# Pan-cancer genomic amplifications underlie a WNT hyperactivation phenotype associated with stem cell-like features leading to poor prognosis

Running title: Prognosis of a 16-gene signature of WNT hyperactivation in six cancers

Wai Hoong Chang and Alvina G. Lai

Nuffield Department of Medicine, University of Oxford,

Old Road Campus, OX3 7FZ, United Kingdom

For correspondence: Alvina.Lai@ndm.ox.ac.uk

## List of Abbreviations

| | |
|---|---|
| TCGA | The Cancer Genome Atlas |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| GO | Gene Ontology |
| ROC | Receiver operating characteristic |
| AUC | Area under the curve |
| HR | Hazard ratio |
| TNM | Tumor, node and metastasis |
| HIF | Hypoxia inducible factor |
| TF | Transcription factor |
| EMT | Epithelial-to-mesenchymal transition |

## Abstract

Cancer stem cells pose significant obstacles to curative treatment contributing to tumor relapse and poor prognosis. They share many signaling pathways with normal stem cells that control cell proliferation, self-renewal and cell fate determination. One of these pathways known as Wnt is frequently implicated in carcinogenesis where Wnt hyperactivation is seen in cancer stem cells. Yet, the role of conserved genomic alterations in Wnt genes driving tumor progression across multiple cancer types remains to be elucidated. In an integrated pan-cancer study involving 21 cancers and 18,484 patients, we identified a core Wnt signature of 16 genes that showed high frequency of somatic amplifications linked to increased transcript expression. The signature successfully predicted overall survival rates in six cancer cohorts (n=3,050): bladder (P=0.011), colon (P=0.013), head and neck (P=0.026), pan-kidney (P<0.0001), clear cell renal cell (P<0.0001) and stomach (P=0.032). Receiver operating characteristic analyses revealed that the performance of the 16-Wnt-gene signature was superior to tumor staging benchmarks in all six cohorts and multivariate Cox regression analyses confirmed that the signature was an independent predictor of overall survival. In bladder and renal cancer, high risk patients as predicted by the Wnt signature had more hypoxic tumors and a combined model uniting tumor hypoxia and Wnt hyperactivation resulted in further increased death risks. Patients with hyperactive Wnt signaling had molecular features associated with stemness and epithelial-to-mesenchymal transition. Our study confirmed that genomic amplification underpinning pan-cancer Wnt hyperactivation and transcriptional changes associated with molecular footprints of cancer stem cells lead to increased death risks.

41 Keywords: Wnt signaling; cancer stem cells; cell adhesion; pan-cancer; genomic amplification;

42 tumor microenvironment

## Introduction

There is a requirement for tumor cells to self-renew and proliferate in order to perpetuate tumorigenesis. It is perhaps not surprising that tumor-initiating cells or cancer stem cells share similar signal transduction processes with normal stem cells[1,2]. The ability for self-renewal and differentiation in both stem cells and cancer stem cells have converged on a common pathway known as Wnt signaling[3,4]. Wnt proteins are highly conserved across the animal kingdom, functioning as developmentally important molecules controlling cell fate specification, cell polarity and homeostatic self-renewal processes in embryonic and adult stem cells[5]. Wnts are a group of glycoproteins serving as ligands for the frizzled receptor to initiate signaling cascades in both canonical and non-canonical pathways[6]. Beyond embryogenesis, Wnt proteins control cell fate determination in adults where they regulate homeostatic self-renewal of intestinal crypts and growth plates[7–9].

Wnt signaling is the product of an evolutionary adaptation to growth control in multicellular organisms, and it has now become clear that aberrations in this pathway contributes to deranged cell growth associated with many disease pathologies including cancer[10]. Loss-of-function mutations in genes that inhibit the Wnt pathway lead to ligand-independent constitutive activation of Wnt signaling in hepatocellular carcinoma[11], colorectal cancer[12], gastric cancer[13] and acute myeloid leukemia[14]. Thus, inhibition of Wnt signaling would hold great promise as therapeutic targets[15]. A small molecule inhibitor ICG-001 functions to inhibit the degradation of the Wnt repressor Axin and treatment of colon cancer cell lines with this inhibitor resulted in increased apoptosis[16]. Antibodies against Wnts and frizzled receptors have also demonstrated antitumor effects[17,18].

Much of the previous research on Wnt genes and cancer have focused on somatic mutations and transcriptional dysregulation of Wnt pathway members. Activating mutations of β-catenin have been implicated in adrenocortical tumorigenesis[19] and multiple gastrointestinal cancers[20]. Downregulation of a Wnt antagonist *DKK1*, a downstream target of β-catenin, is also observed in colorectal cancer[21]. However, there is limited understanding on the role of somatic copy number alterations in Wnt pathway genes as well as their downstream targets on driving tumor progression and patient prognosis. Studies examining the transcriptional dysregulation of Wnt pathway genes offered limited insights into whether differences in transcript abundance were caused by genomic amplifications or losses.

Given the complexity of Wnt signaling in cancer, it is important to investigate genomic alterations alongside transcriptional regulation of *all* genes associated with Wnt signaling in a comparative approach. We hypothesize that pan-cancer transcriptional aberrations in Wnt signaling is caused by genomic amplifications of a group of genes known as Wnt drivers and that transcriptional profiles of driver genes are important predictors of patient outcome. We conducted a pan-cancer analysis on 147 Wnt signaling genes, which involved positive and negative regulators of the pathway alongside their downstream targets. We analyzed 18,484 matched genomic and transcriptomic profiles representing 21 cancer types to determine whether *1)* somatic copy number amplifications are drivers of hyperactive Wnt signaling, *2)* Wnt driver genes harbor clinically relevant prognostic information and *3)* crosstalk exists between Wnt driver genes, tumor hypoxia and signaling pathways associated with stem cell function. We demonstrate that overexpression of Wnt driver genes resulted in significantly poorer survival outcomes in six cancer types involving 3,050 patients. Hyperactivation of Wnt

91 signaling is linked to loss of cell adhesion and molecular features of stemness. Overall, our
92 findings would facilitate the development of improved therapies through the inhibition of Wnt
93 driver genes in a stratified manner.

## Materials and Methods

A total of 147 genes associated with active and inactive Wnt signaling were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database listed in Table S1.

### Study cohorts

Genomic and transcriptomic profiles of 21 cancers were generated by The Cancer Genome Atlas (TCGA) initiative[22] (n=18,484) (Table S2). For transcriptomic profiles, we retrieved Illumina HiSeq rnaseqv2 Level 3 RSEM normalized data from the Broad Institute GDAC Firehose website. For somatic copy number alterations analyses, we retrieved GISTIC datasets[23] using the RTCGAToolbox package to access Firehose Level 4 copy number variation data. Level 4 clinical data were retrieved using RTCGAToolbox for survival analyses.

### Somatic copy number alterations analyses

GISTIC gene-level table provided discrete amplification and deletion indicators for all tumor samples. Amplified genes were denoted as positive numbers: '1' represents amplification above the threshold or low-level gain (1 extra copy) while '2' represents high-level amplification (2 or more extra copies). Deletions were denoted as negative values: '-1' represents heterozygous deletion while '-2' represents homozygous deletion.

### Determining the 16-gene scores and hypoxia scores

16-Wnt-gene scores for each patient were determined from the mean $\log_2$ expression values of 16 genes: *WNT2, WNT3, WNT3A, WNT10B, FZD2, FZD6, FZD10, DVL3, WISP1, TBL1XR1, RUVBL1, MYC, CCND1, CAMK2B, RAC3* and *PRKCG*. Hypoxia scores were computed from the

mean $\log_2$ expression values of 52 hypoxia signature genes[24]. The hypoxia signature genes were: *ESRP1, CORO1C, SLC2A1, UTP11, CDKN3, TUBA1B, ENO1, NDRG1, PGAM1, CHCHD2, SLC25A32, SHCBP1, KIF20A, PGK1, BNIP3, ANLN, ACOT7, TUBB6, MAP7D1, YKT6, PSRC1, GPI, PGAM4, GAPDH, MRPL13, SEC61G, VEGFA, MIF, TPI1, MAD2L2, HK2, AK4, CA9, SLC16A1, KIF4A, PSMA7, LDHA, MRPS17, PNP, TUBA1C, HILPDA, LRRC42, TUBA1A, MRGBP, MRPL15, CTSV, ADM, DDIT4, PFKP, P4HA1, MCTS1* and *ANKRD37.* For analyses in Figures 5 and 7, patients were separated into four groups using median 16-gene scores and median hypoxia scores or median *EZH2* expression values as thresholds. Nonparametric Spearman's rank-order correlation tests were employed to investigate the relationship between 16-gene scores and hypoxia scores or *EZH2* expression values.

### Differential expression analyses

To compare Wnt gene expression between tumor and non-tumor samples, gene expression profiles for both sample types were separated into two files based on TCGA barcode information. RSEM expression values were converted to $\log_2(x + 1)$ scale. To compare changes in gene expression between high- and low-score groups, patients were median dichotomized based on their 16-gene scores in each cancer type. Differential expression analyses were performed using the R limma package employing the linear model and Bayes method. P value adjustments were conducted using the Benjamini-Hochberg false discovery rate method.

### Biological enrichment and transcription factor analyses

To ascertain which biological pathways and signaling processes were significantly enriched as a result of Wnt hyperactivation, differentially expressed genes obtained from comparing high- and low-score patients were mapped against the KEGG and Gene ontology (GO) databases

using GeneCodis[25]. Differentially expressed genes were also mapped against the Reactome database[26]. The Enrichr tool was used to determine whether differentially expressed genes were enriched with binding targets of stem cell-associated transcription factors[27,28]. Genes were mapped against the ChEA and ENCODE databases using Enrichr.

## Survival analysis

The R survminer and survival packages were used for Kaplan-Meier and Cox proportional hazards regression analyses to determine if the expression levels of the 16 signature genes were significantly associated with overall survival. The ability of the 16-gene signature to predict overall survival when used in combination with hypoxia scores or *EZH2* expression levels was also examined. Univariate Cox regression analyses were performed on each of the individual 16 genes in 20 cancer types (where survival information is available) to determine the contribution of each gene in predicting overall survival. Univariate analyses were also performed on the gene set as a signature (by taking the mean expression scores of the 16 genes) to determine its ability in predicting overall survival. Multivariate Cox regression analyses were employed to demonstrate the independence of the signature to tumor staging parameters. Hazard ratios (HR) and confidence intervals were determined from Cox models where HR greater than one ($P<0.05$) indicated that a covariate was positively associated with even probability (increased hazard) and negatively linked to survival length. The non-significant relationship between scaled Schoenfeld residuals and time supported the proportional hazards assumption; this was tested using the R survival package. Kaplan-Meier analyses were employed to confirm results obtained from Cox regression. Patients were first median-separated into low- and high-score groups based on the expression of the 16 genes (detailed above) for Kaplan-Meier analyses. Statistical difference between high- and low-score patient

groups was evaluated using the log-rank test. Receiver operating characteristic analyses were performed using the R survcomp package to assess the predictive performance (sensitivity and specificity) of the signature in relation to tumor stage. Area under the ROC curves (AUCs) were calculated using survcomp. AUC values can fall between 1 (perfect marker) and 0.5 (uninformative marker).

All plots were generated using ggplot2 and pheatmap packages implemented in R[29]. The InteractiVenn tool[30] was employed to generate the Venn diagram in Figure S2.

## Results

### Pan-cancer genomic alterations of Wnt signaling lead to dysregulated transcriptional response in tumors

A list of 147 genes involved in the Wnt signal transduction pathway was retrieved from the KEGG database (Table S1). They include genes in both canonical and non-canonical Wnt pathways along with their downstream targets. A literature search was conducted to manually curate these genes into two categories: 1) genes associated with active Wnt signaling (90 genes) and 2) genes associated with repressed Wnt signaling (50 genes) (Fig. 1A). To systematically evaluate the extent of Wnt dysregulation across cancers, we analyzed genomic and transcriptomic datasets from 18,484 patients representing 21 cancer types[22]. To determine whether genomic alterations were present in the 147 genes, we evaluated the frequency of somatic copy number alterations across all 21 cancers.

Focusing on genomic amplifications that occurred in at least 20% of samples in each cancer type and amplification events that were present in at least one-third of cancer types (> 8 cancers), we observed that 61 genes were recurrently amplified (Fig. 1B). Of these 61 genes, 41 genes were associated with active Wnt signaling while 20 genes were linked to repressed Wnt signaling (Fig. 1A). Some of the most amplified genes found in at least 95% of cancer types included genes from both canonical (*FZD1, FZD9, WNT16, WNT2, SFRP4, CSNK2A1* and *RAC1*) and non-canonical Wnt pathways (*PLCB1, PLCB4, CAMK2B* and *NFATC2*) (Fig. 1B).

When comparing the frequency of Wnt gene amplifications between cancers, interesting associations were observed. Cancers that affect organ systems working together to perform a common function, i.e. gastrointestinal tract, exhibited similar patterns of genomic amplifications where most of the 61 genes were amplified in at least 20% of tumors. Hierarchical clustering on amplification frequencies using Euclidean distance metric revealed that gastrointestinal cancers of the colon (COAD), stomach (STAD), bile duct (CHOL) and liver (LIHC) were clustered together, implying that there was a significant degree of conservation in genetic aberration of Wnt signaling in these cancers (Fig. 1B). In contrast, cancers of the brain and central nervous system (GBMLGG and GBM) had the least number of amplified genes; 11 and 12 genes respectively (Fig. 1B).

We reason that somatic amplification events that were linked with transcriptional overexpression could represent candidate Wnt drivers, given that positive correlation between RNA and DNA levels would imply a gain of function. We performed differential expression analyses on the 90 genes involved in active Wnt signaling (Table S1) using tumor and non-tumor samples from each cancer type (Table S2). We observed that 28 genes were overexpressed (fold change > 1.5) in at least 8 or more cancers. Of the 28 genes, we identified 16 genes that were also recurrently amplified (Fig. 1A, B). These 16 genes were prioritized as core Wnt driver candidates representative of multiple tumors: *WNT2, WNT3, WNT3A, WNT10B, FZD2, FZD6, FZD10, DVL3, WISP1, TBL1XR1, RUVBL1, MYC, CCND1, CAMK2B, RAC3* and *PRKCG* (Fig. 1B).

## Pan-cancer prognostic relevance of the newly identified core Wnt drivers

We rationalize that the gain of function of the core Wnt drivers could influence patient outcome. Univariate Cox proportional hazards regression analyses were performed on the transcriptional profiles of each of the 16 Wnt drivers on 20 cancers where survival information is available. A vast majority of the core Wnt driver genes were significantly associated with poor prognosis (hazard ratio [HR] above 1, $P<0.05$) (Fig. S1). Interestingly, there were variations in the number of prognostic genes between cancers. Esophageal cancer (ESCA) had no prognostic genes and only two genes were prognostic in sarcoma (SARC) and cholangiocarcinoma (CHOL). In contrast, clear cell renal cell carcinoma (KIRC) and the pan-kidney cohort (KIPAN) involving chromophobe renal cell, papillary renal cell and clear cell renal cell carcinoma had 13 and 10 prognostic genes respectively (Fig. S1). To determine whether core Wnt driver genes harbored prognostic information as a gene set, we calculated expression scores for each patient in each cancer type by taking the mean expression of the 16 Wnt drivers. Patients were subsequently median-dichotomized into low- and high-score groups for survival analyses. Remarkably, when the core Wnt drivers were considered as a gene signature, we observed that patients with high scores had significantly poorer survival rates in six cancer cohorts (n=3,050): bladder ($P=0.011$), colon ($P=0.013$), head and neck ($P=0.026$), pan-kidney ($P<0.0001$), clear cell renal cell ($P<0.0001$) and stomach ($P=0.032$) (Fig. 2).

To determine whether the 16-Wnt-gene signature harbored independent prognostic value over current tumor, node and metastasis (TNM) staging system, the signature was evaluated on patients grouped according to tumor stage; early (stages 1 and/or 2), intermediate (stages 2 and/or 3) and late (stages 3 and/or 4). Patients were first separated by tumor stage followed

by median-stratification based on their 16-gene scores into low- and high-score groups within each stage category. Regardless of tumor stage, the signature retained its predictive value where high-score patients consistently had higher risk of death: bladder (P<0.0001), colon (P<0.0001), head and neck (P=0.027), pan-kidney (P<0.0001), clear cell renal cell (P<0.0001) and stomach (P=0.034) (Fig. 3A). Moreover, we observed that the expression of Wnt driver genes increased with tumor stage (Fig. 3B). Taken together, this suggests that another level of patient stratification beyond that of TNM staging is afforded by the 16-gene signature, especially for patients with early stage cancer where tumors are more heterogeneous.

Multivariate Cox regression analyses were performed to further confirm that the 16-Wnt-gene signature was independent of TNM staging. Indeed, in all six cancer types, the signature remained prognostic when controlling for TNM stage (Table S3). High-score patients had significantly higher risk of death even when TNM stage was taken into account: bladder (HR=1.409, P=0.015), colon (HR=1.561, P=0.018), head and neck (HR=1.378, P=0.036), pan-kidney (HR=1.738, P<0.0001), clear cell renal cell (HR=2.146, P<0.0001) and stomach (HR=1.457, P=0.035) (Table S3).

We next employed the receiver operating characteristic (ROC) method to assess the predictive performance (specificity and sensitivity) of the 16-gene signature in determining 5-year overall survival rates. As revealed by the area under the ROC curves (AUCs), we observed that the signature yielded modestly higher AUCs compared to those of TNM staging in all six cohorts: bladder (AUC=0.707 vs. AUC=0.626), colon (AUC=0.673 vs. AUC=0.652), head and neck (AUC=0.624 vs. AUC=0.606), pan-kidney (AUC=0.779 vs. AUC=0.717), clear cell renal cell (AUC=0.740 vs. AUC=0.717) and stomach (AUC=0.754 vs. AUC=0.561) (Fig. 4). Importantly,

when the signature was used as a combined model with TNM staging, we observed a further increase in AUC suggesting that the signature offered incremental predictive value: bladder (AUC=0.713), colon (AUC=0.723), head and neck (AUC=0.663), pan-kidney (AUC=0.833), clear cell renal cell (AUC=0.818) and stomach (AUC=0.757) (Fig. 4).

## Association of Wnt drivers with tumor hypoxia

Poor vascularization in solid tumors results in tumor hypoxia that is frequently associated with very poor prognosis due to reduced effectiveness of chemotherapy and radiotherapy[31]. Furthermore, the stabilization of the hypoxia inducible factor (HIF) in hypoxic tumor microenvironments can promote metastasis and cancer progression leading to poor prognosis[32–34]. An emerging view on cancer stem cells postulates that hypoxic regions could serve as stem cell niches to provide an oxidative DNA damage-buffered zone for cancer stem cells[35,36]. Moreover, crosstalk between HIFs and stem cell signal transduction pathways (Wnt, Notch and *Oct4*) have been reported[37,38]. For instance, HIF-1α can interact with β-catenin to promote stem cell adaptation in hypoxic conditions[39].

Multiple evidence suggests that Wnt signaling may be influenced by the extent of hypoxia within the tumor microenvironment. We reason that hypoxia could further enhance Wnt signaling to allow cancer stem cells to persist, which together contribute to even poorer survival outcomes in patients. Integrating hypoxia information with the 16-Wnt-gene signature would enable the evaluation of the crosstalk between both pathways and its clinical relevance. We predict that patients with more hypoxic tumors would have higher expression of Wnt

driver genes, which may imply that these patients have higher proportions of tumor-initiating cells with hyperactive Wnt signaling. To assess tumor hypoxia levels, we utilized a computationally derived hypoxia gene signature comprising of 52 genes[24]. Hypoxia scores were calculated for each patient as the average expression of the 52 genes. Interestingly, significant positive correlations were observed between the 16-Wnt-gene scores and hypoxia scores in bladder (rho=0.365, P<0.0001) and clear cell renal cell cancers (rho=0.305, P<0.0001), suggesting that in these two cancers, hypoxic tumors had higher expression of core Wnt drivers (Fig. 5A).

To determine the clinical relevance of this positive association, we separated patients into four groups: 1) high scores for both 16-gene and hypoxia, 2) high 16-gene score and low hypoxia score, 3) low 16-gene score and high hypoxia score and 4) low scores for both 16-gene and hypoxia (Fig. 5A). Kaplan-Meier analyses were performed on the four patient groups and we observed that the combined relation of Wnt hyperactivation and hypoxia was significantly associated with overall survival in both cancers: bladder (P=0.009) and clear cell renal cell (P<0.0001) (Fig. 5B). Notably, patients with high hypoxia and high 16-gene scores had significantly higher mortality rates compared to those with low hypoxia and low 16-gene scores: bladder (HR=1.897, P=0.0096) and clear cell renal cell (HR=2.946, P<0.0001) (Fig. 5C). Overall, our results suggest that the joint effect of elevated hypoxia and Wnt signaling is linked to more aggressive disease states.

## Wnt hyperactivation is responsible for epithelial-to-mesenchymal transition properties through decreased cell adhesion

317 Given the poor survival outcomes in patients with high 16-gene scores, we wanted to assess
318 the biological consequences of hyperactive Wnt signaling. Patients were median-stratified into
319 two categories, high- and low-score, for differential expression analyses. For each cancer, the
320 number of differentially expressed genes ($-1 > \log_2$ fold-change $> 1$, $P<0.05$) were 1,543
321 (bladder), 1,164 (colon), 984 (head and neck), 659 (pan-kidney), 943 (clear cell renal cell) and
322 328 (stomach) (Table S4) (Fig. S2). Gene ontology (GO) enrichment analyses revealed
323 enrichment of biological processes consistent with those of cancer stem cells: cell proliferation,
324 cell differentiation, embryo development and cell morphogenesis (Fig. 6A). Moreover, despite
325 their diverse tissue origins, high-score patients from all six cancers exhibited remarkably similar
326 biological alterations (Fig. 6A) (Table S4). For example, high-score patients appear to show a
327 phenotype associated with loss of cell adhesion properties. Genes involved in regulating cell
328 adhesion were downregulated and the 'cell adhesion' GO term was among the most enriched
329 ontologies across all six cancers (Fig. 6A). As a further confirmation, differentially expressed
330 genes were mapped to the KEGG database and enrichments of ontology related to cell
331 adhesion molecules were similarly observed (Fig. 6B). A third database known as Reactome[26]
332 was used in functional enrichment analyses. Comparing results from both KEGG and Reactome
333 analyses revealed enrichments of additional processes related to oncogenesis and Wnt
334 signaling; e.g. altered metabolism, PPAR signaling, MAPK signaling, TGF-β signaling, Hedgehog
335 signaling, calcium signaling, collagen synthesis and degradation, focal adhesion and chemokine
336 signaling (Fig. 6B, C). Within the tumor microenvironment, collagen can modulate extracellular
337 matrix conformation that could paradoxically promote tumor progression[40,41]. Indeed, we
338 observed the enrichment of numerous collagen-related Reactome pathways: assembly of
339 collagen fibrils, collagen biosynthesis, collagen formation, collagen chain trimerization and
340 collagen degradation (Fig. 6C). Overall, our results suggest that elevated mortality risks in high-

score patients could potentially be due to loss of cell adhesion and aggravated disease states exacerbated by Wnt hyperactivation.

To determine the extent of the loss of adhesive properties in tumor cells expressing high levels of Wnt driver genes, we examined the expression profiles of 32 genes from the major cadherin superfamily. Major cadherins are a group of highly conserved proteins that encode at least five cadherin repeats, which include type I and II classical cadherins (*CDH1, CDH2, CDH3, CDH4, CDH5, CDH6, CDH7, CDH8, CDH9, CDH10, CDH11, CDH12, CDH13, CDH15, CDH18, CDH19, CDH20, CDH22, CDH24* and *CDH26*), 7D cadherins (*CDH16* and *CDH17*), desmosomal cadherins (*DSC1, DSC2, DSC3, DSG1, DSG2, DSG3* and *DSG4*) and CELSR cadherins (*CELSR1, CELSR2* and *CELSR3*)[42]. Spearman's correlation analyses between major cadherins and each of the individual Wnt driver genes revealed that the 16 genes exhibited a global pattern of negative correlation with major cadherins across all six cancer types (Fig. 6E). Taken together, these results provide further support to the notion on loss of cadherin-mediated cell adhesion in tumor cells with hyperactive Wnt signaling, which may act in concert to promote neoplastic progression.

## A role for *EZH2* histone methyltransferase in cancer stem cells

When analyzing transcription factor (TF) binding to differentially expressed genes described in the previous section, we observed that these genes were enriched for targets of several notable TFs such as EZH2, SUZ12, Nanog, Sox2 and Smad4 (Fig. 6D). Sox2 and Nanog are well-known stem cell markers[43] while EZH2 and SUZ12 are part of the polycomb repressive complex

2 responsible for epigenetic regulation during embryonic development[44,45] (Fig. 6D). The enrichment of target genes of these TFs supports the hypothesis that Wnt hyperactivation is associated with cancer stem cell properties. Aberrations in *EZH2* and *SUZ12* have been linked to cancer progression[46–50] and overexpression of *EZH2* is associated with poor prognosis[51]. Direct crosstalk between *EZH2* function and Wnt signaling has been reported where *EZH2* was shown to inhibit Wnt pathway antagonists to activate Wnt/β-catenin signaling leading to increased cellular proliferation[52]. Moreover, *EZH2* inhibits E-cadherin expression via lncRNA H19 to promote bladder cancer metastasis[53]. To further confirm that the 16 Wnt drivers represent potential markers of cancer stem cells, we performed correlation analyses between 16-Wnt-gene scores and nine well studied cancer stem cell markers: *CD133*[54,55], *NESTIN*[56,57], *CD200*[58,59], *CD44*[60,61], *CD105*[62,63], *CD24*[64,65], *CD90*[66,67], *CD73*[68,69] and *CD29*[70,71]. Cancer stem cell markers are often cancer type specific. Hence, we would not expect all 9 markers to be positively correlated with the Wnt driver gene scores in every cancer type. Our analyses revealed that a vast majority of these markers were significantly positively correlated with 16-Wnt-gene scores. For example, in head and neck, with the exception of *CD24*, all CSC markers were significant. (Fig. S3). The number of significant positive correlations for the remaining cancer cohorts were as follow: bladder (seven genes), colon (five genes), clear cell renal cell (four genes) and stomach (four genes) and pan-kidney (two genes) (Fig. S3).

Since EZH2 binding targets were enriched among differentially expressed genes (confirmed by both ChEA and ENCODE databases) and given the role of EZH2 in cell adhesion and Wnt signaling, we reason that *EZH2* would be overexpressed in tumors with hyperactive Wnt signaling. Indeed, significant positive correlations were observed between 16-Wnt-gene scores and *EZH2* expression in renal cancers: pan-kidney (rho=0.203, P<0.0001) and clear cell renal

cell (rho=0.233, P<0.0001) (Fig. 7A). Patients were further grouped by their 16-gene scores and *EZH2* expression profiles into four categories: 1) high 16-gene score and high *EZH2* expression, 2) high 16-gene score and low *EZH2* expression, 3) low 16-gene score and high *EZH2* expression and 4) low 16-gene score and low *EZH2* expression (Fig. 7A). Interestingly, patients with high 16-gene score that concurrently had high *EZH2* expression had the poorest survival outcomes compared to the others: pan-kidney (P<0.0001) and clear cell renal cell (P<0.0001) (Fig. 7B). This suggests that Wnt hyperactivation and *EZH2* overexpression could synergize to drive tumor progression resulting in significantly higher death risks: pan-kidney (HR=3.444, P<0.0001) and clear cell renal cell (HR=3.633, P<0.0001) (Fig. 7C).

## Discussion and Conclusion

We performed a comprehensive pan-cancer analysis of 147 Wnt pathway genes in 18,484 patients from 21 different cancer types to unravel the intricacies of Wnt regulation of cancer phenotypes. Taking into account genomic, transcriptomic and clinical data, we demonstrated that overexpression of Wnt genes is underpinned by somatically acquired gene amplifications (Fig. 1). We found that differential Wnt activation contributed to significant heterogeneity in disease progression and survival outcomes. Focusing on 16 core Wnt drivers that were recurrently amplified and overexpressed, our results confirmed that Wnt hyperactivation drove malignant progression that is conserved across diverse cancer types (Fig. 2, 3, 4). Our newly developed 16-Wnt-gene signature could predict patients with more aggressive disease states who may benefit from treatment with small molecule inhibitors of Wnt[16,72,73].

Copy number amplification and concomitant overexpression of WNT driver genes in bladder, colon, head and neck, renal and stomach cancers were significantly associated with stem cell-like molecular features (Fig. 6). The transcriptional profiles of 16 Wnt drivers were negatively correlated with the expression of a vast majority of major cadherin genes involved cell adhesion; a process that may drive epithelial-to-mesenchymal transition (EMT)[74](Fig. 6E). This is consistent with the role of Wnts as inducers of EMT[75]. Patients with high expression of Wnt driver genes exhibited enriched biological processes involving cytokine, TGF-β and Hedgehog signaling (Fig. 6); these components are also implicated in regulating EMT induction[75]. TGF-β activation orchestrates signaling events activating downstream effectors such as Smad proteins that play essential roles in cellular differentiation[76]. Indeed, we observed that dysregulated genes in tumors with hyperactive Wnt signaling were enriched for Smad4 targets

(Fig. 6D). Smads can bind to Zeb proteins to repress E-cadherin expression during the onset of EMT[77,78]. The downregulation of major cadherins in tumors expressing high levels of Wnt drivers (Fig. 6E) could thus be a combined result of aberrant Wnt and TGF-β signaling.

Patients with Wnt hyperactivation exhibited additional molecular features of undifferentiated cancer stem cells. We observed enrichments of stem cell-related TFs such as Nanog, Sox2 and polycomb proteins (SUZ12 and EZH2) as upstream targets of Wnt-associated dysregulated genes; this pattern was consistent across the different cancer types (Fig. 6D). Patients with Wnt hyperactivation phenotypes could have poorly differentiated tumors reminiscent of cancer stem cells given their preferential misexpression of genes normally associated with embryonic stem cell function (Fig. 7). The distinction between cancer stem cells and normal stem cells is of paramount interest. Molecular footprints of stemness identified from analyzing the transcriptional changes between high- and low-16-WNT-gene-score patients could provide additional evidence of cancer stem cell identity in these tumors that is linked to poor overall prognosis.

Our results also demonstrated that Wnt signaling is positively correlated with tumor hypoxia in bladder and clear cell renal cell cancers. Patients with more hypoxic tumors had higher 16-Wnt-gene scores, suggesting that tumor hypoxia may contribute to the activation of Wnt genes. These patients could benefit from the use of hypoxia-modifying drugs such as carbogen and nicotinamide shown to be effective in bladder cancer[79] to reduce tumor hypoxia, which may consequently dampen Wnt signaling. Crosstalk between Wnt signaling and hypoxia has been demonstrated in multiple cancers. β-catenin expression is induced by hypoxia in liver cancer, which contributes to increased EMT, invasion and metastasis[80]. Overexpression of HIF-

1α promoted invasive potential of prostate cancer cells through β-catenin induction, while the silencing of β-catenin in HIF-1α expressing cells resulted in increased and reduced epithelial marker and mesenchymal marker expression respectively[81]. Hypoxia-induced EMT is further enhanced by the addition of recombinant Wnt3a or is repressed by inhibiting β-catenin[82]. Indeed, our results confirmed that increased expression of Wnt driver genes was associated with a global downregulation of major cadherin genes consistent across six cancer types, which may occur through hypoxia-mediated processes (Fig. 6E). We observed that in clear cell renal cell carcinoma, patients with more hypoxic tumors who also had higher Wnt signature scores concomitant with a 2.9-fold higher risk of death (Fig. 5C). Interestingly, renal cancers have a high incidence of VHL mutations[83]. VHL is a protein involved in proteasomal degradation of HIF-1α[84]. VHL antagonizes the Wnt pathway through β-catenin inhibition in renal tumors[85], meaning that VHL mutations would derepress Wnt signaling and create a pseudohypoxic environment to further promote the expression of Wnt pathway genes. Our results will open up new research avenues for investigating the role of the 16 Wnt drivers and potential crosstalk with VHL-mediated HIF signaling in renal cancer.

In summary, we identified Wnt pathway genes that were recurrently amplified and overexpressed across 21 diverse cancer types. A core set of 16 genes known as Wnt drivers were preferentially expressed in high-grade tumors linking to poor overall survival. This signature is a prognostic indicator in six cancer types involving 3,050 patients and is independent and superior to tumor staging parameters, providing additional resolution for patient stratification within similarly staged tumors. We demonstrated clinically relevant relationships between the 16-gene signature, cancer stem cells, cell adhesion, tumor hypoxia and *EZH2* expression. Hence, aggressive tumor behavior and survival outcomes are, in part,

driven by Wnt hyperactivation. Furthermore, we reported evidence for crosstalk between Wnt signaling and other embryonic stem cell pathways (TGF-β signaling, Nanog, Sox2 and polycomb repressive complex 2) confirming that these pathways do not operate in isolation and that interactions between them could add to the complexity of neoplastic progression. Prospective validation in clinical trials and additional functional studies on individual Wnt drivers are needed before they can be harnessed for therapeutic intervention.

**Acknowledgement.** All authors have read the journal's authorship agreement and policy on disclosure of potential conflicts of interest. The authors declare no conflicts of interest. All authors have read the journal's authorship agreement and the manuscript has been reviewed by and approved by all named authors.

**Funding.** None.

**Authors contribution.** WHC and AGL designed the study, analyzed the data and interpreted the data. AGL supervised the research. WHC and AGL wrote the initial manuscript draft. AGL revised the manuscript draft and approved the final version.

**Data statement.** All data supporting the conclusion of this manuscript are included as main and supplementary materials.

## References


1. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001;414(6859):105.
2. Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev cancer*. 2008;8(10):755.
3. Taipale J, Beachy P a. The Hedgehog and Wnt signalling pathways in cancer. *Nature*. 2001;411(May):349-354. doi:10.1038/35077219.
4. Reya T, Clevers H. Wnt signalling in stem cells and cancer. *Nature*. 2005;434(7035):843.
5. Cadigan KM, Nusse R. Wnt signaling: a common theme in animal development. *Genes Dev*. 1997;11(24):3286-3305.
6. Angers S, Moon RT. Proximal events in Wnt signal transduction. *Nat Rev Mol cell Biol*. 2009;10(7):468.
7. van Es JH, Jay P, Gregorieff A, et al. Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nat Cell Biol*. 2005;7(4):381.
8. Andrade AC, Nilsson O, Barnes KM, Baron J. Wnt gene expression in the post-natal growth plate: regulation with chondrocyte differentiation. *Bone*. 2007;40(5):1361-1369.
9. Clevers H. Wnt/β-catenin signaling in development and disease. *Cell*. 2006;127(3):469-480.
10. Nusse R. Wnt signaling in disease and in development. *Cell Res*. 2005;15(1):28.
11. Satoh S, Daigo Y, Furukawa Y, et al. AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat Genet*.

2000;24(3):245.

12. Suzuki H, Watkins DN, Jair K-W, et al. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat Genet*. 2004;36(4):417.

13. Kim MS, Kim SS, Ahn CH, Yoo NJ, Lee SH. Frameshift mutations of Wnt pathway genes AXIN2 and TCF7L2 in gastric carcinomas with high microsatellite instability. *Hum Pathol*. 2009;40(1):58-64.

14. Mart\'\in V, Valencia A, Agirre X, et al. Epigenetic regulation of the non-canonical Wnt pathway in acute myeloid leukemia. *Cancer Sci*. 2010;101(2):425-432.

15. Takebe N, Harris PJ, Warren RQ, Ivy SP. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nat Rev Clin Oncol*. 2011;8(2):97.

16. Emami KH, Nguyen C, Ma H, et al. A small molecule inhibitor of β-catenin/cyclic AMP response element-binding protein transcription. *Proc Natl Acad Sci*. 2004;101(34):12682-12687.

17. He B, Reguart N, You L, et al. Blockade of Wnt-1 signaling induces apoptosis in human colorectal cancer cells containing downstream mutations. *Oncogene*. 2005;24(18):3054.

18. You L, He B, Xu Z, et al. An anti-Wnt-2 monoclonal antibody induces apoptosis in malignant melanoma cells and inhibits tumor growth. *Cancer Res*. 2004;64(15):5385-5389.

19. Tissier F, Cavard C, Groussin L, et al. Mutations of β-catenin in adrenocortical tumors: activation of the Wnt signaling pathway is a frequent event in both benign and malignant adrenocortical tumors. *Cancer Res*. 2005;65(17):7622-7627.

20. White BD, Chien AJ, Dawson DW. Dysregulation of Wnt/β-catenin signaling in gastrointestinal cancers. *Gastroenterology*. 2012;142(2):219-232.

21. Gonzalez-Sancho JM, Aguilera O, Garc\'\ia JM, et al. The Wnt antagonist DICKKOPF-1 gene is a downstream target of β-catenin/TCF and is downregulated in human colon cancer. *Oncogene*. 2005;24(6):1098.

22. Weinstein JN, Collisson EA, Mills GB, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113.

23. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41.

24. Buffa FM, Harris AL, West CM, Miller CJ. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br J Cancer*. 2010;102(2):428-435. doi:10.1038/sj.bjc.6605450.

25. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res*. 2012;40(W1):W478--W483.

26. Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2013;42(D1):D472--D477.

27. Kuleshov M V, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90--W97.

28. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14(1):128.

29. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016. http://ggplot2.org.

30. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based

561 tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*.
562 2015;16(1):169.
563 31. Semenza GL. Hypoxia-inducible factors: mediators of cancer progression and targets
564 for cancer therapy. *Trends Pharmacol Sci*. 2012;33(4):207-214.
565 32. Lu X, Kang Y. Hypoxia and hypoxia-inducible factors (HIFs): master regulators of
566 metastasis. *Clin cancer Res*. 2010:clincanres--1360.
567 33. Chang WH, Forde D, Lai AG. A novel signature derived from immunoregulatory and
568 hypoxia genes predicts prognosis in liver and five other cancers. *J Transl Med*.
569 2019;17(1):14. doi:10.1186/s12967-019-1775-9.
570 34. Chang WH, Forde D, Lai AG. Dual prognostic role for 2-oxoglutarate oxygenases in ten
571 diverse cancer types: Implications for cell cycle regulation and cell adhesion
572 maintenance. *bioRxiv*. 2018. doi:10.1101/442947.
573 35. Heddleston JM, Li Z, McLendon RE, Hjelmeland AB, Rich JN. The hypoxic
574 microenvironment maintains glioblastoma stem cells and promotes reprogramming
575 towards a cancer stem cell phenotype. *Cell cycle*. 2009;8(20):3274-3284.
576 36. Mohyeldin A, Garzón-Muvdi T, Quiñones-Hinojosa A. Oxygen in stem cell biology: a
577 critical component of the stem cell niche. *Cell Stem Cell*. 2010;7(2):150-161.
578 37. Keith B, Simon MC. Hypoxia-inducible factors, stem cells, and cancer. *Cell*.
579 2007;129(3):465-472.
580 38. Holland JD, Klaus A, Garratt AN, Birchmeier W. Wnt signaling in stem and cancer stem
581 cells. *Curr Opin Cell Biol*. 2013;25(2):254-264.
582 39. Kaidi A, Williams AC, Paraskeva C. Interaction between β-catenin and HIF-1 promotes
583 cellular adaptation to hypoxia. *Nat Cell Biol*. 2007;9(2):210.
584 40. Provenzano PP, Inman DR, Eliceiri KW, et al. Collagen density promotes mammary

585 tumor initiation and progression. *BMC Med*. 2008;6(1):11.
586 41. Shintani Y, Maeda M, Chaika N, Johnson KR, Wheelock MJ. Collagen I promotes
587 epithelial-to-mesenchymal transition in lung cancer cells via transforming growth
588 factor-β signaling. *Am J Respir Cell Mol Biol*. 2008;38(1):95-104.
589 42. Hulpiau P, Van Roy F. Molecular evolution of the cadherin superfamily. *Int J Biochem*
590 *Cell Biol*. 2009;41(2):349-369.
591 43. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2
592 coding regions modulate embryonic stem cell differentiation. *Nature*.
593 2008;455(7216):1124.
594 44. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. The polycomb group protein Suz12
595 is required for embryonic stem cell differentiation. *Mol Cell Biol*. 2007;27(10):3769-
596 3779.
597 45. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in
598 human embryonic stem cells. *Cell*. 2006;125(2):301-313.
599 46. Yoo KH, Hennighausen L. EZH2 methyltransferase and H3K27 methylation in breast
600 cancer. *Int J Biol Sci*. 2012;8(1):59.
601 47. Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is
602 involved in progression of prostate cancer. *Nature*. 2002;419(6907):624.
603 48. Zingg D, Debbache J, Schaefer SM, et al. The epigenetic modifier EZH2 controls
604 melanoma growth and metastasis through silencing of distinct tumour suppressors.
605 *Nat Commun*. 2015;6:6051.
606 49. Fan Y, Shen B, Tan M, et al. TGF-β-induced upregulation of malat1 promotes bladder
607 cancer metastasis by associating with suz12. *Clin cancer Res*. 2014.
608 50. Li H, Cai Q, Wu H, et al. SUZ12 promotes human epithelial ovarian cancer by

609 suppressing apoptosis via silencing HRK. *Mol Cancer Res*. 2012.
610 51. Wagener N, Macher-Goeppinger S, Pritsch M, et al. Enhancer of zeste homolog 2
611 (EZH2) expression is an independent prognostic factor in renal cell carcinoma. *BMC*
612 *Cancer*. 2010;10(1):524.
613 52. Cheng ASL, Lau SS, Chen Y, et al. EZH2-mediated concordant repression of Wnt
614 antagonists promotes β-catenin--dependent hepatocarcinogenesis. *Cancer Res*. 2011.
615 53. Luo M, Li Z, Wang W, Zeng Y, Liu Z, Qiu J. Long non-coding RNA H19 increases bladder
616 cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression.
617 *Cancer Lett*. 2013;333(2):213-221.
618 54. Liu G, Yuan X, Zeng Z, et al. Analysis of gene expression and chemoresistance of
619 CD133+ cancer stem cells in glioblastoma. *Mol Cancer*. 2006;5(1):67.
620 55. Ma S, Chan KW, Lee TK-W, et al. Aldehyde dehydrogenase discriminates the CD133
621 liver cancer stem cell populations. *Mol Cancer Res*. 2008;6(7):1146-1153.
622 56. Neradil J, Veselska R. Nestin as a marker of cancer stem cells. *Cancer Sci*.
623 2015;106(7):803-811.
624 57. Zhang M, Song T, Yang L, et al. Nestin and CD133: valuable stem cell-specific markers
625 for determining clinical outcome of glioma patients. *J Exp Clin cancer Res*.
626 2008;27(1):85.
627 58. Kawasaki BT, Farrar WL. Cancer stem cells, CD200 and immunoevasion. *Trends*
628 *Immunol*. 2008;29(10):464-468.
629 59. Moreaux J, Veyrune JL, Reme T, De Vos J, Klein B. CD200: a putative therapeutic target
630 in cancer. *Biochem Biophys Res Commun*. 2008;366(1):117-122.
631 60. Takaishi S, Okumura T, Tu S, et al. Identification of gastric cancer stem cells using the
632 cell surface marker CD44. *Stem Cells*. 2009;27(5):1006-1020.

633 61. Du L, Wang H, He L, et al. CD44 is of functional importance for colorectal cancer stem
634 cells. *Clin cancer Res*. 2008;14(21):6751-6760.
635 62. Bussolati B, Dekel B, Azzarone B, Camussi G. Human renal cancer stem cells. *Cancer*
636 *Lett*. 2013;338(1):141-146.
637 63. Saroufim A, Messai Y, Hasmim M, et al. Tumoral CD105 is a novel independent
638 prognostic marker for prognosis in clear-cell renal cell carcinoma. *Br J Cancer*.
639 2014;110(7):1778.
640 64. Gao MQ, Choi YP, Kang S, Youn JH, Cho NH. CD24+ cells from hierarchically organized
641 ovarian cancer are enriched in cancer stem cells. *Oncogene*. 2010;29(18):2672.
642 65. Hurt EM, Kawasaki BT, Klarmann GJ, Thomas SB, Farrar WL. CD44+ CD24- prostate
643 cells are early cancer progenitor/stem cells that provide a model for patients with poor
644 prognosis. *Br J Cancer*. 2008;98(4):756.
645 66. Yang ZF, Ho DW, Ng MN, et al. Significance of CD90+ cancer stem cells in human liver
646 cancer. *Cancer Cell*. 2008;13(2):153-166.
647 67. Jiang J, Zhang Y, Chuai S, et al. Trastuzumab (herceptin) targets gastric cancer stem
648 cells characterized by CD90 phenotype. *Oncogene*. 2012;31(6):671.
649 68. López J, Valdez-Morales FJ, Benitez-Bribiesca L, Cerbón M, Carrancá AG. Normal and
650 cancer stem cells of the human female reproductive system. *Reprod Biol Endocrinol*.
651 2013;11(1):53.
652 69. McLean K, Gong Y, Choi Y, et al. Human ovarian carcinoma--associated mesenchymal
653 stem cells regulate cancer stem cells and tumorigenesis via altered BMP production. *J*
654 *Clin Invest*. 2011;121(8):3206-3219.
655 70. Geng S, Guo Y, Wang Q, Li L, Wang J. Cancer stem-like cells enriched with CD29 and
656 CD44 markers exhibit molecular characteristics with epithelial--mesenchymal

transition in squamous cell carcinoma. *Arch Dermatol Res*. 2013;305(1):35-47.

71. Moncharmont C, Levy A, Gilormini M, et al. Targeting a cornerstone of radiation resistance: cancer stem cell. *Cancer Lett*. 2012;322(2):139-147.

72. Chen B, Dodge ME, Tang W, et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat Chem Biol*. 2009;5(2):100.

73. Takahashi-Yanaga F, Kahn M. Targeting Wnt signaling: can we safely eradicate cancer stem cells? *Clin cancer Res*. 2010:432-1078.

74. Nelson WJ, Nusse R. Convergence of Wnt, ß-catenin, and cadherin pathways. *Science (80- )*. 2004;303(5663):1483-1487.

75. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene*. 2010;29(34):4741.

76. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-β family signalling. *Nature*. 2003;425(6958):577.

77. Comijn J, Berx G, Vermassen P, et al. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell*. 2001;7(6):1267-1278.

78. Postigo AA, Depp JL, Taylor JJ, Kroll KL. Regulation of Smad signaling through a differential recruitment of coactivators and corepressors by ZEB proteins. *EMBO J*. 2003;22(10):2453-2462.

79. Hoskin PJ, Rojas AM, Bentzen SM, Saunders MI. Radiotherapy with concurrent carbogen and nicotinamide in bladder carcinoma. *J Clin Oncol*. 2010;28(33):4912-4918.

80. Liu L, Zhu X-D, Wang W-Q, et al. Activation of β-catenin by hypoxia in hepatocellular carcinoma contributes to enhanced metastatic potential and poor prognosis. *Clin cancer Res*. 2010:432-1078.

81. Zhao J-H, Luo Y, Jiang Y-G, He D-L, Wu C-T. Knockdown of β-Catenin through shRNA cause a reversal of EMT and metastatic phenotypes induced by HIF-1α. *Cancer Invest*. 2011;29(6):377-382.

82. Zhang Q, Bai X, Chen W, et al. Wnt/β-catenin signaling enhances hypoxia-induced epithelial--mesenchymal transition in hepatocellular carcinoma via crosstalk with hif-1α signaling. *Carcinogenesis*. 2013;34(5):962-973.

83. Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat Genet*. 1994;7(1):85.

84. Haase VH. The VHL/HIF oxygen-sensing pathway and its relevance to kidney disease. *Kidney Int*. 2006;69(8):1302-1307.

85. Chitalia VC, Foy RL, Bachschmid MM, et al. Jade-1 inhibits Wnt signalling by ubiquitylating β-catenin and mediates Wnt pathway inhibition by pVHL. *Nat Cell Biol*. 2008;10(10):1208.

695 Figure legends

696

697 **Figure 1. Pan-cancer core drivers of Wnt signaling. (A)** Schematic diagram depicting the study
698 design and the identification of core Wnt driver genes subsequently representing the 16-gene
699 signature. A total of 147 Wnt signaling genes representing both canonical and non-canonical
700 pathways alongside their downstream targets were obtained from the KEGG database. Genes
701 were grouped into two categories depending on whether they were associated with active or
702 inactive Wnt signaling. Somatic copy number variations in all 147 genes were determined in
703 21 cancer types. A total of 61 genes were recurrently amplified in at least 20% of tumors in
704 each cancer type. They included 41 genes associated with active Wnt signaling. Of the 41
705 genes, 16 genes (core Wnt drivers) were upregulated in tumor compared to non-tumor
706 samples in at least 8 cancer types. Cox proportional hazards regression and Kaplan-Meier
707 analyses were performed using the 16-gene signature, which demonstrated its ability to
708 predict overall survival in at least six cancer types: bladder, colon, head and neck, clear cell
709 renal cell, papillary renal cell, chromophobe renal cell and stomach cancers (n=3,050).
710 Associations of the 16-Wnt-gene signature with cancer stem cell features, tumor hypoxia and
711 cell adhesion were investigated. Potential clinical applications of the signature were proposed.
712 **(B)** Somatic amplification and differential expression profiles of 61 Wnt genes. Cumulative bar
713 chart depicts the number of cancer types with at least 20% of tumors with somatic gains. The
714 heatmap on the left shows the extent of genomic amplifications for each of the 61 genes
715 separated into 'active' and 'inactive' Wnt signaling categories across 21 cancer types. Heatmap
716 intensities indicate the fraction of the cohort in which a given gene is gained or amplified. The
717 columns were ordered using hierarchical clustering with Euclidean distance metric to reveal
718 cancers that have similar somatic amplification profiles. The heatmap on the right

719 demonstrates differential expression values ($log_2$) between tumor and non-tumor samples for
720 each of the 61 genes. Genes marked in red represent the 16 Wnt driver genes. These are genes
721 that were amplified in at least 20% of tumors in at least 8 cancers and genes that were
722 overexpressed (fold-change > 1.5) in at least 8 cancers. Refer to Table S2 for cancer
723 abbreviations.
724
725 **Figure 2. Survival analyses using the 16-Wnt-gene signature in six cancer cohorts.** Kaplan-Meier
726 analyses of overall survival on patients stratified into high- and low-score groups using the 16-
727 gene signature. P values were determined from the log-rank test.
728
729 **Figure 3. The 16-Wnt-gene signature is independent of TNM stage.** (A) Kaplan-Meier analyses
730 were performed on patients categorized according to tumor TNM stages that were further
731 stratified using the 16-gene signature. P values were determined from the log-rank test. TNM:
732 tumor, node, metastasis. (B) Expression of Wnt drivers (16-gene scores) increased with tumor
733 stage. P values were determined from the ANOVA test.
734
735 **Figure 4. Predictive performance of the 16-Wnt-gene signature is superior to TNM staging.**
736 Prediction of five-year overall survival was assessed using the receiver operating characteristic
737 (ROC) analysis to determine specificity and sensitivity of the signature. ROC curves were
738 generated based on the 16-gene signature, TNM stage and a combination of the signature and
739 TNM stage. AUC: area under the curve. TNM: tumor, node, metastasis.
740
741

**Figure 5. Positive associations between the 16-gene signature and tumor hypoxia in bladder and clear cell renal cell cancers. (A)** Scatter plots show significant positive correlation between 16-gene scores and hypoxia scores as determined by Spearman's rank-order correlation analyses. Patients were separated and color-coded into four categories based on median 16-gene and hypoxia scores. **(B)** Kaplan-Meier analyses were performed on the four patient categories to determine the effects of the combined relationship between hypoxia and the Wnt signature on overall survival. **(C)** Univariate Cox proportional hazards analysis of the relation between the 16-gene signature and hypoxia. CI: confidence interval.

**Figure 6. Wnt hyperactivation is associated with a cancer stem cell-like phenotype.** Patients were median separated into high- and low-score groups using the 16-gene signature for differential expression analyses. Enrichments of biological processes on differentially expressed genes were determined by mapping the genes to **(A)** Gene Ontology, **(B)** KEGG and **(C)** Reactome databases. Significantly enriched pathways or ontologies for all six cancer cohorts were depicted. **(D)** Differentially expressed genes were enriched for targets of stem cell-related transcription factors (Nanog, Sox2, Smad4, EZH2 and SUZ12) as confirmed by mapping to ENCODE and ChEA databases. Refer to Table S2 for cancer abbreviations. **(E)** Significant negative correlations between the expression profiles of individual Wnt driver genes and 32 major cadherin genes. Heatmaps were generated based on Spearman's correlation coefficient values.

**Figure 7. Positive associations between the 16-gene signature and *EZH2* expression in renal cancers. (A)** Scatter plots show significant positive correlation between 16-gene scores and *EZH2* expression as determined by Spearman's rank-order correlation analyses. Patients were

766 separated and color-coded into four categories based on median 16-gene score and *EZH2*
767 expression. **(B)** Kaplan-Meier analyses were performed on the four patient categories to
768 determine the effects of the combined relationship between *EZH2* expression and the Wnt
769 signature on overall survival. **(C)** Univariate Cox proportional hazards analysis of the relation
770 between the 16-gene signature and *EZH2* expression. CI: confidence interval.

Supplementary figures and tables

**Figure S1. Prognosis of each of the 16 signature genes in 20 cancer types as determined using Cox regression analyses.** Both columns (cancer types) and rows (Wnt genes) were ordered using hierarchical clustering (Euclidean distance metric). Grey boxes represent non-prognostic genes. Heatmap intensities represent hazard ratios of prognostic genes that were significant (P<0.05).

**Figure S2. Venn diagram depicts a six-way comparison of the differentially expressed genes identified from high-score versus low-score patients in all six cancer cohorts.** Numbers in parentheses represent the number of differentially expressed genes (-1 > log2 fold-change > 1, P<0.05) in each cancer.

**Figure S3. Correlations between the Wnt gene signature and nine cancer stem cell markers.** Scatter plots depict the associations between 16-gene scores and cancer stem cell marker expression levels in six cancer cohorts. P values were determined by Spearman's rank-order correlation analyses. Refer to Table S2 for cancer abbreviations.

**Table S1.** List of 147 genes associated with Wnt signaling.

**Table S2.** Abbreviations and number of tumor and non-tumor samples in TCGA cancers.

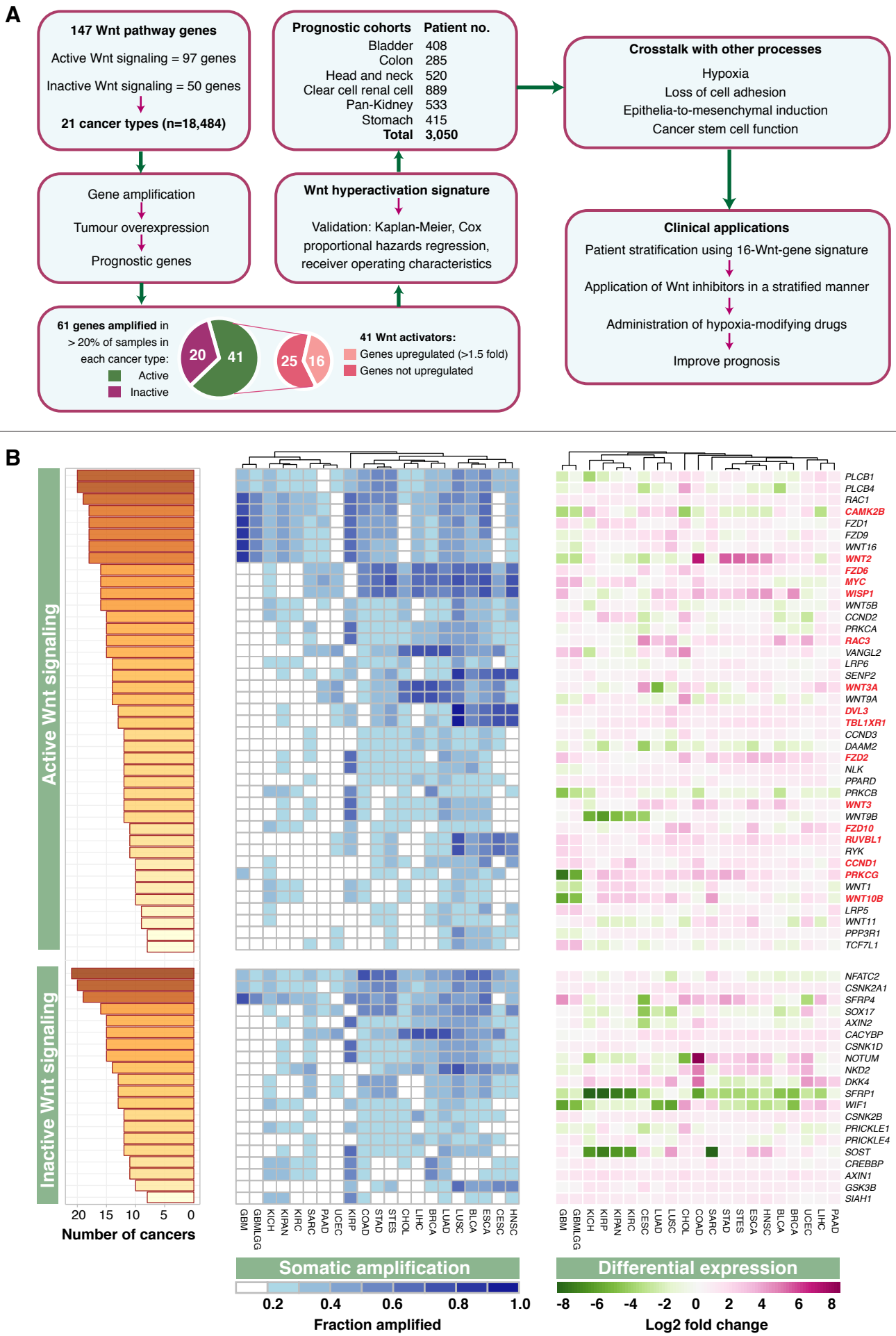
**Table S3.** Univariate and multivariate Cox proportional hazards analysis of risk factors associated with overall survival in multiple cancers.

795

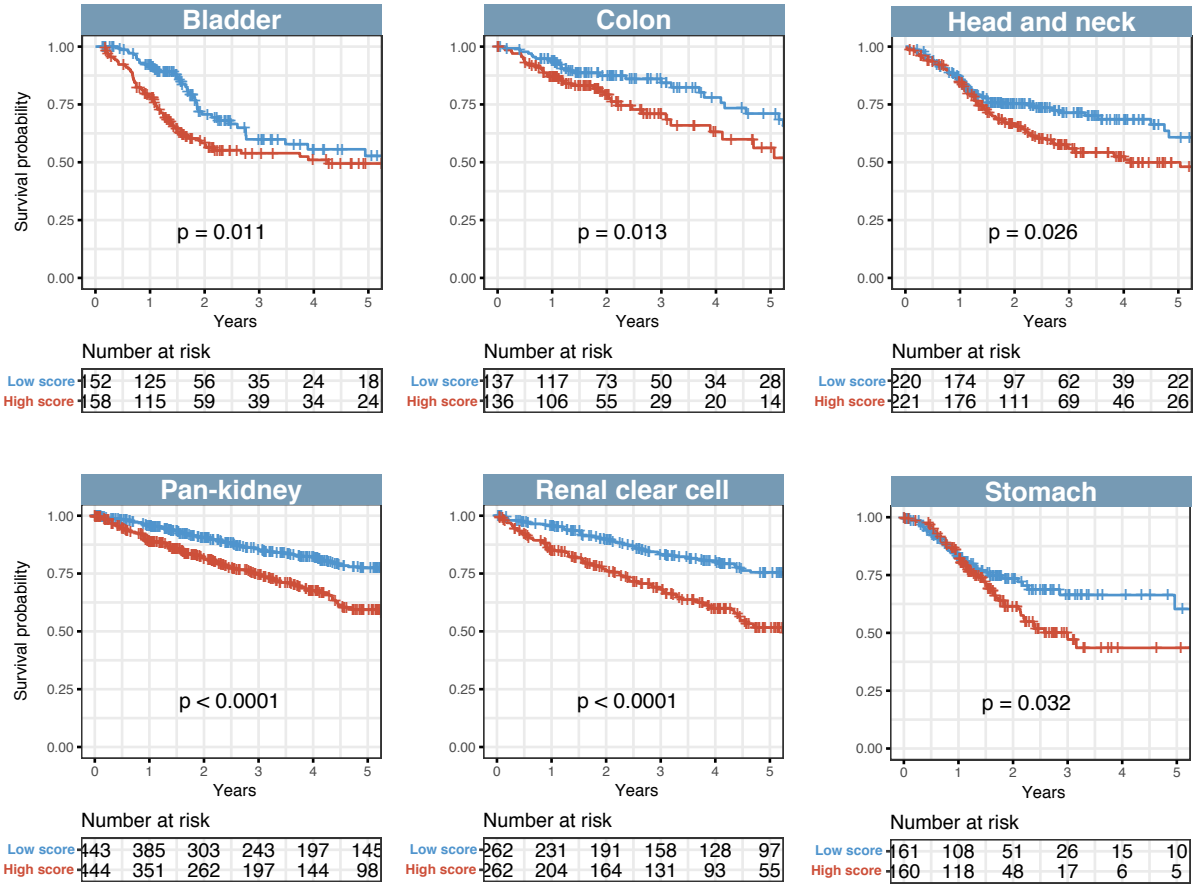
796 **Table S4.** Differentially expressed genes between high- and low 16-Wnt-score patient groups

797 in six cancers.

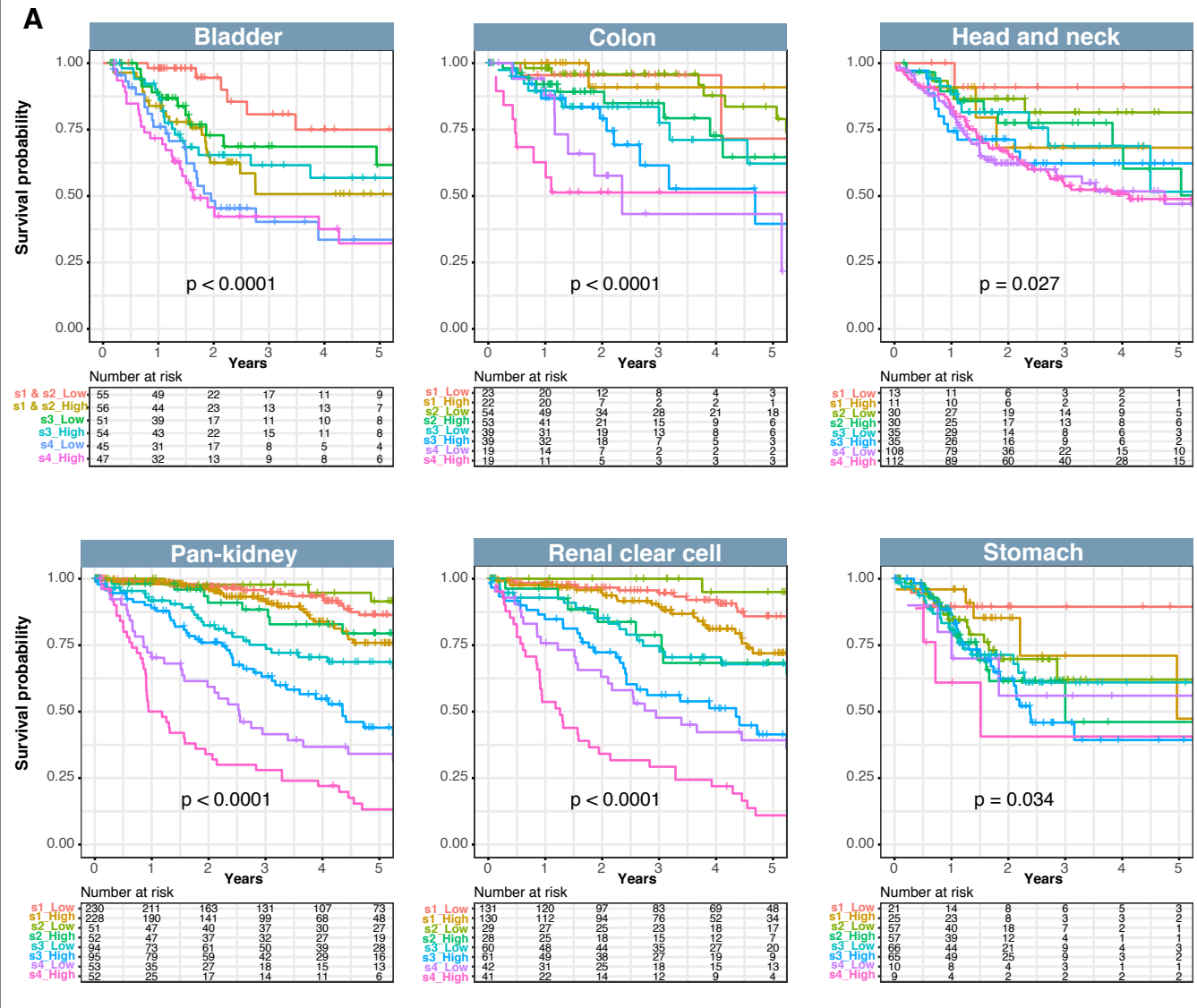
# Figure 1


A
147 Wnt pathway genes
Active Wnt signaling = 97 genes
Inactive Wnt signaling = 50 genes
21 cancer types (n=18,484)
Prognostic cohorts Patient no.
Bladder 408
Colon 285
Head and neck 520
Clear cell renal cell 889
Pan-Kidney 533
Stomach 415
Total 3,050
Crosstalk with other processes
Hypoxia
Loss of cell adhesion
Epithelia-to-mesenchymal induction
Cancer stem cell function
Gene amplification
Tumour overexpression
Prognostic genes
Wnt hyperactivation signature
Validation: Kaplan-Meier, Cox proportional hazards regression, receiver operating characteristics
Clinical applications
Patient stratification using 16-Wnt-gene signature
Application of Wnt inhibitors in a stratified manner
Administration of hypoxia-modifying drugs
Improve prognosis
61 genes amplified in > 20% of samples in each cancer type:
Active
Inactive
20
41
25
16
41 Wnt activators:
Genes upregulated (>1.5 fold)
Genes not upregulated
B
Active Wnt signaling
Inactive Wnt signaling
20 15 10 5 0
Number of cancers
GBM GBMLGG KICH KIPAN KIRC SARC PAAD UCEC KIRP COAD STAD STES CHOL LIHC BRCA LUAD LUSC BLCA ESCA CESC HNSC
Somatic amplification
0.2 0.4 0.6 0.8 1.0
Fraction amplified
PLCB1
PLCB4
RAC1
CAMK2B
FZD1
FZD9
WNT16
WNT2
FZD6
MYC
WISP1
WNT5B
CCND2
PRKCA
RAC3
VANGL2
LRP6
SENP2
WNT3A
WNT9A
DVL3
TBL1XR1
CCND3
DAAM2
FZD2
NLK
PPARD
PRKCB
WNT3
WNT9B
FZD10
RUVBL1
RYK
CCND1
PRKCG
WNT1
WNT10B
LRP5
WNT11
PPP3R1
TCF7L1
NFATC2
CSNK2A1
SFRP4
SOX17
AXIN2
CACYBP
CSNK1D
NOTUM
NKD2
DKK4
SFRP1
WIF1
CSNK2B
PRICKLE1
PRICKLE4
SOST
CREBBP
AXIN1
GSK3B
SIAH1
GBM GBMLGG KICH KIRP KIPAN KIRC CESC LUAD LUSC CHOL COAD SARC STAD STES ESCA HNSC BLCA BRCA UCEC LIHC PAAD
Differential expression
-8 -6 -4 -2 0 2 4 6 8
Log2 fold change

# Figure 2


Bladder
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.011
0 1 2 3 4 5
Years
Number at risk
Low score 152 125 56 35 24 18
High score 158 115 59 39 34 24
Colon
p = 0.013
Years
Number at risk
Low score 137 117 73 50 34 28
High score 136 106 55 29 20 14
Head and neck
p = 0.026
Years
Number at risk
Low score 220 174 97 62 39 22
High score 221 176 111 69 46 26
Pan-kidney
Survival probability
p < 0.0001
Years
Number at risk
Low score 443 385 303 243 197 145
High score 444 351 262 197 144 98
Renal clear cell
p < 0.0001
Years
Number at risk
Low score 262 231 191 158 128 97
High score 262 204 164 131 93 55
Stomach
p = 0.032
Years
Number at risk
Low score 161 108 51 26 15 10
High score 160 118 48 17 6 5

# Figure 3


A
Bladder
Colon
Head and neck
Pan-kidney
Renal clear cell
Stomach
Survival probability
Years
p < 0.0001
p < 0.0001
p = 0.027
p < 0.0001
p < 0.0001
p = 0.034
Number at risk
s1 & s2_Low 55 49 22 17 11 9
s1 & s2_High 56 44 23 13 13 7
s3_Low 51 39 17 11 10 8
s3_High 54 43 22 15 11 8
s4_Low 45 31 17 8 5 4
s4_High 47 32 13 9 8 6
s1_Low 23 20 12 8 4 3
s1_High 22 20 7 2 2 1
s2_Low 54 49 34 28 21 18
s2_High 53 41 21 15 9 6
s3_Low 39 31 19 13 8 6
s3_High 39 32 18 7 5 3
s4_Low 19 14 7 2 2 2
s4_High 19 11 5 3 3 3
s1_Low 13 11 6 3 2 1
s1_High 11 10 6 2 2 1
s2_Low 30 27 19 14 9 5
s2_High 30 25 17 13 8 6
s3_Low 35 29 14 8 6 3
s3_High 35 26 16 9 6 2
s4_Low 108 79 36 22 15 10
s4_High 112 89 60 40 28 15
s1_Low 230 211 163 131 107 73
s1_High 228 190 141 99 68 48
s2_Low 51 47 40 37 30 27
s2_High 52 47 37 32 27 19
s3_Low 94 73 61 50 39 28
s3_High 95 79 59 42 29 16
s4_Low 53 35 27 18 15 13
s4_High 52 25 17 14 11 6
s1_Low 131 120 97 83 69 48
s1_High 130 112 94 76 52 34
s2_Low 29 27 25 23 18 17
s2_High 28 25 18 15 12 7
s3_Low 60 48 44 35 27 20
s3_High 61 49 38 27 19 9
s4_Low 42 31 25 18 15 13
s4_High 41 22 14 12 9 4
s1_Low 21 14 8 6 5 3
s1_High 25 23 8 3 3 2
s2_Low 57 40 18 7 2 1
s2_High 57 39 12 4 1 1
s3_Low 66 44 21 9 4 3
s3_High 65 49 25 9 3 2
s4_Low 10 8 4 3 1 1
s4_High 9 4 2 2 2 2

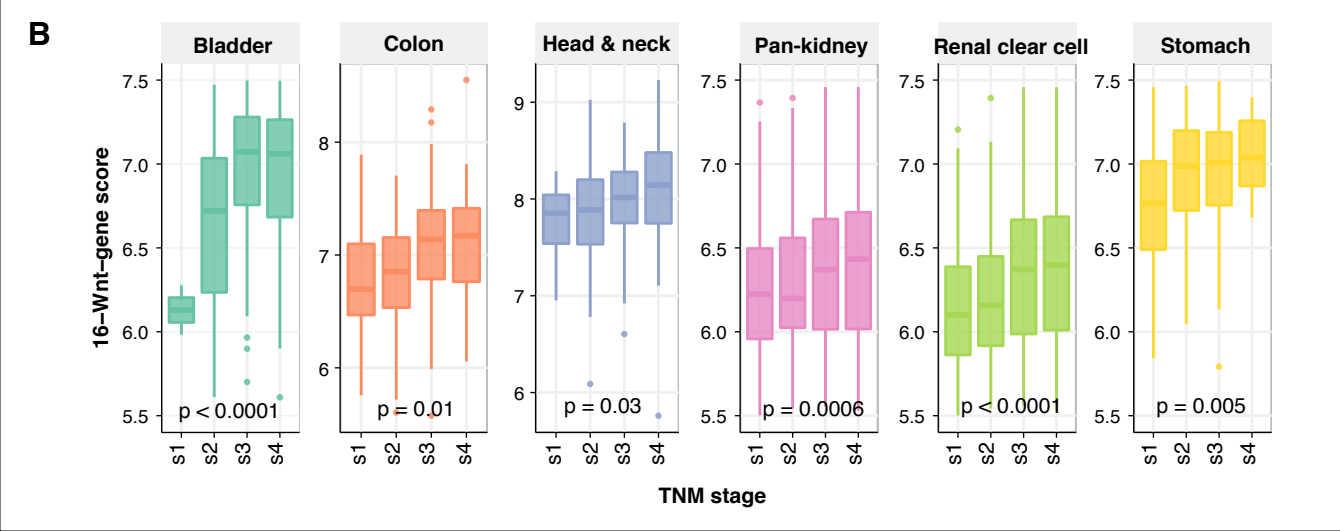
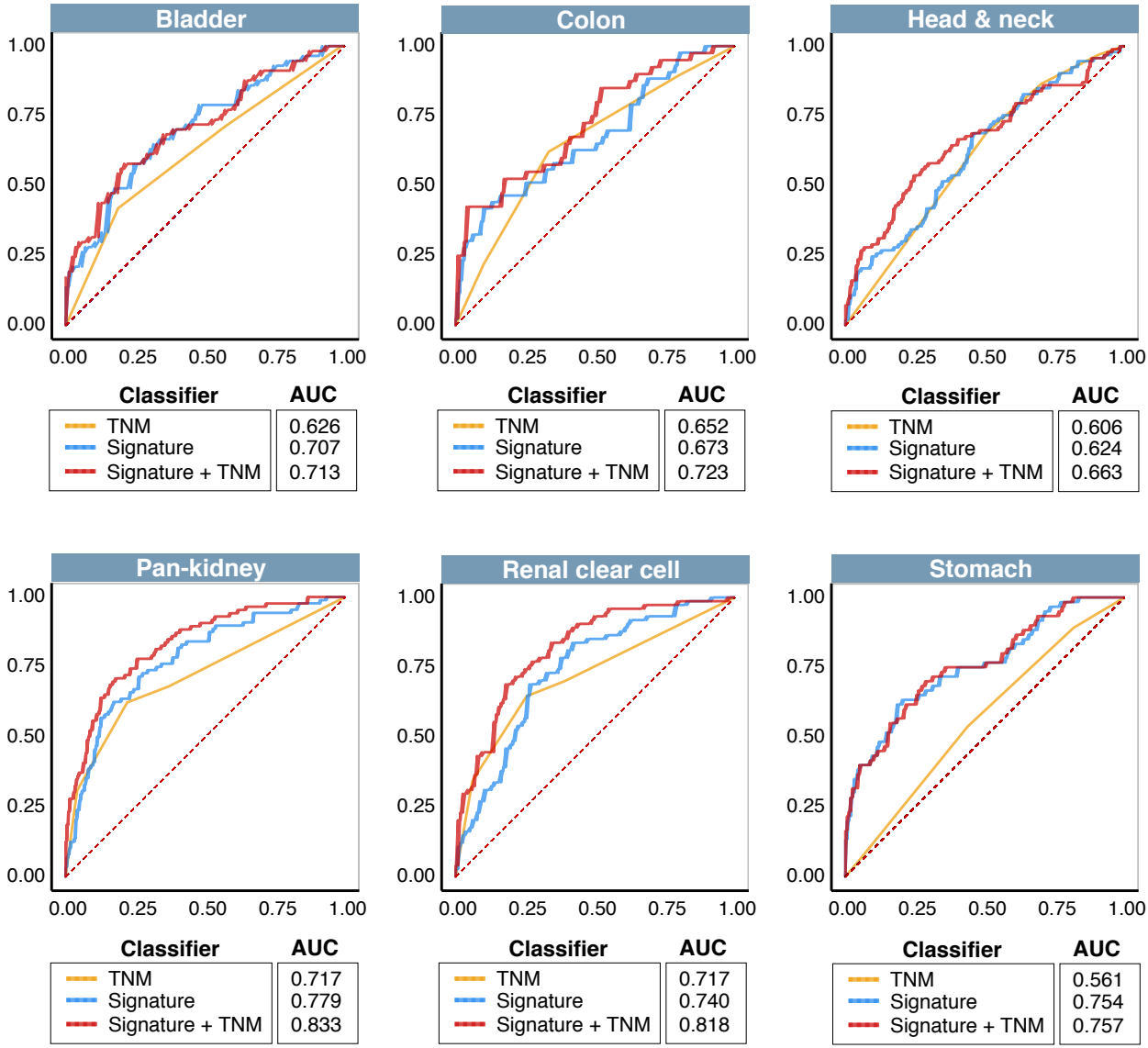
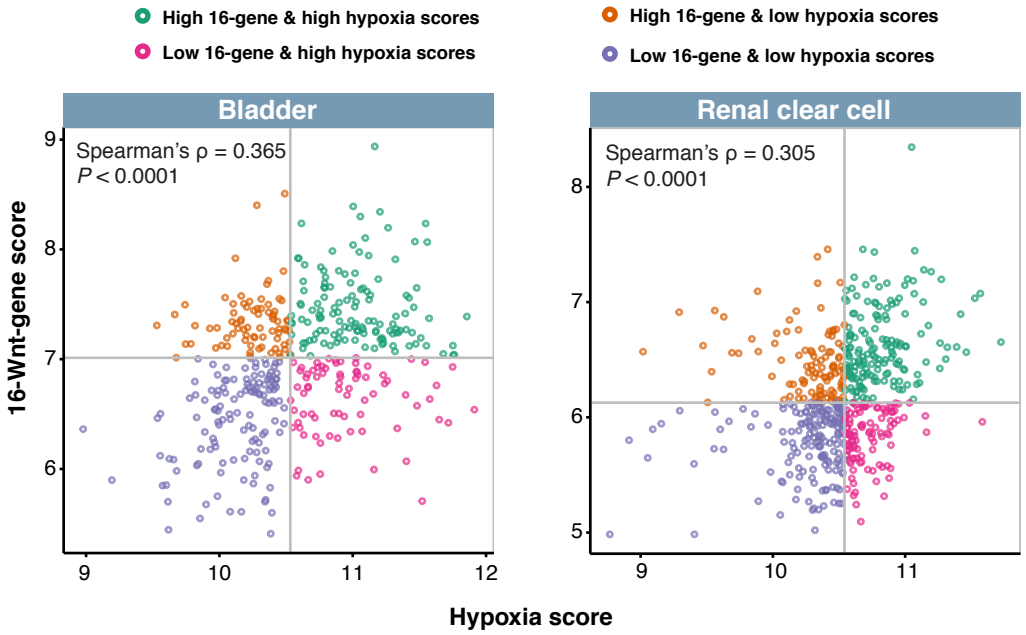

B
Bladder
Colon
Head & neck
Pan-kidney
Renal clear cell
Stomach
16−Wnt−gene score
p < 0.0001
p = 0.01
p = 0.03
p = 0.0006
p < 0.0001
p = 0.005
s1
s2
s3
s4
TNM stage

**Figure 4**

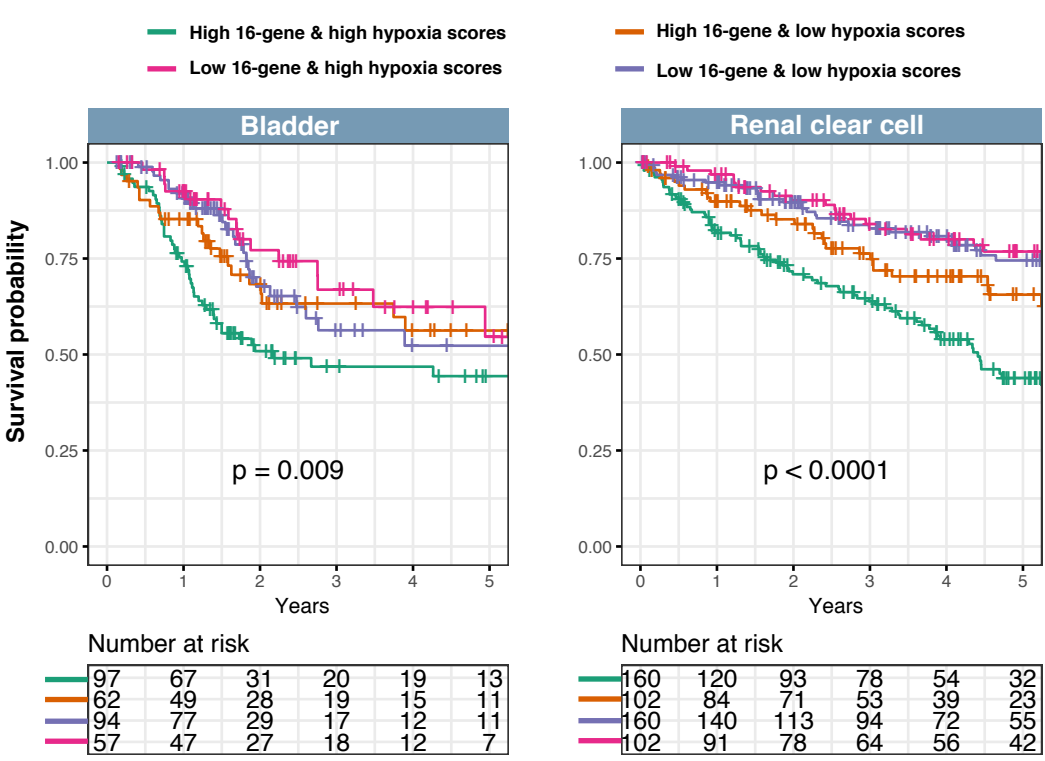

Bladder
Colon
Head & neck
Pan-kidney
Renal clear cell
Stomach
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
Classifier
AUC
TNM
Signature
Signature + TNM
0.626
0.707
0.713
0.652
0.673
0.723
0.606
0.624
0.663
0.717
0.779
0.833
0.717
0.740
0.818
0.561
0.754
0.757

# Figure 5

**A**



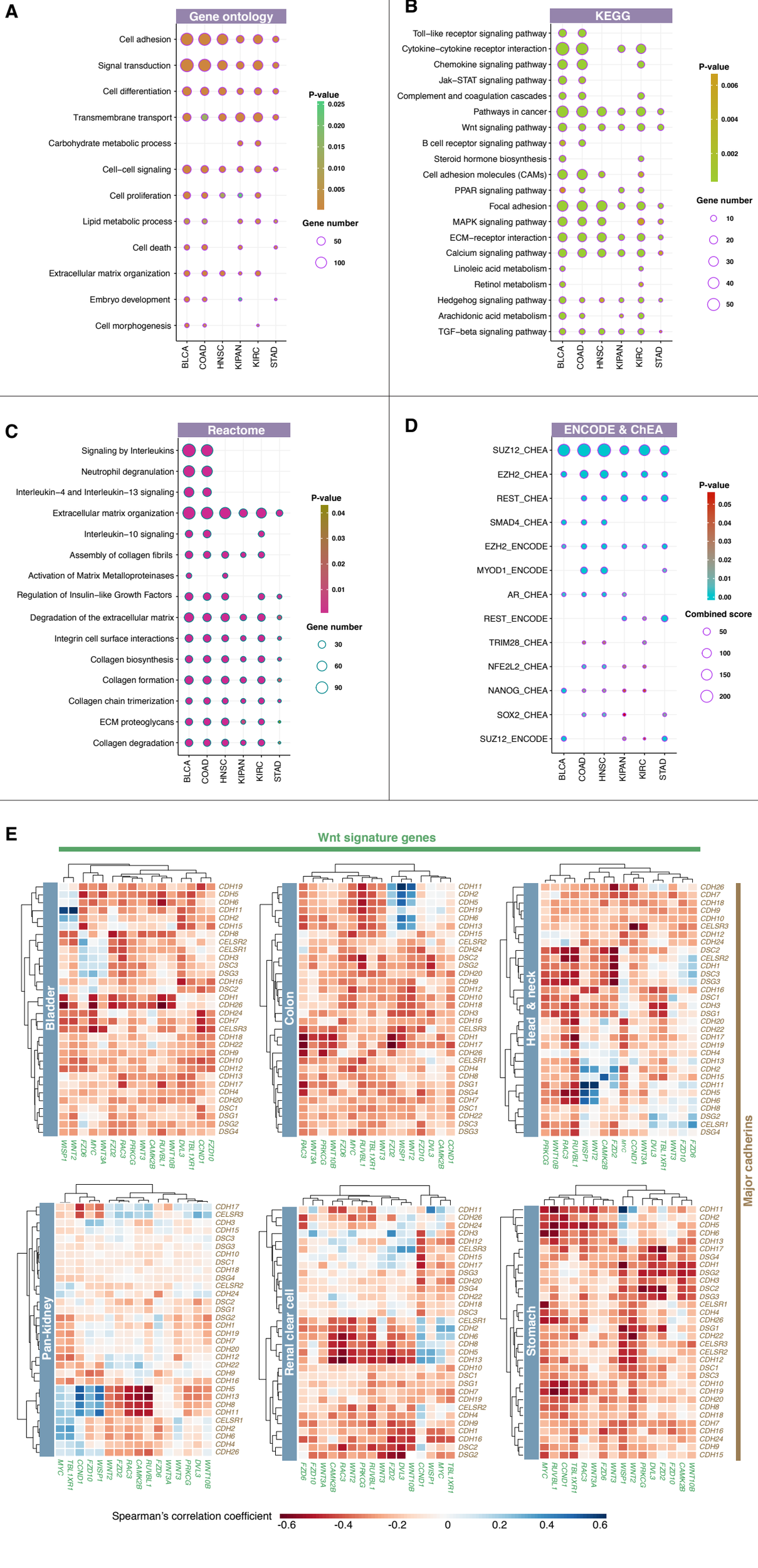

**B**




**C**

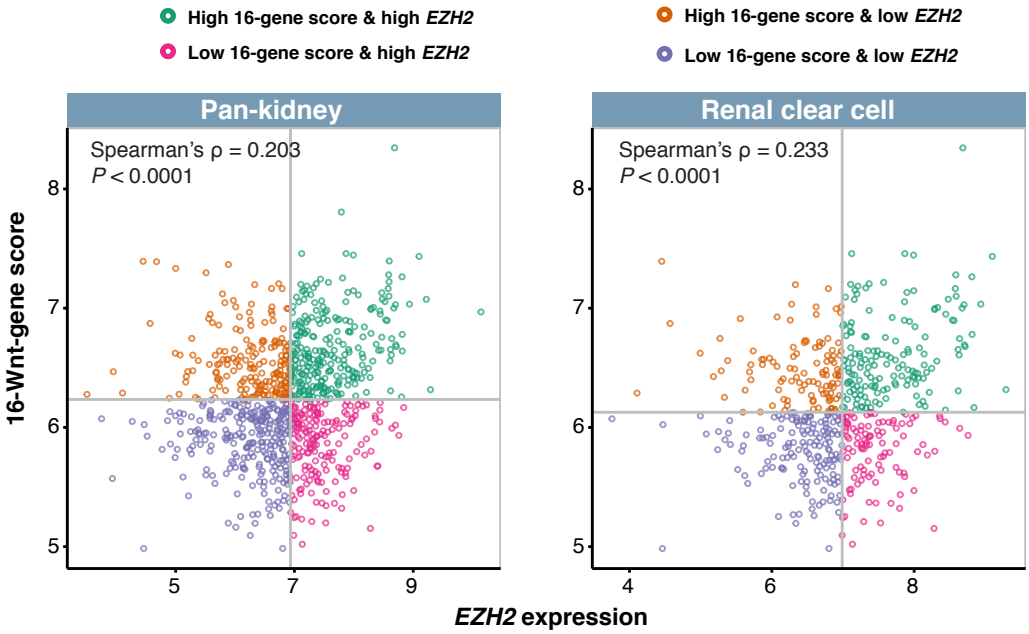
| | Hazard Ratio (95% CI) | P-value |
|---|---|---|
| **Bladder** | | |
| High 16-gene score & high hypoxia score vs. low 16-gene score & low hypoxia score | 1.897 (1.168 - 3.079) | **0.0096** |
| High 16-gene score & low hypoxia score vs. low 16-gene score & low hypoxia score | 1.179 (0.664 - 2.095) | 0.57 |
| Low 16-gene score & high hypoxia score vs. low 16-gene score & low hypoxia score | 0.859 (0.453 - 1.632) | 0.64 |
| **Clear cell renal cell** | | |
| High 16-gene score & high hypoxia score vs. low 16-gene score & low hypoxia score | 2.946 (1.972 - 4.399) | **1.29E-07** |
| High 16-gene score & low hypoxia score vs. low 16-gene score & low hypoxia score | 1.690 (1.052 - 2.714) | **0.03** |
| Low 16-gene score & high hypoxia score vs. low 16-gene score & low hypoxia score | 1.040 (0.628 - 1.723) | 0.88 |

# Figure 6

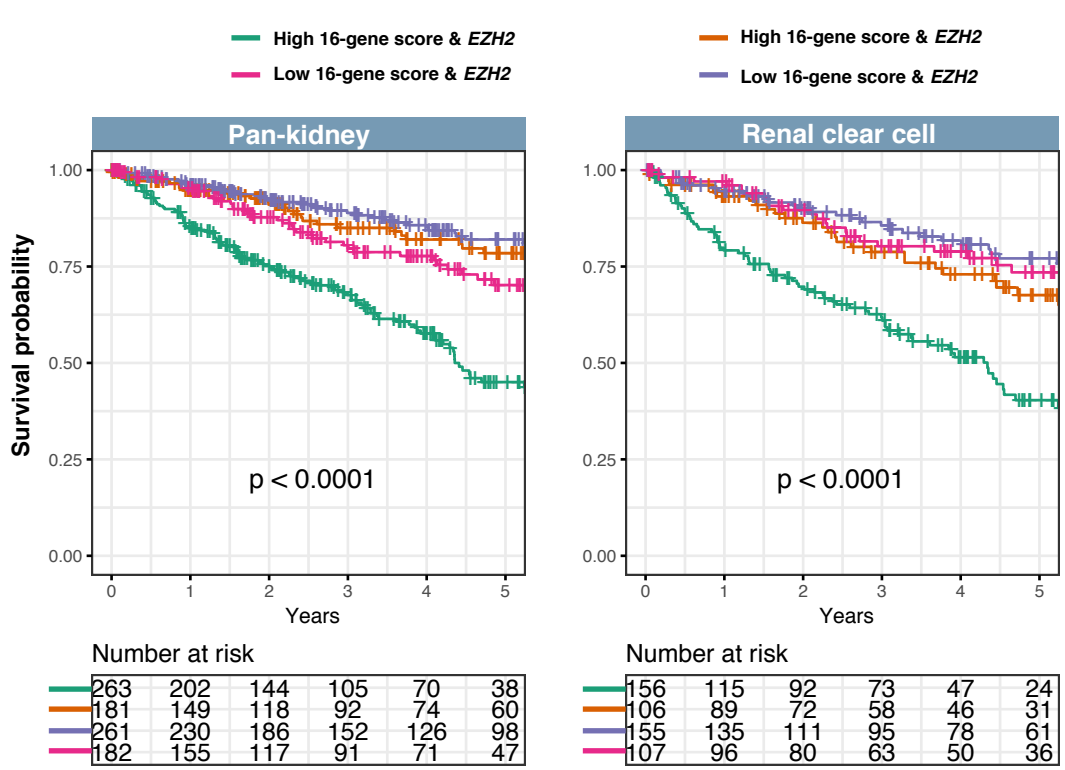

A
Gene ontology
Cell adhesion
Signal transduction
Cell differentiation
Transmembrane transport
Carbohydrate metabolic process
Cell–cell signaling
Cell proliferation
Lipid metabolic process
Cell death
Extracellular matrix organization
Embryo development
Cell morphogenesis
BLCA
COAD
HNSC
KIPAN
KIRC
STAD
P-value
0.025
0.020
0.015
0.010
0.005
Gene number
50
100
B
KEGG
Toll–like receptor signaling pathway
Cytokine–cytokine receptor interaction
Chemokine signaling pathway
Jak–STAT signaling pathway
Complement and coagulation cascades
Pathways in cancer
Wnt signaling pathway
B cell receptor signaling pathway
Steroid hormone biosynthesis
Cell adhesion molecules (CAMs)
PPAR signaling pathway
Focal adhesion
MAPK signaling pathway
ECM–receptor interaction
Calcium signaling pathway
Linoleic acid metabolism
Retinol metabolism
Hedgehog signaling pathway
Arachidonic acid metabolism
TGF–beta signaling pathway
P-value
0.006
0.004
0.002
Gene number
10
20
30
40
50
C
Reactome
Signaling by Interleukins
Neutrophil degranulation
Interleukin–4 and Interleukin–13 signaling
Extracellular matrix organization
Interleukin–10 signaling
Assembly of collagen fibrils
Activation of Matrix Metalloproteinases
Regulation of Insulin–like Growth Factors
Degradation of the extracellular matrix
Integrin cell surface interactions
Collagen biosynthesis
Collagen formation
Collagen chain trimerization
ECM proteoglycans
Collagen degradation
P-value
0.04
0.03
0.02
0.01
Gene number
30
60
90
D
ENCODE & ChEA
SUZ12_CHEA
EZH2_CHEA
REST_CHEA
SMAD4_CHEA
EZH2_ENCODE
MYOD1_ENCODE
AR_CHEA
REST_ENCODE
TRIM28_CHEA
NFE2L2_CHEA
NANOG_CHEA
SOX2_CHEA
SUZ12_ENCODE
P-value
0.05
0.04
0.03
0.02
0.01
0.00
Combined score
50
100
150
200
E
Wnt signature genes
Bladder
Colon
Head & neck
Pan-kidney
Renal clear cell
Stomach
Major cadherins
Spearman's correlation coefficient
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

# Figure 7

**A**




**B**




**C**

| | Hazard Ratio (95% CI) | *P*-value |
|---|---|---|
| **Pan-Kidney** | | |
| High 16-gene score & high *EZH2* vs. low 16-gene score & low *EZH2* | 3.444 (2.430 - 4.882) | **3.66E-12** |
| High 16-gene score & low *EZH2* vs. low 16-gene score & low *EZH2* | 1.075 (0.683 - 1.693) | 0.75 |
| Low 16-gene score & high *EZH2* vs. low 16-gene score & low *EZH2* | 1.665 (1.105 - 2.508) | **0.014** |
| **Clear cell renal cell** | | |
| High 16-gene score & high *EZH2* vs. low 16-gene score & low *EZH2* | 3.633 (2.412 - 5.471) | **6.63E-10** |
| High 16-gene score & low *EZH2* vs. low 16-gene score & low *EZH2* | 1.564 (0.959 - 2.549) | 0.073 |
| Low 16-gene score & high *EZH2* vs. low 16-gene score & low *EZH2* | 1.282 (0.776 - 2.115) | 0.33 |