The TGFB2-AS1 lncRNA Regulates TGF-beta Signaling by Modulating Corepressor Activity

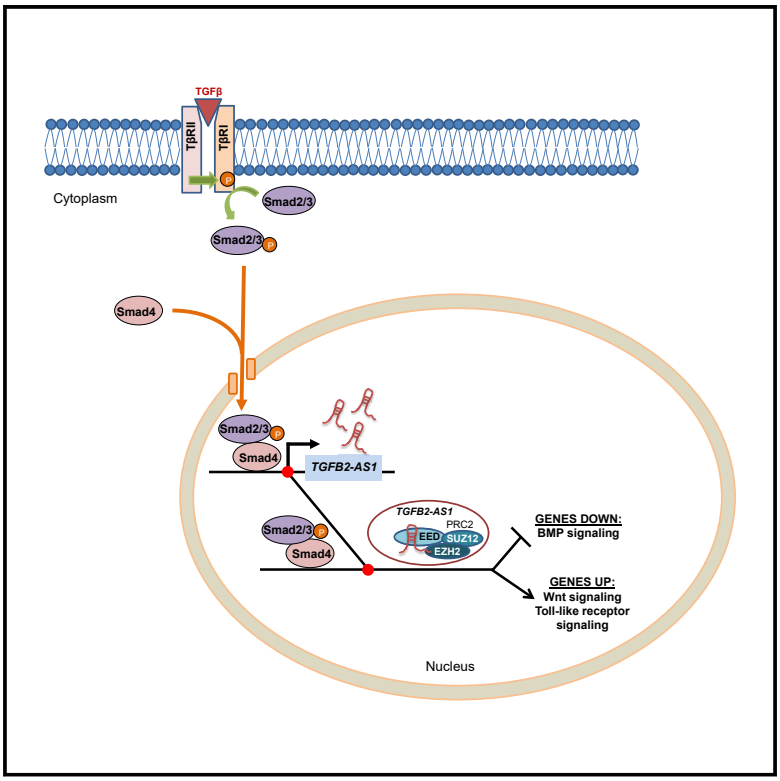
Panagiotis Papoutsoglou, Yutaro Tsubakihara, Laia Caja, Anita Moren, Paris Pallis, Adam Ameur, Carl-Henrik Heldin, Aristidis Moustakas

Article

Cell Reports

The *TGFB2-AS1* lncRNA Regulates TGF-β Signaling by Modulating Corepressor Activity

Graphical Abstract



Authors

Panagiotis Papoutsoglou, Yutaro Tsubakihara, Laia Caja, ..., Adam Ameur, Carl-Henrik Heldin, Aristidis Moustakas

Correspondence

aris.moustakas@imbim.uu.se

In Brief

Papoutsoglou et al. show that TGFB2-antisense RNA1 (*TGFB2-AS1*) is induced by TGF-β, interacts with the EED adaptor of the Polycomb repressor complex 2, and limits the response of target genes to TGF-β signaling.

Highlights

- TGF-β signaling transcriptionally regulates lncRNAs that regulate TGF-β signaling
- *TGFB2-AS1* is induced by TGF-β to negatively regulate Smad transcriptional output
- *TGFB2-AS1* associates with EED, the Polycomb repressor complex 2 adaptor
- *TGFB2-AS1* promotes repressive histone modifications at TGF-β-target genes

Papoutsoglou et al., 2019, Cell Reports *28*, 3182–3198
September 17, 2019
https://doi.org/10.1016/j.celrep.2019.08.028

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

The *TGFB2-AS1* lncRNA Regulates TGF-β Signaling by Modulating Corepressor Activity

Panagiotis Papoutsoglou,[1,3] Yutaro Tsubakihara,[1] Laia Caja,[1] Anita Morén,[1] Paris Pallis,[1,4] Adam Ameur,[2] Carl-Henrik Heldin,[1] and Aristidis Moustakas[1,5,*]

[1]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, and Ludwig Cancer Research Box 582, Biomedical Center, Uppsala University, 751 23 Uppsala, Sweden

[2]Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Box 256, Uppsala University, 751 05 Uppsala, Sweden

[3]Present address: INSERM, University of Rennes, Inra, Institut NuMeCan (Nutrition Metabolisms and Cancer), UMR_S 1241, 35033, Rennes, France

[4]Present address: Department of Microbiology and Immunobiology, 77 Avenue Louis Pasteur, Harvard Medical School, Boston, MA 02115, USA

[5]Lead Contact

*Correspondence: aris.moustakas@imbim.uu.se

https://doi.org/10.1016/j.celrep.2019.08.028

SUMMARY

Molecular processes involving lncRNAs regulate cell function. By applying transcriptomics, we identify lncRNAs whose expression is regulated by transforming growth factor β (TGF-β). Upon silencing individual lncRNAs, we identify several that regulate TGF-β signaling. Among these lncRNAs, TGFB2-antisense RNA1 (*TGFB2-AS1*) is induced by TGF-β through Smad and protein kinase pathways and resides in the nucleus. Depleting *TGFB2-AS1* enhances TGF-β/Smad-mediated transcription and expression of hallmark TGF-β-target genes. Increased dose of *TGFB2-AS1* reduces expression of these genes, attenuates TGF-β-induced cell growth arrest, and alters BMP and Wnt pathway gene profiles. Mechanistically, *TGFB2-AS1*, mainly via its 3′ terminal region, binds to the EED adaptor of the Polycomb repressor complex 2 (PRC2), promoting repressive histone H3K27me^3 modifications at TGF-β-target gene promoters. Silencing EED or inhibiting PRC2 methylation activity partially rescues *TGFB2-AS1*-mediated gene repression. Thus, the TGF-β-induced *TGFB2-AS1* lncRNA exerts inhibitory functions on TGF-β/BMP signaling output, supporting auto-regulatory negative feedback that balances TGF-β/BMP-mediated responses.

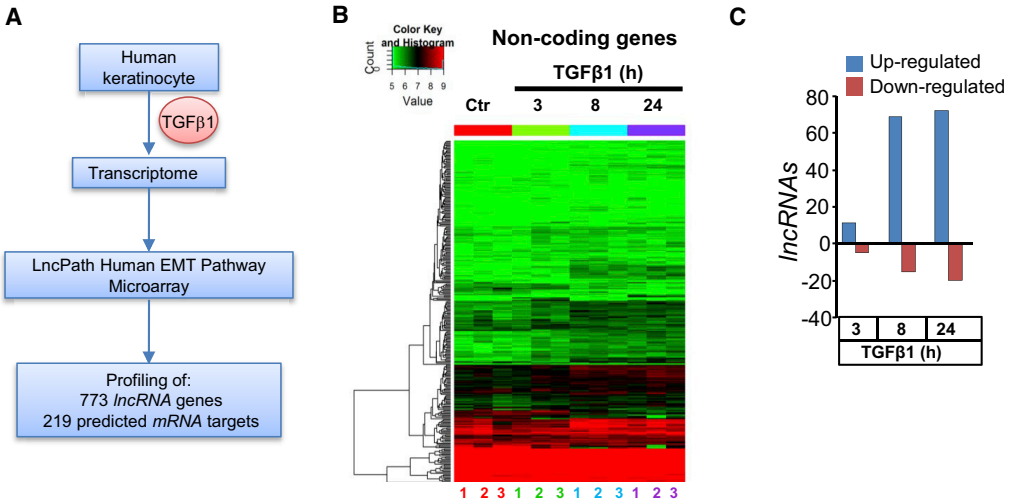
INTRODUCTION

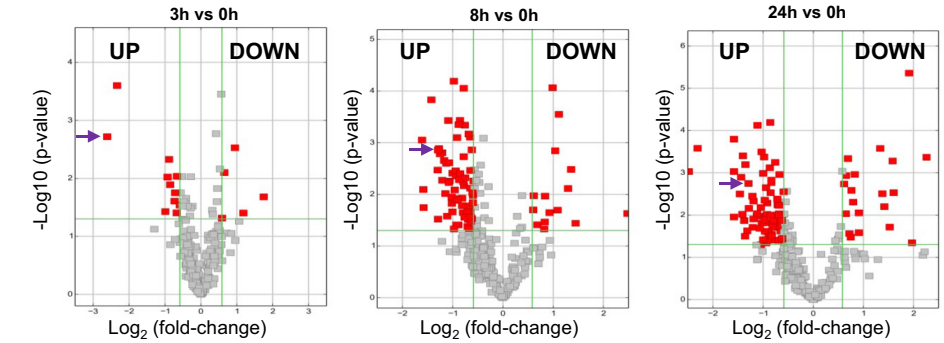
The transforming growth factor β (TGF-β) signaling pathway is conserved in metazoans and controls embryonic tissue morphogenetic potential and adult tissue homeostasis (Morikawa et al., 2016). TGF-β acts in different cells by regulating the expression of 300–500 protein-coding genes (David and Massagué, 2018). Misregulation of tissue-specific gene programs controlled by TGF-β is often seen in diseases, including immunological dysregulation, fibrotic tissue malfunction, and malignant progression of tumors (David and Massagué, 2018; Meng et al., 2016; Pickup et al., 2017).

TGF-β binds to type I (TβRI) and type II (TβRII) protein kinase receptors, which induce phosphorylation and activation of Smad2/Smad3. Smad2/Smad3 form complexes with Smad4 and induce or repress target genes (David and Massagué, 2018; Morikawa et al., 2016). TGF-β receptors also activate MAP-kinase (MAPK), phosphatidyl-inositol-3′ kinase, and other signaling pathways.

Regulation of protein-coding genes by TGF-β explains several biological effects. Recent examples highlight the significance of non-coding RNA genes, encoding microRNAs (miRNAs) or long non-coding RNAs (lncRNAs) (Meng et al., 2016; Wang et al., 2016). lncRNAs are longer than 200 nt and lack protein-coding capacity because of short or suboptimal open reading frames (Tordonato et al., 2015); they are transcribed by RNA polymerase II from intergenic or gene regions, generating in the latter case antisense transcripts relative to an mRNA (Pelechano and Steinmetz, 2013). Bidirectional lncRNAs are transcribed from intronic regions or in the opposite direction relative to a coding gene (Tordonato et al., 2015). Functionally, many lncRNAs scaffold ribonucleoprotein complexes that regulate transcription, chromatin remodeling, splicing, and mRNA translation (Rinn and Chang, 2012).

The TGF-β-induced *HOTAIR*, *H19*, *MALAT1*, *lncRNA-ATB*, and *lncRNA-HIT* regulate epithelial-mesenchymal transition (EMT), stemness, and tumor cell metastasis (Fan et al., 2014; Matouk et al., 2014; Pádua Alves et al., 2013; Richards et al., 2015; Yuan et al., 2014). *HOTAIR* and *MALAT1* associate with the histone H3K27 methyl-transferase Polycomb repressor complex 2 (PRC2) and repress epithelial gene expression (e.g., *CDH1*) (Fan et al., 2014; Pádua Alves et al., 2013). Other lncRNAs regulate TGF-β signaling, yet their expression is not regulated by TGF-β. For example, *MEG3* forms triple helices with Smad genes, repressing their expression and limiting TGF-β activity (Mondal et al., 2015). *GAS5*, via double-stranded Smad-binding RNA elements, associates with and prohibits Smad3 from reaching its DNA targets, thus blocking mesenchymal differentiation (Tang et al., 2017). Cytoplasmic *NORAD* associates with importin-β, which carries Smad3 to





E

Top up-regulated genes (3 vs 0 h)	
Name	Fold-induction
TGFB2-AS1	6.1
XLOC_009485	5.0
CTD-2022H16.3	2.0
MIR100HG	1.9

Top up-regulated genes (8 vs 0 h)	
Name	Fold-induction
MIR100HG	3.0
RPL31P56	3.0
RP11-106A1.3	2.98
AK126698	2.7

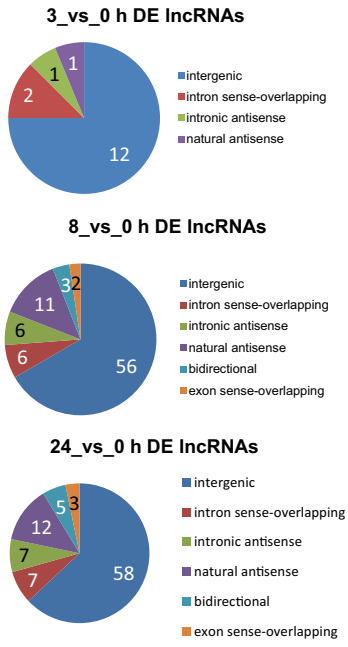
Top up-regulated genes (24 vs 0 h)	
Name	Fold-induction
VIM-AS1	5.6
XLOC_000587	4.9
AK126698	3.0
RP11-520A21.1	3.0

Top down-regulated genes (3 vs 0 h)	
Name	Fold-reduction
RP11-288G3.4	3.4
RP11-19B4.1	2.3
RP11-134K13.2	1.9
RP3-481F12.1	1.6

Top down-regulated genes (8 vs 0 h)	
Name	Fold-reduction
RP11-986E7.6	5.6
HSPD1P1	2.7
HMGN2P5	2.6
GUSBP1	2.4

Top down-regulated genes (24 vs 0 h)	
Name	Fold-reduction
HMGN2P5	4.8
HSPD1P1	3.9
RP1-197O17.2	3.8
LINC00707	3.0

F



(legend on next page)

the nucleus, enhancing TGF-β signaling and EMT (Kawasaki et al., 2018).

Thus, TGF-β responses can be mediated by diverse non-coding RNAs and motivate identification of lncRNAs that control the TGF-β output. After screening for lncRNAs whose expression is regulated by TGF-β and analyzing a subset, whose feedback could regulate TGF-β signaling, we identified *TGFB2-AS1*, a head-to-head antisense nuclear lncRNA, which associates with the PRC2 adaptor protein EED, thus facilitating H3K27 tri-methylation and repression of specific TGF-β-target genes.

RESULTS

TGF-β Preferentially Regulates Intergenic and Natural Antisense lncRNAs

Human HaCaT keratinocytes are a resource for TGF-β signaling studies (Gomis et al., 2006). Previous transcriptomic studies pointed to the time-dependent response of epithelial genomes to TGF-β signaling and defined immediate-early (0–3 h), immediate-intermediate (3–8 h), and indirect-late (8–72 h) responses (Kang et al., 2003; Valcourt et al., 2005; Zavadil et al., 2001). Accordingly, HaCaT cells were stimulated with TGF-β1 (abbreviated as TGF-β) for 3, 8, and 24 h, followed by transcriptomic analysis using a microarray platform enriched in lncRNA and selected mRNA genes; some of the 773 lncRNA genes queried were bioinformatically predicted to regulate the EMT (Figure 1A). Time-dependent response to TGF-β followed a gradual enrichment in expressed genes between 3 and 24 h post-TGF-β stimulation (Figures 1B–1D; Tables S1 and S2). lncRNAs exhibiting immediate-early response to TGF-β were the *TGFB2-AS1* (upregulated) and *RP11-288G3.4* (downregulated), the immediate-early and intermediate responding *MIR100HG* (upregulated), the intermediate-to-late responding *HMGN2P5* (downregulated), and the late responding *VIM-AS1* (upregulated) (Figure 1E). Genome-wide location analysis indicated that the responding lncRNA genes were primarily intergenic and natural antisense genes (Figure 1F). The time-dependent profile of lncRNA gene expression was superimposable to the protein-coding gene profile (Figure S1A; Table S2), which included up- or downregulated mRNAs (Figure S1B). Top examples were previously established mRNAs, such as the immediate-early and intermediate *WNT11* (upregulated), the indirect-late responding *FN1* (upregulated), the immediate-early and intermediate *PTK2* (downregulated), and the indirect-late responding *S100A4* (downregulated) (Figure S1C). Several lncRNA genes were validated by independent qRT-PCR assays using gene-specific primers (Table S3), confirming the accuracy of the microarray-based screen (Figures S1D and S1E). The transcriptomic assay revealed a wealth of lncRNA genes being regulated by TGF-β.

Several TGF-β-Regulated lncRNAs Regulate Smad-Dependent Transcription

We selected a subset of lncRNAs whose expression was induced by TGF-β and designed short interfering RNAs (siRNAs) targeting each lncRNA (Table S4). In pairwise silencing experiments, non-targeting along with lncRNA-specific siRNAs were transfected, and their impact was monitored using the transcriptional output of the $CAGA_{12}$-luc reporter that is sensitive to TGF-β-mediated Smad2/Smad3/Smad4 signaling (Figure 2A). Twenty-three TGF-β-induced lncRNAs were analyzed (Figure 2B) and a positive control siRNA targeting TβRII (*TGFBR2*). Silencing *TGFBR2* reduced luciferase activity by 4- to 5-fold (Figure 2C). Among the 23 lncRNAs (Figure 2B), silencing 17 resulted in enhanced reporter activity, silencing 2 had neutral effects, and silencing 4 resulted in reduced reporter activity (Figure 2C). Silencing of *TGFB2-AS1* resulted in the highest fold enhancement of reporter activity (similar to *MIAT* and *VIM-AS1*; Figure 2C), a highly reproducible effect in independent reporter assays after transient *TGFB2-AS1* silencing (Figure S2A).

To bypass cell-type-specific effects and siRNA-based bias, we designed individual short hairpin RNA (shRNA) vectors and generated stably transfected pools of HaCaT and A549 human lung adenocarcinoma cells. Two representative pools of transfected cells with two independent shRNAs targeting different regions of the human *TGFB2-AS1* lncRNA (Figure 2D) per cell type were analyzed deeper. Expression of endogenous *TGFB2-AS1* was stably and strongly suppressed (Figures S2B and S2C). Scoring for the TGF-β/Smad-sensitive reporter reproducibly showed enhanced activity (Figures 2E and 2F). We conclude that TGF-β induces expression of many lncRNAs; some act in a feedback mechanism and positively or negatively regulate TGF-β/Smad transcriptional output. Seventeen of 23 lncRNAs analyzed exhibited negative regulation of the pathway, and *TGFB2-AS1* was among those with the highest negative effects.

Smad and Protein Kinase Inputs Induce Expression of Nuclear *TGFB2-AS1*

TGFB2-AS1 is on chromosome 1 (https://www.ncbi.nlm.nih.gov/gene/728463 and https://lncipedia.org/db/transcript/TGFB2-AS1:2; Volders et al., 2019), transcribed in an antisense orientation relative to *TGFB2*, encompasses three exons and two introns, generating a 557-nt-long RNA (Figure 2D), with low protein-coding capacity (see Method Details). The first *TGFB2-AS1* exon overlaps with the 5′ UTR of *TGFB2* (Figure 2D). *TGFB2* encodes for one of the three isoforms of TGF-β, which has unique biological functions, such as induction of heart mesenchyme differentiation, but *in vitro* signals indistinguishably from TGF-β1 or TGF-β3 (Morikawa et al., 2016). Time course experiments revealed that *TGFB2-AS1* responded with immediate-early kinetics to all three TGF-β isoforms but not to stimulation by

Figure 1. TGF-β Regulates lncRNA Expression in HaCaT Cells

(A) Schematic outline of the experimental design.

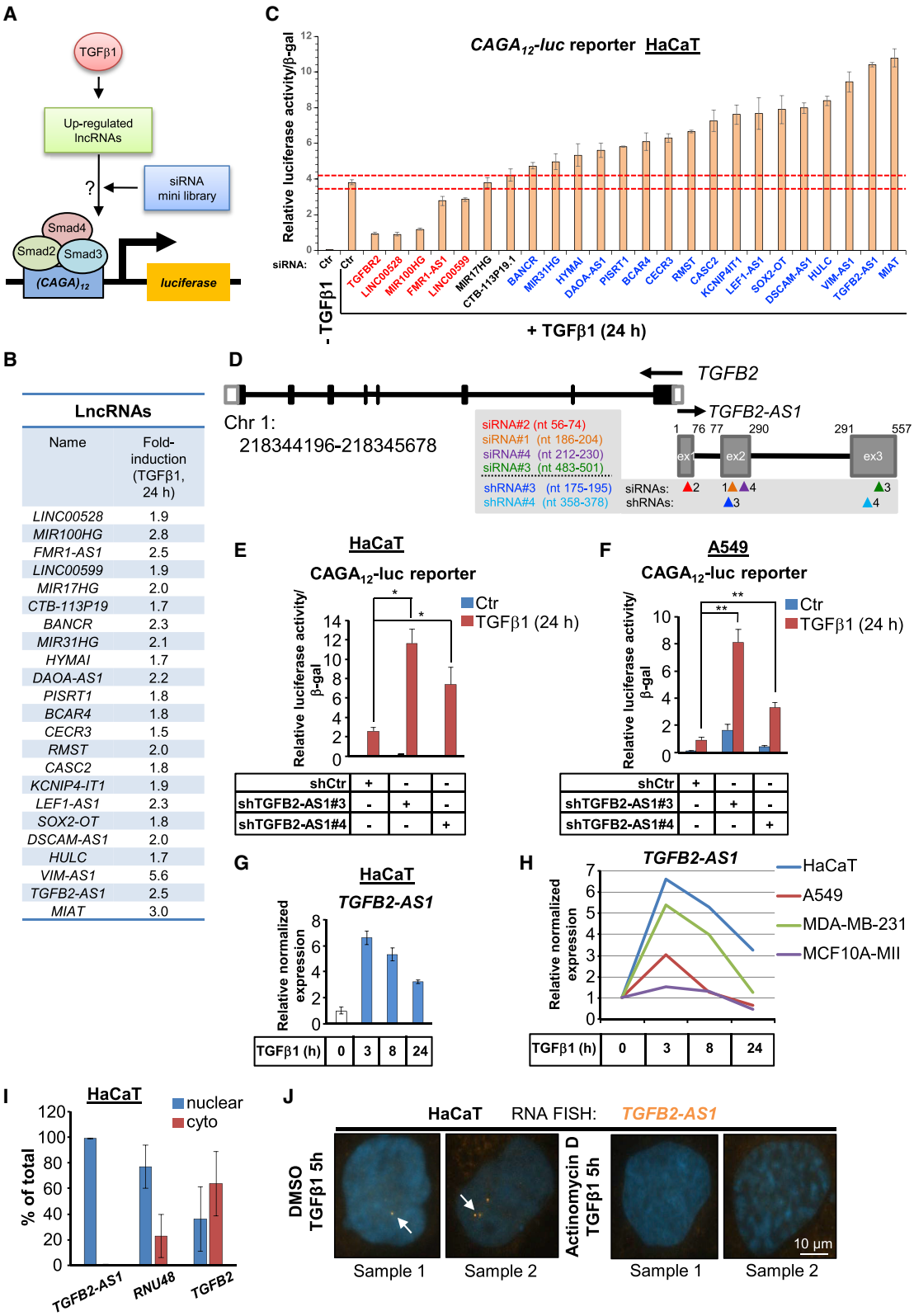
(B) Heatmap of the lncRNAs examined in the microarray in response to TGF-β treatment, with triplicate samples (1–3) per condition.

(C) Number of TGF-β-regulated lncRNAs at different time periods of TGF-β stimulation.

(D) Volcano plots of lncRNA expression. Red dots represent significantly regulated lncRNAs. Arrows mark the *TGFB2-AS1*.

(E) The four top-regulated lncRNAs in response to TGF-β treatment for each time period.

(F) Classification of TGF-β-regulated lncRNAs according to genomic localization relative to neighboring protein-coding genes.



(legend on next page)

bone morphogenetic protein (BMP) ligands (Figures 2G and S2D–S2F). *TGFB2-AS1* was induced by TGF-β in all normal and cancer cell lines tested, albeit with different amplitude but reproducible kinetics (Figure 2H). Cell fractionation followed by qRT-PCR analysis indicated presence of *TGFB2-AS1* in the nuclear fraction with undetectable levels in the cytoplasm (Figure 2I). *RNU48* ncRNA control was enriched in the nucleus, where it exhibits its biological function, whereas *TGFB2* mRNA partitioned roughly equally to cytoplasm and nucleus (Figure 2I). Fluorescent *in situ* hybridization (FISH) analysis after TGF-β stimulation confirmed the pattern of two to four distinct *TGFB2-AS1* spots per individual nuclear section (Figure 2J). The RNA polymerase II inhibitor actinomycin D abolished all RNA signals from the cells, demonstrating dependency on RNA polymerase activity (Figure 2J).

TGF-β signals via TGF-β receptor kinases, Smads, and MAPKs. Chemical inhibition of TβRI kinase activity by GW6604 blocked *TGFB2-AS1* induction, implicating canonical TGF-β receptor signaling (Figure 3A). siRNA-mediated silencing of Smad4 or Smad3 partially inhibited induction, demonstrating involvement of Smads (Figures 3B and 3C). Inhibitors against three MAPK pathways, MEK1, p38, and c-Jun N-terminal kinase (JNK), showed that inhibiting p38 and JNK reduced *TGFB2-AS1* induction (Figure 3D), implicating these two MAPKs together with Smads in the transcriptional regulation of *TGFB2-AS1*.

We cloned the human *TGFB2-AS1* cDNA from HaCaT cells in both orientations, generating vectors that express identical size (557 nt) RNA transcripts, one corresponding to *TGFB2-AS1* and the other to the anti-parallel RNA strand, named *anti-TGFB2-AS1* (Figures S2G and S2H). Upon transient transfection, *TGFB2-AS1* was highly expressed (Figure 3E) and resulted in many large nuclear spots (Figure 3F), demonstrating that overexpression retained proper subcellular distribution. In agreement with the loss-of-function experiments (Figures 2 and S2), overexpression of *TGFB2-AS1* repressed TGF-β-induced promoter reporter activity (Figure 3G). In order to bypass transient transfection variability issues, we selected individual stably transfected HaCaT clones and confirmed that transgenic *TGFB2-AS1* was expressed 4 to 8 times higher relative to induction of endogenous *TGFB2-AS1* (Figures S2I and S2J). The promoter-reporter assay confirmed robust repression of TGF-β/Smad-dependent transcriptional output in such clones (Figure 3H). We verified the specificity of the effects of *TGFB2-AS1* expression on promoter-reporter activity by simultaneously silencing *TGFB2-AS1* in one of the stable clones (Figure 3I); silencing endogenous *TGFB2-AS1* enhanced the Smad transcriptional response (as shown before, Figure 2), whereas silencing the overexpressed *TGFB2-AS1* resulted in basal promoter response to TGF-β (Figure 3I).

A simple interpretation of the above results is that *TGFB2-AS1* may regulate Smad signaling in a direct manner. Smad3 is the major mediator of TGF-β signaling and direct activator of the $CAGA_{12}$-luciferase reporter. However, *TGFB2-AS1* showed no observable effect on Smad3 accumulation in the nucleus upon brief TGF-β stimulation (Figure S3A). Note that the actin staining in these experiments helps identify individual cells in the population that responds to TGF-β but does not rearrange, because of short stimulation. Furthermore, chromatin immunoprecipitation (ChIP) of endogenous Smad2/3 readily revealed binding of the Smad complex on a well-established TGF-β-target gene enhancer, *SERPINE1*, and overexpression of *TGFB2-AS1* did not reduce such binding (Figure S3B). These data support the notion that *TGFB2-AS1* is induced by Smad-MAPK signaling modules, shows nuclear localization, and negatively regulates TGF-β transcriptional responses, without interfering directly with Smad pathway activation.

TGFB2-AS1 Negatively Regulates Biological Responses to TGF-β Signaling

The impact of *TGFB2-AS1* on the Smad-sensitive reporter (Figures 2 and 3) prompted analysis of physiological responses to TGF-β. Because *TGFB2-AS1* is a head-to-head antisense transcript relative to the *TGFB2* gene (Figure 2D), we analyzed the impact on expression of *TGFB2* mRNA (Figure S3C). TGF-β stimulation for 3 h led to a significant upregulation of *TGFB2* mRNA, an effect not observed after 8 or 24 h stimulation (Figure S1C). Silencing of endogenous *TGFB2-AS1* led to a small but not significant reduction of *TGFB2* mRNA levels (Figures S3C and S3D).

We then analyzed cell-cycle arrest, a hallmark response of epithelial cells to TGF-β. *TGFB2-AS1* overexpression significantly limited the number of cells arrested by TGF-β analyzed

Figure 2. *TGFB2-AS1* Regulates TGF-β/Smad Transcriptional Responses

(A) Illustration of the experimental approach.

(B) The 23 TGF-β-induced lncRNAs analyzed for impact on TGF-β signaling (C). Fold-induction in response to TGF-β treatment for 24 h.

(C) Impact of silencing individual TGF-β-induced lncRNAs on the transcriptional activation of the $CAGA_{12}$-luciferase reporter by TGF-β1. lncRNAs with a positive impact on TGF-β signaling are coded in red, without significant impact are coded in black, and with a negative impact are coded in blue letters.

(D) Genomic organization of *TGFB2* and *TGFB2-AS1* (out of scale relative to *TGFB2*) genes. Exons shown in boxes, introns shown in lines and arrows indicate the direction of transcription. The *TGFB2-AS1* transcriptional unit coordinates on the H19 genome sequence (chromosome 1) are shown along with nucleotide (nt) numbering and coordinates of four siRNAs and two shRNAs used in the study, marked by colored and numbered arrowheads.

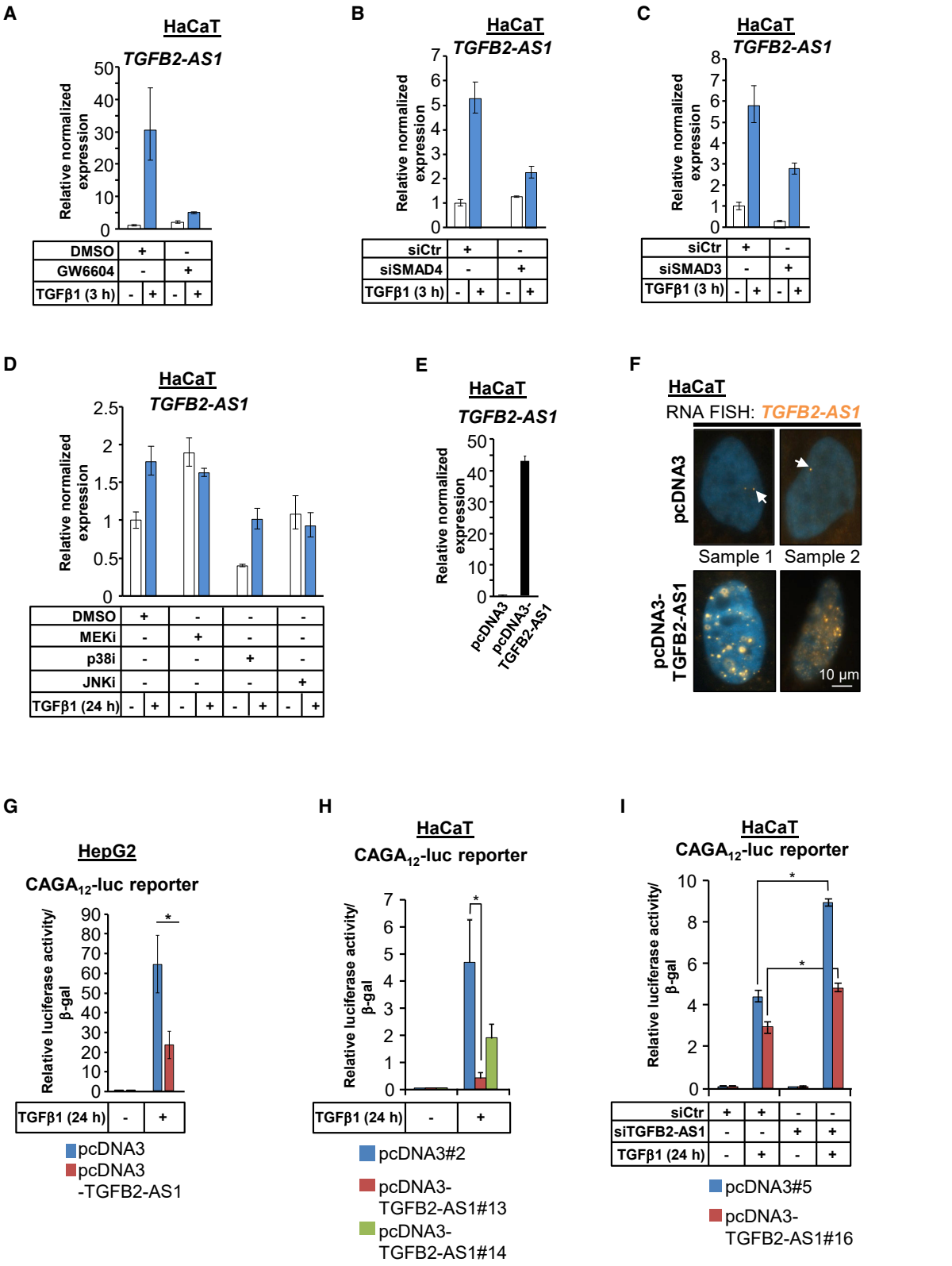
(E and F) TGF-β-responsive $CAGA_{12}$-luciferase reporter assay in HaCaT (E) or A549 (F) cells stably transfected with *shTGFB2-AS1* and treated with TGF-β1 for 24 h. Error bars represent SD from three different experiments (*$p < 0.05$ and **$p < 0.01$).

(G) Quantitative real-time PCR to determine *TGFB2-AS1* expression in response to TGF-β1 stimulation for the indicated time periods in HaCaT cells. Error bars represent SD from three different experiments.

(H) Expression levels of *TGFB2-AS1* in HaCaT, A549, MDA-MB-231, or MCF10A-MII cells in response to TGF-β stimulation for the indicated time periods. Average values from three different experiments.

(I) Expression levels of *TGFB2-AS1*, *RNU48*, and *TGFB2* RNAs in nuclear and cytoplasmic fractions of HaCaT lysates. Error bars represent SD from three different experiments.

(J) RNA FISH for *TGFB2-AS1* in HaCaT cells treated with actinomycin D or DMSO, in combination with TGF-β stimulation for 5 h. Two representative images (samples) out of three independent experiments are shown along with a magnification bar. Arrows point to distinct nuclear RNA puncta.



(legend on next page)

using a thymidine incorporation assay (Figure S4A). Furthermore, using A549 cells that undergo EMT in response to TGF-β, we observed enhanced, more spindle-formed, TGF-β-induced mesenchymal cell morphology, with intense actin stress fibers and much weaker cell-cell contacts, upon silencing of *TGFB2-AS1* (Figure S4B). This result was confirmed by molecular marker analysis, whereby upon *TGFB2-AS1* silencing, A549 cells responded with a much stronger induction of the extracellular matrix components SERPINE1/PAI-1 and FN1, and the negative TGF-β pathway regulator Smad7 (Figure S4C). Because EMT correlates with cancer cell invasion, we analyzed A549 cell invasion through laminin (a physiological basement membrane protein) or not (Figure S4D). TGF-β enhanced invasive movement under both conditions, and silencing of *TGFB2-AS1* was sufficient to increase cell invasion further (Figure S4D). Yet when *TGFB2-AS1* silencing was combined with TGF-β stimulation, we observed the expected trend but not a significant extra increase in invasiveness, possibly reflecting an upper limit of the sensitivity of this assay. In order to confirm that the above effects on EMT and endogenous gene expression were not cell type dependent, the same genes were analyzed in the keratinocyte models with stable *TGFB2-AS1* silencing; enhanced and earlier time-dependent response of PAI-1 and FN1 protein (Figure S4E) and mRNA (Figure S4F) were observed, compared with control keratinocytes expressing non-specific shRNA. Conversely, stable *TGFB2-AS1* overexpression suppressed TGF-β-induced *SERPINE1* and *Smad7* mRNA levels (Figure S4G). Finally, upon stimulation for 72 h, mesenchymal differentiation was observed in the keratinocytes; when *TGFB2-AS1* was stably overexpressed, a limited morphological and actin reorganization response was confirmed, as cells formed fewer stress fibers and remained more adherent to one another (Figure S4H). Overall, these cell-based assays support a negative regulatory role of *TGFB2-AS1* on multiple physiological responses to TGF-β.

BMP Pathway Genes Are Regulated by *TGFB2-AS1*

Because TGF-β exhibits complex and context-dependent modes of regulation of target genes, it was formally possible that *TGFB2-AS1* acted as a negative regulator only for a subset of responses to TGF-β. A robust way to investigate this possibility is to analyze genome-wide transcriptomic responses of cells to TGF-β. Using HaCaT cells stably expressing *TGFB2-AS1*, we performed AmpliSeq RNA sequencing before and after TGF-β stimulation (Figure 4; Table S5). Heatmaps of gene expression profiles revealed clear patterns of change. The higher inter-similarity of the TGF-β-treated samples relative to the untreated samples is clearly observed by the hierarchical clustering in the heatmap (Figure 4A). Expressing *TGFB2-AS1* resulted in repression of many genes and induction of genes previously not expressed in the keratinocytes, while not altering other genes (Figure 4A). Thus, *TGFB2-AS1* can affect gene expression positively and negatively. Stimulating the cells with TGF-β for 24 h, resulted in induction of 216 and repression of 206 genes (Figure 4B), whereas combining TGF-β stimulation with ectopic *TGFB2-AS1* expression revealed that *TGFB2-AS1* inhibited several TGF-β-mediated gene responses (positive or negative; Figure 4A), as predicted from previous results (Figures 2 and 3), but also enhanced some TGF-β-mediated gene responses (Figure 4A), suggesting positive contribution of *TGFB2-AS1* to TGF-β responses. These results are presented cumulatively in Figure 4C and Table S5. There is an overlap between genes regulated by TGF-β and *TGFB2-AS1* alone, indicating that not only *TGFB2-AS1* can counteract TGF-β action, but it can cooperate with TGF-β signaling in order to induce or repress a subset of genes.

Gene Ontology analysis of the differentially regulated genes (Figures 4D–4G, S5A, and S5B) revealed that genes involved in macrophage and Toll-like receptor signaling (*TLR1*, *TLR2*, *TLR4*, *TLR6*, *IRAK2*, *IKBKG*, and *NR1H3*; Figures 4D and 4E) were upregulated. These gene sets were described under multiple ontology groupings, including pattern recognition, lipopolysaccharide signaling, and responses to bacterial or infectious agents (all representing Toll-like receptor pathways). In addition, BMP pathway genes, also members of the TGF-β family (*BMP3*, *BMP7*, *GDF6*, *SOST*, and *DLX5*; Figures 4F and 4G), were downregulated in *TGFB2-AS1*-overexpressing cells, in the presence of TGF-β stimulation. The ontology terms neuronal, kidney, and urogenital development included, among other, these BMP pathway genes. Moreover, *TGFB2-AS1*-overexpressing cells, in the absence of TGF-β, exhibited elevated levels of Wnt pathway genes and decreased levels of genes implicated in protein lipidation, epidermal development, and BMP signaling, compared with control cells (Figures S5A and S5B). TGF-β treatment of control cells led to upregulation of genes involved mainly in extracellular matrix organization (Figure S5A), whereas genes

Figure 3. Regulation of *TGFB2-AS1* Expression by TGF-β Signaling

(A) Quantitative real-time PCR to determine *TGFB2-AS1* expression in HaCaT cells treated or not with the TGF-β receptor I inhibitor GW6604, with or without TGF-β stimulation for 3 h.

(B and C) Quantitative real-time PCR to determine *TGFB2-AS1* expression in HaCaT cells transiently transfected with siRNA targeting *SMAD4* (B) or *SMAD3* (C), with or without TGF-β stimulation for 3 h.

(D) Quantitative real-time PCR to determine *TGFB2-AS1* expression in HaCaT cells treated with MEK (PD184352), p38 (SB203580), or JNK (SP600125) inhibitors (i), with or without TGF-β stimulation for 24 h. Error bars represent SD from three different experiments.

(E) Quantitative real-time PCR for *TGFB2-AS1* expression in HaCaT cells transiently transfected with the pcDNA3-TGFB2-AS1 vector.

(F) RNA FISH for *TGFB2-AS1* in HaCaT cells transiently transfected with the pcDNA3-TGFB2-AS1 vector. Arrows point to individual endogenous *TGFB2-AS1* RNA puncta. A magnification bar is also shown. Two representative images (samples) out of three independent experiments are shown.

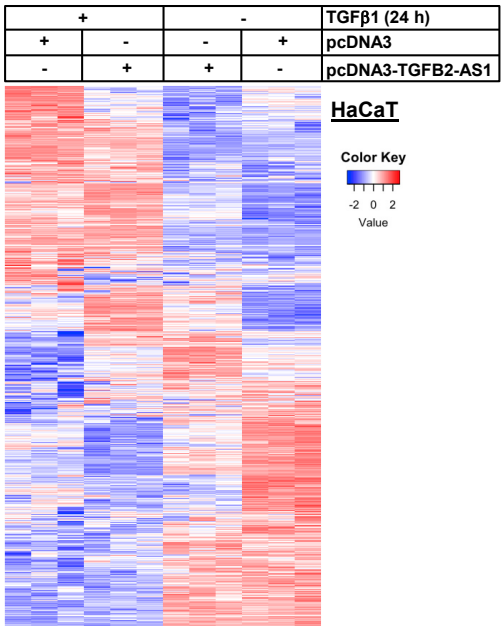
(G) Reporter $CAGA_{12}$-luciferase assay in HepG2 cells transiently transfected with pcDNA3-TGFB2-AS1 and stimulated with TGF-β1 for 24 h.

(H) TGF-β-responsive $CAGA_{12}$-luciferase reporter assay in HaCaT cells stably overexpressing *TGFB2-AS1* and treated with TGF-β1 for 24 h.

(I) TGF-β-responsive $CAGA_{12}$-luciferase reporter assay in HaCaT cells stably overexpressing *TGFB2-AS1*, transiently transfected with *siTGFB2-AS1*, and treated with TGF-β1 for 24 h.

In (G)–(I), error bars represent SD from three different experiments (*$p < 0.05$).

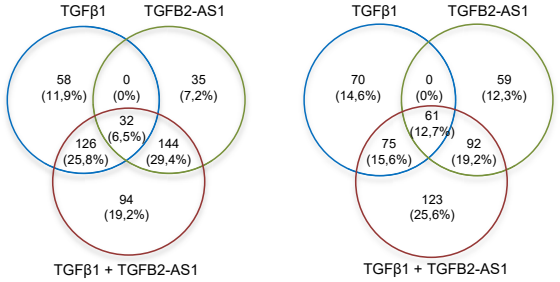
A



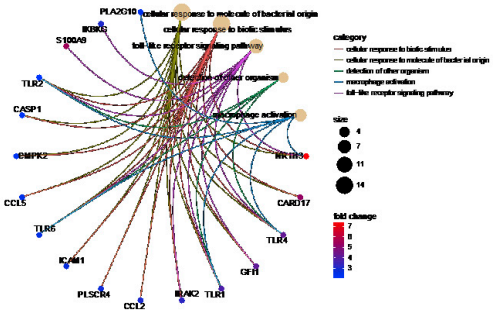
B

Ctr group	Exp. group	Up	Down
pcDNA3	pcDNA3 +TGFβ1	216	206
pcDNA3	TGFB2-AS1	211	212
pcDNA3	TGFB2-AS1 +TGFβ1	396	351
TGFβ1	TGFB2-AS1 +TGFβ1	207	142

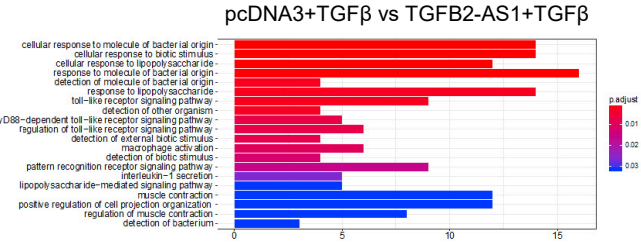
C



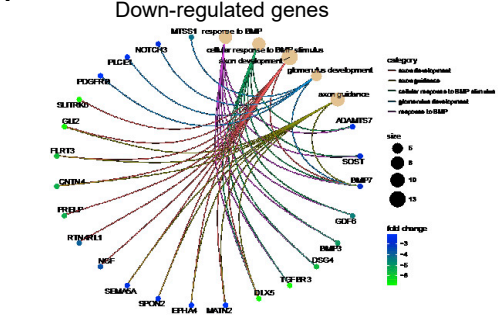
D



E



F



G

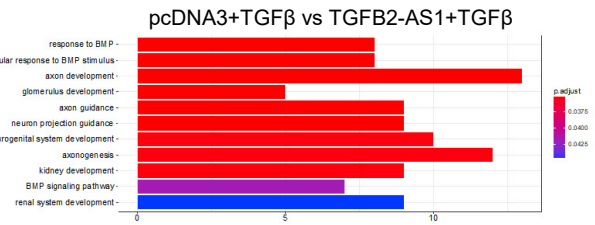


Figure 4. *TGFB2-AS1* Regulates Multiple Signaling Pathways

(A) Heatmap of hierarchically clustered mRNA expression in pcDNA3, pcDNA3+TGF-β, pcDNA3-*TGFB2-AS1*, and pcDNA3-*TGFB2-AS1*+TGF-β cells. Columns are clustered on the basis of similarity of average gene expression among all differentially expressed genes per condition.

(B) Total number of up- or downregulated genes in HaCaT cells transfected or not with *TGFB2-AS1* and stimulated or not with TGF-β for 24 h.

(C) Venn diagrams representing overlapping up- or downregulated genes between the experimental conditions.

(legend continued on next page)

related to cornification and keratinocyte differentiation were repressed (Figure S5B). Validation of genes with robust statistical significance based on the transcriptomic analysis, using qRT-PCR, confirmed strong negative transcriptional regulation of established TGF-β-target genes (*LEFTY1* and *CDKN2B*; Figure S5C).

We then focused on the BMP network as predicted from the Gene Ontology analysis (Figures 4F and 4G). In agreement with the RNA sequencing results, *BMP7*, *BMP3*, and the direct gene target of BMP signaling *ID1* were downregulated and *GDF6* was weakly upregulated in response to TGF-β; their expression was essentially lost in cells overexpressing *TGFB2-AS1* (Figure 5A). Conversely, transient knockdown of *TGFB2-AS1* enhanced the response of the BMP-sensitive and *ID1* gene-derived promoter-reporter (Figure 5B). A more enhanced response of the reporter to BMP7 stimulation was observed in cells stably silencing endogenous *TGFB2-AS1* (Figure 5C). As specificity control, the same promoter-reporter carrying point mutations in its Smad-binding elements (Korchynskyi and ten Dijke, 2002), showed lack of response to BMP7 and lack of effect by *TGFB2-AS1* silencing (Figure 5D). The *ID1*-promoter-based reporter results were also confirmed at the endogenous level; *ID1* mRNA induction by BMP7 stimulation was further enhanced upon silencing *TGFB2-AS1* (Figure 5E), and endogenous ID1 protein showed the same response in time course experiments, whereas Smad1 levels or BMP receptor-phosphorylated Smad1/5/8 levels did not show appreciable changes between control and cells with *TGFB2-AS1* silencing (Figure 5F). It is worth emphasizing that endogenous *TGFB2-AS1* expression did not change in response to BMP7 treatment (Figure S2F), suggesting that although the regulation of *TGFB2-AS1* expression is TGF-β but not BMP dependent, *TGFB2-AS1* can regulate BMP pathway genes. We conclude that the BMP signaling module lies downstream of the TGF-β/*TGFB2-AS1* regulatory pathway.

TGFB2-AS1 Binds to EED and Modulates Histone 3 Methylation during TGF-β-Induced Gene Expression

The nuclear localization and the wide array of gene-specific effects of *TGFB2-AS1* suggested a biochemical involvement of *TGFB2-AS1* in transcription factor or chromatin-dependent (epigenetic) regulation. To address this hypothesis, we analyzed the predicted secondary structure of *TGFB2-AS1*, which revealed stable stem structures, partitioned into a 5′-proximal, a middle, and a 3′-proximal cluster (Figure 6A). Such stem-loop structures often imply interactions with proteins in ribonucleoprotein complexes. We performed RNA pull-down assays using *in vitro* synthesized *TGFB2-AS1* coupled to biotin and captured on streptavidin beads, followed by incubation with a HaCaT nuclear extract and unbiased mass spectrometry analysis of proteins that associated with the biotinylated RNA (Figure 6B). Because *TGFB2-AS1* is nuclear, we tried to "filter out" potential non-specific interacting proteins from the cytoplasm by examining only nuclear extracts. An *in vitro* transcribed luciferase RNA served as negative control (Figure 6C). *TGFB2-AS1* resulted in pull-down of 2,529 identifiable proteins, 2,313 (90%) of which were also identified on the control *F-luc* mRNA (Figure 6D). Specifically, 182 proteins from untreated cells associated with *TGFB2-AS1* and 216 from cells stimulated with TGF-β, resulting in 34 proteins that were unique to the TGF-β-stimulated sample (Figure 6D). Comparing cells stimulated or not with TGF-β could provide a comparison of the impact TGF-β might have on abundance or post-translational modifications of nuclear proteins. These proteins scored significantly and gave high peptide representation (Figure S6A; Table S6). Among the top-scoring proteins showing specific interaction with *TGFB2-AS1* were EED and three components of the Mediator complex, MED1, MED4, and MED21 (Figure 6D). Smad3 of the TGF-β pathway also scored, but the statistics of peptide identification were of low confidence (SMAD3 in brackets; Figure 6D). We confirmed that endogenous EED resided in the nuclei of the keratinocytes and stimulation with TGF-β or overexpression of *TGFB2-AS1* appeared to result in higher accumulation of nuclear EED, yet these effects were not statistically significant (Figure S6B).

Validation of the above results was followed using *in vitro* synthesized full-length (FL) *TGFB2-AS1* and three fragments, each representing roughly one third of the *TGFB2-AS1* RNA (Figure 6A). Full-length *TGFB2-AS1* readily associated with HA-tagged EED protein expressed in HEK293T cells but failed to interact with the anti-parallel synthetic RNA (*anti-TGFB2-AS1*; Figures 6E, 6F, and S6C). The association involved primarily the two 3′ fragments of *TGFB2-AS1* (nt 201–557; Figure 6E). Because EED is a component of the PRC2 complex that includes EZH2 and SUZ12, we tested whether the latter two proteins could form complexes with *TGFB2-AS1* and found that EZH2 (and perhaps even SUZ12; notice the weak association with the control *anti-TGFB2-AS1* RNA, which raises issues of specificity (Figure S6C)), showed preferential binding to *TGFB2-AS1* compared with weak or undetectable binding to the *anti-TGFB2-AS1* control (Figure S6C). In order to test whether EED was responsible for the strong association of EZH2 with *TGFB2-AS1*, we repeated the RNA pull-down assays after silencing endogenous EED in the HEK293T cells (Figures 6G and 6H). Both transfected HA-EED and HA-EZH2 scored positively on the pull-down assay with specificity only for *TGFB2-AS1* RNA (Figure S6H). The association with HA-SUZ12 was not reproducible in this set of experiments (Figure 6H). Furthermore, silencing endogenous EED resulted in a noticeable decrease of association of EED and EZH2 with *TGFB2-AS1* (Figure 6H). Note that the EED siRNAs used target 3′ UTR sequences and do not affect the transfected human HA-EED (Figure 6H, TCL).

Independent RNA immunoprecipitation (RIP) assays followed by qRT-PCR confirmed that EED- and EZH2-specific endogenous immunoprecipitation significantly enriched for *TGFB2-AS1* compared with control immunoglobulin (Figure 6I,

(D and F) Gene Ontology analysis of upregulated (D) and downregulated (F) genes in *TGFB2-AS1*+TGF-β compared with the pcDNA3+TGF-β condition, with representative gene examples. Number of genes per biological process is shown by circle diameter and fold change in expression level by color.
(E and G) Biological processes of upregulated (E) and downregulated (G) genes observed in *TGFB2-AS1*+TGF-β compared with the pcDNA3+TGF-β condition. The data are expressed in the form of adjusted p value (color-coded).

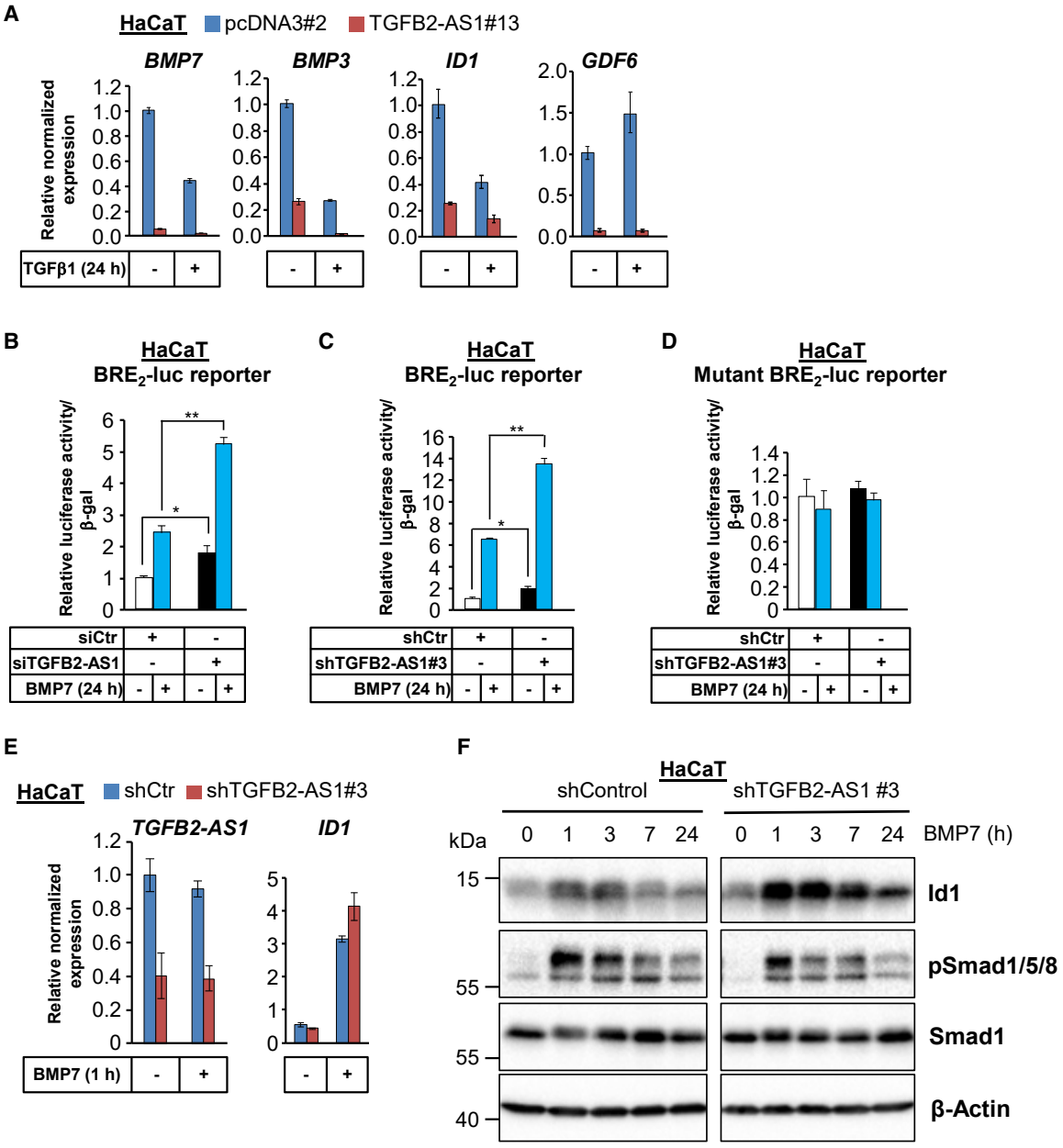


Figure 5. *TGFB2-AS1* Opposes the BMP Signaling Pathway

(A) Quantitative real-time PCR monitoring *BMP7*, *BMP3*, *ID1*, and *GDF6* mRNA expression in HaCaT clones stably expressing *TGFB2-AS1* or control, in the presence or absence of TGF-β1 stimulation for 24 h.

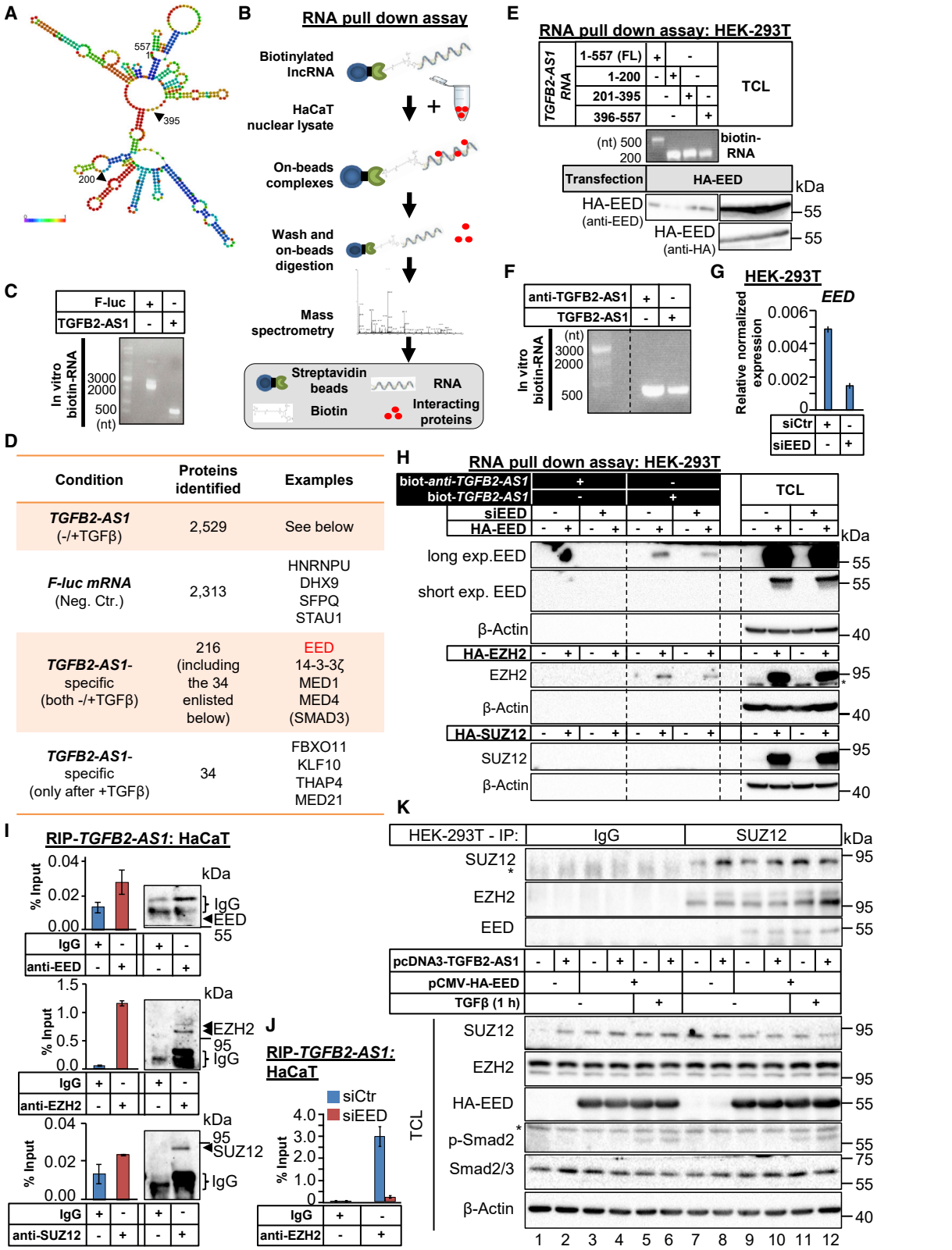
(B–D) BRE_2-luciferase assay using HaCaT cells transiently transfected with *siTGFB2-AS1* (B), HaCaT cells stably transfected with *shTGFB2-AS1* (C), or mutant BRE_2-luciferase assay using HaCaT cells expressing a non-responsive BRE_2-reporter construct, stably transfected with *shTGFB2-AS1* (D), in the presence or absence of BMP7 stimulation for 24 h.

(E) Quantitative real-time PCR to determine*TGFB2-AS1* and *ID1* in HaCaT cells stably transfected with *shTGFB2-AS1* or control, and in the presence or absence of BMP7 stimulation for 1 h.

(F) Protein expression levels of Id1, p-Smad1/5/8, Smad1, and β-actin (loading control, along with molecular size markers [kDa]) in HaCaT cells transfected with *shTGFB2-AS1*, or *shControl*, after BMP7 stimulation for the indicated time periods. Representative immunoblot out of four independent experiments.

In (A)–(E), error bars represent SD from three different experiments (*$p < 0.05$ and **$p < 0.01$).

immunoblot controls verify immunoprecipitation efficiency). On the other hand, SUZ12 immunoprecipitation did not significantly enrich for *TGFB2-AS1* RNA (Figure 6I). Further evidence for the key role of EED was obtained when endogenous EZH2-specific immunoprecipitation was performed after silencing endogenous EED (Figures S6D and 6J). The robust enrichment of *TGFB2-AS1*



(legend on next page)

in the anti-EZH2 RIP almost disappeared upon silencing endogenous EED (Figure 6J). In order to examine whether *TGFB2-AS1* might regulate the assembly of the PRC2 complex, we relied on HEK293T cell experiments in which HA-EED was expressed and endogenous SUZ12 was immunoprecipitated, followed by immunoblot for endogenous EZH2 and transfected HA-EED (Figure 6K). Compared with the negative control IgG, the presence of the PRC2 complex was revealed in the cells (Figure 6K, lane 9). The PRC2 complex was also readily detected upon transfection of *TGFB2-AS1* or 1 h stimulation with TGF-β (Figure 6K, lanes 10 and 11, respectively) and upon combination of TGF-β stimulation with *TGFB2-AS1* transfection, a 1.7- to 2-fold increase of co-precipitating SUZ12 and HA-EED was observed (Figure 6K, lane 12).

Analyzing the association of *TGFB2-AS1* with additional proteins derived from the mass spectrometry screen (e.g., Smad3, MED components, and transcription factor KLF10; Figures 6D and S6A) gave equally positive association results on the basis of the RIP assay (Figures S6E–S6H). The heterogeneous ribonucleoprotein U (hnRNPU) that scored positive for both the negative control F-luc and the *TGFB2-AS1* RNAs (Figure 6D) was also validated; we verified association between hnRNPU and *TGFB2-AS1* but even more with endogenous *HPRT1* mRNA (Figure S6H). We conclude that proteins, such as hnRNPU, known to be involved in RNA maturation exhibit general specificity for RNAs, whereas the nuclear EED, EZH2, Smad3, MED21, or KLF10 associate with higher specificity to the *TGFB2-AS1* RNA.

Further support for a link between *TGFB2-AS1* and the PRC2 complex was obtained by bioinformatic analysis of transcription factors that have been analyzed to bind (via ChIP sequencing [ChIP-seq] experiments) to the genes downregulated after *TGFB2-AS1* overexpression (Figure S6I). Among the most highly enriched transcription-chromatin factors were EED, EZH2, SUZ12, and JARID2, four key components of the PRC2 complex. In addition, we compared the 212 downregulated genes after *TGFB2-AS1* overexpression in HaCaT cells with published ChIP-seq data for identifying PRC2 and PRC1 target genes in DU145 human prostate cancer cells (Cao et al., 2014). Fifty genes were regulated by components of the PRC1 and PRC2 complexes and downregulated by *TGFB2-AS1* overexpression (Figure S6J; Table S7). Members of the BMP signaling pathway (*BMP7* and *GDF6*) scored among the 50 commonly regulated genes (Figure S6K). In addition, 9 of these genes are known to be downregulated and 11 of the genes are known to be upregulated by TGF-β signaling in various cell types, whereas for the other 30, no information on their regulation by TGF-β is available (Figure S6K). Among these 20 TGF-β-regulated genes, for 8 of them, ChIP-derived information on Smad2/Smad3 binding to their promoters is available (Figure S6K).

Association of the PRC2 components SUZ12 and EZH2 with specific nuclear lncRNAs has been established (Davidovich and Cech, 2015). Our evidence highlighted the WD domain-rich adaptor protein EED, to which further functional analysis focused, aiming at uncovering new molecular mechanisms by which an lncRNA affects PRC2 function. Using the *TGFB2-AS1*-overexpressing HaCaT cells, in which TGF-β responses are repressed, transient silencing of endogenous EED led to normalization (i.e., a rescue) of the effects of *TGFB2-AS1* on *SERPINE1* mRNA induction by TGF-β (Figure 7A). The $CAGA_{12}$-luc reporter that is derived from the *PAI-1* promoter exhibited similar but weaker rescue upon EED silencing (Figure S7A), presumably because the transfected reporter did not

Figure 6. *TGFB2-AS1* Interacts with EED

(A) Predicted secondary structure of *TGFB2-AS1* generated by RNAfold. The 5′ and 3′ ends and break-points (arrowheads) of fragments used in RNA pull-down are marked with nucleotide numbers: 1–200, 201–395, and 396–557. Each nucleotide is color-coded (see color scale): dark blue (probability 0) indicates the lowest and dark red (probability 1) the highest base-pairing probability; for unpaired regions, the color highlights the probability of being unpaired.

(B) Schematic illustration of the RNA pull-down assay followed by mass spectrometry analysis.

(C) *In vitro* transcribed biotinylated *F-luciferase* (*F-luc*) mRNA and *TGFB2-AS1* RNA analyzed by agarose gel electrophoresis and molecular size markers (nt).

(D) Total protein numbers and examples of proteins interacting with *TGFB2-AS1* or *F-luc* mRNA, identified by mass spectrometry in nuclear lysates of HaCaT cells, stimulated with or without TGF-β for 1 h. EED is highlighted (red) and Smad3 is in brackets because of low statistical coverage.

(E) RNA pull-down assay using biotinylated full-length (FL) or different *TGFB2-AS1* fragments immobilized on streptavidin beads and lysates from HA-EED overexpressing HEK293T cells. The same cell lysate was applied to each specific RNA. Biotinylated *TGFB2-AS1* RNA fragments were analyzed by agarose gel electrophoresis (top). The proteins retained on the RNA beads and total cell lysates (TCLs) were analyzed by immunoblotting using the indicated antibody. Representative immunoblots out of two independent experiments with molecular size markers (kDa).

(F) *In vitro* transcribed biotinylated *anti-TGFB2-AS1* and *TGFB2-AS1* RNAs were analyzed using agarose gel electrophoresis along with molecular size markers (nt).

(G) Quantitative real-time PCR to determine mRNA levels of *EED* in HEK293T cells transiently transfected with siEED or siControl. The error bars represent SD from three different experiments.

(H) RNA pull-down assay using biotinylated (biot-) *anti-TGFB2-AS1* and *TGFB2-AS1* RNAs (black background) immobilized on streptavidin beads and lysates from HEK293T cells transiently transfected with control (−) or specific siRNA targeting EED (siEED; +) and overexpressing HA-EED, HA-EZH2, and HA-SUZ12. The same cell lysate was applied to each specific RNA. The proteins retained on the RNA beads and total cell lysates (TCLs) were analyzed by immunoblotting using the indicated antibodies. Representative immunoblots out of three independent experiments, with two different exposures (exp.) for HA-EED and molecular size markers (kDa). A star indicates a non-specific protein band. Dotted lines serve orientation and alignment (the immunoblots were not disrupted).

(I) RIP of HaCaT lysates using antibodies against endogenous EED, EZH2, SUZ12, or control IgG. Error bars represent SD from three different experiments. Corresponding immunoblots indicate the immunoprecipitated IgG or specific protein (marked by arrowheads) and molecular size markers (kDa).

(J) RIP of HaCaT lysates after transient transfection with siEED or siControl, using antibody against endogenous EZH2 or control IgG. Error bars represent SD from three different experiments.

(K) Immunoprecipitation of endogenous SUZ12 (or control IgG) in HEK293T cells transiently transfected with the indicated expression constructs and stimulated or not with TGF-β1 for 1 h, followed by immunoblotting for endogenous SUZ12, endogenous EZH2, or transfected HA-EED. Immunoblots of corresponding total cell lysates (TCLs) for the same three proteins in addition to p-Smad2 (indicator of TGF-β stimulation), total Smad2/3, and loading control β-actin, along with molecular size markers (kDa). Stars indicate non-specific protein bands. Representative immunoblots out of three independent experiments are shown.

obtain a fully organized chromatin state, which is necessary for PRC2/EED to exhibit its actions. The relative resistance to cell-cycle arrest by TGF-β observed in cells overexpressing *TGFB2-AS1* (Figure S4A) was also rescued after silencing the endogenous EED (Figure 7B). However, EED silencing had a general negative effect on deoxynucleotide incorporation even in control cells (Figure 7B).

An important marker of the mesenchymal differentiation (EMT) induced by TGF-β is N-cadherin (*CDH2*); its induction was strongly repressed in cells expressing *TGFB2-AS1* (Figure S7B). Depletion of endogenous EED (Figure S7C) resulted in enhancement of both protein (Figure S7D) and mRNA (Figure S7B) levels of *CDH2*. However, the impact especially on *CDH2* mRNA was relatively weak and partial. Inhibiting PRC2/EZH2 methyltransferase activity with the GSK343 chemical inhibitor weakly induced basal *CDH2* mRNA expression and enhanced the response of *CDH2* to TGF-β, whereas the repressive effect of *TGFB2-AS1* was relieved by at least 3-fold, which was significant, yet not adequate to revert the *CDH2* levels to their fully TGF-β-inducible state (Figure S7E). Total tri-methylation levels of H3K27 were indeed lost after treatment with GSK343 (Figure S7F). In accordance with the impact on N-cadherin expression, the GSK343 inhibitor enhanced basal and TGF-β-induced cell invasion through laminin (Figure S7G). Cells overexpressing *TGFB2-AS1* did not exhibit altered invasiveness, relative to control; treatment with GSK343 enhanced the invasiveness of the cells to a degree where the net impact of TGF-β stimulation on invasiveness was not significant anymore, as forced expression of *TGFB2-AS1* plus GSK343 significantly exceeded the effect of TGF-β alone (Figure S7G).

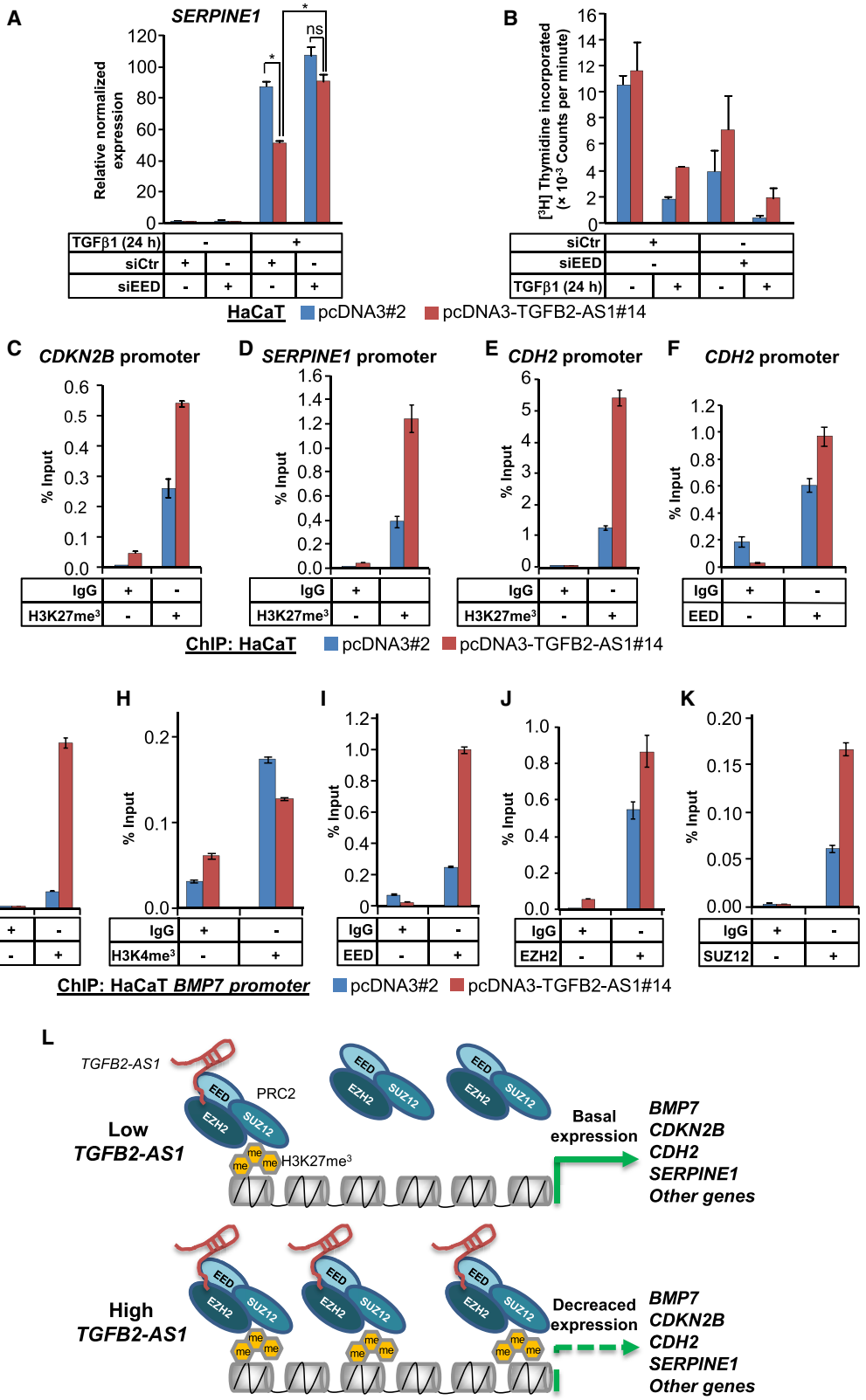
Thus, our findings support the notion that *TGFB2-AS1* enhances recruitment of PRC2 complexes and promotes tri-methylation of H3K27 on genes responding to TGF-β. Indeed, H3K27me^3-specific ChIP-qPCR analysis showed specific recruitment to the *CDKN2B*, *SERPINE1*, and *CDH2* gene promoters, and cells overexpressing *TGFB2-AS1* acquired much higher H3K27me^3 levels to these promoters (Figures 7C–7E). For *CDH2*, the latter correlated with enhanced recruitment of EED to the promoter in cells overexpressing *TGFB2-AS1* (Figure 7F). Analyzing more extensively *BMP7*, H3K27me^3, EED, EZH2, and SUZ12 were identified on the promoter after ChIP-qPCR analysis; *TGFB2-AS1* expression enhanced recruitment of all three PRC2 proteins and of H3K27me^3 levels to the promoter (Figures 7G–7K). The enhancer and/or promoter active chromatin histone mark, H3K4me^3, could be detected on the *BMP7* promoter and cells overexpressing *TGFB2-AS1* exhibited relatively weaker association of this mark with the promoter, but these differences were not robust (Figure 7H). Finally, we selected *TGFB3*, as a gene whose expression was not affected by *TGFB2-AS1* overexpression (Figure 4; Table S5), and confirmed the lack of effect using qRT-PCR (Figure S7H). ChIP analysis revealed abundant H3K27me^3 modification and PRC2 (EZH2 and SUZ12) binding to the *TGFB3* promoter; *TGFB2-AS1* overexpression did not reveal significant differences, and in the case of EZH2, recruitment was even reduced to half after *TGFB2-AS1* overexpression (Figures S7I–S7K). These results suggest that the ability of *TGFB2-AS1* to repress expression of genes that respond to TGF-β signaling depends on the activation of the PRC2 complex through an interaction with EED (and EZH2), which enhances tri-methylation on H3K27, one of the repressive chromatin marks that mediate gene silencing (Figure 7L).

DISCUSSION

Unbiased screening for genes regulated by TGF-β signaling identified many lncRNA genes that are transcriptionally induced or repressed. A proportion of these lncRNAs engage into positive or negative regulation of transcriptional responses to TGF-β. By analyzing the regulation and function of *TGFB2-AS1*, we provide a model (Figure 7L) that explains how such TGF-β-induced lncRNAs can fine-tune gene responses that link to the process of cell-cycle arrest, EMT invasiveness, and BMP signaling. The mechanism involves *TGFB2-AS1* association with the EED adaptor protein of the PRC2 complex. EED is the WD domain adaptor that ''reads'' chromatin modifications and assists PRC2 tethering to specific loci along with the JARID2-dependent methylated cytosine ''reader'' (Davidovich and Cech, 2015).

Many lncRNAs have been shown to interact with the PRC2 complex; among 469 long intergenic non-coding RNAs, 114 (24%) interact with EZH2 or SUZ12 (Khalil et al., 2009). In contrast, <2% of mRNAs can associate with PRC2, suggesting that this chromatin modifier selects for lncRNAs among other types of RNA (Khalil et al., 2009). Although lncRNA association to either EZH2 or SUZ12 is established, to the best of our knowledge, the only known example of an EED-binding lncRNA is the *ROR1-AS1*, which binds to all three EZH2, SUZ12, and EED, in B cell lymphoma (Hu et al., 2017). Unbiased mass spectrometry analysis revealed a specific association of EED with *TGFB2-AS1*; although neither EZH2 nor SUZ12 scored in the mass spectrometry analysis, RNA pull-down and RIP assays demonstrated that EZH2 (but not SUZ12) can reproducibly associate with *TGFB2-AS1* in an EED-dependent manner (Figure 6). EED can suppress the TGF-β pathway in chondrocytes; deletion of EED in mice results in hyper-activated TGF-β signaling, which is responsible for growth and skeletal defects (Mirzamohammadi et al., 2016). This finding is in line with our observations that depletion of EED enhances TGF-β-induced cell-cycle arrest (Figure 7B) and TGF-β/Smad-mediated transcriptional responses (Figures S7A–S7D). In addition, PRC2 is required for maintaining pluripotency in human embryonic stem cells. Deficiency of PRC2 components, in this context, leads to induction of BMP pathway members and consequently meso-endodermal differentiation (Shan et al., 2017). We propose that *TGFB2-AS1* may use a similar mechanism that requires the adaptor protein EED in order to inhibit genes related to the BMP pathway (*BMP7*, *BMP3*, *GDF6*, and *ID1*; Figure 5). For this reason, when exogenous BMP7 stimulates signaling, the impact of *TGFB2-AS1* on target gene *Id1* expression is evident in the absence of any impact on pSmad1/5/8 levels (Figure 5F), which lie upstream of the nuclear or chromatin-bound regulatory module built by *TGFB2-AS1*/PRC2.

Furthermore, *TGFB2-AS1* is exclusively nuclear (Figures 2I and 2J). We suggest that localization is defined by the proteins *TGFB2-AS1* interacts with, which is supported by the nuclear factors identified in our mass spectrometry-based interaction



(legend on next page)

screen (Figures 6 and S6). The speckled nuclear pattern of endogenous and overexpressed *TGFB2-AS1* (Figures 2J and 3F) may be dictated by PRC2 complexes that decorate specific parts of the chromatin (Davidovich and Cech, 2015).

TGF-β signaling regulates several lncRNA genes (Figure 1) (Meng et al., 2016; Wang et al., 2016). Because genomes have evolved mechanisms to control the expression of all genes, irrespective of their coding capacity, we expect a large number of lncRNA genes to be regulated by other TGF-β family members. Furthermore, the transcriptionally regulated lncRNAs that feedback to the TGF-β signaling pathway and regulate its activity are interesting and demand further work. Such feedback mechanism may not only operate by regulation of chromatin modifiers, as demonstrated here for *TGFB2-AS1*, but can also affect Smad protein nuclear accumulation, as *NORAD* does (Kawasaki et al., 2018); Smad3 sequestration away from its binding elements on DNA, as *GAS5* does (Tang et al., 2017); enhancement of Smad3 binding to specific target gene promoters, as *ELIT-1* does (Sakai et al., 2019); or ''sponge''-like regulation of the availability of miRNAs in the cytoplasm, as *XIST* does via sequestration of *miR-141* and *miR-367*, which downregulate downstream EMT effectors, such as ZEB2 (Li et al., 2018; Sripathy et al., 2017). Smad3 also scored as forming complexes with *TGFB2-AS1* (Figures 6D and S6F). A mechanism of gene regulation based on ribonucleoprotein complexes between *TGFB2-AS1* and Smad3 could possibly explain the observed results (i.e., why only specific but not all TGF-β-target genes are regulated by *TGFB2-AS1*). According to such a model, chromatin regions where *TGFB2-AS1* could crosslink Smad3 and PRC2, might define those genes that can be regulated by this lncRNA downstream of TGF-β signaling, a hypothesis worth analyzing deeper. The bioinformatic analysis that compared genes downregulated in cells overexpressing *TGFB2-AS1* and genes occupied by PRC2 complexes (Figures S6I–S6K) suggests that *TGFB2-AS1*/PRC2 may repress genes that are either up- or downregulated by TGF-β and even genes for TGF-β ligand-receptor regulators (Figure S6K). This hypothesis suggests the notion that TGF-β induces *TGFB2-AS1* expression in order to either silence gene responses induced by the same pathway (acting as feedback inhibitory mechanism) or to mediate sustained downregulation of gene expression.

The *TGFB2* locus generates two lncRNAs, *TGFB2-AS1*, via transcription from the opposite strand, and *TGFB2-OT1*, an overlapping transcript via transcription in the same direction as the *TGFB2* mRNA (Figure 2D). *TGFB2-OT1* associates with and inhibits a small family of miRNAs that regulate autophagy and inflammation (Huang et al., 2015). Because many natural antisense transcripts function in *cis* and regulate the expression of their corresponding sense genes (Pelechano and Steinmetz, 2013), we investigated whether *TGFB2-AS1* regulates expression of *TGFB2*. *TGFB2* expression was weakly increased in *TGFB2-AS1*-overexpressing cells (Table S5), which could not withstand qRT-PCR validation (Figures S3C and S3D). Nevertheless, similarly to *VIM-AS1*, an antisense RNA to the vimentin (*VIM*) mRNA (Boque-Sastre et al., 2015), *TGFB2-AS1* might be able to form structures known as R loops, promoting accessibility of transcriptional activators to promoters of genes that are upregulated in *TGFB2-AS1*-overexpressing cells, such as genes of the Wnt pathway. Indeed, we identified interactions between *TGFB2-AS1* and components of the Mediator complex (MED1, MED4, and MED21; Figures 6D and S6G), a general transcriptional facilitator important for RNA polymerase II-dependent transcription (Soutourina, 2018). Thus, Mediator recruitment could contribute to the effect *TGFB2-AS1* has on gene activation. An intriguing possibility can be hypothesized on the basis of evidence of direct interactions between Mediator-dependent regulators, namely, cyclin-dependent kinases 8 and 19, with EZH2 in developing neurons (Fukasawa et al., 2015). It will be most interesting to examine whether *TGFB2-AS1* may provide a molecular bridge between mediator components and the PRC2, a mechanism that might operate downstream of TGF-β, similar to the one demonstrated for retinoic acid signaling (Fukasawa et al., 2015).

The mechanistic evidence provided here (Figures 6 and 7) suggests a role of *TGFB2-AS1* as a regulator of EED-PRC2 function. It is possible that other nuclear lncRNAs regulated by TGF-β may play similar roles, by providing a rheostat-like mechanism that can enhance and/or prolong or silence or decelerate TGF-β-dependent transcriptional events during basic physiological processes.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Virus strains
  - Human cell lines
- METHOD DETAILS
  - Reagents and treatments
  - Transfections with plasmids or siRNAs and lentiviral infection
  - Molecular cloning

Figure 7. TGFB2-AS1 Epigenetically Represses TGF-β-Regulated Genes in an EED/PRC2-Dependent Manner

(A) Quantitative real-time PCR to determine *SERPINE1* expression in HaCaT cells overexpressing *TGFB2-AS1* or control, transiently transfected with siEED or control siRNA, and stimulated with or without TGF-β for 24 h.

(B) Thymidine incorporation assay using HaCaT cells overexpressing *TGFB2-AS1* or control, transiently transfected with siEED or control siRNA, and stimulated with or without TGF-β for 24 h.

(C–F) ChIP-qPCR analysis for $H3K27me^3$ (C–E) occupancy to the promoters of *CDKN2B* (C), *SERPINE1* (D), *CDH2* (E), and for EED occupancy (F) to the *CDH2* promoter in HaCaT cells overexpressing *TGFB2-AS1* or control.

(G–K) ChIP-qPCR analysis for $H3K27me^3$ (G), $H3K4me^3$ (H), EED (I), EZH2 (J), and SUZ12 (K) occupancy to the promoter of *BMP7* in HaCaT cells overexpressing *TGFB2-AS1* or control. In all relevant panels, error bars represent SD from three different experiments (*$p < 0.05$).

(L) Proposed model for epigenetic regulation of transcription by *TGFB2-AS1* via interaction with EED/PRC2.

- RNA structure and coding analysis
- RNA isolation and microarray analysis
- AmpliSeq transcriptome analysis
- *In vitro* RNA transcription and RNA-pull down
- Mass spectrometry
- cDNA synthesis and real-time qPCR
- Nucleo-cytoplasmic fractionation
- RNA immunoprecipitation (RIP)
- RNA FISH
- Immunofluorescence
- Immunoblotting
- Chromatin immunoprecipitation
- Luciferase assays
- Thymidine incorporation assay
- Invasion assays

- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.08.028.

ACKNOWLEDGMENTS

We thank Eleftheria Vasilaki and Anders Sundqvist for reagents and advice, Paula Elhorst for technical assistance, and laboratory members for useful discussions. AmpliSeq analysis was performed at the National Genomics Infrastructure Uppsala (Uppsala Genome Center) of the Science for Life Laboratory, Sweden. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). This work was funded by Ludwig Cancer Research (Uppsala Branch); the Swedish Cancer Society (contract CAN2015/438 to A. Moustakas); the Swedish Research Council (contracts K2013-66X-14936-10-5 to A. Moustakas and 2015-02757 to C.-H.H.); the European Research Council (contract 787472 to C.-H.H.); the Bodossaki Foundation and the Alexander Onassis Foundation to P. Papoutsoglou; and the Japan Society for the Promotion of Science and the Kanae Foundation for the Promotion of Medical Science to Y.T.

AUTHOR CONTRIBUTIONS

A. Moustakas conceived the project. A. Moustakas and P. Papoutsoglou designed the experiments. P. Papoutsoglou, Y.T., L.C., P. Pallis, A. Morén, and A.A. acquired the data. P. Papoutsoglou, Y.T., L.C., and A. Moustakas analyzed the data. P. Papoutsoglou, Y.T., L.C., C.-H.H., and A. Moustakas interpreted the data. P. Papoutsoglou and A. Moustakas drafted the article. All authors critically revised the article for important intellectual content and provided final approval prior to submission for publication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 11, 2018
Revised: May 8, 2019
Accepted: August 5, 2019
Published: September 17, 2019

REFERENCES

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. U S A *112*, 5785–5790.

Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., and Helin, K. (2003). EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J. *22*, 5323–5335.

Cao, Q., Wang, X., Zhao, M., Yang, R., Malik, R., Qiao, Y., Poliakov, A., Yocum, A.K., Li, Y., Chen, W., et al. (2014). The central role of EED in the orchestration of polycomb group complexes. Nat. Commun. *5*, 3127.

David, C.J., and Massagué, J. (2018). Contextual determinants of TGFβ action in development, immunity and cancer. Nat. Rev. Mol. Cell Biol. *19*, 419–435.

Davidovich, C., and Cech, T.R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA *21*, 2007–2022.

Fan, Y., Shen, B., Tan, M., Mu, X., Qin, Y., Zhang, F., and Liu, Y. (2014). TGF-β-induced upregulation of malat1 promotes bladder cancer metastasis by associating with suz12. Clin. Cancer Res. *20*, 1531–1541.

Fukasawa, R., Iida, S., Tsutsui, T., Hirose, Y., and Ohkuma, Y. (2015). Mediator complex cooperatively regulates transcription of retinoic acid target genes with Polycomb Repressive Complex 2 during neuronal differentiation. J. Biochem. *158*, 373–384.

Gomis, R.R., Alarcón, C., He, W., Wang, Q., Seoane, J., Lash, A., and Massagué, J. (2006). A FoxO-Smad synexpression group in human keratinocytes. Proc. Natl. Acad. Sci. U S A *103*, 12747–12752.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36*, W70–W74.

Hu, G., Gupta, S.K., Troska, T.P., Nair, A., and Gupta, M. (2017). Long noncoding RNA profile in mantle cell lymphoma identifies a functional lncRNA ROR1-AS1 associated with EZH2/PRC2 complex. Oncotarget *8*, 80223–80234.

Huang, S., Lu, W., Ge, D., Meng, N., Li, Y., Su, L., Zhang, S., Zhang, Y., Zhao, B., and Miao, J. (2015). A new microRNA signal pathway regulated by long noncoding RNA TGFB2-OT1 in autophagy and inflammation of vascular endothelial cells. Autophagy *11*, 2172–2183.

Kang, Y., Chen, C.R., and Massagué, J. (2003). A self-enabling TGFβ response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. Mol. Cell *11*, 915–926.

Kawasaki, N., Miwa, T., Hokari, S., Sakurai, T., Ohmori, K., Miyauchi, K., Miyazono, K., and Koinuma, D. (2018). Long noncoding RNA NORAD regulates transforming growth factor-β signaling and epithelial-to-mesenchymal transition-like phenotype. Cancer Sci. *109*, 2211–2220.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U S A *106*, 11667–11672.

Korchynskyi, O., and ten Dijke, P. (2002). Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. J. Biol. Chem. *277*, 4883–4891.

Kowanetz, M., Valcourt, U., Bergström, R., Heldin, C.-H., and Moustakas, A. (2004). Id2 and Id3 define the potency of cell proliferation and differentiation responses to transforming growth factor β and bone morphogenetic protein. Mol. Cell. Biol. *24*, 4241–4254.

Li, C., Wan, L., Liu, Z., Xu, G., Wang, S., Su, Z., Zhang, Y., Zhang, C., Liu, X., Lei, Z., and Zhang, H.T. (2018). Long non-coding RNA XIST promotes TGF-β-induced epithelial-mesenchymal transition by regulating miR-367/141-ZEB2 axis in non-small-cell lung cancer. Cancer Lett. *418*, 185–195.

Matouk, I.J., Raveh, E., Abu-lail, R., Mezan, S., Gilon, M., Gershtain, E., Birman, T., Gallula, J., Schneider, T., Barkali, M., et al. (2014). Oncofetal H19 RNA promotes tumor metastasis. Biochim. Biophys. Acta *1843*, 1414–1426.

Meng, X.M., Nikolic-Paterson, D.J., and Lan, H.Y. (2016). TGF-β: the master regulator of fibrosis. Nat. Rev. Nephrol. *12*, 325–338.

Mirzamohammadi, F., Papaioannou, G., Inloes, J.B., Rankin, E.B., Xie, H., Schipani, E., Orkin, S.H., and Kobayashi, T. (2016). Polycomb repressive complex 2 regulates skeletal growth by suppressing Wnt and TGF-β signalling. Nat. Commun. *7*, 12047.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. Nat. Commun. *6*, 7743.

Morikawa, M., Derynck, R., and Miyazono, K. (2016). TGF-β and the TGF-β family: context-dependent roles in cell and tissue physiology. Cold Spring Harb. Perspect. Biol. *8*, a021873.

Pádua Alves, C., Fonseca, A.S., Muys, B.R., de Barros E Lima Bueno, R., Bürger, M.C., de Souza, J.E., Valente, V., Zago, M.A., and Silva, W.A., Jr. (2013). Brief report: the lincRNA Hotair is required for epithelial-to-mesenchymal transition and stemness maintenance of cancer cell lines. Stem Cells *31*, 2827–2832.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J. *23*, 4061–4071.

Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. Nat. Rev. Genet. *14*, 880–893.

Pickup, M.W., Owens, P., and Moses, H.L. (2017). TGF-β, bone morphogenetic protein, and activin signaling and the tumor microenvironment. Cold Spring Harb. Perspect. Biol. *9*, a022285.

Richards, E.J., Zhang, G., Li, Z.P., Permuth-Wey, J., Challa, S., Li, Y., Kong, W., Dan, S., Bui, M.M., Coppola, D., et al. (2015). Long non-coding RNAs (LncRNA) regulated by transforming growth factor (TGF) β: LncRNA-hit-mediated TGFβ-induced epithelial to mesenchymal transition in mammary epithelia. J. Biol. Chem. *290*, 6857–6867.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Sakai, S., Ohhata, T., Kitagawa, K., Uchida, C., Aoshima, T., Niida, H., Suzuki, T., Inoue, Y., Miyazawa, K., and Kitagawa, M. (2019). Long noncoding RNA ELIT-1 acts as a Smad3 cofactor to facilitate TGF-β/Smad signaling and promote epithelial-mesenchymal transition. Cancer Res. *79*, 2821–2838.

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y., et al. (2017). PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. Nat. Commun. *8*, 672.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. *19*, 262–274.

Sripathy, S., Leko, V., Adrianse, R.L., Loe, T., Foss, E.J., Dalrymple, E., Lao, U., Gatbonton-Schwager, T., Carter, K.T., Payer, B., et al. (2017). Screen for reactivation of MeCP2 on the inactive X chromosome identifies the BMP/TGF-β superfamily as a regulator of XIST expression. Proc. Natl. Acad. Sci. U S A *114*, 1619–1624.

Tang, R., Zhang, G., Wang, Y.C., Mei, X., and Chen, S.Y. (2017). The long noncoding RNA GAS5 regulates transforming growth factor β (TGF-β)-induced smooth muscle cell differentiation via RNA Smad-binding elements. J. Biol. Chem. *292*, 14270–14278.

Tordonato, C., Di Fiore, P.P., and Nicassio, F. (2015). The role of non-coding RNAs in the regulation of stem cells and progenitors in the normal mammary gland and in breast tumors. Front. Genet. *6*, 72.

Valcourt, U., Kowanetz, M., Niimi, H., Heldin, C.-H., and Moustakas, A. (2005). TGF-β and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. Mol. Biol. Cell *16*, 1987–2002.

Volders, P.-J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., and Vandesompele, J. (2019). LNCipedia 5: towards a reference set of human long non-coding RNAs. Nucleic Acids Res. *47* (*D1*), D135–D139.

Wang, J., Shao, N., Ding, X., Tan, B., Song, Q., Wang, N., Jia, Y., Ling, H., and Cheng, Y. (2016). Crosstalk between transforming growth factor-β signaling pathway and long non-coding RNAs in cancer. Cancer Lett. *370*, 296–301.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Yuan, J.H., Yang, F., Wang, F., Ma, J.Z., Guo, Y.J., Tao, Q.F., Liu, F., Pan, W., Wang, T.T., Zhou, C.C., et al. (2014). A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. Cancer Cell *25*, 666–681.

Zavadil, J., Bitzer, M., Liang, D., Yang, Y.C., Massimi, A., Kneitz, S., Piek, E., and Böttinger, E.P. (2001). Genetic programs of epithelial cell plasticity directed by transforming growth factor-β. Proc. Natl. Acad. Sci. U S A *98*, 6686–6691.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-β-actin (AC-15)	Santa Cruz Biotechnology Inc.	Cat# sc-69879
anti-N-Cadherin	BD Biosciences-Europe	Cat# 610920, RRID:AB_2077527
anti-EED	Active Motif Europe	Cat# 61203, RRID:AB_2615071
anti-EZH2 (AC-22, immunoblot only grade)	Active Motif Europe	Cat# 39875, RRID: AB_2561022
anti-EZH2 (ChIP grade, AC-22)	Millipore/Merck	Cat# 17-662, RRID: AB_1977568
anti-Fibronectin	Sigma-Aldrich AB	Cat# F3648, RRID:AB_476976
anti-HA	Ludwig Cancer Research Uppsala	Home-made hybridoma
anti-Histone H3	Active Motif Europe	Cat# 39763, RRID:AB_2650522
anti-Histone-H3-tri-methyl-K27	Abcam	Cat# ab6002
anti-Histone-H3-tri-methyl-K4	Abcam	Cat# ab8580, RRID:AB_306649
anti-hnRNPU	Santa Cruz Biotechnology Inc.	Cat# sc-32315
anti-Id1 (Z-8)	Santa Cruz Biotechnology Inc.	Cat# sc-427
anti-mouse IgG dynabeads M-280	Invitrogen, Thermofisher Scientific	Cat# 11202D
anti-PAI-1	BD Biosciences-Europe	Cat# 612025, RRID:AB_399420
anti-p-Smad1/5/8	Cell Signaling Technology	Cat# 9511
anti-Smad1	Abcam	Cat# ab33902, RRID:AB_777975
anti-p-Smad2	Home-made	Not applicable
anti-Smad2/3	BD Biosciences-Europe	Cat# 610843, RRID:AB_39816
anti-Smad3 (ChIP/RIP grade)	Abcam	Cat# ab28379, RRID: AB_2192903
anti-Smad3 (immunolblot grade)	Abcam	Cat# ab40854, RRID: AB_777979
anti-Smad3 (C67H3)	Cell Signaling Technology	Cat# 9523S, RRID:AB_2193182
anti-SRB7 (MED21) (31-C)	Santa Cruz Biotechnology Inc.	Cat# sc-101186, RRID:AB_2142175
anti-SUZ12 (ChIP/immunoblot grade)	Abcam	Cat# ab12073, RRID: AB_442939
anti-TIEG1 (KLF10) (95-D)	Santa Cruz Biotechnology Inc.	Cat# sc-130408, RRID:AB_2296622
normal mouse IgG	Millipore/Merck	N/A
secondary antibodies (Alexa Fluor-488)	Invitrogen, Thermofisher Scientific	N/A
Bacterial and Virus Strains		
MISSION pLKO.1-puro lentivirus targeting *TGFB2-AS1*: shTGFB2-AS1#3: CGTGTCTGCCTTCAACAAAGT	Sigma-Aldrich AB	N/A
MISSION pLKO.1-puro lentivirus targeting *TGFB2-AS1*: shTGFB2-AS1#4: ACTGGAAACTCTGTTGAATGT	Sigma-Aldrich AB	N/A
MISSION pLKO.1-puro-CMV-TurboGFP control lentivirus	Sigma-Aldrich AB	SHC003V
MISSION non-mammalian shRNA control lentivirus	Sigma-Aldrich AB	SHC002V
Biological Samples		
See Experimental Models: Cell Lines		N/A
Chemicals, Peptides, and Recombinant Proteins		
Actinomycin D	Sigma-Aldrich AB	Cat# A9415
BMP7	Sanofi-Genzyme Research	Dr. Kuber Sampath, gift
4′,6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich AB	Cat# D9542
Dharmafect-1	Dharmacon/GE Healthcare	Cat# T-2001-01
Dimethyl-sulfoxide (DMSO)	Sigma-Aldrich AB	Cat# D2650
Dulbecco's modified eagle medium (DMEM)	Sigma-Aldrich AB	Cat# D5796
EcoRI-HF restriction enzyme	New England Biolabs Inc.	Cat# R3101L

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Fetal bovine serum (FBS)	Biowest, Almeco A/S	Cat# S1810
Fluoromount-G	SouthernBiotech, AH Diagnostics	Cat# 0100-01
Fugene HD	Promega	Cat# E2312
Geneticin	Thermofisher Scientific	Cat# 11811-031
GSK343	Sigma-Aldrich AB	Cat# SML0766
GW6604	Ludwig Cancer Research Ltd	Home-made synthesis
Hind III-HF restriction enzyme	New England Biolabs Inc.	Cat# R3104S
Laminin	Sigma-Aldrich AB	Cat# L2020
Lipofectamine 3000	Thermofisher Scientific	Cat# L3000-015
PD184352	Sigma-Aldrich AB	Cat# PZ0181
Penicillin-streptomycin solution	Sigma-Aldrich AB	Cat# P0781
Protease inhibitor cocktail	Roche Diagnostics Scandinavia AB	Cat# 10190300
Superase In RNase inhibitor	Ambion, Thermofisher Scientific	Cat# AM2696
SB203580	Calbiochem-Merck	Cat# 559389
SiLentFect lipid reagent	Bio Rad Laboratories AB	Cat# 170-3362
SP600125	Calbiochem-Merck	Cat# 420119
Tetramethylrhodamine-isothiocyanate-conjugated phalloidin	Sigma-Aldrich AB	Cat# P1951
TGFβ1	PeproTech EC Ltd	Cat# 100-21
TGFβ2	Whitehead Institute for Biomedical Research	Dr. Harvey F. Lodish, gift
TGFβ3	Whitehead Institute for Biomedical Research	Dr. Harvey F. Lodish, gift
Critical Commercial Assays		
Agencourt® AMPure® XP reagent	Beckman Coulter Inc.	N/A
DNF-474 High Sensitivity NGS Fragment Analysis Kit	Advanced Analytical Technologies, INC.	N/A
Enhanced chemiluminescence kit	Merck/Millipore	Cat# WBKLS0500
Firefly and Renilla Dual Luciferase Assay kit	Biotium	Cat# BTIU30003-2
HiScribe™ T7 High Yield RNA Synthesis kit	New England Biolabs, BioNordika Sweden AB	Cat# E2040S
Ion AmpliSeq Transcriptome Human Gene Expression Preparation Kit (Revision A.0)	Life Technologies	N/A
Ion AmpliSeq Transcriptome Human Gene Expression core panel	Life Technologies	N/A
Ion PI Hi-Q Sequencing 200 Kit	Life Technologies	N/A
iScript cDNA synthesis kit	Bio Rad Laboratories AB	Cat# 170-8891
LncPathTM Human Epithelial to Mesenchymal Transition (EMT) Array	Arraystar Inc	Cat# AS-LP-004H
Magna-RIP RNA-binding protein immunoprecipitation kit	Millipore/Merck	Cat# 17-700
Magnetic RNA-Protein Pull-Down kit protocol	Pierce/Thermofisher Scientific	Cat# 20164
NucleoSpin RNA Plus Kit	Macherey-Nagel, AH Diagnostics	Cat# 740984.25
PARIS nucleo-cytoplasmic fractionation kit	Ambion, Thermofisher Scientific	Cat# AM1921
Platinum® PCR SuperMix High Fidelity and Library Amplification Primer Mix	Life Technologies	N/A
PrimeScript 1st strand cDNA synthesis kit	Takara Bio Europe	Cat# 6110A
qPCRBIO SyGreen 2 × Master Mix	PCR Biosystems	Cat# 22-PB20.13-50
Quick Amp Labeling Kit for fluorescent cRNA (version 5.7)	Agilent	N/A
RNA 3′ End Desthiobiotinylation kit	Pierce/Thermofisher Scientific	Cat# 20160
RNeasy kit	QIAGEN AB	Cat# 75144
Stellaris RNA-FISH probes, custom assay	Biosearch Technologies	Cat# SMF-1083-5

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Ampliseq RNA sequencing data	Arrayexpress, EBI	E-MTAB-7773
lncRNA microarray data	Arrayexpress, EBI	E-MTAB-7773
Mass spectrometry primary data	This article	Table S5
Experimental Models: Cell Lines		
A549 human lung adenocarcinoma	Uppsala University	ATCC Cat# CRM-CCL-185, RRID:CVCL_0023
HaCaT human immortalized keratinocytes	Leiden University Medical Center, the Netherlands	Dr. Peter ten Dijke
HaCaT $CAGA_{12}$-Luc/TK-Renilla	The Francis Crick Institute, UK	Dr. Caroline Hill
HEK293T human embryonic kidney	ATCC	ATCC Cat# CRL-3216, RRID:CVCL_0063
HepG2 human hepatoblastoma	ATCC	ATCC Cat# HB-8065, RRID:CVCL_0027
MCF10A-MII	Fred Hutschinson Cancer Center, Seattle, USA	Dr. Dennis Miller
MDA-MB-231 human breast carcinoma	Uppsala University	ATCC Cat# HTB-26, RRID:CVCL_0062
Oligonucleotides		
ON-TARGETplus Human SMAD3 siRNA SMARTpool	Dharmacon/GE Healthcare	L-020067-00
ON-TARGETplus Human SMAD4 siRNA SMARTpool	Dharmacon/GE Healthcare	L-003902-00
ON-TARGETplus Human EED siRNA SMARTpool	Dharmacon/GE Healthcare	L-017581-00
ON-TARGETplus Non-targeting Pool	Dharmacon/GE Healthcare	D-001810-10
Lincode human TGFB2-AS1 siRNA SMARTpool	Dharmacon/GE Healthcare	R-181063-00
Lincode human Non-Targeting Pool	Dharmacon/GE Healthcare	D-001320-10
Additional oligonucleotides are presented in Table S1		N/A
$pBRE_2$-luc	Leiden University Medical Center, the Netherlands	Dr. Peter ten Dijke
Mutant- $pBRE_2$-luc	Leiden University Medical Center, the Netherlands	Dr. Peter ten Dijke
$pCAGA_{12}$-luc	Heldin lab	N/A
pcDNA3	Heldin lab	N/A
pcDNA3-TGFB2-AS1	This study	N/A
pcDNA3-TGFB2-AS11-200	This study	N/A
pcDNA3-TGFB2-AS1201-395	This study	N/A
pcDNA3-TGFB2-AS1396-557	This study	N/A
pCMV-β-gal	Heldin lab	N/A
pCMV-HA-EED	Addgene	Cat# 24231
pCMV-HA-EZH2	Addgene	Cat# 24230
pCMV-HA-SUZ12	Addgene	Cat# 24232
Software and Algorithms		
Enrichr program	http://amp.pharm.mssm.edu/Enrichr/	RRID: SCR_001575
GenePix Pro 6.0 software	Axon, Molecular Devices, LLC	RRID: SCR_010969
Image/J	National Institutes of Health, Bethesda, MD; https://imagej.nih.gov/ij/	RRID: SCR_003070
Proteome Discoverer 1.4 software platform	Thermofisher Scientific	RRID: SCR_014477
QED Camera Plugin v1.1.6 microscopy image acquisition software	QED Imaging Inc.	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R project for statistical computing	https://www.r-project.org/	RRID: SCR_001905
Vienna RNA: RNAfold web server package	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi	RRID: SCR_008550
Stellaris RNA FISH Probe designer	Biosearch Technologies	N/A
Torrent Mapping Alignment Program (TMAP)	Thermofisher Scientific	RRID: SCR_000687
Volocity 1 3D Image Analysis Software	PerkinElmer	RRID: SCR_002668
Other		
Axon GenePix 4000B microarray scanner	Molecular Devices, LLC	N/A
Bio Rad CFX96 cycler	Bio Rad Laboratories AB	N/A
Bio Rad wet, semidry transfer units	Bio Rad Laboratories AB	N/A
CCD digital camera Hamamatsu C4742-95	Carl Zeiss AB	N/A
Diagenode Bioruptor sonicator	Diagenode, Bionordika	N/A
DionexUltiMate 3000 RSLCnano System coupled to a Q-Exactive (QEx) mass spectrometer	Thermofisher Scientific	N/A
Fuorescence microscope Zeiss Axioplan 2 - Zeiss 40 × objective lens	Carl Zeiss AB	N/A
Fragment Analyzer	Advanced Analytical Technologies, INC.	N/A
Ion Proton System	Life Technologies	N/A
NanoDrop ND-1000 spectrophotometer	Thermofisher Scientific	N/A

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and request for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Aristidis Moustakas (aris.moustakas@imbim.uu.se). Plasmids and human cell lines generated in this study are listed in the Key Resources Table and the STAR Methods, respectively, and are available upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Virus strains

MISSION pLKO.1-puro lentiviruses targeting *TGFB2-AS1*, shTGFB2-AS1#3 and #4, pLKO.1-puro-CMV-TurboGFP control lentivirus (SHC003V) and non-mammalian shRNA control lentivirus (SHC002V), were purchased from Sigma-Aldrich AB, Stockholm, Sweden.

Human cell lines

Human immortalized male keratinocytes (HaCaT, Leiden University Medical Center, the Netherlands, Dr. Peter ten Dijke), the derivative cell clone, HaCaT CAGA12-Luc/TK-Renilla (The Francis Crick Institute, UK, Dr. Caroline Hill), human male lung adenocarcinoma A549 cells (ATCC Cat# CRM-CCL-185, RRID:CVCL_0023), human female breast cancer MDA-MB-231 cells (ATCC Cat# HTB-26, RRID:CVCL_0062), human male hepatoblastoma HepG2 cells (ATCC Cat# HB-8065, RRID:CVCL_0027) and human female fetal kidney HEK293T cells (ATCC Cat# CRL-3216, RRID:CVCL_0063) were cultured in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich AB, Stockholm, Sweden) supplemented with 10% fetal bovine serum (FBS, Biowest, Almeco A/S, Esbjerg, Denmark) and 100 U/ml penicillin, 100 μg/ml streptomycin (Sigma-Aldrich AB, Stockholm, Sweden). Human transformed female breast epithelial MCF10A-MII cells (MCF10AneoT, Fred Hutschinson Cancer Center, Seattle, USA, Dr. Dennis Miller) were maintained in DMEM/F12 (Sigma-Aldrich AB, Stockholm, Sweden) supplemented with 5% FBS, 20 ng/ml epidermal growth factor, 100 ng/ml cholera toxin, 0.5 μg/ml hydrocortisone, 10 μg/ml insulin, 100 U/ml penicillin, and 100 μg/ml streptomycin. All cell lines were cultured at 37°C under a 5% CO_2 atmosphere, they were free of mycoplasma (tested every 4 months) and were authenticated using PCR-single-locus-technology (Eurofins, Uppsala, Sweden).

METHOD DETAILS

Reagents and treatments

The growth factors used for cell treatments were: recombinant human TGFβ1 (5 ng/ml, PeproTech EC Ltd, London, UK), recombinant human TGFβ2 and TGFβ3 (5 ng/ml, a gift from Dr. Harvey F. Lodish, Whitehead Institute for Biomedical Research, Cambridge, MA,

USA) and recombinant human BMP7 (30 ng/ml, a gift from K. Sampath, Sanofi-Genzyme Research Center, Framingham, MA, USA). The periods of the treatments with growth factors are indicated in the figures. The GW6604 TβRI kinase inhibitor was synthesized by the Ludwig Cancer Research Ltd (New York, NY, USA) and administered to cells 30 min before TGFβ1 addition at a final concentration of 5 μM. In order to inhibit MAPK pathways, cells were pre-treated with 0.5 μM MAP-kinase kinase (MAPKK/MEK) (PD184352, Sigma-Aldrich AB, Stockholm, Sweden), 10 μM Jun N-terminal kinase (SP600125, Calbiochem-Merck, Stockholm, Sweden) or 10 μM p38 MAP-kinase (SB203580, Calbiochem-Merck, Stockholm, Sweden) inhibitors for 1 h before the stimulation with TGFβ1 for 24 h. RNA polymerase-dependent transcription was blocked by actinomycin D (5 μg/ml, Sigma-Aldrich AB, Stockholm, Sweden), which was administered to cells simultaneously with TGFβ1, for 5 h. The inhibitor of the H3K27 methyltransferase EZH2 (GSK343, Sigma-Aldrich AB, Stockholm, Sweden) was added to cells at a final concentration of 5 μM for 3 days. Dimethyl-sulfoxide (DMSO) served as vehicle control treatment for all the inhibitors, which were dissolved in DMSO.

Transfections with plasmids or siRNAs and lentiviral infection

Transient transfections of HaCaT or A549 cells with plasmids were performed using Lipofectamine 3000 (Invitrogen, Thermo Fisher Scientific, Stockholm, Sweden). Twenty four hours after transfection, the growth medium was changed to starvation medium (0.1% or 1% FBS/DMEM) and cells were treated with growth factors for the indicated time periods. The cells were transfected with plasmids transiently for a period of 48 h. For the generation of stable *TGFB2-AS1* overexpressing clones, HaCaT cells were transfected as described above and 1.2 mg/ml geneticin (Thermo Fisher Scientific, Stockholm, Sweden) was added to the cells 48 h after transfection. The transfected cells were then cultured in the presence of geneticin for two weeks. Then, individual overexpressing clones were selected from the pool of transfected cells, with limiting dilution assay and clones were grown in culture medium, containing geneticin. Transient transfections of HEK293T cells were performed using Fugene HD transfection reagent (Promega Corp., Stockholm, Sweden). The transient transfections with siRNAs targeting mRNAs were performed using the siLentFect lipid reagent (Bio-Rad Laboratories AB, Solna, Sweden), following the protocol by the manufacturer. All recombinant plasmids used for transfections are listed in the Key Resources Table. The reporter plasmid $pBRE_2$-luc and its mutant derivative mutant-BRE_2-luc were provided by Peter ten Dijke (Leiden University Medical Center, Leiden, the Netherlands), as described (Korchynskyi and ten Dijke, 2002). The $CAGA_{12}$-luc reporter and the pCMV-β-gal were also previously described (Kowanetz et al., 2004). The expression vectors pCMV-HA-EED (Addgene plasmid #24231), pCMV-HA-EZH2 (Addgene plasmid #24230) and pCMV-HA-SUZ12 (Addgene plasmid #24232) were gifts from Kristian Helin (Biotech Research and Innovation Center, Copenhagen, Denmark) (Bracken et al., 2003; Pasini et al., 2004).

Cells were transfected with siRNAs targeting mRNAs at a final concentration of 20 nM, in a single transfection, or at a final concentration of 2 × 20 nM, in two sequential transfections, performed on successive days. Transient transfections with siRNAs targeting lncRNAs were performed using Dharmafect 1 reagent (Dharmacon/GE Healthcare, Uppsala, Sweden), according to the manufacturer's instructions, at a final concentration of 25 nM. In the case of simultaneous transfections with siRNAs and plasmids, Lipofectamine 3000 (Thermo Fisher Scientific, Stockholm, Sweden) was used. In all transient siRNA transfections against mRNAs, human siRNA ON-TARGETplus SMARTpools (Dharmacon/GE Healthcare, Uppsala, Sweden) were used: siSMAD3 (ON-TARGETplus human *SMAD3* siRNA SMARTpool, L-020067-00-0005), siSMAD4 (ON-TARGETplus human *SMAD4* siRNA SMARTpool, L-003902-00-0005), siEED (ON-TARGETplus human *EED* siRNA SMARTpool, L-017581-00-0005), non-targeting siRNA (ON-TARGETplus Non-targeting Pool D-001810-10-20) for control transfections. For the transient siRNA transfections against lncRNAs, human siRNA Lincode SMARTpools (Dharmacon/GE Healthcare, Uppsala, Sweden) were used: Lincode human *TGFB2-AS1* siRNA SMARTpool (R-181063-00-0005), Lincode human Non-Targeting Pool siRNAs (D-001320-10-05) for control transfections. All siRNAs listed above are also detailed in the Key Resources Table. The siRNAs used for the silencing the 23 lncRNAs, followed by $CAGA_{12}$ luciferase assay are listed in Table S4.

The infection of HaCaT or A549 cells with short harpin RNAs (shRNAs) targeting *TGFB2-AS1* was performed using MISSION pLKO.1-puro lentivirus constructs (Sigma-Aldrich AB, Stockholm, Sweden). The shRNA sequences targeting the TGFB2-AS1 lncRNA were: *shTGFB2-AS1*#3, CGTGTCTGCCTTCAACAAAGT; and *shTGFB2-AS1*#4, ACTGGAAACTCTGTTGAATGT. Both cell lines were infected at a final multiplicity of infection (MOI) of 10. The efficiency of transduction and the evaluation of the optimal MOI for the lentiviral infection were confirmed using a MISSION pLKO.1-puro-CMV-TurboGFP positive control lentivirus (SHC003V, Sigma-Aldrich AB, Stockholm, Sweden). For control infection, a MISSION non-mammalian shRNA Control (SHC002V, Sigma-Aldrich AB, Stockholm, Sweden) was used. Three days after lentiviral infection, the cells were treated with 2 μg/ml puromycin for two weeks. Then, the efficiency of *TGFB2-AS1* knock-down was confirmed by RT-qPCR.

Molecular cloning

The human *TGFB2-AS1* RNA (NR_046268.1) was reverse transcribed to cDNA and amplified by PCR using total cDNA from HaCaT cells and specific cloning primers synthesized by Eurofins AB (Uppsala, Sweden). Then, the insert was ligated to the pcDNA3 vector between the HindIII and EcoRI restriction enzyme sites, after digestion with the respective enzymes (New England Biolabs Inc. Ipswich, MA, USA). The pcDNA3-TGFB2-AS1 construct was verified by sequencing. The *TGFB2-AS1* cDNA in its inverse orientation, which would lead to an RNA transcript with the anti-sense RNA strand relative to the *TGFB2-AS1* RNA, here named *anti-TGFB2-AS1*, was generated as follows: the cDNA sequence was amplified from pcDNA3-*TGFB2-AS1*, used as DNA template, by PCR using the primer set shown in Table S3 and including specific restriction enzyme cutting sites. The EcoRI-*TGFB2-AS1*-HindIII amplicon was ligated with HindIII/EcoRI sites on pcDNA3 and generated pcDNA3-*anti-TGFB2-AS1*. The same approach was taken in order to

generate the constructs expressing the *TGFB2-AS1* fragments (pcDNA3-*TGFB2-AS1* 1-200, pcDNA3-*TGFB2-AS1* 201-395 and pcDNA3-*TGFB2-AS1* 396-557). For the amplification of the *TGFB2-AS1* fragments by PCR, specific primers were synthesized (Eurofins AB, Uppsala, Sweden). All cloning primers are listed in Table S3. All plasmids were verified by sequencing. The sequencing primers were synthesized by Eurofins AB (Uppsala, Sweden) and were: CMV promoter forward, GTCAATGGGAGTTTGTTTTGG; pcDNA3 BGH polyA R1149 reverse, GCATCGCATTGTCTGAGTAG.

RNA structure and coding analysis

For the analysis of the secondary structure of *TGFB2-AS1* the Vienna RNA/RNAfold WebServer package (Gruber et al., 2008) was used. The optimal secondary structure with a minimum free energy of −147.6 kcal/mol was calculated based on the primary sequence of the *TGFB2-AS1* RNA. The base-pairing probabilities were calculated and are depicted by colored nucleotides in the derived structures: probability 0 is depicted by dark blue color and indicates the lowest probability whereas 1 is the highest base-pairing probability and is depicted as dark red. For unpaired regions the color highlights the probability of being unpaired using the same convention.

NCBI Reference Sequence

NR_046268.1. *Homo sapiens* TGFB2 antisense RNA 1 (head to head) (*TGFB2-AS1*), long non-coding RNA of 557 nt. Alternative transcript names: ENST00000414452.1; RP11-224O19.2-001; OTTHUMT00000095358.1; NONHSAT009517; NR_046268. The txCdsPredict score calculated from the UCSC genome browser (https://genome.mdc-berlin.de/) for *TGFB2-AS1* is 332, significantly low to ensure non protein-coding identity. According to txCdsPredict, a score over 1,000 suggests protein-coding identity almost certainly, and scores over 800 predict protein-coding identity with about 90% certainty. Inspection of LNCipedia (https://lncipedia.org/db/transcript/TGFB2-AS1:2; Volders et al., 2019) also suggests no predicted protein encoded by *TGFB2-AS1*. Another way of predicting the protein coding potential of a transcript is the evaluation of the PhyloCSF (Coding Potential of a multi-species nucleotide sequence alignment) score. A score lower than 60.79 represents a non-coding transcript. According to data derived from LNCipedia, the PhyloCSF score for *TGFB2-AS1* is −228.56, confirming its non-protein coding properties. In addition, the Coding-Potential Assessment Tool (CPAT) algorithm that calculates the coding probability based on the sequence of the transcript, identified a probability of 0.13, much lower to the human coding probability (CP) cutoff, set at 0.364 (CP ≥ 0.364: coding sequence, CP < 0.364: non-coding sequence). Furthermore, the same database does not list any evolutionary conservation among vertebrates or insects. Searches though NCBI indicate that the mouse *Tgfb2* locus encompasses a non-coding RNA, Gm35655, however, BLASTn analysis did not identify any stretch of similarity among the respective nucleotide sequences. In addition, the two lncRNAs have distinct lengths. We therefore conclude that *TGFB2-AS1* is specific for the human *TGFB2* locus.

In the following cDNA sequence: // indicate exon-exon junctions; bold nucleotides predicted start codons; underlined capitalized nucleotides predicted stop codons in-frame with start codons; underlined lower case nucleotides predicted stop codon out of frame with start codons.

5′-CGCAGGGGCTCCCTTGGACTCCCACCCTTTCCCAACCTCTGCTGGCTTCTCCCCAACGTGGAATTGCTCGCTtagG//GAGCTTCTGGAGCTCCGCTCGGAGCGAACCTTCTGCTGCCAGCAGAtaaCATCACGATCTTGCCGGGGAGGGGAGATTTAtaaGTTCTCTCtgaACCACGTGTCTGCCTTCAACAAAGtgaCGTGCTAGGATTACAGTCAGAAGTCCTCTCGTTACTTAGACCACGAGCTCTCCCCGAACCGTTGAGGGAGTGTGGAA**ATG**,AGG,ACC,GCT,GTG,G//TT,ACT,CAA,AGT,CAA,CAT,TTT,CAT,TTC,GGC,**ATG**,ACT,CTG,CAG,CCA,TAC,TGACCAGACAG**ATG**,ATT,TAAACTGGAAACTCTGTTGA**ATG**,TAC,TCC,CAA,ACC,CAA,GAT,CTC,TAT,ACT,GTT,CTC,TAC,TTC,AGA,TGT,TCC,TGAGTTCAGAAAAGTCACAAAGAACTTTCTATCACTTTTGGAATTTATACTTCAGTTGTGTATGGGGCCTAGGTTATTAGTAGAGCACTGAATAAATAAAATAATAATAGCTAACATATACTACACTTAGTA-3′

ORF1: 21 codons, 274 nt from the 5′ end (not optimal, optimal distance from the 5′ end is around 40 nt); the 5′ UTR contains 151 G/C out of 274 or 55% G/C-rich, a very unfavorable sequence. ORF1 has 2 upstream in-frame stop codons and 5 out-of-frame stop codons. The Kozak sequence is very weak based on the nucleotides in positions minus 2, minus 1, plus 1, plus 2 relative to the A of the ATG. **ORF2:** 5 codons, is included into ORF1 and has the same features as ORF1; the Kozak sequence is mildly stronger than that of ORF1, but is preceded by TGA, which makes ribosomal translation highly improbable. **ORF3:** 2 codons, with 8 upstream stop codons and a Kozak sequence with weak context. **ORF4:** 17 codons, with 10 upstream stop codons and a Kozak sequence with the least favorable context among the 4 ORFs. Finally, the splicing pattern does not indicate possibilities to generate a protein-coding ORF encoding a domain of function.

Translation of ORF1

ATGAGGACCGCTGTGGTTACTCAAAGTCAACATTTTCATTTCGGCATGACTCTGCAGCCATAC: **MRTAVVTQSQHFHFGMTLQPY**. BLAST-P search identified the mucin-4 isoform e precursor with 88% amino acid identity and the vitamin D-25-hydroxylase isoform X4 with 41% amino acid identity.

Translation of ORF4

ATGTACTCCCAAACCCAAGATCTCTATACTGTTCTCTACTTCAGATGTTCC: **MYSQTQDLYTVLYFRCS**. BLAST-P search identified the V-type proton ATPase 116 kDa subunit a isoform X2 and the protocadherin alpha-9 isoform 2 precursor with 58% amino acid identity, the peroxisomal leader peptide-processing protease isoform a and the P2Y purinoceptor 12 with 41% amino acid identity.

cDNA sequence of the *anti-TGFB2-AS1* that corresponds to a 557 nt transcript with anti-parallel to *TGFB2-AS1* RNA sequence:
5′-TACTAAGTGTAGTATATGTTAGCTATTATTATTTTATTTATTCAGTGCTCTACTA

ATAACCTAGGCCCCATACACAACTGAAGTATAAATTCCAAAAGTGATAGAAAGTTCTTTGTGACTTTTCTGAACTCAGGAACATCT
GAAGTAGAGAACAGTATAGAGATCTTGGGTTTGGGAGTACATTCAACAGAGTTTCCAGTTTAAATCATCTGTCTGGTCAGTATGGCT
GCAGAGTCATGCCGAAATGAAAATGTTGACTTTGAGTAACCACAGCGGTCCTCATTTCCACACTCCCTCAACGGTTCGGGGAGAG
CTCGTGGTCTAAGTAACGAGAGGACTTCTGACTGTAATCCTAGCACGTCACTTTGTTGAAGGCAGACACGTGGTTCAGAGAGAAC
TTATAAATCTCCCCTCCCCGGCAAGATCGTGATGTTACTGCTGGCAGCAGAAGGTTCGCTCCGAGCGGAGCTCCAGAAGCTCCC
TAAGCGAGCAATTCCACGTTGGGGAGAAGCCAGCAGAGGTTGGGAAAGGGTGGGAGTCAAGGGAGCCCCTGCG-3′.

RNA isolation and microarray analysis

Total RNA from HaCaT cells, treated with TGFβ1 for different time periods (3, 8 and 24 h) or untreated (control) cells was isolated using the RNeasy kit (QIAGEN AB, Sollentuna, Sweden) and subjected to microarray analysis using the LncPath™ Human Epithelial to Mesenchymal Transition (EMT) Array (Arraystar Inc, Rockville, MD, USA). Total RNA from each sample was quantified using the NanoDrop ND-1000 spectrophotometer. For microarray analysis, the Agilent Array platform was employed. The sample preparation and microarray hybridizations were performed based on the manufacturer's standard protocols. Briefly, total RNA from each sample was amplified and transcribed into fluorescent cRNA by using the manufacturer's Agilent's Quick Amp Labeling protocol (version 5.7, Agilent Technologies). The labeled cRNAs were hybridized onto the LncPath™ Human Epithelial to Mesenchymal Transition (EMT) Array (6 × 7K, Arraystar Inc, Rockville, MD, USA). After having washed the slides, the arrays were scanned by the Axon GenePix 4000B microarray scanner (Molecular Devices, LLC, San Jose, CA, USA). Scanned images were then imported into GenePix Pro 6.0 software (Axon, Molecular Devices, LLC, San Jose, CA, USA) for grid alignment and data extraction. Quantile normalization and subsequent data processing were performed using the R software package (https://www.r-project.org/). Differentially expressed LncRNAs/mRNAs with statistical significance between two samples were identified through Volcano Plot filtering. Differentially expressed LncRNAs/mRNAs between two samples were identified through Fold Change filtering (Tables S1 and S2). Hierarchical Clustering was performed to show the distinguishable expression pattern of LncRNAs and mRNAs among samples. Primary data from this analysis is presented in Tables S1 (lncRNA genes) and S2 (mRNA genes).

AmpliSeq transcriptome analysis

Total RNA from HaCaT cells expressing pcDNA3 (empty vector) or pcDNA3-TGFB2-AS1, and treated with TGFβ1 for 24 h or untreated (control), was isolated using the NucleoSpin RNA Plus Kit (Macherey-Nagel, AH Diagnostics, Solna, Sweden). Total RNA (50 ng) was reverse-transcribed to cDNA using the Ion AmpliSeq Transcriptome Human Gene Expression Kit Preparation protocol (Revision A.0, Life Technologies, Stockholm, Sweden). The acquired cDNA was amplified using the Ion AmpliSeq Transcriptome Human Gene Expression core panel (Life Technologies, Stockholm, Sweden) and the primer sequences were partially digested. Adaptors (Ion P1 Adaptor and Ion Xpress Barcode Adaptor, Life Technologies, Stockholm, Sweden) were ligated to the amplicons. Adaptor-ligated amplicons were purified using Agencourt® AMPure® XP reagent (Beckman, Coulter Inc., Brea, CA, USA) and eluted in amplification mix (Platinum® PCR SuperMix High Fidelity and Library Amplification Primer Mix, Life Technologies, Stockholm, Sweden) and amplified. Size-selection and purification was conducted using Agencourt® AMPure® XP reagent (Beckman, Coulter Inc., Brea, CA, USA). The amplicons were quantified using the Fragment Analyzer instrument (Advanced Analytical Technologies, INC., Ankeny, IA, USA) with DNF-474 High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical Technologies, INC., Ankeny, IA, USA). Samples were then pooled (six or less per pool), followed by emulsion PCR on either the Ion OneTouch two System using the Ion PI Hi-Q OT2 Kit (Life Technologies, Stockholm, Sweden), or on the Ion Chef System using the Ion PI Hi-Q Chef Kit (Life Technologies, Stockholm, Sweden). The pooled samples were loaded on Ion PI v3 chips and sequenced on the Ion Proton System using the Ion PI Hi-Q Sequencing 200 Kit chemistry (200 bp read length, Life Technologies, Stockholm, Sweden). Acquired reads were aligned to the hg19 AmpliSeq Transcriptome ERCC v1 using the Torrent Mapping and Alignment Program (TMAP, Thermo Fisher Scientific, Stockholm, Sweden) with default settings. Differentially expressed genes (Table S5) were identified by performing EdgeR analysis (Robinson et al., 2010) using standard parameters. Adjusted p values (padj) for multiple testing, using Benjamini-Hochberg to estimate the false discovery rate (FDR), were calculated for final estimation of differential expression (DE) significance (Table S5). Genes with an FDR < 0.05 and a logFC < −2 or > 2 were selected for further analysis: Venn diagram or Gene Ontology. In order to perform Gene Ontology analysis, the R package Clusterprofiler (Yu et al., 2012) was used. Genes with an FDR < 0.05 and a logFC < −2 (downregulated) in cells overexpressing pcDNA3-TGFB2-AS1, were further selected and analyzed using ChIP Enrichment analysis (ChEA) based on the Enrichr program (http://amp.pharm.mssm.edu/Enrichr/), version 2016 (ChEA 2016) in order to identify transcription factor binding sites on the promoters of the selected genes (Table S7). The expression profiles have been deposited to Array Express with accession number E-MTAB-7773.

In vitro RNA transcription and RNA-pull down

In vitro TGFB2-AS1 full length (nt 1-557) RNA, *anti-TGFB2-AS1* full length RNA, *TGFB2-AS1* (nt 1-200), *TGFB2-AS1* (nt 201-395), *TGFB2-AS1* (nt 396-557) fragment RNAs or *firefly luciferase* mRNA were synthesized using the HiScribe™ T7 High Yield RNA Synthesis kit (New England Biolabs, BioNordika Sweden AB, Stockholm, Sweden), according to the manufacturer's recommendations. The reaction mixture for transcripts longer than 0.3 kb was incubated at 37°C for 2 h. For the synthesis of the short *TGFB2-AS1* fragments the reaction was incubated at 37°C for 16 h. The correct size and the integrity of the *in vitro* transcribed RNAs were verified using denaturing agarose gel electrophoresis. The *in vitro* transcribed RNAs were labeled at the 3′ terminus with a desthiobiotinylated

cytidine biphosphate nucleotide, using the Pierce RNA 3′ End Desthiobiotinylation kit (Pierce/Thermo Fisher Scientific, Stockholm, Sweden), according to the protocol by the manufacturer, and were purified using the RNA cleanup protocol from the RNeasy kit (QIAGEN AB, Sollentuna, Sweden). The *in vitro* biotinylated RNAs were then used for RNA-pull down assays, following the Pierce Magnetic RNA-Protein Pull-Down kit protocol (Pierce/Thermo Fisher Scientific, Stockholm, Sweden). First, biotinylated RNAs were heated at 65°C for 5 min and then kept at room temperature to form their secondary conformation. Then, the biotinylated RNAs were pre-coupled to Nucleic Acid Compatible Streptavidin Magnetic Beads in RNA Capture Buffer (Pierce/Thermo Fisher Scientific, Stockholm, Sweden) for 3 h at 4°C. After washing in 20 mM Tris (pH 7.5), the RNA-beads complex was incubated with protein lysates, supplemented with protease inhibitor cocktail (Roche Diagnostics Scandinavia AB, Bromma, Sweden) and 1 U/μl RNase inhibitors (Superase In, Ambion, Thermo Fisher Scientific, Stockholm, Sweden) for 1 h at 4°C, with rotation. Protein lysates were generated from HEK293T cells expressing endogenous proteins or transiently transfected with the pCMV-HA-EED, pCMV-HA-EZH2 or pCMV-HA-SUZ12 plasmids that express full-length wild-type human EED, EZH2 or SUZ12 proteins, respectively. Protein lysates from parental HaCaT cells expressing endogenous PRC2 proteins were also analyzed using the same protocol. Next, the protein-RNA-beads complexes were washed in 1 × Wash Buffer and proteins were eluted in Elution Buffer (Pierce/Thermo Fisher Scientific, Stockholm, Sweden), boiled at 95°C for 5 min and subjected to SDS-PAGE, followed by immunoblotting.

Mass spectrometry

The mass spectrometry analysis was performed by the Clinical Proteomics Mass Spectrometry facility (Karolinska Institutet, Karolinska University Hospital, Science for Life Laboratory, Stockholm, Sweden), according to the following protocol:

1) On-beads digestion: the 9 protein samples (protein-RNA complexes) bound with magnetic beads were digested with lysC and trypsin in 25 mM ammonium bicarbonate. Briefly, the beads were covered with enough volume (50 μl) of 25 mM ammonium bicarbonate, followed by reduction with dithiothreitol (DTT) and alkylation with iodoacetamide (IAA) at end concentrations of 1 mM and 5 mM, respectively. Then, 5 mM DTT was added to quench the alkylation reaction after IAA incubation. To each sample, 0.1 μg LysC was added and digestion was performed at 37°C overnight. Then, the samples were further digested by 0.1 μg trypsin (each sample) with 37°C overnight incubation.
2) Sp3 peptide enrichment and clean-up: the peptide mixtures generated from the on-beads digestion were subjected to Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) procedure 1. In brief, Sera-Mag SP3 bead mix (5.5 μl) was transferred into the approx. 55 μl sample together with > 96% acetonitrile (ACN). The mix was incubated at room temperature for 8 min. Subsequently, the peptide mixtures were immobilized and rinsed by 100% ACN on the surface of the paramagnetic beads. Then, they were subjected to LC-MS/MS analysis.
3) LC-MS/MS analysis: online LC-MS was performed using a DionexUltiMate 3000 RSLCnano System coupled to a Q-Exactive(QEx)mass spectrometer (Thermo Scientific, Stockholm, Sweden). Samples were trapped on a C18 guard desalting column (Acclaim PepMap 100, 75 μm × 2 cm, nanoViper, C18, 5 μm, 100 Å), and separated on a C18 column (Easy spray PepMap RSLC, C18, 2 μm, 100 Å, 75 μm × 50 cm). The nano capillary solvent A was 95% water, 5% DMSO, 0.1% formic acid; and solvent B was 5% water, 5% DMSO, 95% acetonitrile, 0.1% formic acid. At a constant flow of 0.25 $\mu l \cdot min^{-1}$, the curved gradient went from 2% B up to 40% B in 180 min, followed by a steep increase to 100% B in 5 min. FTMS master scans with 70,000 resolution (and mass range 300-1700 m/z) were followed by data-dependent MS/MS (35,000 resolution) on the top 5 ions using higher energy collision dissociation (HCD) at 30%–40% normalized collision energy. Precursors were isolated with a 2 m/z window. Automatic gain control (AGC) targets were 1e6 for MS1 and 1e5 for MS2. Maximum injection times were 100 ms for MS1 and 150-200 ms for MS2. The entire duty cycle lasted ~2.5 s. Dynamic exclusion was used with 60 s duration. Precursors with unassigned charge state or charge state 1 were excluded. An under-fill ratio of 1% was used.
4) Peptide and protein identification: the MS raw files were searched using Sequest-Percolator (06-10-2017) under the software platform Proteome Discoverer 1.4 (Thermo Scientific, Stockholm, Sweden) against Human Uniprot database (24-07-2017) and filtered to a 1% FDR cut off. We used a precursor ion mass tolerance of 10 ppm, and product ion mass tolerances of 0.02 Da for HCD-FTMS. The algorithm considered tryptic peptides with maximum 2 missed cleavages, carbamido-methylation (C) as a static modification and oxidation (M) as a dynamic modification.

The list of *TGFB2-AS1*-interacting proteins and associated peptide coverage with statistics are shown in Table S6.

cDNA synthesis and real-time qPCR

Total RNA was extracted using the NucleoSpin RNA Plus Kit (Macherey-Nagel, AH Diagnostics, Solna, Sweden). The concentration of RNA was measured using a NanoDrop 2000 instrument (Thermo Fisher Scientific, Stockholm, Sweden), and 1 μg of RNA was used for reverse transcription using the iScript cDNA synthesis kit (Bio-Rad Laboratories AB, Solna, Sweden), according to the protocol by the manufacturer. Real-time qPCR was performed on a Bio-Rad CFX96 cycler (Bio-Rad Laboratories AB, Solna, Sweden) using the qPCRBIO SyGreen 2 × Master Mix (PCR Biosystems, London, UK). The expression levels of target genes were normalized to the expression levels of the reference genes *GAPDH*, *HPRT1* or *18S rRNA* and relative normalized expression was calculated based on the $\Delta\Delta C_t$ method. The results were plotted in graphs as average values of relative normalized expression, with standard deviations of at least three biological experiments. A complete list of the oligonucleotides used for RT-qPCR is shown in the Table S3.

Nucleo-cytoplasmic fractionation

Nucleo-cytoplasmic fractionation was performed using the PARIS kit (Ambion, Thermo Fisher Scientific, Stockholm, Sweden) based on the manufacturer's instructions. Briefly, HaCaT cells were trypsinized, pelleted and washed once in PBS. Then, cells were resuspended in Cell Fractionation Buffer (PARIS kit) on ice and incubated at 4°C for 10 min. Lysates were centrifuged at 500 × g at 4°C for 5 min and the supernatant (cytoplasmic fraction) was transferred to new tubes. Then, the nuclear pellet was washed once in PBS and lysed in ice-cold Cell Disruption Buffer (PARIS kit), with vigorous vortexing. The RNA was isolated by adding a 2 × Lysis/Binding solution (PARIS kit) to each fraction, which was followed by addition of 100% ethanol and capture of the RNA by a filter cartridge. The RNA was eluted in pre-heated Elution solution (PARIS kit) and stored at −70°C.

RNA immunoprecipitation (RIP)

RIP was performed according to the Magna-RIP™ RNA-binding protein immunoprecipitation kit (Millipore/Merck, Stockholm, Sweden). Briefly, cells were scraped in PBS, centrifuged and re-suspended in RIP lysis buffer. After short (5 min) incubation on ice, the lysates were stored at −80°C. In the meantime, magnetic beads carrying protein A/G were incubated with 5 μg anti-EED antibody (Active Motif Europe, La Hulpe, Belgium), anti-EZH2 (Milipore/Merck, Stockholm, Sweden), anti-SUZ12 (Abcam, Cambridge, United Kingdom), anti-Smad3 (Abcam, Cambridge, United Kingdom), anti-SRB7 (MED21), anti-TIEG1 (KLF10), anti-hnRNPU antibodies (all three from Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) or normal mouse IgG (Millipore/Merck, Stockholm, Sweden) for 30 min at room temperature. Lysates were incubated with the bead-antibody complexes in RIP immunoprecipitation buffer at 4°C, overnight. One part (10%) of the lysate from each sample was kept as an input. The next day, the RNA-protein-bead complexes were washed in RIP Wash Buffer (Millipore/Merck, Stockholm, Sweden) and split into two fractions: from the major fraction, bound RNA, together with total RNA from input samples was purified using the RNA cleanup protocol from the RNeasy kit (QIAGEN AB, Sollentuna, Sweden); from the minor fraction, washed lysate was loaded on polyacrylamide gels for immunoblotting using the same antibody with which the RNA immunoprecipitation was performed and described above or a different antibody against the same protein in the case of EZH2 (Active Motif Europe, La Hulpe, Belgium) and Smad3 (ab40854, Abcam, Cambridge, United Kingdom). Using RNA purified from the major fraction, cDNA was generated from the isolated RNA using the PrimeScript™ 1st strand cDNA synthesis kit (Takara Bio Europe, Saint-Germain-en-Laye, France), following the instructions by the manufacturer. The cDNA was subjected to real-time qPCR, using the qPCRBIO SyGreen mix (PCR Biosystems, London, UK). The same protocol was used in the case of parental HaCaT cells that were transiently transfected with control or EED-specific siRNAs prior to the RIP assay.

RNA FISH

In situ hybridization of *TGFB2-AS1* RNA was performed according to the Stellaris RNA-FISH protocol for adherent cells (Biosearch Technologies, Petaluma, CA, USA). Briefly, HaCaT cells were fixed in 3.7% formaldehyde and permeabilized in 70% (v/v) ethanol. Then, hybridization step was performed, using hybridization buffer, containing RNA probes specific for *TGFB2-AS1*. The fixed cells were incubated with the hybridization mixture at 37°C for 16 h and then stained with 5 ng/ml 4′,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich AB, Stockholm, Sweden), mounted in Fluoromount-G (SouthernBiotech, AH Diagnostics, Solna, Sweden) and examined on a Zeiss Axioplan 2 fluorescence microscope with the Zeiss 40 × objective lens (Carl Zeiss AB, Stockholm, Sweden). Images were acquired with a Hamamatsu C4742-95 CCD digital camera (Carl Zeiss AB, Stockholm, Sweden) and the acquisition software QED Camera Plugin v1.1.6 (QED Imaging Inc., Rockville, MD, USA) and Volocity 1 (PerkinElmer, Waltham, MA, USA). The design of the specific probes for detecting *TGFB2-AS1* RNA was performed following the Stellaris RNA FISH Probe designer (Biosearch Technologies, Petaluma, CA, USA) and the probes were coupled to CAL Fluor® Orange 560 Dye.

Immunofluorescence

HaCaT or A549 cells were fixed in 3.7% (w/v) formaldehyde stabilized with 10% (v/v) methanol for 10 min at room temperature. The cells were permeabilized with 0.1% Triton X-100 for 10 min at room temperature and blocked in 5% FBS/PBS for 1 h at room temperature. Then, the samples were incubated with primary antibodies in 5% FBS/PBS overnight at 4°C. The next day, the coverslips were incubated with Alexa Fluor-488-labeled secondary antibody (Invitrogen, Thermo Fisher Scientific, Stockholm, Sweden) at a dilution of 1:1,000 in 5% FBS/PBS for 1 h at room temperature. Then, phalloidin staining was also performed using tetramethylrhodamine-isothiocyanate-conjugated phalloidin (dilution 1:1,000 in 5% FBS/PBS; Sigma-Aldrich AB, Stockholm, Sweden) for 20 min at room temperature, followed by DAPI staining (10 min at room temperature, 1:1,000 dilution). The coverslips were then mounted in Fluoromount-G and examined on a Zeiss Axioplan 2 fluorescence microscope with the Zeiss 40 × objective lens (Carl Zeiss AB, Stockholm, Sweden). Images were acquired with a Hamamatsu C4742-95 CCD digital camera (Carl Zeiss AB, Stockholm, Sweden) and the acquisition software QED Camera Plugin v1.1.6 (QED Imaging Inc., Rockville, MD, USA) and Volocity 1 (PerkinElmer, Waltham, MA, USA). The primary antibodies used were: anti-EED, clone 41D (1:50 dilution; Active Motif, 61203) and anti-Smad3 (C67H3, 1:100 dilution; Cell Signaling Technology, 9523S, Leiden, the Netherlands).

Immunoblotting

Total proteins were extracted using lysis buffer (20 mM Tris-HCl, pH = 8.0, 1% Nonidet P-40, 150 mM NaCl, 2 mM EDTA, and complete protease inhibitor mixture from Roche Diagnostics Scandinavia AB, Bromma, Sweden). The lysates were cleared by centrifugation at maximum speed for 10 min. The supernatant was transferred to new tubes and protein concentration was measured with

the bicinchoninic acid (BCA) assay. Next, 2 × sample buffer (0.12 M Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 0.01% bromophenol blue, 100 mM DTT) was added to the lysates, which were then boiled at 95°C for 5 min and subjected to SDS-PAGE. Equal amount of protein (40 μg) was loaded to each polyacrylamide gel. The resolved proteins were transferred to a nitrocellulose filter using a Bio-Rad wet or semidry transfer unit (Bio-Rad Laboratories AB, Solna, Sweden). Then the filters were incubated with primary antibodies and horseradish peroxidase-conjugated secondary antibodies and enhanced chemiluminescence assays were performed using the Millipore kit (Merck/Millipore, Stockholm, Sweden).

Chromatin immunoprecipitation

HaCaT cells expressing pcDNA3 (empty vector) or pcDNA3-TGFB2-AS1 were grown to 80% confluency in 15-cm dishes and cross-linked in 1% formaldehyde for 10 min at room temperature. The crosslinking was quenched by addition of 125 mM glycine for 5 min at room temperature. The cells were then washed and scraped in ice-cold PBS and centrifuged. The cell pellets were stored at −80°C until further use. The cell pellets were lysed in ChIP lysis buffer (50 mM Tris-HCl, pH 8, 10 mM EDTA, 1% SDS), supplemented with protease inhibitors (Roche Diagnostics Scandinavia AB, Bromma, Sweden). The lysates were subjected to sonication, in a water bath Diagenode Bioruptor sonicator (Diagenode, Bionordika, Stockholm, Sweden), with 30 s pulses (totally 5 min), so that the chromatin was sheared and DNA fragments could be generated. After sonication, the lysates were centrifuged at 14,000 rpm for 10 min at 4°C. The supernatant (90% of the volume) was diluted 10 times in ChIP dilution buffer (20 mM Tris-HCl, pH 8, 2 mM EDTA, 1% Triton X-100, 150 mM NaCl), supplemented with protease inhibitors (cOmplete EDTA-free protease inhibitor cocktail, Roche Diagnostics Scandinavia AB, Bromma, Sweden) and used for immunoprecipitation. The remaining 10% of the lysate was used as an input. Three μg of antibody, already pre-coupled with sheep anti-mouse IgG dynabeads M-280 (Invitrogen, Thermo Fisher Scientific, Stockholm, Sweden), in 0.5% BSA (IgG-free)/PBS solution (with overnight end-over-end rotation at 4°C) were used for the ChIP assay. The lysates with the antibody-beads complexes were incubated at 4°C overnight with end-over-end rotation. The precipitated complexes were washed 5 times in RIPA buffer (50 mM HEPES-KOH, pH 7, 0.5 M LiCl, 1 mM EDTA, 0.7% DOC, 1% NP-40), with an additional final wash in TE buffer (50 mM Tris-HCl, pH 8, 10 mM EDTA). Then, elution buffer was added to precipitated complexes, as well as to input samples, and samples were subjected to reverse crosslinking for 16 h at 65°C. After that, the immunoprecipitated DNA was purified, using the QIAquick PCR Purification kit, according to the protocol by the manufacturer (QIAGEN AB, Sollentuna, Sweden). The antibodies used for ChIP assays were: anti-H3K27me^3 (Abcam, Cambridge, United Kingdom), anti-H3K4me^3 (ab8580, Abcam, Cambridge, United Kingdom), anti-EED (Active Motif Europe, La Hulpe, Belgium), anti-EZH2 (Milipore/Merck, Stockholm, Sweden), anti-SUZ12 (Abcam, Cambridge, United Kingdom), anti-Smad2/3 (BD Biosciences-Europe, Stockholm, Sweden) and normal mouse IgG (Millipore/Merck, Stockholm, Sweden). A complete list of the oligonucleotides used for ChIP-qPCR assays is shown in Table S3.

Luciferase assays

HaCaT, HepG2 or A549 cells transiently transfected with the TGFβ/Smad-responsive CAGA$_{12}$ promoter reporter construct (CAGA$_{12}$-luc) were used for luciferase assays. The HaCaT CAGA$_{12}$-Luc/TK-Renilla cells, which stably express the CAGA$_{12}$-luc construct and the TK-Renilla luciferase reporter plasmid, were also used for luciferase assays. The BMP/Smad responsive promoter reporter (BRE$_2$-luc) construct was transiently transfected to *shControl* or *shTGFB2-AS1* HaCaT cells. The pCMV-β-gal construct, which encodes the β-galactosidase, was co-transfected with the transiently transfected promoter reporter plasmids for normalization of the firefly luciferase measurements. In the case of the HaCaT CAGA$_{12}$-Luc/TK-Renilla cells, the firefly luciferase values were normalized to the Renilla-luc values. Luciferase reporter assays were performed using the Firefly and Renilla Dual Luciferase Assay kit from Biotium, Fremont CA, USA (BTIU30003-2). Relative normalized luciferase activity, which derives from average values from triplicate determinations, with standard deviations, is presented in the graphs. Each experiment was repeated at least twice.

Thymidine incorporation assay

HaCaT cells expressing pcDNA3 (empty vector) or pcDNA3-TGFB2-AS1 were seeded in 12-well plates (30,000 cells per well) and treated with TGFβ1 for 24 h in 1% FBS/DMEM. Additional transfections with siRNAs were performed as indicated in the figure legends. The last 6 h of the 24-h time period, cells were incubated with 3 μCi/ml [^{3}H]thymidine. Then, cells were washed in PBS and treated with 5% trichloroacetic acid for 15 min at room temperature. Next, the cells were washed in distilled H_2O, dehydrated in 70% (v/v) ethanol and lysed in 0.1 M NaOH. The lysates were transferred to scintillation vials and the radioactivity of [^{3}H]thymidine was measured in a scintillation counter. The graphs represent average values with standard deviation of triplicate repeats for each condition. The experiments were repeated twice.

Invasion assays

Prior to the invasion experiment, A549 cells were first starved for 8 h in DMEM/0.1% FBS, then treated with TGFβ1 (5 ng/ml) for 24 h. Cells were trypsinized and seeded on the upper chamber of 24-well plate transwell inserts (6.5 mm diameter, 8 μm pore; Corning, NY, USA). The transwell inserts were either not coated or coated with 10 μg/ml laminin (Sigma-Aldrich Sweden AB, Stockholm, Sweden) and incubated at 37°C for 1 h. After trypsinization, 4 × 10^4 cells were seeded in DMEM/0.1% FBS in the upper chamber with or without TGFβ1 (5 ng/ml), and DMEM/6% FBS was placed in the lower chamber. After 20 h, cells that did not migrate were removed

with a cotton swab and inserts were fixed in methanol and stained with DAPI. For quantification, 10-15 pictures of each insert were taken at 20 × magnification, and the nuclei were counted using the ImageJ software (National Institutes of Health, Bethesda, MD, USA).

HaCaT cell invasion was analyzed as described above for A549 cells with the following alterations: cells were first treated with GSK343 in DMEM/0.1% FBS for 24 h, then media were changed to DMEM/0.1% FBS containing GSK343 (5 μM) with or without TGFβ1 (5 ng/ml) for 24 h. Cells on the upper chamber received DMEM/0.1% FBS with or without TGFβ1 (5 ng/ml) in the presence or absence of GSK343 (5 μM), and DMEM/6% FBS was placed in the lower chamber.

QUANTIFICATION AND STATISTICAL ANALYSIS

The results of RT-qPCRs, luciferase assays and thymidine incorporation assays represent the mean from at least three independent experiments (n = 3, whereby n equals the number of biological experiments). Each biological experiment included triplicate technical repeats. Error bars represent standard deviations (SD). For ChIP-qPCR and RIP-qPCR experiments, one representative experiment out of three biological repeats, is shown in the figures. In this case also, three technical replicates per each independent experiment were used to determine the average values and the SD. For immunoblotting experiments of plain lysates or after RNA pull-down or RIP, one representative result is presented. The exact number of repeated experiments is indicated in the Figure legends. For immunofluorescence and FISH experiments, only single (or two, FISH) representative photomicrographs are shown. Each independent immunofluorescence or FISH experiment was repeated 2 to 4 times (n = 2-4 biological repeats) and included single technical repeats; out of each technical repeat 4-5 independent photomicrographs were collected and compared subjectively for assessment of high reproducibility. Comparisons of quantitative measurements were performed using two-tailed paired Student's t test in Excel. In all experimental conditions statistical significance was accepted based on a p value smaller than 0.05. The exact degree of significance for each assay is indicated in the figure legends using stars, whereby (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The statistics for the invasion assays of Figures S4D and S7G were performed using one-way Anova test in the GraphPAd Prism 7 software, based on the Turkey's multiple comparison test. For the AmpliSeq transcriptome human gene expression analysis (Figure 4), details are given in the Method details. Differentially expressed genes were identified by performing EdgeR analysis using standard parameters. Adjusted p values (padj) for multiple testing, using Benjamini-Hochberg to estimate the false discovery rate (FDR), were calculated for final estimation of differential expression (DE) significance. Genes with an FDR < 0.05 and a logFC < −2 or > 2were selected for further analysis. In order to perform Gene Ontology analysis, the R package Clusterprofiler was used.

DATA AND CODE AVAILABILITY

The lncRNA, mRNA microarray and Ampliseq RNA sequencing data are deposited to Arrayexpress, EBI, UK under accession number E-MTAB-7773; the mass spectrometry primary data are published in this article.