrichment and CTCF binding are dependent on each other as they do not bind to overlapping sites. We speculate that opening up of the chromatin through deposition of the active mark, H3K36me2 allows binding of CTCF and transcription factors such as AP1, which recruits CBP/p300 to mediate deposition of H3K27ac, a well-defined marker of enhancer activity. Indeed, we found that the motif for AP1 is enriched at new H3K27ac peaks including those designated as super-enhancers. Moreover, the gene that encodes this protein is upregulated in NSD2 high cells. H3K27ac enriched regulatory elements and CTCF bound sites are known to participate in cohesin-mediated loop formation and in line with this, we show that new H3K27ac and CTCF peaks overlap with new Rad21 peaks and form new or strengthened contacts, linked to changes in gene expression. An overall increase in CTCF/Rad21 and H3K27ac as well as transcriptional activity likely contributes to the overall increase in intra-TAD interactions as shown by the pairwise and three way-comparisons in **Figure 5**. These changes are also positively correlated with B to A compartment changes. Specifically, we find that 83% of TADs with B to A switching regions overlap with active features in the 3D graph in common TADs with differential changes in CTCF, H3K27ac as well as RNA.

Here we have identified a functional interplay between NSD2-mediated changes in chromatin, 3D organization and transcriptional output. These results underscore the bidirectional relationship between 2D chromatin status and 3D genome organization in gene regulation. In the context of

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

multiple myeloma, a cancer in which many patients do not harbor oncogenic driver mutations, we demonstrate that global chromatin changes can lead to the increased expression of *KRAS* and *SYK*, that contribute to the oncogenic program of the disease as highlighted by the enrichment of KRAS amongst deregulated pathways (**Figure S1B**). Furthermore, our finding that global alterations in chromatin drive changes in transcriptional programs that can be regulated at the level of insulated domains has implications for any disease that involves altered chromatin modifiers. Indeed, this work demonstrates that alterations in chromatin modifiers in disease states enables dissection of the interplay between 2D and 3D chromatin structure and their links to gene regulation

Acknowledgements

The authors thank Ben Ho Park laboratory for sending the NSD2 High and Low cells, Skok lab members for helpful scientific discussions, New York University School of Medicine High Performance Computing (HPC) for computing technical support, and Adriana Heguy and the Genome Technology Center (GTC) core for sequencing efforts.

Funding

This work was supported by NIH 1R35GM122515 (J.S), American Association for Cancer Research Takeda Multiple Myeloma fellowship (P.L), National Cancer Center postdoctoral fellowship (P.L), American Cancer Society (RSG-15-189-01-RMC) (A,T), St. Baldrick's foundation (581357) (A.T) and NIH R35GM128938 (F.A).

Author contributions

Conceptualization & Study Design, P.L and J.S; Investigation, P.L; Formal analysis, P.L, Sa.Ba, J-R.H, G.S, A.K, M.C, So.Bh; Writing – Original Draft, P.L, J.S, Sa.Ba and J-R.H; Supervision, R.B, F.A, A.T and J.S; Funding Acquisition, P.L, J.S, A.T and F.A.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Declaration of Interests

The authors declare no competing interests.

Figure legends

Figure 1. NSD2 overexpression leads to alterations in H3K27ac enrichment linked to changes in gene expression. (A) NSD2 Low and High isogenic cell lines generated from a patient derived KMS11 t(4;14) multiple myeloma cell line: NTKO (non-translocated knockout) and TKO (translocated knockout) cells, which have the endogenous allele or translocated *NSD2* allele, respectively inactivated. (B) Volcano plot showing significant NSD2-mediated changes in gene expression (FDR <0.01). Upregulated genes: 1650 (red, log2 fold change >1); downregulated genes: 303 (blue, log2 fold change <-1). (C) Volcano plot showing significant NSD2-mediated changes in H3K27Ac (FDR <0.01). Decreased peaks: 701 (blue, log2 fold change <-1); increased peaks: 1896 (red, log2 fold change >1). (D) Heatmap showing differential H3K27ac peaks identified using DiffBind analysis of ChIP-seq triplicates in NSD2 High versus Low cells. Increase and decrease in H3K27ac signal is shown in red and blue, respectively. (E) Genomic locations of the differential H3K27ac peaks. (F) Volcano plot showing significant NSD2-mediated changes in super-enhancers (FDR <0.1). Super-enhancers in NSD2 High versus Low cells were called based on H3K27ac levels using ROSE. Increased super-enhancers: 51 (red) and decreased super-enhancers: 40 (blue). (G) Transcription factor motifs identified in 119 ATAC-seq peaks of increased (51) and 102 ATAC-seq peaks of decreased (40) super-enhancers using TRAP. Motifs found in increased (red) and decreased (blue) super-enhancers (-log10 p-value). Red and blue stars indicate that the gene coding for the TF is respectively up- or downregulated in NSD2 High cells. (H) Gene expression changes are associated with H3K27ac changes at promoters (top panel, -/+ 3kb around TSS), distal H3K27acand super-enhancers (middle and bottom panels,

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

respectively, 3-250kb up and downstream from TSSs associated with genes using GREAT). H3K27ac up, down and stable red, blue and grey in NSD2 High versus Low cells.

Figure 2. New CTCF and H3K27ac peaks are located within expanded H3K36me2 domains. (A) Volcano plot showing significant NSD2-mediated changes in CTCF binding (FDR <0.01). Increasing peaks: 1650 (red, log2 fold change >1); decreasing peaks: 295 (blue, log2 fold change <-1). (B) Heatmap showing differential CTCF peaks identified using DiffBind analysis of ChIP-seq triplicates in NSD2 High versus Low cells. Increases and decreases in CTCF binding are shown in red and blue, respectively. (C) Genomic locations of differential CTCF peaks. (D) Gene expression changes are associated with CTCF changes at promoters (top panel, -/+ 3kb around TSS) and distal CTCF (bottom panel, 3-250kb up and downstream from TSSs associated with genes using GREAT). CTCF up, down and stable red, blue and grey in NSD2 High versus Low cells. (E) Heatmaps of H3K27ac, H3K36me2, H3K27me3, ATAC-seq, Rad21 and CTCF signal at the top 1000 increased H3K27ac (top panels) and CTCF (bottom panels) peaks in NSD2 High cells. Peaks are ranked by H3K27ac (top panels) and CTCF (bottom panels) signal in NSD2 High cells. (F) UCSC genome browser screenshots of chromatin features at regions surrounding three oncogenes (*SYK*, left panel; *MET*, middle panel; and *SH3GL3*, right panel).

Figure 3. NSD2 overexpression drives A/B compartment switching. (A) Compartment weakening for half of chromosome 7 is shown in NSD2 High versus Low cells. Top and right: Eigen vector (PC1) for compartments A and B in red and blue, respectively. Switching regions are shown in green and in purple the subtraction of CTCF signal (NSD2 High - Low). Heatmaps represent the Pearson correlation of interactions in NSD2 Low (top) and High (bottom) cells. Positive and negative Pearson correlation between two loci are represented in red and blue, respectively. (B) Total number of A (red) and B (blue) compartments in NSD2 High versus Low cells. (C) Number of regions that switch compartments from B to A (491 regions) or A to B (324

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

regions) in NSD2 High versus Low cells. (D) IGV screenshots show examples of regions that switch from B to A (top panel) and A to B (bottom panel) in NSD2 High versus Low cells. Eigen vectors (PC1), differential compartment switching, subtraction tracks of H3K36me2, H3K27me3, ATAC-seq, H3K27ac, Rad21 and CTCF, and log2 fold-change expression for NSD2 High/Low cells are shown. H-L refers to subtraction High – Low. Regions that significantly switch from B to A and from A to B are indicated in red (top panel, B to A) or in blue (bottom panel, A to B), respectively. (E) Changes in H3K36me2 and H3K27me3 levels within regions that switch compartment from B to A (red), A to B (blue) or are stable (grey) in NSD2 High cells. Log2 fold changes (NSD2 High versus Low cells) for H3K36me2 and H3K27me3 are shown. The median is indicated under the violins.

Figure 4. Changes in TAD boundaries and intra-TAD interactions. (A) Boundary alterations in NSD2 High versus Low cells. NSD2 overexpression is associated with increases (red, 61) and decreases (blue, 5) in TAD boundary strength (cutoffs of absolute Log2 fold change > 0.1 and FDR < 0.05). (B) TAD boundary increases and decreases in NSD2 High versus Low cells are associated with increases and decreases in CTCF and Rad21 binding, respectively. The median is indicated under the violins. (C) Intra-TAD interaction changes in NSD2 High versus Low cells for common TADs (1564). NSD2 overexpression is associated with gain (red, 229) and loss (blue, 30) of intra-TAD interactions (cutoffs of absolute Log2 fold change > 0.3 and FDR < 0.05). (D) Changes in H3K36me2 and H3K27me3 within TADs that have increased (red), decreased (red) or stable (grey) interactions in NSD2 High cells. Log2 Fold changes (NSD2 High/Low cells) for H3K36me2 and H3K27me3 ChIP-seq are shown. The median is indicated under the violins. (E) UCSC tracks showing chromatin features in the region surrounding the *FZD8* gene (*FZD8* gene indicated in red and location highlighted by a yellow stripe) and the new contact that is formed by a strengthened boundary (blue stripe). The graphical representation of interaction between *FZD8* and the super enhancer is shown as a loop below. H-L refers to subtraction High – Low. (F) Hi-C

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

plots of the region surrounding the *FZD8* gene. Top panel: NSD2 Low, bottom panel: NSD2 High. Green arrow identifies the TAD boundary that is strengthened in NSD2 High cells versus Low cells. Black arrow indicates the *FZD8* gene. Circle indicates a new loop between *FZD8* and the boundary.

Figure 5. NSD2 overexpression drives concordant chromatin and transcriptional changes in insulated domains. (A) Scheme illustrating the strategy to identify chromatin and transcriptional changes within TADs (470) or CTCF HiChIP loops (543), that were filtered to have at least one differentially expressed gene, CTCF and H3K27ac peak. (B) Pairwise (2D scatter plots left panel) and three-way (3D scatter plots right panels) comparisons representing significant log2 fold-changes in gene expression, H3K27ac, CTCF, intra-TAD interactions and PC1 values (representing subtraction of NSD2 High and Low levels) within TADs that have at least one significantly differentially expressed gene, CTCF and H3K27ac peak (FDR < 0.05). Concordant increased and decreased changing TADs are colored in red and blue, respectively. TADs that switch from B to A according to HOMER analysis (see STAR Method for details) are highlighted in orange. Pearson correlations are indicated. (C) Pairwise (2D scatter plots left panel) and three-way (3D scatter plots right panels) comparisons representing significant log2 fold-changes of NSD2 High versus Low levels in gene expression, H3K27ac and CTCF within CTCF HiChIP loops that have at least one differentially expressed gene, CTCF and H3K27ac peak (FDR < 0.05). Concordant increased and decreased changing loops are colored in red and blue, respectively. Pearson correlations are indicated.

Figure 6. Significant changes in gene expression and activation of oncogenes occurs in insulated domains. (A) Bar graph showing the proportion of concordant and discordant change in CTCF, H3K27ac and gene expression in TADs and CTCF loops. (B) Scheme showing analysis of changes in CTCF, H3K27ac and gene expression in concordant and discordant TADs and

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CTCF loops. (C) UCSC tracks showing chromatin features and CTCF-mediated loop changes in the region surrounding the *PTPN13* (left panel) and *ETV5* (right panel) genes (*PTPN13* and *ETV5* genes are indicated in red and highlighted with a yellow stripe). H-L refers to subtraction High – Low.

Figure 7. Activation of *SYK* involves increased promoter contacts within a CTCF-mediated loop. (A) Interaction profile of a 4C bait located 27 kb downstream of the SYK promoter (green arrow, bait -27 kb) in a 2.4 Mb region surrounding the *SYK* gene in NSD2 High (red line represents the average between two replicates) and Low (blue line represents the average between two replicates) cells using 4C counts in 20 kb sliding windows. DESeq2 analysis identified significantly different 4C signal in duplicated 4C samples from NSD2 High versus Low cells. Regions with differential interactions are indicated by red and blue dots (p-value < 0.05). (B) UCSC tracks showing chromatin features in the region surrounding *SYK* (*SYK* indicated in red). A graphical representation of interactions from the 4C viewpoint (highlighted by a yellow strip) located 27 kb downstream of the *SYK* promoter is drawn with arcs at the bottom and highlighted by blue stripes. H-L refers to subtraction High – Low. (C) Hi-C plots of the region surrounding *SYK* in NSD2 Low and High cells (left and right panels, respectively).

STAR Methods

Key resource table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CTCF ChIP-grade, from Rabbit, polyclonal	Millipore	Cat# 07-729, lots# 2475698, 2628833, 1350637

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

H3K27ac, Rabbit ChIP grade purified 1mg/ml	Abcam	Cat# ab4729, lots# GR211895-1, GR312658-1
Rad21, Rabbit polyclonal ChIP grade purified 1mg/ml	Abcam	Cat# ab992, lot# GR208564-1
H3K36me2, serum from rabbit, ChIP-grade	Active motif	Cat# 39255, Lot# 8308001
H2Av, Rabbit polyclonal affinity purified	Active motif	Cat# 39716, lot# 08614002
Rabbit IgG control	Abcam	Cat# ab37415, Lot# GR215822-1
Critical Commercial Assays		
NEBNext® Poly(A) mRNA Magnetic Isolation Module	NEB	E7490S
Expand™ Long Template PCR System	Sigma	11681834001
Arima Hi-C kit	Arima	
Deposited Data		
H3K27me3 ChIP-seq	(Popovic et al., 2014)	GSE57977
All other raw sequencing data	This study	Available to public upon publication.
Experimental Models: Cell Lines		
NSD2 High (KMS11 and NTKO) and NSD2 Low (TKO)	(Lauring et al., 2008)	
Oligonucleotides		

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SYK 4C bait primer (DpnII restriction site): AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNNNGAGGGCATTCCC ATTAGATC (NNNNNN: barcode sequence different for each sample)	This paper	N/A
SYK 4C bait primer (Csp6I restriction site): CAAGCAGAAGACGGCATACGAGATA GGTCGCAGTGACTGGAGTTCAGACG TGTGCTCTTCCGATCTtaatctttggataagt ggcc	This paper	N/A
Software and Algorithms		
R	https://www.r-project.org	
Bowtie2	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	(Li et al., 2009)	http://samtools.sourceforge.net/
TopHat2	(Kim et al., 2013)	N/A
Trim Galore!	https://github.com/FelixKrueger/TrimGalore	N/A
UCSC genome browser	https://genome.ucsc.edu/	N/A

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

deepTools/2.3.3	(Ramirez et al., 2016)	N/A
Bowtie 2	(Langmead and Salzberg, 2012)	N/A
BEDTools	(Quinlan and Hall, 2010)	N/A
MACS1.4.2	(Zhang et al., 2008)	N/A
Hi-C bench	(Lazaris et al., 2017)	N/A
GenomicTools	(Tsirigos et al., 2012)	N/A
ICE “correction” algorithm	(Imakaev et al., 2012)	N/A
Crane insulation scores	(Crane et al., 2015)	N/A
Hi-C Pro	(Servant et al., 2015)	N/A
Fit-HiChIP	(Bhattacharyya et al., 2018)	N/A
DESeq2 version 1.4	(Love et al., 2014)	N/A
htseq-counts	(Anders et al., 2015)	N/A
4Cker	(Raviram et al., 2016)	https://github.com/rr1859/R.4Cker
GSEA	(Subramanian et al., 2005)	http://software.broadinstitute.org/gsea/index.jsp
PeaKDEck	(McCarthy and O'Callaghan, 2014)	N/A

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents may be directed to the Lead Contact, Jane Skok (Jane.Skok@nyulangone.org).

Experimental Model and Subject Details

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NSD2 High (one clone of NTKO and KMS11 parental cell line, which is a human myeloma cell line from a female patient) and NSD2 Low cells (two clones of TKO) were obtained from Ben Ho Park (Lauring et al., 2008) who generated the cell lines. Upon reception, the phenotype of the cells was as described (Lauring et al., 2008), and NSD2 High and Low cells harbored expected alterations in NSD2 RNA and protein as confirmed by RNA-seq and Western-blot, respectively. Additional cell authentication was not performed. Cells were maintained in culture as per Ben Ho Park's recommendations. Briefly, cells were grown in RPMI-1640 supplemented with 10% fetal bovine serum, 100 U/mL penicillin and 100 µg/mL streptomycin and subcultured twice a week by dilution. Our PCA revealed that NTKO (two replicates from independent cultures) and KMS11 (one replicate) were identical in terms of gene expression, CTCF and H3K27ac peaks. Only NTKO cells were used for further experiments. For NSD2 low cells, two replicates from independent cultures were used from one clone as well as one replicate from the other clone. Our PCA revealed that TKO clones and replicates were identical in term of gene expression, CTCF and H3K27ac peaks. For downstream experiments, NTKO and KMS11 cells were trypsinized while TKO cells were directly obtained from suspension cultures. Cells were freshly crosslinked for ChIP-seq, Hi-C, Hi-ChIP and 4C (see corresponding sections for detailed protocols), resuspended in RLT buffer for RNA extraction as per the instructions provided by the kit (RNeasy plus kit from QIAGEN), or directly processed for ATAC-seq (see corresponding section for detailed protocol).

Method Details

RNA-seq

RNA was extracted from 5 replicates for each condition, using the RNeasy plus kit from QIAGEN. Poly-adenylated transcripts were positively selected using the NEBNext® Poly(A) mRNA Magnetic Isolation Module following the kit procedure. Libraries were prepared according to the directional RNA-seq dUTP method adapted from

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

http://wasp.einstein.yu.edu/index.php/Protocol:directional WholeTranscript_seq that preserves information about transcriptional direction. Sequencing was performed with Illumina Hi-Seq 2500 using 50 cycles paired-end mode.

ATAC-seq

NTKO and KMS11 cells were trypsinized while TKO cells were directly obtained from suspension cultures. Cell were counted to collect 50,000 cells per replicate. Two different cultures of NTKO cells were used for the NSD2 High condition and one culture from two different clones of TKO for the NSD2 Low condition. The procedure was repeated in three independent days for a total of six replicates for NSD2 High and six replicates for NSD2 Low. The assay was performed as described previously (Buenrostro et al., 2013). Cells were washed in cold PBS and resuspended in 50 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3mM MgCl2, 0.1% IGEPAL CA-630). The tagmentation reaction was performed in 25 µl of TD buffer (Illumina Cat #FC-121-1030), 2.5 µl Nextera Tn5 Transposase, and 22.5 µl of Nuclease Free H2O at 37°C for 30 min. DNA was purified on a column with the Qiagen Mini Elute kit, eluted in 10 µl H2O. Purified DNA (10 µl) was combined with 10 µl of H2O, 2.5 µl of each primer at 25 mM and 25 µl of NEB Next PCR master mix. DNA was amplified for 5 cycles and a monitored quantitative PCR was performed to determine the number of extra cycles needed as per the original ATAC-seq protocol (Buenrostro et al., 2013). DNA was purified on a column with the Qiagen Mini Elute kit. Samples were quantified using Tapestation bioanalyzer (Agilent) and the KAPA Library Quantification Kit and sequenced on the Illumina Hi-Seq 2000 using 50 cycles paired-end mode. The six replicates were sequenced independently and three replicates were pooled together during the data processing in order to get two replicates with sufficient sequencing depth for downstream analysis (see ATAC-seq processing data for details).

ChIPmentation

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cells were fixed in culture medium within 1% formaldehyde at RT for 10 minutes and quenched with 0.125 M glycine. Pellets were washed twice with ice-cold PBS, snap-frozen in liquid nitrogen and stored at -80°C. ChIP-seq was performed as per the original ChIPmentation protocol (Schmidl et al., 2015) in triplicate for CTCF and H3K27ac, and in duplicate for H3K36me2 and Rad21. Briefly, chromatin was lysed during a 10 min rotation in the cold room in 350 µl of lysis buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH8.0 NaOH, 0.5 mM EGTA pH8.0 NaOH, 0.1% sodium deoxycholate, 0.5% N-lauroysarcosine). Lysates were sonicated using Bioruptor (Diagenode) (15 cycles 30 sec ON, 30 sec OFF, an agarose gel was run to make sure that the sonicated DNA smear was in the range of 100-700bp). Triton X-100 1% finale were added and the samples were centrifuged 5 min at 16000 rcf at 4°C. Supernatant was collected. Antibody was combined with protein A magnetic beads for one hour at room temperature and added to chromatin. For CTCF, H3K27ac and Rad21 and IgG as negative control, 10 µl of antibody (Millipore 07-729, Abcam ab4729, Abcam ab992, Abcam ab37415, respectively) was added to 50 µl of protein-A magnetic beads (Dynabeads) and added to the sonicated chromatin from 10 million cells per immunoprecipitation. For H3K36me2, internal spike-in was added for normalization as previously described (Oksuz et al., 2018). Briefly, per immunoprecipitation, 1 µl of H3K36me2 antibody and 0.1 µl of Drosophila-specific H2Av antibody were added to 10 µl of protein-A magnetic beads and added to 100 µg of human sonicated chromatin supplemented with 2 µg of drosophila sonicated chromatin. Of note, chromatin was quantified with Nanodrop at 260 nm. Immunoprecipitation was performed for 3 to 6 hours rotating in the cold room, then washes and tagmentation were performed as per the original ChIPmentation protocol (Schmidl et al., 2015). Briefly, beads were washed twice with 500 µl cold low-salt wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA pH8.0 NaOH, 0.1% SDS, 1% triton X-100), twice with 500 µl cold LiCl-containing wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA pH8.0 NaOH, 1% triton X-100, 0.7% sodium deoxycholate), and twice with 500 µl cold 10 mM cold Tris-Cl, pH 8.0, to remove detergent, salts and EDTA. Subsequently, beads were resuspended in 25 µl of the

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

freshly prepared tagmentation reaction buffer (10 mM Tris-HCl, pH 8.0, 5 mM MgCl2, 10% dimethylformamide) and 1 µl Tagment DNA Enzyme from the Nextera DNA Sample Prep Kit (Illumina) and incubated at 37°C for 1 min in a thermocycler. Following tagmentation, the beads were washed twice with 500 µl cold low-salt wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA pH8.0 NaOH, 0.1% SDS, 1% triton X-100), and twice with 500 µl cold Tris-EDTA-Tween buffer (0.2% tween, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). Chromatin was eluted and decrosslinked by adding 70 µl of freshly prepared elution buffer (0.5% SDS, 300 mM NaCl, 5 mM EDTA pH 8.0, 10 mM Tris-HCl pH 8.0) and 2 µl of proteinase K at 10 mg/ml for 2 hours at 55°C and overnight incubation at 65°C. Supernatant was kept and to recover as much DNA as possible, beads were washed with an additional 30 µl of elution buffer and combined supernatant was incubated an additional hour at 55°C. DNA was purified on a column with the Qiagen Mini Elute kit. Purified DNA (20 µl) was combined with 2.5 µl of each primer at 25 mM and 25 µl of NEB Next PCR master mix and was amplified as per the ChIPmentation protocol (Schmidl et al. 2015) in a thermomixer with the following program: 72°C for 5 min; 98°C for 30 s; 14 cycles of 98°C for 10 s, 63°C for 30 s and 72°C 30 s; and a final elongation at 72°C for 1 min. DNA was purified using two consecutive rounds of SPRI AMPure XP beads: the first one with a beads-to-sample ratio of 0.6:1 to remove potential fragments larger than 700 bp (supernatant kept) and the second one with a beads-to-sample ratio of 1:1 to remove potential primer dimers (beads kept), and eluted in 20 µl of H2O. Samples were quantified using Tapestation bioanalyzer (Agilent) and KAPA Library Quantification Kit and sequenced with Illumina Hi-Seq 2500 using 50 cycles paired-end mode (CTCF and H3K27ac) or single-end mode (Rad21 and H3K36me2).

Hi-C

Hi-C was performed in duplicate, from 0.5 to 1 million cells fixed in culture medium within 1% formaldehyde at RT for 10 minutes and quenched with 0.125M glycine. Hi-C samples were

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

processed using the Arima Hi-C kit as per the kit protocol, and sequenced with Illumina Hi-Seq 2500 using 50 cycles paired-end mode.

CTCF HiChIP

HiChIP was performed in duplicate. Cells were fixed in culture medium with 1% formaldehyde at RT for 10 minutes and quenched with 0.125 M glycine. Pellets were washed twice with ice-cold PBS, snap-frozen in liquid nitrogen and stored at -80°C. HiChIP was performed with 15 million cells, as per the original protocol (Mumbach et al., 2016). Cells were then lysed in 500 µl ice-cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630, protease inhibitor cocktail (Roche complete, EDTA-free)) rotating at 4°C for 30 minutes. Cell pellets were collected, washed once in 500 µl ice-cold lysis buffer and then incubated in 100 µl 0.5% SDS at 62°C for 10 min. SDS was quenched by adding 285 µl of H2O and 50 µl of Triton X-100 10%, and incubating at 37°C for 15 min. Chromatin was then digested by adding 50 µl of NEBuffer 2 10X and 350 units of MboI (NEB R0147M) at 37°C for 2 hours while rotating at 950 rpm. MboI was inactivated by incubating the samples 20 minutes at 62°C. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 1.5µl 10 mM dCTP, 1.5µl 10 mM dGTP, 1.5µl 10 mM dTTP, 37.5µl 0.4 mM biotin-14-dATP (Life Technologies 19524-016), and 10 µL 5U/µl Klenow (DNA polymerase I large fragment, NEB M0210L) were added to each tube, and incubated for 60 minutes at 37°C. Ligation mix was added to the samples (150 µl 10x ligation buffer (NEB B0202S), 7.5 µl 20mg/ml BSA (NEB B9001S), 150 µl Triton X-100 10%, 10 µl 400U/µl T4 DNA ligase (NEB M0202S), and 655.5 µl H2O) for 4 hours at RT with rotation. Following ligation, nuclei were pelleted and resuspended in 350 µl cold Nuclei Lysis Buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS, and 1x Protease Inhibitors) with incubation rotating in the cold room for 10 min. Samples were sonicated on the bioruptor for 15 min (an agarose gel was performed to make sure that the sonicated DNA smear was in the range of 250-600bp), supplemented with 1% Triton X-100 and centrifuged 5 min at 16000 rcf at 4°C. CTCF antibody (5 µl, Millipore 07-729) was combined with

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

50 µl of protein-A magnetic beads (Dynabeads) for one hour at room temperature, and added to sonicated chromatin from 15 million cells. Immunoprecipitation was performed by overnight incubation rotating in cold room and washes were performed. Briefly, beads were washed twice with 500 µl cold Low-salt wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA pH8.0 NaOH, 0.1% SDS, 1% triton X-100), twice with 500 µl cold high-salt wash buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM EDTA pH8.0 NaOH, 0.1% SDS, 1% triton X-100), and twice with 500 µl cold LiCl-containing wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA pH8.0 NaOH, 1% NP-40, 1% sodium deoxycholate). Chromatin was eluted and decrosslinked by adding 100 µl of freshly prepared elution buffer (0.5% SDS, 300 mM NaCl, 5 mM EDTA pH 8.0, 10 mM Tris-HCl pH 8.0) and 10 µl of proteinase K at 10 mg/ml for 45 min at 55°C and at least 1.5 hour at 67°C. DNA was purified on kept a column with the Qiagen Mini Elute kit, eluted in 12 µl of H2O, and quantified using Qubit. Of note, we obtained between 3 to 8 ng of DNA for CTCF HiChIP from 15 million cells. To enrich for ligation events, 5 µl of Streptavidin C-1 beads were washed in Tween Wash Buffer (TWB, 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA pH 8.0, 1M NaCl, 0.05% Tween-20), resuspended in 10 µl of 2X Biotin Binding Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 2 M NaCl), added to 10 µl of the samples and incubated at room temperature for 15 min with rotation. Samples were then washed twice in TWB with 2 min incubation at 55°C shaking. For tagmentation, beads were washed twice in 100 µl of freshly prepared tagmentation reaction buffer (10 mM Tris-HCl, pH 8.0, 5 mM MgCl2, 10% dimethylformamide), and resuspended in 25 µl of the tagmentation reaction buffer and 1 µl Tagment DNA Enzyme from the Nextera DNA Sample Prep Kit (Illumina) and incubated at 55°C for 10 min in a thermocycler with interval shaking. Beads were resuspended in 50 mM EDTA and incubated at 50°C for 30 min to quench the transposase reaction. This was followed by two washes in 50 mM EDTA incubated at 50°C for 3 min, two washes in Tween Wash Buffer incubated at 55°C for 2 min, and one wash in 10 mM Tris-HCl pH 7.5. Beads were resuspended in 50 µl of PCR master mix (1 µl of each primer at 25 mM, 25 µl of NEB Next PCR master mix and 23 µl of H2O) and DNA was amplified in a thermomixer with the

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

following program: 72°C for 5 min; 98°C for 30 s; 10 cycles of 98°C for 10 s, 63°C for 30 s and 72°C for 1 min. DNA was purified using two consecutive rounds of SPRI AMPure XP beads: the first one with a beads-to-sample ratio of 0.6:1 to remove potential fragments larger than 700 bp (supernatant kept) and the second one with a beads-to-sample ratio of 0.18:1 to keep fragments greater than 300 bp (on beads), and eluted in 15 µl of H2O. Samples were quantified using the Tapestation bioanalyzer (Agilent) and the KAPA Library Quantification Kit and sequenced with Illumina Hi-Seq 2500 using 50 cycles paired-end mode.

4C-seq

4C-seq was performed in duplicate and analyzed as previously described (Raviram et al., 2016; Rocha et al., 2016) with minor changes. 10 million cells were fixed in 2% formaldehyde for 10 minutes at room temperature and quenched with glycine (0.125 M final concentration). Nuclei were isolated in lysis buffer (50mM Tris-HCl ph7.5, 150mM NaCl, 5mM EDTA, 0.5% NP-40, 1% TX-100 containing 1X Roche complete Mini protease inhibitors) and dounced 40 times on ice. Nuclei were resuspended in 360 µl H2O and 60 µl warm 10X NEB DpnII buffer. They were permeabilized using 15 µl 10% SDS for 60min at 37°C and then 150 µl 10% Triton X-100 for 60min at 37°C. Chromatin was digested using 500U DpnII (NEB) overnight at 37°C while shaking, and the digestion was repeated with an additional 250U of enzymes for 8 hours, meanwhile digestion was determined by gel electrophoresis. Enzyme was deactivated at 65°C for 20 min. Chromatin samples were divided in 3 tubes, diluted and ligated by adding H2O up to 1.2 ml, 133 µl T4 ligase buffer 10X and 6000U total NEB T4 DNA Ligase (M0202M) per tube and incubating at 16°C overnight while shaking. Ligation efficiency was checked by gel electrophoresis. Chromatin was de-crosslinked with proteinase K overnight at 65°C, and treated with RNase A at 37°C for 1 hour. DNA was extracted by Phenol:Choroform and precipitated with Ethanol. Purified DNA was digested with 50U NEB Csp6I overnight at 37°C while shaking, and digestion was determined by gel electrophoresis. Enzyme was deactivated at 65°C for 20 min. DNA

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

circularization was performed using 4000U NEB T4 DNA Ligase overnight at 16°C. A total of 1 µg DNA was amplified per sample with inverse PCR primers containing Illumina forward and reverse sequencing adapters (see Key resources for sequences). PCR was performed using Expand™ Long Template PCR System (Sigma) with the following thermocycler program: 94°C for 2 min; 94°C for 15 sec; 53-55°C for 1 min; 68°C for 2.30 min; repeat for 29 cycles; 68°C for 7 min; hold at 4°C. 4C-seq libraries were size-selected on gel to remove any potential primer dimers and fragments above 700 bp, then quantified using RT-PCR (KAPA Biosystems) and sequenced using 50bp single-end on Illumina HiSeq 2500.

Quantification and Statistical Analysis

RNA-Seq Data processing and quality control

Paired-end reads were mapped to the hg38 genome using TopHat2 (Kim et al., 2013) (parameters:–no-coverage-search–no-discordant–no-mixed–b2-very-sensitive–N 1). Bigwigs were obtained for visualization on individual as well as merged bam files using Deeptools/2.3.3 (Ramirez et al., 2016) (parameters bamCoverage --binSize 1 --normalizeUsing RPKM). Counts for Refseq genes were obtained using htseq-counts (Anders et al., 2015). Principal Component Analysis was performed using R to check the reproducibility of replicates (See **Supplementary Figure S1A**). DESeq2 version 1.4 (Love et al., 2014) was used to normalize expression counts and get differentially expressed genes (absolute log2 fold-change> 1 and FDR < 0.01). Gene Set Enrichment Analysis of gene expression in NSD2 High versus Low cells were performed using GSEA (Subramanian et al., 2005) desktop application on normalized reads counts for each replicate from the 26586 protein coding genes of hg38 genome.

ATAC-seq Data processing and quality control

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Reads were aligned to hg38 genome with Bowtie2 (Langmead and Salzberg, 2012) (parameters: –no-discordant -p 12 –no-mixed -N 1 -X 2000). Potential PCR duplicates were removed from the reads with Picard-tools. ATAC-seq peaks were called with PeaKDEck (McCarthy and O'Callaghan, 2014) (parameters: -sig 0.0001 -PVAL ON). Bigwigs were obtained for visualization on individual as well as merged bam files using Deeptools/2.3.3 (parameters bamCoverage --binSize 1 --normalizeUsing RPKM).

4C-seq Data processing and quality control

Processing of 4C-seq data was performed using 4Cker pipeline (Raviram et al., 2016). Briefly, mapping was performed using Bowtie2 to a reduced genome consisting of all unique 24-nt-long regions surrounding DpnII sites from the human reference genome (hg38), allowing for zero mismatches. For comparison between conditions, DESeq2 version 1.4 (Love et al., 2014) with default parameters was used to normalize total read count per window between samples and to identify the windows with significant 4C signal differences, using an FDR-adjusted p-value cutoff of 0.05.

ChIP-seq Data processing and quality control

Reads were aligned to hg38 genome with Bowtie2 (Langmead and Salzberg, 2012) (parameters: –no-discordant -p 12 –no-mixed -N 1 -X 2000). Ambiguous reads were filtered to use uniquely mapped reads in the downstream analysis. PCR duplicates were removed using Picard-tools (version 1.88). Bigwigs were obtained for visualization on individual as well as merged bam files using Deeptools/2.3.3 (Ramirez et al., 2016) (parameters bamCoverage --binSize 1 --normalizeUsing RPKM; or bamCompare --verbose --binSize 25 --ratio subtract --scaleFactorsMethod SES for subtraction files). For H3K36me2 ChIP-seq, an internal Drosophila spike-in was added. Reads were aligned to dm6 genome with Bowtie2 (parameters: –no-discordant -p 12 –no-mixed -N 1 -X 2000). Bigwigs were created after normalization with the

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spike-in Drosophila read counts. Heatmaps and average profiles were performed on merged bigwig files using Deeptools/2.3.3. For CTCF, H3K27ac and Rad21 ChIP-seq, MACS version 1.4.2 (Zhang et al., 2008) was used to call peaks (parameters: -p 1e-6 -g hs -B --single-profile). For CTCF and H3K27ac ChIP-seq, a reference list of peaks coordinates was created containing all the peaks present in any replicate, and merging overlapping peaks using Bedtools merge –i. Counts for the reference list of peak coordinates were obtained using htseq-counts (Anders et al., 2015). PCA was performed using R to check the reproducibility of replicates (See **Supplementary Figure S2A and S2D)**. DESeq2 version 1.4 (Love et al., 2014) was used to normalize read counts and get differential peaks (absolute log2 fold-change> 1 and FDR < 0.01).

Hi-C Data processing and quality control

Processing

HiC-Bench (Lazaris et al., 2017) was used to align and filter the Hi-C data, identify TADs, and generate Hi-C heatmaps. To generate Hi-C filtered contact matrices, the Hi-C reads were aligned against the human reference genome (hg38) by bowtie2 (Langmead and Salzberg, 2012) (version 2.3.1) (Settings: --very-sensitive-local –local). Mapped read pairs were filtered by the GenomicTools (Tsirigos et al., 2012) tools-hic filter command integrated in HiC-bench for known artifacts of the Hi-C protocol. The filtered reads include multi-mapped reads ('multihit'), read-pairs with only one mappable read ('single sided'), duplicated read-pairs ('ds.duplicate'), low mapping quality reads (MAPQ < 30), read-pairs resulting from self-ligated fragments, and short-range interactions resulting from read-pairs aligning within 25kb ('ds.filtered'). For the downstream analyses, all the accepted intra-chromosomal read-pairs ('ds.accepted intra') were used. The Hi-C filtered contact matrices were corrected using the ICE "correction" algorithm (Imakaev et al., 2012) built into HiC-bench. TADs and boundaries were identified using the Crane method (Crane et al., 2015) at 40 kb bin resolution with an insulating window of 500kb. HiC heatmaps for regions

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of interest were generated using the ICE corrected contact matrices through the 'hic-plotter-diff' pipeline step integrated in HiC-bench.

Quality Control and TAD / Boundary Stats

Quality assessment analysis shows that the total numbers of reads in the samples ranged from ~120 million to ~185 million (Supplementary Figure S4A). The percentage of reads aligned was always over 98% in all samples. The proportion of accepted reads ('ds-accepted-intra' and 'ds-accepted-inter') were in the range of ~46 - 47%.

The number of TADs identified by the Crane method across replicates ranged between 3460 and 3532 from which ~500 TADs (14%) showed sizes smaller or equal to 80kb in each replicate. We removed such short length TADs from the following metrics analysis, considering this approach more representative. The distribution of TAD sizes showed that 85% of the TADs ranged between 160 kb and 1.04 Mb with a median of 480 kb in both NSD2 High and Low conditions. The mean TAD sizes were 573 kb and 591 kb for NSD2 High and Low conditions, respectively (Supplementary Figure S4G).

R (prcomp, scale=TRUE and center=TRUE) was used to perform a PCA on the Hi-C datasets using the "ratio" insulation scores produced by HiC-Bench (bins = 40kb) after ICE correction (Supplementary Figure S4B).

DOWNSTREAM ANALYSIS: TADs, Boundaries, and Compartments

Screening of Potentially Altered Boundaries

The Hi-C downstream analysis involved a genome-wide screening of TAD boundary insulation changes in NSD2 High cells based on boundary insulation scores (ratio index) and CTCF/Rad21

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

binding enrichment. Individual cases of potentially altered boundaries were confirmed by visual inspection of HiC heatmaps.

Mean Boundary Insulation Scores (Ratio index)

To assess and compare boundary strength alteration in NSD2 High versus Low cells, we included the calculation of the Mean Boundary Score (MBS) for every boundary identified in the NSD2 Low condition (reference boundaries). For this purpose, we used the 'by group' NSD2 Low TADs identified by the Crane method. HiC-Bench performs a 'by group' analysis by merging the DNA interaction information of all the replicates of the same condition, in this case, to identify TADs per condition.

We calculated the MBS as the arithmetic mean of the 'ratio' insulation scores inside the reference boundary coordinates being assessed. HiC-Bench calculates one ratio score per bin (40kb) as explained in Lazaris et al. (2017). As a result, there are generally multiple insulation scores per TAD boundary identified by the Crane algorithm. The number of bins inside boundaries showed a median number of 7 and, accordingly, the boundary median size was 280 kb. The MBS was used to calculate NSD2 High MBS logFC values with respect to the NSD2 Low condition. A differential analysis on the ratio insulation scores inside each boundary was also performed. An unpaired t-test (two-sided) was used by pooling all the ratio insulation scores inside the reference boundary coordinates and adjusting with FDR correction.

CTCF and Rad21 Occupancy in Boundaries: Integration with Insulation Data

The CTCF and Rad21 peaks were mapped to the boundaries to integrate the boundary insulation data obtained with the CTCF and Rad21 binding data. We assigned a peak to a boundary if the peak overlapped with the boundary region (> 0 bp). An extension of the boundary region by 1 bin (40 kb) on either side of the boundary was considered. The CTCF/Rad21 signal of all the peaks

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

assigned to a boundary were aggregated and then NSD2 High versus Low logFC values were calculated. Significant changes in global CTCF/Rad21 occupancy within the boundaries were calculated using a two-sided unpaired t-test by pooling all the CTCF/Rad21 peak intensities assigned in the boundary.

A ranked-table of boundary coordinates with the insulation and CTCF/Rad21 metrics was created. To generate Hi-C heatmaps, the best-ranked boundary cases were selected by taking into account unidirectional significant fold changes in MBS, CTCF and Rad21 (FDR < 0.01). In this approach, we assumed that there would be a positive correlation between boundary insulation scores and CTCF/Rad21 binding in boundaries. In addition, we screened each boundary candidate and the adjacent TADs and boundaries in search of significant deregulated genes, gain / loss of interactions and significant changes in H3K27ac, CTCF and other features.

Compartments

Compartment analysis was carried out using the Homer pipeline (v4.6) (Crane et al., 2015). Homer performs a principal component analysis of the normalized interaction matrices and uses the PC1 component to define regions of active (A compartments) and inactive chromatin (B compartments). HiC filtered matrices were given as input to run Homer with default parameters (50kb resolution). To confirm the proper sign of the A and B compartment, we used the --active parameter to input peaks of the active mark H3K27ac. We use Homer to compare the interaction profiles in both experiments and calculate a correlation. If one region interacts similarly with other regions in both conditions, then the correlation will be high. On the other hand, the correlation will be low if a locus interacts with different regions in both conditions. Using Homer's getHiCcorrDiff.pl, we compared the interaction profiles of both conditions and obtained a correlation difference to identify stable and switching compartments. Altered compartments are named as AB and BA for regions switching from A to B and B to A, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Intra-TAD interactions

To assess statistically significant intra-TAD interactions, we used an algorithm developed by Andreas Kloetgen. As a first step, the algorithm identifies overlapped or positionally consistent TADs (common TADs). This approach establishes a minimum TAD length parameter (default: 10 bins) and extends either side of the TAD by 3 bins (+/-120 kb in 40kb resolution). TADs across two samples are considered positionally consistent if their boundaries are as close as 3 bins. The boundaries of the common TAD are then set to those which yield the largest TAD. The set of common TADs between any two samples s_1 and s_2 is denoted as T. In the next step, a paired two-sided t-test is performed on each single interaction bin within each common TAD between the two samples. It calculates the difference between the average scores of all interaction intensities within such TADs. A multiple testing correction by calculating the false-discovery rate per common TAD (using the R function p.adjust with method="fdr") is also calculated.

$$(1)\ TAD\ interactions\ change(t) = \left(\frac{\sum_{i \in I_t} s_{2_i}}{\#I_t}\right) - \left(\frac{\sum_{i \in I_t} s_{1_i}}{\#I_t}\right)$$

for each t $\in$ T, and I_t being all intra-TAD interactions for TAD t.

We classified the common TADs in terms of Loss, Gain or Stable intra-TAD interactions by using FDR < 0.1 and absolute TAD interactions change > 0.3. A minimum common TAD length of 200 kb was considered in the intra-TAD interactions differential analysis (5 bins). We used the intra-TAD interactions fold change of each common TADs with a TAD length higher or equal to 80 kb (2 bins) in the Local Analysis in Common TADs (See '**Data Integration: Local Analysis in Common TADs' section**).

Data Integration: Global Analysis.

AB and BA Compartment Changes.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To show the correlation between the different measurements and the compartment changes, we mapped the peaks obtained in H3K27ac2, CTCF, RAD21 and ATAC-seq to the AB, BA and Stable regions. We used the peak intensity values to calculate the peak intensity fold change between NSD2 High and Low and the mean fold change of all the peaks assigned to a compartment region. Similarly, in the case of RNA-seq data, genes were mapped to the compartment regions and the mean fold change of all the genes assigned to a region was calculated by using the DESeq2 fold change data.

We assigned a peak or gene to a compartment region when the complete peak or gene coordinate was found inside the compartment coordinates. For sparse measurements as in H3K36me2 and H3K27me3 chromatin marks, we used HTSeq to obtain read counts on the AB, BA and Stable regions. Next, we used DESeq2 to normalize the read counts across NSD2 High and Low replicates and to obtain fold change values.

To show the correlation between the different measurements and the compartment changes, we used the mean fold change values to generate boxplots. Statistical significance was assessed using a paired two-sided Wilcoxon rank-sum test.

Intra-TAD Interactions. Gain, Loss and Stable TADs.

We used the same method as described in the previous subsection to integrate the H3K27ac2, CTCF, RAD21, ATAC-seq, RNA-seq data, H3K36me2, H3K27me3 measurements with the different intra-TAD interactions subgroups described above (Gain, Loss and Stable TADs).

Data Integration: Local Analysis in Common TADs.

CTCF binding, H3K27ac and RNA expression.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To compute the total number of differential changes in CTCF binding, RNA expression, and H3K27ac mark that fall within common TADs. We overlapped all CTCF peaks, genes, and H3K27ac peaks with common TADs. Volcano plots were generated using the log2 fold changes and –log10 (p-value) for all three data types (CTCF-ChIP, RNA-seq, H3K27ac-ChIP). Overlapping features with log2 fold change greater than 1 and a FDR less than 0.01 were colored red, while overlapping features with log2 fold change smaller than -1 and a FDR smaller than 0.01 were colored in blue, and all others were colored in black.

To assess the correlation of CTCF binding, H3K27ac, and RNA expression inside each Common TAD, we assigned a peak to each Common TAD when the complete peak coordinate was found inside the TAD coordinates. Genes were assigned to common TADs when the promoter overlapped with each TAD (overlap > 1 bp). Common TADs lacking either CTCF, H3K27ac, or RNA expression features were filtered in the three-way analysis (3D scatter-plots). The mean fold change of each feature (CTCF, H3K27ac and RNA expression) inside each Common TAD was computed by two methods. One method considered only the differential peaks / genes found inside a Common TAD (FDR < 0.05) in the mean fold change calculation (filtered analysis). The second method considered all the peaks / genes assigned to a Common TAD (unfiltered analysis).

Intra-TAD Interactions and Compartment Alteration.

One mean log fold change value for each feature (CTCF, H3K27ac and RNA expression) was assigned to every Common TAD, together with the intra-TAD interactions fold change value previously calculated (see '**Intra-TAD interactions'** subsection). Compartment alteration was also assessed by calculating the mean PC1 value of each Common TAD (by 50kb bins) and computing the PC1 mean difference between conditions (NSD2 High – NSD2 low). A positive value of the PC1 mean difference indicates that in the NSD2 High condition, that Common TAD

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

had become more active. The higher the PC1 mean difference the stronger the compartment alteration change. To confirm if the PC1 mean difference correlated with real compartment changes, we looked at whether the significant AB and BA regions identified by Homer overlapped with the Common TADs. We then computed the overlap length. One common TAD was assigned as an AB or BA switching region when the overlap length of the BA or AB in a TAD was higher or equal to 100kb (2 PC1 bins). Common TADs assigned to a BA switching region were colored orange in **Figure 5C**. The median overlap lengths in the three-way concordant common TADs were 400 kb and 450 kb for common TADs assigned to B to A and A to B regions, respectively.

To show the intra-TAD association of all the five features (CTCF, H3K27ac, RNA expression, Intra-TAD interactions and PC1 mean difference) we classified the Common TADs in 'positive', 'negative' or 'no correlated' groups, by taking into account the direction of CTCF, H3K27ac and RNA expression (positive correlation). The Pearson Correlation Coefficient (PCC) was calculated by using R ('Cor' and 'pairs' command). 3D plots were also generated using R ('plot_ly' command of the 'plotly' library).

HiChIP Data processing and quality control

HiChIP paired-end reads were aligned to hg38 genome using the HiC-Pro pipeline (Servant et al., 2015). Default settings were used to remove duplicate reads, assign reads to MboI restriction fragments, filter for valid pairs, and generate 10kb binned interaction matrices. FitHiChIP (Bhattacharyya et al., 2018) was used to identify statistically significant chromosomal interactions from the CTCF HiChIP experiments, using a bin size of 10kb. FitHiChIP allows users to use a reference list of peaks from ChIP-seq data. Therefore, we used the ChIP-seq data from NSD2 high and low cells as described above. Statistically significant HiChIP interactions ($q<0.01$) were called separately in each experiment with a minimum distance of 20kb and a maximum distance of 3MB. The UseP2PBackgrnd parameter was set to 1 for estimating contact probability

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

background only from peak-to-peak interactions (i.e., stringent mode). Statistical significance was assigned for all HiChIP interactions that involve at least one anchor with a CTCF ChIP-seq peak (peak-to-all interactions). This resulted in 8,651 statistically significant CTCF peak-to-all interactions in NSD2 High and 4,914 CTCF peak-to-all interactions in NSD2 low. Common CTCF high confidence loops in NSD2 High and NSD2 Low were identified by overlapping loop anchors on both sides. The median size for common loops was calculated to be 210 kb.

HiChIP Data integration with CTCF ChIP-seq, RNA-seq and H3K27ax ChIP-seq

To compute the total number of differential changes in CTCF binding, RNA expression, and H3K27ac marks that fall within common CTCF HiChIP loops, we overlapped all CTCF peaks, genes (5kb upstream of TSS), and H3K27ac peaks within common HiChIP loops (full loop plus anchors). Volcano plots were generated using the log2 fold changes and –log10 (p-value) for all three data types (CTCF-ChIP, RNA-seq, H3K27ac-ChIP). Overlapping features with log2 fold change greater than 1 and FDR less than 0.01 were colored red, while overlapping features with log2 fold change less -1 and a FDR less than 0.01 were colored in blue, and all others were colored in black.

To assess the correlation of CTCF binding, H3K27ac, and RNA expression inside each Common CTCF Loops, we applied the same approach used to assess the correlation of these features in Common TADs. First, all CTCF peaks were overlapped with all the common loops, and the mean log2 fold change for all peaks found within a common loop were calculated. The same approach was applied to calculate the mean log2 fold change for all genes and for all H3K27ac peaks within each common loop.

To correlate the mean log2 fold changes of all three features, we considered only the common loops that had at least one overlapping differential CTCF and H3K27ac peak as well as one gene

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expression change. Pairwise correlation plots were generated for the three comparisons, so that each individual point in the scatterplot represents a common CTCF loop. Loops that had positive mean log2 fold change for all three features were colored in red, while loops that had negative log2 fold change for all three features were colored in blue.

To determine if there is a correlation between the significant alterations of CTCF, RNA expression, and H3K27ac, we performed the same analysis, but first filtered for significant changes in CTCF, RNA expression, and H3K27ac peaks, with a FDR of less than 0.05 and an absolute value of the log2 fold change greater than 1. Next, the mean log2 fold change was computed for all the differential CTCF, gene expression changes, and H3K27ac peaks that overlapped common HiChIP loops. Pearson correlation was calculated for all three comparisons.

Supplemental Information

Figure S1. RNAseq analysis in NSD2 High versus Low cells, related to Figure 1. (A) PCA of RNA-seq replicates for NSD2 High and Low cells. (B) Gene Set Enrichment Analysis of gene expression in NSD2 High versus Low cells on normalized genes reads counts for each replicate.

Figure S2. Transcription factor motifs identified in differential H3K27Ac peaks, related to Figures 1 and 2. (A) PCA of H3K27ac ChIP-seq replicates for NSD2 High and Low cells. (B) Identification of super-enhancers in NSD2 Low (upper panel) and High (lower panel) using H3K27ac with 'ROSE' (rank ordering of super-enhancers (Whyte et al., 2013)). (C) Transcription factor motifs identified in increased (1650) and decreased (303) H3K27ac peaks using TRAP, related to Figure 1G. Motifs also presents in Figure 1G are in bold. Motifs found in increased (red) and decreased (blue) H3K27ac peaks (-log10 p-value). (D) PCA of CTCF ChIP-seq replicates for NSD2 High and Low cells.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S3. Chromatin landscape of the regions surrounding CTCF and H3K27ac changes, related to Figure 2. Heatmaps (A) and average profiles (B) of H3K27ac, H3K36me2, H3K27me3, ATAC-seq, Rad21 and CTCF signal at the top 1000 increased, 100 stable and 1000 decreased H3K27ac peaks in NSD2 High versus Low cells. Heatmaps (C) and average profiles (D) of CTCF, H3K36me2, H3K27me3, ATAC-seq, Rad21 and H3K27ac signal at top 1000 increased, 100 stable and 1000 decreased CTCF peaks in NSD2 High versus Low cells. Top 1000 increased, 100 stable and 1000 decreased CTCF and H3K27ac peaks are based on Log2 Fold changes in reads counts in NSD2 High versus Low cells. Peaks are ranked by H3K27ac (A) and CTCF (C) signal in NSD2 High cells.

Figure S4. Hi-C analysis in NSD2 High and Low cells, related to Figure 3 and Figure 4. (A) Histogram representing read filtering of Hi-C replicates processed using Hi-C bench (Lazaris et al., 2017). “Ds accepted-intra” and “ds accepted inter” reads were retained for further analysis. (B) PCA of Hi-C replicates for NSD2 High and Low cells using the “ratio” insulation scores produced by HiC-Bench (bins = 40kb) after ICE correction. Histograms representing the total size (C) and average size (D) of A, B and switching compartments in NSD2 High and Low cells. (E) Bar plot representing the proportion of A (red) and B (blue) compartments in the genome for NSD2 High and Low cells. (F) Histogram representing the number of TADs for each replicate of NSD2 High and Low cells. (G) Boxplots representing the distribution of TAD sizes (in megabases, Mb), for each replicate of NSD2 High and Low cells.

Figure S5. CTCF and H3K27ac peaks participate in new or strengthened chromosomal loops, related to figure 4. (A) UCSC tracks showing chromatin features in the region surrounding the *FGF13* gene (*FGF13* gene indicated in red and location highlighted by a yellow stripe) and the new contacts that are formed by stronger CTCF and Rad21 binding (blue stripes and shown as a loop below). H-L refers to subtraction High – Low. (B) Hi-C plots of the region surrounding

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the *FGF13* gene. Top panel: NSD2 Low, bottom panel: NSD2 High. Black arrow indicates the *FGF13* gene. Circle indicates a loop between *FGF13* and surrounding increased CTCF and Rad21 regions highlighted by a blue strip in panel (B). (C) UCSC tracks showing chromatin features in the region surrounding the *KRAS* gene (*KRAS* gene indicated in red and highlighted by a yellow stripe) and a downstream super enhancer (indicated by a green arrow and highlighted by a blue stripe). Left and right boundaries surrounding intra-TAD interaction gain are highlighted by blue stripes and indicated by green arrows labelled as boundary '1' and '2', respectively. A graphical representation of the interaction between *KRAS* and the super enhancer is shown as a loop below. H-L refers to subtraction High – Low. (D) Hi-C plots of the region surrounding the *KRAS* gene in NSD2 Low versus High cells (top and bottom panels, respectively). Black arrow indicates the location of *KRAS*. Green arrows indicate the super enhancer (also labelled as 'SE'), and left and right boundaries surrounding intra-TAD interaction gain as labelled as boundary '1' and '2', respectively. Circle indicates a loop between *KRAS* and the super enhancer.

Figure S6. NSD2 overexpression drives concordant chromatin and transcriptional changes in insulated domains, related to Figure 5. (A) Bar plot (left panel) and Volcano plots (right panel) showing significant NSD2-mediated changes in CTCF binding (left panels), gene expression (middle panels) and H3K27ac signal (right panels) within TADs (top panels) or CTCF HiChIP loops (bottom panels). Increases of log2 fold change >1 are shown in red and decreases, log2 fold change <-1 are shown in blue (FDR <0.01). (B) Percentage of significantly differential features in TADs. (C) Density plot of CTCF HiChIP loop sizes.

Figure S7. NSD2 overexpression drives concordant chromatin and transcriptional changes in insulated domains, related to Figure 5. (A) Pairwise (2D scatter plots left panel) and three-way (3D scatter plots right panels) comparisons representing log2 fold-changes of gene expression, H3K27ac, CTCF, intra-TAD interactions and PC1 values (representing subtraction of

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NSD2 High and Low levels within TADs) in NSD2 High versus Low cells. Concordant increased and decreased changing TADs are colored in red and blue, respectively. TADs that switch from B to A as determined by HOMER analysis (see method for details) are highlighted in orange. Pearson correlations are indicated. (B) Pairwise (2D scatter plots left panel) and three-way (3D scatter plots right panels) comparisons representing log2 fold-changes of gene expression, H3K27ac and CTCF within CTCF HiChIP loops in NSD2 High versus Low cells. Concordant increased and decreased changing loops are colored in red and blue, respectively. Pearson correlations are indicated.

References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Baljevic, M., Zaman, S., Baladandayuthapani, V., Lin, Y.H., de Partovi, C.M., Berkova, Z., Amini, B., Thomas, S.K., Shah, J.J., Weber, D.M.*, et al.* (2017). Phase II study of the c-MET inhibitor tivantinib (ARQ 197) in patients with relapsed or relapsed/refractory multiple myeloma. Ann Hematol *96*, 977-985.

Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2018). FitHiChIP: Identification of significant chromatin contacts from HiChIP data. bioRxiv, 412833.

Bublik, D.R., Bursac, S., Sheffer, M., Orsolic, I., Shalit, T., Tarcic, O., Kotler, E., Mouhadeb, O., Hoffman, Y., Fuchs, G.*, et al.* (2017). Regulatory module involving FGF13, miR-504, and p53 regulates ribosomal biogenesis and supports cancer cell survival. Proc Natl Acad Sci U S A *114*, E496-E505.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Chang-Yew Leow, C., Gerondakis, S., and Spencer, A. (2013). MEK inhibitors as a chemotherapeutic intervention in multiple myeloma. Blood Cancer J *3*, e105.

Chen, R., Zhao, H., Wu, D., Zhao, C., Zhao, W., and Zhou, X. (2016). The role of SH3GL3 in myeloma cell migration/invasion, stemness and chemo-resistance. Oncotarget *7*, 73101-73113.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240-244.

Croonquist, P.A., Linden, M.A., Zhao, F., and Van Ness, B.G. (2003). Gene profiling of a myeloma cell line reveals similarities and unique signatures among IL-6 response, N-ras-activating mutations, and coculture with bone marrow stromal cells. Blood *102*, 2581-2592.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K.*, et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374-387.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep *15*, 2038-2049.

Fulciniti, M., Lin, C.Y., Samur, M.K., Lopez, M.A., Singh, I., Lawlor, M.A., Szalat, R.E., Ott, C.J., Avet-Loiseau, H., Anderson, K.C.*, et al.* (2018). Non-overlapping Control of Transcriptome by Promoter- and Super-Enhancer-Associated Dependencies in Multiple Myeloma. Cell Rep *25*, 3693-3705 e3696.

Gomez-Marin, C., Tena, J.J., Acemel, R.D., Lopez-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E.*, et al.* (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc Natl Acad Sci U S A *112*, 7542-7547.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Gong, Y., Lazaris, C., Sakellaropoulos, T., Lozano, A., Kambadur, P., Ntziachristos, P., Aifantis, I., and Tsirigos, A. (2018). Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. Nat Commun *9*, 542.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C.*, et al.* (2011). CTCF-binding elements mediate control of V(D)J recombination. Nature *477*, 424-430.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Hollenhorst, P.C., Ferris, M.W., Hull, M.A., Chae, H., Kim, S., and Graves, B.J. (2011). Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. Genes Dev *25*, 2147-2157.

Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S.H., Xu, J., and Yuan, G.C. (2018). Dissecting super-enhancer hierarchy based on chromatin interactions. Nat Commun *9*, 943.

Huang, Z., Wu, H., Chuai, S., Xu, F., Yan, F., Englund, N., Wang, Z., Zhang, H., Fang, M., Wang, Y.*, et al.* (2013). NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma. Cancer Res *73*, 6277-6288.

Hudlebusch, H.R., Santoni-Rugiu, E., Simon, R., Ralfkiaer, E., Rossing, H.H., Johansen, J.V., Jorgensen, M., Sauter, G., and Helin, K. (2011a). The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors. Clin Cancer Res *17*, 2919-2933.

Hudlebusch, H.R., Skotte, J., Santoni-Rugiu, E., Zimling, Z.G., Lees, M.J., Simon, R., Sauter, G., Rota, R., De Ioris, M.A., Quarto, M.*, et al.* (2011b). MMSET is highly expressed and associated with aggressiveness in neuroblastoma. Cancer Res *71*, 4226-4235.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods *9*, 999-1003.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., *et al.* (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262-275.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290-294.

Keats, J.J., Reiman, T., Belch, A.R., and Pilarski, L.M. (2006). Ten years and counting: so what do we know about t(4;14)(p16;q32) multiple myeloma. Leuk Lymphoma *47*, 2289-2300.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Koerber, R.M., Held, S.A.E., Heine, A., Kotthoff, P., Daecke, S.N., Bringmann, A., and Brossart, P. (2015). Analysis of the anti-proliferative and the pro-apoptotic efficacy of Syk inhibition in multiple myeloma. Exp Hematol Oncol *4*, 21.

Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A., *et al.* (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Mol Cell *44*, 609-620.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lauring, J., Abukhdeir, A.M., Konishi, H., Garay, J.P., Gustin, J.P., Wang, Q., Arceci, R.J., Matsui, W., and Park, B.H. (2008). The multiple myeloma associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity. Blood *111*, 856-864.

Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I., and Tsirigos, A. (2017). HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. BMC Genomics *18*, 22.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, L., Zhang, X., Cao, L., An, Q., Hao, J., Zhang, Y., Jin, R., Chang, Y., Huang, X., Lu, J.*, et al.* (2017). Reelin promotes adhesion of multiple myeloma cells to bone marrow stromal cells via integrin beta1 signaling. J Cancer *8*, 2212-2222.

Liu, D., and Mamorska-Dyga, A. (2017). Syk inhibitors in clinical development for hematological malignancies. J Hematol Oncol *10*, 145.

Lorenz, J., Waldschmidt, J., Wider, D., Follo, M., Ihorst, G., Chatterjee, M., May, A.M., Duyster, J., Rosenwald, A., Wasch, R.*, et al.* (2016). From CLL to Multiple Myeloma - Spleen Tyrosine Kinase (SYK) influences multiple myeloma cell survival and migration. Br J Haematol *174*, 985-989.

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

Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320-334.

Mahtouk, K., Moreaux, J., Hose, D., Reme, T., Meissner, T., Jourdan, M., Rossi, J.F., Pals, S.T., Goldschmidt, H., and Klein, B. (2010). Growth factors in multiple myeloma: a comprehensive analysis of their expression in tumor cells and bone marrow environment using Affymetrix microarrays. BMC Cancer *10*, 198.

Martinez-Garcia, E., Popovic, R., Min, D.J., Sweet, S.M., Thomas, P.M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L.M.*, et al.* (2011). The MMSET histone methyl transferase switches

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. Blood *117*, 211-220.

McCarthy, M.T., and O'Callaghan, C.A. (2014). PeaKDEck: a kernel density estimator-based peak calling program for DNaseI-seq data. Bioinformatics *30*, 1302-1304.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods *13*, 919-922.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-1021.

Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu Rev Genet *35*, 673-745.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944 e922.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc Natl Acad Sci U S A *115*, E6697-E6706.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A.*, et al.* (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. Mol Cell *70*, 1149-1162 e1145.

Oyer, J.A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C.I., Patel, J.P., Melnick, A.*, et al.* (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. Leukemia *28*, 198-201.

Phillip, C.J., Stellrecht, C.M., Nimmanapalli, R., and Gandhi, V. (2009). Targeting MET transcription as a therapeutic strategy in multiple myeloma. Cancer Chemother Pharmacol *63*, 587-597.

Popovic, R., Martinez-Garcia, E., Giannopoulou, E.G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M.Y., Zheng, Y., Will, C.M., Small, E.C.*, et al.* (2014). Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. PLoS Genet *10*, e1004566.

Qin, X., Lin, L., Cao, L., Zhang, X., Song, X., Hao, J., Zhang, Y., Wei, R., Huang, X., Lu, J.*, et al.* (2017). Extracellular matrix protein Reelin promotes myeloma progression by facilitating tumor cell proliferation and glycolysis. Sci Rep *7*, 45305.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D.*, et al.* (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320 e324.

Raviram, R., Rocha, P.P., Muller, C.L., Miraldi, E.R., Badri, S., Fu, Y., Swanzey, E., Proudhon, C., Snetkova, V., Bonneau, R.*, et al.* (2016). 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. PLoS Comput Biol *12*, e1004780.

Rocha, P.P., Raviram, R., Fu, Y., Kim, J., Luo, V.M., Aljoufi, A., Swanzey, E., Pasquarella, A., Balestrini, A., Miraldi, E.R.*, et al.* (2016). A Damage-Independent Role for 53BP1 that Impacts Break Order and Igh Architecture during Class Switch Recombination. Cell Rep *16*, 48-55.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A *112*, E6456-6465.

Sardina, J.L., Lopez-Ruano, G., Prieto-Bermejo, R., Sanchez-Sanchez, B., Perez-Fernandez, A., Sanchez-Abarca, L.I., Perez-Simon, J.A., Quintales, L., Sanchez-Yague, J., Llanillo, M.*, et al.* (2014). PTPN13 regulates cellular signalling and beta-catenin function during megakaryocytic differentiation. Biochim Biophys Acta *1843*, 2886-2899.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods *12*, 963-965.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., C, H.H., Mirny, L.*, et al.* (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51-56.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol *16*, 259.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Snetkova, V., and Skok, J.A. (2018). Enhancer talk. Epigenomics *10*, 483-498.

Spaan, I., Raymakers, R.A., van de Stolpe, A., and Peperzak, V. (2018). Wnt signaling in multiple myeloma: a central player in disease with therapeutic potential. J Hematol Oncol *11*, 67.

Stafford, J.M., Lee, C.H., Voigt, P., Descostes, N., Saldana-Meyer, R., Yu, J.R., Leroy, G., Oksuz, O., Chapman, J.R., Suarez, F.*, et al.* (2018). Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma. Sci Adv *4*, eaau5935.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tonjes, M., Sill, M., Bender, S.*, et al.* (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell *22*, 425-437.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Sun, F., Chronis, C., Kronenberg, M., Chen, X.F., Su, T., Lay, F.D., Plath, K., Kurdistani, S.K., and Carey, M.F. (2019). Promoter-Enhancer Communication Occurs Primarily within Insulated Neighborhoods. Mol Cell *73*, 250-263 e255.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B.*, et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611-1627.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'Keeffe, S., Masri, N.E., Roider, H.G., Manke, T., and Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat Protoc *6*, 1860-1869.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tsirigos, A., Haiminen, N., Bilal, E., and Utro, F. (2012). GenomicTools: a computational platform for developing high-throughput analytics in genomics. Bioinformatics *28*, 282-283.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep *10*, 1297-1309.

Walker, B.A., Wardell, C.P., Melchor, L., Hulkki, S., Potter, N.E., Johnson, D.C., Fenwick, K., Kozarewa, I., Gonzalez, D., Lord, C.J.*, et al.* (2012). Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. Blood *120*, 1077-1086.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M.*, et al.* (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet *44*, 251-253.

Xiang, Y., Zhou, X., Hewitt, S.L., Skok, J.A., and Garrard, W.T. (2011). A multifunctional element in the mouse Igkappa locus that specifies repertoire and Ig loci subnuclear location. J Immunol *186*, 5356-5366.

Xu, J., Pfarr, N., Endris, V., Mai, E.K., Md Hanafiah, N.H., Lehners, N., Penzel, R., Weichert, W., Ho, A.D., Schirmacher, P.*, et al.* (2017). Molecular signaling in multiple myeloma: association of RAS/RAF mutations and MEK/ERK pathway activation. Oncogenesis *6*, e337.

Zaman, S., Shentu, S., Yang, J., He, J., Orlowski, R.Z., Stellrecht, C.M., and Gandhi, V. (2015). Targeting the pro-survival protein MET with tivantinib (ARQ 197) inhibits growth of multiple myeloma cells. Neoplasia *17*, 289-300.

bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

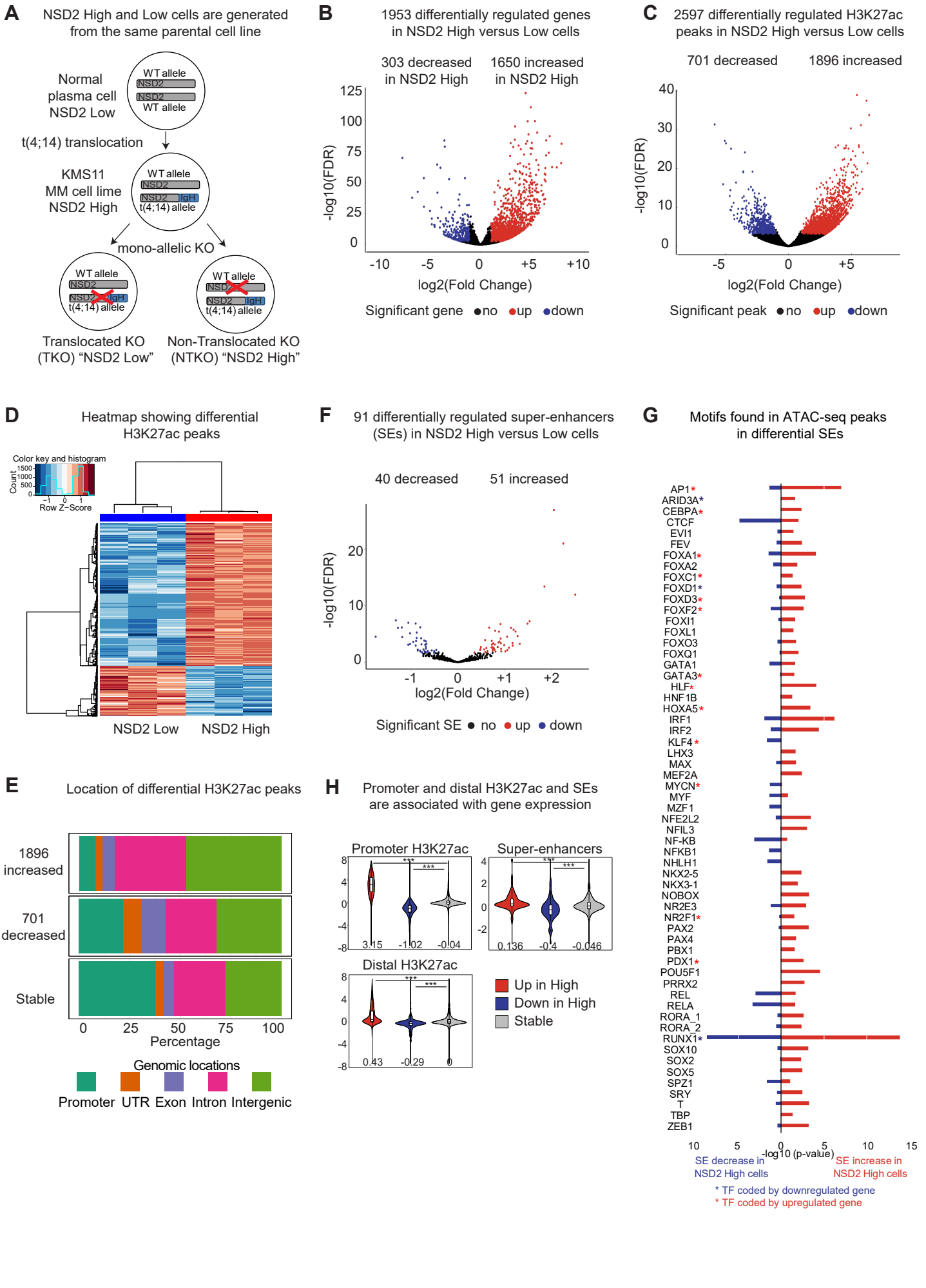
Zhang, Z., Newton, K., Kummerfeld, S.K., Webster, J., Kirkpatrick, D.S., Phu, L., Eastham-Anderson, J., Liu, J., Lee, W.P., Wu, J.*, et al.* (2017). Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. Proc Natl Acad Sci U S A *114*, 3903-3908.

Zhou, Y., Barlogie, B., and Shaughnessy, J.D., Jr. (2009). The molecular characterization and clinical management of multiple myeloma in the post-genome era. Leukemia *23*, 1941-1956.

Zhuang, L., Shou, T., Li, K., Gao, C.L., Duan, L.C., Fang, L.Z., Zhang, Q.Y., Chen, Z.N., Zhang, C., Yang, S.T.*, et al.* (2017). MicroRNA-30e-5p promotes cell growth by targeting PTPN13 and indicates poor survival and recurrence in lung adenocarcinoma. J Cell Mol Med *21*, 2852-2862.

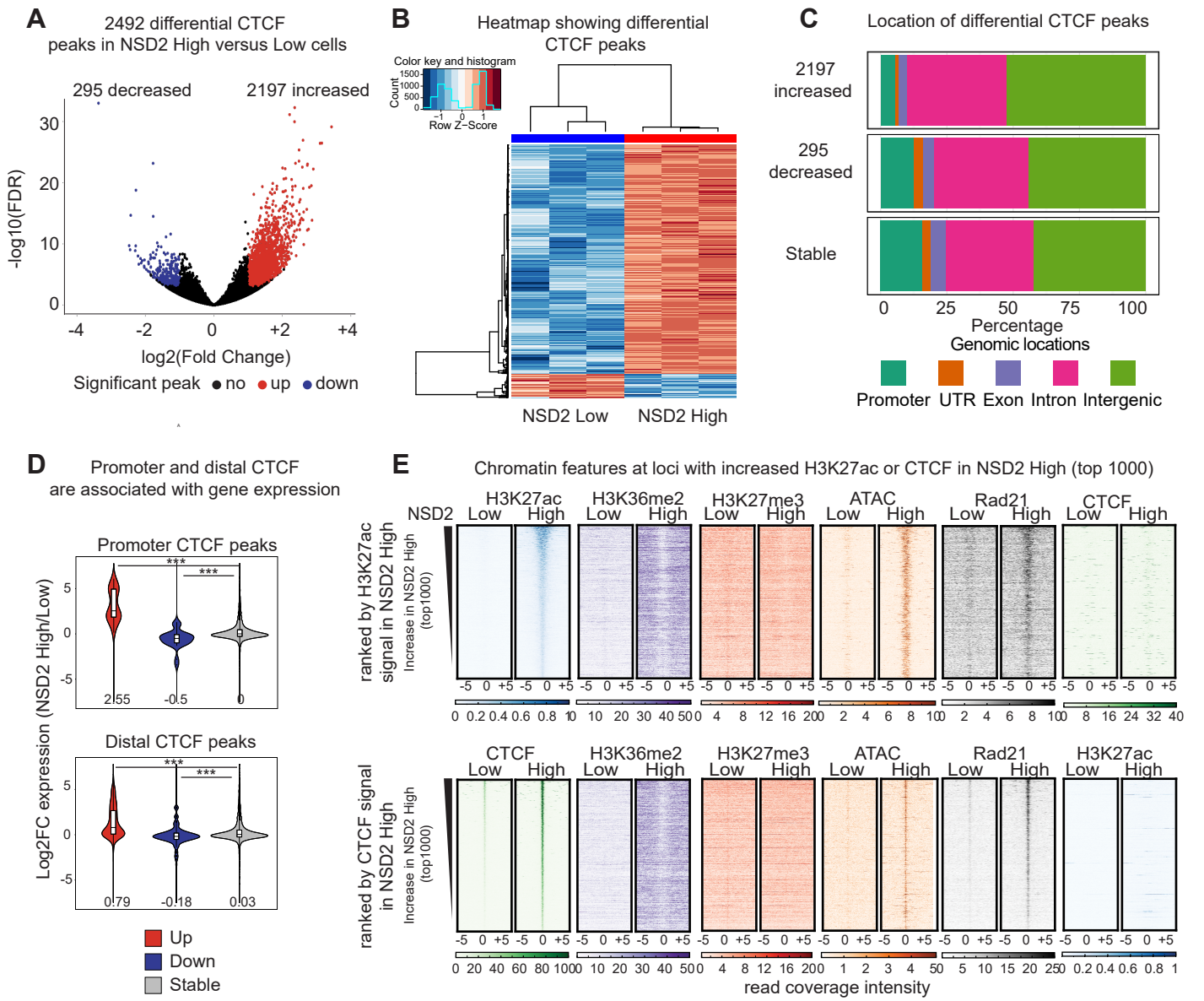
bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1



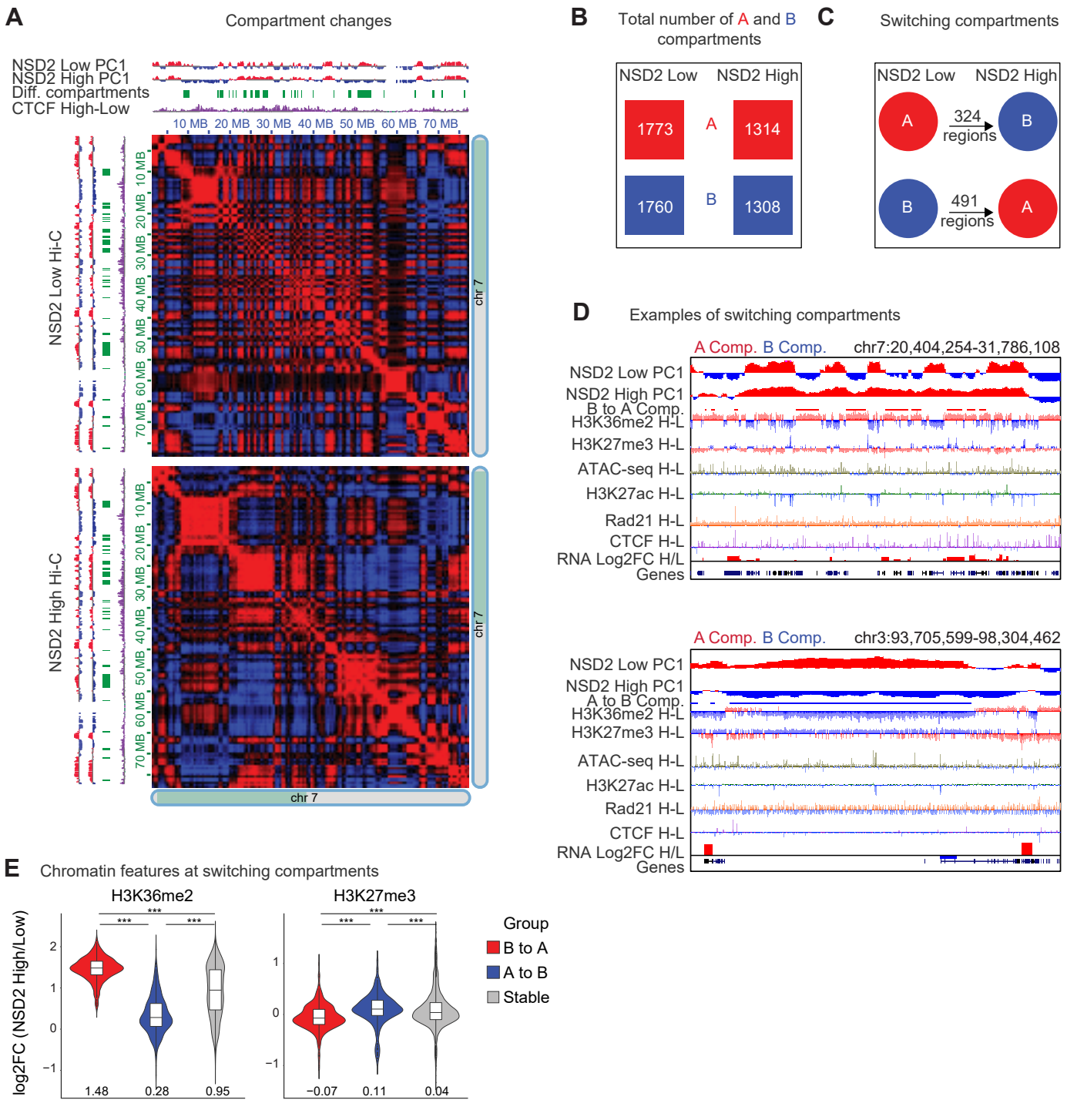
bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2



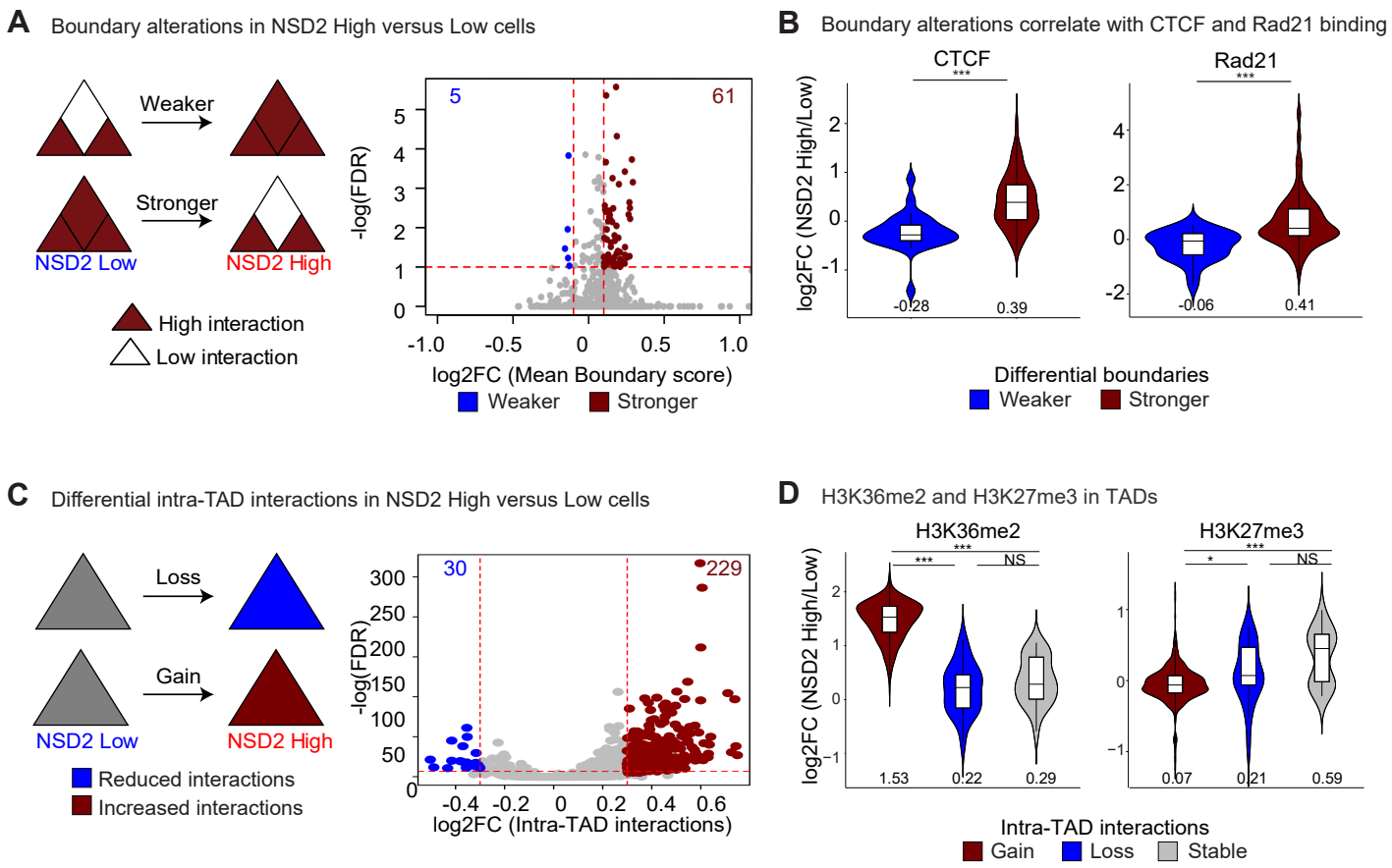
bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3



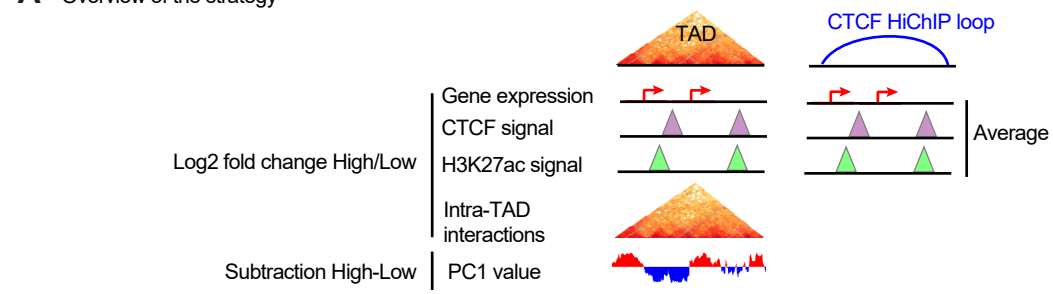
bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

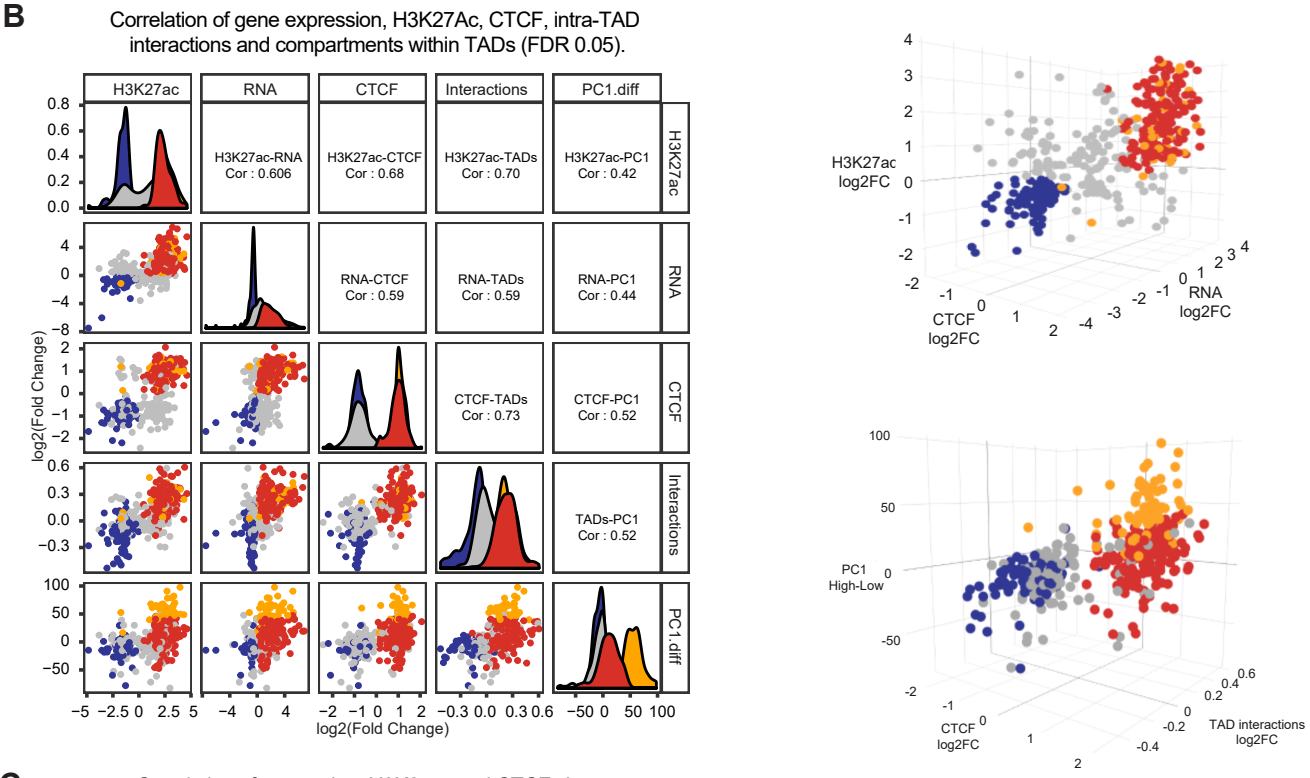
Figure 4

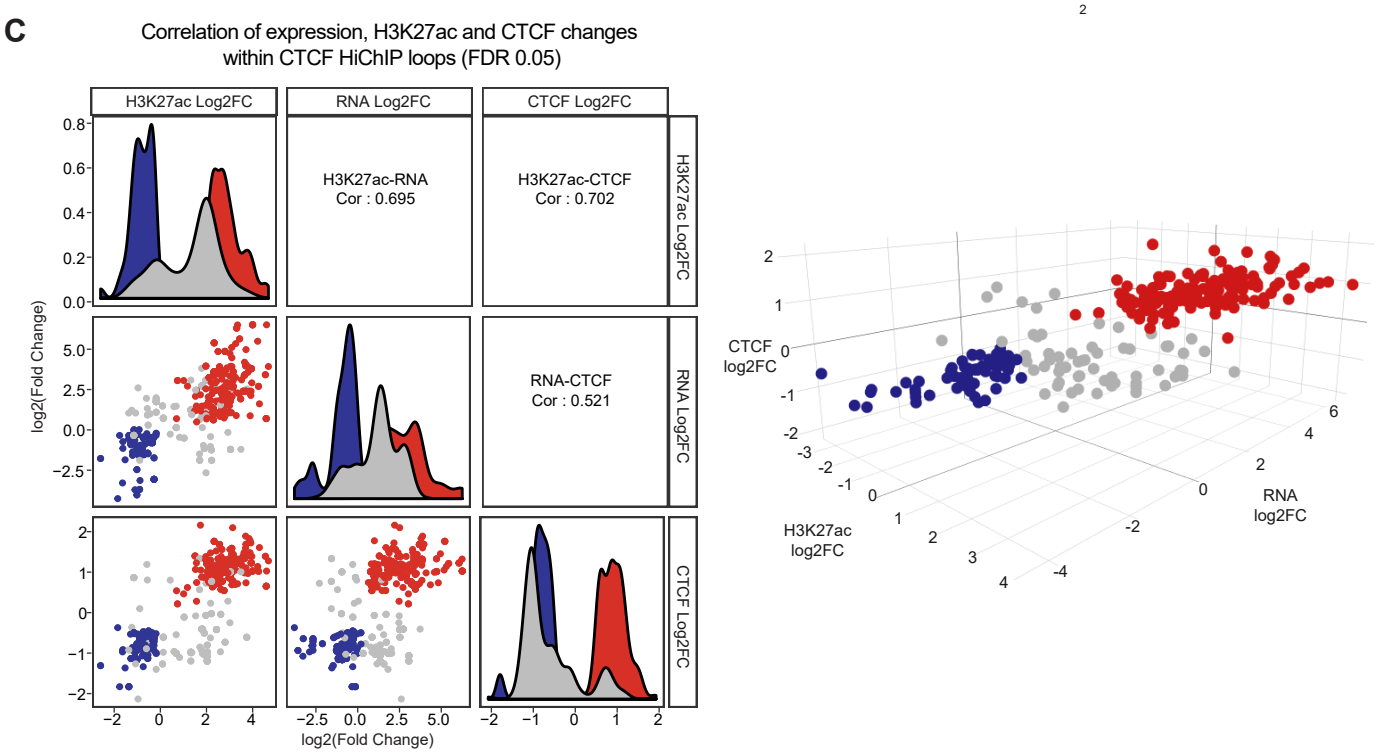


bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5

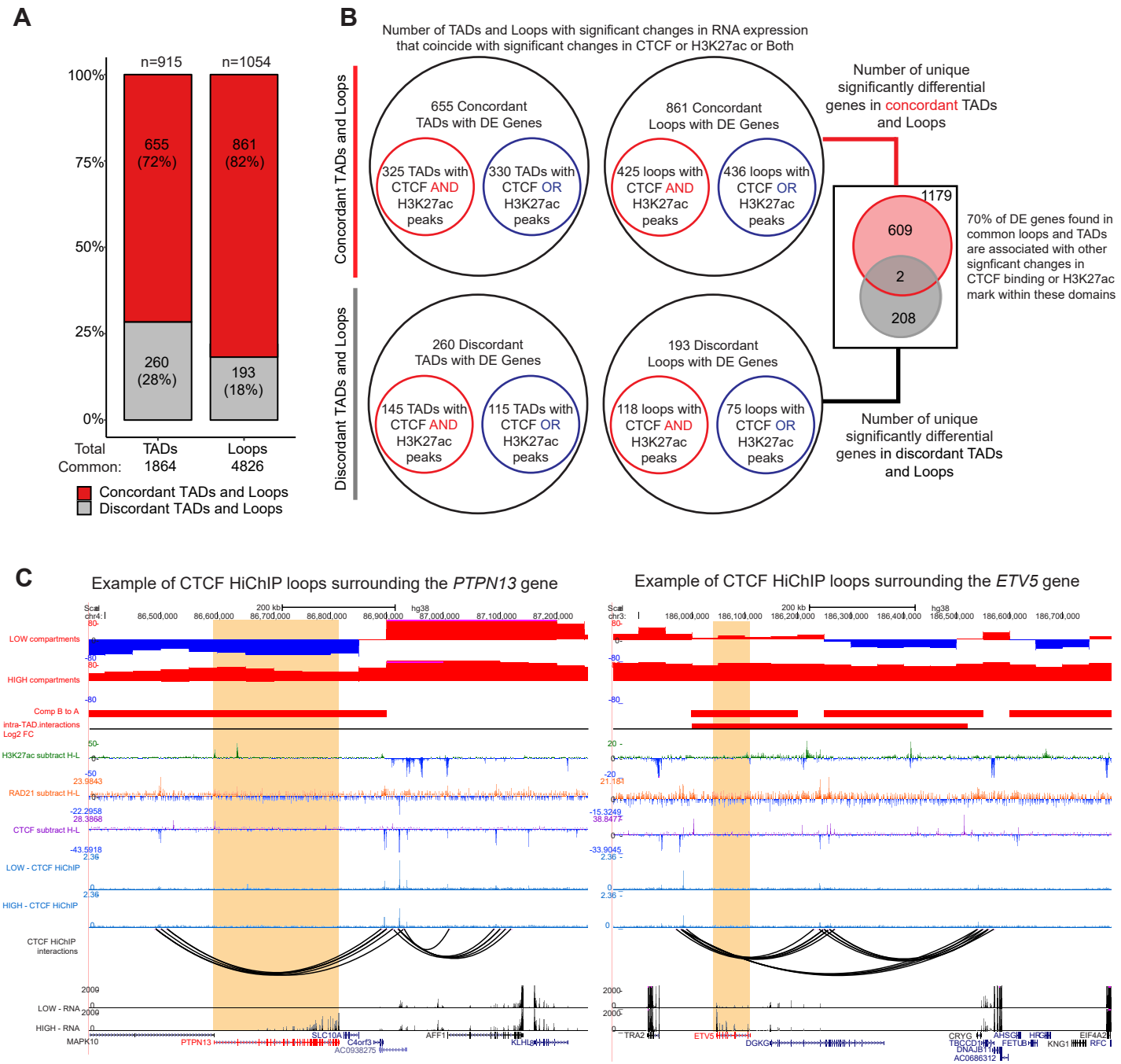






bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6



bioRxiv preprint doi: https://doi.org/10.1101/587931; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

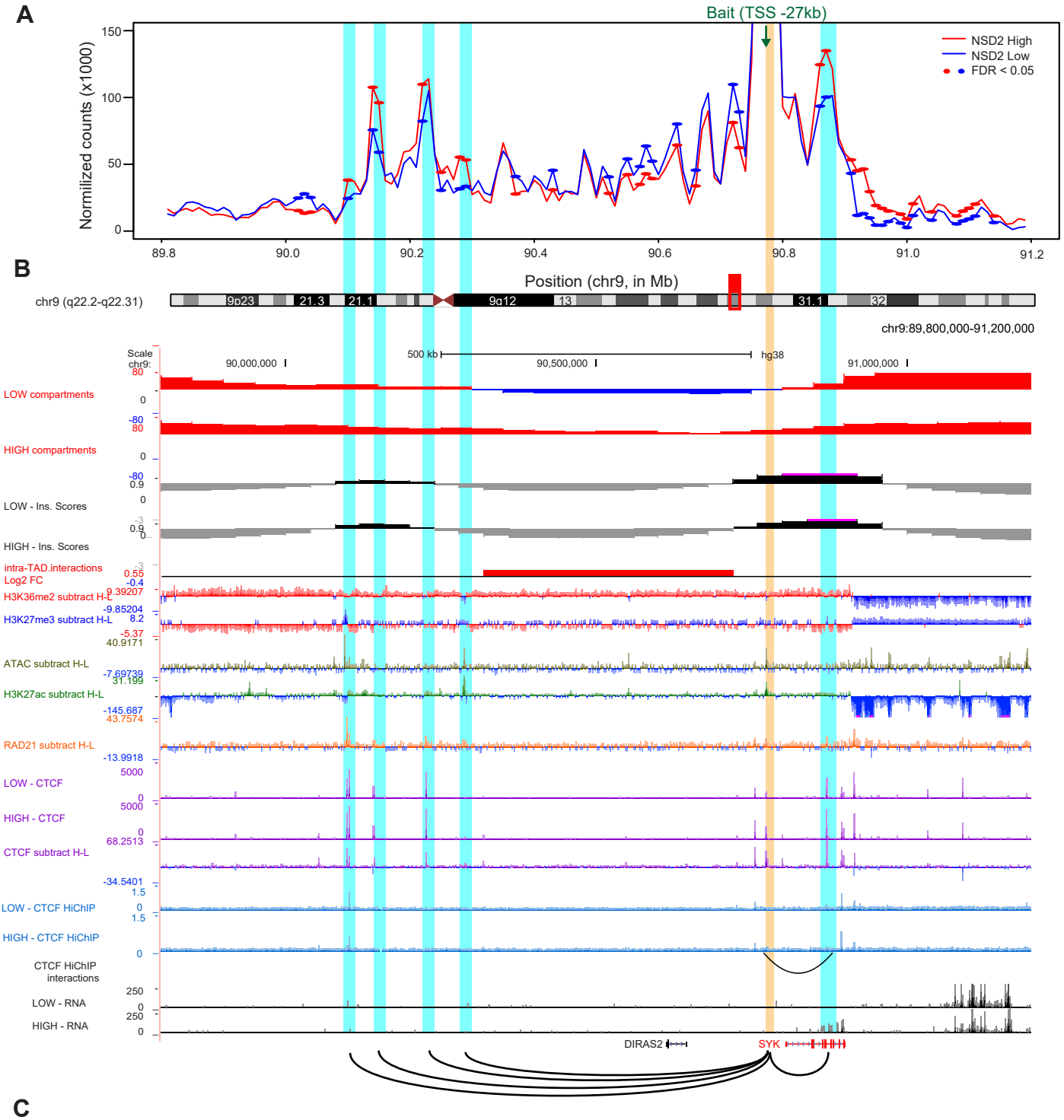



Figure 7