Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs

Graphical Abstract

Highlights
- Two promoter transplantation strategies elucidate the regulatory role of LAD chromatin
- LADs are generally repressive but also highly heterogeneous
- LADs can impede both promoter activity and transcription elongation
- Promoters vary intrinsically in their sensitivity to LAD repression

Authors
Christ Leemans, Marloes C.H. van der Zwalm, Laura Brueckner, ..., Ludo Pagie, Joris van Arensbergen, Bas van Steensel

Correspondence
b.v.steensel@nki.nl

In Brief
A systematic look at mammalian promoters reveals that lamina-associated domains inherently repress transcription and also gives the first clues as to what dictates whether a gene can escape such silencing.

Leemans et al., 2019, Cell 177, 852–864
May 2, 2019 © 2019 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2019.03.009
Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs

Christ Leemans,1 Marloes C.H. van der Zwalm,1 Laura Brueckner,1 Federico Comoglio,1 Tom van Schaik,1 Ludo Pagie,1 Joris van Arensbergen,1 and Bas van Steensel1,2,*

1Division of Gene Regulation and Oncode Institute, Netherlands Cancer Institute, Amsterdam, the Netherlands
2Lead Contact
*Correspondence: b.v.steensel@nki.nl
https://doi.org/10.1016/j.cell.2019.03.009

SUMMARY

It is largely unclear whether genes that are naturally embedded in lamina-associated domains (LADs) are inactive due to their chromatin environment or whether LADs are merely secondary to the lack of transcription. We show that hundreds of human promoters become active when moved from their native LAD position to a neutral context in the same cells, indicating that LADs form a repressive environment. Another set of promoters inside LADs is able to "escape" repression, although their transcription elongation is attenuated. By inserting reporters into thousands of genomic locations, we demonstrate that escape promoters are intrinsically less sensitive to LAD repression. This is not simply explained by promoter strength but by the interplay between promoter sequence and local chromatin features that vary strongly across LADs. Enhancers also differ in their sensitivity to LAD chromatin. This work provides a general framework for the systematic understanding of gene regulation by repressive chromatin.

INTRODUCTION

Heterochromatin is generally defined as a compacted chromatin state and is thought to repress the transcriptional activity of genes and mobile elements (Allshire and Madhani, 2018). For many individual genes in a variety of species, this effect on transcription has been demonstrated. The evidence consists typically of two observations: (1) in its inactive state, the gene is naturally embedded in heterochromatin and (2) the gene becomes active when a key component of the heterochromatin protein complex is removed. Despite many compelling anecdotal examples of genes that fit these criteria, genome-wide studies have rarely found that all genes within a certain heterochromatin type are activated when key heterochromatin proteins are removed or mutated. Rather, only a fraction of the heterochromatin genes become active after such perturbations (Bulut-Karslioglu et al., 2014; King et al., 2018; Pengelly et al., 2015; Penke et al., 2016; Santoni de Sio et al., 2012; Yokochi et al., 2009).

One possible explanation for this conundrum could lie in the redundancy of heterochromatin proteins, which may make it difficult to disable heterochromatin completely by depletion of a single protein. An alternative possibility is that many genes in heterochromatin are inactive due to other causes, for example, due to the absence of an essential transcription factor (TF) in the cell type that is studied. If this is true, then lifting the heterochromatic state would not be sufficient to release gene activity. It is generally not known which fraction of human genes in heterochromatin is actively repressed by the heterochromatic context and which fraction is passively inactive due to lack of an essential activating signal.

Lamina-associated domains (LADs) are among the most prominent heterochromatin domains in metazoan genomes (Gonzalez-Sandoval and Gasser, 2016; Luperchio et al., 2017; van Steensel and Belmont, 2017). In mammalian cells, LADs are about 10 kb–10 Mb in size and collectively cover about 30%–40% of the genome (Guelen et al., 2008; Reddy et al., 2010). LADs interact closely with the nuclear lamina (NL) and are enriched in the histone modifications H3K9me2 and H3K9me3. In some cases, H3K27me3. In mammals, several thousands of genes are located within LADs, and the majority of these genes are transcriptionally inactive. It is believed that LADs form a repressive chromatin state. This notion is supported by experiments involving artificial tethering of genomic loci to the NL, which caused reduced expression of some, but not all, genes in the tethered loci (Finnlan et al., 2008; Kumanan and Specter, 2008; Reddy et al., 2008). However, the regulation of mammalian genes naturally embedded in LADs is poorly understood. In particular, for very few of these genes, it has been conclusively demonstrated that the lack of transcriptional activity is due to the heterochromatic state of LADs. This may be because critical protein components of LAD chromatin still have not been identified. Additionally, if the heterochromatic state of LADs indeed inhibits transcription, it may do so through several mechanisms: it may block initiation at the promoter but also transcription elongation, and it may silence enhancers that are essential for promoter activity. Which of these mechanisms may play a role within LADs is not known.

Although the majority of genes in LADs are inactive, about 10% are expressed (Guelen et al., 2008; Wu and Yao, 2017; Zheng et al., 2015). Analysis of such exceptions to the rule may provide valuable mechanistic insights. Possibly, these genes somehow are able to override the (putative) repressive LAD environment, for example, if their promoters are extremely strong. Alternatively, LADs may be heterogeneous, and the local chromatin context of these active genes may be less repressive or even stimulatory. It...
is not known which of these mechanisms—which are not mutually exclusive—allow genes to be active inside LADs.

Here, we report a detailed analysis of gene repression mechanisms in LADs. In particular, we studied how promoter activity is controlled by LADs. For this purpose, we made use of two orthogonal high-throughput genomic transplantation strategies. We systematically moved promoters from their native LAD location to a more neutral chromatin environment, and we also transplanted them to a wide range of locations, including many LAD contexts. Combined with extensive computational analysis, the results reveal that indeed hundreds of promoters inside LADs are repressed by the local chromatin state. Moreover, the data demonstrate that features encoded in the promoter sequence, as well as strong variation in local LAD composition, determine the expression level of genes inside LADs.

RESULTS

Frequent Repression of Promoter Activity in LADs

One approach to investigate whether a promoter embedded in a LAD in a certain cell type is repressed by the LAD chromatin context is to insert the promoter into a reporter plasmid and transfect it into the same cells. This transfer to an episomal context may lead to increased activity of the promoter due to its release from the LAD environment.

We aimed to test this hypothesis for many promoters in LADs. We recently reported Survey of Regulatory Elements (SuRE), a massively parallel reporter assay in which we assayed for promoter activity in a transiently transfected plasmid (van Arensbergen et al., 2017). These genome-wide SuRE data include quantitative measurements of the activity of all annotated human promoters (Data S1), thus offering an opportunity to systematically investigate whether LAD promoters are activated when moved into a plasmid reporter.

As a measure of promoter activity in the native chromatin context, we used data obtained with the GRO-cap (global run-on sequencing with 5' cap selection) method (Core et al., 2014). We focused on 31,043 well-annotated promoters (see STAR Methods) in human K562 cells, a widely studied leukemia cell line for which both SuRE and GRO-cap data are available. In inter-LAD regions (iLADs), promoter expression measured by the two methods shows a substantial correlation (Figure 1A; Pearson’s r = 0.70; Spearman’s rank correlation rho = 0.76), in line with what was previously observed genome-wide (van Arensbergen et al., 2017). This correlation is weaker in LADs (Figure 1B; r = 0.53; rho = 0.52). More importantly, LAD promoters exhibit on average ~10-fold lower GRO-cap signals than iLAD promoters with matching SuRE activity (sliding window curves in Figure 1B). Furthermore, 58% of the 2,048 LAD promoters with SuRE activity (SuRE log10 expression values > 0) exhibited no detectable GRO-cap activity, compared to 24% of a set of 3,289 iLAD promoters that were selected to have a closely matching SuRE activity (Figures S1A and S1B). Evidence for generally low activity of promoters inside LADs was obtained by similar analyses using PRO-seq (precision nuclear run-on sequencing) and CAGE (cap analysis gene expression) data as measures of endogenous promoter activity (Figures S1C and S1D). Thus, promoters with the same intrinsic activity tend to be much less active when embedded in LADs compared to iLADs. This indicates that LADs are generally poorly conducive to promoter activity compared to iLADs.

Three Classes of Promoters in LADs

Despite this clear general trend, not all LAD promoters respond similarly to their environment. Some exhibit high GRO-cap activity, suggesting that they can somehow escape the repressive influence of their LAD environment. To investigate this heterogeneity further, we defined three distinct categories of LAD promoters (Figure 1C): (1) repressed LAD promoters, which show...
>10-fold lower GRO-cap activity than typical iLAD promoters with similar SuRE activity; (2) escaper LAD promoters, which exhibit GRO-cap activity that is similar to or even higher than that of iLAD promoters with matching SuRE activity; and (3) inactive LAD promoters, which show very low SuRE and GRO-cap signals.

We initially interpreted these three classes of LAD promoters as follows. Repressed promoters have the potential to be active (i.e., all required transcriptional activators are present in the cell) but appear to be repressed by their native LAD context. In contrast, escaper promoters may carry features that render them less sensitive to the repressive effects of LADs or they may reside in a LAD sub-region that is not repressive. Finally, inactive promoters appear to lack an essential activating signal, both in their native environment and in the plasmid context. This could be due to the absence of a critical activating TF in the cell type studied. Alternatively, it could be because the promoter requires a distal enhancer that is not included in the SuRE reporter plasmid and that is also not functional in the native LAD context.

Of the 2,444 thus classified promoters in LADs, 33% are repressed, 16% are escapers, and 52% are inactive. Analysis of PRO-seq and CAGE data generally supports the classification repressed, 16% are escapers, and 52% are inactive. Analysis of PRO-seq and CAGE data generally supports the classification of these three types of promoters (Figures S1E and S1F). Interestingly, the three promoter classes show a difference in their overall tissue specificity (Figure 1D). Escaper LAD promoters are broadly expressed and, based on this, many may be classified as promoters of housekeeping genes. In contrast, inactive LAD promoters are mostly highly tissue specific, and repressed LAD promoters exhibit an intermediate pattern of tissue specificity. These observations underscore that the three promoter classes are of distinct nature.

**Escaper Promoters Are Locally Detached from the NL**

Although escaper promoters are by definition located inside LADs, close inspection of the DamID data suggested that they are locally detached from the NL. On average, the DamID signals around these promoters are about 4-fold lower than for their upstream and downstream regions (Figure 2A). The detached region typically extends from about 10 kb upstream to 10 kb downstream of the transcription start site (TSS), although even beyond these distances, the DamID signals remain somewhat lower than the average signal inside LADs. Virtually no such detachment was observed for the repressed and inactive promoter classes (Figure 2A). Of note, the ∼20-kb region around escaper promoters that appears to be detached from the NL is substantially larger than the size of the promoter itself or of the nucleosome-free region that is detected by DNase-seq (Figure S2A).

NL interactions are in part controlled by the histone modifications H3K9me2 and H3K9me3 (Bian et al., 2013; Harr et al., 2015; Kind et al., 2013; Towbin et al., 2012). Analysis of chromatin immunoprecipitation sequencing (ChIP-seq) data shows that escaper promoters indeed show a local decrease of these marks (Figures S2B and S2C). However, the other promoter classes show a nearly similar local decrease of H3K9me2 and H3K9me3, suggesting that the detachment of escaper promoters from the NL cannot be solely explained by the absence of marks. The local detachment from the NL might facilitate the transcriptional activity of escaper promoters, although we cannot rule out that it is a secondary consequence of their activity (Chuang et al., 2006; Therizols et al., 2014).

**Partially Impaired Transcription Elongation from Escaper Promoters in LADs**

Because escaper promoters are active, we expected that the corresponding genes would produce similar amounts of mRNA as their iLAD counterparts. Surprisingly, we found that this is not the case. Escaper genes produce on average 5-fold less mRNA than a set of iLAD genes with the same promoter activity distribution as measured by GRO-cap (Figures 2B and S2D). Escaper genes also produced less mRNA than a set of iLAD genes matched for promoter activity as measured by SuRE (Figures S2E and S2F).

To investigate this discrepancy between promoter activity and mRNA yield, we examined marks of transcription elongation. The amount of nascent RNA as detected by the TT-seq (transient transcriptome sequencing) method (Schwalb et al., 2016) was lower along the gene bodies of escaper genes (Figure 2C), as was the occupancy of RNA polymerase II (Pol II) (Figure 2D) and PRO-seq signals (Figure S2G). Also H3K36me3, another mark of elongation (Wagner and Carpenter, 2012), was lower along escaper gene transcription units (Figure 2E). In each of these analyses, we compared the escaper genes to a set of iLAD genes with matched promoter activities according to GRO-cap (Figure S2D). We note that these elongation marks are reduced along the entire length of the genes; although there is some indication of progressive loss of TT-seq signals toward the 3′ ends, this is only marginally more pronounced than in matching iLAD genes (Figure 2C). Together, these results point to inefficient transition of Pol II from initiation to elongation rather than to random abortion of elongation along the gene body.

This prompted us to investigate available K562 ChIP-seq data of various proteins previously linked to regulation of elongation (Figures 2F and S2H–S2L). We found that c-Myc binding is nearly 2-fold weaker at escaper promoters compared to activity-matched iLAD promoters (Figure 2F). It has previously been reported that c-Myc stimulates the release of paused polymerase from promoters (Rahl and Young, 2014). We also observed reduced occupancy at escaper promoters of Brd4 (Figure S2H), which is thought to promote assembly of the transcription elongation complex (Tyler et al., 2017; Winter et al., 2017). Other investigated proteins (GTF2F1, LARP7, NELFε, and ENL; Durham et al., 2012) showed only minor or non-significant differences (Figures S2I–S2L). The reduced binding of c-Myc and Brd4 may in part explain the relatively poor elongation of escaper genes, but it is also possible that association with the NL (or a chromatin feature linked to this association) impedes the initiation-to-elongation transition.

**Testing Repressed and Escaper Promoters in Many Chromatin Contexts**

One possibility is that the different activity levels of repressed and escaper promoters inside LADs are encoded in their proximal sequence. If this is true, repressed promoters should remain inactive when transplanted to a different LAD, and escaper promoters should remain active when moved to a different LAD. Alternatively, the differences between repressed and escaper...
promoters may reflect local differences in LAD context, with some sub-LAD regions being more repressive than others. If this is the case, then the activity of both repressed and escaper promoters should strongly depend on the precise LAD context in which they are located.

To discriminate between these two hypotheses, we inserted several escaper and repressed promoters into a large number of different genomic locations (both LAD and iLAD) by means of the thousands of reporters integrated in parallel (TRIP) method (Akhtar et al., 2013). Specifically, we cloned representative promoters of each class into a common reporter construct (Figure S3A) and generated K562 cell pools in which the reporters were integrated randomly into hundreds of genomic locations, including many LADs. We mapped the integration sites by inverse-PCR and high-throughput sequencing (Akhtar et al., 2014). Because we marked each copy of the reporter with a unique random barcode in its transcription unit, we could determine the expression level of all integrations in parallel by high-throughput sequencing and counting of each barcode in mRNA from the cell pools. This strategy enabled us to study the effect of many different LAD and iLAD contexts on individual promoters.

We performed these TRIP experiments with three repressed promoters and three escaper promoters (Table S1). In addition,
we included the promoter of the PGK gene, a housekeeping gene that is located in an iLAD. For these seven promoters combined, we obtained 12,812 integrations (629-3,514 per promoter) that could be mapped to unique genomic locations (Figures S3B–S3H). Of these, 23% were located inside a LAD. This is less than may be expected by random chance, because LADs occupy 43% of the genome in K562 cells. This may reflect an integration bias of the PiggyBac vector (de Jong et al., 2014) but also a relatively poor mappability of integrations inside LADs due to a higher repeat content. Nevertheless, we obtained 123–737 integrations in LADs per promoter, providing ample statistical power, as will be shown below. Expression of the barcoded reporter transcripts in the cell pools carrying these integrations were analyzed in two independent biological replicates each (Data S2).

Repressed and Escaper Promoters Differ in Sensitivity to LAD Context

All six promoters originating from LADs showed strong variation in expression levels, depending on their integration site (discussed in more detail below). For the three promoters of the repressed class, the median expression level was about 43- to 130-fold lower within LADs than within iLADs (Figures 3A–3C), underscoring the repressive potential of LADs. For the escaper promoters, however, the difference between LAD and iLAD integrations was much less pronounced (3.7- to 20-fold; Figures 3D–3F). Indeed, a linear regression model of reporter expression as function of Lamin B1 DamID signal and promoter class (escaper or repressed) showed a significant interaction between promoter class and Lamin B1 DamID levels (Figure 3H), indicating that repressed promoters and escaper promoters respond differently to the LAD environment.

Surprisingly, the PGK promoter showed only a 1.9-fold difference in median expression between LAD and iLAD contexts (Figure 3G), indicating that it is largely refractory to LAD environments. This underscores that promoters differ in their sensitivity to LAD contexts and suggests that the PGK promoter has strong escaper-like properties.

Promoter Strength Does Not Explain Differential Sensitivity to LAD Context

We considered the possibility that escaper promoters are intrinsically stronger than repressed promoters and thereby may overrule the repressive environment of LADs more effectively. To test this hypothesis, we measured the activity of the promoters in episomal plasmid context, i.e., not embedded in any genomic chromatin environment. We considered this to be the most neutral environment that could reasonably be obtained. For the most precise estimate of promoter activity, we constructed mini-libraries of ~100 barcoded reporter plasmids for each promoter, mixed these in equal amounts, and transiently transfected the entire mix into K562 cells. Barcodes were then counted in mRNA and normalized to barcode counts in the plasmid DNA pool. The representation of each promoter by ~100 random barcodes ensures average expression levels that are not biased by barcode sequence and that are minimally affected by random noise, providing an accurate estimate of the relative activities of the promoters.

The results of these measurements show that the three escaper promoters are not systematically stronger than the repressed promoters (Figure 4). Thus, the intrinsic promoter strength does not account for the different responses of escaper and repressed promoters to LAD contexts. However, the PGK promoter is by far the strongest of the seven promoters; possibly this contributes to its insensitivity to LAD environments.

We note that these transient transfection experiments also provide an estimate of the variance in expression levels that may be caused by differences in barcode sequences and other sources of noise. In the transient transfections, we observed unimodal distributions with SD = 0.35 ± 0.25 (n = 7 promoters). This contrasts starkly with the multi-modal and much broader distributions (SD = 3.66 ± 0.87; n = 7) obtained with integrated reporters (Figures 3A–3G). Thus, most of the variance observed with integrated reporters is not explained by barcode differences or technical noise but rather by chromatin context.

Chromatin Features that Correlate with Promoter Activity within LADs

Aside from the overall differences between LADs and iLADs, there is substantial variation in the TRIP reporter expression levels within LADs. This implies that LADs are functionally heterogeneous structures and that local chromatin features within LADs can substantially affect reporter activity. We note that, within LADs, the three repressed promoters showed a broader distribution of reporter expression levels than the three escaper promoters (Figures 3A–3F). Thus, compared to escaper promoters, repressed promoters are not only more sensitive to global LAD/iLAD differences but also to local chromatin differences within LADs.

In order to investigate which features may be linked to the variation in reporter activity within LADs, we took a statistical learning approach. We compared our maps of reporter activity to a collection of available high-quality epigenomic maps from K562 cells (Figure S4; Table S2). We considered both the signal strength of these epigenome features at the integration sites and the distance to the nearest peaks or domains of these features. Note that these features were mapped in the absence of the integrations and can therefore not be the consequence of the integrations; conversely, it was previously shown that integrated reporters generally adopt the local chromatin state (Corrales et al., 2017), although exceptions cannot be ruled out. To identify the most likely candidate chromatin features linked to our reporter activities, we applied a feature selection algorithm that combines a lasso linear regression model with bootstrapping. This approach identified a set of features that statistically explain the reporter expression levels, with an estimate of their relative contribution to the predictive power of the model.

We first restricted this analysis to integrations inside LADs. Collectively, chromatin features could explain nearly half of the variance in reporter expression for repressed promoters (R² = 0.49 ± 0.05; mean ± SD across 100 samplings) and significantly less for escaper promoters (R² = 0.35 ± 0.05; p = 8e–31; two-sided Wilcoxon test of R² values; Figure 5A). These results are in agreement with our previous conclusion that escaper promoters are less sensitive to their chromatin environment than repressed promoters.
The chromatin features that were most predictive of reporter expression in LADs were generally shared between the two promoter classes, although we observed some quantitative differences in predictive power of individual features (Figure 5B). The LMNB1-DamID signal at the site of integration was one of the strongest predictors and negatively correlated with reporter activity. This signal was previously shown to be tightly linked to NL contact frequency (Kind et al., 2015), suggesting that reporters are more potently repressed when they are inserted in regions that are more stably associated with the NL. The local level of the histone variant H2A.Z is among the strongest positive predictors of reporter activity for both promoter classes.
Figure 4. Precise Measurement of Promoter Activities in Episomal Plasmid Context

Each indicated promoter was cloned in the same promoter-less reporter plasmid with about 100 different random barcodes. A pool of the resulting ~700 plasmids was transiently transfected into K562 cells. Barcodes were counted in cDNA and plasmid DNA isolated from these cells after 2 days. The plot shows the distribution of expression levels of all barcodes sorted by promoter; horizontal lines depict medians. Data are average of at least two independent experiments.

Mammalian H2A.Z generally marks active enhancers and promoters and is thought to promote binding of TFs and cofactors by destabilizing nucleosomes (Hu et al., 2013; Ku et al., 2012; Li et al., 2012). Other positive predictive features are well-known active chromatin marks, such as H3K4me1, H3K36me3, and Pol II, suggesting that the integrated reporters are generally more active when inserted near active transcription units. Although the active chromatin marks identified in the model are less prevalent in LADs, they are found near reporter integration sites, with signal intensities in the range of those found in iLADs (Figures S4A and S4B).

Although many of the feature importance scores of the model differ significantly (p < 0.001) between the escaper and repressed promoters, only a few marks show substantial differences (Figure 5B). Noteworthy is H3K122ac, proximity of which is predicted to have a positive effect on escaper promoters and a negative effect on repressed promoters, and H3K27ac proximity shows a higher positive score for repressed promoters than for escaper promoters. Both marks have been found on enhancers but in a largely mutually exclusive pattern (Pradeepa et al., 2016). Our results suggest that H3K122ac-marked enhancers preferentially activate escaper promoters, and H3K27ac-marked enhancers preferentially activate repressed promoters.

Because many of the predictive features are abundantly present in iLADs, we expected that the overall differential responsiveness of escaper and repressed promoters to chromatin context could also be observed in iLAD integrations. Indeed, this was the case: R^2 values of a model fitted only to iLAD integrations were again higher for repressed promoters (R^2 = 0.29 ± 0.028; mean ± SD across 100 samplings) than for escaper promoters (R^2 = 0.20 ± 0.022; p = 3e−33; two-sided Wilcoxon test of R^2 values; Figure S5A), although the R^2 values indicate that the model performed less well in iLADs than in LADs. Finally, to ensure that the differences observed between repressed and escaper promoters are not driven by a single promoter, we fitted six promoter-specific models. In this instance, we used integrations in LADs and iLADs combined in order to maintain sufficient statistical power. The results confirm a systematic difference in R^2 values between the promoters of the escaper and repressed classes (Figure S5B).

Together, these results underscore that escaper promoters are generally less responsive to chromatin heterogeneity inside LADs, and the modeling identifies the most likely chromatin features that may affect the activity of overlapping or nearby promoters.

Promoter Repression in H3K27me3 Domains

We wondered whether the differential sensitivity of the seven promoters is specific for LADs or also applies to other types of heterochromatin. We focused on domains of H3K27me3, which represent heterochromatin associated with polycomb repressive complexes. In K562 cells, 80% of the H3K27me3 domains are located outside of LADs. Interestingly, within these regions, the integrated promoters also show striking quantitative differences in repression (Figures 6A–6G). Generally, these differences correlate with those observed in LADs, but the degree of repression is systematically about 2- to 5-fold less in H3K27me3 domains compared to LADs (Figure 6H). Like in LADs, in H3K27me3 domains, the three escaper promoters are less repressed than the three promoters of the LAD-repressed category, and the PGK promoter is essentially insensitive to H3K27me3 chromatin. In particular, the ADAMTS1 and BRINP1 promoters show a broad range of expression levels inside H3K27me3 domains, pointing to a strong heterogeneity of these domains.

To complement these TRIP results, we revisited the SuRE and GRO-cap data and investigated the activity of iLAD promoters that are naturally located in H3K27me3 chromatin (Figure 6I). On average, promoters in H3K27me3 domains show a roughly 10-fold lower GRO-cap activity than SuRE-matched promoters located outside H3K27me3 domains (and outside LADs). This is similar to what we observed in LADs (see Discussion).

Effects of LADs on Enhancer Activity

Finally, we explored the effects of LAD context on enhancers. We previously reported that SuRE can also detect enhancer activity, by virtue of the fact that enhancers typically act as transcription start sites and produce RNA in proportion to their activity (van Aarenberg et al., 2017). Similarly, in the native context, enhancers generate transcripts that can be detected by GRO-cap (Core et al., 2014). Hence, to investigate whether LADs might affect enhancer activity, we repeated the SuRE versus GRO-cap analysis, but now for enhancers (Figure 7A; Data S3). This showed that LAD enhancers exhibit on average ~3-fold lower GRO-cap signals than SuRE enhancers with matching SuRE activity. Thus, like promoters, enhancers inside LADs appear generally repressed. It is possible that we underestimate the average magnitude of repression, because for enhancers, the
SuRE and GRO-cap signals are much closer to background levels than for promoters.

**Sequence Motifs Linked to Enhancer Sensitivity to LAD Repression**

Similar to escaper promoters, a subset of enhancers in LADs shows high GRO-cap signals that are more typical of enhancers in iLADs. Furthermore, some enhancers with very low GRO-cap signals were highly active in SuRE, and others were virtually inactive in SuRE. We therefore took a similar approach as for promoters and defined three classes of enhancers: inactive; repressed; and escaper enhancers (Figure 7B). The large numbers of enhancers in each class provided sufficient statistical power to search for sequence motifs that may explain the difference in activity of escaper and repressed enhancers (a similar analysis on the much lower numbers of promoters lacked statistical power). *De novo* motif analysis comparing these two enhancer classes, corrected for SuRE activity, identified three prominent motifs enriched in escaper enhancers that all carry CpG dinucleotides and one infrequent motif of unknown nature (Figure 7C). The CpG dinucleotide enrichment cannot be attributed to CpG islands, because very few of the escaper and repressed enhancers overlap with CpG islands (22 out of 927 and 15 out of 1,433, respectively). No motifs were found that could be linked to specific TFs.

These results suggest that no single TF can account for the different responses of escaper and repressed enhancers to LADs. Instead, a broad class of TFs may contribute collectively to this difference. Prominent candidates are pioneer TFs, which are able to activate promoters inside condensed chromatin (Zaret and Carroll, 2011). We therefore tested whether cognate motifs for known and predicted pioneer TFs (Zaret and Carroll, 2011) collectively are enriched in escaper enhancers compared to repressed enhancers. Indeed, we observed a modest but statistically significant enrichment (Figure 7D), suggesting that pioneer TFs collectively help escaper enhancers to overcome the repressive LAD environment.

**DISCUSSION**

In the past two decades, a wide diversity of genome-wide mapping methods (Dirks et al., 2016; Kelsey et al., 2017; Rivera and Ren, 2013; Zentner and Henikoff, 2014) has yielded a wealth of descriptive epigenome maps. Analysis of these data has uncovered extensive correlations between gene activity and many chromatin features. A current challenge is to move from these genome-wide correlations to a detailed understanding of causal relationships, while maintaining a genome-wide perspective. This requires systematic perturbation approaches (Catarino and Stark, 2018; Stricker et al., 2017). Here, we developed a general framework, centered on two genome-wide perturbation methods, to dissect causal relationships between local chromatin state and gene activity. The combined application of SuRE and TRIP provides detailed insights into the interplay between promoters and the local chromatin environment, here illustrated for LADs.

Previous TRIP experiments with two promoters (one from an iLAD and one synthetic) had suggested that LADs form a repressive environment (Akhtar et al., 2013). Furthermore, it has been shown that some genes can be downregulated by artificial tethering to the NL (Dialynas et al., 2010; Finlan et al., 2008; Kumaran and Spector, 2008; Reddy et al., 2008). It remained unclear, however, whether genes that are naturally located in LADs are also repressed. Here, we identified hundreds of promoters that are repressed in their native context inside LADs, as evidenced by their activation upon transfer to an episomal plasmid in the same cells. Furthermore, for three promoters from the repressed class, we confirm by TRIP that they indeed are extremely sensitive to a typical LAD environment, leading to ~40- to 130-fold...
Figure 6. Effects of H3K27me3 Domain Context on Promoter Activity

(A–G) Distribution of expression levels of barcoded reporters in iLAD regions that overlap with H3K27me3 domains (green), or not (blue). Data are from the same experiments as in Figures 3A–3G, with promoters of the repressed (purple; A–C) or escaper (yellow; D–F) class, or the iLAD-derived PGK promoter (blue; G). Horizontal black bars indicate medians; the fold-difference between the medians inside and outside of H3K27me3 regions is indicated for each promoter. n denotes the number of integrated reporters assayed. Data are averages of two independent experiments.

(H) Correlation between repressive effect of LADs (fold difference between LAD and iLAD median expression; Figures 3A–3G) and H3K27me3 domains (fold difference between H3K27me3 domain and non-H3K27me3 domain median expression; Figures 6A–6G) for the seven promoters tested by TRIP.

(I) SuRE versus GRO-cap analysis as in Figure 1B for iLAD promoters located in H3K27me3 domains (green) compared to iLAD promoters not located in H3K27me3 domains (blue). Red curve shows LAD promoter trend line as in Figure 1B.

See also Figure S3 and Data S2.
reduction in their median expression levels compared to iLAD contexts.

Escapers, i.e., promoters that are active inside LADs, have been observed before (Guelen et al., 2008; Wu and Yao, 2017), but it was not known how they overcome the repressive LAD environment. Our results identify two complementary determinants. First, we find that escaper promoters are intrinsically more resistant to the repressive LAD context than repressed promoters. This resistance is not generally due to a higher intrinsic transcriptional activity, although in some instances, this may contribute. Instead, we suggest that escaper promoters contain sequence motifs that recruit a class of TFs that are somehow more effective in overcoming LAD repression. In addition, our analysis of escaper enhancers revealed a general enrichment of motifs that bind known and predicted pioneer TFs. It is possible that escaper promoters also rely on pioneer TFs, but we may have lacked statistical power to detect any enrichment of pioneer TF motifs in escaper promoters. Future experimental studies may investigate the contribution of pioneer TFs more directly. This may be challenging, however, because the escaper effect most likely does not depend on individual TFs but rather on the combined effect of multiple TFs.

Our TRIP data also show that the activity of escaper promoters is higher in LAD regions where NL interactions as detected by DamID are weaker and when various active marks are nearby. Escaper promoters were previously found to carry such features (Wu and Yao, 2017), and our data are in agreement with this. DamID signal strength is known to correlate with NL contact frequencies (Kind et al., 2015), suggesting that escaper promoters may be active only when not contacting the NL. Importantly, our TRIP experiments help to interpret such correlative data in terms of causality, because these experiments show that the activity of escaper promoters is higher when they are inserted in or near regions that carried these features prior to integration of the reporters. It is thus highly likely that the activity of escapers in the native context is also at least partially facilitated by these features.

About half of all promoters located in LADs cannot be activated by transfer to an episomal plasmid. The lack of activity of these promoters is most likely explained by the absence of an activator. This missing activator may be either a promoter-binding TF that is not expressed in K562 cells or a distal enhancer that is inactive or unable to contact the promoter. The repressive environment of LADs may provide an additional lock to keep these genes inactive (Peric-Hupkes et al., 2010). However, it is also possible that some inactive genes form LADs as a consequence of their lack of activity. In support of this, it has been shown that forced activation of genes in LADs can lead to their detachment from the NL (Therizols et al., 2014).

We emphasize that our experimental strategy was not designed to address whether NL contacts, rather than the heterochromatic state of LADs, contribute to the repressive nature of LADs. Artificial anchoring of some genes to the NL can lead to reduced transcription (Finlan et al., 2008; Kumaran and Spector, 2008; Reddy et al., 2008), as can tethering of lamins to promoters of transfected reporter plasmids (Lee et al., 2009). The local detachment of escaper promoters from the NL might indicate that NL contacts are important for repression, as does our observation that local NL contacts are partially predictive of the degree of repression of integrated reporters. However, it is likely that the heterochromatin packaging of LADs also plays an important role. Future studies may unravel the precise links between NL contacts and heterochromatin formation.

Finally, our analysis of SuRE and GRO-cap data indicate that H3K27me3 domains are generally equally repressive domains as LADs. Our TRIP data confirm the repressive effect of H3K27me3, but surprisingly, the promoters tested by TRIP

![Figure 7. LAD Repression of Enhancer RNA Production; Effect of Pioneer TF Binding Sites](image-url)
appear less sensitive to H3K27me3 contexts than to LAD contexts. One interesting possibility is that promoters naturally embedded in H3K27me3 domains have evolved to be particularly sensitive to this chromatin environment. It is also possible that H3K27me3 heterochromatin cannot be established easily on a promoter that is already active at the moment of integration. We cannot rule out that the PiggyBac transposon sequences used in TRIP partially counteract the repressive effect of H3K27me3 (more than the repressive effect of LAD chromatin), for example, by specifically impeding the spreading of H3K27me3 chromatin into the promoters. Future studies may unravel the molecular architecture of LADs and the molecular mechanisms of gene repression and escape inside these domains.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - Promoter and enhancer definitions
  - Endogenous and plasmid promoter/enhancer activity
  - LAD promoter and enhancer classifications
  - Selection of promoters for TRIP
  - TRIP plasmid construction and generation of TRIP cell pools
  - Analysis of TRIP expression data
  - Mapping of TRIP reporter integrations
  - Multiplexed measurement of promoter activity in plasmid context
  - LMNB1 DamID-seq
  - DamID-seq data processing
  - Chromatin immunoprecipitation data analysis
  - Calling domains of TRIP integrations
  - Calling of endogenous domains
  - TT-seq
  - Statistical modeling of TRIP and epigenome data
  - Sequence motif analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND SOFTWARE AVAILABILITY**
- **ADDITIONAL RESOURCES**

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2019.03.009.

**ACKNOWLEDGMENTS**

We thank the NKI Genomics Core Facility for technical support. We thank Harmen Bussemaker, Vincent FitzPatrick, members of our laboratory, the Division of Gene Regulation, and the 4DN Center for Nuclear Cytomics for helpful discussions. This work was supported by NIH Common Fund 4D Nucleome Program (grant U54DK107965), ERC advanced grants 293662 and 694466, and ZonMW TOP to B.v.S. and an early postdoc mobility fellowship from the Swiss National Science Foundation to F.C. The Oncode Institute is supported by KWF Dutch Cancer Society.

**REFERENCES**

Abegessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H., and Kasukawa, T. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. Sci. Data 4, 170107.

Achter, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F.A., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell 154, 914–927.

Achter, W., Pindyurin, A.V., de Jong, J., Pagie, L., Ten Hoeve, J., Berns, A., Wessels, L.F.A., van Steensel, B., and van Lohuizen, M. (2014). Using TRIP for genome-wide position effect analysis in cultured cells. Nat. Protoc. 9, 1255–1281.

Altshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. 19, 229–244.

Amendola, M., Venneri, M.A., Biffi, A., Vigna, E., and Naldini, L. (2005). Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. Nat. Biotechnol. 23, 108–116.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27, 1653–1659.

Bian, Q., Khanna, N., Alvikas, J., and Belmont, A.S. (2013). β-globin cis-elements determine differential nuclear targeting through epigenetic modifications. J. Cell Biol. 203, 767–783.

Brown, G. (2018). GreyListChIP: Grey Lists – Mask Artefact Regions Based on ChIP Input. R package version 1.14.0.

Bulut-Karslioglu, A., De La Rosa-Veláquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G.E., Engist, B., Gerle, B., et al. (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol. Cell 55, 277–290.

Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. 32, 202–223.

Chuang, C.H., Carpenter, A.E., Fuchsova, B., Johnson, T., de Lanerolle, P., and Belmont, A.S. (2006). Long-range directional movement of an interphase chromosome resident. Curr. Biol. 16, 825–831.

Comoglio, F., and Paro, R. (2014). Combinatorial modeling of chromatin features quantitatively predicts DNA replication timing in Drosophila. PloS Comput. Biol. 10, e1003419.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet. 46, 1311–1320.

Corrales, M., Rosado, A., Cortini, R., van Arensbergen, J., van Steensel, B., and Fillon, G.J. (2017). Clustering of Drosophila housekeeping promoters facilitates their expression. Genome Res. 27, 1153–1161.
de Jong, J., Akhtar, W., Badhaj, J., Rust, A.G., Rad, R., Hilkens, J., Berns, A., van Lohuizen, M., Wessels, L.F.A., and de Ridder, J. (2014). Chromatin land- 
scapes of retroviral and transposon integration profiles. PLoS Genet. 10, e1004250.

Diaíynas, G., Speese, S., Budnik, V., Geyer, P.K., and Wallrath, L.L. (2010). The role of Drosophila Lamin C in muscle function and gene expression. Develop- 
ment 137, 3067–3077.

Dirks, R.A.M., Stunnenberg, H.G., and Marks, H. (2016). Genome-wide epige- 
nomic profiling for biomarker discovery. Clin. Epigenetics 8, 122.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Ep- 
stein, C.B., Fritze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Ehsani, R., Bahrami, S., and Drabiles, F. (2016). Feature-based classification of human transcription factors into hypothetical sub-classes related to regulatory 
function. BMC Bioinformatics 17, 459.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet. 4, e1000399.

Fishilevich, S., Nadal, R., Rappaport, N., Hadar, R., Plaschkes, I., Iry Stein, T., Rosen, N., Kohan, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. (Oxford) 2017, bx0268.

Follenzi, A., Alles, L.E., Bakovic, S., Geuna, M., and Naldini, L. (2000). Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. Nat. Genet. 25, 217–222.

Forrest, A.R., Kawaji, H., Rehl, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Roh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGTF) (2014). A promoter-level mammalian expression atlas. Nature 507, 462–470.

Gonzalez-Sandoval, A., and Gasser, S.M. (2016). On TADs and LADs: spatial control over gene expression. Trends Genet. 32, 485–495.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eus- 
sen, B.H., de Kleijn, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453, 948–951.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin and a-type lamins. J. Cell Biol. 208, 33–52.

Harwood, J., Frankish, A., Gonzalez, J.M., Tapanainen, E., Diekhaus, M., Kocokin- 
ski, F., Aken, B.L., Barrett, D., zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and Zhao, K. (2013). H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell 12, 180–192.

Kelsey, G., Stegle, O., and Reik, W. (2017). Single-cell epigenomics: reading the past and predicting the future. Science 358, 69–75.

Kind, J., Pagie, L., Ortaboykoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. Cell 153, 178–192.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., DeY., S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell 163, 134–147.

King, H.W., Fursova, N.A., Blackledge, N.P., and Kloze, R.J. (2018). Polycomb repressive complex 1 shapes the nucleosome landscape but not accessibility at target genes. Genome Res. 28, 1494–1507.

Ku, M., Jaffe, J.D., Koche, R.P., Rheinbay, E., Endoh, M., Koseki, H., Carr, S.A., and Bernstein, B.E. (2012). H2A.Z landcsapes and dual modifications in pluripotent and multipotent stem cells underline complex genome regulatory functions. Genome Biol. 13, R85.
## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Bacterial and Virus Strains** | | |
| CloneCatcher DH5G electrocompetent *Escherichia coli* | Genlantis | Cat#C810111 |
| E. coli 10 g | Lucigen | Ca#60107-1 |
| pCCL.sin.cPPT.hPGK,ΔLNGFR,Wpre lentiviral construct | Amendola et al., 2005 | N/A |
| VSV-pseudotyped third-generation lentivirus | Follenzi et al., 2000 | N/A |
| **Chemicals, Peptides, and Recombinant Proteins** | | |
| Lipofectamine 2000 | ThermoFisher | Cat#11668019 |
| ISOLATE II Genomic DNA Kit | Bioline | Cat#BIO-52067 |
| CutSmart buffer | New England Biolabs | Cat#B7204S |
| NEBuffer DpnII | New England Biolabs | Cat#B0543S |
| DpnI | New England Biolabs | Cat#R0176L |
| DpnII | New England Biolabs | Cat#R0543L |
| ISOLATE II PCR and Gel Kit | Bioline | Cat#BIO-52060 |
| End-It DNA end-repair kit | Lucigen | Cat#ER81050 |
| T4DNA ligase | Roche | Cat#10909246103 |
| 2x MyTaq Mix (200 rxn) | Bioline | Cat#BIO-25041 |
| CleanPCR beads | CleanNA | Cat#CPCR-0050 |
| Klenow | New England Biolabs | Cat#NEBM02125 |
| **Deposited Data** | | |
| SuRE data re-mapped to hg38 | Open Science Framework | https://osf.io/6qwj2/ |
| Processed TRIP data | Open Science Framework | https://osf.io/6qwj2/ |
| Raw sequence data TRIP | NCBI Sequence Read Archive | https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA504533 |
| LaminB1 Dam-ID | 4D Nucleome Project | https://data.4dnucleome.org/experiments-damid/4DNEXZKHQXYY/ |
| Published ChIP-Seq datasets used in this study, see Table S2 | this paper | N/A |
| Lab notebook extract of TRIP experiments | Open Science Framework | https://osf.io/6qwj2/ |
| **Experimental Models: Cell Lines** | | |
| K-562 Homo sapiens bone marrow chronic myelogenous leukemia (CML). Cells were tested for mycoplasma every 2-3 months. | ATCC | ATCC CCL-243; RRID:CVCL_0004 |
| **Oligonucleotides** | | |
| Adr-PCR-Rand1: 5'-GGTCGCGGCGCGAGGATC-3' | IDT | N/A |
| AdRt: 5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGAGGATC-3' | IDT | N/A |
| AdRb: 5'-TCCTCGGGCGGAGGATC-3' | IDT | N/A |
| Y-adaptor-top: 5'-ACACTCTTCTTCCCTACAGACGCTTTCAGAATCT-3' | IDT | N/A |
| Y-adaptor-bottom: 5'-pGATCGGAAGACACGCTGCTCGGAGGTAAAGACTTTCAGAATCT-3' | IDT | N/A |
| P5-Illumina universal PCR adaptor: 5'-AATGATACGGGTGTTCGGCAGCGACACGCTTCTTCGATCT-3' | Illumina | N/A |
| **Recombinant DNA** | | |
| TRIP vector pPTK-Gal4-tet-Off-Puro-ires-eGFP-snRP-pA | GenBank | accession KC10228 |

(Continued on next page)
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact. Bas van Steensel (b.v.steensel@nki.nl).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

K562 cells are a human chronic myelogenous leukemia (CML) cell line from a 53 year old female. These cells were obtained from American Type Culture Collection (ATCC) and cultured according to 4D Nucleome guidelines (<https://www.4dnucleome.org/cell-lines.html>) in suspension at 37°C in Iscove’s Modified Dulbecco’s medium with 10% fetal bovine serum and 1% penicillin/streptomycin.

METHOD DETAILS

Promoter and enhancer definitions
For the comparison of GRO-cap and SuRE, promoters were taken from GENCODE (Harrow et al., 2012) version 27, with the following additional requirements. First, to focus on annotated promoters with likely functional relevance, we required that the promoters are active in at least one cell type or tissue according to the FANTOM5 database, reprocessed for hg38, using CAGE peaks lifted over from hg19 passing QC filters and CAGE peaks newly identified in hg38 (fair + new set), version 4 (Abugessaisa et al., 2017). For this criterion, only promoters with at least one CAGE peak within 50 bp were used. Second, to avoid overlapping data due to the resolution of SuRE, promoters were required to be at least 500 bp apart. For alternative promoters from the same gene located within 500 bp from each other, we selected the most active promoter, using the sum of the expression levels normalized by read depth for each experiment in the FANTOM5 database.

Enhancer coordinates were taken from the genehancer v4.8 database (downloaded from https://genecards.weizmann.ac.il/geneloc/index.shtml). We defined enhancer regions as 300 bp windows centered around the center of these enhancers. We only used enhancers that were at least 5 kb away from any promoter as listed by Gencode v27.

Endogenous and plasmid promoter/enhancer activity
GRO-cap data (Core et al., 2014) was aligned to the hg38 genome using Bowtie2. Similar to Core et al. (2014), reads were trimmed to 30 bases and first mapped to the human ribosomal DNA complete repeat unit (GenBank ID: U13369.1). Unaligned reads were then aligned to hg38 (without alternative haplotypes). Reads were aligned using Bowtie2. The bamcoverage tool from the deeptools package was subsequently used to create separate coverage tracks for forward and reverse strands. For this, using an offset and binsize of 1.

For SuRE data, reads were aligned to hg38 and processed as previously reported (van Arensbergen et al., 2017), but re-mapped to the hg38 version of the human genome.

Continued
For PRO-seq, similar processing was used as described (Core et al., 2014). Adapters were removed using cutadapt (Martin, 2011) and reads greater than 15bp were retained and first mapped to the human ribosomal DNA complete repeat unit (GenBank ID: U13369.1). Unaligned reads were subsequently aligned to hg38 (without alternative haplotypes). Bowtie2 was used for alignment. Reads with mapping quality of 42 and up to 2 mismatches were retained. The bamcoverage tool from the deeptools package was subsequently used to create separate coverage tracks of reads for forward and reverse strands using an offset and binsize of 1.

For CAGE data, strand specific coverage tracks for hg38 were downloaded directly from the FANTOM database (Abuigussaisa et al., 2017; Lizio et al., 2015).

A pseudocount equal to half the minimum expression was added before log10 transformation.

LAD promoter and enhancer classifications
The three classes of LAD promoters were defined based on a combination of GRO-cap signals, SuRE signals, and a “LAD Repression Score” (LRS). The LRS is the deviation of the measured GRO-cap signal of promoters in LADs from the average GRO-cap signal of promoters in iLADs with a matching SuRE signal. This was calculated as follows. First, all promoters were sorted based on their SuRE expression. Next, for 60 windows of 501 promoters, equally distributed across the sorted list, the average log10(GRO-cap) and log10(SuRE) values for the iLAD promoters were calculated. A complete curve was then obtained by linear interpolation of the 60 points (blue curve in Figures 1A and 1B). Subsequently, for all LAD promoters the LRS was calculated by subtracting the predicted GRO-cap score according to this curve from the measured GRO-cap score. Finally, LAD promoters with a measured log10(SuRE) value < −0.3 and a log10(GRO-cap) value < −2 were classified as inactive; LAD promoters with a log10(SuRE) value > 0.3, log10(GRO-cap) < −2 and LRS < −1 were classified as repressed. LAD promoters with a log10(GRO-cap) value > −2, a log10(SuRE) value > 0 and LRS > −0.5 were classified as escaper. The cutoff values are shown as dotted lines in Figure 1C.

A similar strategy was used to define the three classes of enhancers. Specifically, LAD enhancers with a measured log10(SuRE) value < −0 and a log10(GRO-cap) value < −2.8 were classified as inactive; LAD enhancers with a log10(SuRE) value > 0, log10(GRO-cap) < −2.8 were classified as repressed; LAD promoters with a log10(GRO-cap) value > −2 and a log10(SuRE) value > 0 were classified as escaper.

Selection of promoters for TRIP
The escaper promoters for TRIP were selected to have at least log10(SuRE) > 0.5 and log10(GROcap) > −0.5 and to show clearly detectable expression in mRNA-seq data from K562 cells (ENCODE accession ENCSR000CPH (Dunham et al., 2012)). We excluded promoters with other nearby known active elements such as other promoters. Similar criteria were used to select repressed promoters except that these promoters were selected for low GRO-cap and mRNA-seq levels. Detailed information on the selected promoters is provided in Table S1.

TRIP plasmid construction and generation of TRIP cell pools
TRIP was essentially performed as described (Akhtar et al., 2014). The piggyBac reporter construct of Akhtar et al. (2013) was used, except that the 14 GAL4 repeats, the mouse phosphoglycerate kinase (mPGK) promoter, the Puromycin resistance cassette (PuroR) and the internal ribosome entry site (IRES) were replaced by the promoter of interest. For each promoter, libraries of plasmids with 16 bp random barcodes were generated by electroporation into CloneCatcher DH5G electrocompetent Escherichia coli (#C810111; Genlantis) (PGK, ADAMTS1, ARHGEF9, THEM106B) or into E. cloni 10G (#60107-1; Lucigen) (MED30, ZNF300 and BRINP1). Plasmid libraries had a complexity in the range of 40000–352500 (estimated by colony counting).

Next, 15 µg of plasmid DNA from each library was co-transfected with 5 µg mCherry-expressing plasmid in 4 million K562 cells using Lipofection with Lipofectamine 2000 (ThermoFisher #11668019). Successfully transfected cells were obtained by FACS sorting of mCherry-positive cells, and cell pools carrying random reporter integrations were obtained as described (Akhtar et al., 2014).

Analysis of TRIP expression data
For each promoter, two independent TRIP experiments were done with two independent cell pools each. Quantification of barcodes in cDNA and genomic DNA was performed as described (Akhtar et al., 2014), using Illumina HiSeq2500 sequencing with read length 75 bp.

Barcodes were extracted from cDNA and gDNA reads using an in-house script using functions from cutadapt. After demultiplexing, the start of the sequence was matched to the expected sequence GTCAACAAGGCGCAGCACAACCTCGAG, allowing for an error-rate of less than 0.1 (1 mismatch per 10 bp). If this sequence matched the read, the subsequent 16 basepairs were classified as barcode and stored only if the sequence after this barcode matched the expected sequence TGAATCTGGGTG with the same maximum error-rate as the preceding sequence. IUPAC codes in sequences are also considered as matching (e.g., N matches any nucleotide).

To identify the total set of genuine barcodes in each cell pool we used the barcodes extracted from the gDNA reads. We required each barcode to be present in at least 5 reads. To eliminate aberrant barcodes arising from mutations during PCR and sequencing, we applied starcode (Zorita et al., 2015), using sphere clustering with a maximum Levenshtein distance of 2.
Next, for the resulting set of genuine barcodes, the counts in both cDNA and gDNA reads were determined. A pseudocount of 1 was added to the cDNA counts. Subsequently, counts for both cDNA and gDNA were normalized to the total respective barcode counts, and resulting cDNA values were divided by gDNA values. After log₂-transformation these ratios were averaged across the two replicate experiments to obtain the expression score for each barcode. Data from two independent cell pools (each with different sets of integrations) were then combined.

Mapping of genomic integration sites together with barcode identities was done by inverse PCR (iPCR) followed by 2 × 75 bp paired-end sequencing on an Illumina HiSeq2500 as described (Akhtar et al., 2014). Parsing of the reads was done as follows. The first read in each pair was used to extract the barcode. This was done by first removing GTCAAGGGCCGCCCACAC constant sequence and subsequently matching a regular expression TCGAG[ACGT][16]TGATC. From this sequence, the 16 bp barcode was extracted. Barcodes were matched to the genuine set of barcodes obtained from gDNA sequencing reads and variations upon this set of barcodes likely caused by errors were discarded.

The second read of each pair was used to locate the site of integration after removing the GTAGTCACAATGATTTATCTTC region which matches the transposon arm. The flanking sequence was aligned to hg38 using bowtie2 using the very-sensitive-local option (20 seed extension attempts, up to 3 re-seed attempts for repetitive seeds, 0 mismatches per seed, with a seed-length of 20 and using a multi-seed function \( f(x) = 1 + 0.5 \cdot \sqrt{x} \)). Locations of integration sites were required to be supported by at least 5 reads with an average mapping quality larger than 10 at the primary location, having at least 70% of the reads located at this locus, with not more than 10% of the reads at a secondary location.

Multiplexed measurement of promoter activity in plasmid context

For each promoter, the barcoded TRIP plasmid library was re-transformed into DH5-alpha E. coli and plated; 100 colonies were picked, pooled, and expanded to produce plasmid mini-libraries with 100 different barcodes each. The barcodes in these mini-libraries were PCR amplified and sequenced by Illumina MiSeq to identify the barcodes belonging to each promoter. Next, the plasmid mini-libraries of the 7 promoters were mixed in equal proportions and transfected into K562 cells by the same nucleofection method as used for SuRE (van Arensbergen et al., 2017). Two days after transfection, barcodes were counted in cDNA by PCR amplification and Illumina sequencing, similar to TRIP. The same was done for the input plasmid DNA. For each barcode, cDNA counts were then normalized to plasmid counts to obtain the relative expression. The resulting values were log₂-transformed, and results from independent experiments were averaged. We performed a total of six independent experiments of which two included the complete set of barcoded promoters, two included all promoters, but PGK was not marked by barcodes; two experiments were done without PGK, of which 1 was performed in unequal ratios 1:2 (BRINP1, TMEM106B and ZNF300);(ADAMTS1, ARHGEF9 and MED30).

**LMNB1 DamID-seq**

K562 cells were grown according to the ATCC culture protocol (https://www.lgcstandards-atcc.org). Cells were lentiviral transduced with Dam or Dam-LMN1 constructs, both fused to a destabilization domain and under control of the PGK1 promoter (Amendola et al., 2005; Kind et al., 2013). A GFP sample was added to estimate transduction efficiency. To keep expression low, Dam fusion proteins were not stabilized with Shield1. Three days after transduction, gDNA was isolated using the Bioline Isolate II genomic DNA kit (BIO-52067). Additional RNase A (4 μL of 10 mg/μL) was added during cell lysis.

DamID was performed similar to Vogel et al., 2007, but modified for next generation sequencing as described below. 500 ng of DNA was digested with DpnI (0.5 μl DpnI (20U/μL New England Biolabs #R0176L), 1 μl 10x CutSmart) in 10 μL total volume for 8 hours at 37°C, followed by 20 minutes of heat inactivation at 80°C. Adapters were ligated to the digested fragments by adding 10 μL of ligation mixture (0.5 μL T4 ligase (Roche #10909246103), 2 μL 10x ligation buffer, 0.25 μL DamID adaptor (dsAdR 5’-GGATC-3’ and 5’-Y-adaptor (50 μM) and 1.5 μL H2O). After an overnight incubation at 16°C, ligase is heat inactivated for 10 minutes at 65°C. Unmethylated GATC sequences were cleaved with DpnII (addition of 1 μl DpnII (10U/μL New England Biolabs R0543L), 5 μl DpnII buffer and 24 μl H2O) by incubating 1 hour at 37°C. During these steps, controls without DpnII and ligase were included to assess specificity of the amplification. 8 μL of the final digestion was added to a PCR mixture (20 μl 2xMyTaq (Bioline Bio-25041), 1 μl primer (Adr-PCR-Rand1, 50 μM) and 11 μL H2O) and amplified with the following PCR protocol: 8 minutes at 720°C, cycles of 20 s 94°C, 30 s 580°C and 20 s 720°C, and finally 2 minutes at 720°C. 23 and 22 amplification cycles were used for replicates 1 and 2, respectively. 4 μL of the PCR mixture was put on gel to verify specific amplification, while the rest was column purified using Bioline Isolate II PCR and Gel kit (Bioline BIO-52060). Sequencing libraries were prepared by a series of end-repair in 50 μL (End-It, Lucigen, ER81050), column purification (Bioline BIO-52060), Klenov 3A-overhang in 50 μL (Klenov 3’ → 5’ exon., New England Biolabs, M02125) and bead purification (1.8x CleanPCR beads, CleanNA, CPCR-0050), all following provider’s protocols. DNA concentration was measured by Nanodrop and ~250 ng of DNA was resolved in 6.5 μL H2O. Y-shaped adapters were ligated overnight at 160°C, by adding 3.5 μL ligation mixture (0.5 μl T4DNA ligase (5U/μL Roche #10909246103), 1 μl 10x ligation buffer, 0.5 μl Y-adaptor (50 μM) and 1.5 μL H2O), followed by 10 minutes
heat inactivation at 650°C and another round of bead purification resolving in 20 μL H2O (1.8x CleanPCR beads, CleanNA, CPCR-0050). Illumina indices were added with amplification, using 8 μL bead purified DNA in a 20 μL PCR reaction (10 μL 2xMyTaq (Bioline BIO-25041), 0.5 μL Illumina P5 primer (5.0 μM), 0.5 μL Illumina P7 indexing primer (5.0 μM) and 1 μL H2O). The following PCR protocol was used: 1 minute 940°C, cycles of 30 s 940°C, 30 s 580°C and 30 s on 720°C, followed by 2 minutes at 720°C. 10 and 11 cycles were used for replicates 1 and 2, respectively, and 4 μL was put on gel. Based on smear intensities, samples were pooled and cleaned with bead purification (1.6x CleanPCR beads, CleanNA, CPCR-0050). Samples were sequenced on a HiSeq 2500 with High Output Mode of single-end 65 bp reads.

**DamID-seq data processing**

First, the constant sequence of the DamID adaptor was trimmed from the 65 bp single-end reads (with cutadapt 1.11 and custom scripts). The remaining gDNA starting with GATC was mapped to a combination of GRCh38 v15 (without alternative haplotypes) and a ribosomal model (GenBank: U13369.1) with bwa mem 0.7.17. Reads with a mapping quality of at least 10 were counted to GATC fragments. The middle of the GATC fragment was used to combine these counts into bins of various sizes. Only bins with at least 10 reads (combined target + Dam-only) were subsequently normalized. These bins were first normalized to 1M reads and with a pseudocount of 1 a log2-ratio over the Dam-only control was calculated. LADs were defined by running a hidden markov model over the normalized values (using the R-package HMMit; https://github.com/gui11aume/HMMit). BigWigs tracks were generated of the normalized counts per GATC-fragment. These normalized BigWigs were used to calculate average count in 100bp bins around the TSS using the computeMatrix function from deeptools. These bins were used to calculate a running sum using the sum of 5 bins for each 100bp bin. These running sums were then used to calculate a fold-change of the LmnB1-Dam over Dam signal.

**Chromatin immunoprecipitation data analysis**

We re-processed published ChIP-seq data from various sources (Table S2) for consistency. Raw sequencing data was obtained from the sequence read archive (SRA). Reads were aligned to the human genome hg38 using bowtie2 with default options. Replicates for sample data were processed separately, while the sequences from the input were combined. After alignment, reads were filtered on a minimum mapping quality of 30 and duplicate reads were removed except for the reads coming from experiments using augmentation, in which duplicates were kept. After this, regions were masked based on blacklist regions identified by the ENCODE project (ENCFF419RSJ) (Dunham et al., 2012) and artifact regions identified based on the input reads using chipseq-greylist, a python implementation of GreyListChIPs (Brown, 2018). We considered ChIP-seq datasets to be of sufficient quality for our analyses if there was well annotated input and sample data available and consistent read lengths were used. Filtered alignments were used to call domains with significant enrichment by hiddenDomains with a binsize of 1kb and minimum posterior of 0.9. Overlapping domains were selected between replicate experiments. For the calling of endogenous genes located in H3K27me3 regions, binsize was set to 20kb and a posterior of 0 was used.

Mean Chip-seq signals for specific regions (e.g., TSSs, genebodies, TRIP integration sites) were calculated by taking the sum of the reads in the region, scaling input and sample counts by the smallest library size, adding a pseudo count of 1 and subsequently dividing sample over input normalized counts. After this, replicate experiments were averaged. Enrichment signal along a window around sites of interest (e.g., TSS) was calculated by first using computeMatrix function of the deeptools package to calculate read coverage in 100bp bins centered around the site of interest (Ramirez et al., 2016). After calculating coverage, smoothed average signal for each group of interest was calculated by using running mean with a width of 9 bins. Coverage was then normalized by library size and enrichment over input was calculated.

**Calling domains of TRIP integrations**

TRIP integration sites were called as either LAD or H3K27me3 localized when the TTAA sequence of integration falls within either domain. Overlap was calculated using the intersect function of bedtools (Quinlan and Hall, 2010). For H3K27me3 domains both datasets by Schmidl et al. (2015) were used and integrations overlapping with either dataset were classified as H3K27me3 domain integrations.

**Calling of endogenous domains**

Endogenous genes were called as either LAD or H3K27me3 localized when the complete gene-body fall within respective domains and are at least 1 binsize (5kb and 20kb respectively) from the border. Overlap was calculated using the intersect function of bedtools (Quinlan and Hall, 2010). For H3K27me3 domains, unlike previous TRIP data, overlap was calculated based on individual replicates, genes completely overlapping with at least 2 replicates were called as located inside H3K27me3 domain.

**TT-seq**

TT-seq data from K562 cells (Schwalb et al., 2016) was kindly provided by B. Schwalb and P. Cramer as 4 BigWig files, two for each replicate (one file for forward and one for reverse orientation relative to the hg38 reference genome). Windows around the TSS were created similarly to the approach used for the DamID-seq and ChIP-seq data by first using the computeMatrix function. For
each TSS, the bins in sense orientation with its gene were used to calculate running means with a width of 3 bins. Subsequently the average for each class and across 2 replicates was calculated.

**Statistical modeling of TRIP and epigenome data**

For statistical modeling of TRIP expression, mean ChIP-seq and DamID signals were calculated within a region centered on the integration site (5kb and 10kb respectively). Mean values were averaged between experiments. Proximity to the nearest domain was calculated as $-\log_{10}(\text{distance} + 1)$. For experiments with the same target, the minimum distance was used. For integrations inside LADs, the proximity to nearest border was calculated.

To predict barcode expression values (log2), L1-regularized linear regression models (lasso) were trained using mean ChIP-seq/DamID signals at the integration sites along with proximity features. Only TRIP integrations within LADs were considered for each promoter class (escaper or repressed). A total of 100 models were fitted for each class. For each model, data points were first randomly split into a training set (80%) and a test set (20%). The model was trained with 10-fold cross-validation using the glmnet package v2.0-16. The value of the regularization parameter that minimized the cross-validated mean squared error was used for model selection. The prediction accuracy (R²) of the selected model was computed on the test set. For promoter-specific models, both LAD and iLAD integrations were considered.

Feature importance analysis was performed using bootstrap-Lasso essentially as described in Comoglio and Paro (2014). The importance of a feature corresponds to its selection probability (normalized frequency of non-zero coefficients) across all bootstrap-Lasso models. All features with a selection probability > 0.7 were considered.

**Sequence motif analysis**

For the motif analysis, enhancers identified as escapers were compared to a subset of enhancers of the repressed class, matched on SuRE activity. All escaper and repressed enhancers were sorted on SuRE expression and subsequently all enhancers ranking right before and after each escaper enhancer in the list were selected. Of this group of enhancers, the repressed enhancers were used for the matching set.

*DREME* was used to discover denovo motifs enriched in escapers with repressed enhancers as background (Bailey, 2011). The number of most significant words passed in the beam search was set to 1000 and the maximum word-size was set to 10. The other settings were left to default (E-value Threshold 0.05, minimum word-size of 3). Both forward and reverse strand were processed.

For each enhancer the 300bp region around the center was extracted from the human genome version hg38 using the *getfasta* command from *bedtools* (Quinlan and Hall, 2010). Subsequently, for a well curated set of motifs, we approximated position specific affinity matrices (PSAM) by transforming the position weight matrix (PWM) by dividing each column, representing 1bp, by it’s maximum weight value. AffinityProfile from the *REDUCE* Suite was used to calculate the sum of affinities calculated along both orientations of the 300bp sequence (Roven and Bussemaker, 2003). Motifs were classified as either pioneer or non-pioneer based according to Ehsani et al. (2016). For every motif a median affinity was calculated for each enhancer class and the median for escapers was divided by the median for repressed. Only TFs that are detectably expressed in K562 cells (RNA FPKM > 1 in ENCODE mRNA-seq dataset ENCSR000CPh) were included.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Pearson’s r and Spearman rank correlations for all iLAD promoters (n = 26968) and LAD promoters (n = 4075) were reported in the main text referencing Figure 1. Calculations were done in base R. P value depicted in Figure S1 was performed by Fisher’s exact test between 3289 iLAD and 2048 LAD promoters GRO-cap using the *fisher.test* function in R. P values shown in Figure 2 and S2 between 238 escaper promoters and 236 expression matched iLAD promoters were calculated using *wilcox.test* inside *stat_compare_means* in the *ggplot2* package in R. The linear model depicted in Figure 3H was calculated in the *lm* function in base R with a total of 4798 repressed barcode intergrations and 4500 escaper promoters.

R² values depicted in Figure 5 and described in the main text are explained in the “Statistical modeling of TRIP and epigenome data” in the Method Details section. LAD specific models were performed using 1151 escaper and 1107 barcode integrations in LADs. For promoter specific models a total of 679 ADAMTS1, 2567 ARHGEF9, 1552 BRINP1, 2177 MED30, 629 TMEM106B and 1694 ZNF300 barcode intergrations were used.

*DREME* was used to calculate de-novo motif enrichment in Figure 7 on 927 escaper and 1433 SuRE expression matched repressed enhancers. *DREME* reports corrected p values of Fisher’s exact tests.

Largest vertical lines in categorical dot plots represent median values, smaller lines represent 25% and 75% quantiles. Fold changes were calculated between median values.

**DATA AND SOFTWARE AVAILABILITY**

Custom scripts were made to perform TRIP data analysis, ChIP pipeline and the R scripts for statistical analysis and figure generation, are available at [https://github.com/vansteensellab/Promoters_in_LADs](https://github.com/vansteensellab/Promoters_in_LADs).

Raw sequence reads of TRIP experiments were deposited in the Sequence Read Archive [https://www.ncbi.nlm.nih.gov/sra], BioProject ID: PRJNA504533. Processed data are available as Data S1, S2, and S3, and can also be downloaded from [https://osf.io/6qwj2/].

**ADDITIONAL RESOURCES**

An extract of the lab notebook records describing the TRIP experiments with escaper and repressed promoters is available at [https://osf.io/6qwj2/>. 
Supplemental Figures

Figure S1. Promoters with SuRE Activity Inactive in Endogenous Context, Related to Figure 1
(A) Percentage of promoters active in episomal plasmid context (log_{10}(SuRE) > 0) with no measurable activity in native context (log_{10}(GRO-cap) < -3), for LAD promoters and a subset of iLAD promoters matched on SuRE activity.
(B) Distributions of expression levels in episomal plasmid context (log_{10}(SuRE)) of matched promoter sets used in (A).
(C) Comparison between endogenous (log_{10}(CAGE)) and episomal (log_{10}(SuRE)) expression for all promoters in LADs (red) and iLAD regions (blue). Similar to Figure 1B, but with CAGE instead of GRO-cap. Blue and red lines are sliding window averages with a bin size of 501 promoters along the SuRE expression axis.
(D) Same analysis as in Figures 1B and S1C, but with log_{10}(PRO-seq) as a measure of endogenous expression.
(E) Comparison between CAGE and SuRE for the three classes of LAD promoters as defined in Figure 1C.
(F) Comparison between PRO-seq and SuRE for the three classes of LAD promoters as defined in Figure 1C.
Figure S2. Properties of Escaper Promoters and Their Genes in Native Context, Related to Figure 2
(A) Average DNase-seq profile of each promoter class. Data are from (Dunham et al., 2012).
(B and C) Average H3K9me2 (B) and H3K9me3 (C) signals around each promoter class according to ChIP-seq. Data are from (Salzberg et al., 2017).
(D) Distribution of GRO-cap signals of the matched sets of iLAD and escaper promoters used in Figures 2B–2F, S2A–S2C, and S2G–S2L. Promoters in iLADs were matched on GRO-cap activity of escaper promoters.
(E) Same as Figure 2B, but with iLAD promoters matched on SuRE activity.
(F) Distribution of SuRE values of the matched sets of iLAD and escaper promoters analyzed in (E).
(G) PRO-seq signals at TSS and gene bodies of matched gene sets as in (D).
(H–L) ChIP-seq signals of proteins that have been implicated in regulation of elongation, on matched promoter sets as in (D). Data in (H) are from (Tyler et al., 2017); in (I–L) from (Dunham et al., 2012).
Figure S3. TRIP Construct and Mapping of Integrations, Related to Figures 3, 5, and 6

(A) TRIP construct design. PiggyBac: terminal repeat regions of PiggyBac transposon; GFP: green fluorescent protein open reading frame; sNRP-1 pA: poly-A signal.

(B–H) Locations of TRIP reporter integrations that could be uniquely mapped to the genome and assigned to a unique barcode, for each promoter. Red: integrations in LAD, blue: integrations in iLADs.
Figure S4. Chromatin Features Used for Modeling TRIP Expression, Related to Figure 5

(A) Distributions of log₂(signal/control) values (ChIP or DamID) for 5 kb regions centered on TRIP integration sites, for each feature used as input for the modeling in Figure 5.

(B) Distributions of the proximities \(-\log_{10}(\text{distance [bp]} + 1)\) of TRIP integrations to the nearest peak or domain of each feature used as input for the modeling in Figure 5.
Figure S5. Modeling TRIP Expression Levels as Function of Chromatin Features, Related to Figure 5
(A) Distribution of R² values of 100 lasso regression models of TRIP data for escaper and repressed promoters using integrations in iLADs only. Black lines represent the median R² values.
(B) Distribution of R² values of 100 lasso regression models of TRIP data for individual promoters using all available integrations (both LAD and iLAD). Black lines represent the median R² values.