ATRX binds to atypical chromatin domains at the 3' exons of ZNF genes to preserve H3K9me3 enrichment

David Valle-García[1,2], Domhnall S. McHugh[1], Zulekha A. Qadeer[1,3], Flávia G. Ghiraldini[1], Dan Hasson[1], Michael A. Dyer[4], Félix Recillas-Targa[2] and Emily Bernstein[1,3,5]

[1]Departments of Oncological Sciences and Dermatology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[2]Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, México DF 04510, México.

[3]Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[4]Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

[5]Corresponding Author

Correspondence and requests for materials should be addressed to Emily Bernstein, Icahn School of Medicine at Mount Sinai, 1470 Madison Avenue, Room S6-114, New York, NY 10029, Phone: 212-824-9335, Emily.Bernstein@mssm.edu

Running title: ATRX binds to the 3' exons of ZNFs

Key words: ATRX, C2H2 zinc finger, zinc finger genes, KRAB-ZNFs, ZNF274, SETDB1, ESET, TRIM28, KAP1, H3K9me3

Reviewer's link to deposited GEO data:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=gfuzakysfpqbpsj&acc=GSE70920

ABSTRACT

ATRX is a SWI/SNF chromatin remodeler proposed to govern genomic stability through the regulation of repetitive sequences such as rDNA, retrotransposons, and pericentromeric and telomeric repeats. However, few direct ATRX target genes have been identified and high-throughput genomic approaches are currently lacking for ATRX. Here we present a comprehensive ChIP-sequencing study of ATRX in the human erythroleukemic cell line K562, in which we identify the 3' exons of zinc finger genes (ZNFs) as a new class of ATRX targets. These 3' exonic regions encode the zinc finger motifs, which can range from 1-40 copies per ZNF gene and share large stretches of sequence similarity. These regions often contain an atypical chromatin signature: they are transcriptionally active, contain high levels of H3K36me3 and are paradoxically enriched in H3K9me3. We find that these ZNF 3' exons are also co-occupied by SETDB1, TRIM28 and ZNF274. CRISPR/Cas9-mediated loss-of-function studies demonstrate a significant reduction of H3K9me3 at the ZNF 3' exons in the absence of ATRX and that H3K9me3 levels at atypical chromatin regions are particularly sensitive to ATRX loss compared to other H3K9me3-occupied regions. As ZNFs are one of the most rapidly expanding gene families in primates and genomic rearrangements are believed to be responsible for ZNF duplications, our results suggest that ATRX binds to the 3' exons of ZNFs to maintain their genomic stability through preservation of H3K9me3.

INTRODUCTION

Chromatin remodeling proteins act through shifting, sliding, deposition and eviction of nucleosomes and histones. Members of the SWI/SNF family of chromatin remodelers are fundamental in many cellular processes such as transcription, replication, DNA repair and recombination (Sinha and Peterson 2009; Hargreaves and Crabtree 2011; Euskirchen et al. 2012; Gospodinov and Herceg 2013; Längst and Manelyte 2015). One notable chromatin remodeler involved in all of these processes is ATRX. Increasing evidence supports that ATRX acts as a sentinel of genome integrity by maintaining heterochromatin at repetitive sequences (Clynes et al. 2013). Interestingly, ATRX mutations are responsible for a complex genetic disorder called ATR-X (Alpha Thalassemia, Mental Retardation X-linked) syndrome while mutations, deletions, and mechanisms of altered ATRX expression are highly prevalent in a wide variety of cancers (Ratnakumar and Bernstein 2013).

ATRX contains two highly conserved domains: the ADD (ATRX-DNMT3-DNMT3L) and the SWI/SNF helicase domain (Park et al. 2004). The ADD domain contains a PHD finger that binds H3K9me3/H3K4me0 (Eustermann et al. 2011; Iwase et al. 2011), whereas the SWI/SNF domain is an ATP-dependent helicase responsible for the chromatin remodeling capacity of ATRX (Xue et al. 2003; Park et al. 2004). Despite the fact that ATRX binds H3K9me3/H3K4me0 *in vitro*, ATRX binds to only a subset of H3K9me3-containing regions *in vivo* (McDowell et al. 1999; Law et al. 2010; Eustermann et al. 2011; Iwase et al. 2011). In particular, ATRX is highly enriched at certain H3K9me3-containing repetitive regions such as telomeric and pericentromeric repeats as well as some retrotransposon families (Drané et al. 2010; Goldberg et al. 2010; Wong

et al. 2010; Lewis et al. 2010; Sadic et al. 2015; Elsässer et al. 2015; He et al. 2015). Furthermore, ATRX physically interacts with other H3K9me3 binding proteins such as HP1α (Bérubé et al. 2000; Lechner et al. 2005). Altogether, these pieces of evidence suggest that ATRX is involved in the regulation of particular H3K9me3-modified chromatin.

A well characterized role of ATRX is deposition of histone variants into the chromatin template. For example, ATRX and DAXX (death-domain associated protein) act together as a histone chaperone complex for the H3 variant H3.3. ATRX is required for the localization of H3.3 at telomeres and pericentromeric repeats (Goldberg et al. 2010; Drané et al. 2010; Wong et al. 2010; Lewis et al. 2010), retrotransposons (Sadic et al. 2015; Elsässer et al. 2015) and imprinted loci (Voon et al. 2015), which all contain H3K9me3. This ability appears to be unique for ATRX, as the HIRA complex deposits H3.3 only at euchromatic regions (Goldberg et al. 2010; Ray-Gallet et al. 2011; Schneiderman et al. 2012; Filipescu et al. 2013). In addition to promoting H3.3 deposition, our group showed that ATRX negatively regulates the deposition of histone variant macroH2A at the α-globin locus (Ratnakumar et al. 2012).

ATRX has also been implicated in resolving aberrant secondary DNA structures, called G-quadruplexes, which form in guanine-rich regions during replication and transcription (Law et al. 2010; Clynes et al. 2013). G-quadruplexes are a common feature of some families of repetitive sequences and tandem repeats such as those found in telomeres. Intriguingly, ATRX mutations in cancer have been linked to the Alternative Lengthening of Telomeres (ALT) pathway (Heaphy et al. 2011; Schwartzentruber et al. 2012; Lovejoy et al. 2012; Bower et al. 2012). Although the precise role of ATRX in ALT remains

unclear, it has been suggested that ATRX prevents Homologous Recombination (HR) between telomeric sequences through the resolution of stalled replication forks in G-rich regions (Clynes and Gibbons 2013; Clynes et al. 2015). In accordance with its role as a regulator of genome stability, several reports demonstrated that ATRX depletion causes telomere dysfunction, increased fork stalling and increased sensitivity to replicative stress across different cellular and *in vivo* models (Wong et al. 2010; Huh et al. 2012; Watson et al. 2013; Leung et al. 2013; Clynes et al. 2014; He et al. 2015).

Despite these important functions, surprisingly few direct ATRX target genes have been identified. To address this, we utilized an unbiased approach using the ENCODE Tier 1 human erythroleukemic cell line K562 as a model system to analyze ATRX genomic occupancy. Through comprehensive ChIP-seq analyses, we identified an unexpected binding pattern of ATRX at the 3' exons of Zinc Finger Genes (ZNFs). ZNFs represent the largest family of putative transcription factors in the human genome with more than seven hundred identified members (Tadepally and Aubry 2010; Nowick et al. 2011).

Notably, the 3' exons of ZNFs are enriched in chromatin that is permissive to transcription yet contains high levels of H3K9me3 and H3K36me3 (Blahnik et al. 2011). These atypical chromatin regions do not possess the characteristics of any known regulatory region (i.e. promoter, enhancer, insulator) and their functional significance remains unclear (Blahnik et al. 2011). Here we show that ATRX co-occupies 3' ZNF exons containing an H3K9me3/H3K36me3 chromatin signature, together with the H3K9 methyltransferase SETDB1 (also known as ESET), the co-repressor TRIM28 (also known as KAP1), and the transcription factor ZNF274. Deletion of ATRX leads to a significant reduction of H3K9me3, particularly at 3' ZNF exons and other

H3K9me3/H3K36me3-containing regions, as well as increased DNA damage, and defects in the cell cycle. Taken together, our studies suggest that ATRX binds the 3' exons of ZNFs to maintain genomic stability by regulating H3K9me3 levels.

RESULTS

ATRX binds to the 3' exons of ZNF genes in K562 cells

In order to perform an unbiased search for novel direct ATRX target genes, we examined its genomic distribution by ChIP-seq analysis in the human erythroleukemic cell line K562 using two independent antibodies (see Methods for details). We chose K562 as a model system for two reasons: first, it has been established that ATRX has important roles in the regulation of the erythroid lineage (Law et al. 2010; Ratnakumar et al. 2012); second, K562 is a Tier 1 ENCODE cell line that has been extensively analyzed using a wide array of genomic and epigenomic methodologies that are publicly available.

To determine the global ATRX binding pattern in relation to other chromatin modifications, we analyzed the available ChIP-seq ENCODE datasets for K562 and performed a correlation analysis of their binding profiles. As previously reported, ATRX correlates with H3K9me3 (rho = 0.46) and moderately with macroH2A (rho = 0.19), consistent with its role as a macroH2A regulator (**Supp. Fig. S1A**). Furthermore, we examined the genomic distribution of ATRX significant peaks and found that, consistent with previous reports (Law et al. 2010; Elsässer et al. 2015; Sadic et al. 2015; He et al. 2015), ATRX is bound mainly to repetitive sequences (~56% of ATRX peaks overlap with repeats) (**Supp. Fig. S1B**). In order to understand ATRX distribution at a functional

level, we analyzed its distribution across Hidden Markov Model-derived chromatin states (Ernst et al. 2011). While ATRX binds significantly to repressed and repetitive regions (**Fig. 1A**), ATRX is significantly enriched in transcribed regions as well (**Fig. 1A**). In order to further investigate the significance of ATRX occupancy at these transcriptionally active regions, we performed Gene Ontology (GO) analysis with significant ATRX-bound genes (n=374, **Supp. Table S1**). Strikingly, Zinc Finger genes (ZNFs) were the most overrepresented gene family and comprised one quarter of the ATRX–bound genes, many of which contain the repressive KRAB domain (**Fig. 1B**). We next analyzed ATRX binding at promoters and gene bodies and found that enrichment of ATRX at gene bodies of ZNFs was highly significant as compared to non-ZNF genes, but that promoter regions had minimal binding in either group of genes (**Fig. 1C**).

ZNFs represent the fastest expanding gene family in the primate lineage. Frequent gene duplications and rapid divergence of paralogs are characteristic of ZNFs (Tadepally and Aubry 2010; Nowick et al. 2011). Because of this, ZNFs are often arranged in large continuous clusters in the human genome and share stretches of highly similar DNA sequence, particularly at their 3' exons where the DNA sequence encoding the zinc finger motifs is contained (Tadepally and Aubry 2010). Chromosome 19 contains the majority of ZNF clusters in the human genome. By examining ATRX enrichment on chromosome 19, we found that the ZNFs clusters are demarcated by ATRX occupancy (**Fig. 1D,E**). We next analyzed the binding pattern of ATRX over individual ZNFs. Interestingly, ATRX is preferentially enriched at the 3' exons of ZNFs (**Fig. 1F,G)**. These results were confirmed by ChIP-seq with a second ATRX antibody, which showed nearly identical enrichment patterns at ZNF genes (**Supp. Fig. S2**). Overall, our ChIP-seq

studies demonstrate that ZNFs are a novel set of ATRX targets and that ATRX is highly enriched at their 3' exons.

ATRX is enriched at ZNF genes harboring an atypical chromatin signature and distinctive epigenetic and genomic features

A large proportion of ZNFs contain an atypical chromatin signature at their 3' exons (Blahnik et al. 2011). This includes high levels of H3K9me3, permissibility to transcription, and enrichment of H3K36me3, a mark associated with transcriptional elongation. Our analysis of ENCODE ChIP-seq data corroborated these observations (**Fig. 1E-G; Fig. 2A**).

To investigate the epigenetic and genetic characteristics of the ATRX-enriched ZNFs and their relationship with the above atypical chromatin signature, we categorized all ZNF genes into three classes based on their ATRX enrichment levels: Class I represents ZNFs highly enriched for ATRX, Class II contains ZNFs moderately enriched, and Class III for ZNFs depleted of ATRX enrichment (**Fig. 2A,** top). We next quantified the average ChIP-seq signals of ATRX, H3K9me3 and H3K36me3 over the gene bodies of the ZNF classes. As shown in **Figure 2A**, Class I ZNFs show high levels of both H3K9me3 and H3K36me3. In contrast, Class III genes are largely depleted of H3K9me3 and show less enrichment of H3K36me3. To analyze if ATRX enrichment is correlated with the presence of H3K9me3 and H3K36me3 at the same loci, we calculated the Spearman correlation of these marks in Class I and Class III ZNFs. H3K9me3 is correlated with ATRX and H3K36me3 in Class I ZNFs whereas a poor correlation was observed in

Class III ZNFs (**Fig. 2B**). These data suggest that ATRX is specifically enriched at those ZNFs displaying an atypical chromatin signature.

To further understand ATRX recruitment and function at ZNFs, we analyzed the ZNF-related genomic features in the three classes. ZNF genes can be classified as transcriptional activators or KRAB-containing repressors. KRAB is a potent repressor domain that is contained in about half of the ZNFs and is generally encoded in two exons independent of the 3' exon containing the zinc finger motifs (Tadepally and Aubry 2010). As KRAB-containing genes were enriched in our Gene Ontology analysis of ATRX-bound genes (**Fig. 1B**), we plotted the number of KRAB domains contained by each ZNF ordered by class. Most of the ATRX-enriched Class I and Class II ZNFs contained KRAB domains, while very few of Class III ZNF genes contain this feature (**Fig. 2C,** left). Since the KRAB domain is not encoded within the 3' exons where ATRX binds, the KRAB domain likely aids indirectly in the recruitment of ATRX to KRAB-containing ZNFs.

Because the DNA sequence encoding C2H2-like zinc finger motifs is similar between ZNFs, it has been proposed that ZNF genes are prone to homologous recombination (HR), particularly those with more zinc finger motifs (Tadepally and Aubry 2010; Blahnik et al. 2011). Therefore, the presence of H3K9me3 at the ZNF 3' exons has been suggested to protect against HR (Vogel et al. 2006; Blahnik et al. 2011). To support this idea, we plotted the number of predicted C2H2-like zinc finger motifs per ZNF gene. On average, human ZNF genes contained ~9 zinc finger motifs per gene. In contrast, Class I ZNF genes contained significantly more motifs with an average of ~14, while ATRX depleted Class III genes contained only ~6 domains per gene (**Fig. 2C,** left and top right). These results suggest that ATRX enrichment at the ZNF 3' exons positively

correlates with the number of C2H2-like zinc finger motifs. This is in accordance with a study that reported H3K9me3 enrichment at 3' ZNF exons positively associated with the number of zinc finger motifs (Blahnik et al. 2011).

A genomic feature proposed to be important for ATRX binding is the Guanine DNA content (G-content). ATRX binds to G-quadruplexes with high affinity *in vitro* and facilitates polymerase elongation through deposition of H3.3 specifically in G-rich regions that have a tendency to form these structures (Law et al. 2010; Levy et al. 2015). Based on these observations, we measured the G-content of the ZNFs and predicted the potential of G-quadruplex formation at the 3' ends of the ZNF genes. Surprisingly, ATRX enrichment levels negatively correlated with both G-content and the potential to form G-quadruplexes (**Fig. 2C,** left and right middle panel). This strongly indicates that ATRX recruitment to ZNF 3' exons is not mediated by its ability to recognize G-quadruplexes, but by an alternative mechanism(s).

We then investigated whether ATRX enrichment and the presence of the atypical chromatin signature correlates with ZNF transcriptional levels. Thus, we analyzed the ENCODE RNA-seq datasets for K562 and plotted the normalized RPKM signal for the three ZNFs classes. Intriguingly, there was no evident correlation between RNA-seq levels and ATRX enrichment. This suggests that neither ATRX binding nor the formation of the atypical chromatin signature have a direct effect on ZNF expression levels (**Fig. 2C,** left and bottom right).

As ATRX regulates late stalled replication forks and H3K9me3-marked chromatin is often late replicating (Huh et al. 2012; Leung et al. 2013; Julienne et al. 2013; Clynes et

al. 2014), we queried whether ATRX binds to late-replicating ZNFs. To address this, we analyzed the K562 Repli-seq data from ENCODE and quantified the signal for the ZNF classes throughout S phase. Interestingly, we found that ZNF Classes I and II tend to be late replicating while Class III ZNFs replicate early (**Supp. Fig. S3A**).

In summary, we have established that ATRX levels positively correlate with H3K9me3 and H3K36me3 at atypical chromatin found at the 3' of ZNF genes. ATRX enrichment at ZNFs is independent of transcriptional levels. Moreover, ATRX-enriched ZNFs tend to be late replicating, KRAB-containing ZNFs with a larger number of zinc finger motifs than the genomic average. Such ZNFs also contain low levels of G-content and low potential for G-quadruplex formation. These trends are all statistically significant (**Supp. Tables S2-S3**).

SETDB1, TRIM28 and ZNF274 co-localize at 3' exons of ATRX-bound ZNF genes

In order to find additional chromatin factors that bind the ZNF 3' exons, we performed metagene analyses at the ZNF gene bodies with available K562 ChIP-seq datasets of known H3K9-regulating proteins. We found that the enrichment levels of the H3K9me3 methyltransferase SETDB1 (also known as ESET) and the SETDB1-interacting protein TRIM28 (also known as KAP1) correlated appreciably with ATRX and H3K9me3 at the 3' exons of ZNF genes (**Fig. 2D**). TRIM28 is a co-repressor that interacts with KRAB-containing ZNF transcription factors and recruits HDACs and H3K9 methyltransferases to enforce silencing (Iyengar and Farnham 2011). From this collective data arises the possibility that KRAB-ZNFs guide the TRIM28/SETDB1/ATRX complex to ZNF genes. In fact, the KRAB-containing transcription factor ZNF274 has been shown to bind the 3'

region of ZNF genes and recruit SETDB1 through its interaction with TRIM28 (Frietze et al. 2010). We therefore analyzed ENCODE ChIP-seq data of ZNF274 in K562 and found that ZNF274 indeed co-localizes with TRIM28/SETDB1/ATRX, suggesting they form a ZNF gene-regulatory complex (**Fig. 2D**).

We then performed a correlation analysis of the ChIP-seq signals of the ZNF binding factors and H3K9me3 to determine if the ZNF274/TRIM28/SETDB1/ATRX complex is ZNF-specific. Genome-wide, ATRX associates only with H3K9me3 and TRIM28 (**Fig. 2E**, left). In striking contrast, when focusing only on the ZNF genes, the correlation between ZNF274/TRIM28/SETDB1/ATRX significantly increases, along with presence of H3K9me3 signal (**Fig. 2E**, right).

Because ZNF274 is the only factor of the complex that can recognize specific DNA motifs, we hypothesized that ZNF274 recognizes sequences present at the ZNF 3' exons, which then tethers TRIM28/SETDB1/ATRX to these regions. Thus, we further analyzed ZNF274 genomic binding sites in K562 and performed a comprehensive motif analysis search. We found 3 significant DNA motifs that were highly enriched at the 3' end of ATRX enriched Class I and Class II ZNF genes, but not at ATRX depleted Class III ZNFs (**Supp. Fig. S3B**). This implicates ZNF274 as the transcription factor that guides ATRX and associated factors to ZNF 3' exons.

Next, we investigated if this set of proteins is bound at ZNF 3' exons in other cell types. While ATRX, SETDB1 and TRIM28 ChIP-seq data sets are not available for the majority of cell lines, we utilized those cell lines with publicly available ChIP-seq data for ZNF274 and H3K9me3 to analyze their pattern over ZNFs (e.g. HeLa-S3, H1-hESC, etc.).

Interestingly, ZNF274 binds to the 3' of ATRX bound Class I and Class II ZNF genes (as defined in K562) in the majority of cell lines analyzed (**Supp. Fig. S3C**). In accordance with previous studies (Blahnik et al. 2011), the H3K9me3 pattern at the 3' region of ZNFs was also conserved across cell lines (**Supp. Fig. S3C**).

Collectively, we have identified a ZNF-specific regulatory complex (ZNF274/TRIM28/SETDB1/ATRX) that binds the 3' region of ZNF genes and correlates with H3K9me3 enrichment in K562 cells. Furthermore, the enrichment pattern of ZNF274 and H3K9me3 at ZNF 3' exons is conserved in other cell lines of diverse lineages, including non-tumorigenic cells, suggesting a more general mechanism of ZNF gene regulation.

ATRX enrichment at ZNFs and repetitive regions is cell-type specific

Based on our ChIP-seq analyses in K562, we next hypothesized that ZNF274 binding to 3' ZNFs would predict ATRX enrichment in other cell types. To test this, we chose the human hepatocellular carcinoma cell line HepG2, a Tier 2 ENCODE cell line that shows similar ZNF274 and H3K9me3 enrichment patterns at ZNFs as K562 (**Fig. 3A**). We performed ATRX ChIP-qPCR in HepG2 and analyzed eleven ATRX-bound Class I ZNFs distributed across different chromosomes as well as two ATRX-depleted Class III ZNF genes as negative controls. As predicted, we observed distinct ATRX enrichment at most of the Class I ZNFs and none at Class III (**Fig. 3B**). However, one Class I ZNF, ZNF441, was devoid of ATRX enrichment in HepG2 (**Fig. 3B)**. Interestingly, this ZNF also displayed low levels of ZNF274 ChIP-seq enrichment (data not shown). Importantly,

these results suggest that ATRX binds to ZNF 3' exons in diverse cell types and its binding correlates with that of ZNF274.

To further investigate ATRX binding in other cell types, we analyzed a published human ATRX ChIP-seq dataset generated in primary human erythroblasts and compared it with the K562 ATRX ChIP-seq data (Law et al. 2010). Despite the fact that K562 cells are of erythroid lineage, the genome-wide distribution of ATRX in K562 is remarkably different from that of erythroblasts. While ATRX is distributed throughout genes, promoters and intergenic regions in K562 cells, ATRX is significantly enriched at promoters in erythroblasts (**Supp. Fig. S4A,B**). Because a significant proportion of ATRX is bound to repetitive sequences (**Fig. 1B, Supp. Fig. S1B**), we next compared the ATRX binding pattern at these regions. To address this, we estimated the ATRX enrichment at different repeat families in K562 and erythroblasts (Day et al. 2010). We then plotted the estimated enrichments for four different repetitive classes bound by ATRX in humans: telomeric, rDNA, satellite and simple repeats (McDowell et al. 1999; Goldberg et al. 2010; Law et al. 2010; Wong et al. 2010; Voon et al. 2015). As previously reported, erythroblasts showed considerable ATRX enrichment in all repeat categories, particularly telomeres and rDNA. In contrast, K562 cells showed modest ATRX enrichment in telomeric and rDNA repeats, low enrichment levels at satellite repeats and was absent from simple repeats (**Supp. Fig. S4C**). These results suggest that ATRX occupancy at distinct repetitive sequences may be cell-type specific.

We next analyzed the ATRX enrichment pattern at ZNFs in erythroblasts. In contrast to K562 cells, erythroblasts lack ATRX binding at ZNF gene bodies, but instead show significant enrichment at their promoters (**Fig. 3C, Supp. Fig. S4D-F**). Thus, it appears

that ATRX is not bound to ZNF gene clusters in erythroblasts. While we cannot completely exclude experimental variations, the observed differences may be result of cell type-specific and highly regulated ATRX binding.

To assess ATRX binding at ZNF genes in other organisms, we analyzed two recent ATRX ChIP-seq datasets from mouse embryonic stem cells (mESC) and mouse embryonic fibroblasts (MEF), as well as available ChIP-seq data for H3K9me3 in these same cell types. Similar to ATRX binding in erythroblasts, ATRX was highly enriched at ZNF promoters in mESC, while MEFs displayed low ATRX enrichment at ZNFs (**Fig. 3D**). Interestingly, H3K9me3 enrichment at ZNFs does not co-localize with ATRX, and only mESC showed H3K9me3 enrichment at the 3' exons of ZNFs.

Taken together, our comparative analyses provide evidence of cell type-specific ATRX binding patterns, which may be reflective of developmental stage, lineage or tumor type. Thus, ATRX binding to 3' exons of ZNFs may be a highly regulated phenomenon present only in certain cell types.

ATRX deficient cells have reduced H3K9me3 enrichment at 3' exons of ZNFs

To functionally investigate the role of ATRX in K562 cells, we generated two clonal ATRX knock out (KO) cell lines using CRISPR/Cas9 genome editing. Our KO lines (ATRX KO1 and KO2) were validated in detail (**Supp. Fig. S5**). As a control, we used a clonal cell line overexpressing Cas9 alone (referred to as V2). Western blot analysis showed that ATRX KO1 cells retain residual ATRX, while KO2 cells are completely devoid of ATRX protein (**Fig. 4A**). We next performed ATRX ChIP-qPCR in these lines

for the described panel of Class I and Class III ZNFs. In accordance with our ChIP-seq data, Class III ZNFs lacked ATRX enrichment in all lines while Class I ZNFs were enriched for ATRX in the control, reduced in KO1 and ablated in KO2 cells (**Fig. 4B**).

Loss of ATRX has been shown to promote genomic instability and defects in cell cycle (Huh et al. 2012; Watson et al. 2013; Clynes et al. 2014; Leung et al. 2013). We tested these functional readouts in our KO cell lines and found that indeed ATRX KO K562 cells have increased DNA damage as compared to control cells as assessed by Comet assay (**Supp. Fig. S6A,B**). Furthermore, ATRX KO cells displayed a slight but reproducible defect in the G1/S transition of the cell cycle (**Supp. Fig. S6C,D**).

Then, we queried whether ATRX deficiency alters the chromatin state of the atypical chromatin domains at ZNF genes. To address this question, we performed native ChIP for H3K9me3 in control and ATRX KO cells, followed by qPCR analysis for ZNFs. Strikingly, H3K9me3 levels decreased at Class I ZNFs in the ATRX KO cells compared to control cells, and displayed a similar pattern to that of ATRX KO (**Fig. 4B,C**). This data suggests that ATRX explicitly regulates H3K9me3 levels at ATRX-bound ZNF genes.

Next, we investigated if ATRX KO affected H3K36me3 levels at ATRX-enriched ZNFs by native ChIP-qPCR. In contrast to H3K9me3, H3K36me3 levels remain stable upon ATRX loss (**Supp. Fig. S7A**). We also assessed if ATRX depletion affected the transcriptional levels of ZNF genes. We did not observe changes in the transcriptional levels of ZNFs in the ATRX KO cell lines (**Supp. Fig. S7B**), consistent with our ChIP-seq findings that ATRX enrichment at ZNFs does not correlate with their transcriptional activity (**Fig. 2C**).

We then analyzed if the binding of the ZNF274/TRIM28/SETDB1 complex was altered at Class I ZNF genes upon ATRX KO (**Supp. Figs. S7C-E**). Because ZNF274 binds DNA and recruits TRIM28, which in turn recruits SETDB1, we speculate that the ZNF274/TRIM28/SETDB1 complex serves as a scaffold for ATRX at the 3' exons of ZNFs. Therefore, we posit that ATRX KO would not affect the complex binding to the chromatin, but rather its capacity to deposit or maintain H3K9me3. As expected, TRIM28 and ZNF274 binding remains largely unchanged at ZNF genes after ATRX KO (**Supp. Figs. S7C,D**). While ATRX KO1 cell line did not affect SETDB1 binding, SETDB1 occupancy was increased in ATRX KO2 (**Supp. Fig. S7E**). This may reflect a compensatory effect for complete loss of ATRX and concomitant H3K9me3 depletion. These data broadly favors a model in which ZNF274/TRIM28/SETDB1 is able to bind to the 3' exons of ZNFs independently of ATRX, but ATRX is required to establish or maintain H3K9me3 at atypical chromatin of ZNF genes.

To investigate the genome-wide alterations of H3K9me3 after ATRX depletion, we performed native ChIP-seq for H3K9me3 in control and ATRX KO cells. We first analyzed the H3K9me3 enrichment at ZNF genes. In agreement with our ChIP-qPCR results, we found H3K9me3 levels significantly reduced in KO cells, specifically in ATRX-enriched Class I and II ZNFs (**Fig. 4D**). These results were confirmed by metagene analysis of ZNFs (**Fig. 4E**) and visualization of H3K9me3 peaks at the ZNF clusters on chromosome 19 (**Supp. Fig. S8A,B**). This demonstrates that ATRX is required for establishing and/or maintaining H3K9me3 levels at the 3' end of a subset of ZNFs.

H3K9me3 within atypical chromatin domains is sensitive to ATRX depletion

To further understand the nature of H3K9me3 reduction upon ATRX depletion, we analyzed all genomic regions that show a significant loss of H3K9me3 levels in the ATRX KO cell line. We divided such regions based on their ATRX content in ATRX-bound (n = 1,458) and ATRX-unbound (n = 4,597). We hypothesized that ATRX-bound regions loose H3K9me3 as a direct effect of ATRX depletion whereas reduction of H3K9me3 at ATRX-unbound regions are due to secondary effects. We then calculated and plotted the average ChIP-seq enrichment signal of SETDB1, TRIM28 and ZNF274 derived from published ChIP-seq data in K562 over the K9me3 reduced regions. We observed that the three factors are enriched genome-wide in ATRX-bound regions as compared to the ATRX unbound regions (**Fig. 4F**). We also analyzed the H3K36me3 content in H3K9me3 reduced regions. Strikingly, H3K36me3 is highly enriched at ATRX-bound regions while ATRX-unbound regions had significantly lower levels (**Fig. 4F**, bottom right). These data suggest that H3K9me3 loss at ATRX targets occurs mainly in atypical chromatin regions.

Finally, we analyzed the genomic distribution of H3K9me3 reduced regions. The majority of ATRX-bound and ATRX-unbound H3K9me3 reduced regions are within repetitive sequences and intergenic regions (**Fig. 4G**). Interestingly, ZNF genes account for ~10% of the ATRX-bound H3K9me3 reduced regions. These results suggest that ATRX is a general regulator of H3K9me3 levels at atypical chromatin regions, and that ZNF genes comprise a subset of these affected regions.

Collectively, we propose a model in which ATRX is tethered to the 3' exons of ZNF genes by the ZNF274/TRIM28/SETDB1 complex to establish or maintain/protect

H3K9me3 at these transcriptionally active regions (**Fig. 5**, left). ATRX depletion leads to reduction of H3K9me3 levels in atypical chromatin domains, particularly at ZNF 3' exons and may explain, at least in part, the genomic instability associated with loss of ATRX. Furthermore, we propose that impairing ATRX function has important consequences for the genomic stability and evolution of ZNFs clusters, as the loss of H3K9me3 at ZNF 3' exons may increase probability of HR between them (**Fig. 5**, right; see discussion below).

DISCUSSION

ATRX is an important chromatin regulator involved in diverse cellular processes such as transcriptional regulation, maintenance of imprinted loci, replication, genome stability and chromatin looping (Bérubé 2011; Ratnakumar and Bernstein 2013; Clynes et al. 2013). Here we report that ATRX binds the 3' exons of ZNFs with an atypical chromatin signature to establish or maintain high levels of H3K9me3 (**Fig. 5**). ZNFs represent the fastest expanding gene family encoding transcription factors in humans, with the greatest diversity of target sequences (Tadepally and Aubry 2010; Nowick et al. 2011; Najafabadi et al. 2015). This property of ZNFs to recognize a multitude of motifs is generated through diverse combinations of their zinc finger motifs (Najafabadi et al. 2015). The average ZNF gene contains 9 independent zinc finger motifs and the number of motifs varies from gene to gene. Interestingly, the DNA sequences that code for these motifs are almost always located at the 3' exons of ZNF genes and harbor an atypical chromatin signature consisting of both H3K9me3 and H3K36me3 (**Fig. 5**) (Tadepally and Aubry 2010; Blahnik et al. 2011).

Through an unbiased ChIP-seq approach and analysis of ENCODE data, we found that ATRX preferentially binds to the 3' exons of a subset of ZNF genes containing this atypical chromatin signature. These ZNF genes are distinguished by the presence of a KRAB repressor domain, a higher than average number of zinc finger motifs, low levels of G-content and low probability of forming G-quadruplexes. Previous studies have shown that ATRX recognizes and resolves G-quadruplexes, relevant in the context of gene regulation (Law et al. 2010). However, our data suggests that ATRX binds and regulates these atypical chromatin regions by an alternative mechanism.

Intriguingly, we found that ATRX co-localizes with the previously reported ZNF274/TRIM28/SETDB1 complex (Frietze et al. 2010) at ZNF 3' exons with an atypical chromatin signature. Interestingly, it was demonstrated that ATRX binds to TRIM28 and SETDB1 to regulate the silencing of ERV repeats that contain an atypical H3.3/H3K9me3 chromatin signature in the mouse genome, although the transcription factor(s) that tether the TRIM28/SETDB1/ATRX complex to those regions remains unknown (Sadic et al. 2015; Elsässer et al. 2015). The fact that we find atypical chromatin regions that are enriched in TRIM28/SETDB1/ATRX but do not have ZNF274 enrichment (data not shown) strongly suggests that other ZNF transcription factors are involved in targeting ATRX to specific loci. Curiously, some ZNFs have co-evolved to specifically recognize sub-families of repetitive sequences, particularly retrotransposons (Thomas and Schneider 2011; Jacobs et al. 2014; He et al. 2015). Because ATRX is a known regulator of retrotransposons, it is possible that ATRX is tethered to such sequences by ZNFs. Furthering this notion, several ZNFs have tissue-specific expression profiles (Vogel et al. 2006; Najafabadi et al. 2015) and this may explain, at least in part, the observed variability of ATRX binding across different cell lines.

Our loss-of-function studies revealed that H3K9me3 levels at atypical chromatin regions, in particular those at ZNF 3' exons, are significantly decreased after ATRX loss. Other studies have reported similar effects upon ATRX loss at known ATRX target regions such as ERVs and imprinted loci (Sadic et al. 2015; Elsässer et al. 2015; Voon et al. 2015). However, the mechanisms underlying H3K9me3 loss in ATRX-deficient cells remain unclear. Based on our observations, we hypothesize three non-exclusive scenarios: 1) ATRX may facilitate SETDB1-mediated H3K9me3 deposition by promoting an optimal nucleosome structure for SETDB1 function, 2) ATRX binds H3K9me3 and blocks demethylase activity, and 3) ATRX may help to reestablish H3K9me3 after transcription or replication.

Functionally, loss of H3K9me3 in ATRX KO cells may impact on the stability of ZNFs. It was recently shown that regions containing high levels of H3K36me3 are prone to resolve double strand breaks (DSB) through the HR pathway (Aymard et al. 2014; Pfister et al. 2014). As ZNFs contain not only long stretches of highly similar sequences in their 3' exons, but also high levels of H3K36me3, these regions may be prone to recombine. As it has been shown that H3K9me3-rich heterochromatin regions are refractory to repair of DSB and HR (Murray et al. 2012; Kalousi et al. 2015), ATRX-mediated H3K9me3 enrichment at ZNF 3' exons may protect them from HR. In this regard, recent studies demonstrated that ATRX can act as a suppressor of recombination, particularly in the context of ALT (Napier et al. 2015; Clynes et al. 2015; He et al. 2015; Ramamoorthy and Smith 2015). For instance, ATRX mutations are associated with ALT-positive cancers that present active recombination at telomeres, which are known ATRX targets (Heaphy et al. 2011; Lovejoy et al. 2012; Bower et al. 2012; Schwartzentruber et

al. 2012). The mechanisms by which ATRX suppresses recombination remain unclear. We speculate that ATRX may indirectly suppress recombination by maintaining H3K9me3 levels in regions prone to recombination. Another hypothesis is that ATRX suppresses recombination by resolving stalled forks prone to DSB in specific late-replicating loci (Clynes and Gibbons 2013; Clynes et al. 2015). In agreement with this, we found that ZNFs with high levels of ATRX enrichment also tend to be late replicating. Furthermore, it has been shown that ATRX mutant cells are hypersensitive to inhibition of the master regulator of the DSB response ATR (Flynn et al. 2015). The functional link between ATR and ATRX remains unclear, but it adds to the idea that ATRX is an important player of the DSB/HR pathway.

Interestingly, ZNF gene duplications through HR are thought to be a common and important phenomenon for the evolution of the human genome and its transcription factor network (Tadepally and Aubry 2010; Nowick et al. 2011). Therefore, ATRX may be a key regulator of ZNF stability and thus an important driver of human genome evolution. Also, an altered recombination level between ZNFs due to the loss of H3K9me3 may be responsible, at least in part, for the defects in cell cycle and an increased percentage of DNA damage we observed in ATRX KO cells. However, we cannot exclude the possibility that this is induced by telomeric or pericentromeric dysfunction.

Several recent studies have found that ATRX deposits the euchromatin-associated H3.3 histone variant at several H3K9me3-containing regions (Sadic et al. 2015; Elsässer et al. 2015; Voon et al. 2015). Interestingly, the H3K9me3 levels in those regions are frequently decreased upon ATRX inactivation. How similar these H3.3/H3K9me3 regions are to the H3K9me3/H3K36me3 atypical chromatin signature remains an open question.

However, there is growing evidence to suggest that ATRX is a global regulator of H3K9me3 regions containing euchromatic marks (Sadic et al. 2015; Elsässer et al. 2015; Voon et al. 2015).

In summary, we demonstrate here that ATRX regulates H3K9me3 levels at the 3' exons of ZNFs and other loci containing an atypical chromatin signature. This unexpected function sheds light onto the complex genomic regulatory pathways that ATRX participates in, and may be important for the future understanding of diseases in which ATRX is mutated or altered.

METHODS

XL-MNase ChIP

Cross-linked-MNase ChIP was performed with Cell Signaling SimpleChIP Enzymatic Chromatin IP Kit (cat. #9003) following manufacturer's instructions with modifications. Briefly, K562 ($4X10^6$) cells were cross-linked with 1% formaldehyde in PBS, 100mM NaCl, 1mM EDTA pH 8.0, 50mM HEPES pH 8.0, for 10 minutes at room temperature. Reaction was quenched with 125 mM glycine. Cells were lysed to obtain nuclei and chromatin was digested with Micrococal Nuclease (MNase) (NEB, cat. #M0247S) at 37°C for 20 minutes. Nuclei were disrupted by brief sonication (4 cycles, 20 sec ON/OFF, high power) in a Bioruptor Twin. Chromatin was quantified and 40-80 ug was incubated with antibody at 4°C overnight with 1% taken as input sample. After incubation, Magna protein A/G magnetic beads (Millipore, cat. #16-663) were added for 3 hours at 4°C. Beads were washed following the manufacturer's protocol, followed by an extra LiCl buffer wash (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 1% sodium deoxicholate, 1% Igepal, 250mM LiCl). DNA bound to complexes was eluted at 65°C for 30 mins, treated with RNAse A (10mg/ml) for 1 hour at 37°C, with Proteinase K (20mg/ml) for 3 hours at 55°C and then cross-linking was reversed for 4-6 hours at 65°C. DNA was purified using the Qiagen MinElute PCR purification kit and subsequently analyzed and quantified using an Agilent 2100 Bioananalyzer High Sensitivity Kit.

Native ChIP

Native ChIP was performed as previously described (Hasson et al. 2013) with minor changes. Briefly, nuclei isolated from K562 ($6x10^6$) cells were treated with MNase (Affymetrix Cat #70196Y) and ~80ug of digested chromatin was immunoprecipitated with

specific antibodies. The immunoprecipitated material was treated with Proteinase K for 3h at 56 °C and purified using the Qiagen MinElute PCR purification kit.

Antibodies

See Supp. Table S4 for a full list of antibodies and concentrations used for each assay.

Library preparation and ChIP-seq

ChIP-seq libraries were prepared as previously described (Hasson et al. 2013) and libraries were sequenced single-end on Illumina HiSeq 2500. See Supp. Table S5 for detailed descriptions (number and size of reads, etc.) of the sequenced samples.

ChIP-seq analysis

Sequenced reads were aligned to the GRCh37 (hg19) assembly using Bowtie 1.0.0 (Langmead et al. 2009). Redundant reads were eliminated using the MACS2 (2.1.0) (Zhang et al. 2008) filterdup option with default parameters. The estimated background reads and the optimal normalization factor between ChIP and Input samples was calculated with the R NCIS package 1.0.1 (Liang and Keleş 2012) using a shift size of 75bp. Peak calling was performed with MACS2 callpeak using the --ratio option with the estimated value from NCIS. ChIP/Input fold enrichment pileups were created with the macs2 bdgcmp tool using the -m FE option and converted to bigWig files using the bedGraphToBigWig program (v4) from the UCSC binaries. For some samples, a second peak calling was performed using SICER (Zang et al. 2009) 1.1 using a window of 200bp, allowing gaps of 400pb and filtering for q-values $< 1x10^{-8}$. All peaks overlapping with the Encode blacklisted regions were eliminated. When available, a final list of peaks overlapping the MACS2 and SICER peaks was obtained from the intersectBed program

from bedtools 2.17.0. See Supp. Table S6 for GEO accessions, details of analysis parameters (q-values, etc.), and results for all the datasets analyzed.

Gene and coordinate analyses

All analyses were performed using Ensembl genes (putative genes, pseudogenes and genes unmapped to chromosomes were excluded). For human the Ensembl genes 75 version (GRCh37.p13) was used. For mouse the Ensembl genes 77 version (GRCm38.p3) was used. The lists of ZNF genes were downloaded from Biomart, using the InterPro id for the Zinc Finger C2H2 domain (IPR007087) as a filter, manually analyzed and curated. Repeats coordinates were obtained from the repmask table from the UCSC Genome Browser for hg19. Bed files for subsequent analysis were generated using the following coordinates: promoters (-3kb to TSS), intergenic (regions falling outside genes or promoters).

Correlation analysis

Correlation heatmaps were generated with the bigwigCorrelate program from the DeepTools suite (v 1.5.9.1) (Ramírez et al. 2014) using the spearman correlation method. Fold enrichment over input bigwig files were used as inputs. For genome-wide analysis, all chromosomes were divided in 10kb non-overlapping bins. Bins falling in the blacklisted regions from Encode were excluded. For gene-specific analysis, a bed file with the coordinates from TSS to TES was used.

Analysis of genomic distribution and correlation with chromatin states

Bed files with the coordinates of the chromatin states for K562 calculated in Ernst et al. 2011 were generated. Similar chromatin states were merged into single categories

(states 1-3 were fused as promoters, states 4-7 as enhancers, 10-11 as transcription, 14-15 as repetitive). The probability of overlap between ATRX peaks and the HMM states, genes, promoters and intergenic regions was calculated with The Genomic HyperBrowser (Sandve et al. 2013). ATRX peaks were randomized 10,000 times preserving its segment length and inter-segment gaps. Observed/expected values were calculated by dividing the overlap of the ATRX peaks over the overlap of the randomized regions.

Gene Ontology analysis

Genes overlapping ATRX significant peaks were obtained using intersectBed from the Bedtools suite. Gene Ontology analysis was performed with DAVID (Huang et al. 2009) using default parameters. See Supp. Table S1 for the list of genes overlapping ATRX peaks.

Generation of ZNF classes

Average ATRX ChIP/Input enrichment per gene was obtained with the computeMatrix program from the DeepTools suite. The genes were then clustered by kmeans into 3 groups using R (v 3.0.1). Bed files from each ZNF Class are provided in Supp. File S1.

Protein domain analysis

The protein sequence of the ZNFs was obtained with Biomart. The sequences were matched to all the Prosite motifs using the ScanProsite tool with default parameters. In-house scripts were used to parse the ScanProsite output and count the number of motifs per ZNF gene.

G-content and g-quadruplex analysis

G-content at the 3' region of ZNF genes (last 3kb) was calculated with in-house scripts. The probability of G-quadruplex formation was calculated using the quadparser program (Huppert and Balasubramanian 2005) with default parameters.

Average enrichment analyses

The color plots showing the distribution of genetic features among ZNF classes were drawn using the image function of R. Darker colors represent presence of KRAB domains, a higher number of zinc finger motifs, high G content at the C-terminal ZNF region (last 3kb of the gene) and presence of sequences predicted to form G-quadruplexes. For RNA-seq data, the Z-score of the RNA-seq normalized signal (log2(RPKM+1)) from K562 ENCODE data was calculated and plotted. Red colors match high expression signals, while blue colors match low expression signals. ZNFs are sorted from high to low ATRX enrichment from top to bottom. The calculated values used to generate color plots of Figure 2C are provided in Supp. File S2.

Metagene analysis

ZNF genes were ordered by ATRX enrichment from high to low. The metagene plots were generated with DeepTools using the ordered gene file and the ChIP/Input bigwig tracks. All genes from TSS to TES were scaled to a 5kb region +/- 1kb with sliding windows of 100 bp. Metagene values were calculated using computeMatrix. Values were plotted with heatmapper and the average enrichment profiles were plotted with profiler. The matrices for all heatmaps and profile plots are provided in Supp. File S3.

Motif analysis

The ZNF274-bound regions were obtained from the summit peaks generated by MACS2. A fasta file from these coordinates was created using in-house scripts and used as input for MEME-ChIP (v. 4.1.0.0) (Machanick and Bailey 2011). The sequences were compared to the JASPAR vertebrate database; other options were set as defaults. The regions matching the motifs (fimo output) were parsed, sorted by class and counted using in-house scripts.

Repli-seq analysis

Repli-seq bigwig files generated by the ENCODE project for K562 were analyzed using DeepTools. The average signal per ZNF was calculated with computeMatrix and then plotted in R using the heatmap.2 function. Calculated Repli-seq values are provided in Supp. File S4.

Repeat enrichment analysis

Fastq files (trimmed to 50bp when reads are >50bp) were aligned against the human hs36 repeat database from repbase using the Repeat Enrichment Estimator program (Day et al. 2010) with default parameters. The output was parsed by repeat type and the average and SEM Mle values for each repeat type were calculated and plotted.

Generation of ATRX KO cell lines by CRISP/Cas9

Single guide RNAs (sgRNAs) targeting the exons of the ATRX gene were designed (Ran et al. 2013) and cloned into the lentiCRISPR V2 vector (addgene #52961, Sanjana et al. 2014). Lentiviral production using HEK293T cells was performed using standard laboratory protocols. To generate stable cell lines, K562 (~$1x10^6$) cells were infected with virus for each CRISPR guide and the empty lentiCRISPR V2 was used as a control.

Infected cells were selected (puromycin 2ug/ml; 3 weeks) and subsequently single clones were sorted into 96-well plates using an IMI5L cell sorter (BD Biosciences). Clones were grown in selection for 3-4 weeks and tested for KO by DNA sequencing and western blot analysis. Two individual clones from two independent sgRNAs were selected and expanded (KO1 and KO2) for further analyses. One clone from the CAS9-only infection (V2) was randomly selected as a control. The four most likely off-target loci of each sgRNA were also sequenced to assure specificity of the KO (data not shown).

Comet assay

The alkaline comet assay was performed as described with modifications (Singh et al. 1988). K562 ($1X10^5$) cells were washed with cold PBS, resuspended, and diluted in 100 uL of 0.5% low melting point agarose to be pipetted onto slides covered with 1.5% agarose. Cells were lysed (2.5M NaCl, 100 mM EDTA, 10mM Tris, pH 10, 1% Triton and 10% DMSO) for 24h at 4^oC, incubated in electrophoresis buffer (300 mM NaOH, pH 13,1 mM EDTA) for 30 min and subjected to electrophoresis in the dark for 25 min at 25 V and 300 mA. Slides were then neutralized 3x with Tris buffer (0.4M Tris, pH 7.5) for 5 min, dried with 100% ethanol and stained with ethidium bromide (20 ug/mL). Cells were imaged on a Nikon Eclipse® microscope and ≥100 random cells were analyzed with CellProfiler software. Cell images were segmented using pixel intensity of 0.5 as threshold to generate masks matching the nucleoid. The comet tail was calculated by subtracting the nucleoid-integrated intensity from the comet-integrated intensity. For each sample, a positive control with cells treated with hydrogen peroxide (H_2O_2) (100uM for 30 min at 25^oC) was analyzed concurrently. Experimental analysis was performed in a blinded fashion.

Cell cycle analysis

K562 ($3x10^6$) cells were pulsed with 10 uM of BrdU for 45 min, washed with PBS, and fixed in 70% cold ethanol for 1 hour on ice. The cells were then washed 2x with cold PBS and incubated in 2M hydrochloric acid for 30 min at room temperature. After incubation, cells were washed 2x with cold PBS and once with cold PBS-T solution (PBS, 0.2% Tween, 1% BSA), then stained with anti-BrdU FITC PBS-T (eBioscience, cat #11-5071-41) for 30 min in the dark. Cells were washed once with cold PBS-T, once more with cold PBS and then stained for 30 min on ice with a PI/RNAse solution (PBS, 20ug/ml PI, 10ug/ml RNAse). FACS analysis (10,000 cells per sample) was performed using BD FACS Canto II and FlowJo 10.

H3K9me3 enrichment analysis

The significantly reduced H3K9me3 domains in the K562 ATRX KO2 H3K9me3 ChIP-seq were calculated with the SICER-df program (Zang et al. 2009). Only regions overlapping significant H3K9me3 peaks in the V2 control were taken into account. The average ATRX content of the reduced regions was calculated with computeMatrix. The regions were k-mean clustered into two groups by their average ATRX content in R. Average ChIP-seq profiles of KAP1, TRIM28, ZNF274 and H3K36me3 were calculated and drawn with computeMatrix and profiler. Bed files with the reduced regions coordinates are provided in Supp. File S5.

Statistical analyses

The hypergeometric test and permutation tests implemented in the R coin package (v 1.0-24) were the primary statistical tests utilized (See Supp. Table 2 for details). See Supp. Table S3 for the results of all the performed tests.

ChIP-qPCR and RT-qPCR

ChIP-qPCR and RT-qPCR were performed as described (Vardabasso et al. 2015). See Supp. Table S7 for the list of primers used in our ChIP-qPCR and RT-qPCR experiments.

DATA ACCESS

All datasets were deposited to Gene Expression Omnibus (GEO) with accession number GSE70920.

ACKNOWLEDGMENTS

The authors thank the Genomics Core Facility and the Flow Cytometry Center at Mount Sinai for their technical assistance and Matthew O'Connell and Kajan Ratnakumar for helpful discussions and guidance. Funding was provided by the DGAPA-PAPIIT, UNAM (IN209403, IN203811 and IN201114), and CONACyT (42653-Q, 128464 and 220503) to FR-T, Graduate fellowship from CONACyT (239663, CVU 257385) to DV-G, NCI T32-CA078207 to ZAQ, NIH EY014867, EY018599 and CA168875, Cancer Center Support from the NCI (CA21765), support from the American Lebanese Syrian Associated Charities (ALSAC) and a grant from Alex's Lemonade Stand Foundation for Childhood Cancer to MAD, and NCI/NIH R01CA154683 to EB. This article is part of the doctoral thesis of DV-G from the Doctorado en Ciencias Biomedicas, UNAM.

DISCLOSURE DECLARATION

The authors declare no conflict of interest.

FIGURE LEGENDS

Figure 1. ATRX binds to the 3' regions of ZNF genes. (A) Observed over expected random distribution of significant ATRX peaks within HMM chromatin categories (Ernst et al. 2011). In (A) and (C) error bars represent standard deviation. Asterisks represent significantly overrepresented regions (*= pvalue < 0.05; ****= pvalue < $1x10^{-4}$) assessed by the hypergeometric test (see Supp. Tables 2 and 3 for details of statistical tests). (B) Gene Ontology analysis of genes that overlap with significant ATRX peaks in their gene bodies. (C) Observed over expected random distribution of significant ATRX peaks in ZNFs and non-ZNF (other) promoters and gene bodies. The red line represents the expected value of a random distribution. (D) Hilbert curve plot of chromosome 19 showing the ZNF clusters (red, left), highly enriched ATRX regions (fold enrichment over input >5, blue, middle), and overlap (right). The blue dot and the red arrow mark the start (5') and end (3') of the chromosome, respectively (see diagram on left). The green dots indicate the start and end of centromere, which is excluded from the analysis. (E) UCSC Genome browser screenshot of a typical ZNF cluster (genes in blue) on chromosome 19. The enrichment over input signal for ATRX, H3K9me3 and H3K36me3 ChIP-seq is shown. Significant peaks represented as bars below enrichment tracks. Normalized RPKM signal for RNA-seq is also shown. (F) Zoomed in snapshot of two ZNFs contained within the ZNF cluster shown in (E). (G) Average enrichment ChIP-seq profiles of ATRX, H3K9me3 and H3K36me3 over all ZNF gene bodies +/- 1kb (n = 736).

Figure 2. ATRX and the ZNF274/TRIM28/SETDB1 complex bind to ZNF genes with an atypical chromatin signature and distinctive genomic and epigenetic features. (A) Average enrichment ChIP-seq profiles of ATRX, H3K9me3 and H3K36me3 at ZNFs

classified by their ATRX content. Class I contains high levels of ATRX (n = 91), Class II contains medium to low levels of ATRX (n = 303) and Class III is devoid of ATRX enrichment (n = 342). (B) Spearman correlation heatmap of ChIP-seq signal at ZNF Class I genes (left) and ZNF Class III genes (right). (C) Left: Distribution of genetic features among ZNF classes (sorted from high to low ATRX enrichment from top to bottom). Dashed lines show separation of the three classes. Colors represent presence of KRAB domains (black), number of zinc finger motifs (pink), G content at the C-terminal ZNF region (last 3kb of the gene) (gray) and presence of sequences predicted to form G-quadruplexes (brown). RNA-seq bar shows the Z score of the normalized RPKM signal (log2(RPKM+1)); red = high expression signal and blue = low expression signal. For statistical tests between the classes see Supp. Table S3. Right: Box plots displaying the number of ZNF motifs, G-content at the ZNF region and RNA-seq values per ZNF Class. Asterisks show significant differences (pvalue < $1x10^{-4}$). (D) Metagene analysis of ChIP-seq enrichment over input profiles at ZNF gene bodies +/- 1kb. (E) Spearman correlation heatmaps between ChIP-seq profiles genome-wide (left) and at ZNF genes (right). Black boxes indicate the significant correlations.

Figure 3. ATRX binding to ZNF 3' exons is cell-type specific. (A) Metagene analysis of ZNF274 and H3K9me3 ChIP-seq enrichments over ZNFs +/- 1kb in HepG2; ZNF classes represented as in Figure 2. (B) ATRX ChIP-qPCR over ZNF genes in HepG2 cells (average of at least 3 independent biological replicates). Error bars represent the SEM. Asterisks depict statistically significant ATRX enrichments as compared to IgG (for details see Supp. Table S3) (C) ATRX ChIP-seq metagene profiles of ATRX enrichment over ZNFs in K562 cells and human erythroblasts. (D) ATRX and H3K9me3 ChIP-seq metagene profiles over ZNFs in two mouse cell types. Mouse ZNF genes were sorted

from top to bottom by their ATRX enrichment in Embryonic Stem Cells (mESC) from high to low.

Figure 4. ATRX deficient cells have decreased levels of H3K9me3 at 3' exons of ZNFs and other atypical chromatin regions. (A) Western blot of ATRX in chromatin preparations from control (V2) and two CRISPR ATRX KO K562 cell lines. Amido Black staining of histones is shown as a loading control. (B) ATRX ChIP-qPCR over ZNF genes in control and ATRX KO K562 cell lines. (C) Same as in (B) with H3K9me3 native ChIP-qPCR. In both (B) and (C), bars represent average of at least 3 biological replicates and the error bars represent the SEM. (D) Box plot of the H3K9me3 changes (represented in % with respect to the control) observed in the ZNF Classes and non-ZNF H3K9me3-bound genes upon ATRX KO. Asterisks show significant changes with respect to the non-ZNF genes (*= pvalue <0.05; ****= pvalue < $1x10^{-4}$). (E) ChIP-seq metagene profiles of H3K9me3 in control (V2) and ATRX KO (KO2) K562 cell lines over ZNFs and non-ZNF H3K9me3-bound genes. (F) Average enrichment ChIP-seq profiles of SETDB1, TRIM28, ZNF274 and H3K36me3 in regions that significantly lost H3K9me3 upon ATRX KO, separated by their ATRX enrichment status. All observed differences are statistically significant (pvalue < $1x10^{-4}$); see Supp. Table S3 for details. (G) Genomic distribution of H3K9me3 decreased regions in ATRX KO.

Figure 5. Model of ATRX regulation at ZNF 3' exons. Left: ATRX regulates H3K9me3 at ZNF 3' exons by facilitating SETDB1 or blocking the activity of demethylases. The presence of the mark establishes an atypical H3K9me3/H3K36me3 domain. Right: When ATRX is depleted, the H3K9me3 and the atypical chromatin domains at ZNF 3'

exons are lost. Loss of ATRX induced altered cell cycle, increased DNA damage and possibly recombination between ZNFs.

REFERENCES

Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM, Jackson SP, et al. 2014. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**: 366–74.

Bérubé NG. 2011. ATRX in chromatin assembly and genome architecture during development and disease. *Biochem Cell Biol* **89**: 435–44.

Bérubé NG, Smeenk C a, Picketts DJ. 2000. Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. *Hum Mol Genet* **9**: 539–547.

Blahnik KR, Dou L, Echipare L, Iyengar S, O'Geen H, Sanchez E, Zhao Y, Marra M a, Hirst M, Costello JF, et al. 2011. Characterization of the contradictory chromatin signatures at the 3' exons of zinc finger genes. *PLoS One* **6**: e17121.

Bower K, Napier CE, Cole SL, Dagg R a, Lau LMS, Duncan EL, Moy EL, Reddel RR. 2012. Loss of wild-type ATRX expression in somatic cell hybrids segregates with activation of Alternative Lengthening of Telomeres. *PLoS One* **7**: e50062.

Clynes D, Gibbons RJ. 2013. ATRX and the replication of structured DNA. *Curr Opin Genet Dev* **23**: 289–294.

Clynes D, Higgs DR, Gibbons RJ. 2013. The chromatin remodeller ATRX: a repeat offender in human disease. *Trends Biochem Sci* **38**: 461–6.

Clynes D, Jelinska C, Xella B, Ayyub H, Scott C, Mitson M, Taylor S, Higgs DR, Gibbons RJ. 2015. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat Commun* **6**: 7538.

Clynes D, Jelinska C, Xella B, Ayyub H, Taylor S, Mitson M, Bachrati CZ, Higgs DR, Gibbons RJ. 2014. ATRX dysfunction induces replication defects in primary mouse cells. *PLoS One* **9**: e92915.

Day DS, Luquette LJ, Park PJ, Kharchenko P V. 2010. Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol* **11**: R69.

Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. 2010. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* **24**: 1253–1265.

Elsässer SJ, Noh K-M, Diaz N, Allis CD, Banaszynski L a. 2015. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature*.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43–9.

Euskirchen G, Auerbach RK, Snyder M. 2012. SWI/SNF chromatin-remodeling factors: Multiscale analyses and diverse functions. *J Biol Chem* **287**: 30897–30905.

Eustermann S, Yang J-C, Law MJ, Amos R, Chapman LM, Jelinska C, Garrick D, Clynes D, Gibbons RJ, Rhodes D, et al. 2011. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat Struct Mol Biol* **18**: 777–82.

Filipescu D, Szenker E, Almouzni G. 2013. Developmental roles of histone H3 variants and their chaperones. *Trends Genet* **29**: 630–640.

Flynn RL, Cox KE, Jeitany M, Wakimoto H, Bryll AR, Ganem NJ, Bersani F, Pineda JR, Suva ML, Benes CH, et al. 2015. Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science (80-)* **347**: 273–277.

Frietze S, O'Geen H, Blahnik KR, Jin VX, Farnham PJ. 2010. ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. *PLoS One* **5**: e15082.

Goldberg AD, Banaszynski L a, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**: 678–91.

Gospodinov A, Herceg Z. 2013. Chromatin structure in double strand break repair. *DNA Repair (Amst)* **12**: 800–810.

Hargreaves DC, Crabtree GR. 2011. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* **21**: 396–420.

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. 2013. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat Struct Mol Biol* **20**: 687–95.

He Q, Kim H, Huang R, Lu W, Tang M, Shi F, Yang D, Zhang X, Huang J, Liu D, et al. 2015. The Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by Promoting H3K9 Trimethylation. *Cell Stem Cell* **17**: 273–86.

Heaphy CM, de Wilde RF, Jiao Y, Klein AP, Edil BH, Shi C, Bettegowda C, Rodriguez FJ, Eberhart CG, Hebbar S, et al. 2011. Altered telomeres in tumors with ATRX and DAXX mutations. *Science* **333**: 425.

Huang DW, Sherman BT, Lempicki R a. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.

Huh M, O'Dea T, Ouazia D. 2012. Compromised genomic integrity impedes muscle growth after Atrx inactivation. *J Clin …* **122**.

Huppert JL, Balasubramanian S. 2005. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**: 2908–2916.

Iwase S, Xiang B, Ghosh S, Ren T, Lewis PW, Cochrane JC, Allis CD, Picketts DJ, Patel DJ, Li H, et al. 2011. ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol* **18**: 769–776.

Iyengar S, Farnham PJ. 2011. KAP1 protein: An enigmatic master regulator of the genome. *J Biol Chem* **286**: 26267–26276.

Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*.

Julienne H, Zoufir A, Audit B, Arneodo A. 2013. Human Genome Replication Proceeds through Four Chromatin States. *PLoS Comput Biol* **9**.

Kalousi A, Hoffbeck A-S, Selemenakis PN, Pinder J, Savage KI, Khanna KK, Brino L, Dellaire G, Gorgoulis VG, Soutoglou E. 2015. The Nuclear Oncogene SET Controls DNA Repair by KAP1 and HP1 Retention to Chromatin. *Cell Rep* **11**: 149–163.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Längst G, Manelyte L. 2015. Chromatin Remodelers: From Function to Dysfunction. *Genes (Basel)* **6**: 299–324.

Law MJ, Lower KM, Voon HPJJ, Hughes JR, Garrick D, Viprakasit V, Mitson M, De Gobbi M, Marra M, Morris A, et al. 2010. ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner. *Cell* **143**: 367–378.

Lechner MS, Schultz DC, Negorev D, Maul GG, Rauscher FJ. 2005. The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochem Biophys Res Commun* **331**: 929–937.

Leung JW-CC, Ghosal G, Wang W, Shen X, Wang J, Li L, Chen J. 2013. Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *J Biol Chem* **288**: 6342–6350.

Levy M a., Kernohan KD, Jiang Y, Bérubé NG. 2015. ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions. *Hum Mol Genet* **24**: 1824–35.

Lewis PW, Elsaesser SJ, Noh K-M, Stadler SC, Allis CD. 2010. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* **107**: 14075–14080.

Liang K, Keleş S. 2012. Normalization of ChIP-seq data with control. *BMC Bioinformatics* **13**: 199.

Lovejoy C a, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, De S, Petrini JHJ, Sung P a, Jasin M, et al. 2012. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet* **8**: e1002772.

Machanick P, Bailey TL. 2011. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696–1697.

McDowell TL, Gibbons RJ, Sutherland H, O'Rourke DM, Bickmore W a, Pombo a, Turley H, Gatter K, Picketts DJ, Buckle VJ, et al. 1999. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **96**: 13983–8.

Murray JM, Stiff T, Jeggo P a. 2012. DNA double-strand break repair within heterochromatic regions. 173–178.

Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM, et al. 2015. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol* **advance on**.

Napier CE, Huschtscha LI, Harvey A, Bower K, Jane R, Hendrickson E a, Reddel RR. 2015. ATRX represses alternative lengthening of telomeres. *Oncotarget*.

Nowick K, Fields C, Gernat T, Caetano-Anolles D, Kholina N, Stubbs L. 2011. Gain, loss and divergence in primate Zinc-Finger genes: A rich resource for evolution of gene regulatory differences between species. *PLoS One* **6**.

Park DJ, Pask AJ, Huynh K, Renfree MB, Harley VR, Marshall Graves J a. 2004. Comparative analysis of ATRX, a chromatin remodeling protein. *Gene* **339**: 39–48.

Pfister SX, Ahrabi S, Zalmas LP, Sarkar S, Aymard F, Bachrati CZ, Helleday T, Legube G, LaThangue NB, Porter ACG, et al. 2014. SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell Rep* **7**: 2006–2018.

Ramamoorthy M, Smith S. 2015. Loss of ATRX Suppresses Resolution of Telomere Cohesion to Control Recombination in ALT Cancer Cells. *Cancer Cell* **28**: 357–369.

Ramírez F, Dündar F, Diehl S, Grüning B a., Manke T. 2014. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: 187–191.

Ran F, Hsu P, Wright J, Agarwala V. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281–308.

Ratnakumar K, Bernstein E. 2013. ATRX: the case of a peculiar chromatin remodeler. *Epigenetics* **8**: 3–9.

Ratnakumar K, Duarte LF, LeRoy G, Hasson D, Smeets D, Vardabasso C, Bönisch C, Zeng T, Xiang B, Zhang DY, et al. 2012. ATRX-mediated chromatin association of histone variant macroH2A1 regulates α-globin expression. *Genes Dev* **26**: 433–8.

Ray-Gallet D, Woolfe A, Vassias I, Pellentz C, Lacoste N, Puri A, Schultz DC, Pchelintsev N a., Adams PD, Jansen LET, et al. 2011. Dynamics of Histone H3 Deposition In Vivo Reveal a Nucleosome Gap-Filling Mechanism for H3.3 to Maintain Chromatin Integrity. *Mol Cell* **44**: 928–941.

Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, Theis FJ, Schotta G. 2015. Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep* 1–15.

Sandve GK, Gundersen S, Johansen M, Glad IK, Gunathasan K, Holden L, Holden M, Liestøl K, Nygård S, Nygaard V, et al. 2013. The Genomic HyperBrowser: an analysis web server for genome-scale data. *Nucleic Acids Res* **41**: 133–141.

Sanjana NE, Shalem O, Zhang F. 2014. *Improved vectors and genome-wide libraries for CRISPR screening*.

Schneiderman JI, Orsi G a, Hughes KT, Loppin B, Ahmad K. 2012. Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. *Proc Natl Acad Sci U S A* **109**: 19721–6.

Schwartzentruber J, Korshunov A, Liu X-Y, Jones DTW, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang D-AK, Tönjes M, et al. 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**: 226–31.

Singh NP, McCoy MT, Tice RR, Schneider EL. 1988. A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp Cell Res* **175**: 184–191.

Sinha M, Peterson CL. 2009. Chromatin dynamics during repair of chromosomal DNA double-strand breaks. *Epigenomics* **1**: 371–385.

Tadepally HD, Aubry M. 2010. Evolution of C2H2 Zinc Finger Gene Families in Mammals. *eLS* 1–9.

Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Res* **21**: 1800–1812.

Vardabasso C, Gaspar-Maia A, Hasson D, Pünzeler S, Valle-Garcia D, Straub T, Keilhauer EC, Strub T, Dong J, Panda T, et al. 2015. Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. *Mol Cell* 75–88.

Vogel MJ, Guelen L, De Wit E, Peric-Hupkes D, Lodén M, Talhout W, Feenstra M, Abbas B, Classen AK, Van Steensel B. 2006. Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res* **16**: 1493–1504.

Voon HPJ, Hughes JR, Rode C, De La Rosa-Velázquez IA, Jenuwein T, Feil R, Higgs DR, Gibbons RJ. 2015. ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Rep* 405–418.

Watson LA, Solomon L a., Li JR, Jiang Y, Edwards M, Shin-Ya K, Beier F, Bérubé NG. 2013. Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. *J Clin Invest* **123**: 2049–2063.

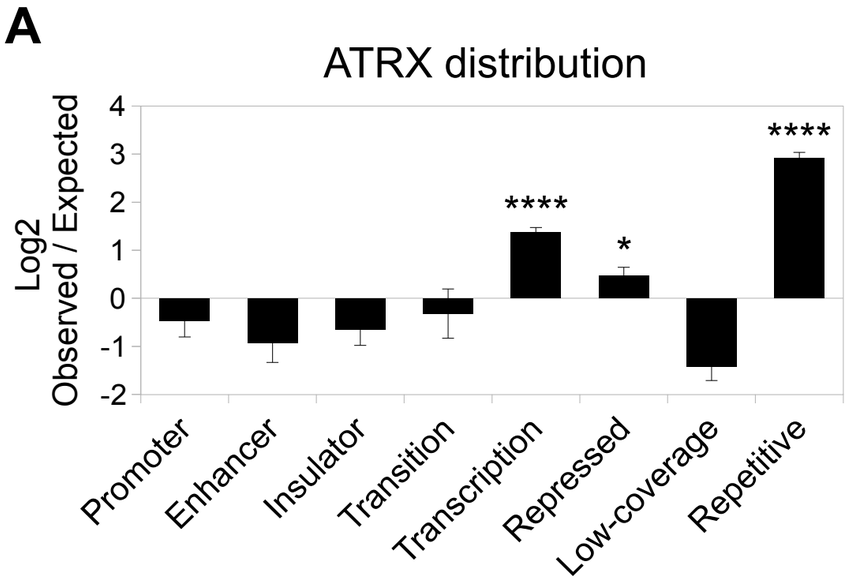
Wong LH, McGhie JD, Sim M, Anderson M a, Ahn S, Hannan RD, George AJ, Morgan K a, Mann JR, Choo KHA. 2010. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* **20**: 351–60.

Xue Y, Gibbons R, Yan Z, Yang D, McDowell TL, Sechi S, Qin J, Zhou S, Higgs D, Wang W. 2003. The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc Natl Acad Sci U S A* **100**: 10635–40.

Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**: 1952–8.

Zhang Y, Liu T, Meyer C a, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

A

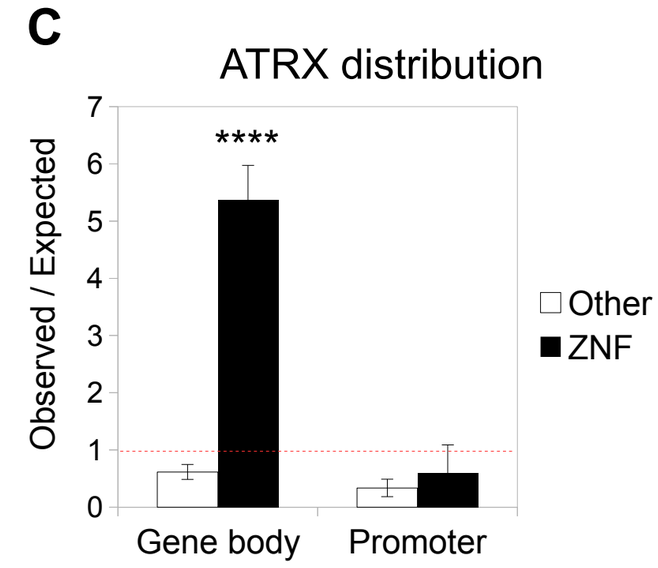


B

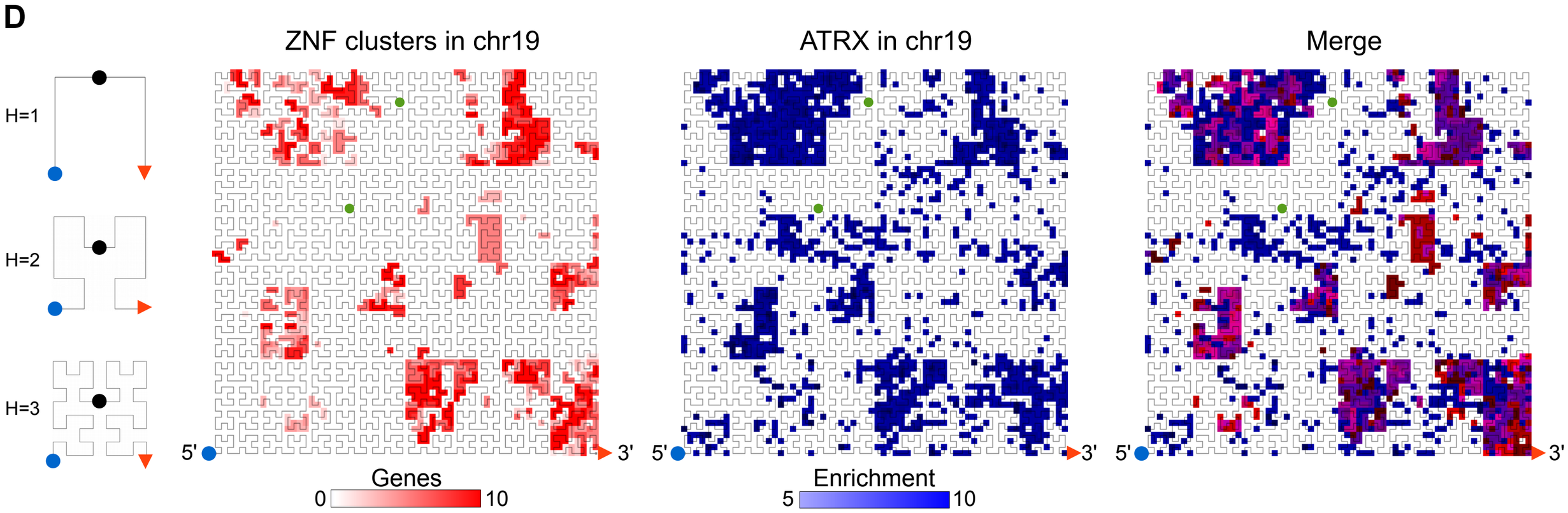
Gene Ontology of ATRX-bound genes

GO Term	Genes	FDR
Domain: KRAB	67	6.1E-49
Zinc Finger, C2H2-like	83	4.3E-37
Regulation of transcription, DNA dependent	96	1.0E-20
Transcription	102	8.3E-19
Regulation of transcription	109	1.3E-15

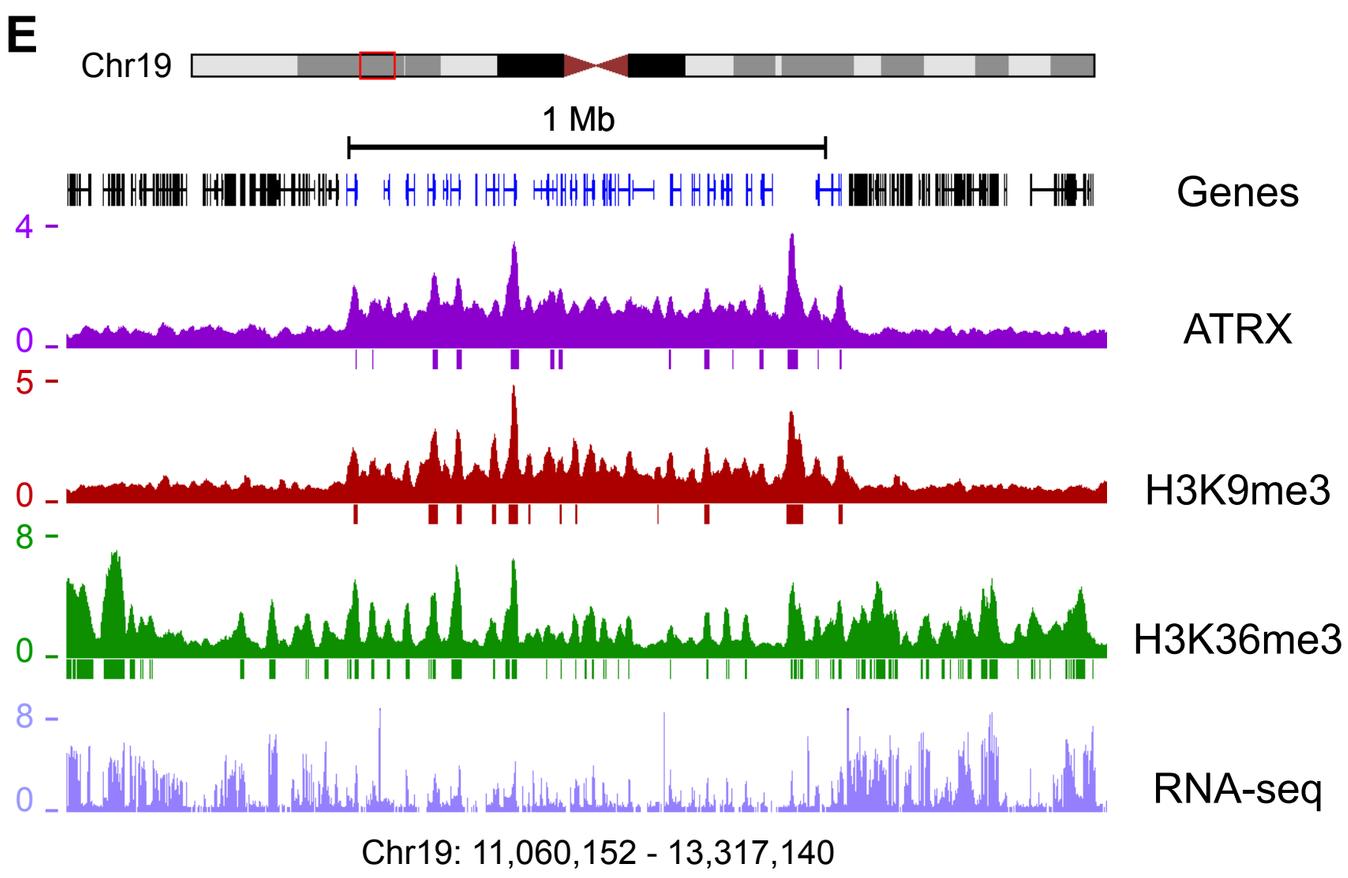
C



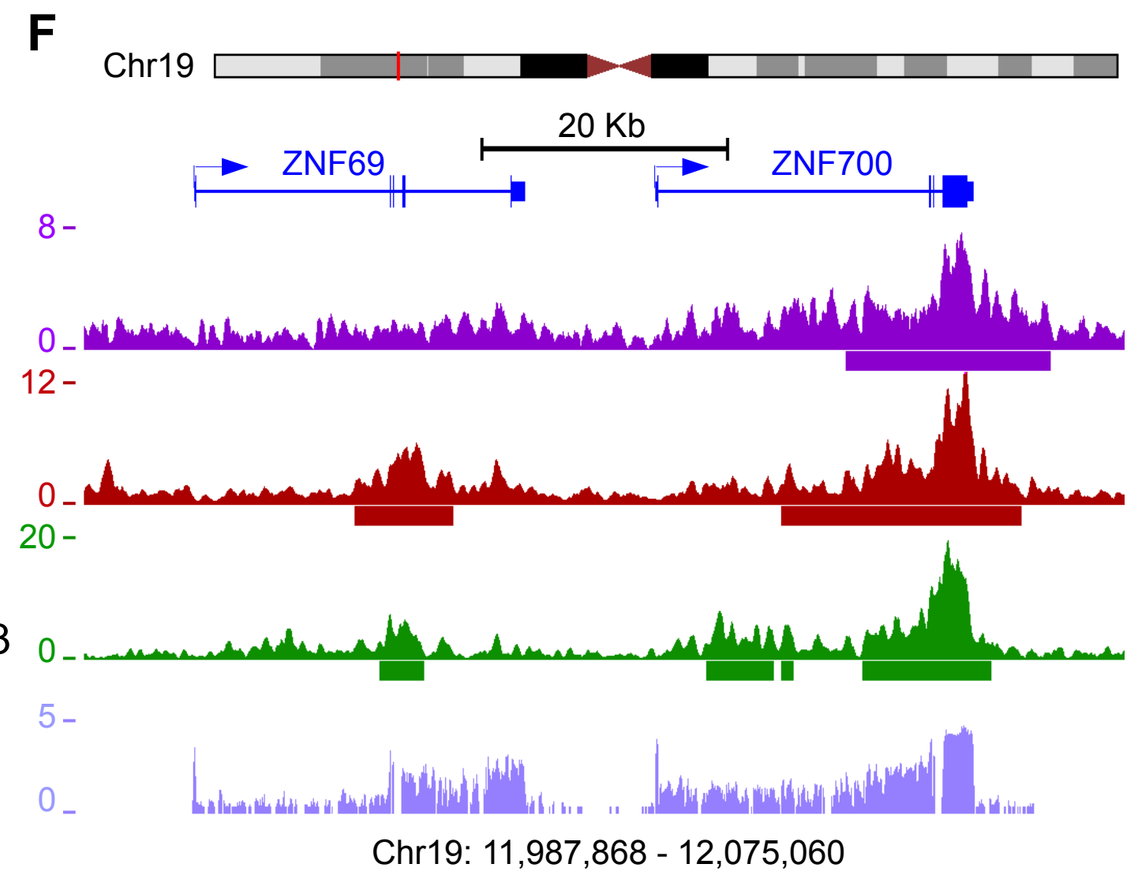
D



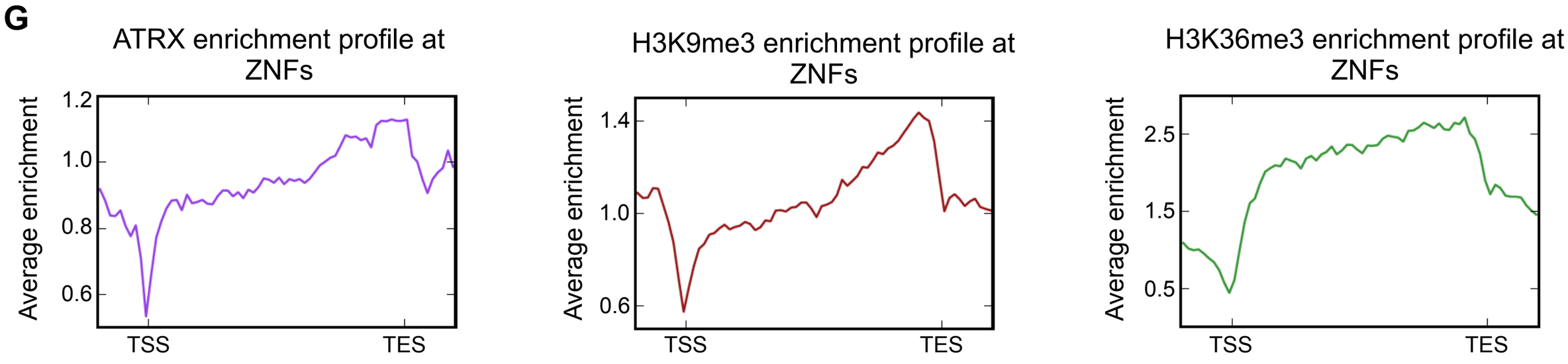
E

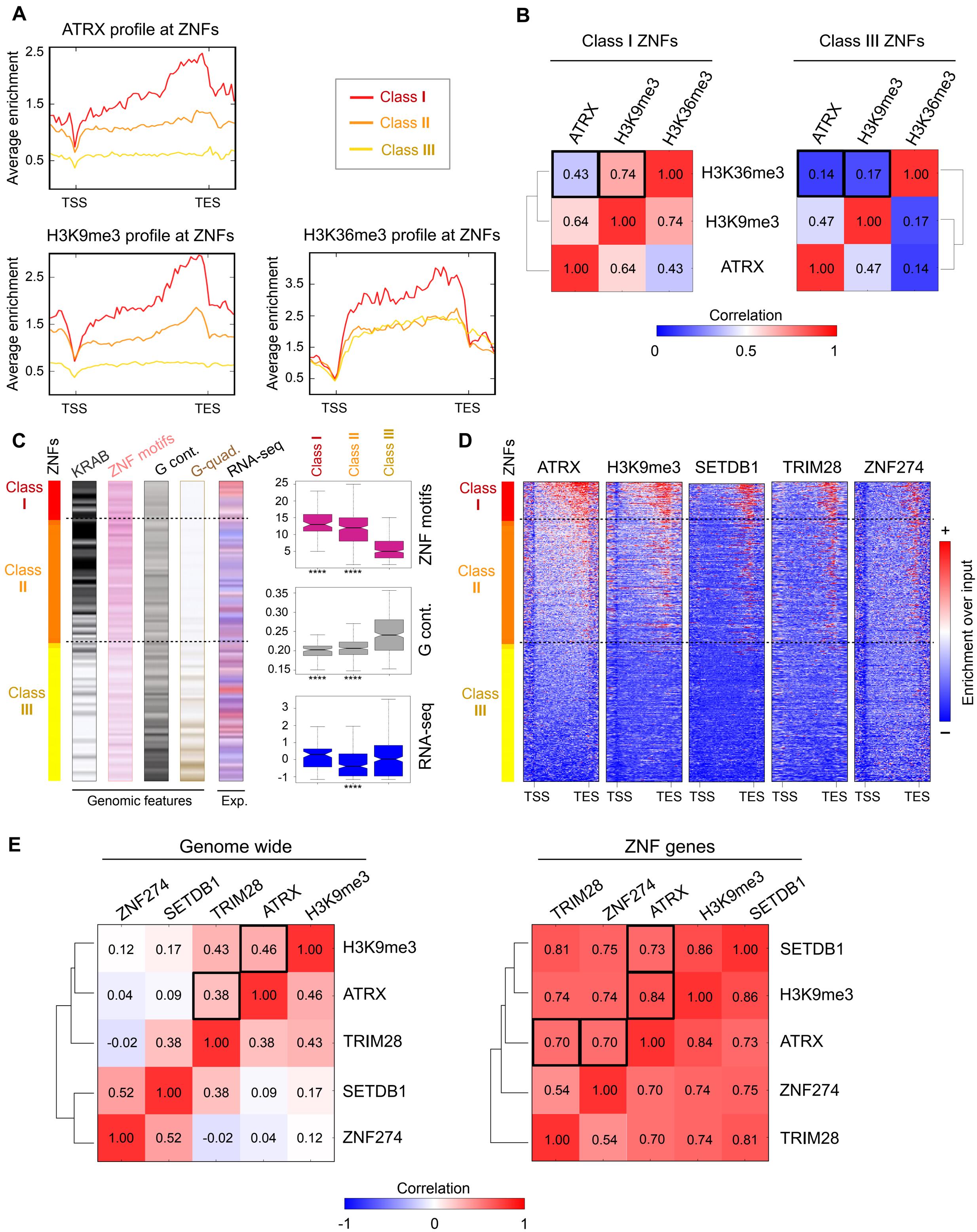


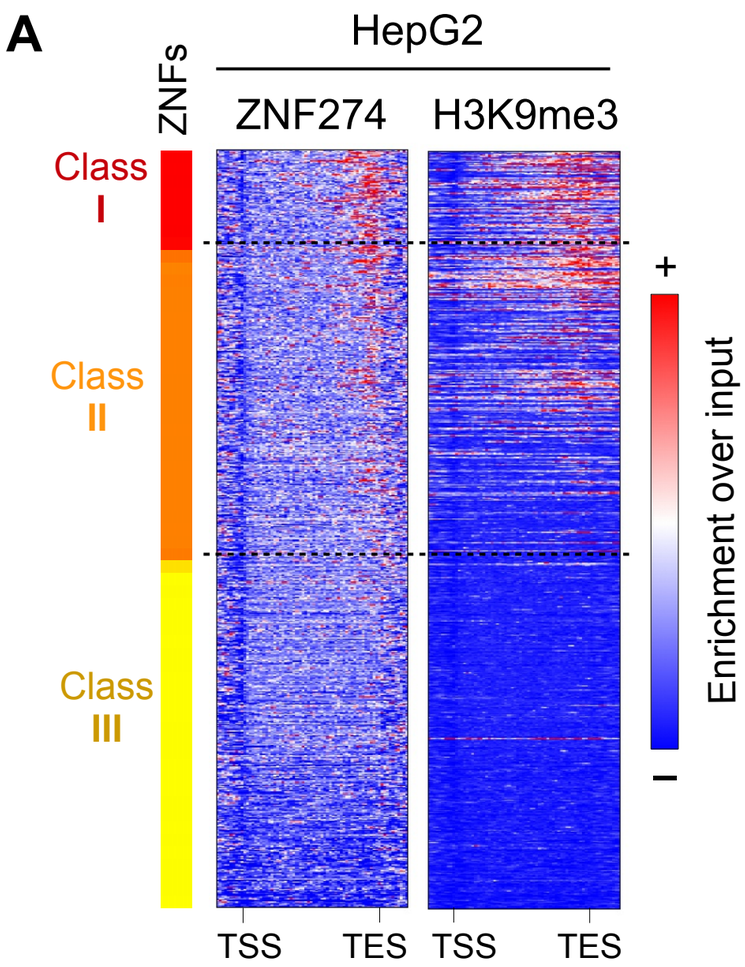
F

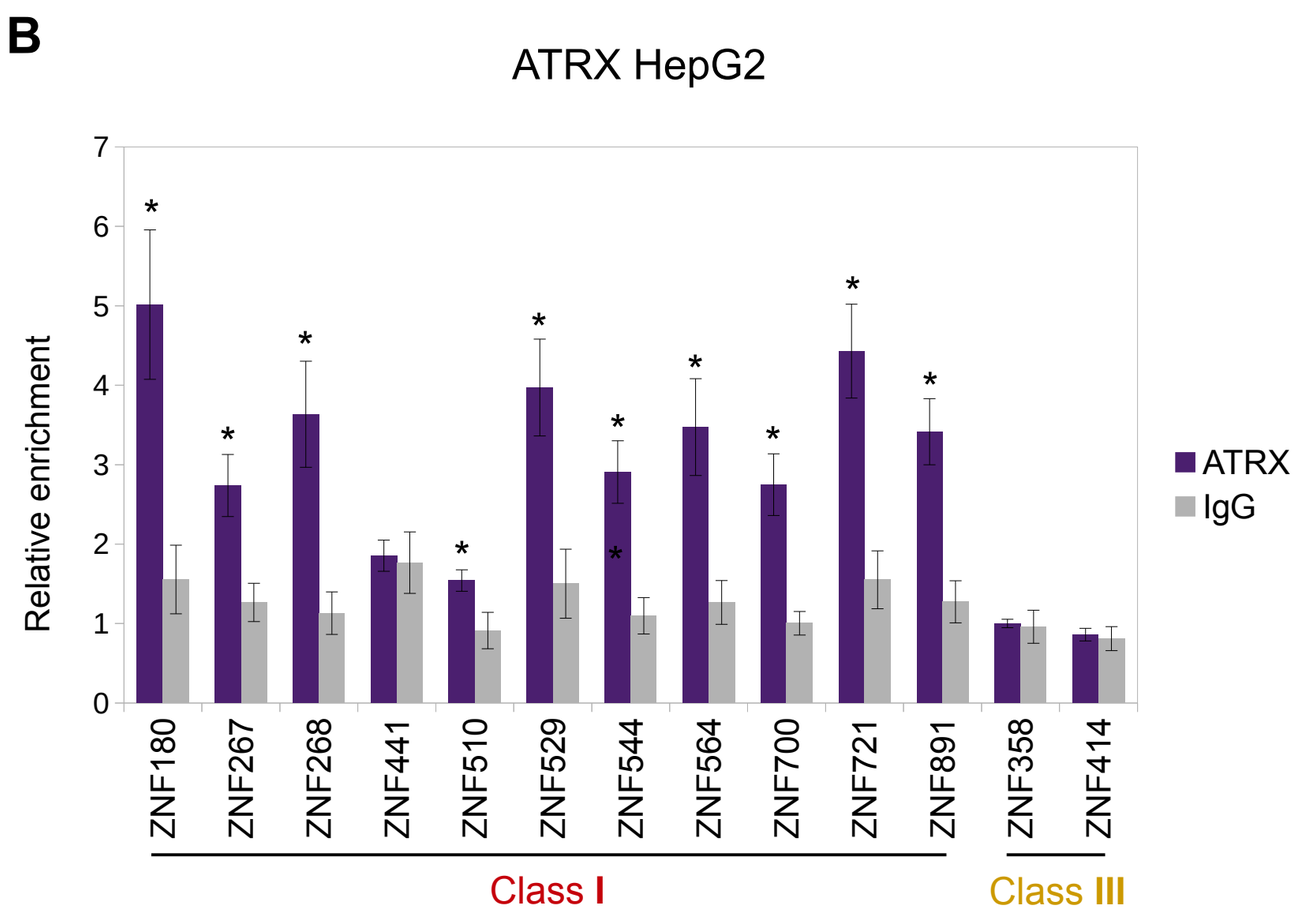


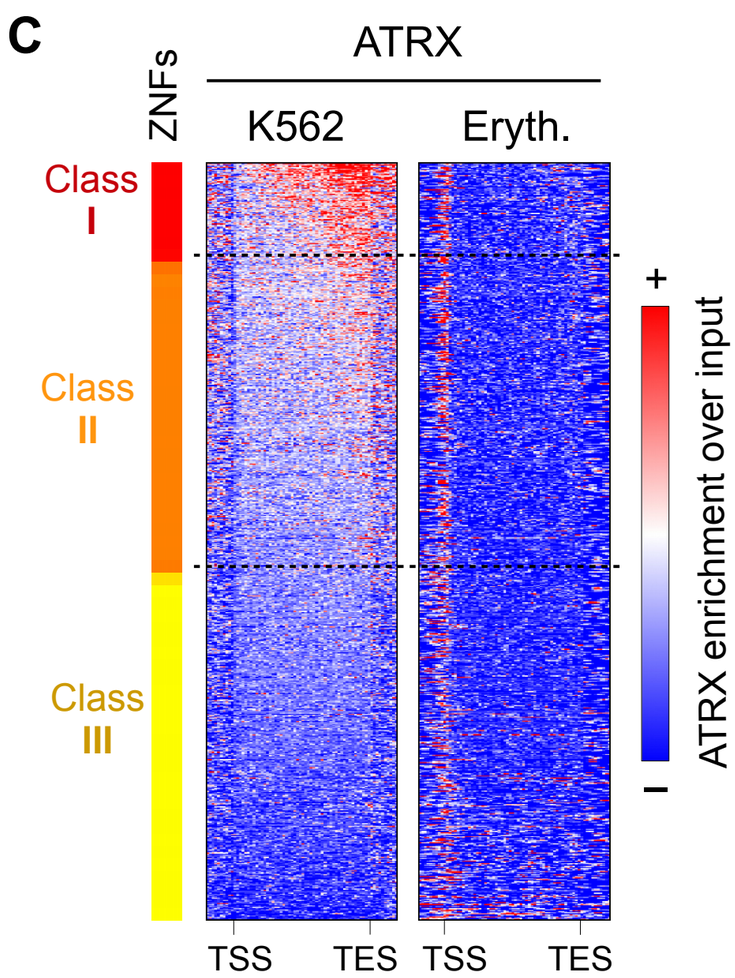
G

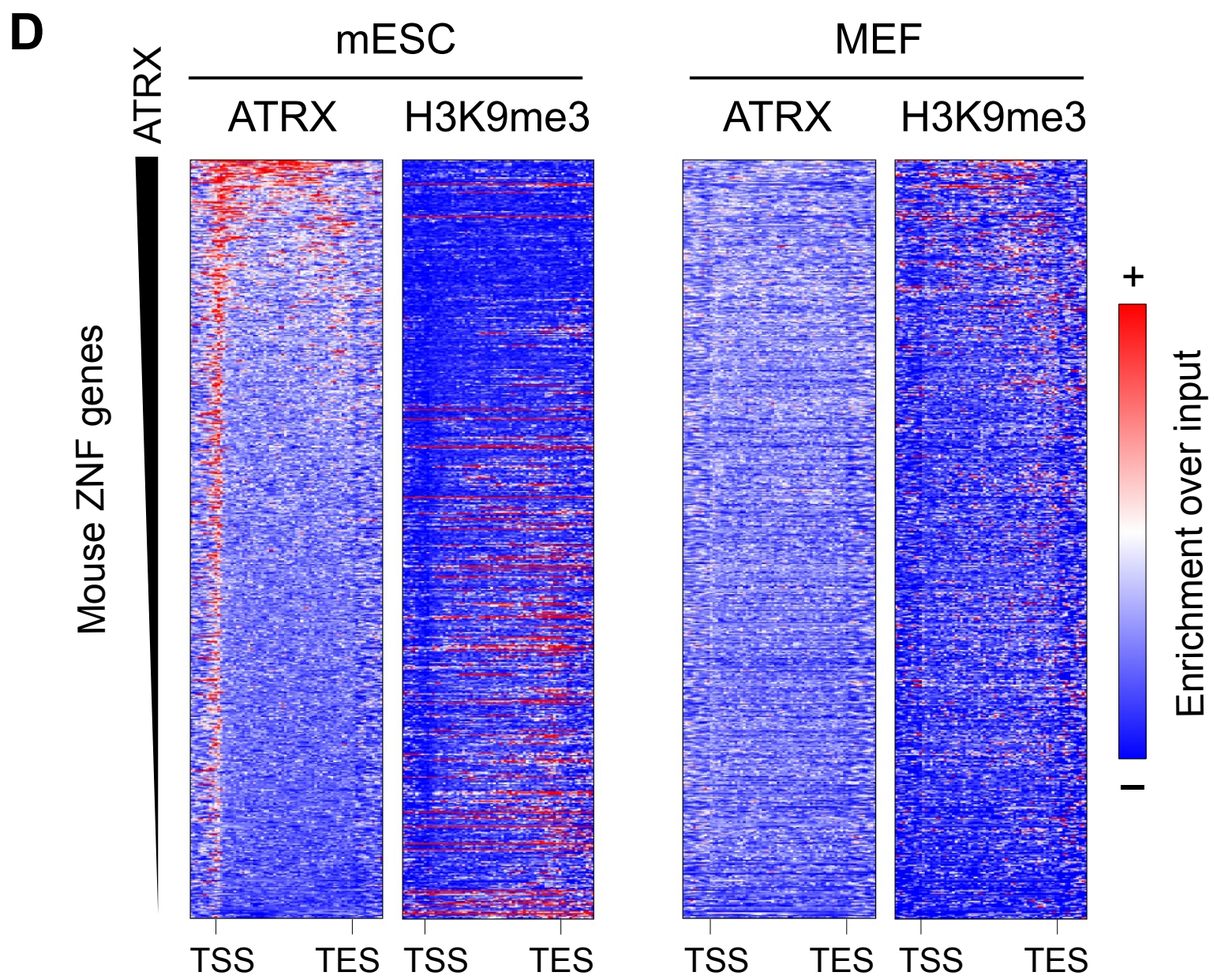


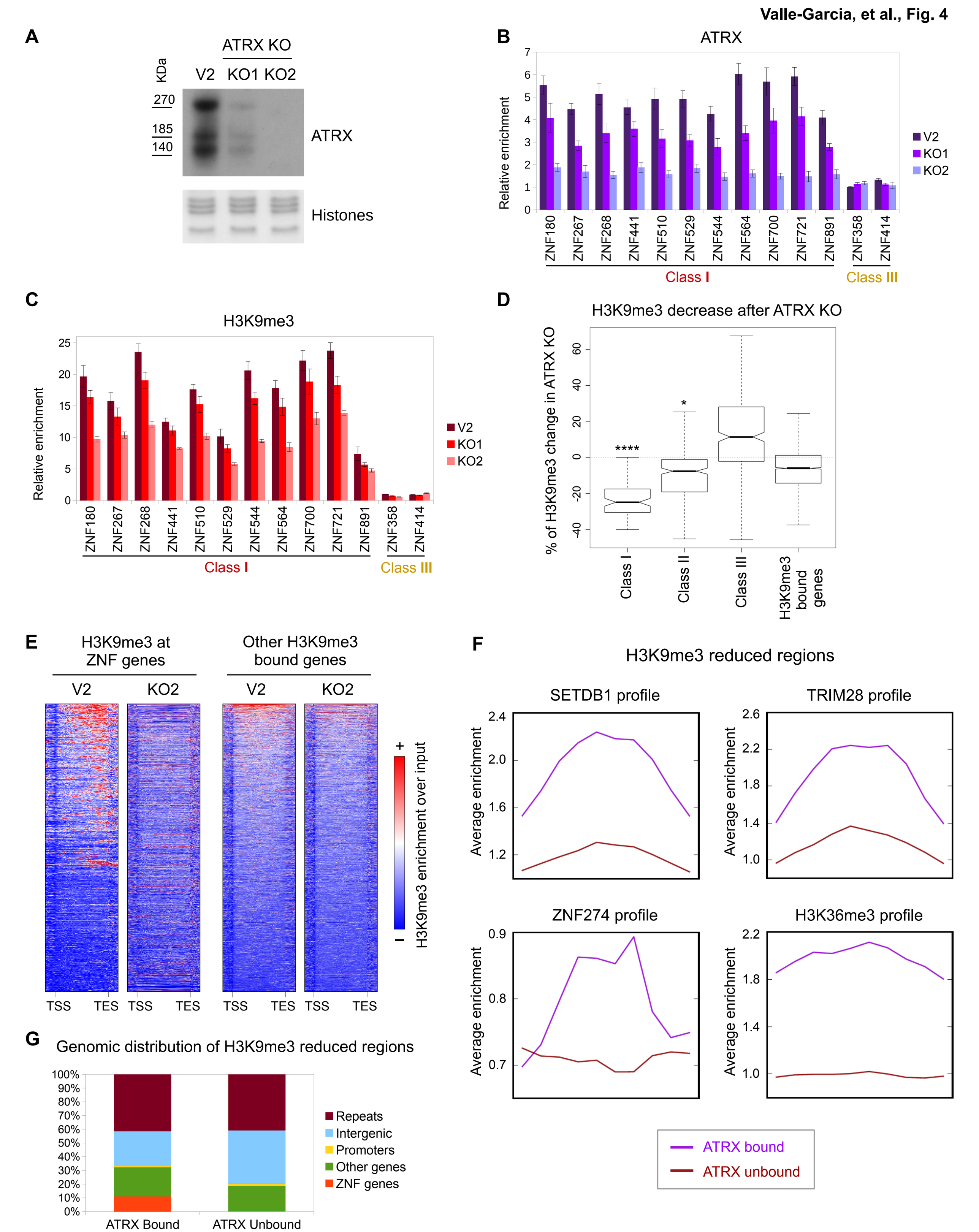

A
ATRX profile at ZNFs
H3K9me3 profile at ZNFs
H3K36me3 profile at ZNFs
Average enrichment
TSS
TES
Class I
Class II
Class III
B
Class I ZNFs
Class III ZNFs
ATRX
H3K9me3
H3K36me3
0.43 0.74 1.00
0.64 1.00 0.74
1.00 0.64 0.43
0.14 0.17 1.00
0.47 1.00 0.17
1.00 0.47 0.14
Correlation
0
0.5
1
C
ZNFs
KRAB
ZNF motifs
G cont.
G-quad.
RNA-seq
Genomic features
Exp.
ZNF motifs
G cont.
RNA-seq
****
D
ATRX
H3K9me3
SETDB1
TRIM28
ZNF274
Enrichment over input
E
Genome wide
ZNF genes
ZNF274
SETDB1
TRIM28
ATRX
H3K9me3
0.12 0.17 0.43 0.46 1.00
0.04 0.09 0.38 1.00 0.46
-0.02 0.38 1.00 0.38 0.43
0.52 1.00 0.38 0.09 0.17
1.00 0.52 -0.02 0.04 0.12
0.81 0.75 0.73 0.86 1.00
0.74 0.74 0.84 1.00 0.86
0.70 0.70 1.00 0.84 0.73
0.54 1.00 0.70 0.74 0.75
1.00 0.54 0.70 0.74 0.81
Correlation
-1
0
1

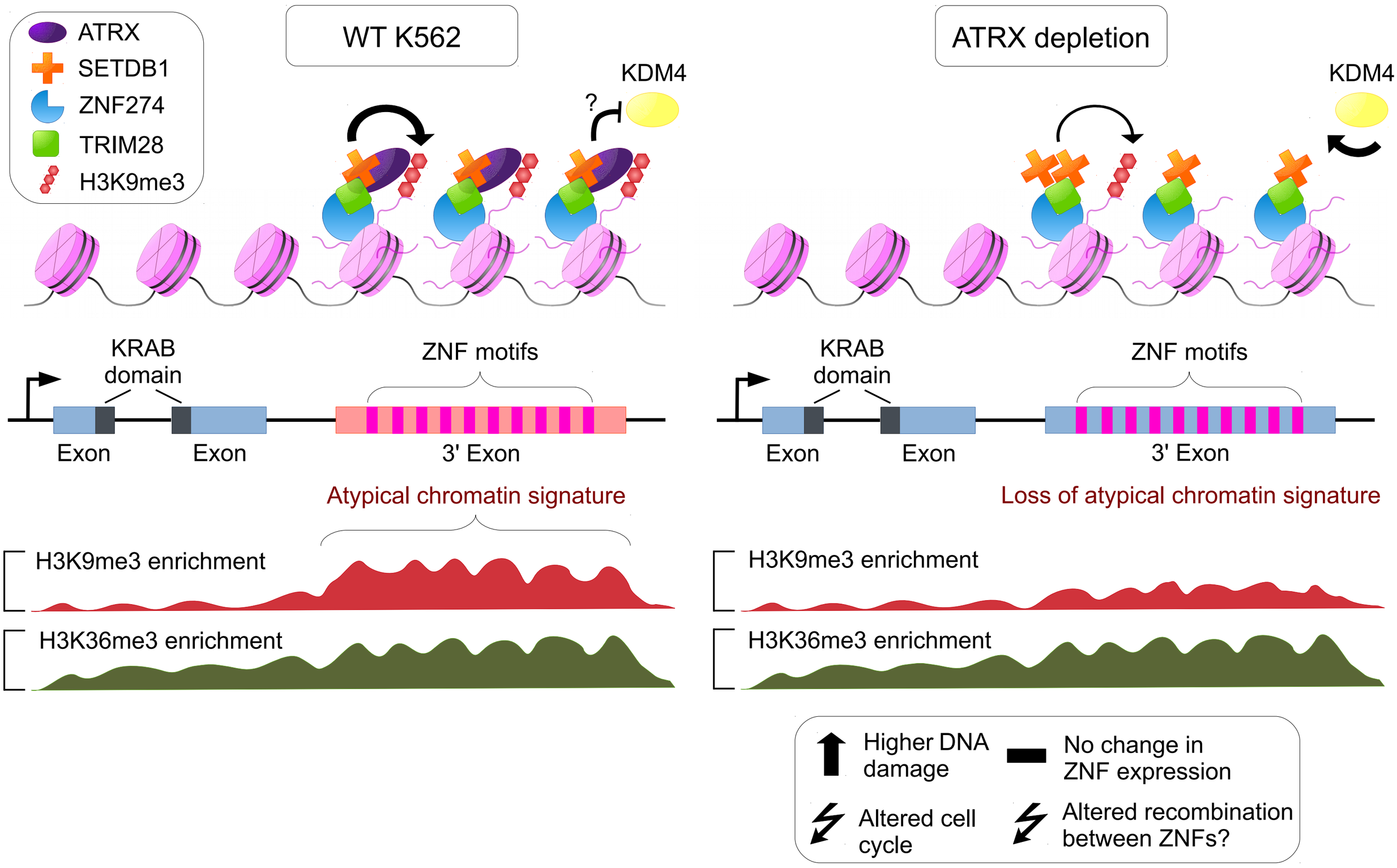

A
HepG2
ZNFs
ZNF274
H3K9me3
Class I
Class II
Class III
+
−
Enrichment over input
TSS
TES
TSS
TES


B
ATRX HepG2
Relative enrichment
0
1
2
3
4
5
6
7
ZNF180
ZNF267
ZNF268
ZNF441
ZNF510
ZNF529
ZNF544
ZNF564
ZNF700
ZNF721
ZNF891
ZNF358
ZNF414
Class I
Class III
ATRX
IgG

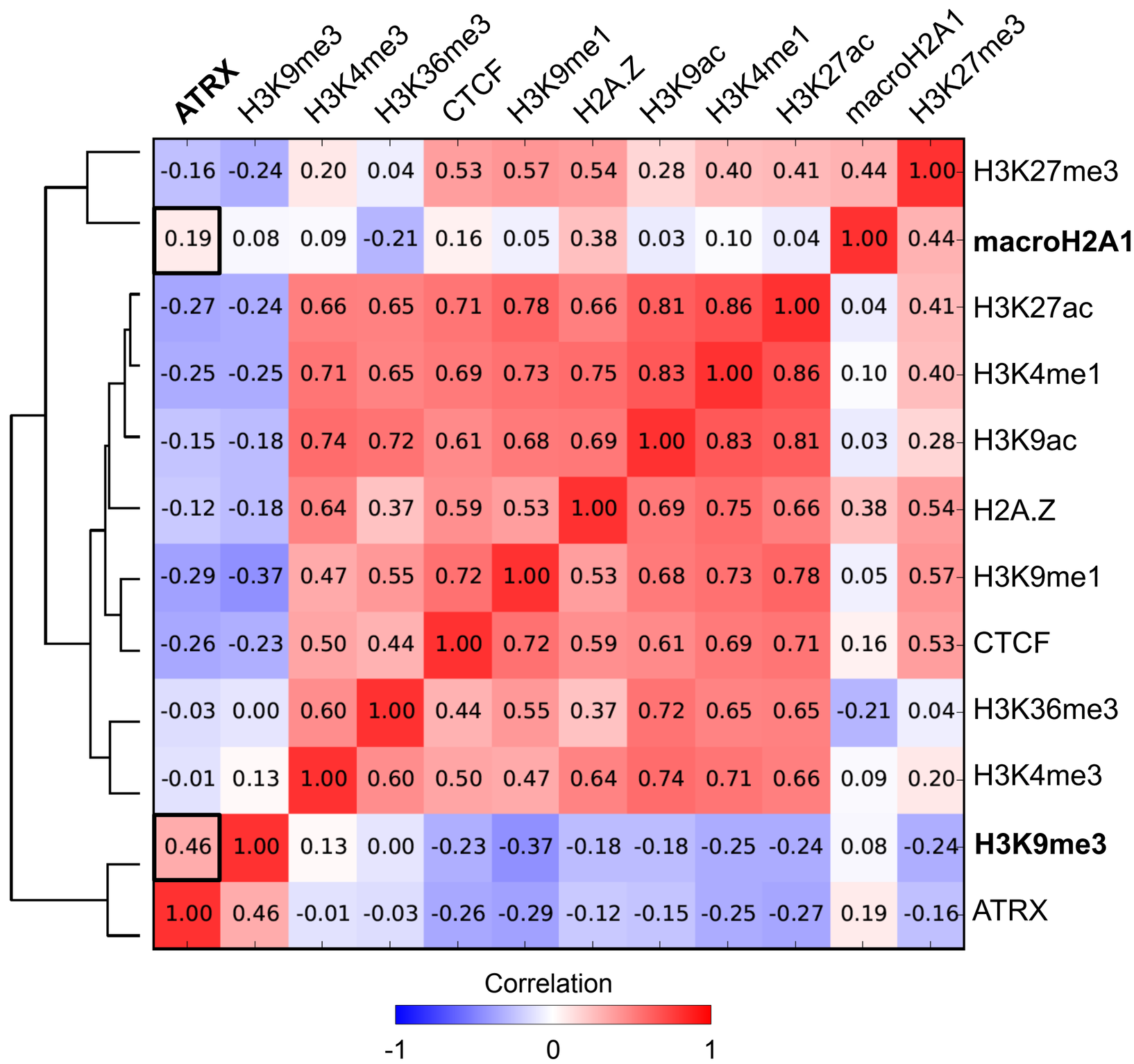

C
ATRX
ZNFs
K562
Eryth.
Class I
Class II
Class III
+
−
ATRX enrichment over input
TSS
TES
TSS
TES


D
mESC
MEF
ATRX
ATRX
H3K9me3
ATRX
H3K9me3
Mouse ZNF genes
+
−
Enrichment over input
TSS
TES
TSS
TES
TSS
TES
TSS
TES

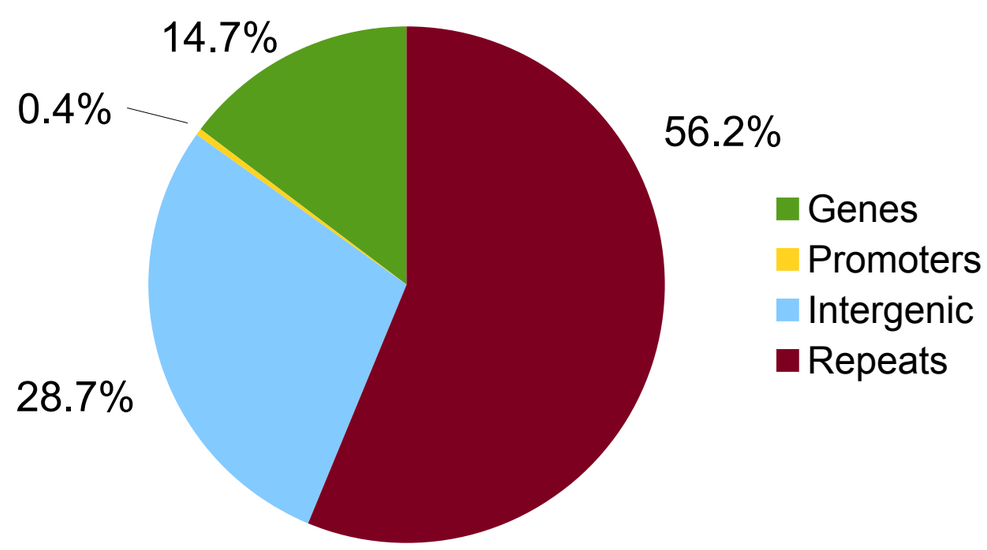

A
ATRX KO
V2
KO1
KO2
KDa
270
185
140
ATRX
Histones
B
ATRX
Relative enrichment
V2
KO1
KO2
ZNF180
ZNF267
ZNF268
ZNF441
ZNF510
ZNF529
ZNF544
ZNF564
ZNF700
ZNF721
ZNF891
ZNF358
ZNF414
Class I
Class III
C
H3K9me3
Relative enrichment
V2
KO1
KO2
Class I
Class III
D
H3K9me3 decrease after ATRX KO
% of H3K9me3 change in ATRX KO
****
*
Class I
Class II
Class III
H3K9me3 bound genes
E
H3K9me3 at ZNF genes
Other H3K9me3 bound genes
V2
KO2
TSS
TES
H3K9me3 enrichment over input
F
H3K9me3 reduced regions
SETDB1 profile
TRIM28 profile
ZNF274 profile
H3K36me3 profile
Average enrichment
ATRX bound
ATRX unbound
G
Genomic distribution of H3K9me3 reduced regions
Repeats
Intergenic
Promoters
Other genes
ZNF genes
ATRX Bound
ATRX Unbound


ATRX
SETDB1
ZNF274
TRIM28
H3K9me3
WT K562
KDM4
?
KRAB domain
ZNF motifs
Exon
Exon
3' Exon
Atypical chromatin signature
H3K9me3 enrichment
H3K36me3 enrichment
ATRX depletion
KDM4
KRAB domain
ZNF motifs
Exon
Exon
3' Exon
Loss of atypical chromatin signature
H3K9me3 enrichment
H3K36me3 enrichment
Higher DNA damage
No change in ZNF expression
Altered cell cycle
Altered recombination between ZNFs?

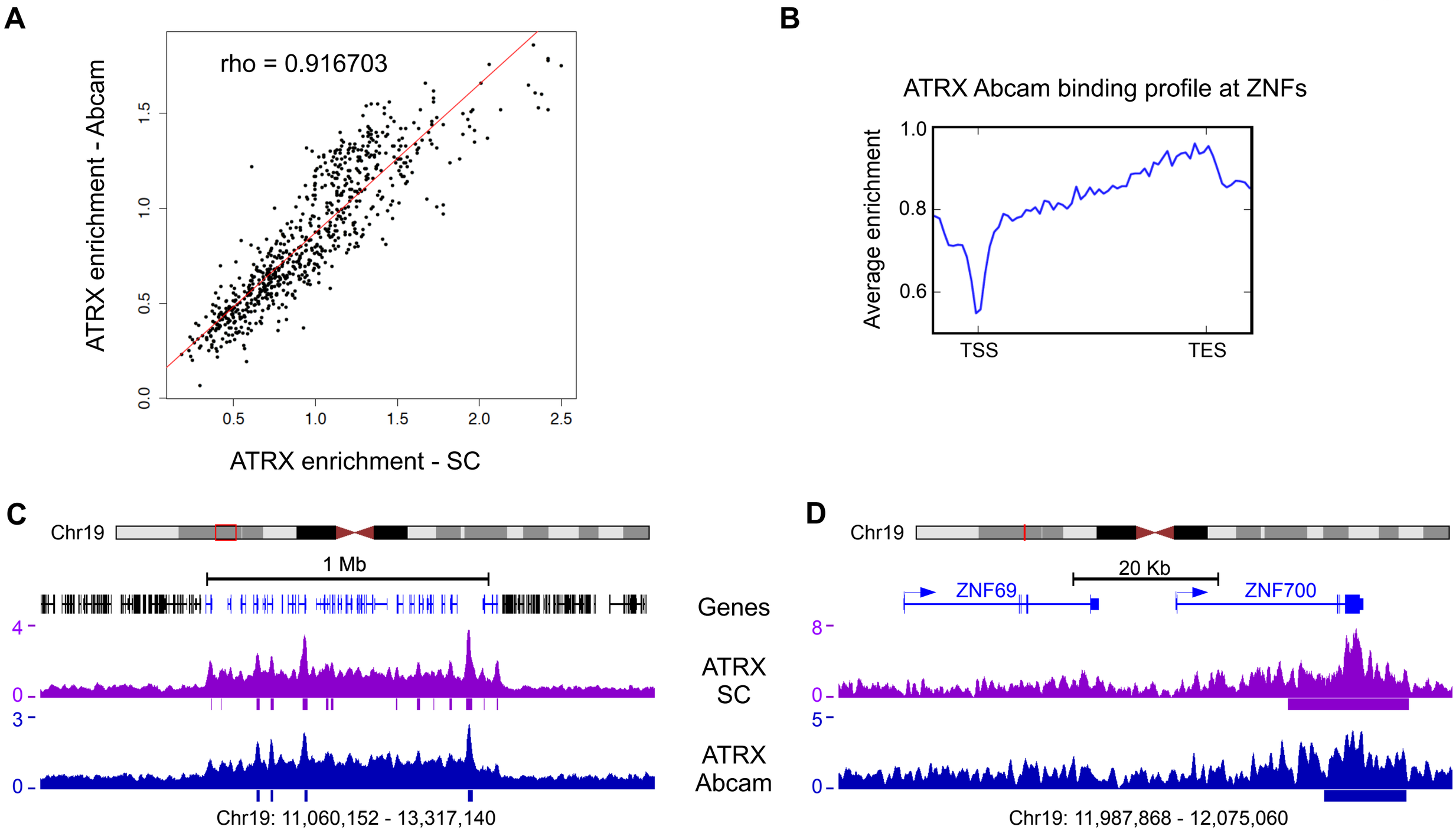
A



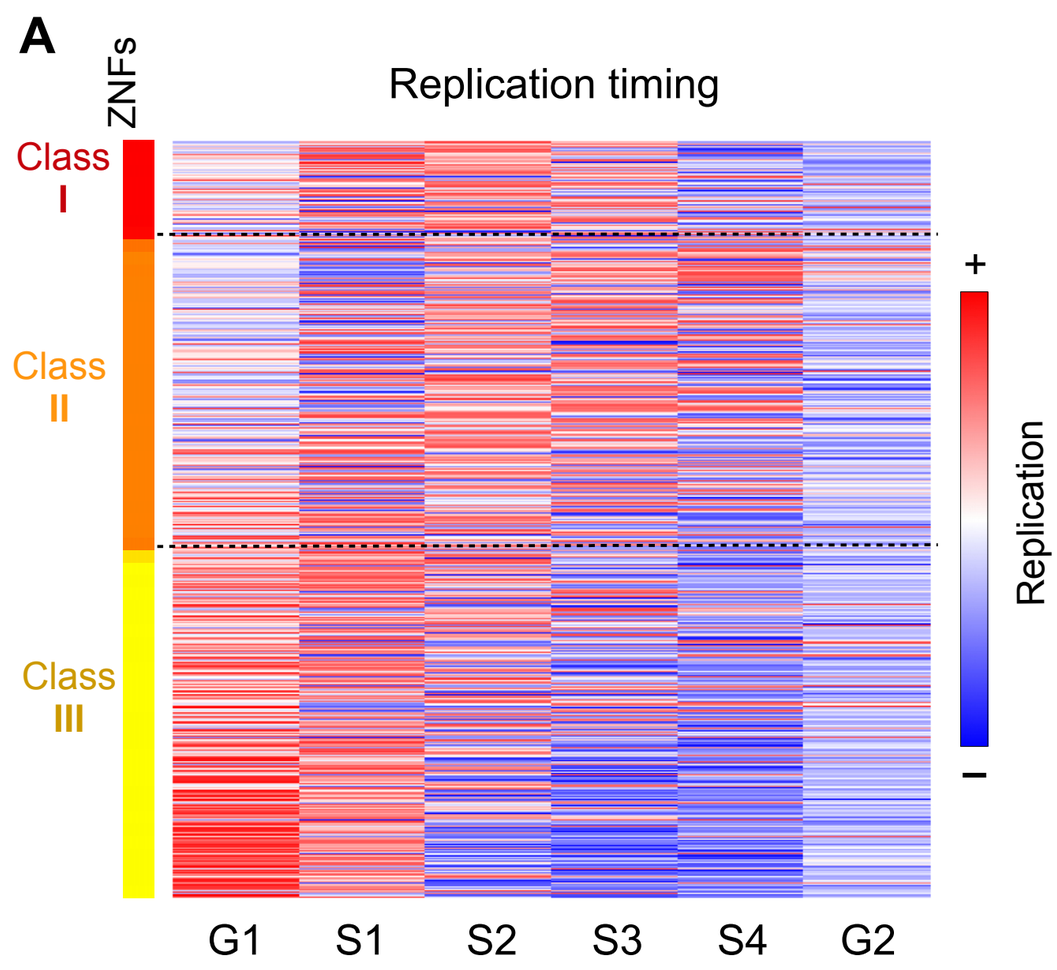
B



Supplementary Figure S1. ATRX correlates with H3K9me3 genome-wide. (A) Correlation plot of ATRX ChIP-seq enrichment with histone modifications, TFs and histone variants in K562 cells. (B) Distribution of ATRX significant peaks in genes (TSS-TES), promoters (-3kb to TSS), intergenic and repetitive (regions masked by repeat masker) sequences.

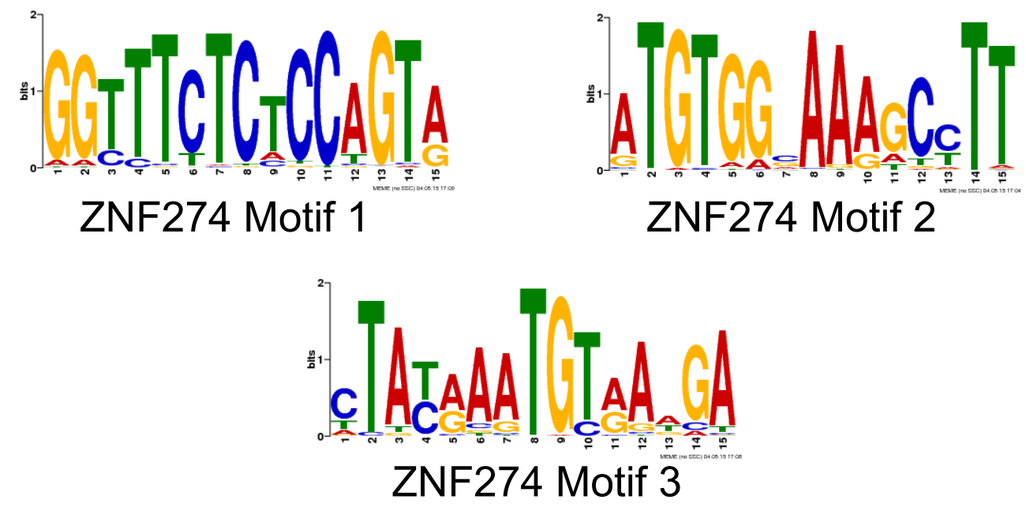


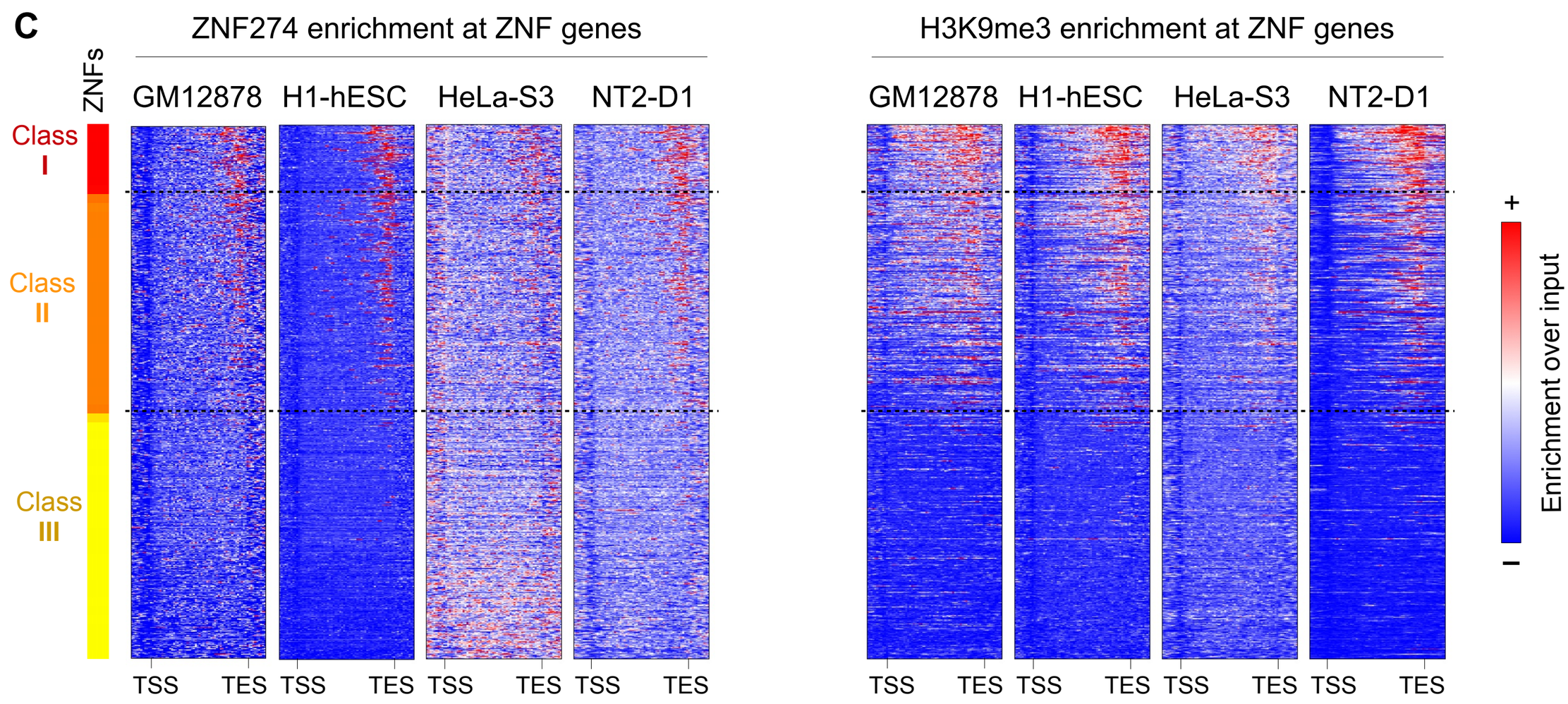
Supplementary Figure S2. Independent antibodies for ATRX ChIP-seq analysis show similar enrichment patterns. (A) Correlation of enrichment over ZNF genes for two independent antibodies. Each dot represents one ZNF gene. X axis = the average enrichment over input for the ATRX Santa Cruz (SC) antibody; Y axis = the average enrichment over input for the ATRX Abcam antibody. Antibodies show a high correlation of enrichment. (B) Average enrichment ChIP-seq profile of ATRX Abcam over all ZNF gene bodies +/- 1kb shows enrichment at the 3' region. The pattern is very similar to the one observed for ATRX SC (see Fig. 1G, left). (C) UCSC Genome browser screenshot of a typical ZNF cluster in chromosome 19 (blue genes). The enrichment over input signals for both ATRX antibodies are shown. Significant peaks are shown as bars below enrichment tracks. (D) Zoomed in snapshot of two genes contained in the ZNF cluster shown in (C).



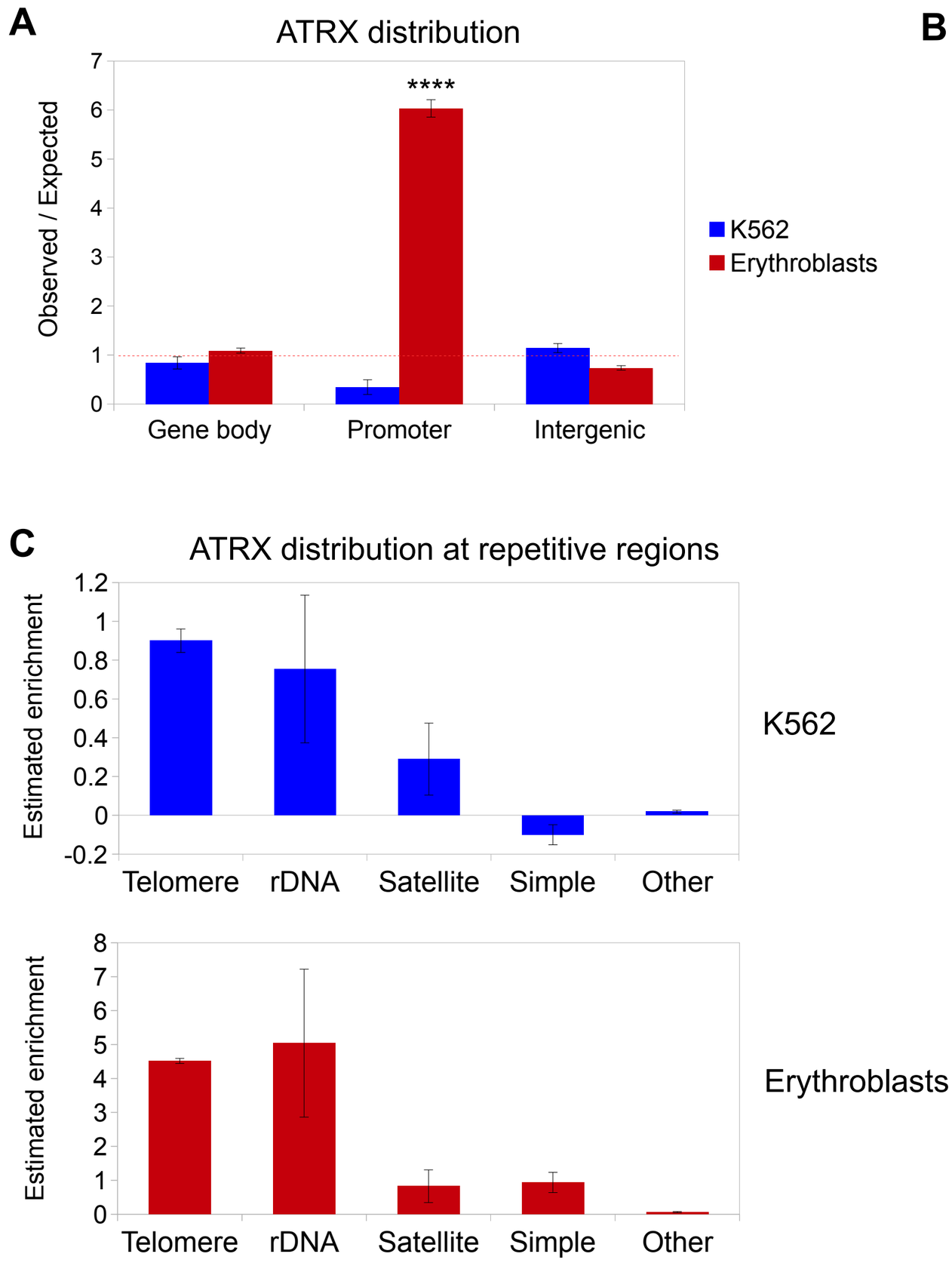
B

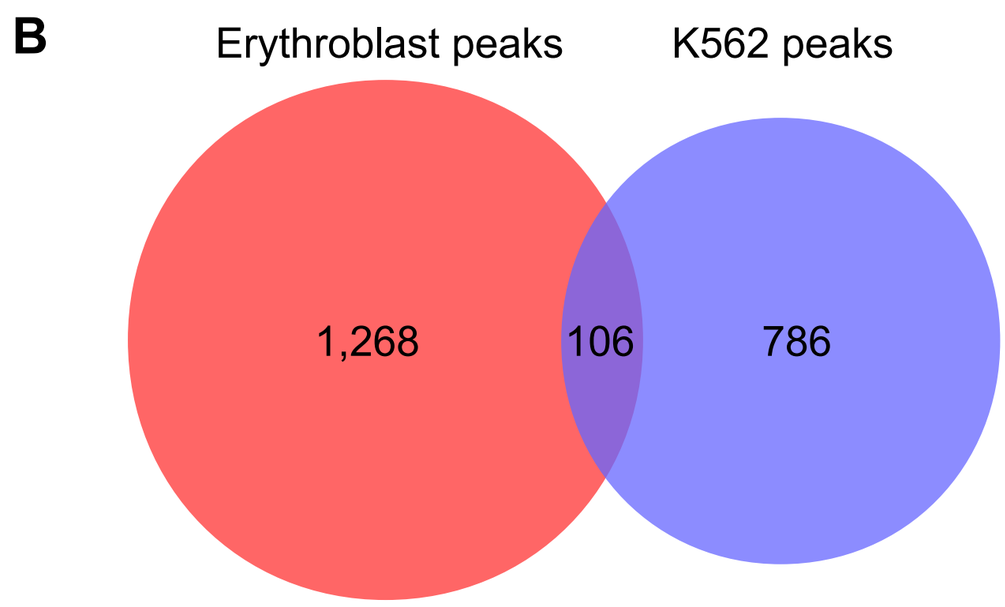
ZNF class	ZNFs with at least one ZNF274 motif	P-value
Class I (high ATRX)	78%	6.94E-29
Class II (medium ATRX)	39%	1.7E-10
Class III (no ATRX)	2.3%	1

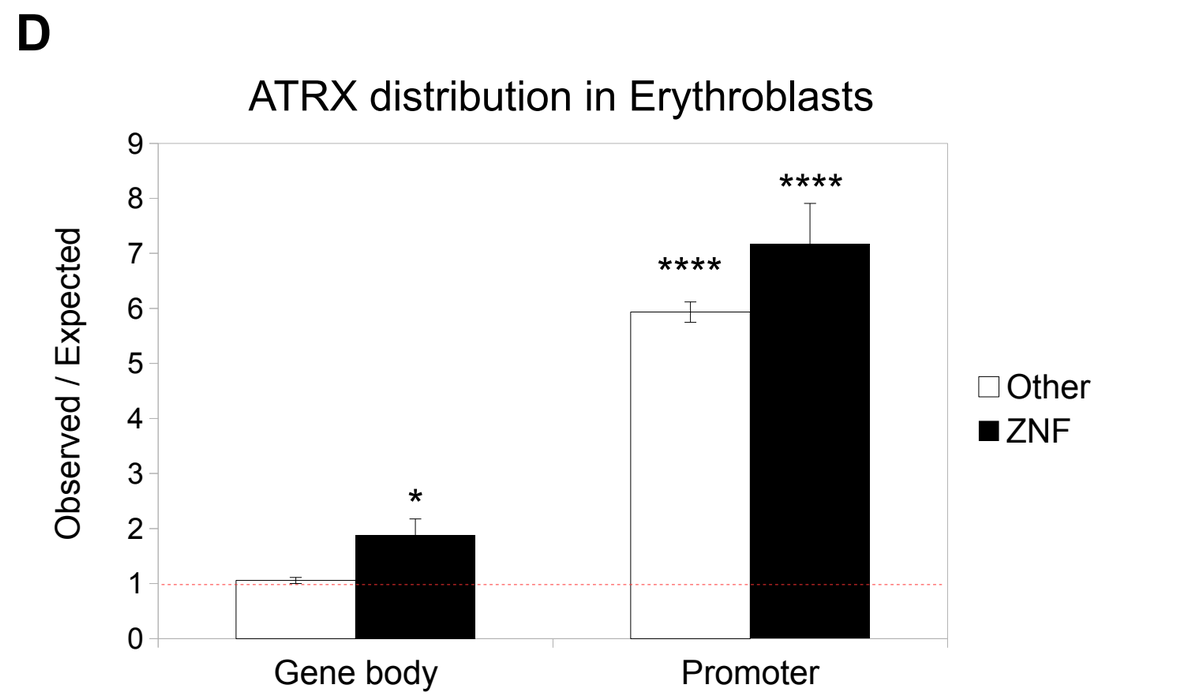


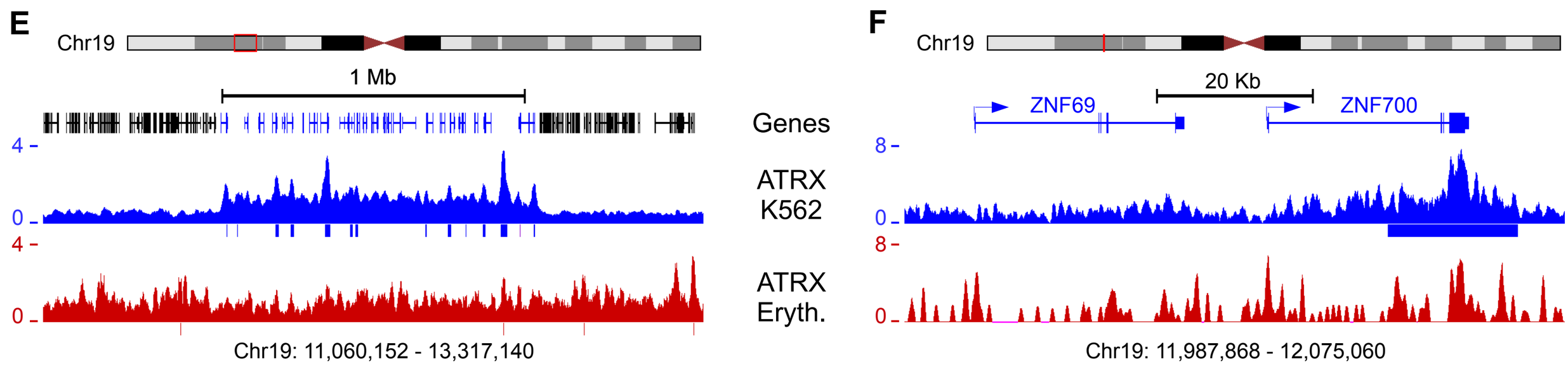


Supplementary Figure S3. ATRX-bound ZNFs show particular genetic and epigenetic features. (A) Repli-seq average enrichments over ZNFs at distinct points of the cell cycle. S1 and S2 are early S phase, whereas S3 and S4 represent the late stages of S phase. Genes are sorted by Class (ATRX enrichment from high to low, top to bottom, respectively). (B) Enrichment of ZNF274 binding motifs in different ZNF Classes (top). Logos of the three identified ZNF274 binding motifs (bottom). (C) Metagene plots over ZNFs of ZNF274 and H3K9me3 in different cell lines show that the pattern observed in K562 is maintained in most but not all cell lines. The genes are sorted as in (A).

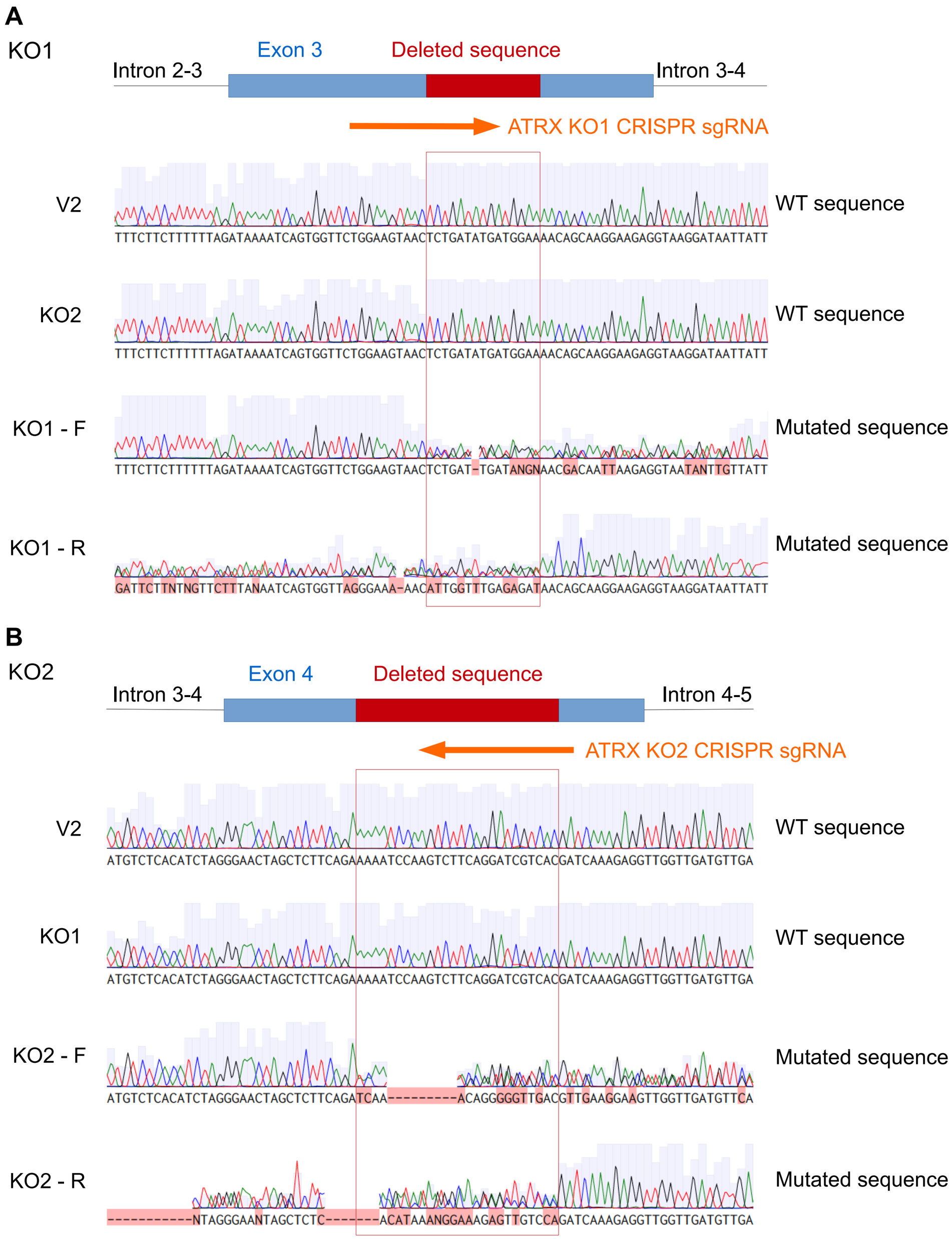




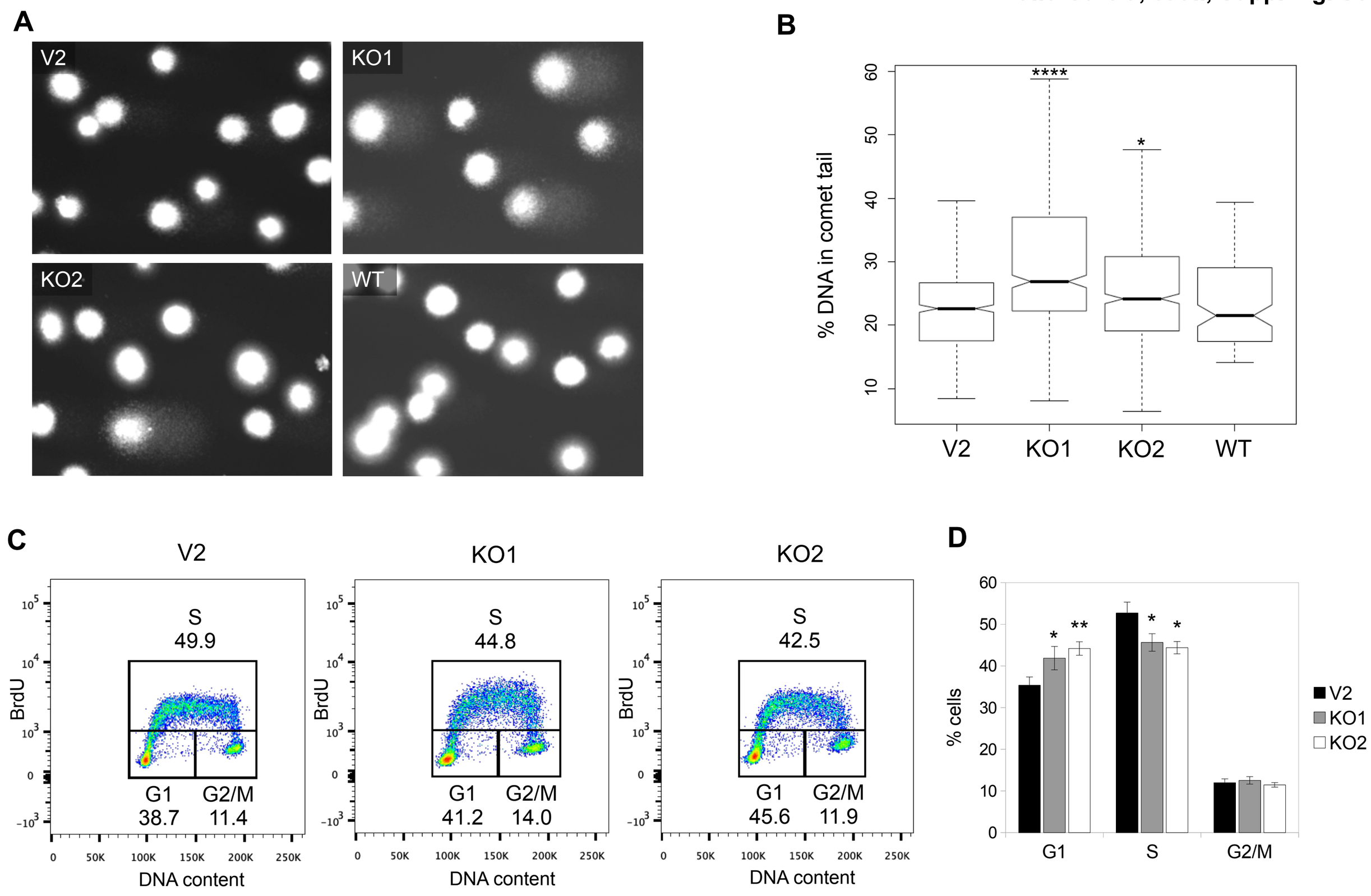




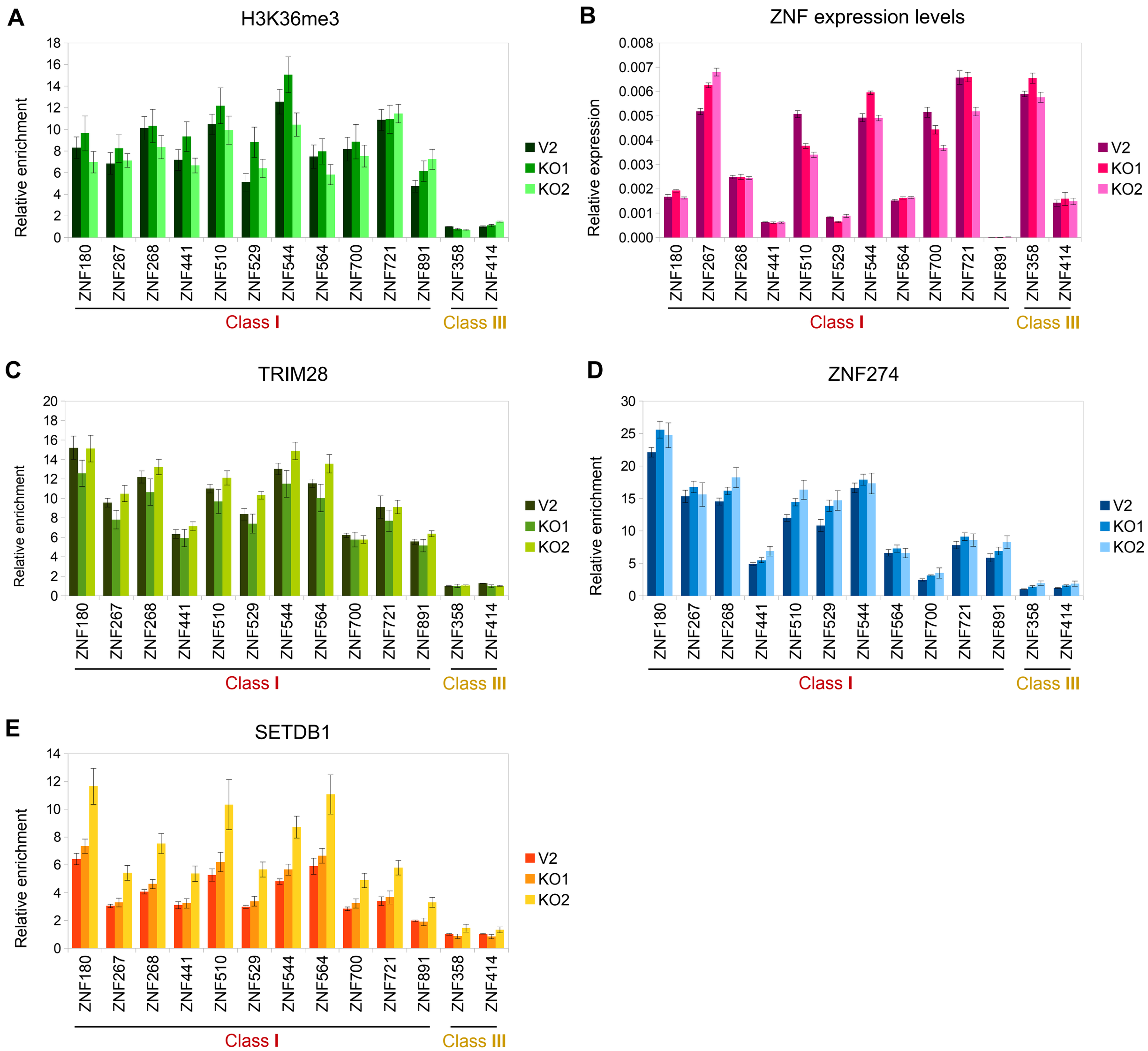
Supplementary Figure S4. Comparison of ATRX occupancy between human primary erythroblasts and K562 cells. (A) Observed over expected distribution of ATRX significant peaks in K562 (blue) and erythroblasts (red) among gene bodies, promoters (-3kb-TSS) and intergenic regions. Error bars in (A), (C) and (D) represent standard deviation. Asterisks represent significantly overrepresented regions (*= pvalue <0.05; ****= pvalue < 1x10-4). (B) Venn diagram of ATRX peaks between K562 and erythroblasts. (C) ATRX enrichment at repetitive sequences in K562 and erythroblasts. Error bars represent the standard deviation. (D) Observed over expected distribution of ATRX in erythroblasts at ZNFs and other non-ZNF genes and promoters. (E) UCSC genome browser screenshot of a typical ZNF cluster in chromosome 19 (blue genes). The enrichment over input signals for ATRX in K562 (blue) and erythroblasts (red) is shown. Significant peaks are shown as bars below enrichment tracks. (F) Zoomed in snapshot of two genes contained in the ZNF cluster shown in (E).



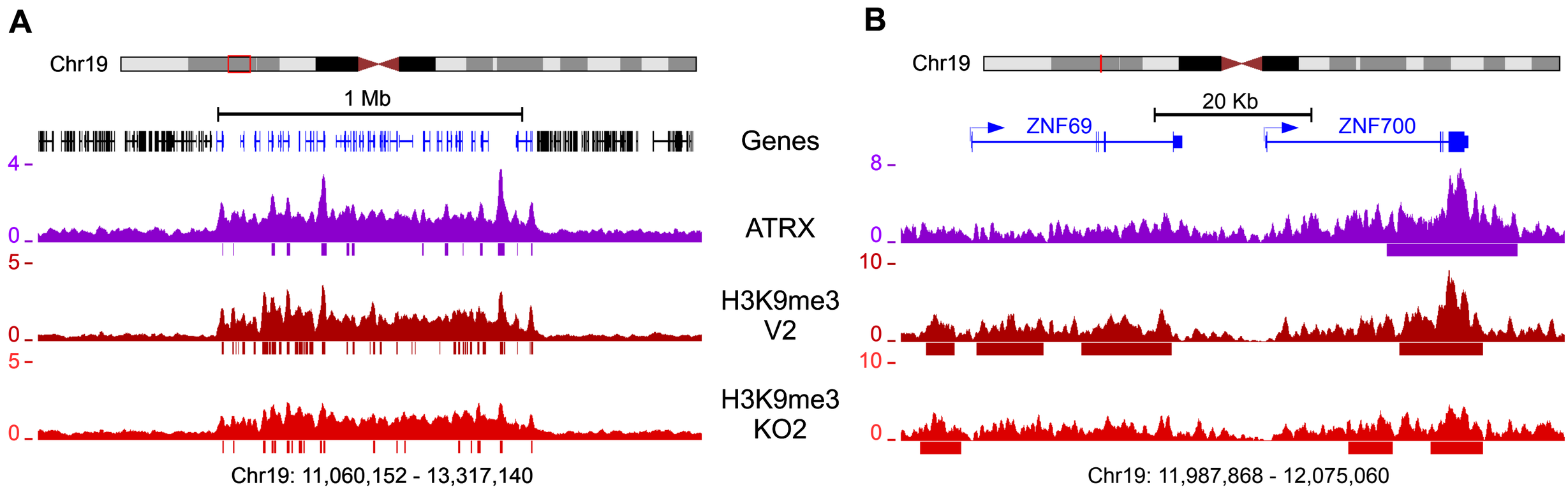
Supplementary Figure S5. Characterization of ATRX KO K562 cell lines. (A) Genomic sequence at ATRX exon 3 of the two CRISPR ATRX KO lines (KO1 and KO2) and the control V2 cell line. The red box and red square depict the region that is mutated only in KO1. The position and direction of the CRISPR sgRNA is shown as an orange arrow. KO1 was sequenced using forward and reverse primers to find the exact breakpoints of the deletion. (B) Same as in (A) but for exon 4, which is mutated only in KO2.



Supplementary Figure S6. ATRX KO K562 cells have higher percentage of DNA damage and defects in cell cycle. (A) Representative slides of the comet assay performed in WT, control V2, ATRX KO1 and ATRX KO2 cells. (B) Quantification of 4 independent comet assay replicates (>100 cells analyzed per replicate). Asterisks show significant differences compared to the V2 control (*= pvalue <0.05; ****= pvalue < 1x10-4). (C) Representative cell cycle profiles of control V2 and ATRX KO1 and KO2 cells assessed by BrdU/PI staining profiles. (D) Quantification of at least 5 independent BrdU/PI profile replicates. The bars show the average % of cells in each phase, error bars depict SEM. Asterisks show significant changes compared to the control V2 (*= pvalue < 0.05; **= pvalue < 0.01).



Supplementary Figure S7. Effect of ATRX KO at ZNFs. (A) Native ChIP-qPCR of H3K36me3 over the 3' exon region of ZNF genes in control (V2) and ATRX KO cell lines. (B) RT-qPCR of the panel of ZNFs from (A) in V2, KO1 and KO2 cells. (C-E) ChIP-qPCR of TRIM28 (C), ZNF274 (D) and SETDB1 (E) in V2, KO1 and KO2 cell lines using the same ZNF panel as in (A). All graphs show the average of at least three independent biological replicates. Error bars represent SEM. For details about statistical comparisons see Supp. Table S3.



Supplementary Figure S8. H3K9me3 loss in ATRX KO cell line. (A) UCSC genome browser screenshot of a typical ZNF cluster in chromosome 19 (blue genes). The enrichment over input signals for ATRX and H3K9me3 in control V2 and KO2 cells is shown. Significant peaks are shown as bars below the enrichment tracks. (B) Zoomed in screenshot of two genes contained in the ZNF cluster shown in (A).