Original Research

Reinstatement of CDX2 as a differentiation therapy for colorectal cancers

Authors: Saptarshi Sinha[1], Joshua Alcantara[1, 2], Kevin Perry[1, 2], Vanessa Castillo[1], Celia R. Espinoza[1], Sahar Taheri[3], Eleadah Vidales[1, 2], Courtney Tindle[1, 2], Adel Adel[1], Siamak Amirfakhri[5], Joseph R. Sawires[6], Jerry Yang[6], Michael Bouvet[5,7], Debashis Sahoo[3,4,7] and Pradipta Ghosh[1,2,7,8‡]

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 (CoRE), 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 Paediatrics, School of Medicine, University of California San Diego, La Jolla, CA 92093.

[5]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.

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

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

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

‡**Corresponding author**: 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, Rm 232, 239; La Jolla, CA 92093. Phone: 858-822-7633. Email: prghosh@ucsd.edu.

Abstract

Background: Lack of expression of the transcription factor CDX2 identifies a subset of poorly differentiated colorectal cancers (CRCs) that are associated with lower overall (OS) and disease-free (DFS) survival, independent of ethnicity, MSI status, or stage. Reinstatement of CDX2 is predicted to lower the risk of death and recurrence by 50% but such a therapeutic strategy is yet to emerge.

Methods: A network-based computational model, validated using healthy colon and CRC tissues in diverse gene expression datasets (~5,000 human and >300 mouse samples), was used to identify therapeutic targets that can augment CDX2 expression. The top candidate with available clinical-grade drug (PRKAB1) was tested on 3 models: CRC lines *in vitro*, xenografts in mice, and in a prospective cohort of healthy (n = 3) and CRC (n = 23) patient-derived organoids (PDOs).

Results: In all 3 models, PRKAB1-agonism predictably shifted the network, induced *CDX2* and crypt differentiation signatures and showed selective cytotoxicity towards *CDX2*-negative CRCs. Xenotransplantation studies in mice reduced tumour volume by 68% and abolished mortality (Hazard ratio; H.R = 0.09). Effective pairing of therapeutic efficacy (IC50 for PRKAB1-agonism) and biomarker (*CDX2*-low state) was validated in an independent cohort of PDOs. A multivariate analysis revealed *CDX2*-low state as the key determinant of therapeutic efficacy. A 50-gene signature of therapeutic response tested on 9 cohorts (n = 1701 patients) showed that *CDX2*-reinstatement therapy will reduce the risk of mortality/recurrence by ~50%.

Conclusions: CDX2-reinstatement therapy selectively triggers cell differentiation and death in CDX2-low CRCs. Realization of the benefit of such therapy-biomarker pairing requires confirmatory studies in randomized clinical trials.

Introduction

CDX2 is a nuclear homeobox transcription factor and a tumour suppressor that belongs to the caudal-related family of CDX homeobox genes[1-4]. In 2016, an unbiased search using a novel bioinformatic approach (Boolean logic[5,6]) looking for a biomarker of tumour stemness/differentiation status identified CDX2 as the statistically top ranked candidate biomarker of mature colon epithelial tissues, and by that token, of well-differentiated tumours [7]. The study predicted *CDX2* as the top developmentally regulated pro-differentiation candidate gene whose expression in colorectal cancer (CRCs) was inversely related to the "activated leukocyte cell adhesion molecule" (*ALCAM*/CD166); the latter is a marker of stem-like colon epithelial cells at the bottom of colon crypts[8,9] and on human colon cancer cells with high tumorigenic capacity in mouse xenograft models[10]. Validation studies using immunohistochemistry (IHC) showed that compared to patients with CDX2-pos tumors, those with a loss of CDX2 (accounting for approximately 9% of patients) had lower disease-free survival (DFS) and benefited from an adjuvant chemotherapy, even in stage II CRC.

Numerous groups have since reproduced the prognostic claims [11-24]. Thirty-two independent studies, involving >13,000 unique subjects, most that used IHC and one that used IHC in combination with quantitative mass spectrometry [25], have determined that CDX2-neg CRCs are associated with lower overall survival (OS) and DFS independent of ethnicity, DNA mismatch repair (MMR) status or stage [7,20,24,26] [**Figure 1**-*Step 1*]. Absence of CDX2 in tumours 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 [26-28]. Presence of CDX2 lowers the risk of death and recurrence by 50% and 52%, respectively[24]; this finding is more pronounced in stage II and III CRCs, with up to 70% reduction in the risk of death. These findings drove consensus that pharmacologic reinstatement of CDX2 in CRCs is a clinically important therapeutic goal [see **Figure 1**-*Step 1*], one that is also challenging because promising candidates have not emerged.

Here we adopted the same Boolean-logic-based bioinformatics approach[5,6] to screen for therapeutic targets that can induce CDX2 and promote tumour differentiation.

Methods

Detailed computational and experimental methods are presented in ***Supplementary Materials*** and mentioned briefly here.

Bioinformatics analysis of gene-expression array databases. A complete catalogue of all the NCBI-GEO sample number identifiers (GSEIDs) of individual gene-expression array experiments used for model training and validation are provided (**Supplemental Information 1**).

Patient derived organoids. To enhance reproducibility and rigor, all CRC PDOs used in this work were acquired from the American Type Culture Collection (ATCC), which acquired these from diverse clinical sources [The Human Cancer Models Initiative (HCMI) US National Cancer Institute (NCI), Cancer Research UK (CRUK), the Wellcome Sanger Institute (WSI), and the foundation Hubrecht Organoid Technology (HUB)]. Relevant demographic, clinical and genetic information is provided (**Table 1**).

Statistical tests. Optimal cut-off values of gene expression levels are derived using a threshold defined by *StepMiner* algorithm in the individual dataset[29]. Gene signature is used to classify sample categories and the performance of the multi-class classification is measured by ROC-AUC (receiver operating characteristics area under the curve) values. Standard t-tests were performed using python scipy.stats.ttest_ind package (version 0.19.0) with Welch's two sample t-test (unpaired, unequal variance (equal_var = False), and unequal sample size) parameters. Multiple hypothesis correction were performed by adjusting p-values with statsmodels.stats.multitest.multipletests (fdr_bh: Benjamini/Hochberg principles). Differential expression analysis was performed in DESeq2 in R (version 1.16.1). Kaplan–Meier analysis and determination of statistical significance by Log Rank test was performed using GraphPad Prism 9.1. Violin and Bubble plots are created using python seaborn package version 0.12. BooleanNet algorithm is used to statistically determine Boolean Implication relationships between genes (statistic > 3 and error-rate < 0.1). The source code for Boolean analysis framework is https://github.com/sinha7290/Prodiff[30]. The best estimate of effect size (ratio between the difference in mean and standard deviation) is 2.5 in PDO and 1.0 in animal experiment for CDX2 induction using PRKAB1-agonist. The sample size in each group of at least 4 PDOs and 17 animals should suffice ($\alpha = 0.05$ and power = 0.80).

Data availability. All transcriptomic datasets generated in this work has been deposited at NCBI GEO [GSE237623, GSE237624, GSE237625]. They can be accessed with password now (see Cover Letter) and will be released immediately upon publication.

Study Design. The study was designed by the first and last author. All authors contributed to the gathering, analysis, and interpretation of the data. The first draft of the manuscript was written by the first and last author and was subsequently edited based on comments from all authors.

Ethics statement. This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All animal studies were approved by the San Diego Veterans Administration Medical Center Institutional Animal Care and Use Committee (protocol A17-020; PI Bouvet). All colorectal cancer PDOs were obtained from the American Type Culture Collection (ATCC). The latter is a nonprofit, global biological resource centre and a distributor of organoid models curated as part of Human Cancer Models Initiative (HCMI) for enhancing reproducibility. Deidentified case-associated data, including that of molecular characterization of the models, the tissues from which they were derived were used here, whereas available harmonized data are accessible through NCI's Genomic Data Commons. All healthy colon PDOs were obtained from individuals presenting for routine colonoscopy as part of colon cancer screening at UC San Diego, with their informed consent to participate in a study protocol approved by an Institutional Review Board (IRB; protocol #190105; PI Ghosh).

Role of funders

Funders of the study had no role in study design, data collection, data analyses, interpretation, or writing of report.

Results

A network-rationalized approach

We used the computational platform, Boolean Network Explorer (*BoNE*)[31], and *CDX2* as a 'seed' gene to create a model of progressive gene regulatory events that occur during differentiation in the colon (**Figure 1**-*Step 2;* **Figure S1A-B**). The model was trained on a pooled heterogeneous dataset that included maximum possible diversity (1662 CRCs, 68 adenomas and 170 normal tissues) (**Figure S1C-D**). Machine learning on the *CDX2*-proximal immediate gene clusters (clusters #2 and #3, **Figure S1D**) helped statistically rank 5 genes as top targets (**Figure 1**-*Step 2;* **Figure S1E**). By virtue of how these genes relate to *CDX2* within the network, upregulation of any of these 5 genes (presumably via target/pathway agonists) is predicted to also upregulate *CDX2*. Validation studies on an additional 2,937 human (**Figure S1F**) and 318 mouse (**Figure S2**) colon samples showed that these 5 genes remain relevant across species and diverse cohorts.

In search of potent and specific drugs to modulate the 5 targets, we searched the PDB [Protein Data Bank; https://www.rcsb.org] for structure resolved selective agonistic modulators and found 25 molecules/structures for PRKAB1, but none for the remaining targets. To determine the most clinically advanced candidate, the PDB-derived hits were cross-searched through the Clinical Trials.gov portal, which led to the identification of a clinical-grade drug, PF-06409577 (henceforth, PF). Although lacking clinical indication at present, the drug was deemed safe in a randomized, double-blind, placebo-controlled Phase I trial [*ClinicalTrials.gov* ID: NCT02286882]. Furthermore, PF was recently shown to demonstrate target specificity and the ability to upregulate the *PRKAB1* transcript and the translated protein in the colonic epithelium ex-vivo and mouse colon in vivo[31].

The network model predicted that agonism of the protein that is translated by the *PRKAB1* gene, i.e., Protein Kinase AMP-activated Non-Catalytic Subunit Beta-1 will-- (i) upregulate *PRKAB1* expression and, by virtue of their "*equivalent*" relationship, also upregulate the other genes that co-cluster with it (**Figure S1D**); (ii) upregulate *CDX2* and the *CDX2*-proximal gene clusters (#2 and #3) by virtue of "hi=>hi" relationship (**Figure S1D**); and (iii) downregulate stemness genes by virtue of "hi=>lo" relationship (*CCDC88A* and others in C#9; **Figure S1D**). Notably, *PRKAB1* has a universally conserved relationship with *CDX2*, across diverse human and mouse CRC datasets, including early onset CRCs (**Figure S2A-D**). That PRKAB1 (but not the other β-subunit, PRKAB2) is strongly expressed in the GI tract[31] and that PRKAB1-agonists trigger cell polarity in the gut epithelium[32-35], served as additional justifications for its rationalized selection as target.

For target validation, we prioritized three models, based on their *CDX2*-low (target tumour type) and high (expected to be negative control) status. Alongside the traditional readouts of anti-cancer effects, a set of additional end points were *a priori* formulated as metrics of success; therapeutic goal was set such that all three models must pass all three end points (see **Figure 1**-*Step 3*). Prior attempts that successfully induced *CDX2* temporarily[36], but ultimately failed to emerge as differentiation therapy, had mixed impact on these metrics of success (**Figure S3**). Finally, to track successful *CDX2*-reinstatement and assess its impact on

survival, a 50-gene signature of therapeutic response is derived and tested on numerous independent cohorts (see **Figure 1**-*Step 4*).

Target validation on CRC cell lines and xenograft models

Next, we prioritized two poorly differentiated CRC cell lines (HCT116 and Sw480 cells; both exhibited the lowest levels of *CDX2* transcripts among a panel of 26 sequenced CRC cell lines that were profiled by microarrays; GSE10843) as the optimal models for hypothesis testing, and one of the two well-differentiated cells with high *CDX2* levels (DLD1) as the negative control (**Figure 2A**). PF was capable of inducing cell death within 48 h in *CDX2*-low HCT116 and Sw480 cells but not the *CDX2*-high DLD1 cells (IC50's in 6-7 µM range vs > 40 µM, respectively; **Figure 2B**; **Figure S4A**). In both *CDX2*-low cell lines, PF-induced cell death was preceded by induction of differentiation and suppression of stemness markers, at both transcript (**Figure 2C**) and protein (**Figure S4B-E**) levels in a dose-dependent manner. These effects were associated also with dose-dependent increase in apoptosis (**Figure S5A-H**). PF failed to reinstate *CDX2*, change other markers of stemness and differentiation (**Figure 2D**) or induce cell death (**Figure 2E**) in *PRKAB1*-depleted cells, thereby demonstrating that its anti-cancer action is target specific. Finally, Metformin which is a global, non-specific and indirect agonist of all forms of the trimeric AMP-kinase (regardless of whether PRKAB1 or PRKAB2 is the β-subunit within the αβγ trimer) failed to trigger cell death or induce *CDX2* expression even at very high doses (**Figure 2B**; **Figure S5I-K**), suggesting that PRKAB1-specific effect is required for therapeutic potency.

When implanted subcutaneously as xenografts in nude mice (**Figure 2F**), HCT116 tumors revealed a 68% reduction in tumour volume (total AUC for vehicle and PF time series were 51280.23 and 26676.77, respectively; **Figure 2G-I**) and an ~62.48% increase in survival (**Figure 2J**; Mantel-Haenszel Hazard Ratio of 0.09 for reduction in death rates). PF promoted glandular structures within tumour (see legend **Figure S6A**), reinstated *CDX2*, induced differentiation markers and suppressed stemness markers, as determined by qPCR (**Figure 2K**). Reinstatement of CDX2 protein and its nuclear localization, alongside upregulation of PRKAB1 protein were confirmed also by immunohistochemistry (IHC; **Figure 2L-M; Figure S6B**). Consistent with the notion that colon stem cells divide very slowly [37], PF triggered a ~3-fold increase in mitotic bodies (**Figure S6C**; **Figure 2M**). Conventional markers of tumour differentiation were also observed, such as a reduction of nuclear β-Catenin (**Figure S6D**; **Figure 2M**) with concomitant increase in junctional β-Catenin, where it co-localized with E-Cadherin (**Figure 2N**; **Figure S6E**).

Target validation in patient-derived organoids

We next assessed the impact of PF on 23 CRC PDOs (recruited as two prospective independent cohorts) and 3 healthy colon PDOs (see **Table** and *Methods*; **Figure 3A**). The anti-cancer action (as determined by low IC50 values) was limited only to *CDX2*-low CRC PDOs, whereas the drug was virtually inert (with IC50>40 µM) against *CDX2*-high PDOs as well as healthy colon PDOs (**Figure 3B**). Both *CDX2* ('seed' gene in the

network) and *PRKAB1* (network-derived therapeutic target) showed significant positive correlations with IC50 (**Figure 3B**). By contrast, *CCDC88A* (a network-derived stemness marker; see **Figure S1D**), showed a significant inverse correlation with IC50 (**Figure 3B**). Hierarchical unsupervised clustering performed using just 3 network-derived genes (*CDX2*, *PRKAB1* and *CCDC88A*) perfectly classified sensitive from resistant PDOs (**Figure 3C**). PF treatment enhanced differentiation and suppressed stemness (**Figure 3D**), as observed in CRC cell lines (**Figure 2C**) and in xenografts (**Figure 2K**) previously, but these impacts were restricted only to *CDX2*-low CRC PDOs. A multivariate analysis performed on the cohort of CRC PDOs (see **Supplemental Information 3** for all variables included) showed that *CDX2*-status was the strongest variable that determined IC50 (**Figure 3E**). In addition, two additional co-variates emerged: (i) histological grade and $KRAS^{mut}$ status, which indicates that PF was less likely to work on moderate/well-differentiated tumors that tend to frequently harbor oncogenic *KRAS* mutations (**Figure 3E**; **Table**).

The relationship between CDX2-induction therapy and KRAS-status suggested that CDX2 may not be prognostic among $KRAS^{mut}$ group because its expression remains high (i.e., not lost). A univariate analysis on a large cohort of CRCs with known driver mutations and patient outcome confirmed that *CDX2* was indeed a determinant of outcome in tumors with $KRAS^{WT}$ (**Figure 3F**-*left*), but not in those with $KRAS^{mut}$ (**Figure 3F**-*right*). These findings suggest that although relatively small, our PDO cohort is sufficiently diverse and representative of the heterogeneity in the clinic to have revealed the interplay between *CDX2* and $KRAS^{mut}$.

Indicators of therapeutic efficacy and target cell specificity

RNA seq on all 3 models─ CRC cell lines, xenografts, and PDOs—confirmed that PRKAB1 agonism met 3 *a priori* formulated metrics of success: (i) *CDX2* was reinstated, which was the primary therapeutic goal (bubble plots in **Figure 4A**-*top row*; **4B**; violin plots in **Figure S7A-C**), (ii) markers of crypt top enterocytes and Goblet cells, which are established surrogate markers of differentiation in the colon crypt [6,38] were induced (**Figure 4A**-*middle row*; **Figure S7D-F**), and (iii) the network was perturbed predictably, i.e., genes in clusters #2 and #3 were induced (**Figure 4A**-*bottom row*; **Figure S7G-I**). Of note, all three metrics were fulfilled selectively in *CDX2*-low CRC cell lines, xenografts and PDOs, but not in *CDX2*-high PDOs or in healthy PDOs (**Figure 4A**-*columns #5 and #6*; **Figure S7C, S7F, S7I**).

A recent effort, using an innovative machine-learning algorithm extracted transcriptomic and epigenetic feature sets derived from healthy (normal) stem cells and their differentiated progeny; the work released novel gene signatures of oncogenic dedifferentiation[39]. Those signatures revealed that *CDX2*-reinstatement therapy had no effect on normal stem cell-related genes in healthy and CRC PDOs (**Figure 4C**) but specifically reduced the expression of genes that impart cancer-related stemness (**Figure 4D**), those that enable somatic cell reprogramming (i.e., the Yamanaka factors which have been implicated in the origin of cancer stem cells (CSCs[40]; **Figure 4E**) and those that define CRC-specific CSCs (**Figure 4F**).

Taken together these findings show that *CDX2*-reinstatement therapy selectively downregulates CSC-programs that are unique to *CDX2*-low tumors, thereby sparing normal stem cells in the healthy colon tissue as well as *CDX2*-high tumour cells.

Potential for synergy with other therapeutic modalities

Lack of CDX2 expression in Stage II-III CRCs has been previously shown to be associated with benefits from adjuvant chemotherapy [7]; CDX2-positive tumours lack such benefit. We asked how therapeutic reinstatement of CDX2 in *CDX2*-low tumours may impact response to chemotherapy. PF treatment consistently altered a panel of patient-derived gene signatures across all models tested; it induced signatures of favourable response and suppressed signatures of non-response to diverse chemotherapeutic agents (e.g., 5-FU, Oxaliplatin, FOLFIRI) (**Figure 4G-i**). Consistent with these gene expression changes, PF showed synergy with a panel of chemotherapeutic agents (see legend, **Figure 4H**).

Similar observations were made when we extended the analysis to include other treatment modalities; PF consistently induced gene signatures of response to anti-EGFR (Cetuximab; **Figure 4G-ii**) and other tyrosine kinase inhibitors (TKIs; **Figure 4G-iii**), with a few notable exceptions (Afatinib and Gefitinib; **Figure 4G-iii**). PF also induced gene signatures of response and suppressed signatures of non-response to immunotherapy (**Figure 4G-iv**). The latter is consistent with the observation that PF also robustly induced a consensus molecular subtype (cms)-1 in all three model systems tested (**Figure 4G-v**); the immune landscape in cms1 subtype is known to favour response to immune checkpoint inhibitors[41].

A molecular signature of therapeutic response and selectivity

To define the nature of the therapeutic response, we leveraged RNA seq data from all 3 models to conduct an integrated differential expression analysis (see *Methods*). Briefly, differentially expressed genes (DEGs) from two independent models (cell lines and xenografts; **Supplemental Information 4-5**) were further refined on a third model, the CRC PDOs (**Figure 5A**). A set of 50 DEGs emerged; these genes were expressed in *CDX2*-low CRC PDOs at levels significantly higher than healthy or *CDX2*-high CRC PDOs (**Figure 5B**; **Figure S8A-B**). We used this 50-gene signature as a computational tool to track the changes along the stemness→differentiation axis that occurs due to reinstatement of *CDX2* and found that PF suppressed these genes (**Figure 5B**; see **Supplemental Information 6**) to levels seen in healthy PDOs and *CDX2*-high CRC PDOs (**Figure S8A-B**).

This 50-gene signature was notable for the presence of several unique CSC-associated genes, such as, *AGR2* [42], *ASCL2* [43,44], *ALDH3A1* [45], and *TM4SF1* [46,47], that were not parts of CSC signatures identified by machine-learning (**Figure 4C-F**). The 50 genes were enriched for pathways that are known to augment CSC survival [48-51], e.g., PTK6-dependent JAK/STAT signaling, numerous growth factor receptor and non-receptor TK and the Hippo/YAP pathways and loss of TGFβ signaling (**Figure 5C**). The selective upregulation of these pro-survival pathways in *CDX2*-low CRCs (but not healthy PDOs or *CDX2*-high CRCs) suggests that

their concomitant and catastrophic withdrawal in the setting of *CDX2* reinstatement could be the reason for the observed target-cell selectivity.

Impact of CDX2 reinstatement therapy on survival

To assess the impact of differentiation therapy on disease outcome, we analyzed the ability of the 50-gene signature of therapeutic response to stratify relapse/disease-free survival (RFS/DFS) in diverse cohorts of CRC patients. Compared to patients whose tumour had a high 50-gene score (like the untreated *CDX2*-low PDOs), those with a low 50-gene score consistently had a better outcome across all datasets analyzed (**Figure 5D, G-I**). The odds ratio (O.R) was ~2.0 (**Supplemental Information 7**-*tab #1*), implying that high expression of the 50-gene signature carries a double risk of death/recurrence and that therapeutic reinstatement of *CDX2* will reduce that risk by 50%. A univariate analysis revealed that the 50-gene signature and *CDX2* were significant deterministic covariates of RFS (**Figure 5E**-*left*) but emerged as the better of the two in *KRAS*^{WT} (**Figure 5E**-*middle*). The 50-gene score also outperformed *CDX2* as a prognostic indicator, consistently across all cohorts tested (**Figure S8C-F**; **Supplemental Information 7**). These findings show that the positive impact of CDX2 reinstatement is better represented in the 50-gene score (compared to *CDX2* alone). Consistent with our observations earlier in the case of *CDX2* (**Figure 3E**), the 50-gene signature was not a factor in *KRAS*^{mut} tumors (**Figure 5E**-*right*). These results provide objective evidence that therapeutic suppression of the 50-gene signature is a desirable therapeutic goal that could translate to improved survival.

Discussion

The major discovery we report here is a network-rationalized strategy to push differentiation of CRC stem cells by reinstating *CDX2* expression and define the cellular, molecular, and translational impact of such reinstatement. Much like immunotherapy acts by reinvigorating a physiologic response, we show that *CDX2*-reistatement acts by reinvigorating yet another physiologic program, that of differentiation. We identify and validate one target (PRKAB1) and confirm the utility of a first-in-class agent with proven safety (PF-06409577) to exert a target- and cell-type specific pro-differentiation and anti-tumour effect. *CDX2*-reinstatement therapy also fulfilled a set of rigorous and objective metrics of success that were formulated *a priori*, and its efficacy can be objectively tracked using a set of companion biomarkers (*CDX2* expression and a 50-gene signature of therapeutic response). Mechanistically, we show that *CDX2*-reinstatement is associated with the suppression of key stemness genes and numerous pro-survival pathways to which *CDX2*-low CRC stem cells (but not normal stem cells) may be uniquely addicted. Finally, we provide evidence that *CDX2*-reinstatement therapy is expected to synergize with most existing therapeutic modalities and translate to a reduction in death/relapse by ~50%.

Translational potential and key considerations: We advocate for the findings to be translated into clinical trials, prioritized as follows. The study pinpointed *CDX2*-low CRCs, and more specifically, those with a high 50-gene signature, as the optimal target group that will maximally benefit from this form of therapy. That *CDX2*-high tumors were found to be more differentiated and more likely to harbor mutant *KRAS* (but not *BRAF*) is consistent with the fact that *KRAS*^{mut} promotes hyperplasia and mucinous differentiation[52], but not stemness[53] and that loss of CDX2 (as determined by IHC) is positively associated with *BRAF* ^{mut}[54] but inversely associated with *KRAS*^{mut}[54]. This intriguing distinction could be attributed to cell-type specific roles of *KRAS*/*BRAF*^{mut} reported recently[55]. *KRAS*^{mut} status, however, does not appear to have any bearing on the determination of therapeutic candidacy because the two cell lines that were prioritized here based on low *CDX2* expression levels are also *KRAS*^{mut}, [HCT116 (1 mutant allele; G13D) and Sw480 (both alleles mutated; G12V)], and are sensitive to PF. These findings are consistent with a population-based study on metastatic CRCs (mCRCs; n = 796) that showed loss of CDX2 (as determined by IHC) defines a small subgroup of *KRAS*^{mut} cases (~9%) with a worse prognosis[54], and hence, could benefit from differentiation therapy. We conclude that *KRAS*^{mut} status is a confounding co-variate that is associated with the differentiation status of a tumour; however, it should not override low-CDX2 and/or high-50-gene signature in the determination of who may benefit from differentiation therapy.

Because loss of *CDX2* is an adverse prognostic factor in stage IV and stage III/*BRAF*^{mut} CRCs[26,56], these subgroups can be prioritized for *CDX2*-reinstatement therapy with PRKAB1 agonist as single agent or together with other modalities.

Although *CDX2* is prognostic in stage II(MSS)[7,20,24]and stage III CRCs[7,57], the use of *CDX2*-reinstatement therapy in these subgroups deserves additional considerations, and even caution. In stage III

disease, adjuvant chemotherapy has been a standard of care since the 90's based on the Moertel trial[58]; however, many in this subgroup relapse despite chemotherapy (variable ~15-75% based on T-N substage[59]) and die of disease[60]. Because CDX2 expression is associated with increased expression of the multi-drug resistance gene, MDR1, which is implicated in the chemoresistance of *CDX2*-pos tumours[61,62], the use of PRKAB1 agonist in combination with conventional chemotherapeutic agents must wait despite demonstrable synergy in our studies. Considering these factors, PRKAB1 agonist as a single agent or with other modalities can instead be prioritized for those with *CDX2*-low tumors who relapse and/or progress to metastatic disease. This is feasible because CDX2 expression status in CRCs is not affected by metastasis or chemotherapy[63]. If those studies show benefit, PRKAB1 agonist as single agent could be considered as a first-line option for patients with *CDX2*-low tumours. Because stage II/III early-onset CRC is a subgroup that show worse outcome (poorer DFS) than its late-onset counterparts[64], and the *CDX2-PRKAB1* relationship is strongly conserved in this subgroup (**Figure S2A**), we advocate the inclusion of this group in future studies considering PRKAB1 agonists in stage II/III CRCs.

Finally, because the degree of differentiation dictates the susceptibility to *BRAF*-mut-driven tumourigenesis[65,66], early-stage patients with histopathologic criteria not qualifying for chemotherapy might be reconsidered for *CDX2*-reinstatement therapy if their tumour has focal or global loss of CDX2 and mutated *BRAF*.

Expected impact on pre-neoplastic lesions of the colon:

There is sufficient literature on how *CDX2*-loss (and hence reduced differentiation) is permissive to CRC initiation, especially along the serrated pathway[65-67] and in the setting of inflammatory bowel disease-associated CRCs[68-70]. A recent study using mosaic *CDX2*-/- mouse models showed that *CDX2* serves as a tumour suppressor gene in a non–cell-autonomous manner; *CDX2*-/- metaplastic pre-cancerous cells could induce tumourigenesis from neighbouring non-metaplastic cells without themselves becoming cancerous. Focused studies on PDOs derived from such high-risk population are expected to reveal if PRKAB1 agonists such as PF can serve as prophylactic therapy for those at highest risk of CRCs.

Study limitations:

How PRKAB1-agonism translates to *CDX2*-reinstatement was not studied; future efforts will assess whether aberrant promoter methylation that has been implicated previously[71,72] is reversed. Given the relatively small size of PDO cohorts, these results will need to be further validated. We advocate for these findings to be confirmed within the framework of randomized clinical trials, in conjunction with RNA sequencing studies.

Conclusions

Despite level 1 evidence showing the prognostic role of CDX2, its independent validation through numerous studies[7,20,21,23-25,36,73,74], and a compelling pharmacoeconomic justification (i.e., cost $ 50 000 per QALY) for a CDX2-stratified decision to offer chemotherapy[75], prospective trials demonstrating the superiority of such a stratified decision has not emerged. Consequently, current National Comprehensive Cancer Network (NCCN), European Society for Medical Oncology (ESMO), National Institute for Health and Care Excellence (NICE), and Cancer council Australia (CCA) guidelines do not recommend CDX2 be used in prognostication in CRCs [76-80]. Our finding that CDX2 is a companion biomarker for differentiation therapy with PF [a drug that has been deemed as clinically safe in Phase 1 on healthy volunteers (NCT02286882) and can be immediately repurposed], provides a compelling impetus for setting up prospective trials that can rigorously test both claims.

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 and UH3 TR002968 (to P.G), R01-GM138385 (to DS), and 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. Other sources of support include: R01-AI155696 (P.G and D.S), R01-AI141630 (P.G), and UG3TR003355 (D.S. and P.G), a Torey Coast Foundation Award (P.G, D.S, M.B, J.Y), and by three Padres Pedal the Cause awards #PTC2017, #PTC2021 and #PTC2022. D.S and P.G were also supported by the Leona M. and Harry B. Helmsley Charitable Trust. Funders had no role in study design or conclusions.

Author Contributions

SS, DS and PG conceptualized the study; SS and PG designed and analysed all parts of the study; JA, KP, EV and CT acquired, expanded, biobanked and conducted all experiments with patient-derived organoids; VC conducted immunocytochemistry studies; CRE conducted qPCR studies and assisted SS, VC and PG on key parts of the study; ST and DS built and validated the network model; SS and PG conducted all other computational and bioinformatic analyses; SA and MB conducted the xenotransplantation studies; JRS and JY synthesized the pharmacophore; MB, JY, DS 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:

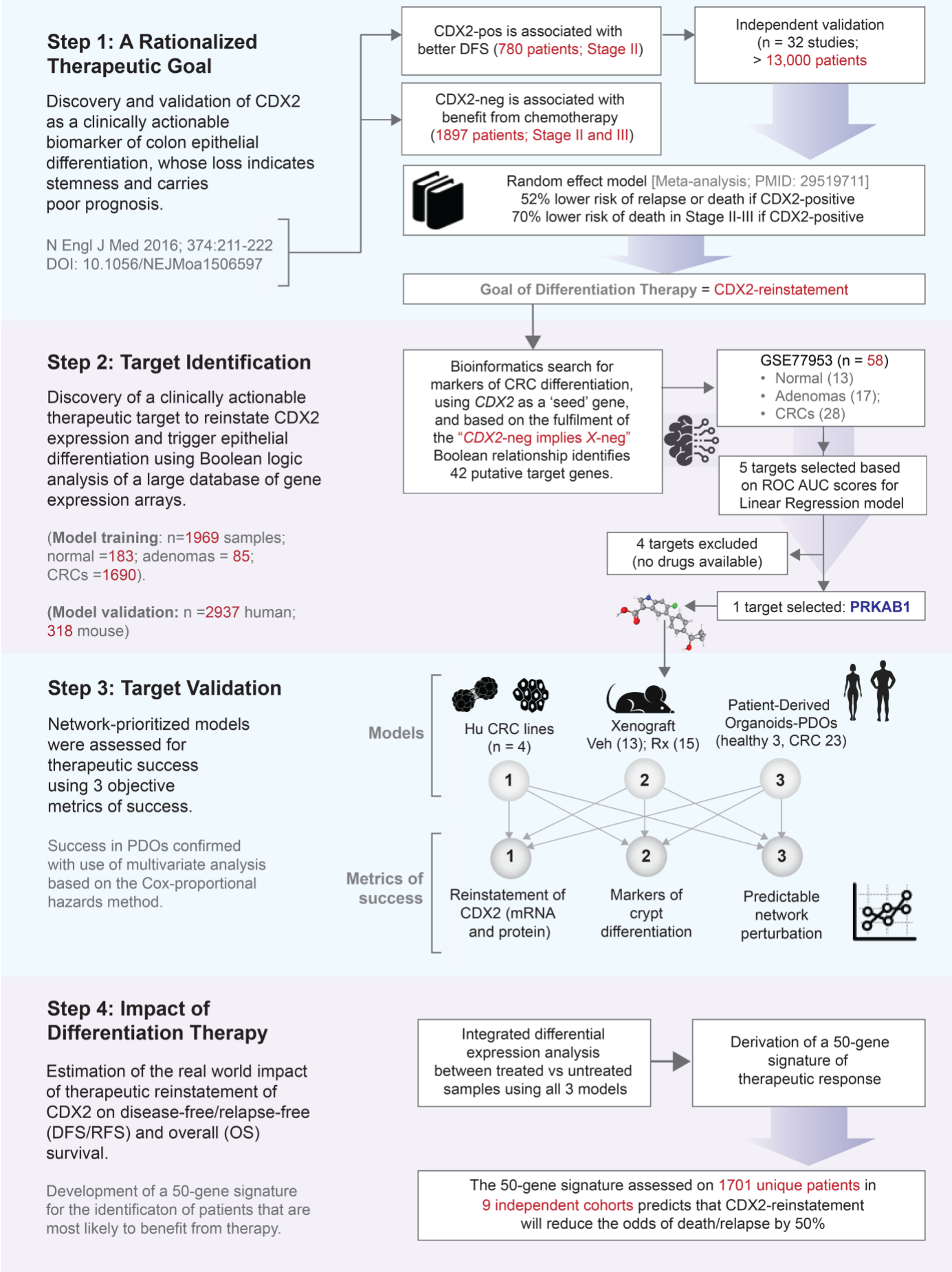

Figure 1. Study Design: Identification and validation of a therapeutic strategy to reinstate CDX2.

Step 1: Flow charts presents the emerging evidence, in chronological order, which rationalized reinstatement of CDX2 as a desirable end goal for differentiating cancer stem cells in the most the aggressive subtypes of CRCs.

Step 2. The key steps towards the network-based identification of PRKAB1 as a potent and actionable target, whose agonism is predicted to reinstate CDX2 expression in those CRCs where its expression is lost. **Supplemental Information 1** for a catalogue of all datasets used for model training and validation and **Supplemental Information 2** for gene clusters, beyond *CDX2*, that are predicted to be co-induced upon PRKAB1 agonism.

Step 3. Three model systems and objective metrics of success that are used to validate PRKAB1 agonism as a strategy to reinstate CDX2 expression and induce differentiation.

Step 4. A 50-gene signature of therapeutic response derived from all 3 models is used to objectively measure the impact of CDX2 reinstatement therapy and identify patients who are most likely to benefit from the same.

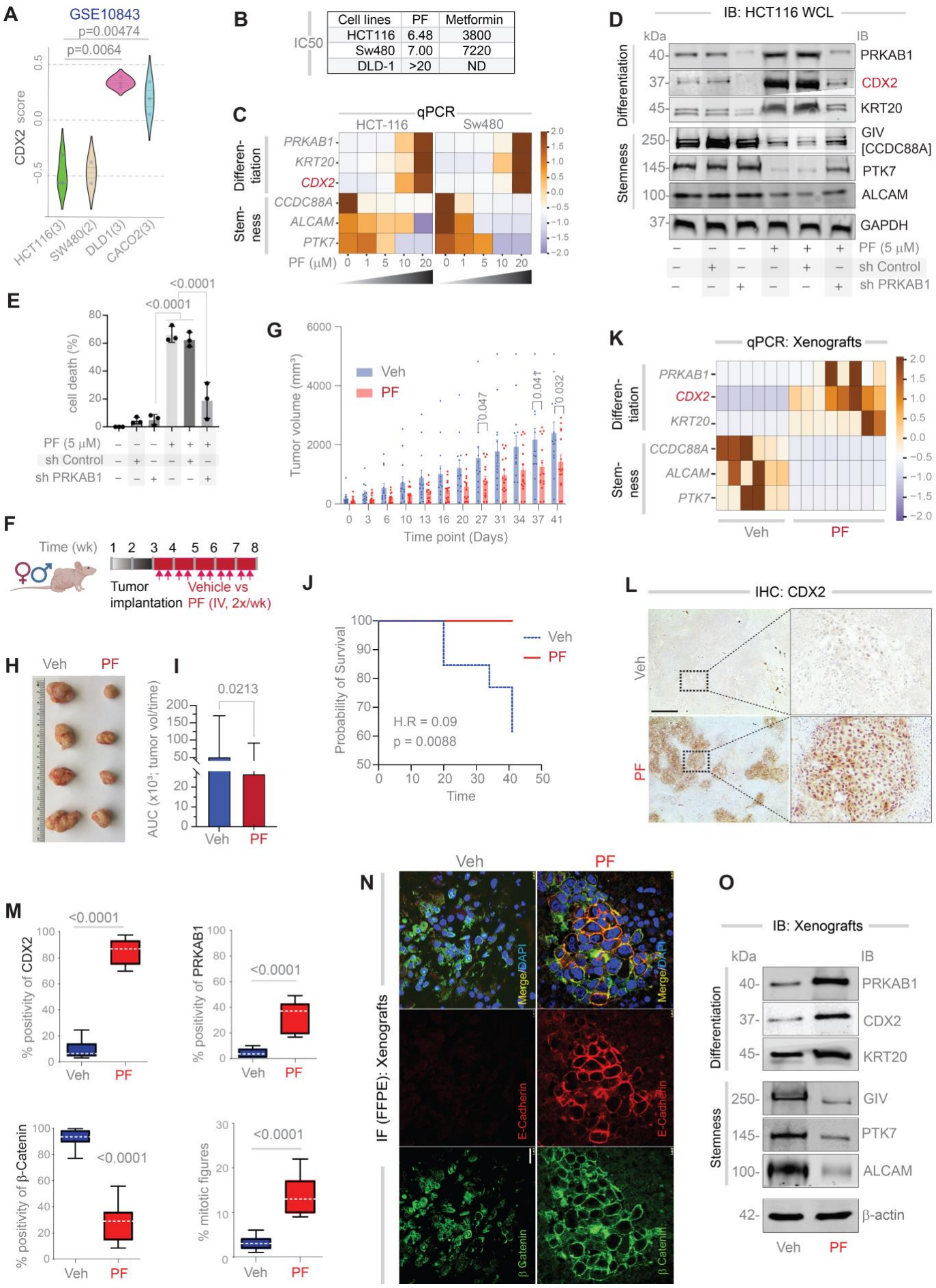

Figure 2. Target validation in CRC cell lines and in xenotransplantation studies in mice. A. Violin plot displays *CDX2* expression (transcripts/million; tpm) in 4 common CRC cell lines. p-values are calculated using Welch's

t-test. **B**. Table shows IC50 values for specific (PF) or non-specific (Metformin) PRKAB1-agonists on CRC cell lines, as determined by MTT assays at 48 h. **C**. RT-qPCR on viable cells (>95%, at 16 h) of differentiation and stemness related markers in the colon is displayed as a heatmap of Z-score normalized values. **D-E**. Immunoblots (D) and MTT based survival assays (E) show the impact of PRKAB1 depletion (in HCT116 cells by shRNA) on the anti-cancer effect of PF. p-values are calculated using one-way ANOVA. Error bars indicate S.E.M of 3 independent biological experiments. See **Figure S4** and **S5** for additional studies on HCT116 and Sw480 cells challenged with increasing doses of PF (**S4** and **S5A-H**) or Metformin (**S5I-K**). **F**. Schematic showing experimental design and workflow. **G**. Graphs display growth kinetics of vehicle (Veh; n = 13) and PF-treated (n = 15) tumours, measured using callipers (2/wk), collated from 3 independent experimental cohorts. p-values are calculated using multiple paired t-test following Benjamini, Krieger, and Yekutieli method. **H**. Images of tumour xenografts at 8-wk post-implantation. **I**. Bar plots compare the area under the curve (AUC) of treatment and control arms [95% Confidence Interval was 31788.29 to 70772.18 (Veh) and 16949.29 to 36404.26 (PF); Error bars represent S.E.M., and p-value was determined by unpaired 2-tailed t-test using GraphPad]. **J**. Kaplan-Meier analysis shows time to the fulfilment of IACUC-mandated criteria for euthanasia (death and/or signs of discomfort). p-value was determined by Log Rank test. H.R., hazard ratio. **K**. RT-qPCR on xenografts displayed as a heatmap of Z-score normalized values. **L**. Representative fields from CDX2-stained FFPE-xenografts are shown. Scale bar = 100 µm. **M**. Quantification of CDX2, PRKAB1, and nuclear β-catenin staining by ImageJ (IHC profiler application) and mitotic bodies (manually counted in H&E-stained sections). p-values are calculated by unpaired 2-tailed t-test. See **Figure S6B-D** for images from representative fields. **N**. Representative fields from FFPE tumors co-stained for E-cadherin (red), β-catenin (green) and DAPI (nuclei; blue) are shown. Scale bar = 20 µm. See also **Figure S6K** for colocalization analysis by ImageJ. **O**. Immunoblots of lysates of viable tumour tissues. See **Figure S6F** for quantification.

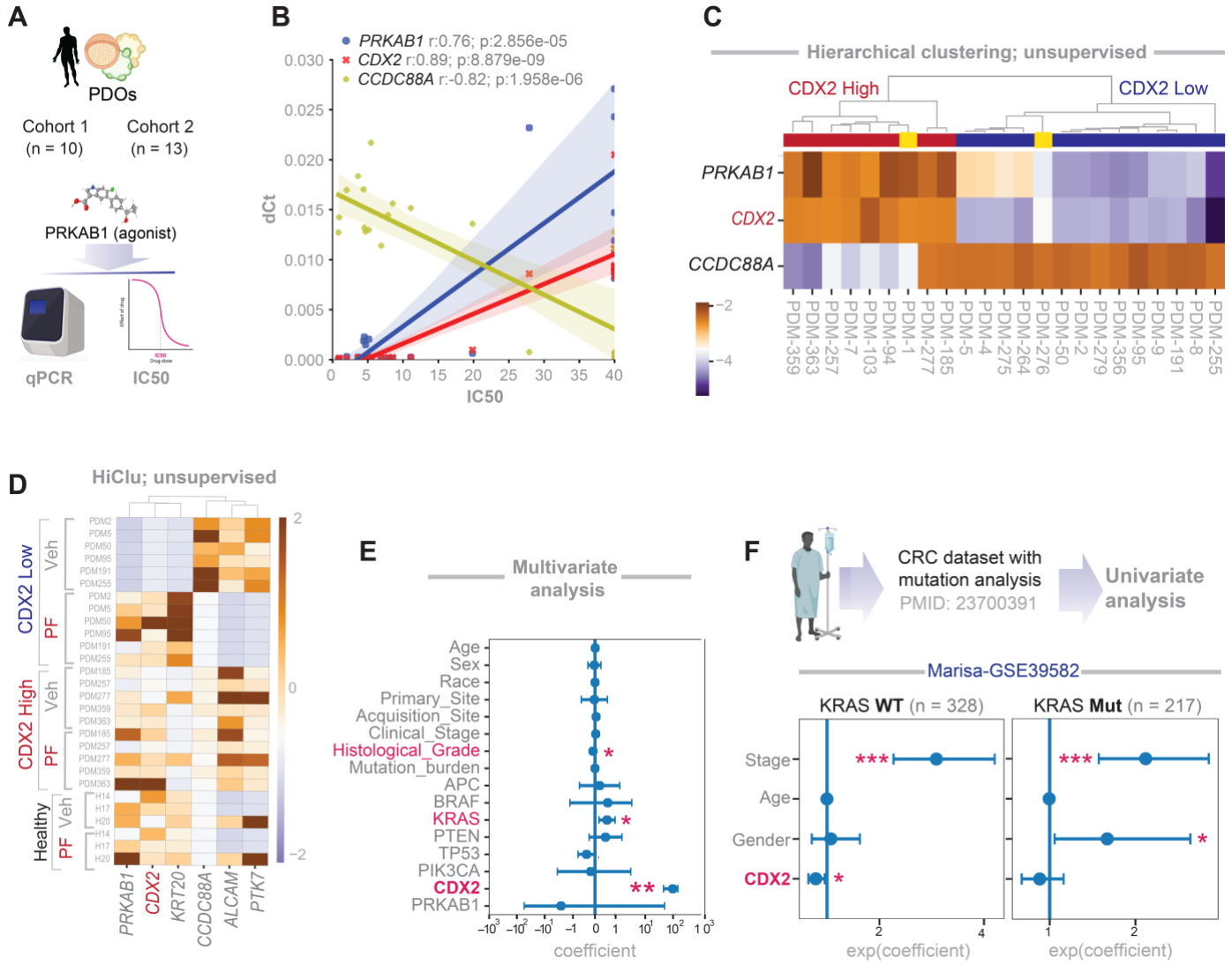

Figure 3. Differentiation and anti-cancer effects of PRKAB1 agonism are restricted to *CDX2*-low PDOs. **A**. Schematic of workflow for sequential analyses in two independent patient cohorts. See **Table 1** for patient demographics, tumour characteristics and mutation status of key genes. **B**. Correlation plots between IC50 and the levels of expression of 3 genes, as determined by qPCR. **C**. Heatmap displays the results of unsupervised clustering of the CRC PDOs using Z-score normalized qPCR estimation of 3 genes. **D**. Heatmap generated as above of the vehicle (Veh) and PF-treated CRC PDOs, using a panel of stemness and differentiation markers. **E**. Multivariate analysis of IC50 with PF as a linear combination of all variables in the PDOs used in this study. The coefficient of each variable (at the centre) with their upper and lower bounds of 95% confidence interval (as error bars) and the p-values from t-tests are illustrated in the bar plot. The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). Red = significant co-variates. *$p \leq 0.05$; **$p \leq 0.01$. See also **Supplemental Information 3** for source data. **F**. Univariate analysis model of relapse-free survival (RFS) in KRAS-WT vs KRAS-mutant tumours in GSE39582. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$.

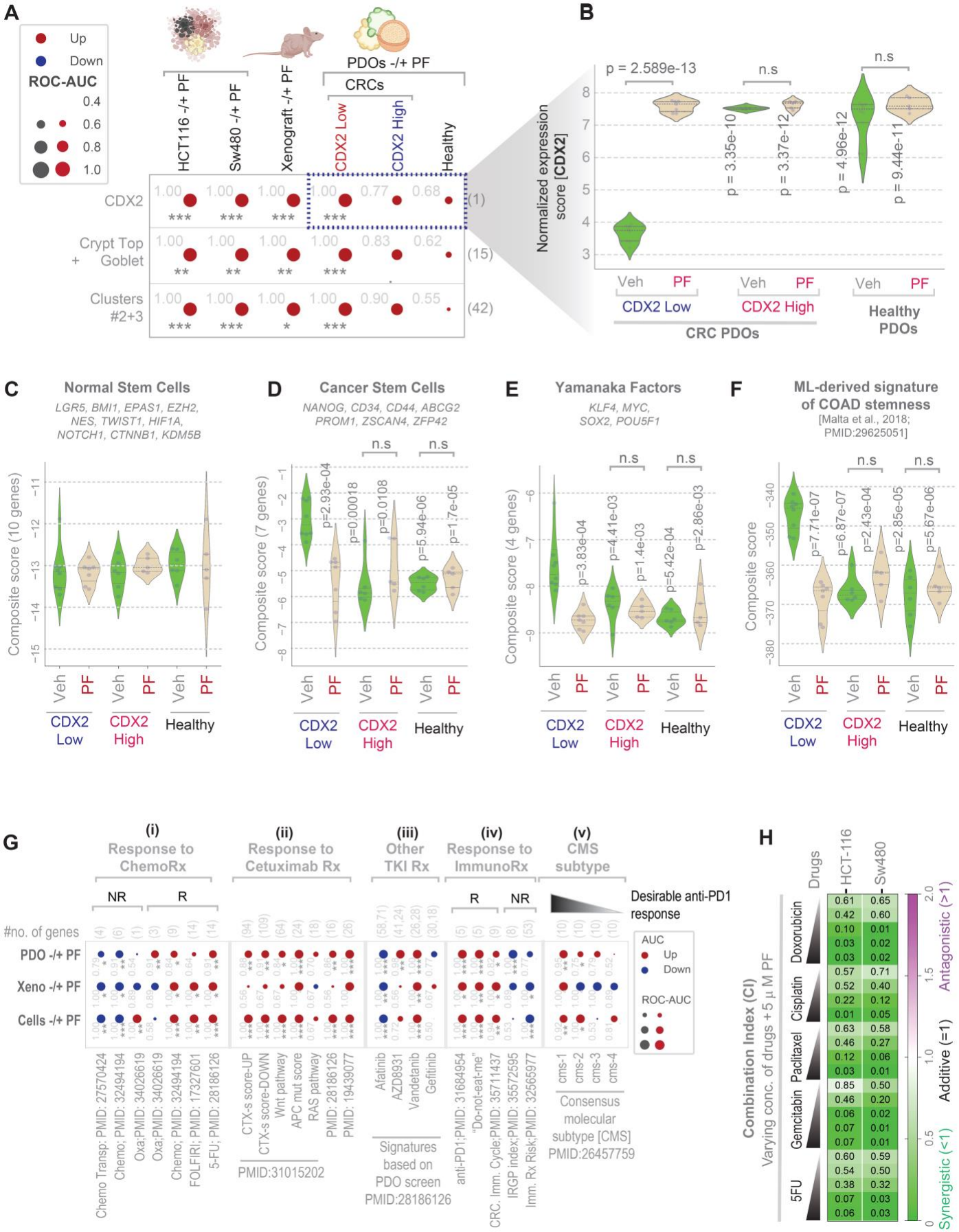

Figure 4. Indicators of success, target-cell specificity, and the potential for synergy with other treatment modalities. A. Expression of *CDX2* (top row), crypt top and Goblet cell markers (middle row) and genes in clusters #2-3 in the network (bottom row) gene/gene signatures in HCT116, Sw480, xenografts and PDOs is

visualized as bubble plots of ROC-AUC values (radius of circles are based on the ROC-AUC) for the classification of vehicle (Veh) vs PF-treated samples. Colors indicate the direction of gene regulation (Up, red; Down, blue) in the treated samples. The significance of AUC is determined based on Welch's t-test and displayed using standard codes (''p>0.1, '.'p<=0.1, *p<=0.05, **p<=0.01, ***p<=0.001) next to the ROC-AUC values. See **Figure S7** for the display of the same data as violin plots with actual p-values. **B**. Violin plots show *CDX2* expression in PDOs. p-values are based on Welch's t-test compared to vehicle (Veh)-treated *CDX2*-low PDOs. **C-F.** Violin plots display the composite score of stemness markers recently reported in a pan-cancer effort that used machine learning. p-values were determined by Welch's t-test, comparing each sample against vehicle (Veh)-treated *CDX2*-low PDOs. **G**. Expression of gene signatures of response (default or indicated with 'R') or non-response (N.R.) to multiple therapeutic modalities in CRC cell lines, xenografts and PDO is visualized as bubble plots of ROC-AUC values (radius of circles are based on the ROC-AUC) for the classification of vehicle (Veh) vs PF-treated samples. Colors indicate the direction of gene regulation (Up, red; Down, blue). p-values displayed are based on significance of AUC using Welch's t-test and using standard code (*p<=0.05, **p<=0.01, ***p<=0.001) next to the ROC-AUC. **H**. IC50 values for combinations of 5 µM PF (constant) with increasing conc. of the indicated chemotherapeutic agents, as determined on HCT116 and sw480 cell lines by MTT assays were used to derive a combination index (CI) using a previously established formula [81]. Results are displayed as a heatmap.

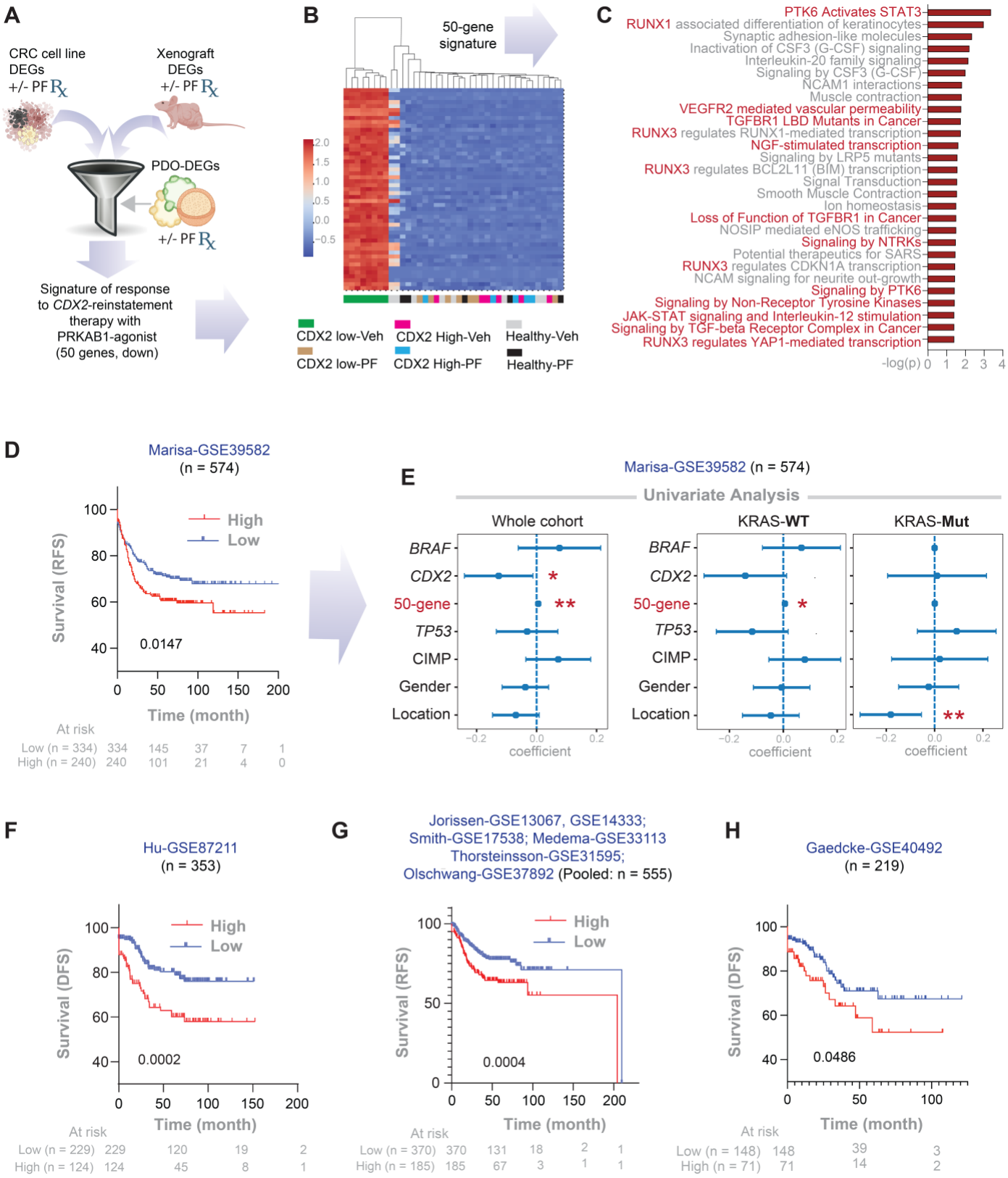

Figure 5. A gene signature for tracking therapeutic response and estimating impact on survival.

A. Computational workflow used in the derivation of a 50-gene signature of therapeutic response using an integrated differential expression analysis (DEA) approach. Differentially expressed genes (DEGs) from CRC cell lines and xenografts (see **Supplemental Information 4-6**) were further refined based on their ability to accurately classify PDO samples with/without PF treatment (see **Supplemental Information 6** for gene list rank-ordered by ROC-AUC score). **B**. Heatmap displays the results of unsupervised clustering of PDOs based on the 50-gene signature. Veh, vehicle. See

Figure S8A-B for visualization of findings as violin plots. **C**. Reactome pathway analysis of the 50-gene signature. **D**, **F-H**. Kaplan-Meier (KM) plots show recurrence-free (RFS) or disease-free (DFS) survival over time on various cohorts of CRC patients, stratified by high vs low expression scores of the 50-gene signature. p-values were determined by Log Rank test. See **Figure S8C** for KM plots on the same cohorts based on *CDX2*-based stratification. **E**. Univariate analysis models of RFS as a linear combination of all variables in a large CRC cohort, including the 50-gene signature. The coefficient of each variable (at the centre) with 95% CI (as error bars) and the p values are illustrated. The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). Left, entire cohort; middle, *KRAS*-WT CRCs; right, *KRAS*-mut CRCs. *$p \leq 0.05$; **$p \leq 0.01$.

ATCC CAT No.	COHORT #	AGE	SEX	RACE	Primary Site	Acquisition Site	Clinical Stage	*CDX2* expression (qPCR)	Histo Grade	APC	BRAF	KRAS	TP53	PTEN	PIK3CA	Tumour Mutation Burden	Neoadj. Therapy	IC50 (PF; μM)
PDM-255	1	73	F	White	Colon	Cecum	IV C	1.88E-05	GB	■			■			218	yes	0.84
PDM-356	2	58	M	Unk	Colon	Sigmoid colon	Unknown	0.000156	G2	■			■			151	no	0.95
PDM-8	2	75	M	Asian	Colon	Sigmoid colon	II A	0.000106	G2	■			■			343	no	2
PDM-191	1	56	M	White	Rectum	Sigmoid Colon	III C	0.000222	GX		■		■			219	yes	3.61
PDM-264	2	67	F	White	Rectosigmoid junction	Rectosigmoid junction	III B	0.000147	G2	■			■			203	no	4.4
PDM-5	2	60	F	White	Colon	Sigmoid colon	III B	0.000209	G2	■			■			131	no	4.5
PDM-275	1	73	F	White	Colon	Lymph node(s)	Unknown	0.000226	Unknown							211	yes	4.67
PDM-4	1	50	M	White	Colon	Sigmoid colon	III B	0.00023	G2	■			■			102	no	5.14
PDM-95	2	61	M	Black	Colon	Sigmoid colon	III B	0.000219	GB	■			■			116	no	5.45
PDM-279	1	51	M	Black	Colon	Cecum	0	0.000229	GX	■	■				■	173	no	6.9
PDM-2	1	68	M	White	Colon	Sigmoid colon	IV A	0.000217	GB	■			■			129	yes	7.86
PDM-50	1	78	M	White	Colon	Ascending colon	III B	0.000216	G3							0	no	8.53
PDM-9	1	63	M	Asian	Colon	Liver	IV A	0.00019	G2	■			■			139	no	11.1
PDM-276	2	54	M	Black	Colon	Transverse colon	Unknown	0.000965	Unknown	■		■				109	no	19.85
PDM-1	1	75	M	White	Colon	Cecum	I	0.00859	G2	■		■	■			191	no	27.9
PDM-185	1	71	M	White	Colon	Cecum	IIA	0.00933	G2	■		■				154	no	>>40
PDM-94	2	67	M	White	Colon	Sigmoid colon	Unknown	0.0111	G2	■			■	■		168	no	>40
PDM-7	2	75	M	Black	Colon	Sigmoid colon	III A	0.00941	G2	■		■	■			156	no	>40
PDM-257	2	53	F	Unk	Rectosigmoid junction	Rectosigmoid junction	III B	0.00832	G2	■		■	■	■		155	no	>40
PDM-363	2	72	M	Unk	Colon	Cecum	Unknown	0.00878	GX	■	■		■	■	■	2701	no	>40
PDM-277	2	76	F	White	Colon	Lymph node(s)	IV A	0.00899	G2	■		■	■			209	no	>40
PDM-359	2	64	F	Unk	Colon	Ascending colon	Unknown	0.00922	G2	■	■				■	1502	no	>40
PDM-103	2	51	F	Black	Colon	Liver	Unknown	0.0205	Unknown	■		■	■			171	yes	>40

Table 1. Demographic and clinicopathologic features of the colorectal cancer PDOs used in this study. Two independent cohorts of PDOs were recruited to this study from the ATCC biobank. *CDX2* expression and IC50 for PF are colour coded using default setting on Microsoft Excel. M, male; F, female; HISTO, histological; NEOADJ, neoadjuvant; IC50, Half-maximal Inhibitory concentration 50.

References

1. Bonhomme, C., *et al.* The Cdx2 homeobox gene has a tumour suppressor function in the distal colon in addition to a homeotic role during gut development. *Gut* **52**, 1465-1471 (2003).
2. Silberg, D.G., Swain, G.P., Suh, E.R. & Traber, P.G. Cdx1 and cdx2 expression during intestinal development. *Gastroenterology* **119**, 961-971 (2000).
3. Chawengsaksophak, K., James, R., Hammond, V., Köntgen, F. & Beck, F. Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* **386**, 84-87 (1997).
4. Hryniuk, A., Grainger, S., Savory, J.G. & Lohnes, D. Cdx1 and Cdx2 function as tumor suppressors. *Journal of Biological Chemistry* **289**, 33343-33354 (2014).
5. Sahoo, D., Dill, D.L., Gentles, A.J., Tibshirani, R. & Plevritis, S.K. Boolean implication networks derived from large scale, whole genome microarray datasets. *Genome biology* **9**, R157 (2008).
6. Dalerba, P., *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* **29**, 1120-1127 (2011).
7. Dalerba, P., *et al.* CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *New England Journal of Medicine* **374**, 211-222 (2016).
8. Levin, T.G., *et al.* Characterization of the intestinal cancer stem cell marker CD166 in the human and mouse gastrointestinal tract. *Gastroenterology* **139**, 2072-2082 e2075 (2010).
9. Weichert, W., Knosel, T., Bellach, J., Dietel, M. & Kristiansen, G. ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. *Journal of clinical pathology* **57**, 1160-1164 (2004).
10. Dalerba, P., *et al.* Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A* **104**, 10158-10163 (2007).
11. Derangère, V., *et al.* 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 (2022).
12. Bennedsen, A.L.B., *et al.* An exploration of immunohistochemistry-based prognostic markers in patients undergoing curative resections for colon cancer. *BMC Cancer* **22**, 62 (2022).

13. Azcue, P., *et al.* A Novel Prognostic Biomarker Panel for Early-Stage Colon Carcinoma. *Cancers (Basel)* **13**(2021).

14. den Uil, S.H., *et al.* Quantitative analysis of CDX2 protein expression improves its clinical utility as a prognostic biomarker in stage II and III colon cancer. *Eur J Cancer* **144**, 91-100 (2021).

15. Tarazona, N., *et al.* Detection of postoperative plasma circulating tumour DNA and lack of CDX2 expression as markers of recurrence in patients with localised colon cancer. *ESMO Open* **5**, e000847 (2020).

16. Hestetun, K.E., *et al.* Mismatch repair phenotype determines the implications of tumor grade and CDX2 expression in stage II-III colon cancer. *Mod Pathol* **34**, 161-170 (2021).

17. Asgari-Karchekani, S., Karimian, M., Mazoochi, T., Taheri, M.A. & Khamehchian, T. CDX2 Protein Expression in Colorectal Cancer and ItsCorrelation with Clinical and Pathological Characteristics, Prognosis, and Survival Rate of Patients. *J Gastrointest Cancer* **51**, 844-849 (2020).

18. Ma, C., *et al.* SATB2 and CDX2 are prognostic biomarkers in DNA mismatch repair protein deficient colon cancer. *Mod Pathol* **32**, 1217-1231 (2019).

19. Ben-Aharon, I., *et al.* Sidedness Matters: Surrogate Biomarkers Prognosticate Colorectal Cancer upon Anatomic Location. *Oncologist* **24**, e696-e701 (2019).

20. Hansen, T.F., *et al.* Prognostic impact of CDX2 in stage II colon cancer: results from two nationwide cohorts. *British Journal of Cancer* **119**, 1367-1373 (2018).

21. Pilati, C., *et al.* CDX2 prognostic value in stage II/III resected colon cancer is related to CMS classification. *Annals of Oncology* **28**, 1032-1035 (2017).

22. Nolte, S., *et al.* Construction and analysis of tissue microarrays in the era of digital pathology: a pilot study targeting CDX1 and CDX2 in a colon cancer cohort of 612 patients. *J Pathol Clin Res* **3**, 58-70 (2017).

23. Zhang, B.Y., *et al.* Lack of caudal-type homeobox transcription factor 2 expression as a prognostic biomarker in metastatic colorectal cancer. *Clinical colorectal cancer* **16**, 124-128 (2017).

24. Tomasello, G., *et al.* Association of CDX2 expression with survival in early colorectal cancer: a systematic review and meta-analysis. *Clinical colorectal cancer* **17**, 97-103 (2018).

25. Sjoerd, H., *et al.* Quantitative analysis of CDX2 protein expression improves its clinical utility as a prognostic biomarker in stage II and III colon cancer. *European Journal of Cancer* **144**, 91-100 (2021).

26. Bruun, J., *et al.* Prognostic, predictive, and pharmacogenomic assessments of CDX 2 refine stratification of colorectal cancer. *Molecular oncology* **12**, 1639-1655 (2018).

27. Baba, Y., *et al.* Relationship of CDX2 loss with molecular features and prognosis in colorectal cancer. *Clinical cancer research* **15**, 4665-4673 (2009).

28. Olsen, J., *et al.* CDX2 downregulation is associated with poor differentiation and MMR deficiency in colon cancer. *Experimental and molecular pathology* **100**, 59-66 (2016).

29. Sahoo, D., Dill, D.L., Tibshirani, R. & Plevritis, S.K. Extracting binary signals from microarray time-course data. *Nucleic Acids Res* **35**, 3705-3712 (2007).

30. Sahoo, D., *et al.* MiDReG: a method of mining developmentally regulated genes using Boolean implications. *Proc Natl Acad Sci U S A* **107**, 5732-5737 (2010).

31. Sahoo, D., *et al.* Artificial intelligence guided discovery of a barrier-protective therapy in inflammatory bowel disease. *Nat Commun* **12**, 4246 (2021).

32. Ghosh, P. The stress polarity pathway: AMPK 'GIV'-es protection against metabolic insults. *Aging (Albany NY)* **9**, 303-314 (2017).

33. Ghosh, P., *et al.* The stress polarity signaling (SPS) pathway serves as a marker and a target in the leaky gut barrier: implications in aging and cancer. *Life Sci Alliance* **3**(2020).

34. Aznar, N., *et al.* AMP-activated protein kinase fortifies epithelial tight junctions during energetic stress via its effector GIV/Girdin. *Elife* **5**(2016).

35. Gough, N.R. Maintaining Polarity in the Face of Energetic Stress. *Science's STKE* **2007**, tw166-tw166 (2007).

36. Graule, J., *et al.* CDX2 in colorectal cancer is an independent prognostic factor and regulated by promoter methylation and histone deacetylation in tumors of the serrated pathway. *Clinical epigenetics* **10**, 1-12 (2018).

37. Boman, B.M. & Huang, E. Human colon cancer stem cells: a new paradigm in gastrointestinal oncology. *J Clin Oncol* **26**, 2828-2838 (2008).

38. Kalisky, T., *et al.* Analysis of human colon tissue cell composition using single-cell gene-expression PCR. *Journal of biomolecular techniques: JBT* **24**, S11 (2013).

39. Malta, T.M., *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354 e315 (2018).

40. Friedmann-Morvinski, D. & Verma, I.M. Dedifferentiation and reprogramming: origins of cancer stem cells. *EMBO Rep* **15**, 244-253 (2014).

41. Guinney, J., *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356 (2015).

42. Valladares-Ayerbes, M., *et al.* Evaluation of the adenocarcinoma-associated gene AGR2 and the intestinal stem cell marker LGR5 as biomarkers in colorectal cancer. *Int J Mol Sci* **13**, 4367-4387 (2012).

43. Li, Y., *et al.* AT7867 Inhibits the Growth of Colorectal Cancer Stem-Like Cells and Stemness by Regulating the Stem Cell Maintenance Factor Ascl2 and Akt Signaling. *Stem Cells Int* **2023**, 4199052 (2023).

44. Jubb, A.M., *et al.* Achaete-scute like 2 (ascl2) is a target of Wnt signalling and is upregulated in intestinal neoplasia. *Oncogene* **25**, 3445-3457 (2006).

45. Durinikova, E., *et al.* ALDH1A3 upregulation and spontaneous metastasis formation is associated with acquired chemoresistance in colorectal cancer cells. *BMC Cancer* **18**, 848 (2018).

46. Chen, G., *et al.* Targeting TM4SF1 exhibits therapeutic potential via inhibition of cancer stem cells. *Signal Transduct Target Ther* **7**, 350 (2022).

47. Tang, Q., *et al.* TM4SF1 promotes EMT and cancer stemness via the Wnt/β-catenin/SOX2 pathway in colorectal cancer. *J Exp Clin Cancer Res* **39**, 232 (2020).

48. Matsui, W.H. Cancer stem cell signaling pathways. *Medicine (Baltimore)* **95**, S8-S19 (2016).

49. Yang, L., *et al.* Targeting cancer stem cell pathways for cancer therapy. *Signal Transduct Target Ther* **5**, 8 (2020).

50. Gierut, J., *et al.* Disruption of the mouse protein tyrosine kinase 6 gene prevents STAT3 activation and confers resistance to azoxymethane. *Gastroenterology* **141**, 1371-1380, 1380.e1371-1372 (2011).

51. Liu, C., *et al.* Pharmacological targeting PTK6 inhibits the JAK2/STAT3 sustained stemness and reverses chemoresistance of colorectal cancer. *J Exp Clin Cancer Res* **40**, 297 (2021).

52. Li, X., *et al.* Colorectal carcinomas with mucinous differentiation are associated with high frequent mutation of KRAS or BRAF mutations, irrespective of quantity of mucinous component. *BMC Cancer* **20**, 400 (2020).

53. Feng, Y., *et al.* Mutant KRAS promotes hyperplasia and alters differentiation in the colon epithelium but does not expand the presumptive stem cell pool. *Gastroenterology* **141**, 1003-1013 e1001-1010 (2011).

54. Aasebø, K., *et al.* CDX2: A Prognostic Marker in Metastatic Colorectal Cancer Defining a Better. *Front Oncol* **10**, 8 (2020).

55. Brandt, R., *et al.* Cell type-dependent differential activation of ERK by oncogenic KRAS in colon cancer and intestinal epithelium. *Nat Commun* **10**, 2919 (2019).

56. Aasebø, K., *et al.* CDX2: a prognostic marker in metastatic colorectal cancer defining a better BRAF mutated and a worse KRAS mutated subgroup. *Frontiers in oncology* **10**, 8 (2020).

57. Slik, K., *et al.* CDX2 loss with microsatellite stable phenotype predicts poor clinical outcome in stage II colorectal carcinoma. *The American journal of surgical pathology* **43**, 1473-1482 (2019).

58. Moertel, C.G., *et al.* Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine* **322**, 352-358 (1990).

59. Sobrero, A.F., *et al.* A new prognostic and predictive tool for shared decision making in stage III colon cancer. *European Journal of Cancer* **138**, 182-188 (2020).

60. Taieb, J., André, T. & Auclin, E. Refining adjuvant therapy for non-metastatic colon cancer, new standards and perspectives. *Cancer Treatment Reviews* **75**, 1-11 (2019).

61. Delhorme, J.-B., *et al.* CDX2 controls genes involved in the metabolism of 5-fluorouracil and is associated with reduced efficacy of chemotherapy in colorectal cancer. *Biomedicine & Pharmacotherapy* **147**, 112630 (2022).

62. Takakura, Y., *et al.* CDX2 regulates multidrug resistance 1 gene expression in malignant intestinal epithelium. *Cancer research* **70**, 6767-6778 (2010).

63. Shigematsu, Y., *et al.* CDX2 expression is concordant between primary colorectal cancer lesions and corresponding liver metastases independent of chemotherapy: a single-center retrospective study in Japan. *Oncotarget* **9**, 17056 (2018).

64. Fontana, E., *et al.* Early-Onset Colorectal Adenocarcinoma in the IDEA Database: Treatment Adherence, Toxicities, and Outcomes With 3 and 6 Months of Adjuvant Fluoropyrimidine and Oxaliplatin. *Journal of Clinical Oncology* **39**, 4009-4019 (2021).

65. Tong, K., *et al.* Degree of tissue differentiation dictates susceptibility to BRAF-driven colorectal cancer. *Cell reports* **21**, 3833-3845 (2017).

66. Sakamoto, N., *et al.* BRAFV600E cooperates with CDX2 inactivation to promote serrated colorectal tumorigenesis. *Elife* **6**, e20331 (2017).

67. Felipe De Sousa, E.M., *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nature medicine* **19**, 614-618 (2013).

68. Coskun, M. The role of CDX2 in inflammatory bowel disease. *Danish medical journal* **61**, B4820-B4820 (2014).

69. Coskun, M., *et al.* TNF-α-induced down-regulation of CDX2 suppresses MEP1A expression in colitis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1822**, 843-851 (2012).

70. Yin, F., *et al.* Double negativity for expression of YAP1 and CDX2 defines an aggressive type of colitis-associated cancer. *Anticancer research* **40**, 5411-5416 (2020).

71. Kawai, H., *et al.* Promoter methylation downregulates CDX2 expression in colorectal carcinomas. *Oncology reports* **13**, 547-551 (2005).

72. Sun, X., Yang, Q., Rogers, C.J., Du, M. & Zhu, M.-J. AMPK improves gut epithelial differentiation and barrier function via regulating Cdx2 expression. *Cell Death & Differentiation* **24**, 819-831 (2017).

73. Bae, J.M., Lee, T.H., Cho, N.-Y., Kim, T.-Y. & Kang, G.H. Loss of CDX2 expression is associated with poor prognosis in colorectal cancer patients. *World journal of gastroenterology: WJG* **21**, 1457 (2015).

74. Reggiani Bonetti, L., Lionti, S., Vitarelli, E. & Barresi, V. Prognostic significance of CDX2 immunoexpression in poorly differentiated clusters of colorectal carcinoma. *Virchows Archiv* **471**, 731-741 (2017).

75. Alarid-Escudero, F. CDX2 testing+ FOLFOX for CDX2-negative colon cancer patients cost-effective strategy. *PharmacoEconomics & Outcomes News* **892**, 9-27 (2021).

76. Argilés, G., *et al.* Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **31**, 1291-1305 (2020).

77. Benson, A.B., *et al.* Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network* **19**, 329-359 (2021).

78. Committee, A.C.N.C.C.G.R. Guidelines for the prevention, early detection and management of colorectal cancer. *Sydney: The Cancer Council Australia and Australian Cancer Network* (2005).

79. Van Cutsem, E., *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Annals of Oncology* **27**, 1386-1422 (2016).

80. Baum, M., Fallowfield, L., Farewell, V., Macbeth, F. & Treasure, T. NICEGuidelines: management of colorectal cancer metastases. *British Journal of Surgery* **107**, e357-e357 (2020).

81. Foucquier, J. & Guedj, M. Analysis of drug combinations: current methodological landscape. *Pharmacol Res Perspect* **3**, e00149 (2015).