The Transforming Growth Factor-β (TGFβ) signaling pathway controls transcription by regulating enhancer activity. How TGFβ-regulated enhancers are selected and what chromatin changes are associated with TGFβ-dependent enhancers regulation are still unclear. Here we report that TGFβ treatment triggers fast and widespread increase in chromatin accessibility in about 80% of the enhancers of normal mouse mammary epithelial-gland cells, irrespective of whether they are activated, repressed or not regulated by TGFβ. This enhancer opening depends on both the canonical and non-canonical TGFβ pathways. Most TGFβ-regulated genes are located around enhancers regulated in the same way, often creating domains of several co-regulated genes that we term TGFβ regulatory domains (TRD). CRISPR-mediated inactivation of enhancers within TRDs impairs TGFβ-dependent regulation of all co-regulated genes, demonstrating that enhancer targeting is more promiscuous than previously anticipated. The area of TRD influence is restricted by topologically associating domains (TADs) borders, causing a bias towards co-regulation within TADs.
**Results**

**TGFβ provokes widespread enhancer chromatin opening.** To better understand the mechanism by which TGFβ signaling controls transcription in epithelial cells, we investigated early-chromatin changes provoked by TGFβ treatment. TGFβ signals through several transduction pathways, most of which are stimulated by growth factors present in the serum. Therefore, to clearly determine all the changes promoted by TGFβ, NMuMG cells were depleted of serum and insulin (an additive of the NMuMG cell culture medium) for 6 h, prior to TGFβ treatment. Then, samples were taken after 2 h to monitor fast chromatin modifications, and after 12 h (when mesenchymal-like morphological changes are first observed) to monitor late chromatin modifications (Fig. 1a). To systematically identify functional enhancers, we first determined chromatin accessibility by assay for transposase-accessible chromatin followed by high-throughput sequencing (ATAC-seq)\(^{29}\). A total of 32,626, 100,459, and 105,334 significant peaks (IDR (irreproducible discovery rate) <0.05) were identified from vehicle, 2 h and 12 h TGFβ samples, respectively (see correlations between replicates, Supplementary Fig. 1a). Peaks with summits closer than 1000 bp were considered as overlapping and merged. All ATAC-seq peaks found in any of the three tested conditions were then merged, resulting in 39,432 regions of open chromatin. Of these, 10,362 were coincident with previously defined transcription start sites (TSS), and 29,071 were considered to be putative enhancers. Notably, TGFβ treatment (at 2 h and 12 h) provokes a widespread increase of chromatin accessibility at both TSS and putative enhancers, concomitant with a decrease of ATAC signal in the rest of the genome (Supplementary Fig. 1b). Pervasive increase of chromatin accessibility was confirmed by DNasel-seq (Supplementary Fig. 1c, d). Thus, 88.3% of the DNasel-seq peaks (20,908) were coincident with ATAC peaks and 76.8% of the DNasel-ATAC common peaks increased their signal at 2 h of TGFβ treatment. Furthermore, a very good correlation was observed between DNasel-seq and ATAC-seq fold changes (FCs).

We next imaged the transposase-accessible chromatin in situ using ATAC-seq\(^{28}\), which uses microscopy to visualize accessible chromatin at the single-cell level, by inserting fluorophores by the Tn5 transposase at open-chromatin sites. First, we confirmed that TGFβ treatment promoted a general increase of the ATAC-seq signal (Fig. 1b). Signal was visible as a general nuclear background, but also concentrated into puncta. TGFβ treatment increased the total intensity of the ATAC-seq signal per nuclei by about fourfold, with slightly fewer puncta per nucleus, as shown by 3D confocal microscopy image analysis using TANGO software\(^{30}(Fig. 1c–f)\). Notably, 2 h of TGFβ treatment strongly increased the volume and fluorescence intensity of the puncta (Fig. 1g, h). Kinetic analysis revealed that general chromatin opening occurred as early as 10 min after TGFβ addition (Fig. 1i, j), coinciding with the reported accumulation of SMAD2/3 and SMAD4 in the nucleus of TGFβ treated cells\(^{31}\) and with the activation of the ERK pathway (Supplementary Fig. 2a). We verified that TGFβ-dependent widespread chromatin opening also occurred in cells not deprived of growth factors prior to TGFβ stimulation (Supplementary Fig. 2b–e) or with severe serum starvation for 24 h or 48 h prior to TGFβ treatment (Supplementary Fig. 2f). TGFβ-dependent generalized chromatin opening also occurred in other epithelial cell lines, such as the human MCF7 and RPE1 (Supplementary Fig. 2g), although with much slower kinetics, which is consistent with the slower EMT process in these cell lines\(^{23,39}\). Critically, knockdown of SMAD4 by either of two different siRNAs abolished the increase of chromatin accessibility (Fig. 1k and Supplementary Fig. 2h). Furthermore, increased accessibility after TGFβ treatment was strongly impaired not only by the ALK4/5/7 TGFβ receptors inhibitor SB431542, but also by the TGFβ–activated kinase 1 (TAK1) inhibitor TAKinib\(^{34}\) or the MEK inhibitor U0126, which prevents ERK activation (Fig. 1l). Thus, both canonical and non-canonical TGFβ-related signaling pathways are required for the widespread chromatin opening observed in TGFβ-treated NMuMG cells.
canonical TGFβ pathways are required for widespread chromatin opening.

The putative enhancers were first classified into four categories depending on differential accessibility measured by changes in the average ATAC-seq signal provoked by TGFβ after 2 h (Fig. 2a). TGFβ treatment led to increased accessibility in 75.0% and 79.7% of enhancers at 2 h and 12 h, respectively (linear FC ≥ 1.5-fold). Proportional DNaseI accessibility gain was also observed in the same four categories (Supplementary Fig. 1e). Widespread increases of chromatin accessibility was even more evident at the early timepoint of 10 min, as analyzed by ATAC-seq (88.67% of the peaks had FC ≥ 1.5) (Fig. 2b), confirming the ATAC-seq kinetics.

We next investigated the role of TFs in chromatin opening by genomic TF footprint analysis using ATAC-seq data28, which detects actual TF occupancy to a single-nucleotide resolution, based on the steric hindrance between TF complexes and the Tn5 transposase used for ATAC experiments. AP-1 footprints were strongly overrepresented in most enhancers (Fig. 2c, p-values in Supplementary Fig. 3a). AP-1 is a TF comprising dimers of proteins belonging to the JUN, FOS, ATF, MAF and CREB subfamilies, with important roles in differentiation, apoptosis, and proliferation35. Notably, AP-1, SMAD4 and SMAD3 footprints were differentially enriched in enhancers that displayed a strong gain of accessibility (FC ≥ 4) 2 h and 12 h after TGFβ, but not at the 10 min timepoint, which instead were enriched for...
footprints related to ETS transcription factors (e.g., ELK4, ETV, ELF1, and GABPA) (Fig. 2c–f and Supplementary Fig. 3a, b). Indeed, at the 2 h and 12 h timepoints of TGFβ, SMAD footprints were negatively enriched (underrepresented) in enhancers with little or no change of accessibility (FC < 1.5). These results were confirmed by analyzing recently published ChIP-seq data of SMAD2/3 in NMuMG cells, at 1.5 h after TGFβ-treatment (Supplementary Fig. 3c, d). Many enhancers that strongly gained accessibility at 2 h and 12 h after TGFβ were co-occupied by SMAD2/3 and AP-1, and these displayed about 25% more accessibility than enhancers containing only one of the two factors (Supplementary Fig. 3e, f). Enhancers with little or no change of accessibility (FC < 1.5) at 2 h and 12 h after TGFβ were notably enriched in CTFCF and CTFCFL1 sites (Fig. 2c, d, f and Supplementary Fig. 3a, b), suggesting that at least part of these accessible regions have a structural function.

Both Jun and Fos, two classical AP-1 factors, were induced by TGFβ after 2 h, but not after 10 min of treatment, as shown by western blotting (Fig. 2g). Furthermore, RNA-seq transcriptomic analysis confirmed that several genes encoding components of AP-1 dimers, including Jun, Fos, JunB, Jund, Fosl1, Fosl2, Idp2 and Atfβ, were also significantly upregulated by 2 h and 12 h of TGFβ treatment (Fig. 2h and Supplementary Fig. 4a, Supplementary Data 1). Thus, TGFβ causes a fast and widespread chromatin opening, which depends on both the canonical and the non-canonical TGFβ pathways and which correlates to increased occupancy of AP-1 and SMAD TFs at enhancers 2 h and 12 h, but not 10 min, after TGFβ addition. In the next sections we will concentrate in the chromatin changes that occurs 2 h and 12 h after TGFβ.

Identification and characterization of TGFβ-regulated enhancers. Enhancers are often classified based on their histone post-transcriptional modifications. Active enhancers contain histone H3 acetylated at lysine K27 (H3K27ac), histone H3 monomethylated at lysine 4 (H3K4me1)37–39, and RNA polymerase II which often synthesizes enhancer RNAs (eRNAs)40. Primed enhancers contain only H3K4me1 and present a stable pre-activated situation. Finally, latent enhancers are genomic regions that lack enhancer-associated marks (i.e., H3K4me1 and H3K27ac) but that acquire these features upon induction by cellular signaling pathways17. We determined H3K4me1 and H3K27ac modifications genome wide by ChIP-seq in NMuMG cells, under the same conditions and timepoints described for Fig. 1a, and then performed chromatin-RNA sequencing (ChromRNA-seq), which gives an enriched fraction of nascent RNAs that has been previously shown to be enriched in eRNAs41,42. H3K4me3, a typical active promoter mark, but also present in some active enhancers43,44 was also determined. The presence or not of histone marks in the 29,071 putative enhancers determined by ATAC-seq allowed us to identify 4576 primed enhancers (only H3K4me1), 21,794 active enhancers (H3K4me1 and H3K27ac), and 2697 regions with neither H3K4me1 nor H3K27ac (likely regions with levels of marks below our threshold, which we have considered as undefined), prior to TGFβ stimulation in NMuMG cells (see Supplementary Data 2). Since the level of H3K27ac is an excellent indicator of enhancer activation or repression17,45,46, we used this mark to classify the enhancers upon TGFβ treatment. We found the following categories (Fig. 3aa–n and Supplementary Fig. 5): early-activated (n = 2166), late-activated (n = 2242), early-repressed (n = 1059) late-repressed (n = 2654) and TGFβ-independent enhancers (n = 20,820). While we also observed transient-activated (n = 109) and transient-repressed (n = 21) enhancers, we did not analyze these categories further due to too small number of components for statistically significant conclusions. Thus, TGFβ treatment alters the epigenetic marks of about 28% (8251) of enhancers. Of the activated enhancers, 2.3% (102) were latent, 10.5% (462) were primed, and 87.2% (3844) were active (which would imply that H3K27Ac was present and increased upon TGFβ treatment). As expected, 100% of the TGFβ-repressed enhancers were active before TGFβ treatment. Supplementary Data 2 contains information about the status of all identified enhancers from NMuMG cells, before and after TGFβ treatment. The levels of H3K4me1 and H3K4me3 marks followed a pattern similar to H3K27ac, although the changes were smaller and delayed. For example, H3K4me3 levels decreased (Log2FC ≤ −1.5) in only 1.5% of the early-repressed enhancers after 2 h of treatment but in 70% of them after 12 h (Fig. 3a and Supplementary Fig. 5), indicating that enhancer decommissioning follows enhancer repression with a slower kinetics.

ChromRNA-seq transcripts associated with a 400 bp window surrounding the ATAC-seq peak (summit ± 200 bp) of intergenic enhancers were considered as eRNA (see examples in Supplementary Fig. 6a). No correlation between the H3K27ac and the eRNAs levels prior to TGFβ treatment was observed (Supplementary Fig. 6b). However, following TGFβ treatment, there was a strong correlation between the FC of H3K27ac and that of eRNA (Supplementary Fig. 6c, d), confirming that change of H3K27ac signal is a good indicator of enhancer activity. Consistently, expression of eRNAs increased in early- and late-activated enhancers and decreased in early- and late-repressed enhancers (Fig. 3b).

We investigated the relationship between enhancer activation/repression (via changes in H3K27ac levels) and changes in chromatin accessibility (via ATAC-seq signal changes). While most of the enhancers of all categories displayed a TGFβ-dependent increase of accessibility (Fig. 3c), those enhancers with a larger increase of accessibility (FC ≥ 4) also displayed a more
pronounced increase of H3K27 acetylation (Fig. 3d). Examples of increased chromatin accessibility at activated, repressed or TGFβ-independent enhancers are shown in Fig. 3e–g. Thus, widespread chromatin opening occurs irrespective of enhancer activation or repression, with a positive correlation between strong activation and strong increase of accessibility.

Next, we computed enrichment of genomic TF footprints at each category of regulated enhancers. Activated enhancers were enriched in SMAD3, SMAD4 and AP-1 family footprints (Fig. 3h). A strong SMAD2/3 occupancy at early-activated enhancers (Fig. 3i and Supplementary Fig. 7) was also observed by ChIP-seq. In contrast, repressed enhancers were positively
enriched in TEAD, HNF1A and HNF1B footprints, and depleted of AP-1 family footprints (Fig. 3h). Notably, we found no enrichment of SMAD3 or SMAD4 footprints at repressed enhancers; however, SMAD2/3 binding was detected by ChIP-seq at late-repressed enhancers, suggesting that SMAD2/3 requires interaction with other TFs to bind these repressed enhancers (Fig. 3i and Supplementary Fig. 7b, c).

Most TGFβ-dependent genes are in the TGFβ-regulated enhancers neighborhood. We next investigated the location of TGFβ-regulated enhancers with respect to the TGFβ-regulated genes (as determined by RNA-seq). We first labeled genes surrounding a regulated enhancer as E_g (Enhancer gene) ±n, whereby n = 1, 2, 3 or 4, according to gene’s upstream (−) or downstream (+) position and chromosome order with respect to the enhancer. For every TGFβ-regulated enhancer, changes of mRNA levels of the neighboring genes (E_g ±1 to E_g ±4) were plotted. Genes surrounding an enhancer showed a strong tendency to be regulated in the same way as the enhancer. This effect was not observed when enhancer order positions were randomized (Fig. 4a and Supplementary Fig. 8a). Specifically, about 60% of upregulated and 40% of downregulated TGFβ-dependent genes are in a neighborhood of ±3 genes from an enhancer regulated in the same way (p = 7.63 × 10−6 and 7.05 × 10−14, respectively). A significant regulatory effect of the enhancer on the neighboring genes was observed up to about 150 kb upstream or downstream of the enhancers (Fig. 4b and Supplementary Fig. 8b). Approximately 79% of upregulated and 56% of downregulated genes are in a neighborhood of 150 kb from an enhancer regulated in the same way (p = 9.6 × 10−31 and 5.5 × 10−19, respectively). Importantly, a strong linear relationship was observed between changes of the enhancer H3K27Ac levels and the change of gene expression of the enhancer’s closest gene. This correlation is maintained with the second, third and fourth genes closer to the enhancer (Fig. 4c), albeit with a decreasing regression line slope, indicating that, on average, the enhancer effect is greater at proximal genes and decreases with the distance and/or the presence of genes in between. As enhancers can regulate genes at large distances by looping mechanisms, we wondered whether they operate in a continuous manner or skip over interposed genes. Indeed, enhancers create a continuous influence over their neighborhood: when the second-closest gene (up- or downstream, E_g±2) is robustly regulated (|log2FC| >1; adjusted p < 0.05) in the same way as the enhancer, the interposing gene (E_g-1 or E_g+1, respectively) also tends to be regulated that way (Fig. 4d and Supplementary Fig. 9).

TGFβ-regulated genes are often in co-regulated clusters. Next we analyzed the genomic distribution of genes regulated by TGFβ. First, genes robustly regulated by TGFβ treatment (adjusted p-value < 0.05 and |log2FC| >1) were divided into six gene categories: early-upregulated (411), early-downregulated (194), late-upregulated (542), late-downregulated (1129), transient-upregulated (49) and transient-downregulated (36) (Supplementary Fig. 4b and Supplementary Data 1). Altogether, TGFβ promoted robust expression changes in 2361 genes, affecting about 11% of protein-coding genes. As transient categories presented a small number of genes, they were not considered for further analysis. We frequently found several contiguous genes regulated in the same way (Fig. 5a, b). In fact, we found more pairs of co-regulated genes than randomly expected (Supplementary Fig. 10a). To verify this trend, we investigated the transcriptional behavior of genes in the vicinity of a robustly TGFβ-regulated gene by first numbering genes according to their 5′-3′ chromosomal order (1, 2, …, n), and then plotting the FC of genes n ±1 to n ±4 for every robustly TGFβ-regulated gene. Indeed, upstream and downstream neighboring genes of a robustly TGFβ-regulated gene (but not of a random gene) tend to be co-regulated (Fig. 5c and Supplementary Fig. 10b). Co-regulation decreased with the number of interposed genes, but significant co-regulation was observed up to the n ±3 position in early categories, and to the n ±2 position in late categories. Significant co-regulation was observed for up to about 100 kb upstream and downstream of a robustly TGFβ-regulated gene (Fig. 5d and Supplementary Fig. 10c, d). These data suggest the existence of domains in the genome where genes tend to be co-regulated by TGFβ. Notably, co-regulation also influenced intergenic transcription (Fig. 5e and Supplementary Fig. 10e), and the chromatin landscape (Fig. 5f and Supplementary Fig. 10f, Supplementary Fig. 11). In summary, our co-regulation analyses suggest that TGFβ regulation often operates in clusters of genes.

Characterization of TGFβ regulatory domains. The results exposed in the last two sections suggest that TGFβ-dependent enhancers create genomic regions of influence that often expand to several co-regulated genes. Analysis of co-regulation demonstrated that robustly regulated genes are often surrounded by differentially expressed genes with a lower FC. In order to identify most regions of co-regulation, we looked for regions of the genome containing at least two contiguous TGFβ-co-regulated genes, using a less stringent but still significant threshold (|log2FC| ≥0.5; p < 0.05). We found 372 of these clusters (86 early-upregulated, 66 late-upregulated, 49 early-downregulated, 173 late-downregulated) (Supplementary Data 3), which we termed TGFβ regulatory domains (TRDs). TRDs have an average size of 115 kb (with a range from 1 kb to 1.1 Mb) and an average of 2.2 genes (with a range from 2 to 7). About 80% of the upregulated and 53% of the downregulated TRDs contained at least one co-regulated enhancer or had one in its vicinity (±50 kb) (p = 3.3 × 10−19 and p = 7.7 × 10−10, respectively). Meta-analysis of the TRDs demonstrated coordinated regulation of gene mRNA,
intergenic RNA and H3K27Ac levels as well as a sharp difference of the TRD with respect to the surrounding regions (Fig. 5g). Similar gene ontology (GO) categories were found both in TRD and non-TRD TGFβ−regulated genes (Supplementary Fig. 12a). TRDs tend to be present in regions of high gene density (Supplementary Fig. 12b). Furthermore, the distance between a gene and the closest co-regulated enhancer was shorter in TRD genes than in non-TRD genes. However, there were no differences when distance between a gene and the closest anti-regulated enhancer were computed (Supplementary Fig. 12c, d). Total gene expression and TGFβ-dependent FCs were similar between TRD and non-TRD genes (Supplementary Fig. 12e, f).

Enhancers control expression of several genes within a TRD.

We next speculated that co-regulation within a TRD is created by the promiscuous activity of an enhancer. To test this hypothesis, we selected three TRDs (the Amigo1-Cyb561d1-Atxn712, Arid5a-
Kans3 and Ladi-Tmmrt2-Pkpl clusters) and inhibited enhancer activity of one enhancer inside of each cluster by CRISPR interference (CRISPRi)47 (Fig. 6a, c, e), using two distinct small guide RNAs (sgRNAs) for each enhancer (see Supplementary Fig. 13, for sgRNA targeting). Indeed, enhancer inactivation impaired TGFβ-dependent upregulation of expression of all genes within the TRD (Fig. 6b, d, f). As controls, we showed that neither the TGFβ-non regulated gene Gna13 (contiguous to Amigo1-Cybb611-Atnx712) nor the TGFβ-repressed gene Neurll3 (contiguous to Arid6a-Kans3) were affected by enhancer inactivation (Supplementary Fig. 14). In summary, our data demonstrate that a single enhancer can control the expression of several co-regulated linked genes within a TRD.

Relationship between TRD and chromatin 3D organization. Numerous data indicate that enhancers interact with promoters through a loop mechanism48. By analyzing Hi-C data from TGFβ treated NMuMG cells49, we detected a higher frequency of contacts between TRD enhancers with promoters of the same TRD than with other close promoters (Supplementary Fig. 15a). However, only 0.9% of TRDs correspond to Topologically Associating Domains (TADs) (which are characterized by preferential self-interaction in Hi-C experiments50). Further, a strong directionality index50 change was found at TAD borders (as expected) but not at TRD borders (Supplementary Fig. 15b). The average size of TADs (1 Mb) does not correspond to the average size of TRDs (115 kb) either. Finally, while TRDs are defined by gene co-regulation, most TADs contain genes differently affected by TGFβ (Fig. 7a–d). These data indicate that TRDs are not TADs.

However, we also found a number of TADs with all their expressed genes affected by TGFβ in the same way (93 at 2 h and 70 at 12 h; only TADs with 3 expressed genes were considered) (see example in Fig. 7d). To determine whether this reflected an statistically significant tendency of coordinated regulation within TADs, we calculated the percentage of genes per TAD with a positive or negative FC after TGFβ treatment51. We observed more TADs than randomly expected with all their genes activated (p = 7.7 × 10−6 for 2 h and 1.7 × 10−5 for 12 h) or repressed (p = 1.7 × 10−28 for 2 h and 1.2 × 10−7 for 12 h) by TGFβ, while we observed fewer TADs than randomly expected with around 50% of genes activated or repressed (p = 5.9 × 10−15 for 2 h and 2.5 × 10−7 for 12 h) (Fig. 7e and Supplementary Fig. 15c, Supplementary Data 5). Taken together, these results suggest that while TRDs are distinct from TADs, TAD organization influences co-regulation of the genes by TGFβ.

Next, we investigated the role of TAD boundaries on gene-gene and enhancer-gene co-regulation. We observed a clear correlation between TGFβ-dependent change of mRNA levels of every two contiguous neighboring genes that disappeared when gene positions were randomized (Fig. 7f and Supplementary Fig. 15d). Notably, contiguous genes separated by a TAD boundary had little or no co-regulation (Fig. 7f and Supplementary Fig. 15d), indicating that TAD boundaries are critical for suppressing co-regulation between contiguous genes. Similarly, the strong correlation that we found between H3K27ac and gene expression changes of close enhancer-gene pairs (Fig. 4c) was lost when enhancer-gene pairs separated by a TAD border were analyzed (Fig. 7g and Supplementary Fig. 15e). In contrast, if TAD borders were randomized, the linear correlation was recovered, demonstrating that, as expected, TAD borders insulate TGFβ-regulated enhancers from TGFβ-regulated genes.

Finally, we have computed the occupancy of CTCF at TAD and TRD borders by using mouse mammary ChIP-seq data and our CTCF ATAC-seq footprints. Both methods clearly demonstrated high and very significant density of CTCF at TAD borders (Supplementary Fig. 16), as expected. However, although some increase was observed at TRD borders using ChIP-seq data, it was not significant with respect to the background, suggesting that TRD limits are mostly not determined by CTCF.

Discussion
In this study, we reveal that TGFβ causes a fast and widespread increase of accessibility at enhancers and TSS, using both canonical and non-canonical TGFβ pathways. We also describe the enhancer repertoire of NMuMG cells and characterize its regulation by TGFβ, suggesting a role of AP-1 together with SMAD proteins in TGFβ—dependent enhancer regulation. We extensively investigated the genomic organization of TGFβ-regulated enhancers and genes. Most TGFβ-regulated genes are close to TGFβ-regulated enhancers. Often, TGFβ-regulated genes are grouped into what we term TGFβ regulatory domains (TRDs), which differ from TADs but are constrained by TAD boundaries. Enhancers display a promiscuous activity within TRDs, with one enhancer able to regulate several genes.

Signal-driven TFs (SDTFs) can activate specific enhancers, mostly among the repertoire of cell-type-specific accessible enhancers,6,10,15 but, although de novo activation from latent enhancers has been also reported16,17. Indeed, we show that most TGFβ-regulated enhancers were already active or primed before TGFβ treatment. Unexpectedly, however, we also found that shortly after addition of TGFβ, a massive chromatin opening occurs at most active and primed enhancers of the genome and in a number of latent enhancers, irrespectively of whether the enhancer will be activated, repressed, or not regulated. This result clearly establishes and separates two different regulatory events:
Fig. 4 Most TGFβ-dependent genes are in the TGFβ-regulated enhancers neighborhood. a Change of mRNA levels (RNA-seq signal of 2 h TGFβ versus vehicle) for genes upstream or downstream of an early-activated (left) or early-repressed (right) enhancer. Plots for late-enhancer categories are shown in Supplementary Fig. 8b. b Change of mRNA levels (RNA-seq signal of 2 h TGFβ versus vehicle) of genes located at the indicated distance (kb) upstream or downstream of an early-activated (top left) or early-repressed (top right) enhancer. Mean change of gene expression is shown. c Correlation plot between changes of H3K27ac (log2 FC) versus H3K27ac (log2 FC) of the enhancer (ChIPseq signal of 2 h TGFβ versus vehicle). Data were binned into 20 intervals. Spearman correlation coefficient and slope of the regression line are shown. d Change of mRNA levels (RNA-seq signal of 2 h TGFβ versus vehicle) for genes that are located around an early-activated enhancer. Only enhancers for which the second-closest gene (marked in red) upstream (Eg-2) or downstream (Eg+2) was robustly upregulated (log2 FC > 1; adjusted p < 0.05) were considered. Gray color boxes correspond to change of mRNA levels of genes around a random enhancer. Plots for other enhancer categories are shown in Supplementary Fig. 9. a, b, c, d Statistical significance between real and random distributions were determined with the two-tailed Mann–Whitney test (*p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001). Sample size of each distribution and exact p-values are provided in Supplementary Data 5. The horizontal black line of the boxplot represents the median value, the box spans the 25th to 75th percentiles, and whiskers indicate 5th and 95th percentiles. Source data are provided as a Source Data file.

the change of chromatin accessibility and the enhancer activation or repression decision. Thus, we speculate that, for some stimuli, the cell first uses an indiscriminate strategy to make most of the active and primed enhancers accessible, prior to moving to a more specific regulation that specifies, which enhancers will be activated or repressed. However, we cannot rule out that pervasive chromatin opening is part of a rapid stress response that might not be specifically related to TGFβ. More ATAC-seq dynamic studies are required to verify whether widespread opening is a prerequisite for signal-dependent enhancer selection triggered by other stimuli.

Which factors are responsible for this widespread enhancer opening? We showed that silencing SMAD4 impaired widespread chromatin opening. Furthermore, we found that SMAD3/4...
occupancy is enriched in enhancers that strongly gained accessibility at 2 h and 12 h after TGFβ addition. However, SMAD2/3 was only present in about 25% of all enhancers that gain accessibility. In fact, AP-1 footprints were strongly enriched in most enhancers, with a strong correlation between AP-1 occupancy and the increase of enhancer accessibility at 2 h and 12 h after TGFβ addition. Direct interaction and transcriptional cooperation between SMAD proteins and components of AP-1 (such as JUN, JUNB and FOS) have been reported. Consistently, a fraction of SMAD2/3-containing enhancers that strongly gained accessibility, also contained footprints for AP-1, suggesting cooperation of these factors for enhancer opening. An important role of AP-1 in enhancer opening and BAF chromatin remodeling complex recruitment has been demonstrated in different cell types.

In addition to the canonical TGFβ pathway, inhibition of kinases of the non-canonical pathway (MEK and TAK1) abolished chromatin opening. MEK activates ERK, a well-known...
player of the TGFβ signaling. ERK may regulate chromatin opening in several ways, including AP-1 and other transcription factors regulation, phosphorylation of histone H3 at serine 10 through MKS1 and direct regulation of the chromatin and transcription machinery. TAK1 is also a component of the TGFβ non-canonical signaling pathway that acts upstream of JNK, which regulates AP-1. Therefore, our data point to both SMAD2/3/4 and AP-1 TFs as being mainly responsible for pervasive enhancer chromatin opening at 2 h and 12 h after TGFβ. However, AP-1 and SMAD2/3/4 footprints were not enriched in highly opened chromatin after 10 min. Furthermore, at this early time, Jun and Fos protein levels have not yet been induced by TGFβ. At this time ETS TF footprints were enriched in enhancers that strongly gained accessibility. As many ETS factors are direct targets of ERK, the ERK-ETS pathway might lead a first wave of accessibility, at least in a subset of enhancers. Even though about 80% of enhancers in a cell gained accessibility upon TGFβ addition, only about 28% of them presented increased or decreased H3K27ac levels (indicating activation or repression, respectively, by TGFβ). Activated enhancers were strongly enriched in SMAD3, SMAD4, and AP-1 footprints and displayed a stronger chromatin opening. In sharp contrast, repressed enhancers were not enriched in SMAD3 or SMAD4 footprints and were negatively enriched in AP-1 footprints as compared to other enhancers. However, ChIP-seq data demonstrate an important SMAD2/3 occupancy of a number of repressed enhancers. Early-repressed enhancers presented a significant enrichment in footprints of TEAD factors. In human embryonic stem cells, a complex of TEAD factors, SMAD2 and SMAD4 has been shown to repress mesodermal genes through binding and repression of enhancers. Consistent with a role of TEAD factors in TGFβ-dependent enhancer repression also in mammary epithelium, we found upregulation of several TEAD factors (and especially of TEAD2) at 2 h and 12 h of TGFβ treatment (1.6- and 4-fold upregulation, respectively; Supplementary Data 1). SNAI1 is one of the master regulators of the EMT process and the Snail gene is strongly upregulated by the TGFβ treatment (see Supplementary Data 1). Notably, we did not find SNAI1 motifs or footprints in the enhancers identified in this work. Several works have shown binding of SNAI1 to promoters, suggesting that this factor is more likely associated with promoters rather than with enhancers.

Enhancers can regulate genes from large distances through genomic loops. Our data indicate, however, that most genes robustly regulated by TGFβ (79% of the upregulated, and 56% of the downregulated) were in the close vicinity of an enhancer that is also regulated by TGFβ in the same direction. This tendency for parallel regulation often extended to several genes around the enhancer in a continuous manner (e.g., jumping of interposed genes is not frequent), indicating that enhancers generate regions of gene co-regulation. Indeed, we found 372 clusters of co-regulated genes, which we termed TRDs. Most TRDs contain at least one enhancer within their limits or in their close proximity, and their average size is about 115 kb. Previous evidences of regulatory domains somewhat similar to TRDs have been reported. For instance, after analyzing the reporter activity of hundreds of transposon integrations in the mouse genome, Symmons and co-workers identified 46 regions of the genome (median size of 359 kb) at which adjacent integrations presented similar expression patterns, suggesting that enhancers act in a largely undiscriminating manner within these regions. Based on cohesins ChIA-PET data, Young and collaborators created the concept of insulated neighborhoods as specific DNA-loop structures, including one enhancer and one or two genes regulated by that enhancer. Insulated neighborhoods promote enhancer-promoter interactions, are delimited by CTCF-binding sites and are within TADs but are different from TADs. Although some TRD characteristics are consistent with those of insulated neighborhoods, we did not observe a high density of CTCF at TRD borders. Within TRDs, not all genes are regulated with the same intensity by the enhancer. While we found a strong dependence of enhancer-gene proximity (Fig. 4c), there are also multiple cases in which genes at positions ±2 with respect to the enhancer were regulated by TGFβ more strongly than the interposed gene at ±1, indicating that the strength of the enhancer-promoter influence depends not only on the distance and gene order with respect to the enhancer, but also on other promoter-enhancer specificity factors.

Clustering of co-regulated or co-expressed genes has been previously reported, and enhancers have been proposed as probable regulatory elements responsible for co-regulation. However, to our knowledge, functional experiments that demonstrate this statement have not been reported. We show that inactivation of three different enhancers from three TRDs impairs TGFβ-dependent regulation of the two to three genes in the neighborhood of the enhancers, demonstrating that enhancers have a more unrestrained activity than previously anticipated. We favor the view that, at least a subset of TGFβ-dependent enhancers display a rather promiscuous preference for surrounding genes. To which extent TADs control transcription is still open to discussion. Our data indicate that TRDs differ from TADs, and that TADs are not units of co-regulation. Indeed, most TADs contain genes that are differently affected by TGFβ. However, our analysis of the frequency of co-regulated genes in TADs showed a
small but significant tendency of the genes of a TAD to be co-regulated, as also reported by others. It is well-established that TAD borders are enriched in CTCF-binding sites and have enhancer-promoter insulation activity. We confirmed in our system that gene co-regulation and enhancer-gene regulation drastically decrease across TAD borders. Extending the concept to other regulatory domains (RDs), we favor a model in which RDs are the building block of co-regulation. However, the insulating activity of TAD borders obligates most RDs to reside within single TADs, resulting in a bias towards co-regulation of the genes within TADs (see model in Fig. 7h and real examples in Fig. 7a–c). As RDs have 2 to 3 genes on average, 100% co-
Fig. 6 CRISPR-mediated enhancer inactivation demonstrates promiscuous activity of enhancers on neighboring genes. a, c, d Screenshot of H3K27ac, H3K4me1, H3K4me3, RNA-seq, and ATAC-seq tracks at three different genomic regions. Position of a TGFβ-regulated enhancer is shown. For the inactivation of the enhancers by CRISPR, two different sgRNAs (pink) that target each enhancer were used. b, f Determination of mRNA levels of the indicated genes by RT-qPCR analysis. Values were normalized to the Gapdh mRNA level. Values are average ± SEM of at least six determinations from three independent experiments. Statistical significance of the values with respect to the same timepoint of the negative control (a non-targeting sgRNA, sgGal4) were determined with the two-tailed Mann-Whitney non-parametric test. *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001. Sample size (n) of each distribution and exact p-values are provided in Supplementary Data 5. Source data are provided as a Source Data file.

regulation of the genes of a TAD is mostly observed in small TADs that only harbor a single TRD. We think that this view can reconcile opposing results describing fundamental or marginal roles of TADs in gene expression.66

In summary, our work provides a valuable catalog of enhancers and their regulation by TGFβ in the model cell line NMuMG. In addition, our results also shed light about fundamental aspects of genome organization and enhancer biology.

Methods

Cell culture and treatments. Normal murine mammary gland NMuMG cells (provided by José Antonio Pintor-Toro, CABIMER, in 2014) were cultured in Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum (FBS) and 10 μg/ml insulin (complete medium), until reaching 70–80% confluency. HEK293 cells were grown in Dulbecco’s modified Eagle medium (DMEM) containing 10% fetal bovine serum (FBS). Human epithelial MCF7 cells were cultured in RPMI medium containing 10% FBS and supplemented with L-glutamine 2 mM. RPE1 cells were grown in DMEM-F12 medium containing 10% FBS and supplemented with L-glutamine 2 mM. All the cell culture media included penicillin (100 U/ml) and streptomycin (100 μg/ml). TGFβ treatments were performed after 6 h of serum starvation, except when indicated. All serum starvation treatments of the NMuMG cells also include deprivation of insulin. In some experiments, TGFβ treatments were performed in normal complete medium (Supplementary Fig. 2b–e) or after 24 or 48 h of serum and insulin starvation (Supplementary Fig. 2f).

For TGFβ treatments, 5 ng/ml TGFβ1 diluted in 4 mM HC1, 1 mg/ml BSA (240-B, R&D Systems) or 4 mM HCl 1 mg/ml BSA (vehicle, as control), was added to the medium for the indicated time.

ATAC-seq experiments. For ATAC-seq experiments, cells were grown on coverslips and processed according to ref. 28. For kinetic analyses, NMuMG cells were changed to serum- and insulin-free medium for 6 h and then were treated with vehicle or TGFβ for 5, 10 min, 1, 2, and 12 h before fixation with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 min. For the study of the canonical pathway, cells were transfected with siCT or siDSMAD4 (IDT, Leuven, Belgium) (previously used in ref. 27), using Lipofectamine RNAiMAX (#13778-150, Thermo Fisher) according to manufacturer instructions. After 48 h, cells were serum- and insulin-depleted for 6 h and treated with vehicle or TGFβ for 10 min before fixation. For the study of the non-canonical pathway, cells were first serum- and insulin-depleted for 4 h. Then, cells were treated with DMSO or SB431542 (S4317, Sigma-Aldrich), U0126 (S1102, Selleckchem) and TAK1inB (SML2126, Sigma-Aldrich) inhibitors at 10 μM for 2 additional hours and finally cells were treated with vehicle or TGFβ for 10 min. Other ATAC-seq experiments were performed under the conditions detailed in figure legends. After fixation, cells were permeabilized with lysis buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.01% Igepal CA-630) for 10 min at room temperature. Next, coverslips were rinsed in PBS twice and put in a humidity chamber box. After 30 min, the transposase mixture solution (25 μl 2x TD buffer, final concentration of 100 nM Tn5-ATTO-590N, adding dH2O up to 50 μl) was added on the coverslips and incubated for 30 min at 37 °C. As a negative control 50 μM EDTA was added to the Tn5-ATTO590N reaction buffer. After the transposase reaction, coverslips were washed with PBS containing 0.1% sodium dodecyl sulfate (SDS) and 50 μM EDTA for 15 min three times at 35 °C. Then, slides were stained with DAPI for 5 min and mounted using Vectashield for imaging.

DNAseq. DNAseq 1 hypersensitivity mapping was performed according to ref. 98 with brief modifications. NMuMG cells were either vehicle or TGFβ-treated for 2 h. The cells were trypsinized and pelleted before washing and resuspension in buffer A (15 mM Tris–Cl (pH 8.0), 15 mM NaCl, 60 mM KCI, 1 mM EDTA (pH 8.0), 0.5 mM EGTA (pH 8.0), 0.5 mM spermidine, 0.15 mM spermine). Nuclei were extracted by adding buffer A containing NP-40. The nuclei were washed with buffer A and resuspended in pre-warmed lysis buffer at a concentration of 5 × 10^6/ml and then digested with 75 units of DNAse I (Roche) for 5 min at 37 °C. The reactions were terminated by the addition of an equal volume of stop buffer and incubated at 53 °C. After 15 min, proteinase K (final concentration of 20 μg/ml) was added to each digestion reaction and incubated for 16 h at 55 °C. DNA was extracted by careful phenol-chloroform purification using phase-lock gel. DNA fragments of 50–300 bp were sequenced using low-melting agarose gel and was purified using GenElute Gel Extraction Kit (Sigma-Aldrich). DNA was again purified using MinElute PCR Purification kit and samples were sequenced using NextSeq500 system (Illumina) with 75 bp paired-end reads at the Genomic Unit of CABIMER (Sevilla, Spain).

RNA sequencing. Total RNA from NMuMG cells, treated with either vehicle or TGFβ for 2 or 12 h, was extracted using the RNeasy kit (74106, QIAGEN). Then, libraries were prepared with the TruSeq Stranded TOTAL RNA kit (Illumina). The sequencing was performed with a HiSeq 2000 system (Illumina) with 30 bp single-end reads at the Genomics Core Facility (EMBL Heidelberg).

ChromRNA sequencing. Cellular fractionation was carried out according to ref. 99, based on the protocol described in ref. 100. Cell pellets were resuspended in 400 μl cold 10 mM Tris–HCl (pH 7.4) buffer (40 μg/ml Proteinase K) and incubated on ice for 5 min. The lysates were layered onto 1 ml cold sucrose buffer (10 mM Tris pH 7.5, 150 mM NaCl, 24% sucrose w/v), and centrifuged in microtube at 3500 g for 10 min. The nuclear pellets were gently resuspended into 250 μl cold glycero buffer (20 mM Tris pH 7.9, 75 mM Na3C, 0.5 mM EDTA, 50% glycero). An additional 250 μl of cold nuclei buffer (20 mM HEPES pH 7.6, 7.5 mM MgCl2, 0.2 mM EDTA, 0.3 M NaCl, 1 M urea, 1% NP-40, 1 mM DTT) was added to the samples, followed by a pulsed vortexing and incubation on ice for 2 min. Samples were then spun in microtube for 2 min at 13,000 g. Fifty microliters of cold PBS was added to the remaining chromatin pellet, and gently resuspended by pipetting down over the pellet, followed by brief incubation on ice. Now a sample was extracted using TRIZOL and treated with TURBO DNA-free kit (AM1907, Invitrogen) for DNA removal according to manufacturer’s instructions. Library preparation and sequencing was as described in the RNA sequencing section.

ChIP sequencing. ChIP assays were performed according to ref. 101 with some modifications. Briefly, NMuMG cells treated with either vehicle for 2 or TGFβ for 2 or 12 h were crosslinked in 1% formaldehyde for 10 min at room temperature followed by addition of glycine (125 mM final concentration) for 5 min. Nuclei were isolated using lysis buffer 1 (5 mM P Pipes pH 8, 85 mM KCI, 0.5% NP-40, and complete protease inhibitor cocktail (Roche)) and were lysed using lysis buffer 2 (1% SDS, 10 mM EDTA, 50 mM Tris–HCl pH 8.1, and complete protease inhibitor cocktail (Roche)). Chromatin was sheared into an average size of 500 bp by eight pulses of 30 s (30 s pause between pulses) at 4 °C in the water bath sonicator (Bioruptor (Diagenode, Liège, Belgium). Thirty micrograms of chromatin were digested 1:10 10 μl IP buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris–HCl pH 8.1, 167 mM NaCl, 1% sodium deoxycholate) and incubated overnight at 4 °C in rotation with 3 g of the respective antibodies: anti-H3K27ac (Z0474, Abcam), anti-H3K4me1 (ab8895, Abcam), anti-H3K4me3 (ab85880, Abcam). Immunoprecipitations were incubated for 2 h at 4 °C in rotation with protein A Dynabeads (Invitrogen) and then washed with wash buffer 1 (0.1% SDS, 1% Triton X-100, 20 mM EDTA, 20 mM Tris–HCl pH 8.1, and 150 mM NaCl, 1% sodium deoxycholate), wash buffer 2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl pH 8.1, 500 mM NaCl, 1% sodium deoxycholate), wash buffer 3 (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris–HCl pH 8.1), and twice with TE1x buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA). The complexes were eluted twice from the beads with elution buffer (1% SDS in TE1x buffer) by incubating 10 min at 65 °C. The eluates and the inputs were incubated overnight at 65 °C for de-crosslinking and treated with Proteinase K for 1 h at 37 °C. Chipped DNA was purified with phenol-chloroform extraction followed by ethanol precipitation and resuspended in miliQ water. The sequencing was
Fig. 7 TRDs Differ from TADs but are Constrained by TAD Borders. a–d Changes of mRNA levels (2 h TGFβ versus vehicle) of genes contained at the indicated TADs. Co-regulated genes at TRDs are shown. e Top panel: ratio of observed versus expected frequencies of TADs with distinct proportions of genes with upregulated or downregulated FC (FC > 1 or FC < 1; 2 h TGFβ versus vehicle). Values are means ± SD. Bottom: histogram of the frequencies of TADs for the observed (blue) or randomized (orange) position of genes. TADs (n = 688) were binned into ten intervals depending on the percentage of up- versus downregulated genes. Significance was determined by comparing the real value with 500 randomizations of the gene order (see Methods). Probabilities (p) of the real number considering Normal distribution are provided. **p ≤ 0.001; ***p ≤ 0.0001. Exact p-values are provided in Supplementary Data 5. Data for 12 h are given in Supplementary Fig. 15 c. f Correlation between change of expression (12 h TGFβ versus vehicle) of every pair of expressed contiguous genes (x, y) of the genome using real chromosomal order (left), random gene order (middle) or pairs of contiguous genes separated by a TAD border (right). Spearman correlation coefficient and p-values are shown. Data were binned into 20 intervals. Data for 2 h are given in Supplementary Fig. 15d. g Correlation plot between change of H3K27ac signals of enhancers (ChIP-seq signal 2 h TGFβ versus vehicle) and change of mRNA level (RNA-seq signal 2 h TGFβ versus vehicle) of their closest gene separated by real (left), or by random TAD borders (right). Data were binned into 20 intervals. Data for 12 h are given in Supplementary Fig. 15e. h Model of influence of TGFβ-regulatory domains (TRD) and other regulatory domains (RD) constrained by the insulating activity of TAD borders. Small TADs can harbor a single RD. Source data are provided as a Source Data file.
performed with the HiSeq 2000 system (Illumina) with 50 bp single-end reads at the Genomics Core Facility (EMBL Heidelberg).

ChiP qPCR. ChiP assays were performed as described above for ChiP sequencing, using anti-H3K27ac (Ab4729, Abcam) and anti-CHIRPSR-Cas9 (C1531058-100, Diagenode). Rabbit IgG (Sigma) was used as a control for non-specific interactions. Chipped DNA was purified using ChiP DNA Clean & Concentrator kit (Zymo Research), and eluted in 50 µl of elution buffer. Input was prepared with 10% of the chromatin material used for immunoprecipitation. Input material was diluted 1:5 before PCR amplification. Quantification of immunoprecipitated DNA was performed by real-time PCR (qPCR) with the Applied Biosystems 7500 FAST real-time PCR system, using Applied Biosystems Power SYBR green master mix.

Determination of mRNA by quantitative reverse transcription PCR (RT-qPCR). Total RNA was prepared by using the RNeasy Kit (Qiagen), as described in the manufacturer’s instructions, including DNase I digestion to avoid potential DNA contamination. Complementary DNA (cDNA) was generated from 3 µg of total RNA using SuperScript First Strand Synthesis System (Invitrogen). Two 2 µl of generated cDNA solution was used as a template for real-time PCR (qPCR). Gene products were quantified by qPCR with the Applied Biosystems 7500 FAST Real-Time PCR System, using Applied Biosystems Power SYBR Green Master Mix. Values were normalized to the expression of the Gapdh housekeeping gene. At least three biological replicates and three technical determinations were performed in each case. All oligonucleotide sequences used are listed in Supplementary Data 4.

Motif-binding analysis with CentriMo. Motif-binding analyses were performed with CentriMo (v0.2.0) from MEME suite against all 579 binding motifs from JASPAR vertebrates (v1.12.0) with default options. CentriMo performs Central Motif Enrichment Analysis (CMEA), which better identifies putative enhancers. Correlation between responses was calculated using multi-BannSummary and plotCorrelation from deepTools (v3.1.3) using default options. ShuSi package (v1.16.0) was also performed in full-length ATAC peaks overlapping TSS (promoters), putative enhancers and the whole genome outside ATAC-seq peaks (rest) using regionCounr(s) with same parameters as above.

Analysis of ATAC-seq using TANGO. Cofacial microscopy images were analyzed using TANGO© plugin from imageJ104. To determine nuclei area we used standard protocol for segmentation and post-filters obtaining 203 nuclei for vehicle condition and 194 for TGFβ2h from two independent experiments. To determine open-chromatin domains we used as pre-filters standard fast filters 3D and LoG 3D (Big) filter. Then 3D spot detector was used for segmentation and region adjustment and merged region was finally used as post-filters. Standard measure geometrical simple and signal quantification were performed to analyze open-chromatin domains features and signal distribution.

ATAC-seq signal quantification. Fluorescence microscopy images were analyzed using an ImageJ macro as follow: first images were sliced by color. Then nuclei were segmented on DAPI max projection previous preprocessing and size selection. ATTO390 signal was quantified in max projection images in nucleus-exclusive area after background subtraction. All experiments were median-normalized. Three independent experiments were performed. Nuclei number quantified (n) per condition including the three replicates are provided in Supplementary Data 5.
in real footprints and background sequences with FIMO (v5.0.0) using default option and p-value < 0.0001 as cutoff. Enrichment was calculated as:

\[
\text{Enr} = \frac{\text{footprints[motif]/footprints[total]}}{\text{footprints_bg[motif]/footprints_bg[total]}}
\]

where footprints[motif] is the number of footprints with a specific motif-binding sequence; footprints[total] is the number of total footprints; footprints_bg[motif] is the number of footprints with a specific motif-binding sequence in the background set of sequences; and footprints_bg[total] is the number of total footprints in the background set of sequences. p-values were calculated using fisher.test() from R stats package. Enrichment heatmap was plotted using giotto(v2.3.1) including most representative motives.

Measurement of Tn5 integrations was calculated in Bioconductor using GenomicAlignments(v1.20.1), Biostrings(v2.52.0) and soGGi(v1.16.0) packages. Briefly, all ATAC-seq fragment smaller than 100 bp (nucleosome-free regions) were extracted from TGFβ2h BAM file. Then we resized our reads to 1bp and make the shift of 4 or – 4bp depending on strand to adjust for expected shift from insertion of Tn5 transposase to produce cut-sites.

Footprinting analysis in regulated enhancers was performed as above but using as a category background all enhancers sequences that are not included in that category. TGFβ2h ATAC-seq data were used for footprinting identification in early categories of regulated enhancers, and TGFβ1h2 for late categories.

**SMAD2/3 CHIP-seq analysis.** ChiP-seq data for SMAD2/3 in NMuMG cells were downloaded from GSE121524, and aligned against mm9 mouse reference genome using align() function from Rsubread package (v1.28.1) with type = 2, TH1 = 2 and unique = TRUE parameters. Peak calling was performed using macs2 call peak using input as control file and default options. Then, peaks were filtered using –log10(p-value) > 5 and those peaks overlap with blacklisted region of the genome were removed using bedtools subtract -A options resulting in 8401 peaks. Peaks that overlap with the TSS ± 1000 bp of any transcript from mm9 UCSC KnownGene were discarded obtaining 7503 peaks, of which 5906 overlap with peaks in ChromRNA-seq and/or AP-1 footprints were performed using vecsets(v2.1.2) package. p-values were calculated using dhyper() function from stats R package. ATAC-seq signal density graphs (Supplementary Fig. 3f) were plotted for TGFβ2h using compute-Matrix reference-point from dewtoots with -a 2000 -b 2000 options in different sets of regions: AP-1 footprinted enhancers, SMAD2/3 bound enhancers and both enhancers with AP-1 footprints and SMAD2/3 binding.

**RNA-seq analysis.** Two independent biological replicates for each condition of RNA-seq data were primarily filtered using the FASTQ Toolkit v1.0.0 program. Data were aligned using subjunc function from Rsubread package, to map reads to the mm9 mouse reference genome using TH1 = 2 and unique = TRUE parameters. The downstream analysis was performed on bamfiles with duplicates removed using the samtools(v0.19) mpdmp command. FeatureCounts() function from Rsubread package was used to assign reads to UCSC mm9 KnownGenes (miRNAs were discarded from the analysis) using GTF.attributeType = "exon", GTF.attributeType - "gene_id" and strandSpecific = 2 parameters on duplicate removed bamfiles. Then differential gene expression analysis was performed using the voom/limma (v3.3.4.9) and edgeR (v3.20.9) Bioconductor packages158. Genes that were expressed at >2 counts per million counted reads in >2 replicates were analyzed, resulting in 11081 genes out of 21240 UCSC KnownGene. CalCnormFactors() function using TMM method was used to normalize samples. Differentially expressed genes with p-value<0.05 (empirical RAYES moderated t-statistics test with adjustments for multiple comparison,) were classified according to:

- Early-downregulated: log2FC(TGFβ2h_vs_veh) < -1 and not log2FC(TGFβ1h2_vs_TGFβ2h) < -1
- Early-upregulated: log2FC(TGFβ2h_vs_veh) > 1 and not log2FC(TGFβ1h2_vs_TGFβ2h) < -1
- Late-downregulated: log2FC(TGFβ2h_vs_veh) < -1 and not log2FC(TGFβ2h_vs_veh) < -1
- Late-upregulated: log2FC(TGFβ1h2_vs_veh) > 1 and not log2FC(TGFβ2h_vs_veh) > 1
- Transient-downregulated: log2FC(TGFβ2h_vs_veh) < -1 and log2FC(TGFβ1h2_vs_TGFβ2h) < -1
- Transient-upregulated: log2FC(TGFβ2h_vs_veh) > 1 and log2FC(TGFβ1h2_vs_TGFβ2h) > 1
- TGFβ-independent: expressed genes that are not included in any of the categories above.

Because of low number of genes in transient categories, they were not taken into account for downstream analyses. For MA plots (Supplementary Fig. 4a), log2FC were plotted against average expression (log2CPM) for all samples for each comparison. For boxplots of Supplementary Fig. 4b, distribution of log2CPM of the two independent replicates were plotted. For expression heatmap of Fig. 2h log2FC were plotted for selected genes using giotto.

**Histone Chip-seq analysis.** Two biological replicates of the Chip-seq data for each condition and for each histone modification (H3K27ac, H3K4me1, and H3K4me3) were primarily filtered using the FASTQ Toolkit program. Data were aligned using align function from Rsubread package, to map reads to the mm9 mouse reference genome using TH1 = 2 and unique = TRUE parameters. The downstream analysis was performed on bamfiles with duplicates replaced using the samtools rmdup command. Information about read depth can be found in the reporting summary. Shushu package (v1.16.0) was used for data visualization.

**Enhancer classification.** For creating the NMuMG cells enhancer repertoire first H3K27ac and H3K4me1 signal was measured by counting reads in full-length ATAC-seq peaks (29071 non-TSS peaks) using regionCounts() from csaw package. For H3K4me1, coordinates were increased ≥500 bp because H3K4me1 showed a broad distribution. Then, CPM and RPMK were calculated using all counted reads in ATAC-seq peaks as library size and combining replicates. Peaks whose RPMK were larger than percentile 25 of all RPMK for H3K27ac or 10 for H4K4me1 were considered candidates as having the mark. Enhancers without H3K4me1 nor H3K27ac in vehicle condition but that acquire these marks with TGFβ treatment were classified as latent. Enhancers with H3K4me1 but no H3K27ac were classified as primed and those with both marks, as active. To classify enhancers according to its feature by TGFβ, we analyze H3K27ac dynamics in vehicle. 2h and 12 h TGFβ-treated ChiP-seq data, considering both replicates independently and using a pipeline, including caw and edgeR(v3.20.9) packages. Reads were counted in full-length ATAC-seq peaks as above and normalization factors were calculated as normalisSet () function with option type = "loess" parameter. For differential binding analysis scaleOffset(), estimateDisp(), glmQLFiti() and glmRLi() functions was used. Then enrichment was classified in each condition, early-regulated, late-regulated and late-repressed, following same criteria as differentially expressed genes. For each comparison we also consider a p-value < 0.05 and average log2CPM > 0. Because of low number of enhancers in transient categories, they weren’t taking into account for downstream analyses.

**EnrichedHeatmap package was used for drawing heatmaps.** For Fig. 3a, rows were ordered according to decreasing H3K27ac CPM for TGFβ2h and all enhancers were plotted as summit ±500 bp increasing ±1500 bp from start and end. Boxplots of Supplementary Fig. 5, represents average CPM of the two replicates for each condition. Profiles screen-shot was constructed using sushu package (v1.16.0) and bedgraph constructed by bamCoverage from deepTools previous merging of replicates with samtools merge.

**ChromRNA-seq.** Two biological replicates for each condition of ChromRNA-seq data were primarily filtered using the FASTQ Toolkit program. Data were aligned using subjunc function from Rsubread package, to map reads to the mm9 mouse reference genome using TH1 = 2 and unique = TRUE parameters. The downstream analysis was performed on bamfiles with duplicates replaced using the samtools(v0.19) mpdmp command. After read counting, CPM of all enhancers in ENA file was calculated CPM of all enhancers in ENA file was calculated. Correlation between H3K27ac levels and eRNAs levels shown. For creating the NMuMG cells enhancer repertoire first H3K27ac and increase in eRNA shown in Supplementary Fig. 6c, d, first we filtered by expression, obtaining 3636 eRNA expressing enhancers. Then, differential expression analysis using caw and edgeR was computed.

**Analysis of the location of TGFβ-regulated genes with respect to TGFβ-regulated enhancers.** To study the regulation of genes in the neighborhood of TGFβ-regulated enhancers, first we plotted log2FC distribution of genes adjacent to regulated enhancers. To determine position of genes with respect to enhancers, distances between gene TSSs and enhancers ATAC-seq peaks summit were calculated. Then genes were grouped according to gene position relative to the enhancers (upstream (-) or downstream(+)) and gene order with respect to the enhancer (1st to 4th). According to this classification, genes were called: enhancer gene n, Eg π, n with π = 1, 2, 3, and 4. Non-expressed genes were not taken into account for analysis. Randomization of enhancers was performed by shifting or by random sampling. For shifting randomization, lists of numbered enhancers for each recircuclarized chromosome were constructed and random enhancers were obtained by shifting enhancers 50 positions. So, if the regulated enhancer in position n, the random enhancer is at position n = 50. For sampling randomization, all enhancers of each chromosome were randomly sampled. We obtained similar results with both randomization methods, however, we choose shifting method for considering it more conservative with genome structure. To study the dependency on enhancer-gene distance, we grouped genes located in 50 kb non-overlapping distance intervals (i.e., from summit to 50 kb, from 50 kb to 100 kb... upstream and downstream until 250 kb) from TGFβ-regulated enhancers of each enhancer category. The same was done for randomized enhancers. To study the correlation between change of activity of enhancers (measured as log2FC(H3K27ac)) and differential expression of the closest genes (log2FC), pairs of enhancer-closest
gene were ordered according to increasing log2FC(H3K27ac) and grouped into 20 bins. Then, least squares fit analysis were performed to obtain A and slope. To study if enhancer-promoter interactions, in a continuous manner, first we retrieve those enhancer regions with one of their neighborhood genes (for example, Eg1-2) robustly regulated (log2FC > 1 and adjusted p-value < 0.05). Then, log2FC distribution of the rest of genes in the neighborhood region (Eg-2, Eg-1, and Eg+1) were plotted. We only tested until Eg-2 and Eg+2 because of the small number of enhancers whose 3rd or 4th closest upstream or downstream gene is robustly regulated.

Identification of clusters of co-regulated genes. To identify clusters of co-regulated genes, first we sorted all genes according to 5′−3′ chromosomal order using gene TSS coordinates and we looked for robustly regulated pairs of genes from the same category (early-upregulated, early-downregulated, late-upregulated, or late-downregulated) that are strictly contiguous. To determine if the number of co-regulated pairs of genes was higher than expected by chance, we computed a random distribution of this score by randomly sampling a number of genes identical to the number of robustly regulated genes for each category and looking for pairs of contiguous genes. This process was repeated 5000 times, obtaining a distribution of random pairs of co-regulated genes. Probability of obtaining the real number of pairs of co-regulated genes given the random distributions was obtained using (https://keisan.casio.com/exec/system/1180573188).

Analysis of gene co-regulation. Similar to previous analysis, we numbered genes according to 5′−3′ chromosomal order using gene TSS coordinates (1, 2, …, n). Then, for every robustly TGFβ-regulated gene (log2FC > 1 and adjusted p-value < 0.05), log2FC of genes n ± k with k = 1, 2, 3, or 4 were plotted. Non-expressed genes were not taken into account for random genes lists, per chromosome (bins) upstream and downstream from a TGFβ-regulated gene, were obtained by random sampling or by shifting 50 positions, as described above, with similar results. To study the effect of distance on co-regulation, we grouped genes into non-overlapping 50 kb distance intervals from robustly regulated gene TSSs (i.e., from TSS to 50 kb, from 50 kb to 100 kb, …, until 250 kb upstream and downstream for each category) and log2FC distributions were plotted. Similar analysis was done for genes randomly modified by shifting or by sampling with similar results.

Study of the influence of co-regulation in intergenic transcription. To study the influence of co-regulation in intergenic transcription we determined the ChromRNA-seq signal in the neighborhood of TGFβ-regulated genes, previously all knownGene transcript bodies (from the 2kb TSS to avoid termination read-through and divergent promoter transcription) were removed. Then, we counted reads of 250 kb (divided into 10 kb non-overlapping bins) upstream and downstream from a TGFβ robustly regulated gene TSS. For that, we used region.counts(), winCount() with bin = TRUE and width = 10,000 options. Normalize() function was used to calculate normalization factors in libraries. Then we performed differential analysis bin to bin using estimateDisp() and glmQLFit() functions to obtain log2FC. Similar analysis was performed using random genes (as described above) instead of TGFβ-regulated genes. Finally, log2FC of random genes was subtracted from log2FC of TGFβ-regulated genes, bin to bin, to obtain log2FC over random values that were finally plotted.

Study of the influence of co-regulation in histone modifications. log2FC of H3K27ac, H3K4me3, and H3K4me1 ChIP-seq signal in the neighborhood of TGFβ-regulated genes was performed as was the previous section. In this case 5 kb upstream and downstream of the regulated gene TSS (or random gene) were removed, in order to eliminate promoter signal of the regulated gene. UCSC KnownGene transcript bodies coordinates were not removed from this analysis. Differential analysis was performed as in the previous section.

Identification of TGFβ-regulatory domains. To identify TGFβ-regulatory domains (TRD) we identified clusters of co-regulated genes, as above, using a different threshold of differentially expressed genes (log2FC > 0.5 and p-value < 0.05) to include genes that are significantly regulated with TGFβ, although not robustly affected. Metaplots were created combining late and early TRD for upregulation and downregulation and using computeMatrix scale-regions with -m very-sensitive. Then, raw and ICE-normalized matrices for TRD or random TRD coordinates were extracted using extractRegion() function. Random TRD coordinates were calculated inverting the start coordinate for each TRD and keeping the same length.

To study the effect of TAD borders in enhancer-promoter interaction, we first retrieved bins corresponding to TSS and enhancer midpoint coordinates. Then we obtained the interaction for each enhancer-promoter pair for random matrix and grouped then according to: if the interaction occurs within the TRD (enhancer and promoter are located in the same TRD) or outside the TRD at different windows (enhancer is located within the TRD and genes are located at 0−100 kb, 100−250 kb, and 250−500 kb from the TRD upstream and downstream boundaries).

Directionality Index (DI) was calculated for each bin according to ref. 30.

$$DI = \frac{(B - A) - (A + B / 2)}{E} = \frac{(A - E)^2}{E} + \frac{(B - E)^2}{E}.$$

where A is the number of reads that map from a given 10 kb bin to the upstream 250 kb, B is the number of reads that map from the same 10 kb bin to the downstream 250 kb, and E, the expected number of reads under the null hypothesis, is equal to (A + B / 2). Then, directionality index metaplots were constructed for TADs and TRDs using proportional number of bin for each structure: 100 bins for TADs (as average size of TAD is 1Mb) and 10 bins for TRDs. DI in boundaries was also plotted for TADs and TRDs using surrounding 100 bins (50 upstream and 50 downstream) and 30 bins respectively.

CTCF metaplot analysis. First, CTCF ChIP-seq data for mouse mammary gland tissue was downloaded from GSE74826 and was processed as described above for SMAD2/3 ChIP-seq data, finding 35360 CTCF peaks. Then we looked for motives in those peaks using fimo what allowed us to divide CTCF peaks according to its strand orientation: 17,800 in the forward and 17,850 in the reverse orientation. To identify CTCF footprints across the genome using our ATAC-seq data PiQ software was used. CTCF density metaplots was constructed using ChIP-seq and footprints informations in TADs and TRDs using proportional number of bin for each structure: 100 bins for TADs (as average size of TAD is 1Mb) and 10 bins for TRDs. CTCF density in boundaries was also plotted for TADs and TRDs using surrounding 50 bins (25 upstream and 25 downstream).

Statistical analysis. Statistical and graphical data analyses were performed using either Prism 5 (Graphpad) software or R package. To determine the significance between two groups, comparisons were made using two-tailed Mann–Whitney non-parametric test by using wilcox.test() function form Stats R package. For correlation, the non-parametric Spearman coefficient was calculated and significance was calculated using Fisher exact test using fisher.test() function from R. Probabilities of overlapping were calculated using the hypergeometric distribution using dhyper() function from the stats package. Other statistical tests are described above. The horizontal black line of the boxplot represents the median value, the box spans the 25th and 75th percentiles, and whiskers indicate 5th and 95th percentiles.
Data availability

Sequencing data (raw data and processed files) that support the findings of this study have been deposited in NCBI GEO with the accession code "GSE140552", SMAD2/3 ChIP-seq data was downloaded as raw data from NCBI GEO accession code "GSE121254”. HiC data was obtained from NCBI GEO accession code "GSE96033”. CTCF ChIP-seq data was downloaded as a raw data from "GSE74826”. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information file or from the corresponding author upon reasonable request. Source data are provided with this paper.

Received: 28 November 2019; Accepted: 5 November 2020; Published online: 03 December 2020

References

1. David, C. J. & Massague, J. Contextual determinants of TGFbeta action in development, immunity and cancer. Nat. Rev. Mol. Cell Biol. 19, 419–435 (2018).

2. Colak, S. & Ten Dijke, P. Targeting TGF-beta signaling in cancer. Trends Cancer 3, 56–71 (2017).

3. Derynck, R. & Buds, E. H. Specificity, versatility, and control of TGF-β family signaling. Sci. Signal. 12, eaav5183 (2019).

4. Hill, C. S. Transcriptional Control by the SMADs. Cell 147, 565–576 (2011).

5. Mullen, A. C. et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell 147, 577–589 (2011).

6. Tannahill, E. et al. Lineage specific transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell 147, 577–589 (2011).

7. Sundqvist, A. et al. JUNB governs a feed-forward network of TGFbeta signaling that aggravates breast cancer invasion. Nuclear Acids Res. 46, 1180–1195 (2018).

8. Raghunandan, A. et al. TGF-beta uses a novel mode of receptor activation to phosphorylate SMAD1/5 and induce epithelial-to-mesenchymal transition. elife 7, e31756(2018).

9. Fueyo, R. et al. Lineage specific transcription factors and epigenetic regulators mediate TGFbeta-dependent enhancer activation. Nuclear Acids Res. 46, 3351–3366 (2018).

10. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell 167, 1170–1187 (2017).

11. Rickels, R. & Shilatifard, A. Enhancer logic and mechanics in development and disease. Trends Cell Biol. 28, 608–630 (2018).

12. Shvedova, A. D., Stampfli, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 15, 272–286 (2014).

13. Plank, J. L. & Dean, A. Enhancer function: mechanistic and genome-wide insights come together. Mol. Cell 55, 5–14 (2014).

14. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 25, 2227–2241 (2011).

15. Heintz, S., Romanski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. 16, 144–154 (2015).

16. Verbuchen, T. et al. AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. Mol. Cell 68, 1067–1082 e1012 (2017).

17. Ostuni, R. et al. Latent enhancers activated by stimulation in differentiated cells. Cell 152, 157–171 (2013).

18. Nieto, M. A. Epithelial plasticity: a common theme in embryonic and cancer cells. Science 342, 1234850 (2013).

19. Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. Trends Cell Biol. 26, 1149–1157 (2015).

20. Zheng, B. et al. Dynamic H3K27ac signature identifies VEGA-stimulated endothelial enhancers and requires EP300 activity. Genome Res. 23, 917–927 (2013).

21. Malik, A. N. et al. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. Nat. Neurosci. 17, 1330–1339 (2014).

22. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451 (2013).

23. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. 20, 437–455 (2019).

24. Pascual-Reguant, L. et al. Lamin B1 mapping reveals the existence of dynamic and functional euchromatin lamin B1 domains. Nat. Commun. 9, 3420 (2018).

25. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380 (2012).

26. Le Dily, F. et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 28, 2151–2162 (2014).

27. Zhang, Y., Feng, X. H. & Derynck, R. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. Nature 394, 909–913 (1998).

28. Liberati, N. T. et al. Smads bind directly to the Jun family of AP-1 transcription factors. Proc. Natl Acad. Sci. USA 96, 4844–4849 (1999).

29. Phanstiel, D. H. et al. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. Mol. Cell 67, 1037–1048 e1036 (2017).
Maturi, V., Moren, A., Enroth, S., Heldin, C. H. & Moustakas, A. Genomewide
Baulida, J., Diaz, V. M. & Herreros, A. G. Snail1: a transcriptional factor
Ye, X. et al. Distinct EMT programs control normal mammary stem cells and

Hurst, L. D., Pal, C. & Lercher, M. J. The evolutionary dynamics of eukaryotic
gene order. Nat. Rev. Genet. 5, 299–310 (2004).

Purmann, A. et al. Genomic organization of transcriptioners in mammals:
regulation and cofunctionality. Genomics 89, 580–587 (2007).

Michalak, P. Coexpression, coregulation, and cofunctionality of neighboring
genes in eukaryotic genomes. Genomics 91, 243–248 (2008).

Despag, A. et al. Functional dissection of the Sox9-Kcnj2 locus identifies
nonessential and instructive roles of TAD architecture. Nat. Genet. 51,
1236–1237 (2019).

Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling
between genome topology and gene expression. Nat. Genet. 51, 1272–1282
(2019).

Finn, E. H. & Misteli, T. A genome disconnect. Nat. Genet. 51, 1205–1206
(2019).

Neems, D. S., Garza-Gonzaga, A. G., Smith, E. D. & Kosak, S. T. Topologically
associated domains enriched for lineage-specific gene reveal expression-
dependent nuclear topologies during myogenesis. Proc. Natl Acad. Sci. USA
113, E1691–E1700 (2016).

Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-
inactivation centre. Nature 485, 381–385 (2012).

Dohna-Schindler, O. G. et al. Dynamics of topological chromatin domains cause
pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025 (2015).

Rodriguez-Mateo, C., Torres, B., Gutierrez, G. & Pintor-Toro, J. A.
Downregulation of Lnc-Spry1 mediates TGF-beta-induced epithelial-
mesenchymal transition by transcriptional and posttranscriptional regulatory
mechanisms. Cell Death Differ. 24, 785–797 (2017).

He, H. et al. Refined Dnase-seq protocol and data analysis reveals intrinsic
bias in transcription factor footprint identification. Nat. Methods 11, 73–78
(2014).

Conrad, T. & Orom, U. A. Cellular fractionation and isolation of chromatin-
associated RNA. Methods Mol. Biol. 1468, 1–9 (2017).

Wuarrin, J. & Schiller, U. Physical isolation of nascent RNA chains transcribed
by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell Biol.
14, 7219–7225 (1994).

Subit-Rodríguez, A. & Reyes, J. C. BRG1 helps RNA polymerase II to overcome
a nucleosomal barrier during elongation, in vivo. EMBO Rep. 11,
751–757 (2010).

Ceballos-Chavez, M. et al. Control of neuronal differentiation by sumoylation
of BRF35, a subunit of the LSD1-CoREST histone demethylase complex.
Proc. Natl Acad. Sci. USA 109, 8085–8090 (2012).

Lun, A. T. & Smyth, G. K. caw: a Bioconductor package for differential
analysis of ChIP-seq data using sliding windows. Nucleic Acids Res.
44, e45 (2016).

Schneider, C. A., Rashband, W. S. & Elseceri, K. W. NIH image to Image: 25 years of
image analysis. Nat. Methods 9, 671–675 (2012).

Bailey, T. L. & Machanick, P. Inferring direct DNA binding from ChIP-seq.
Nucleic Acids Res. 40, e128 (2012).

Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-
access database of transcription factor binding profiles. Nucleic Acids Res.
44, D110–D115 (2016).

Martin-Malpartida, P. et al. Structural basis for genome wide recognition of 5-
bp GC motifs by SMAD transcription factors. Nat. Commun. 8, 2070 (2017).

Piper, J. et al. Wellington: a novel method for the accurate identification of
digital genomic footprints from Dnase-seq data. Nucleic Acids Res. 41,e201
(2013).

Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a
given motif. Bioinformatics 27, 1017–1018 (2011).

Law, C. W. et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and
edgeR. F1000Res 5, 1SCB Comm J-1408 (2016).

Servant, N. et al. HiC-Pre: an optimized and flexible pipeline for Hi-C data
processing. Genome Biol. 16, 259 (2015).

Sherwood, R. L. et al. Discovery of directional and nondirectional pioneer
transcription factors by modeling Dnase profile magnitude and shape. Nat.
Biotechnol. 32, 171–174 (2014).

Acknowledgements

We thank J. A. Pintor-Toro, F. Cortés-Ledesma, A. Lopez-Rivas, M. García-Domínguez, and M. Rodríguez-Paredes for reagents, critical reading of the manuscript, and
to continuous discussion. We thank P. Domínguez-Giménez from the CABIMER Microscopy Unit and E. Andújar and M. Pérez from the CABIMER Genomic Unit for technical
support and continuous discussion. We thank P. Domínguez-Giménez from the CABIMER Microscopy Unit and E. Andújar and M. Pérez from the CABIMER Genomic Unit for technical
support and continuous discussion. We thank P. Domínguez-Giménez from the CABIMER Microscopy Unit and E. Andújar and M. Pérez from the CABIMER Genomic Unit for technical
support and continuous discussion.
supported by the FERO Foundation, La Caixa Foundation (LCF/PR/PR12/S1070001) and Cellex Foundation.

**Author contributions**
Conceptualization and methodology, J.A.G.-M. and J.C.R.; Investigation, J.A.G.-M. and M.C.-G.; Resources, F.K., S.P.; Writing, J.C.R. and J.A.G.-M.; Supervision, project administration, and funding acquisition: J.C.R.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-19877-5.

**Correspondence** and requests for materials should be addressed to J.C.R.

**Peer review information** Nature Communications thanks Xiaomin Bao, Alvaro Rada-Iglesias and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2020