Pan-cancer methylation and expression profiling of adenocarcinomas revealed epigenetic silencing in the WNT signaling pathway

J. LI[1,‡], K. L. HUANG[2,‡], T. ZHANG[1], H. LI[1,3], J. ZHAO[1,4,5], H. WANG[1,4,5,*]

[1]International Joint Cancer Institute, the Second Military Medical University, Shanghai, 200433, China; [2]The Genome Institute, Washington University in St. Louis, Missouri 63108, USA; [3]Tumor Immunology and Gene Therapy Center, Eastern Hepatobiliary Surgery Hospital affiliated to the Second Military Medical University, Shanghai 200433, China; [4]State Key Laboratory of Antibody Medicine and Targeted Therapy, Shanghai Key Laboratory of Cell Engineering, Shanghai, China; [5]PLA General Hospital Cancer Center; PLA Postgraduate School of Medicine, Beijing, China

*Correspondence: wanghao@smmu.edu.cn
‡Contributed equally to this work.

Adenocarcinomas are tumors of glandular characteristics. While tissues of common origins have been known to undergo similar epigenetic changes, it is unclear whether adenocarcinomas of different cancer types would exhibit similar DNA methylation and epigenetic regulation profiles. Herein, we studied global methylation and mRNA expression levels in 1214 lung, prostate, colon, and rectal cancer samples from The Cancer Genome Atlas (TCGA). We identified 602 candidate epigenetically silenced genes shared across these cancer types, and 835 associated CpG sites. The shared candidate genes are enriched in developmental processes. Specifically, 15 of these genes were found in the WNT signaling pathway (enrichment test p-value=1.53×10^{-6}). Notably, the subset of silenced WNT pathway genes in each sample may be different, and both WNT activating or inhibiting genes could be suppressed. Clustering analysis showed that each tumor type contained a similar hyper-methylated subset of samples showing strong epigenetic silencing in the WNT pathway genes, and other fractions of samples expressing subset of the genes. Overall, our results showed that aberration in epigenetic regulation of the WNT signaling pathway is a common signature in adenocarcinomas.

Key words: methylation, epigenetics, adenocarcinoma, WNT signaling

Adenocarcinoma is defined as tumors that occur in tissues of glandular origins or characteristics. Prevalent types of adenocarcinomas include lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), as well as colon and rectal adenocarcinoma (CRAD). TCGA and other studies has conducted molecular profiling study of LUAD [1] and CRAD[2,3] separately, and discovered the genomic and epigenetic aberrations that may have driven each cancer type. These studies identified distinct methylation subtypes within each cancer type and found hypermethylation signatures within some samples. For example, a subset of both colon and rectal cancers were hyper-methylated and was characterized as CpG island methylator phenotype-high (CIMP-H) and the other subset as CIMP-low (CIMP-L), with CIMP-H tumors showing hypermutation in MLH1. In lung cancer, there were largely three subsets, CIMP-H, CIMP-L and the other intermediate group. CIMP-H lung tumors showed DNA hypermethylation in CD-KN2A, GATA2, GATA4 among other genes. Methylation and epigenetic regulation of cancer genes are considered to largely contribute to tumorigenesis, and identifying these key events would advance our understanding of tumor etiology [4].

Despite the known signatures in each cancer type, it is unclear whether different types of adenocarcinoma share defining epigenetic signature. Shared signatures within adenocarcinoma would be of great interest, as they present opportunities to develop therapies that may be applied to treat multiple cancer types. The pan-cancer initiative of TCGA aims to gain additional insights from jointly analyzing tumor profiles from different cancer types, and has resulted in many fruitful genomic findings [5]. However, relatively few studies have investigated methylation at a pan-cancer level. By reviewing methylomes across cancer types from single-tumor-type studies, Witte et al.[6] pinpointed some shared and distinct methylation signatures between cancer types.

Recently, Gevaert et al. [7] developed MethylMix and evaluated differentially methylated genes across 12 cancer types. They identified 10 major clusters through the pan-cancer analysis, and suggested there may be similar epigenetic regulation on certain cancer types.

DNA methylation is highly involved in developmental processes and has been widely established to be dependent on tissue types and origins [8]. However, to our knowledge, no TCGA pan-cancer studies have combined and examined methylome according to their tissue similarities. In this study, we hypothesized that different types of adenocarcinoma may share methylation patterns that drive tumorigenesis. We conducted a comprehensive analysis on three types of common adenocarcinoma in the TCGA: LUAD, PRAD and CRAD. We identified enrichment of likely epigenetic silenced genes in developmental processes and the WNT signaling pathway, and dissected the epigenetic regulation landscape in associated WNT pathway genes across samples from different types of adenocarcinomas.

Materials and methods

Samples and methylation, RNA expression data. Level 3 DNA methylation data of LUAD (451), CRAD (393) and PRAD (498) was downloaded from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/). We extracted the methylation β-value as methylation level for downstream analysis.

Processed mRNA-seq data of LUAD (451), CRAD (370) and PRAD (498) was downloaded from the same source. RPKM (Reads Per Kilobase per Million mapped reads) values that quantify at the gene level were calculated for further analysis in the R 2.15 environment. RPKM for a given Gene1 is calculated by: (raw read counts × 10^9) / (total reads × length of Gene1). We used gene level read counts to calculate RPKM.

421 LUAD, 314 CRAD, and 479 PRAD samples, or a total of 1214 adenocarcinoma samples, have both DNA methylation and expression data and were used in the final analysis. We merged colon and rectal adenocarcinomas into one group, CRAD, as molecular profiles of these two showed no distinguishing difference [3].

Candidate functional CpG sites and epigenetically silenced genes. We determined candidate functional CpG sites and epigenetically silenced genes in all three cancer types by adopting a previously developed strategy [9] .The methylation status of CpG site and its associated gene's expression level is evaluated jointly. We defined four criteria each with a relaxed threshold and a stringent threshold. We required candidate epigenetically silenced genes to pass all 4 relaxed thresholds, and at least 3 out of 4 stringent thresholds: (1) The mean DNA methylation β-value in non-tumor adjacent colonic tissue < 0.5 (relaxed) and < 0.4 (stringent); (2) The difference in DNA methylation β-value between the 90th percentile tumor and mean adjacent-normal > 0.1 (relaxed) and > 0.3 (stringent);

(3) The fold expression change between mean adjacent-normal and mean of the 10% of tumor samples with the highest DNA methylation > 1.5 (relaxed) and > 3 (stringent); (4) Spearman's correlation coefficient between DNA methylation and gene expression calculated jointly across tumor and adjacent- normal tissue samples < –0.2 (relaxed) and < –0.3 (stringent).

Location analysis. CpG islands, RefSeq gene coding loci feature were all downloaded from the UCSC Genome Browser (https://genome.ucsc.edu/), and introns, exons were defined according to the coordinates. We defined the promoter region as 500bp upstream of transcription start site to 500bp downstream of transcription start site.

Enrichment analysis. Enrichment of biological processes was evaluated through the Gene Ontology website (http://geneontology.org). We retained all biological processes where the enrichment FDR is below 0.01. Enrichment of biological pathways was determined through the InnateDB Pathway Analysis (http://www.innatedb.ca/index.jsp). We used the Chi Square test and Bonferroni p-value correction. We retained all pathways where the enrichment FDR is below 0.01.

Genome browser data assembly. We calculated the mean methylation value of samples in each cancer type, and then upload these to WashU Epigenome Browser (http://epigenomegateway.wustl.edu/) and captured in the methylation values in the display track.

Clustering of epigenetically silenced genes in the WNT pathway. We assessed the DNA methylation and mRNA expression of each candidate epigenetically silenced genes in the WNT signaling pathway in each sample. Then we picked the probes with the highest absolute Spearman's Rho to assess the methylation level of the chosen gene. We conducted k-means clustering (K=2) based on the two-dimensional space of DNA methylation and gene expression data to classify the epigenetically silenced group and non-epigenetically silenced group of sample. Then, according to the gene silencing status, we performed hierarchical clustering to find samples and WNT pathway genes with similar epigenetic regulation profiles.

Statistical analysis. All analysis on candidate functional CpG sites and epigenetically silenced genes were conducted in the R 2.15 environment. Enrichment tests were conducted through web implementations of Gene Ontology Analysis (http://geneontology.org) and InnateDB Pathway Analysis (http://www.innatedb.ca/index.jsp).

Results

Candidate epigenetically silenced genes and functional CpG sites shared across all three types of adenocarcinomas. Through analyzing methylation and RNA expression data with pre-defined criteria (see Material and Methods), we identified 1752, 2246, and 2621 candidate epigenetically silenced genes in LUAD, CRAD and PRAD, respectively (Figure 1C). These genes exhibited high levels of methylation that correlated with low mRNA expressions. 602 of these genes were shared between all 3 cancer types, suggesting a sizable extent of
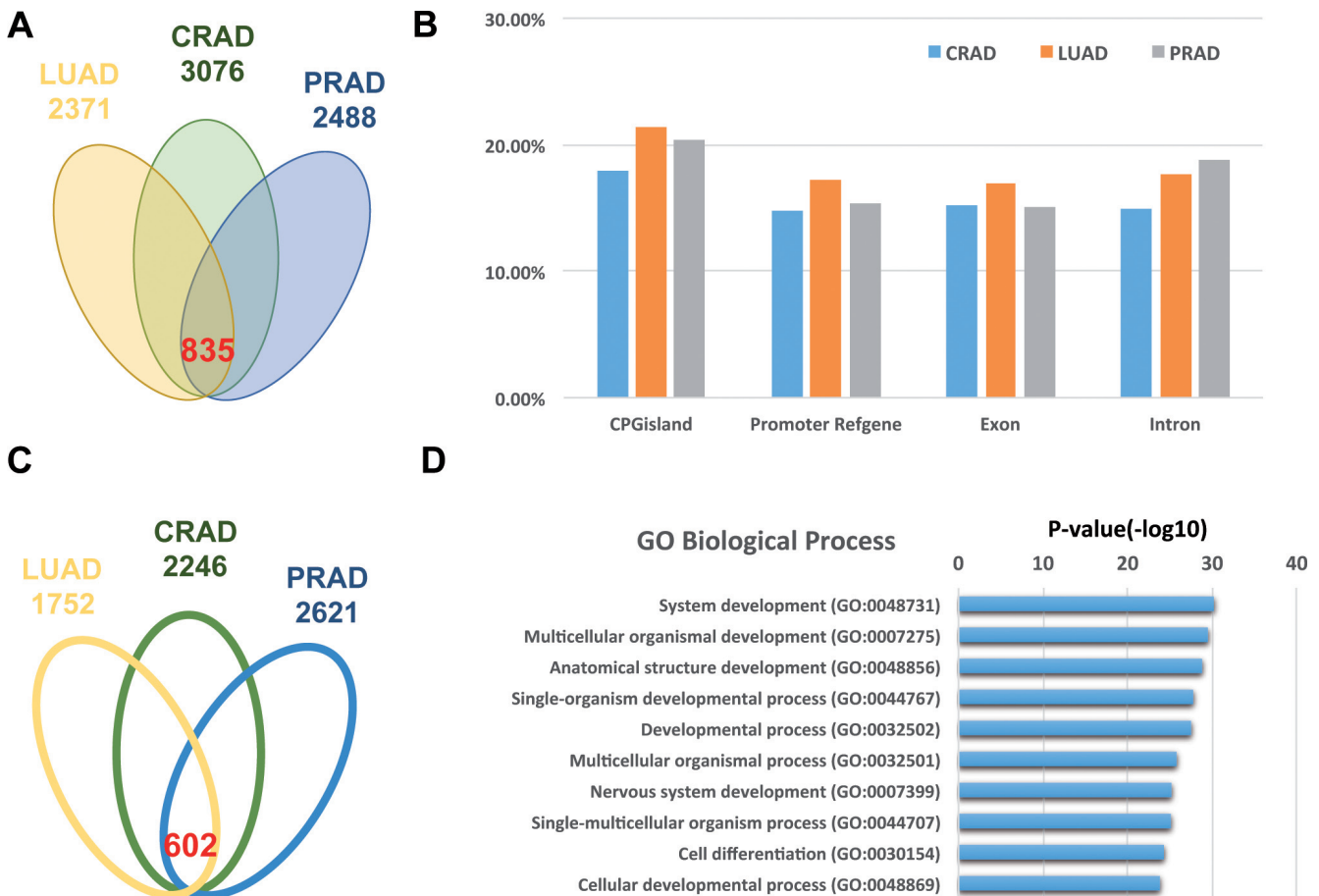
Figure 1. Landscape of candidate functional CpG sites and epigenetically silenced genes in adenocarcinoma. (A) Number of candidate functional CpG sites associated with candidate epigenetically silenced genes in each cancer types. 835 CpGs are shared in all three adenomcarcinoma. **(B)** Distribution of candidate functional CpG sites across genomic features. **(C)** Number of candidate epigenetically silenced genes in each cancer types. 602 genes are shared in all three adenomcarcinoma. **(D)** Enrichment of biological processes of 602 shared candidate epigenetically silenced genes in adenocarcinoma (X-axis denotes negative log10-transformed p-value).

similarity between epigenetic regulation in adenocarcinoma (Supplemental table 1).

Then we identified 2371, 3076, and 2488 candidate functional CpG sites associated with the candidate epigenetically silenced genes in LUAD, CRAD and PRAD, respectively (Figure 1A and Supplemental table 2). Most of the candidate functional CpG sites resided in CpG islands. Promoter, exon, and intron regions each harbor around 15% to 20% of these sites, and the distributions between these three cancer types were consistent (Figure 1B). This indicated that CpG sites in gene regions other than the promoter could also regulate gene expression. Additionally, this analysis confirmed that we are not observing batch effect in individual cancer type, and gave us confidence to pursue further analysis on the shared epigenetically silenced gene set.

Enrichment analyses of candidate epigenetically silenced genes. Through gene ontology analysis, the 602 candidate epigenetically silenced genes were found to be enriched in

multiple development-related biological processes (Figure 1D). The top enriched processes include system development, multicellular organismal development, and anatomical structure development. This indicated that the genes in these developmental programs were likely suppressed through epigenetic regulation in adenocarcinomas.

Then we conducted pathway enrichment analysis and found the 602 genes to be enriched in the KEGG basal cell carcinoma pathway ($p=7.58 \times 10^{-7}$) and WNT signaling pathway ($p=1.53 \times 10^{-6}$) at FDR levels lower than 1×10^{-3} (Supplemental table 3). Basal cell carcinoma involved mis-regulation of the WNT signaling pathway, and those that were present in the WNT signaling pathway largely account for the enrichment of 602 genes in the basal cell carcinoma pathway. This suggested that epigenetic silencing of the WNT pathway components may contribute to adenocarcinomas. Interestingly, the epigenetically repressed genes include WNT pathway activators, such as WNT3A, WNT10A, AXIN2, as well as negative
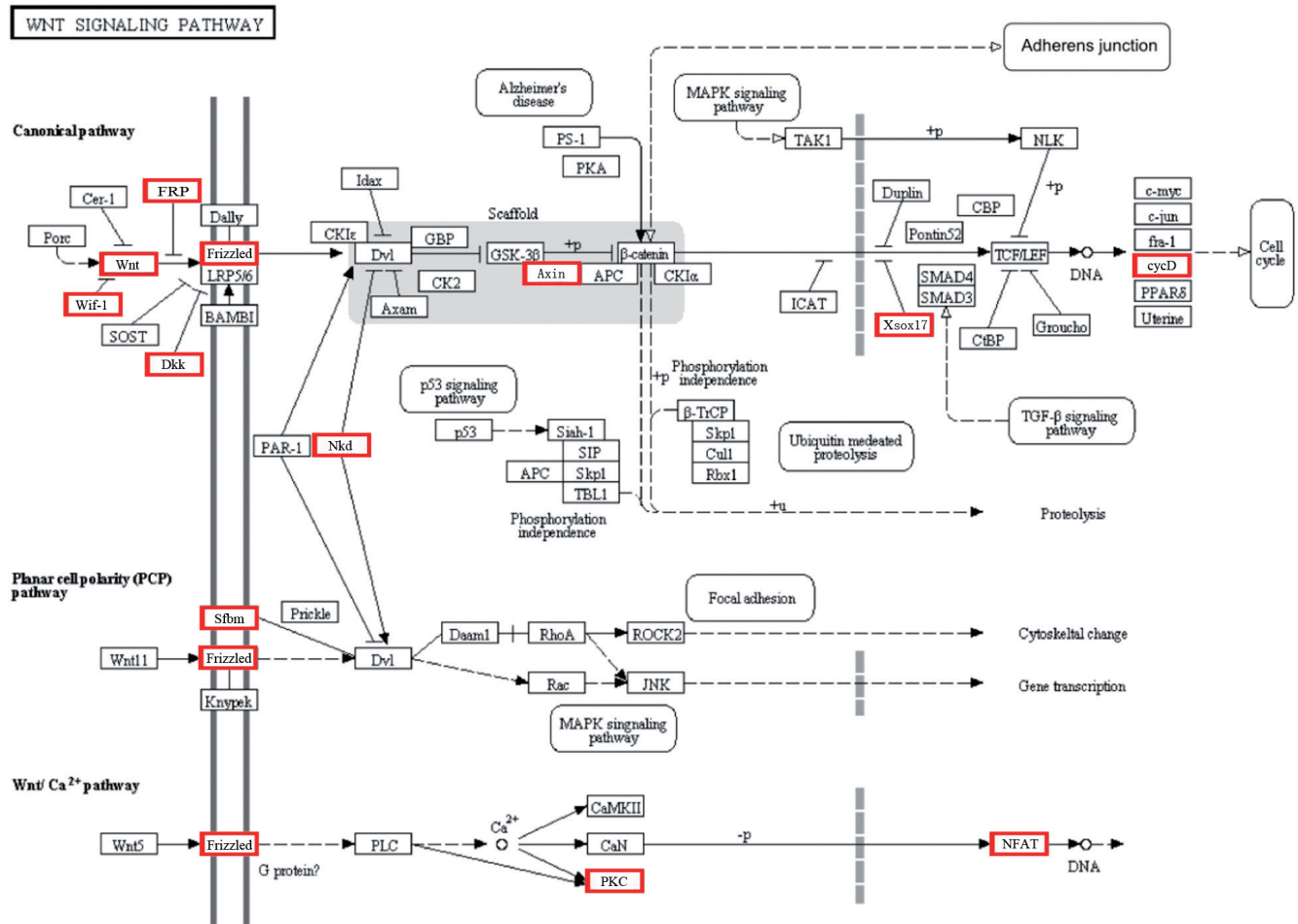
Figure 2. The KEGG WNT Signaling Pathway (hsa04310) was enriched for candidate epigenetically silenced genes. The candidate epigenetically silenced genes in the pathway were colored in red.

regulators, such as KDD1, SFRP1, and WIF1 (Figure 2). While the methylation and expression of the WNT pathway genes were largely mis-regulated in adenocarcinomas, the mechanism may not be a uniform activation or repression of the pathway in all tumors.

Functional CpG sites and methylation levels in epigenetically silenced genes. While the candidate epigenetically silenced genes overlapped between three types of adenocarcinomas, it was unclear whether the candidate functional CpG sites and the pattern of methylation across these genes would be consistent across all three tumor-types. Specifically, we examined methylation patterns of three WNT pathway genes that showed similar or different extent of epigenetic silencing across cancer types (Supplemental table 4): PRKCB (epigenetically-silenced in 75.8% of LUAD, 85.6% of PRAD, and 93.6% of COAD samples), NKD1 (epigenetically-silenced in 99.8% of LUAD, 99.8% of PRAD, and 84.7% of COAD samples), and FZD10 (epigenetically-silenced in 97.2% of LUAD, 80.0% of PRAD, and 95.9% of COAD samples). We

focused on the promoter region, and examined methylation levels at sites that are within -3kb to +3kb of the transcription start site (TSS).

Each gene showed different extents of shared methylation and functional CpG site profiles across 3 cancer types (Figure 3). In the promoter region of PRKCB, there were consistently multiple adjacent CpG sites where the level of methylation in tumor was higher than that in adjacent normal tissue. And the candidate functional CpG site (shaded in yellow) was consistent across LUAD, PRAD, CRAD and their corresponding controls. For NKD1, there was only one obvious differentially methylated CpG site that was also determined to be a candidate functional site. Like PRKCB, the pattern of this site was also consistent across three tumor types. In FZD10, interestingly, the methylation patterns varied across tumor types. LUAD and CRAD shared multiple candidate functional CpG sites that were not significantly differentially methylated in PRAD upstream of TSS. Thus, while the candidate epigenetically silenced genes may be
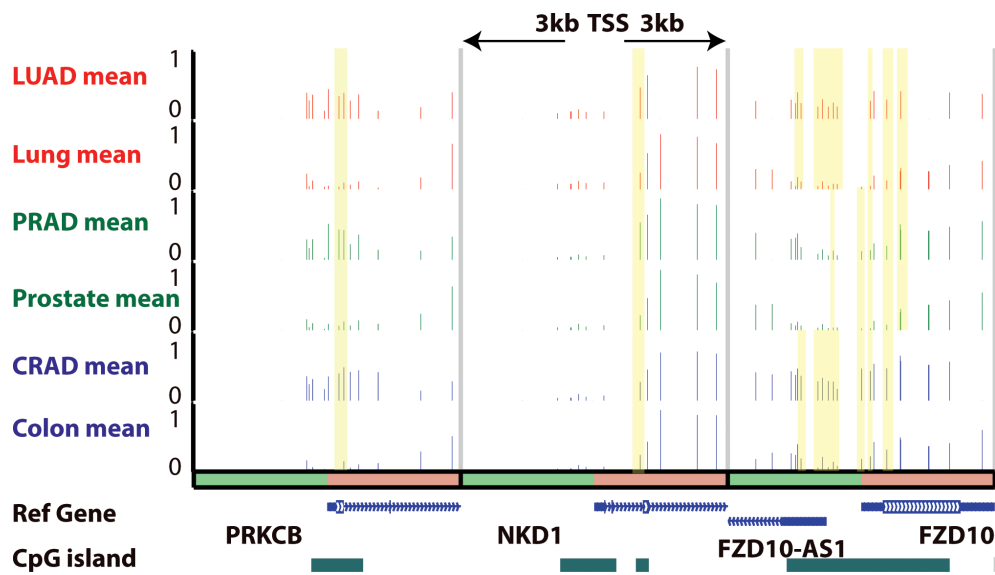
Figure 3. Landscape of methylation in three epigenetically silenced WNT pathway genes in the WNT pathway in the WashU Epigenome Browser. The tracks showed average methylation beta-values across LUAD, PRAD, CRAD and those in corresponding lung, prostate, and colon and rectal control tissues. Candidate functional CpG sites were highlighted by yellow shading.

shared in three adenocarcinomas, the likely functional CpG sites and underlying epigenetic regulation may still be different depending on the tumor type.

Patient-specific methylation and expression signatures in genes of the WNT signaling pathway. In addition to cross-tumor type differences, we were interested in whether there would be differences in methylation and epigenetic regulation of the WNT signaling pathway between different samples of the same tissue type. We conducted a k-means (k=2) clustering on the two-dimensional space of DNA methylation and gene expression to define whether each WNT pathway gene was epigenetically repressed or expressed in each sample. Then, we carried out hierarchical clustering to determine which samples and genes showed similar epigenetic regulation signatures.

Interestingly, the samples did not cluster solely based on tumor type (Figure 4). Each tumor type contained a fraction
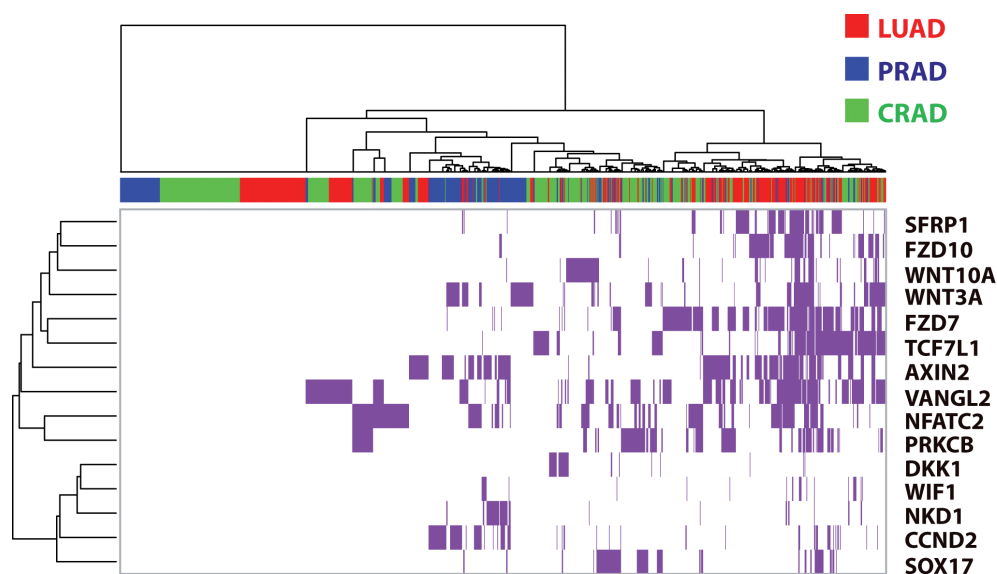

Figure 4. Unsupervised clustering of expression and epigenetic silencing patterns of the WNT pathway genes in 3 types of adenocarcinomas. Each column represents a sample, and red, blue, and green color represents LUAD, PRAD, and CRAD respectively. The white color represents epigenetically silenced expression, whereas the purple color represents low methylation and gene expression.

of samples showing strong epigenetic silencing in all of the WNT pathway genes, and other fractions of samples expressing subset of the WNT pathway genes. This was consistent with TCGA's previous single-tumor type study on LUAD [1] and CRAD [3] where a CIMP-H and a CIMP-L groups were defined. Our result not only validated this finding in the WNT signaling pathway, but also demonstrated that CIMP-H tumors may resemble CIMP-H tumors of different adenocarcinomas to a higher extent than they resemble CIMP-L tumors of the same tumor type. The samples that expressed a subset of the WNT pathway genes showed varied combination of epigenetic silencing and expression across WNT pathway genes. Thus, the epigenetic regulation of each of these adenocarcinomas may be different.

On the gene level, each of the WNT pathway genes displayed different methylation profiles across samples (Figure 4). Of note, we identified a cluster of genes--- including DKK1, WIF1, and NKD1--- that were epigenetically silenced in almost all adenocarcinomas. These three genes were all negative regulators of the WNT signaling pathway (Figure 2), implicating repressing the negative regulation of the WNT pathway may contribute to adenocarcinoma. However, multiple key activating components of the WNT signaling pathway were also significantly repressed in many samples. Epigenetic regulation of the WNT pathway genes may be complex and patient-specific.

Discussion

While methylation has been investigated in cohorts of single or several tumor types [7], methylation in tissues of common developmental origins, such as the glandular tissues, have not been examined globally. To our knowledge, this is the first study to systematically evaluate the methylation and epigenetic silencing landscape across multiple types of adenocarcinoma. We successfully identified 602 epigenetically silenced genes in adenocarcinoma through analyzing methylation and mRNA expression data across 1214 samples of LUAD, PRAD, and CRAD (Figure 1A). These genes were significantly enriched in developmental processes (Figure 1D), implicating the importance of epigenetic repression of certain developmental programs in tumorigenesis. While somatic mutations have been widely considered as main drivers of tumors, amounting evidence has suggested that tumorigenesis is largely affected by surrounding environments and driven by aberrant epigenetic regulation during development [10,11] Our results provided supporting evidence that epigenetic aberration in development may be associated with adenocarcinoma.

The WNT signaling pathway is essential in development and controls cell growth and survival [12]. Previous studies have shown that mutations or mis-regulation of WNT pathway genes may contribute to tumorigenesis in multiple cancer types [13,14,15]. We identified WNT pathway as one of the most enriched pathways for our candidate epigenetically silenced genes; out of the 602 genes we identified, 15 genes

were in the WNT pathway (Supplemental table 3, enrichment test p-value=1.53 x10^{-6}). Some of these genes, including SFRP1, SOX17 and WIF1, have previously been shown to be epigenetically silenced in LUAD [1]. Our result suggested that epigenetic silencing of these genes might not be limited to LUAD, but common to multiple adenocarcinomas.

Beyond what previous studies have found, we identified both activating components and negative regulators of the WNT pathway that were epigenetically silenced in adenocarcinoma (Figure 2), indicating that the pathway was neither uniformly activated nor suppressed in tumors. Cancer stem cells (CSC) are thought to drive tumorigenesis and progression. They have the ability to cell-renew and give rise to malignant tumor cells [16] CSCs may require an intricate balance in the WNT pathway activity: on the one hand, they may require WNT activation to trigger self-renewal [17]; on the other hand, they may need to inhibit WNT pathway to suppress developmental processes to retain their stem-cell state [18]. This could potentially explain why we observed methylation and epigenetic silencing in known WNT inhibitors, such as SFRP1, WIF1, and DKK1, as well as WNT activating components, such as WNT3A, FZD10, and PRKCB, in adenocarcinomas.

Since CSCs play a crucial role in tumor expansion, cancer treatments suppressing CSCs could potentially deter tumor relapse and metastasis. Multiple therapeutic strategies targeting the WNT signaling pathway in CSCs are under active development [19,20]. Given the differences in the epigenetic profiles of the WNT pathway genes between samples (Figure 4), it will be important to examine methylation of key targeted genes before devising treatment. Hierarchical clustering of methylation and expression of WNT pathway genes showed that each of the tumor types had a CIMP-H and CIMP-L cluster. This is consistent with earlier TCGA finding in LUAD [1] and CRAD [3]. Additionally, our clustering analysis showed that CIMP-H samples of different cancer types may be more similar to each other than CIMP-L samples of the same cancer type. Thus, WNT-targeting treatments that were developed for CIMP-H samples of a certain adenocarcinoma may also be applied to CIMP-H samples of another adenocarcinoma. Overall, while this study provides strong evidence showing that aberrant methylation and epigenetic regulation of the WNT signaling pathway genes associate with adenocarcinomas, the detailed mechanisms still require further elucidation. The functional consequences of these epigenetic changes would also need to be experimentally validated before this knowledge can be applied to developing WNT-targeting therapies in adenocarcinoma.

References

[1] CANCER GENOME ATLAS RESEARCH N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014; 511: 543–550. http://dx.doi.org/10.1038/nature13385

[2] HINOUE T, WEISENBERGER DJ, LANGE CP, SHEN H, BYUN HM et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res 2012; 22: 271–282. http://dx.doi.org/10.1101/gr.117523.110

[3] CANCER GENOME ATLAS N. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012; 487: 330–337. http://dx.doi.org/10.1038/nature11252

[4] CLARK SJ, MELKI J. DNA methylation and gene silencing in cancer: which is the guilty party?. Oncogene 2002; 21: 5380–5387. http://dx.doi.org/10.1038/sj.onc.1205598

[5] CANCER GENOME ATLAS RESEARCH N, WEINSTEIN JN, COLLISSON EA, MILLS GB, SHAW KR et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013; 45: 1113–1120. http://dx.doi.org/10.1038/ng.2764

[6] WITTE T, PLASS C, GERHAUSER C. Pan-cancer patterns of DNA methylation. Genome Med 2014; 6: 66. http://dx.doi.org/10.1186/s13073-014-0066-6

[7] GEVAERT O, TIBSHIRANI R, PLEVRITIS SK. Pancancer analysis of DNA methylation-driven genes using MethylMix. Genome Biol 2015; 16: 17. http://dx.doi.org/10.1186/s13059-014-0579-8

[8] MOORE LD, LE T, FAN G. DNA methylation and its basic function. Neuropsychopharmacology 2013; 38: 23–38. http://dx.doi.org/10.1038/npp.2012.112

[9] CANCER GENOME ATLAS RESEARCH N. Integrated genomic analyses of ovarian carcinoma. Nature 2011; 474: 609–615. http://dx.doi.org/10.1038/nature10166

[10] SONNENSCHEIN C, SOTO AM, RANGARAJAN A, KULKARNI P. Competing views on cancer. J Biosci 2014; 39: 281–302. http://dx.doi.org/10.1007/s12038-013-9403-y

[11] BAKER SG. A cancer theory kerfuffle can lead to new lines of research. J Natl Cancer Inst 2015; 107.

[12] CADIGAN KM, NUSSE R. Wnt signaling: a common theme in animal development. Genes Dev 1997; 11: 3286–3305. http://dx.doi.org/10.1101/gad.11.24.3286

[13] Polakis P. Wnt signaling and cancer. Genes Dev 2000; 14: 1837–1851.

[14] KLARMANN GJ, DECKER A, FARRAR WL. Epigenetic gene silencing in the Wnt pathway in breast cancer. Epigenetics 2008; 3: 59–63. http://dx.doi.org/10.4161/epi.3.2.5899

[15] REYA T, CLEVERS H. Wnt signalling in stem cells and cancer. Nature 2005; 434: 843–850. http://dx.doi.org/10.1038/nature03319

[16] LOBO NA, SHIMONO Y, QIAN D, CLARKE MF. The biology of cancer stem cells. Annu Rev Cell Dev Biol 2007; 23: 675–699. http://dx.doi.org/10.1146/annurev.cellbio.22.010305.104154

[17] TAIPALE J, BEACHY PA. The Hedgehog and Wnt signalling pathways in cancer. Nature 2001; 411: 349–354. http://dx.doi.org/10.1038/35077219

[18] HSIEH IS, CHANG KC, TSAI YT, KE JY, LU PJ et al. MicroRNA-320 suppresses the stem cell-like characteristics of prostate cancer cells by downregulating the Wnt/beta-catenin signaling pathway. Carcinogenesis 2013; 34: 530–538. http://dx.doi.org/10.1093/carcin/bgs371

[19] TAKEBE N, HARRIS PJ, WARREN RQ, IVY SP. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. Nat Rev Clin Oncol 2011; 8: 97–106. http://dx.doi.org/10.1038/nrclinonc.2010.196

[20] ANASTAS JN, MOON RT. WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer 2013; 13: 11–26.