TGF-β uses a novel mode of receptor activation to phosphorylate SMAD1/5 and induce epithelial-to-mesenchymal transition

Anassuya Ramachandran[1], Pedro Vizán[1,6], Debipriya Das[1,7], Probir Chakravarty[2], Janis Vogt[3], Katherine W. Rogers[4], Patrick Müller[4], Andrew P. Hinck[5], Gopal P. Sapkota[3] and Caroline S. Hill[1,8]

[1]Developmental Signalling Laboratory, [2]Bioinformatics and Biostatistics Facility, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK.

[3]Medical Research Council Protein Phosphorylation and Ubiquitylation Unit at the University of Dundee, Dundee DD1 5EH, UK.

[4]Friedrich Miescher Laboratory of the Max Planck Society, 72076, Tübingen, Germany.

[5]Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Avenue, Pittsburgh, PA 15260, USA.

[6]Present address: Center for Genomic Regulation, E-08003 Barcelona, Spain.

[7]Present address: Flow Cytometry, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK.

[8]Corresponding author: caroline.hill@crick.ac.uk

Keywords: ACVR1; Epithelial-to-mesenchymal transition; ID genes, SMAD1/5; TGF-β

Running title: The SMAD1/5 arm of TGF-β signaling

Abstract

The best characterized signaling pathway downstream of the transforming growth factor β (TGF-β) pathway is through SMAD2 and SMAD3. However, TGF-β also induces phosphorylation of SMAD1 and SMAD5, but the mechanism of this phosphorylation and its functional relevance is not known. Here, we show that TGF-β-induced SMAD1/5 phosphorylation requires members of two classes of type I receptor, TGFBR1 and ACVR1, and establish a new paradigm for receptor activation where TGFBR1 phosphorylates and activates ACVR1, which phosphorylates SMAD1/5. We demonstrate the biological significance of this pathway by showing that approximately a quarter of the TGF-β-induced transcriptome depends on SMAD1/5 signaling, with major early transcriptional targets being the *ID* genes. Finally, we show that TGF-β-induced epithelial-to-mesenchymal transition requires signaling via both the SMAD3 and SMAD1/5 pathways, with SMAD1/5 signaling being essential to induce ID1. Therefore, combinatorial signaling via both SMAD pathways is essential for the full TGF-β-induced transcriptional program and physiological responses.

Introduction

Members of the transforming growth factor β (TGF-β) family of ligands, which includes the TGF-βs, Activins, NODAL, BMPs and GDFs, have pleiotropic effects on cell behavior ranging from germ layer specification and patterning in embryonic development, to tissue homoestasis and regeneration in adults (Massague, 2012, Morikawa et al., 2016, Wu and Hill, 2009). TGF-β family signaling is also deregulated in a number of human diseases through mutation or altered expression of either the ligands or downstream signaling pathway components (Miller and Hill, 2016). In this context, the most widely studied pathology is cancer (Bellomo et al., 2016, Massague, 2008, Meulmeester and Ten Dijke, 2011, Wakefield and Hill, 2013), where TGF-β itself has both tumor suppressive and tumor promoting effects (Massague, 2008). At early stages of cancer TGF-β's tumor suppressive effects dominate, such as its cytostatic and pro-apoptotic functions (Padua and Massague, 2009). As tumors develop, however, mutations in key components of the pathway or downstream target genes allow the tumor to evade TGF-β's tumor suppressive effects, whilst remaining sensitive to its tumor promoting activities. TGF-β directly promotes the oncogenic potential of tumor cells, for example by driving epithelial-to-mesenchymal transition (EMT), a hallmark of cancer that enhances cell invasion and migration, and also increases the frequency of tumor-initating cancer stem cells (Massague, 2008, Ye and Weinberg, 2015). TGF-β's dual role in cancer thus provides an excellent example of how diverse responses can be elicited by a single ligand.

The TGF-β family ligands all signal via a common mechanism, initiated by ligand binding to two cell surface serine/threonine kinase receptors, the type II and type I receptors. In the receptor complex, the type II receptors phosphorylate and activate the type I receptors (Wrana et al., 1994). These in turn phosphorylate the downstream effectors of the pathway, the receptor-regulated SMADs (R-SMADs) on two serines in an SXS motif at their extreme C-terminus. Phosphorylated R-SMADs form complexes with the common SMAD, SMAD4, which accumulate in the nucleus and directly regulate the transcription of target genes, leading to new programs of gene expression (Shi and Massague, 2003). In the classic view of TGF-β family signaling there are two branches, characterized by distinct combinations of type II and type I receptors, and the recruitment of specific R-SMADs to particular type I

receptors (Wakefield and Hill, 2013, Shi and Massague, 2003). One branch is activated by TGF-β, Activins and NODAL and is mediated via the type I receptors TGFBR1, ACVR1B and ACVR1C (also known as ALK5, ALK4 and ALK7 respectively), which phosphorylate SMAD2 and 3. The other is activated by BMPs and GDFs, and is mediated via ACVRL1, ACVR1, BMPR1A and BMPR1B (also known as ALK1, ALK2, ALK3 and ALK6 respectively), which phosphorylate SMAD1, 5 and 9 (Miller and Hill, 2016).

In general, while this pairing between type I receptors and R-SMADs broadly fits the assignment of specific ligands to the different branches of TGF-β family signaling, it is an oversimplification. For example, ACVR1 is now described as a BMP receptor, but early work indicated that it could bind Activin and TGF-β (Massague, 1996, Miettinen et al., 1994), and very recently it has been shown to signal downstream of Activin in the context of the disease, fibrodysplasia ossificans progressiva (Hatsell et al., 2015, Hino et al., 2015). Furthermore, ACVRL1, a type I receptor that recognizes BMP9 and 10, also transduces TGF-β signals in endothelial cells (Pardali et al., 2010) by phosphorylating SMAD1/5 in parallel with the canonical TGF-β-induced phosphorylation of SMAD2/3 (Goumans et al., 2002, Goumans et al., 2003). This SMAD1/5 arm of TGF-β signaling has also been shown to occur in a wide range of other cell types, including epithelial cells, fibroblasts and cancer cell lines, which do not express ACVRL1 (Liu et al., 1998, Daly et al., 2008, Liu et al., 2009, Wrighton et al., 2009).

Important questions concerning this noncanonical TGF-β-induced SMAD1/5 phosphorylation remain unanswered. First, the mechanism by which TGF-β induces SMAD1/5 phosphorylation, in particular, the type I receptors involved, is not known. Some studies have concluded that the canonical TGF-β receptors TGFBR1 and TGFBR2 are sufficient for phosphorylation of both SMAD2/3 and SMAD1/5 (Liu et al., 2009, Wrighton et al., 2009). In contrast, others demonstrated that one of the classic BMP type I receptors (ACVR1 or BMPR1A), or in endothelial cells, ACVRL1, is additionally required (Daly et al., 2008, Goumans et al., 2002, Goumans et al., 2003). The second crucial issue concerns the biological relevance of TGF-β-induced SMAD1/5 signaling. Nothing is known about the transcriptional program activated by this arm of TGF-β signaling, or indeed, the specific SMAD complexes

involved. It is also not known to what extent this pathway is required for the physiological responses to TGF-β.

Here we dissect the SMAD1/5 arm of TGF-β signaling and define the underlying mechanism and its biological function. We show that TGF-β-induced SMAD1/5 phosphorylation requires both TGFBR1 and ACVR1 and using biosensors, and an optogenetic approach, we establish a new paradigm for TGF-β receptor activation. We have mapped the binding sites on chromatin of nuclear phosphorylated SMAD1/5 (pSMAD1/5) genome-wide, which led us to define the target genes regulated by this arm of TGF-β signaling. We go on to show that this arm of signaling is required for TGF-β-induced EMT. Our data reveal that the full transcriptional programme activated in response to TGF-β requires integrated combinatorial signaling via both the SMAD2/3 and SMAD1/5 pathways.

Results

The kinetics of TGF-β-mediated SMAD1/5 phosphorylation

To begin to dissect which receptors are required for TGF-β-induced SMAD1/5 phosphorylation, we compared the kinetics of SMAD1/5 and SMAD2 phosphorylation in response to TGF-β. Using the human breast cancer cell line MDA-MB-231 and the mouse mammary epithelial cell line NMuMG as model systems we found that TGF-β induced only transient phosphorylation of SMAD1/5 that peaked at 1 hr and then returned to baseline (Figure 1A). This was in contrast to a more sustained TGF-β-induced SMAD2 phosphorylation, or SMAD1/5 phosphorylation in response to BMP4. However, transient SMAD1/5 phosphorylation is not a defining characteristic of this arm of TGF-β signaling, as another human breast cancer line, BT-549, exhibited sustained SMAD1/5 phosphorylation that is still readily detectable 8 hr after TGF-β stimulation (Figure 1 – figure supplement 1A). Furthermore, when BT-549 cells were grown as non-adherent spheres, TGF-β-induced SMAD1/5 phosphorylation did not attenuate at all in the first 8 hr of signaling (Figure 1 – figure supplement 1A). Thus, the kinetics of TGF-β-induced SMAD1/5 phosphorylation is cell type-specific and dependent on the culture conditions and are independent of the kinetics of TGF-β-induced SMAD2/3 phosphorylation, suggesting a distinct receptor complex may be involved.

To address whether new protein synthesis was required for the transient nature of TGF-β-induced SMAD1/5 phosphorylation, cells were induced with TGF-β in the presence of either a translation inhibitor, cycloheximide or a transcription inhibitor, actinomycin D. Inhibition of translation was uninformative because it also led to a very rapid loss of TGFBR2 and TGFBR1, due to their short half-lives (Vizan et al., 2013). Use of actinomycin D, however, circumvented this problem, as *TGFBR2* and *TGFBR1* mRNAs are relatively stable (Figure 1 - figure supplement 1B) and their translation was unimpeded. In these conditions SMAD1/5 phosphorylation was sustained (Figure 1B; Figure 1 - figure supplement 1C). Thus, the rapid loss of pSMAD1/5 at later time points after TGF-β stimulation requires new transcription, suggesting that it is mediated by a component whose expression is induced by TGF-β.

Acute TGF-β stimulation results in the rapid internalization of the receptors, which is sufficient to deplete almost all of the type II receptor TGFBR2 from the cell surface (Vizan et al., 2013). As a result, cells are refractory to further acute TGF-β stimulation, read out by SMAD2 phosphorylation (Vizan et al., 2013). Cells in this refractory state were also unable to induce SMAD1/5 phosphorylation in response to TGF-β, although they remained responsive to BMP4 (Figure 1C, Figure 1 - figure supplement 1D). This suggested that TGFBR2 is required for TGF-β-induced SMAD1/5 activation.

TGF-β-induced SMAD1/5 phosphorylation requires the kinase activity of two different type I receptors

The distinct kinetics of TGF-β-induced SMAD1/5 phosphorylation compared with SMAD2/3 phosphorylation suggested that different receptor complexes are likely involved. To explore this further, we used combinations of well-characterized small molecule inhibitors of the type I receptor kinases in MDA-MB-231 cells. SB-431542, a selective TGFBR1/ACVR1B/ACVR1C inhibitor (Inman et al., 2002), completely inhibited the phosphorylation of both SMAD1/5 and SMAD2 in response to TGF-β when used at 10 μM (Figure 1D), indicating that the kinase activity of TGFBR1 is essential for TGF-β-induced SMAD1/5 phosphorylation. Interestingly, a 40-fold lower dose also substantially inhibited SMAD1/5 phosphorylation, whilst having less effect on SMAD2 phosphorylation (Figure 1D). TGF-β-induced SMAD1/5 phosphorylation was also partially inhibited by the BMP type I receptor inhibitor LDN-193189 (Cuny et al., 2008) (Figure 1D), establishing a requirement for a member of this class of type I receptors. Strikingly, the combination of the low dose SB-431542 and LDN-193189 completely inhibited TGF-β-dependent SMAD1/5 phosphorylation, without affecting phosphorylation of SMAD2 (Figure 1D). Analogous results were obtained in NMuMG cells (Figure 1 – figure supplement 1E).

We conclude that the kinase activity of both classes of type I receptor is required for maximal SMAD1/5 phosphorylation downstream of TGF-β. Taking these results together with the receptor expression profiles of these cells and receptor knockdown experiments (Daly et al., 2008), we deduce that the receptors involved are TGFBR1, a canonical BMP type I receptor (ACVR1 and/or BMPR1A) and TGFBR2.

SMAD1 is primarily phosphorylated by ACVR1

We next used an *in vitro* approach to explore why TGF-β-induced phosphorylation of SMAD1 requires two different type I receptors. We focused on ACVR1 as a representative of the BMP type I receptor class, as it is the most homologous to ACVRL1 that responds to TGF-β in endothelial cells (Chen and Massague, 1999). Moreover, in some cell types, knockdown of ACVR1 was sufficient to block TGF-β-induced pSMAD1/5 (Daly et al., 2008).

SMAD1 is known to be a poor substrate for TGFBR1 *in vivo* (Kretzschmar et al., 1997, Hoodless et al., 1996). We demonstrated that SMAD1 is also a poor substrate for TGFBR1 *in vitro,* although it is efficiently phosphorylated by both ACVR1 and BMPR1A as expected (Figure 1 – figure supplement 2A, B). As a control we showed that TGFBR1 could potently phosphorylate SMAD2, and surprisingly, ACVR1 was also able to phosphorylate SMAD2 (Figure 1 – figure supplement 2A, B).

Given that SMAD1 is a poor substrate for TGFBR1, it is intriguing that the kinase activity of TGFBR1 is essential for TGF-β-induced SMAD1 phosphorylation. We hypothesized that TGFBR1 might catalyze a priming phosphorylation on SMAD1, which then serves as a substrate for ACVR1, or *vice versa*. To address this we mapped the sites phosphorylated by ACVR1 on full length SMAD1. We identified three species of C-terminal SMAD1 phosphorylation by ACVR1 – a dually phosphorylated S[pS]V[pS] and the singly phosphorylated [pS]SVS and S[pS]VS (Figure 1 – figure supplement 2C). From this it was clear that ACVR1 could phosphorylate both serines in the critical SVS motif and we deduced that the order of phosphorylation is the penultimate serine of the motif, followed by the terminal one. Moreover, if the preceding serine was phosphorylated, it prevented the phosphorylation of the other sites.

Taking all these results together, we conclude that in response to TGF-β, the receptor kinase that phosphorylates SMAD1 is ACVR1 and not TGFBR1, and it does so on both serines in the SVS motif in a defined order.

ACVR1 is activated by TGFBR1 *in vitro* and *in vivo*

The absence of a role for the TGFBR1 kinase activity in phosphorylating SMAD1 left open the question of why it is required *in vivo* for TGF-β-induced SMAD1/5

phosphorylation. We postulated that it might be necessary for ACVR1 activation, and therefore investigated whether TGFBR1 could directly phosphorylate ACVR1. Both TGFBR1 and ACVR1 exhibit significant autophosphorylation activity *in vitro*, which was inhibited by SB-505124 (DaCosta Byfield et al., 2004) and LDN-193189 respectively (Figure 2A). Crucially, when TGFBR1 and ACVR1 were co-incubated, ACVR1 was phosphorylated, even in the presence of LDN-193189, indicating that ACVR1 *is* a substrate of TGFBR1 (Figure 2A).

To determine whether TGFBR1 could activate ACVR1 *in vivo* we used an optogenetic approach. To this end we fused the light-oxygen-voltage (LOV) domain of aureochrome1 from *Vaucheria frigida*, which dimerizes upon blue light stimulation (Sako et al., 2016), to the C-terminal ends of the intracellular domains of a constitutively-activated TGFBR1 (mutation T204D) (Wieser et al., 1995) and of wild type ACVR1, along with an N-terminal myristoylation motif to anchor them to the plasma membrane (Figure 2B; Supplementary Files 1 and 2). We refer to these constructs as Opto-TGFBR1* and Opto-ACVR1, respectively. We tested their ability, alone or in combination, to induce phosphorylation of SMAD1/5 in NIH-3T3 cells co-transfected with FLAG-SMAD1 to increase the range of the assay. Transfection of the Opto-ACVR1 alone resulted in no phosphorylation of co-transfected FLAG-SMAD1, either in the absence or presence of blue light. However, when Opto-ACVR1 and Opto-TGFBR1* were co-transfected, a robust light-inducible phosphorylation of FLAG-SMAD1 was observed (Figure 2C). Importantly, this was inhibited by both SB-505124 and LDN-193189, confirming the involvement of both receptors (Figure 2D). This directly demonstrates that TGFBR1 can activate ACVR1 *in vivo*. As a control, we showed that Opto-TGFBR1* phosphorylated co-expressed GFP-SMAD3 in the presence of light, which was inhibited by SB-505124, but to a much lesser extent by LDN-193189 (Figure 2E). As a further control to ensure that the activation of ACVR1 by TGFBR1 required the kinase activity of the latter, we made a kinase-dead version of Opto-TGFBR1. This construct was unable to induce the activity of ACVR1 in a light inducible manner and was also unable to induce phosphorylation of GFP-SMAD3 (Figure 2 – figure supplement 1).

To confirm that the light-inducible phosphorylation of FLAG-SMAD1 observed with the combination of Opto-ACVR1 and Opto-TGFBR1* genuinely resulted from activation of Opto-ACVR1 by Opto-TGFBR1*, we generated a mutant version of Opto-ACVR1, in which the serines and threonines of the GS domain were

mutated to alanine and valine respectively. Since phosphorylation of these serines and threonines is required for type I receptor activation, we would expect this mutant to be uninducible (Wieser et al., 1995). Indeed, we found that light-inducible phosphorylation of FLAG-SMAD1 was inhibited when this GS domain mutant of Opto-ACVR1 was used instead of the wild type Opto-ACVR1 (Figure 2F, G).

We therefore conclude that the requirement of the kinase activity of both TGFBR1 and ACVR1 for TGF-β-induced phosphorylation of SMAD1/5 reflects a requirement for activation of ACVR1 by TGFBR1 through phosphorylation of the ACVR1 GS domain.

TGF-β leads to clustering of ACVR1 and TGFBR1

Having shown that both type I receptors are required, we next tested whether they were components of the same tetrameric receptor complex, or whether they resided in separate receptor complexes that clustered at the cell membrane in response to ligand stimulation (compare model I and model II, Figure 3A). To distinguish between these possibilities we used previously published recombinant versions of TGF-β3, designated WW and WD (Huang et al., 2011). TGF-$\beta3^{WW}$, the wildtype TGF-β3 dimer, is composed of two identical monomeric TGF-β3 subunits, whereas TGF-$\beta3^{WD}$ contains one wildtype subunit of TGF-β3 and one mutated subunit that cannot bind to either TGFBR2 or TGFBR1 (Huang et al., 2011). Thus, while the TGF-$\beta3^{WW}$ ligand engages two type II:type I pairs in the tetrameric complex, the TGF-$\beta3^{WD}$ ligand can only engage one pair. In addition, TGF-$\beta3^{WW}$ does not bind ACVR1, and by inference, neither does TGF-$\beta3^{WD}$ (data not shown). It was previously demonstrated that TGF-$\beta3^{WD}$ binding to a single type II:type I receptor pair is sufficient to induce phosphorylation of SMAD3 (Huang et al., 2011). We therefore reasoned that if model I was correct then only TGF-$\beta3^{WW}$ would induce phosphorylation of SMAD1/5, as the heterotetrameric complex would not be able to be assembled with TGF-$\beta3^{WD}$. If model II was correct, however, then both TGF-$\beta3^{WW}$ and TGF-$\beta3^{WD}$ would be competent to induce pSMAD1/5. Treatment of MDA-MB-231 or NMuMG cells with either TGF-$\beta3^{WW}$ or TGF-$\beta3^{WD}$ led to a dose-dependent increase in both SMAD1 and SMAD2 phosphorylation (Figure 3B; Figure 3 – figure supplement 1). Thus, TGF-β stimulation is unlikely to lead to formation of a heterotetrameric complex comprising TGFBR2/TGFBR1/ACVR1, but instead,

leads to the formation of a higher order receptor cluster at the cell surface that includes TGFBR2/TGFBR1 complexes and ACVR1.

TGF-β induces ACVR1 activation *in vivo* in a TGFBR1-dependent manner

To obtain direct evidence that TGF-β activates ACVR1, we generated an ACVR1 biosensor that fluoresces when activated. In this construct ACVR1 is fused to the conformation-sensitive circularly permutated yellow fluorescent protein (cpYFP) core of the InversePericam Ca^{2+} sensor and FKBP1A (formerly FKBP12) to make ACVR1-InversePericam-FKBP1A (ACVR1-IPF) (Michel et al., 2011). When the receptor is inactive, the FKBP1A moiety binds to the GS domain of the receptor, which suppresses cpYFP fluorescence. Upon ligand induction, phosphorylation of the GS domain releases FKBP1A, allowing the cpYFP to adopt a fluorescent conformation (Michel et al., 2011). We first showed that ACVR1-IPF is functional in that it is able to induce phosphorylation of SMAD1/5 when overexpressed (Figure 4 – figure supplement 1A). We then stably expressed this biosensor in a number of cell lines (Figure 4 – figure supplement 1B). In the polarized epithelial cell line, MDCKII and in NIH-3T3 fibroblasts, ACVR1-IPF is readily detectable at the cell membrane, as well as in internal structures, and had no adverse effect on the inducibility of these cells in response to TGF-β or BMP4 (Figure 4 – figure supplement 1B–D). As a control we showed that ACVR1-IPF was activated in response to FK506 which binds FKBP1A and releases it from the GS domain of ACVR1 (Wang et al., 1994) (Figure 4 – figure supplement 1E). Treatment of the MDCKII ACVR1-IPF cells with TGF-β resulted in a significant increase in fluorescence that was inhibited by SB-431542 (Figure 4A and B; Videos 1–3). Furthermore, using flow cytometry for a more quantitative approach we demonstrated that the TGF-β-induced increase in fluorescence was blocked by both SB-431542 and a TGF-β neutralizing antibody and was independent of BMP signaling, as it was unaffected by the BMP antagonist Noggin (Figure 4C). Similarly, TGF-β also activated ACVR1 in NIH-3T3 ACVR1-IPF cells in a TGFBR1-dependent manner (Figure 4 – figure supplement 1F, G; Videos 4–6).

Mapping the binding sites on chromatin for TGF-β-induced pSMAD1/5 reveals that *ID* genes are major transcriptional targets of this pathway

Although the existence of TGF-β-induced pSMAD1/5 has been known for some time, its transcriptional role has never been addressed. Earlier experiments had suggested that TGF-β-induced pSMAD1/5 could only be found in complex with pSMAD2/3 (Daly et al., 2008), but using optimized immunoprecipitation conditions it was clear that TGF-β-induced pSMAD1/5 can also be part of pSMAD1/5–SMAD4 complexes (Figure 5A). We therefore used chromatin immunoprecipitation sequencing (ChIP-seq) for pSMAD1/5 to explore where in the genome pSMAD1/5 binds in response to TGF-β. We also wanted to determine which SMAD complexes were primarily responsible for regulating transcription in addition to the canonical pSMAD2/3–SMAD4 complexes (Figure 5A).

ChIP-seq in MDA-MB-231 cells for pSMAD1/5 and SMAD3 (as a control) resulted in 2378 pSMAD1/5 peaks and 2440 SMAD3 peaks identified in response to TGF-β after filtering (Figure 5 - Source data 1, sheet 1). The majority of the pSMAD1/5 peaks (2287) were also bound by SMAD3. To identify binding sites preferentially bound by pSMAD1/5 versus SMAD3 we calculated the ratio of the number of tags in the pSMAD1/5 peaks versus the SMAD3 peaks, and focused on the 100 peaks with the highest pSMAD1/5:SMAD3 tag ratio (Figure 5 - Source data 1, sheet 2). Interrogating the nearest genes to these peaks we found a significant enrichment of both TGF-β and BMP target genes (Figure 5 - Source data 1, sheets 2 and 3). Strikingly, 8 of the top 10 peaks flanked known BMP target genes (*ID1*, *ID3*, *ID4*, *ATOH8*, *BIRC3*) (Figure 5B; Figure 5 – figure supplement 1A; Figure 5 - Source data 1, sheet 2) (Gronroos et al., 2012). In contrast, classical TGF-β target genes like *JUNB*, *BHLHE40*, *PMEPA1*, *SERPINE1* (Levy and Hill, 2005) were not in this top 100 list, but were amongst those with the highest enrichment for SMAD3 (Figure 5B; Figure 5 – figure supplement 1A; Figure 5 - Source data 1, sheet 1). Using ChIP-qPCR, we validated these different binding patterns (Figure 5C; Figure 5 – figure supplement 1A). For pSMAD1/5, the binding in response to TGF-β was transient, peaking at 1 hr and thereafter decreasing, whilst SMAD3 binding at *JUNB* and *PMEPA1* was sustained. A subset of the peaks were also validated in BT-549 cells (Figure 5 – figure supplement 1B).

We performed motif enrichment analyses on the top 50 and 100 peaks with the highest pSMAD1/5:SMAD3 tag ratio. In both cases a SMAD1/5 binding motif GGCGCC was found (Figure 5D and E; Figure 5 – figure supplement 1C)

(Gaarenstroom and Hill, 2014). In addition, in the top 50 peaks the composite SMAD1/5–SMAD4 site was clearly identified (GGCGCC(N_5)GTCT) (Gaarenstroom and Hill, 2014, Morikawa et al., 2011) (Figure 5 – figure supplement 1C), with a slightly more degenerate version being present in the top 100 peaks (Figure 5D). This strongly suggests that TGF-β-induced SMAD1/5–SMAD4 complexes are responsible for regulating the genes with the highest enrichment of pSMAD1/5.

The enrichment of pSMAD1/5 on the *ID* genes in response to TGF-β suggests that they are *bona fide* target genes of this arm of TGF-β signaling. We confirmed this using siRNAs to deplete specific SMADs. TGF-β induction of *ID1* and *ID3* in MDA-MB-231 cells depended on SMAD1/5 and SMAD4, but not SMAD3 (Figure 5 – figure supplement 2A and B). In contrast, the induction of *JUNB* required SMAD3 and SMAD4, but was independent of SMAD1/5 (Figure 5 – figure supplement 2A and B). We further corroborated these observations using the drug dosing strategy that selectively inhibits SMAD1/5 phosphorylation in response to TGF-β (Figure 1D). The combination of low dose SB-431542 and LDN-193189 greatly decreased *ID* gene induction without impacting on the induction of *JUNB* in both MDA-MB-231 and NMuMG cells (Figure 5 – figure supplement 2C and D). The induction of target gene expression was also examined after treatment of cells with TGF-β3^{WW} or TGF-β3^{WD}. As expected both TGF-β ligands induced the expression of the *ID*s and *JUNB* (Figure 5 – figure supplement 2E).

The results in this section reveal that pSMAD1/5–SMAD4 complexes formed in response to TGF-β are responsible for regulating the genes with the highest enrichment of pSMAD1/5, and that the *ID*s are major early downstream targets.

The SMAD1/5 arm of TGF-β signalling is required for TGF-β-induced EMT

The ID proteins have been implicated in many processes involved in oncogenesis (Lasorella et al., 2014), and importantly, ID1 was shown to be upregulated by TGF-β in tumor cells isolated from pathological pleural fluids from patients with ER- and ER+ metastatic breast cancer, and also in patient-derived glioblastomas (Anido et al., 2010, Padua et al., 2008). Since we have now shown that the pSMAD1/5 arm of TGF-β signaling is responsible for TGF-β-induced *ID1* induction, this prompted us to explore further the biological relevance of the pSMAD1/5 arm of TGF-β signaling in oncogenic processes, and to gain a comprehensive view on the relative contribution of

this arm of signaling to longer term TGF-β responses. We decided to focus on the process of EMT, as this is a key step in tumorigenesis that confers a migratory phenotype, acquisition of stem cell properties and resistance to chemotherapeutic agents (Ye and Weinberg, 2015). For these studies we primarily used the NMuMG cell model, as we have shown above that these cells show a robust phosphorylation of SMAD1/5 in response to TGF-β and are well known to undergo a TGF-β-induced EMT within 48 hr (Piek et al., 1999).

CRISPR/Cas9 was used to generate clones of NMuMG cells deleted for SMAD1 and SMAD5 (Figure 6A; Figure 6 – figure supplement 1A–C). We compared the TGF-β-induced transcriptome at 48 hr of the parental clone with one deleted for SMAD1/5 using RNA-sequencing (RNA-seq). Of the 5798 genes that are significantly up- or down-regulated by TGF-β in this time frame we found that approximately a quarter (1398) were dependent on the SMAD1/5 branch of signaling (see Materials and Methods for the cut-offs used) (Figure 6 - Source data 1, sheets 1 and 2). This demonstrates that this arm of TGF-β signaling plays a crucial role in long term downstream transcription responses. To corroborate the RNA-seq results we validated a subset of them by qPCR, measuring levels of mRNA over time in response to TGF-β (Figure 6 – figure supplement 2).

Gene set enrichment analysis revealed that the TGF-β target genes that depend on this arm of signaling were involved in processes such as regulation of the cytoskeleton, focal adhesions, adherens and tight junctions, as well as TGF-β signaling in EMT (Figure 6 - Source data 1, sheet 3). We therefore next investigated whether TGF-β-induced EMT required SMAD1/5 signaling. Using delocalization of the adherens junction marker CDH1 (also called E-Cadherin) together with loss of the tight junction marker TJP1 (also called ZO-1) as a measure of EMT, we could readily demonstrate that signaling through SMAD1/5 was crucial for this process in two separate ΔSMAD1/5 clones (Figure 6B; Figure 6 – figure supplement 1D). In addition, we observed that two mesenchymal markers, *Acta2* (also called smooth muscle actin) and *Fn1* were more weakly induced in the ΔSMAD1/5 clone compared with the wild type (Figure 6 – figure supplement 2). We also used an siRNA knockdown approach, and showed that EMT was dependent on SMAD1/5, SMAD4 and SMAD3, but independent of SMAD2 (Figure 6 – figure supplement 3A and B). Furthermore, treatment of the cells with the BMP type 1 receptor inhibitor, LDN-

193189 also inhibited EMT either alone or when combined with low dose SB-431542 which we have shown is sufficient to inhibit TGF-β-induced SMAD1/5 signaling, but not signaling through SMAD2/3 (Figure 6C and D; Figure 6 – figure supplement 3C). Moreover DMH1, another BMP type 1 receptor inhibitor, had a similar effect (Figure 6 – figure supplement 3C and D). Finally, to confirm that that the dependence of TGF-β induced EMT on SMAD1/5 signaling was not unique to NMuMG cells, we used another mouse mammary cell line, EpRas that also undergoes a TGF-β-induced EMT (Daly et al., 2010, Grunert et al., 2003). SMAD1/5 signaling in this line was also essential for EMT (Figure 6E and F). Thus, we conclude that TGF-β-induced EMT requires the SMAD1/5 arm of the signaling pathway, as well as the canonical pathway through SMAD3.

Taking our ChIP-seq and RNA-seq analyses together, we found that the *ID* genes are major early transcriptional targets of the SMAD1/5 arm of the TGF-β pathway. Of these, ID1 was the prominent family member up regulated by TGF-β in NMuMGs (Figure 7 – figure supplement 1A). We hypothesized that the dependency on the SMAD1/5 arm of the TGF-β pathway could reflect a requirement of ID1 for EMT. We tested this by knocking down *Id1* with siRNAs, both as a pool and as individual siRNAs and found that cells depleted of ID1 were indeed unable to undergo TGF-β-induced EMT (Figure 7A and B; Figure 7 figure supplement 1B and C). Thus, we conclude that TGF-β-induced up-regulation of ID1 is essential for EMT.

Discussion

Combinatorial signaling downstream of TGF-β

Here we have defined both the mechanism whereby TGF-β induces the phosphorylation of SMAD1/5, and its functional role. We have shown that two type I receptors are required, the canonical TGF-β receptor TGFBR1, and additionally, one of the classical BMP type I receptors, ACVR1. Using *in vitro* kinase assays, an optogenetic approach and an ACVR1 receptor fluorescent biosensor, we have uncovered a new mechanism for receptor activation whereby one type I receptor activates another. We show that in response to TGF-β, TGFBRI phosphorylates and activates ACVR1, which phosphorylates SMAD1/5. To address the functional significance of this arm of TGF-β signaling we used genome-wide ChIP-seq and RNA-seq and show that approximately a quarter of the TGF-β-regulated transcriptome is dependent on SMAD1/5, with major early targets being the ID transcriptional regulators. Finally, we have also demonstrated that the SMAD1/5 pathway is essential for TGF-β-induced EMT, and this reflects a requirement for ID1.

Taking these results together with previous work (Liu et al., 2009, Daly et al., 2008) we propose a model of combinatorial signaling that is essential for the TGF-β cellular program (Figure 7C). In most cells tested the induction of pSMAD1/5 is more transient than the pSMAD2/3 induction (Liu et al., 2009, Daly et al., 2008). Thus, the initial transcriptional program is regulated by both SMAD pathways and is refined at later time points by the SMAD2/3 pathway. Therefore, the full TGF-β-induced transcriptional program requires combinatorial signaling via both SMAD pathways. With respect to the functional relevance of TGF-β-induced SMAD1/5 phosphorylation, we have now shown that complete EMT requires both SMAD pathways. TGF-β-induced anchorage-independent growth, migration and invasion have also been shown to require SMAD1/5 signaling, whilst TGF-β-induced growth arrest is only dependent on SMAD2/3 signaling (Liu et al., 2009, Daly et al., 2008) (Figure 7C).

Since we have now demonstrated that TGF-β induces the formation of SMAD1/5–SMAD4 complexes that regulate canonical BMP target genes, it is important to ask what discriminates TGF-β signaling from BMP signaling as it is well

known that BMP and TGF-β functional responses are distinct (Itoh et al., 2014, Miyazono et al., 2010). The answer lies in the combinatorial signaling, and likely also in the signaling dynamics. In contrast to TGF-β, BMP stimulation leads to a sustained phosphorylation of SMAD1/5 in the absence of SMAD2/3 activation (Gronroos et al., 2012, Daly et al., 2008). As a result, although the gene expression program downstream of BMP shares some common targets with that downstream of TGF-β at early time points, it will be completely distinct at later time points as a result of the sustained SMAD1/5 signaling and the absence of SMAD2/3-driven transcription (Figure 7C).

Receptor requirements for TGF-β-induced SMAD1/5 phosphorylation

We have shown that two classes of type I receptors are necessary for TGF-β-induced SMAD1/5 phosphorylation, the canonical TGF-β receptor, TGFBR1 and one of the BMP type I receptors, of which we have focused on ACVR1. Our results demonstrate that the kinase activity of TGFBR1 is essential for activation of ACVR1, whereas the kinase activity of ACVR1 is necessary to phosphorylate SMAD1/5. Surprisingly, we found that inhibition of TGF-β-induced SMAD1/5 phosphorylation by LDN-193189, which inhibits the BMP type I receptors, is incomplete, even though the same LDN-193189 concentration is sufficient to inhibit BMP-induced SMAD1/5 phosphorylation. This same result was also previously seen when the BMP type I receptor inhibitor dorsomorphin was used (Daly et al., 2008). A complete inhibition of TGF-β-induced pSMAD1/5 is achieved by combining LDN-193189 with a sub-optimal dose of SB-431542. This is likely explained by the fact that LDN-193189-inhibited ACVR1 is still able to efficiently recruit SMAD1/5, where it may be inefficiently phosphorylated by TGFBR1, which is sensitive to the sub-optimal dose of SB-431542. The requirement for two distinct type I receptors fits well with what was shown for TGF-β responses in endothelial cells, where ACVLR1 and TGFBR1 were both required (Goumans et al., 2003, Goumans et al., 2002).

Our optogenetic experiments revealed that activated TGFBR1 phosphorylates and activates ACVR1 *in vivo*. We previously hypothesized that TGFBR1 and ACVR1 could be in the same receptor complex (Daly et al., 2008), but our use of the mutant TGF-β3 ligands here indicated that these two type I receptors are not part of an obligate heterotetrameric receptor, but rather that TGFBR1, activated by TGFBR2

as a result of TGF-β stimulation can phosphorylate and activate ACVR1 in the membrane as a result of receptor clustering (Figure 7C). We were surprised to see in our optogenetic experiments that light-induced dimers of the activated kinase domain of TGFBR1 were much more active than the monomeric domains, as this suggests that TGFBR1 is able to autophosphorylate and auto-activate in the absence of type II receptors, if brought into close proximity. In fact, a similar observation was made in early studies using chimeric receptors with the extracellular domain of the erythropoietin receptor and the intracellular domain of constitutively active TGFBR1 (Luo and Lodish, 1996). This chimeric receptor could only mediate a growth arrest after stimulation with erythropoietin, indicating that clustering was important for receptor activity *in vivo*.

Dynamics of TGF-β-induced pSMAD1/5 signaling

We and others have observed that in most cell types TGF-β-induced SMAD1/5 phosphorylation is transient compared with SMAD2/3 phosphorylation (Daly et al., 2008, Liu et al., 2009, Wrighton et al., 2009). Using pSMAD2 as a readout, we previously showed that pSMAD2 levels attenuate over time, and remain at a low steady state level that depends on receptors replenishing the cell surface, for as long as ligand is available (Vizan et al., 2013). Our demonstration that levels of fluorescence of the ACVR1-IPF biosensor steadily increase over a number of hours indicates that ACVR1 can also be continuously activated for as long as ligand is present. We have shown that the transience of SMAD1/5 phosphorylation requires new protein synthesis, indicating that SMAD1/5 phosphorylation is likely to be actively terminated by an inhibitor induced by the pathway. Given the prolonged activation of ACVR1-IPF in response to ligand, we hypothesize that such an inhibitor is unlikely to target the receptors, but might be a TGF-β-induced phosphatase that targets phosphorylated SMAD1/5 directly. The transience of SMAD1/5 phosphorylation is not a defining characteristic of this arm of TGF-β signaling as BT-549 breast cancer cells exhibit a more sustained response, which is even more pronounced when the cells are grown as spheres. Comparing TGF-β target genes in BT-549s versus MDA-MB-231s where the response is transient, might shed light on the identity of the putative inhibitor.

TGF-β-induced pSMAD1/5 is transcriptionally active and required for a subset of TGF-β-induced target genes

Our ChIP-seq analysis demonstrated for the first time that TGF-β-induced pSMAD1/5 accumulates in the nucleus and binds to chromatin. These experiments revealed that the peaks with the highest pSMAD1/5 enrichment flanked classical BMP target genes, such as *ID1*, *ID3* and *ATOH8*. Analysis of the binding sites led us to the discovery that the SMAD complexes responsible for inducing these target genes downstream of TGF-β were pSMAD1/5–SMAD4 complexes. The ChIP-seq analysis also revealed widespread co-binding of pSMAD1/5 and SMAD3, which was surprising. For the classical BMP targets, the ratio of pSMAD1/5:SMAD3 in the peaks was high, whereas at classical TGF-β targets like *JUNB*, *PMEPA1*, *SERPINE1* and *BHLHE40* (Kang et al., 2003, Levy and Hill, 2005), this ratio was less than 1. We do not fully understand the functional significance of the pSMAD1/5 and SMAD3 co-binding. We previously demonstrated that at least in some contexts, pSMAD3–pSMAD1/5 complexes are inhibitory (Gronroos et al., 2012), and this is evident in the work presented here for *ID3* induction. However, for *JUNB* we found that knockdown of SMAD1/5 had no effect on TGF-β-induced transcription, suggesting that pSMAD1/5 is not contributing to its transcriptional regulation. This may also be true of other genes with a similar pattern of SMAD3/pSMAD1/5 binding.

TGF-β-induced SMAD1/5 signaling is required for EMT through induction of ID1.

We have now shown that SMAD1/5 signaling in response to TGF-β is required for a complete TGF-β-induced EMT in NMuMG cells and in EpRas cells. This accounts for a previously unexplained observation that overexpression of dominant negative ACVR1 in NMuMGs caused a partial loss of EMT in response to TGF-β (Miettinen et al., 1994). In an earlier study using siRNAs we had concluded that the SMAD1/5 arm of the TGF-β pathway was not required for EMT in EpRas cells (Daly et al., 2008). The likely explanation for this discepency is the poor SMAD1/5 knockdown we achieved in those cells compared with the very effective strategy of inhibiting this arm of TGF-β signaling using the combined small molecule inhibitors that we have employed here.

We have gone on to show that TGF-β-induced ID1 is required for EMT. Importantly, although ID1 is necessary for EMT, it is clearly not sufficient, as BMP cannot induce EMT in NMuMGs (Kowanetz et al., 2004). Consistent with this we have also shown that the SMAD3 pathway is essential for EMT. This arm of the pathway is likely required for the induction of some or all of the so-called EMT-associated transcription factors, most notably SNAI1, SNAI2, ZEB1, ZEB2 and BHLH proteins such as TWIST1 and E47 (now called TCF3), some of which are known direct TGF-β targets (Peinado et al., 2007, Diepenbruck and Christofori, 2016).

Our finding that EMT depends on TGF-β-induced ID1 expression has implications for the role of SMAD1/5 and the IDs in cancer. The prevailing view is that ID1 is downregulated by TGF-β in non-tumorigenic human epithelial lines, but upregulated by TGF-β in established tumor cell lines, as we have observed here in MDA-MB231 and BT-549s, and also in patient-derived tumor cells (Anido et al., 2010, Padua et al., 2008, Lasorella et al., 2014). Furthermore, ID proteins are overexpressed in many different tumor types, and are implicated in the maintenance of tumor stem cells and for some cancer-related phenotypes (Lasorella et al., 2014). ID1 was also found in a lung metastatic gene signature of breast cancer (Minn et al., 2005). The role of ID1 in EMT is context dependent. In a recent study of breast cancer, ID1 was shown to be expressed in tumor cells that had already undergone an EMT, and it contributed to the growth of the primary tumor by inducing a stem cell-like phenotype. At the metastatic site however, TGF-β-induced ID1 was proposed to induce an mesenchymal-to-epithelial transition (MET) by interferring with the activity of TWIST (Stankic et al., 2013). In light of our current data it will be important to investigate in what tumor contexts ID1 is required for EMT, and more broadly how the TGF-β–SMAD1/5 pathway contributes to different aspects of tumorigenesis.

Materials and Methods

Cell line origin, authentication and maintenance

MDA-MB-231 cells were obtained from the ECACC/HPA culture collection, BT-549 cells were obtained from the Francis Crick Institute Cell Services, NMuMG cells were obtained from ATCC, MDCKII cells were obtained from Sigma, NIH-3T3 cells were obtained from Richard Treisman (Francis Crick Institute) and EpRas cells were obtained from Martin Oft and Hartmut Beug (IMP, Vienna). All cell lines have been banked by the Francis Crick Institute Cell Services, and certified negative for mycoplasma. In addition, MDA-MB-231 and BT-549 cells were authenticated using the short tandem repeat profiling, while MDCKII, NIH-3T3 and EpRas cells had species confirmation at the Francis Crick Institute Cell Services. Their identity was also authenticated by confirming that their responses to ligands and their phenotype were consistent with published history.

MDA-MB-231, BT-549, EpRas, NIH-3T3 and MDCKII cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin. NMuMG cells were grown in the same medium, but supplemented with 10 µg/ml insulin. MDA-MB-231 and MDCKII cells were starved overnight in OptiMEM prior to ligand stimulation; NMuMG cells were starved overnight in OptiMEM with 10 µg/ml insulin; NIH-3T3 cells were starved in DMEM with 0.5% FCS. For ligand stimulation experiments, BT-549 cells were plated in the mammosphere culture media (Dontu et al., 2003) (MEBM (PromoCell) with B27 (Thermo Fisher), 20 ng/ml EGF (PeproTech), 20 ng/ml bFGF (PeproTech) and 4 µg/ml heparin (Sigma)).

Ligands and chemicals

All recombinant ligands were reconstituted in 4.4 mM HCl supplemented with 0.1% BSA. Cells were treated with recombinant TGF-β1 (PeproTech, 100-21C; 2 ng/ml), BMP4 (PeproTech, 120-05ET; 20 ng/ml) and Noggin (PeproTech, 250-38; 300 ng/ml). TGF-$\beta3^{WW}$ and TGF-$\beta3^{WD}$ were as described (Huang et al., 2011). SB-431542 (Tocris) was used at the concentrations indicated, SB-505124 (Tocris) at 10 or 50 µM, LDN-193189 (a gift from Paul Yu) at 1 or 0.5 µM, DMH1 (Selleck Chemicals) at 1 µM, cyclohexamide (Sigma) at 20 µg/ml and actinomycin D (Sigma)

at 1 µg/ml. For TGF-β blocking experiments, the pan-TGF-β blocking antibody (1D11) and the control antibody (13C4) were used at 30 µg/ml (Nam et al., 2008).

CRISPR/Cas9 knockout of SMAD1/5 and ACVR1/BMPR1A in NMuMG cells

From the wildtype NMuMG cells, a parental clone was selected that expressed robust junctional markers (TJP1 and CDH1) and underwent an efficient EMT in response to TGF-β. Two guide RNAs (see Key Resources Table) targeting the MH1 domain (SMAD1) and MH2 domain (SMAD5) were expressed from the plasmid pSpCas9(BB)-2A-GFP (PX458) (Ran et al., 2013) and used to knockout SMAD1 and SMAD5. NMuMG parental clone cells were simultaneously transfected with both plasmids, sorted for GFP expression, plated as single cells in 96-well plates and screened by sequencing to verify mutations in SMAD1 and SMAD5. Two knockout clones, ΔSMAD1/5 clone 1 and 2, were used in these studies. The same parental clone of NMuMG cells was also used to generate a line knocked out for ACVR1 and BMPR1A. The strategy was as described for the SMAD1/5 knockout and the guides are given in the Key Resources Table.

Generation of cell lines stably expressing ACVR1-IPF

The inverse pericam FKBP1A (IPF) fusion protein was amplified by PCR from the pCS2+zALK3-IPF (Michel et al., 2011) and cloned in-frame downstream of the human ACVR1 cDNA sequence in the pcDNA3.1 Hygro + vector (Thermo Fisher). MDCKII and NIH-3T3 cells were transfected with the ACVR1-IPF construct and selected with 400 µg/ml hygromycin or 40 µg/ml hygromycin respectively. After selection, cells were FACS sorted for GFP expression. MDCKII ACVR1-IPF cells were maintained as a pool, while a single clone was isolated for NIH-3T3 cells. To test the functionality of ACVR1-IPF, NMuMG cells knocked out for ACVR1 and BMPR1A were transfected with empty pcDNA3.1 Hygro (+), ACVR1-IPF or FLAG-ACVR1 (Daly et al., 2008) as a positive control, and activity was monitored by their ability to induce phosphorylation of SMAD1/5.

Generation and LED light photoactivation of Opto-receptors

The general design of the Opto receptors was a previously described (Sako et al., 2016). Opto-TGFBR1* and Opto-ACVR1 were generated by overlapping PCR

(Horton et al., 1990) to include a N-terminal myristyolation domain, the intracellular domain of either human TGFBR1 (residues 149-503) or human ACVR1 (residues 147-509), a light-oxygen-voltage (LOV) domain from *Vaucheria frigida* (Takahashi et al., 2007) and a C-terminal HA-tag and cloned into the pCS2 expression plasmid (see Supplementary Files 1 and 2). In the case of TGFBR1, the T204D point mutation was introduced that renders the kinase constitutively active (Wieser et al., 1995), thus generating the construct Opto-TGFBR1*. A kinase dead version of Opto-TGFBR1 was also generated in which K232 was mutated to R (Wrana et al., 1994). Furthermore, the GS-domain of ACVR1 (^{189}TSGSGS^{194}) was mutated to VAGAGA to generate Opto-ACVR1 GS-mut. NIH-3T3 cells were transfected with a total of 2 μg of plasmid DNA that included either 5 ng of GFP-SMAD3 (Nicolas et al., 2004) or 25 ng of Flag-SMAD1 (Lechleider et al., 2001) alone or in combination with 25 ng of Opto-TGFBR1* and/or 50 ng of Opto-ACVR1 (WT or GS-mut). We co-transfected the SMADs with the Opto-receptors to increase the range of the assay. Twenty-four hours post transfection, cells were starved overnight in DMEM with 0.5% FCS. Cells were then left untreated or pre-treated with 0.5 μM LDN-193189 or 50 μM SB-505124 and then exposed to blue light from an LED array for 1 hr at 37^{o}C in a humidified incubator. Control cells (i.e. in the dark) were wrapped in aluminium foil and placed in the same incubator.

siRNAs and transfections

All siRNAs were purchased from Dharmacon/GE Health Care Life Sciences and are listed in Supplementary file 3. MDA-MB-231 and NMuMG cells were transfected with siRNAs at a final concentration of 20 nM with Interferin (Polyplus). Twenty four hours post transfection, cells were starved overnight, and the following day cells were treated with TGF-β or BMP-4 for 1 hr and RNA and/or protein extracted. NMuMG cells were also treated with TGF-β for a further 24–48 hr to assess the effects of target gene knockdown on EMT.

EMT assay

NMuMG or EpRas cells were plated on glass coverslips in 6-well plates (200,000 or 75,000 cells respectively). For NMuMG cells treated with small molecule inhibitors, the media was changed the day after plating to OptiMEM with 10 μg/ml insulin and

the cells treated with 2 ng/ml TGF-β alone or in combination with 0.125 μM SB-431542, 1 μM LDN-193189 or DMH1 for the durations indicated. For knockdown experiments, NMuMG cells were transfected the day after plating with the indicated siRNAs. Twenty-four hours after transfection, the media was changed to OptiMEM with 10 μg/ml insulin and the following day, cells were treated with TGF-β for the durations indicated. For EpRas, cells were treated with 2 ng/ml TGF-β alone or in combination with 0.125 μM SB-431542 and 1 μM LDN-193189 the day after plating. EpRas cells were then split and re-plated at the initial splitting density in the presence of 2 ng/ml TGF-β alone or in combination with SB-431542 and LDN-193189 every three days.

Antibodies, immunoblotting, immunoprecipitations and indirect immunofluorescence

All primary and secondary antibodies used are listed in the Key Resources Table. Western blots using whole cell extracts and immunoprecipitations followed by Western blotting were as previously described (Germain et al., 2000, Daly et al., 2008). Indirect immunofluorescence of the ACVR1-IPF was performed after fixing cells in 4% formaldehyde for 5 minutes. Indirect immunofluorescence for CDH1 and TJP1 was performed after fixation in methanol:acetone (1:1) as previously described (Nicolas and Hill, 2003). Nuclei were counter stained with DAPI (0.1 μg/ml). Imaging was performed on a Zeiss Upright 780 confocal microscope. Z-stacks were acquired for all channels and maximum intensity projection images are shown.

Live cell imaging

Live cell imaging was performed for MDCKII ACVR1-IPF and NIH-3T3 ACVR1-IPF cells on a Zeiss Invert 780 confocal microscope. Cells were plated on 35 mm Matek dishes and starved overnight in phenol-free, HEPES-buffered DMEM with 0.5% FCS. During imaging, the temperature was maintained at 37^{o}C. Data were acquired every 15 minutes over a time course. At each time point, a z-stack was acquired, and maximum intensity z-projections were quantified with ImageJ.

Flow cytometry

MDCKII ACVR1-IPF and NIH-3T3 ACVR1-IPF cells were treated with ligand ± inhibitors. Twenty four hours post treatment, cells were trypsinized, washed and analyzed for GFP/YPF fluorescence on a LSRII flow cytometer (BD Biosciences), gated for viable, single cells. Treatment with FK506 (Sigma) was performed for 4 hr prior to analysis.

Recombinant proteins, *in vitro* kinase assays and mapping of phospho-sites

Recombinant SMAD proteins were expressed in *E. coli* and purified as previously described (Ross et al., 2006). Recombinant intracellular domains of ACVR1, BMPR1A and TGFBR1 which were expressed in insect cells were purchased from Carna Biosciences Inc (see Key Resources Table). Radioactive kinase reactions were performed with varying amounts of receptor (25–200 ng) at 37^{o}C for 1 hr in a 20 µl reaction volume with 50 mM Tris-Cl (pH 7.5), 50 mM NaCl, 5 mM $MnCl_2$ (ACVR1 and TGFBR1) or $MgCl_2$ (BMPR1A), 16.5 nM ^{32}P-γ-ATP (Perkin Elmer; NEG502A500UC) and either 200 µM or 50 µM cold ATP. Substrates were either the receptors themselves (autophosphorylation) or 2 µg of recombinant SMAD proteins. Reactions were stopped by adding Laemmli sample buffer and heating to 95°C for 5 minutes. Proteins were resolved on a NuPAGE Novex 4-12% Bis-Tris gradient gel (Thermo Fisher) and stained with Colloidal Blue (Thermo Fisher). Gels were destained, dried and radioactivity measured by autoradiography.

To map phosphorylated residues on SMAD1, radioactive kinase reactions were performed in triplicate with 200 ng ACVRI, 2 µg recombinant SMAD1, 200 µM cold ATP, 0.33 µM ^{32}P-γ-ATP. For phospho-residue mapping, ^{32}P-labelled SMAD1 was digested with trypsin, the peptides were resolved by HPLC with an acetonitrile gradient and the ^{32}P-labelled peptides eluted. Edman sequencing and mass-spectrometry (Orbitrap Classic, Thermo Fisher) were then used to confirm phospho-residues, as described previously (Campbell and Morrice, 2002), with the addition of multi-stage activation during the MS2 analysis.

Chromatin immunoprecipitations, ChIP-Seq and motif enrichment

Four million MDA-MB-231 or BT-549 cells were plated; 24 hr later, cells were starved overnight and the following day treated with TGF-β or BMP-4. One 15 cm plate was used per immunoprecipitation. Chromatin immunoprecipitations, ChIP-seq

library preparation, next generation sequencing and data analysis were performed in biological duplicate essentially as previously described (Coda et al., 2017). In brief, ChIP-seq was performed on an Illumina HiSeq2500 generating 50 bp single end reads. Reads were aligned to the human GRCh37/hg19 genome assembly using BWA version 0.6 (Li and Durbin, 2009) with a maximum mismatch of 2 bases. Picard tools version 1.81 (http://sourceforge.net/projects/picard/) was used to sort, mark duplicates and index the resulting alignment bam files. Normalized tdf files for visualization purposes were created using IGVtools software (Robinson et al., 2011) (http://software.broadinstitute.org/software/igv/) by extending reads by 50 bp and normalizing to 10 million mapped reads per sample. Peaks were called by comparing stimulated samples to the respective untreated samples using MACS version 1.4.2 (Zhang et al., 2008), using mfold change parameters of between 5 and 30. Peaks called by MACS were annotated using the annotatepeaks command in the Homer software (Heinz et al., 2010); http://homer.salk.edu/homer/).

Peaks with less than 20 tags in the pSMAD1/5 IP after TGF-β treatment or less than 30 tags in the SMAD3 IP after TGF-β treatment were excluded from the analysis. In addition, peaks that had less than 1 tag per 10 bp in either of the above conditions were also excluded. Finally a ratio was taken between the number of tags in the pSMAD1/5 IP and the number of tags in the SMAD3 IP after TGF-β treatment to determine the top 100 peaks with preferential SMAD1/5 binding. Of these, the top 50 peaks with the highest density of tags per 10 bases in the pSMAD1/5 IP after TGF-β treatment were used for more refined motif enrichment analysis and gene annotation.

Motif enrichment was performed using MEME (http://meme-suite.org/) with default parameters (zero or one occurrence per sequence, motifs between 6-50 bases in width).

RNA-sequencing analysis in the NMuMG parental clone and ΔSMAD1/5 clone 1

NMuMG parental and ΔSMAD1/5 clone 1 were plated, starved the next day in OptiMEM with 10 µg/ml insulin and treated for a further 48 hr with 2 ng/ml TGF-β. Total RNA was extracted as previously described (Gronroos et al., 2012), DNase I (Qiagen) treated and cleaned up with RNeasy columns (Qiagen). Biological triplicate libraries were prepared using the TruSeq RNA Library Prep Kit (Illumina) and were

single-end sequenced on Illumina HiSeq 2500 platform. Sequencing yield was typically ~80 million strand-specific reads per sample. The RSEM package (version 1.2.31) (Li and Dewey, 2011) in conjunction with the STAR alignment algorithm (version 2.5.2a) (Dobin et al., 2013) was used for the mapping and subsequent gene-level counting of the sequenced reads with respect to Ensembl mouse GRCm.38.86 version genes. Normalization of raw count data and differential expression analysis was performed with the DESeq2 package (version 1.10.1) (Love et al., 2014) within the R programming environment (version 3.2.3) (R Development Core Team, 2009). Genes were first identified as differentially expressed in the parental clone if they had more than 10 reads in either the untreated or TGF-β treated samples and a fold change between untreated and TGF-β induced of > 1.5 or < 0.75 and FDR < 0.05. An interaction contrast was then used to determine differentially regulated genes after TGF-β treatment in the parental clone versus ΔSMAD1/5 clone 1. The resulting gene lists ranked by the Wald statistic were used to look for pathway and biological process enrichment using the Broad's GSEA Tool (Subramanian et al., 2005). Genes with a fold difference between the two clones after TGF-β treatment of > 1.5 or < 0.75 and an FDR < 0.05 were judged to be dependent on SMAD1/5.

Public availability of data

The ChIP-seq data have been submitted to the NCBI Gene Expression Omnibus (GEO) under the accession number GSE92443. The RNA-seq data has been submitted to GEO under the accession number GSE103372.

qPCR

Oligonucleotides used are listed in Supplementary file 3. Total RNA extraction and reverse transcription were performed as previously described (Gronroos et al., 2012). The cDNA was diluted 10-fold and then used for quantitative PCR (qPCR). All qPCRs were performed with Express Sybr Greener (Thermo Fisher) with 300 nM of each primer and 2 µl of diluted cDNA or eluted immunoprecipitated chromatin. Fluorescence acquisition was performed on a 7500 FAST machine (Thermo Fisher). Quantification for relative gene expression was done using the comparative Ct method with target gene expression normalized to *GAPDH*. Quantification for ChIPs was performed using a standard curve and presented normalized to input.

Statistical analysis

Western blots, immunofluorescence experiments and ChIP-PCRs are representative of at least two biological replicate experiments. All qPCRs are the mean and SEM of three independent biological experiments except gene expression after actinomycin D treatment and stimulation with TGF-$\beta3^{WW}$ and TGF-$\beta3^{WD}$ and validation of RNA-sequencing results that are a representative of two independent experiments. Statistical analyses were performed with the unpaired Students T-Test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns, non significant.

References

Acton SE, Farrugia AJ, Astarita JL, Mourao-Sa D, Jenkins RP, Nye E, Hooper S, Van Blijswijk J, Rogers NC, Snelgrove KJ, Rosewell I, Moita LF, Stamp G, Turley SJ, Sahai E, Reis E Sousa C. 2014. Dendritic cells control fibroblastic reticular network tension and lymph node expansion. *Nature* **514**: 498-502. 10.1038/nature13814.

Anido J, Saez-Borderias A, Gonzalez-Junca A, Rodon L, Folch G, Carmona MA, Prieto-Sanchez RM, Barba I, Martinez-Saez E, Prudkin L, Cuartas I, Raventos C, Martinez-Ricarte F, Poca MA, Garcia-Dorado D, Lahn MM, Yingling JM, Rodon J, Sahuquillo J, Baselga J, Seoane J. 2010. TGF-β Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell* **18**: 655-668. 10.1016/j.ccr.2010.10.023.

Bellomo C, Caja L, Moustakas A. 2016. Transforming growth factor β as regulator of cancer stemness and metastasis. *Br J Cancer* **115**: 761-769. 10.1038/bjc.2016.255.

Campbell DG, Morrice NA. 2002. Identification of protein phosphorylation sites by a combination of mass spectrometry and solid phase Edman sequencing. *J Biomol Tech* **13**: 119-130.

Chen YG, Massague J. 1999. Smad1 recognition and activation by the ALK1 group of transforming growth factor-β family receptors. *J Biol Chem* **274**: 3672-3677.

Coda DM, Gaarenstroom T, East P, Patel H, Miller DSJ, Lobley A, Matthews N, Stewart A, Hill CS. 2017. Distinct modes of SMAD2 chromatin binding and remodeling shape the transcriptional response to Nodal/Activin signaling. *eLife* **6**: e22474. 10.7554/eLife.22474.

Cuny GD, Yu PB, Laha JK, Xing X, Liu JF, Lai CS, Deng DY, Sachidanandan C, Bloch KD, Peterson RT. 2008. Structure-activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors. *Bioorg Med Chem Lett* **18**: 4388-4392. 10.1016/j.bmcl.2008.06.052.

Dacosta Byfield S, Major C, Laping NJ, Roberts AB. 2004. SB-505124 is a selective inhibitor of transforming growth factor-β type I receptors ALK4, ALK5, and ALK7. *Mol Pharmacol* **65**: 744-752. 10.1124/mol.65.3.744.

876 Daly AC, Randall RA, Hill CS. 2008. Transforming growth factor β-induced
877 Smad1/5 phosphorylation in epithelial cells is mediated by novel receptor
878 complexes and is essential for anchorage-independent growth. *Mol Cell Biol*
879 **28**: 6889-6902. 10.1128/MCB.01192-08.
880 Daly AC, Vizan P, Hill CS. 2010. Smad3 protein levels are modulated by Ras activity
881 and during the cell cycle to dictate transforming growth factor-β responses. *J*
882 *Biol Chem* **285**: 6489-6497. 10.1074/jbc.M109.043877.
883 Diepenbruck M, Christofori G. 2016. Epithelial-mesenchymal transition (EMT) and
884 metastasis: yes, no, maybe? *Curr Opin Cell Biol* **43**: 7-13.
885 10.1016/j.ceb.2016.06.002.
886 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson
887 M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner.
888 *Bioinformatics* **29**: 15-21. 10.1093/bioinformatics/bts635.
889 Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, Wicha
890 MS. 2003. In vitro propagation and transcriptional profiling of human
891 mammary stem/progenitor cells. *Genes Dev* **17**: 1253-1270.
892 10.1101/gad.1061803.
893 Gaarenstroom T, Hill CS. 2014. TGF-β signaling to chromatin: how Smads regulate
894 transcription during self-renewal and differentiation. *Semin Cell Dev Biol* **32**:
895 107-118. 10.1016/j.semcdb.2014.01.009.
896 Germain S, Howell M, Esslemont GM, Hill CS. 2000. Homeodomain and winged-
897 helix transcription factors recruit activated Smads to distinct promoter
898 elements via a common Smad interaction motif. *Genes Dev* **14**: 435-451.
899 Goumans MJ, Valdimarsdottir G, Itoh S, Lebrin F, Larsson J, Mummery C, Karlsson
900 S, Ten Dijke P. 2003. Activin receptor-like kinase (ALK)1 is an antagonistic
901 mediator of lateral TGFβ/ALK5 signaling. *Mol Cell* **12**: 817-828.
902 Goumans MJ, Valdimarsdottir G, Itoh S, Rosendahl A, Sideras P, Ten Dijke P. 2002.
903 Balancing the activation state of the endothelium via two distinct TGF-β type I
904 receptors. *EMBO J* **21**: 1743-1753. 10.1093/emboj/21.7.1743.
905 Gronroos E, Kingston IJ, Ramachandran A, Randall RA, Vizan P, Hill CS. 2012.
906 Transforming growth factor β inhibits bone morphogenetic protein-induced
907 transcription through novel phosphorylated Smad1/5-Smad3 complexes. *Mol*
908 *Cell Biol* **32**: 2904-2916. 10.1128/MCB.00231-12.

909 Grunert S, Jechlinger M, Beug H. 2003. Diverse cellular and molecular mechanisms
910 contribute to epithelial plasticity and metastasis. *Nat Rev Mol Cell Biol* **4**: 657-
911 665. 10.1038/nrm1175.
912 Hatsell SJ, Idone V, Wolken DM, Huang L, Kim HJ, Wang L, Wen X, Nannuru KC,
913 Jimenez J, Xie L, Das N, Makhoul G, Chernomorsky R, D'ambrosio D,
914 Corpina RA, Schoenherr CJ, Feeley K, Yu PB, Yancopoulos GD, Murphy AJ,
915 Economides AN. 2015. ACVR1R206H receptor mutation causes
916 fibrodysplasia ossificans progressiva by imparting responsiveness to activin A.
917 *Sci Transl Med* **7**: 303ra137. 10.1126/scitranslmed.aac4358.
918 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C,
919 Singh H, Glass CK. 2010. Simple combinations of lineage-determining
920 transcription factors prime cis-regulatory elements required for macrophage
921 and B cell identities. *Mol Cell* **38**: 576-589. 10.1016/j.molcel.2010.05.004.
922 Hino K, Ikeya M, Horigome K, Matsumoto Y, Ebise H, Nishio M, Sekiguchi K,
923 Shibata M, Nagata S, Matsuda S, Toguchida J. 2015. Neofunction of ACVR1
924 in fibrodysplasia ossificans progressiva. *Proc Natl Acad Sci U S A* **112**:
925 15438-15443. 10.1073/pnas.1510540112.
926 Hoodless PA, Haerry T, Abdollah S, Stapleton M, O'connor MB, Attisano L, Wrana
927 JL. 1996. MADR1, a MAD-related protein that functions in BMP2 signaling
928 pathways. *Cell* **85**: 489-500.
929 Horton RM, Cai ZL, Ho SN, Pease LR. 1990. Gene splicing by overlap extension:
930 tailor-made genes using the polymerase chain reaction. *Biotechniques* **8**: 528-
931 535.
932 Huang T, David L, Mendoza V, Yang Y, Villarreal M, De K, Sun L, Fang X, Lopez-
933 Casillas F, Wrana JL, Hinck AP. 2011. TGF-β signalling is mediated by two
934 autonomously functioning TβRI:TβRII pairs. *EMBO J* **30**: 1263-1276.
935 10.1038/emboj.2011.54.
936 Inman GJ, Nicolas FJ, Callahan JF, Harling JD, Gaster LM, Reith AD, Laping NJ,
937 Hill CS. 2002. SB-431542 is a potent and specific inhibitor of transforming
938 growth factor-β superfamily type I activin receptor-like kinase (ALK)
939 receptors ALK4, ALK5, and ALK7. *Mol Pharmacol* **62**: 65-74.
940 Itoh F, Watabe T, Miyazono K. 2014. Roles of TGF-β family signals in the fate
941 determination of pluripotent stem cells. *Semin Cell Dev Biol* **32**: 98-106.
942 10.1016/j.semcdb.2014.05.017.

943 Kang Y, Chen CR, Massague J. 2003. A self-enabling TGFβ response coupled to
944 stress signaling: Smad engages stress response factor ATF3 for Id1 repression
945 in epithelial cells. *Mol Cell* **11**: 915-926.
946 Kowanetz M, Valcourt U, Bergstrom R, Heldin CH, Moustakas A. 2004. Id2 and Id3
947 define the potency of cell proliferation and differentiation responses to
948 transforming growth factor β and bone morphogenetic protein. *Mol Cell Biol*
949 **24**: 4241-4254.
950 Kretzschmar M, Liu F, Hata A, Doody J, Massague J. 1997. The TGF-β family
951 mediator Smad1 is phosphorylated directly and activated functionally by the
952 BMP receptor kinase. *Genes Dev* **11**: 984-995.
953 Lasorella A, Benezra R, Iavarone A. 2014. The ID proteins: master regulators of
954 cancer stem cells and tumour aggressiveness. *Nat Rev Cancer* **14**: 77-91.
955 10.1038/nrc3638.
956 Lechleider RJ, Ryan JL, Garrett L, Eng C, Deng C, Wynshaw-Boris A, Roberts AB.
957 2001. Targeted mutagenesis of Smad1 reveals an essential role in
958 chorioallantoic fusion. *Dev Biol* **240**: 157-167. 10.1006/dbio.2001.0469.
959 Levy L, Hill CS. 2005. Smad4 dependency defines two classes of transforming
960 growth factor β (TGF-β) target genes and distinguishes TGF-β-induced
961 epithelial-mesenchymal transition from its antiproliferative and migratory
962 responses. *Mol Cell Biol* **25**: 8108-8125. 10.1128/MCB.25.18.8108-
963 8125.2005.
964 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data
965 with or without a reference genome. *BMC Bioinformatics* **12**: 323.
966 10.1186/1471-2105-12-323.
967 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
968 transform. *Bioinformatics* **25**: 1754-1760. 10.1093/bioinformatics/btp324.
969 Liu IM, Schilling SH, Knouse KA, Choy L, Derynck R, Wang XF. 2009. TGFβ-
970 stimulated Smad1/5 phosphorylation requires the ALK5 L45 loop and
971 mediates the pro-migratory TGFβ switch. *EMBO J* **28**: 88-98.
972 10.1038/emboj.2008.266.
973 Liu X, Yue J, Frey RS, Zhu Q, Mulder KM. 1998. Transforming growth factor β
974 signaling through Smad1 in human breast cancer cells. *Cancer Res* **58**: 4752-
975 4757.

976 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
977 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
978 10.1186/s13059-014-0550-8.
979 Luo K, Lodish HF. 1996. Signaling by chimeric erythropoietin-TGF-β receptors:
980 homodimerization of the cytoplasmic domain of the type I TGF-β receptor and
981 heterodimerization with the type II receptor are both required for intracellular
982 signal transduction. *EMBO J* **15**: 4485-4496.
983 Massague J. 1996. TGFβ signaling: receptors, transducers, and Mad proteins. *Cell* **85**:
984 947-950.
985 Massague J. 2008. TGFβ in Cancer. *Cell* **134**: 215-230. 10.1016/j.cell.2008.07.001.
986 Massague J. 2012. TGFβ signalling in context. *Nat Rev Mol Cell Biol* **13**: 616-630.
987 10.1038/nrm3434.
988 Meulmeester E, ten Dijke P. 2011. The dynamic roles of TGF-β in cancer. *J Pathol*
989 **223**: 205-218. 10.1002/path.2785.
990 Michel M, Raabe I, Kupinski AP, Perez-Palencia R, Bokel C. 2011. Local BMP
991 receptor activation at adherens junctions in the Drosophila germline stem cell
992 niche. *Nat Commun* **2**: 415. 10.1038/ncomms1426.
993 Miettinen PJ, Ebner R, Lopez AR, Derynck R. 1994. TGF-β induced
994 transdifferentiation of mammary epithelial cells to mesenchymal cells:
995 involvement of type I receptors. *J Cell Biol* **127**: 2021-2036.
996 Miller DSJ, Hill CS 2016. TGF-β Superfamily Signaling. *In:* BRADSHAW, R. A. &
997 STAHL, P. D. (eds.) *Encycopedia of Cell Biology*. Elsevier Inc.
998 Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB,
999 Gerald WL, Massague J. 2005. Genes that mediate breast cancer metastasis to
1000 lung. *Nature* **436**: 518-524. 10.1038/nature03799.
1001 Miyazono K, Kamiya Y, Morikawa M. 2010. Bone morphogenetic protein receptors
1002 and signal transduction. *J Biochem* **147**: 35-51. 10.1093/jb/mvp148.
1003 Morikawa M, Derynck R, Miyazono K. 2016. TGF-β and the TGF-β Family:
1004 Context-Dependent Roles in Cell and Tissue Physiology. *Cold Spring Harb*
1005 *Perspect Biol* **8**: a021873. 10.1101/cshperspect.a021873.
1006 Morikawa M, Koinuma D, Tsutsumi S, Vasilaki E, Kanki Y, Heldin CH, Aburatani
1007 H, Miyazono K. 2011. ChIP-seq reveals cell type-specific binding patterns of
1008 BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* **39**: 8712-
1009 8727. 10.1093/nar/gkr572.

Nam JS, Terabe M, Mamura M, Kang MJ, Chae H, Stuelten C, Kohn E, Tang B, Sabzevari H, Anver MR, Lawrence S, Danielpour D, Lonning S, Berzofsky JA, Wakefield LM. 2008. An anti-transforming growth factor β antibody suppresses metastasis via cooperative effects on multiple cell compartments. *Cancer Res* **68**: 3835-3843. 10.1158/0008-5472.CAN-08-0215.

Nicolas FJ, De Bosscher K, Schmierer B, Hill CS. 2004. Analysis of Smad nucleocytoplasmic shuttling in living cells. *J Cell Sci* **117**: 4113-4125. 10.1242/jcs.01289.

Nicolas FJ, Hill CS. 2003. Attenuation of the TGF-β-Smad signaling pathway in pancreatic tumor cells confers resistance to TGF-β-induced growth arrest. *Oncogene* **22**: 3698-3711. 10.1038/sj.onc.1206420.

Padua D, Massague J. 2009. Roles of TGFβ in metastasis. *Cell Res* **19**: 89-102. 10.1038/cr.2008.316.

Padua D, Zhang XH, Wang Q, Nadal C, Gerald WL, Gomis RR, Massague J. 2008. TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**: 66-77. 10.1016/j.cell.2008.01.046.

Pardali E, Goumans MJ, Ten Dijke P. 2010. Signaling by members of the TGF-β family in vascular morphogenesis and disease. *Trends Cell Biol* **20**: 556-567. 10.1016/j.tcb.2010.06.006.

Peinado H, Olmeda D, Cano A. 2007. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* **7**: 415-428. 10.1038/nrc2131.

Piek E, Moustakas A, Kurisaki A, Heldin CH, Ten Dijke P. 1999. TGF-β type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. *J Cell Sci* **112** 4557-4568.

R Development Core Team 2009. *R: A language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281-2308. 10.1038/nprot.2013.143.

Reichert S, Randall RA, Hill CS. 2013. A BMP regulatory network controls ectodermal cell fate decisions at the neural plate border. *Development* **140**: 4435-4444. 10.1242/dev.098707.

1044 Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G,
1045 Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24-26.
1046 10.1038/nbt.1754.
1047 Ross S, Cheung E, Petrakis TG, Howell M, Kraus WL, Hill CS. 2006. Smads
1048 orchestrate specific histone modifications and chromatin remodeling to
1049 activate transcription. *EMBO J* **25**: 4490-4502. 10.1038/sj.emboj.7601332.
1050 Sako K, Pradhan SJ, Barone V, Ingles-Prieto A, Muller P, Ruprecht V, Capek D,
1051 Galande S, Janovjak H, Heisenberg CP. 2016. Optogenetic Control of Nodal
1052 Signaling Reveals a Temporal Pattern of Nodal Signaling Regulating Cell Fate
1053 Specification during Gastrulation. *Cell Rep* **16**: 866-877.
1054 10.1016/j.celrep.2016.06.036.
1055 Shi Y, Massague J. 2003. Mechanisms of TGF-β signaling from cell membrane to the
1056 nucleus. *Cell* **113**: 685-700.
1057 Stankic M, Pavlovic S, Chin Y, Brogi E, Padua D, Norton L, Massague J, Benezra R.
1058 2013. TGF-β-Id1 signaling opposes Twist1 and promotes metastatic
1059 colonization via a mesenchymal-to-epithelial transition. *Cell Rep* **5**: 1228-
1060 1242. 10.1016/j.celrep.2013.11.014.
1061 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA,
1062 Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set
1063 enrichment analysis: a knowledge-based approach for interpreting genome-
1064 wide expression profiles. *Proc Natl Acad Sci U S A* **102**: 15545-15550.
1065 10.1073/pnas.0506580102.
1066 Takahashi F, Yamagata D, Ishikawa M, Fukamatsu Y, Ogura Y, Kasahara M,
1067 Kiyosue T, Kikuyama M, Wada M, Kataoka H. 2007. AUREOCHROME, a
1068 photoreceptor required for photomorphogenesis in stramenopiles. *Proc Natl*
1069 *Acad Sci U S A* **104**: 19625-19630. 10.1073/pnas.0707692104.
1070 Vizan P, Miller DSJ, Gori I, Das D, Schmierer B, Hill CS. 2013. Controlling long-
1071 term signaling: receptor dynamics determine attenuation and refractory
1072 behavior of the TGF-β pathway. *Sci Signal* **6**: ra106.
1073 10.1126/scisignal.2004416.
1074 Wakefield LM, Hill CS. 2013. Beyond TGFβ: roles of other TGFβ superfamily
1075 members in cancer. *Nat Rev Cancer* **13**: 328-341. 10.1038/nrc3500.
1076 Wang T, Donahoe PK, Zervos AS. 1994. Specific interaction of type I receptors of
1077 the TGF-β family with the immunophilin FKBP-12. *Science* **265**: 674-676.

Wieser R, Wrana JL, Massague J. 1995. GS domain mutations that constitutively activate T β R-I, the downstream signaling component in the TGF-β receptor complex. *EMBO J* **14**: 2199-2208.

Wrana JL, Attisano L, Wieser R, Ventura F, Massague J. 1994. Mechanism of activation of the TGF-β receptor. *Nature* **370**: 341-347. 10.1038/370341a0.

Wrighton KH, Lin X, Yu PB, Feng XH. 2009. Transforming Growth Factor β Can Stimulate Smad1 Phosphorylation Independently of Bone Morphogenic Protein Receptors. *J Biol Chem* **284**: 9755-9763. 10.1074/jbc.M809223200.

Wu MY, Hill CS. 2009. TGF-β superfamily signaling in embryonic development and homeostasis. *Dev Cell* **16**: 329-343. 10.1016/j.devcel.2009.02.012.

Ye X, Weinberg RA. 2015. Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. *Trends Cell Biol* **25**: 675-686. 10.1016/j.tcb.2015.07.012.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137. 10.1186/gb-2008-9-9-r137.

Acknowledgements

We thank Lalage Wakefield for providing the TGF-β neutralizing antibody and the isotype-matched control, Paul Yu for LDN-193189, Christian Bökel for the pCS2+zALK3-IPF expression plasmid and Bob Lechleider for the FLAG-SMAD1 expression plasmid. We thank Nik Mathews, Greg Elgar and the Advanced Sequencing Facility for the next generation sequencing. We are grateful to the Francis Crick Institute Light Microscopy and Flow Cytometry facilities and to the Genomics Equipment Park. We thank Alex Bullock for very fruitful discussions and all the members of the Hill lab for useful comments on the manuscript. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001095), the UK Medical Research Council (FC001095), and the Wellcome Trust (FC001095). The development and characterization of TGF-$\beta3^{WD}$ in the Hinck laboratory was enabled by support provided by the NIH (GM58670 and CA172886).

Conflicts of Interest

The authors declare that they have no conflict of interest.

Figure legends

Figure 1. Characterization of SMAD1/5 phosphorylation by TGF-β.

(A) MDA-MB-231 and NMuMG cells were treated with TGF-β or BMP4 for the times indicated.

(B) MDA-MB-231 cells were treated with TGF-β for the times shown either alone or after 5 minutes pre-treatment with cyclohexamide (CHX) or actinomycin D (Act D).

(C) MDA-MB-231 cells were treated with TGF-β for 1 or 8 hr, and after 8 hr, cells were re-stimulated with TGF-β or BMP4 for 1 hr as shown in the scheme. For comparison, cells were stimulated for 1 hr with BMP4.

(D) MDA-MB-231 cells were induced or not with TGF-β or BMP4 in the presence of either 0.25 μM or 10 μM SB-431542 (SB) or 1 μM LDN-193189 (LDN) or a combination of 0.25 μM or 10 μM SB-431542 and 1μM LDN-193189. In all panels Western blots are shown probed with the antibodies indicated. B, BMP4, Un, unstimulated. In B, SERPINE1, whose expression is induced by TGF-β, provides a control for the efficacy of the CHX and Act D.

Figure 2. ACVR1 is activated by TGFBR1 *in vitro* and *in vivo.*

(A) The kinase domains of TGFBR1 and ACVR1 were analyzed alone or together in an *in vitro* kinase reaction. SB-505124 and LDN-193189 were included as shown to inhibit the activity of TGFBR1 and ACVR1 respectively. The autoradiograph is shown in the top panel, with the Coomassie-stained gel below as a loading control.

(B) Schematic to show the domain organization of the Opto receptors. In Opto-TGFBR1*and Opto-ACVR1, the kinase domains of TGFBR1 and ACVR1 are fused to the light sensitive LOV domain. At the N-terminus there is a myristylation domain (indicated by the red zig zag). At the C-terminus there is an HA tag. The kinase domain of TGFBR1 contains the activating mutation T204D. These Opto receptors dimerize in the presence of blue light.

(C) NIH-3T3 cells were untransfected or transfected with FLAG-SMAD1 together with either Opto-TGFBR1*, Opto-ACVR1 or both receptors together. Post transfection, cells were either kept in the dark or exposed to blue light for 1 hr. Whole cell extracts were Western blotted using antibodies against pSMAD1/5 (which detects endogenous and FLAG pSMAD1/5), SMAD1 (which detects endogenous and FLAG SMAD1), HA (to detect the Opto receptors) and Tubulin as a loading control.

(D) NIH-3T3 cells were untransfected or transfected with FLAG-SMAD1 together with either Opto-TGFBR1*, Opto-ACVR1 or both receptors together. Post transfection, cells were either kept in the dark or exposed to blue light for 1 hr. The inductions were performed in the absence or presence of 0.5 µM LDN-193189 or 50 µM SB-505124 as indicated. Whole cell extracts were blotted as in (C).

(E) The experimental set up was as in (D) except that GFP-SMAD3 was used instead of FLAG-SMAD1 to assess the activity of Opto-TGFBR1*.

(F) As in (C), except that an ACVR1 mutant in which all the threonines and serines of the GS domain were mutated to valine or alanine respectively, was also assayed.

(G) As in (F), except that GFP-SMAD3 was used instead of FLAG-SMAD1.

Note that in all cases that the 1 hr induction with blue light led to reduced levels of the transfected receptors and substrates.

Figure 3. TGFBR1 and ACVR1 are present in distinct receptor complexes upon TGF-β stimulation.

(A) Alternative models of receptor clustering mediated by TGF-β3 derivatives capable of interacting with two pairs (TGF-$\beta3^{WW}$) or one pair (TGF-$\beta3^{WD}$) of type II:type I receptors. If an obligate heterotetramer of two type I:type II pairs is required for SMAD1/5 phosphorylation (Model I), then only TGF-$\beta3^{WW}$ would lead to SMAD1/5 phosphorylation. If TGF-β induces higher order receptor clustering at the cell surface (Model II), then both TGF-$\beta3^{WW}$ and TGF-$\beta3^{WD}$ would lead to SMAD1/5 phosphorylation.

(B) MDA-MB-231 cells were treated with different concentrations of TGF-$\beta3^{WW}$ or TGF-$\beta3^{WD}$ for 1 hr as indicated. As a control, cells were either untreated (Un) or treated with TGF-β1 (T) or BMP4 (B) for 1 hr. Whole cell lysates were Western blotted using the antibodies shown.

Figure 4. ACVR1 is activated by TGF-β in a TGFBR1-dependent manner.

(A and B) MDCKII ACVR1-IPF cells were imaged at 15 min intervals for 60 min before the addition (arrow) of either media alone or media containing TGF-β ± 10 μM SB-431542 for a further 150 min. The panels in (A) are stills of the maximum intensity projections at the times shown. The quantifications are shown in (B). The fluorescence at the 60-min time point was taken as the reference that was subtracted from all other time points. Data presented are the mean ± SD of three independent fields. Statistical significance is shown for the indicated pairs of conditions at the 210 min timepoint.

(C) Fluorescence in MDCKII ACVR1-IPF cells assayed by flow cytometry 24 hr after treatment. Each panel shows an overlay of the indicated treatment conditions. The black line indicates the median of the untreated (Un) sample. Quantifications are shown on the right. For each group, the percentage of cells greater than the median fluorescence intensity of the untreated sample was quantified. Data are the mean ± SEM of three independent experiments. SB, SB-431542 at 10 μM; Ab, antibody; Nog, noggin; C, control antibody; B, blocking antibody.

Figure 4 - Source data 1. Source data for ACVR1-IPF fluorescence (panel B)

Figure 4 - Source data 2. Source data for ACVR1-IPF fluorescence by flow cytometry (panel C)

Figure 5. pSMAD1/5 is recruited to chromatin in response to TGF-β and is most highly enriched at GGCGCC motifs.

(A) MDA-MB-231 cells were either untreated (-) or treated with TGF-β (+) for 1 hr. Whole cell extracts were immunoprecipitated (IP) with the antibodies (Ab) indicated or beads alone (Be). The IPs were Western blotted using the antibodies shown. Inputs are shown on the left.

(B) IGV browser displays over the *ID1*, *ID3* and *JUNB* loci after ChIP-Seq of MDA-MB-231 untreated (Un) and TGF-β-treated samples. IPs were performed with antibodies against pSMAD1/5 (pS1/5), SMAD3 (S3) or with beads alone as a negative control. Inputs are also shown. Red lines indicate regions validated in (C). U; upstream peak; D1, downstream peak 1.

(C) Genomic regions were validated by ChIP-qPCR after treatment of MDA-MB-231 cells with TGF-β (T) or BMP4 (B) for the times shown. IPs were as in (B). A representative experiment of two performed in triplicate is shown with means ± SD.

(D) The most enriched motif obtained from a MEME-ChIP analysis of the top 100 pSMAD1/5 peaks.

(E) Proportion of variants of the GGCGCC motif identified in the top 100 pSMAD1/5 peaks.

Figure 5 - Source data 1. ChIP-seq datasets

Figure 5 - Source data 2. ChIP-PCR data for graphs in panel C

Figure 6. SMAD1/5 is required for TGF-β-induced EMT.

(A) The parental NMuMG clone and the ΔSMAD1/5 clone 1 were treated with TGF-β or BMP4 for the times shown. Whole cell extracts were immunoblotted with the antibodies indicated.

(B) Parental NMuMG clone and the ΔSMAD1/5 clone 1 cells were left untreated or treated with TGF-β for 48 hr and imaged after indirect immunofluorescence (IF) using antibodies against TJP1 and CDH1. A merge of the two with DAPI in blue is also shown.

(C) NMuMG cells were left untreated (Un) or treated with TGF-β alone or in combination with 1 μM LDN-193189 (LDN) ± 0.125 μM SB-431542 (SB) for 48 hr. Panels show cells imaged under either phase contrast (left panels) or by indirect immunofluorescence (IF) using antibodies against TJP1 and CDH1. A merge of the two with DAPI in blue is also shown.

(D) NMuMG cells were left untreated or treated with TGF-β alone or in combination with either 1 μM LDN-193189 ± 0.125 μM SB-431542 for 48 hr. Whole cell lysates were immunoblotted with the indicated antibodies.

(E) EpRas cells were left untreated (Un) or treated with TGF-β alone or in combination with 1 μM LDN-193189 (LDN) ± 0.125 μM SB-431542 (SB) for 9 days, then imaged after indirect immunofluorescence (IF) using antibodies against TJP1 and CDH1 or a merge of the two with DAPI in blue.

(F) EpRas cells were left untreated or treated with TGF-β for 1 hr alone or with combinations of 1 μM LDN-193189, 0.125 μM SB-431542 or 10 μM SB-431542 as indicated. Whole cell lysates were immunoblotted with the indicated antibodies

In (B), (C) and (E) the indirect IF images are maximum intensity projections of a z-stack in each channel.

Figure 6 - Source data 1. RNA-seq datasets

Figure 7. TGF-β-induced ID1 via pSMAD1/5 is required for EMT.

(A) NMuMG cells were transfected with siRNAs against *ID1* or NT control, then left untreated or treated with TGF-β for 24 hr. Cells were imaged after indirect IF with antibodies against TJP1 and CDH1 or a merge of the two with DAPI in blue. All indirect IF images are maximum intensity projections of a z-stack in each channel.

(B) Western blots to show knockdown efficiency of the *ID1* siRNA. NMuMG cells were treated with TGF-β (T) or BMP4 (B) for 1 hr.

(C) The model shows combinatorial signaling by TGF-β utilizing complexes containing two different type I receptors. Type II receptors are shown in blue, TGFBR1 in orange and ACVR1 in green as in Figure 3A. P denotes phosphorylation. S1/5, SMAD1/5; S2/3, SMAD2/3; S4, SMAD4. The question mark indicates that we do not yet know the function of the mixed R-SMAD complexes in the physiological responses. For discussion, see text.

Figure 1 – figure supplement 1. SMAD1 phosphorylation kinetics in response to TGF-β.

(A, C–E) Western blots are shown probed with the indicated antibodies.

(A) BT-549 cells were grown as a monolayer on plastic (2D) or as spheres in low attachment plates (phase contrast images on the right) and treated with TGF-β or BMP4 for the times indicated. BT-549 cells show sustained SMAD1/5 phosphorylation in response to TGF-β.

(B) qPCR of the indicated genes in MDA-MB-231 cells treated with actinomycin D for the times shown. Data are presented as fold change relative to 0 hr. A representative experiment performed in triplicate is shown with means ± SD. Transcripts of both *TGFBRI* and *TGFBR2* are relatively stable.

(C) NMuMG cells were treated with TGF-β for the times shown either alone or after 5 minutes pre-treatment with cyclohexamide (CHX) or actinomycin D (Act D). Act D prolongs, while CHX terminates both SMAD1/5 and SMAD2 phosphorylation in response to TGF-β. Un, untreated.

(D) NMuMG cells were treated with TGF-β for 1 or 8 hr and after 8 hr, cells were restimulated with 10 or 20 ng/ml BMP4 as shown in the scheme. Cells were also treated for 1 hr with 10 or 20 ng/ml BMP4 as a control. Cells pre-treated with TGF-β can still be stimulated with BMP4.

(E) NMuMG cells were left untreated or treated with TGF-β ± SB-431542 (SB; 0.125 μM or 10 μM) ± 1 μM LDN-193189 (LDN) or BMP4 ± 1 μM LDN-193189 for 1 hr. The kinase activity of both classes of type I receptors is required for SMAD1/5 phosphorylation by TGF-β.

Figure 1 – figure supplement 1 - Source data 1. Source data for qPCRs (panel B)

Figure 1 - figure supplement 2. SMAD1 is efficiently phosphorylated by ACVR1 and BMPR1A, but poorly phosphorylated by TGFBR1.

(A) *In vitro* kinase assays using the kinase domains of ACVR1, BMPR1A, and TGFBR1 at 200, 100, 50, 25 ng with recombinant SMAD1 (S1) or SMAD2 (S2) as substrates. Top panels, autoradiograph; bottom panels, Coomassie-stained gel.

(B) Incorporation of ^{32}P into SMAD1 and SMAD2 catalysed by ACVR1 and TGFBR1 using different specific activities of [γ–^{32}P]-ATP. A constant amount of [γ–^{32}P]-ATP was added into the kinase reaction with either 200 or 50 µM cold ATP. Top panels, autoradiograph; bottom panels, Coomassie-stained gel. Numbers underneath indicate the fold changes relative to the ^{32}P incorporation in SMAD1 (upper) or SMAD2 (lower) catalyzed by TGFBR1 using 200 µM cold ATP. The phosphorylation of SMAD1 and 2 by ACVR1 and TGFBR1 was dependent on the specific activity of the [γ–^{32}P]-ATP, whilst the apparent phosphorylation of SMAD1 by TGFBR1 is not, suggesting that it is non-specific.

(C) Mapping ACVR1 phosphorylation sites on SMAD1. Full length SMAD1 phosphorylated by ACVR1 was digested with trypsin. Peptides were resolved by reverse phase HPLC (left panel). The C-terminal peptide of SMAD1 existed in three different phosphorylation states (peptides a, b, and c); the three subsequent peaks are tryptic miscleavage products. The phosphorylation sites in the peptides were mapped using solid phase Edman sequencing (panels labelled a, b and c). The deduced phosphorylation sites in the SSVS motif in the individual peptides are shown in red.

Figure 2 - figure supplement 1. Kinase dead Opto-TGFBR1 cannot activate Opto-ACVR1.

(A) NIH-3T3 cells were untransfected or transfected with FLAG-SMAD1 together with Opto-TGFBR1*, Opto-ACVR1 or both receptors together or a kinase-dead version of Opto-TGFBR1 (Opto-TGFBR1-KR) alone or together with Opto-ACVR1. Post transfection, cells were either kept in the dark or exposed to blue light for 1 hr. Whole cell extracts were Western blotted using antibodies against pSMAD1/5 (which detects endogenous and FLAG pSMAD1/5), SMAD1 (which detects endogenous and FLAG SMAD1), HA (to detect the Opto receptors) and Tubulin as a loading control.

(B) The experimental set up was as in (A) except that GFP-SMAD3 was used instead of FLAG-SMAD1 to assess the activity of Opto-TGFBR1* and Opto-TGFBR1-KR. The band marked with an asterisk is a background band.

Figure 3 – figure supplement 1. NMuMG cells respond to both TGF-β3^{WW} and TGF-β3^{WD}.

NMuMG cells were untreated (Un) or treated with TGF-β1 (T), BMP4 (B) or the indicated concentrations of TGF-β3^{WW} and TGF-β3^{WD} for 1 hr. Whole cell lysates were immunoblotted with the antibodies shown. Both TGF-β3^{WW} and TGF-β3^{WD} induce phosphorylation of pSMAD1/5, although the latter is less potent.

Figure 4 – figure supplement 1. Characterization of cells stably transfected with the ACVR1-IPF.

(A) NMuMG cells knocked out for ACVR1 and BMPR1A were transfected with either empty vector (pcDNA3.1 Hygro +), ACVR1-IPF or ACVR1-FLAG. Whole cell extracts were Western blotted for pSMAD1/5, SMAD1 and GFP and FLAG. Actin is a loading control.

(B) MDCKII ACVR1-IPF or NIH-3T3 ACVR1-IPF cells imaged by indirect IF with an antibody against GFP (red) with nuclei stained with DAPI (blue). The controls were a secondary antibody only sample (left) and the matched empty vector-transfected cells stained with the GFP antibody (middle). Scale bar equates to 20 μM. ACVR1-IPF localizes to the membrane in both cell types, with basolateral localisation in the MDCKII cells.

(C and D) The MDCKII or NIH-3T3 cell lines shown in (B) were treated with TGF-β or BMP4 for the times indicated. Whole cell lysates were immunoblotted with the indicated antibodies. Stable transfection of ACVR1-IPF does not affect the phosphorylation kinetics of SMAD1/5 in response to TGF-β or BMP.

(E) Median fluorescence intensity (y-axis) as measured by flow cytometry of MDCKII ACVR1-IPF cells treated with the indicated concentrations of FK506 for 4 hr. FK506 activates ACVR1-IPF fluorescence in a dose-dependent manner.

(F) Quantification of NIH-3T3 ACVR1-IPF cells imaged at 30-minute intervals for a total of 210 min after the addition of either media alone or media containing TGF-β ± 10 μM SB-431542. Data presented are the mean ± SD of three independent fields. Statistical significance is shown for the indicated pairs of conditions at the 210 min timepoint.

(G) Fluorescence in NIH-3T3 ACVR1-IPF cells assayed by flow cytometry 24 hr after being treated with media alone or with TGF-β ± 10 μM SB-431542 (SB). The percentage of cells with fluorescence greater than the median fluorescence intensity of the untreated sample (-) was quantified. Data are presented as the mean ± SEM of three independent experiments.

Figure 4 - figure supplement 1 - Source data 1. Source data for ACVR1-IPF fluorescence by flow cytometry (panel E)

Figure 4 - figure supplement 1 - Source data 2. Source data for ACVR1-IPF fluorescence (panel F)

Figure 4 - figure supplement 1 - Source data 3. Source data for ACVR1-IPF fluorescence by flow cytometry (panel G)

Figure 5 – figure supplement 1. Chromatin binding of pSMAD1/5 and SMAD3.

(A) IGV browser displays around the *ID4*, *ATOH8*, *PMEPA1*, *BHLHE40* and *SERPINE1* loci after ChIP-Seq for SMAD3 (S3) and pSMAD1/5 (pS1/5) in MDA-MB-231 cells. Untreated and TGF-β-treated samples and the inputs are shown. IPs were performed with beads as a negative control. Red lines indicate genomic regions validated by ChIP-qPCR after treatment of MDA-MB-231 cells with TGF-β (T) or BMP4 (B) for the times indicated (right panels). U, upstream peak; U1, upstream peak 1. In response to TGF-β, pSMAD1/5 bound transiently at the *ID4* and *ATOH8* loci while SMAD3 bound stably to *PMEPA1* over the same time course.

(B) ChIP-qPCR of the indicated loci after treatment of BT-549 cells with TGF-β (T) or BMP4 (B) for 1 hr. IPs were as in (A). In response to TGF-β, pSMAD1/5 bound strongly around the *ID* loci while SMAD3 bound strongly to the *JUNB* upstream locus. In A and B a representative experiment of two performed in triplicate is shown with means ± SD.

(C) The most enriched motif obtained from a MEME-ChIP analysis of the top 50 pSMAD1/5 peaks. The canonical SMAD1/5:SMAD4 binding element is strongly enriched in these peaks.

Figure 5 – figure supplement 1 - Source data 1. ChIP-PCR data for graphs in panel A

Figure 5 – figure supplement 1 - Source data 2. ChIP-PCR data for graphs in panel B

Figure 5 – figure supplement 2. *ID1* and *ID3* are TGF-β-induced target genes that require the pSMAD1/5 signaling arm.

(A) Western blots showing knockdown efficiency in MDA-MB-231s of the siRNAs shown. S3, SMAD3; S4, SMAD4; S1/5, SMAD1/5; NT, non targeting. Cells were untreated (U) or treated with TGF-β (T) or BMP4 (B) for 1 hr. Lysates were immunoblotted using the antibodies shown.

(B) MDA-MB-231 cells were transfected with siRNAs against the indicated *SMADs* or a non-targeting control (NT) and then treated with TGF-β (T) or BMP4 (B) for 1 hr. Un, untreated.

(C) MDA-MB-231 cells were left untreated or treated with TGF-β ± SB-431542 (SB; 0.25 µM or 10 µM) ± 1 µM LDN-193189 (LDN) or BMP4 ± 1 µM LDN-193189 for 1 hr. In B and C, gene expression was measured by qPCR. Data are presented as fold change relative to the untreated NT sample in B and to the (-) sample in (C) and are the means ± SEM of three independent experiments. Statistical significance is shown for selected comparisons.

(D) NMuMG cells were treated with TGF-β or BMP4 ± the inhibitors indicated. Gene expression was measured by qPCR. The combination of 0.125 µM SB-431542 (SB) and 1 µM LDN-193189 (LDN) inhibited TGF-β-induced *Id1* and *Id3* expression without affecting *JunB* expression. The data are means ± SEM of at least two independent experiments. Statistical significance is shown for selected comparisons.

(E) MDA-MB-231 cells were treated with TGF-β1 (T), BMP4 (B) or different concentrations of TGF-$\beta3^{WW}$ and TGF-$\beta3^{WD}$ for 1 hr as in Figure 3B. Gene expression was measured by qPCR. Both TGF-$\beta3^{WW}$ and TGF-$\beta3^{WD}$ led to the induction of *ID1*, *ID3* and *JUNB*, although the induction by TGF-$\beta3^{WD}$ was weaker. A representative experiment of two, performed in triplicate is shown with means ± SD.

Figure 5 - figure supplement 2 - Source data 1. qPCR data for graphs in panel B

Figure 5 - figure supplement 2 - Source data 2. qPCR data for graphs in panel C

Figure 5 - figure supplement 2 - Source data 3. qPCR data for graphs in panel D

Figure 5 - figure supplement 2 - Source data 4. qPCR data for graphs in panel E

Figure 6 – figure supplement 1. Characterization of the NMuMG ΔSMAD1/5 clones.

(A) Sequences of SMAD1 and SMAD5 in the regions around the guides in NMuMG ΔSMAD1/5 clone 1. From our sequencing we conclude that there are two alleles of SMAD1 and 3 alleles of SMAD5 in NMuMG cells. The protein sequence for the wild type (WT) is shown in red above the DNA sequence. Frame shifts are evident in both SMAD1 mutant alleles, and all three SMAD5 alleles.

(B) The parental NMuMG clone and the ΔSMAD1/5 clone 2 were treated with TGF-β or BMP4 for the times shown. Whole cell extracts were immunoblotted with the antibodies indicated.

(C) Sequences of SMAD1 and SMAD5 in the regions around the guides in NMuMG ΔSMAD1/5 clone 2. The protein sequence for the wild type (WT) is shown in red above the DNA sequence. This clone exhibits no pSMAD1/5 in reponse to either TGF-β or BMP4 (see panel B), despite having a single allele of SMAD1 and SMAD5 with an in-frame deletion of a single amino acid. This is readily explained by the nature of those mutations, which likely lead to unfolded proteins. Mutant SMAD1 allele 1 is deleted for the conserved amino acid Y125 in the MH1 domain, which is adjacent to H126, that is responsible for chelating a Zn ion in the zinc finger (BabuRajendran et al., 2010). Mutant SMAD5 allele 1 is deleted for the conserved amino acid V283 which is in β-sheet 2 of the MH2 domain, which is critical for folding of this domain (Wu et al., 2001, Qin et al., 2001). In SMAD1 allele 1 the insert length is 124 bp and in SMAD5 allele 3, it is 59 bp.

(D) Parental NMuMG clone and the ΔSMAD1/5 clone 2 cells were treated with or without TGF-β for 48 hr, fixed and imaged following indirect immunofluorescence (IF) using antibodies against TJP1 and CDH1. A merge of the two with DAPI in blue is also shown. The indirect IF images are maximum intensity projections of a z-stack in each channel.

Figure 6 – figure supplement 2. Validation of SMAD1/5-dependent TGF-β induced genes.

NMuMG parental clone and NMuMG ΔSMAD1/5 clone 1 cells were untreated or treated with TGF-β for the times shown. Total RNA was extracted and qPCR was used to assay the levels of mRNA for the genes shown. The data shown are from a representative experiment (means ± SD).

Figure 6 - figure supplement 2 - Source data 1. qPCR data for all graphs shown.

Figure 6 – figure supplement 3. The pSMAD1/5 signaling arm is required for TGF-β-mediated EMT.

(A) NMuMG cells were transfected with non-targeting (NT) or siRNAs against the SMADs as indicated. Cells were then left untreated or treated with TGF-β for 24 hr. Cells were imaged after indirect IF with antibodies against TJP1 and CDH1. A merge of the two with DAPI in blue is also shown. All indirect IF images are maximum intensity projections of a z-stack in each channel. SMAD1/5, SMAD3 and SMAD4 are all required for TGF-β-induced EMT.

(B) Western blots to show knockdown efficiency of the siRNAs. NMuMG cells were untreated (Un) or treated with TGF-β (T) or BMP4 (B) for 1 hr. Lysates were immunoblotted using the antibodies shown.

(C) NMuMG cells were treated with ligands or inhibitors as indicated. Cell lysates were immunoblotted using the antibodies shown. A combination of 0.125 µM SB-431542 and 1 µM LDN-193189 or 0.125 µM SB-431542 and 1 µM DMH1 was sufficient to abolish the TGF-β-induced phosphorylation of SMAD1/5.

(D) NMuMG cells were either untreated (Un) or treated with TGF-β ± 1 µM DMH1 ± 0.125 µM SB-431542 (SB) for 48 hr. Cells were fixed and stained for TJP1 and CDH1. In the merge, DAPI (blue) marks the nuclei. DMH1 is sufficient to inhibit TGF-β-induced EMT.

Figure 7 – figure supplement 1. TGF-β-induced ID1 is required for EMT.

(A) NMuMG, BT-549 and MDA-MB-231s were treated with ligands as shown and lysates were immunoblotted using the antibodies indicated. The induction of ID1 correlates with pSMAD1/5 phosphorylation.

(B) NMuMG cells were transfected with non-targeting (NT) or individual siRNAs against the *Id1* as indicated. Cells were then left untreated or treated with TGF-β for 48 hr. Cells were imaged after indirect IF with antibodies against TJP1 and CDH1. A merge of the two with DAPI in blue is also shown. All indirect IF images are maximum intensity projections of a z-stack in each channel. Knockdown of ID1 by any of the siRNAs inhibits TGF-β-induced EMT.

(C) NMuMG cells were transfected with siRNAs as in (B). They were uninduced or induced with BMP4 for 1 hr to induce expression of ID1 in control-transfected cells. Whole cell extracts were immunoblotted for the proteins indicated.

Supplementary Movie legends

Video 1. Fluorescence in MDCKII ACVR1-IPF cells treated with media alone. MDCKII ACVR1-IPF cells were imaged for 1 hr prior to the addition of media alone followed by imaging for a further 2.5 hr. Very little increase in fluorescence was observed over the time course.

Video 2. Fluorescence in MDCKII ACVR1-IPF cells treated with TGF-β. MDCKII ACVR1-IPF cells were imaged for 1 hr prior to the addition of 2 ng/ml TGF-β followed by imaging for a further 2.5 hr. Significant increase in fluorescence was observed over the time course with intracellular puncta of fluorescence becoming more evident over time.

Video 3. Fluorescence in MDCKII ACVR1-IPF cells treated with TGF-β and SB-431542. MDCKII ACVR1-IPF cells were imaged for 1 hr prior to the addition of 2 ng/ml TGF-β + 10 μM SB-431542 to the cells followed by imaging for a further 2.5 hr. Very little increase in fluorescence was observed over the time course.

Video 4. Fluorescence in NIH-3T3 ACVR1-IPF cells treated with media alone. NIH-3T3 ACVR1-IPF cells were imaged for 3.5 hr after the addition of media alone. A modest and gradual increase in fluorescence was observed over the time course.

Video 5. Fluorescence in NIH-3T3 ACVR1-IPF cells treated with TGF-β. NIH-3T3 ACVR1-IPF cells were imaged for 3.5 hr after the addition of 2 ng/ml TGF-β. A significant increase in fluorescence was observed over the time course with fluorescence becoming more evident on membrane projections and intracellular vesicles over time.

Video 6. Fluorescence in NIH-3T3 ACVR1-IPF cells treated with TGF-β and SB-431542. NIH-3T3 ACVR1-IPF cells were imaged for 3.5 hr after the addition of 2 ng/ml TGF-β + 10 μM SB-431542. A modest and gradual increase in fluorescence was observed over the time course.

Supplementary Files

Supplementary File 1. Sequence of Opto-TGFBR1*

Supplementary File 2. Sequence of Opto-ACVR1

Supplementary File 3. List of oligonucleotides and siRNAs

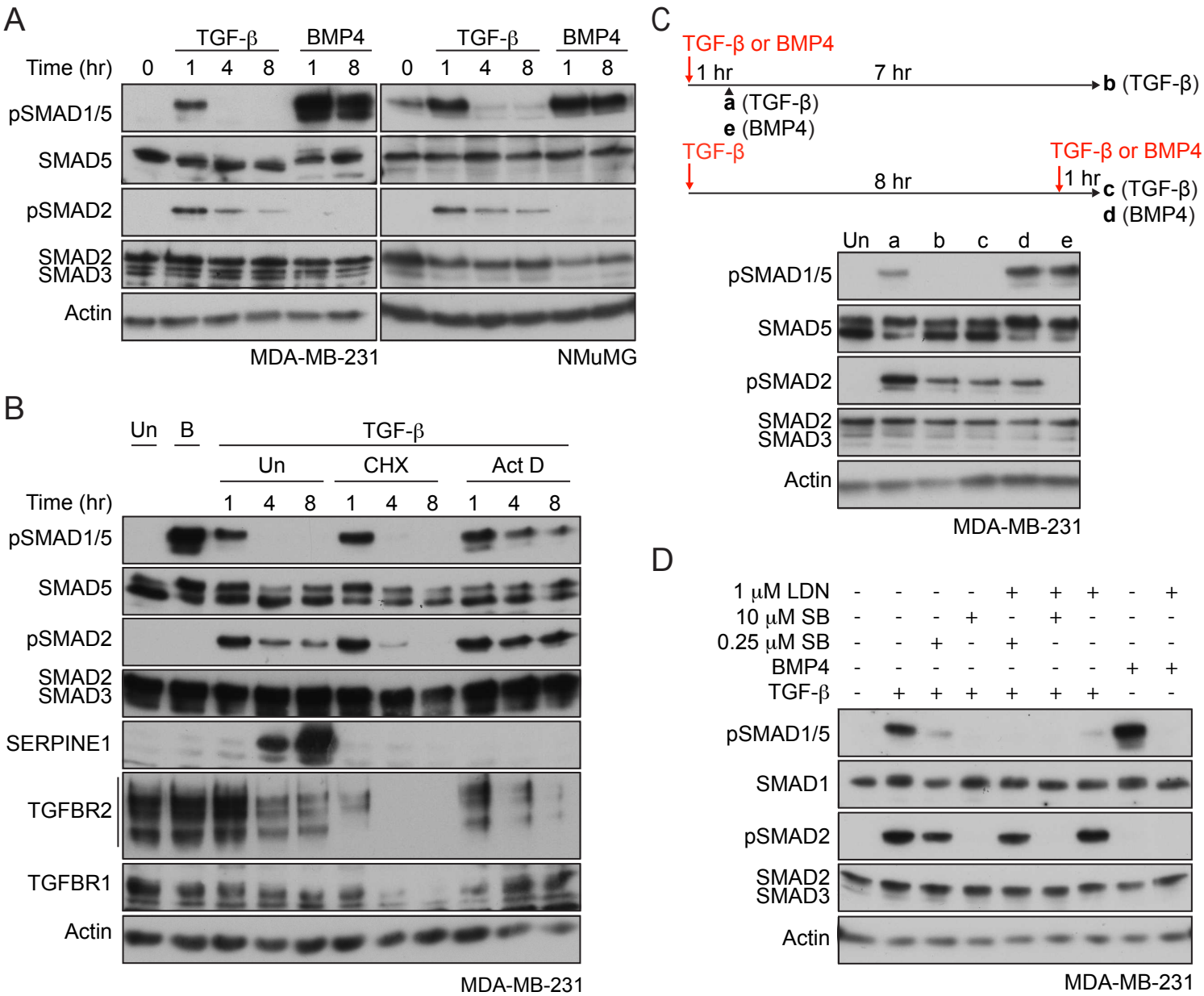

A
TGF-β
BMP4
TGF-β
BMP4
Time (hr)
0 1 4 8 1 8 0 1 4 8 1 8
pSMAD1/5
SMAD5
pSMAD2
SMAD2
SMAD3
Actin
MDA-MB-231
NMuMG
B
Un
B
TGF-β
Un
CHX
Act D
Time (hr)
1 4 8 1 4 8 1 4 8
pSMAD1/5
SMAD5
pSMAD2
SMAD2
SMAD3
SERPINE1
TGFBR2
TGFBR1
Actin
MDA-MB-231
C
TGF-β or BMP4
1 hr
7 hr
b (TGF-β)
a (TGF-β)
e (BMP4)
TGF-β
TGF-β or BMP4
8 hr
1 hr
c (TGF-β)
d (BMP4)
Un a b c d e
pSMAD1/5
SMAD5
pSMAD2
SMAD2
SMAD3
Actin
MDA-MB-231
D
1 μM LDN - - - - + + + - +
10 μM SB - - - + - + - - -
0.25 μM SB - - + - + - - - -
BMP4 - - - - - - - + +
TGF-β - + + + + + + - -
pSMAD1/5
SMAD1
pSMAD2
SMAD2
SMAD3
Actin
MDA-MB-231

Figure 1

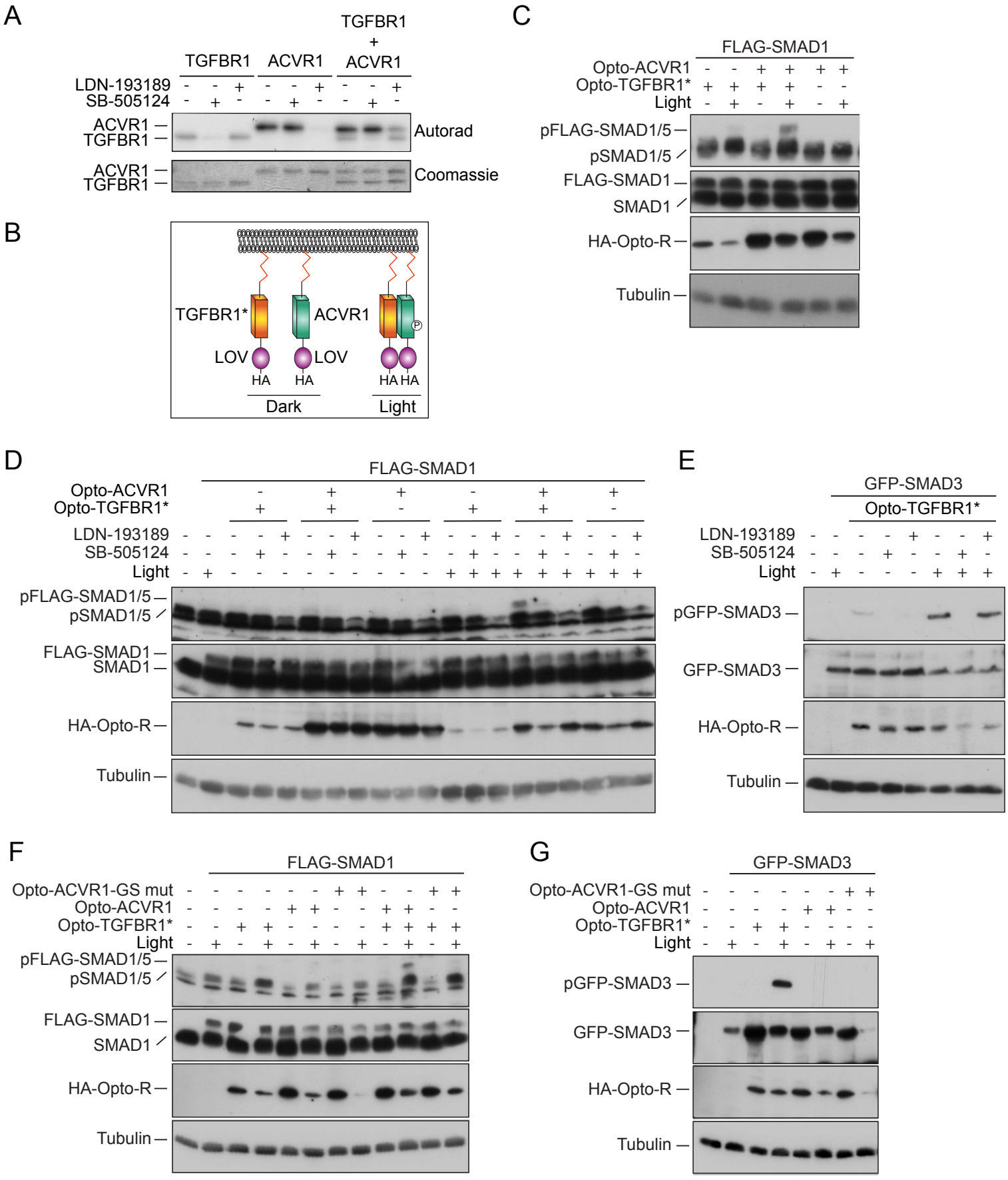
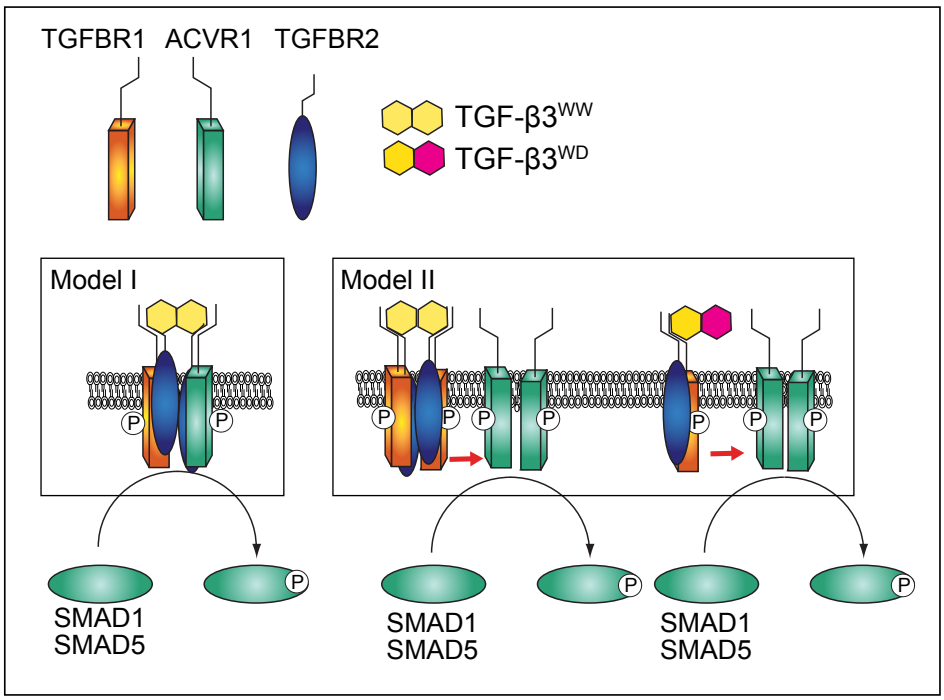

A
TGFBR1
ACVR1
TGFBR1 + ACVR1
LDN-193189
SB-505124
ACVR1
TGFBR1
Autorad
Coomassie
B
TGFBR1*
ACVR1
LOV
HA
Dark
Light
C
FLAG-SMAD1
Opto-ACVR1
Opto-TGFBR1*
Light
pFLAG-SMAD1/5
pSMAD1/5
FLAG-SMAD1
SMAD1
HA-Opto-R
Tubulin
D
FLAG-SMAD1
Opto-ACVR1
Opto-TGFBR1*
LDN-193189
SB-505124
Light
pFLAG-SMAD1/5
pSMAD1/5
FLAG-SMAD1
SMAD1
HA-Opto-R
Tubulin
E
GFP-SMAD3
Opto-TGFBR1*
LDN-193189
SB-505124
Light
pGFP-SMAD3
GFP-SMAD3
HA-Opto-R
Tubulin
F
FLAG-SMAD1
Opto-ACVR1-GS mut
Opto-ACVR1
Opto-TGFBR1*
Light
pFLAG-SMAD1/5
pSMAD1/5
FLAG-SMAD1
SMAD1
HA-Opto-R
Tubulin
G
GFP-SMAD3
Opto-ACVR1-GS mut
Opto-ACVR1
Opto-TGFBR1*
Light
pGFP-SMAD3
GFP-SMAD3
HA-Opto-R
Tubulin

Figure 2

A


TGFBR1
ACVR1
TGFBR2
TGF-β3WW
TGF-β3WD
Model I
Model II
P
SMAD1
SMAD5
SMAD1
SMAD5
SMAD1
SMAD5

B

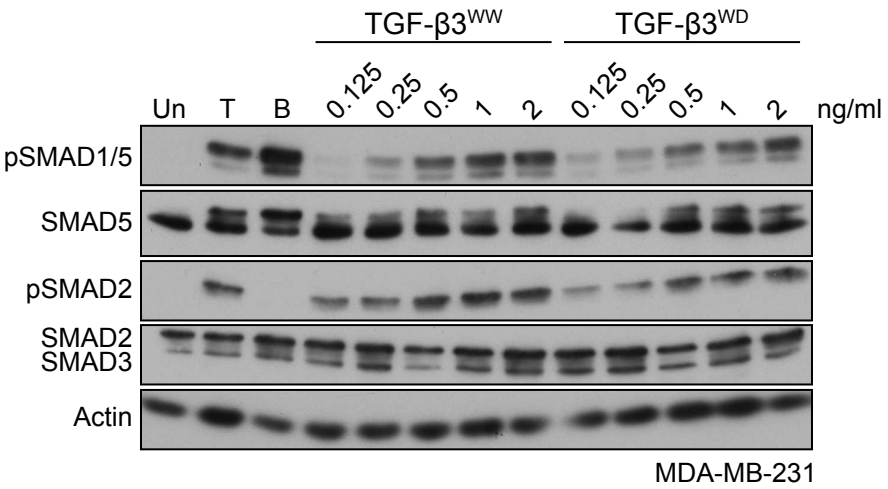

TGF-β3WW
TGF-β3WD
Un
T
B
0.125
0.25
0.5
1
2
0.125
0.25
0.5
1
2
ng/ml
pSMAD1/5
SMAD5
pSMAD2
SMAD2
SMAD3
Actin
MDA-MB-231

Figure 3

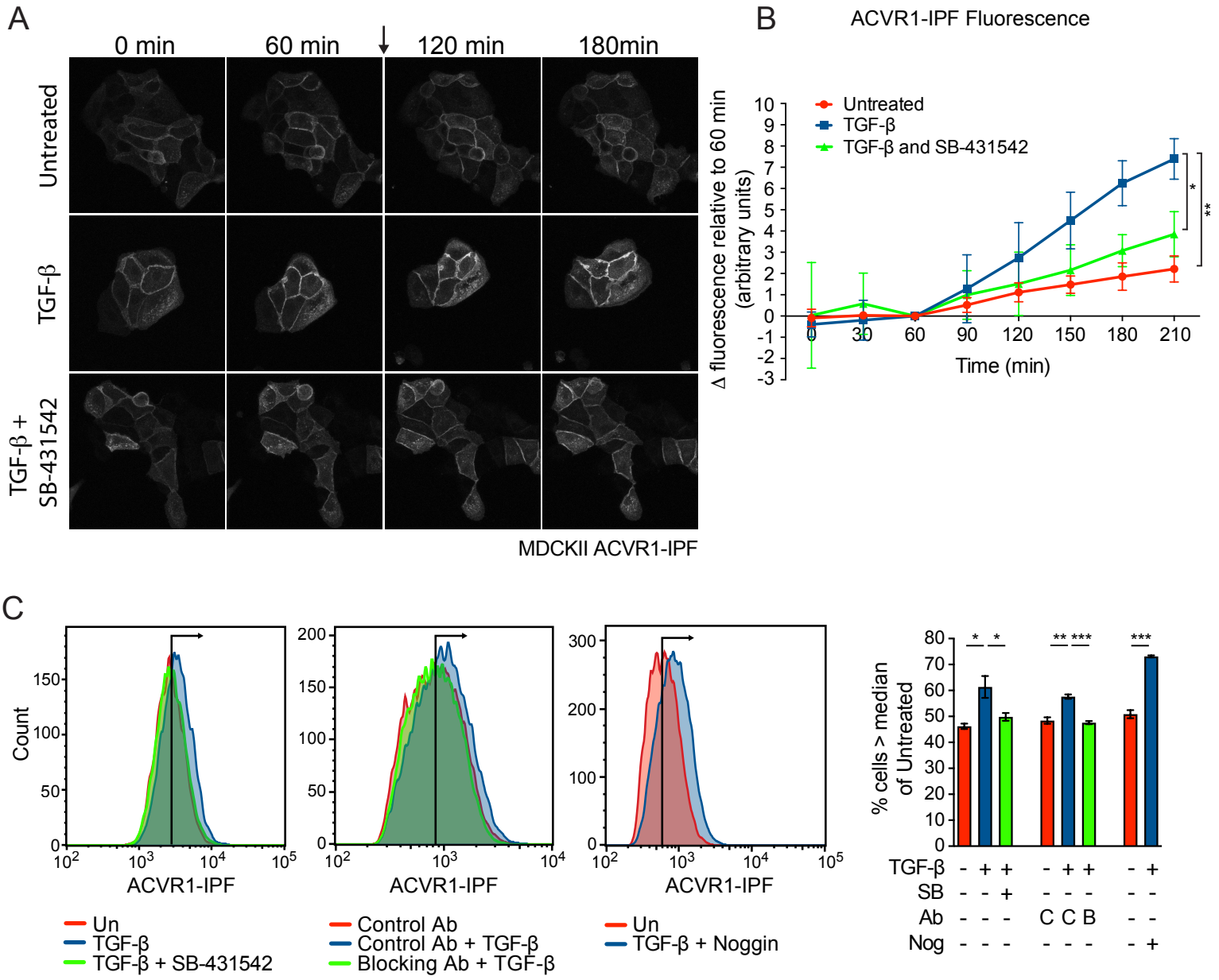

A
0 min
60 min
120 min
180min
Untreated
TGF-β
TGF-β + SB-431542
MDCKII ACVR1-IPF
B
ACVR1-IPF Fluorescence
Untreated
TGF-β
TGF-β and SB-431542
Δ fluorescence relative to 60 min (arbitrary units)
Time (min)
C
Count
ACVR1-IPF
Un
TGF-β
TGF-β + SB-431542
Control Ab
Control Ab + TGF-β
Blocking Ab + TGF-β
Un
TGF-β + Noggin
% cells > median of Untreated
TGF-β
SB
Ab
Nog

Figure 4

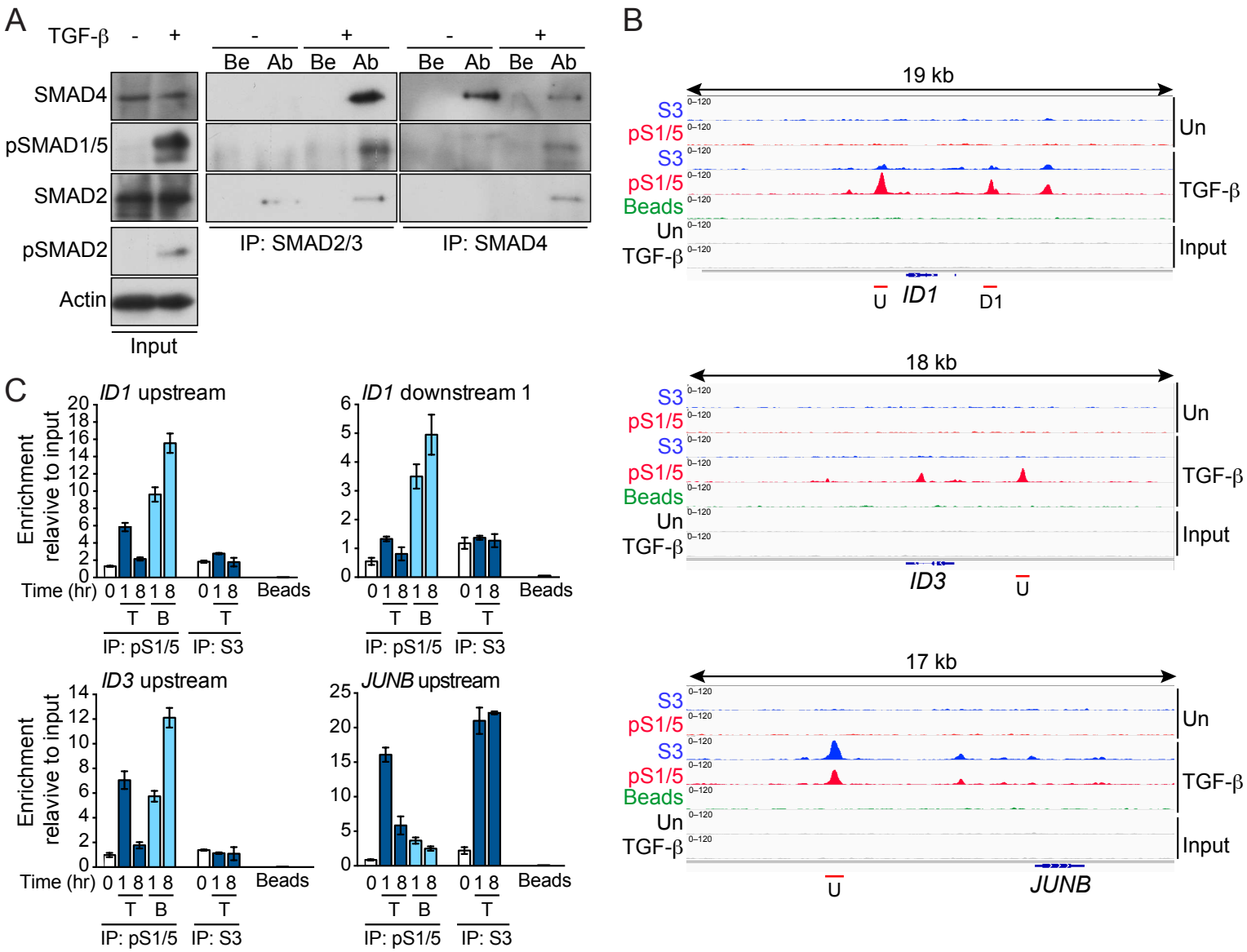

A
TGF-β - +
- + - +
Be Ab Be Ab Be Ab Be Ab
SMAD4
pSMAD1/5
SMAD2
pSMAD2
Actin
IP: SMAD2/3
IP: SMAD4
Input
B
19 kb
S3
pS1/5
Beads
Un
TGF-β
Un
TGF-β
Input
ID1
D1
18 kb
ID3
17 kb
JUNB
C
ID1 upstream
ID1 downstream 1
ID3 upstream
JUNB upstream
Enrichment relavive to input
Time (hr) 0 1 8 1 8 0 1 8 Beads
T B T
IP: pS1/5 IP: S3
D
bits
E
GGCGCC
GG[A/T]GCC
G[A/T/C]CGCC
GGG[G/T]CC
Other

Figure 5

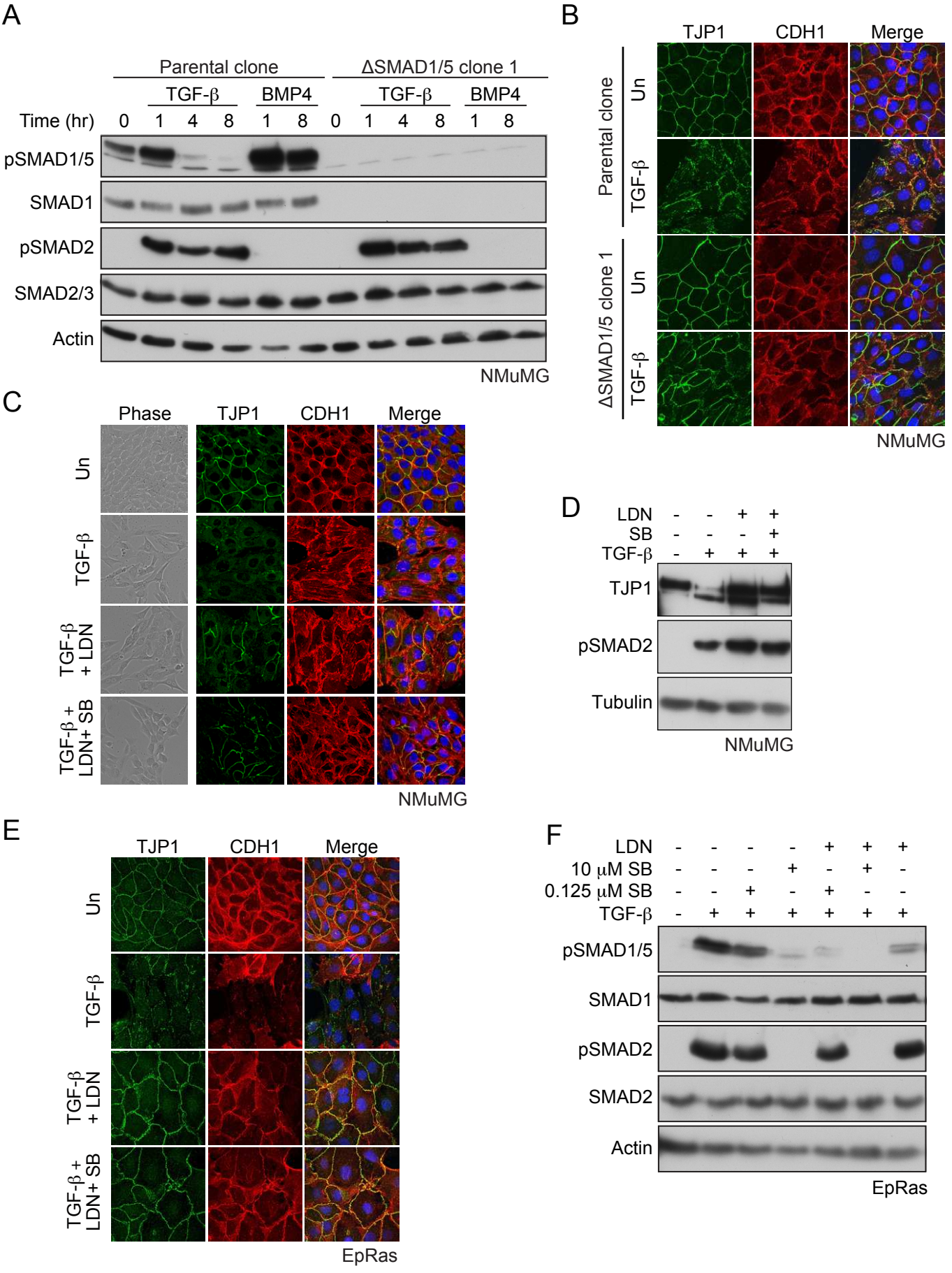

A
Parental clone
ΔSMAD1/5 clone 1
TGF-β
BMP4
TGF-β
BMP4
Time (hr)
0 1 4 8 1 8 0 1 4 8 1 8
pSMAD1/5
SMAD1
pSMAD2
SMAD2/3
Actin
NMuMG
B
TJP1
CDH1
Merge
Parental clone
Un
TGF-β
ΔSMAD1/5 clone 1
Un
TGF-β
NMuMG
C
Phase
TJP1
CDH1
Merge
Un
TGF-β
TGF-β + LDN
TGF-β + LDN+ SB
NMuMG
D
LDN - - + +
SB - - - +
TGF-β - + + +
TJP1
pSMAD2
Tubulin
NMuMG
E
TJP1
CDH1
Merge
Un
TGF-β
TGF-β + LDN
TGF-β + LDN+ SB
EpRas
F
LDN - - - - + + +
10 μM SB - - - + - + -
0.125 μM SB - - + - + - -
TGF-β - + + + + + +
pSMAD1/5
SMAD1
pSMAD2
SMAD2
Actin
EpRas

Figure 6

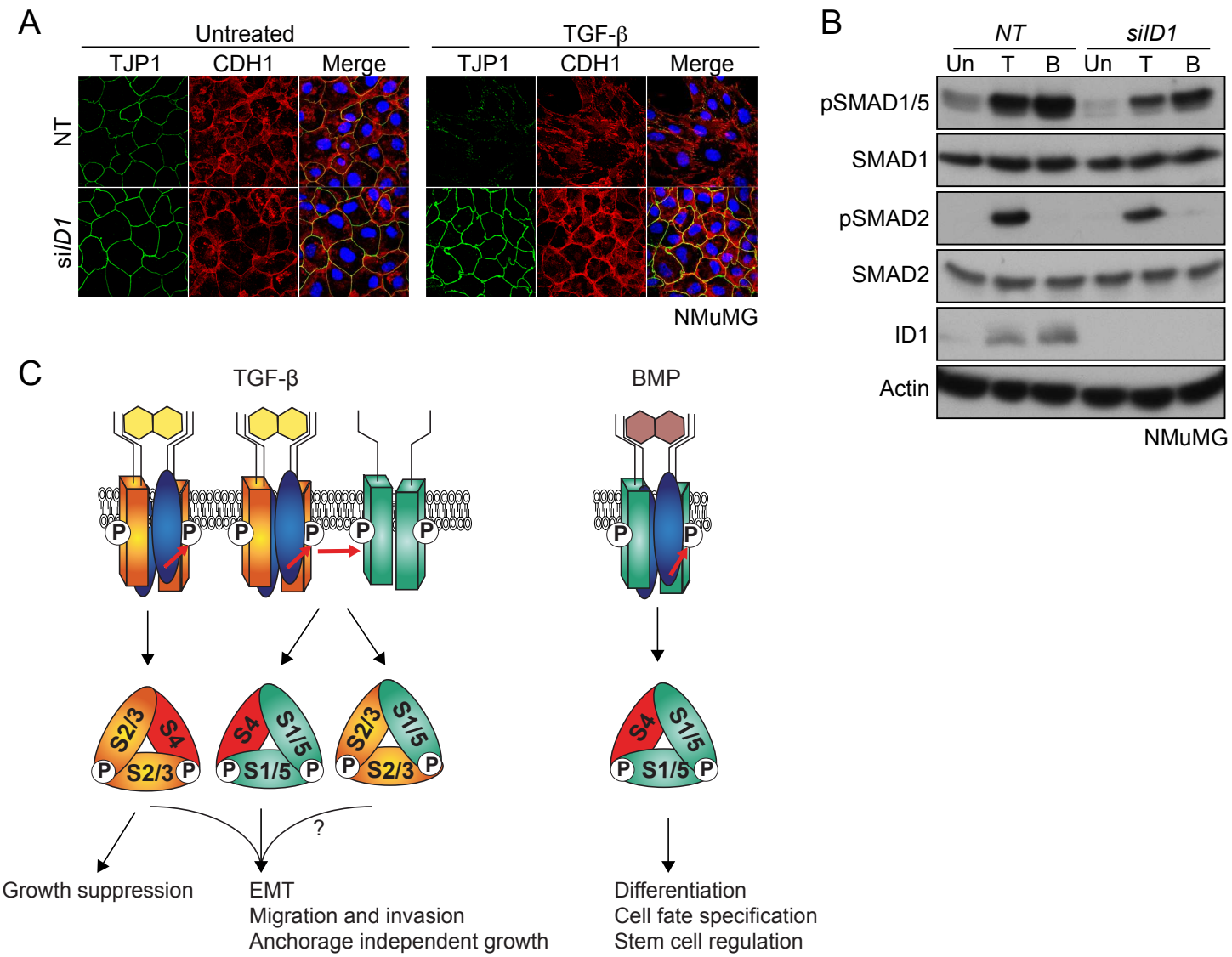

A
Untreated
TGF-β
TJP1
CDH1
Merge
NT
siID1
NMuMG
B
NT
siID1
Un
T
B
pSMAD1/5
SMAD1
pSMAD2
SMAD2
ID1
Actin
NMuMG
C
TGF-β
BMP
S2/3
S4
S1/5
P
?
Growth suppression
EMT
Migration and invasion
Anchorage independent growth
Differentiation
Cell fate specification
Stem cell regulation

Figure 7

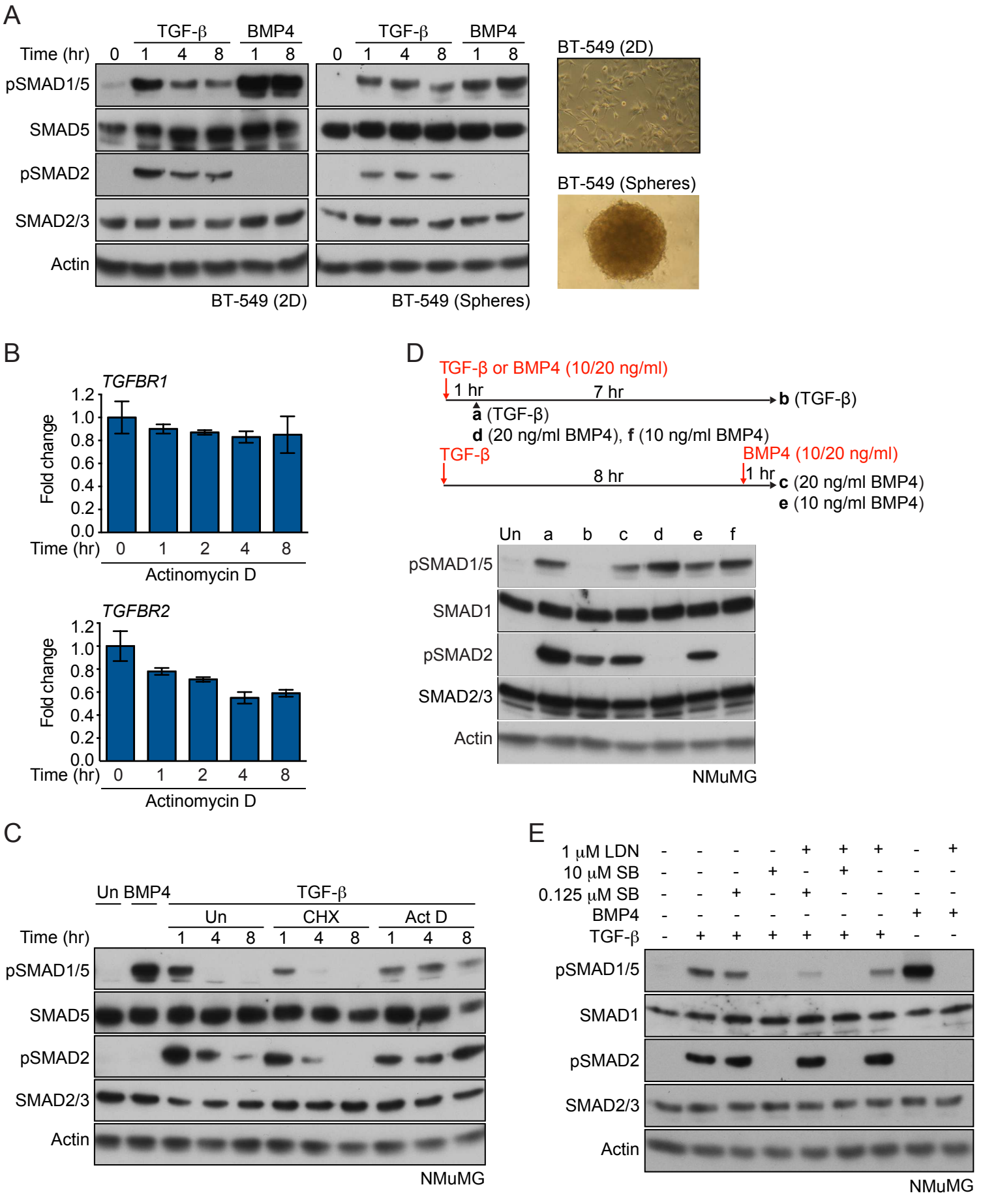


Figure 1 - figure supplement 1

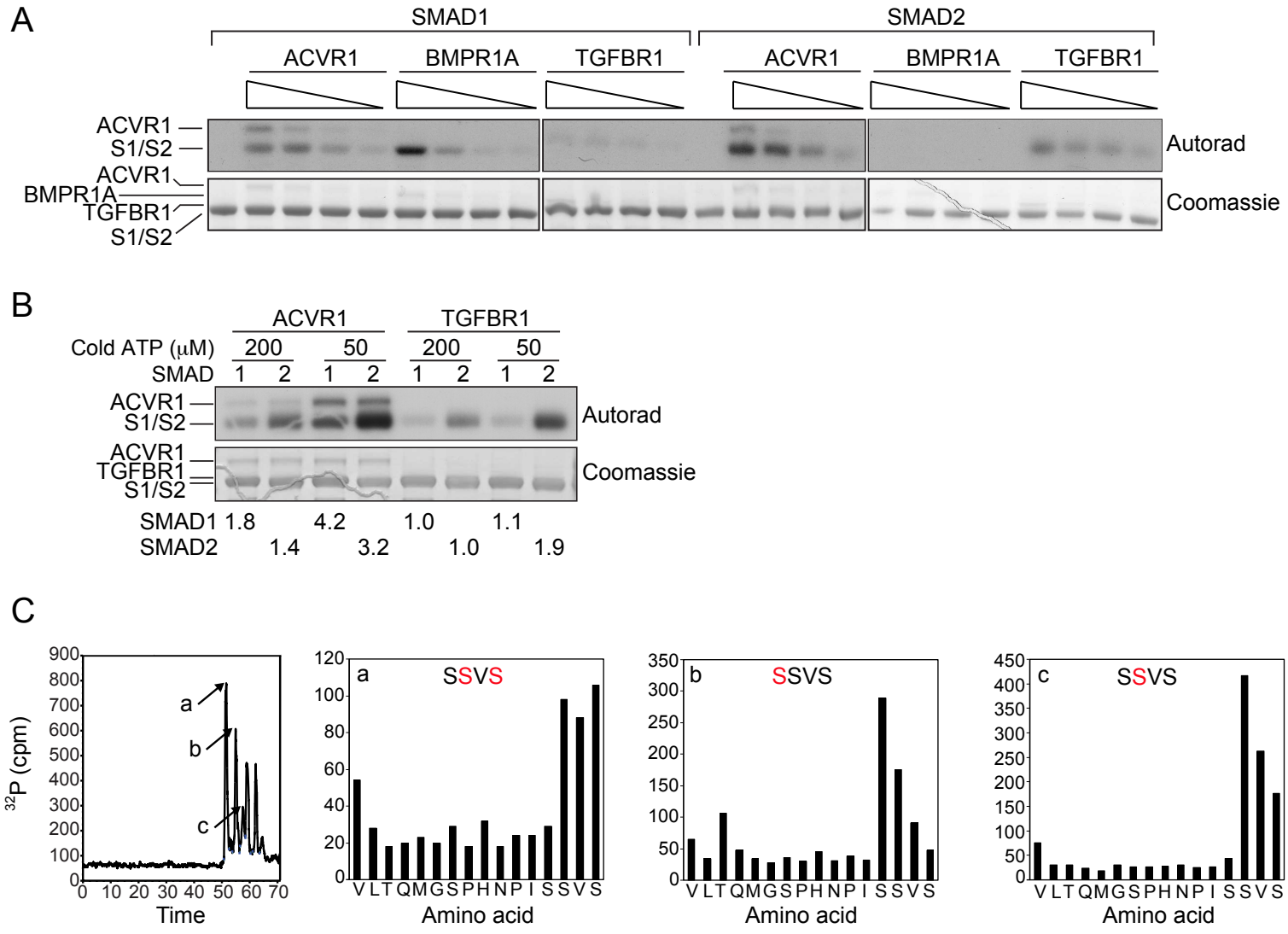


Figure 1 - figure supplement 2

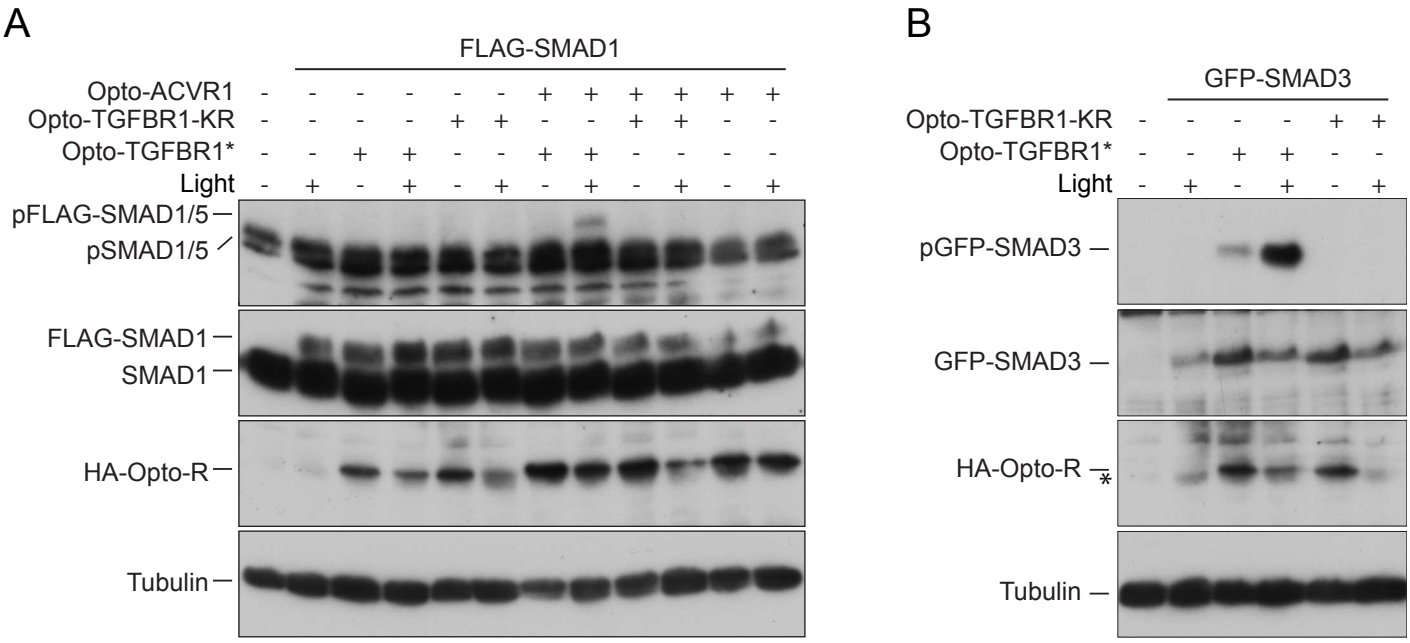


Figure 2 - figure supplement 1

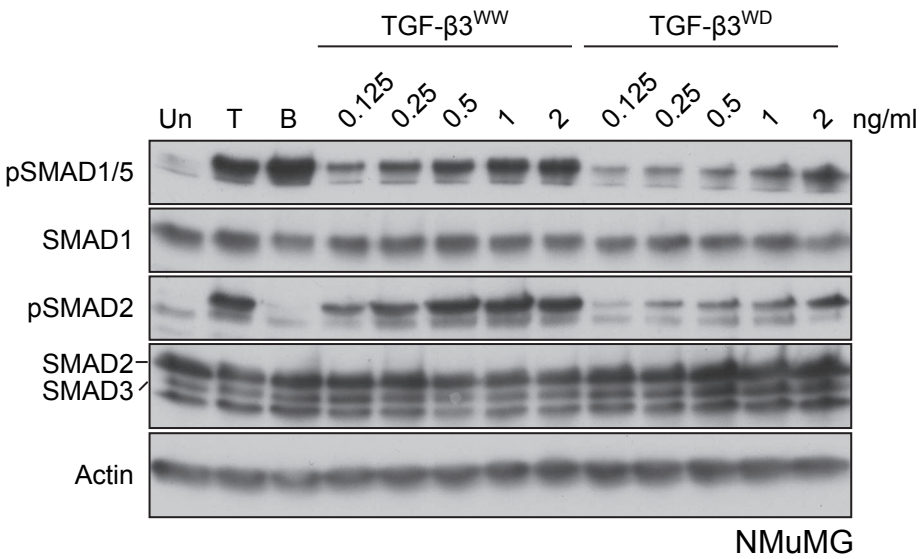


Figure 3 - figure supplement 1

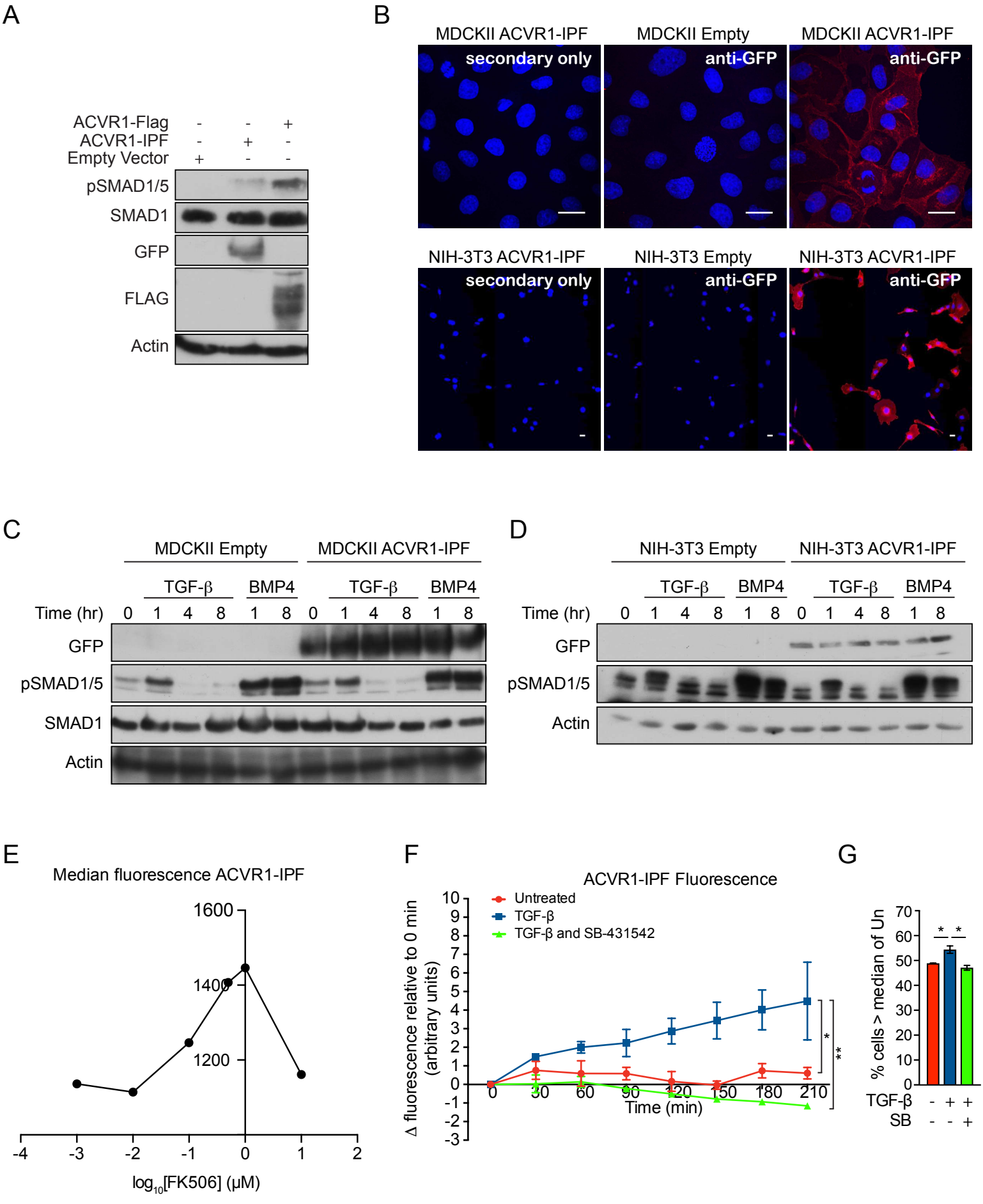


Figure 4 - figure supplement 1

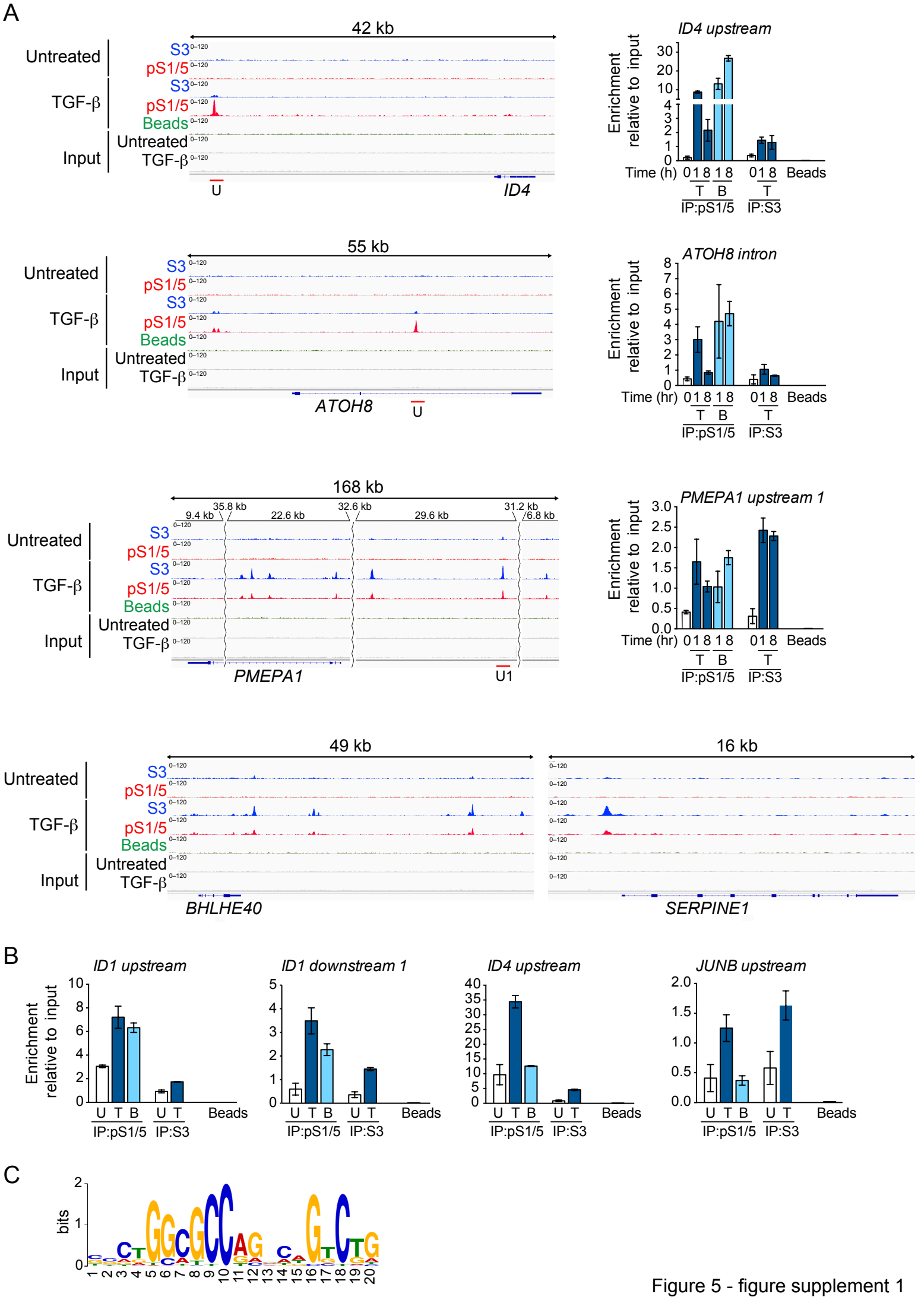


Figure 5 - figure supplement 1

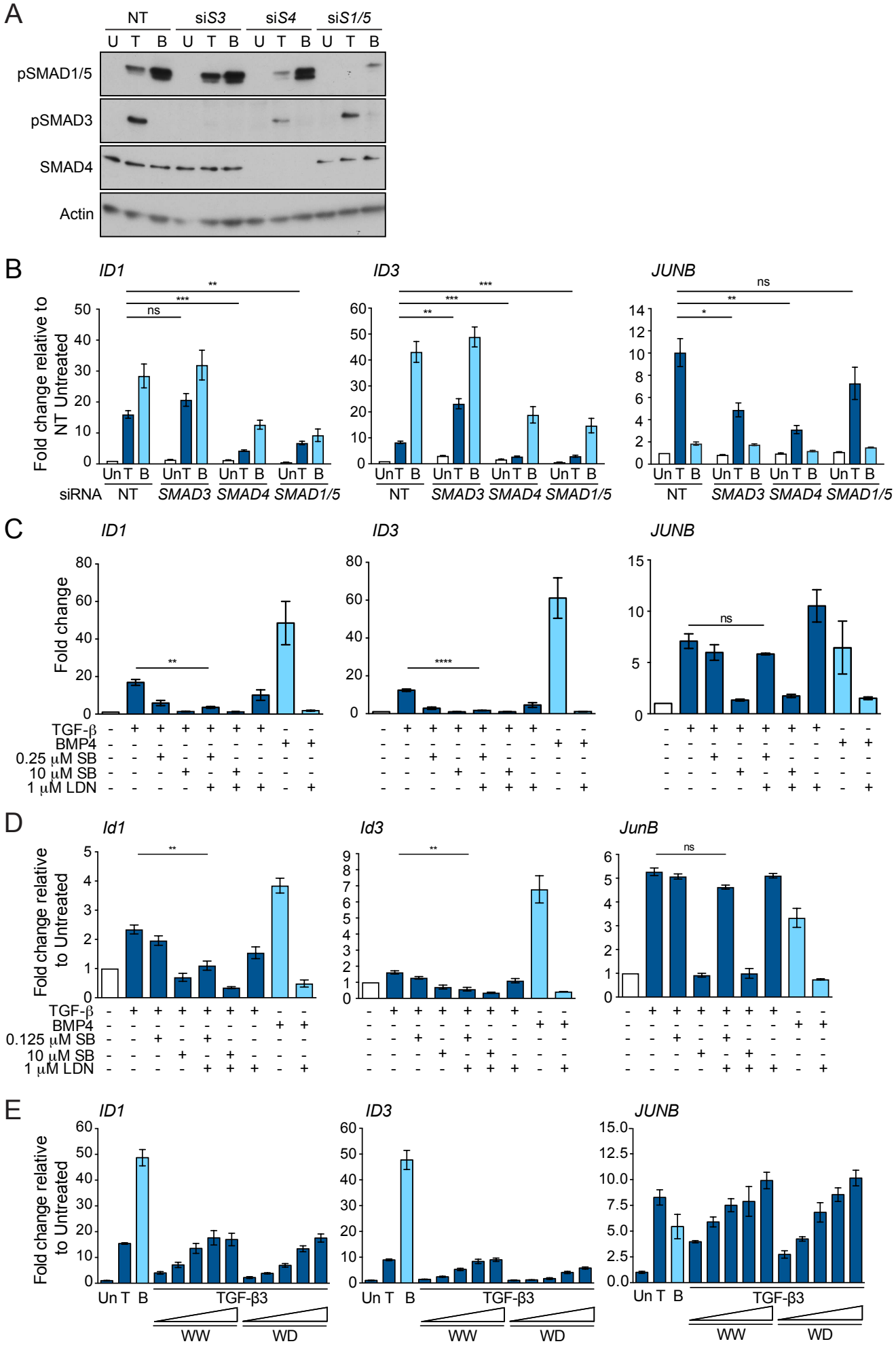


Figure 5 - figure supplement 2

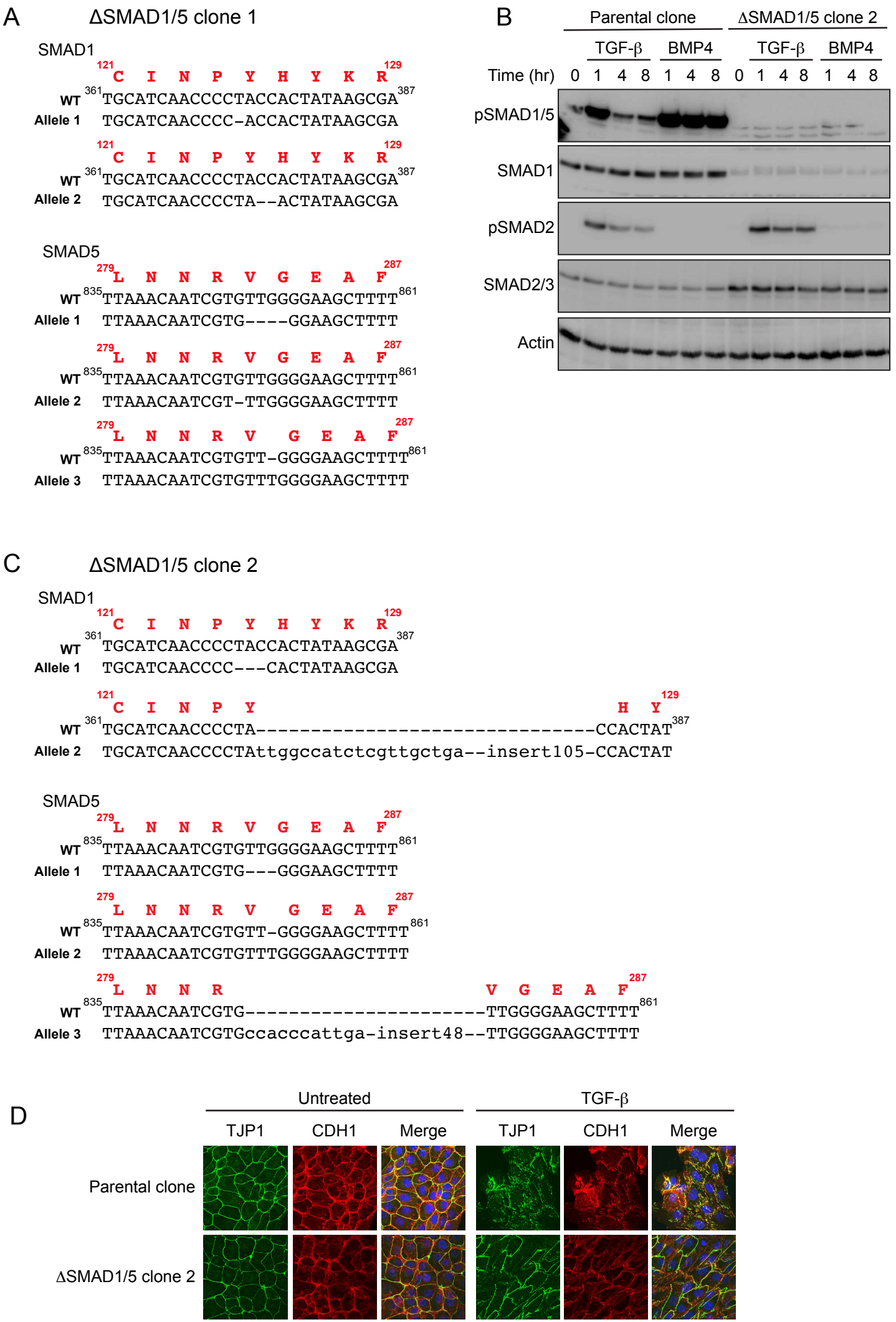


Figure 6 - figure supplement 1

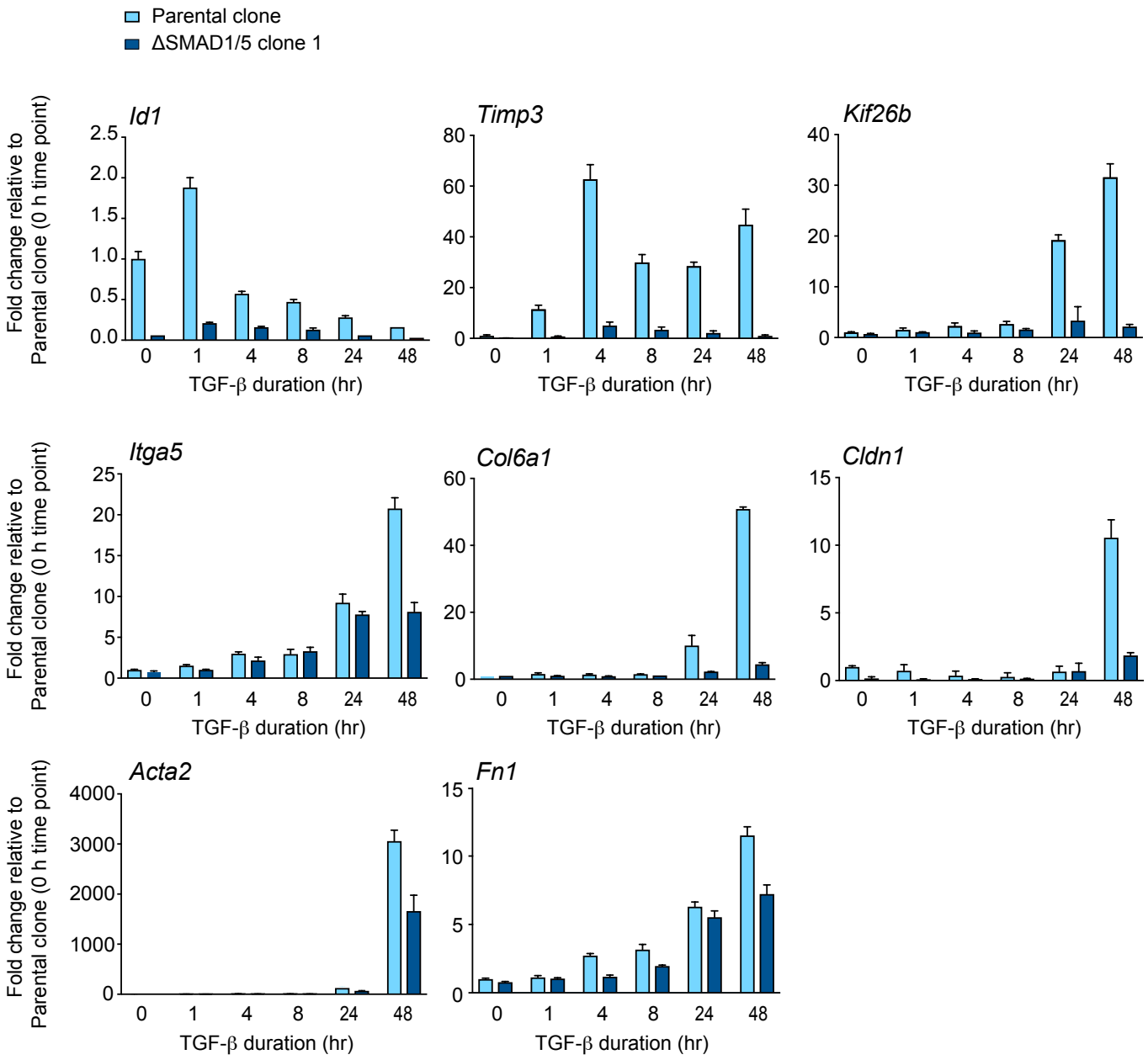


Figure 6 - figure supplement 2

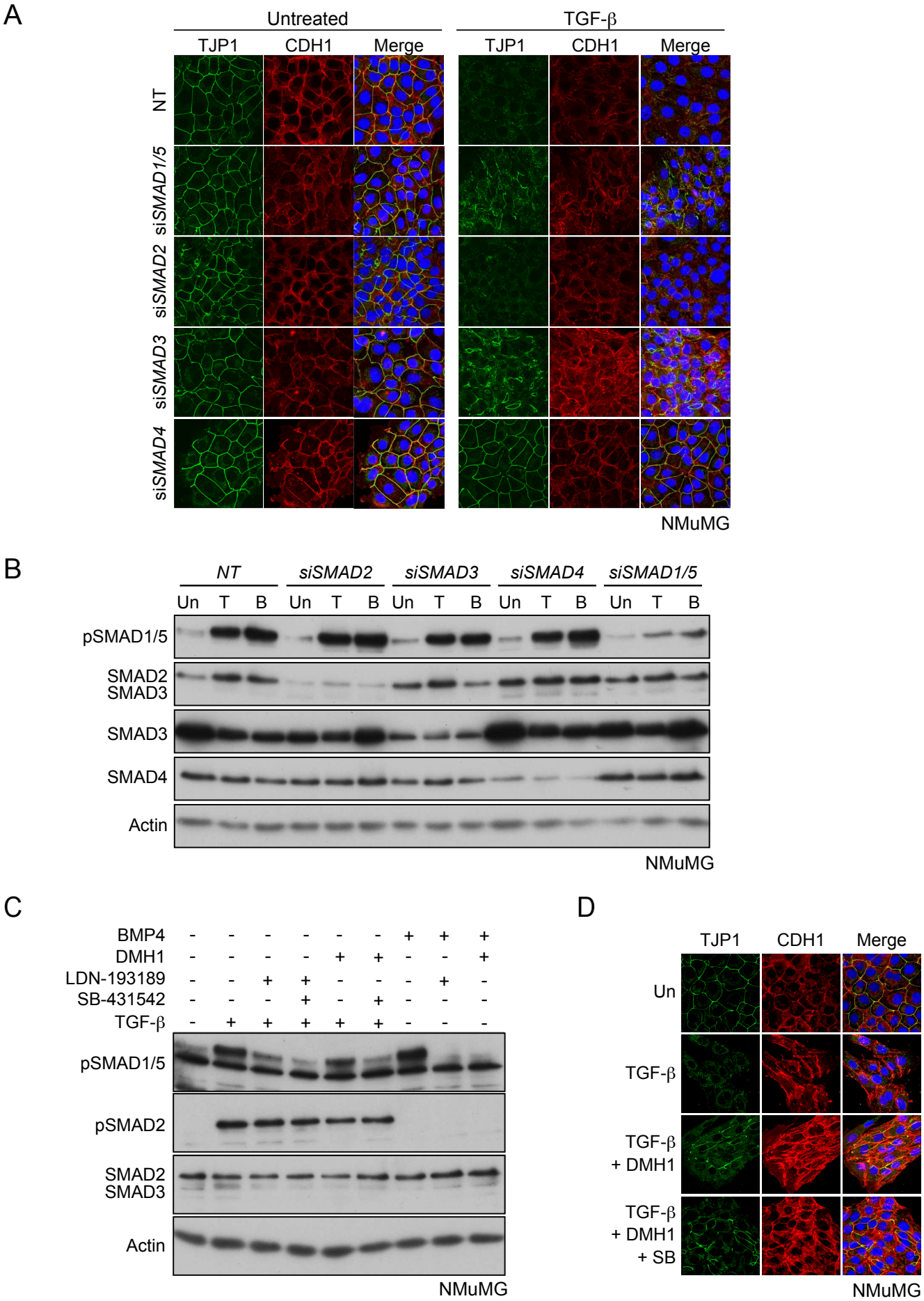


Figure 6 - figure supplement 3

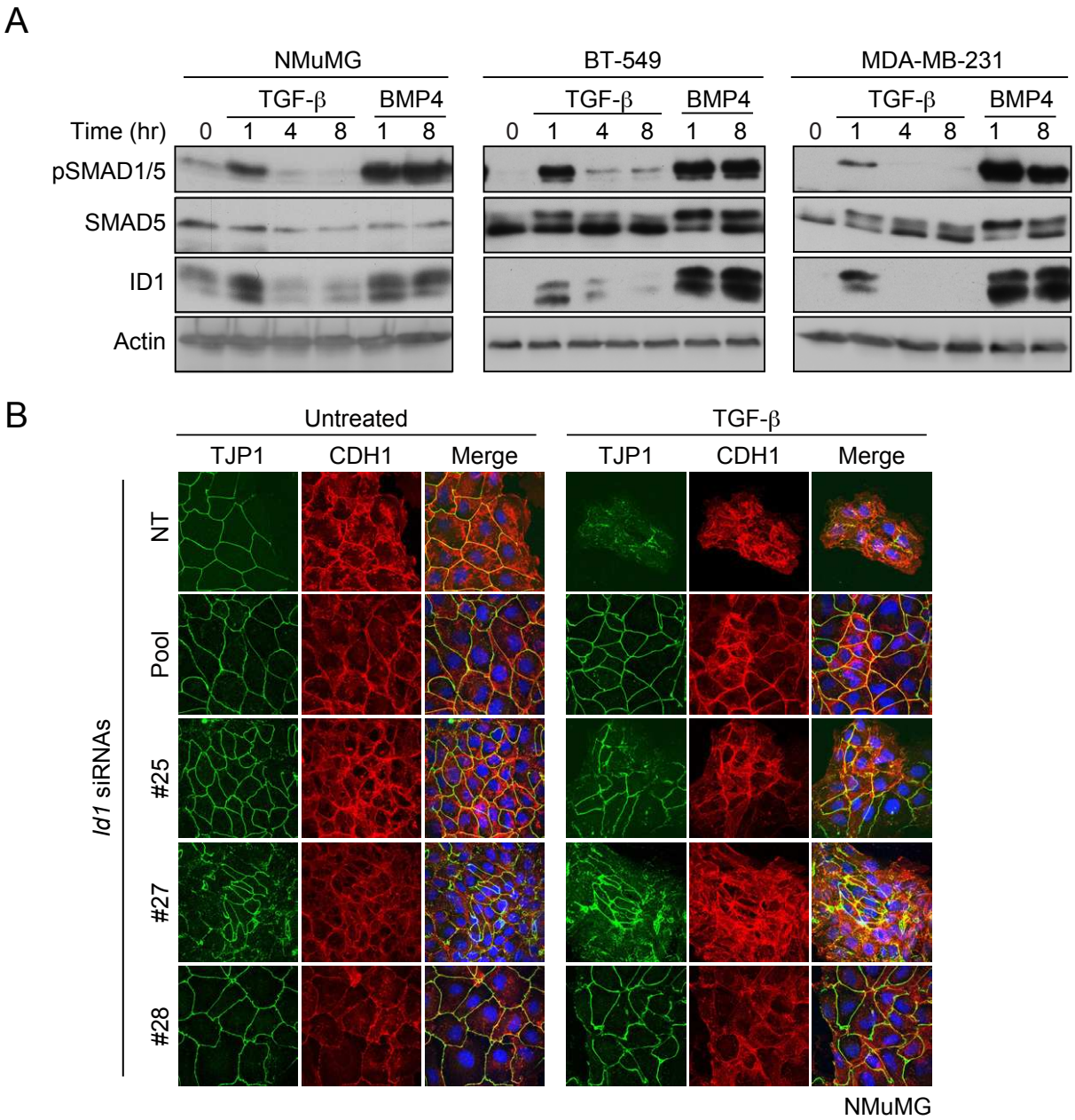


Figure 7 - figure supplement 1