



TGF β promotes widespread enhancer chromatin opening and operates on genomic regulatory domains

Jose A. Guerrero-Martínez¹, María Ceballos-Chávez ¹, Florian Koehler^{2,3}, Sandra Peiró⁴ & Jose C. Reyes ¹✉

The Transforming Growth Factor- β (TGF β) signaling pathway controls transcription by regulating enhancer activity. How TGF β -regulated enhancers are selected and what chromatin changes are associated with TGF β -dependent enhancers regulation are still unclear. Here we report that TGF β 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 TGF β . This enhancer opening depends on both the canonical and non-canonical TGF β pathways. Most TGF β -regulated genes are located around enhancers regulated in the same way, often creating domains of several co-regulated genes that we term TGF β regulatory domains (TRD). CRISPR-mediated inactivation of enhancers within TRDs impairs TGF β -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.

¹Centro Andaluz de Biología Molecular y Medicina Regenerativa-CABIMER, Consejo Superior de Investigaciones Científicas-Universidad de Sevilla-Universidad Pablo de Olavide (CSIC-USE-UPO), Avenida Americo Vesputio 24, 41092 Seville, Spain. ²Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany. ³Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. ⁴Vall d'Hebron Institute of Oncology (VHIO), 08035 Barcelona, Spain. ✉email: jose.reyes@cabimer.es

TGF β and related ligands are believed to be the family of cytokines with the most diverse and widespread functions in metazoans¹. TGF β cytokines have crucial roles in development, proliferation, tissue homeostasis, differentiation, and immune regulation. Consequently, alterations in TGF β signaling underlie numerous diseases, including cancer^{1,2}. The TGF β canonical pathway involves ligand binding to a transmembrane tetrameric complex of dual-specificity kinase receptors, which becomes activated and can phosphorylate receptor-regulated (R-) SMAD proteins (SMAD2 and SMAD3, among others) at carboxy-terminal serine residues³. R-SMAD factors shuttle between the nucleus and the cytoplasm; upon phosphorylation, they oligomerize together with SMAD4 and accumulate in the nucleus, where they bind chromatin⁴. Chromatin immunoprecipitation (ChIP)-seq experiments demonstrated that SMAD proteins mostly bind to distant *cis*-regulatory regions, commonly called enhancers^{5–9}. Enhancers are regions for sequence-specific transcription factors (TFs) binding, which potentiate transcription of their target genes, independently of their relative distance, location or orientation relative to them^{10–13}. The human genome has more than one million enhancers; however, only a subset of these are accessible and, eventually, active in each cell-type. Chromatin can regulate enhancer function by allowing or preventing TF binding to its target motifs within an enhancer. However, the so-called pioneer TFs can bind DNA that is wrapped around a histone octamer and trigger enhancer regulation by recruiting chromatin remodeling complexes and histone modifying enzymes¹⁴. The current model proposes that most lineage-determining TFs (LDTFs) are pioneer TFs that cooperate to establish the cell-type-specific landscape of open enhancers^{5,6,10,15}. This mechanism would restrict the number of enhancers at which signal-driven TFs (SDTFs) can bind. Nonetheless, *de novo* enhancer chromatin opening by SDTFs alone, or in combination with LDTFs, have also been reported^{16,17}. Thus, how SDTFs lead to enhancer selection and activation is still an open question.

The epithelial-to-mesenchymal transition (EMT) switches cell phenotypes from epithelial to mesenchymal during development^{18–20} and has been linked to dissemination and migration of epithelial tumor cells²¹ and metastasis formation^{22–25}. TGF β is one of the most potent inducers of EMT in normal and oncogenic epithelial cells from different origins²⁶; however, little is known about the genomic repertoire of enhancers activated by TGF β , the chromatin dynamics during this process or the SMAD partners in epithelial cells.

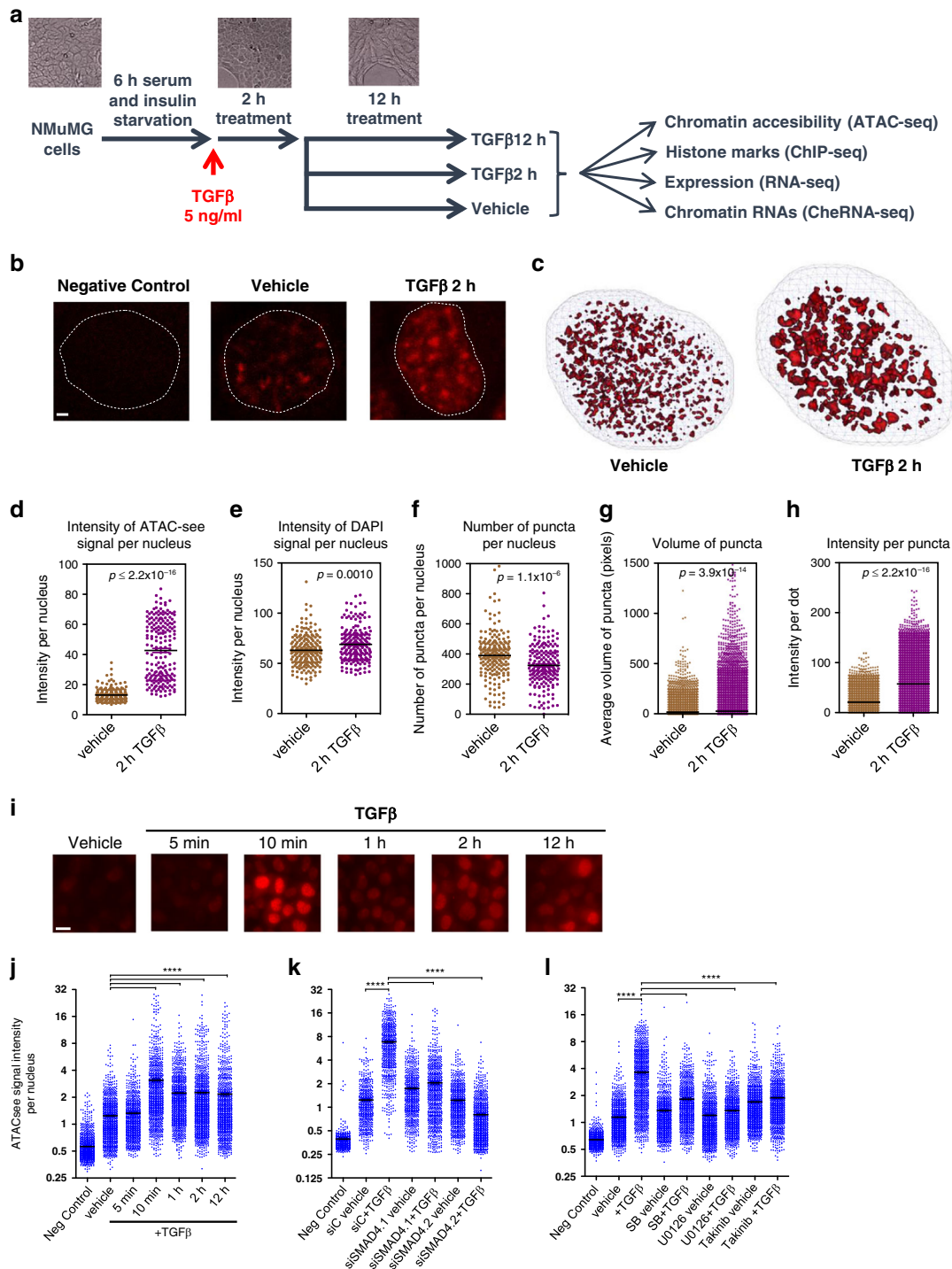
We have now addressed these outstanding questions about the genomic regulation mediated by TGF β in epithelial cells by determining and characterizing the enhancer atlas of the TGF β response in normal murine mammary gland (NMuMG) epithelial cells, a well-established model for TGF β -dependent EMT²⁷. We show that TGF β promotes a fast and widespread increase of chromatin accessibility (assayed by ATAC-seq, ATAC-see and DNaseI-seq) in most enhancers of the cell line, irrespectively of whether the enhancer will become activated or repressed. Pervasive chromatin opening depends on both the canonical and the non-canonical TGF β pathways. Activated enhancers are strongly enriched for SMAD2/3/4 and AP-1 footprints. Strikingly, analyses of the regulated genes around TGF β -regulated enhancers revealed TGF β regulatory domains that can encompass several genes and that are constrained by three-dimensional (3D) chromatin conformation. In these domains enhancer targeting is more promiscuous than previously anticipated.

Results

TGF β provokes widespread enhancer chromatin opening. To better understand the mechanism by which TGF β signaling

controls transcription in epithelial cells, we investigated early-chromatin changes provoked by TGF β treatment. TGF β 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 TGF β , NMuMG cells were depleted of serum and insulin (an additive of the NMuMG cell culture medium) for 6 h, prior to TGF β 1 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)²⁸. A total of 52,626, 100,459, and 105,334 significant peaks (IDR (irreproducible discovery rate) <0.05) were identified from vehicle, 2 h and 12 h TGF β 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, TGF β 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 DNaseI-seq (Supplementary Fig. 1c, d). Thus, 88.3% of the DNaseI-seq peaks (20,908) were coincident with ATAC peaks and 76.8% of the DNaseI-ATAC common peaks increased their signal at 2 h of TGF β treatment. Furthermore, a very good correlation was observed between DNaseI-seq and ATAC-seq fold changes (FCs).

We next imaged the transposase-accessible chromatin *in situ* using ATAC-see²⁹, 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 TGF β treatment promoted a general increase of the ATAC-see signal (Fig. 1b). Signal was visible as a general nuclear background, but also concentrated into puncta. TGF β treatment increased the total intensity of the ATAC-see signal per nuclei by about fourfold, with slightly fewer puncta per nucleus, as shown by 3D confocal microscopy image analysis using TANGO software³⁰ (Fig. 1c–f). Notably, 2 h of TGF β 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 TGF β addition (Fig. 1i, j), coinciding with the reported accumulation of SMAD2/3 and SMAD4 in the nucleus of TGF β treated cells³¹ and with the activation of the ERK pathway (Supplementary Fig. 2a). We verified that TGF β -dependent widespread chromatin opening also occurred in cells not deprived of growth factors prior to TGF β stimulation (Supplementary Fig. 2b–e) or with severe serum starvation for 24 h or 48 h prior to TGF β treatment (Supplementary Fig. 2f). TGF β -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^{32,33}. 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 TGF β treatment was strongly impaired not only by the ALK4/5/7 TGF β receptors inhibitor SB431542, but also by the TGF β -activated kinase 1 (TAK1) inhibitor Takinib³⁴ or the MEK inhibitor U0126, which prevents ERK activation (Fig. 1l). Thus, both canonical and non-



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

Fig. 1 TGF β provokes widespread enhancer chromatin opening. **a** Flow diagram depicting experimental design. Vehicle-treated cells (control) were taken after 2 h of treatment. **b** Confocal microscope images of ATAC-seq signals from NMuMG nuclei of cells cultured under the indicated conditions. Negative control shows ATAC-seq signal obtained in the presence of 50 mM EDTA, which inhibits Tn5 transposome activity. Bar, 1 μ m. Representative images out of twelve independent experiments are shown. **c** Representative images showing processing of ATAC-seq 3D signal by TANGO software in cells treated with TGF β or vehicle for 2 h. **d–h** Quantification of ATAC-seq images by using TANGO software. Scatter dot plots showing ATAC-seq signal intensity (**d**), DAPI signal intensity (**e**), and number of puncta per nucleus (**f**) from 203 (vehicle) and 194 (2 h TGF β) nuclei from two independent experiments. **g, h** Scatter dot plots showing volume of puncta expressed as pixels (**g**), and intensity of ATAC-seq signal per puncta (**h**). Number of puncta analyzed: 80,839 for vehicle and 63,001 for 2 h TGF β from 203 (vehicle) or 194 (2 h TGF β) cells, from two independent experiments. Two-tailed Mann-Whitney test p -values are provided. **i** Fluorescence microscope images of ATAC-seq signals from cells cultured under the indicated conditions. Bar, 10 μ m. Representative images out of three independent experiments are shown. **j–l** Quantification of ATAC-seq signal intensity of nuclei at the indicated times (**j**) or 10 min after TGF β or vehicle addition (**k, l**). **k** Cells were transfected 48 h before treatment with siControl (siC) or with two different siRNAs against SMAD4 (siSMAD4.1 or siSMAD4.2). **l** Cells were treated with 10 μ M of SB431542, U0126, or Takinib for 2 h before TGF β addition. Merged data from three independent experiments are shown. Statistical significance between indicated samples was determined by using the two-tailed Mann-Whitney test, **** $p \leq 0.0001$. Quantified nuclei (n) in each scatter dot plot and exact p -values are provided in Supplementary Data 5. **d–h, j–l** The horizontal black line of the scatter dot plots represents the mean value. Source data are provided as a Source Data file.

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 TGF β , 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 TGF β -treatment³⁶ (Supplementary Fig. 3c, d). Many enhancers that strongly gained accessibility at 2 h and 12 h after TGF β 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 TGF β were notably enriched in CTCF and CTCFL1 sites (Fig. 2c, d, f and Supplementary Fig. 3a, b), suggesting that at least part of these accessible regions have a structural function.

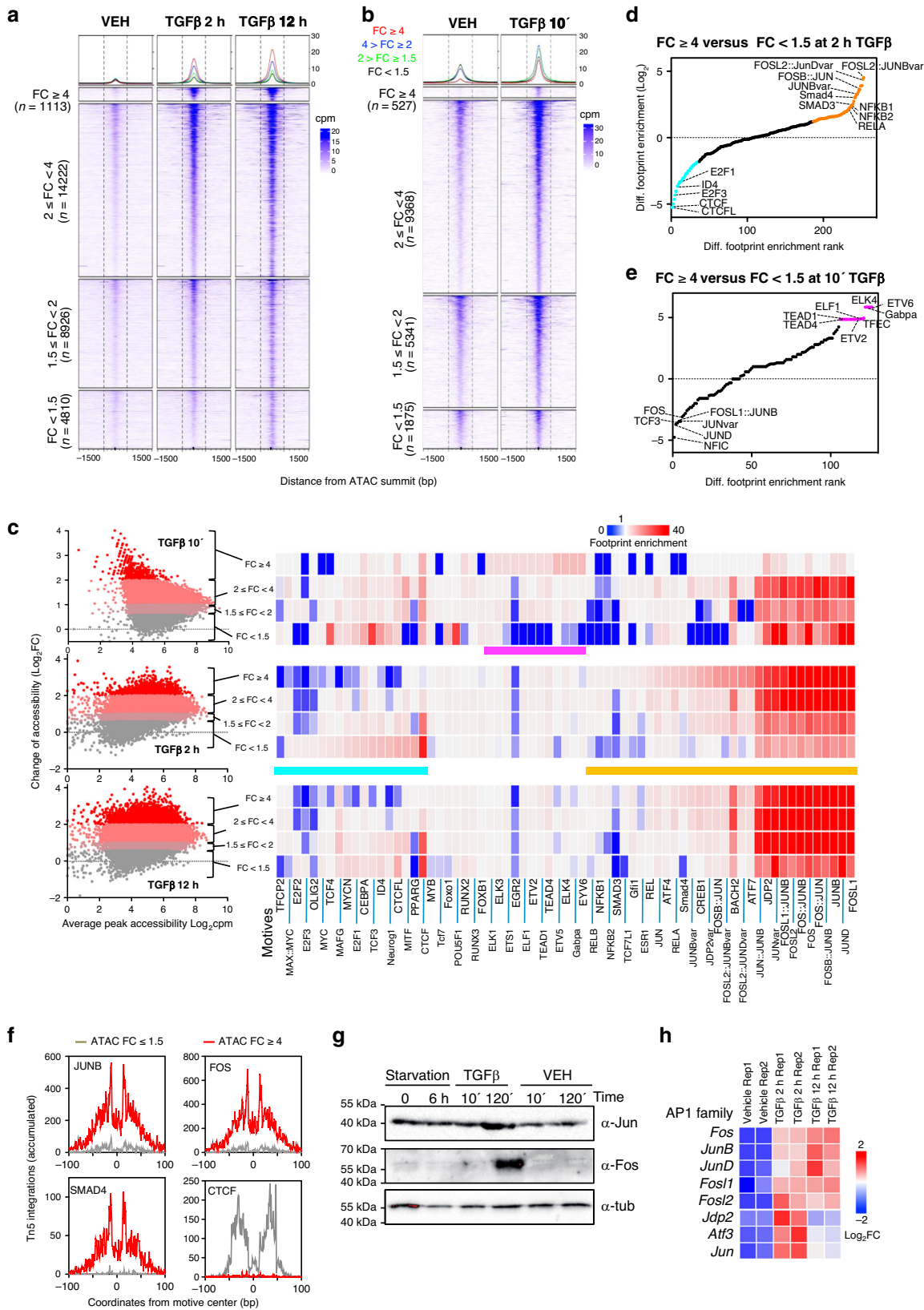
Both Jun and Fos, two classical AP-1 factors, were induced by TGF β after 2 h, but not after 10 min of treatment, as shown by western blotting (Fig. 2g). Furthermore, RNA-seq transcriptomic analysis confirmed that several genes encoding components of AP-1 dimers, including *Jun*, *Fos*, *JunB*, *JunD*, *Fos1*, *Fos2*, *Jdp2* and *Atf3*, were also significantly upregulated by 2 h and 12 h of TGF β treatment (Fig. 2h and Supplementary Fig. 4a, Supplementary Data 1). Thus, TGF β causes a fast and widespread chromatin opening, which depends on both the canonical and the non-canonical TGF β 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 TGF β addition. In the next sections we will concentrate in the chromatin changes that occurs 2 h and 12 h after TGF β .

Identification and characterization of TGF β -regulated enhancers. Enhancers are often classified based on their histone post-transcriptional modifications^{12,15}. 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 eRNAs^{41,42}. H3K4me3, a typical active promoter mark, but also present in some active enhancers^{43,44} 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 TGF β stimulation in NMuMG cells (see Supplementary Data 2). Since the level of H3K27ac is an excellent indicator of enhancer activation or repression^{17,45,46}, we used this mark to classify the enhancers upon TGF β treatment. We found the following categories (Fig. 3a and Supplementary Fig. 5): early-activated ($n = 2166$), late-activated ($n = 2242$), early-repressed ($n = 1059$) late-repressed ($n = 2654$) and TGF β -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, TGF β 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 TGF β treatment). As expected, 100% of the TGF β -repressed enhancers were active before TGF β treatment. Supplementary Data 2 contains information about the status of all identified enhancers from NMuMG cells, before and after TGF β treatment. The levels of H3K4me1 and H3K4me3 marks followed a pattern similar to H3K27ac, although the changes were smaller and delayed. For example, H3K4me1 levels decreased (Log₂FC ≤ -1) 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 TGF β treatment was observed (Supplementary Fig. 6b). However, following TGF β 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 TGF β -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

Fig. 2 ATAC-seq analysis of TGF β -treated cells. **a, b** Heatmaps showing alignment of ATAC-seq peaks signal from cells treated with vehicle, or TGF β for 2 h or 12 h (**a**) or with vehicle or TGF β for 10 min (**b**). Peaks are divided into four categories ($FC < 1.5$, $1.5 \leq FC < 2$, $2 \leq FC < 4$ and $FC \geq 4$) based in change of ATAC-seq signal between TGF β and vehicle for 2 h (**a**) or for 10 min (**b**). **c** TF occupancy depending on changes in accessibility at the indicated timepoint. Left: scatter plot of changes of ATAC-seq signal (y-axis, \log_2FC) versus average signal (x-axis). ATAC-seq peaks were divided into four categories according to FC as in (**a** and **b**). Right: TF genomic footprints enrichments were determined in the indicated category of ATAC-seq peaks. TF footprints were determined by calculating the protection from transposition observed in the ATAC-seq signal. TF motifs (from JASPAR) were identified in footprinted regions. Seventy of 550 are shown. **d, e** Differential footprint enrichment between $FC \geq 4$ and $FC < 1.5$ of TGF β or vehicle for 2 h (**d**) or for 10 min (**e**). Dots correspond to the same colored bars in **c**. **f** Example of TF footprint profile at 2 h after TGF β . Accumulated Tn5 integrations (y-axis) along the indicated TF binding sites at nucleotide resolution (x-axis, bp from center of each TF motif) in ATAC peaks with $FC \geq 4$ (red) and $FC < 1.5$ (gray). **g** Western blotting showing levels of Jun and Fos proteins under the indicated conditions. A representative image out of three independent biological replicates is shown. **h** Heatmap showing changes of mRNA levels by RNA-seq (\log_2FC) of genes encoding members of the AP-1 family. Source data are provided as a Source Data file.

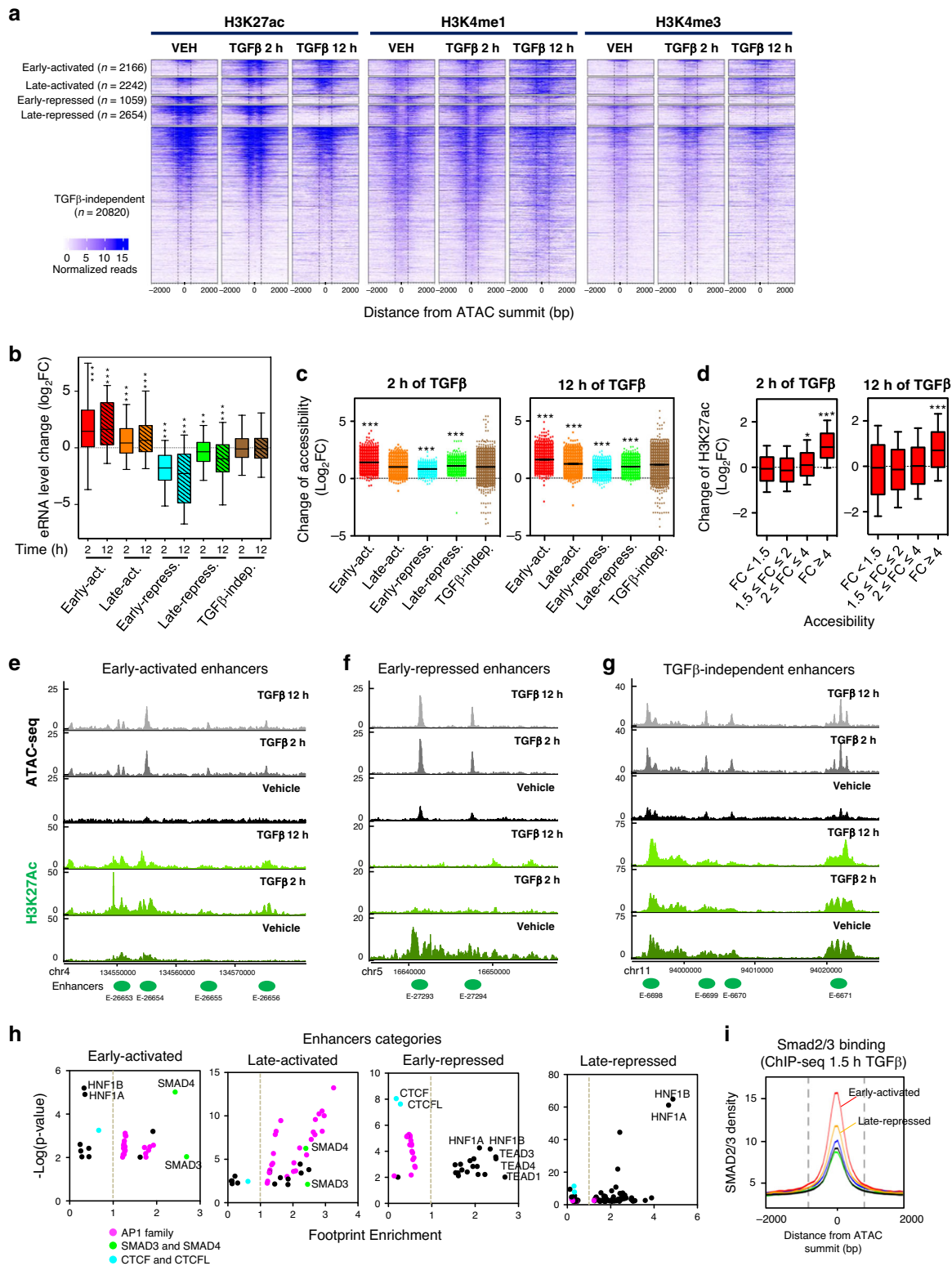
enriched in TEAD, HNF1A and HNF1B footprints, and depleted of AP-1 family footprints (Fig. 3h). Notably, we found no enrichment of SMAD3 or SMAD4 footprints at repressed enhancers; however, SMAD2/3 binding was detected by ChIP-seq at late-repressed enhancers, suggesting that SMAD2/3 requires interaction with other TFs to bind these repressed enhancers (Fig. 3i and Supplementary Fig. 7b, c).

Most TGF β -dependent genes are in the TGF β -regulated enhancers neighborhood. We next investigated the location of TGF β -regulated enhancers with respect to the TGF β -regulated genes (as determined by RNA-seq). We first labeled genes surrounding a regulated enhancer as Eg (Enhancer gene) $\pm n$, 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 TGF β -regulated enhancer, changes of mRNA levels of the neighboring genes (Eg ± 1 to Eg ± 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 TGF β -dependent genes are in a neighborhood of ± 3 genes from an enhancer regulated in the same way ($p = 7.63 \times 10^{-6}$ 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 ($p = 9.6 \times 10^{-31}$ 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, Eg ± 2) is robustly regulated ($|\log_2FC| > 1$; adjusted $p < 0.05$) in the same way as the enhancer, the interposing gene (Eg-1 or Eg+1, respectively) also tends to be regulated that way (Fig. 4d and Supplementary Fig. 9).

TGF β -regulated genes are often in co-regulated clusters. Next we analyzed the genomic distribution of genes regulated by TGF β . First, genes robustly regulated by TGF β treatment

(adjusted p -value < 0.05 and $|\log_2FC| > 1$) 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, TGF β 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 TGF β -regulated gene by first numbering genes according to their 5'-3' chromosomal order (1, 2, ..., n), and then plotting the FC of genes $n \pm 1$ to $n \pm 4$ for every robustly TGF β -regulated gene. Indeed, upstream and downstream neighboring genes of a robustly TGF β -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 $n \pm 3$ position in early categories, and to the $n \pm 2$ position in late categories. Significant co-regulation was observed for up to about 100 kb upstream and downstream of a robustly TGF β -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 TGF β . 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 TGF β regulation often operates in clusters of genes.

Characterization of TGF β regulatory domains. The results exposed in the last two sections suggest that TGF β -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 TGF β -co-regulated genes, using a less stringent but still significant threshold ($|\log_2FC| \geq 0.5$; $p < 0.05$). We found 372 of these clusters (86 early-upregulated, 66 late-upregulated, 49 early-downregulated, 173 late-downregulated) (Supplementary Data 3), which we termed TGF β 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) ($p = 3.3 \times 10^{-19}$ and $p = 7.7 \times 10^{-10}$, 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-*

Fig. 3 Identification of distinct classes of enhancers regulated by TGF β . **a** Heatmaps showing ChIP-seq signal of H3K27ac, H3K4me1, and H3K4me3 at distinct categories of enhancers and at the indicated times. **b** Change of eRNA levels of different categories of enhancers shown in **(a)**, at the indicated time. Intergenic eRNA-expressing enhancers were selected as described in Methods. Early-activated, $n = 328$; Late-activated, $n = 313$; Early-repressed, $n = 131$; late-repressed $n = 418$; TGF β -independent, $n = 2414$ enhancers. **c** Change of accessibility (ATAC-seq signal) of different categories of enhancers shown in **a**, 2 h (left panel) and 12 h (right panel) after TGF β treatment. Each dot corresponds to an enhancer: early-activated ($n = 2166$), late-activated ($n = 2242$), early-repressed ($n = 1059$) late-repressed ($n = 2654$) and TGF β -independent enhancers (20820). **d** Change of H3K27ac levels, 2 h (left panel) and 12 h (right panel) after TGF β treatment, of the different enhancers classified according to change of accessibility as in Fig. 2a. $FC < 1.5$ ($n = 4810$), $1.5 \leq FC < 2$ ($n = 8926$), $2 \leq FC < 4$ ($n = 14222$) and $FC \geq 4$ ($n = 1113$). Statistical significance in **b-d** are given with respect to the TGF β -independent category of enhancers (**b, c**) or the category $FC < 1.5$ (**d**) and were determined with the two-tailed Mann-Whitney non-parametric test, $*p \leq 0.01$; $**p \leq 0.001$; $***p \leq 0.0001$. Exact p -values are provided in Supplementary Data 5. **b, d** The horizontal black line of the boxplot and scatter plot represent the median and the mean value, respectively. The box spans the 25th to 75th percentiles, and whiskers indicate 5th and 95th percentiles. **e-g** Screen shot of H3K27ac and ATAC-seq profiles at three different genomic regions. While ATAC signal increased in all enhancers shown, H3K27ac signal increases (**e**), decreases (**f**), or remained unchanged (**g**), upon TGF β treatment. **h** Enrichment of TF footprints at the indicated categories of enhancers with respect to the rest of categories. On the y -axes, $-\log_{10}(p\text{-value})$ of enrichment (Fisher's exact test) for each TF is represented. **i** SMAD2/3 ChIP-seq signal density in the different categories of enhancers. Source data are provided as a Source Data file.

Kansl3 and *Lad1-Tmmt2-Pkp1* 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 TGF β -dependent upregulation of expression of all genes within the TRD (Fig. 6b, d, f). As controls, we showed that neither the TGF β -non regulated gene *Gnai3* (contiguous to *Amigo1-Cyb561d1-Atxn712*) nor the TGF β -repressed gene *Neurl3* (contiguous to *Arid5a-Kansl3*) 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 TGF β 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 TGF β (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 TGF β 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 a statistically significant tendency of coordinated regulation within TADs, we calculated the percentage of genes per TAD with a positive or negative FC after TGF β treatment⁵¹. We observed more TADs than randomly expected with all their genes activated ($p = 7.7 \times 10^{-6}$ for 2 h and 1.7×10^{-5} for 12 h) or repressed ($p = 1.7 \times 10^{-28}$ for 2 h and 1.2×10^{-7} for 12 h) by TGF β , while we observed fewer TADs than randomly expected with around 50% of genes activated or repressed ($p = 5.9 \times 10^{-15}$ 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 TGF β .

Next, we investigated the role of TAD boundaries on gene-gene and enhancer-gene co-regulation. We observed a clear correlation

between TGF β -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 TGF β -regulated enhancers from TGF β -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 TGF β causes a fast and widespread increase of accessibility at enhancers and TSS, using both canonical and non-canonical TGF β pathways. We also describe the enhancer repertoire of NMuMG cells and characterize its regulation by TGF β , suggesting a role of AP-1 together with SMAD proteins in TGF β -dependent enhancer regulation. We extensively investigated the genomic organization of TGF β -regulated enhancers and genes. Most TGF β -regulated genes are close to TGF β -regulated enhancers. Often, TGF β -regulated genes are grouped into what we term TGF β 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^{5,6,10,15}, although de novo activation from latent enhancers has been also reported^{16,17}. Indeed, we show that most TGF β -regulated enhancers were already active or primed before TGF β treatment. Unexpectedly, however, we also found that shortly after addition of TGF β , 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:

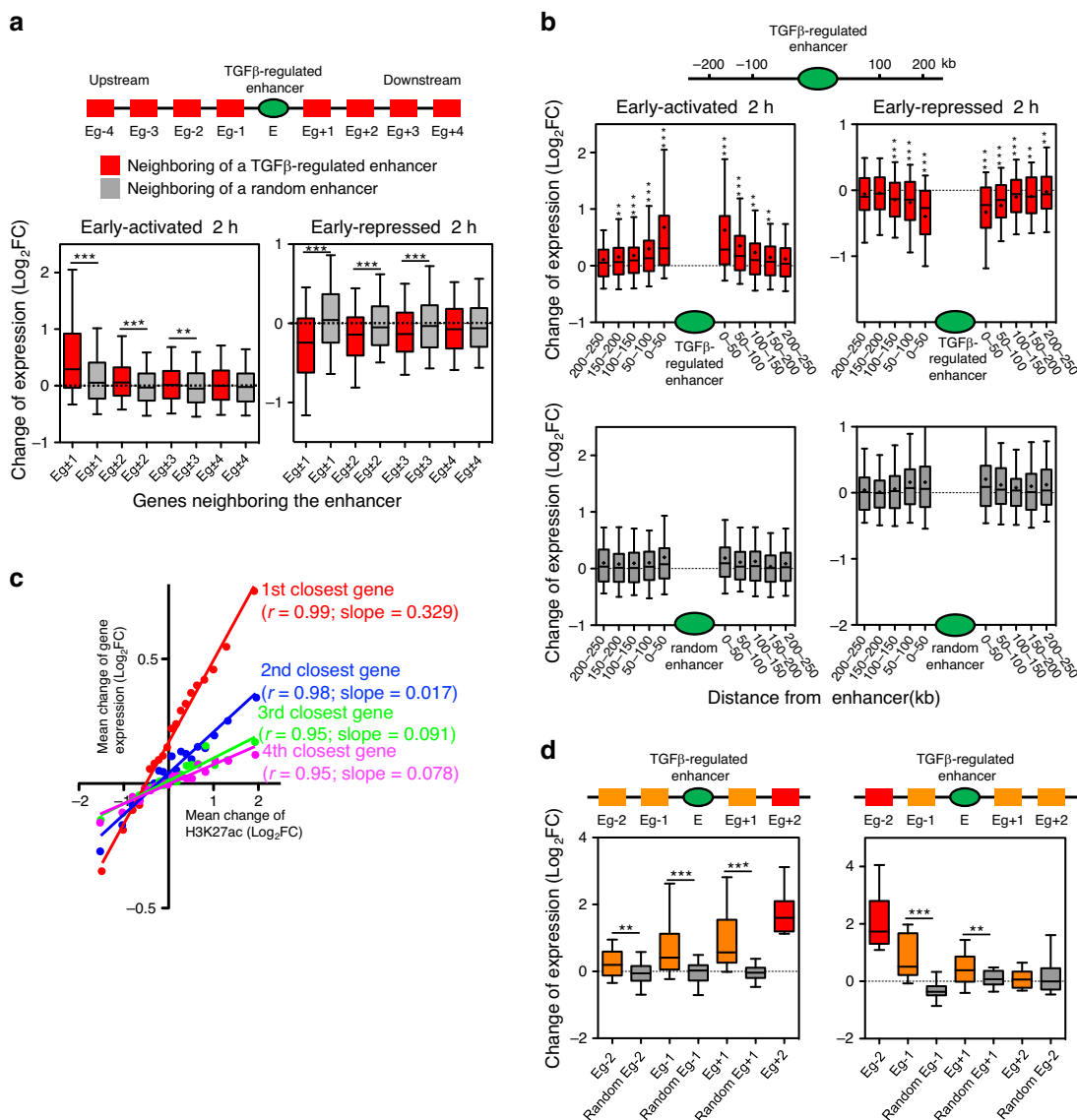
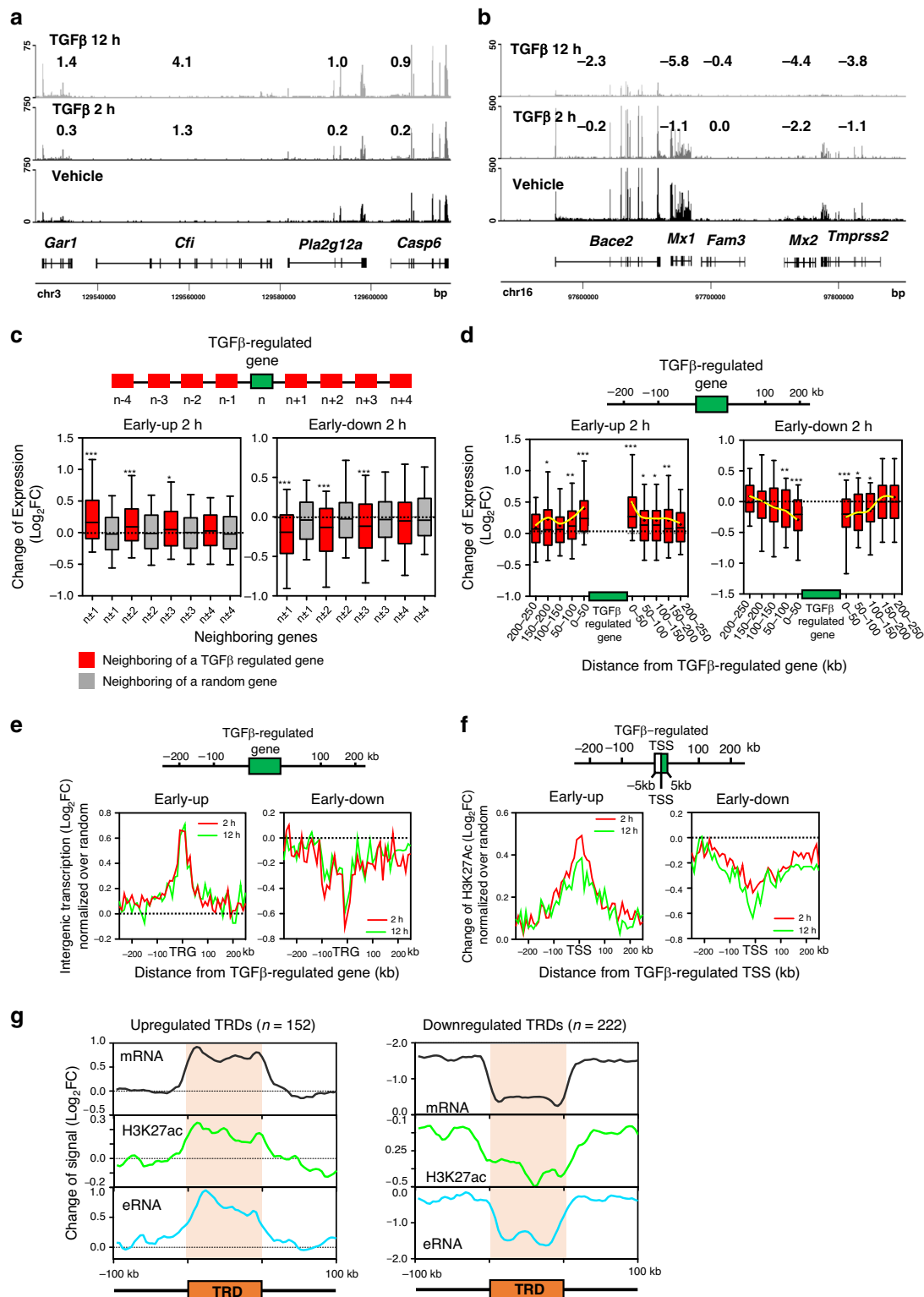


Fig. 4 Most TGFβ-dependent genes are in the TGFβ-regulated enhancers neighborhood. **a** Change of mRNA levels (RNA-seq signal of 2 h TGFβ versus vehicle) for genes upstream or downstream of an early-activated (left) or early-repressed (right) enhancer, as in the scheme (red boxes), or of a randomly selected enhancer (gray boxes). For randomization, see Methods. Eg±n (whereby n = 1, 2, 3 or 4) depicts genes that occupy the first, second, third, or fourth positions, respectively, in the chromosomal order, upstream (–) or downstream (+) of the enhancer. Plots for late-enhancer categories are shown in Supplementary Fig. 8a. **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. Inclusion of the genes in the interval was determined by the position of their TSSs. Lower panels represent the corresponding data when enhancers were randomized. Plots for other enhancer categories are shown in Supplementary Fig. 8b. **c** Correlation plot between changes of mRNA levels (RNA-seq signal of 2 h TGFβ versus vehicle) of the first, second, third, or fourth closest genes respect to an enhancer, versus the change of H3K27ac of the enhancer (ChIP-seq 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 ($\log_2 FC > 1$; 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, d** Statistical significance between real and random distributions were determined with the two-tailed Mann-Whitney test, * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. 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 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^{7,52,53}. 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^{16,54–57}.

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

Fig. 5 Co-regulated clusters of TGF β -regulated genes. **a, b** Examples of clusters of TGF β co-regulated genes. Screenshot of RNA-seq tracks at two different genomic regions. Numbers show Log₂FC of the indicated gene at the corresponding timepoint of TGF β versus vehicle. **c** Changes of mRNA levels (RNA-seq) of the four genes located upstream ($n - 1$ to $n - 4$) and downstream ($n + 1$ to $n + 4$) of a robustly TGF β -regulated gene ($|\log_2FC| > 1$; adjusted $p < 0.05$) at position n . The central gene (n) is early-upregulated (left) or early-downregulated (right). Gray boxes correspond to the changes of mRNA levels of genes around a random gene. For randomization, see Methods. **d** Change of mRNA levels (RNA-seq) of genes located at the indicated distance (kb) upstream or downstream of a TGF β -regulated gene. Left: the central gene is early-upregulated. Right: the central gene is early-downregulated. Randomizations are shown in Supplementary Fig. 10d. **e, f** Statistical significance between real and random distributions were determined with the two-tailed Mann-Whitney non-parametric test, * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. Sample size (n) 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. A black dot in the boxplot represent the mean. **e** Changes of intergenic transcription in the neighborhood (± 250 kb binned in 10 kb bins) of a TGF β -regulated gene (TRG) normalized to random. ChromRNA-seq data were used. To avoid termination read-through and promoter-divergent transcription, regions of 2 kb upstream and downstream of the TSS and transcription termination site were not considered. **f** Change of H3K27ac (ChIP-seq signal) in the neighborhood (± 250 kb binned in 10 kb bins) of a TGF β -regulated TSS normalized to random. To avoid histone modifications of the TGF β -regulated TSS, regions of 5 kb upstream and downstream of the TSS were not considered. **c-f** Late-upregulated and late-downregulated categories are shown in Supplementary Fig. 10b, c, e, f. **g** Meta-analysis of TRDs. Density plots of changes of mRNA, eRNA, and H3K27ac levels are shown. Data from early- and late-upregulated or downregulated TRDs were pooled. Source data are provided as a Source Data file.

player of the TGF β signaling⁵⁸. ERK may regulate chromatin opening in several ways, including AP-1 and other transcription factors regulation^{59–63}, phosphorylation of histone H3 at serine 10 through MSK1^{64,65} and direct regulation of the chromatin and transcription machinery^{66,67}. TAK1 is also a component of the TGF β non-canonical signaling pathway that acts upstream of JNK^{58,68}, 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 TGF β . 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 TGF β . At this time ETS TF footprints were enriched in enhancers that strongly gained accessibility. As many ETS factors are direct targets of ERK⁶⁹, the ERK-ETS pathway might lead a first wave of accessibility, at least in a subset of enhancers.

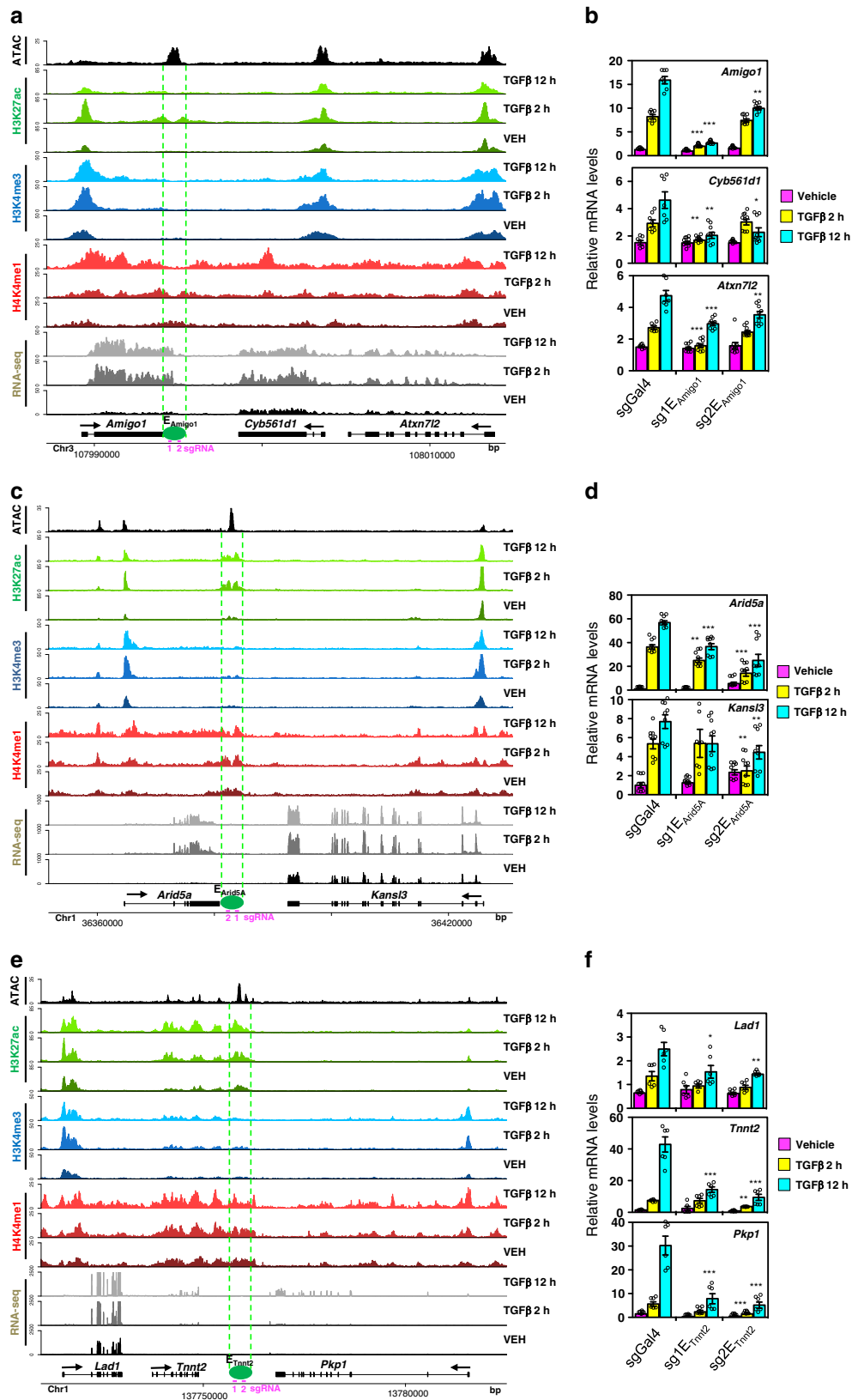
Even though about 80% of enhancers in a cell gained accessibility upon TGF β addition, only about 28% of them presented increased or decreased H3K27ac levels (indicating activation or repression, respectively, by TGF β). Activated enhancers were strongly enriched in SMAD3, SMAD4, and AP-1 footprints 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 enhancers⁷⁰. Consistent with a role of TEAD factors in TGF β -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 TGF β treatment (1.6- and 4-fold upregulation, respectively; Supplementary Data 1). SNAI1 is one of the master regulators of the EMT process⁷¹ and the *Snai1* gene is strongly upregulated by the TGF β 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^{72–79}, suggesting that this factor is more likely associated with promoters rather than with enhancers.

Enhancers can regulate genes from large distances through genomic loops⁴⁸. Our data indicate, however, that most genes robustly regulated by TGF β (79% of the upregulated, and 56% of the downregulated) were in the close vicinity of an enhancer that is also regulated by TGF β 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-workers⁸⁰ 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 indiscriminating 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^{81–83}. 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 TGF β 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^{84–87}, and enhancers have been proposed as probable regulatory elements responsible for co-regulation^{88–90}. 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 TGF β -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 TGF β -dependent enhancers display a rather promiscuous preference for surrounding genes.

To which extent TADs control transcription is still open to discussion^{91–93}. 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 TGF β . 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^{51,94,95}. It is well-established that TAD borders are enriched in CTCF-binding sites and have enhancer-promoter insulation activity⁵⁰. 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, e Screenshot of H3K27ac, H3K4me1, H3K4me3, RNA-seq, and ATAC-seq tracks at three different genomic regions. Position of a TGF β -regulated enhancer is shown. For the inactivation of the enhancers by CRISPRi, two different sgRNAs (pink) that target each enhancer were used. **b, d, f** Determination of mRNA levels of the indicated genes by RT-qPCR analysis. Values were normalized to the *Gapdh* mRNA level. Values are average \pm 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. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. 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 expression⁹⁶.

In summary, our work provides a valuable catalog of enhancers and their regulation by TGF β 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 μ g/ml insulin (complete medium), until reaching 70–80% confluency. HEK293T cells were grown in Dulbecco's modified Eagle medium (DMEM) with 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 μ g/ml). TGF β 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, TGF β 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 TGF β treatments, 5 ng/ml TGF β 1 diluted in 4 mM HCl, 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-sequencing. The ATAC experiments were performed according to ref. 28. NMuMG cells were treated with either vehicle or TGF β for 2 or 12 h (after 6 h of serum and insulin starvation), vehicle or TGF β -treated for 10 min (after 6 h of serum and insulin starvation) or vehicle or TGF β -treated for 2 h (in complete medium); and then collected and treated with transposase Tn5 (Nextera DNA Library Preparation Kit, Illumina). DNA was purified using MinElute PCR Purification Kit (Qiagen). All samples were then amplified by PCR using NEBNextHigh-Fidelity 2 \times PCR Master Mix (New England Labs) with primers containing a barcode to generate libraries. DNA was again purified using MinElute PCR Purification kit and samples were sequenced using Illumina NextSeq 500 system (Illumina) with 75 bp paired-end reads at the Genomic Unit of CABIMER (Sevilla, Spain)

ATAC-see experiments. For ATAC-see experiments, cells were grown on coverslips and processed according to ref. 29. For kinetic analyses, NMuMG cells were changed to serum- and insulin-free medium for 6 h and then were treated with vehicle or TGF β 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 DsiSMAD4 (IDT, Leuven, Belgium) (previously used in ref. 97), 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 TGF β 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 TAKINIB (SML2216, Sigma-Aldrich) inhibitors at 10 μ M for 2 additional hours and finally cells were treated with vehicle or TGF β for 10 min. Other ATAC-see 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 at 37 °C. The transposase mixture solution (25 μ l 2 \times TD buffer, final concentration of 100 nM Tn5-ATTO-590N, adding dH₂O up to 50 μ l) was added on the coverslips and incubated for 30 min at 37 °C. As a negative control 50 mM EDTA was added to the Tn5-ATTO590N reaction buffer. After the transposase reaction, coverslips were washed with PBS containing 0.01% sodium dodecyl sulfate (SDS) and 50 mM EDTA for 15 min three times at 55 °C. Then, slides were stained with DAPI for 5 min and mounted using Vectashield for imaging.

DNaseI-seq. DNase I hypersensitivity mapping was performed according to ref. 98 with brief modifications. NMuMG cells were either vehicle or TGF β -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 KCl, 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 \times 10⁶/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 55 °C. After 15 min, proteinase K (final concentration of 20 μ g/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 selected 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 Illumina NextSeq 500 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 TGF β 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 50 bp single-end reads at the Genomics Core Facility (EMBL Heidelberg).

ChromRNA sequencing. Cellular fractionation was carried out according to ref. 99, based in the protocol described in ref. 100. Cell pellets were resuspended in 400 μ l cold cytoplasmic lysis buffer (0.15% NP-40, 10 mM Tris pH 7.5, 150 mM NaCl) 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 microfuge tubes at 3500 g for 10 min. The nuclear pellets were gently resuspended into 250 μ l cold glycerol buffer (20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol). An additional 250 μ l of cold nuclei lysis 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 microfuge tubes for 2 min at 13,000 \times g. Fifty microliters of cold PBS was added to the remaining chromatin pellet, and gently pipetted up and down over the pellet, followed by a brief vortex. Then RNA was extracted using TRIZOL and treated with TURBO DNA-free kit (AM1907, Invitrogen) for DNA removal according to manufacturer's instructions. Libraries 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 h or TGF β 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 Pipes pH 8, 85 mM KCl, 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 diluted 1:10 in 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 μ g of the respective antibodies: anti-H3K27ac (Ab4729, Abcam), anti-H3K4me1 (Ab8895, Abcam) and anti-K3m4me3 (Ab8580, 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, 2 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 milliQ water. The sequencing was

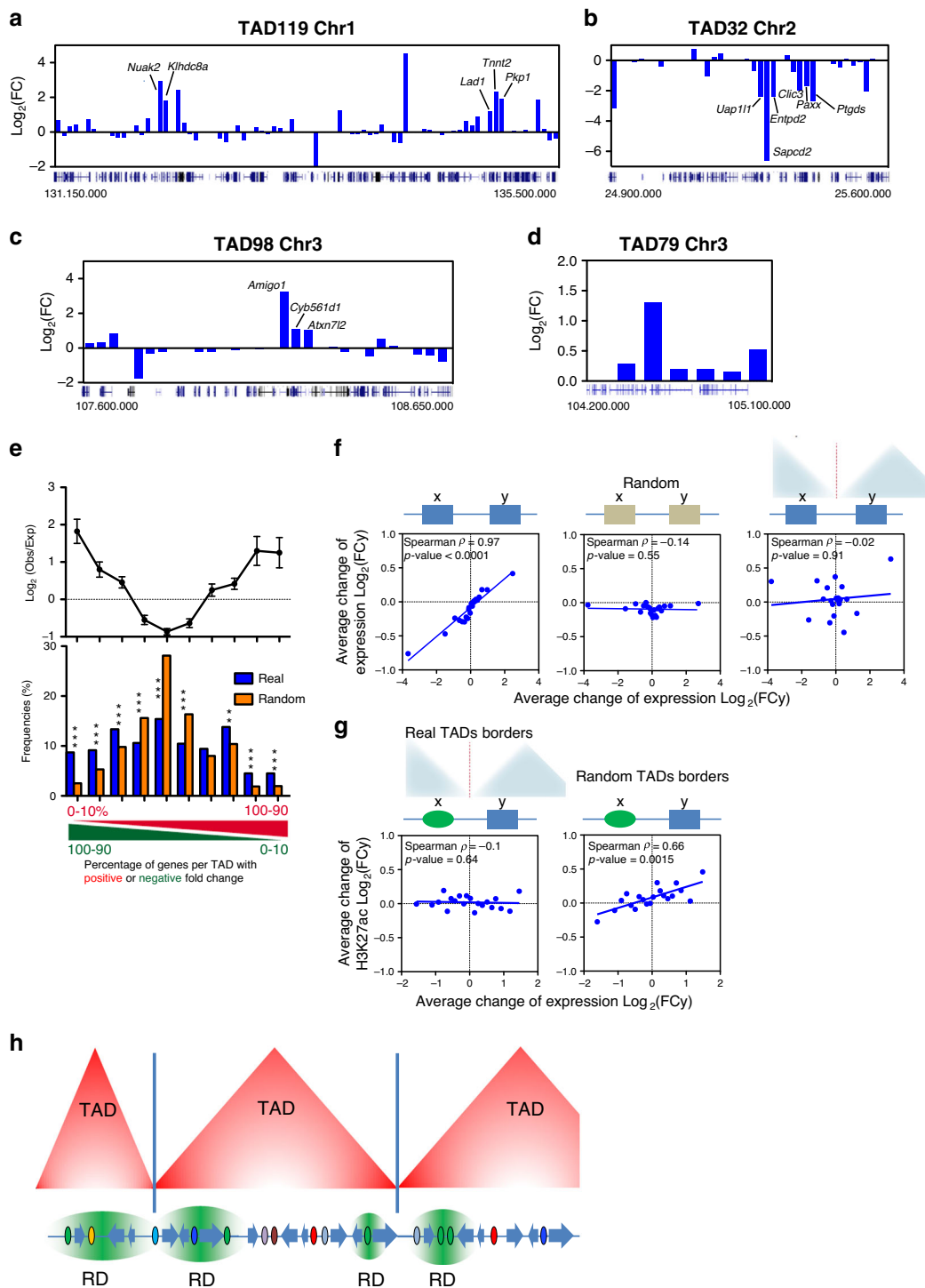


Fig. 7 TRDs Differ from TADs but are Constrained by TAD Borders. **a–d** Changes of mRNA levels (2 h TGF β versus vehicle) of genes contained at the indicated TADs. Co-regulated genes at TRDs are shown. **e** Top panel: ratio of observed versus expected frequencies of TADs with distinct proportions of genes with upregulated or downregulated FC (FC > 1 or FC < 1; 2 h TGF β versus vehicle). Values are means \pm 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. $**p \leq 0.001$; $***p \leq 0.0001$. 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 TGF β 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. 15 d. **g** Correlation plot between change of H3K27ac signals of enhancers (ChIP-seq signal 2 h TGF β versus vehicle) and change of mRNA level (RNA-seq signal 2 h TGF β 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. 15 e. **h** Model of influence of TGF β -regulatory domains (TRD) and other regulatory domains (RD) constrained by the insulating activity of TAD borders. Small TADs often 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-CRIPSR-Cas9 (C15310258-100, 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 μ l 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. Sequences of all oligonucleotides are listed on Supplementary Data 4. Data are the average of two or three biological independent replicates and three technical replicates of each biological replicate.

Western blotting. Western blotting were performed according to ref. 102, using the following antibodies: anti-cFos (2250, Cell Signaling, Dilution 1/1000), anti-cJun (60A8; 9165, Cell Signaling, Dilution 1/1000), anti- α -Tubulin (DM1A; T9026, Sigma, Dilution 1/1000), anti-Erk1/2 (P28482, Millipore, Dilution 1/1000), and anti-Phospho-Erk1/2 (thr202/Tyr204) (9101, Cell Signaling, Dilution 1/1000), anti-SMAD4 (SC-7996, Santa Cruz, Dilution 1/250). As secondary antibodies anti-mouse-HRP and anti-rabbit-HRP (Sigma-Aldrich, Dilution 1/1000) were used.

CRISPR interference. For enhancer inactivation we used CRISPRi technology described in ref. 47. First the NMuMG-dCas9 cell line was established as follows. For lentiviral particles preparation 2×10^6 HEK293T cells were transfected with Eugene 6 (Promega) using 7.5 μ g of the transfer vector pHR-SFFV-dCas9-BFP-KRAB (this plasmid was a gift from S. Qi & J. Weissman (Addgene plasmid # 46911; <http://n2t.net/addgene:46911>; RRID:Addgene_46911)⁴⁷ with 5 and 2.5 μ g of the packaging plasmids pCMVDR8.91 and pVSVG, respectively. Lentiviruses were harvested 72 h post-transfection, passed through a 0.45- μ m filter and concentrated by ultracentrifugation at 100,000 \times g for 90 min. Virus particles were resuspended in DMEM, snap frozen in liquid nitrogen and stored at -80°C . After titrating, NMuMG cells were infected with pHR-SFFV-dCas9-BFP-KRAB lentiviruses using polybrene 8 μ g/ml. Single BFP-positive cells were sorted 72 h after infection into individual wells of a 96-well plate, containing a mixture of fresh and conditioned medium (1:1), and then one clone was chosen after BFP signal and EMT functional validation. Then, NMuMG cell pools expressing specific sgRNAs were established as follows. To design sgRNAs we selected neighboring region for ATAC-seq peak submit and use IDT (https://eu.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM) and CRISPOR (<http://crispor.tefor.net/>) tools. Designed sgRNA for selected enhancers (Supplementary Data 4) were inserted into sgRNA expressing lentiviral vectors pU6-sgGAL4-1 (this plasmid was a gift from S. Qi & J. Weissman (Addgene plasmid # 46915; <http://n2t.net/addgene:46915>; RRID:Addgene_46915)⁴⁷). Then lentiviral particles were produced and purified as above. After titrating, NMuMG-dCas9 cells were infected with sgRNA expressing lentiviruses and selected with puromycin 1 μ g/ml for one week before working with the selected pool.

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 μ g of total RNA using SuperScript First Strand Synthesis System (Invitrogen). Two 2 μ l 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.

Computational methods and statistical analysis. Most of analyses were performed using R (v3.4.4), Rstudio (v0.99.879) and Bioconductor (v2.38).

Statement on the use of the mm9 assembly. We used the GRCm37 (mm9) assembly to map all sequencing reads from mouse origin in this study (ATAC-seq, RNA-seq, ChromRNA-seq, DNaseI-seq and ChIP-seq) because many programs did not support yet the mm10 build when we started our study. As our study compares samples across different treatments and does not perform absolute analyses, re-aligning the reads to GRCm38 (mm10) should not significantly affect our conclusions.

Processing of ATAC-seq data. ATAC-seq data were processed (trimmed, aligned, filtered, and quality controlled) for each condition independently, using the ATAC-seq pipeline from the Kundaje lab (<https://github.com/ENCODE-DCC/atac-seq-pipeline>). Information about read depth can be found in the reporting summary. The model-based analysis of ChIP-seq macs2 (v2.1.2) was used to identify the peak regions with options -B, -q 0.01-nomodel, -f BAM, -g mm9. The Irreproducible

Discovery Rate (IDR) method was used to identify reproducible peaks between two biological replicates. Only peaks reproducible between the two biological replicates (with IDR ≤ 0.05) were retained for downstream analyses. Peaks for all conditions of each experiment were then merged together into a standard peak list. For the experiment described in Figs. 1a and 2a, 52,626, 100,459, and 105,334 peaks were identified for vehicle, TGF β 2h and TGF β 12h conditions, respectively. Then sets of peaks for each condition were refined. All peaks smaller than 150 bp were discarded, and all peaks closer than 1000 bp were merged together using csaw (v1.12.0)¹⁰³, obtaining 16,769, 32,545, and 33,077 peaks for each condition. Then, peaks for the three different conditions closer than 1000 bp were combined in a single set of peaks obtaining 39,432 unique regions. Similar procedures were used for the experiment described in Fig. 2b and Supplementary Fig. 2c–e. In all the cases summits were re-calculated using MACS2 refinepeak with default options. Peaks coordinates for downstream analysis were calculated as summit ± 500 bp. For enhancer assignment, those peaks that overlap with TSS ± 1000 bp of any transcript from mm9 UCSC KnownGene, were discarded using bedtools (v.2.27.1) subtract with -A options, obtaining 29,071 peaks, which were considered as putative enhancers. Correlation between replicates were calculated using multiBamSummary and plotCorrelation from deeptools (v3.1.3) using default options. Shushi package (v.1.16.0) was used for data visualization.

Processing of DNaseI-seq data. DNaseI-seq data was processed following the same pipeline as ATAC-seq data but using specific options for DNaseI-seq. Information about read depth can be found in the reporting summary.

Classification of enhancers according to increase of accessibility. To classify enhancers according to the increase of accessibility, reads were counted in full-length of ATAC-seq peaks using regionCounts() from csaw with minq = 40 option. To normalize data, first we count reads in full genome using windowCounts() with bin = TRUE, width = 2000 and minq = 40 options; then normalization factors were calculated using normOffsets(). After that, estimateDisp(), glmQLFit() and glmLRT() from edgeR package (v3.20.9) was used for calculating log₂FC and statistics. Enhancers were then classified into four groups: FC < 1.5, 1.5 \leq FC < 2, 2 \leq FC < 4 and FC ≥ 4 , obtaining 4810, 8926, 14,222, and 1113 for TGF β 2h vs veh comparison; and 3988, 7010, 15,362, and 2771 for TGF β 12h vs. veh. EnrichedHeatmap (v1.12.0) package was used for drawing heatmaps. In Fig. 2a rows were ordered according to decreasing ATAC-seq cpm for TGF β 2h or TGF β 12h (merged replicates). Differential ATAC-seq signal analysis (Supplementary Fig. 1b) was also performed in full-length ATAC peaks overlapping TSS (promoters), putative enhancers and the whole genome outside ATAC-seq peaks (rest) using regionCounts() with same parameters as above.

Analysis of ATAC-seq using TANGO. Confocal microscopy images were analyzed using TANGO³⁰ 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 TGF β 2h 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 regions were 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 imageJ¹⁰⁴ macro as follow: first images were splitted by color. Then nuclei were segmented on DAPI max projection previous preprocessing and size-selection. ATTO590 signal was quantified in max projection images in nucleus-exclusive area after background subtraction. All experiments were median-normalized. Three independent experiments were performed. Number of nuclei quantified (*n*) per condition including the three replicates are provided in Supplementary Data 5.

Motif-binding analysis with CentriMo. Motif-binding analyses was performed with CentriMo (v5.0.2)¹⁰⁵ from MEME suite against all 579 binding motifs from JASPAR vertebrates¹⁰⁶ with default options. CentriMo perform Central Motif Enrichment Analysis (CMEA), which better identified binding motifs involved in the regulation of ChIP-seq or ATAC-seq peaks. Position Weight Matrixes (PWM) for SNAI1 and SMAD5GC¹⁰⁷ binding motifs were also constructed and included in the analysis. For each category, enhancer sequences were shuffled five times for obtaining 5x random sequences with the same %CG content using scrambleFasta.pl from HOMER (v4.10.3) suite, and use that set of sequences as a background.

Footprinting analysis. To identify footprints in enhancer sequences wellington_footprints.py script from pyDNase suite (v0.2.4)¹⁰⁸ was used with -A option. Footprints were detected in TGF β 2h and TGF β 12h ATAC-seq data. Both ATAC-seq replicates for each condition were merged together for increasing the strength of footprint identification. For footprints enrichment analyses in different categories of chromatin accessibility first scrambleFasta.pl was used for obtaining 5x background sequences for each category. Then we look for motif-binding sites

in real footprints and background sequences with FIMO (v5.0.0)¹⁰⁹ using default option and p -value < 0.0001 as cutoff. Enrichment was calculated as:

$$\text{Enr} = (\text{footprints}[\text{motif}]/\text{footprints}[\text{total}])/(\text{footprints}_{\text{bg}}[\text{motif}]/\text{footprints}_{\text{bg}}[\text{total}]) \quad (1)$$

where $\text{footprints}[\text{motif}]$ is the number of footprints with a specific motif-binding sequence; $\text{footprints}[\text{total}]$ is the number of total footprints; $\text{footprints}_{\text{bg}}[\text{motif}]$ is the number of footprints with a specific motif-binding sequence in the background set of sequences; and $\text{footprints}_{\text{bg}}[\text{total}]$ 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 `gitools` (v2.3.1) including most representatives motifs.

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 TGF β 2h BAM file. Then we resized our reads to 1 bp and make the shift of 4 or -5 bp 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. TGF β 2h ATAC-seq data were used for footprints identification in early categories of regulated enhancers, and TGF β 12h for late categories.

SMAD2/3 ChIP-seq analysis. ChIP-seq data for SMAD2/3 in NMuMG cells were downloaded from GSE121254³⁶, and aligned against mm9 mouse reference genome using `align()` function from `Rsubread` package (v1.28.1) with `type = 1`, `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 $-\log_{10}(\text{FDR}) \geq 5$ criteria and those peaks that overlap with blacklisted region of the genome were removed using `bedtools` `subtract` with `-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 our enhancers. Overlapping analysis between SMAD2/3 bound enhancers and enhancers classified according to its increase in chromatin accessibility at TGF β 2h and/or AP-1 footprints were performed using `vecsets` (v1.2.1) package. p -values were calculated using `dhyper()` function from `stats` R package. ATAC-seq signal density graphs (Supplementary Fig. 3f) were plotted for TGF β 2h using `computeMatrix` reference-point from `deeptools` 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` (v0.1.19) `rmdup` command. `FeatureCounts()` function from `Rsubread` package was used to assign reads to UCSC mm9 `KnownGenes` (miRNAs were discarded from the analysis) using `GTF.featureType = "exon"`, `GTF.attrType = "gene_id"` and `strandSpecific = 2` parameters on duplicate removed bamfiles. Then differential gene expression analysis was performed using the `voom/limma` (v3.34.9) and `edgeR` (v3.20.9) `Bioconductor` packages¹¹⁰. 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 BAYES moderated t -statistics test with adjustments for multiple comparison.) were classified according to:

Early-downregulated: $\log_2\text{FC}(\text{TGF}\beta 2\text{h_vs_veh}) < -1$ and not $\log_2\text{FC}(\text{TGF}\beta 12\text{h_vs_TGF}\beta 2\text{h}) > 1$

Early-upregulated: $\log_2\text{FC}(\text{TGF}\beta 2\text{h_vs_veh}) > 1$ and not $\log_2\text{FC}(\text{TGF}\beta 12\text{h_vs_TGF}\beta 2\text{h}) < -1$

Late-downregulated: $\log_2\text{FC}(\text{TGF}\beta 12\text{h_vs_veh}) < -1$ and not $\log_2\text{FC}(\text{TGF}\beta 2\text{h_vs_veh}) < -1$

Late-upregulated: $\log_2\text{FC}(\text{TGF}\beta 12\text{h_vs_veh}) > 1$ and not $\log_2\text{FC}(\text{TGF}\beta 2\text{h_vs_veh}) > 1$

Transient-downregulated: $\log_2\text{FC}(\text{TGF}\beta 2\text{h_vs_veh}) < -1$ and $\log_2\text{FC}(\text{TGF}\beta 12\text{h_vs_TGF}\beta 2\text{h}) < -1$

Transient-upregulated: $\log_2\text{FC}(\text{TGF}\beta 2\text{h_vs_veh}) > 1$ and $\log_2\text{FC}(\text{TGF}\beta 12\text{h_vs_TGF}\beta 2\text{h}) > 1$

TGF β -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), $\log_2\text{FC}$ were plotted against average expression ($\log_2\text{CPM}$) for all samples for each comparison. For boxplots of Supplementary Fig. 4b, distribution of $\log_2\text{CPM}$ of the two independent replicates were plotted. For expression heatmap of Fig. 2h $\log_2\text{FC}$ were plotted for selected genes using `gitools`.

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 `type = 1`, `TH1 = 2` and `unique = TRUE` parameters. The downstream analysis was performed on bamfiles with duplicates removed using the `samtools` `rmdup` command. Information about read depth can be found in the reporting summary. `Shushi` 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 RPKM were calculated using all counted reads in ATAC-seq peaks as library size and combining replicates. Peaks whose RPKM were larger than percentile 25 of all RPKM for H3K27ac or 10 for H4K4me1 were considered as having the mark. Enhancers without H3K4me1 nor H3K27ac in vehicle condition but that acquire these marks with TGF β 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 regulation by TGF β , we analyze H3K27ac dynamics in vehicle, 2 h and 12 h TGF β -treated ChIP-seq data, considering both replicates independently and using a pipeline, including `csaw` and `edgeR` (v3.20.9) packages. Reads were counted in full-length ATAC-seq peaks as above and normalization factors were calculated as `normOffset()` function with option `type = "loess"` parameter. For differential binding analysis `scaleOffset()`, `estimateDisp()`, `glmQLFit()` and `glmLRT()` functions was used. Then enhancers were classified in early-activated, early-repressed, late-activated and late-repressed, following same criteria as differentially expressed genes. For each comparison we also consider a p -value < 0.05 and average $\log_2\text{CPM} > 0$. 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 TGF β 2h 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 `sushi` package (v1.16.0) on 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 removed using the `samtools` `rmdup` command. As short size of eRNAs, ChromRNA-seq reads were counted in summit ± 200 bp for intergenic enhancers using `regionCounts()`. From 29,071 enhancers, those which overlap with any UCSC `KnownGene` transcript bodies (5 kb extended from the Transcription termination site (TTS) to avoid termination read-through reads) were discarded, resulting in 12857 intergenic enhancers. After read counting, we calculated CPM of eRNA for each eRNA-expressing enhancer. Correlation between H3K27ac levels and eRNAs levels shown in Supplementary Fig. 6b was calculated using H3K27ac levels of the 12857 enhancers where eRNAs were counted. For correlation between increase in H3K27ac and increase in eRNA shown in Supplementary Fig. 6c, d, first we filtered by expression, obtaining 3636 eRNA-expressing intergenic enhancers. Then, differential expression analysis using `csaw` and `edgeR` was computed.

Analysis of the location of TGF β -regulated genes with respect to TGF β -regulated enhancers. To study the regulation of genes in the neighborhood of TGF β -regulated enhancers, first we plotted $\log_2\text{FC}$ 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 with respect 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 $\pm n$, $Eg \pm n$, with $n = 1, 2, 3$, 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 circularized chromosome were constructed and random enhancers were obtained by shifting enhancers 50 positions. So, if the regulated enhancer in at 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 TGF β -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 $\log_2\text{FC}(\text{H3K27ac})$) and differential expression of the closest genes ($\log_2\text{FC}$), pairs of enhancer-closest

gene were ordered according to increasing $\log_2\text{FC}(\text{H3K27ac})$ and grouped into 20 bins. Then, least squares fit analysis were performed to obtain r and slope. To study if enhancers operate in a continuous manner, first we subset those enhancers with one of their neighborhood genes (for example, Eg+2) robustly regulated ($|\log_2\text{FC}| > 1$ and adjusted p -value < 0.05). Then, $\log_2\text{FC}$ 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 iterated 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 TGF β -regulated gene ($|\log_2\text{FC}| > 1$ and adjusted p -value < 0.05), $\log_2\text{FC}$ of genes $n \pm k$ with $k = 1, 2, 3, \text{ or } 4$, were plotted. Non-expressed genes were not taken into account for analysis. Random genes lists, per chromosome, 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 $\log_2\text{FC}$ distributions were plotted. Similar analysis was done for genes randomized 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 TGF β -regulated genes, previously all UCSC KnownGene transcript bodies (2 kb extended from the TSS and TTS to avoid termination read-through reads 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 TGF β robustly regulated gene TSS. For that, we used `region.counts().windowCounts()` with `bin = TRUE` and `width = 10,000` options. `NormOffsets()` was used to calculate normalization factors in libraries. Then we performed differential analysis bin to bin using `estimateDisp()` and `glmQLFit()` functions to obtain $\log_2\text{FC}$. Similar analysis was performed using random genes (as described above) instead of TGF β -regulated genes. Finally, $\log_2\text{FC}$ of random genes was subtracted from $\log_2\text{FC}$ of TGF β -regulated genes, bin to bin, to obtain $\log_2\text{FC}$ over random values that were finally plotted.

Study of the influence of co-regulation in histone modifications. $\log_2\text{FC}$ of H3K27ac, H3K4me3, and H3K4me1 ChIP-seq signal in the neighborhood of TGF β -regulated genes was performed as in 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 TGF β -regulatory domains. To identify TGF β -regulatory domains (TRD) we identified clusters of co-regulated genes, as above, using a different threshold of differentially expressed genes ($|\log_2\text{FC}| > 0.5$ and p -value < 0.05) to include genes that are significantly regulated with TGF β , although not robustly affected. Metaplots were created combining late and early TRDs for upregulation and downregulation and using `computeMatrix` scale-regions with `-m 200,000 -bs 4000 -a 10,000 -b 100,000` options. `-bl` option was used for calculating intergenic transcription (ChromRNA-seq) to avoid counting in all UCSC knownGene transcript bodies.

Analysis of the relationship between TRDs and chromatin 3D organization.

We used TADs coordinates from NMuMG cells after 8 h of TGF β treatment, published in ref. ⁴⁹ after conversion to mm9 mouse gene assembly using UCSC liftover tool. Random TADs lists for each chromosome were obtained by recirculating the chromosome and shifting TAD borders by 10 Mb. To determine within TAD co-regulation, first TADs with less than three expressed genes were removed. Then, TADs were classified according to its percentage of expressed upregulated/downregulated genes ($\log_2\text{FC} > 0$ or $\log_2\text{FC} < 0$). For randomization, we entirely randomized positions of genes within TADs, conserving the same number of genes per TAD. Five hundred iterations were performed to obtain

random distributions, for statistical analysis. Observed/expected plot was created using the ratio real over the mean of the random distribution.

Analysis of the role of TAD boundaries on gene co-regulation and enhancer-gene regulation. To determine the expression correlation between two strictly contiguous genes (x, y), we sorted genes according to their 5'→3' chromosomal order using gene TSSs and then we obtain all expression $\log_2\text{FC}$ pairs ($\log_2\text{FC}(x), \log_2\text{FC}(y)$) for all expressed genes, obtaining 6265 pairs. Then genes were sorted according to $\log_2\text{FC}(x)$ and grouped into 20 bins. For random plots we randomized position of all expressed genes and proceed in a similar way. To analyze the effect of TAD borders we extracted pairs ($\log_2\text{FC}(x), \log_2\text{FC}(y)$), where x and y genes were separated by a TAD border, obtaining 393 pairs. Then genes were sorted according to $\log_2\text{FC}(x)$ and grouped into 20 bins.

To study the effect of TAD borders in enhancer-gene regulation, enhancer-closest gene pairs separated by a TAD border were selected (1230 pairs). Then we extracted $\log_2\text{FC}(\text{H3K27ac})$ of the enhancer and $\log_2\text{FC}(\text{expression})$ of its closest gene that is separated by a TAD border, sorted by $\log_2\text{FC}(\text{expression})$ and grouped into 20 bins. As random borders, we used the borders of random TADs that we describe in the previous section.

Processing Hi-C data. NMuMG TGF β -treated for 8 h HiC data was obtained from GSE96033. Paired-end reads were aligned to the mm9 using bowtie2 (global parameters: `-very-sensitive -L 30 -score-min L,-0.6,-0.2 -end-to-end-reorder`; local parameters: `-very-sensitive -L 20 -scoremin L,-0.6,-0.2 -end-to-end-reorder`) through HiC-Pro software (v2.11.1)¹¹¹. Unmapped reads, non-uniquely mapped reads and PCR duplicates were filtered and uniquely aligned reads were paired. Genome fragment distribution was obtained for DpnII restriction enzyme using as ligation sequence GATCGATC. Raw and ICE-normalized cis-contact matrices were assembled by binning paired reads into uniform 10-kb bins. HiC Bioconductor package (v1.30.0) was used to analyze HiC interactions matrices. To determine intra-TRD interactions, interactions from raw and ICE-normalized matrices for TRD or random TRD coordinates were obtained using `extractRegion()` function. Random TRD coordinates were calculated inverting the start coordinate for each TRD and keeping the same length.

To analyze enhancer-promoter interactions first, we retrieve bins corresponding to TSS and enhancer midpoint coordinates. Then we obtained the interaction for each enhancer-promoter pair for raw 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. ⁵⁰:

$$DI = \left(\frac{B - A}{|B - A|} \right) \cdot \left(\frac{(A - E)^2}{E} + \frac{(B - E)^2}{E} \right), \quad (2)$$

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 1 Mb) 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 motifs 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 software¹¹² was used. CTCF density metaplots was constructed using ChIP-seq and footprints information in TADs and TRDs using proportional number of bin for each structure: 100 bins for TADs (as average size of TAD is 1 Mb) 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 from `Stats` R package. For correlation, the non-parametric Spearman coefficient was calculated and significance was calculated using Fisher exact test. Significance of enrichments were calculated with the Fisher exact test using `fisher.test()` function from R. Probabilities of overlapping were calculated using the hypergeometric distribution using `dhyper()` function from the `stats` R package. Other statistical methods 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.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

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 files 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

- David, C. J. & Massague, J. Contextual determinants of TGFbeta action in development, immunity and cancer. *Nat. Rev. Mol. Cell Biol.* **19**, 419–435 (2018).
- Colak, S. & Ten Dijke, P. Targeting TGF-beta signaling in cancer. *Trends Cancer* **3**, 56–71 (2017).
- Derynck, R. & Budi, E. H. Specificity, versatility, and control of TGF- β family signaling. *Sci. Signal.* **12**, eaav5183 (2019).
- Hill, C. S. Transcriptional Control by the SMADs. *CSH Perspect. Biol.* **8**, a022079 (2016).
- Mullen, A. C. et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**, 565–576 (2011).
- Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577–589 (2011).
- Sundqvist, A. et al. JUNB governs a feed-forward network of TGFbeta signaling that aggravates breast cancer invasion. *Nucleic Acids Res.* **46**, 1180–1195 (2018).
- Ramachandran, 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).
- Fueyo, R. et al. Lineage specific transcription factors and epigenetic regulators mediate TGFbeta-dependent enhancer activation. *Nucleic Acids Res.* **46**, 3351–3365 (2018).
- Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
- Rickels, R. & Shilatifard, A. Enhancer logic and mechanics in development and disease. *Trends Cell Biol.* **28**, 608–630 (2018).
- Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
- Plank, J. L. & Dean, A. Enhancer function: mechanistic and genome-wide insights come together. *Mol. Cell* **55**, 5–14 (2014).
- Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
- Heinz, S., Romanoski, 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).
- Vierbuchen, T. et al. AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. *Mol. Cell* **68**, 1067–1082 e1012 (2017).
- Ostuni, R. et al. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013).
- Nieto, M. A. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science* **342**, 1234850 (2013).
- Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
- Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).
- Beerling, E. et al. Plasticity between epithelial and mesenchymal states unlinks EMT from metastasis-enhancing stem cell capacity. *Cell Rep.* **14**, 2281–2288 (2016).
- Krebs, A. M. et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat. Cell Biol.* **19**, 518–529 (2017).
- Ye, X. & Weinberg, R. A. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol.* **25**, 675–686 (2015).
- Ye, X. et al. Upholding a role for EMT in breast cancer metastasis. *Nature* **547**, E1–E3 (2017).
- Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168**, 670–691 (2017).
- Moustakas, A. & Heldin, C. H. Mechanisms of TGFbeta-Induced Epithelial-Mesenchymal Transition. *J. Clin. Med.* **5**, 63 (2016).
- Miettinen, P. J., Ebner, R., Lopez, A. R. & Derynck, R. TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *J. Cell Biol.* **127**, 2021–2036 (1994).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Chen, X. et al. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
- Ollion, J., Cochenne, J., Loll, F., Escude, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).
- Inman, G. J., Nicolas, F. J. & Hill, C. S. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. *Mol. Cell* **10**, 283–294 (2002).
- Gamulescu, M. A. et al. Transforming growth factor beta2-induced myofibroblastic differentiation of human retinal pigment epithelial cells: regulation by extracellular matrix proteins and hepatocyte growth factor. *Exp. Eye Res.* **83**, 212–222 (2006).
- Moreno-Bueno, G. et al. The morphological and molecular features of the epithelial-to-mesenchymal transition. *Nat. Protoc.* **4**, 1591–1613 (2009).
- Totzek, J. et al. Takinib, a selective TAK1 inhibitor, broadens the therapeutic efficacy of TNF-alpha inhibition for cancer and autoimmune disease. *Cell Chem. Biol.* **24**, 1029–1039 e1027 (2017).
- Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* **4**, E131–E136 (2002).
- Itoh, Y. et al. A comparative analysis of Smad-responsive motifs identifies multiple regulatory inputs for TGF-beta transcriptional activation. *J. Biol. Chem.* **294**, 15466–15479 (2019).
- Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
- Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
- Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403 (2015).
- Bhatt, D. M. et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
- Russ, B. E. et al. Regulation of H3K4me3 at transcriptional enhancers characterizes acquisition of virus-specific CD8(+) T cell-lineage-specific function. *Cell Rep.* **21**, 3624–3636 (2017).
- Chen, K. et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47**, 1149–1157 (2015).
- Zhang, B. et al. A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity. *Genome Res.* **23**, 917–927 (2013).
- Malik, A. N. et al. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **17**, 1330–1339 (2014).
- Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
- Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
- Pascual-Reguant, L. et al. Lamin B1 mapping reveals the existence of dynamic and functional euchromatin lamin B1 domains. *Nat. Commun.* **9**, 3420 (2018).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- 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).
- 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).
- 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).
- 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).

55. Madrigal, P. & Alasoo, K. AP-1 takes centre stage in enhancer chromatin dynamics. *Trends Cell Biol.* **28**, 509–511 (2018).
56. Zancanato, F. et al. Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat. Cell Biol.* **17**, 1218–1227 (2015).
57. Biddie, S. C. et al. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43**, 145–155 (2011).
58. Zhang, Y. E. Non-smad signaling pathways of the TGF-beta family. *Cold Spring Harb. Perspect. Biol.* **9**, a022129 (2017).
59. Courcelles, M. et al. Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions. *Mol. Syst. Biol.* **9**, 669 (2013).
60. Whitmarsh, A. J., Shore, P., Sharrocks, A. D. & Davis, R. J. Integration of MAP kinase signal transduction pathways at the serum response element. *Science* **269**, 403–407 (1995).
61. Murphy, L. O., Smith, S., Chen, R. H., Fingar, D. C. & Blenis, J. Molecular interpretation of ERK signal duration by immediate early gene products. *Nat. Cell Biol.* **4**, 556–564 (2002).
62. Monje, P., Hernandez-Losa, J., Lyons, R. J., Castellone, M. D. & Gutkind, J. S. Regulation of the transcriptional activity of c-Fos by ERK. A novel role for the prolyl isomerase PIN1. *J. Biol. Chem.* **280**, 35081–35084 (2005).
63. Monje, P., Marinissen, M. J. & Gutkind, J. S. Phosphorylation of the carboxyl-terminal transactivation domain of c-Fos by extracellular signal-regulated kinase mediates the transcriptional activation of AP-1 and cellular transformation induced by platelet-derived growth factor. *Mol. Cell Biol.* **23**, 7030–7043 (2003).
64. Soloaga, A. et al. MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J.* **22**, 2788–2797 (2003).
65. Thomson, S. et al. The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. *EMBO J.* **18**, 4779–4793 (1999).
66. Tee, W. W., Shen, S. S., Oksuz, O., Narendra, V. & Reinberg, D. Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. *Cell* **156**, 678–690 (2014).
67. Hamilton, W. B. et al. Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature* **575**, 355–360 (2019).
68. Massague, J. TGFbeta signalling in context. *Nat. Rev. Mol. Cell Biol.* **13**, 616–630 (2012).
69. Maroulakou, I. G. & Muise-Helmericks, R. C. Guide to techniques in creating transgenic mouse models using SV40 T antigen. *Methods Mol. Biol.* **165**, 269–286 (2001).
70. Beyer, T. A. et al. Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells. *Cell Rep.* **5**, 1611–1624 (2013).
71. Baulida, J., Diaz, V. M. & Herreros, A. G. Snail1: a transcriptional factor controlled at multiple levels. *J. Clin. Med.* **8**, 757 (2019).
72. Maturi, V., Moren, A., Enroth, S., Heldin, C. H. & Moustakas, A. Genomewide binding of transcription factor Snail1 in triple-negative breast cancer cells. *Mol. Oncol.* **12**, 1153–1174 (2018).
73. Ye, X. et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
74. Beyes, S. et al. Genome-wide mapping of DNA-binding sites identifies stemness-related genes as directly repressed targets of SNAIL1 in colorectal cancer cells. *Oncogene* **8**, 6647–6661 (2019).
75. Javid, S. et al. Dynamic chromatin modification sustains epithelial-mesenchymal transition following inducible expression of Snail-1. *Cell Rep.* **5**, 1679–1689 (2013).
76. Batlle, E. et al. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* **2**, 84–89 (2000).
77. Cano, A. et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol.* **2**, 76–83 (2000).
78. Vincent, T. et al. A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition. *Nat. Cell Biol.* **11**, 943–950 (2009).
79. Lin, Y. et al. The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1. *EMBO J.* **29**, 1803–1816 (2010).
80. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
81. Downen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
82. Hnisz, D., Day, D. S. & Young, R. A. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* **167**, 1188–1200 (2016).
83. Ji, X. et al. 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
84. Kikuta, H. et al. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545–555 (2007).
85. Soler-Oliva, M. E., Guerrero-Martinez, J. A., Bachetti, V. & Reyes, J. C. Analysis of the relationship between coexpression domains and chromatin 3D organization. *PLoS Comput. Biol.* **13**, e1005708 (2017).
86. Bortoluzzi, S. et al. A comprehensive, high-resolution genomic transcript map of human skeletal muscle. *Genome Res.* **8**, 817–825 (1998).
87. Kosak, S. T. et al. Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol.* **5**, e309 (2007).
88. Hurst, L. D., Pal, C. & Lercher, M. J. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310 (2004).
89. Purmann, A. et al. Genomic organization of transcriptomes in mammals: coregulation and cofunctionality. *Genomics* **89**, 580–587 (2007).
90. Michalak, P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* **91**, 243–248 (2008).
91. Despang, A. et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
92. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
93. Finn, E. H. & Misteli, T. A genome disconnect. *Nat. Genet.* **51**, 1205–1206 (2019).
94. Neems, D. S., Garza-Gongora, A. G., Smith, E. D. & Kosak, S. T. Topologically associated domains enriched for lineage-specific genes reveal expression-dependent nuclear topologies during myogenesis. *Proc. Natl Acad. Sci. USA* **113**, E1691–E1700 (2016).
95. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
96. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
97. 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).
98. He, H. H. et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73–78 (2014).
99. Conrad, T. & Orom, U. A. Cellular fractionation and isolation of chromatin-associated RNA. *Methods Mol. Biol.* **1468**, 1–9 (2017).
100. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell Biol.* **14**, 7219–7225 (1994).
101. Subtil-Rodriguez, A. & Reyes, J. C. BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. *EMBO Rep.* **11**, 751–757 (2010).
102. Ceballos-Chavez, M. et al. Control of neuronal differentiation by sumoylation of BRAF35, a subunit of the LSD1-CoREST histone demethylase complex. *Proc. Natl Acad. Sci. USA* **109**, 8085–8090 (2012).
103. Lun, A. T. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45 (2016).
104. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
105. Bailey, T. L. & Machanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* **40**, e128 (2012).
106. 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 (2015).
107. 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).
108. 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).
109. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
110. Law, C. W. et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* **5**, ISCB Comm J-1408 (2016).
111. Servant, N. et al. Hi-C-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
112. Sherwood, R. I. et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (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 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 assistance. This work was funded by the Spanish Ministry of Economy and Competitiveness (BFU2014-53543-P and BFU2017-85420-R to J.C. Reyes), the Junta de Andalucía (BIO-321), “Fundación Vencer El Cáncer” (VEC) and the European Union (FEDER). CABIMER is a Center partially funded by the Junta de Andalucía. S.P. was

supported by the FERO Foundation, La Caixa Foundation (LCF/PR/PR12/51070001) and Cellex Foundation.

Author contributions

Conceptualization and methodology, J.A.G.-M. and J.C.R.; Investigation, J.A.G.-M. and M.C.-C.; 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020