

Transcriptional landscape of human cancers

Mengyuan Li¹, Qingrong Sun¹ and Xiaosheng Wang²

¹School of Science, China Pharmaceutical University, Nanjing 211198, China

²Department of Basic Medicine, School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 211198, China

Correspondence to: Xiaosheng Wang, **email:** xiaosheng.wang@cpu.edu.cn

Keywords: gene expression profiles, pathways, human cancers, TCGA, intertumor homogeneity and heterogeneity

Received: December 27, 2016

Accepted: February 08, 2017

Published: March 02, 2017

Copyright: Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The homogeneity and heterogeneity in somatic mutations, copy number alterations and methylation across different cancer types have been extensively explored. However, the related exploration based on transcriptome data is lacking. In this study we explored gene expression profiles across 33 human cancer types using The Cancer Genome Atlas (TCGA) data. We identified consistently upregulated genes (such as E2F1, EZH2, FOXM1, MYBL2, PLK1, TTK, AURKA/B and BUB1) and consistently downregulated genes (such as SCARA5, MYOM1, NKAPL, PEG3, USP2, SLC5A7 and HMGCLL1) across various cancers. The dysregulation of these genes is likely to be associated with poor clinical outcomes in cancer. The dysregulated pathways commonly in cancers include cell cycle, DNA replication, repair, and recombination, Notch signaling, p53 signaling, Wnt signaling, TGF β signaling, immune response etc. We also identified genes consistently upregulated or downregulated in highly-advanced cancers compared to lowly-advanced cancers. The highly (low) expressed genes in highly-advanced cancers are likely to have higher (lower) expression levels in cancers than in normal tissue, indicating that common gene expression perturbations drive cancer initiation and cancer progression. In addition, we identified a substantial number of genes exclusively dysregulated in a single cancer type or inconsistently dysregulated in different cancer types, demonstrating the intertumor heterogeneity. More importantly, we found a number of genes commonly dysregulated in various cancers such as PLP1, MYOM1, NKAPL and USP2 which were investigated in few cancer related studies, and thus represent our novel findings. Our study provides comprehensive portraits of transcriptional landscape of human cancers.

INTRODUCTION

It has been recognized that cancer is associated with the genetic and genomic changes [1]. With the advance of microarray and next-generation sequencing technology, gene expression profiling has been widely used for identifying molecular biomarkers for cancer diagnosis, treatment and prognosis [2–6]. In addition, as the biology of cancer is extremely complicated, a simple genetic or genomic perspective is insufficient to understand it. The exploration of pathway perturbations in cancer is critical in comprehending the disease [7–9].

Cancers originating from different tissues or cell types vary in terms of their genomic profiles. Lawrence

et al analyzed 27 cancer types and found that the median frequency of non-synonymous mutations varied by more than 1,000-fold across different cancer types [10]. The variation in mutation frequencies is mostly associated with cancer tissue type of origin where haematological and paediatric cancers have the lowest mutation frequencies while melanoma and lung cancers the highest mutation frequencies [10]. Zack et al analyzed the copy number profiles of 4,934 primary cancer specimens across 11 cancer types and found that the mean rate of somatic copy number alterations (SCNAs) varied across different cancer types with ovarian, cervix, breast and bladder cancers having a large number of SCNAs while leukemia and kidney cancers very few SCNAs [11]. Previous studies

have also shown that numbers of methylomes and patterns of DNA methylation varied across different cancer types [12, 13].

Due to the varied genomic profiles, different cancer types may show different prognosis. Among all cancer types, pancreatic, lung, liver and esophageal cancers have the worst survival prognosis while prostate, thyroid and skin cancers have the best survival prognosis [14]. Moreover, different cancer types with the same genomic or genetic profiles may exhibit different responses to the same treatment strategies. For example, melanoma with the BRAF V600E mutation is highly responsive to the small-molecule inhibitor vemurafenib while colon cancers with the same mutation show a very limited response to this drug [15]. On the other hand, different cancer types show the major homogeneity. It has been appreciated that the multistep development of human tumors depends on the eight biological capabilities acquired [8, 16]. They include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, reprogramming of energy metabolism and evading immune destruction [8, 16]. Although intertumor and intratumor heterogeneity extensively exists in genomic profiles [10, 17, 18], all the cancer driver genes are associated with 12 pathways that confer a selective growth advantage [9]. The 12 pathways include APC, Hedgehog, NOTCH, chromatin modification, transcriptional regulation, DNA damage control, TGF- β , MAPK, STAT, PI3K, RAS, cell cycle/apoptosis [9].

With the emergence of large-scale cancer genomics projects such as the International Cancer Genome Consortium (ICGC, <http://icgc.org/>) [19] and The Cancer Genome Atlas (TCGA, <https://gdc-portal.nci.nih.gov/>), the homogeneity and heterogeneity across different cancer types have been extensively explored [10–13, 17, 18, 20, 21]. The TCGA datasets cover 33 different cancer types and more than 10,000 cancer cases in total. Each TCGA cancer type contains different types of “omics” data, including: whole exome or genome sequencing; genomic DNA copy number arrays; DNA methylation; mRNA expression array and RNA-Seq data; microRNA sequencing; reverse-phase protein arrays; and clinical metadata. Based on the TCGA datasets, the tumor homogeneity and heterogeneity have been studied in various genomic profiles including somatic mutations [10, 18], SCNAs [11], and methylation [12, 13]. However, the exploration of homogeneity and heterogeneity across different cancer types specifically based on transcriptome data from large-scale cancer genomics projects such as TCGA is lacking.

In this study we explored gene expression profiles across 33 human cancer types in TCGA (Table 1). We identified dysregulated genes and pathways across different cancer types, and performed survival analyses based on expression profiles of the dysregulated genes.

Tumors can be classified based on stage and grade. Tumor stage refers to the size and/or extent of the primary tumor and whether or not tumor cells have spread in the body [22]. Tumor grade refers to how abnormal the tumor cells and the tumor tissue look under a microscope compared to normal cells, indicative of how quickly a tumor is likely to grow and spread [22]. We identified differentially expressed (DE) genes and pathways between different stages and different grades of cancers, respectively. Furthermore, we explored the transcriptional homogeneity and heterogeneity across different cancer types. This study would bring additional insights into the biology of human cancers.

RESULTS

Identification of DE genes between cancer and normal samples

We compared gene expression levels between cancer and normal samples in 18 cancer types each of which contains at least five normal samples (Table 1). Supplementary Tables 1 and 2 list genes whose expression is significantly higher and lower in cancer than in normal samples (fold change > 1.5, false discovery rate (FDR) < 0.05), respectively. The numbers of DE genes vary across different cancer types (Figure 1). The most number (5,755) of genes are more highly expressed in CHOL, and the least number (1,780) in PRAD. The most number (6,404) of genes are more lowly expressed in KICH, and the least number (2,797) in ESCA. The median number of genes with higher and lower expression levels in cancers is 3,626 and is 4,055, respectively.

Identification of genes consistently upregulated in different types of cancer

There are 51 genes consistently upregulated in all the 18 cancer types, and 52 genes consistently upregulated in 17 of the 18 cancer types compared to normal tissue (Supplementary Tables 3 and 4). Many of these genes have been reported to be overexpressed in cancers and associated with cancer progression and aggression. For example, E2F1 encodes a member of the E2F family of transcription factors (TFs) which play a crucial role in the control of cell cycle [23]. E2F1 has been shown to be upregulated in various cancers [24–26]. EZH2 encodes a member of the Polycomb-group (PcG) family that regulates cell generations. High EZH2 expression has been associated with different types of cancer [27]. Interestingly, the collaboration of EZH2 and E2F1 in transcriptional regulation has been observed in various cancers [28, 29]. The other TF genes such as FOXM1 [30–33], MYBL2 [34], NFE2L3 [35], and UHRF1 [36, 37] have been involved in various malignancies with overexpression. Among the 103 genes that are consistently

Table 1: 33 TCGA cancer types in which gene expression profiles were analyzed

Cancer ^a	Full name	# cancer samples	# normal samples	Stage ^b		Grade ^c	
				# early-stage	# late-stage	# low-grade	# high-grade
<u>BLCA</u>	bladder urothelial carcinoma	408	19	132	274	21	384
<u>BRCA</u>	breast invasive carcinoma	1100	112	801	270	NA	NA
<u>CHOL</u>	cholangiocarcinoma	36	9	28	8	16	20
<u>COAD</u>	colon adenocarcinoma	287	41	155	119	NA	NA
<u>ESCA</u>	esophageal carcinoma	185	11	97	65	95	49
<u>GBM</u>	glioblastoma multiforme	166	5	NA	NA	NA	NA
<u>HNSC</u>	head and neck squamous cell carcinoma	522	44	118	388	366	132
<u>KICH</u>	kidney chromophobe	66	25	46	20	NA	NA
<u>KIRC</u>	kidney renal clear cell carcinoma	534	72	324	207	243	282
<u>KIRP</u>	kidney renal papillary cell carcinoma	291	32	193	67	NA	NA
<u>LIHC</u>	liver hepatocellular carcinoma	373	50	257	90	232	134
<u>LUAD</u>	lung adenocarcinoma	517	59	397	110	NA	NA
<u>LUSC</u>	lung squamous cell carcinoma	501	51	406	91	NA	NA
<u>PRAD</u>	prostate adenocarcinoma	498	52	NA	NA	NA	NA
<u>READ</u>	rectum adenocarcinoma	95	10	38	46	NA	NA
<u>STAD</u>	stomach adenocarcinoma	415	35	180	212	160	246
<u>THCA</u>	thyroid carcinoma	509	59	334	165	NA	NA
<u>UCEC</u>	uterine corpus endometrial carcinoma	370	11	272	98	185	185
ACC	adrenocortical carcinoma	79	0	46	31	NA	NA
CESC	cervical squamous-cell carcinoma and endocervical adeno-carcinoma	306	3	231	66	153	119
DLBC	lymphoid neoplasm diffuse large B-cell lymphoma	48	0	25	17	NA	NA
LAML	acute myeloid leukemia	173	0	NA	NA	NA	NA
LGG	brain lower-grade glioma	530	0	NA	NA	249	265
OV	ovarian serous cystadenocarcinoma	307	0	22	280	34	262
PAAD	pancreatic adeno-carcinoma	179	4	168	8	126	50
PCPG	pheochromocytoma and paraganglioma	184	3	NA	NA	NA	NA
SARC	sarcoma	263	2	NA	NA	NA	NA
SKCM	skincutaneous melanoma	472	1	217	193	NA	NA
TGCT	testicular germ-cell tumors	156	0	110	15	NA	NA

(Continued)

Cancer ^a	Full name	# cancer samples	# normal samples	Stage ^b		Grade ^c	
				# early-stage	# late-stage	# low-grade	# high-grade
UCS	uterine carcino-sarcoma	57	0	27	30	NA	NA
UVM	uveal melanoma	80	0	36	44	NA	NA
MESO	mesothelioma	87	0	26	61	NA	NA
THYM	thymoma	120	2	97	21	NA	NA

^a 18 cancer types are underlined in which gene expression were compared between cancer and normal samples.

^b Numbers of early-stage (stage I-II) and late-stage (stage III-IV) cancer samples are shown.

^c Numbers of low-grade (G1-2) and high-grade (G3-4) cancer samples are shown.

* "NA" indicates no related information available.

upregulated in at least 17 of the 18 cancer types, there are 11 protein kinase encoding genes including PLK1, TTK, AURKA, AURKB, BUB1, BUB1B, GSG2, MELK, NEK2, PBK, and PKMYT1. Most of these kinase genes have been shown to be overexpressed in various cancers such as PLK1 [38–43], TTK [44–47], and AURKA/B [48–51]. These protein kinases are of particular interest because kinase inhibitors have been intensively investigated as a key class of anticancer drugs in clinical use or trials [52].

Using Gene Set Enrichment Analysis (GSEA) software [53], we identified 41 Rectome pathways [54] significantly associated with the set of 103 genes (FDR<0.05, Figure 2 and Supplementary Table 5). Obviously, these gene products are significantly involved in cancer related pathways such as cell cycle, DNA replication and repair, and immune response. Network analysis of the gene set composed of the aforementioned seven TF genes (E2F1, EZH2, FOXM1, MYBL2, NFE2L3 and UHRF1) and the 11 protein kinase genes by STRING [55] shows that the protein products of these genes interact to each other (Figure 3). PLK1, a hub node in the network, interacts with 11 of the other 17 proteins (BUB1, BUB1B, PKMYT1, AURKA, AURKB, FOXM1, MYBL2, NEK2, TTK, PBK, and GSG2). As another hub node in the network, BUB1 interacts with 12 of the other 17 proteins (BUB1B, PLK1, MELK, MYBL2, PKMYT1, AURKA, AURKB, FOXM1, NEK2, TTK, PBK, and GSG2). The TF FOXM1 regulates seven protein kinases (BUB1, BUB1B, PLK1, MELK, AURKA, AURKB, and NEK2). The three TFs E2F1, EZH2 and MYBL2 interact with each other. This demonstrates that interactions of these oncoproteins may play an important role in the initiation and progression of various cancers.

We compared overall survival (OS) and disease-free survival (DFS) between patients with higher expression levels and patients with lower expression levels of the 11 protein kinase genes and seven TF genes in 30 cancer types (three cancer types were excluded from the analysis due to lack of survival data). Kaplan-Meier survival

curves show that higher expression levels of these genes are associated with significantly worse OS and DFS prognoses in various cancers (Figure 4 and Supplementary Figure 1). For example, patients with higher expression levels of BUB1 have worse OS prognoses than those with lower expression levels of BUB1 in 10 cancer types (ACC, HNSC, KICH, KIRC, KIRP, LGG, LUAD, PAAD, and SKCM), and worse DFS prognoses in nine cancer types (ACC, KIRC, KIRP, LGG, LIHC, LUAD, PAAD, SARC and UVM) (Figure 4A, log-rank test, unadjusted P-value < 0.05). Patients with higher expression levels of FOXM1 have worse OS prognoses than those with lower expression levels of FOXM1 in 11 cancer types (ACC, BRCA, KICH, KIRC, KIRP, LGG, LUAD, PAAD, SKCM, UCEC and UVM), and worse DFS prognoses in seven cancer types (ACC, KIRC, KIRP, LIHC, SARC, SKCM and UVM) (Figure 4B, log-rank test, unadjusted P-value < 0.05). Kaplan-Meier survival curves show that higher expression of NEK2 and MYBL2 is also associated with worse clinical outcomes in various cancers (Figure 4C, 4D). These results are consistent with previous studies showing that overexpression of BUB1, FOXM1, and NEK2 correlated with poor prognosis of cancers [56–58].

Identification of genes consistently downregulated in different types of cancer

We identified 11 genes (SCARA5, PLP1, MYOM1, ADH1B, NKAPL, SYNE1, PEG3, USP2, PCDH9, SLC5A7 and HMGCLL1) which are consistently downregulated in all the 18 cancer types compared to normal tissue. Among the 11 genes, SCARA5 expression has been shown to be frequently downregulated in various cancers [59], and has been proposed as a novel tumor suppressor gene [60]. ADH1B was shown to be downregulated in colorectal cancer [61] and lung cancer [62]. SYNE1 has been shown to be downregulated in various human cancers [63]. PEG3 encodes a tumor suppressor and was downregulated in several cancer types

[64–66]. The downregulation of PCDH9 has been shown to contribute to the development of various human cancers [67–69]. These previous studies confirm the results we obtained from the TCGA data analysis. However, few studies have shown that PLP1, MYOM1, NKAPL, and USP2 were consistently downregulated in various cancers. Thus, this analysis provides novel findings about these genes whose downregulation may play an important role

in carcinogenesis. In addition, HMGCLL1 has been shown to be upregulated in several human cancers such as breast cancer [70] and brain cancer [71]. SLC5A7 had elevated mRNA expression in breast cancer cells compared with mammary epithelial cells [72]. These observations conflict with our results from the TCGA data analysis. Therefore, the roles of HMGCLL1 and SLC5A7 in carcinogenesis remain to be clarified.

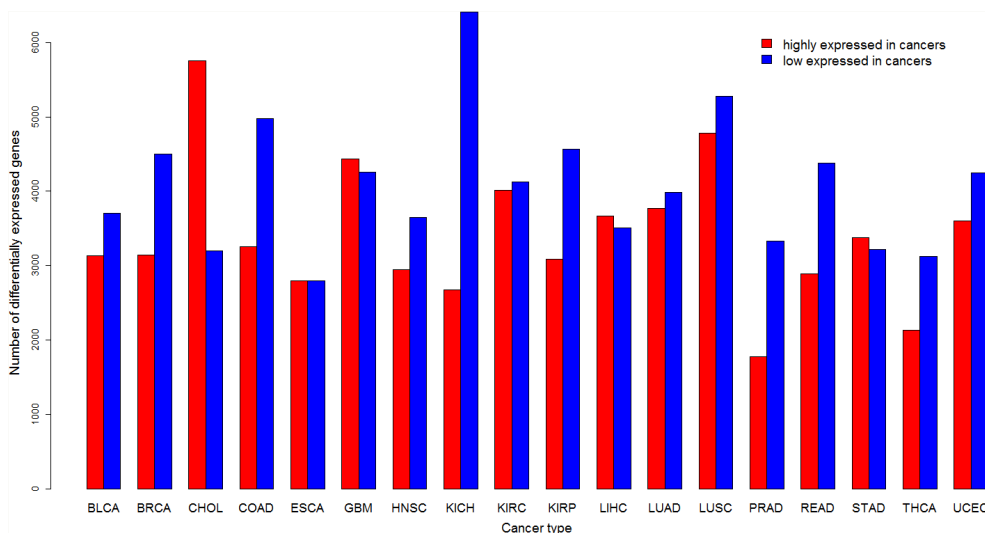


Figure 1: Number of differentially expressed (DE) genes between cancer and normal samples identified in each of the 18 cancer types.

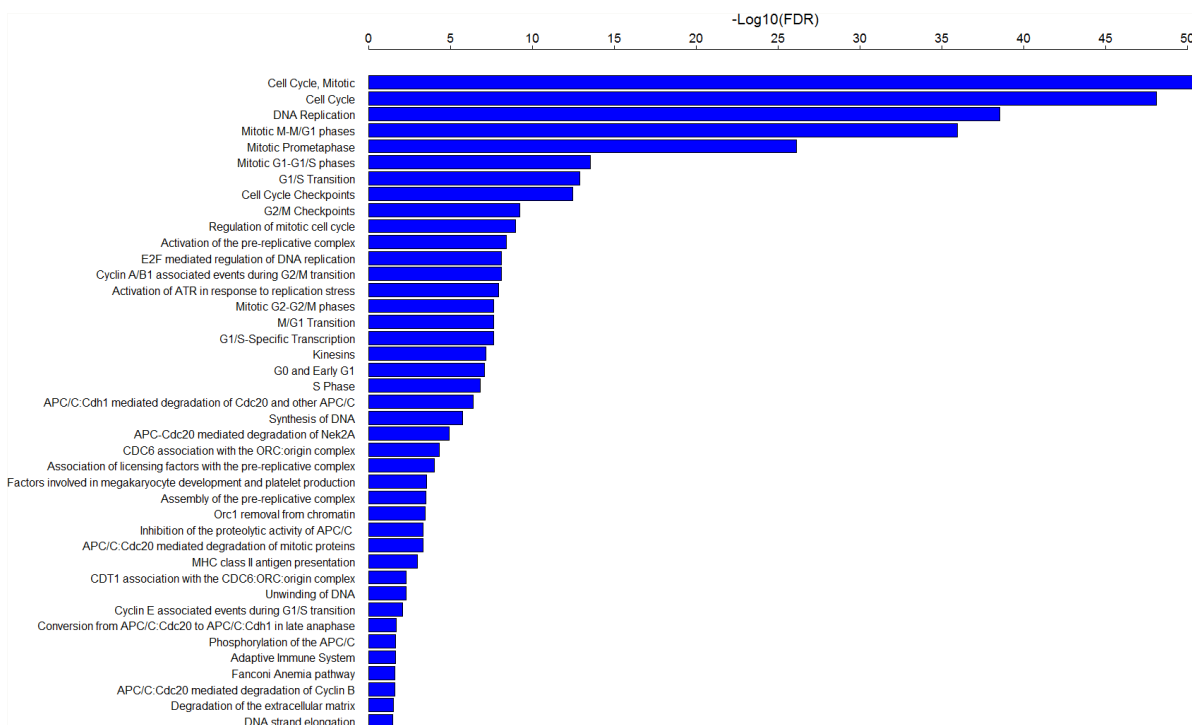


Figure 2: 41 Rectome pathways significantly associated with the 103 genes upregulated in various cancers.

We compared OS and DFS between patients with higher expression levels and patients with lower expression levels of the 11 genes in the 30 cancer types. Kaplan-Meier survival curves show that higher expression levels of most of these genes are associated with better OS and DFS prognoses in multiple cancer types (Figure 5 and Supplementary Figure 1). For example, patients with higher expression levels of NKAPL have better OS prognoses than those with lower expression levels of NKAPL in four cancer types (ACC, KIRP, LGG, and PAAD), and better DFS prognoses in three cancer types (ACC, KIRP, and THYM) (Figure 5A, log-rank test, unadjusted P-value < 0.05). Patients with higher expression levels of USP2 have better OS prognoses than those with lower expression levels of USP2 in three cancer types (ACC, KIRC, and PAAD), and better DFS prognoses in three cancer types (ACC, KIRC, and THCA) (Figure 5B, log-rank test, unadjusted P-value < 0.05).

Kaplan-Meier survival curves show that higher expression of PEG3 and SLC5A7 is also associated with better clinical outcomes in several cancer types (Figure 5C, 5D). Surprisingly, a literature survey shows that few studies have revealed the correlation between overexpression of these genes and beneficial clinical outcomes in cancers.

In addition, we identified 53 genes which are consistently downregulated in 17 of the 18 cancer types (Supplementary Tables 4 and 6). Pathway analysis did not find any significant pathway associated with the set of 64 genes consistently downregulated in at least 17 of the 18 cancer types (FDR<0.05).

Identification of DE genes between highly-advanced and lowly-advanced cancers

We compared gene expression levels between early-stage (stage I-II) and late-stage (stage III-IV) cancers, and

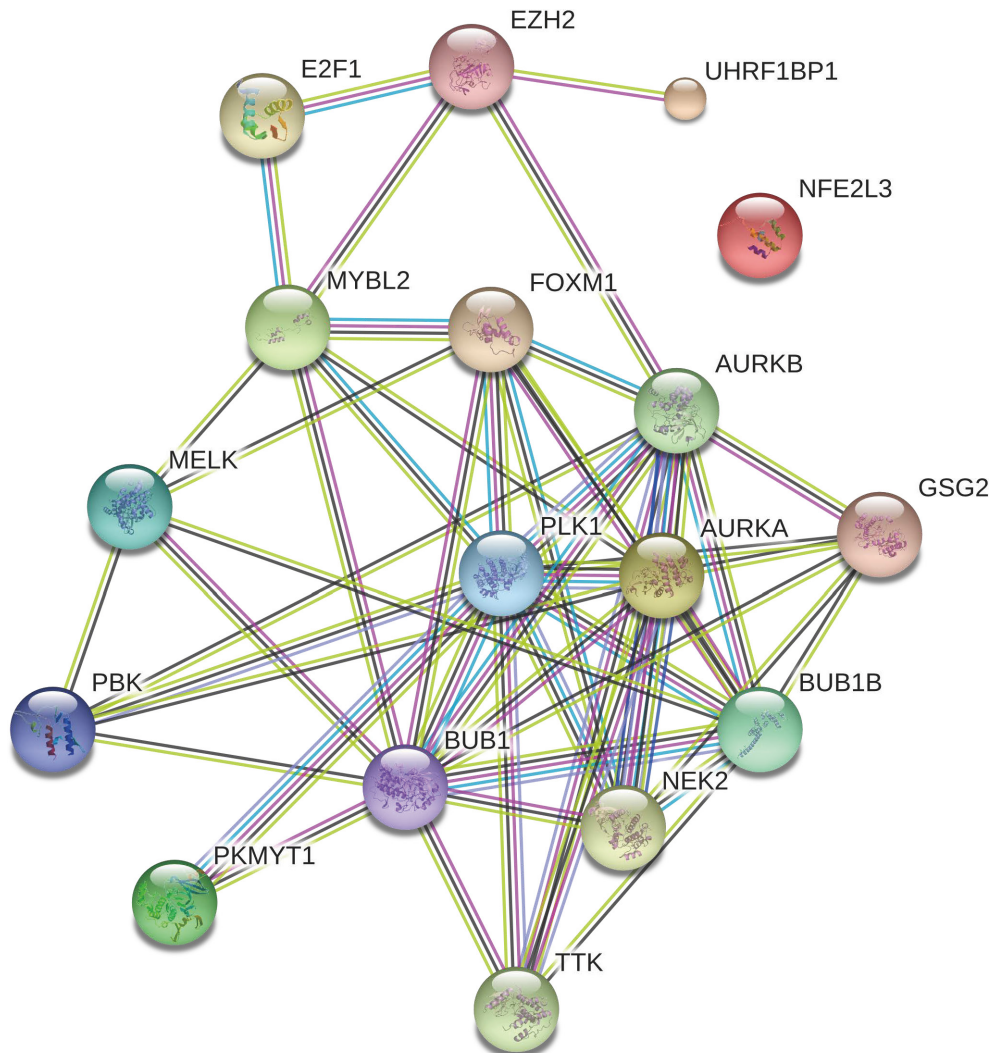


Figure 3: Interaction network associated with the seven transcriptional genes and the 11 protein kinase genes overexpressed in various cancers.

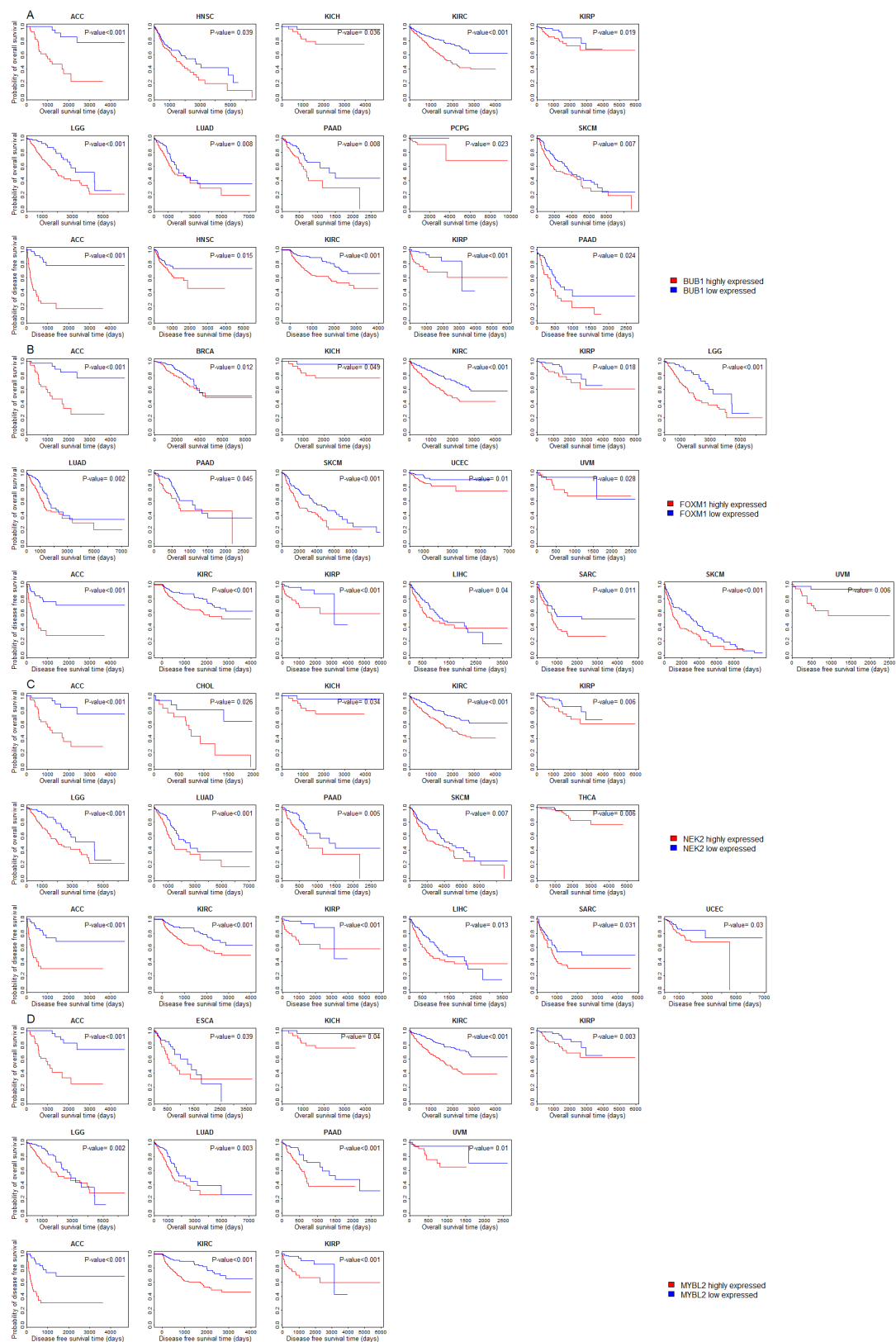


Figure 4: Survival analyses of cancer patients based on expression of the upregulated genes in cancers (log-rank test, unadjusted P-value < 0.05). (A) Compare survival time between BUB1 higher-expression-level and BUB1 lower-expression-level cancers; (B) Compare survival time between FOXM1 higher-expression-level and FOXM1 lower-expression-level cancers; (C) Compare survival time between NEK2 higher-expression-level and NEK2 lower-expression-level cancers; (D) Compare survival time between MYBL2 higher-expression-level and MYBL2 lower-expression-level cancers.

between low-grade (G1-2) and high-grade (G3-4) cancers, respectively. We refer to early-stage or low-grade cancers as lowly-advanced cancers, and late-stage or high-grade cancers highly-advanced cancers. There are 27 and 12 cancer types whose clinical data contain stage and grade information, respectively (Table 1). Supplementary Table 7 presents the numbers of DE genes between highly-advanced and lowly-advanced cancers. Supplementary Tables 8 and

9 list genes whose expression is higher and lower in late-stage than in early-stage cancers, respectively. In 13 of the 27 cancer types there are DE genes between different stages of cancers (fold change > 1.5, FDR < 0.05). The numbers of DE genes between different stages of cancers vary across different cancer types (Figure 6A). In KIRP the most number (1,318) of genes are more highly expressed in late-stage than in early-stage cancers, and in the same

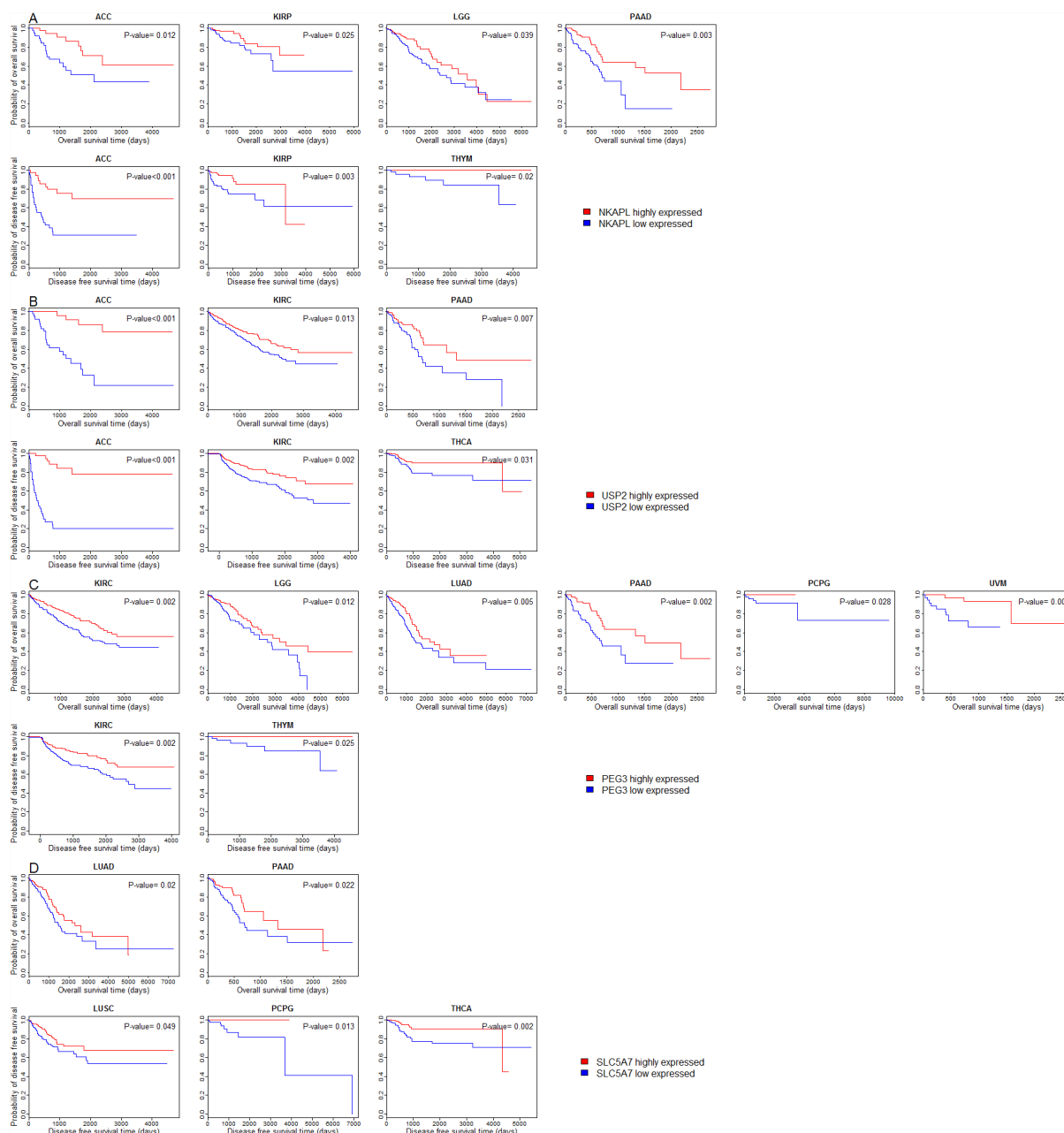


Figure 5: Survival analyses of cancer patients based on expression of the downregulated genes in cancers (log-rank test, unadjusted P-value < 0.05). (A) Compare survival time between NKAPL higher-expression-level and NKAPL lower-expression-level cancers; (B) Compare survival time between USP2 higher-expression-level and USP2 lower-expression-level cancers; (C) Compare survival time between PEG3 higher-expression-level and PEG3 lower-expression-level cancers; (D) Compare survival time between SLC5A7 higher-expression-level and SLC5A7 lower-expression-level cancers.

cancer type the most number (975) of genes are more highly expressed in early-stage than in late-stage cancers. In contrast, in some other cancer types such as BRCA, only two and six genes were upregulated and downregulated in late-stage compared to early-stage cancers, respectively. In some cancer types such as KIRP, BLCA, KIRC, ACC, and COAD, the number of genes with higher expression levels in late-stage than in early-stage cancers are much higher than that of genes with lower expression levels in late-stage than in early-stage cancers. In some other cancer types such as THYM, LIHC, HNSC, and LUAD, the situation is just the opposite.

Supplementary Tables 10 and 11 list genes whose expression is higher and lower in high-grade than in low-grade cancers, respectively. In nine of the 12 cancer types there are DE genes between different grades of cancers (fold change > 1.5, FDR < 0.05). The numbers of DE genes between different grades of cancers also vary across different cancer types (Figure 6B). In BLCA the most number (3,572) of genes are more highly expressed in high-grade than in low-grade cancers, and in the same cancer type the most number (1,819) of genes are more highly expressed in low-grade than in high-grade cancers. In contrast, in ESCA only 79 and 118 genes were upregulated and downregulated in high-grade compared to low-grade cancers, respectively. In some cancer types such as STAD, the number of genes with higher expression levels in high-grade than in low-grade cancers are much higher than that of genes with lower expression levels in high-grade than in low-grade cancers (1,420 versus 573). In some other cancer types such as CESC, we observed the opposite situation (35 versus 156).

These results indicate that during cancer progression some cancers such as BLCA, KIRP, KIRC, HNSC, LIHC and UCEC exhibit expression disturbances in a large number of genes, while some other cancers such as BRCA, PAAD, LUAD, LUSC, and SKCM exhibit expression disturbances in a small number of genes. Interestingly, although both stage and grade indicate the degree of cancer progression, in some cancer types such as STAD, ESCA, and CESC, the markedly different numbers of DE genes between lowly-advanced and highly-advanced cancers were identified in the stage and grade phenotype comparisons (Figure 6A, 6B).

Identification of genes upregulated in highly-advanced cancers

We identified 71 genes which are upregulated in late-stage compared to early-stage cancers in more than three cancer types (Supplementary Table 12). We call the 71 genes late-stage-activated (LSA) genes. Pathway analysis of the 71 LSA genes identified four significant Rectome pathways: extracellular matrix organization (P-value = 6.9×10^{-15}), collagen formation (P-value = 2.1×10^{-14}), cell surface interactions at the vascular wall (P-value = 9.4×10^{-6}), and integrin cell surface interactions (P-value = 0.0002). Obviously, these pathways are associated with cancer progression and metastasis characteristics such as cell interaction, cell adhesion, and cell motility. Some of the LSA genes have been shown to be overexpressed in advanced cancers and be associated with unfavorable clinical outcomes such as SOX11 [73, 74], PTPRN [75], PNCK [76] and HMGA2 [77–79]. In addition, we

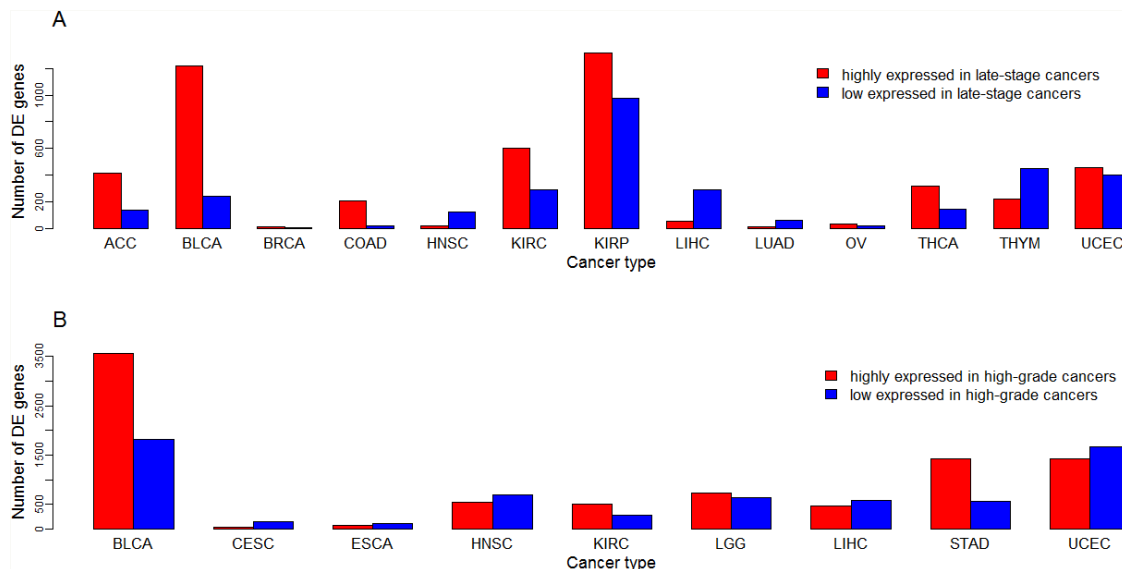


Figure 6: Number of differentially expressed (DE) genes between lowly-advanced and highly-advanced cancers across different cancer types. (A) Number of DE genes between early-stage and late-stage cancers; (B) Number of DE genes between low-grade and high-grade cancers.

identified 212 genes which are upregulated in high-grade compared to low-grade cancers in more than three cancer types (Supplementary Table 13). We call the 212 genes high-grade-activated (HGA) genes. Pathway analysis of these HGA genes identified 63 significant Rectome pathways (Supplementary Table 14). These pathways are mainly involved in cell cycle, DNA replication, and immune system whose dysregulations are the leading causes of cancer development [8, 9].

There are much more HGA genes upregulated in cancers than HGA genes downregulated in cancers compared to normal tissue. For example, in more than nine (50%) of the 18 cancer types, 128 (60%) of the 212 HGA genes are upregulated in cancers, compared to 15 (7%) of the 212 HGA genes downregulated in cancers (Fisher's exact test, P -value $< 2.2 \times 10^{-16}$). Similarly, there are much more LSA genes upregulated in cancers than LSA genes downregulated in cancers compared to normal tissue. For example, in more than nine (50%) of the 18 cancer types, 26 (37%) of the 71 LSA genes are upregulated in cancers, compared to 9 (13%) of the 71 LSA genes downregulated in cancers (Fisher's exact test, P -value = 0.0016). These results indicate that cancer initiation (normal tissue evolving into cancer tissue) and cancer progression (low-grade cancers evolving into high-grade cancers, or early-stage cancers evolving into late-stage cancers) may depend on many common changes in gene expression profiles. In fact, we found a number of genes whose expression follows this pattern:

late-stage cancers $>$ early-stage cancers $>$ normal tissue. For example, in the eight cancer types (BLCA, BRCA, COAD, KIRC, LIHC, LUAD, THCA and UCEC) with both stage phenotype information and normal control samples, SOX11 expression follows the pattern in four cancer types (BLCA, KIRC, LIHC and THCA) (Figure 7A). Similarly, we found a number of genes whose expression follows this pattern: high-grade cancers $>$ low-grade cancers $>$ normal controls. For example, in the seven cancer types (BLCA, ESCA, HNSC, KIRC, LIHC, STAD and UCEC) with both grade phenotype information and normal control samples, expression of AURKB, BUB1, FOXM1, HMMR, MYBL2, and PLK1 follows the pattern in five cancer types (BLCA, HNSC, KIRC, LIHC, and UCEC) (Figure 7B). These results confirm that overexpression of PLK1 [38–42], BUB1 [80], AURKB [50], HMMR [81], FOXM1 [30–33], MYBL2 [34] and SOX11 [73, 74] are associated with both cancer onset and cancer progression.

Surprisingly, very few (only seven) genes are common between the LSA gene list and the HGA gene list. In fact, the gene expression profiling alteration (GEPA) from normal tissue to cancers is closer to the GEPA from low-grade to high-grade cancers than to that from early-stage to late-stage cancers. For example, 42 (82%) of the 51 genes overexpressed in all the 18 cancer types are included in the HGA gene list, compared to one (2%) of the 51 genes included in the LSA gene list (Fisher's exact test, P -value $< 2.2 \times 10^{-16}$). In more than nine (50%)

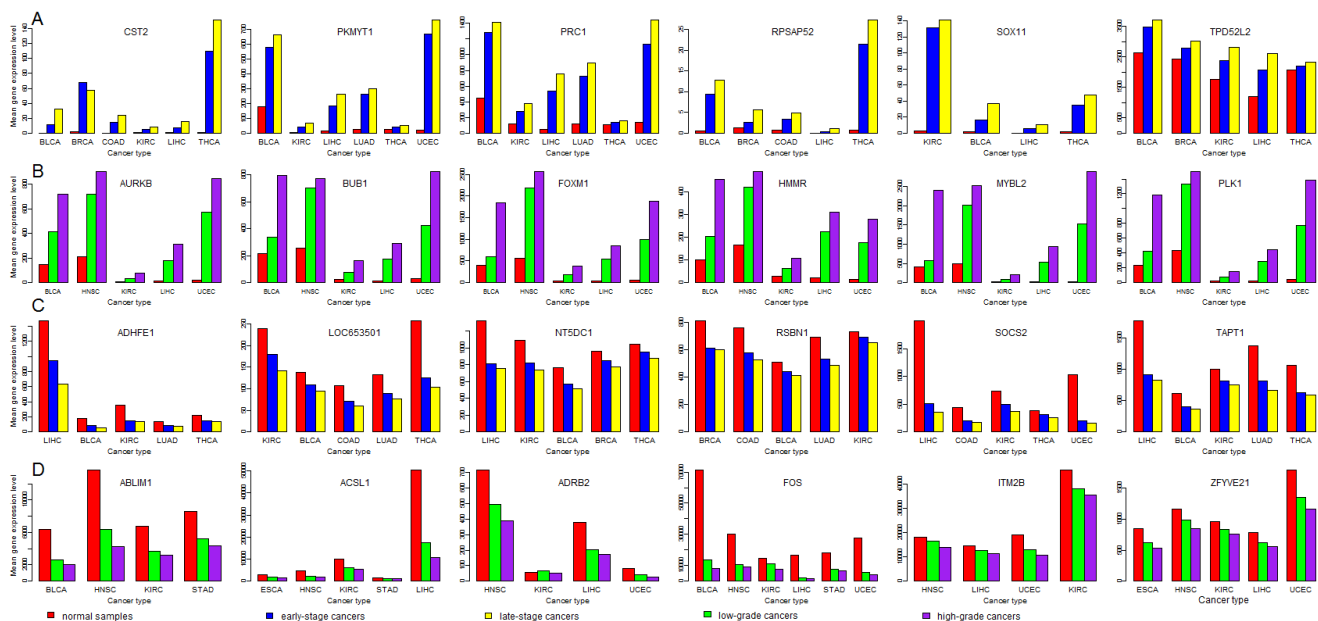


Figure 7: Genes have increased or decreased expression alterations from normal tissue to lowly-advanced cancers, and to highly-advanced cancers (Student's t test, P -value < 0.05). (A) Genes whose expression follows the pattern: late-stage cancers $>$ early-stage cancers $>$ normal tissue; (B) Genes whose expression follows the pattern: high-grade cancers $>$ low-grade cancers $>$ normal tissue; (C) Genes whose expression follows the pattern: late-stage cancers $<$ early-stage cancers $<$ normal tissue; (D) Genes whose expression follows the pattern: high-grade cancers $<$ low-grade cancers $<$ normal tissue.

of the 18 cancer types, 128 (60%) of the HGA genes are more highly expressed in cancers than in normal tissue, compared to 26 (37%) of the LSA genes (Fisher's exact test, P-value = 0.0006). It is consistent with a recent study showing that cancer grade, but not stage, was driven by transcriptional alterations [82].

Identification of genes downregulated in highly-advanced cancers

We identified 13 genes downregulated in late-stage compared to early-stage cancers in more than three cancer types (Supplementary Table 15). We call the 13 genes (DNASE1L3, CD1E, SLC44A4, PLIN5, IYD, RORC, GGT6, FBP1, ALDH1L1, PIGR, SPATA18, ARPP21 and CWH43) late-stage-inactivated (LSiA) genes. Pathway analysis of these LSiA genes did not find any significant Rectome pathway associated with them. In addition, we identified 58 genes downregulated in high-grade compared to low-grade cancers in more than three cancer types (Supplementary Table 16). We call the 58 genes high-grade-inactivated (HGIA) genes. Pathway analysis of these HGIA genes identified three significant Rectome pathways: biological oxidations, phase 1 - functionalization of compounds, and cytochrome P450 - arranged by substrate type. The associations between these pathways and cancer progression are unclear and remain to be investigated.

There are more HGIA genes downregulated in cancers than HGIA genes upregulated in cancers compared to normal tissue, although the difference is not significant. For example, in more than nine (50%) of the 18 cancer types, 9 (16%) of the 58 HGIA genes are downregulated in cancers, compared to 3 (5%) of the 58 HGIA genes upregulated in cancers (Fisher's exact test, P-value = 0.12). Similarly, there are more LSiA genes downregulated in cancers than LSiA genes upregulated in cancers, although the difference is not significant. Likewise, we found a number of genes whose expression follows this pattern: late-stage cancers < early-stage cancers < normal tissue (Figure 7C), such as ADHFE1, LOC653501, NT5DC1, RSBN1, SOCS2 and TAPT1 in five cancer types. We also found a number of genes whose expression follows this pattern: high-grade cancers < low-grade cancers < normal tissue (Figure 7D), such as ALDH1L1, WIF1, ACSL1, FOS, and ABLIM1 in at least four cancer types. ALDH1L1 [83, 84] and WIF1 [85, 86] have been shown to be ubiquitously downregulated in cancers, and their downregulation was associated with poor clinical outcomes in cancer.

Again, very few genes (only three genes ALDH1L1, SLC44A4 and IYD) are common between the LSiA gene list and the HGIA gene list. It strongly suggests that although both stage and grade reflect the status of cancer advancement in phenotype, they are markedly different in molecular levels.

Pathway analyses of DE genes

We used GSEA software [53] to perform pathway analyses of the DE genes between cancer and normal samples in each of the 18 cancer types. Supplementary Tables 17 and 18 show significant KEGG pathways [87] associated with the upregulated and downregulated genes in cancers (FDR < 0.05), respectively. There are 22 pathways only associated with upregulated genes, and six pathways only associated with downregulated genes in at least one cancer type (Supplementary Table 19). Remarkably, the cell cycle pathway is consistently upregulated in all the 18 cancer types (Figure 8), suggesting that hyperactivation of this pathway is a common mechanism underlying cancer initiation and progression. The other dysregulated pathways in cancer such as DNA replication, repair, and recombination, and Notch signaling [9] were also identified to be aberrantly activated in various cancers in this study (Supplementary Table 19).

Pathway analyses of the DE genes between early-stage and late-stage cancers in each of the 13 cancer types identified 28 and 48 significant KEGG pathways associated only with the upregulated genes and only with the downregulated genes in late-stage cancers in at least one cancer type, respectively (FDR < 0.05, Supplementary Table 20). Pathway analyses of the DE genes between low-grade and high-grade cancers in each of the nine cancer types identified 44 and 29 significant KEGG pathways associated only with the upregulated genes and only with the downregulated genes in high-grade cancers in at least one cancer type, respectively (FDR < 0.05, Supplementary Table 21). Supplementary Tables 20 and 21 show that a number of pathways are upregulated in highly-advanced cancers such as cell cycle, ECM receptor interaction, DNA replication, DNA mismatch repair, homologous recombination, antigen processing and presentation, and nicotinate and nicotinamide metabolism. Among them, the pathways cell cycle, DNA replication, DNA mismatch repair and homologous recombination have also been identified to be consistently upregulated in various cancers (Supplementary Table 19). It suggests that hyperactivation of these pathways drives both cancer onset and cancer progression. The pathways significantly downregulated in highly-advanced cancers are mainly involved in metabolism regulation such as ether lipid metabolism, alpha linolenic acid metabolism, glycolysis gluconeogenesis, histidine metabolism, butanoate metabolism, beta alanine metabolism, propanoate metabolism, pyruvate metabolism, and phenylalanine metabolism.

Intertumor homogeneity and heterogeneity in gene expression profiles

We identified a number of genes which are consistently upregulated or downregulated in various cancers. For example, there are 51 genes consistently

upregulated and 11 genes consistently downregulated in all the 18 cancer types (Supplementary Tables 22, 23). Moreover, the cell cycle pathway is consistently upregulated in all the 18 cancer types (Supplementary Table 19). When comparing highly-advanced with lowly-advanced cancers, we also identified a number of genes consistently upregulated or downregulated in highly-advanced cancers (Supplementary Tables 24, 25, 26, 27). For example, there are 70 and 12 genes consistently upregulated and consistently downregulated in late-stage cancers in at least four cancer types, respectively. There are 49 and six genes consistently upregulated and consistently downregulated in high-grade cancers in at least five cancer types, respectively. These results demonstrate that there exist common genes and pathways whose dysregulations lead to the development of different types of cancer.

In addition, we identified a number of genes which are upregulated in some cancer types while downregulated in other cancer types (Supplementary Table 28). For example, there are 171 genes which are upregulated in at least six cancer types while downregulated in other at least six cancer types, respectively. We also identified a number of genes which are upregulated in highly-advanced cancers in some cancer types while downregulated in other cancer types (Supplementary Tables 29, 30). For example, there are 15 genes which are upregulated in late-stage cancers in at least two cancer types while downregulated in other at least two cancer types (Supplementary Table 29), and 110 genes which are upregulated in high-grade cancers in at least two cancer types while downregulated in other at least two cancer types (Supplementary Table 30). Moreover, we identified a number of genes exclusively

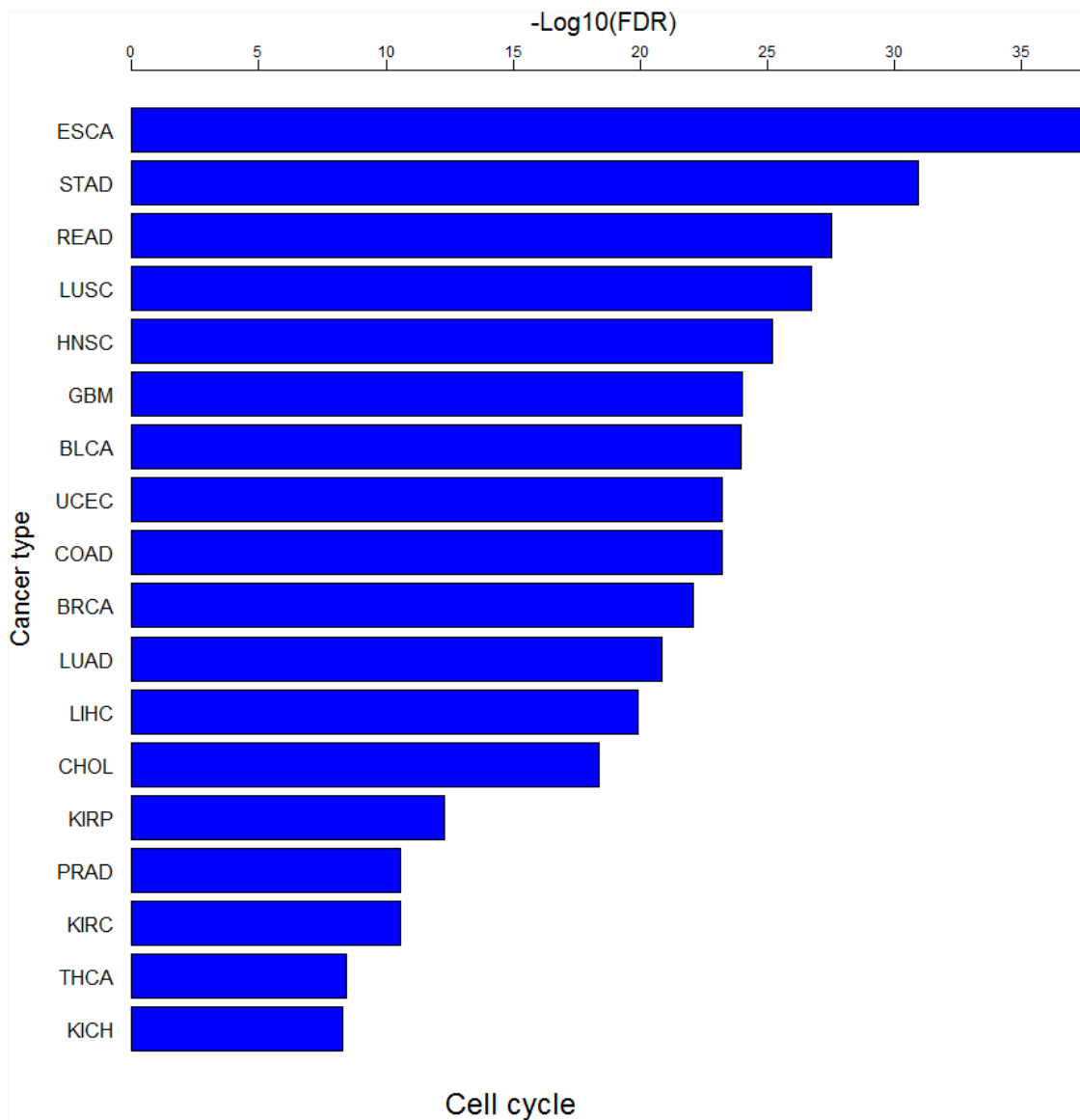


Figure 8: The cell cycle pathway is consistently upregulated in all the 18 cancer types.

dysregulated in a single cancer type (Supplementary Table 31, Figure 9). For example, there are 178 and 186 genes upregulated and downregulated in GBM, respectively, but not in the other 17 cancer types. The number of exclusively differentially expressed (EDE) genes in GBM is the most among the 18 cancer types, suggesting the major specificity in transcriptional dysregulations that underly the development of GBM. In fact, a recent study has shown that GBM is different from other cancers in that TP53-mutated GBM has a better prognosis than TP53-wildtype GBM while most of other cancers have worse prognoses when TP53 mutated [88]. Interestingly, although few genes are exclusively differentially expressed between BLCA and normal tissue, a large number of genes are exclusively differentially expressed between highly-advanced and lowly-advanced BLCA. It may suggest that the BLCA progression not BLCA onset is associated with substantial specific gene expression disturbances.

DISCUSSION

In this study we performed extensive analyses of gene expression and clinical data from 33 TCGA cancer

type-specific datasets. We identified upregulated and downregulated genes and pathways commonly across different cancer types. Many TF genes (such as E2F1, EZH2, FOXM1 and MYBL2) and protein kinase genes (such as PLK1, TTK, AURKA, AURKB, BUB1, MELK, NEK2, PBK and PKMYT1) are overexpressed in various cancers, and their overexpression is associated with poor clinical outcomes in cancer (Supplementary Figure 1). Clearly, these genes are oncogenes whose hyperactivation leads to cancer initiation and progression. In contrast, lower expression of the downregulated genes (such as SCARA5, MYOM1, NKAPL, PEG3, USP2, SLC5A7 and HMGCLL1) in various cancers is associated with poor clinical outcomes in cancer (Supplementary Figure 1). These genes are tumor suppressor genes whose hypoactivation leads to cancer initiation and progression. Although many of the identified genes such as E2F1, EZH2, FOXM1, PLK1, TTK, AURKA, AURKB, and BUB1 have been revealed to be dysregulated in various cancers by previous studies, many other genes such as PLP1, MYOM1, NKAPL, and USP2 were investigated in few cancer-related studies, and thus represent our novel findings. Pathway analyses show that the cell cycle

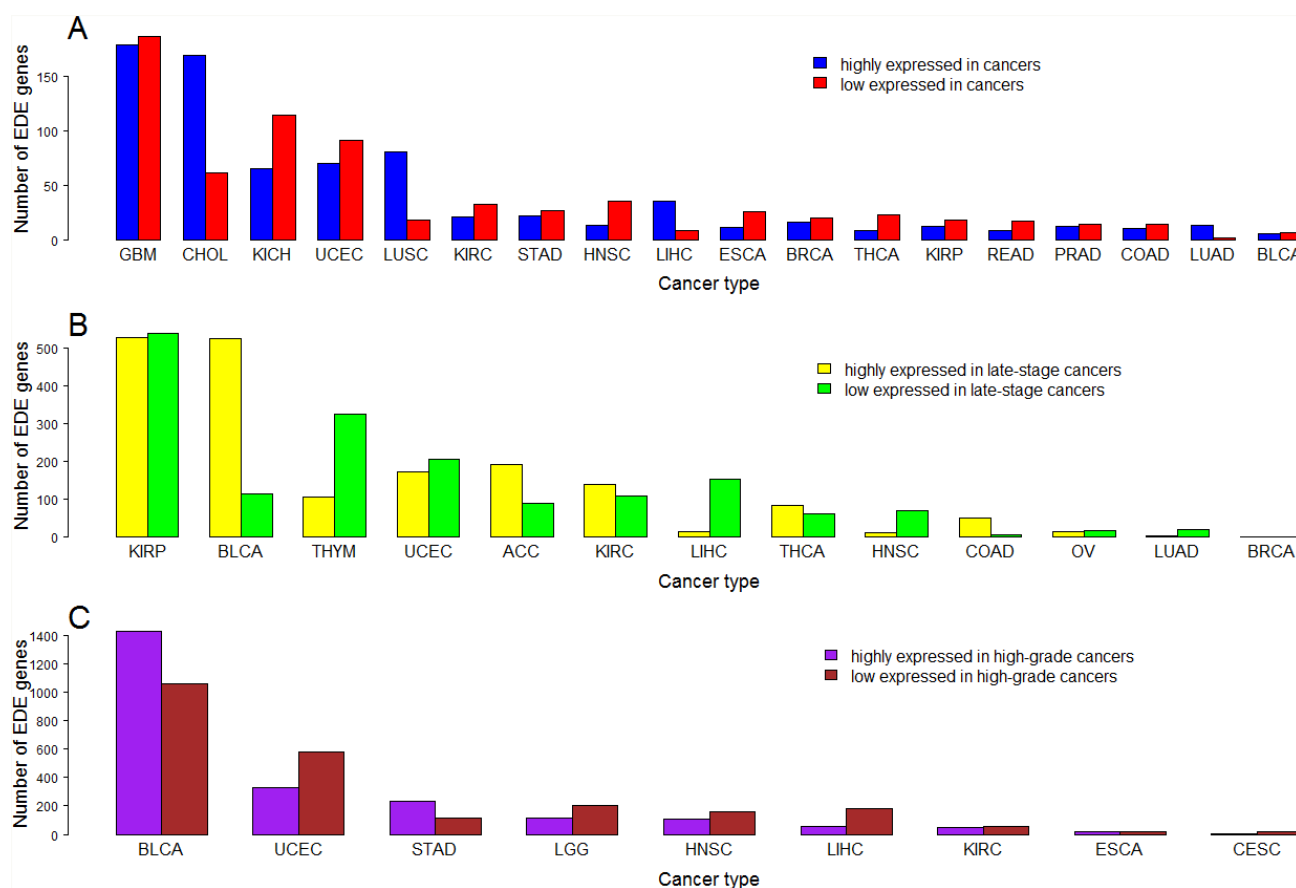


Figure 9: Number of exclusively differentially expressed (EDE) genes identified in a single cancer type. (A) Number of EDE genes between cancer and normal samples; **(B)** Number of EDE genes between early-stage and late-stage cancers; **(C)** Number of EDE genes between low-grade and high-grade cancers.

pathway is upregulated commonly in all the cancer types. The other dysregulated pathways in various cancers include DNA replication, repair, and recombination, Notch signaling etc. Moreover, we identified a number of genes consistently upregulated or downregulated in highly-advanced relative to lowly-advanced cancers. An interesting finding is that those genes with higher (lower) expression levels in highly-advanced than in lowly-advanced cancers are likely to have higher (lower) expression levels in cancers than in normal tissue. It suggests that common molecular perturbations drive cancer evolution from normal tissue to early cancer, and from early cancer to late cancer. However, we also found many genes which are upregulated in some cancer types while downregulated in other cancer types, and many genes which are exclusively dysregulated in a single cancer type. It suggests that there exists extensive intertumor heterogeneity in genomic profiles.

The commonly dysregulated genes and pathways identified in various cancers may involve attractive therapeutic targets for cancer. For example, since the cell cycle pathway is consistently hyperactivated in cancers, development of cell-cycle inhibitors may be effective in treatment of a wide type of cancers. We have identified a number of protein kinase genes which are upregulated commonly in cancers and are involved in the cell cycle regulation such as PLK1, TTK, BUB1, BUB1B, and PKMYT1. Development of small molecule inhibitors targeting these protein kinases could be a promising direction for curing cancer. On the other hand, the identification of a number of genes exclusively dysregulated in a single cancer type indicates that an individual cancer type may need its own specific therapeutic strategies in addition to the common strategies in cancer therapy. Furthermore, the identification of a considerable number of genes commonly dysregulated across various cancers while with different directions indicates the complexities of cancer therapy unless the dysregulation of these genes is a passenger event.

A limitation of the present study is that a small number of normal samples in some cancer types such as GBM and CHOL could compromise the validity of the results from the analyses of DE genes between normal and cancer samples. To overcome the limitation, the method of merging samples based on the body sites of cancer initiation (such as brain cancer, lung cancer, gastrointestinal cancer, kidney cancer, blood cancer, etc.) can be used. This is a direction for our future studies.

In addition, many upstream factors may affect expression of mRNAs (genes) in cancers such as gene mutations, DNA copy number alterations, DNA methylation, microRNA expression, and expression change of regulators. Combined analysis of other genomic profiles with gene expression profiles may gain more in-depth insights into the mechanism underlying oncogenesis [89]. Besides, the study of downstream products (proteins) of genes is crucial in cancer research since proteins

directly determine cell function and fate [90]. Integration of different “omics” data to explore oncogenesis in a wide type of cancers represents a promising direction for cancer research.

MATERIALS AND METHODS

Materials

We downloaded RNA-Seq gene expression data (Level 3), and clinical data for all of the 33 cancer types for which data are available from the TCGA data portal (<https://gdc-portal.nci.nih.gov/>). For survival analyses we used clinical data from FireBrowse (<http://gdac.broadinstitute.org/>).

Class comparison to identify differentially-expressed genes

We first normalized the gene expression data by base-2 log transformation, and then identified DE genes between two classes of samples using Student's *t* test. We used FDR to adjust for multiple tests. The FDR was estimated using the Benjamini and Hochberg (BH) method [91]. We used the threshold of FDR < 0.05 and mean gene-expression fold-change > 1.5 to identify the DE genes.

Gene-set enrichment analysis

We performed pathway analysis of gene sets using the GSEA tool [53], and network analysis of gene sets by STRING [55].

Survival analyses

We performed survival analyses of TCGA patients based on gene expression data. Kaplan-Meier survival curves were used to show the survival (OS or DFS) differences between gene higher-expression-level patients and lower-expression-level patients. Gene higher-expression-level and lower-expression-level patients were determined by the median values of gene expression levels. If the gene expression level in a patient was higher than the median value, the patient was classified as gene higher-expression-level; otherwise as gene lower-expression-level. We used the log-rank test to calculate the significance of survival-time differences between two classes of patients with a threshold of P-value < 0.05.

CONCLUSION

The present study provides comprehensive portraits of transcriptional landscape of human cancers, showing extensive intertumor homogeneity and heterogeneity in genomic profiles. This work would bring new insights into the biology of human cancers.

Author contributions

ML performed data analyses. QS performed data analyses. XW conceived the research, designed analysis strategies, performed data analyses, and wrote the manuscript. All the authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments which helped us to improve the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

GRANT SUPPORT

This work was supported by startup funds awarded to XW by the China Pharmaceutical University.

REFERENCES

1. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet.* 2003; 33:238–44.
2. Abba MC, Lacunza E, Butti M, Aldaz CM. Breast cancer biomarker discovery in the functional genomic age: a systematic review of 42 gene expression signatures. *Biomark Insights.* 2010; 5:103–18.
3. Liang Y, Diehn M, Watson N, Bollen AW, Aldape KD, Nicholas MK, Lamborn KR, Berger MS, Botstein D, Brown PO, Israel MA. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA.* 2005; 102:5814–19.
4. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA.* 2003; 100:10393–98.
5. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, Mohsin S, Osborne CK, Chamness GC, Allred DC, O'Connell P. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet.* 2003; 362:362–69.
6. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002; 415:530–36.
7. Burotto M, Chiou VL, Lee JM, Kohn EC. The MAPK pathway across different malignancies: a new perspective. *Cancer.* 2014; 120:3446–56.
8. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74.
9. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA JR, Kinzler KW. Cancer genome landscapes. *Science.* 2013; 339:1546–58.
10. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013; 499:214–18.
11. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng CZ, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet.* 2013; 45:1134–40.
12. Witte T, Plass C, Gerhauser C. Pan-cancer patterns of DNA methylation. *Genome Med.* 2014; 6:66.
13. Yang X, Gao L, Zhang S. Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Brief Bioinform.* 2016; 17:063.
14. Ryerson AB, Ehemann CR, Altekruze SF, Ward JW, Jemal A, Sherman RL, Henley SJ, Holtzman D, Lake A, Noone AM, Anderson RN, Ma J, Ly KN, et al. Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring the increasing incidence of liver cancer. *Cancer.* 2016; 122:1312–37.
15. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, Beijersbergen RL, Bardelli A, Bernards R. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature.* 2012; 483:100–03.
16. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000; 100:57–70.
17. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, Ji HP, Maley CC. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med.* 2016; 22:105–13.
18. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, Miller CA, Welch JS, et al. Mutational landscape and significance across 12 major cancer types. *Nature.* 2013; 502:333–39.
19. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, et al. International network of cancer genome projects. *Nature.* 2010; 464:993–98.
20. Akbani R, Ng PK, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, Liu W, Yang JY, Yoshihara K, Li J, Ling S, Seviour EG, Ram PT, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun.* 2014; 5:3887.
21. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart

- JM, and Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; 45:1113–20.
22. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 7th ed. New York (NY): Springer; 2010.
 23. Attwooll C, Lazzarini Denchi E, Helin K. The E2F family: specific functions and overlapping interests. *EMBO J.* 2004; 23:4709–16.
 24. Ma X, Gao Y, Fan Y, Ni D, Zhang Y, Chen W, Zhang P, Song E, Huang Q, Ai Q, Li H, Wang B, Zheng T, et al. Overexpression of E2F1 promotes tumor malignancy and correlates with TNM stages in clear cell renal cell carcinoma. *PLoS One.* 2013; 8:e73436.
 25. Huang CL, Liu D, Nakano J, Yokomise H, Ueno M, Kadota K, Wada H. E2F1 overexpression correlates with thymidylate synthase and survivin gene expressions and tumor proliferation in non small-cell lung cancer. *Clin Cancer Res.* 2007; 13:6938–46.
 26. Eymin B, Gazzeri S, Brambilla C, Brambilla E. Distinct pattern of E2F1 expression in human lung tumours: E2F1 is upregulated in small cell lung carcinoma. *Oncogene.* 2001; 20:1678–87.
 27. Jiang T, Wang Y, Zhou F, Gao G, Ren S, Zhou C. Prognostic value of high EZH2 expression in patients with different types of cancer: a systematic review with meta-analysis. *Oncotarget.* 2016; 7:4584–97. doi: 10.18632/oncotarget.6612.
 28. Lee SR, Roh YG, Kim SK, Lee JS, Seol SY, Lee HH, Kim WT, Kim WJ, Heo J, Cha HJ, Kang TH, Chung JW, Chu IS, Leem SH. Activation of EZH2 and SUZ12 Regulated by E2F1 Predicts the Disease Progression and Aggressive Characteristics of Bladder Cancer. *Clin Cancer Res.* 2015; 21:5391–403.
 29. Xu H, Xu K, He HH, Zang C, Chen CH, Chen Y, Qin Q, Wang S, Wang C, Hu S, Li F, Long H, Brown M, Liu XS. Integrative Analysis Reveals the Transcriptional Collaboration between EZH2 and E2F1 in the Regulation of Cancer-Related Gene Expression. *Mol Cancer Res.* 2016; 14:163–72.
 30. Bao B, Wang Z, Ali S, Kong D, Banerjee S, Ahmad A, Li Y, Azmi AS, Miele L, Sarkar FH. Over-expression of FoxM1 leads to epithelial-mesenchymal transition and cancer stem cell phenotype in pancreatic cancer cells. *J Cell Biochem.* 2011; 112:2296–306.
 31. Kong FF, Qu ZQ, Yuan HH, Wang JY, Zhao M, Guo YH, Shi J, Gong XD, Zhu YL, Liu F, Zhang WY, Jiang B. Overexpression of FOXM1 is associated with EMT and is a predictor of poor prognosis in non-small cell lung cancer. *Oncol Rep.* 2014; 31:2660–68.
 32. Wen N, Wang Y, Wen L, Zhao SH, Ai ZH, Wang Y, Wu B, Lu HX, Yang H, Liu WC, Li Y. Overexpression of FOXM1 predicts poor prognosis and promotes cancer cell proliferation, migration and invasion in epithelial ovarian cancer. *J Transl Med.* 2014; 12:134.
 33. Zhang HG, Xu XW, Shi XP, Han BW, Li ZH, Ren WH, Chen PJ, Lou YF, Li B, Luo XY. Overexpression of forkhead box protein M1 (FOXM1) plays a critical role in colorectal cancer. *Clin Transl Oncol.* 2016; 18:527–32.
 34. Sala A. B-MYB, a transcription factor implicated in regulating cell cycle, apoptosis and cancer. *Eur J Cancer.* 2005; 41:2479–84.
 35. Chevillard G, Blank V. NFE2L3 (NRF3): the Cinderella of the Cap'n'Collar transcription factors. *Cell Mol Life Sci.* 2011; 68:3337–48.
 36. Cui L, Chen J, Zhang Q, Wang X, Qu J, Zhang J, Dang S. Up-regulation of UHRF1 by oncogenic Ras promoted the growth, migration, and metastasis of pancreatic cancer cells. *Mol Cell Biochem.* 2015; 400:223–32.
 37. Wan X, Yang S, Huang W, Wu D, Chen H, Wu M, Li J, Li T, Li Y. UHRF1 overexpression is involved in cell proliferation and biochemical recurrence in prostate cancer after radical prostatectomy. *J Exp Clin Cancer Res.* 2016; 35:34.
 38. Tut TG, Lim SH, Dissanayake IU, Descallar J, Chua W, Ng W, de Souza P, Shin JS, Lee CS. Upregulated Polo-Like Kinase 1 Expression Correlates with Inferior Survival Outcomes in Rectal Cancer. *PLoS One.* 2015; 10:e0129313.
 39. Ramani P, Nash R, Sowa-Avugrah E, Rogers C. High levels of polo-like kinase 1 and phosphorylated translationally controlled tumor protein indicate poor prognosis in neuroblastomas. *J Neurooncol.* 2015; 125:103–11.
 40. Weichert W, Schmidt M, Gekeler V, Denkert C, Stephan C, Jung K, Loening S, Dietel M, Kristiansen G. Polo-like kinase 1 is overexpressed in prostate cancer and linked to higher tumor grades. *Prostate.* 2004; 60:240–45.
 41. Ito Y, Miyoshi E, Sasaki N, Kakudo K, Yoshida H, Tomoda C, Uruno T, Takamura Y, Miya A, Kobayashi K, Matsuzuka F, Matsuura N, Kuma K, Miyauchi A. Polo-like kinase 1 overexpression is an early event in the progression of papillary carcinoma. *Br J Cancer.* 2004; 90:414–18.
 42. Takahashi T, Sano B, Nagata T, Kato H, Sugiyama Y, Kunieda K, Kimura M, Okano Y, Saji S. Polo-like kinase 1 (PLK1) is overexpressed in primary colorectal cancers. *Cancer Sci.* 2003; 94:148–52.
 43. Liu Z, Sun Q, Wang X. PLK1, A Potential Target for Cancer Therapy. *Transl Oncol.* 2017; 10:22–32.
 44. Kaistha BP, Honstein T, Müller V, Bielak S, Sauer M, Kreider R, Fassan M, Scarpa A, Schmees C, Volkmer H, Gress TM, Buchholz M. Key role of dual specificity kinase TTK in proliferation and survival of pancreatic cancer cells. *Br J Cancer.* 2014; 111:1780–87.
 45. Maia AR, de Man J, Boon U, Janssen A, Song JY, Omerzu M, Sterrenburg JG, Prinsen MB, Willemsen-Seegers N, de Roos JA, van Doornmalen AM, Uitdehaag JC, Kops GJ, et al. Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. *Ann Oncol.* 2015; 26:2180–92.

46. Miao R, Wu Y, Zhang H, Zhou H, Sun X, Csizmadia E, He L, Zhao Y, Jiang C, Miksad RA, Ghaziani T, Robson SC, Zhao H. Utility of the dual-specificity protein kinase TTK as a therapeutic target for intrahepatic spread of liver cancer. *Sci Rep*. 2016; 6:33121.
47. Xie Y, Wang A, Lin J, Wu L, Zhang H, Yang X, Wan X, Miao R, Sang X, Zhao H. Mps1/TTK: a novel target and biomarker for cancer. *J Drug Target*. 2017; 25:112–18.
48. Chung CM, Man C, Jin Y, Jin C, Guan XY, Wang Q, Wan TS, Cheung AL, Tsao SW. Amplification and overexpression of aurora kinase A (AURKA) in immortalized human ovarian epithelial (HOSE) cells. *Mol Carcinog*. 2005; 43:165–74.
49. Kasap E, Gerceker E, Boyacioglu SÖ, Yuceyar H, Yıldırım H, Ayhan S, Korkmaz M. The potential role of the NEK6, AURKA, AURKB, and PAK1 genes in adenomatous colorectal polyps and colorectal adenocarcinoma. *Tumour Biol*. 2016; 37:3071–80.
50. Lucena-Araujo AR, de Oliveira FM, Leite-Cueva SD, dos Santos GA, Falcao RP, Rego EM. High expression of AURKA and AURKB is associated with unfavorable cytogenetic abnormalities and high white blood cell count in patients with acute myeloid leukemia. *Leuk Res*. 2011; 35:260–64.
51. Veerakumarasivam A, Goldstein LD, Saeb-Parsy K, Scott HE, Warren A, Thorne NP, Mills IG, Venkitaraman A, Neal DE, Kelly JD. AURKA overexpression accompanies dysregulation of DNA-damage response genes in invasive urothelial cell carcinoma. *Cell Cycle*. 2008; 7:3525–33.
52. Gharwan H, Groninger H. Kinase inhibitors and monoclonal antibodies in oncology: clinical implications. *Nat Rev Clin Oncol*. 2016; 13:209–27.
53. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50.
54. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, et al. The Reactome pathway Knowledgebase. *Nucleic Acids Res*. 2016; 44:D481–87.
55. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41:D808–15.
56. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J Clin Invest*. 2005; 115:1503–21.
57. Ito T, Kohashi K, Yamada Y, Iwasaki T, Maekawa A, Kuda M, Hoshina D, Abe R, Furue M, Oda Y. Prognostic Significance of Forkhead Box M1 (FOXM1) Expression and Antitumor Effect of FOXM1 Inhibition in Angiosarcoma. *J Cancer*. 2016; 7:823–30.
58. Ning Z, Wang A, Liang J, Liu J, Zhou T, Yan Q, Wang Z. Abnormal expression of Nek2 in pancreatic ductal adenocarcinoma: a novel marker for prognosis. *Int J Clin Exp Pathol*. 2014; 7:2462–69.
59. Yan N, Zhang S, Yang Y, Cheng L, Li C, Dai L, Dai L, Zhang X, Fan P, Tian H, Wang R, Chen X, Su X, et al. Therapeutic upregulation of Class A scavenger receptor member 5 inhibits tumor growth and metastasis. *Cancer Sci*. 2012; 103:1631–39.
60. Huang J, Zheng DL, Qin FS, Cheng N, Chen H, Wan BB, Wang YP, Xiao HS, Han ZG. Genetic and epigenetic silencing of SCARA5 may contribute to human hepatocellular carcinoma by activating FAK signaling. *J Clin Invest*. 2010; 120:223–41.
61. Kropotova ES, Zinovieva OL, Zyryanova AF, Dybovaya VI, Prasolov VS, Beresten SF, Oparina NY, Mashkova TD. Altered expression of multiple genes involved in retinoic acid biosynthesis in human colorectal cancer. *Pathol Oncol Res*. 2014; 20:707–17.
62. Mutka SC, Green LH, Verderber EL, Richards JP, Looker DL, Chlipala EA, Rosenthal GJ. ADH IB expression, but not ADH III, is decreased in human lung cancer. *PLoS One*. 2012; 7:e52995.
63. Marmé A, Zimmermann HP, Moldenhauer G, Schorpp-Kistner M, Müller C, Keberlein O, Giersch A, Kretschmer J, Seib B, Spiess E, Hunziker A, Merchán F, Möller P, et al. Loss of Drop1 expression already at early tumor stages in a wide range of human carcinomas. *Int J Cancer*. 2008; 123:2048–56.
64. Dowdy SC, Gostout BS, Shridhar V, Wu X, Smith DI, Podratz KC, Jiang SW. Biallelic methylation and silencing of paternally expressed gene 3 (PEG3) in gynecologic cancer cell lines. *Gynecol Oncol*. 2005; 99:126–34.
65. Jiang X, Yu Y, Yang HW, Agar NY, Frado L, Johnson MD. The imprinted gene PEG3 inhibits Wnt signaling and regulates glioma growth. *J Biol Chem*. 2010; 285:8472–80.
66. Nye MD, Hoyo C, Huang Z, Vidal AC, Wang F, Overcash F, Smith JS, Vasquez B, Hernandez B, Swai B, Onoko O, Mlay P, Obure J, et al. Associations between methylation of paternally expressed gene 3 (PEG3), cervical intraepithelial neoplasia and invasive cervical cancer. *PLoS One*. 2013; 8:e56325.
67. Chen Y, Xiang H, Zhang Y, Wang J, Yu G. Loss of PCDH9 is associated with the differentiation of tumor cells and metastasis and predicts poor survival in gastric cancer. *Clin Exp Metastasis*. 2015; 32:417–28.
68. Wang C, Yu G, Liu J, Wang J, Zhang Y, Zhang X, Zhou Z, Huang Z. Downregulation of PCDH9 predicts prognosis for patients with glioma. *J Clin Neurosci*. 2012; 19:541–45.
69. Zhu P, Lv J, Yang Z, Guo L, Zhang L, Li M, Han W, Chen X, Zhuang H, Lu F. Protocadherin 9 inhibits epithelial-mesenchymal transition and cell migration through

- activating GSK-3 β in hepatocellular carcinoma. *Biochem Biophys Res Commun.* 2014; 452:567–74.
70. Pavlides S, Tsirigos A, Migneco G, Whitaker-Menezes D, Chiavarina B, Flomenberg N, Frank PG, Casimiro MC, Wang C, Pestell RG, Martinez-Outschoorn UE, Howell A, Sotgia F, Lisanti MP. The autophagic tumor stroma model of cancer: role of oxidative stress and ketone production in fueling tumor cell metabolism. *Cell Cycle.* 2010; 9:3485–505.
 71. Montgomery C, Pei Z, Watkins PA, Mizioro HM. Identification and characterization of an extramitochondrial human 3-hydroxy-3-methylglutaryl-CoA lyase. *J Biol Chem.* 2012; 287:33227–36.
 72. Glunde K, Bhujwalla ZM, Ronen SM. Choline metabolism in malignant transformation. *Nat Rev Cancer.* 2011; 11:835–48.
 73. Shepherd JH, Uray IP, Mazumdar A, Tsimelzon A, Savage M, Hilsenbeck SG, Brown PH. The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget.* 2016; 7:13106–21. doi: 10.18632/oncotarget.7437.
 74. Weigle B, Ebner R, Temme A, Schwind S, Schmitz M, Kiessling A, Rieger MA, Schackert G, Schackert HK, Rieber EP. Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas. *Oncol Rep.* 2005; 13:139–44.
 75. Xu H, Cai T, Carmona GN, Abuhatzira L, Notkins AL. Small cell lung cancer growth is inhibited by miR-342 through its effect of the target gene IA-2. *J Transl Med.* 2016; 14:278.
 76. Deb TB, Zuo AH, Barndt RJ, Sengupta S, Jankovic R, Johnson MD. PncK overexpression in HER-2 gene-amplified breast cancer causes Trastuzumab resistance through a paradoxical PTEN-mediated process. *Breast Cancer Res Treat.* 2015; 150:347–61.
 77. Shi Z, Wu D, Tang R, Li X, Chen R, Xue S, Zhang C, Sun X. Silencing of HMGA2 promotes apoptosis and inhibits migration and invasion of prostate cancer cells. *J Biosci.* 2016; 41:229–36.
 78. Wu J, Zhang S, Shan J, Hu Z, Liu X, Chen L, Ren X, Yao L, Sheng H, Li L, Ann D, Yen Y, Wang J, Wang X. Elevated HMGA2 expression is associated with cancer aggressiveness and predicts poor outcome in breast cancer. *Cancer Lett.* 2016; 376:284–92.
 79. Zhao XP, Zhang H, Jiao JY, Tang DX, Wu YL, Pan CB. Overexpression of HMGA2 promotes tongue cancer metastasis through EMT pathway. *J Transl Med.* 2016; 14:26.
 80. Ricke RM, Jeganathan KB, van Deursen JM. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *J Cell Biol.* 2011; 193:1049–64.
 81. Koelzer VH, Huber B, Mele V, Iezzi G, Trippel M, Karamitopoulou E, Zlobec I, Lugli A. Expression of the hyaluronan-mediated motility receptor RHAMM in tumor budding cells identifies aggressive colorectal cancers. *Hum Pathol.* 2015; 46:1573–81.
 82. Peña-Llopis S, Wan Y, Martinez ED. Unique epigenetic gene profiles define human breast cancers with poor prognosis. *Oncotarget.* 2016; 7:85819–31. doi: 10.18632/oncotarget.13334.
 83. Chen XQ, He JR, Wang HY. Decreased expression of ALDH1L1 is associated with a poor prognosis in hepatocellular carcinoma. *Med Oncol.* 2012; 29:1843–49.
 84. Krupenko SA, Oleinik NV. 10-formyltetrahydrofolate dehydrogenase, one of the major folate enzymes, is down-regulated in tumor tissues and possesses suppressor effects on cancer cells. *Cell Growth Differ.* 2002; 13:227–36.
 85. Ramachandran I, Thavathiru E, Ramalingam S, Natarajan G, Mills WK, Benbrook DM, Zuna R, Lightfoot S, Reis A, Anant S, Queimado L. Wnt inhibitory factor 1 induces apoptosis and inhibits cervical cancer growth, invasion and angiogenesis *in vivo*. *Oncogene.* 2012; 31:2725–37.
 86. Ramachandran I, Ganapathy V, Gillies E, Fonseca I, Sureban SM, Houchen CW, Reis A, Queimado L. Wnt inhibitory factor 1 suppresses cancer stemness and induces cellular senescence. *Cell Death Dis.* 2014; 5:e1246.
 87. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28:27–30.
 88. Wang X, Sun Q. TP53 mutations, expression and interaction networks in human cancers. *Oncotarget.* 2017; 8:624–43. doi: 10.18632/oncotarget.13483.
 89. Wang X, Guda C. Integrative exploration of genomic profiles for triple negative breast cancer identifies potential drug targets. *Medicine (Baltimore).* 2016; 95:e4321.
 90. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, Davies SR, Wang S, Wang P, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature.* 2014; 513:382–87.
 91. Benjami Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995; 57:289–300.