Identification of a Nervous System Gene Expression Signature in Colon Cancer Stem Cells Reveals a Role for Neural Crest Regulators *EGR2* and *SOX2* in Tumorigenesis

Joseph L. Regan[1,2,13*], Dirk Schumacher[3,4], Stephanie Staudte[1,2], Andreas Steffen[1], Ralf Lesche[1,5], Joern Toedling[1,5], Thibaud Jourdan[1], Johannes Haybaeck[6,7], Nicole Golob-Schwarzl[7,8], Dominik Mumberg[1], David Henderson[1], Balázs Győrffy[9,10], Christian R.A. Regenbrecht[3,11,12], Ulrich Keilholz[2], Reinhold Schäfer[2,3,4], Martin Lange[1,5]

Affiliations

[1]Bayer AG, Drug Discovery, Pharmaceuticals, 13342 Berlin, Germany

[2]Charité Comprehensive Cancer Center, Charité - Universitätsmedizin Berlin, 10117 Berlin, Germany

[3]Laboratory of Molecular Tumor Pathology, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany

[4]German Cancer Consortium (DKTK), DKFZ, 69120 Heidelberg, Germany

[5]Nuvisan ICB GmbH, 13353 Berlin, Germany

[6]Department of Pathology, Neuropathology and Molecular Pathology, Medical University of Innsbruck, Austria

[7]Diagnostic & Research Center for Molecular Biomedicine, Institute of Pathology, Medical University of Graz, Austria

[8]Department of Dermatology and Venereology, Medical University of Graz, Austria

[9]Department of Bioinformatics, Semmelweis University, 1094 Budapest, Hungary

[10]TTK Cancer Biomarker Research Group, Institute of Enzymology, 1117 Budapest, Hungary

[11]CELLphenomics GmbH, 13125 Berlin, Germany

[12]Institute of Pathology, University Medical Center Göttingen, 37075 Göttingen, Germany

[13]Lead Contact

*Correspondence: joseph.regan@charite.de (J.L.R.)

SUMMARY

Recent data support a hierarchical model of colon cancer driven by a population of cancer stem cells (CSCs). Greater understanding of the mechanisms that regulate CSCs may therefore lead to more effective treatments. Serial limiting dilution xenotransplantation assays of colon cancer patient-derived tumors demonstrated ALDH^{Positive} cells to be enriched for tumorigenic self-renewing CSCs. In order to identify CSC modulators, we performed RNA-sequencing analysis of ALDH^{Positive} CSCs from a panel of colon cancer patient-derived organoids (PDOs) and xenografts (PDXs). These studies demonstrated CSCs to be enriched for embryonic and neural development gene sets. Functional analyses of genes differentially expressed in both ALDH^{Positive} PDO and PDX CSCs demonstrated the neural crest stem cell (NCSC) regulator and wound response gene *EGR2* to be required for CSC tumorigenicity and to control expression of homeobox superfamily embryonic master transcriptional regulator *HOX* genes and the embryonic and neural stem cell regulator *SOX2*. In addition, we identify *EGR2*, *HOXA2*, *HOXA4*,

HOXA5, *HOXA7*, *HOXB2*, *HOXB3* and the tumor suppressor *ATOH1* as new prognostic biomarkers in colorectal cancer.

INTRODUCTION

Colorectal cancer (CRC), the third most common cancer and fourth most common cause of cancer deaths worldwide[1], is a heterogeneous tumor driven by a subpopulation of CSCs, that may also be the source of relapse following treatment[2–5]. Elucidation of the mechanisms that regulate CSC survival and tumorgenicity may therefore lead to novel treatments and improved patient outcomes.

CSCs are undifferentiated cancer cells that share many of the attributes of stem cells, such as multipotency, self-renewal and the ability to produce daughter cells that differentiate[2,6,7]. Stem cells are controlled by core gene networks that include the embryonic master transcriptional regulator *HOX* genes[8,9] and *SOX2*[10], whose misregulation can result in aberrant stem cell function, developmental defects and cancer[11,12]. These genes are crucial for embryonic development and their expression is maintained in adult tissue stem cells, where they regulate self-renewal and differentiation[9,13–15]. *HOX* genes and *SOX2* are aberrantly expressed in several cancers, including CRC, and emerging evidence demonstrates their involvement in the transformation of tissue stem cells into CSCs[11,16–23]. Modulation of *HOX* genes and *SOX2* could therefore provide novel therapeutic strategies to block tumorigenesis and overcome therapy resistance in CRC and other CSC driven cancers.

69 During embryonic development of the neural crest, which gives rise to the peripheral
70 nervous system (PNS) and several non-neuronal cell types[24], *HOX* and *SOX* genes are
71 regulated by retinoic acid[25,26], a product of the normal tissue stem cell and CSC marker
72 aldehyde dehydrogenase (*ALDH1A1, ALDH1A2, ALDH1A3*)[8,26–29], and by the neural
73 crest stem cell (NCSC) zinc finger transcription factor and wound response gene *EGR2*
74 (*KROX20*)[30–41].
75
76 Here we carried out whole transcriptome analysis of functionally tested ALDH^{Positive} CSCs
77 from a panel of colon PDOs and PDX models and show that colon CSCs and Lgr5^{Positive}
78 intestinal stem cells (ISCs) are highly enriched for nervous system development and
79 neural crest genes. Furthermore, we demonstrate that the neural crest stem cell (NCSC)
80 gene *EGR2* is a marker of poor prognosis in CRC and modulates expression of *HOX*
81 genes and *SOX2* in CSCs to regulate tumorigenicity and differentiation.
82
83 **RESULTS**
84 **Colon cancer PDOs are heterogeneous and enriched for ALDH^{Positive} self-renewing**
85 **CSCs**
86 Colon cancer PDO models were established from freshly isolated primary tumors and
87 metastases from colon cancer patients (Table S1) by embedding in growth-factor reduced
88 Matrigel and cultivating in serum free media, as previously described[42–44].
89 Immunostaining of PDOs for the structural proteins EZRIN and EPCAM demonstrated
90 that PDOs retain the apical-basal polarity and structural adhesion of the normal intestine
91 (Figure 1A). Immunostaining of PDOs and equivalent PDX models for stem cell regulator

Wnt signaling protein BETA-CATENIN demonstrated differences in nuclear localization of BETA-CATENIN and confirmed previous data demonstrating heterogeneous Wnt signaling activity within the tumors[43] (Figure 1B). Increased aldehyde dehydrogenase (ALDH) activity, as measured using the Aldefluor™ assay, is a marker of CSCs in colon cancer and many other cancer types[29]. We previously carried out limiting dilution serial xenotransplantation of $ALDH^{Negative}$ and $ALDH^{Positive}$ cells and demonstrated that colon CSCs are $ALDH^{Positive}$ and enriched for Wnt signaling activity[43]. However, $ALDH^{Negative}$ cells also gave rise to tumors when transplanted at higher cell numbers. In order to determine if $ALDH^{Negative}$ and $ALDH^{Positive}$ cells maintained their self-renewal and tumorigenic capacity, we performed additional rounds of limiting dilution serial xenotransplantation of $ALDH^{Negative}$ and $ALDH^{Positive}$ cells (Figure 1E). These data confirmed that PDOs are enriched for $ALDH^{Positive}$ cells compared to equivalent PDX models (Figure 1C and D) and that $ALDH^{Positive}$ CSCs self-renew to maintain their tumorigenic capacity over extended rounds of xenotransplantation, but that $ALDH^{Negative}$ cells do not (Figure 1E).

Colon CSCs are enriched for embryonic and nervous system development gene expression signatures

In order to identify modulators of colon CSCs, $ALDH^{Negative}$ cells and $ALDH^{Positive}$ CSCs were isolated from PDO and PDX models and subjected to whole transcriptome analysis by RNA-sequencing. *ALDH1A1* is a marker of poor prognosis in several cancer types[27,29,45–49] and has been reported to be responsible for the aldehyde dehydrogenase activity that defines the $ALDH^{Positive}$ cell fraction in the Aldefluor™ assay[50]. However,

nineteen different isoforms of ALDH exist and several of these, including *ALDH1A2*, *ALDH1A3* and *ALDH2* have also been reported to be involved in the Aldefluor™ assay[51–53]. Here we show that *ALDH1A1* expression is enriched in ALDH^{Positive} CSCs compared to ALDH^{Negative} cells (Figure 2A and S1). GSEA of ALDH^{Positive} and ALDH^{Negative} cells isolated from PDO and PDX models demonstrated that ALDH^{Positive} CSCs are enriched for nervous system development, TNFα via NFκB signaling, epithelial mesenchymal transition (EMT), embryonic development and Wnt signaling transcripts (Figure 2B).

Differential gene expression analysis identified 218 genes upregulated in PDOs and 250 genes upregulated in PDX models compared to ALDH^{Negative} cells. Of these, 30 genes were found to be differentially expressed in both ALDH^{Positive} PDO and PDX cells (Figure 2C). Interestingly, many of these differentially expressed and common PDO-PDX genes are expressed during embryogenesis and have a role in neural crest cell (NCC) and central nervous system (CNS) development. Of these 30 common genes (Figure S2) 10, *ALDH1A1*[50], *EGR2*[31,38,41,54–60], *EGR3*[56,58,61–66], *HDGFRP3*[67–72], *OLFM2*[73–76], *OLFML3*[75,77,78], *PCP4*[79–82], *PEG10*[83–91], *PRKACB*[92,93], and *THBS1*[94–99], were selected for functional analysis based on their tissue expression and roles in development and cancer (Figure 2D, S2 and S3).

***EGR2* is required for colon CSC survival in non-adherent cell culture**

The ability of CSCs to survive and form spheroids in non-adherent cell culture is the gold standard assay for the assessment of normal stem cells and CSCs *in vitro*[100,101]. In order to test 10 of the differentially expressed genes common to ALDH^{Positive} PDO-PDX models,

cells were transfected with siRNAs against *ALDH1A1, EGR2, EGR3, HDGFRP3, OLFM2, OLFML3, PCP4, PEG10, PRKACB* and *THBS1* (Figure 3B), serially plated at limiting dilution into low-attachment plates and assessed for spheroid formation. siRNA *EGR2* caused a significant decrease in spheroid formation and proliferation in all models (Figure 3A, C and D). Immunostaining of PDO, PDX and clinical samples demonstrated *EGR2* to be ubiquitously expressed, with increased cytoplasmic and nuclear expression in cancer compared to normal mucosa (Figure S4).

shRNA EGR2 cells are less tumorigenic, more differentiated and have decreased expression of *HOX* genes and *SOX2*

Limiting dilution xenotransplantation of control virus transduced and shRNA EGR2 transduced 195-CB-P cells was carried out to determine if *EGR2* regulates tumorigenesis *in vivo*. Control virus transduced cells generated xenografts at each cell dilution tested but shRNA EGR2 transduced cells were significantly impaired in their ability to generate tumors when transplanted at low cell number (Figure 3E). In addition, shRNA EGR2 tumors grew more slowly than control transduced cells (Figure 3F). These data demonstrate that loss of *EGR2* in CSCs significantly decreased their tumorigenic capacity. Quantitative RT-PCR analysis of three shRNA EGR2 tumors confirmed that the shRNA EGR2 knockdown was present (Figure 3G). Significantly, expression of proliferation (*MKI67, MYC*), intestinal stem cell genes (*ALCAM, ALDH1A1, BMI1, EPHA4, EPHB2, LRIG1, OLFM4, PROM1*) and Wnt signaling genes (*AXIN2, CTNNB1, LGR5, RUNX2*) were decreased, while the expression of differentiation markers, including the tumor suppressor and Wnt signaling target *ATOH1*, were strongly increased (Figure 3G).

Interestingly, *ATOH1* is also essential for neuronal differentiation during embryonic development[102–109].

During embryogenesis *EGR2* has a conserved role in regulating embryonic master transcriptional regulator HOX genes and the stem cell regulator *SOX2*[30–32,34–41]. In addition, several HOX genes and *SOX2* have recently been shown to be enriched in and to regulate colon CSCs[17–20,23]. We therefore investigated whether these genes were similarly regulated by *EGR2* in colon PDX tumors. Notably, we found that *SOX2* and several *HOX* genes, namely *HOXA2*, *HOXA4*, *HOXA5*, *HOXA7*, *HOXB2*, *HOXB3* and *HOXD10*, were downregulated in shRNA EGR2 tumors (Figure 3G).

EGR2, ATOH1, HOXA2, HOXA4, HOXA5, HOXA7, HOXB2 and _HOXB3_ are predictors of patient outcome in colorectal cancer

To characterize *EGR2, ATOH1, HOXA2, HOXA4, HOXA5, HOXA7, HOXB2, HOXB3 HOXD10* and *SOX2* expression in clinical samples, we analyzed expression across different colorectal tumor stages (Figure 4A). These data demonstrated that *EGR2* (p-value 0.027), *HOXA2* (p-value 0.026), *HOXA4* (p-value 0.000075) *HOXA5* (p-value 0.001), *HOXA7* (p-value 0.009), *HOXB3* (p-value 0.0016) and HOXD10 (p-value 0.043) expression are more enhanced in late stage T4 clinical tumors. Of these, *HOXA4*, *HOXA5*, *HOXA7*, and *HOXB3* are significant at FDR < 5%. Analysis of Kaplan-Meier survival curves showed that patients with higher *EGR2, HOXA2, HOXA4, HOXA5, HOXA7, HOXB2 and HOXB3* expression have a poorer clinical outcome (p-values 0.00017, 0.0028, 0.0006, 0.0043, 0.0022, 0.00025 and 0.019, respectively). Of these,

184 higher *EGR2*, *HOXA2*, *HOXA4*, *HOXA5* and *HOXA7* are significant at FDR < 5%.
185 Furthermore, these data demonstrated that high levels of *ATOH1* are predictive of good
186 prognosis (p-value 0.0013). These data support *ATOH1*, *EGR2* and its target genes
187 *HOXA2, HOXA4, HOXA5, HOXA7 and HOXB3* as potential new biomarkers for CRC
188 prognosis.
189
190 **DISCUSSION**
191 We previously demonstrated that colon cancer PDOs are enriched for CSCs and
192 preserve the functional and molecular heterogeneity found *in vivo*, thus making them
193 excellent models for the study of CSCs[43]. However, the defined conditions of the PDO
194 culture media results in reduced cell type diversity[42]. Conversely, the *in vivo* environment
195 promotes differentiation and reduces CSCs to a minority population. Therefore, in order
196 to identify genes that regulate CSC survival and differentiation we carried out whole
197 transcriptome analyses of functionally defined $ALDH^{Negative}$ cells and $ALDH^{Positive}$ CSCs
198 from colon cancer PDO and PDX models and performed functional analyses of genes
199 differentially expressed and common to $ALDH^{Positive}$ CSCs from both models.
200
201 Interestingly, these analyses revealed transcripts associated with nervous system
202 development and NCSCs to be highly enriched in both PDO and PDX CSCs. Recent
203 studies have demonstrated that solid tumors, including CRC, contain nerve fibers that
204 promote tumor growth and metastasis, indeed, neurogenesis in CRC is an independent
205 indicator of poor clinical outcome[110,111], but their origin and mechanism of innervation is
206 unknown[112–117].

A growing body of evidence has demonstrated a gut-neural axis[118–123] in which various intestinal cells, including stem cells, interact with the autonomic nervous system (ANS), either directly[124–130] or via the enteric nervous system (ENS)[131–133], a network of neurons and glia within the bowel wall that regulates most aspects of intestinal function[134], to control stem cell proliferation and differentiation[135,136]. For example, ISCs express ANS-associated alpha2A adrenoreceptor (Adra2a) and acetylcholine (ACh) receptors implicated in controlling intestinal epithelial proliferation[130,137–140]. In addition, differentiated cell types, such as intestinal enterochromaffin (EC) cells have been found to be electrically excitable and modulate serotonin-sensitive primary afferent nerve fibers via synaptic connections, enabling them to detect and transduce environmental, metabolic, and homeostatic information from the gut directly to the nervous system[141]. Recent studies have also demonstrated that enteroendocrine cells form neuroepithelial circuits by directly synapsing with vagal neurons and called for a renaming of these cells from enteroendocrine to neuropod cells[129,142]. Neuropod cells and EC cells, like all differentiated intestinal cells (enteroendocrine, enterocyte, goblet, paneth) and CSCs, derive from multipotent Lgr5^{Positive} crypt stem cells[143,144]. Significantly, colorectal CSCs themselves have been shown to be capable of generating neurons when transplanted intraperitoneally in nude mice[145]. Intestinal stem cells and CSCs should therefore possess the capacity to express nervous system genes, since they are the progenitors of cells with neural function. However, until now, no previous study had directly reported nervous system gene enrichment in ISCs or CSCs.

We therefore carried out gene ontology analysis of Lgr5^{Positive} crypt stem cell transcriptomes from earlier studies[146–148]. In agreement with our CSC data (Figure 2), this analysis revealed normal ISCs to also be enriched for nervous system genes (Figure S5). In addition, the PDOs showed ubiquitous staining for the epithelial cell marker EPCAM (Figure 1A), demonstrating that they do not contain a separate non-epithelial neural cell lineage that could be the origin of the nervous system gene expression. Overall, these data suggest that CSCs may be a source of the neural connections that interact with the ANS and peripheral nervous system (PNS) to drive tumor growth and metastasis[112–116]. Denervation of the ANS and PNS, which causes loss of autonomic neurotransmitters in the gut, results in loss of crypt stem cell proliferation and suppression of tumorigenesis[124–128,131,132,149–151]. The inhibition of nervous system gene transcription in CSCs and their progeny may therefore provide a novel therapeutic strategy in colorectal cancer, with results similar to denervation[149,150,152].

During embryonic development, the PNS, of which the ENS is a part, arises from NCSCs, multipotent and highly migratory stem cells that move throughout the embryo to colonize multiple organ primordia and differentiate into numerous cell types[24,153–155]. Recently, self-renewing NCSCs have been discovered in post-natal tissue[156–160], including the adult gut[161,162], although the degree to which these cells contribute to the adult tissue is not yet known.

EGR2 is a conserved regulator and marker of NCSCs that acts upstream of several *HOX* genes and *SOX2* to control cell fate in embryonic and nervous system stem cells[30–41].

253 Interestingly, its expression is also rapidly activated after wounding in the embryonic and
254 adult mouse[33], suggesting a role in adult tissue stem cells, which contribute to tissue
255 regeneration and wound repair[163]. However, no previous study has identified a role for
256 *EGR2* in CRC. Here, we demonstrate that *EGR2* is enriched in colon CSCs and is
257 required for tumorigenicity and to maintain CSCs in an undifferentiated state by regulating
258 *HOX* genes and *SOX2*.
259
260 *SOX2* is one of the early genes activated in the developing neural crest and has a broad
261 role as a transcriptional regulator in embryonic and adult stem cells[15,164–169]. In embryonic
262 and adult neural stem cells, it is required for the maintenance of neural stem cell
263 properties, including proliferation, survival, self-renewal and neurogenesis[170–174]. In the
264 intestine, its expression results in cell fate conversion and redirects the intestinal
265 epithelium to a more undifferentiated phenotype[175–177]. In addition, *SOX2* has been
266 associated with a stem cell state in several cancer types[178–180] and is aberrantly
267 expressed in CRC[176,181,182]. Overall, these data, combined with our own, support a role
268 for *SOX2* in CRC tumor initiation and progression, possible by promoting neural
269 specification in CSCs and their descendants.
270
271 *HOX* genes have been reported to be enriched in and required for the maintenance of
272 normal stem cells and CSCs in various adult tissues[11,13,16,183–189]. Recently, *HOXA4,*
273 *HOXA9* and *HOXD10* were shown to be selectively expressed in ALDH^{Positive} intestinal
274 crypt stem cells and colon CSCs, to promote self-renewal and regulate expression of
275 stem cell markers[17,18]. Here, we demonstrate that the same *HOX* genes that are regulated

by *EGR2* in NCSCs are also regulated by *EGR2* in colon CSCs and that several of these, *HOXA2*, *HOXA4*, *HOXA5*, *HOXA7*, *HOXB2*, *HOXB3*, along with *EGR2*, are indicators of poor prognosis in CRC.

These data demonstrate that colon CSCs are enriched for neural crest and nervous system development genes, including the NCSC regulator *EGR2*, which controls *SOX2* and *HOX* genes to maintain CSCs in an undifferentiated state required for tumorigenesis. Targeting *EGR2* to induce differentiation and block potential intestinal-neural cell specification, e.g. by downregulating the neural stem cell regulator SOX2, may offer a novel therapeutic strategy to eliminate colon CSCs and prevent nervous system driven proliferation and metastasis.

EXPERIMENTAL PROCEDURES

Human tissue samples and establishment of patient-derived cancer organoid cell cultures

Tumor material was obtained with informed consent from CRC patients under approval from the local Institutional Review Board of Charité University Medicine (Charité Ethics Cie: Charitéplatz 1, 10117 Berlin, Germany) (EA 1/069/11) and the ethics committee of the Medical University of Graz and the ethics committee of the St John of God Hospital Graz (23-015 ex 10/11). Tumor staging was carried out by experienced and board-certified pathologists (Table S1). Cancer organoid cultures were established and propagated as described[42,44].

299

300 **Limiting dilution xenotransplantation**

301 Housing and handling of animals followed European and German Guidelines for
302 Laboratory Animal Welfare. Animal experiments were conducted in accordance with
303 animal welfare law, approved by local authorities, and in accordance with the ethical
304 guidelines of Bayer AG. PDO derived PDX models were processed to single cells and
305 sorted by FACS (BD FACS Aria II) for ALDH activity (Aldefluor assay) and DAPI to
306 exclude dead cells. Cells were then re-transplanted at limiting dilutions by injected
307 subcutaneously in PBS and Matrigel (1:1 ratio) at limiting cell dilutions into female 8 – 10-
308 week-old nude$^{-/-}$ mice.

309

310 **Histology and immunohistochemistry**

311 Tumors were fixed in 4% paraformaldehyde overnight for routine histological analysis and
312 immunohistochemistry. Immunohistochemistry was carried out via standard techniques
313 with non-phospho (Active) β-Catenin (#8814, rabbit monoclonal, Cell Signaling
314 Technology; diluted 1:200) and EGR2 (ab43020, Abcam, rabbit IgG, polyclonal, diluted
315 1:1000) antibodies. Negative controls were performed using the same protocols with
316 substitution of the primary antibody with IgG-matched controls (ab172730, rabbit IgG,
317 monoclonal [EPR25A], Abcam). Colorectal cancer tissue microarrays from the
318 OncoTrack patient cohort[44] were obtained from The Institute of Pathology, Medical
319 University Graz, Austria and analyzed using Aperio TMALab and Image software (Leica
320 Biosystems).

321

Immunofluorescence staining of PDOs

For immunofluorescence imaging, cancer organoid cultures were fixed in 4% paraformaldehyde for 30 min at room temperature and permeabilized with 0.1% Triton X-100 for 30 min and blocked in phosphate-buffered saline (PBS) with 10% bovine serum albumin (BSA). Samples were incubated with primary antibodies overnight at 4°C. Antibodies used were Non-phospho (Active) β-Catenin (#8814, rabbit monoclonal, Cell Signaling Technology; diluted 1:200), EZRIN (ab40839, rabbit monoclonal, Abcam, diluted 1:200), EPCAM (#2929, mouse monoclonal, Cell Signaling Technology, diluted 1:500) and EGR2 (ab43020, rabbit polyclonal, Abcam, diluted 1:1000). Samples were stained with a conjugated secondary antibody overnight at 4°C. F-actin was stained with Alexa Fluor® 647 Phalloidin (#A22287, Thermo Fisher; diluted 1:20) for 30 min at room temperature. Nuclei were counterstained with DAPI. Negative controls were performed using the same protocol with substitution of the primary antibody with IgG-matched controls. Cancer organoids were then transferred to microscope slides for examination using a Zeiss LSM 700 Laser Scanning Microscope.

Aldefluor Assay

Organoids and xenografts were processed to single cells and labelled using the Aldefluor Assay according to manufacturer's (Stemcell Technologies) instructions. ALDH levels were assessed by FACS on a BD LSR II analyzer.

RNA Sequencing

Cells were lysed in RLT buffer and processed for RNA using the RNeasy Mini Plus RNA extraction kit (Qiagen). Samples were processed using Illumina's TrueSeq RNA protocol and sequenced on an Illumina HiSeq 2500 machine as 2x125nt paired-end reads. The raw data in Fastq format were checked for sample quality using our internal NGS QC pipeline. Reads were mapped to the human reference genome (assembly hg19) using the STAR aligner (version 2.4.2a). Total read counts per gene were computed using the program "featureCounts" (version 1.4.6-p2) in the "subread" package, with the gene annotation taken from Gencode (version 19). The "DESeq2" Bioconductor package was used for the differential-expression analysis.

siRNA transfection

Cells were seeded in 100 µl volumes of Accell™ Delivery Media (Dharmacon™) at 1.0 x 10^5 cells per well in ultra-low attachment 96-well plates and transfected with 2 µM concentrations of Accell™ siRNAs (Table S2) and control siRNA (Accell™ non-targeting siRNA control) (Dharmacon™) by incubating for up to 72 h in Accell siRNA Delivery Media.

Viral transduction

Cells were seeded in 100 µl volumes of antibiotic free culture media at 1.0 x10^5 cells per well in ultra-low attachment 96-well plates. Control and shRNA lentiviruses were purchased from Sigma-Aldrich (Table S3). Viral particles were added at a multiplicity of infection of 1. Cells were transduced for up to 96 h or until GFP positive cells were

observed before being embedded in Matrigel for the establishment of lentiviral transduced cancer organoid cultures. Puromycin (2 µg/ml) was used to keep the cells under selection.

Limiting dilution spheroid assays

For siRNA spheroid assays, transfected live ($DAPI^{Negative}$) cells were sorted at 10 cells per well into 96-well ultra-low attachment plates. 20 days later wells containing spheroids were counted and used to calculate CSC frequency using ELDA software. Proliferation was measured using the CellTiter-Glo® Luminescent Cell Viability Assay.

Gene expression analysis

For quantitative real-time RT-PCR analysis RNA was isolated using the RNeasy Mini Plus RNA extraction kit (Qiagen). cDNA synthesis was carried out using a Sensiscript RT kit (Qiagen). RNA was transcribed into cDNA using an oligo dTn primer (Promega) per reaction. Gene expression analysis was performed using TaqMan® Gene Expression Assays (Applied Biosystems) (Table S4) on an ABI Prism 7900HT sequence detection system (Applied Biosystems). GAPDH was used as an endogenous control and results were calculated using the Δ-ΔCt method. Data were expressed as the mean fold gene expression difference in three independently isolated cell preparations over a comparator sample with 95% confidence intervals. Pairwise comparison of gene expression was performed using R[190] together with package ggplot2[191] on log2 transformed RNAseq data from 533 patients with clinical data (n=378 colon adenocarcinomas, n=155 rectal carcinomas staged T1-T4) extracted from the cBioPortal for Cancer Genomics (cbioportal.org)[192,193]. Survival curves were generated using the Kaplan-Meier Plotter

(www.kmplot.com/analysis)[194]. Gene ontology enrichment analysis was carried out using the Gene Ontology Resource (www.geneontology.org)[195,196].

Statistical analysis

GraphPad Prism 6.0 was used for data analysis and imaging. All data are presented as the means ± SD, followed by determining significant differences using the two-tailed t test. Significance of RT-PCR data was determined by inspection of error bars as described by Cumming *et al*. (2007)[197]. Limiting-dilution frequency and probability estimates were analyzed by the single-hit Poisson model and pairwise tests for differences in stem cell frequencies using the ELDA software (http://bioinf.wehi.edu.au/software/elda/index.html, Hu and Smyth, 2009)[198]. Gene set enrichment analysis was carried out using pre-ranked feature of the Broad Institute GSEA software version 2 using msigdb v5.1 gene sets[199,200]. The ranking list was derived from the fold changes (1.5fold upregulated) calculated from the differential gene expression calculation and nominal p-values. P-values <0.05 were considered as statistically significant. For the final list of significant genes, False Discovery Rate was computed using the Benjamini-Hochberg method[201].

Acknowledgements

We thank Dorothea Przybilla, and Cathrin Davies (Laboratory of Molecular Tumor Pathology, Charité Universitätsmedizin Berlin, Germany) for technical and cell culture assistance. The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement 115234 (OncoTrack), the resources of which are composed of financial contribution from the

European Union Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution. A.S., T.J., D.M. and. D.H. are employees of Bayer AG. R.L., J.T. and M.L. are employees of Nuvisan ICB GmbH. C.R.A.R. is a co-founder of CELLphenomics.

Authors Contribution

Conceptualization, J.L.R.; Methodology, J.L.R. and M.L.; Investigation, J.L.R., D.S., S.S., A.S., R.L., J.T., T.J., J.H., N.G.-S., and M.L.; Writing, J.L.R.; Visualization, J.L.R; Data Curation, A.S., J.T.; Resources, J.H., U.K., C.R.A.R. and B.G.; Supervision, J.L.R., D.M., D.H., R.S., and M.L.

Accession Numbers

Array data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-5209 and E-MTAB-8927.

References

1. Siegel, R., DeSantis, C. & Jemal, A. Colorectal cancer statistics, 2014. *CA. Cancer J. Clin.* **64**, 104–117 (2014).
2. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
3. Vermeulen, L., Sprick, M. R., Kemper, K., Stassi, G. & Medema, J. P. Cancer

435 stem cells--old concepts, new insights. *Cell Death Differ.* **15**, 947–58 (2008).
436 4. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating
437 cells. *Nature* **445**, 111–115 (2007).
438 5. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell
439 capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110
440 (2007).
441 6. Clarke, M. F. & Fuller, M. Stem cells and cancer: two faces of eve. *Cell* **124**,
442 1111–1115 (2006).
443 7. Wicha, M. S., Liu, S. & Dontu, G. Cancer stem cells: an old idea--a paradigm shift.
444 *Cancer Res* **66**, 1883–1886 (2006).
445 8. Nolte, C., Ahn, Y. & Krumlauf, R. B. T.-R. M. in B. S. Hox Genes Expression☆. in
446 (Elsevier, 2014). doi:https://doi.org/10.1016/B978-0-12-801238-3.04638-9
447 9. Gouti, M. & Gavalas, A. Hox Genes and Stem Cells. in *HOX Gene Expression*
448 111–120 (Springer New York, 2007). doi:10.1007/978-0-387-68990-6_8
449 10. Sarkar, A. & Hochedlinger, K. The Sox Family of Transcription Factors: Versatile
450 Regulators of Stem and Progenitor Cell Fate. *Cell Stem Cell* **12**, 15–30 (2013).
451 11. Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nat. Rev.*
452 *Cancer* **10**, 361–371 (2010).
453 12. Chew, L.-J. & Gallo, V. The Yin and Yang of Sox proteins: Activation and
454 repression in development and disease. *J. Neurosci. Res.* **87**, 3277–3287 (2009).
455 13. Seifert, A., Werheid, D. F., Knapp, S. M. & Tobiasch, E. Role of Hox genes in
456 stem cell differentiation. *World J. Stem Cells* **7**, 583–595 (2015).
457 14. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and

differentiation. *Development* **140**, 4129 LP – 4144 (2013).

15. Arnold, K. *et al.* Sox2+ Adult Stem and Progenitor Cells Are Important for Tissue Regeneration and Survival of Mice. *Cell Stem Cell* **9**, 317–329 (2011).

16. Bhatlekar, S., Fields, J. Z. & Boman, B. M. Role of HOX Genes in Stem Cell Differentiation and Cancer. *Stem Cells Int.* **2018**, 3569493 (2018).

17. Bhatlekar, S., Viswanathan, V., Fields, J. Z. & Boman, B. M. Overexpression of HOXA4 and HOXA9 genes promotes self-renewal and contributes to colon cancer stem cell overpopulation. *J. Cell. Physiol.* **233**, 727–735 (2018).

18. Bhatlekar, S. *et al.* Identification of a developmental gene expression signature, including HOX genes, for the normal human colonic crypt stem cell niche: overexpression of the signature parallels stem cell overpopulation during colon tumorigenesis. *Stem Cells Dev.* **23**, 167–179 (2014).

19. Takeda, K. *et al.* Sox2 is associated with cancer stem-like properties in colorectal cancer. *Sci. Rep.* **8**, 17639 (2018).

20. Novak, D. *et al.* SOX2 in development and cancer biology. *Semin. Cancer Biol.* (2019). doi:https://doi.org/10.1016/j.semcancer.2019.08.007

21. Basu-Roy, U. *et al.* Sox2 maintains self renewal of tumor-initiating cells in osteosarcomas. *Oncogene* **31**, 2270–2282 (2012).

22. Lundberg, I. V *et al.* SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. *BMC Cancer* **16**, 471 (2016).

23. Schaefer, T. & Lengerke, C. SOX2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond. *Oncogene* **39**, 278–292 (2020).

481 24. Bronner, M. E. & LeDouarin, N. M. Development and evolution of the neural crest:
482 an overview. *Dev. Biol.* **366**, 2–9 (2012).
483 25. Duester, G. Retinoic Acid Synthesis and Signaling during Early Organogenesis.
484 *Cell* **134**, 921–931 (2008).
485 26. Tremblay, R. *et al.* Retinoic acid regulates Sox2 expression during neuronal and
486 glial differentiation in mouse P19 cells. *Retinoic acid: structure, mechanisms and*
487 *roles in disease* 165–174 (2012).
488 27. Vassalli, G. Aldehyde Dehydrogenases: Not Just Markers, but Functional
489 Regulators of Stem Cells. *Stem Cells Int.* **2019**, 3904645 (2019).
490 28. Douville, J., Beaulieu, R. & Balicki, D. ALDH1 as a Functional Marker of Cancer
491 Stem and Progenitor Cells. *Stem Cells Dev.* **18**, 17–26 (2008).
492 29. Huang, E. H. *et al.* Aldehyde dehydrogenase 1 is a marker for normal and
493 malignant human colonic stem cells (SC) and tracks SC overpopulation during
494 colon tumorigenesis. *Cancer Res.* **69**, 3382–3389 (2009).
495 30. Nonchev, S. *et al.* The conserved role of Krox-20 in directing Hox gene
496 expression during vertebrate hindbrain segmentation. *Proc. Natl. Acad. Sci.* **93**,
497 9339 LP – 9345 (1996).
498 31. Ghislain, J., Desmarquet-Trin-Dinh, C., Gilardi-Hebenstreit, P., Charnay, P. &
499 Frain, M. Neural crest patterning: autoregulatory and crest-specific elements co-
500 operate for Krox20 transcriptional control. *Development* **130**, 941—953 (2003).
501 32. Vesque, C. *et al.* Hoxb-2 transcriptional activation in rhombomeres 3 and 5
502 requires an evolutionarily conserved cis-acting element in addition to the Krox-20
503 binding site. *EMBO J.* **15**, 5383–5396 (1996).

504 33. Grose, R., Harris, B. S., Cooper, L., Topilko, P. & Martin, P. Immediate early
505 genes krox-24 and krox-20 are rapidly up-regulated after wounding in the
506 embryonic and adult mouse. *Dev. Dyn.* **223**, 371–378 (2002).
507 34. Nonchev, S. *et al.* Segmental expression of Hoxa-2 in the hindbrain is directly
508 regulated by Krox-20. *Development* **122**, 543 LP – 554 (1996).
509 35. Kim, M. H., Cho, M. & Park, D. Sequence Analysis of the 5′-Flanking Region of
510 the Gene Encoding Human HOXA-7. *Somat. Cell Mol. Genet.* **24**, 371–374
511 (1998).
512 36. Manzanares, M. *et al.* Krox20 and kreisler co-operate in the transcriptional control
513 of segmental expression of Hoxb3 in the developing hindbrain. *EMBO J.* **21**, 365–
514 376 (2002).
515 37. Chavrier, P. *et al.* The segment-specific gene Krox-20 encodes a transcription
516 factor with binding sites in the promoter region of the Hox-1.4 gene. *EMBO J.* **9**,
517 1209–1218 (1990).
518 38. Topilko, P. *et al.* Krox-20 controls myelination in the peripheral nervous system.
519 *Nature* **371**, 796 (1994).
520 39. Jang, S.-W. *et al.* Locus-wide identification of Egr2/Krox20 regulatory targets in
521 myelin genes. *J. Neurochem.* **115**, 1409–1420 (2010).
522 40. Sham, M. H. *et al.* The zinc finger gene Krox20 regulates
523 HoxB2 (Hox2.8) during hindbrain segmentation. *Cell* **72**, 183–196
524 (1993).
525 41. Desmazières, A., Charnay, P. & Gilardi-Hebenstreit, P. Krox20 Controls the
526 Transcription of Its Various Targets in the Developing Hindbrain According to

Multiple Modes. *J. Biol. Chem.* **284**, 10831–10840 (2009).

42. Sato, T. *et al.* Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

43. Regan, J. L. *et al.* Non-Canonical Hedgehog Signaling Is a Positive Regulator of the WNT Pathway and Is Required for the Survival of Colon Cancer Stem Cells. *Cell Rep.* **21**, 2813–2828 (2017).

44. Schütte, M. *et al.* Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. *Nat. Commun.* **8**, (2017).

45. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of a poor clinical outcome. *Cell Stem Cell* **1**, 555–567 (2007).

46. Deng, Y. *et al.* ALDH1 is an independent prognostic factor for patients with stages II-III rectal cancer after receiving radiochemotherapy. *Br. J. Cancer* **110**, 430–434 (2014).

47. Flahaut, M. *et al.* Aldehyde dehydrogenase activity plays a Key role in the aggressive phenotype of neuroblastoma. *BMC Cancer* **16**, 781 (2016).

48. Kuroda, T. *et al.* ALDH1-high ovarian cancer stem-like cells can be isolated from serous and clear cell adenocarcinoma cells, and ALDH1 high expression is associated with poor prognosis. *PLoS One* **8**, e65158–e65158 (2013).

49. Wakamatsu, Y. *et al.* Expression of cancer stem cell markers ALDH1, CD44 and CD133 in primary tumor and lymph node metastasis of gastric cancer. *Pathol. Int.*

62, 112–119 (2012).

50. Tomita, H., Tanaka, K., Tanaka, T. & Hara, A. Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget* **7**, 11018–11032 (2016).

51. Puttini, S. *et al.* ALDH1A3 Is the Key Isoform That Contributes to Aldehyde Dehydrogenase Activity and Affects in Vitro Proliferation in Cardiac Atrial Appendage Progenitor Cells. *Front. Cardiovasc. Med.* **5**, 90 (2018).

52. Moreb, J. S. *et al.* The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. *Chem. Biol. Interact.* **195**, 52–60 (2012).

53. Marcato, P., Dean, C. A., Giacomantonio, C. A. & Lee, P. W. K. Aldehyde dehydrogenase: Its role as a cancer stem cell marker comes down to the specific isoform. *Cell Cycle* **10**, 1378–1384 (2011).

54. Chandra, A., Lan, S., Zhu, J., Siclari, V. A. & Qin, L. Epidermal Growth Factor Receptor (EGFR) Signaling Promotes Proliferation and Survival in Osteoprogenitors by Increasing Early Growth Response 2 (EGR2) Expression. *J. Biol. Chem.* **288**, 20488–20498 (2013).

55. Liu, C.-J. *et al.* Differential gene expression signature between primary and metastatic head and neck squamous cell carcinoma. *J. Pathol.* **214**, 489–497 (2007).

56. To, S. Q., Knower, K. C. & Clyne, C. D. NFκB and MAPK signalling pathways mediate TNFα-induced Early Growth Response gene transcription leading to aromatase expression. *Biochem. Biophys. Res. Commun.* **433**, 96–101 (2013).

573 57. Young, E. *et al.* EGR2 mutations define a new clinically aggressive subgroup of
574 chronic lymphocytic leukemia. *Leukemia* **31**, 1547 (2016).
575 58. Yao, T., Wang, Q., Zhang, W., Bian, A. & Zhang, J. Identification of genes
576 associated with renal cell carcinoma using gene expression profiling analysis.
577 *Oncol. Lett.* **12**, 73–78 (2016).
578 59. Dillon, R. L., Brown, S. T., Ling, C., Shioda, T. & Muller, W. J. An EGR2/CITED1
579 Transcription Factor Complex and the 14-3-3σ Tumor Suppressor Are Involved in
580 Regulating ErbB2 Expression in a Transgenic-Mouse Model of Human Breast
581 Cancer. *Mol. Cell. Biol.* **27**, 8648 LP – 8657 (2007).
582 60. O'Donovan, K. J., Tourtellotte, W. G., Millbrandt, J. & Baraban, J. M. The EGR
583 family of transcription-regulatory factors: progress at the interface of molecular
584 and systems neuroscience. *Trends Neurosci.* **22**, 167–173 (1999).
585 61. Baron, V. T., Pio, R., Jia, Z. & Mercola, D. Early Growth Response 3 regulates
586 genes of inflammation and directly activates IL6 and IL8 expression in prostate
587 cancer. *Br. J. Cancer* **112**, 755–764 (2015).
588 62. Xi, H. & Kersh, G. J. Early growth response gene 3 regulates thymocyte
589 proliferation during the transition from CD4-CD8- to CD4+CD8+. *J Immunol* **172**,
590 964–971 (2004).
591 63. Inoue, A., Omoto, Y., Yamaguchi, Y., Kiyama, R. & Hayashi, S. I. Transcription
592 factor EGR3 is involved in the estrogen-signaling pathway in breast cancer cells.
593 *J. Mol. Endocrinol.* **32**, 649–661 (2004).
594 64. Shin, H., Kwon, S., Song, H. & Lim, H. J. The Transcription Factor Egr3 Is a
595 Putative Component of the Microtubule Organizing Center in Mouse Oocytes.

PLoS One **9**, e94708 (2014).

65. Quach, D. H., Oliveira-Fernandes, M., Gruner, K. A. & Tourtellotte, W. G. A Sympathetic Neuron Autonomous Role for Egr3-Mediated Gene Regulation in Dendrite Morphogenesis and Target Tissue Innervation. *J. Neurosci.* **33**, 4570–4583 (2013).

66. Tourtellotte, W. G. & Milbrandt, J. Sensory ataxia and muscle spindle agenesis in mice lacking the transcription factor Egr3. *Nat. Genet.* **20**, 87 (1998).

67. Ortega-Paino, E., Fransson, J., Ek, S. & Borrebaeck, C. A. K. Functionally associated targets in mantle cell lymphoma as defined by DNA microarrays and RNA interference. *Blood* **111**, 1617–1624 (2008).

68. Fok, K. L. *et al.* STK31 Maintains the Undifferentiated State of Colon Cancer Cells. *Carcinogenesis* **33**, 2044–2053 (2012).

69. Yun, H. S. *et al.* Depletion of hepatoma-derived growth factor-related protein-3 induces apoptotic sensitization of radioresistant A549 cells via reactive oxygen species-dependent p53 activation. *Biochem. Biophys. Res. Commun.* **439**, 333–339 (2013).

70. Xiao, Q. *et al.* HDGF-related protein-3 is required for anchorage-independent survival and chemoresistance in hepatocellular carcinomas. *Gut* **62**, 440 LP – 451 (2013).

71. Cai, H. *et al.* HRP-3 protects the hepatoma cells from glucose deprivation-induced apoptosis. *Int. J. Clin. Exp. Pathol.* **8**, 14383–14391 (2015).

72. El-Tahir, H. M., Abouzied, M. M., Gallitzendoerfer, R., Gieselmann, V. & Franken, S. Hepatoma-derived Growth Factor-related Protein-3 Interacts with Microtubules

and Promotes Neurite Outgrowth in Mouse Cortical Neurons. *J. Biol. Chem.* **284**, 11637–11651 (2009).

73. Shi, N., Guo, X. & Chen, S.-Y. Olfactomedin 2, a novel regulator for transforming growth factor-β-induced smooth muscle differentiation of human embryonic stem cell-derived mesenchymal cells. *Mol. Biol. Cell* **25**, 4106–4114 (2014).

74. Shi, N. & Chen, S.-Y. From nerve to blood vessel: a new role of Olfm2 in smooth muscle differentiation from human embryonic stem cell-derived mesenchymal cells. *J. Biomed. Res.* **29**, 261–263 (2015).

75. Anholt, R. Olfactomedin proteins: central players in development and disease . *Frontiers in Cell and Developmental Biology* **2**, 6 (2014).

76. Lee, J.-A., Anholt, R. R. H. & Cole, G. J. Olfactomedin-2 mediates development of the anterior central nervous system and head structures in zebrafish. *Mech. Dev.* **125**, 167–181 (2008).

77. Miljkovic-Licina, M. *et al.* Targeting Olfactomedin-like 3 Inhibits Tumor Growth by Impairing Angiogenesis and Pericyte Coverage. *Mol. Cancer Ther.* **11**, 2588 LP – 2599 (2012).

78. Torres, S. *et al.* Proteome Profiling of Cancer-Associated Fibroblasts Identifies Novel Proinflammatory Signatures and Prognostic Markers for Colorectal Cancer. *Clin. Cancer Res.* **19**, 6006 LP – 6019 (2013).

79. Harashima, S., Wang, Y., Horiuchi, T., Seino, Y. & Inagaki, N. Purkinje cell protein 4 positively regulates neurite outgrowth and neurotransmitter release. *J. Neurosci. Res.* **89**, 1519–1530 (2011).

80. Yoshimura, T. *et al.* PCP4/PEP19 promotes migration, invasion and adhesion in

642 human breast cancer MCF-7 and T47D cells. *Oncotarget* **7**, 49065–49074 (2016).
643 81. Hamada, T. *et al.* Anti-apoptotic effects of PCP4/PEP19 in human breast cancer
644 cell lines: a novel oncotarget. *Oncotarget* **5**, 6076–6086 (2014).
645 82. Erhardt, J. A., Legos, J. J., Johanson, R. A., Slemmon, J. R. & Wang, X.
646 Expression of PEP-19 inhibits apoptosis in PC12 cells. *Neuroreport* **11**, (2000).
647 83. Xie, T. *et al.* PEG10 as an oncogene: expression regulatory mechanisms and role
648 in tumor progression. *Cancer Cell Int.* **18**, 112 (2018).
649 84. Li, C.-M. *et al.* PEG10 Is a c-MYC Target Gene in Cancer
650 Cells. *Cancer Res.* **66**, 665 LP – 672 (2006).
651 85. Okabe, H. *et al.* Involvement of PEG10 in Human Hepatocellular Carcinogenesis
652 through Interaction with SIAH1. *Cancer Res.* **63**, 3043 LP – 3048 (2003).
653 86. Peng, Y.-P. *et al.* PEG10 overexpression induced by E2F-1 promotes cell
654 proliferation, migration, and invasion in pancreatic cancer. *J. Exp. Clin. Cancer*
655 *Res.* **36**, 30 (2017).
656 87. Kainz, B. *et al.* Overexpression of the paternally expressed gene 10 (PEG10)
657 from the imprinted locus on chromosome 7q21 in high-risk B-cell chronic
658 lymphocytic leukemia. *Int. J. Cancer* **121**, 1984–1993 (2007).
659 88. LI, X. *et al.* PEG10 promotes human breast cancer cell proliferation, migration and
660 invasion. *Int. J. Oncol.* **48**, 1933–1942 (2016).
661 89. Tsou, A.-P. *et al.* Overexpression of a novel imprinted gene,PEG10, in human
662 hepatocellular carcinoma and in regenerating mouse livers. *J. Biomed. Sci.* **10**,
663 625–635 (2003).
664 90. Akamatsu, S. *et al.* The Placental Gene PEG10 Promotes Progression of

Neuroendocrine Prostate Cancer. *Cell Rep.* **12**, 922–936 (2015).

91. Ono, R. *et al.* Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**, 101–106 (2006).

92. Furuta, K. *et al.* Integrated analysis of whole genome exon array and array-comparative genomic hybridization in gastric and colorectal cancer cells. *Cancer Sci.* **103**, 221–227 (2011).

93. Livshits, A., Git, A., Fuks, G., Caldas, C. & Domany, E. Pathway-based personalized analysis of breast cancer expression data. *Mol. Oncol.* **9**, 1471–1483 (2015).

94. Kazerounian, S., Yee, K. O. & Lawler, J. Thrombospondins in cancer. *Cell. Mol. Life Sci.* **65**, 700–712 (2008).

95. Huang, T., Sun, L., Yuan, X. & Qiu, H. Thrombospondin-1 is a multifaceted player in tumor progression. *Oncotarget* **8**, 84546–84558 (2017).

96. Iruela-Arispe, M. L., Liska, D. J., Sage, E. H. & Bornstein, P. Differential expression of thrombospondin 1, 2, and 3 during murine development. *Dev. Dyn.* **197**, 40–56 (2018).

97. Adams, J. C. & Tucker, R. P. The thrombospondin type 1 repeat (TSR) superfamily: Diverse proteins with related roles in neuronal development. *Dev. Dyn.* **218**, 280–299 (2000).

98. Ostankova, Y. V *et al.* Expression of Thrombospondin-1 Gene mRNA and Protein in the Placenta in Gestosis. *Bull. Exp. Biol. Med.* **151**, 215–218 (2011).

99. Daubon, T. *et al.* Deciphering the complex role of thrombospondin-1 in

688 glioblastoma development. *Nat. Commun.* **10**, 1146 (2019).
689 100. Weiswald, L.-B., Bellet, D. & Dangles-Marie, V. Spherical Cancer Models in
690 Tumor Biology. *Neoplasia* **17**, 1–15 (2015).
691 101. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating
692 cells. *Nature* **445**, 111–5 (2007).
693 102. Maricich, S. M. *et al.* Atoh1-lineal neurons are required for hearing and for the
694 survival of neurons in the spiral ganglion and brainstem accessory auditory nuclei.
695 *J. Neurosci.* **29**, 11123–11133 (2009).
696 103. Aragaki, M. *et al.* Proteasomal degradation of Atoh1 by aberrant Wnt signaling
697 maintains the undifferentiated state of colon cancer. *Biochem Biophys Res*
698 *Commun* **368**, 923–929 (2008).
699 104. Yang, Q., Bermingham, N. A., Finegold, M. J. & Zoghbi, H. Y. Requirement of
700 Math1 for Secretory Cell Lineage Commitment in the Mouse Intestine. *Science*
701 *(80-.).* **294**, 2155 LP – 2158 (2001).
702 105. Ben-Arie, N. *et al.* Evolutionary Conservation of Sequence and Expression of the
703 bHLH Protein Atonal Suggests a Conserved Role in Neurogenesis. *Hum. Mol.*
704 *Genet.* **5**, 1207–1216 (1996).
705 106. Bossuyt, W. *et al.* Atonal homolog 1 is a tumor suppressor gene. *PLoS Biol.* **7**,
706 e39–e39 (2009).
707 107. Ishibashi, M. Molecular mechanisms for morphogenesis of the central nervous
708 system in mammals. *Anat Sci Int* **79**, 226–234 (2004).
709 108. Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S. & Kageyama, R. A
710 Mammalian Helix-Loop-Helix Factor Structurally Related to the Product of

Drosophila Proneural Gene atonal Is a Positive Transcriptional Regulator Expressed in the Developing Nervous System. *J. Biol. Chem.* **270**, 8730–8738 (1995).

109. VanDussen, K. L. & Samuelson, L. C. Mouse atonal homolog 1 directs intestinal progenitors to secretory cell rather than absorptive cell fate. *Dev. Biol.* **346**, 215–223 (2010).

110. Albo, D. *et al.* Neurogenesis in colorectal cancer is a marker of aggressive tumor behavior and poor outcomes. *Cancer* **117**, 4834–4845 (2011).

111. Liebl, F. *et al.* The Severity of Neural Invasion Is Associated with Shortened Survival in Colon Cancer. *Clin. Cancer Res.* **19**, 50 LP – 61 (2013).

112. Amit, M. *et al.* Loss of p53 drives neuron reprogramming in head and neck cancer. *Nature* **578**, 449–454 (2020).

113. Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573**, 532–538 (2019).

114. Magnon, C. *et al.* Autonomic Nerve Development Contributes to Prostate Cancer Progression. *Science (80-.).* **341**, 1236361 (2013).

115. Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81 (2014).

116. Rademakers, G. *et al.* The role of enteric neurons in the development and progression of colorectal cancer. *Biochim. Biophys. Acta - Rev. Cancer* **1868**, 420–434 (2017).

117. Monje, M. *et al.* Roadmap for the Emerging Field of Cancer Neuroscience. *Cell* **181**, 219–222 (2020).

734 118. Han, W. *et al.* A Neural Circuit for Gut-Induced Reward. *Cell* **175**, 665-678.e23
735 (2018).
736 119. Bohórquez, D. V & Liddle, R. A. The gut connectome: making sense of what you
737 eat. *J. Clin. Invest.* **125**, 888–890 (2015).
738 120. Clemmensen, C. *et al.* Gut-Brain Cross-Talk in Metabolic Control. *Cell* **168**, 758–
739 774 (2017).
740 121. de Araujo, I. E., Ferreira, J. G., Tellez, L. A., Ren, X. & Yeckel, C. W. The gut-
741 brain dopamine axis: a regulatory system for caloric intake. *Physiol. Behav.* **106**,
742 394–399 (2012).
743 122. Mayer, E. A. Gut feelings: the emerging biology of gut-brain communication. *Nat.*
744 *Rev. Neurosci.* **12**, 453–466 (2011).
745 123. Sharon, G., Sampson, T. R., Geschwind, D. H. & Mazmanian, S. K. The Central
746 Nervous System and the Gut Microbiome. *Cell* **167**, 915–932 (2016).
747 124. Callaghan, B. D. The effect of pinealectomy and autonomic denervation on crypt
748 cell proliferation in the rat small intestine. *J. Pineal Res.* **10**, 180–185 (1991).
749 125. Kennedy, M. F. G., Tutton, P. J. M. & Barkla, D. H. Adrenergic factors involved in
750 the control of crypt cell proliferation in jejunum and descending colon of mouse.
751 *Clin. Exp. Pharmacol. Physiol.* **10**, 577–586 (1983).
752 126. Lachat, J.-J. & Goncalves, R. Influence of autonomic denervation upon the
753 kinetics of the ileal epithelium of the rat. *Cell Tissue Res.* **192**, 285–297 (1978).
754 127. Musso, F., Lachat, J.-J., Cruz, A. R. & Gon□alves, R. P. Effect of denervation on
755 the mitotic index of the intestinal epithelium of the rat. *Cell Tissue Res.* **163**, 395–
756 402 (1975).

128. Tutton, P. J. M. & Helme, R. D. The influence of adrenoreceptor activity on cell proliferationin the rat jejunum. *Cell Prolif.* **7**, 125–136 (1974).

129. Bohórquez, D. V *et al.* Neuroepithelial circuit formed by innervation of sensory enteroendocrine cells. *J. Clin. Invest.* **125**, 782–786 (2015).

130. Davis, E. A., Zhou, W. & Dailey, M. J. Evidence for a direct effect of the autonomic nervous system on intestinal epithelial stem cell proliferation. *Physiol. Rep.* **6**, e13745–e13745 (2018).

131. Hernandes, L., Zucoloto, S. & Parisi Alvares, E. Effect of myenteric denervation on intestinal epithelium proliferation and migration of suckling and weanling rats. *Cell Prolif.* **33**, 127–138 (2000).

132. Zucoloto, S. *et al.* The relationship between myenteric neuronal denervation, smooth muscle thickening and epithelial cell proliferation in the rat colon. *Res. Exp. Med.* **197**, 117–124 (1997).

133. Puzan, M., Hosic, S., Ghio, C. & Koppes, A. Enteric Nervous System Regulation of Intestinal Stem Cell Differentiation and Epithelial Monolayer Function. *Sci. Rep.* **8**, 6313 (2018).

134. Sasselli, V., Pachnis, V. & Burns, A. J. The enteric nervous system. *Dev. Biol.* **366**, 64–73 (2012).

135. Davis, E. A. & Dailey, M. J. A direct effect of the autonomic nervous system on somatic stem cell proliferation? *Am. J. Physiol. Integr. Comp. Physiol.* **316**, R1–R5 (2018).

136. Lundgren, O., Jodal, M., Jansson, M., Ryberg, A. T. & Svensson, L. Intestinal epithelial stem/progenitor cells are controlled by mucosal afferent nerves. *PLoS*

One **6**, e16295–e16295 (2011).

137. Greig, C. J. & Cowles, R. A. Muscarinic acetylcholine receptors participate in small intestinal mucosal homeostasis. *J. Pediatr. Surg.* **52**, 1031–1034 (2017).

138. Schaak, S. *et al.* Alpha(2) adrenoceptors regulate proliferation of human intestinal epithelial cells. *Gut* **47**, 242–250 (2000).

139. Valet, P. *et al.* Characterization and distribution of alpha 2-adrenergic receptors in the human intestinal mucosa. *J. Clin. Invest.* **91**, 2049–2057 (1993).

140. Takahashi, T. *et al.* Non-neuronal acetylcholine as an endogenous regulator of proliferation and differentiation of Lgr5-positive stem cells in mice. *FEBS J.* **281**, 4672–4690 (2014).

141. Bellono, N. W. *et al.* Enterochromaffin Cells Are Gut Chemosensors that Couple to Sensory Neural Pathways. *Cell* **170**, 185-198.e16 (2017).

142. Kaelberer, M. M. *et al.* A gut-brain neural circuit for nutrient sensory transduction. *Science (80-.).* **361**, eaat5236 (2018).

143. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).

144. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).

145. Lu, R. *et al.* Neurons generated from carcinoma stem cells support cancer progression. *Signal Transduct. Target. Ther.* **2**, 16036 (2017).

146. Merlos-Suárez, A. *et al.* The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse. *Cell Stem Cell* **8**, 511–524 (2011).

803 147. Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of
804 proposed quiescent '+4' cell markers. *EMBO J.* **31**, 3079–3091 (2012).
805 148. Yan, K. S. *et al.* Intestinal Enteroendocrine Lineage Cells Possess Homeostatic
806 and Injury-Inducible Stem Cell Activity. *Cell Stem Cell* **21**, 78-90.e6 (2017).
807 149. Zhao, C.-M. *et al.* Denervation suppresses gastric tumorigenesis. *Sci. Transl.*
808 *Med.* **6**, 250ra115-250ra115 (2014).
809 150. Saloman, J. L., Albers, K. M., Rhim, A. D. & Davis, B. M. Can Stopping Nerves,
810 Stop Cancer? *Trends Neurosci.* **39**, 880–889 (2016).
811 151. Saloman, J. L. *et al.* Ablation of sensory neurons in a genetic model of pancreatic
812 ductal adenocarcinoma slows initiation and progression of cancer. *Proc. Natl.*
813 *Acad. Sci. U. S. A.* **113**, 3078–3083 (2016).
814 152. Rabben, H.-L., Zhao, C.-M., Hayakawa, Y., Wang, T. C. & Chen, D. Vagotomy
815 and Gastric Tumorigenesis. *Curr. Neuropharmacol.* **14**, 967–972 (2016).
816 153. Nagy, N. & Goldstein, A. M. Enteric nervous system development: A crest cell's
817 journey from neural tube to colon. *Semin. Cell Dev. Biol.* **66**, 94–106 (2017).
818 154. Le Douarin, N. M. & Teillet, M.-A. The migration of neural crest cells to the wall of
819 the digestive tract in avian embryo. *J. Embryol. Exp. Morphol.* **30**, 31 LP – 48
820 (1973).
821 155. Yntema, C. L. & Hammond, W. S. The origin of intrinsic ganglia of trunk viscera
822 from vagal neural crest in the chick embryo. *J. Comp. Neurol.* **101**, 515–541
823 (1954).
824 156. Morrison, S. J., White, P. M., Zock, C. & Anderson, D. J. Prospective
825 identification, isolation by flow cytometry, and in vivo self-renewal of multipotent

826 mammalian neural crest stem cells. *Cell* **96**, 737–49 (1999).
827 157. Jinno, H. *et al.* Convergent genesis of an adult neural crest-like dermal stem cell
828 from distinct developmental origins. *Stem Cells* **28**, 2027–2040 (2010).
829 158. Adameyko, I. *et al.* Schwann Cell Precursors from Nerve Innervation Are a
830 Cellular Origin of Melanocytes in Skin. *Cell* **139**, 366–379 (2009).
831 159. Dyachuk, V. *et al.* Parasympathetic neurons originate from nerve-associated
832 peripheral glial progenitors. *Science (80-.).* **345**, 82 LP – 87 (2014).
833 160. Espinosa-Medina, I. *et al.* Parasympathetic ganglia derive from Schwann cell
834 precursors. *Science (80-.).* **345**, 87 LP – 90 (2014).
835 161. Uesaka, T., Nagashimada, M. & Enomoto, H. Neuronal Differentiation in Schwann
836 Cell Lineage Underlies Postnatal Neurogenesis in the Enteric Nervous System. *J.*
837 *Neurosci.* **35**, 9879–9888 (2015).
838 162. Kruger, G. M. *et al.* Neural crest stem cells persist in the adult gut but undergo
839 changes in self-renewal, neuronal subtype potential, and factor responsiveness.
840 *Neuron* **35**, 657–69 (2002).
841 163. Ge, Y. *et al.* Stem Cell Lineage Infidelity Drives Wound Repair and Cancer. *Cell*
842 **169**, 636-650.e14 (2017).
843 164. Rogers, C. D., Harafuji, N., Archer, T., Cunningham, D. D. & Casey, E. S.
844 Xenopus Sox3 activates sox2 and geminin and indirectly represses Xvent2
845 expression to induce neural progenitor formation at the expense of non-neural
846 ectodermal derivatives. *Mech. Dev.* **126**, 42–55 (2009).
847 165. Wakamatsu, Y., Endo, Y., Osumi, N. & Weston, J. A. Multiple roles of Sox2, an
848 HMG-box transcription factor in avian neural crest development. *Dev. Dyn.* **229**,

849 74–86 (2004).
850 166. Aquino, J. B. *et al.* In vitro and in vivo differentiation of boundary cap neural crest
851 stem cells into mature Schwann cells. *Exp. Neurol.* **198**, 438–449 (2006).
852 167. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend
853 on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
854 168. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4
855 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007).
856 169. Favaro, R. *et al.* Hippocampal development and neural stem cell maintenance
857 require Sox2-dependent regulation of Shh. *Nat. Neurosci.* **12**, 1248–1256 (2009).
858 170. Pevny, L. H. & Nicolis, S. K. Sox2 roles in neural stem cells. *Int. J. Biochem. Cell*
859 *Biol.* **42**, 421–424 (2010).
860 171. Suh, H. *et al.* In Vivo Fate Analysis Reveals the Multipotent and Self-Renewal
861 Capacities of Sox2+ Neural Stem Cells in the Adult Hippocampus. *Cell Stem Cell*
862 **1**, 515–528 (2007).
863 172. Feng, R. *et al.* Sox2 protects neural stem cells from apoptosis via up-regulating
864 survivin expression. *Biochem. J.* **450**, 459–468 (2013).
865 173. Thiel, G. How Sox2 maintains neural stem cell identity. *Biochem. J.* **450**, e1–e2
866 (2013).
867 174. Ellis, P. *et al.* SOX2, a Persistent Marker for Multipotential Neural Stem Cells
868 Derived from Embryonic Stem Cells, the Embryo or the Adult. *Dev. Neurosci.* **26**,
869 148–165 (2004).
870 175. Hagey, D. W. *et al.* SOX2 regulates common and specific stem cell features in the
871 CNS and endoderm derived organs. *PLOS Genet.* **14**, e1007224 (2018).

176. Raghoebir, L. *et al.* SOX2 redirects the developmental fate of the intestinal epithelium toward a premature gastric phenotype. *J. Mol. Cell Biol.* **4**, 377–385 (2012).

177. Kuzmichev, A. N. *et al.* Sox2 Acts through Sox21 to Regulate Transcription in Pluripotent and Differentiated Cells. *Curr. Biol.* **22**, 1705–1710 (2012).

178. Bareiss, P. M. *et al.* SOX2 Expression Associates with Stem Cell State in Human Ovarian Carcinoma. *Cancer Res.* **73**, 5544 LP – 5555 (2013).

179. Herreros-Villanueva, M. *et al.* SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis* **2**, e61–e61 (2013).

180. Lee, S. H. *et al.* SOX2 regulates self-renewal and tumorigenicity of stem-like cells of head and neck squamous cell carcinoma. *Br. J. Cancer* **111**, 2122–2130 (2014).

181. Fang, X. *et al.* ChIP-seq and Functional Analysis of the SOX2 Gene in Colorectal Cancers. *Omi. A J. Integr. Biol.* **14**, 369–384 (2010).

182. Park, E. T. *et al.* Aberrant expression of SOX2 upregulates MUC5AC gastric foveolar mucin in mucinous cancers of the colorectum and related lesions. *Int. J. Cancer* **122**, 1253–1260 (2008).

183. Liedtke, S. *et al.* The HOX Code as a “biological fingerprint” to distinguish functionally distinct stem cell populations derived from cord blood. *Stem Cell Res.* **5**, 40–50 (2010).

184. Leucht, P. *et al.* Embryonic origin and Hox status determine progenitor cell fate

during adult bone regeneration. *Development* **135**, 2845 LP – 2854 (2008).

185. Hassan, M. Q. *et al.* HOXA10 Controls Osteoblastogenesis by Directly Activating Bone Regulatory and Phenotypic Genes. *Mol. Cell. Biol.* **27**, 3337 LP – 3352 (2007).

186. Eoh, K. J. *et al.* Upregulation of homeobox gene is correlated with poor survival outcomes in cervical cancer. *Oncotarget* **8**, (2017).

187. Ben Khadra, Y., Said, K., Thorndyke, M. & Martinez, P. Homeobox Genes Expressed During Echinoderm Arm Regeneration. *Biochem. Genet.* **52**, 166–180 (2014).

188. Shah, M. *et al.* HOXC8 regulates self-renewal, differentiation and transformation of breast cancer stem cells. *Mol. Cancer* **16**, 38 (2017).

189. Tabuse, M. *et al.* Functional analysis of HOXD9 in human gliomas and glioma cancer stem cells. *Mol. Cancer* **10**, 60 (2011).

190. R Core Team. A Language and Environment for Statistical Computing. (2020).

191. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. *Springer-Verlag New York* (Springer-Verlag New York, 2016).

192. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11 (2018).

193. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).

194. Nagy, Á., Lánczky, A., Menyhárt, O. & Győrffy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* **8**, 9227 (2018).

195. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

196. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2018).

197. Cumming, G., Fidler, F. & Vaux, D. L. Error bars in experimental biology. *J Cell Biol* **177**, 7–11 (2007).

198. Hu, Y. & Smyth, G. K. ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).

199. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

200. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).

201. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat. Soc B* **57**, 289–300 (1995).

FIGURE LEGENDS

Figure 1. Colon cancer PDOs are heterogeneous and enriched for self-renewing ALDH^{Positive} CSCs

(A) Immunofluorescence staining of colon cancer PDOs for EZRIN (green) and EPCAM (red). Nuclei are stained blue with DAPI (Bars = 20 μm). (B) Immunofluorescence staining

941 of a PDO for BETA-CATENIN (green) and F-ACTIN (red) (left hand side) and
942 immunostaining of a PDX model for BETA-CATENIN (right hand side) (Bars = 20 μm).
943 (C) Representative Aldefluor Assay FACS plots of cells derived from PDO model 195-
944 CB-P (data from 10 independent experiments). DEAB (diethylaminobenzaldehyde) is a
945 specific inhibitor of ALDH and is used to control for background fluorescence. (D)
946 Frequency (±SD) of $ALDH^{Positive}$ cells in PDOs and corresponding PDX models (data from
947 10 independent experiments). (E) Tables show results of two rounds of limiting dilution
948 serial xenotransplantation of $ALDH^{Positive}$ and $ALDH^{Negative}$ cells from previously
949 established PDO derived xenograft models. The number of successfully established
950 tumors as a fraction of the number of animals transplanted is given. P-values for pairwise
951 tests of differences in CSC frequencies between $ALDH^{Positive}$ versus $ALDH^{Negative}$ cells in
952 151-ML-M, 278-ML-P, 302-CB-M and 195-CB-P in serial transplant round one tumors are
953 1.12×10^{-4}, 1.37×10^{-1}, 8.39×10^{-14} and 2.92×10^{-17} respectively and in 278-ML-P, 302-
954 CB-M and 195-CB-P serial transplant round two tumors are 3.82×10^{-7}. 3.67×10^{-22} and
955 3.78×10^{-15}, respectively. (See also Figure S1 and Table S1).
956
957 **Figure 2. PDO and PDX $ALDH^{Positive}$ CSCs are enriched for nervous system**
958 **development gene sets and neural crest stem cell genes**
959 (A) RNA sequencing generated FPKM values for ALDH1A1 (n = 3 separate cell
960 preparations). (B) Gene set enrichment analysis for nervous system development
961 (nominal p-values = <0.0005), TNFα signaling via NFkB (nominal p-value = <0.0005),
962 epithelial to mesenchymal transition (nominal p-values = <0.0005 and 0.002), embryonic
963 development (nominal p-value = <0.0005), and Wnt β-Catenin signaling (nominal p-

values= <0.0005) in ALDH^{Positive} cells (compared to ALDH^{Negative} cells) from PDO models (top panels) and PDX models (bottom panels). (C) Venn diagram shows the number of RNA-sequencing generated transcripts upregulated in PDO ALDH^{Positive} cells (218 genes) and PDX ALDH^{Positive} cells (250 genes) and upregulated in both PDO ALDH^{Positive} cells and PDX ALDH^{Positive} cells (30 genes) n = 4 separate cell preparations, basemean greater than or equal to 100, log2 fold change = 1.5 fold upregulated, p-value <0.05). (D) Table shows 10 genes upregulated in both PDO ALDH^{Positive} cells and PDX ALDH^{Positive} cells selected for functional analysis by RNA-interference (relevant literature is cited in brackets below gene names). (See also Figure S2 and S3).

Figure 3. *EGR2* is required for CSC tumorigenicity and differentiation and regulates expression of NCSC *HOX* genes and *SOX2*

(A) Proliferation of siRNA transfected patient-derived colon cancer cells in non-adherent cell culture compared to control cells (mean ± SD; data from three independent experiments). *p-value < 0.05; ***p-value < 0.001 (t test). (B) Fold expression of *ALDH1A1, EGR2, EGR3, HDGFRP3 OLFML2, PCP4, PEG10, PRKACB* and *THBS1* RT-PCR gene expression data (±95% confidence intervals) in siRNA transfected 278-ML-P cells (n=3 independent cell preparations) over the comparator population (control siRNA transfected 278-ML-P cells) (see also Table S2 and S4). (C) Frequency of siRNA EGR2 spheroid formation in non-adherent cell culture compared to control transfected cells (mean ± SD; data from three independent experiments). ns = not significant; *p-value < 0.05; **p-value < 0.01 (t test). (D) Representative images of a 278-ML-P control spheroid (LHS) and a siRNA EGR2 spheroid (RHS) in non-adherent cell culture (Bars = 100 μm).

(E) Table shows results of limiting dilution transplantation of control virus transduced and shRNA EGR2 transduced 195-CB-P cells. The number of established tumors as a fraction of the number of animals transplanted is given. P-values for pairwise tests of differences in CSC frequencies between control virus versus shRNA EGR2 1, shRNA EGR2 2 and shRNA EGR2 3 195-CB-P cells are 6.9 $x10^{-9}$, 4.9 x 10^{-6} and 6.92 x 10^{-8}, respectively. (F) Growth curves for xenografts derived from control virus transduced cells and shRNA EGR2 transduced cells. (G) Fold expression of *EGR2,* proliferation, differentiation, stem cell genes, Wnt signaling and EGR2 NCSC target genes RT-PCR gene expression data (±95% confidence intervals) in four separate 195-CB-P shRNA EGR2 tumors over the comparator population (four control virus transduced 195-CB-P xenografts). Significant differences are as follows: $*p < 0.05$, $**p < 0.01$. (see also Table S3).

Figure 4. *EGR2, HOXA2, HOXA4, HOXA5, HOXA7* and *HOXB3* are increased in late stage tumors and are indicators of poor prognosis in clinical samples

(A) Expression *of EGR2, ATOH1, HOXA2, HOXA4, HOXA5, HOXA7, HOXB2, HOXB3, HOXD10* and *SOX2* in colorectal cancer patients across different tumor stages (T1 v T4, p-value = 0.027, 0.53, 0.026, 0.000075, 0.001, 0.009, 0.075, 0.0016, 0.043 and 0.1, respectively). Of these, *HOXA4*, *HOXA5*, *HOXA7*, and *HOXB3* are significant at FDR < 5%. RNAseq and clinical data of 533 patients (n=378 colon adenocarcinoma, n=155 rectal adenocarcinoma) was extracted from cBioPortal. (B) Kaplan-Meier survival curves for *EGR2, ATOH1, HOXA2, HOXA4, HOXA5, HOXA7, HOXB2, HOXB3, HOXD10* and *SOX2* in colorectal cancer patients comparing lower third percentile to upper third percentile (logrank p-values = 0.00017, 0.0013, 0.0028, 0.0006, 0.0043, 0.0022, 0.00025,

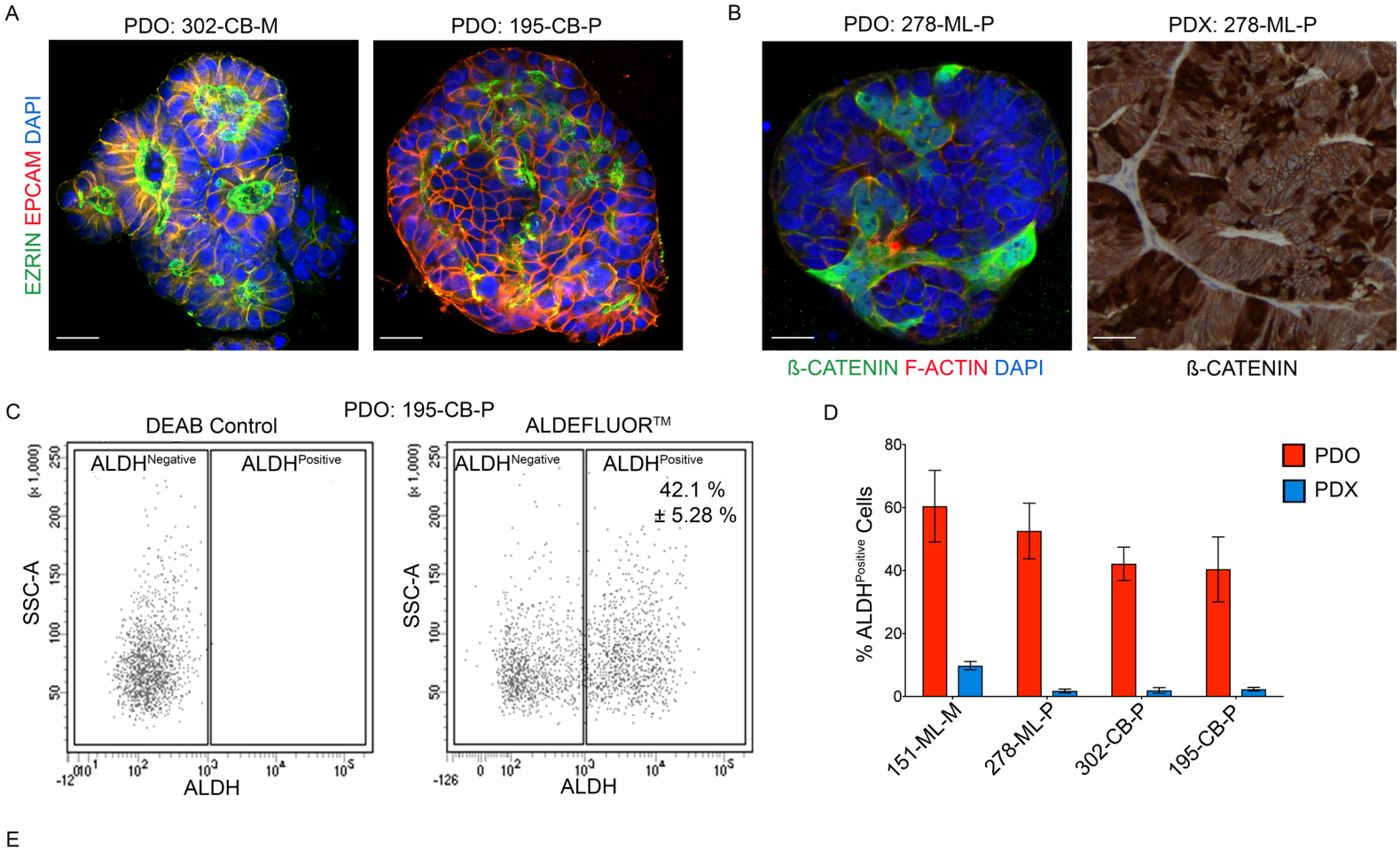
1010 0.019, 0.11 and 0.21, respectively. Of these, higher *EGR2*, *HOXA2*, *HOXA4*, *HOXA5* and

1011 *HOXA7* are significant at FDR < 5%. Results based upon data generated by the Kaplan-

1012 Meier Plotter (www.kmplot.com/analysis)[194].

1013

Figure 1



E

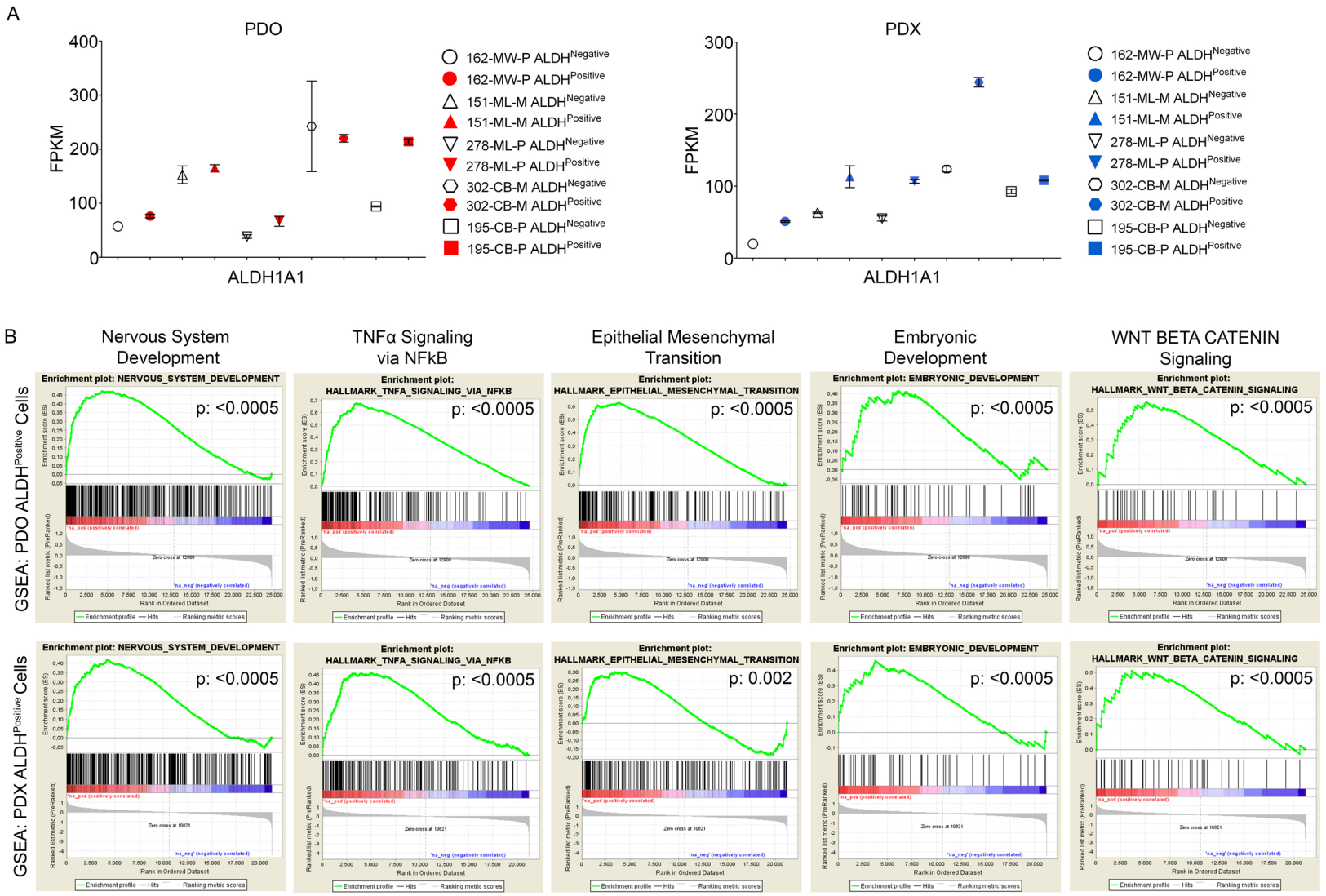
Serial PDX Transplant 1

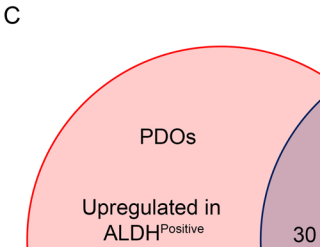
Cell dilution	151-ML-M		278-ML-P		302-CB-M		195-CB-P	
	ALDH^Negative	ALDH^Positive	ALDH^Negative	ALDH^Positive	ALDH^Negative	ALDH^Positive	ALDH^Negative	ALDH^Positive
100 cells	0/0	2/6	0/0	1/6	0/0	5/6	0/0	5/6
1000 cells	0/6	1/6	1/6	0/6	1/6	6/6	0/6	6/6
10000 cells	0/6	0/0	0/6	0/0	3/6	0/0	1/6	0/0
Frequency of Cancer Stem Cells	0	1 in 1,986	1 in 65,499	1 in 6,550	1 in 12,113	1 in 55.8	1 in 60,863	1 in 55.8
(95% confidence limits)	NA	(1 in 591 - 1 in 6,667)	(1 in 8,664 - 1 in 495,152)	(1 in 866 - 1 in 49,515)	(1 in 4,350 - 1 in 33,733)	(1 in 20.6 - 1 in 151)	(1 in 8,614 - 1 in 43,0042)	(1 in 20.6 - 1 in 151)

Serial PDX Transplant 2

Cell dilution	278-ML-P		302-CB-M		195-CB-P	
	ALDH^Negative	ALDH^Positive	ALDH^Negative	ALDH^Positive	ALDH^Negative	ALDH^Positive
100 cells	0/0	3/6	0/0	6/6	0/0	5/6
1000 cells	1/6	3/6	0/6	6/6	0/6	5/6
10000 cells	0/6	0/0	0/6	0/0	0/6	0/0
Frequency of Cancer Stem Cells	1 in 65,499	1 in 772	0	1 in 1	0	1 in 247
(95% confidence limits)	(1 in 8,664 - 1 in 495,152)	(1 in 307 - 1 in 1,940)	NA	(1 in 1 - 1 in 107)	NA	(1 in 90 - 1 in 686)

Figure 2

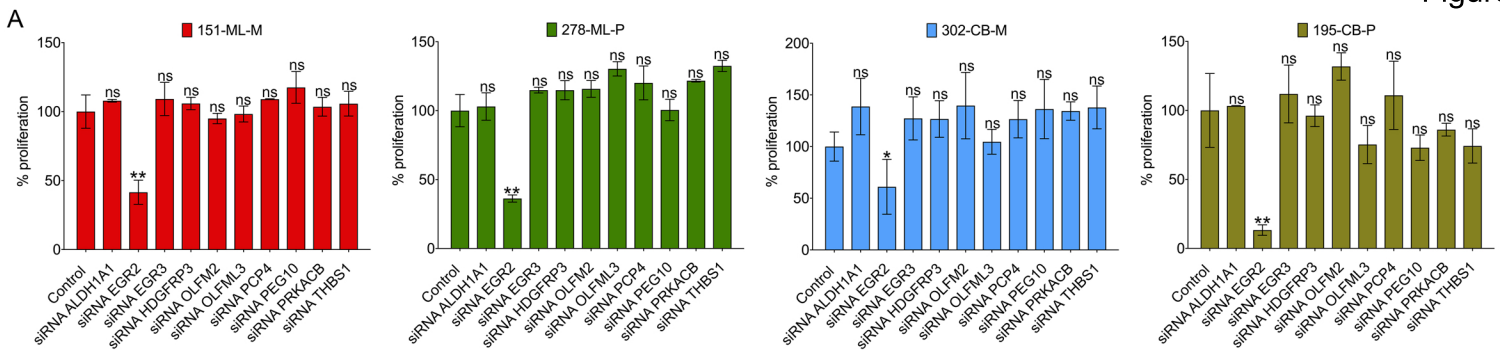


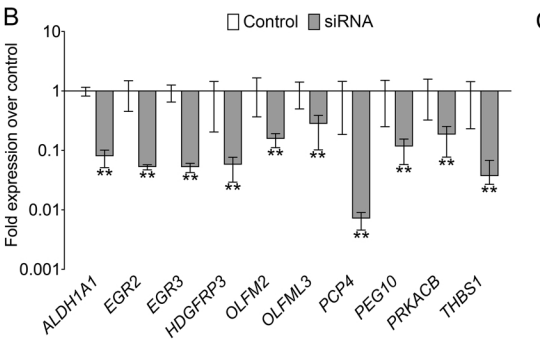


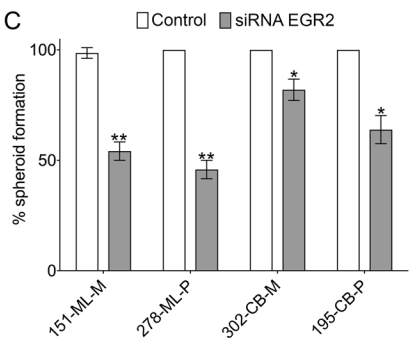
D

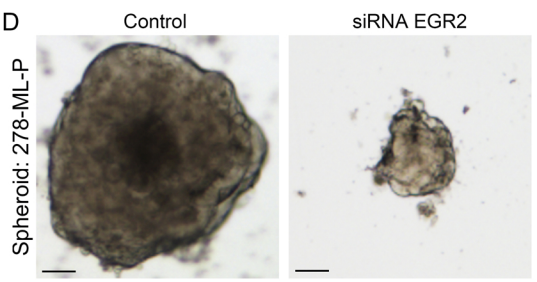
Gene Name	Expression	Development	Cancer
ALDH1A1 (52)	Liver, Kidney, Stomach, Colon, Lung, Breast, Nervous system, Pancreas, Testis, Prostate, Ovary	Embryogenesis, Hematopoiesis, CNS, NCC	Liver, Lung, Colorectal, Breast, Stomach, Ovarian
EGR2 (31,38,41,56–62)	Thyroid, Testis, Uterine, Prostate, Nervous system	Embryogenesis, CNS, NCC	Breast, Leukemia, Gastric, Uterine, Glioblastoma, Prostate
EGR3 (58,60,63–68)	Nervous system, Adrenal gland, Gall bladder	Embryogenesis, CNS, NCC	Renal cell, Colorectal, Breast, Prostate
HDGFRP3 (69–74)	Testis, Nervous system, Endometrium, Adrenal gland, Colon	Embryogenesis, CNS, Muscle	Lung, Hepatocellular
OLFM2 (75–78)	Nervous system, Skin, Smooth muscle	Embryogenesis, CNS, NCC	Unknown
OLFML3 (77,79,80)	Uterine, Endometrium, Placental, Smooth muscle	Embryogenesis, CNS, NCC	Numerous
PCP4 (81–84)	Nervous system	CNS, NCC	Breast
PEG10 (85–93)	Placenta, Nervous system, Testis, Ovary, Adrenal gland	Placenta	Hepatocellular, Pancreatic, Breast, Prostate, Leukemia, Colorectal
PRKACB (94,95)	Nervous system, Colon, Adrenal gland	CNS, NCC	Prostate, Colorectal, Gastric, Adrenal
THBS1 (96–101)	Appendix, Gall bladder, Placenta, Endometrium, Smooth muscle	Embryogenesis, CNS	Numerous

Figure 3





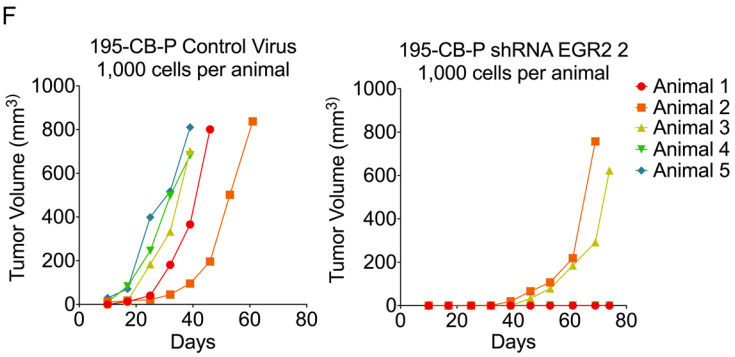


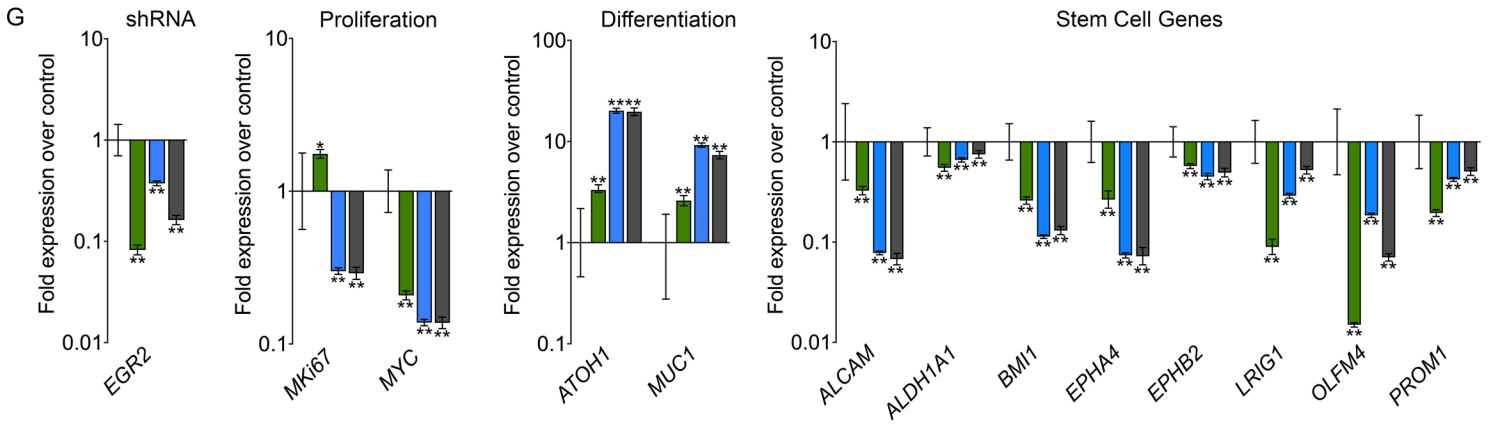


E

195-CB-P

Cell Dilution	Control Virus	shRNA EGR2 1	shRNA EGR2 2	shRNA EGR2 3
100 cells	3/5	0/5	2/5	1/5
1,000 cells	5/5	1/5	2/5	2/5
10,000 cells	5/5	3/5	4/5	3/5
Frequency of Tumor Initiating Cells	1 in 106	1 in 9,341	1 in 3,150	1 in 4,947
(95% confidence limits)	(1 in 34.1 - 1 in 348)	(1 in 3,301 - 1 in 26,436)	(1 in 1,137 - 1 in 8,722)	(1 in 1,886 - 1 in 12,974)





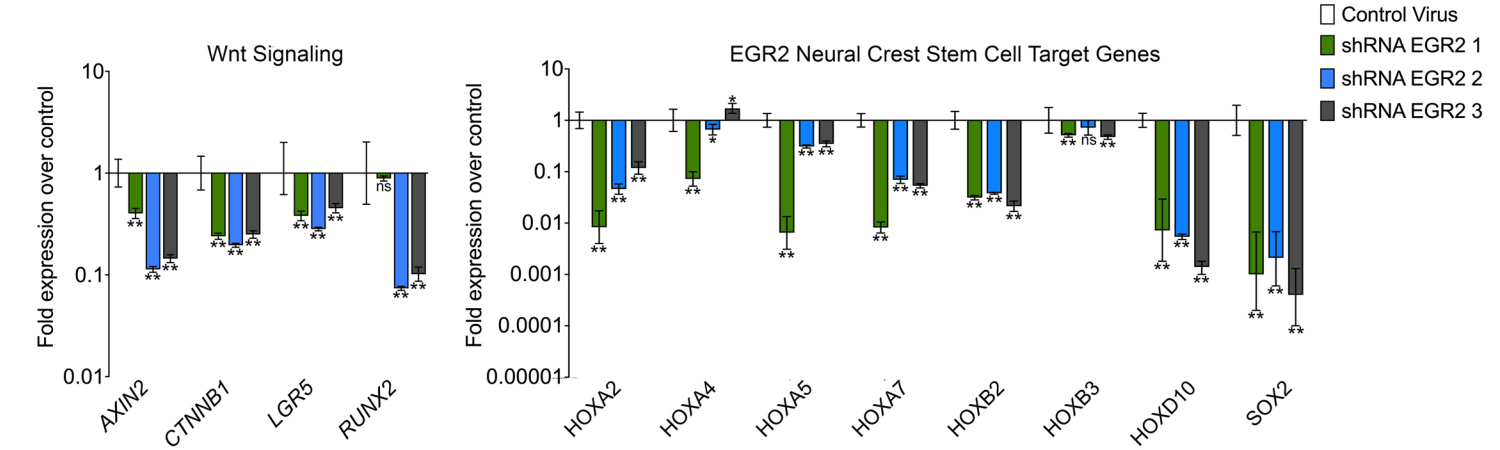
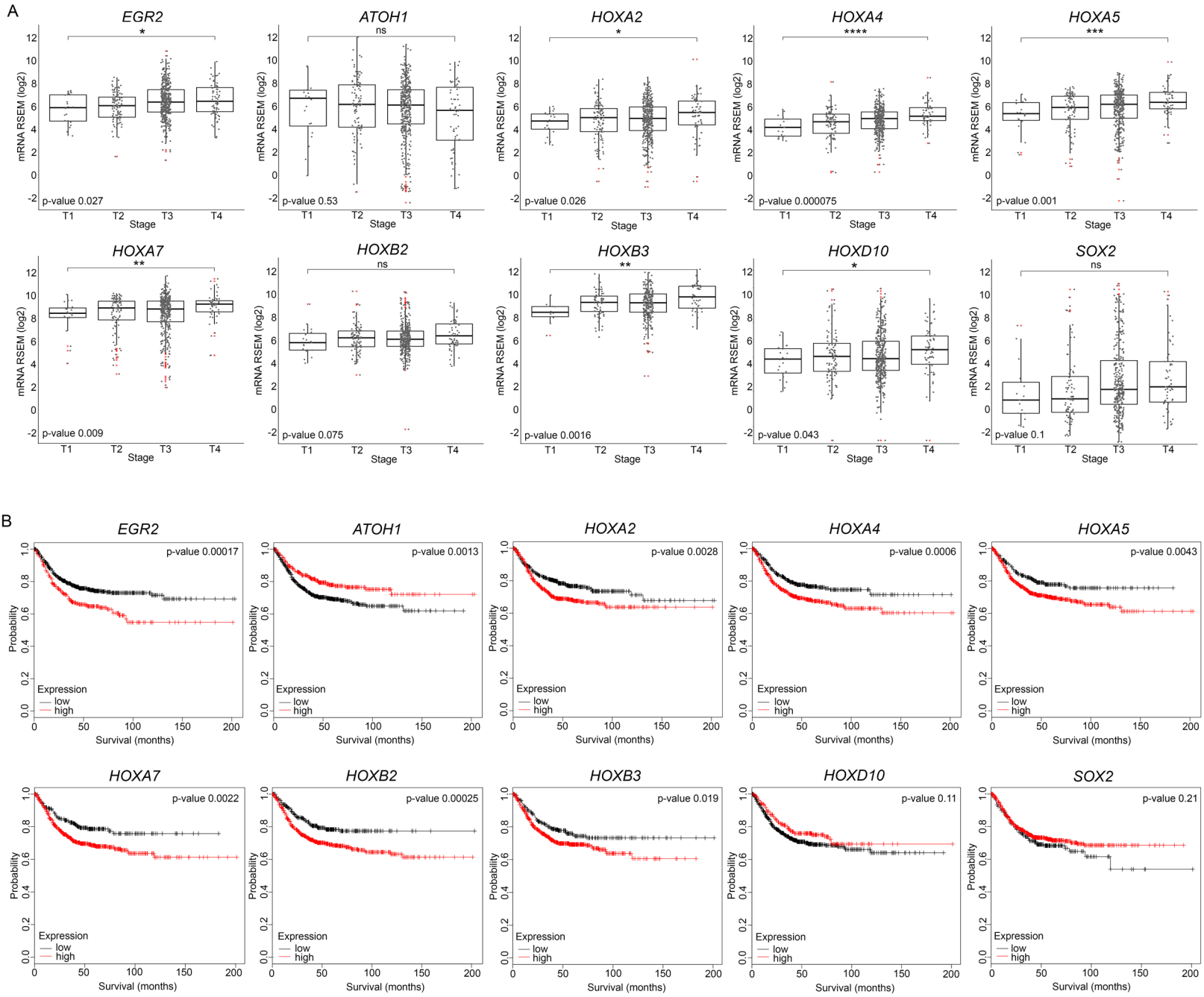


Figure 4



Supplementary Information

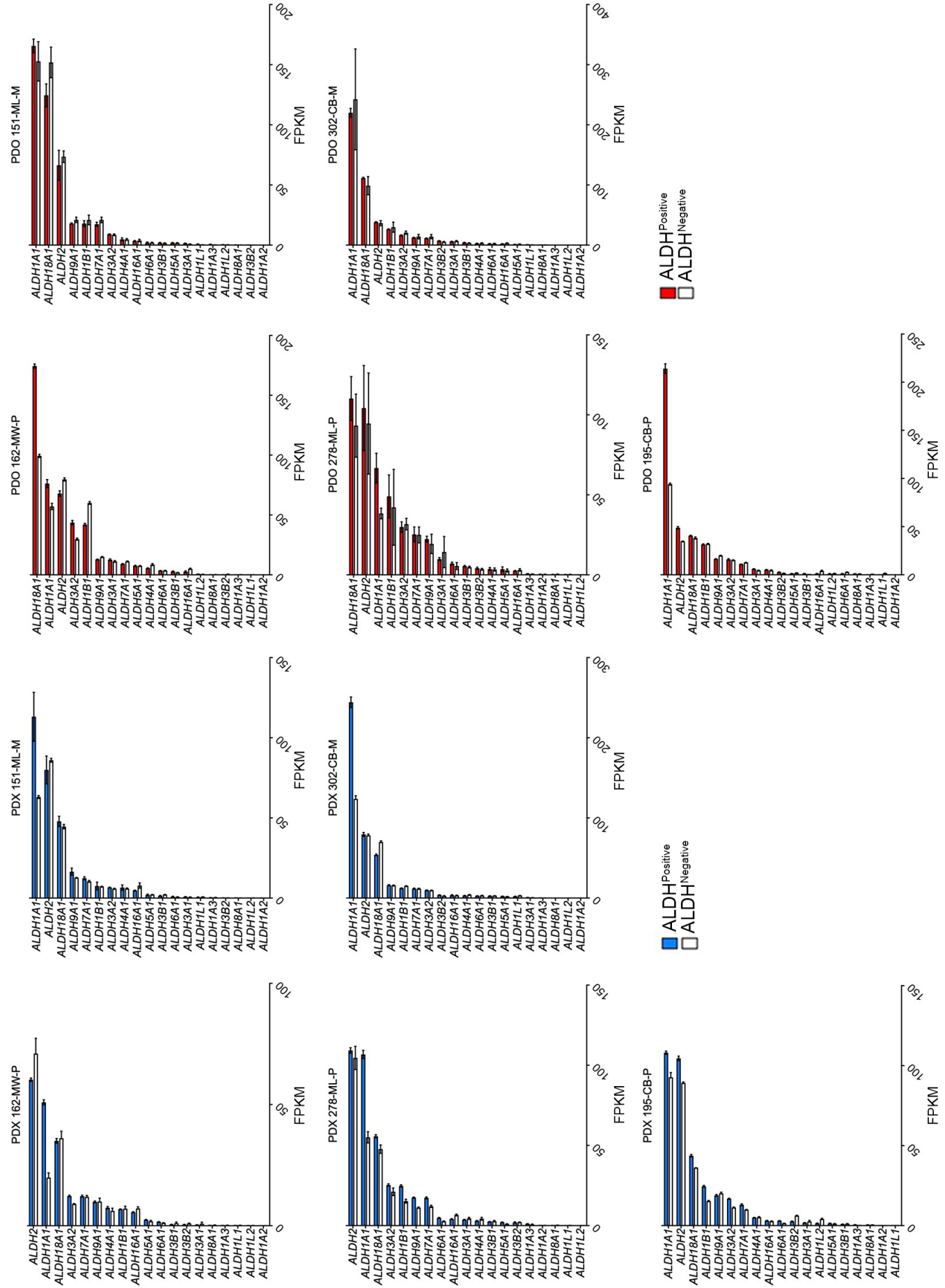


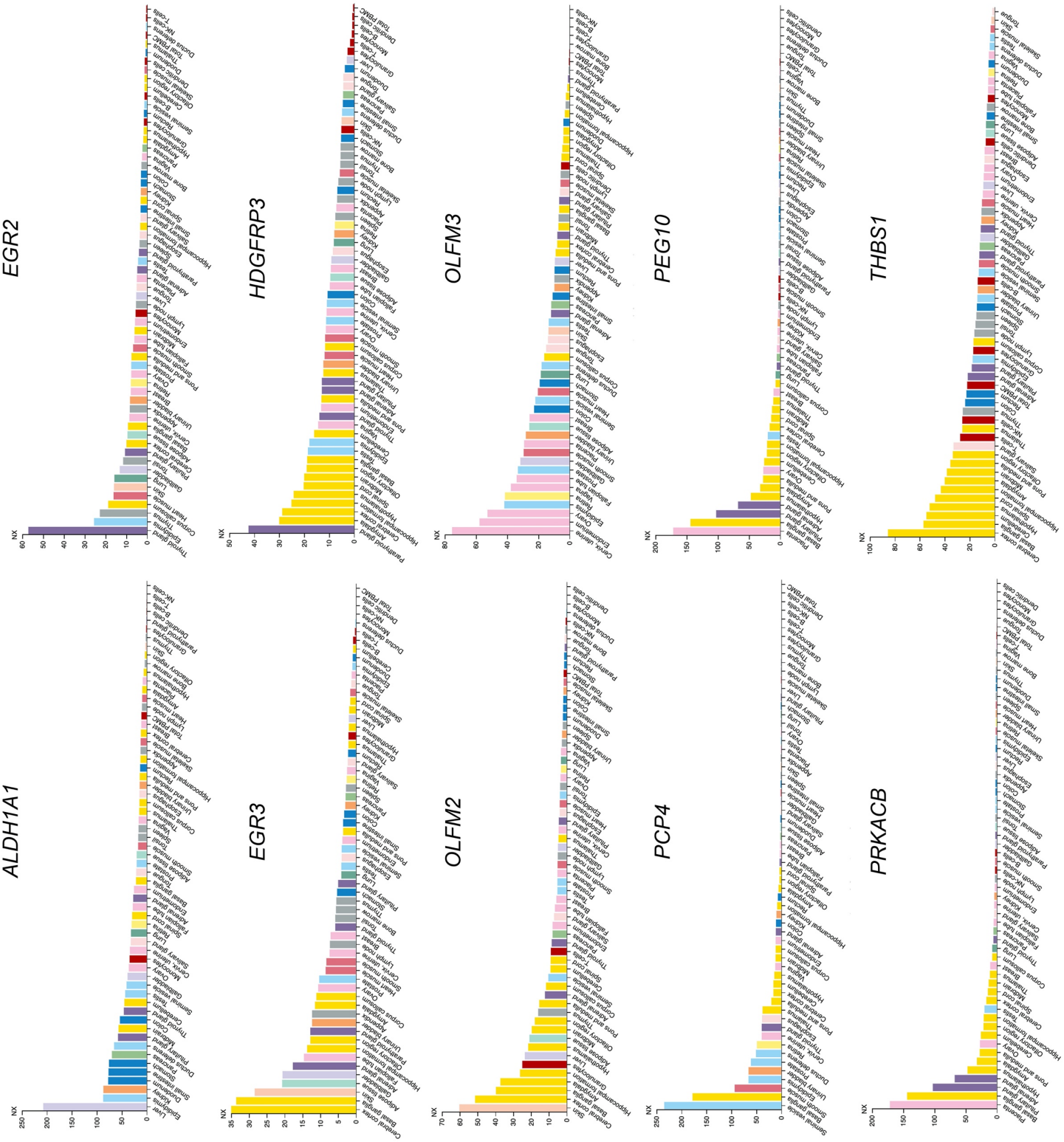
Figure S1: ALDH isoform expression in ALDH^{Positive} PDO (LHS) and PDX (RHS)cells. Related to Figure 1 and 2.

Gene Name	Description	Expression	Function
LINGO1	Leucine rich repeat and Immunoglobin domain-containing protein 1	nervous system, placenta, testis	regulates myelination, oligodendrocyte differentiation, axon regeneration, and neuronal survival
EGR2	Early growth response 2	thyroid, nervous system, placenta, testis	early myelination of the peripheral nervous system
RASL11B	RAS like family 11 member B	ovary, heart, kidney, nervous system, placenta	member of the small GTPase protein family with a high degree of similarity to RAS
DEFA6	defensin alpha 6	intestine, colon	antimicrobial and cytotoxic peptide thought to be involved in host defense
PRKACB	protein kinase cAMP-activated catalytic subunit beta	nervous system, placenta, testis	mediates cAMP-dependent signaling; regulates diverse cellular processes including cell proliferaton and differentiation
THBS1	thrombospondin 1	placenta, endometrium, smooth muscle	mediates cell-to-cell and cell-to-matrix interactions, roles in platelet aggregation, angiogenesis, and tumorigenesis
PCP4	Purkinje cell protein 4	nervous system	development of the central nervous system, may play a role in neuronal differentiation through activation of calmodulin-dependent kinase signaling pathways
EGR3	early growth response 3	nervous system, adrenal gland, gall bladder	early growth response gene induced by mitogenic stimulation; regulates biological rhythm, endothelial cell growth and migration, and neuronal development
FBLN1	fibulin 1	placenta, uterine,	mediates platelet adhesion via binding fibrinogen
CFI	complement factor I	liver, kidney, gallbladder	encodes a serine proteinase that is essential for regulating the complement cascade
ADAMTSL2	ADAMTS like 2	adenal gland, kidney, nervous system	a secreted glycoprotein that binds the cell surface and extracellular matrix
PHYHIPL	phytanoyl-CoA 2-hydroxylase interacting protein like	nervous system, adrenal gland, intestine, placenta, testis	May play a role in the development of the central system
OSBPL6	oxysterol binding protein like 6	nervous system, skeletal muscle, intestine	intracellular lipid receptor
HDGFRP3	Hepatoma-Derived Growth Factor 2	testis, nervous system, endometrium, placenta	nnhances DNA synthesis and may play a role in cell proliferation
OLFML3	olfactomedin like 3	uterine, endometrium, placental, smooth muscle	facilitates protein-protein interactions, cell adhesion, and intercellular interactions, scaffold protein and pro-angiogenic vascular tissue remodeler
FRMPD1	FERM and PDZ domain containing 1	testis, skin, nervous system	a regulatory binding partner for AGS3, which stabilizes GDP-bound inhibitory G proteins. FRMPD1 binds to and secures AGS3 localization at cell membranes
LMCD1	LIM and cysteine rich domains 1	skeletal ann smooth muscle, adipose tissue, placenta	protein-protein interactions, may act as a co-regulator of transcription
DACT1	dishevelled binding antagonist of beta catenin 1	nervous system, ovary, testis, placenta, colon	regulates dishevelled-mediated signaling pathways during development; may play role in the regulation of Wnt signaling via degradation of CTNNB1
PEG10	paternally expressed 10	placenta, nervous tissue, testis, ovary, adrenal	cell proliferation, differentiation and apoptosis; Overexpression of this gene has been associated with several malignancies
TENT5C	Terminal Nucleotidyltransferase 5C	bone marrow, colon, testis, stomach	Probable nucleotidyltransferase that may act as a non-canonical poly(A) RNA polymerase; Seems to enhance replication of some viruses
ERVMER34-1	endogenous retrovirus group MER34 member 1, envelope	Placenta, kidney, heart	retroviral envelope proteins, which mediate receptor recognition and membrane fusion during early infection.
SULT1B1	sulfotransferase family 1B member 1	intestine, liver, spleen	catalyze the sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds
OLFM2	olfactomedin 2	nervous system, skin	involved in early eye development and function; plays a role in TGFB-mediated differentiation of smooth muscle cells from mesenchymal stem cells
MAP1B	microtubule associated protein 1B	nervous system	microtubule assembly, required for neurogenesis
PCDHB16	protocadherin beta 16	parathyroid gland, nervous system	specific functions are unknown; likely plays a critical role in the establishment and function of specific cell-cell neural connections
ALDH1A1	aldehyde dehydrogenase 1 family member A1	liver, stomach, intestine, testis, ovary, adrenal	conversion/oxidization of retinaldehyde to retinoic acid
SLC2A12	solute carrier family 2 member 12	protate, endometrium, intestine, nervous system	Facilitative glucose transporter
UGT2A3	UDP glucuronosyltransferase family 2 member A3	Instestine, kidney	important in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds
PCSK5	proprotein convertase subtilisin/kexin type 5	Intestine, adipose, placenta, uterine, ovary, testis	likely functions in the secretory pathways; essential role in pregnancy establishment by proteolytic activation factors such as BMP2, CALD1 and alpha-integrins
RAB30	RAB30, member RAS oncogene family	nervous system, ovary, breast, placenta	regulator of intracellular membrane trafficking; required for maintaining the structural integrity of the Golgi apparatus,

Figure S2: List of genes differentially expressed and common in ALDH^{Positive} PDO and ALDH^{Positive} PDX CSCs. Related to Figure 2.

Figure S3: RNA tissue expression of differentially expressed and common ALDH^{Positive} PDO and ALDH^{Positive} PDX cells. Related to Figure 2.

Consensus RNA-seq dataset from the Human Protein Atlas (HPA) the Genotype-Tissue Expression (GTEx) project and CAGE data from FANTOM5 project. Color-coding is based on tissue groups, each consisting of tissues with functional features in common.



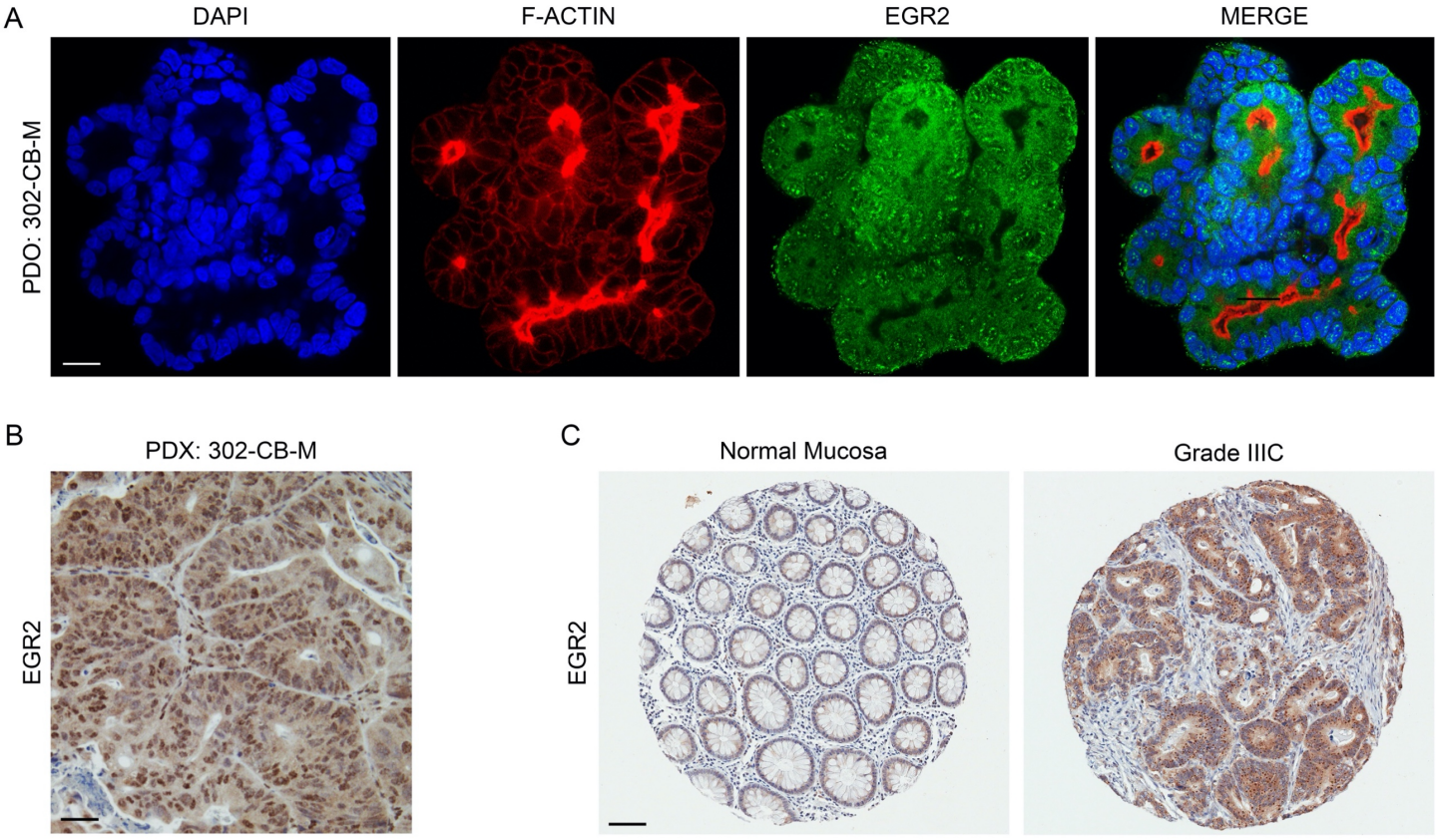


Figure S4: EGR2 immunostaining in PDO, PDX and clinical samples.

(A) Immunofluorescence staining of PDO for EGR2 (green) and F-ACTIN (red). Nuclei are stained blue with DAPI (Bars = 20 μm). Immunostaining of PDX tissue (B) and tissue microarrays of normal intestinal mucosa and colorectal cancer patient tissue (C) for EGR2. (Bars = 200 μm). Related to Figure 3.

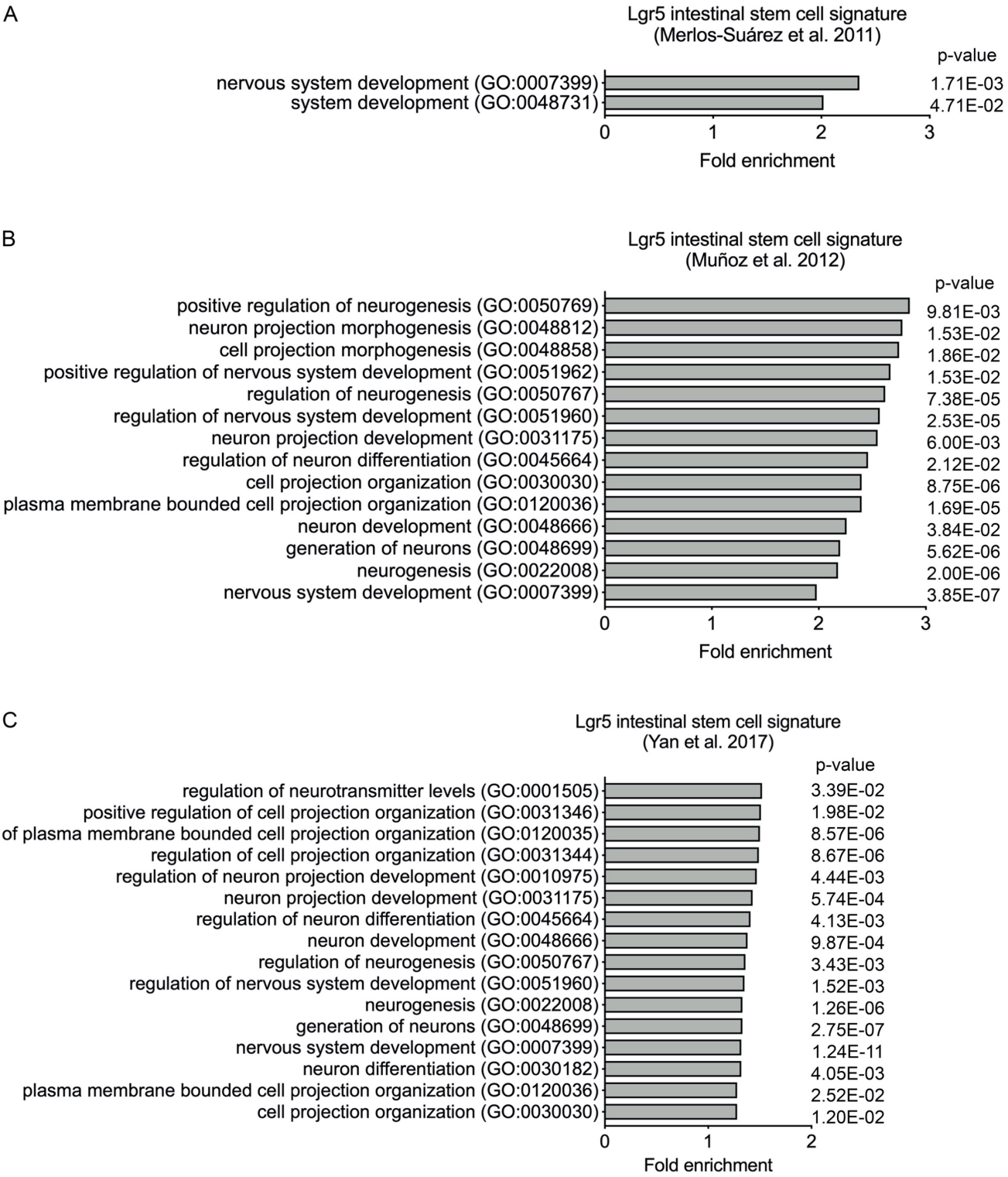


Figure S5: Lgr5 intestinal crypt-base stem cells are enriched for nervous system genes.

Gene ontology analysis of Lgr5 intestinal stem cell gene signatures from (A) Merlos-Suarez, *et al*., 2011), (B) Munoz, *et al*., (2012) and (C) Yan, *et al*., (2017). Related to Figure 2.

Patient Model	Origin	TNM stage	Stage
162-MW-P	Sigmoid colon & descending colon	T3 N0 M0	IIA
151-ML-M	Liver	T2 N0 M0 , M1a	IVA
278-ML-P	Sigmoid colon & descending colon	T4a N0 M0	IIB
302-CB-M	Liver	T3 N1a M1a	IVA
195-CB-P	Sigmoid colon	T4a N2b M1a	IVA

Table S1 Tissue Origin and TNM Classification of Malignant Tumors (TNM). Related to Figure 1.

T: primary tumor size, N: regional lymph nodes involved, M: distant metastasis.

siRNA	Dharmacon™ Product	Product Number
ALDH1A1	Accell Human (216) siRNA - SMARTpool	E-008722-00-5
EGR2	Accell Human (1959) siRNA - SMARTpool	E-006527-01-5
EGR3	Accell Human (1960) siRNA - SMARTpool	E-006528-00-5
HDGFRP3	Accell Human (50810) siRNA - SMARTpool	E-017093-00-5
OLFM2	Accell Human (93145) siRNA - SMARTpool	E-015212-00-5
OLFML3	Accell Human (56944) siRNA - SMARTpool	E-020325-00-5
PCP4	Accell Human (5121) siRNA - SMARTpool	E-020122-00-5
PEG10	Accell Human (23089) siRNA - SMARTpool	E-032579-00-5
PRKACB	Accell Human (5567) siRNA - SMARTpool	E-004650-00-5
THBS1	Accell Human (7057) siRNA - SMARTpool	E-019743-00-5

Table S2: Dharmacon™ Smartpool siRNAs. Related to Figure 3.

LENTIVIRUS	SIGMA PRODUCT	PRODUCT NAME	VECTOR	TRC NUMBER
Control	SHC003V	MISSION® tGFP™ Positive Control Transduction Particles	–pLKO.1-puro-CMV-tGFP	NA
shEGR2 1	SHCLNV-NM_000399	EGR2 MISSION shRNA Lentiviral Transduction Particles	–hPGK-Puro-CMV-tGFP	TRCN0000013839
shEGR2 2	SHCLNV-NM_000399	EGR2 MISSION shRNA Lentiviral Transduction Particles	–hPGK-Puro-CMV-tGFP	TRCN0000013840
shEGR2 3	SHCLNV-NM_000399	EGR2 MISSION shRNA Lentiviral Transduction Particles	–hPGK-Puro-CMV-tGFP	TRCN0000013841

Table S3. Lentiviral Transduction Particles. Related to Figure 3.

Symbol	Gene Name	UniGene ID	TaqMan® Gene Expression Assay
ATOH1	atonal BHLH transcription factor 1	Hs.532680	Hs00245453_s1
AXIN2	axin 2	Hs.156527	Hs00610344_m1
BMI1	BMI1 proto-oncogene, polycomb ring finger	Hs.380403	Hs00180411_m1
CTNNB1	catenin beta 1	Hs.476018	Hs00355049_m1
EGR2	early growth response 2	Hs.1395	Hs00166165_m1
EPHA4	EPH receptor A4	Hs.371218	Hs00953178_m1
EPHB2	EPH receptor B2	Hs.523329	Hs00362096_m1
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	Hs.544577	Hs02758991_g1
HOXA2	homeobox A2	Hs.445239	Hs00534579_m1
HOXA5	homeobox A5	Hs.655218	Hs00430330_m1
HOXA7	homeobox A7	Hs.610216	Hs00600844_m1
HOXB2	homeobox B2	Hs.514289	Hs01911167_s1
HOXB3	homeobox B3	Hs.654560	Hs05048382_s1
HOXD10	homeobox D10	Hs.123070	Hs00157974_m1
LGR5	leucine rich repeat containing G protein-coupled receptor 5	Hs.658889	Hs00969422_m1
MKI67	MKI67	Hs.689823	Hs04260396_g1
MUC1	mucin 1, cell surface associated	Hs.89603	Hs00159357_m1
MYC	v-myc avian myelocytomatosis viral oncogene homolog	Hs.202453	Hs00153408_m1
RUNX2	runt related transcription factor 2	Hs.535845	Hs01047973_m1
SOX2	SRY-box 2	Hs.518438	Hs01053049_s1
HDGFRP3	hepatoma-derived growth factor, related protein 3	Hs.513954	Hs00274988_m1
OLFM2	olfactomedin 2	Hs.169743	Hs01017934_m1
OLFML3	olfactomed like 3	Hs.9315	Hs01113293_g1
PCP4	Purkinje cell protein 4	Hs.80296	Hs01113638_m1
PEG10	paternally expressed 10	Hs.147492	Hs00248288_s1
PRKACB	protein kinase cAMP-activated catalytic subunit beta	Hs.487325	Hs01086757_m1
THBS1	thrombospondin 1	Hs.164226	Hs00962908_m1

Table S4. Taqman Gene Expression Assays. Related to Figure 3.