Experimental design

1. Sample size

Describe how sample size was determined. The sample sizes are determined based on the prevailing and widely accepted practice and methods. In addition, the sample sizes for each experiment have been detailed in the Figure legend and Method.

2. Data exclusions

Describe any data exclusions. For IP-MS: A contamination protein database in Maxquant was used to exclude peptides that match this database. The parameters for database searching were set as: 1) Carbamidomethylation of cysteines was set as fixed modification and protein N-terminal acetylations as well as oxidation of methionines were set as variable modification for the peptide search; 2) A maximum mass deviation of 7.5 ppm was allowed for precursor’s identification and 0.5 Da was set as match tolerance for fragment identification (acquisition in orbitrap); 3) Up to two missed cleavages was allowed for trypsin digestion; 4) Peptide length of seven amino acids was considered for protein identification and quantification; 5) The minimum number of total peptides a protein group was set to 1 and the minimum number of razor+ unique peptides was set to 1; 6) The minimum Andromeda score for accepting an MS/MS identification for modified peptides was set to 40; 7) Using unique and razor peptides for protein quantification; 8) The false discovery rates (FDR) for peptide and protein identifications were set to a maximum 1% and 5% respectively.

For ChIP-seq: Raw reads (56-72 Million reads per sample) were first cleaned for adapter sequences using Trim Galore (version 0.4.4, a wrapper tool of Cutadapt, www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with parameters “--paired --stringency 1 --dont_gzip --length 20 -q 20 --trim1”, in which reads with length less than 20 (-length 20) or average quality score less than 20 (-q 20) were removed for downstream analysis. Cleaned reads were aligned to mm10 using Bowtie (version 1.2, Ben Langmead, Cole Trapnell, Mihai Pop and Steven L Salzberg Ultrafast and memory-efficient alignment of short DNA sequences to the human genome Genome Biology 2009 10:R25 with the parameters – “-p 10 -m 1 -v 2 -X 1500 –S”. These parameters ensured that fragments up to 1500bp (-X1500) and mismatch up to 2 (-v2) were allowed to align, and that only unique aligning reads were collected (-m1).

For RNA-seq: Reads (32-45 Million reads per sample) were first cleaned for adapter sequences using Trim Galore (version 0.4.4) with parameters “--paired --stringency 1 --dont_gzip --length 20 -q 25 --trim1”. Cleaned reads were aligned to the UCSC mm10 reference genome using STAR (Dobin A et al. STAR: ultrafast universal RNA-seq aligner Bioinformatics. 2013 Jan 1;29(1):15-21) with default parameters. The reads count for each gene was calculated by featureCounts (Liao Y, Smyth GK and Shi W 2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 30(7):923-30 with parameter “-T 6 -t exon -g gene_name -M”. The quantification results were then analyzed with the Bioconductor package DESeq2 (Love MI, Huber W and Anders S 2014). “Moderated estimation of
3. Replication
Describe whether the experimental findings were reliably reproduced.
The experimental findings were reliably reproduced.

4. Randomization
Describe how samples/organisms/participants were allocated into experimental groups.
Mice with needed genotypes and ages were chosen randomly. For example, different group of mice (e.g. WT and KO, littermates if possible) were mixed in one cage, and treated equally without knowing the identity of the mice. The mice were then identified upon the completion of the experiments.

5. Blinding
Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
Samples were labeled with simple numbers. The person who performed the experiments did not know the exact samples until after data were analyzed.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>☒</td>
</tr>
</tbody>
</table>

The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)

A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.

A statement indicating how many times each experiment was replicated.

The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section).

A description of any assumptions or corrections, such as an adjustment for multiple comparisons.

The test results (e.g. \(p\) values) given as exact values whenever possible and with confidence intervals noted.

A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range).

Clearly defined error bars.

See the web collection on statistics for biologists for further resources and guidance.

Software
Policy information about availability of computer code

7. Software
Describe the software used to analyze the data in this study.

Flowjo, MAC version, 9.3.2, Graghpad Prism 6.0c, XCalibur software version 2.1, MaxQuant version 1.5.0.25, Trim Galore version 0.4.4, Bowtie version 1.2, STAR, featureCounts, Bioconductor package DESeq2, HOMER Version 4, GREAT Version 3

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents
Policy information about availability of materials

8. Materials availability
Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No restricted material was used.
9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Antibodies including anti-CD4 (GK1.5), anti-CD25 (PC61.5), anti-CD62L (MEL-14), anti-CD44 (IM7), anti-CD45.1 (A20), anti-CD45.2 (104), anti-IFN-γ 188 (XMG1.2), anti-Thy1.1 (OX-7) and anti-IL-17A (TC11-18H10.1) from Biolegend; anti-CD3 (145-2C11), anti-CD28 (37.51), anti-IFN-γ (XMG1.2) and anti-IL-4 (11B11) from BioXcell; anti-Ski (H-329), anti-217 RORγt (B2D), anti-β-actin (I-19) and anti-IgG (sc-2027) from Santa Cruz, anti-Smad4 (D3M6U) and anti-pSmad3(C25A9) from Cell Signaling Technology, anti-Smad4 (EP618Y) from Abcam were obtained from indicated commercial vendors with ensured quality. In addition, all the antibodies has been used in multiple experiments to detect intended markers and proteins in control samples with expected results to validate their effectiveness in our study. Moreover, Smad4, RORγt, and Ski deficient T cells have been used to further validate related antibodies in our study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

b. Describe the method of cell line authentication used.

c. Report whether the cell lines were tested for mycoplasma contamination.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used in the presented data.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Smad4fl/fl, Tgfbbr2fl/fl, Cd4cre, ERCre, Rag1-/-, Rorc-/-, Cre-dependent-Cas9 knockin (CdC) and CD45.1 congenic wild-type mice were on the C57BL/6 background. 6-18 weeks old same sex (both males and females) mice were used. Primary T cells or HSC were isolated from the mice and then used for the subsequent experiments.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Study did not involve human research participants
ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

### Data deposition

1. For all ChIP-seq data:
   - a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.
   - b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. (The entry may remain private before publication.)

   - track type=bigWig name="Smad4_WT1" db=mm10 visibility=full alwaysZero=on maxHeightPixels=40 color=0,0,153 bigDataUrl=http://snpinfo.niehs.nih.gov/ucscview/takaku/Smad4_MT-384.singleFrag.Normalized.bigWig
   - track type=bigWig name="Smad4_WT2" db=mm10 visibility=full alwaysZero=on maxHeightPixels=40 color=0,0,153 bigDataUrl=http://snpinfo.niehs.nih.gov/ucscview/takaku/Smad4_MT-385.singleFrag.Normalized.bigWig
   - track type=bigWig name="Smad4_KO1" db=mm10 visibility=full alwaysZero=on maxHeightPixels=40 color=0,0,153 bigDataUrl=http://snpinfo.niehs.nih.gov/ucscview/takaku/Smad4_MT-386.singleFrag.Normalized.bigWig
   - track type=bigWig name="Smad4_KO2" db=mm10 visibility=full alwaysZero=on maxHeightPixels=40 color=0,0,153 bigDataUrl=http://snpinfo.niehs.nih.gov/ucscview/takaku/Smad4_MT-387.singleFrag.Normalized.bigWig

3. Provide a list of all files available in the database submission.

4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

   - ChIP-seq data for WT#1, WT#2, KO#1, KO#2

### Methodological details

5. Describe the experimental replicates.

   - We used two replicates for ChIP-seq. (Two independent experiments)
6. Describe the sequencing depth for each experiment.  
Total numbers of reads are as follows.  
Smad4 ChIP-seq WT1: 72563366, Smad4 ChIP-seq WT2: 62982483  
Smad4 ChIP-seq KO1: 60135618, Smad4 ChIP-seq KO2: 61833161,  
input WT1: 56967653, input WT2: 64144763, input KO1: 63174257, input KO2: 66197177.  
For the analysis, we used uniquely-mapped and deduplicated reads. (Smad4 ChIP-seq WT1: 33955608, Smad4 ChIP-seq WT2: 39070319, Smad4 ChIP-seq KO1: 35688532, Smad4 ChIP-seq KO2: 33577809, input WT1: 32558640, input WT2: 36817054, input KO1: 32535178, input KO2: 35203178)

7. Describe the antibodies used for the ChIP-seq experiments.  
Anti-Smad4 (EP618Y, Abcam)

8. Describe the peak calling parameters.  
We used program called findPeaks from Homer (Version 4) to call all reported ChIP-seq peaks in this manuscript. The basic idea is to identify regions in the genome where we find more sequencing reads than we would expect to see by chance. Two input data was pooled together before to do peak calling. findPeaks was run default parameters. All significant peaks were associated with genes using GREAT Version 3.

9. Describe the methods used to ensure data quality.  
We observed significant enrichment at many loci including previously known Smad4 binding sites. We prepared ChIP-seq libraries from Smad4 KO cells using the same protocol, and the signals were significantly lower in the Smad4 KO ChIP-seq data. More than 20,000 peaks were identified in the Smad4 WT cells while the KO cells only showed ~1,200 peaks.

10. Describe the software used to collect and analyze the ChIP-seq data.  
HOMER Version 4  
GREAT Version 3
Flow Cytometry Reporting Summary

Data presentation

For all flow cytometry data, confirm that:

1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
3. All plots are contour plots with outliers or pseudocolor plots.
4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.
   Naive CD4+ T cells were harvested from mouse spleen and lymph nodes, then MACS purified with CD4-beads or with FACS sorter. After activation, cells were collected and Ficol was used to remove cell debris. Spinal cord lymphocytes: mice were euthanized, then perfused with 50-60ml PBS containing heparin. Spinal cord removed and digested with 2mg/ml collagenase D. for 45min at 37C with brief vortexing every 15min, then filtered with 40ml nylon mesh. Cells were washed with PBS and then re-suspended in 6ml of 38% Percoll solution. Cells were separated at 2000rpm for 20m. Cell pellets were washed with 1xPBS and then re-suspend in PBS or culture media for subsequent procedures.

6. Identify the instrument used for data collection.
   BD Biosciences, FACSCanto II.

7. Describe the software used to collect and analyze the flow cytometry data.
   Flowjo, MAC version,v9.3.2

8. Describe the abundance of the relevant cell populations within post-sort fractions.
   The purity of sorted cell purity was >95%, determined by flow-cytometry.

9. Describe the gating strategy used.
   Preliminary FSC/SSC gating was used to gate on lymphocytes population. Specific T cell populations were gated based on the specific antibody staining as described for each experiment. Positive and negative populations were defined and gated clearly by the negative control sample and samples with control IgG staining.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒