Oncogene (2016), 1–12

www.nature.com/onc

ORIGINAL ARTICLE

A CDK2 activity signature predicts outcome in CDK2-low cancers

SR McCurdy[1,2], M Pacal[1], M Ahmad[1] and R Bremner[1,2]

The role of cyclin-dependent kinase 2 (CDK2) in cancer is controversial. A major hurdle is the availability of tools to easily assess its activity across many samples. Here, we introduce a transcriptional signature to specifically track CDK2 activity. It responds to genetic and chemical perturbations in the CDK-RB-E2F axis, correlates with mitotic rate *in vitro* and *in vivo* and reacts rapidly to changes in CDK2 activity during cell cycle progression. We find that CDK2 activity is specifically elevated in human testes, mirroring its critical function in mice, and report very distinct profiles across human cancers. Increased CDK2 activity decreases risk in colon cancer, but elevates poor outcome two- to fivefold in specific tumors, including low grade glioma, kidney, thyroid, adrenocortical and prostate cancer. These are typically 'CDK2-low' cancers, suggesting that above a certain threshold CDK2 promotes progression, but further increases do not influence outcome. Multivariate analysis revealed that the CDK2 signature is the most important predictive feature in these cancers versus dozens of other clinical parameters, such as tumor grade or mitotic index. Thus, transcriptome data provides a novel, straightforward method to monitor CDK2 activity, implicates key roles for the kinase in a subset of human tissues and tumors and enhances cancer risk prediction. The strategy used here for CDK2 could be applied to other kinases that influence transcription.

Oncogene advance online publication, 7 November 2016; doi:10.1038/onc.2016.409

INTRODUCTION

The cyclin-dependent kinase (CDK) family regulates multiple processes including cell cycle regulation, transcription, survival and development.[1] It includes 20 CDKs, 4 CLKs (Cdc2-like kinases), 5 CDKLs (CDK-like) and >29 Cyclins.[2] CDK1/2/4/6 control cell division through inhibitory and stimulatory phosphorylation of targets such as the retinoblastoma protein (RB), which binds and represses E2F transcription factors that regulate many genes involved in cell cycle progression.[3] Non-canonical CDKs like CDK7-9 regulate transcription. For example, CDK7 phosphorylates the C-terminal domain (CTD) heptad repeats of RNA polymerase II, driving RNA polymerase II from transcriptional pre-initiation to initiation.[4] CDK8 (which binds to Cyclin C) and CDK9 (which binds to Cyclin T1/2) also phosphorylate the CTD repeat of RNA polymerase II, but drive initiation as well as elongation. Similar to CDK4, CDK8 is amplified in human cancer, although specifically in colon.[5,6] CDK2 also targets proteins linked to transcription,[7–9] but the extent to which it regulates gene expression and whether this could be used to track activity is unclear.

Although many CDKs have been linked to cancer,[10,11] resolving their role is hindered by the practical constraints of assessing kinase activity in multiple contexts. The current gold standard is to measure phosphorylation using a kinase assay or to immunoblot for a phospho-site. For CDK2, Cyclin E S384 can be used, but this marks Cyclin E for degradation and is only useful when *Fbw7* is inactivated.[12,13] More efficient kinase assays remain costly,[14] and elegant live cell tracking systems cannot be applied on a broad scale or to primary human tumors.[15]

The role of CDK2 in cancer is hotly debated,[16–23] likely because CDK2 dependency is context-specific, but also because different assays to block activity can have distinct effects. For example, knockdown/out assays free Cyclins and provide an opportunity for compensation by other CDKs, whereas pharmaceutical inhibition does not, and as dominant negative approaches do not displace cyclins, overexpression may have non-physiological consequences. Some studies indicate that CDK activity predicts outcome, although the approaches are laborious.[24–27] Genetic models of murine retinoblastoma revealed that retinal cells only become tumor-prone once CDK2 activity rises above a certain threshold.[28] Whether this concept applies to other cancers is unknown.

No study has systematically examined CDK2 activity across multiple cell lines and clinical samples. CDK2 influences the cell cycle, thus there have been considerable efforts to define tools that accurately proxy cell cycle activity. Some include staining for cell cycle markers (for example, PH3, KI67, PCNA), transfecting live cell markers (for example, FUCCI),[29] incorporating probes into DNA (for example, EdU, ^{3}H-thymidine) and identifying mitotic figures. None is suitable for a meta-analysis across multiple studies and contexts in a streamlined fashion. Here, we develop a new methodology to examine CDK2 activity. We leveraged RNAi-transcriptome studies to identify a signature, and validated the signature in multiple contexts (Supplementary Figure 1). Applying the signature to large datasets uncovered new insights in CDK2 biology and cancer prediction.

RESULTS

A CDK2 gene signature

To identify a CDK2 signature three CDK2 knockdown datasets from three cell lines were mined: A375 melanoma (GSE31534),[30]

[1]Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, Sinai Health Systems, Toronto, Ontario, Canada and [2]Department of Laboratory Medicine and Pathobiology, Department of Ophthalmology and Vision Science, University of Toronto, Toronto, Ontario, Canada. Correspondence: Dr R Bremner, Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital Sinai Health Systems, Room 840 600 University Avenue, Toronto, Ontario, Canada M5G1X5.
E-mail: bremner@lunenfeld.ca
Received 17 June 2016; revised 30 August 2016; accepted 23 September 2016

IMR32 neuroblastoma (GSE16480)[31] and MCF7 breast cancer cells (GSE31912).[30] To generate a signature with 10–250 genes, we filtered differentially expressed genes with a *z*-score $\leqslant -1.96$ in at least two two cell lines, which identified 208 genes (Figure 1a). Strikingly, analysis of expression data in the Cancer Cell Line Encyclopedia (CCLE)[32] revealed that these genes correlate with CDK2 gene expression across >1000 cell lines (Figure 1a). A total of 10 000 random gene signatures of equal size were generated and used to calculate the absolute correlation (ρ) between CDK2 expression and the expression of genes contained within a signature. The query CDK2 signature had a significantly higher median absolute correlation of 0.23 ($z = -38.025$, $P = 0$) compared with the random population (median = 0.07, interquartile range = 0.00827). To simplify analysis, only genes positively correlated to CDK2 gene expression were used. There were also more induced (142/208) genes, providing a more robust higher performance signature than the fewer down-regulated genes (66/208). To further enhance robustness, a correlational matrix was created using expression of the 142-genes across 8411 patient samples from The Cancer Genome Atlas (TCGA).[33] TCGA is multi-institutional effort to comprehensively examine cancer genomics with matched clinical data. Leveraging this database, we detected a striking correlation across this huge sample size, and refer to the 97 most highly correlated (Figure 1b; Supplementary Table 1) as the 'CDK2 signature'. One of the genes is CDK2 itself, indicating the CDK2 mRNA correlates with CDK2 activity across thousands of human samples.

We mined GeneMania (www.genemania.org, Toronto, ON, Canada) for experimentally validated protein–protein interactions. A total of 73/97 genes in the CDK2 signature formed a singular PPI network around CDK1, CDK2, PLK1 and AURKB, hubs with the highest connections/degree (Figure 2a). The genes were enriched for gene sets relating to prophase/metaphase (Figure 2b), suggesting that CDK2 induces genes to poise the cell for mitosis. Indeed, several prophase/metaphase-specific RNAs are first transcribed during S/G2, but only translated once the cell passes

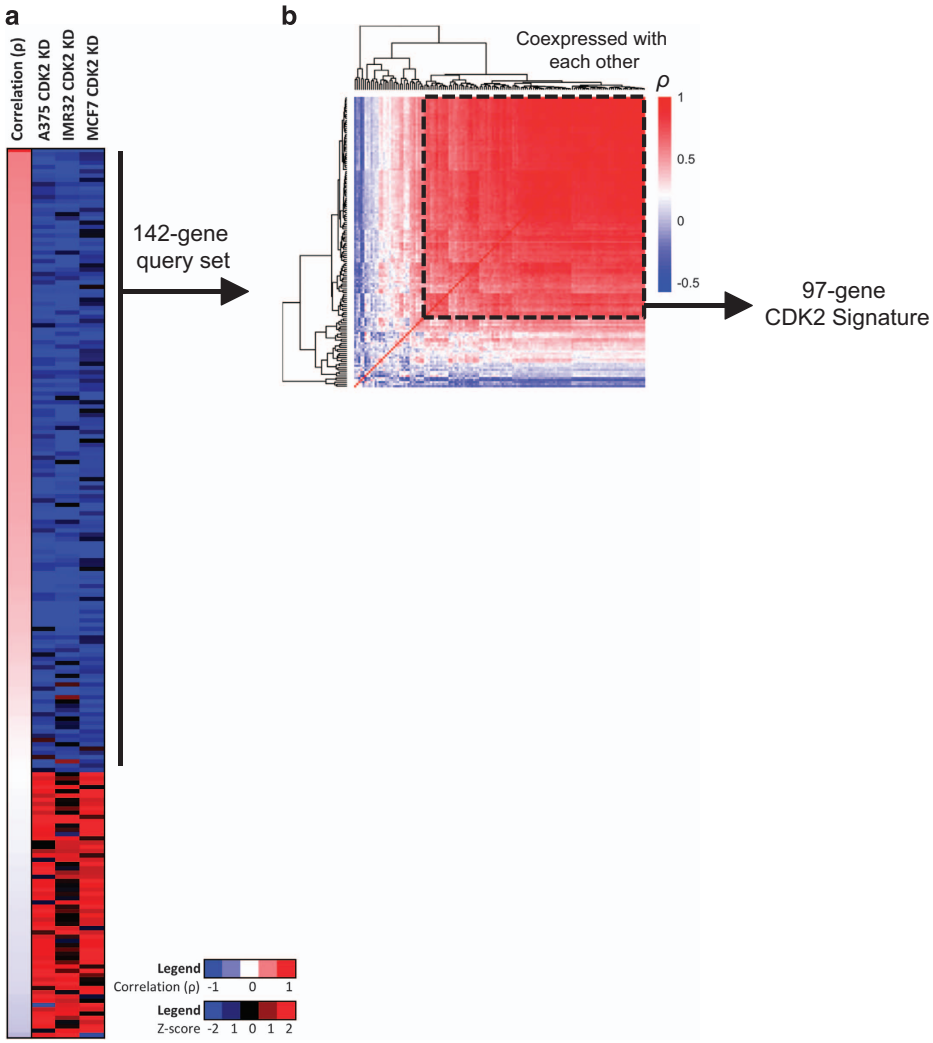


Figure 1. Identification of a CDK2 signature. (**a**) Microarray data was mined from three CDK2 knockdown studies: MCF7 (GSE31912), IMR32 (GSE16480) and A375 (GSE31534). Each dataset was quantile normalized and the log2 difference between CDK2 and control was converted to *z*-scores. The list only contains genes that were significantly down-regulated in at least two studies. We performed a Spearman correlation between the signature genes and CDK2 across 1037 cell lines in the CCLE and ordered the genes in descending order. Genes that are down- or up-regulated on Cdk2 knockdown are positively or negatively correlated with CDK2, respectively. (**b**) Transcriptome data from over 8000 wild-type and tumor samples was mined from the TCGA. The data was filtered to contain only genes that are positively correlated to CDK2 gene expression from **a**. A correlational matrix was created and arranged using hierarchical clustering. The highest correlated cluster was selected to be the 97-gene CDK2 signature (dotted box).

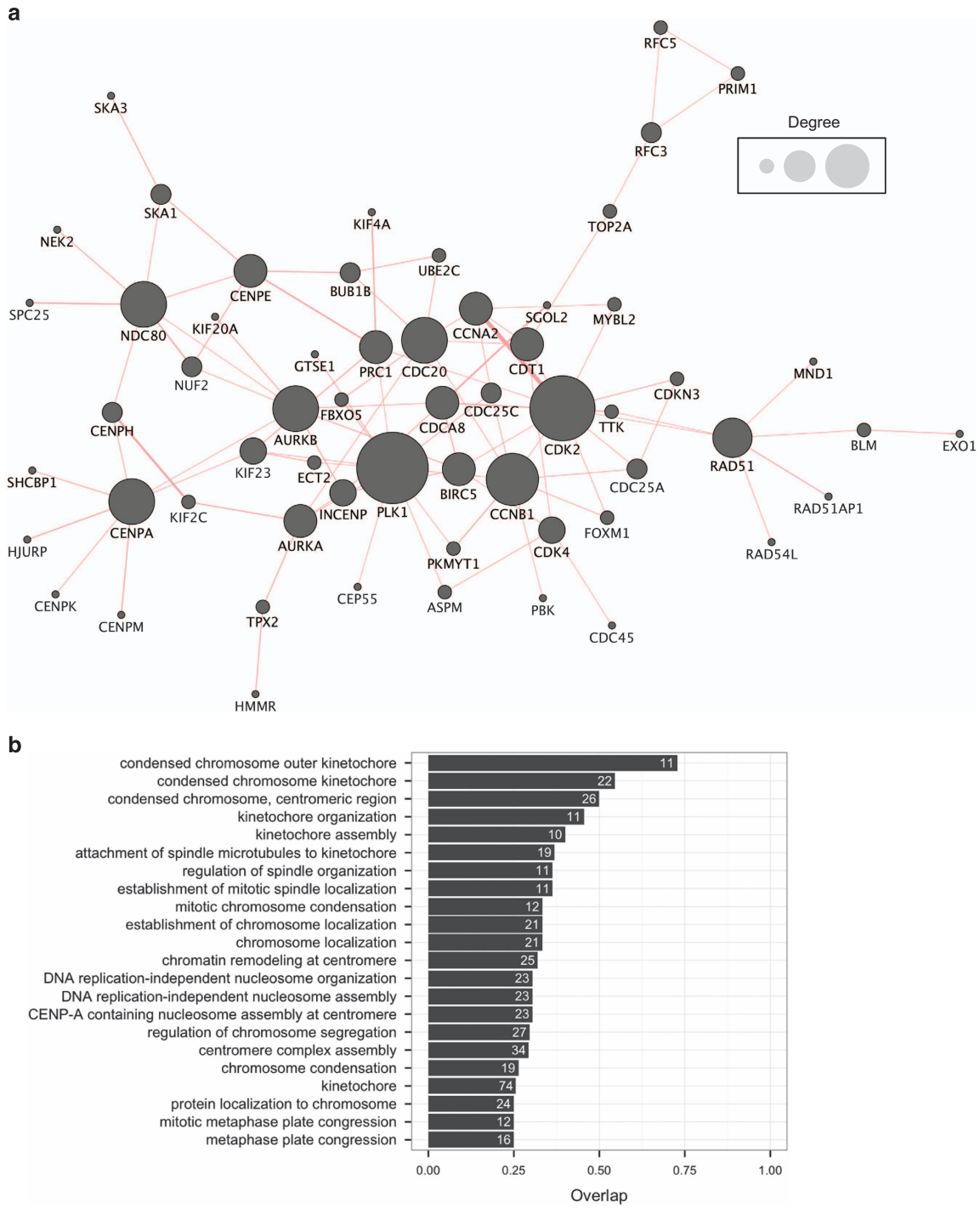


Figure 2. The CDK2 signature produces a highly interconnected protein network. (**a**) A force-directed protein–protein interaction graph derived from the CDK2 signature genes identified in Figure 6. The size of each node correlates to its degree (# of connections). Each line represents a reported physical protein–protein interaction between nodes. The thickness of each of these edges is proportional to the amount of evidence supporting the interaction. (**b**) Functional (GO) annotation for the CDK2 signature with an overlap of greater than 25%. Overlap indicates the fraction of genes in the CDK2 signature present in each gene set. The number with each bar indicates the total size of the respective gene set. All gene sets included are significant ($P < 0.05$).

a G2/M checkpoint.[34] Thus, CDK2 appears to have a major role in preparing the cell at a transcriptional level for mitosis.

CDK2 signature correlates to CDK2 activity in multiple settings

To quantify the CDK2 signature, we normalized gene expression using *z*-scores across samples, which overcomes differences in expression ranges, and exposes which genes are different than the average of all the samples. If all the signature genes are higher on average, then the signature scores is higher. To summarize into a single metric, the median is taken of all signature gene *z*-scores for a given sample (Supplementary Figure 2).

Next we assessed the relationship between the CDK2 signature and the activity of E2F, a transcription factor family that cooperates

with CDK2 to regulate cell cycle progression. Of 130 E2F targets mined from reference,[35] only 19 were part of the CDK2 signature (Figure 3a). In spite of this modest overlap, the common function of E2f and CDK2 function in cell cycle regulation suggest that expression of their targets may correlate. To test this notion, we used the 130 E2F target gene set to generate an E2F signature, as we did for CDK2 above, and found that there was a strong correlation between the CDK2 and E2F signatures across 1037 cell lines in the CCLE database ($\rho = 0.8721$, Figure 3b). Thus the CDK2 signature tracks with E2F activity in a huge number of cell lines derived from multiple tissue types.

RB protein binds and represses E2F to block induction of genes, such as Cyclins A and E that bind and activate CDK2. Thus next we assessed whether the CDK2 signature responds as expected to manipulating RB and E2F. In the developing murine retina, loss of the *Rb* gene increases in Cdk2 activity, which is further enhanced by removing the *Rb* relative *p107* (*Rbl1*), and these effects are blocked by deleting *E2f1*.[28] To test whether the CDK2 signature is Rb and/or E2f-dependent, we performed microarray on E15 retina with different genetic dosages of *Rb*, the Rb family member *p107* and *E2f1*, and then calculated the CDK2 signature score for each of the samples. Progressive loss of *Rb* family members led to incremental increases in the median CDK2 signature ($z_{WT} = -0.0296$, $z_{Rb-/-} = 0.4871$, $z_{Rb-/-p107-/-} = 1.5859$; Figure 3c) and removing *E2f1* decreased the signature in both *Rb*- and *Rb/p107* null retinas (Figure 3c). Finally, these observations were recapitulated using an E2F signature (Figure 3c, right panel). Thus, the CDK2 signature, like CDK2, is Rb-E2f-dependent. Notably, analysis of copy number and mutation data from 462 cancer-associated genes revealed enrichment of *RB1* mutation in CDK2-High cell lines (3.2-fold increase, $P = 4.03e-04$; Supplementary Figures 3A and B). Other significant mutations included amplification of *NOTCH4* (18.3-fold increase, $P = 7.11e-06$), *ERC1* (4.6-fold increase, $P = 8.76e-04$) and *MAPK1* (4.6-fold increase, $P = 8.76e-04$; (Supplementary Figure 3A). Whether these three genes and/or other loci in their respective amplicons influence CDK2 will require additional work, but irrespective, the *RB1* link further validates the signature.

To assess directly whether the signature proxies CDK2 activity we generated microarray data from P8 wild-type, *αCre;Rb*$^{f/f}$ and *αCre;Rb*$^{f/f}$*;p107*$^{-/-}$ retina and compared it with prior data on CDK2 enzyme activity in these contexts.[28] Spearman correlation analysis revealed an excellent correlation ($\rho = 0.9487$, Figure 3d). To test the drug responsiveness of the signature, gene expression data was mined from a study using the pan-CDK inhibitor R547.[36] The signature identified dose-dependent decreases and increased inhibition over time in two cancer cell lines, and in non-dividing PBMCs the CDK2 signature was low, as expected (Figure 3e).

Next, we asked whether the signature captures cell cycle dynamics. Previously, HeLa cells were synchronized using either thymidine-nocodazole block (TN) or double thymidine block (TT), and every 1–2 h RNA was collected for microarray analysis.[37] Strikingly, the CDK2 signature score rose during S–G2 and plummeted at the M–G1 boundary (Figure 3f), exactly as reported using a live cell probe for CDK2 activity.[15] This was observed in multiple cell cycles, clones and synchronization methods, thus the CDK2 signature rapidly proxies cell cycle events.

In summary, the CDK2 signature mimics enzyme activity following genetic or pharmaceutical perturbations, *in vitro* and *in vivo*, and in the rapidly changing context of cell cycle progression. It is thus a robust tool to track CDK2 activity in multiple settings.

Specificity of the CDK2 signature versus other CDK signatures

CDK2 is functionally redundant with CDK1, but mainly binds Cyclin A and E when CDK2 is deleted, rather than in a normal setting.[38,39] The CDK2 signature was conceived using acute knockdown rather than knockout data, and was further honed using correlation analysis across thousands of cell lines and tissue/tumor samples (Figure 1). To examine whether this approach captures specificity we used microarray data from knockdown studies of CDK1 (A375 and MCF7),[30] CDK4 (A375 and MCF7)[30] and/or CDK8 (HT-29 colon cancer cells, HCT116 colon cancer cells and HeLaS3 cervical cancer cells).[40–42] The 97-gene CDK2 signature was generated using a cutoff *z*-score of $\leqslant -1.96$ in at least two cell lines, but this yielded too few genes for other CDKs, which risks skewing subsequent analysis of samples where one or a few genes were missing. Thus, to generate CDK1/4/8 signatures with >10 genes, we filtered genes with a *z*-score of $\leqslant -1.5$ in at least two cell lines. There was no overlap between any CDK signature genes, providing initial evidence for specificity (Figure 4a, bottom Venn Diagram; Supplementary Table 1 and 2). Although 76% of CDK2 signature genes formed a protein–protein interactions network consisting mainly of M-phase functional annotations (Figure 2), genes reduced by CDK1, CDK4 or CDK8 knockdown formed little-to-no protein–protein interaction networks (Supplementary Figure 4), and the affected genes were linked to various processes, with only a few directly connected to cell cycle regulation (Supplementary Table 3). Thus, the transcriptional responses to acute CDK depletion are distinct.

To directly examine CDK signature specificity, we assessed the response to pharmaceutical inhibitors with distinct target profiles. A recent study surveyed 178 kinase inhibitors at 0.5 μM against 300 kinases using a radioactive kinase assay.[43] Roscovitine (aka CYC202) potently inhibited CDK2-Cyclin E/A (76–80% inhibition), but not CDK1-Cyclin A/B1, or CDK4-Cyclin D1/D3 (3–18% inhibition; Supplementary Figures 5 and 6). In line with these data, we found that Roscovitine had a greater effect on the CDK2 versus CDK1 or CDK4 signatures using data from a study in LNCaP cells (GSE20433)[44] (Figure 4c). CDK8 activity was not assessed in the kinase/Roscovitine study,[43] but our data suggest this drug is not a potent CDK8 inhibitor ($P = 0.8254$, Figure 4c). Unlike Roscovitine, R547 inhibits many CDKs with high potency (Supplementary Figure 6). In agreement, R547 had dose-dependent effects on all CDK signatures in two cancer lines (Figure 4d). A limit with this approach is that there are no CDK inhibitors that are specific to CDK2 or, indeed, any single CDK assessed in our study. However, in light of the extensive genetic data above and cell cycle data below, the pharmaceutical analysis, whereas insufficient on its own, adds weight to the notion that the CDK2 signature tracks CDK2 activity.

CDK2 signature predicts cell cycle dynamics in tumors

To assess whether CDK2 activity predicts cell cycle activity in tumors we first compared CDK2 signature scores for NCI60 cell lines with matched population doubling times (dtp.nci.nih.gov/docs/misc/common_files/cell_list.html),[45] and found that the scores predict doubling times ($t_{CDK2LOW} = 38.4$ h, $n = 15$; $t_{CDK2HIGH} = 26.1$ h, $n = 9$; $P = 0.025$;) (Figure 5a). To address *in vivo* relevance, we studied clinical samples. Mitotic rate is a well-established indicator of mortality in patients with adrenocortical carcinoma.[46–49] When transcriptome and mitotic data was mined from adrenocortical tumors (GSE33371),[50] the CDK2 signature score predicted mitotic rate ($\rho = 0.6866$, $n = 106$; Figure 5b). We replicated this finding using TCGA adrenocortical carcinoma datasets,[51] and further observed that CDK2 activity correlates with atypical mitotic figures (Supplementary Figures 7A and B). Thus, the CDK2 signature predicts proliferative index in many contexts, supporting a broad role for CDK2 in the cancer cell cycle.

CDK2 activity in normal tissues

Next we applied the signature to define CDK2 activity in many human tissues and tumors. To our knowledge, this is the first systematic comparison of CDK2 activity across adult tissues.

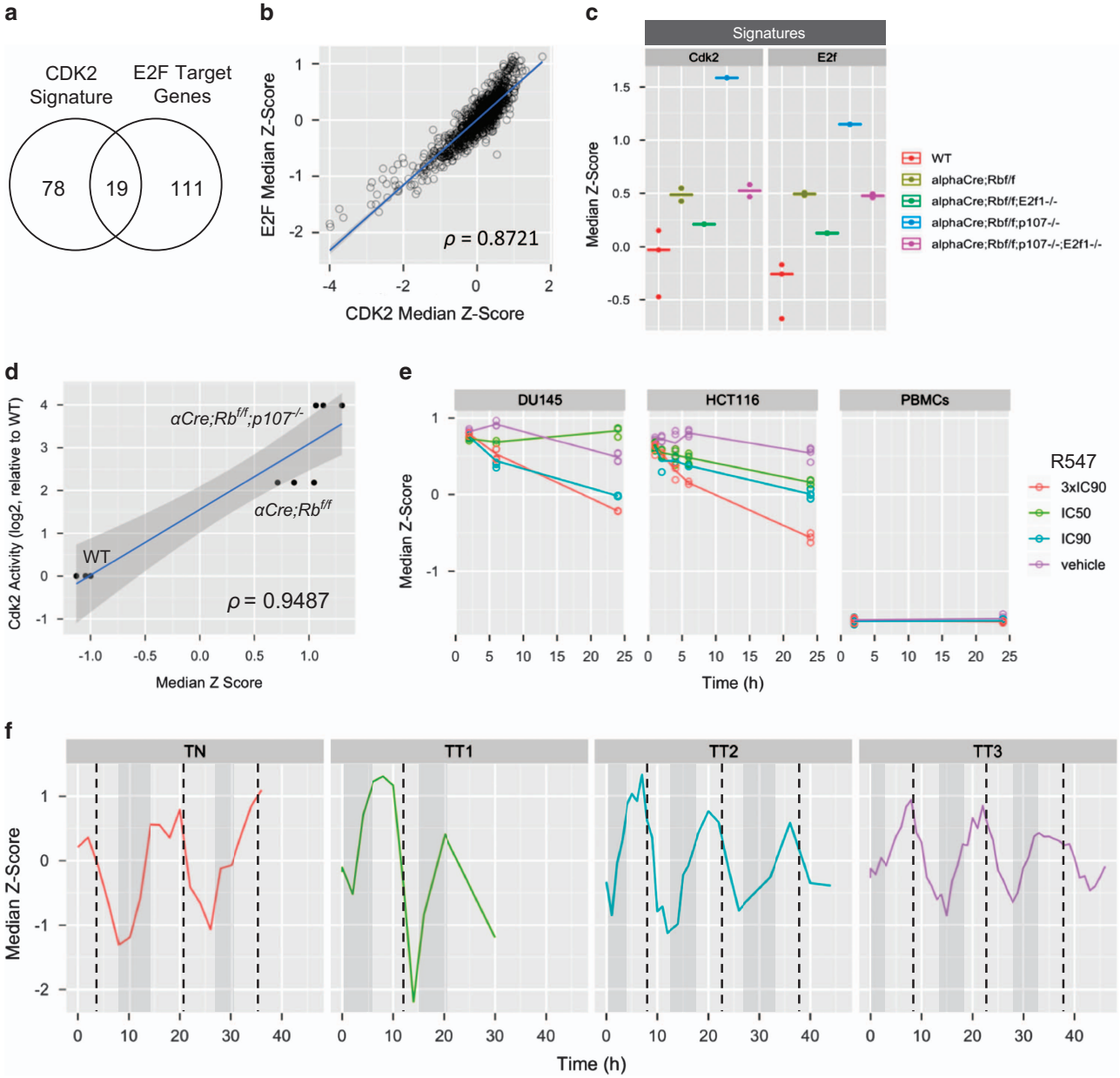


Figure 3. CDK2 signature is correlated to Cdk2 activity and is *E2f1*-dependent. (**a**) A Venn diagram showing the overlap between E2F target genes[35] and genes in the CDK2 signature. (**b**) A scatterplot of the CDK2 signature score versus E2F signature scores calculated from 1037 CCLE cancer cell lines. (**c**) RNA samples were collected from murine retina of the indicated genotypes. Subsequently, microarray was performed and the transcriptome data was used to calculate the CDK2 signature scores for each sample. (**d**) A scatterplot depicting the correlation between the CDK2 signature score ($n = 3$, median *z*-score) and CDK2 kinase activity in P8 retina ($n \geqslant 3$, but mean displayed only) with varying alterations in the RB family (three models). CDK2 activity was mined from.[28] Gray highlight around regression line indicates the 95% confidence intervals. Each dot represents a replicate from the microarray. (**e**) Transcriptome data was mined from GSE15396[36] in which cells were treated with either vehicle (DMSO) or different dosages of the pan-CDK inhibitor R547. RNA was collected at different timepoints and microarray was performed. The CDK2 signature scores were calculated across all samples and then organized by cell line. The R547 concentrations used are as follows: for HCT116 cells ($IC_{50} = 0.1$ μM, $IC_{90} = 0.2$ μM and $3 \times IC_{90} = 0.6$ μM); for DU145 cells ($IC_{50} = 0.1$ μM, $IC_{90} = 1.7$ μM and $3 \times IC_{90} = 5.1$ μM); for PBMCs ($IC_{90} = 0.2$ μM, $3xIC_{90} = 0.6$ μM). PBMCs = human peripheral blood mononuclear cells. (**f**) Transcriptome data was mined from.[37] HeLa cells were synchronized using either thymidine-nocodazole block (TN) or double thymidine block (TT). Poly(a) RNA was collected at intervals (every 1–2 h), reverse transcribed and hybridized to a custom microarray. The CDK2 signature scores were calculated across all conditions and separated by blocking method. The number after each TT indicates the clone number (for example, biological repeats). The dotted lines indicate mitosis and the gray bars indicate S-phase length.

Most adult tissues assessed in the Genotype-Expression Project data[52] had low CDK2 activity, consistent with their mainly post-mitotic state, and testes were the only normal tissue with a high signature (Figure 6). CDK1, CDK4 and CDK8 signatures correlated with CDK2 ($\rho_{CDK1} = 0.7120$, $\rho_{CDK4} = 0.8136$, $\rho_{CDK8} = 0.7680$), but in testes the CDK2 signature was at least one standard deviation higher than the others (Figure 6, green line). One of the few defects observed in *Cdk2*-deficient mice is testicular atrophy.[53,54] Our data now also implies a unique role for CDK2 in human testes.

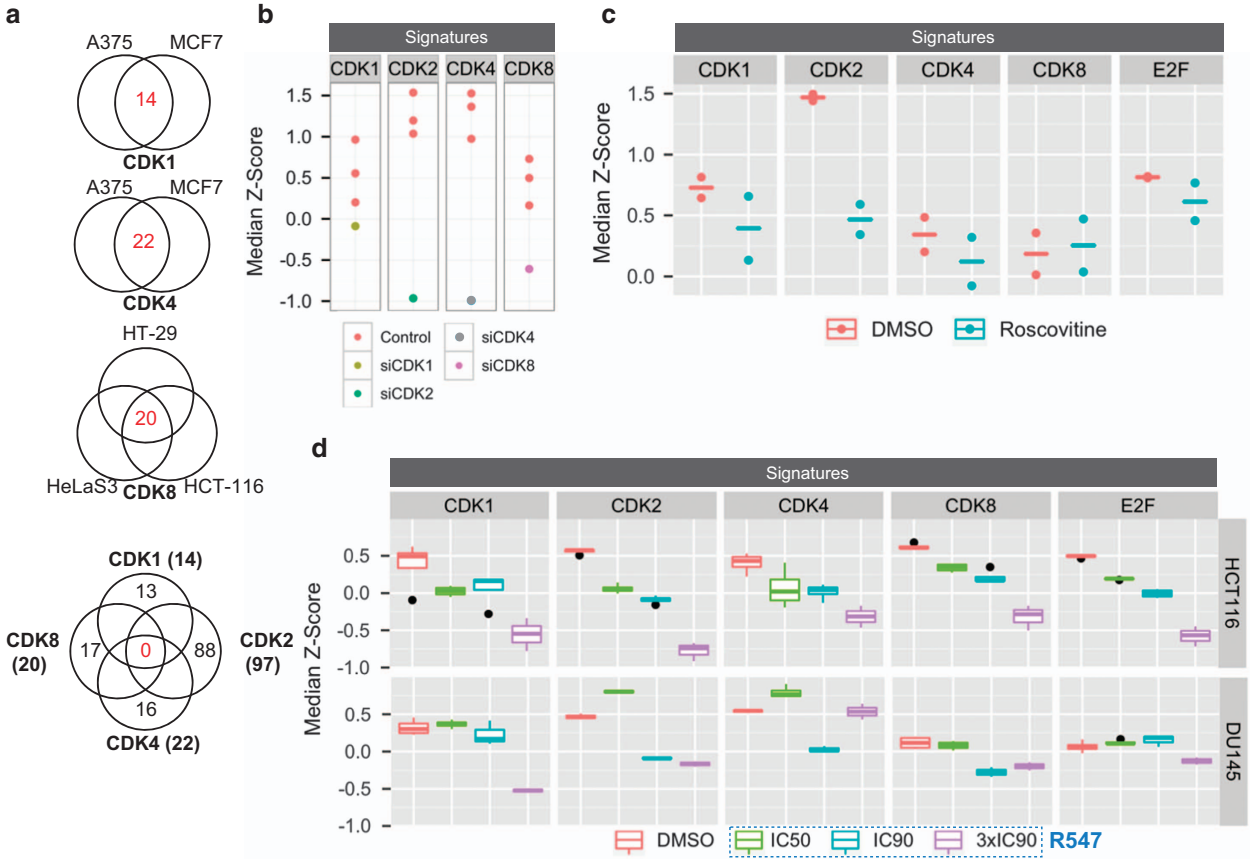


Figure 4. Specificity of different CDK signatures. (**a**) Microarray data was mined from multiple CDK knockdown studies to develop signatures for CDK1, CDK4 and CDK8. Each dataset was normalized and the log2 median difference was calculated between each siCDK condition and control. These differences were converted to *z*-scores. Only genes with a *z*-score >1.5 were included in the Venn Diagram. The brackets indicate the signature gene set size. (**b**) Transcriptome data from **a** was re-mined and processed using other CDK signatures. Each respective siRNA for each signature is shown along with controls. (**c**) Transcriptome data was mined from GSE20433[44] in which LNCaP cells were treated with either DMSO (0.1%) or CDK inhibitor roscovitine (10 μM) for 24 h after which the RNA was collected and sent for microarray. The signature scores (*y*-axis) for CDK1, CDK2, CDK4 and CDK8 were calculated for all samples in the study. The results are summarized as a dot plot with the crossbar indicating the median. Each dot represents a biological replicate ($n=2$). (**d**) Transcriptome data was mined from GSE15396[36] in which cells were treated with either vehicle (DMSO) or different dosages of the pan-CDK inhibitor R547. RNA was collected after 24 h of treatment and microarray was performed. The CDK signature scores were calculated across all samples and then organized by cell line ($n \geqslant 4$).

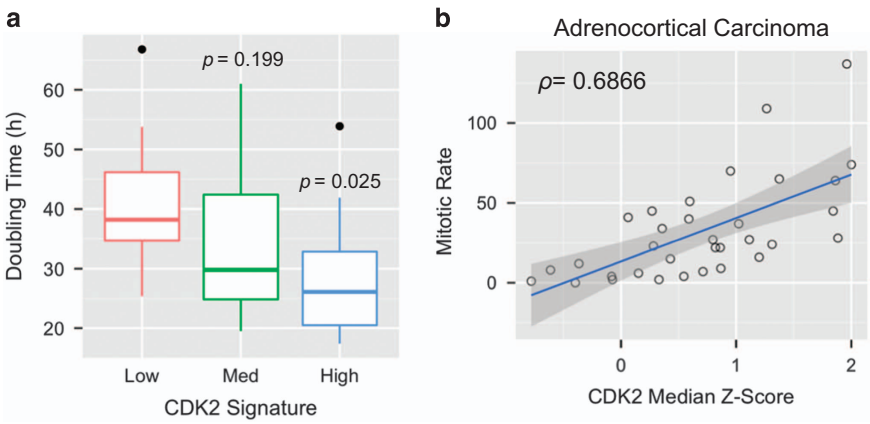


Figure 5. CDK2 signature predicts cell cycle dynamics. (**a**) Doubling times were curated from NCI60 cell lines along with their microarray data from GSE2003.[45] The CDK2 signature scores were calculated for each cell line and divided into three bins based on percentile. *P*-values indicate the pairwise Mann–Whitney Test with a false discovery rate correction for multiple testing. The data is summarized as boxplots. (**b**) Microarray data was mined from GSE33371[50] along with the mitotic rate. The CDK2 signature scores were calculated for all adrenocortical cancer samples and correlated to mitotic rate using Spearman's ρ ($n=33$).

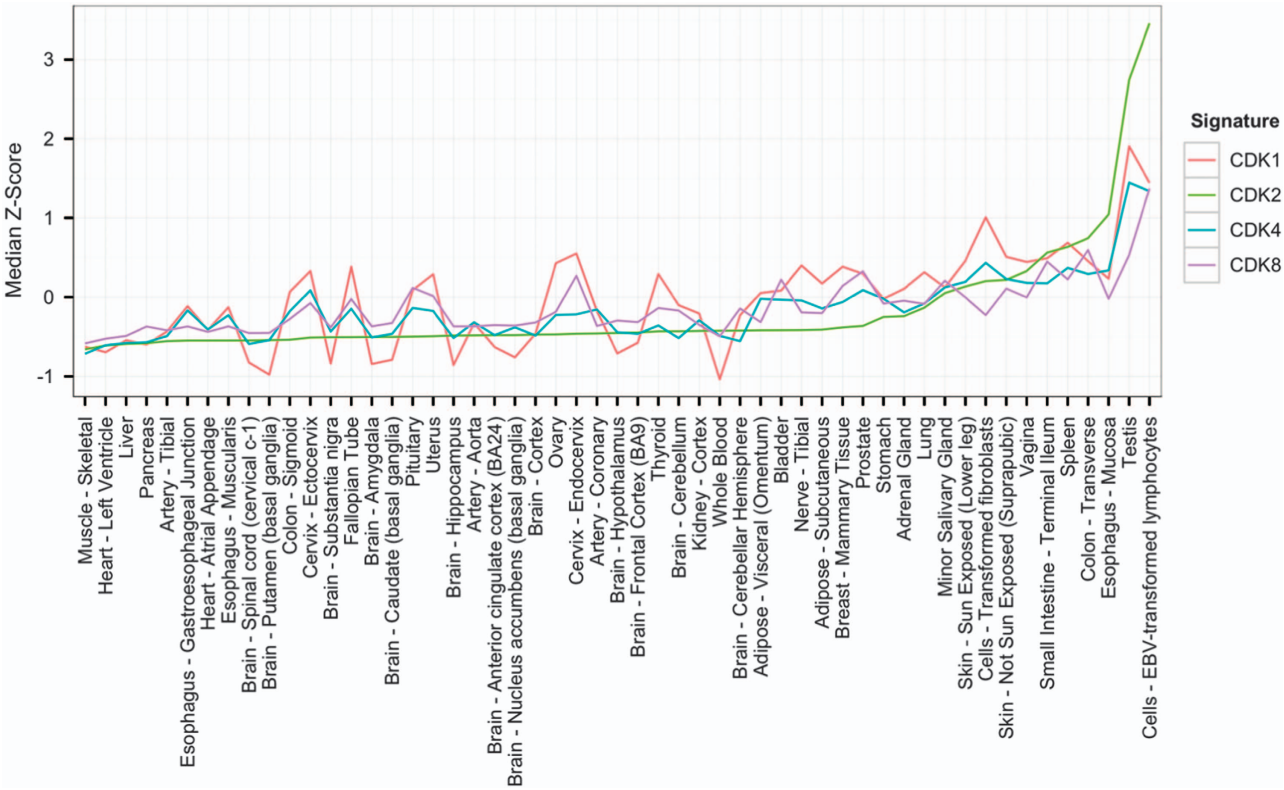


Figure 6. Distribution of CDK2 signature across multiple cell lines and clinical samples. Transcriptome data of 2921 normal human samples were mined from Genotype-Tissue Expression Project (GTEx). All data was log2 transformed and normalized between samples. Subsequently, the CDK signature scores were calculated for each sample. Since each tissue consists of multiple samples (ranging from 3 to 191), only the median of each signature is depicted for a given tissue.

CDK2 activity predicts outcome in a subset of cancers

To compare CDK2 activity globally across cancer we curated RNAseq data from >7000 clinical tumor samples (TCGA) and microarray data from >1000 cell lines (CCLE) using cutoffs at the 20th and 80th percentiles (Supplementary Figure 8). Consistent with selection for the most robustly dividing cells *in vitro* there were fairly minor differences in the distribution of CDK2-low/medium/high cases across cancer cell lines. In stark contrast, many primary tumors exhibited only medium or high but not low CDK2 activity (for example, cervical, colon), or only low or medium but not high CDK2 activity (for example, prostate, thyroid (Figures 7b and c; Supplementary Table 4). These data expose markedly distinct CDK2 activity across different cancers.

Next, we employed TCGA data on recurrence-free survival (RFS) or overall survival (OS) to assess whether CDK2 activity influences risk. As opposed to the global CDK2 signature scores that were used to compare across cancers, we recalculated tumor-specific scores to derive hazard ratios (HRs) for each tumor type. High versus low activity was compared with cutoffs at the 66th and 33rd percentile, respectively. Supplementary Figure 9 shows KM curves for each tumor, significant HRs are summarized to the right of the graph in Figure 7b, the heat map in Figure 7c indicates median and gene-specific HRs, and Supplementary Table 5 provides numerical details.

High tumor-specific CDK2 activity predicted improved RFS in patients with colon cancer, which agrees with prior work.[55] However, among the 26 other cancers, high tumor-specific CDK2 activity predicted poorer OS and/or RFS in 11 cancer types. In general, individual signature genes generated HRs similar to the median of the entire signature (Figure 7c), consistent with their highly correlated expression (Figure 1b). Cases where CDK2 activity predicted OS but not RFS or vice versa reflects limited samples in some cases.

Figure 7b places side by side global CDK2 activity (bar graph) and the risk analysis (to the right of the bar graph). Intriguingly, the CDK2 signature predicted poor outcome in 11/13 cancers that had CDK2-low cases, but in 0/14 tumors with no or very few such cases. These data imply that low CDK2 activity is protective, but medium or high activity confers similar increased risk. Other pieces of data also support the notion of a two-state model for the link between CDK2 and cancer progression (see 'Discussion' section).

CDK2 activity outperforms other predictors

Next, we compared the importance of CDK2 activity relative to other predictive parameters. In adrenocortical cancer, the CDK2 signature generated greater HRs than a few other selected clinical features (Figure 8a), suggesting it might be a better predictive tool. As a multivariate strategy to compare predictors, we implemented random forests. This unbiased approach is an ensemble machine learning algorithm, which utilizes predictive variables to generate hundreds of randomized decision trees and averages the results to create a model of lower variance.[56] Random forests also estimate variable importance by determining how often a feature appears at the top of decision trees across the forest. Of multiple clinical characteristics recorded by TCGA for four cancers with low CDK2 cases, the CDK2 signature was the feature with the highest (adrenocortical carcinoma OS, thyroid cancer RFS) or second highest importance (kidney renal papillary cell carcinoma OS, low grade glioma OS/RFS; Figure 8b; Supplementary Table 6). As a control we analyzed melanoma, in which the CDK signature is not predictive (HR = 1.1), and indeed

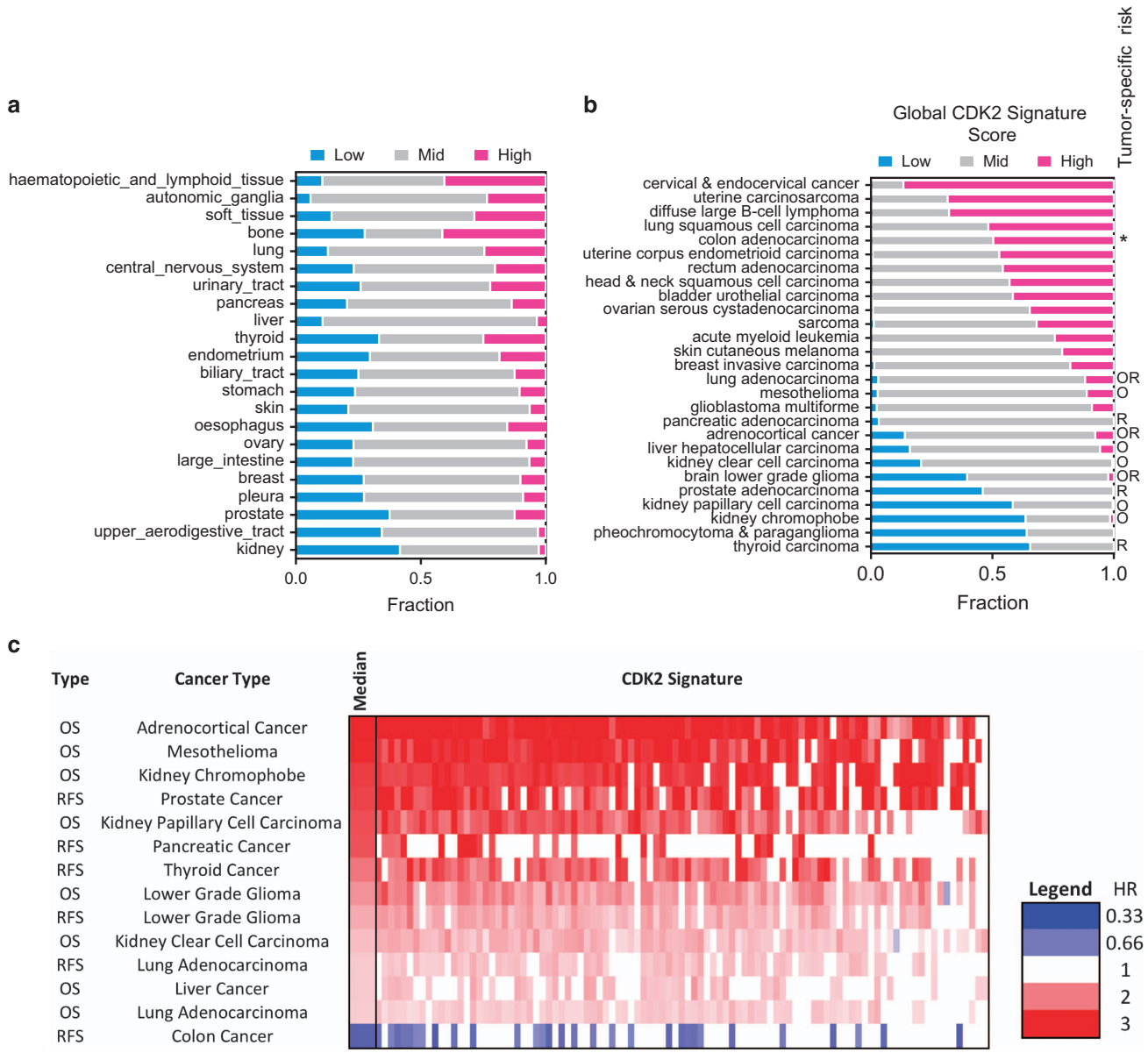
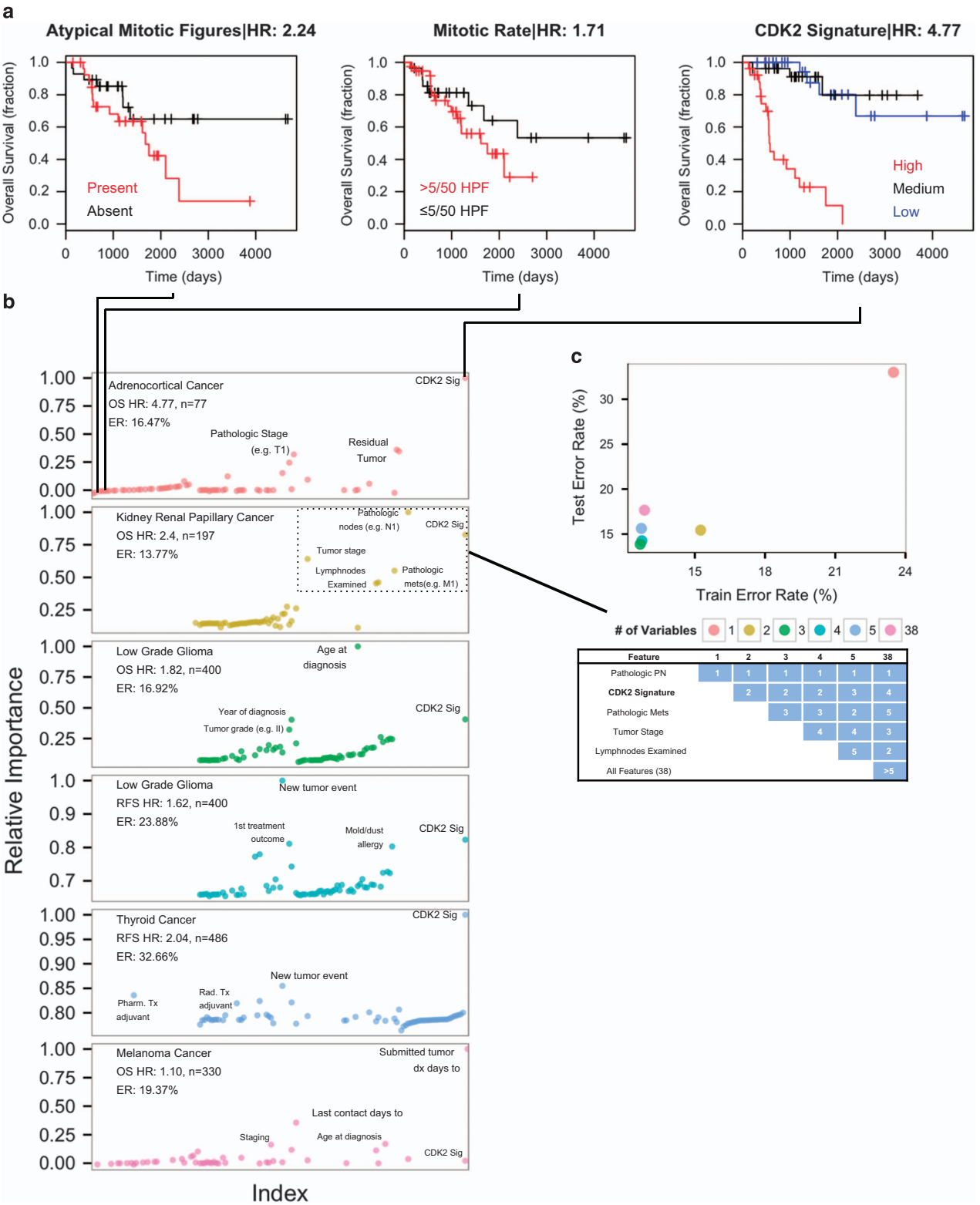


Figure 7. High CDK2 signature predicts poor prognosis in patients in CDK2-Low tumors. (**a**) A total of >1000 cancer cell lines (mined from the CCLE) arranged by their CDK2 signature score. Dotted lines indicate the 20th and 80th percentile. A stacked bar plot is used to depict the distribution of cell lines in with a high, medium or low CDK2 signature score by tissue type. (**b**) Same as **a** except using patient tumor samples mined from the TCGA ($n=7746$). Symbols to the right are based on KM curve analysis on patient outcomes (Supplementary Figure 9; Supplementary Table 5): ^{O} = high tumor-specific CDK2 score significantly increases risk in overall survival. ^{R} = high tumor-specific CDK2 score significantly increases recurrence. * = high tumor-specific CDK2 score significantly decreases risk of recurrence. Global refers to the CDK2 activity calculated relative to the entire TCGA dataset, whereas tumor-specific refers to CDK2 activity calculated using only that tumor type. (**c**) Within each cancer, the normalized *z*-scores for each gene within the CDK2 signature was divided into three bins based on percentile. Using a Cox proportional hazards regression, the hazard ratios were calculated between the high and low groups within each cancer type. The median hazard ratio is also displayed for reference. The top 14 cancer types are shown with colored blocks indicating a significant difference between high and low expressing groups within that cancer type ($P \leqslant 0.05$).

Figure 8. CDK2 activity is an important predictor of patient outcomes compared with other clinical features in CDK2-Low tumors. (**a**) Transcriptome data was mined from adrenocortical carcinoma clinical samples (TCGA) along with clinical features. The CDK2 signature scores were calculated. Using the clinical information, the Kaplan–Meier curves were graphed stratifying by atypical mitotic figures (presence/absence, left), mitotic rates (assessed by the presence/absence of at least five mitotic figures in 50 high powered fields (HPF, middle), or CDK2 signature scores (divided into thirds based on percentile, right). (**b**) Random forests was run on all patients with clinical information to determine the importance of each feature towards its ability to predict outcomes in tumors where CDK2 activity predicted outcome: CDK2 activity is not prognostic in melanoma and was used as a control. ER: error rate. (**c**) To determine the minimum number of features, which could predict survival outcomes in kidney renal papillary cell cancer data was split into a training set (60%) and test set (40%). Random forests was performed on the training group to generate a model of given error and used to predict in the test set. This process was repeated by sequentially eliminating least important features and graphing the response in out of bag error rate. The table below indicates the rank in importance and the selection of each feature per iteration.

random forests analysis ranked it 16th of 47 features (Figure 8b; Supplementary Table 6). Surprisingly the CDK2 signature had a much higher feature importance than indicators like mitotic index, staging and resection status (Figure 8b; Supplementary Table 6).

Next, we utilized the kidney renal papillary cell cancer data to create training and test sets and performed random forests on the training group to generate a model of given error to predict in the test set. The process was repeated and less important features sequentially eliminated. Notably, 3 features, which included the CDK2 signature, predicted outcome at least as reliably as all 38 features, and there was a large increase in error when pathologic PN (stages based on the cancer nodes) was used alone versus it



plus the CDK signature (Figure 8c). These results support an important functional role for rising CDK2 activity in CDK2-low tumors, and expose a novel, enhanced strategy to predict outcome in these cancers.

DISCUSSION

A new tool to track CDK2 and predict cancer outcome

We report a 97-gene transcriptional signature that accurately proxies CDK2 activity, providing a simple method to assess this kinase in different settings and samples. It was derived by coupling transcriptome data from CDK2 depletion studies coupled with coexpression of hits across thousands of tumor samples (Figure 1). The signature includes many functional nodes required for M phase (Figure 2), which suggests that CDK2 prepares the cell for mitosis. This potentially key role of CDK2 in setting up mitosis has been largely overlooked, likely because CDK1 may compensate for this role in CDK2-null animals. Multiple pieces of data validate the signature. It responds in a dose-dependent fashion to RB/E2F disruption or small molecule CDK inhibitors (Figure 3), and shows specificity relative to other CDK signatures (Figure 4). It correlates with cell cycle length, mitotic rate and rapidly tracks changes in CDK2 activity during cell cycle progression (Figures 3 and 5). Indeed, the signature showed the same undulating pattern as that observed with a fusion protein that tracks CDK2 activity in live cells, rising through S-phase and plummeting at the M/G1 boundary.[15] This rapid transcriptional response to CDK2 is in line with data showing that CDK2 phosphorylates RNAPII CTD[7,8] and that 44/180 CDK2-Cyclin A proteomic targets regulate mRNA transcription and processing.[9] The correlation between CDK2 activity and mitotic rate in tumors implies an important role for this enzyme in cancer proliferation. We exposed unusually high CDK2 activity in human testes, matching CDK2 function in mice[53,54] and markedly distinct CDK2 activity in different cancers (Figures 6 and 7b). Intriguingly, high CDK2 activity was associated with improved outcome in colon cancer, but worse outcome in other cancers, all of which contain several CDK2-low tumors (Figure 7b). Machine learning analysis revealed that the signature is more important than numerous other clinical parameters in identifying at-risk patients (Figure 8). Thus, rising CDK2 activity may be oncogenic in CDK2-low tumors. The signature could be applied to better predict when these cancers require aggressive treatment.

CDK2 in cancer: 'All or nothing'?

If the relationship between CDK2 activity and tumorigenesis was linear or exponential, samples with progressively higher CDK2 activity should have increased tumor progression and worse prognosis. Instead, the CDK2 signature could decipher risk in tumors with a mix of low and medium activity, but not in those with medium and high activity (Figure 7b). Thus, above a certain threshold of CDK2 activity, there may be no further advantage to the cancer cell, and indeed a disadvantage in colon cancer. Some cancers may transition through the lower CDK2 state at early, possibly benign periods of growth. In support of this 'two-state' hypothesis, other studies find that only complete ablation of *Cdk2* blocks tumorigenesis, whereas heterozygous *Cdk2* is insufficient.[16,17,22,23] Moreover, a quantum increase in CDK2 activity correlates with the transition from a tumor-resistant to tumor-prone state in *Rb* versus *Rb/p107* or *Rb/p27* null retina, respectively.[28]

Technical advances towards understanding the kinome

To validate our signature, we focused principally on cell cycle contexts because this is by far that most studied aspect of CDK2 function. Having extensively validated the signature, it could now be deployed to track CDK2 activity in other contexts, such as during DNA damage, degeneration, embryonic development and/or drug treatment for cancer or other diseases. The process employed here to study CDK2 could be leveraged to generate other signatures, thus applying publically available transcriptome data to assess the entire kinome. CDKs may be a special case because many of the family members have direct roles in transcription. Nevertheless, efforts to define and validate kinase signatures may hold promise to uncover their relevance in cancer development, and to generate new more powerful predictive tools to guide treatment.

MATERIALS AND METHODS

CDK2 signature data mining, refinement and visualization

To develop a CDK2 signature, microarray data was downloaded from three studies curated on Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/): GSE31534 (siRNA), GSE16480 (shRNA) and GSE31912 (siRNA). Each of these studies used the Affymetrix Human Genome U133 Plus 2.0 Array platform (Affymetrix Inc., Santa Clara, CA, USA), minimizing the cross-platform variability. Each dataset was background corrected using Robust Multi-Array Average (RMA) algorithm,[57] then quantile normalized to batch correct (if necessary). Probes targeting the same gene were aggregated using median. The difference between knockdown CDK2 samples and Control (for example, siScrambled, baseline) was determined and then centered/scaled using *z*-scores. The initial signature only included genes that were significantly up-regulated or down-regulated in at least two of the three studies ($|z| \geqslant 1.96$). To test whether these genes are co-expressed, mRNA expression data was downloaded from TCGA using Cancer Browser[33,58] and correlated to each gene in the signature using Spearman's ρ. Only co-expressed gene were selected for the CDK2 signature based on a highly correlated cluster determined by hierarchical clustering. The CDK2 signature genes are summarized in Supplementary Table 1. Level 3 RNAseq data with file names using the '*.rsem.genes.normalized_results' tag were downloaded from The Cancer Genome Atlas (TCGA) via Cancer Browser (genome-cancer.ucsc.edu/proj/site/hgHeatmap/). The data was then log2(x+1) transformed and quantile normalized across 8411 samples. Only co-expressed genes were inputted to the GeneMania App (Version: 3.3.4, 2014-08-12 core; http://genemania.org/) in Cytoscape (3.2.0, San Diego, CA, USA) for PPI network analysis.

Signature score

Signature scores are used to approximate the activity or level of a signature in a given sample. The score is created by first downloading gene expression data and ensuring samples have appropriate background/transformation for subsequent processing. For most studies mined, RMA background correction was used with log2 transformation. In cases where RMA normalized data was unavailable, quantile normalization was used. After preprocessing, the gene expression scores were filtered based on an input gene signature and converted to human homologs if necessary using Homologene (NCBI, http://www.ncbi.nlm.nih.gov/homologene). After scaling and centering the filtered list with *z*-scores, the gene expression is summarized for a given sample using the median. For global comparisons of CDK2 scores, patient samples or cell lines in the top fifth or bottom fifth quantile were labeled as global signature high or signature low, respectively (Figures 7a and b). However, for tumor-specific comparisons, the top third was considered high and bottom third was considered low. These cutoffs permit sufficient numbers for statistical testing in the majority of cases. The methodology is graphically summarized in Supplementary Figure 2.

Copy number/mutation enrichment analysis

To determine which mutations are enriched in high CDK2 signature cell lines, 1037 cell lines from the CCLE[32] were divided into thirds based on the percentile of their CDK2 signature score. Next, the copy number and mutation status of 462 oncogenes and tumor suppressors was downloaded from cBioPortal (www.cbioportal.org/public-portal/) and matched to the signature high and signature low groups. Alterations were summarized as MUT (exonic mutation), AMP (amplification) or HOMDEL (homozygous deletion). The mutations analyzed were present in at least 20 cell lines in both the high signature and low signature groups with at least a twofold change. A Fisher's exact test was used to determine the significance of pairwise comparisons between the groups. A 20% false

discovery rate was set to correct for multiple testing (for example, fail to reject null hypothesis when $P \leqslant 0.00222$). Lollipop plots were generated using MutationMapper (v1.0) via cBioPortal.[59]

Survival analysis in clinical samples

Clinical data and RNAseq V2 data was mined using TCGA's pan-cancer normalized data via Cancer Browser. The data was matched by TCGA barcodes. The RNAseq data was quantile normalized first then the CDK2 signature score (median *z*-score) was calculated for each patient. The survival data between the signature low and signature high was used to calculate the hazard ratios for the OS and RFS across multiple tumors types ($n = 27$) using a Cox proportional hazards regression model. The same analysis was also performed at an individual gene level.

Mouse microarray

The RNA samples were collected using RNeasy Mini kit (Qiagen, Venlo, Netherlands) from E15.5 and P8 murine retina from mixed C57BL/6 × 129SvJ mice with deletions of *Rb*, *p107* and/or *E2f1*. E15.5 samples were isolated using laser capture microdissection. RNA quality was determined using Agilent 2100 Bioanalyzer (TCAG, Toronto, ON, Canada). Only samples with a RNA Integrity score (RIN) of >7 were used. Reverse transcription and hybridization was performed by TCAG. P8 samples were hybridized to Affymetrix Mouse Gene ST 2.0 array probes whereas E15.5 samples were hybridized to Illumina Mouse WG6 (v1.1, San Diego, CA, USA). Probe intensity scores were processed using RMA background correction and log2 transformation. Biological sample numbers were as follows: WT:3; *Rb* null: 2, *Rb/p107* null: 1; *Rb/E2f1* null: 1; *Rb/p107/E2f1* triple null: 2.

Software and statistical analysis

Data was arranged and organized using R statistical programming language and Microsoft Excel. All statistical tests were two-sided. To determine the statistical significance, R was used graph and test the significance of data from Kaplan–Meier curves, RT–PCR analysis, mined studies and *in vitro* assays. Multiple testing was corrected for using either Bonferroni's correction or false discovery rate. Heatmaps were generated using Microsoft Excel. R package 'survival' was used to estimate hazard ratios and *P*-values of survival data. Most graphs were created using 'ggplot2' graphics package in R. Random forests was performed using the 'randomForestSRC' package. Code is available at https://github.com/seanmccurdy/medical_research.

Significance

CDK2 targets proteins that affect cell cycle, transcription and cancer. Measuring its activity across multiple samples could reveal new roles, yet current methods are expensive, cumbersome and/or are not applicable to large sample sizes. Leveraging transcriptome data we show that a 97-gene set accurately and rapidly tracks CDK2 activity in multiple settings. We reveal new insight into CDK2 activity in human tissues and thousands of tumors. The signature exposes remarkably high CDK2 activity in human testes, and a unique set of 'CDK2-low' tumors where switching to higher activity predicts poor outcome. Our work provides an enhanced strategy to analyze CDK2 and predict cancer risk, and the method could be applied to study other enzymes that rapidly affect transcription.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank G Bader, F Roth and J Wrana for helpful comments. This project was funded by grants to RB from the Canadian Institutes for Health Research (CIHR), Foundation Fighting Blindness Canada, Krembil Foundation, Ontario Institute for Cancer Research through funding provided by the Government of Ontario and the Terry Fox Research Institute. SRM was supported by scholarships from Peterborough KM Hunter Foundation, CIHR (Vanier Graduate Scholarship) and the CIHR Biological Therapeutics Program.

AUTHOR CONTRIBUTIONS

SRM and RB designed the experiments and drafted the manuscript. MP and MA gathered samples and performed microarray on retinal samples. The data analysis was performed by SRM. SRM and RB discussed and revised the manuscript. All authors read and approved the final manuscript.

REFERENCES

1 Lim S, Kaldis P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Dev Camb Engl* 2013; **140**: 3079–3093.
2 Malumbres M, Harlow E, Hunt T, Hunter T, Lahti JM, Manning G *et al.* Cyclin-dependent kinases: a family portrait. *Nat Cell Biol* 2009; **11**: 1275–1276.
3 Burkhart DL, Sage J. Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer* 2008; **8**: 671–682.
4 Palancade B, Bensaude O. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem* 2003; **270**: 3859–3870.
5 Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I *et al.* CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 2008; **455**: 547–551.
6 Morris EJ, Ji J-Y, Yang F, Di Stefano L, Herr A, Moon N-S *et al.* E2F1 represses β-catenin transcription and is antagonized by both pRB and CDK8. *Nature* 2008; **455**: 552–556.
7 Deng L, Ammosova T, Pumfery A, Kashanchi F, Nekhai S. HIV-1 Tat Interaction with RNA polymerase II C-terminal domain (CTD) and a dynamic association with CDK2 induce CTD phosphorylation and transcription from HIV-1 promoter. *J Biol Chem* 2002; **277**: 33922–33929.
8 Nekhai S, Zhou M, Fernandez A, Lane WS, Lamb NJC, Brady J *et al.* HIV-1 Tat-associated RNA polymerase C-terminal domain kinase, CDK2, phosphorylates CDK7 and stimulates Tat-mediated transcription. *Biochem J* 2002; **364**: 649.
9 Chi Y, Welcker M, Hizli AA, Posakony JJ, Aebersold R, Clurman BE. Identification of CDK2 substrates in human cell lysates. *Genome Biol* 2008; **9**: R149.
10 Asghar U, Witkiewicz AK, Turner NC, Knudsen ES. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat Rev Drug Discov* 2015; **14**: 130–146.
11 Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* 2009; **9**: 153–166.
12 Welcker M, Singer J, Loeb KR, Grim J, Bloecher A, Gurien-West M *et al.* Multisite phosphorylation by Cdk2 and GSK3 controls cyclin E degradation. *Mol Cell* 2003; **12**: 381–392.
13 Zhang W, Koepp DM. Fbw7 isoform interaction contributes to cyclin E proteolysis. *Mol Cancer Res MCR* 2006; **4**: 935–943.
14 Ishihara H, Yoshida T, Kawasaki Y, Kobayashi H, Yamasaki M, Nakayama S *et al.* A new cancer diagnostic system based on a CDK profiling technology. *Biochim Biophys Acta Mol Basis Dis* 2005; **1741**: 226–233.
15 Spencer SL, Cappell SD, Tsai F-C, Overton KW, Wang CL, Meyer T. The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell* 2013; **155**: 369–383.
16 Akli S, Pelt CSV, Bui T, Meijer L, Keyomarsi K. Cdk2 is required for breast cancer mediated by the low-molecular-weight isoform of cyclin E. *Cancer Res* 2011; **71**: 3377–3386.
17 Campaner S, Doni M, Hydbring P, Verrecchia A, Bianchi L, Sardella D *et al.* Cdk2 suppresses cellular senescence induced by the c-myc oncogene. *Nat Cell Biol* 2010; **12**: 54–59.
18 Gillam MP, Nimbalkar D, Sun L, Christov K, Ray D, Kaldis P *et al.* MEN1 tumorigenesis in the pituitary and pancreatic islet requires Cdk4 but not Cdk2. *Oncogene* 2014. ;**124**: 1056–1061.
19 Macias E, Kim Y, Marval PLM de, Klein-Szanto A, Rodriguez-Puebla ML. Cdk2 deficiency decreases ras/CDK4-dependent malignant progression, but not myc-induced tumorigenesis. *Cancer Res* 2007; **67**: 9713–9720.
20 Martín A, Odajima J, Hunt SL, Dubus P, Ortega S, Malumbres M *et al.* Cdk2 is dispensable for cell cycle inhibition and tumor suppression mediated by p27Kip1 and p21Cip1. *Cancer Cell* 2005; **7**: 591–598.
21 Padmakumar VC, Aleem E, Berthet C, Hilton MB, Kaldis P. Cdk2 and Cdk4 activities are dispensable for tumorigenesis caused by the loss of p53. *Mol Cell Biol* 2009; **29**: 2582–2593.
22 Puyol M, Martín A, Dubus P, Mulero F, Pizcueta P, Khan G *et al.* A synthetic lethal interaction between K-Ras oncogenes and Cdk4 unveils a therapeutic strategy for non-small cell lung carcinoma. *Cancer Cell* 2010; **18**: 63–73.
23 Ray D, Terao Y, Christov K, Kaldis P, Kiyokawa H. Cdk2-null mice are resistant to ErbB-2-induced mammary tumorigenesis. *Neoplasia* 2011; **13**: 439–444.
24 Hongo F, Takaha N, Oishi M, Ueda T, Nakamura T, Naitoh Y *et al.* CDK1 and CDK2 activity is a strong predictor of renal cell carcinoma recurrence. *Urol Oncol* 2014; **32**: 1240–1246.

25 Kim SJ, Nakayama S, Miyoshi Y, Taguchi T, Tamaki Y, Matsushima T *et al.* Determination of the specific activity of CDK1 and CDK2 as a novel prognostic indicator for early breast cancer. *Ann Oncol* 2008; **19**: 68–72.
26 Kim SJ, Nakayama S, Shimazu K, Tamaki Y, Akazawa K, Tsukamoto F *et al.* Recurrence risk score based on the specific activity of CDK1 and CDK2 predicts response to neoadjuvant paclitaxel followed by 5-fluorouracil, epirubicin and cyclophosphamide in breast cancers. *Ann Oncol Off J Eur Soc Med Oncol* 2012; **23**: 891–897.
27 Kubo H, Suzuki T, Matsushima T, Ishihara H, Uchino K, Suzuki S *et al.* Cyclin-dependent kinase-specific activity predicts the prognosis of stage I and stage II non-small cell lung cancer. *BMC Cancer* 2014; **14**: 755.
28 Sangwan M, McCurdy SR, Livne-bar I, Ahmad M, Wrana JL, Chen D *et al.* Established and new mouse models reveal E2f1 and Cdk2 dependency of retinoblastoma, and expose effective strategies to block tumor initiation. *Oncogene* 2012; **31**: 5019–5028.
29 Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* 2008; **132**: 487–498.
30 Wang L, Hurley DG, Watkins W, Araki H, Tamada Y, Muthukaruppan A *et al.* Cell cycle gene networks are associated with melanoma prognosis. *PLoS One* 2012; **7**: e34247.
31 Molenaar JJ, Ebus ME, Geerts D, Koster J, Lamers F, Valentijn LJ *et al.* Inactivation of CDK2 is synthetically lethal to MYCN over-expressing cancer cells. *Proc Natl Acad Sci* 2009; **106**: 12968–12973.
32 Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012; **483**: 570–575.
33 Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D *et al.* Exploring TCGA Pan-Cancer Data at the UCSC Cancer Genomics Browser. *Sci Rep* 2013; **3**: 2652.
34 Kronja I, Orr-Weaver TL. Translational regulation of the cell cycle: when, where, how and why? *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 3638–3652.
35 Bracken AP, Ciro M, Cocito A, Helin K. E2F target genes: unraveling the biology. *Trends Biochem Sci* 2004; **29**: 409–417.
36 Berkofsky-Fessler W, Nguyen TQ, Delmar P, Molnos J, Kanwal C, DePinto W *et al.* Preclinical biomarkers for a cyclin-dependent kinase inhibitor translate to candidate pharmacodynamic biomarkers in phase I patients. *Mol Cancer Ther* 2009; **8**: 2517–2525.
37 Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 2002; **13**: 1977–2000.
38 Aleem E, Kiyokawa H, Kaldis P. Cdc2–cyclin E complexes regulate the G1/S phase transition. *Nat Cell Biol* 2005; **7**: 831–836.
39 Satyanarayana A, Kaldis P. Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene* 2009; **28**: 2925–2939.
40 Adler AS, McCleland ML, Truong T, Lau S, Modrusan Z, Soukup TM *et al.* CDK8 maintains tumor dedifferentiation and embryonic stem cell pluripotency. *Cancer Res* 2012; **72**: 2129–2139.
41 Galbraith MD, Allen MA, Bensard CL, Wang X, Schwinn MK, Qin B *et al.* HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* 2013; **153**: 1327–1339.
42 Tsutsui T, Fukasawa R, Tanaka A, Hirose Y, Okhuma Y. Identification of target genes for the CDK subunits of the Mediator complex. *Genes Cells Devoted Mol Cell Mech* 2011; **16**: 1208–1218.
43 Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* 2011; **29**: 1039–1045.
44 Chen S, Bohrer LR, Rai AN, Pan Y, Gan L, Zhou X *et al.* Cyclin-dependent kinases regulate epigenetic gene silencing through phosphorylation of EZH2. *Nat Cell Biol* 2010; **12**: 1108–1114.
45 Lee JK, Havaleshko DM, Cho H, Weinstein JN, Kaldjian EP, Karpovich J *et al.* A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. *Proc Natl Acad Sci USA* 2007; **104**: 13086–13091.
46 Stojadinovic A, Ghossein RA, Hoos A, Nissan A, Marshall D, Dudas M *et al.* Adrenocortical carcinoma: clinical, morphologic, and molecular characterization. *J Clin Oncol Off J Am Soc Clin Oncol* 2002; **20**: 941–950.
47 Stojadinovic A, Brennan MF, Hoos A, Omeroglu A, Leung DHY, Dudas ME *et al.* Adrenocortical adenoma and carcinoma: histopathological and molecular comparative analysis. *Mod Pathol* 2003; **16**: 742–751.
48 Weiss LM. Comparative histologic study of 43 metastasizing and nonmetastasizing adrenocortical tumors. *Am J Surg Pathol* 1984; **8**: 163–169.
49 Weiss LM, Medeiros LJ, Vickery AL. Pathologic features of prognostic significance in adrenocortical carcinoma. *Am J Surg Pathol* 1989; **13**: 202–206.
50 Heaton JH, Wood MA, Kim AC, Lima LO, Barlaskar FM, Almeida MQ *et al.* Progression to adrenocortical tumorigenesis in mice and humans through insulin-like growth factor 2 and β-catenin. *Am J Pathol* 2012; **181**: 1017–1033.
51 Assié G, Letouzé E, Fassnacht M, Jouinot A, Luscap W, Barreau O *et al.* Integrated genomic characterization of adrenocortical carcinoma. *Nat Genet* 2014; **46**: 607–612.
52 GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013; **45**: 580–585.
53 Berthet C, Aleem E, Coppola V, Tessarollo L, Kaldis P. Cdk2 knockout mice are viable. *Curr Biol* 2003; **13**: 1775–1785.
54 Ortega S, Prieto I, Odajima J, Martín A, Dubus P, Sotillo R *et al.* Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. *Nat Genet* 2003; **35**: 25–31.
55 Li J-Q, Miki H, Ohmori M, Wu F, Funamoto Y. Expression of cyclin E and cyclin-dependent kinase 2 correlates with metastasis and prognosis in colorectal carcinoma. *Hum Pathol* 2001; **32**: 945–953.
56 Breiman L. Random forests. *Mach Learn* 2001; **45**: 5–32.
57 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostat Oxf Engl* 2003; **4**: 249–264.
58 Zhu J, Sanborn JZ, Benz S, Szeto C, Hsu F, Kuhn RM *et al.* The UCSC Cancer Genomics Browser. *Nat Methods* 2009; **6**: 239–240.
59 Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; **6**: pl1–pl1.

Supplementary Information accompanies this paper on the *Oncogene* website (http://www.nature.com/onc)