Iro/Irx transcription factors negatively regulate Dpp/TGF-β pathway activity during intestinal tumorigenesis

Martorell et al.

Supplementary Methods

Genotypes. *yw hsp70-flp; esg Gal4 UAS-GFP UAS-Ras^{V12}/Cyo; FRT82B Gal80/TM6b* were crossed with *yw hsp70-flp; Sp/Cyo; FRT82B Apc2^{N175K}Apc^{Q8}/TM6b*, *yw hsp70-flp;UAS-Tkv^{Q253}/Cyo; FRT82B Apc2^{N175K}Apc^{Q8} /TM6b*, *yw hsp70-flp; Sp/Cyo; FRT82B Apc2^{N175K}Apc^{Q8} UAS-Mirr^{RNAi} /TM6b*, *yw hsp70-flp; UAS-Tkv^{DN}/Cyo; FRT82B Apc2^{N175K}Apc^{Q8} /TM6b and yw hsp70-flp; UAS-Tkv^{DN}/Cyo; FRT82B Apc2^{N175K}Apc^{Q8} UAS-Mirr^{RNAi} /TM6b* to generate Apc-Ras, Apc-Ras-Tkv*, Apc-Ras-Mirr^{RNAi}, Apc-Ras-Tkv^{DN} and Apc-Ras-Tkv^{DN}-Mirr^{RNAi} clones respectively. Control clones were generated by crossing *yw UAS flp ; esg Gal4 UAS-GFP/CyO; FRT82B Gal80/TM6b* with *yw hsp70-flp;Sp/Cyo; FRT82B/TM6b* flies. *yw hsp70-flp; esg Gal4 UAS-GFP/Cyo; Tub>Gal80^{ts}/TM6b* were crossed with *w; UAS-Tkv^{Q253}/Cyo* or *w; Sp/CyO; UAS-Mirr RNAi/TM3* to express Tkv^{Q253} or Mirr RNAi in the progenitor cells. Adult flies were kept at 29ºC for 10 days to allow the expression of the UAS constructs. Apc2^{N175K} is a loss-of-function allele, Apc^{Q8} is a null allele, UAS-Ras^{V12} and UAS-Tkv^{Q253} are gain-of-function transgenes. UAS-Tkv^{DN} is a dominant negative transgene. Stocks were obtained form Bloomington Stock Center and VDRC.

Staining and antibodies. Adult female flies were dissected in PBS. All the digestive tract was removed and fixed in PBS and 4% electron microscopy grade paraformaldehyde (Polysciences, USA) for 35 minutes. Samples were rinsed 3 times with PBS, 4% BSA, 0.1% Triton X-100 (PBT-BSA), incubated with the primary antibody overnight at 4°C and with the secondary antibody for 2 hours at room temperature. Finally, the samples were rinsed 3 times with PBT-BSA and mounted in DAPI-containing media (Vectashield, USA). All the steps were performed without mechanical agitation. Primary antibody mouse α-Dl (1:10), was obtained from the Developmental Studies Hybridoma Bank (DSHB), rabbit α-PH3 (1:100) from Cell Signalling (USA), rabbit α-Pdm1 (1:1000) was a gift of Dr. Yang Xiaohang (Institute of Molecular and Cell Biology, Singapore), rabbit α-Mirror (1:1000) was a gift of Dr. Helen McNeill (Samuel Lunenfeld Research Institute, Toronto). Secondary antibodies

were from Invitrogen (USA). Images were obtained on a Leica SPE or Leica SP5 confocal microscopy and processed in Photoshop CS5 (Adobe, USA).

Clone gene expression analysis. *Drosophila* midguts were dissected from control, Apc, Apc-Ras, Apc-Ras-Tkv* and Apc-Ras-Mirr^{RNAi} flies. In order to obtain single cell suspensions, samples were incubated in PBS containing 0,4mg/ml dispase (Gibco) for 15 min at 37ºC, syringed using a 27G needle and washed twice with PBS. After selecting for the viable population (propidium iodide negative), GFP^{+} cells were isolated by fluorescence activated cell sorting (FACS) using a FACSAria 2.0 (BD Biosciences) and RNA was extracted and amplified as described[46]. Gene expression levels were assessed using Power Sybr Green quantitative PCR (Applied Biosystems) following the manufacturer's instructions. Actin 5C was used as an endogenous control for normalization and differences in target gene expression were determined using the StepOne 2.0 software (Applied Biosystems). All measurements were performed in triplicate from three independent sorting experiments.

Bioinformatics analysis. The matScan software[25] was used to scan promoter regions of 2000 nucleotides upstream of the genes of interest for possible binding sites for known transcription factors. Genomic coordinates were obtained by the biomaRt package of Bioconductor[47] with the dm3 version of the *Drosophila melanogaster* genome. We used 123 position weight matrices (PWM) from the JASPAR database[48]. Results were filtered by the default value of the matScan software (hits above 80% of the maximum possible score value). To further refine the candidate hits, we performed a permutation test by randomly selecting 500 regions of 2000 nucleotides from the dm3 genome and repeating the analysis for all PWMs. We used a cutoff of 0.005 for the resulting p-values. Finally, we downloaded the evolutionary conservation track from UCSC[49] for 14 species close to *Drosophila melanogaster* and filtered out hits with mean scores lower than 0.9. The matScan software[25]was also used to scan promoter regions of 2000 nucleotides upstream of the main components of the TGF-β pathway (*TGFBR1*, *TGFBR2*, *SMAD2*, *SMAD3*, *SMAD4*, *SMAD6*, *SMAD7*) and BMP pathway for possible binding sites for Irx transcription factors. Genomic coordinates were obtained by the biomaRt package of Bioconductor[47] with the hg19 version of the *Homo sapiens* genome. We used 6 position weight matrices (PWM) for IRX genes from the uniProbe database[50]. In particular, we found matrices for IRX2, IRX3 (2 matrices), IRX4, IRX5 and IRX6. Results were filtered using the default value set by matScan (hits above 80% of the maximum possible score). To further refine the

candidate hits, we performed a permutation test by randomly selecting 1000 regions of 2000 nucleotides from the hg19 genome and repeating the analysis for all PWMs. We used a cutoff of 0.002 for the resulting p-values. Finally, we downloaded the evolutionary conservation track from UCSC[49] for 46 species close to *Homo sapiens* and filtered out hits with mean conservation scores lower than 0.7 (ref.[51]).

To assess correlation between genes we selected the most variable probe set and calculated the correlation between z-scores. We used Spearman correlations to account for possible outliers. P-values were calculated using the "cor.test" function as implemented in R. The expression of gene groups was summarized in the first principal component and the resulting values were used to compute correlation against z-scores of the expression of the gene of interest. Before calculating correlations in datasets GSE50115 and GSE8671 expressions were corrected for processing batch, taking the chip scan date as identifier.

To define a list of TGF-β targets we used expression data from epithelial cells before and after being stimulated with TGF-β. We used a p-value cutoff of 0.05 and fold change smaller than -2. The resulting list of genes is reported in Table S1. The transcriptomic data set generated for this study has been deposited in GEO with the accession number GSE59771.

Cell lines, transfections and treatments. Human adenocarcinoma cell lines from rectum (SW837), cervix (HeLa) or mammary gland (MDA-MB-231) were cultured under standard conditions in DMEM 10% FBS, and transfections were performed using PEI (polyethylenimine) reagent (Polysciences Inc.). To generate TGF-β responsive cells, the colon carcinoma cell line LS174T was transfected with a doxycycline-inducible vector containing the cDNA for TGFBRII, as well as the rtTA doxycycline responsive transactivator. Stable clones were derived by selection in blasticidin and hygromycin (Invivogen) and the TBR19 clone was established. TBR19 cells were then infected with doxycycline-inducible lentiviral constructs containing EGFP or Hs IRX5 cDNAs and cells with stable lentiviral integrations were selected by Puromycin treatment (1 ug / mL) (Invivogen). TGF-β (Preprotech) was added to cultures as stated in a final concentration of 5 ng/mL.

Cell cycle profile. TBR19 GFP or IRX5 cells (3×10^5) were seeded and induced with doxycycline for 72h with or without TGF-β. After this period, EdU was added to the

media to a final concentration of 10 μM and cells were incubated for 90 min. Staining was performed following the EdU Flow cytometry protocol according to manufacturer´s instructions (Life technologies). Cells were analyzed in a Gallios flow cytometer (BD) and FACS profile was analyzed with Flowjo software.

Cell competition assay. TBR19 GFP and IRX5 cells (1 x 10^5 each) were co-cultured and induced with doxycycline for 72h with or without TGF-β. Cells were then analyzed by FACS to quantify the % of GFP^+ and $IRX5^+$ cells. 2 x 10^5 cells were then re-seeded in the same conditions. The quantification of GFP^+ and $IRX5^+$ cells in the culture was repeated at 6 and 9 days. Analysis was performed in a Gallios flow cytometer (BD) and analyzed with Flowjo software.

Quantitative RT-PCR of human genes. All human qRT-PCRs were performed using TaqMan Gene Expression Assays (Applied Biosystems). Irx3: Hs01124217_g1, Irx5: Hs00373920_g1, Smad2: Hs00183425_m1, Smad3: Hs00969210_m1, Smad4: Hs00232068_m1, Cdkn2b: Hs_00394703_m1, Myc: Hs00153408_m1. B2M (Hs_99999907_m1), PPIA (Hs99999904_m1) and GAPDH (Hs99999905_m1) were used as endogenous controls for normalization.

Irx knock down. Down-regulation of human Irx3 or Irx5 was achieved by transfecting HeLa or MDA cells with shRNA constructs from the MISSION TRC-Hs 1.0 shRNA library (clones TRCN0000016900 for Irx3 and TRCN0000018302 for Irx5). Cells transfected with a non-targeting shRNA (SHC002) were used as a control. Knock down efficiency was tested by qRT-PCR.

Generation of the Ls174T gene expression signature. TGFBR2 doxycycline inducible expressing Ls174T cells (LS^{TGFBR2}) were generated by the Invitrogen T-Rex system following manufacturer's instructions. LS^{TGFBR2} cells were seeded in DMEM 2% FBS in the presence or absence of doxycycline (Sigma; 1 μm/ml). 16 hours post-seeding recombinant human TGF-β1 (Peprotech; 5 ng/ml) was added to the media of doxycycline-induced cells. Total RNA was extracted using the TRIzol® Plus RNA Purification Kit (Life Technologies) 16 hours after treatment. Briefly, CRC cell lines were scraped from cell culture dishes (Costar) and homogenized by pipetting in TRIzol solution (Life Technologies, 15596-018). After phase separation with chloroform, the upper aqueous phase was then mixed with 70% ethanol and bound to RNA columns (PureLink™ RNA Mini kit, Life Technologies) according to the protocol provided by the manufacturer. RNA was quantified using a Nanodrop spectrophotometer, quality was assessed using an electrophoresis gel (Bioanalyzer, Agilent Technologies), and

high quality samples (RIN≥9) were used as a template. cDNA synthesis and labeling were performed following Affymetrix (Santa Clara, CA) protocols using the Affymetrix kit "One-cycle Target Labeling kit". Labeled cDNA was hybridized in duplicate to a HG-U133 plus 2.0 Affymetrix chip in our core facility (IRB Functional Genomics) using standard techniques. We used RMA background correction, quantile normalization and RMA summarization. A TGF-β response signature was obtained by selecting genes with limma P-value<0.05 and at least two fold up-regulation in TGF-β treated LS^{TGFBR2} cells.

TGF-β reporter assay. SW837 cells were cotransfected with the pCAGA-Luc12X Firefly luciferase TGF-β reporter[40], the Renilla luciferase plasmid pRL-TK (Promega) and human Irx3 or human Irx5 cDNAs. In the case of TBR19 GFP or IRX5, cells were induced for 24 h with doxycycline prior to the transfection. 36h after transfection recombinant TGF-β was added to cells for 16h and luciferase activity was measured using the Dual Luciferase Reporter Assay System (Promega). TGF-β pathway activation was determined by normalizing the Firefly luciferase activity with Renilla luciferase activity.

Supplementary Figure legends

Supplementary Figure S1. a, Control clones (green) stained with Mirr (red). **b**, box-plot graph of the clone area (GFP^+) per anterior midgut area. Box-plot represents data shown in Table 1. The length of the box represents the distance between the 25th and 75th percentiles, the interior horizontal line represents the group median and the whiskers extend to the group minimum and maximum values. **c**, Expression of activated Tkv* in progenitor cells (esg>Tkv*, GFP) induce the expression of the EC marker Pdm1 (red) in ISCs, marked by Dl (blue), in contrast to control clones (esg>GFP), which do not express high levels of Pdm1 in ISCs.

Supplementary Figure S2. a, Relative expression of Irx3 and Irx5 expression in mouse normal mucosa or $Apc^{-/-}$ adenomas. Notice that both Irx proteins are up-regulated in adenomas. **b**, Graphs showing that the decrease in Irx3 and Irx5 upon short hairpin transfection correlates with an increase in *SMAD3* transcription in two different human cell lines. Note that Irx3 and Irx5 short hairpins hit each other. **c**, SW837 cells show no cytostatic response in the presence of TGF-β. p15/Cdkn2b levels measured by qRT-PCR in SW837 cells in the absence or presence of TGF-β. SW837 cells were treated for 24h with TGF-β and the levels of p15/Cdkn2b were measured by qRT-PCR. **d**, Anti-phospho-SMAD2/3 Western Blot shows that, as opposed to parental cells, the expression of wild type TGFBR2 receptor in LS174T CRC cells results in phosphorylation of the intracellular mediators smads upon the addition of TGF-β. **e**, TGF-β addition results in 400 fold activation of the CAGA12X luciferase reporter gene. **f**, BrdU cell cycle analysis. In response to TGF-β LS^{TBR2} entered cell cycle arrest when compared to LS174T cells lacking the receptor (LS).

Supplementary Figure S3. Negative correlation between TGF-β targets or SMAD3 and IRX3/IRX5 expression in the cohort of adenoma samples from GSE50115.