Original Article

Machine Learning–Guided Differentiation Therapy Targets Cancer Stem Cells in Colorectal Cancers

Authors: Saptarshi Sinha[1], Joshua Alcantara[1, 2], Kevin Perry[1, 2], Vanessa Castillo[1], Annelies Ondersma[1, 2], Satarupa Banerjee[1, #], Ella McLaren[1, 2], Celia R. Espinoza[1], Sahar Taheri[3], Eleadah Vidales[1, 2], Courtney Tindle[1, 2], Adel Adel[1], Siamak Amirfakhri[4], Joseph R. Sawires[5], Jerry Yang[5], Michael Bouvet[4,6], Pradipta Ghosh[1,2,6,7‡]

Affiliations:

[1]Department of Cellular and Molecular Medicine, School of Medicine, University of California San Diego, La Jolla, CA 92093.

[2]HUMANOID™ Center of Research Excellence, University of California San Diego, La Jolla, CA 92093.

[3]Department of Computer Science and Engineering, Jacob's School of Engineering, University of California San Diego, CA 92093.

[4]Department of Surgery, University of California San Diego, San Diego, CA 92093, USA; Department of Surgery, VA San Diego Healthcare System, San Diego, CA, USA.

[5]Department of Chemistry and Biochemistry, University of California San Diego, San Diego, CA, 92093, USA.

[6]Moores Cancer Center, University of California San Diego, San Diego, CA, 92093, USA.

[7]Department of Medicine, University of California San Diego, San Diego, CA, 92093, USA.

Additional Affiliations: [#]Department of Biotechnology, School of Biosciences and Technology, Vellore Institute of Technology, Tamil Nadu, India 632014.

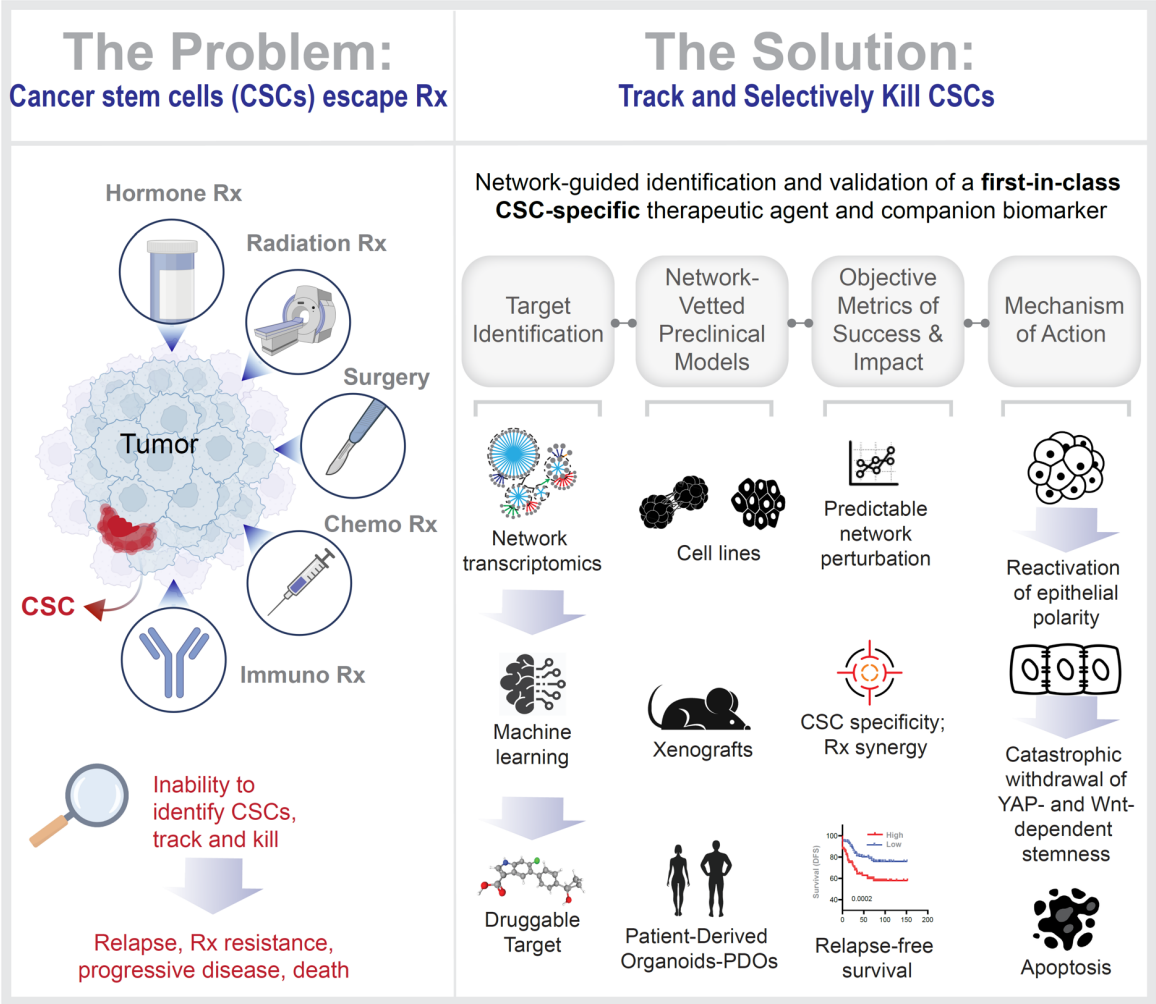
Keywords: Cancer stem cell, colorectal cancer, differentiation therapy, CDX2 restoration, PRKAB1 agonist.

[‡]**Corresponding author and Lead Contact**: Pradipta Ghosh, M.D.; Professor, Departments of Medicine and Cellular and Molecular Medicine, University of California San Diego; 9500 Gilman Drive (MC 0651), George E. Palade Building, Room 232, 239; La Jolla, CA 92093. Phone: 858-822-7633. Email: prghosh@ucsd.edu.

Abstract (212 words)

Despite advances in artificial intelligence (AI) within cancer research, its application toward realizing differentiation therapy in solid tumors remains limited. Using colorectal cancer (CRC) as a model, we developed a machine learning (ML) framework, ***CANDiT*** (*Cancer Associated Nodes for Differentiation Targeting*), to selectively induce differentiation and death of cancer stem cells (CSCs)—a key obstacle to durable response. Centering on one node, *CDX2*, a master differentiation factor lost in high-risk, poorly differentiated CRCs, we built a transcriptomic network to identify therapeutic strategies for CDX2 restoration. Network-based prioritization identified *PRKAB1*, a stress polarity sensor, as a top target. A clinical-grade PRKAB1 agonist reprogrammed transcriptional networks, induced crypt differentiation, and selectively eliminated CDX2-low CSCs in CRC cell lines, xenografts and patient-derived organoids (PDOs). Multivariate analyses in PDOs revealed a strong therapeutic index, linking efficacy (IC_{50}) to the biomarker-defined CDX2-low state. A 50-gene response signature—derived from an integrated analyses of all three models and trained across multiple datasets—revealed that CDX2 restoration therapy may translate into a ~50% reduction in recurrence and mortality risk. Mechanistically, treatment activated a differentiation-associated stress polarity signaling axis while dismantling Wnt and YAP-driven stemness programs essential to CSC survival. Thus, *CANDiT* offers a scalable path to CSC–directed therapy in solid tumors by translating transcriptomic vulnerabilities into precision treatments.

Graphic Abstract:



One sentence summary: In this work, Sinha et al. introduce a machine learning–guided framework to identify and target transcriptomic vulnerabilities in colorectal cancer, demonstrating that differentiation therapy selectively eliminates cancer stem cells and reduces recurrence risk.

Highlights:

- An ML framework (*CANDiT*) identifies target for differentiation therapy for CRCs
- Therapy induces crypt differentiation and CSC-specific cytotoxicity
- CDX2-low state predicts therapeutic response; restoration improves prognosis
- Therapy dismantles stemness via reactivation of stress polarity signaling

Introduction

Poor differentiation, marked by elevated stemness, is a hallmark of cancers. Differentiation therapy, which targets cancer stem cells (CSCs) to induce maturation, has shown success in hematologic malignancies—most notably with all-trans-retinoic acid (ATRA) in acute promyelocytic leukemia (APML)[1-10]. While other agents have also shown promise in leukemias[11,12], differentiation therapy has not yet translated to carcinomas. A major obstacle is the profound intra- and inter-tumoral heterogeneity that obscures identification of CSCs[13].

To address this, computational methods have been developed to interrogate transcriptomic data, construct gene networks, and identify therapeutic targets[14-19]. Traditional symmetric frameworks—such as correlation[20-25], mutual information[17], linear regression[26], dimension reduction[27], and clustering[28,29] —often fail in the face of real-world biological complexity. In contrast, Boolean implication-based network transcriptomics[30,31] uses asymmetric, invariant gene relationships to construct directed networks that map evolving cellular states. This approach has identified translationally relevant states and targets across tissues and contexts[32-45] with high degrees of precision, including a first-in-class therapeutic to protect gut barrier function in inflammatory bowel diseases (IBD)[30].

Here we applied this method to identify targets that restore expression of CDX2, a caudal-related homeobox transcription factor and tumor suppressor[46-49]. In 2016, an unbiased search (using the same Boolean logic approach[50,51]) identified CDX2 as the top marker of colonic epithelial differentiation[52]. Its expression inversely correlates with "activated leukocyte cell adhesion molecule" (*ALCAM*/CD166), a stem cell marker that is present at the crypt base[53,54] and on highly tumorigenic human CRC cells[55]. CDX2 loss—seen in ~9% of CRCs—correlates with poor differentiation, worse prognosis, and enhanced chemotherapy benefit even in stage II disease[52]. Since then, 32 independent studies involving >13,000 patients have confirmed CDX2-low CRCs are associated with worse overall (OS) and disease-free (DFS) survival, independent of stage, mismatch repair (MMR) status, or ethnicity[56-70] [**Figure 1**-*Step 1*]. One study integrated IHC with quantitative mass spectrometry[71]. Absence of CDX2 in tumors is associated with several adverse prognostic features, such as poor differentiation, advanced stage, vascular invasion, right-sided location, CpG island methylator phenotype (CIMP), and *BRAF* mutation[72-74]. Conversely, CDX2 presence lowers recurrence and mortality risk by ~50% and 52%, respectively[69] —with even greater benefit (~70%) in stage II–III CRCs. Collectively, these findings establish the restoration of CDX2 as a clinically meaningful and high-priority therapeutic goal [see **Figure 1**-*Step 1*]. While its biological importance is widely recognized, no pharmacologic approach has yet succeeded in reliably inducing *CDX2*. Because *CDX2* alone is insufficient to fully activate the intestinal lineage program[75] —and that its precise regulation is critical to preserve mucosal architecture—exogenous expression approaches are impractical.

Here, we used network-transcriptomic analysis[50,51] to identify therapeutic targets that upregulate CDX2, triggering a network-wide, domino-like cascade that drives differentiation in CRC. We selected CDX2 as a biologically validated master regulator whose reactivation enables reprograms CSC fate[76], and bridges cellular differentiation states to clinical outcomes, linking cancer stem cell behavior to patient prognosis.

Results

A network-rationalized approach, CANDiT

We applied a machine learning–based computational platform, the Boolean Network Explorer (*BoNE*)[30], and *CDX2* as a 'seed' gene (i.e., node) to model progressive gene regulatory events during colonic differentiation (**Figure 1**-*Step 2;* **Figure S1A-B**). The resulting framework—*CANDiT* (*Cancer-Associated Nodes for Differentiation Targeting*)—prioritizes actionable differentiation nodes in cancer networks to enable therapeutic reprogramming. The model was trained initially on a maximally heterogeneous dataset comprising 1,662 colorectal cancers (CRCs), 68 adenomas, and 170 normal tissues (**Figure S1C-D**). The model was refined subsequently using statistical learning on another independent dataset (GSE77953). Classification accuracy identified five statistically top-ranked target genes from *CDX2*-proximal clusters (#2 and #3; **Figure S1D**) that are predicted to upregulate *CDX2* when activated (**Figure 1**-*Step 2;* **Figure S1E**). These predictions were validated in 2,937 additional human colon samples (**Figure S1F**) and 318 mouse samples (**Figure S1G**), confirming cross-species and cross-cohort robustness.

To identify potent, selective modulators of the five targets, we queried the Protein Data Bank [PDB; https://www.rcsb.org] for structure-resolved agonists. Selective ligands were identified only for PRKAB1 (25 molecules/structures); none were found for the other targets. Cross-referencing these hits in ClinicalTrials.gov yielded PF-06409577 (PF), a clinical-grade PRKAB1 agonist. Although not yet approved for a clinical indication, PF has been deemed safe in a randomized, double-blind, placebo-controlled Phase I trial (NCT02286882[77]). PF was recently shown to specifically upregulate PRKAB1 at both transcript and protein levels in the colonic epithelium ex vivo and in mouse colon in vivo[30].

The *CANDiT*-inferred network predicted that agonizing PRKAB1, the protein product of *PRKAB1* (Protein Kinase AMP-activated Non-Catalytic Subunit Beta-1), would: (i) upregulate *PRKAB1* itself and co-clustered genes (via "equivalent" relationships); (ii) induce *CDX2* and its proximal clusters (#2 and #3) via "hi⇒hi" links; and (iii) repress stemness-associated genes (e.g., *CCDC88A*, cluster #9) via "hi⇒lo" links (**Figure S1D**). This regulatory logic holds consistently across diverse human and mouse CRC datasets, including early-onset CRCs (**Figure S2A-D**). Further supporting its selection, *PRKAB1*—but not its paralog *PRKAB2*—is highly expressed in the gastrointestinal tract[30] and PRKAB1 agonists promote epithelial polarity in the gut[78-81].

For validation studies *in cellulo* (CRC cell lines), *in vivo* (murine xenografts) and *ex vivo* (a 'living biobank' of patient derived organoids; PDOs), we prioritized models with contrasting CDX2 expression (high vs. low, as determined based on a threshold that was determined using *StepMiner*[82]; see *Methods*): CDX2-low (target phenotype) and CDX2-high (negative control). In addition to conventional anti-tumor readouts, three pre-specified success metrics were used to evaluate therapeutic efficacy; all three models were required to meet all three endpoints (see **Figure 1**-*Step 3*). Previous attempts to induce CDX2 transiently[83], failed to produce durable differentiation and showed inconsistent impact across these same metrics (**Figure S2E-H**). Finally, through an integrated transcriptomic approach validated in tumor tissues, we define the molecular mechanism of action and translational impact of successful *CDX2* reinstatement via PRKAB1 agonism (see **Figure 1**-*Step 4*).

174 ***Target validation on CRC cell lines and xenograft models***

175 We prioritized two poorly differentiated CRC cell lines—HCT116 and SW480—which exhibited the lowest
176 CDX2 transcript levels among 26 sequenced CRC cell lines profiled by microarray (GSE10843)—as optimal
177 models for hypothesis testing. Two well-differentiated, *CDX2*-high cell lines (DLD1, *KRAS*^{mut} and Caco2,
178 *KRAS*^{WT}) were selected as a negative control (**Figure 2A**). PF induced cell death within 48 hours in CDX2-
179 low HCT116 and SW480 cells (IC_{50} ~6-7 μM), but not in *CDX2*-high DLD1 or Caco2 cells (IC_{50} >20 μM;
180 **Figure 2B**; **Figure S3A-C**). In both *CDX2*-low cell lines, PF-induced cell death was preceded by dose-
181 dependent induction of differentiation and suppression of stemness markers, at both transcript (**Figure 2C**)
182 and protein (**Figure S3D-G**) levels. These effects coincided with a dose-dependent increase in apoptosis
183 (**Figure S4A-H**). Importantly, PF failed to reinstate CDX2, alter differentiation or stemness markers (**Figure**
184 **2D**), or induce cell death (**Figure 2E**) in *PRKAB1*-depleted cells, confirming target specificity of its anti-cancer
185 effects. In contrast, Metformin—a non-specific, indirect agonist of AMP-kinase (activating both PRKAB1- and
186 PRKAB2-containing complexes)—neither induced CDX2 expression nor triggered cell death, even at high
187 doses (**Figure 2B**; **Figure S4I-K**), underscoring the necessity of PRKAB1-specific activation for therapeutic
188 efficacy.

189 In vivo, HCT116 xenografts implanted subcutaneously in nude mice exhibited a 68% reduction in
190 tumor volume with PF treatment (AUC: vehicle = 51,280.23; PF = 26,676.77; **Figure 2G-I**) and a 62.5%
191 increase in survival (**Figure 2J**; Mantel-Haenszel Hazard Ratio = 0.09 for reduction in death rates). PF-
192 treated tumors developed glandular structures (**Figure S5A**), reinstated *CDX2* expression, and showed
193 increased differentiation and reduced stemness markers by qPCR (**Figure 2K**). IHC confirmed nuclear
194 localization of CDX2 and upregulation of PRKAB1 protein (IHC; **Figure 2L-M; Figure S5B**). Reflecting the
195 slow division rate of colonic stem cells[84], PF triggered a ~3-fold increase in mitotic bodies (**Figure S5C**;
196 **Figure 2M**). Additional signs of tumor differentiation included a reduction in nuclear β-catenin (**Figure S5D)**
197 and a corresponding increase in junctional β-catenin (**Figure 2N**), which co-localized with E-cadherin (**Figure**
198 **2N; Figure S5E**).

199 Together, these findings demonstrate that CDX2-reinstatement therapy drives differentiation and
200 apoptosis selectively in *CDX2*-low CRC models, both in vitro and in vivo, through PRKAB1-specific activation.

201

202 ***Target validation in patient-derived organoids***

203 We next evaluated the therapeutic efficacy of PF in PDOs. To evaluate potential toxicity, we first tested PF
204 on healthy colon PDOs using eTOX Red staining, which selectively labels dead cells based on membrane
205 permeability (**Figure 3A**-*left*). At 5 μM and 20 μM, PF, there was no significant increase in cell death
206 compared to vehicle control. As expected, hydrogen peroxide (H_2O_2), a positive control, induced significant
207 cytotoxicity relative to both PF-treated and control groups (**Figure 3A**-*right*; **Figure 3B**). We then tested PF
208 across two prospective cohorts comprising 23 CRC PDOs and 3 healthy colon PDOs (see **Table** and
209 *Methods*; **Figure 3C**). PF exhibited selective anti-cancer activity in CDX2-low CRC PDOs (as determined
210 based on a threshold that was determined using *StepMiner*[82] cutoff; see *Methods*), reflected by low IC_{50}
211 values. In contrast, CDX2-high CRC PDOs and all healthy colon PDOs were resistant (IC_{50} > 40 μM; **Figure**
212 **3D**). Notably, IC_{50} values positively correlated with expression of CDX2 (seed gene) and PRKAB1 (network-

213 derived target), while showing an inverse correlation with CCDC88A, a network-linked stemness marker
214 (**Figure S1D**; **Figure 3D**).
215 Using only these three network-derived genes (*CDX2, PRKAB1*, and *CCDC88A*), hierarchical
216 unsupervised clustering accurately classified sensitive versus resistant PDOs (**Figure 3E**). As in prior
217 models, PF promoted differentiation and suppressed stemness markers in CRC PDOs (**Figure 3F**) —effects
218 that were again restricted to CDX2-low tumors, consistent with findings from CRC cell lines (**Figure 2C**) and
219 xenografts (**Figure 2K**).
220 Multivariate analysis of CRC PDOs (see **Supplemental Information 3** for covariates) identified CDX2
221 expression as the most significant predictor of IC_{50} (**Figure 3G**). Two additional variables also emerged:
222 histological grade and *KRAS* mutation status. Specifically, PF was less effective in moderately or well-
223 differentiated tumors that more frequently harbor KRAS mutations (**Figure 3G**). The interplay between CDX2
224 induction and *KRAS* status suggested a potential ceiling effect—i.e., CDX2 may remain constitutively high in
225 $KRAS^{mut}$ tumors, limiting the therapeutic benefit of PF. A univariate analysis in a large CRC cohort with known
226 driver mutations and patient outcomes confirmed this relationship: CDX2 was prognostic in $KRAS^{WT}$ tumors
227 (**Figure 3H**-*left*), but not in $KRAS^{mut}$ tumors (**Figure 3H**-*right*).
228 These findings indicate that despite its modest size, our PDO cohort captures clinically relevant
229 heterogeneity and reveals a key interaction between *CDX2* status and *KRAS* mutation in determining
230 therapeutic response.
231

232 ***Indicators of therapeutic efficacy and target cell specificity***

233 RNA sequencing across all three model systems—CRC cell lines, xenografts, and PDOs—confirmed that
234 PRKAB1 agonism met the three a priori defined therapeutic success metrics: (i) reinstatement of CDX2, the
235 primary therapeutic goal (bubble plots in **Figure 4A**-*top row*; **4B**; violin plots in **Figure S6A-C**), (ii) induction
236 of markers for crypt-top enterocytes and goblet cells, established surrogates of differentiation in the colon
237 crypt[51,85] (**Figure 4A**-*middle row*; **Figure S6D-F**), and (iii) predictable perturbation of the network, specifically
238 induction of genes in clusters #2 and #3 (**Figure 4A**-*bottom row*; **Figure S6G-I**). Importantly, these effects
239 were observed exclusively in *CDX2*-low CRC models—cell lines, xenografts, and PDOs—but not in *CDX2*-
240 high PDOs or healthy colon PDOs (**Figure 4A**-*columns #5 and #6*; **Figure S6C, S6F, S6I**).
241 A recent study using a novel machine learning algorithm defined transcriptomic and epigenetic
242 signatures that distinguish normal stem cells from their differentiated progeny, unveiling CSC-specific gene
243 sets indicative of oncogenic dedifferentiation[86]. Applying these signatures, we found that CDX2-reinstatement
244 therapy did not affect normal stem cell–related genes in either healthy or CRC PDOs (**Figure 4C**). In contrast,
245 it selectively downregulated: (i) genes associated with cancer-related stemness (**Figure 4D**), (ii) Yamanaka
246 factors that promote somatic cell reprogramming and are implicated in CSC origins[87] (**Figure 4E**), and (iii)
247 CRC-specific CSC gene signatures (**Figure 4F**).
248 These findings demonstrate that *CDX2*-reinstatement acts as a trigger for broader transcriptomic
249 reprogramming that selectively suppresses CSC programs unique to *CDX2*-low tumors, while sparing normal
250 stem cells in healthy colon and *CDX2*-high tumor cells.
251

252 ***Potential for synergy with other therapeutic modalities***

253 CDX2-low Stage II–III CRCs have previously been shown to derive benefit from adjuvant chemotherapy,
254 whereas CDX2-high tumors do not[52]. We therefore asked whether therapeutic reinstatement of CDX2 in
255 CDX2-low tumors would enhance responsiveness to chemotherapy. Across all models tested, PF treatment
256 consistently altered patient-derived gene expression signatures—inducing those associated with favorable
257 response and suppressing those linked to resistance to various chemotherapeutics (e.g., 5-FU, Oxaliplatin,
258 FOLFIRI) (**Figure 4G-i**). In line with these transcriptional changes, PF (5 μM, 24 h) exhibited synergy with
259 multiple chemotherapeutic agents (see **Figure 4H** legend).

260 We extended this analysis to other treatment modalities. PF similarly induced gene signatures
261 predictive of response to anti-EGFR therapy (Cetuximab; **Figure 4G-ii**) and several tyrosine kinase inhibitors
262 (TKIs; **Figure 4G-iii**), with a few notable exceptions (Afatinib and Gefitinib; **Figure 4G-iii**). It also enhanced
263 signatures associated with immunotherapy response while suppressing those of non-response (**Figure 4G-**
264 **iv**). Consistent with this, PF robustly induced the consensus molecular subtype CMS1 across all three model
265 systems (**Figure 4G-v**). This subtype is characterized by a favorable immune landscape and enhanced
266 sensitivity to immune checkpoint inhibitors[88]. In vivo xenograft studies also revealed a treatment-associated
267 reduction in the consensus molecular subtype CMS4, which is linked to poor prognosis and characterized by
268 high stromal content and TGF-β signaling[88,89]. Consistent with this, PF treatment significantly reduced stromal
269 infiltration, as determined by two independent gene expression signatures (see **Figure 4I-J** legend).

270 These findings suggest that *CDX2*-reinstatement acts as a trigger for broader transcriptomic
271 reprogramming that favors existing treatment modalities, including chemotherapy, targeted therapies, and
272 immunotherapy.

273

274 ***A molecular signature of therapeutic response and selectivity***

275 To characterize the therapeutic response to CDX2-reinstatement and its impact on the broader
276 transcriptome, we performed integrated differential expression analysis using RNA-seq data from all three
277 models (CRC cell lines, xenografts, and PDOs; see *Methods*). Differentially expressed genes (DEGs) from
278 two independent models (cell lines and xenografts; **Supplemental Information 4-5**) were refined using data
279 from the third (CRC PDOs), yielding a 50-gene signature (**Figure 5A**). These genes were significantly
280 upregulated in *CDX2*-low CRC PDOs compared to healthy and *CDX2*-high CRC PDOs (**Figure 5B**; **Figure**
281 **S7A-B**). Upon PF treatment, this signature was downregulated to levels comparable to healthy and *CDX2*-
282 high CRC PDOs (**Figure 5B**; see **Supplemental Information 6**), suggesting a shift from a stem-like to a
283 differentiated state.

284 This 50-gene signature included several unique cancer stem cell (CSC)-associated genes—*AGR2*[90],
285 *ASCL2*[91,92], *ALDH3A1*[93], and *TM4SF1* [94,95] —not previously captured in machine-learning-derived CSC
286 signatures (**Figure 4C-F**). Pathway enrichment analysis revealed activation of pro-survival programs known
287 to sustain CSCs[96-99], including PTK6-dependent JAK/STAT signaling, growth factor receptor and non-
288 receptor tyrosine kinases, Hippo/YAP signaling, and TGFβ suppression (**Figure 5C**). Notably, these
289 pathways were selectively downregulated in *CDX2*-low CRCs, but not in healthy or *CDX2*-high CRCs. Similar
290 pathway enrichment analyses of the upregulated genes showed genes associated with epithelial structure
291 maintenance (**Figure S7C**).

292 ***Impact of CDX2 reinstatement therapy on survival***

293 To evaluate clinical relevance, we tested whether the 50-gene signature could stratify relapse-free/disease-
294 free survival (RFS/DFS) in multiple CRC patient cohorts. Across all datasets, patients with low 50-gene
295 scores—resembling PF-treated or *CDX2*-high tumors—had significantly better outcomes than those with high
296 scores (**Figure 5D, F-K**). The associated odds ratio was ~2.0 (**Supplemental Information 7-*tab #1***),
297 indicating that elevated expression of the signature doubles the risk of recurrence or death—risk that may be
298 mitigated by CDX2-reinstatement therapy.

299 Univariate analysis showed that both *CDX2* expression and the 50-gene signature were significant
300 predictors of RFS (**Figure 5E-*left***), but the signature performed better, especially in *KRAS*^{WT} tumors (**Figure**
301 **5E-*middle***). Across all patient cohorts, the 50-gene score consistently outperformed *CDX2* as a prognostic
302 indicator (**Figure S7D-G**; **Supplemental Information 7**). As seen previously for *CDX2* (**Figure 3E**), the 50-
303 gene signature did not correlate with outcome in *KRAS*^{mut} tumors (**Figure 5E-*right***), suggesting that *KRAS*
304 mutation may bypass the CDX2 axis. In fact, within the *KRAS*^{WT} group, patients with low 50-gene scores
305 resembling PF-treated tumors had significantly better outcomes than those with high scores regardless of
306 microsatellite stability status (**Figure 5J-L**).

307 Together, these findings identify a robust, target-derived 50-gene signature that not only captures the
308 molecular response to *CDX2*-reinstatement but also stratifies survival risk. Therapeutic suppression of this
309 signature emerges as a measurable and meaningful objective, with potential to improve patient outcomes.

310

311 ***State-specific cell fates dictate selectivity of CDX2-restorative therapy***

312 Hallmark pathway analysis revealed that *CDX2*-restoration via PF treatment consistently activated a
313 transcriptional program across all three models centered on the restoration of epithelial integrity, including
314 polarity and cell junction assembly (**Figure 6A**). Integration of downregulated genes across models identified
315 a core set of commonly suppressed transcripts (**Figure 6B**), and transcription factor (TF) enrichment analysis
316 implicated key stemness-associated TFs, e.g., *TEAD4*, *SP1*, and *KLF1/4* as top regulators of these genes
317 (**Figure 6C**).

318 Histological analysis of xenograft tumors reinforced these transcriptional findings. PF-treated tumors
319 displayed a significantly increased number of asymmetrically stretched nuclei, consistent with chromatin
320 reorganization during epithelial reprogramming (**Figure 6D**). These morphological changes were
321 accompanied by pronounced alterations in Hippo signaling: PF reduced nuclear localization of YAP (**Figure**
322 **6E**) while increasing its inactive, phosphorylated (p)-Ser^{127} form (**Figure 6F**). This is consistent with pSer^{127}-
323 mediated cytosolic retention and phosphodegron-dependent degradation of YAP oncogenic activity[100],
324 explaining its near-total loss in treated tumors (**Figure 6E**).

325 PRKAB1 is also known to activate the stress polarity signaling (SPS) pathway[78-80]; it is a sentinel
326 system that integrates metabolic stress with epithelial polarity, coordinating cell–cell junction restoration
327 under bioenergetic challenge (see **Figure S8A**). In normal epithelial cells, SPS is inducible and supports
328 junctional reassembly, anchorage-dependent growth as monolayer, and epithelial barrier integrity[30,79]. In
329 contrast, transformed cells lose this pathway[79], enabling anchorage-independent growth and junctional
330 disassembly (**Figure S8B-C**). Notably, (re)activation of SPS permits anchorage-independent growth of
331 normal cells but not transformed cells[80]—highlighting its potential to discriminate between cell states and a

332 plausible reason for PF's ability to selectively target CSCs, but spare healthy stem cells. We found that PF
333 induced the SPS pathway, as evidenced by upregulation of a canonical biomarker pSer[245]GIV[30,79,80] (**Figure**
334 **6G**).

335 Next, we hypothesized that SPS reactivation preserves the function of normal stem cells while
336 selectively dismantling the survival programs of CSCs, which rely on SPS pathway silencing. To test this, we
337 analyzed expression scores for core transcriptional targets of oncogenic stemness in PF-treated healthy and
338 CRC PDOs. PF significantly reduced the expression of *Annexin A1* (**Figure 6H**), *MYC* (**Figure 6I**), and
339 transcriptional programs of Wnt/β-catenin (**Figure 6J**) and YAP (**Figure 6K**)—all of which were elevated in
340 *CDX2*-low CSC-rich PDOs but not in healthy- or *CDX2*-high CSC-poor PDOs.

341 Together, these findings demonstrate that CDX2-reinstatement reprograms CSCs by restoring
342 epithelial polarity through SPS pathway reactivation. This reactivation triggers the collapse of two key
343 junction-informed transcriptional programs—YAP and Wnt/β-catenin—on which CDX2-low CSCs are
344 critically dependent (see **Figure 7A-B** legend). This mechanism underpins the selectivity and efficacy of PF,
345 driving differentiation in all cells but apoptotic fate specifically in CSCs (**Figure 7C**).
346

347 ## Discussion

348 The central discovery of this study is a network-guided differentiation therapy that reinstates CDX2
349 expression to selectively eliminate CSCs in CRCs. Much like immunotherapy reawakens immune
350 surveillance, CDX2-restorative therapy reactivates a physiologic lineage commitment. We identify *PRKAB1*
351 as a key upstream node and show that PF-06409577, a clinically safe first-in-class agonist, triggers *CDX2*-
352 restoration, cell-type specific differentiation and anti-tumor responses. The therapeutic effect meets rigorous,
353 pre-defined benchmarks for efficacy: *CDX2* induction, a robust 50-gene companion biomarker signature,
354 suppression of stemness and survival pathways, and sparing of normal stem cells. Additional analyses of
355 human CRCs using a panel of treatment-perturbed genes promises a ~50% reduction in relapse and mortality
356 risk in *CDX2*-low CRCs.
357

358 *A Scalable Framework for Differentiation Therapy*:

359 Our use of the BoNE enabled the development of *CANDiT*, a rational, systems-level framework for inducing
360 differentiation in solid tumors. By using *CDX2*—a master regulator of intestinal lineage—as a seed node,
361 BoNE traced upstream regulatory hubs capable of reinstating epithelial identity while suppressing CSC-
362 associated programs in colorectal cancers. This approach is generalizable: master regulators of
363 differentiation in other tissues can similarly be used to drive lineage restoration, examples of which are
364 emerging[101].

365 Although *CDX2* is insufficient to fully activate the lineage program[75], our findings suggest that
366 PRKAB1 agonism by PF enables broader commitment by sustaining *CDX2* induction and engaging
367 additional, underexplored and incompletely mapped pathways essential for terminal differentiation. *CANDiT*
368 distinguishes itself from conventional screening methods by revealing hidden transcriptomic vulnerabilities in
369 dedifferentiated CSCs and pinpointing high-value targets to exploit them. In the PRKAB1-activated state,
370 *CDX2* plays a central—but not solitary—role in driving differentiation and CSC-specific death. Like *CDX2*,
371 *PRKAB1* is likely to be necessary, but not sufficient to impose such broader commitment to differentiation.

372 Our findings exemplify how *CANDiT* offers a scalable, mechanistically grounded blueprint for developing
373 lineage-based therapies across solid tumors, guided by tissue-specific master regulators.
374
375 *Mechanism of Selectivity*:
376 We show that CDX2-restorative therapy exploits a selective vulnerability in CSCs by reactivating the SPS
377 pathway, a circuit integrating bioenergetic stress with epithelial polarity in 3D growth contexts. While this
378 pathway is silenced in cancer, it remains functional in healthy tissues[78,79]. PRKAB1 agonism restores
379 epithelial polarity through SPS activation, disrupting anchorage-independent growth and silencing junction-
380 informed oncogenic programs (e.g., YAP, Wnt/β-catenin; **Figure 7A-B**), leading to a collapse of stemness
381 signaling in CSCs. In contrast, in normal epithelial and stem cells, SPS activation supports barrier function
382 and preserves homeostasis. This cell-state–specific response explains the therapeutic selectivity of PF,
383 enabling CSC-targeted killing while sparing healthy stem cells (see **Figure 7C**).
384 It is noteworthy that the Cancer Dependency Map (DepMap; https://depmap.org/portal/home/)
385 includes data for HCT116 and SW480, but not for key lines used here such as DLD1 and Caco2. Among the
386 cell lines represented, PF IC_{50} values do not correlate with CDX2 transcript abundance. Our mechanistic
387 data— highlighting the role of 3D context and anchorage-independent signaling in PF response— offer a
388 plausible explanation, aligning with DepMap's own push to include 3D IC_{50} datasets[102].
389
390 *Translational Potential and Clinical Prioritization*:
391 We advocate translating these findings into clinical trials with a focus on CDX2-low CRCs, especially those
392 with a high 50-gene response signature. CDX2-low status, as assessed by IHC, is prognostic in stage III/IV
393 disease and is frequently associated with *BRAF* mutations[103]. Conversely, *KRAS*^{mut} tumors tend to retain
394 CDX2, reflecting the role of *KRAS*^{mut} in promoting hyperplasia[104] rather than stemness[105]. This intriguing
395 distinction could be attributed to cell-type specific roles of *KRAS*/*BRAF*^{mut} reported recently[106].
396 However, KRAS status alone should not preclude PF therapy: both HCT116 and SW480 cells (CDX2-
397 low and *KRAS*^{mut}) were PF-sensitive. Notably, population studies confirm that a subset (~9%) of *KRAS*^{mut}
398 metastatic CRCs with low CDX2 have poor prognosis[103] and could benefit from differentiation therapy. Thus,
399 CDX2-low status and high therapeutic response signature, not KRAS or BRAF status, should drive clinical
400 candidacy. Because loss of *CDX2* is an adverse prognostic factor in stage IV and stage III/*BRAF*^{mut}
401 CRCs[74,107], these subgroups can be prioritized for *CDX2*-reinstatement therapy with PRKAB1 agonist as
402 single agent or together with other modalities.
403 Although *CDX2* is prognostic in stage II(MSS)[52,65,69] and stage III CRCs[52,108], caution is warranted
404 when combining PF with conventional chemotherapy. CDX2 expression is associated with MDR1
405 upregulation, which could contribute to chemoresistance[109,110]. Until further data are available, PF should be
406 prioritized as monotherapy or in combination with non-cytotoxic agents in patients with relapsed or metastatic
407 *CDX2*-low CRC. Finally, early-onset CRCs is a subgroup that show worse outcome (poorer DFS) than its
408 late-onset counterparts[111]. Because the *CDX2-PRKAB1* relationship is conserved in this subgroup (**Figure**
409 **S2A**), we propose including this group in future PF-focused trials.
410
411 *Implications for Pre-Neoplastic Disease:*

412 It has been recently shown that CDX2 loss plays a permissive role in CRC initiation, particularly along the
413 serrated pathway[112-114] and in inflammation-associated CRCs[115-117]. Mouse models have shown that *CDX2*-
414 deficient metaplastic cells can induce tumorigenesis non–cell-autonomously[118,119]. Our findings provide an
415 impetus for future studies to evaluate whether *PRKAB1* agonists can serve as prophylactic agents in high-
416 risk populations with pre-neoplastic lesions or chronic inflammation.
417

418 ## Conclusions

419 We present *CANDiT*, a machine learning–guided, network-based differentiation strategy that fills a critical
420 gap in cancer therapy. *CANDiT* differs fundamentally from traditional pipelines in three key ways: 1) Network-
421 transcriptomics and ML-guided target discovery; 2) Validation studies on network-vetted cell line and
422 xenograft models; 3) Validation in PDOs utilizing biomarker-driven stratification and adjusting for confounding
423 variables as done in clinical trial analysis. This strategy enables selective CSC targeting and spares normal
424 stem cells. Despite strong evidence for CDX2 as a prognostic biomarker[52,65,66,68,69,71,83,120,121], and a
425 compelling pharmacoeconomic justification (i.e., cost $ 50 000 per QALY) for a CDX2-stratified decision to
426 offer chemotherapy[122], its clinical utility has not been adopted into guidelines due to lack of prospective
427 validation. Our demonstration that CDX2 serves as both a predictive biomarker and therapeutic effector for
428 PF—a drug already deemed safe in humans (NCT02286882)— provides compelling justification for CDX2-
429 guided clinical trials. More broadly, *CANDiT* opens the door to rational, scalable differentiation therapy across
430 solid tumors.
431
432

433 ## Limitations of the study

434 The exact mechanism(s) linking PRKAB1, SPS activation, *CDX2* restoration and crypt differentiation was not
435 fully delineated. Future studies will examine whether SPS (re)activation reverses aberrant *CDX2* promoter
436 methylation[123,124]. The PDO cohort was limited in size but reflective of real-world diversity, capturing the
437 interplay between tumor differentiation, CDX2 expression, and *KRAS*^{mut} status; larger cohorts and
438 randomized clinical trials with RNA-seq endpoints is warranted. Epigenetic profiling and cell fate trajectory
439 analysis will be essential to uncover the basis for the divergent cell fates—differentiation versus apoptosis—
440 triggered by PF in normal vs CSCs.

Acknowledgements

We are grateful to the staffs of the UC San Diego HUMANOID™ Center of Research Excellence, for support in the logistical support in conducting the studies on human PDOs. This work was supported by the National Institutes for Health (NIH) grants UG3TR002968, UH3 TR002968, R01-CA238042, R01-CA100768 and R01-CA160911 (to P.G) and VA Merit awards 1 I01 BX003856-01A1 and 1 I01 BX004494-01 (M.B). S.S was supported in part by a American Association of Immunologists (AAI) Intersect Fellowship Program for Computational Scientists and Immunologists. SB was supported by SERB International Research Experience Fellowship (SIR/2022/001374) by Govt. of India. Other sources of support include: R01-AI155696, R01-AI141630 and UG3TR003355 (P.G), a Torrey Coast Foundation Award (P.G, M.B, J.Y), Curebound Foundation award 23DG04, and by three Padres Pedal the Cause awards #PTC2021, #PTC2022 and #PTC2024. P.G were also supported by the Leona M. and Harry B. Helmsley Charitable Trust. Funders had no role in study design or conclusions. Authors acknowledge the instrumentation resources at the US San Diego Agilent Center of Excellence in Cellular Intelligence and Debashis Sahoo (UC San Diego) for providing access to computational resources.

Author Contributions

SS and PG conceptualized the study; PG designed and analysed all parts of the study; JA, KP, AO, EV and CT acquired, expanded, biobanked and conducted all experiments with patient-derived organoids and CRC cell lines; VC conducted immunocytochemistry (IHC) and immunofluorescence (IF) and Hematoxylin and Eosin (H&E) staining studies; CRE conducted qPCR studies and assisted SS, VC and PG on key parts of the study; SB, EM, KP, SS and VC conducted IHC, IF and H&E quantifications. ST built and validated the network model; SS and PG conducted all other computational, statistical and bioinformatic analyses; SA and MB conducted the xenotransplantation studies; JRS and JY synthesized the pharmacophore; MB, JY and PG secured funding; SS and PG prepared all display items; PG wrote the first draft of the manuscript; all authors edited and approved the final version.

Conflicts of Interest Statement: The authors have declared that no conflict of interest exists.

Figures and Legends:

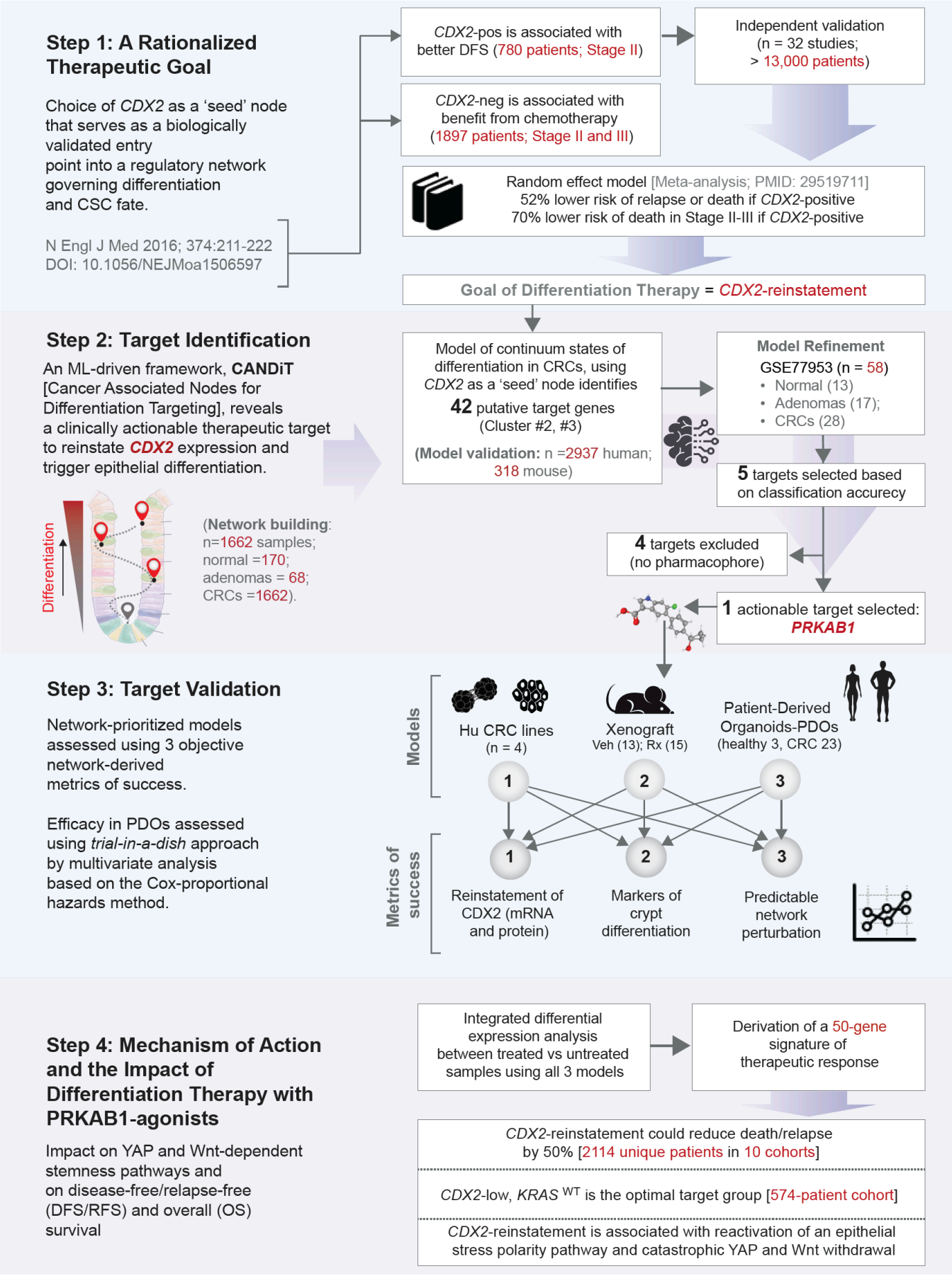


Figure 1. Study Design: Logical Network Perturbation to Reinvigorate CDX2 Expression as a Therapeutic Strategy in CRCs.

472 **Step 1**: Flowchart outlines the chronological accumulation of evidence supporting *CDX2*'s biological role as
473 a master regulator of intestinal epithelial differentiation and its consistent association with poor outcomes[76].
474 **Step 2**. Key steps in the network-based identification of PRKAB1 as a potent, actionable target, whose
475 agonism is predicted to restore CDX2 expression. See **Supplemental Information 1** for datasets used in
476 model training and validation, and **Supplemental Information 2** for gene clusters co-induced with *CDX2*
477 upon PRKAB1 activation.
478 **Step 3.** Three experimental models and objective criteria used to validate PRKAB1 agonism as a strategy
479 for CDX2 reinstatement and induction of cellular differentiation.
480 **Step 4**. Computationally driven insights into underlying mechanism of action and therapeutic impact. CDX2-
481 reinstatement may halve the risk of death/relapse and is associated with a catastrophic collapse of signals
482 that sustain cancer stemness---programs to which CSCs, but not other cancer cells are addicted--and with
483 the reactivation of an epithelial stress polarity pathway that is selectively lost in cancers.
484
485

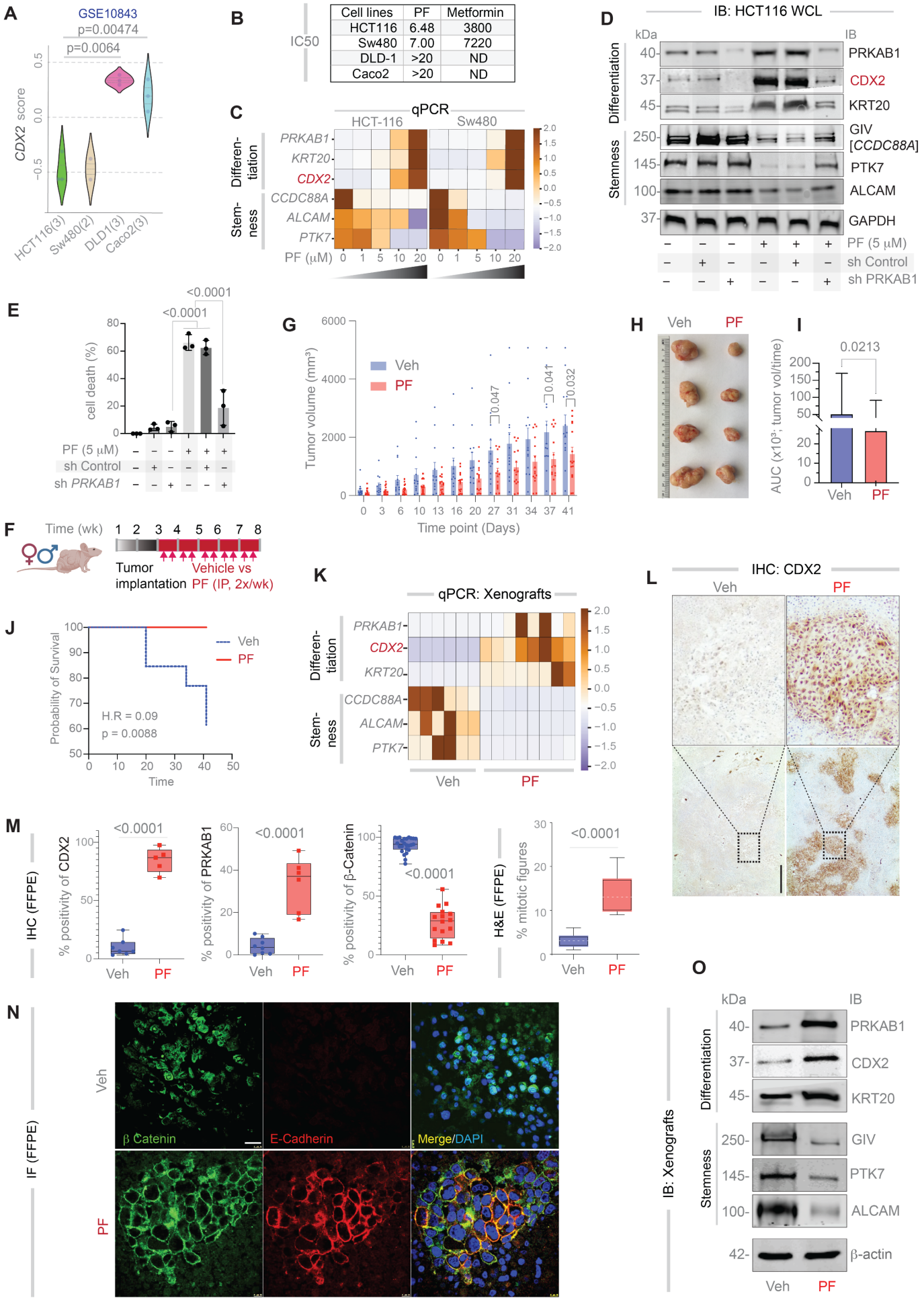


Figure 2. Target validation in CRC cell lines and xenograft models

488 **A**. Violin plot of normalized *CDX2* expression in CRC cell lines. *p*-values are calculated using Welch's *t*-test.
489 **B**. Table summarizing IC_{50} values (μM) for a specific (PF) or a non-specific (Metformin) PRKAB1-agonist in
490 CRC cell lines, measured by MTT assays at 48 h.
491 **C**. RT-qPCR analysis of differentiation and stemness markers in >95% viable cells (16 h post-treatment),
492 displayed as a Z-score heatmap.
493 **D-E**. Immunoblots (**D**) and MTT-based survival assays (**E**) showing the effects of PRKAB1 depletion (via
494 shRNA in HCT116 cells) on the anti-cancer efficacy of PF. *P*-values via one-way ANOVA; error bars
495 represent S.E.M. of three independent biological replicates. See **Figures S3** and **S4** for additional dose-
496 response studies using PF (**S3**, **S4A–H**) and Metformin (**S4I–K**).
497 **F**. Schematic of the xenograft experimental design and workflow.
498 **G**. Tumor growth kinetics in vehicle-treated (Veh; *n* = 13) vs. PF-treated (*n* = 15) mice, measured twice weekly
499 with calipers across three experimental cohorts. *P*-values by multiple paired *t*-tests with Benjamini, Krieger,
500 and Yekutieli correction.
501 **H**. Representative images of xenograft tumors at 8 weeks post-implantation.
502 **I**. Bar plots comparing area under the curve (AUC) for tumor growth: Veh (95% CI: 31,788–70,772) vs. PF
503 (95% CI: 16,949–36,404). *P*-value by unpaired two-tailed *t*-test; error bars indicate S.E.M.
504 **J**. Kaplan–Meier analysis of time to IACUC-mandated endpoints (euthanasia due to death or discomfort). *P*-
505 value was derived via log-rank test; H.R., hazard ratio.
506 **K**. RT-qPCR analysis of differentiation and stemness markers in xenografts displayed as a heatmap of Z-
507 score normalized values.
508 **L**. Representative images of CDX2 immunostaining in FFPE xenografts. Scale bar = 100 μm.
509 **M**. Quantification of CDX2, PRKAB1, and nuclear β-catenin (via IHC Profiler in ImageJ) and mitotic bodies
510 (manually counted in H&E sections). p-values are calculated by unpaired 2-tailed *t*-test. See **Figure S5B-D**
511 for images from representative fields.
512 **N**. Representative fields from FFPE tumors co-stained for E-cadherin (red), β-catenin (green) and DAPI
513 (nuclei; blue) are shown. Scale bar = 20 μm. See also **Figure S5K** for colocalization analysis using ImageJ.
514 **O**. Immunoblots from viable tumor lysates. See **Figure S5F** for quantification.

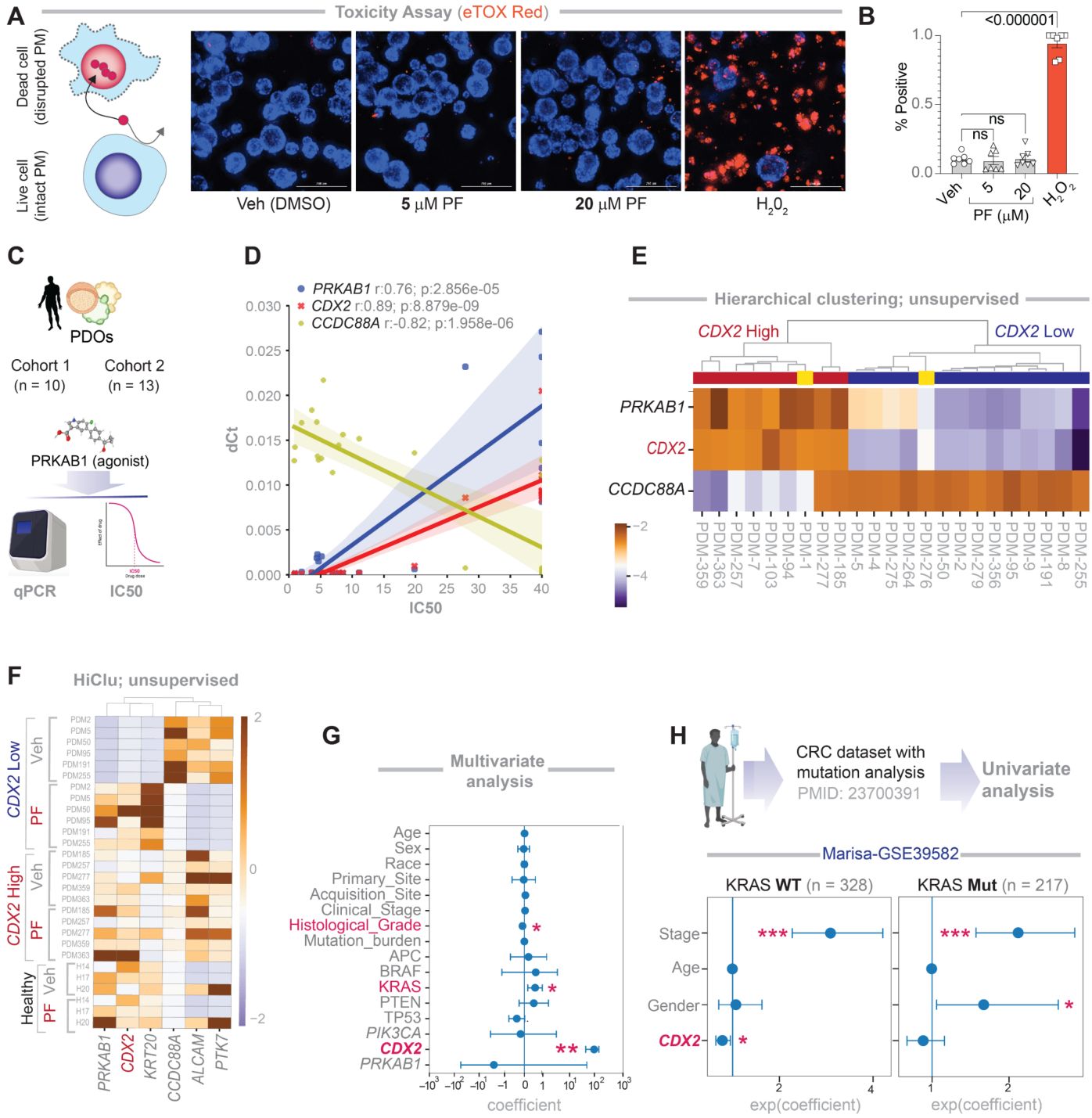


Figure 3. Differentiation and anti-cancer effects of PRKAB1 agonism are restricted to *CDX2*-low PDOs.

A. Schematic illustrating the principle of eTOX Red staining, which selectively labels dead cells based on increased membrane permeability (left). Representative immunofluorescence images show healthy PDOs treated with vehicle (control), 5 μM PF, 10 μM PF, or H_2O_2 (positive control for cell death) (right). Scale bar = 200 μm.

B. Quantification of the % of marker-positive (dead) organoids from IF images shown in panel A (right). Statistical comparisons between treatment groups were performed using one-way ANOVA.

525 **C**. Schematic of the experimental workflow showing sequential analyses conducted in two independent
526 patient cohorts. See **Supplemental Information 3** for patient demographics, tumor characteristics and
527 mutation status of key genes.
528 **D**. Correlation plots showing the relationship between IC_{50} values for PF and mRNA expression levels of 3
529 network-derived genes, as determined by qPCR.
530 **F**. Heatmap showing unsupervised clustering of CRC PDOs based on Z-score-normalized qPCR estimation
531 of the 3 genes in D.
532 **F**. Heatmap comparing expression profiles of stemness and differentiation markers in vehicle- versus PF-
533 treated CRC PDOs.
534 **G**. Multivariate regression analysis of IC_{50} values using a linear model incorporating all measured variables.
535 Bar plot shows the coefficient estimates (center values), 95% confidence intervals (error bars), and p-values
536 for each variable. The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no
537 effect). Red = statistically significant covariates. *$p \leq 0.05$; **$p \leq 0.01$. See **Supplemental Information 3** for
538 source data.
539 **H**. Univariate analysis of relapse-free survival (RFS) stratified by KRAS mutation status (WT vs. mutant)
540 using dataset GSE39582. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$.
541

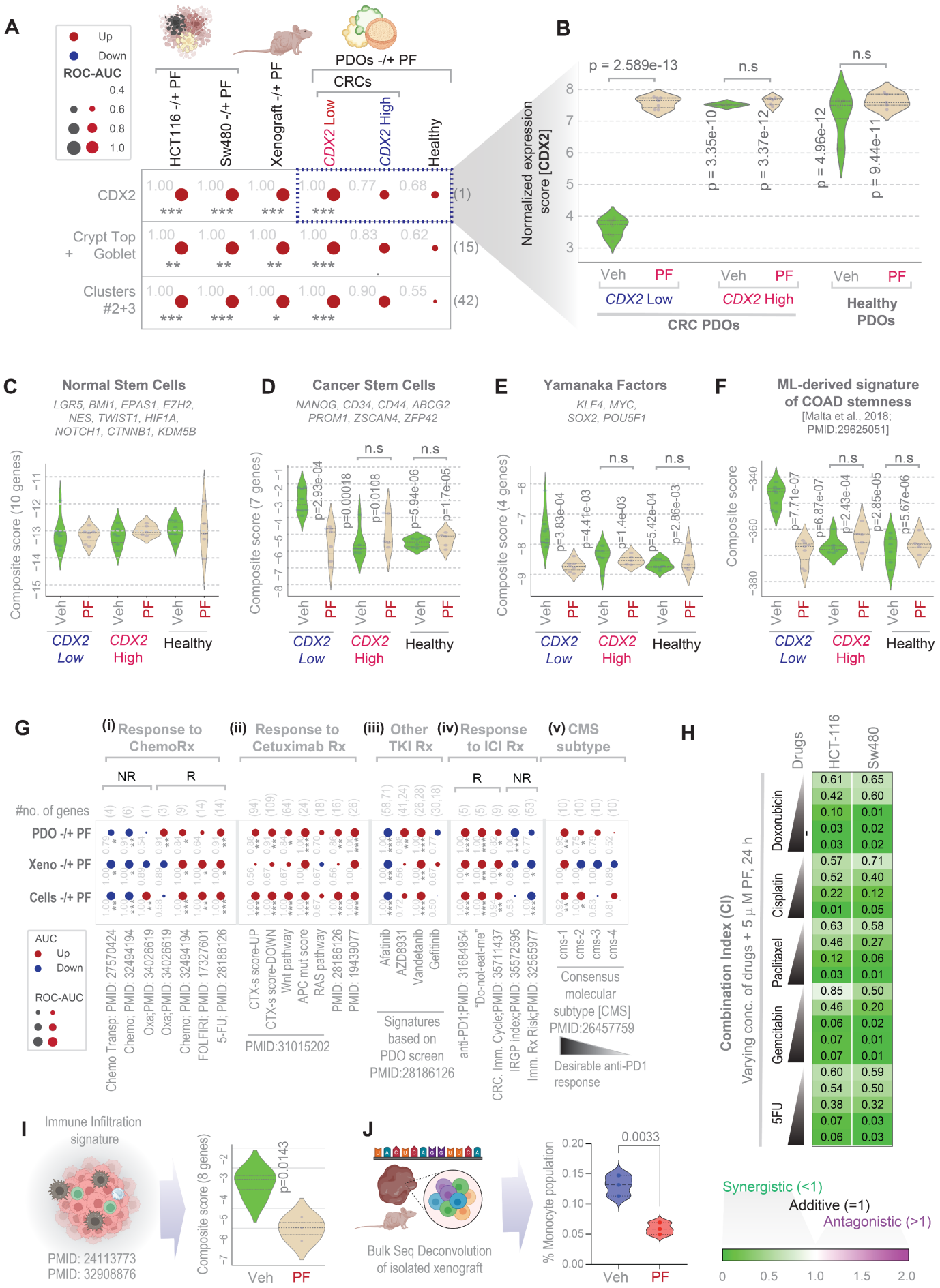


Figure 4. Indicators of therapeutic success, target-cell specificity, and potential for synergy with other treatment modalities.

545 **A**. Bubble plots display ROC-AUC values for the classification of vehicle (Veh) vs. PF-treated samples across
546 multiple models (HCT116, SW480, xenografts, and PDOs). Shown are *CDX2* expression patters (top row),
547 crypt-top and goblet cell markers (middle row), and genes from network clusters #2–3 (bottom row). Circle
548 size reflects the magnitude of the ROC-AUC; color indicates direction of regulation in treated samples (red =
549 upregulated, blue = downregulated). Statistical significance was calculated using Welch's t-test and
550 annotated using standard codes (''p > 0.1; '.'p ≤ 0.1; *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001). See **Figure S7** for
551 the same data shown as violin plots with actual p-values.
552 **B**. Violin plots show *CDX2* expression levels in PDOs. Comparisons are made against vehicle (Veh)-treated
553 *CDX2*-low PDOs using Welch's t-test.
554 **C-F.** Violin plots display the composite score of stemness markers derived from a recent pan-cancer machine
555 learning study[125]. Comparisons were made against vehicle (Veh)-treated *CDX2*-low PDOs using Welch's t-
556 test.
557 **G**. Bubble plots display ROC-AUC values for gene signatures associated with response (default or indicated
558 with "R") and non-response ("N.R.") to multiple therapeutic modalities, across CRC cell lines, xenografts, and
559 PDOs. Circle size reflects ROC-AUC; colors indicate direction of gene regulation. Significance was assessed
560 using Welch's t-test and annotated as in panel A.
561 **H**. Heatmap shows combination index (CI) values derived from MTT assays in HCT116 and SW480 cell lines
562 treated with a fixed concentration of PF (5 μM, 24 h) and increasing doses of chemotherapeutic agents. CI
563 values were calculated using a previously established formula[126].
564 **I**. Violin plot shows the composite score of an 8-gene immune infiltration signature[127] [derived from a large
565 pan-cancer study[128] (https://bioinformatics.mdanderson.org/estimate/) after training on outcome-annotated
566 datasets] in PF-treated and vehicle (Veh) treated xenografts. p-value was derived using Welch's t-test.
567 **J**. Violin plot shows the % of monocyte population in PF-treated vs. vehicle (Veh)-treated xenografts,
568 estimated via bulk RNA-seq deconvolution. p-value was derived using Welch's t-test.

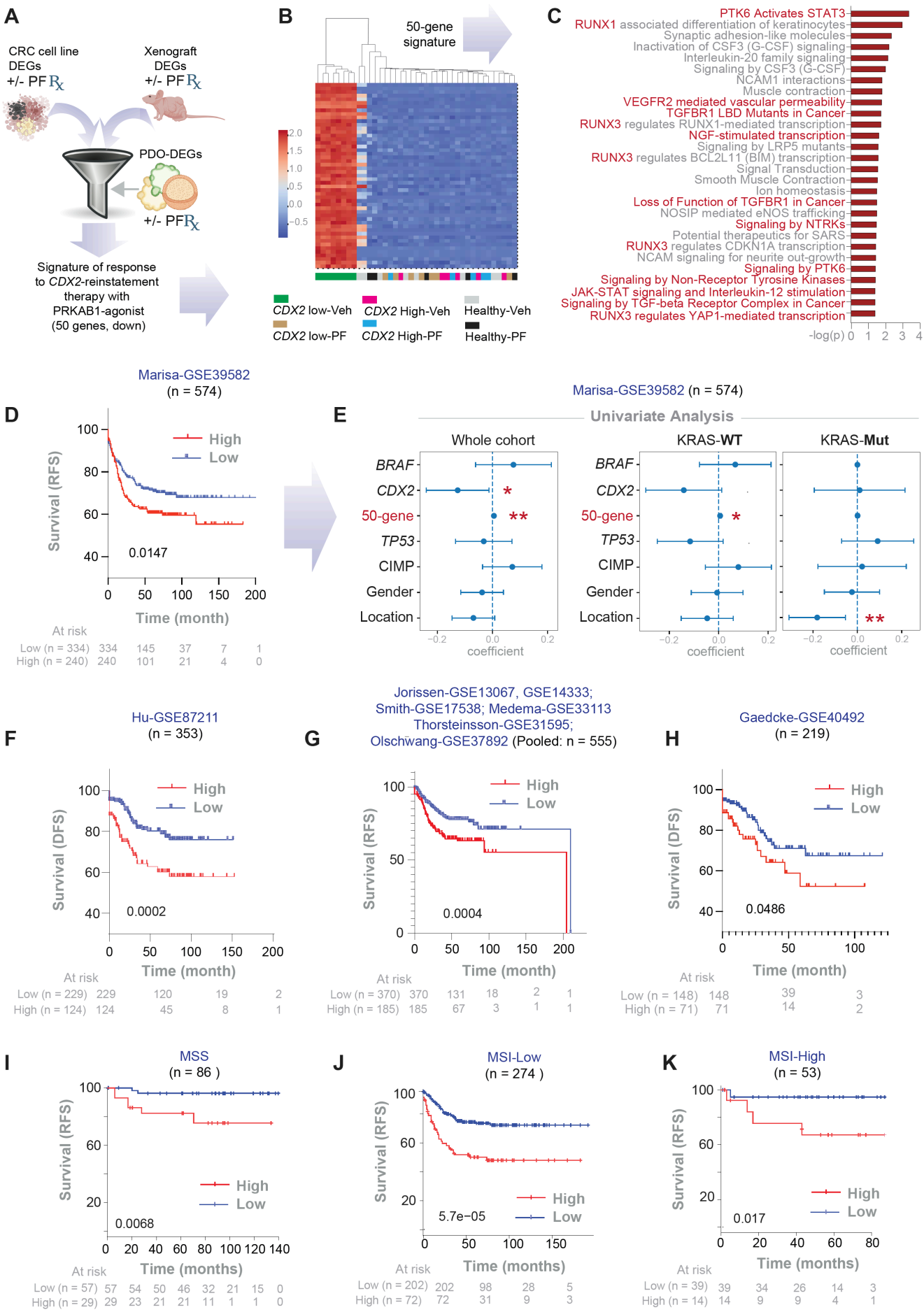


Figure 5. A 50-gene signature for tracking therapeutic response and estimating impact on survival.

572 **A**. Computational workflow used to derive a 50-gene signature of therapeutic response through an integrated
573 differential expression analysis (DEA) approach. Differentially expressed genes (DEGs) from CRC cell lines
574 and xenografts (see **Supplemental Information 4-6**) were further refined based on their ability to classify
575 accurately PDO samples treated with or without PF (see **Supplemental Information 6** for gene list).
576 **B**. Heatmap shows the results of unsupervised clustering of PDOs based on the 50-gene signature. Veh,
577 vehicle. See **Figure S8A-B** for violin plot visualizations.
578 **C**. Reactome pathway analysis of the 50-gene signature, highlighting key biological processes. See **Figure**
579 **S8C** for enrichment of KEGG pathways.
580 **D**, **F-H**. Kaplan-Meier (KM) plots show recurrence-free (RFS) or disease-free (DFS) survival in various CRC
581 patient cohorts, stratified by high vs. low expression scores of the 50-gene signature based on *StepMiner*
582 algorithm[82] computed within each cohort. p-values were calculated using the Log-rank test. For comparison,
583 see **Figure S8C** for KM plots on the same cohorts based on *CDX2* expression, computed using the *StepMiner*
584 algorithm[82] within each cohort.
585 **I-K.** KM plots show recurrence-free (RFS) survival in all *KRAS*^{WT} CRC patients available in KM-plotter
586 database[129], grouped by their microsatellite stability status. The patient groups are stratified by high vs. low
587 expression scores of the 50-gene signature based on *auto selected best percentile cutoff* algorithm used in
588 KM plotter[129]. p-values were calculated using the Log-rank test.
589 **E**. Univariate Cox regression analysis of RFS in a large CRC cohort. Plots display the coefficient for each
590 variable (center point) with corresponding 95% confidence intervals (error bars) and associated p-values.
591 The analysis includes the 50-gene signature along with other clinical variables. Left: entire cohort; middle:
592 KRAS wild-type CRCs; right: KRAS mutant CRCs. *$p \leq 0.05$; **$p \leq 0.01$.
593

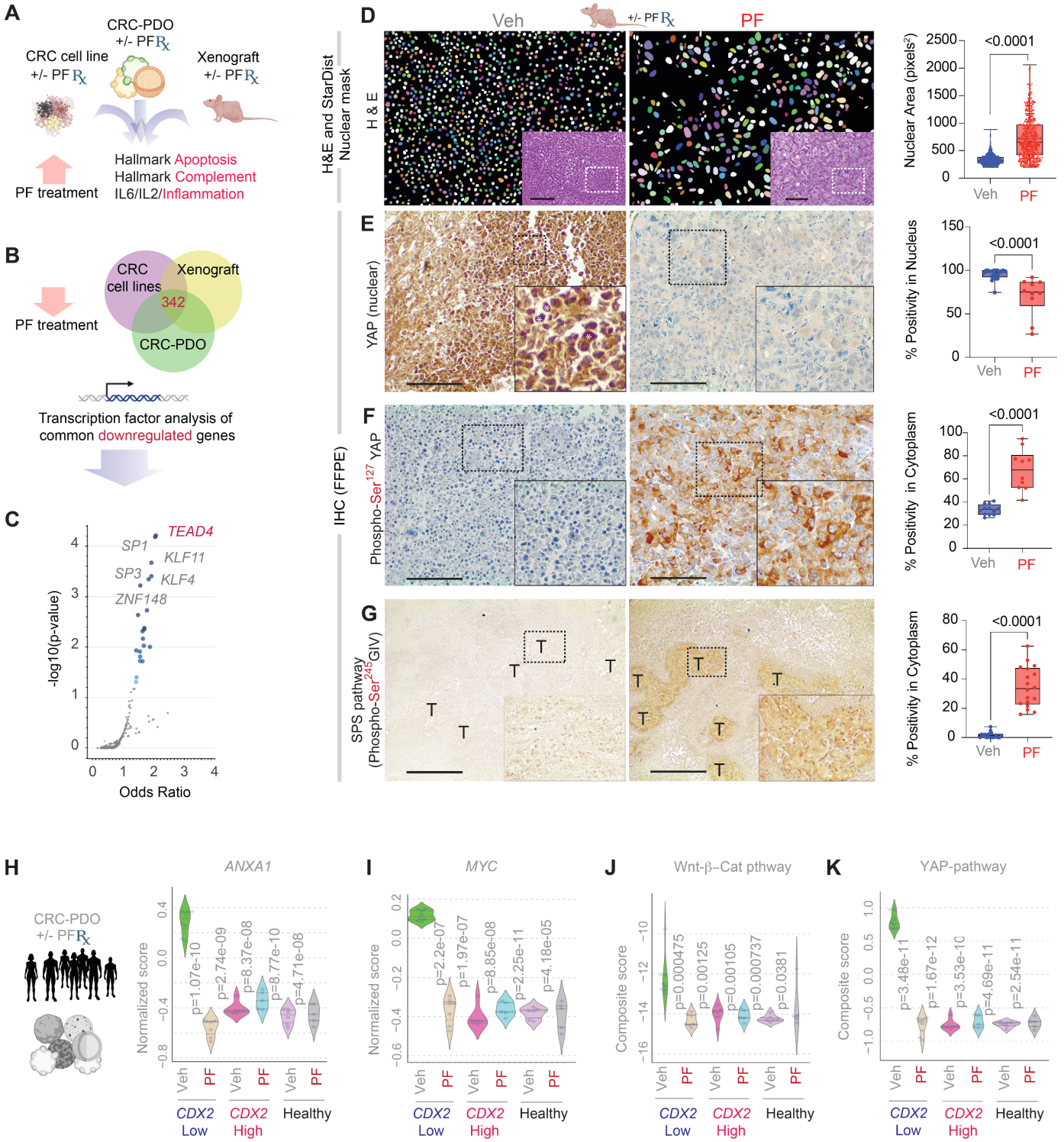


Figure 6. CDX2-reinstatement reprograms CRC stem cells by restoring epithelial polarity

A. Schematic summarizes top hallmark pathways identified using integrated differential expression analysis (DEA) approach that are induced upon PF treatment in three model systems.

B-C. Venn diagram (B) depicts downregulated genes common in all the three models used in this work. Transcription factor (TF) analysis (C) identified top TFs associated with common downregulated genes upon PF treatment in all the models.

603 **D**. Representative H&E-stained sections (left, with inset) of FFPE xenograft tumors from vehicle (Veh) and
604 PF-treated groups, with corresponding quantification of nuclear area (right) using StarDist-based
605 segmentation in ImageJ.
606 **E-G**. Immunohistochemistry (IHC) images (left) and quantification (right) of xenografts showing expression
607 of: nuclear YAP (E; white asterisk, inset); phospho-Ser^{127} YAP (F); and phospho-Ser^{245} GIV (a biomarker of
608 the SPS pathway[78-80]; see also **Figure S8A**). Scale bar = 100 μm.
609 **H-K**. Violin plots showing normalized expression scores in vehicle (Veh)- and PF-treated PDOs for
610 transcriptional markers of oncogenic stemness: Anexin (H), MYC (I), composite Wnt/β-catenin (J) and YAP
611 (K) target genes (listed in **Figure 7**-*left*). p-values were calculated using Welch's t-test, comparing each
612 sample to vehicle-treated *CDX2*-low PDOs.
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627

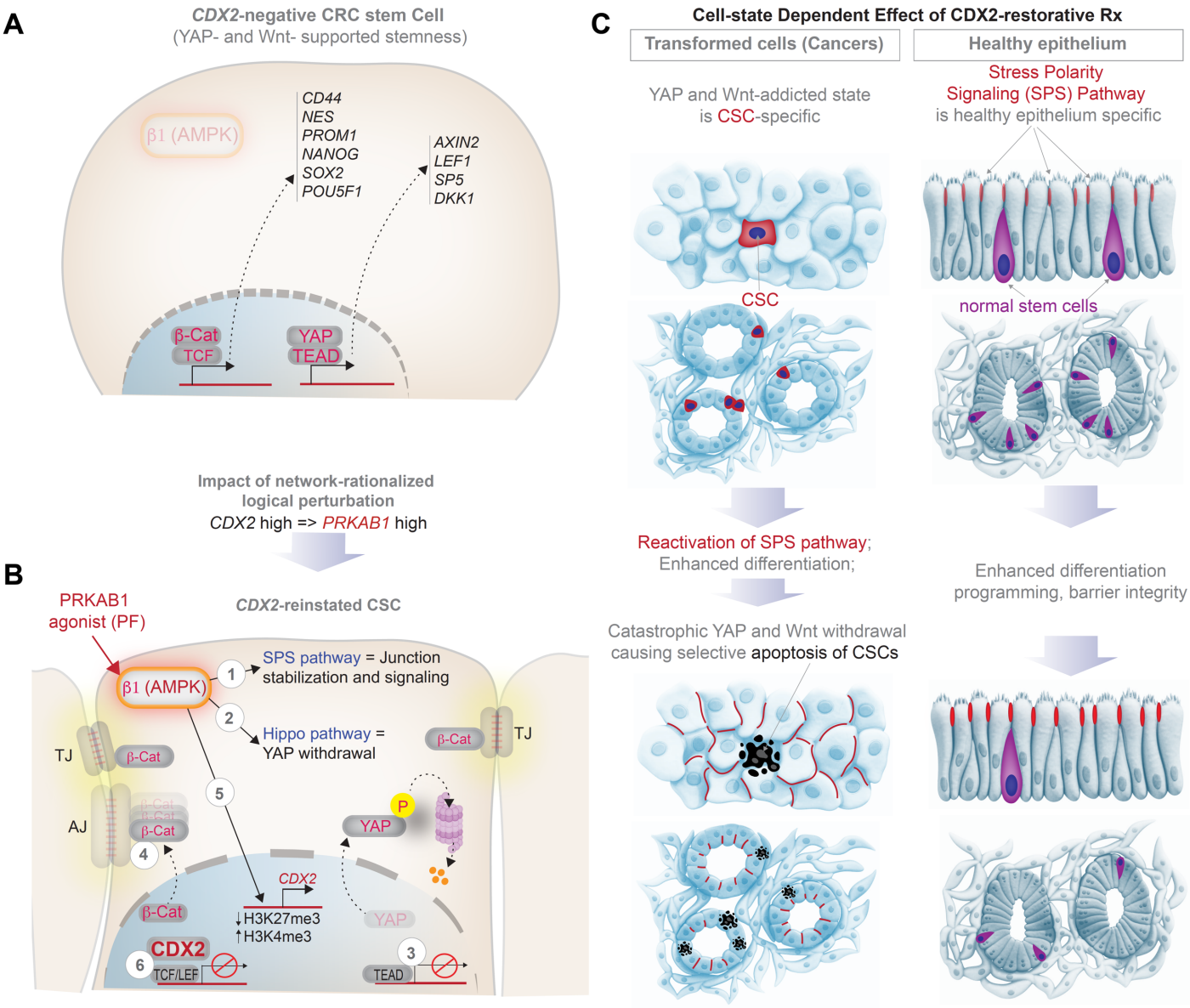


Figure 7. Summary and working model of how CDX2-reinstament selectively targets CRC stem cells.

A. CRC stem cells, but not normal stem cells, are characterized by enhanced transcription of Wnt/β-catenin and YAP target genes that are established markers of cancer stem cells. Both pathways are known to collaboratively enhance stemness in the colon[130,131].

B. PF treatment activates β1-specific AMPK, reactivates a tumor suppressor pathway (SPS pathway; 1) which attempts to restore tight- (TJ) and adherens (AJ) junctions in the setting of bioenergetic stress. Bioenergetic stress (AMPK activation) and junction restoration reshapes two junction-informed stemness programs, leading to their catastrophic withdrawal: *(i)* the Hippo-YAP pathway is *activated* (2), resulting in YAP phosphorylation followed by proteasomal degradation, and consequently, its nuclear exclusion (3) and withdrawal of the YAP/TEAD transcriptional program; *(ii)* junctional sequestration of β-Catenin and consequently, its nuclear exclusion (4) and termination of the β-Catenin•TCF/LEF transcriptional program. *CDX2*-reinstatement, and differentiation programs that accompanies the same, is likely via AMPK-dependent epigenetic events[123,124] (5). Once translated into protein, CDX2 is known to complete and displace β-catenin

644 from TCF/LEF complexes[132] (6), which further accentuates the impact of the SPS pathway on the β-
645 Catenin•TCF/LEF transcriptional program.
646 **C.** Schematic illustrates the cell state-specific impact of CDX2-reinstatement therapy and the impact of
647 restoration of the stress polarity signaling (SPS) pathway. Pre- (top) and post- (bottom) treatment states of
648 cancers (left) and healthy (right) epithelium are shown. Although activation of the SPS pathway enhances
649 differentiation and junction formation in both cell states, the apoptotic fate is selectively seen in *CDX2*-low
650 CSCs.
651
652

653 **Star Methods**

654 **Key Resource Table**

655

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples and Cell Lines		
HCT-116	ATCC	CCL-247
Sw480	ATCC	CCL-228
DLD1	ATCC	CCL-221
Caco2	ATCC	HTB-37
Athymic nude male mice	Jackson Laboratory	002019
PDM-255	ATCC	HCM-CSHL-0245-C18-A
PDM-2	ATCC	HCM-CSHL-0057-C18
PDM-4	ATCC	HCM-CSHL-0060-C18
PDM-50	ATCC	HCM-SANG-0273-C18
PDM-255	ATCC	HCM-CSHL-0245-C18-A
PDM-275	ATCC	HCM-CSHL-0245-C18-B
PDM-1	ATCC	HCM-CSHL-0056-C18
PDM-279	ATCC	HCM-CSHL-0384-D37
PDM-185	ATCC	HCM-CSHL-0238-C18
PDM-191	ATCC	HCM-CSHL-0246-C19
PDM-9	ATCC	HCM-CSHL-0065-C20
PDM-264	ATCC	HCM-CSHL-0382-C19
PDM-94	ATCC	HCM-CSHL-0141-C18
PDM-356	ATCC	HCM-CSHL-0257-C18
PDM-7	ATCC	HCM-CSHL-0063-C18
PDM-257	ATCC	HCM-CSHL-0248-C19
PDM-363	ATCC	HCM-CSHL-0426-C18
PDM-276	ATCC	HCM-CSHL-0461-D12
PDM-277	ATCC	HCM-CSHL-0247-C18
PDM-359	ATCC	HCM-CSHL-0317-C18
PDM-5	ATCC	HCM-CSHL-0061-C18
PDM-103	ATCC	HCM-CSHL-0164-C20
PDM-8	ATCC	HCM-CSHL-0064-C18
PDM-95	ATCC	HCM-CSHL-0142-C18
Human colon-derived healthy organoids	UC San Diego HUMANOID™ Center	n/a
Chemicals and Reagents		
eTox red	Agilent	Cat#8711009
Guava Cell Cycle Reagent	Millipore Sigma	Cat#SKU4500-0220
Dead Cell Apoptosis Kits with Annexin V	Thermo Fisher Scientific	Cat# V13242
TRIzol™ Reagent	Life Technologies	Cat#15596018
Quick-RNA MiniPrep Kit	Zymo Research	Cat#R2052
qScript™ cDNA SuperMix	QuantaBio	Cat#101414
PowerUp™ SYBR™ Green Master Mix	Applied Biosciences	Cat#A25741
Direct-zol RNA Miniprep Kit	Zymo Research	Cat#R1051
Zinc Formalin Fixative	Sigma-Aldrich	Cat#Z2902
DMEM/High glucose with L-glutamine; without sodium pyruvate	Cytiva	Cat#SH30022.FS
Fetal Bovine Serum	Sigma-Aldrich	Cat#F8067-500ML
Penicillin Streptomycin 100X Solution	Cytiva	Cat#SV30010
Trypsin-EDTA (0.25%), phenol red	Fisher Scientific	Cat#25-200-114
Dulbecco's Phosphate Buffered Saline (DPBS), 1X, without calcium, magnesium	Cytiva	Cat#16750-076
Matrigel	Corning	Cat#354234
Prochlorperazine dimaleate salt	Selleck Chemicals	Cat#S4631
Mayer's hematoxylin	BioGenex	Cat#HK1005K-GP
Two-component DAB brown	BioGenex	Cat#HK542XAKE-GPE

Collagenase I	GIBCO	Cat#17100-017
DNase I	Zymo Research	Cat#E1011-A
Human TruStain FcX™	BioLegend	Cat#422302
Poly-L-lysine coated slides	Fisher Scientific	Cat#63410-02
3% hydrogen peroxide	Fisher Scientific	Cat#S25359
Fluid Thioglycollate Medium	BD Difco	Cat#225650
2% paraformaldehyde	Electron Microscopy Services	Cat#15700
0.2% gluteraldehyde	Electron Microscopy Services	Cat#16200
0.1% Probumin® BSA	Millipore Sigma	Cat#820452
2% gelatin	Sigma-Aldrich	Cat# ES-006-B
Protease inhibitor cocktail	Roche Diagnostics	Cat#5056489001
Glutathione Sepharose beads	Sigma-Aldrich	Cat#17-0756-04
RPMI 1640 medium	Thermo Scientific	Cat#11875093
Gibco TrypLE express enzyme (1X)	Thermo Fisher Scientific	Cat#12563011
Intestigro™	HUMANOID	Cat#HUM2019-125
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat#D2650-100ML
Thiazolyl Blue Tetrazolium Bromide	Sigma-Aldrich	Cat#M5655-1G
Sodium Dodecyl Sulfate	Fisher Scientific	Cat#BP166-500
Dead Cell Apoptosis Kit with Annexin V Alexa Flour 488 & Propidium Iodide (PI)	Thermo Fischer Scientific	Cat#V13245
Chemicals and Reagents (For biosynthesis of AMPKβ1 agonist)		
1,4-Dibromobenzene	Fisher Scientific	Cat#AC112732500
Tetrahydrofuran (THF, anhydrous)	Fisher Scientific	Cat#T425-1
n-Butyllithium (n-BuLi, 2.5 M in hexane)	Fisher Scientific	Cat#AC213351000
Cyclobutanone	Fisher Scientific	Cat#AC110930250
Ammonium chloride (NH_4Cl, saturated aqueous)	Fisher Scientific	Cat#A2037300G
Ethyl acetate (EtOAc)	Fisher Scientific	Cat#AA31344M4
Brine (saturated NaCl solution)	Fisher Scientific	Cat#S271-10
Magnesium sulfate ($MgSO_4$)	Fisher Scientific	Cat#M65-3
Sodium tert-butoxide	Fisher Scientific	Cat#S0450100G
Potassium acetate	Fisher Scientific	Cat#P171-500
XPhos Pd G2 (palladium catalyst)	Sigma-Aldrich	Cat#741825-1G
XPhos (ligand)	Sigma-Aldrich	Cat#638064-1G
Tetrahydroxydiboron	Sigma-Aldrich	Cat#754242-25G
Argon gas	Matheson Tri-Gas	SKU: AR 200
5-Bromo-6-chloro-1H-indole	Combi-Blocks	Cat# QB-9262
4-Dimethylaminopyridine (DMAP)	Fisher Scientific	Cat#D1450100G
Pyridine	Fisher Scientific	Cat#P368-500
Trichloroacetyl chloride	Fisher Scientific	Cat#AAB2167518
Sodium methoxide (25% in methanol)	Fisher Scientific	Cat#AA46585AP
2-Methyltetrahydrofuran (2-MeTHF)	Fisher Scientific	Cat#AC396621000
Potassium carbonate (K_2CO_3, aqueous)	Fisher Scientific	Cat#P208-500
Sodium hydroxide (NaOH, 3 M aqueous)	Fisher Scientific	Cat#S318-1
Hydrochloric acid (HCl, 1 M aqueous)	Fisher Scientific	Cat#A144-212
Acetonitrile	Fisher Scientific	Cat#A998-4
Dichloromethane (DCM)	Fisher Scientific	Cat#D37-20
Trifluoroacetic acid (TFA, 0.1% in water for HPLC)	Fisher Scientific	Cat#AA44630AP

Antibodies		
Anti-CDX2 antibody	Abcam	Cat#76541
AMPK β1 Polyclonal antibody	Proteintech	Cat#10308-1-AP
Anti-pSer^{245}-GIV antibody	21st Century Biochemicals (Marlboro, MA)	Customized
Anti p-YAP antibody	Cell Signaling Technology	Cat#13008S
YAP antibody	Santa Cruz	Cat#101199
Anti-beta catenin	Santa Cruz	Cat#7963
Goat anti Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 647	Invitrogen	Cat#A-21245
F(ab')2-Goat anti-Rabbit IgG (H+L) Cross-Absorbed Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat#A-11072
Software Programs and Algorithms		
ImageJ	National Institute of Health	https://imagej.net/Welcome
Prism 9	GraphPad	https://www.graphpad.com/scientific-software/prism/
Molsoft	Molsoft, LLC	https://www.molsoft.com/index.html
LAS-X	Leica	www.leica-microsystems.com/products/microscope-software/p/leica-las-x-ls
Illustrator	Adobe	https://www.adobe.com/products/illustrator.html
ImageStudio Lite	LI-COR	https://www.licor.com/bio/image-studio-lite/
ClueGO	Cytoscape	https://academic.oup.com/bioinformatics/article/25/8/1091/324247
NetworkX	Python	https://networkx.org
NumPy	Python	https://numpy.org/
Pandas	Python	https://pandas.pydata.org/
Seaborn	Python	https://seaborn.pydata.org/
Gephi	Gephi	https://gephi.org
Photoshop	Adobe	https://www.adobe.com/products/photoshop.html
FlowJo	Flow Jo V10, BD BioSciences	RRID:SCR_008520
Deposited data		
GSE237624	Induction of CDX2 as a differentiation therapy in colorectal cancers [cell lines]	This work
GSE237625	Induction of CDX2 as a differentiation therapy in colorectal cancers [xenograft]	This work
GSE237623	Induction of CDX2 as a differentiation therapy in colorectal cancers [CRC organoids]	This work
Other		
6-well Tissue Culture Plate	Genesee Scientific	Cat# 25-105
12-well Tissue Culture Plate	CytoOne	Cat# CC7682-7512
Transwell Inserts (6.5 mm, 0.4 µm pore size)	Corning	Cat# 3470
Cell Scraper	Millipore Sigma	Cat# C5981-100EA
Millicell EZ Slide 8-Well Chamber	Millipore Sigma	Cat# PEZGS0816
Trypan Blue Stain	Invitrogen	Cat# T10282
70 µm Cell Strainer	Thermo Fisher Scientific	Cat# 22-363-548
Noyes Spring Scissors - Angled	Fine Science Tools	Cat# 15013-12
gentleMACS™ C Tubes	Miltenyi Biotec	Cat# 130-093-237

DAPI	Invitrogen	Cat# D1306 RRID: AB_2629482
Phalloidin, Alexa Fluor 594	Invitrogen	Cat# A12381 RRID: AB_2315633
Propidium Iodide	Invitrogen	Cat# V-13245 B
Countess II Automated Cell Counter	Thermo Fisher Scientific	AMQAX1000
Epithelial Volt-Ohm (TEER) Meter	Millipore	MERS00002
Automated TEER measurement system [REMS AutoSampler]	World Precision Instruments (WPI)	N/A
MESO QuickPlex SQ 120	MSD	N/A
Leica TCS SPE Confocal	Leica Microsystems	TCS SPE
Power Pressure Cooker XL	Tristar Products	N/A
Canon Rebel XS DLSR	Canon	N/A
MiniAmp Plus Thermal Cycler	Applied Biosystems	Cat# A37835
QuantStudio5	Applied Biosystems	Cat# A28140 RRID:SCR_020240
Light Microscope (brightfield images)	Carl Zeiss LLC	Axio Observer, Inverted; 491917-0001-000
Fisherbrand™ 150 Handheld Homogenizer	Fisher Scientific	Cat# 15340168
Spark 20M Multimode Microplate Reader	Tecan	N/A
NanoQuant Infinite M200	Tecan	N/A
Guava® easyCyte Benchtop Flow Cytometer	Millipore	Guava easyCyte 6 2L
MESO QuickPlex instrument	Mesoscale Discovery Inc.	SQ 120
gentleMACS™ Dissociator	Miltenyi Biotec	Cat# 130-093-235
Transwell Inserts (6.5 mm, 0.4 µm pore size)	Corning	Cat# 3470
Cell Scraper	Millipore Sigma	Cat# C5981-100EA
Millicell EZ Slide 8-Well Chamber	Millipore Sigma	Cat# PEZGS0816
Trypan Blue Stain	Invitrogen	Cat# T10282
70 µm Cell Strainer	Thermo Fisher Scientific	Cat# 22-363-548
Noyes Spring Scissors - Angled	Fine Science Tools	Cat# 15013-12
16% Paraformaldehyde	Electron Microscopy Sciences	Cat#15710
ProLong Glass Antifade Mountant	Thermo Fisher Scientific	Cat#P36984
2.5% Normal Goat Serum	Vector Laboratories	Cat#S-1012-50
2.5% Normal Horse Serum	Vector Laboratories	Cat#S-2012-50

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Pradipta Ghosh, prghosh@ucsd.edu.

Materials Availability

This study has generated CRC organoid biobank, RNA and DNA from the organoids. These materials are available from the lead contact with a completed Materials Transfer Agreement and patented technology agreement following the guidelines of the University of California, San Diego.

Data and Code Availability

All transcriptomic datasets generated in this work has been deposited at NCBI GEO [GSE237623, GSE237624, GSE237625]. The data underlying all the figures and tables are available in the article and its online supplemental materials. The Boolean analysis framework is available at https://github.com/sinha7290/Prodiff.

Method Details

Computational Approaches

Data Processing: We built a human colon tissue database (n = 1,911) derived from the previously published "Human Colon Global Database," refined by filtering for EpCAM and Albumin expression and restricted to the Affymetrix Human U133 Plus 2.0 platform (GPL570). The dataset was obtained from the NCBI Gene Expression Omnibus (GEO) repository[133-135], as described previously[52], and was expanded by including 68 additional adenoma samples and FACS-purified human colon crypt samples (GSE31255). This dataset contained experiments from 23 independent NCBI-GEO data-series (GSEs) from the same platform (GPL570). A list of the 23 NCBI-GEO GSEs contained within the "Human colon tissue database" is provided in (**Supplemental Information 1**)**.** Then, the dataset was prepared for Boolean analysis by selecting genes that had the best dynamic range (identified by the biggest percentile range between the 10th and 90th percentile of low and high expression value based on the *StepMiner*[82] threshold for each gene in this dataset). All training and validation dataset (**Supplemental Information 1**) were downloaded from the NCBI GEO website[133-135]. All gene expression datasets (**Supplemental Information 1**) were processed separately using the Hegemon data analysis framework[40,51,52]. We did not combine datasets that belong to two different platforms.

Boolean Analysis: *Boolean logic* is a simple mathematical relationship of two values, i.e., high/low, 1/0, or positive/negative. The Boolean analysis of gene expression data requires first the conversion of expression levels into two possible values. Here, *StepMiner*[82] algorithm is used to perform Boolean analysis of gene expression data[7]. StepMiner is an algorithm designed to detect stepwise transitions in time-series gene expression data[82]. It fits step functions to expression profiles by identifying the sharpest changes in signal, corresponding to gene expression switching events. The algorithm evaluates all possible step positions and calculates the average expression on either side of each step to define constant segments. An adaptive regression approach is then used to select the step position that minimizes the sum of squared errors between the observed and fitted data. The selected step is used as the *StepMiner* threshold. *This* threshold is used to convert gene expression values into Boolean values. A noise margin of 2-fold change is applied around the threshold to determine intermediate values, and these values are ignored during Boolean analysis. Finally, a regression test statistic is computed to assess the significance of the identified step transition as follows:

$$F\ stat = \frac{\sum_{i=1}^{n}(\hat{X}_i - \bar{X})^2/(m-1)}{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2/(n-m)}$$

Where X_i for $i = 1$ to n are the values, $\hat{X}_i$ for $i = 1$ to n are fitted values. m is the degrees of freedom used for the adaptive regression analysis. $\bar{X}$ is the average of all the values: $\bar{X} = \frac{1}{n} * \sum_{j=1}^{n} X_j$. For a step position at k, the fitted values $\hat{X}_i$ are computed by using $\frac{1}{k} * \sum_{j=1}^{n} X_j$ for $i = 1$ to k and $\frac{1}{(n-k)} * \sum_{j=k+1}^{n} X_j$ for $i = k+1$ to n.

In a scatter plot, there are four possible quadrants based on Boolean values: (low, low), (low, high), (high, low), (high, high). In all our experimental settings, we applied the same logic to classify samples into CDX2-high and CDX2-low groups based on the dynamic range of CDX2 expression levels.

Invariant Boolean implication relationships: A Boolean implication relationship is observed if any one of the four possible quadrants or two diagonally opposite quadrants are sparsely populated. Based on this rule, there are six different kinds of Boolean implication relationships. Two of them are symmetric: equivalent (corresponding to the highly positively correlated genes), opposite (corresponding to the highly negatively correlated genes). Four of the Boolean relationships are asymmetric, and each corresponds to one sparse quadrant: (low => low), (high => low), (low => high), (high => high). BooleanNet statistics (Equations listed below) is used to assess the sparsity of a quadrant and the significance of the Boolean implication relationships [32,50] . Given a pair of genes A and B, four quadrants are identified by using the *StepMiner* thresholds on A and B by ignoring the Intermediate values defined by the noise margin of 2-fold change (+/- 0.5 around *StepMiner* threshold). Number of samples in each quadrant are defined as a_{00}, a_{01}, a_{10}, and a_{11}. Total number of samples where gene expression values for A and B are low is computed using following equations.

$nA_{low} = (a_{00} + a_{01}), nB_{low} = (a_{00} + a_{10}),$

Total number of samples considered is computed using following equation.

$total = a_{00} + a_{01} + a_{10} + a_{11}$

Expected number of samples in each quadrant is computed by assuming independence between A and B. For example, expected number of samples in the bottom left quadrant $e_{00} = \hat{n}$ is computed as probability of A low ((a_{00} + a_{01})/total) multiplied by probability of B low ((a_{00} + a_{10})/total) multiplied by total number of samples.

Following equation is used to compute the expected number of samples for the low-low quadrant.

$n = a_{ij}, \hat{n} = (nA_{low}/total * nB_{low}/total) * total$

To check whether a quadrant (here low-low) is sparse, a statistical test for ($e_{00} > a_{00}$) or ($\hat{n} > n$) is performed by computing S_{00} and p_{00} using following equations. A quadrant is considered sparse if S_{00} is high ($\hat{n} > n$) and p_{00} is small.

$$S_{ij} = \frac{\hat{n} - n}{\sqrt{\hat{n}}}$$

$$p_{00} = \frac{1}{2}\left(\frac{a_{00}}{(a_{00} + a_{01})} + \frac{a_{00}}{(a_{00} + a_{10})}\right)$$

A threshold of S_{00} > sthr and p_{00} < pthr to check sparse quadrant. A Boolean implication relationship is identified when a sparse quadrant is discovered using following equation.

Boolean Implication = (S_{ij} > sthr, p_{ij} < pthr)

A relationship is called Boolean equivalent if top-left and bottom-right quadrants are sparse.

Equivalent = $(S_{01} > sthr,\ P_{01} < pthr, S_{10} > sthr, P_{10} < pthr)$

Boolean opposite relationships have sparse top-right (a_{11}) and bottom-left (a_{00}) quadrants.

$$Opposite = (S_{00} > sthr,\ P_{00} < pthr, S_{11} > sthr, P_{11} < pthr)$$

Boolean equivalent and opposite are symmetric relationship because the relationship from A to B is same as from B to A. Asymmetric relationship forms when there is only one quadrant sparse (A low => B low: top-left; A low => B high: bottom-left; A high=> B high: bottom-right; A high => B low: top-right). These relationships are asymmetric because the relationship from A to B is different from B to A. For example, A low => B low and B low => A low are two different relationships.

A low => B high is discovered if bottom-left (a_{00}) quadrant is sparse and this relationship satisfies following conditions.

A low => B high = $(S_{00} > sthr,\ P_{00} < pthr)$

Similarly, A low => B low is identified if top-left (a_{01}) quadrant is sparse.

A low => B low = $(S_{01} > sthr,\ P_{01} < pthr)$

A high => B high Boolean implication is established if bottom-right (a_{10}) quadrant is sparse as described below.

A high => B high = $(S_{10} > sthr,\ P_{10} < pthr)$

Boolean implication A high => B low is found if top-right (a_{11}) quadrant is sparse using following equation.

A high => B low = $(S_{11} > sthr,\ P_{11} < pthr)$

For each quadrant, a statistic S_{ij} and an error rate p_{ij} is computed. $S_{ij} > 10$ and $p_{ij} < 0.15$ are the thresholds used on the BooleanNet statistics to identify Boolean implication relationships (BIRs). False discovery rate is computed by randomly shuffling each gene and computing the ratio of the number of Boolean implication relationship discovered in the randomized dataset and original dataset. The false discovery rate for our dataset was less than 0.001.

Boolean analysis on the large human colon tissue database (n = 1911; Global 2018 GPL570 Colon Purged), uses a threshold of sThr = 10 and pThr = 0.15. We increased sThr and relax the pThr to focus on reasonable number of gene candidates. Boolean Implication analysis looks for invariant relationship across all the different types of samples regardless of the conditions and treatment protocols. Therefore, it does not distinguish the sample types when discovering Boolean implication relationships. We assume that there are fundamental invariant Boolean implication formula that are satisfied by every sample regardless of their type.

<u>CANDiT (*Cancer Associated Nodes for Differentiation Targeting*)</u>: We developed a machine learning framework, *CANDiT*, designed to identify actionable nodes that drive differentiation within cancer networks to guide therapeutic intervention. *CANDiT* is agnostic to tissue of origin and begins with a single 'seed' gene that meets two essential criteria: (i) it must be a lineage-defining transcriptional determinant along the stemness-to-differentiation axis in the cancer's tissue of origin, and (ii) it must hold prognostic relevance in human tissue datasets, linking its expression to patient outcomes. Thus, the seed gene must bridge two levels of biological scale—cellular differentiation and clinical prognosis.

CANDiT constructs a Boolean implication network (BIN) by identifying all statistically significant pairwise Boolean implication relationships (BIRs)[8] within a curated expression compendium, the Human

786 Colon Global Database, as previously described[52]. The resulting BIN is a directed graph in which nodes
787 represent genes and edges represent one of six possible Boolean relationships (e.g., high ⇒ high, high ⇒
788 low, low ⇒ high, low ⇒ low, Equivalent, Opposite). While Equivalence relationships often form a scale-free
789 network, other asymmetric BIRs do not.
790 Boolean analysis was performed on genes with sufficient dynamic range to permit clear binary
791 classification (high/low) via the *StepMiner* algorithm. Based on over a decade of experience with Boolean
792 implication methods, BIRs are most reliable in datasets with >200 samples[136]. Our dataset was prepared for
793 Boolean analysis by filtering genes that had a reasonable dynamic range of expression values. Genes with
794 fewer than 5% of samples in either the high or low state were excluded to reduce noise and improve
795 robustness. This filtering step ensures the resulting network highlights biologically meaningful and robust
796 relationships.
797
798 Generation of Clustered Boolean Implication network: Clustering was performed in the Boolean implication
799 network to dramatically reduce the complexity of the network. A clustered Boolean implication network (CBIN)
800 was created by clustering nodes in the original BIN by following the equivalent BIRs. One approach is to build
801 connected components in an undirected graph of Boolean equivalences. However, because of noise, the
802 connected components become internally inconsistent e.g., two genes opposite to each other become part
803 of the same connected component. In addition, the size of clusters became unusually big with almost
804 everything in one cluster. To avoid such a situation, we need to break the component by removing the weak
805 links. To identify the weakest links, we first computed a minimum spanning tree for the graph and computed
806 the Jaccard similarity coefficient for every edge in this tree. Ideally if two members are part of the same
807 cluster, they should share as many connections as possible. A threshold of 0.7 is considered for the Jaccard
808 similarity coefficient and if they share less than 70% of their total individual connections (Jaccard similarity
809 coefficient less than 0.7) the edges are dropped from further analysis. Thus, many weak equivalences were
810 dropped using the above algorithm leaving the clusters internally consistent. We removed all edges that have
811 Jaccard similarity coefficient less than the selected threshold and built the connected components with the
812 rest. The connected components were used to cluster the BIN which is converted to the nodes of the CBIN.
813 The choice of the threshold on the Jaccard similarity coefficient play an important role in determining the size
814 and the number of clusters as well as whether they are internally consistent. A new graph was built that
815 connected the individual clusters to each other using Boolean relationships. The link between two clusters
816 (A, B) was established by using the top representative node from A that was connected to most of the
817 members of A and sampling 6 nodes from cluster B and identifying the overwhelming majority of BIRs
818 between the nodes from each cluster.
819 Here, CBIN was created using our pooled dataset (human colon tissue database (n = 1911) to capture
820 the differentiation events in colon tissue (**Figure S1C**). The edges between the clusters represented the
821 Boolean relationships that are color-coded as follows: orange for low => high, dark blue for low => low, green
822 for high => high, red for high => low, light blue for the equivalent and black for the opposite. A subnetwork is
823 selected using low=>low (blue), high => low (red) and opposite (merged with high=>low as red) edges among
824 the top 10 clusters (**Figure S1D**).
825

Charting Boolean paths: Boolean paths have been previously leveraged to predict the underlying time series events in biological processes such as B cell differentiation[32,33] and early differentiation events in cancer stem cell[40,43,51,52]. The core algorithm enabling this, *MiDReG* (Mining Developmentally Regulated Genes), uses two seed genes infer intermediate genes and transitional states within a developmental or differentiation process. *MiDReG* identifies these intermediates through a series of asymmetric BIRs [10].

In this study, we apply the *MiDReg* framework to traverse the clustered Boolean Implication network, identifying gene clusters whose start and end points may define biologically meaningful differentiation trajectories in colon tissue. The asymmetric BIRs offer a directional, causal logic that distinguishes this network from conventional co-expression or correlation-based gene networks.

A Boolean path refers to a directed sequence of BIR-linked nodes (genes) within the BIN. A simple Boolean path includes two nodes and a single directed edge, while a complex Boolean path involves multiple nodes and edges, representing multi-step transitions between regulatory states. These paths allow for a high-resolution mapping of potential developmental hierarchies or transitions embedded in steady-state transcriptomic data.

Ordering samples based on composite scores along Boolean paths: Each Boolean path comprises one or more clusters. To assign a composite score to each sample, we first computed a normalized average expression value for the genes within each cluster. Normalization was performed using a modified Z-score centered around the *StepMiner* threshold (SThr), defined as: (formula = (expr - SThr)/3*stddev).

Next, a weighted linear combination of these cluster averages was used to generate a final composite score for each sample. Weights along the Boolean path were assigned to either monotonically increase or decrease—ensuring that the sample ordering remained consistent with the logical sequence dictated by the asymmetric BIRs. Clusters highly expressed in disease samples were assigned a weight of +1, while those predominant in healthy samples received a weight of -1. The directionality of the Boolean path—from healthy to disease states—guided this weighting scheme. Samples were then ordered based on their final composite score, enabling biologically meaningful stratification.

The genes contained within each cluster of the Boolean network are detailed in **Supplemental Information 2**.

Measurement of Classification Strength or Prediction Accuracy: To evaluate classification strength and prediction accuracy, Receiver Operating Characteristic (ROC) curves were generated for each gene. These curves assess the performance of a binary classifier, e.g., high vs. low *StepMiner* normalized gene expression levels, across varying discrimination thresholds. ROC curves plot the True Positive Rate (TPR) against the False Positive Rate (FPR) at multiple threshold levels. The Area Under the Curve (AUC) quantifies the classifier's ability to correctly distinguish between the various types of samples, in this instance, CRC or healthy and treated or not with drug. ROC-AUC values were computed using the Python Scikit-learn package.

Correlation Coefficients Analysis: Linear correlation analysis was performed between ΔCT values of *CCDC88A*, CDX2 and *PRKAB1* of all the CRC PDOs with their IC50 values. Regression analysis was

performed using SciPy Stat package. Scatter plot with regression lines were generated using a Seaborn package (0.12.1).

Integrated Differential Expression Analysis: Differential gene expression analysis was performed using RNAseq raw count data processed through DESeq2[137]. Genes with an absolute log2 fold change [FC] > 10 and an adjusted p-value <0.05 were considered differentially expressed genes (DEGs). Pathway analysis of gene lists was carried out via the Reactome database and algorithm [138] . Reactome identifies signalling and metabolic molecules and organizes their relations into biological pathways and processes. A complete catalogue of DEGs and reactome pathway enrichments are provided in **Supplemental Information 4** and **Supplemental Information 5**, respectively. The 50-gene signature that emerged from the integration of all DEGs is provided in **Supplemental Information 6**.

Survival Analysis: *StepMiner* based survival analysis was performed by dividing the patients into two groups, based on high or low levels of expression of the 50-genes signature of therapeutic response [composite score expression values], as determined by implementing the *StepMiner*-derived threshold for each cohort (https://github.com/sinha7290/Prodiff). Log-rank analysis and visualization were performed using GraphPad Prism version 9.1. Default settings of GraphPad Prism ignore events at time zero (t0) for the assessment of the impact of treatment modalities on outcome (https://www.graphpad.com/support/faq/events-deaths-at-time-zero-in-survival-analysis/). These default settings were manually modified to include all samples (without exception) in the cohort for which outcome was known. Survival analysis on $KRAS^{WT}$ patient groups based on microsatellite stability status was performed using KM-plotter[129] based on *auto selected best percentile cutoff* of 50 genes.

Univariate and Multivariate Analyses: To assess which factor(s) may influence the IC50 of the PDOs, multivariate regression has been performed over all the clinical conditions. Here, the statsmodels module from python has been used to perform Ordinary least-squares (OLS) regression analysis of each of the variables. The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect).

We also performed univariate analysis on GSE39582 using the same OLS regression analysis to assess the impact of each variables individually, such as tumor stage, age, gender, tumor location, CIMP status, mutation status (*BRAF* and *TP53*) and gene expression [either *CDX2* alone or the composite score of the 50-gene signature of therapeutic response, both converted into a binary value, high vs low, as determined by *StepMiner* threshold] on survival. The analysis was done on the entire cohort, as well as on *KRAS* WT and *KRAS* mutant patients as two independent sub-cohorts. Both the univariate and multivariate analysis data are represented based on the coefficient values of associated variables and their upper and lower bounds of 95% confidence interval as the error bars using Matplotlib. The significance of the variables was determined by t-test; * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$.

Unsupervised Clustering and Heatmap: Expression patterns of the genes that are differentially expressed in *CDX2*-low and -high groups, with or without PF treatment, are clustered without bias based on their z-

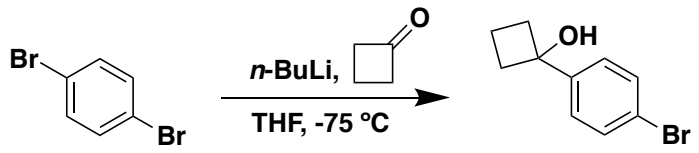
normalized cpm expression values, in all the samples. The data is visualized using the seaborn clustermap package (v 0.12) in python.

Bulk RNAseq Deconvolution: *In silico* deconvolution of bulk RNA-seq data from xenograft models was performed using the Granulator R package[139] to estimate monocyte abundance. Cell-type abundance estimates were normalized using the immune cell signature matrix developed by Monaco et al[140].

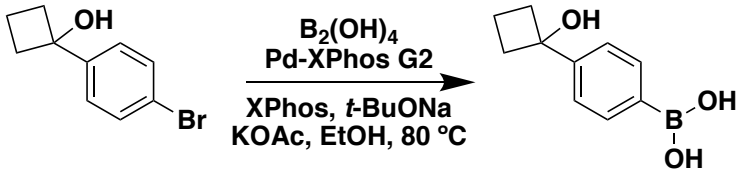
Statistical Tests: Optimal gene expression cut-off values were determined using the *StepMiner* algorithm within each individual dataset[82]. Gene signatures were used to classify sample categories, and multi-class classification performance was evaluated using ROC-AUC (receiver operating characteristic area under the curve) values. Statistical comparisons were performed using Welch's two-sample t-test (unpaired, unequal variance and sample size) via the scipy.stats.ttest_ind function in Python (version 0.19.0; equal_var=False). Multiple hypothesis correction was applied using the Benjamini-Hochberg method (fdr_bh) via statsmodels.stats.multitest.multipletests. Differential expression analysis was conducted using DESeq2 in R (version 1.16.1). Kaplan–Meier survival analysis and log-rank tests for statistical significance were performed using GraphPad Prism 9.1. Violin and bubble plots were generated using Seaborn (Python, version 0.12). Boolean implication relationships between genes were determined using the BooleanNet algorithm (criteria: statistic > 3 and error rate < 0.1). Power calculations estimated effect sizes (mean difference divided by standard deviation) of 2.5 in PDOs and 1.0 in animal experiments for CDX2 induction using the PRKAB1 agonist. A minimum of 4 PDOs and 17 animals per group was determined to provide adequate power (α = 0.05, power = 0.80).

Experimental Approaches:

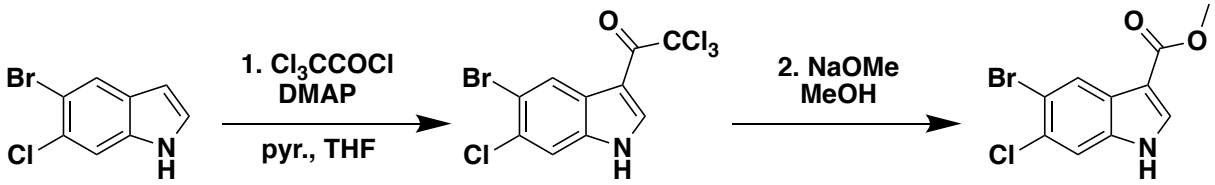
Chemical Synthesis of PRKAB1 Agonist:



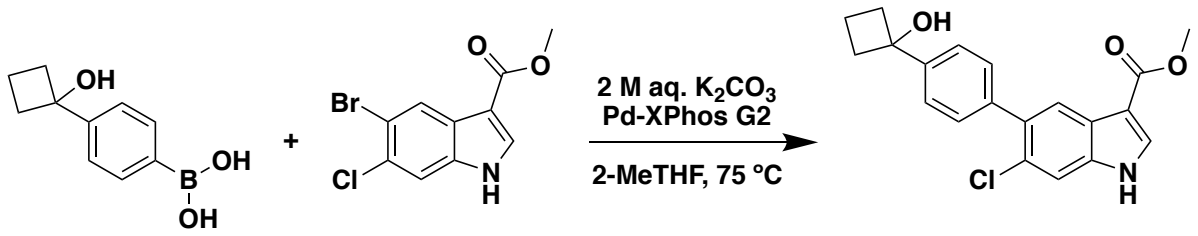
1-(4-Bromophenyl)cyclobutan-1-ol. An oven-dried 50 mL round-bottom flask was charged with 1,4-dibromobenzene (3.43 g, 14.6 mmol, 1.2 equiv) and anhydrous tetrahydrofuran (THF, 14 mL) and was cooled in a dry ice-acetone bath (−75 °C). *n*-Butyllithium (*n*-BuLi, 2.5 M in hexane, 5.8 mL, 14.6 mmol, 1.2 equiv) was added dropwise. After stirring for 30 min., cyclobutanone (0.850 g, 12.1 mmol, 1 equiv) was added dropwise. The reaction mixture was stirred for 60 min at -75ºC. The reaction was quenched by the addition of a saturated NH_4Cl (aq) solution (~7 mL). Water was added (~30 mL) and the solution mixture was extracted twice with Ethyl Acetate (EtOAc, ~100 mL x 2), the combined organics were washed with saturated brine (~50 mL), dried over MgSO4, filtered, and concentrated *in vacuo*. The product was purified via SiO2 column chromatography (using a gradient of 0% to 10% to 15% EtOAc in hexanes as eluent) to give the title compound as a white solid (2.40 g, **87% yield**). 1H NMR (300 MHz, $CDCl_3$) δ (ppm) = 7.49-7.46 (d, 2H), 7.37-7.34 (d, 2H), 2.54-2.46 (m, 2H), 2.39-2.29 (m, 2H), 2.28 (s, 1H), 2.08-1.94 (m, 1H), 1.75-1.61 (m, 1H).



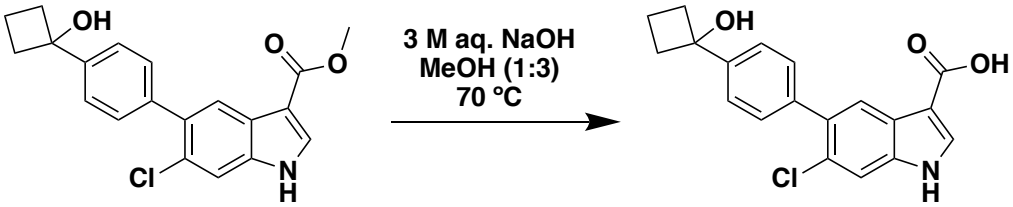
(4-(1-Hydroxycyclobutyl)phenyl)boronic acid: An oven-dried 100 mL round-bottom flask was charged with 1-(4-bromophenyl)cyclobutan-1-ol (2.00 g, 8.81 mmol, 1 equiv) and anhydrous ethanol (25 mL). Sodium tert-butoxide (8.5 mg, 0.088 mmol, 0.01 equiv), oven-dried potassium acetate (1.82 g, 18.5 mmol, 2.1 equiv), XPhos Pd G2 (chloro(2-dicyclohexylphosphino-2′,4′,6′-triisopropyl-1,1′-biphenyl)- [2-(2′-amino-1,1′-biphenyl)]palladium(II), 36 mg, 0.044 mmol, 0.005 equiv), and XPhos (2-dicyclohexylphosphino-2′,4′,6′-triisopropylbiphenyl, 42 mg, 0.088 mmol, 0.01 equiv) were added and the mixture was sparged with Argon. Tetrahydroxydiboron (1.58 g, 17.6 mmol, 2.0 equiv) was added and was stirred at 80 °C for 3 hr. The mixture was cooled to room temperature, and the solids were filtered through celite to give a light-yellow solution. Water was added (~20 mL) and the solution mixture was extracted twice with EtOAc (~50 mL x 2), the combined organics were washed with saturated brine (~20 mL), dried over MgSO4, filtered, and concentrated *in vacuo*. The product was purified via SiO_2 column chromatography (using a gradient of 50% to 60% to 70% EtOAc in hexanes as eluent) to give the title compound as a white solid (0.900 g, **53% yield**). 1H NMR (300 MHz, DMSO-d6) δ (ppm) = 7.96 (s, 2H), 7.77-7.74 (d, 2H), 7.45-7.43 (d, 2H), 5.45 (s, 1H), 2.42−2.33 (m, 2H), 2.30−2.24 (m, 2H), 1.98−1.84 (m, 1H), 1.71−1.56 (m, 1H).



Methyl 5-Bromo-6-chloro-1H-indole-3-carboxylate: An oven-dried 100 mL round-bottom flask was charged with 5-bromo-6-chloro-1H-indole (2.10 g, 9.11 mmol, 1 equiv), DMAP (Dimethylaminopyridine, 111 mg, 0.911 mmol, 0.1 equiv), pyridine (1.87 g, 23.7 mmol, 2.6 equiv), and anhydrous THF (14 mL) and then cooled to 0 °C. Trichloroacetyl chloride (3.98 g, 21.9 mmol, 2.4 equiv) was then added dropwise, and then stirred for 3 days at room temperature. The mixture was then cooled to 0 °C, and methanol (3.7 mL, 10 equiv) was added dropwise, then 25% sodium methoxide in methanol solution (4.2 mL, 2 equiv) was added dropwise. The mixture was stirred at 55 °C for 2 hr. Water was added (~20 mL) and the solution mixture was extracted twice with EtOAc (~50 mL x 2). The combined organics were washed with saturated brine (~20 mL), dried over MgSO4, filtered, and concentrated *in vacuo*. The product was purified via SiO_2 column chromatography (using a gradient of 20% to 30% EtOAc in hexanes as eluent) to give the title compound as an off-white solid (1.67 g, **63% yield**). 1H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.52 (br s, 1H), 8.45 (s, 1H), 7.92 (d, 1H), 7.56 (s, 1H), 3.93 (s, 3H).



Methyl 6-Chloro-5-(4-(1-hydroxycyclobutyl)phenyl)-1H-indole-3-carboxylate: An oven-dried 10 mL round-bottom flask was charged with methyl 5-bromo-6-chloro-1H-indole-3-carboxylate (280 mg, 0.970 mmol, 1 equiv), (4-(1-hydroxycyclobutyl)phenyl)boronic acid (205 mg, 1.07 mmol, 1.1 equiv), 2-MeTHF (2 mL), XPhos Pd G2 (chloro(2-dicyclohexyl-phosphino-2′,4′,6′-triisopropyl-1,1′-biphenyl)- [2-(2′-amino-1,1′-biphenyl)] palladium(II), 16 mg, 0.019 mmol, 0.02 equiv). Aqueous K_2CO_3 (362 mg, 2.62 mmol, 2.7 equiv in 1.1mL water, ~2.4 M) was added, and the mixture was sparged with Argon. The reaction mixture was heated to 75ºC overnight (18 hrs). After cooling, the solvent was evaporated to dryness. Water was added (~10 mL) and the solution mixture was extracted three times with EtOAc (~10 mL x 3), the combined organics were washed with saturated brine (~10 mL), dried over MgSO4, filtered, and concentrated *in vacuo*. The product was purified via SiO2 column chromatography (using a gradient of 30% to 40% to 50% EtOAc in hexanes as eluent) to give the title compound as a white solid (170 mg, **49% yield**). 1H NMR (300 MHz, DMSO-d6): δ (ppm) = 12.08 (br s, 1H), 8.17 (s, 1H), 7.93 (s, 1H), 7.66 (s, 1H), 7.58-7.55 (d, 2H), 7.41-7.39 (d, 2H), 5.55 (s, 1H), 3.79 (s, 3H), 2.48−2.41 (m, 2H), 2.35−2.26 (m, 2H), 2.00−1.91 (m, 1H), 1.76−1.64 (m, 1H).



6-Chloro-5-(4-(1-hydroxycyclobutyl)phenyl)-1H-indole-3-carboxylic acid (PF06409577): A 4-dram vial under Argon was charged with methyl 6-chloro-5-(4-(1-hydroxycyclobutyl)phenyl)-1H-indole-3-carboxylate (140 mg, 0.393 mmol, 1 equiv), 3M NaOH (aq) solution (0.787 mL, 2.36 mmol, 6 equiv), and methanol (MeOH, 2.36 mL). The reaction mixture was heated to 70ºC overnight (18 hrs). After cooling, the solvent was evaporated to dryness. Water was added (~10 mL), the pH was adjusted to ~2 using 1M HCl (aq) solution, and the solution mixture was extracted three times with EtOAc (~10 mL x 3), the combined organics were washed with saturated brine (~10 mL), dried over MgSO4, filtered, and concentrated *in vacuo*. The product was purified via SiO2 column chromatography (using a gradient of 30% to 60% Acetonitrile in Dichloromethane as eluent) to give the title compound as a white solid (54 mg, **40% yield**). 1H NMR (300 MHz, DMSO-d6): δ (ppm) = 11.94 (br s, 1H), 8.05 (s, 1H), 7.96 (s, 1H), 7.63 (s, 1H), 7.58-7.55 (d, 2H), 7.41-7.39 (d, 2H), 5.53 (s, 1H), 2.48−2.40 (m, 2H), 2.35−2.26 (m, 2H), 2.00−1.89 (m, 1H), 1.74−1.63 (m, 1H).

HPLC characterization was conducted using the Agilent 1260 Infinity II Quaternary Pump System with a 150 mm x 3 mm, 2.7 mm particle size C18 column (693975-302T). Solvent conditions (Solvent A: H2O (0.1%

008 TFA), Solvent B: Acetonitrile): 0-2 min isocratic 5% B, 2-20 min gradient to 95% B, 20-22 min isocratic 95%
009 B, at a flow rate of 1 mL/min. The chromatogram was monitored at 254 nm wavelength and the compound
010 had a retention time of 11.956 min. The purity of this final product was 96% by HPLC.
011
012 **Cell Culture:**
013 HCT-116: This human colon cancer cell line was obtained from the American Type Culture Collection (ATCC)
014 and maintained in Roswell Park Memorial Institute 1640 (RPMI-1640) medium (Gibco-BRL). The medium
015 was supplemented with 10% fetal calf serum (Hyclone), 1% L-Glutamine, and 1% penicillin/streptomycin
016 (Gibco-BRL). The cells were incubated at 37°C in a 5% CO2 incubator and were routinely passaged at a
017 dilution of 1:5 to 1:10.
018 SW480: This human colon cancer cell line was obtained from the ATCC and maintained in DMEM/F12,
019 supplemented with 10% FBS. The cells were incubated at 37°C in a 5% CO2 incubator and were routinely
020 passaged at a dilution of 1:2 to 1:5.
021 DLD1: This human colon cancer cell was obtained from the ATCC and were cultured using RMPI media
022 containing 10% FBS. The cells were incubated at 37°C in a 5% CO2 incubator and were routinely passaged
023 at a dilution of 1:5 to 1:10.
024 Caco2: This human colorectal cancer cell line was obtained from the ATCC and was cultured in high glucose
025 DMEM/F12 supplemented with 10% FBS and 1% penicillin/streptomycin (Cytiva). The cells were incubated
026 at 37°C in a 5% CO2 incubator and were routinely passaged at a dilution of 1:3 to 1:8.
027
028 **MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) assays on CRC cell lines:** Cell
029 survival was measured using the MTT reagent and cells cultured in 96-well plates. HCT116 and Sw480 cells
030 were cultured and treated with different concentrations of PF compound (0, 1, 5, 10, 20 μM) using DMSO as
031 a negative control for 16-48 h in various assays. Then the cell lines were incubated with MTT for 4 hr at 37°C.
032 After incubation, culture media was removed, and replaced with phosphate buffered saline (PBS), and 150
033 μl of DMSO was added in order to solubilize the MTT formazan crystals. Optical density was determined at
034 590 nm using a TECAN plate reader. At least three independent experiments were performed.
035
036 **Cell Cycle and Apoptosis Analyses:** Cell cycle distribution and apoptosis were assessed using the Guava
037 Cell Cycle Reagent (Millipore Sigma) and the Annexin V/Propidium Iodide (PI) staining kit (Thermo Fisher
038 Scientific), respectively, following the manufacturers' protocols. Samples were analyzed on a BD LSR II flow
039 cytometer, and data were processed using FlowJo software (FlowJo LLC, Ashland, OR).
040
041 **Cell Toxicity Assays:** Confluent flasks of HCT116 and Caco2 cells were rinsed with PBS and detached
042 using 1X trypsin-EDTA in PBS. Cells were resuspended in 1 mL of their respective medias (HCT116:
043 McCoy's 5A modified medium with 10% FBS and 1% Pen/Strep; Caco2: High glucose DMEM with 10% FBS,
044 1% NEAA, 1% P/S). Prior to seeding, 50 μL of media was added to each well of an Agilent xCELLigence 96-
045 well impedance plate, and baseline cell index readings (2 sweeps) were recorded using the Agilent eSight
046 live-cell analysis system. Cells were then counted and seeded at 10,000 cells/well in 50 μL of media. Cell

047 index measurements were collected at 15-minute intervals for 24 hours to monitor attachment and
048 proliferation.
049 After 24 hours, the assay was paused to allow treatment with compounds or vehicle. Each well
050 received 100 µL of treatment media containing 2X concentrations of the PF compound or DMSO (vehicle),
051 along with 2X Annexin V. The plate was returned to the instrument and continuous monitoring resumed. Cell
052 index measurements and live-cell images (4/well) were collected throughout the remainder of the 126-hour
053 assay. Cell index values were normalized to the final pre-treatment time point. Dose-response curves and
054 Annexin-based toxicity were assessed at the 126-hour endpoint.
055
056 **PDO Toxicity Assay**: PDOs were cultured for 72 hours prior to treatment. Organoids were then exposed to
057 complete medium containing either PF compound (5 µM or 20 µM), DMSO (vehicle), or hydrogen peroxide
058 (H_2O_2, positive control), with 500 nM eTox Red added to detect cell death. After 72 hours of treatment,
059 Hoechst 33342 (5 µg/mL) was added to label nuclei, followed by a PBS wash and replacement with phenol
060 red- and serum-free DMEM. Organoids were imaged using a BioTek Cytation 10 system. eTox Red-positive
061 cells (defined by red fluorescence within nuclei) were quantified using BioTek Gen5 software, and the
062 percentage of positive cells was calculated per condition. Data were analyzed and visualized using GraphPad
063 Prism.
064
065 **Animal Model:** Athymic nude male mice ages 4–6 weeks purchased from Jackson Laboratories (Bar
066 Harbor, ME) were utilized for this study. Mice were maintained in a barrier facility with high-efficiency
067 particulate air filtration and fed an autoclaved laboratory diet. Prior to surgical procedures, mice were
068 anesthetized with an intraperitoneal injection of ketamine and xylazine reconstituted in phosphate-buffered
069 saline (PBS). When the study concluded or if tumor burden became too large, defined as tumor volume >
070 1500 cm^3, mice were euthanized with CO2 inhalation followed by cervical dislocation. This study was
071 carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory
072 Animals of the National Institutes of Health. All animal studies were approved by the San Diego Veterans
073 Administration Medical Center Institutional Animal Care and Use Committee (protocol A17-020).
074 Tumor establishment with cell injection: Subcutaneous injection of HCT-116 cells (1 x 10^6)
075 reconstituted in PBS and Matrigel Matrix (Corning, NY) was performed on the bilateral shoulders and flanks
076 of nude mice. Tumors were allowed to grow until 50 mm^3 in diameter. The tumors were then resected and
077 divided into 1 mm^3 pieces for subcutaneous implantation.
078 Tumor establishment with tumor implantation: Subcutaneous models were established by surgical
079 implantation of 1 mm^3 HCT-116 tumor fragment into the bilateral flank of nude mice. A small incision was
080 made on the back of the mice and tumor fragment was directly implanted in the bilateral flank. The skin was
081 closed with 6-0 nylon suture (Ethicon Inc., Sommerville. NJ). Tumors were allowed to grow until 150 mm^3 in
082 diameter.
083
084 **Immunohistochemistry:** Tumor sections of 4 µm thickness were cut and placed on glass slides coated with
085 poly-L-lysine, followed by deparaffinization and hydration. Heat-induced epitope retrieval was performed
086 using sodium citrate buffer (pH 6) or Tris-EDTA (pH9) in a pressure cooker. Tumor sections were incubated

087 with 3% hydrogen peroxide for 5-15 min to block endogenous peroxidase activity, followed by incubation with
088 primary antibody overnight in a humidified chamber at 4°C. Antibodies used for immunostaining were anti-
089 CDX2 (1:300, rabbit host, cat no. ab76541), AMPK β1 (1:200, rabbit host, cat no. 10308-1-AP), anti-
090 pSer[245]GIV (1:50, anti-rabbit), anti-YAP (clone 63.7; Santa Cruz: sc-101199), anti-phospho(p)YAP Ser[127]
091 (Cell Signaling Technology Cat#13008S) and anti-β-Catenin (1:300, mouse host, SC-7963). Antigen retrieval
092 protocols were optimized based on manufacturer's protocols. Immunostaining was visualized with a labelled
093 streptavidin–biotin using 3,3′-diaminobenzidine as a chromogen and counterstained with hematoxylin.
094 Immunohistochemistry (IHC) images were randomly sampled at different regions of interest (ROI). The ROIs
095 were analysed using IHC Profiler[141]. IHC Profiler uses a spectral deconvolution method of DAB/hematoxylin
096 colour spectra using optimized optical density vectors of the colour deconvolution plugin to properly
097 segregate the DAB colour spectra. The histogram of the DAB intensity was divided into 4 zones: high positive
098 (0–60), positive (61–120), low positive (121–180), and negative (181–235). High positive, positive, and low
099 positive percentages were combined to compute the final percentage positive for each ROI. The range of
100 values for the percent positive was compared among the different experimental groups.
101
102 **Immunofluorescence:** Tumor sections of 4 µm thickness were cut and placed on glass slides coated with
103 poly-L-lysine, followed by deparaffinization and hydration. Heat-induced epitope retrieval was performed
104 using Tris-EDTA (pH9) in a pressure cooker. Samples were blocked for 1 hr using an in-house blocking buffer
105 (1% BSA and 0.01% Tween-20 in TBS). Primary antibodies were diluted in blocking buffer and allowed to
106 incubate overnight at 4°C; antibodies used were E Cadherin (H-108) cat no. SC7870, 1:50, and Anti-Beta
107 Catenin cat no. SC7963, 1:50. Secondary antibodies, Goat anti-mouse IgG H&L secondary antibody, Alexa
108 Fluor 488 and Goat anti-rabbit IgG H&L secondary antibody, Alexa Fluor 594, were diluted in blocking buffer
109 and allowed to incubate for 2 hr in the dark. ProLong Glass was used as a mounting medium. #1 Thick
110 Coverslips were applied to slides and sealed. Samples were stored at 4°C until imaged.
111
112 **Quantification of Nuclear Area:** Formalin-fixed paraffin-embedded (FFPE) xenograft sections were stained
113 with hematoxylin and eosin (H&E) using standard protocols. Whole slide images were acquired using a high-
114 resolution slide scanner. Representative fields from vehicle- and PF-treated xenografts were exported as
115 individual TIFF files for image analysis. To quantify nuclear morphology, images were processed using
116 ImageJ with the StarDist plugin (v0.3.0), trained on an H&E-compatible model for nuclear detection. Each
117 image was converted to 8-bit grayscale, and contrast was adjusted as needed. StarDist-generated nuclear
118 segmentation masks were used to define regions of interest, and nuclear area was quantified for all detected
119 nuclei in each field. Data visualization and statistical analysis were performed using GraphPad Prism.
120
121 **Patient-derived Organoid (PDO) Culture:** Twenty-three colorectal carcinoma (CRC) human patient-derived
122 model (PDM) organoids were acquired from the Human Cancer Models Initiative (HCMI) catalogue of the
123 American Type Culture Collection (ATCC). Five healthy colon PDOs were used from the biobank of the UC
124 San Diego HUMANOID Center of Research Excellence (CoRE) at the University of California, San Diego,
125 USA[142]. Organoids were cultured and passaged according to previously described methods[79,142-147]. For
126 maintenance, organoids were cultured in Matrigel in 12- or 24-well tissue culture plates incubated at 37ºC,

127 5% CO_2. Organoids were provided 50% L-WRN-conditioned medium supplemented with differentiation
128 inhibitors (Intestigrow; Cat#: HUM2019-125, HUMANOID™, San Diego, CA). Organoids were subcultured
129 every 7-14 days, dissociated by non-enzymatic digestion in TrypLE and passaged to achieve the desired
130 density. All relevant demographic, clinical and tumor mutation related information is provided in
131 **Supplemental Information 3**.
132
133 **Estimation of IC_{50} by MTT Assays on PDOs:** To attain IC50 values for CRC PDM lines and healthy colon
134 organoid lines in response to PF-06409577, thiazolyl blue tetrazolium bromide (MTT) reduction assays were
135 performed, based on a previous protocol [148] with modifications. Prior to experiment, $5x10^3$-$1x10^4$ dissociated
136 single cells were seeded in Matrigel in 96-well tissue culture plates. After 3 days in culture, organoid images
137 were captured on EVOS XL Core Imaging System. Organoids were administered complete medium or
138 complete medium treated on alternate days with 1 μM, 5 μM, 10 μM, 20 μM, or 40 μM PF-06409577 (final
139 concentration), while ensuring that only 50% of the media was replenished during subsequent repeated
140 treatments. After 7-9 days of treatment, when apoptosis was most apparent, organoids were imaged and
141 MTT reagent was added to each well for a final MTT concentration of 588 μg/ml. Organoids were incubated
142 in MTT for 4 hours at 37ºC, 5% CO_2, after which images were acquired to document the extent of formazan
143 formation. To prepare the plate for quantification, Matrigel was dissolved in 2% SDS, followed by formazan
144 dissolution in dimethyl sulfoxide. The optical density of each well was measured at 562 nm using Tecan
145 Spark Multimode Microplate Reader[148]. The measured optical densities at each concentration were used to
146 calculate an IC50 for each CRC PDO and healthy colon line.
147
148 **Immunoblotting:** To verify the expression of markers of stemness and differentiation, equal aliquots of
149 whole-cell lysates (prepared using RIPA buffer) were loaded on a 8% SDS PAGE gel, and immunoblotting
150 was carried out for various targets as described previously[149]. Immunoblots were analyzed using a Gel Doc
151 system (BioRad, Hercules, CA).
152
153 **Flow Cytometry:** To produce single-cell suspension from CRC PDOs for fluorescence-activated single cell
154 sorting (FACS), dissociated single cells were seeded in Matrigel in 24-well tissue culture plates. After 3 days
155 in culture, organoids were administered complete medium or complete medium treated with 1 μM, 5 μM, 10
156 μM, 20 μM, or 40 μM PF-06409577. After 7-9 days of treatment, organoids were collected in PBS-EDTA.
157 Collected organoids were dissociated in TrypLE for subsequent analysis by flow. Apoptotic cell quantification
158 was performed using the annexin V/propidium iodide (PI) staining kit (Thermo Fisher Scientific), according to
159 the manufacturer's instructions. Cells were quantified on a BD LSR II flow cytometer and analyzed using
160 FlowJo software (FlowJo, Ashland, OR, USA).
161
162 **RNA Extraction Transcriptomic Studies:** To collect RNA from CRC PDOs, dissociated single cells were
163 seeded in Matrigel in 24-well tissue culture plates. After 3 days in culture, organoids were administered
164 complete medium or complete medium treated with 10 μM PF-06409577. Prior to treatment with PF-
165 06409577 and after 2 days of treatment, whole organoids were collected in Dulbecco's phosphate-buffered
166 saline with 0.5 mM ethylenediaminetetraacetic acid (PBS-EDTA). Organoids were subsequently incubated

167 in cell recovery solution, pelleted, and stored at -80ºC. Total RNA was extracted using the Quick-RNA
168 Microprep Kit and concentration was measured using Nanodrop One^{C} Spectrophotometer. For qPCR, first-
169 strand cDNA synthesis was performed using qScript cDNA SuperMix.
170 For RNA sequencing, isolated RNA has been processed for RNA sequencing in the Illumina NovaSeq
171 6000 platform. Fastq sequence files have been mapped using the human GRCh38 genome. Log normalized
172 CPM expression files are submitted to NCBI GEO [GSE237623, GSE237624, GSE237625].
173
174 **Ethics Statement:** All animal studies were conducted in strict accordance with the NIH *Guide for the Care*
175 *and Use of Laboratory Animals* and were approved by the Institutional Animal Care and Use Committee of
176 the San Diego Veterans Administration Medical Center (protocol A17-020; PI Bouvet). Colorectal cancer
177 patient-derived organoids (PDOs) were obtained from the American Type Culture Collection (ATCC), a
178 nonprofit global biorepository and distributor of organoid models curated under the Human Cancer Models
179 Initiative (HCMI) to enhance reproducibility. Deidentified case-associated data, including molecular
180 characterizations and tissue origins, were used in this study; harmonized datasets are publicly accessible via
181 the NCI's Genomic Data Commons. Healthy colon PDOs were derived from individuals undergoing routine
182 colonoscopy for cancer screening at UC San Diego, with informed consent under an IRB-approved protocol
183 (#190105; PI Ghosh).

184 ## References

185 1. Stone, R.M., Maguire, M., Goldberg, M.A., Antin, J.H., Rosenthal, D.S., and Mayer, R.J. (1988).
186 Complete remission in acute promyelocytic leukemia despite persistence of abnormal bone marrow
187 promyelocytes during induction therapy: experience in 34 patients. Blood *71*, 690-696.
188 2. Kantarjian, H.M., Keating, M.J., Walters, R.S., Estey, E.H., McCredie, K.B., Smith, T.L., Dalton,
189 W.T., Jr., Cork, A., Trujillo, J.M., and Freireich, E.J. (1986). Acute promyelocytic leukemia. M.D.
190 Anderson Hospital experience. Am J Med *80*, 789-797.
191 3. Cunningham, I., Gee, T.S., Reich, L.M., Kempin, S.J., Naval, A.N., and Clarkson, B.D. (1989).
192 Acute promyelocytic leukemia: treatment results during a decade at Memorial Hospital. Blood *73*,
193 1116-1122.
194 4. Tallman, M.S., Andersen, J.W., Schiffer, C.A., Appelbaum, F.R., Feusner, J.H., Ogden, A.,
195 Shepherd, L., Willman, C., Bloomfield, C.D., Rowe, J.M., and Wiernik, P.H. (1997). All-trans-retinoic
196 acid in acute promyelocytic leukemia. N Engl J Med *337*, 1021-1028.
197 10.1056/NEJM199710093371501.
198 5. Tallman, M.S., Andersen, J.W., Schiffer, C.A., Appelbaum, F.R., Feusner, J.H., Woods, W.G.,
199 Ogden, A., Weinstein, H., Shepherd, L., Willman, C., et al. (2002). All-trans retinoic acid in acute
200 promyelocytic leukemia: long-term outcome and prognostic factor analysis from the North American
201 Intergroup protocol. Blood *100*, 4298-4302. 10.1182/blood-2002-02-0632.
202 6. De Botton, S., Dombret, H., Sanz, M., Miguel, J.S., Caillot, D., Zittoun, R., Gardembas, M.,
203 Stamatoulas, A., Conde, E., Guerci, A., et al. (1998). Incidence, clinical features, and outcome of all
204 trans-retinoic acid syndrome in 413 cases of newly diagnosed acute promyelocytic leukemia. The
205 European APL Group. Blood *92*, 2712-2718.
206 7. Testi, A.M., Biondi, A., Lo Coco, F., Moleti, M.L., Giona, F., Vignetti, M., Menna, G., Locatelli, F.,
207 Pession, A., Barisone, E., et al. (2005). GIMEMA-AIEOPAIDA protocol for the treatment of newly
208 diagnosed acute promyelocytic leukemia (APL) in children. Blood *106*, 447-453. 10.1182/blood-
209 2004-05-1971.
210 8. Fenaux, P., Chevret, S., Guerci, A., Fegueux, N., Dombret, H., Thomas, X., Sanz, M., Link, H.,
211 Maloisel, F., Gardin, C., et al. (2000). Long-term follow-up confirms the benefit of all-trans retinoic
212 acid in acute promyelocytic leukemia. European APL group. Leukemia *14*, 1371-1377.

9. Sanz, M.A., Martin, G., Rayon, C., Esteve, J., Gonzalez, M., Diaz-Mediavilla, J., Bolufer, P., Barragan, E., Terol, M.J., Gonzalez, J.D., et al. (1999). A modified AIDA protocol with anthracycline-based consolidation results in high antileukemic efficacy and reduced toxicity in newly diagnosed PML/RARalpha-positive acute promyelocytic leukemia. PETHEMA group. Blood *94*, 3015-3021.

10. Mandelli, F., Diverio, D., Avvisati, G., Luciano, A., Barbui, T., Bernasconi, C., Broccia, G., Cerri, R., Falda, M., Fioritoni, G., et al. (1997). Molecular remission in PML/RAR alpha-positive acute promyelocytic leukemia by combined all-trans retinoic acid and idarubicin (AIDA) therapy. Gruppo Italiano-Malattie Ematologiche Maligne dell'Adulto and Associazione Italiana di Ematologia ed Oncologia Pediatrica Cooperative Groups. Blood *90*, 1014-1021.

11. Bots, M., Verbrugge, I., Martin, B.P., Salmon, J.M., Ghisi, M., Baker, A., Stanley, K., Shortt, J., Ossenkoppele, G.J., Zuber, J., et al. (2014). Differentiation therapy for the treatment of t(8;21) acute myeloid leukemia using histone deacetylase inhibitors. Blood *123*, 1341-1352. 10.1182/blood-2013-03-488114.

12. Yen, K., Travins, J., Wang, F., David, M.D., Artin, E., Straley, K., Padyana, A., Gross, S., DeLaBarre, B., Tobin, E., et al. (2017). AG-221, a First-in-Class Therapy Targeting Acute Myeloid Leukemia Harboring Oncogenic IDH2 Mutations. Cancer Discov *7*, 478-493. 10.1158/2159-8290.CD-16-1034.

13. Medema, J.P. (2017). Targeting the Colorectal Cancer Stem Cell. N Engl J Med *377*, 888-890. 10.1056/NEJMcibr1706541.

14. Butte, A.J., and Kohane, I.S. (1999). Unsupervised knowledge discovery in medical databases using relevance networks. Proc AMIA Symp, 711-715.

15. Shameer, K., Readhead, B., and Dudley, J.T. (2015). Computational and experimental advances in drug repositioning for accelerated therapeutic stratification. Curr Top Med Chem *15*, 5-20.

16. Shen, Y., Alvarez, M.J., Bisikirska, B., Lachmann, A., Realubit, R., Pampou, S., Coku, J., Karan, C., and Califano, A. (2017). Systematic, network-based characterization of therapeutic target inhibitors. PLoS Comput Biol *13*, e1005599. 10.1371/journal.pcbi.1005599.

17. Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., and Califano, A. (2006). Reverse engineering cellular networks. Nat Protoc *1*, 662-671. 10.1038/nprot.2006.106.

18. van Someren, E.P., Wessels, L.F., Backer, E., and Reinders, M.J. (2002). Genetic network modeling. Pharmacogenomics *3*, 507-525. 10.1517/14622416.3.4.507.

19. Kusonmano, K. (2017). Gene Expression Analysis Through Network Biology: Bioinformatics Approaches. Adv Biochem Eng Biotechnol *160*, 15-32. 10.1007/10_2016_44.

20. Allocco, D.J., Kohane, I.S., and Butte, A.J. (2004). Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics *5*, 18. 10.1186/1471-2105-5-18.

21. Arkin, A., and Ross, J. (1995). Statistical Construction of Chemical Reaction Mechanisms from Measured Time-Series. The Journal of Physical Chemistry *99*, 970-979. 10.1021/j100003a020.

22. Jordan, I.K., Marino-Ramirez, L., Wolf, Y.I., and Koonin, E.V. (2004). Conservation and coevolution in the scale-free human gene coexpression network. Mol Biol Evol *21*, 2058-2070. 10.1093/molbev/msh222.

23. Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J., and Pavlidis, P. (2004). Coexpression analysis of human genes across many microarray data sets. Genome Res *14*, 1085-1094. 10.1101/gr.1910904.

24. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nat Genet *22*, 281-285. 10.1038/10343.

25. Butte, A.J., Tamayo, P., Slonim, D., Golub, T.R., and Kohane, I.S. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proc Natl Acad Sci U S A *97*, 12182-12186. 10.1073/pnas.220392197.

26. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med *351*, 2817-2826. 10.1056/NEJMoa041588.

27. Witten, D.M., and Tibshirani, R. (2010). Survival analysis with high-dimensional covariates. Stat Methods Med Res *19*, 29-51. 10.1177/0962280209105024.

28. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503-511. 10.1038/35000501.

29. Zhao, H., Ljungberg, B., Grankvist, K., Rasmuson, T., Tibshirani, R., and Brooks, J.D. (2006). Gene expression profiling predicts survival in conventional renal cell carcinoma. PLoS Med *3*, e13. 10.1371/journal.pmed.0030013.

30. Sahoo, D., Swanson, L., Sayed, I.M., Katkar, G.D., Ibeawuchi, S.R., Mittal, Y., Pranadinata, R.F., Tindle, C., Fuller, M., Stec, D.L., et al. (2021). Artificial intelligence guided discovery of a barrier-

protective therapy in inflammatory bowel disease. Nat Commun *12*, 4246. 10.1038/s41467-021-24470-5.

31. Ghosh, P., Sinha, S., Katkar, G.D., Vo, D., Taheri, S., Dang, D., Das, S., and Sahoo, D. (2023). Machine learning identifies signatures of macrophage reactivity and tolerance that predict disease outcomes. EBioMedicine *94*, 104719. 10.1016/j.ebiom.2023.104719.

32. Sahoo, D., Seita, J., Bhattacharya, D., Inlay, M.A., Weissman, I.L., Plevritis, S.K., and Dill, D.L. (2010). MiDReG: a method of mining developmentally regulated genes using Boolean implications. Proc Natl Acad Sci U S A *107*, 5732-5737. 10.1073/pnas.0913635107.

33. Inlay, M.A., Bhattacharya, D., Sahoo, D., Serwold, T., Seita, J., Karsunky, H., Plevritis, S.K., Dill, D.L., and Weissman, I.L. (2009). Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. Genes Dev *23*, 2376-2381. 10.1101/gad.1836009.

34. Chen, J.Y., Miyanishi, M., Wang, S.K., Yamazaki, S., Sinha, R., Kao, K.S., Seita, J., Sahoo, D., Nakauchi, H., and Weissman, I.L. (2016). Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature *530*, 223-227. 10.1038/nature16943.

35. Chan, C.K., Seo, E.Y., Chen, J.Y., Lo, D., McArdle, A., Sinha, R., Tevlin, R., Seita, J., Vincent-Tompkins, J., Wearda, T., et al. (2015). Identification and specification of the mouse skeletal stem cell. Cell *160*, 285-298. 10.1016/j.cell.2014.12.002.

36. Dimov, I.K., Lu, R., Lee, E.P., Seita, J., Sahoo, D., Park, S.M., Weissman, I.L., and Lee, L.P. (2014). Discriminating cellular heterogeneity using microwell-based RNA cytometry. Nat Commun *5*, 3451. 10.1038/ncomms4451.

37. Chan, C.K., Lindau, P., Jiang, W., Chen, J.Y., Zhang, L.F., Chen, C.C., Seita, J., Sahoo, D., Kim, J.B., Lee, A., et al. (2013). Clonal precursor of bone, cartilage, and hematopoietic niche stromal cells. Proc Natl Acad Sci U S A *110*, 12643-12648. 10.1073/pnas.1310212110.

38. Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich, L.I., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene Expression Commons: an open platform for absolute gene expression profiling. PLoS One *7*, e40321. 10.1371/journal.pone.0040321.

39. Cheah, M.T., Chen, J.Y., Sahoo, D., Contreras-Trujillo, H., Volkmer, A.K., Scheeren, F.A., Volkmer, J.P., and Weissman, I.L. (2015). CD14-expressing cancer cells establish the inflammatory and

301 proliferative tumor microenvironment in bladder cancer. Proc Natl Acad Sci U S A *112*, 4725-4730.
302 10.1073/pnas.1424795112.
303 40. Volkmer, J.P., Sahoo, D., Chin, R.K., Ho, P.L., Tang, C., Kurtova, A.V., Willingham, S.B.,
304 Pazhanisamy, S.K., Contreras-Trujillo, H., Storm, T.A., et al. (2012). Three differentiation states
305 risk-stratify bladder cancer into distinct subtypes. Proc Natl Acad Sci U S A *109*, 2078-2083.
306 10.1073/pnas.1120605109.
307 41. Shin, K., Lim, A., Zhao, C., Sahoo, D., Pan, Y., Spiekerkoetter, E., Liao, J.C., and Beachy, P.A.
308 (2014). Hedgehog signaling restrains bladder cancer progression by eliciting stromal production of
309 urothelial differentiation factors. Cancer Cell *26*, 521-533. 10.1016/j.ccell.2014.09.001.
310 42. Sin, M.L.Y., Mach, K.E., Sinha, R., Wu, F., Trivedi, D.R., Altobelli, E., Jensen, K.C., Sahoo, D., Lu,
311 Y., and Liao, J.C. (2017). Deep Sequencing of Urinary RNAs for Bladder Cancer Molecular
312 Diagnostics. Clin Cancer Res *23*, 3700-3710. 10.1158/1078-0432.CCR-16-2610.
313 43. Sahoo, D., Wei, W., Auman, H., Hurtado-Coll, A., Carroll, P.R., Fazli, L., Gleave, M.E., Lin, D.W.,
314 Nelson, P.S., Simko, J., et al. (2018). Boolean analysis identifies CD38 as a biomarker of
315 aggressive localized prostate cancer. Oncotarget *9*, 6550-6561. 10.18632/oncotarget.23973.
316 44. Bhamre, S., Sahoo, D., Tibshirani, R., Dill, D.L., and Brooks, J.D. (2009). Temporal changes in
317 gene expression induced by sulforaphane in human prostate cancer cells. Prostate *69*, 181-190.
318 10.1002/pros.20869.
319 45. Ghosh, P., Campos, V.J., Vo, D.T., Guccione, C., Goheen-Holland, V., Tindle, C., Mazzini, G.S.,
320 He, Y., Alexandrov, L.B., Lippman, S.M., et al. (2022). AI-assisted discovery of an ethnicity-
321 influenced driver of cell transformation in esophageal and gastroesophageal junction
322 adenocarcinomas. JCI Insight *7*. 10.1172/jci.insight.161334.
323 46. Bonhomme, C., Duluc, I., Martin, E., Chawengsaksophak, K., Chenard, M., Kedinger, M., Beck, F.,
324 Freund, J., and Domon-Dell, C. (2003). The Cdx2 homeobox gene has a tumour suppressor
325 function in the distal colon in addition to a homeotic role during gut development. Gut *52*, 1465-
326 1471.
327 47. Silberg, D.G., Swain, G.P., Suh, E.R., and Traber, P.G. (2000). Cdx1 and cdx2 expression during
328 intestinal development. Gastroenterology *119*, 961-971.
329 48. Chawengsaksophak, K., James, R., Hammond, V., Köntgen, F., and Beck, F. (1997). Homeosis and
330 intestinal tumours in Cdx2 mutant mice. Nature *386*, 84-87.

49. Hryniuk, A., Grainger, S., Savory, J.G., and Lohnes, D. (2014). Cdx1 and Cdx2 function as tumor suppressors. Journal of Biological Chemistry *289*, 33343-33354.

50. Sahoo, D., Dill, D.L., Gentles, A.J., Tibshirani, R., and Plevritis, S.K. (2008). Boolean implication networks derived from large scale, whole genome microarray datasets. Genome Biol *9*, R157. 10.1186/gb-2008-9-10-r157.

51. Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol *29*, 1120-1127. 10.1038/nbt.2038.

52. Dalerba, P., Sahoo, D., Paik, S., Guo, X., Yothers, G., Song, N., Wilcox-Fogel, N., Forgó, E., Rajendran, P.S., Miranda, S.P., et al. (2016). CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. N Engl J Med *374*, 211-222. 10.1056/NEJMoa1506597.

53. Levin, T.G., Powell, A.E., Davies, P.S., Silk, A.D., Dismuke, A.D., Anderson, E.C., Swain, J.R., and Wong, M.H. (2010). Characterization of the intestinal cancer stem cell marker CD166 in the human and mouse gastrointestinal tract. Gastroenterology *139*, 2072-2082 e2075. 10.1053/j.gastro.2010.08.053.

54. Weichert, W., Knosel, T., Bellach, J., Dietel, M., and Kristiansen, G. (2004). ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. Journal of clinical pathology *57*, 1160-1164.

55. Dalerba, P., Dylla, S.J., Park, I.K., Liu, R., Wang, X., Cho, R.W., Hoey, T., Gurney, A., Huang, E.H., Simeone, D.M., et al. (2007). Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A *104*, 10158-10163.

56. Derangère, V., Lecuelle, J., Lepage, C., Aoulad-Ben Salem, O., Allatessem, B.M., Ilie, A., Bouché, O., Phelip, J.M., Baconnier, M., Pezet, D., et al. (2022). Combination of CDX2 H-score quantitative analysis with CD3 AI-guided analysis identifies patients with a good prognosis only in stage III colon cancer. Eur J Cancer *172*, 221-230. 10.1016/j.ejca.2022.05.040.

57. Bennedsen, A.L.B., Cai, L., Hasselager, R.P., Özcan, A.A., Mohamed, K.B., Eriksen, J.O., Eiholm, S., Bzorek, M., Fiehn, A.K., Hviid, T.V.F., and Gögenur, I. (2022). An exploration of immunohistochemistry-based prognostic markers in patients undergoing curative resections for colon cancer. BMC Cancer *22*, 62. 10.1186/s12885-022-09169-0.

360 58. Azcue, P., Guerrero Setas, D., Encío, I., Ibáñez-Beroiz, B., Mercado, M., Vera, R., and Gómez-
361 Dorronsoro, M.L. (2021). A Novel Prognostic Biomarker Panel for Early-Stage Colon Carcinoma.
362 Cancers (Basel) *13*. 10.3390/cancers13235909.
363 59. den Uil, S.H., de Wit, M., Slebos, R.J.C., Delis-van Diemen, P.M., Sanders, J., Piersma, S.R.,
364 Pham, T.V., Coupé, V.M.H., Bril, H., Stockmann, H.B.A.C., et al. (2021). Quantitative analysis of
365 CDX2 protein expression improves its clinical utility as a prognostic biomarker in stage II and III
366 colon cancer. Eur J Cancer *144*, 91-100. 10.1016/j.ejca.2020.10.029.
367 60. Tarazona, N., Gimeno-Valiente, F., Gambardella, V., Huerta, M., Roselló, S., Zuniga, S., Calon, A.,
368 Carbonell-Asins, J.A., Fontana, E., Martinez-Ciarpaglini, C., et al. (2020). Detection of postoperative
369 plasma circulating tumour DNA and lack of CDX2 expression as markers of recurrence in patients
370 with localised colon cancer. ESMO Open *5*, e000847. 10.1136/esmoopen-2020-000847.
371 61. Hestetun, K.E., Aasebø, K., Rosenlund, N.B., Müller, Y., Dahl, O., and Myklebust, M.P. (2021).
372 Mismatch repair phenotype determines the implications of tumor grade and CDX2 expression in
373 stage II-III colon cancer. Mod Pathol *34*, 161-170. 10.1038/s41379-020-0634-9.
374 62. Asgari-Karchekani, S., Karimian, M., Mazoochi, T., Taheri, M.A., and Khamehchian, T. (2020).
375 CDX2 Protein Expression in Colorectal Cancer and ItsCorrelation with Clinical and Pathological
376 Characteristics, Prognosis, and Survival Rate of Patients. J Gastrointest Cancer *51*, 844-849.
377 10.1007/s12029-019-00314-w.
378 63. Ma, C., Olevian, D., Miller, C., Herbst, C., Jayachandran, P., Kozak, M.M., Chang, D.T., and Pai,
379 R.K. (2019). SATB2 and CDX2 are prognostic biomarkers in DNA mismatch repair protein deficient
380 colon cancer. Mod Pathol *32*, 1217-1231. 10.1038/s41379-019-0265-1.
381 64. Ben-Aharon, I., Goshen-Lago, T., Sternschuss, M., Morgenstern, S., Geva, R., Beny, A., Dror, Y.,
382 Steiner, M., Hubert, A., Idelevich, E., et al. (2019). Sidedness Matters: Surrogate Biomarkers
383 Prognosticate Colorectal Cancer upon Anatomic Location. Oncologist *24*, e696-e701.
384 10.1634/theoncologist.2018-0351.
385 65. Hansen, T.F., Kjær-Frifeldt, S., Eriksen, A.C., Lindebjerg, J., Jensen, L.H., Sørensen, F.B., and
386 Jakobsen, A. (2018). Prognostic impact of CDX2 in stage II colon cancer: results from two
387 nationwide cohorts. British Journal of Cancer *119*, 1367-1373.

388 66. Pilati, C., Taieb, J., Balogoun, R., Marisa, L., De Reyniès, A., and Laurent-Puig, P. (2017). CDX2
389 prognostic value in stage II/III resected colon cancer is related to CMS classification. Annals of
390 Oncology *28*, 1032-1035.
391 67. Nolte, S., Zlobec, I., Lugli, A., Hohenberger, W., Croner, R., Merkel, S., Hartmann, A., Geppert, C.I.,
392 and Rau, T.T. (2017). Construction and analysis of tissue microarrays in the era of digital pathology:
393 a pilot study targeting CDX1 and CDX2 in a colon cancer cohort of 612 patients. J Pathol Clin Res
394 *3*, 58-70. 10.1002/cjp2.62.
395 68. Zhang, B.Y., Jones, J.C., Briggler, A.M., Hubbard, J.M., Kipp, B.R., Sargent, D.J., Dixon, J.G., and
396 Grothey, A. (2017). Lack of caudal-type homeobox transcription factor 2 expression as a prognostic
397 biomarker in metastatic colorectal cancer. Clinical colorectal cancer *16*, 124-128.
398 69. Tomasello, G., Barni, S., Turati, L., Ghidini, M., Pezzica, E., Passalacqua, R., and Petrelli, F.
399 (2018). Association of CDX2 expression with survival in early colorectal cancer: a systematic review
400 and meta-analysis. Clinical colorectal cancer *17*, 97-103.
401 70. Bruun, J., Sveen, A., Barros, R., Eide, P.W., Eilertsen, I., Kolberg, M., Pellinen, T., David, L.,
402 Svindland, A., Kallioniemi, O., et al. (2018). Prognostic, predictive, and pharmacogenomic
403 assessments of CDX2 refine stratification of colorectal cancer. Mol Oncol *12*, 1639-1655.
404 10.1002/1878-0261.12347.
405 71. Sjoerd, H., de Wit, M., Slebos, R.J., Delis-van Diemen, P.M., Sanders, J., Piersma, S.R., Pham,
406 T.V., Coupé, V.M., Bril, H., and Stockmann, H.B. (2021). Quantitative analysis of CDX2 protein
407 expression improves its clinical utility as a prognostic biomarker in stage II and III colon cancer.
408 European Journal of Cancer *144*, 91-100.
409 72. Baba, Y., Nosho, K., Shima, K., Freed, E., Irahara, N., Philips, J., Meyerhardt, J.A., Hornick, J.L.,
410 Shivdasani, R.A., and Fuchs, C.S. (2009). Relationship of CDX2 loss with molecular features and
411 prognosis in colorectal cancer. Clinical cancer research *15*, 4665-4673.
412 73. Olsen, J., Eiholm, S., Kirkeby, L., Espersen, M.L.M., Jess, P., Gögenür, I., and Troelsen, J. (2016).
413 CDX2 downregulation is associated with poor differentiation and MMR deficiency in colon cancer.
414 Experimental and molecular pathology *100*, 59-66.
415 74. Bruun, J., Sveen, A., Barros, R., Eide, P.W., Eilertsen, I., Kolberg, M., Pellinen, T., David, L.,
416 Svindland, A., and Kallioniemi, O. (2018). Prognostic, predictive, and pharmacogenomic
417 assessments of CDX 2 refine stratification of colorectal cancer. Molecular oncology *12*, 1639-1655.

418 75. Kumar, N., Tsai, Y.H., Chen, L., Zhou, A., Banerjee, K.K., Saxena, M., Huang, S., Toke, N.H., Xing,
419 J., Shivdasani, R.A., et al. (2019). The lineage-specific transcription factor CDX2 navigates dynamic
420 chromatin to control distinct stages of intestine development. Development *146*.
421 10.1242/dev.172189.
422 76. Fearon, E.R., and Huang, E.H. (2016). CDX2: Linking Cell and Patient Fates in Colon Cancer. Cell
423 Stem Cell *18*, 168-169. 10.1016/j.stem.2016.01.011.
424 77. Fouqueray, P., Bolze, S., Dubourg, J., Hallakou-Bozec, S., Theurey, P., Grouin, J.M., Chevalier, C.,
425 Gluais-Dagorn, P., Moller, D.E., and Cusi, K. (2021). Pharmacodynamic effects of direct AMP
426 kinase activation in humans with insulin resistance and non-alcoholic fatty liver disease: A phase 1b
427 study. Cell Rep Med *2*, 100474. 10.1016/j.xcrm.2021.100474.
428 78. Ghosh, P. (2017). The stress polarity pathway: AMPK 'GIV'-es protection against metabolic insults.
429 Aging (Albany NY) *9*, 303-314. 10.18632/aging.101179.
430 79. Ghosh, P., Swanson, L., Sayed, I.M., Mittal, Y., Lim, B.B., Ibeawuchi, S.R., Foretz, M., Viollet, B.,
431 Sahoo, D., and Das, S. (2020). The stress polarity signaling (SPS) pathway serves as a marker and
432 a target in the leaky gut barrier: implications in aging and cancer. Life Sci Alliance *3*.
433 10.26508/lsa.201900481.
434 80. Aznar, N., Patel, A., Rohena, C.C., Dunkel, Y., Joosen, L.P., Taupin, V., Kufareva, I., Farquhar,
435 M.G., and Ghosh, P. (2016). AMP-activated protein kinase fortifies epithelial tight junctions during
436 energetic stress via its effector GIV/Girdin. Elife *5*. 10.7554/eLife.20795.
437 81. Gough, N.R. (2007). Maintaining Polarity in the Face of Energetic Stress. Science's STKE *2007*,
438 tw166-tw166. doi:10.1126/stke.3862007tw166.
439 82. Sahoo, D., Dill, D.L., Tibshirani, R., and Plevritis, S.K. (2007). Extracting binary signals from
440 microarray time-course data. Nucleic Acids Res *35*, 3705-3712. 10.1093/nar/gkm284.
441 83. Graule, J., Uth, K., Fischer, E., Centeno, I., Galván, J.A., Eichmann, M., Rau, T.T., Langer, R.,
442 Dawson, H., and Nitsche, U. (2018). CDX2 in colorectal cancer is an independent prognostic factor
443 and regulated by promoter methylation and histone deacetylation in tumors of the serrated pathway.
444 Clinical epigenetics *10*, 1-12.
445 84. Boman, B.M., and Huang, E. (2008). Human colon cancer stem cells: a new paradigm in
446 gastrointestinal oncology. J Clin Oncol *26*, 2828-2838. 10.1200/JCO.2008.17.6941.

447 85. Kalisky, T., Rajendran, P.S., Sahoo, D., Sim, S., Okamoto, J., Miranda, S.P., Johnston, D.M.,
448 Clarke, M.F., Quake, S.R., and Dalerba, P. (2013). Analysis of human colon tissue cell composition
449 using single-cell gene-expression PCR. Journal of biomolecular techniques: JBT *24*, S11.
450 86. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kaminska, B.,
451 Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine Learning Identifies Stemness
452 Features Associated with Oncogenic Dedifferentiation. Cell *173*, 338-354 e315.
453 10.1016/j.cell.2018.03.034.
454 87. Friedmann-Morvinski, D., and Verma, I.M. (2014). Dedifferentiation and reprogramming: origins of
455 cancer stem cells. EMBO Rep *15*, 244-253. 10.1002/embr.201338254.
456 88. Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L.,
457 Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of
458 colorectal cancer. Nat Med *21*, 1350-1356. 10.1038/nm.3967.
459 89. Isella, C., Terrasi, A., Bellomo, S.E., Petti, C., Galatola, G., Muratore, A., Mellano, A., Senetta, R.,
460 Cassenti, A., Sonetto, C., et al. (2015). Stromal contribution to the colorectal cancer transcriptome.
461 Nat Genet *47*, 312-319. 10.1038/ng.3224.
462 90. Valladares-Ayerbes, M., Blanco-Calvo, M., Reboredo, M., Lorenzo-Patiño, M.J., Iglesias-Díaz, P.,
463 Haz, M., Díaz-Prado, S., Medina, V., Santamarina, I., Pértega, S., et al. (2012). Evaluation of the
464 adenocarcinoma-associated gene AGR2 and the intestinal stem cell marker LGR5 as biomarkers in
465 colorectal cancer. Int J Mol Sci *13*, 4367-4387. 10.3390/ijms13044367.
466 91. Li, Y., Yuan, Y., Yang, L., Chen, H., Zhang, X., Wen, T., Liao, W., Zhao, M., Zhao, Z., and Hu, Q.
467 (2023). AT7867 Inhibits the Growth of Colorectal Cancer Stem-Like Cells and Stemness by
468 Regulating the Stem Cell Maintenance Factor Ascl2 and Akt Signaling. Stem Cells Int *2023*,
469 4199052. 10.1155/2023/4199052.
470 92. Jubb, A.M., Chalasani, S., Frantz, G.D., Smits, R., Grabsch, H.I., Kavi, V., Maughan, N.J., Hillan,
471 K.J., Quirke, P., and Koeppen, H. (2006). Achaete-scute like 2 (ascl2) is a target of Wnt signalling
472 and is upregulated in intestinal neoplasia. Oncogene *25*, 3445-3457. 10.1038/sj.onc.1209382.
473 93. Durinikova, E., Kozovska, Z., Poturnajova, M., Plava, J., Cierna, Z., Babelova, A., Bohovic, R.,
474 Schmidtova, S., Tomas, M., Kucerova, L., and Matuskova, M. (2018). ALDH1A3 upregulation and
475 spontaneous metastasis formation is associated with acquired chemoresistance in colorectal cancer
476 cells. BMC Cancer *18*, 848. 10.1186/s12885-018-4758-y.

477 94. Chen, G., She, X., Yin, Y., Ma, J., Gao, Y., Gao, H., Qin, H., and Fang, J. (2022). Targeting
478 TM4SF1 exhibits therapeutic potential via inhibition of cancer stem cells. Signal Transduct Target
479 Ther *7*, 350. 10.1038/s41392-022-01177-7.
480 95. Tang, Q., Chen, J., Di, Z., Yuan, W., Zhou, Z., Liu, Z., Han, S., Liu, Y., Ying, G., Shu, X., and Di, M.
481 (2020). TM4SF1 promotes EMT and cancer stemness via the Wnt/β-catenin/SOX2 pathway in
482 colorectal cancer. J Exp Clin Cancer Res *39*, 232. 10.1186/s13046-020-01690-z.
483 96. Matsui, W.H. (2016). Cancer stem cell signaling pathways. Medicine (Baltimore) *95*, S8-S19.
484 10.1097/MD.0000000000004765.
485 97. Yang, L., Shi, P., Zhao, G., Xu, J., Peng, W., Zhang, J., Zhang, G., Wang, X., Dong, Z., Chen, F.,
486 and Cui, H. (2020). Targeting cancer stem cell pathways for cancer therapy. Signal Transduct
487 Target Ther *5*, 8. 10.1038/s41392-020-0110-5.
488 98. Gierut, J., Zheng, Y., Bie, W., Carroll, R.E., Ball-Kell, S., Haegebarth, A., and Tyner, A.L. (2011).
489 Disruption of the mouse protein tyrosine kinase 6 gene prevents STAT3 activation and confers
490 resistance to azoxymethane. Gastroenterology *141*, 1371-1380, 1380.e1371-1372.
491 10.1053/j.gastro.2011.06.071.
492 99. Liu, C., Pan, Z., Chen, Q., Chen, Z., Liu, W., Wu, L., Jiang, M., Lin, W., Zhang, Y., Zhou, R., and
493 Zhao, L. (2021). Pharmacological targeting PTK6 inhibits the JAK2/STAT3 sustained stemness and
494 reverses chemoresistance of colorectal cancer. J Exp Clin Cancer Res *40*, 297. 10.1186/s13046-
495 021-02059-6.
496 100. Zhao, B., Li, L., Tumaneng, K., Wang, C.Y., and Guan, K.L. (2010). A coordinated phosphorylation
497 by Lats and CK1 regulates YAP stability through SCF(beta-TRCP). Genes Dev *24*, 72-85.
498 10.1101/gad.1843810.
499 101. Zage, P.E., Huo, Y., Subramonian, D., Le Clorennec, C., Ghosh, P., and Sahoo, D. (2023).
500 Identification of a novel gene signature for neuroblastoma differentiation using a Boolean implication
501 network. Genes Chromosomes Cancer *62*, 313-331. 10.1002/gcc.23124.
502 102. Arafeh, R., Shibue, T., Dempster, J.M., Hahn, W.C., and Vazquez, F. (2025). The present and
503 future of the Cancer Dependency Map. Nat Rev Cancer *25*, 59-73. 10.1038/s41568-024-00763-x.
504 103. Aasebø, K., Dragomir, A., Sundström, M., Mezheyeuski, A., Edqvist, P.H., Eide, G.E., Ponten, F.,
505 Pfeiffer, P., Glimelius, B., and Sorbye, H. (2020). CDX2: A Prognostic Marker in Metastatic
506 Colorectal Cancer Defining a Better. Front Oncol *10*, 8. 10.3389/fonc.2020.00008.

507 104. Li, X., Sun, K., Liao, X., Gao, H., Zhu, H., and Xu, R. (2020). Colorectal carcinomas with mucinous
508 differentiation are associated with high frequent mutation of KRAS or BRAF mutations, irrespective
509 of quantity of mucinous component. BMC Cancer *20*, 400. 10.1186/s12885-020-06913-2.
510 105. Feng, Y., Bommer, G.T., Zhao, J., Green, M., Sands, E., Zhai, Y., Brown, K., Burberry, A., Cho,
511 K.R., and Fearon, E.R. (2011). Mutant KRAS promotes hyperplasia and alters differentiation in the
512 colon epithelium but does not expand the presumptive stem cell pool. Gastroenterology *141*, 1003-
513 1013 e1001-1010. 10.1053/j.gastro.2011.05.007.
514 106. Brandt, R., Sell, T., Luthen, M., Uhlitz, F., Klinger, B., Riemer, P., Giesecke-Thiel, C., Schulze, S.,
515 El-Shimy, I.A., Kunkel, D., et al. (2019). Cell type-dependent differential activation of ERK by
516 oncogenic KRAS in colon cancer and intestinal epithelium. Nat Commun *10*, 2919. 10.1038/s41467-
517 019-10954-y.
518 107. Aasebø, K., Dragomir, A., Sundström, M., Mezheyeuski, A., Edqvist, P.-H., Eide, G.E., Ponten, F.,
519 Pfeiffer, P., Glimelius, B., and Sorbye, H. (2020). CDX2: a prognostic marker in metastatic
520 colorectal cancer defining a better BRAF mutated and a worse KRAS mutated subgroup. Frontiers
521 in oncology *10*, 8.
522 108. Slik, K., Turkki, R., Carpén, O., Kurki, S., Korkeila, E., Sundström, J., and Pellinen, T. (2019). CDX2
523 loss with microsatellite stable phenotype predicts poor clinical outcome in stage II colorectal
524 carcinoma. The American journal of surgical pathology *43*, 1473-1482.
525 109. Delhorme, J.-B., Bersuder, E., Terciolo, C., Vlami, O., Chenard, M.-P., Martin, E., Rohr, S., Brigand,
526 C., Duluc, I., and Freund, J.-N. (2022). CDX2 controls genes involved in the metabolism of 5-
527 fluorouracil and is associated with reduced efficacy of chemotherapy in colorectal cancer.
528 Biomedicine & Pharmacotherapy *147*, 112630.
529 110. Takakura, Y., Hinoi, T., Oue, N., Sasada, T., Kawaguchi, Y., Okajima, M., Akyol, A., Fearon, E.R.,
530 Yasui, W., and Ohdan, H. (2010). CDX2 regulates multidrug resistance 1 gene expression in
531 malignant intestinal epithelium. Cancer research *70*, 6767-6778.
532 111. Fontana, E., Meyers, J., Sobrero, A., Iveson, T., Shields, A.F., Taieb, J., Yoshino, T., Souglakos, I.,
533 Smyth, E.C., and Lordick, F. (2021). Early-Onset Colorectal Adenocarcinoma in the IDEA Database:
534 Treatment Adherence, Toxicities, and Outcomes With 3 and 6 Months of Adjuvant Fluoropyrimidine
535 and Oxaliplatin. Journal of Clinical Oncology *39*, 4009-4019.

536 112. Sakamoto, N., Feng, Y., Stolfi, C., Kurosu, Y., Green, M., Lin, J., Green, M.E., Sentani, K., Yasui,
537 W., and McMahon, M. (2017). BRAFV600E cooperates with CDX2 inactivation to promote serrated
538 colorectal tumorigenesis. Elife *6*, e20331.
539 113. Tong, K., Pellón-Cárdenas, O., Sirihorachai, V.R., Warder, B.N., Kothari, O.A., Perekatt, A.O.,
540 Fokas, E.E., Fullem, R.L., Zhou, A., and Thackray, J.K. (2017). Degree of tissue differentiation
541 dictates susceptibility to BRAF-driven colorectal cancer. Cell reports *21*, 3833-3845.
542 114. Felipe De Sousa, E.M., Wang, X., Jansen, M., Fessler, E., Trinh, A., De Rooij, L.P., De Jong, J.H.,
543 De Boer, O.J., Van Leersum, R., and Bijlsma, M.F. (2013). Poor-prognosis colon cancer is defined
544 by a molecularly distinct subtype and develops from serrated precursor lesions. Nature medicine *19*,
545 614-618.
546 115. Coskun, M. (2014). The role of CDX2 in inflammatory bowel disease. Danish medical journal *61*,
547 B4820-B4820.
548 116. Coskun, M., Olsen, A.K., Holm, T.L., Kvist, P.H., Nielsen, O.H., Riis, L.B., Olsen, J., and Troelsen,
549 J.T. (2012). TNF-α-induced down-regulation of CDX2 suppresses MEP1A expression in colitis.
550 Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease *1822*, 843-851.
551 117. Yin, F., Xie, H., Lai, J., Chen, Y., Dong, J., Zhang, X., and Liu, X. (2020). Double negativity for
552 expression of YAP1 and CDX2 defines an aggressive type of colitis-associated cancer. Anticancer
553 research *40*, 5411-5416.
554 118. Lourenço, F.C., and Winton, D.J. (2018). Metaplastic Cdx2-depleted cells can be very disruptive
555 neighbors. J Exp Med *215*, 725-727. 10.1084/jem.20180215.
556 119. Balbinot, C., Armant, O., Elarouci, N., Marisa, L., Martin, E., De Clara, E., Onea, A., Deschamps, J.,
557 Beck, F., Freund, J.N., and Duluc, I. (2018). The Cdx2 homeobox gene suppresses intestinal
558 tumorigenesis through non-cell-autonomous mechanisms. J Exp Med *215*, 911-926.
559 10.1084/jem.20170934.
560 120. Bae, J.M., Lee, T.H., Cho, N.-Y., Kim, T.-Y., and Kang, G.H. (2015). Loss of CDX2 expression is
561 associated with poor prognosis in colorectal cancer patients. World journal of gastroenterology:
562 WJG *21*, 1457.
563 121. Reggiani Bonetti, L., Lionti, S., Vitarelli, E., and Barresi, V. (2017). Prognostic significance of CDX2
564 immunoexpression in poorly differentiated clusters of colorectal carcinoma. Virchows Archiv *471*,
565 731-741.

566 122. Alarid-Escudero, F. (2021). CDX2 testing+ FOLFOX for CDX2-negative colon cancer patients cost-
567 effective strategy. PharmacoEconomics & Outcomes News *892*, 9-27.
568 123. Kawai, H., Tomii, K., Toyooka, S., Yano, M., Murakami, M., Tsukuda, K., and Shimizu, N. (2005).
569 Promoter methylation downregulates CDX2 expression in colorectal carcinomas. Oncology reports
570 *13*, 547-551.
571 124. Sun, X., Yang, Q., Rogers, C.J., Du, M., and Zhu, M.-J. (2017). AMPK improves gut epithelial
572 differentiation and barrier function via regulating Cdx2 expression. Cell Death & Differentiation *24*,
573 819-831.
574 125. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B.,
575 Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine Learning Identifies Stemness
576 Features Associated with Oncogenic Dedifferentiation. Cell *173*, 338-354.e315.
577 10.1016/j.cell.2018.03.034.
578 126. Foucquier, J., and Guedj, M. (2015). Analysis of drug combinations: current methodological
579 landscape. Pharmacol Res Perspect *3*, e00149. 10.1002/prp2.149.
580 127. Wu, B., Tao, L., Yang, D., Li, W., Xu, H., and He, Q. (2020). Development of an Immune Infiltration-
581 Related Eight-Gene Prognostic Signature in Colorectal Cancer Microenvironment. Biomed Res Int
582 *2020*, 2719739. 10.1155/2020/2719739.
583 128. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño,
584 V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and
585 immune cell admixture from expression data. Nat Commun *4*, 2612. 10.1038/ncomms3612.
586 129. Győrffy, B. (2024). Integrated analysis of public datasets for the discovery and validation of survival-
587 associated genes in solid tumors. Innovation (Camb) *5*, 100625. 10.1016/j.xinn.2024.100625.
588 130. Deng, F., Peng, L., Li, Z., Tan, G., Liang, E., Chen, S., Zhao, X., and Zhi, F. (2018). YAP triggers
589 the Wnt/β-catenin signalling pathway and promotes enterocyte self-renewal, regeneration and
590 tumorigenesis after DSS-induced injury. Cell Death Dis *9*, 153. 10.1038/s41419-017-0244-8.
591 131. Park, H.W., Kim, Y.C., Yu, B., Moroishi, T., Mo, J.S., Plouffe, S.W., Meng, Z., Lin, K.C., Yu, F.X.,
592 Alexander, C.M., et al. (2015). Alternative Wnt Signaling Activates YAP/TAZ. Cell *162*, 780-794.
593 10.1016/j.cell.2015.07.013.

132. Guo, R.J., Funakoshi, S., Lee, H.H., Kong, J., and Lynch, J.P. (2010). The intestine-specific transcription factor Cdx2 inhibits beta-catenin/TCF transcriptional activity by disrupting the beta-catenin-TCF protein complex. Carcinogenesis *31*, 159-166. 10.1093/carcin/bgp213.

133. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

134. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res *41*, D991-995. 10.1093/nar/gks1193.

135. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res *35*, D760-765. 10.1093/nar/gkl887.

136. Sahoo, D. (2012). The power of boolean implication networks. Front Physiol *3*, 276. 10.3389/fphys.2012.00276.

137. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.

138. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. Nucleic Acids Res *48*, D498-D503. 10.1093/nar/gkz1031.

139. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J.E., Mestdagh, P., and De Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun *11*, 5650. 10.1038/s41467-020-19015-1.

140. Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y.Y., Carré, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., et al. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell Rep *26*, 1627-1640.e1627. 10.1016/j.celrep.2019.01.041.

141. Varghese, F., Bukhari, A.B., Malhotra, R., and De, A. (2014). IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. PLoS One *9*, e96801. 10.1371/journal.pone.0096801.

622 142. Sayed, I.M., Suarez, K., Lim, E., Singh, S., Pereira, M., Ibeawuchi, S.R., Katkar, G., Dunkel, Y.,
623 Mittal, Y., Chattopadhyay, R., et al. (2020). Host engulfment pathway controls inflammation in
624 inflammatory bowel disease. FEBS J *287*, 3967-3988. 10.1111/febs.15236.
625 143. Sayed, I.M., El-Hafeez, A.A.A., Maity, P.P., Das, S., and Ghosh, P. (2021). Modeling colorectal
626 cancers using multidimensional organoids. Adv Cancer Res *151*, 345-383.
627 10.1016/bs.acr.2021.02.005.
628 144. Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H.,
629 Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus
630 structures in vitro without a mesenchymal niche. Nature *459*, 262-265. 10.1038/nature07935.
631 145. Sato, T., Stange, D.E., Ferrante, M., Vries, R.G., Van Es, J.H., Van den Brink, S., Van Houdt, W.J.,
632 Pronk, A., Van Gorp, J., Siersema, P.D., and Clevers, H. (2011). Long-term expansion of epithelial
633 organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium.
634 Gastroenterology *141*, 1762-1772. 10.1053/j.gastro.2011.07.050.
635 146. Sayed, I.M., Chakraborty, A., Abd El-Hafeez, A.A., Sharma, A., Sahan, A.Z., Huang, W.J.M.,
636 Sahoo, D., Ghosh, P., Hazra, T.K., and Das, S. (2020). The DNA Glycosylase NEIL2 Suppresses.
637 Cells *9*. 10.3390/cells9091980.
638 147. Sayed, I.M., Sahan, A.Z., Venkova, T., Chakraborty, A., Mukhopadhyay, D., Bimczok, D., Beswick,
639 E.J., Reyes, V.E., Pinchuk, I., Sahoo, D., et al. (2020). infection downregulates the DNA glycosylase
640 NEIL2, resulting in increased genome damage and inflammation in gastric epithelial cells. J Biol
641 Chem *295*, 11082-11098. 10.1074/jbc.RA119.009981.
642 148. Grabinger, T., Luks, L., Kostadinova, F., Zimberlin, C., Medema, J.P., Leist, M., and Brunner, T.
643 (2014). Ex vivo culture of intestinal crypt organoids as a model system for assessing cell death
644 induction in intestinal epithelial cells and enteropathy. Cell Death Dis *5*, e1228.
645 10.1038/cddis.2014.183.
646 149. Smith, M.C., Luker, K.E., Garbow, J.R., Prior, J.L., Jackson, E., Piwnica-Worms, D., and Luker,
647 G.D. (2004). CXCR4 regulates growth of both primary and metastatic breast cancer. Cancer Res
648 *64*, 8604-8612. 10.1158/0008-5472.CAN-04-1844.
649