Comparative transcriptomes of adenocarcinomas and squamous cell carcinomas reveal molecular similarities that span classical anatomic boundaries

Eric W. Lin[1,2], Tatiana A. Karakasheva[1,2], Dong-Jin Lee[3], Ju-Seog Lee[3], Qi Long[4], Adam J. Bass[5], Kwok K. Wong[6], Anil K. Rustgi[1,2,7] *

1 Department of Medicine, Division of Gastroenterology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 2 Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 3 Department of Systems Biology, Division of Cancer Medicine, UT MDACC, Houston, Texas, United States of America, 4 Department of Biostatistics, Epidemiology and Bioinformatics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 5 Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts, United States of America, 6 Department of Medicine, Division of Hematology and Medical Oncology, New York University Langone Medical Center, New York, New York, United States of America, 7 Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

* anil2@mail.med.upenn.edu

Abstract

Advances in genomics in recent years have provided key insights into defining cancer subtypes "within-a-tissue"—that is, respecting traditional anatomically driven divisions of medicine. However, there remains a dearth of data regarding molecular profiles that are shared across tissues, an understanding of which could lead to the development of highly versatile, broadly applicable therapies. Using data acquired from The Cancer Genome Atlas (TCGA), we performed a transcriptomics-centered analysis on 1494 patient samples, comparing the two major histological subtypes of solid tumors (adenocarcinomas and squamous cell carcinomas) across organs, with a focus on tissues in which both subtypes arise: esophagus, lung, and uterine cervix. Via principal component and hierarchical clustering analysis, we discovered that histology-driven differences accounted for a greater degree of inherent molecular variation in the tumors than did tissue of origin. We then analyzed differential gene expression, DNA methylation, and non-coding RNA expression between adenocarcinomas and squamous cell carcinomas and found 1733 genes, 346 CpG sites, and 42 microRNAs in common between organ sites, indicating specific adenocarcinoma-associated and squamous cell carcinoma-associated molecular patterns that were conserved across tissues. We then identified specific pathways that may be critical to the development of adenocarcinomas and squamous cell carcinomas, including Liver X receptor activation, which was upregulated in adenocarcinomas but downregulated in squamous cell carcinomas, possibly indicating important differences in cancer cell metabolism between these two histological subtypes of cancer. In addition, we highlighted genes that may be common drivers of adenocarcinomas specifically, such as *IGF2BP1*, which suggests a possible link between embryonic development and tumor subtype. Altogether, we demonstrate the need to consider

biological similarities that transcend anatomical boundaries to inform the development of novel therapeutic strategies. All data sets from our analysis are available as a resource for further investigation.

Author summary

In clinical practice, the organ in which a cancer arises typically classifies it. However, developments in our understanding of cancer have revealed that this method overlooks key aspects of cancer biology relevant to both disease prevention and treatment. In fact, work characterizing the genetic make-up of cancers arising in a single organ has revealed that a shared organ of origin does not necessarily imply biological similarity (i.e. not all lung cancers share similar biological and molecular properties). While this approach, known as "within-a-tissue subtyping," identifies key differences between cancers that arise in a single organ, a broader perspective may highlight important biological similarities between cancers across organs. Here we utilize this second approach, or "across-tissue subtyping," to gain insight into similarities between cancers (of different organs) that share the same histology—or appear similarly under a microscope. Using publicly available data from The Cancer Genome Atlas (TCGA), we compare gene expression of two major classes of solid tumors—adenocarcinomas (which arise from cells that form glands) and squamous cell carcinomas (which arise from flattened cells that form physical barriers). We identify several genes and biological pathways that may be common to adenocarcinomas and serve as targets for highly versatile therapies.

Introduction

The classification of cancers for specific and tailored clinical management has been a topic of ongoing study. Historically, cancers have been classified primarily by the organs in which they originate—a convenient strategy that aligns with traditional anatomical divisions of medicine. Accordingly, many currently available treatments, such as surgery and radiation, are anatomically driven, and tumors arising in different sites are often managed in different ways (and by different subspecialty divisions), even if they possess histological similarity (i.e. squamous cell carcinomas arising in the head and neck, as opposed to the anogenital region). Meanwhile, systemic treatments such as chemotherapy largely do not discriminate between histologically distinct tumors arising within the same organ (i.e. squamous cell carcinomas and adenocarcinomas of the esophagus). These tumors, on the whole, are commonly treated in the same manner. In fact, current guidelines suggest similar treatment of squamous cell carcinomas (SCCs) and adenocarcinomas (ADCs) in the esophagus as well as the cervix, despite observations of differences in prognosis, risk factors, patterns of recurrence, and even response to treatment [1–3].

Recent efforts to re-classify cancers for both clinical and research purposes have addressed this latter issue by utilizing "within-a-tissue" subtyping, identifying molecular signatures that can distinguish between tumors arising within the same organ [4–7]. This has facilitated the development of biologically relevant targeted therapies for particular cancer subtypes, such as molecular subtypes of lung cancer. However, a less common approach has been "across-tissues" analysis [8], which focuses on more global molecular patterns that may transcend traditional anatomic boundaries. This latter approach may provide key insights into common

cancer-associated genes and pathways that span several organ sites, which may facilitate drug development by allowing for the generation of therapeutics that can impact a wider population. For example, categorizing cancers based on mutational burden and immunogenicity has been critical to the broad application of immunotherapies, such as immune checkpoint inhibitors, across a wide array of solid malignancies. In addition, recent perspectives have called for a unified approach to the management and study of SCCs, based on observed similarities in risk factors, genomics, and functional biology [9].

Expanding on this concept, we sought to characterize further the molecular and functional similarities between SCCs and ADCs. ADCs represent the most common type of cancer as a whole, comprising over two-thirds of all cases (of solid and hematologic malignancies combined) in the United States from 2009 to 2013 [10]. Typically originating in organs containing native glandular tissue, adenocarcinomas are the predominant subtype of cancers arising in the breast, colon, prostate, lung, pancreas, stomach, ovary, kidney, and, in the western hemisphere, the esophagus [11]. To our knowledge, there is no literature explicitly comparing adenocarcinomas across tissues of origin at the genomic and transcriptomic levels.

In this study, we utilized data generated by The Cancer Genome Atlas (TCGA) Research Network to study molecular and functional differences between SCCs and ADCs arising within the same organ site. By focusing on organs in which both these subtypes arise, namely the esophagus, lung and uterine cervix, we aimed to determine whether tissue of origin or another property, such as histology, contributes more significantly to observed variation in molecular signatures among tumors. Using a transcriptome-centered approach, we demonstrate here that differences in histology account for more variation than anatomic or tissue origin. Furthermore, via differential gene expression, DNA methylation, and microRNA analysis, we identify 1733 genes, 346 CpG sites, and 42 miRNAs that distinguish ADCs and SCCs between organs, and show that histology correlates with distinct transcriptional and epigenetic programs that are largely consistent across organ sites. Lastly, utilizing pathway and survival analysis, we highlight common markers of ADCs that not only have prognostic value but may also have functional roles in tumorigenesis, thus serving as potential targets for broad therapeutic intervention.

Results

Histology accounts for more transcriptome variation than tissue of origin

To identify the most significant drivers of variation in gene expression, we first performed a principal component analysis (PCA) of gene expression data from 80 esophageal adenocarcinomas (EAC), 81 esophageal squamous cell carcinomas (ESCC), 533 lung adenocarcinomas (LUAD), and 502 lung squamous cell carcinomas (LUSC) (Fig 1A). After ruling out significant batch effects (S1 Fig), we discovered that the greatest component driving gene expression variation (PC1) accounted for significantly more variation (34% of total) than other components (Fig 1B). Using multiple regression, we next determined the relative importance of histology and organ site to predicting PC1. We found that of the variance that our regression model could explain, histology accounted for 90.7%, while organ site only accounted for 9.3% (Fig 1C). We built on these findings using hierarchical clustering, which revealed two main clusters that appeared to be defined by histology: EAC tended to cluster with LUAD, while ESCC tended to cluster with LUSC (Fig 1D). Interestingly, while the majority of EAC samples formed its own subcluster, this pattern was less apparent in ESCC. When we determined the relative contribution of histology or organ site to cluster number via multiple regression, we noted a much stronger contribution from histology (Fig 1E), with histology accounting 98.9% of the variance explained by the regression model (organ site accounted for 1.1%). These
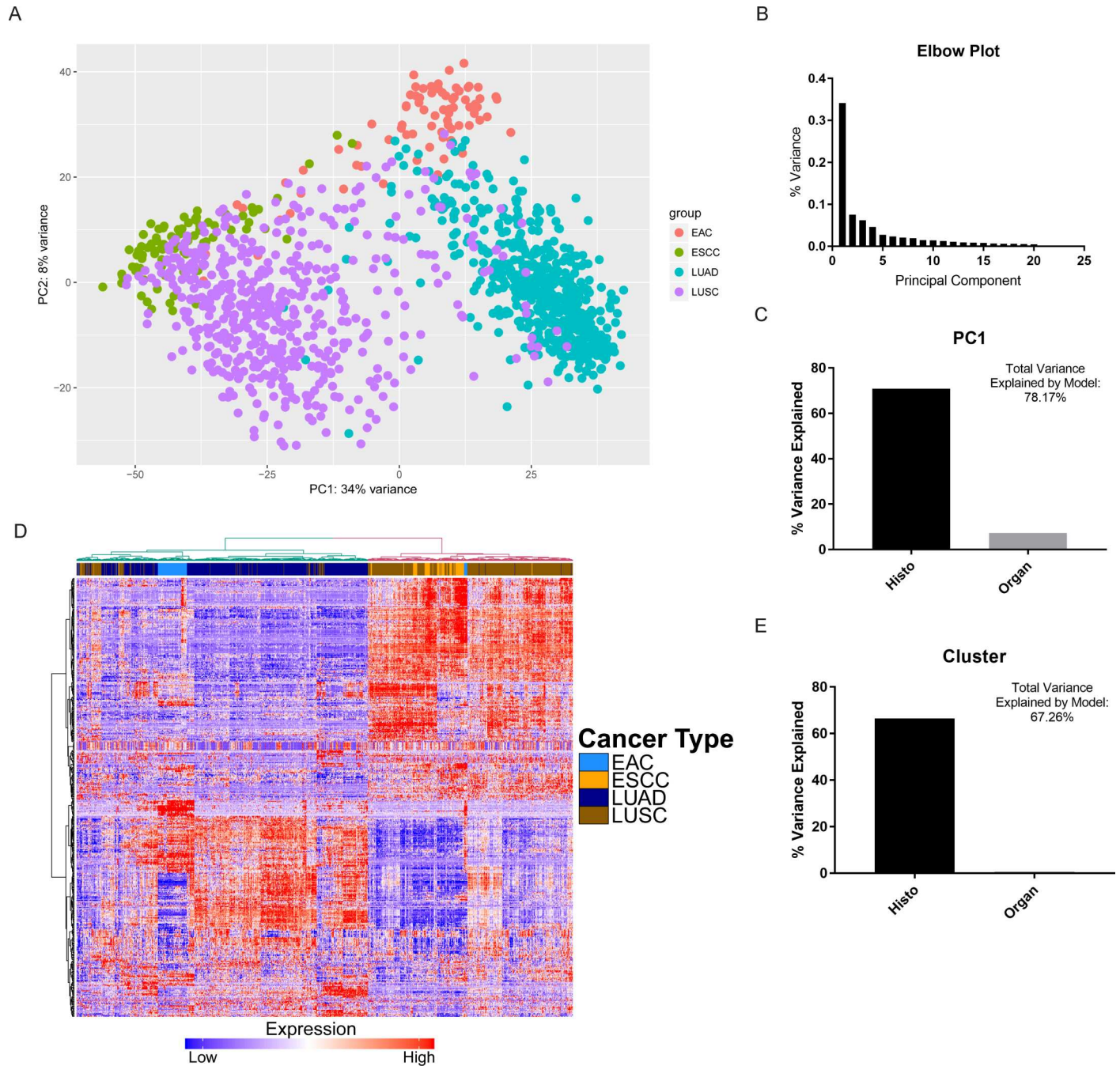
Fig 1. Histology explains a greater degree of gene expression variation than organ site. (A) Principal Component Analysis (PCA) plot depicting two largest components of variance in gene expression in four cancer types. (B) Elbow plot showing the proportion of variance explained by the first 20 principal components (PC1 explains 0.34 or 34% of total variance). (C) Bar chart depicting relative importance (percent variance explained) by histology and organ site, respectively, based on a multiple linear regression model with PC1 as the response variable. Regression model created was [PC1 = 47.2 × Histo + 21.66 × Organ − 42.85], where explanatory variables were demarcated as Histo (0 = SCC, 1 = ADC) and Organ (0 = Esophagus, 1 = Lung). (D) Heatmap of differentially expressed genes between esophageal cancers and lung cancers, with hierarchical clustering. (E) Bar chart depicting relative importance (percent variance explained) by histology and organ site, respectively, based on a multiple linear regression model with Cluster number (1 or 2) as the response variable. Regression model created was [Cluster = 0.80 × Histo + 0.12 × Organ + 1.01], where explanatory variables were demarcated as Histo (0 = ADC, 1 = SCC) and Organ (0 = Lung, 1 = Esophagus).

https://doi.org/10.1371/journal.pgen.1006938.g001

findings suggest that histology may be of greater importance to a tumor's molecular profile than its tissue of origin.

## Histology-driven gene expression patterns are consistent across organs

To understand how consistent histology-driven molecular patterns were across organs, we first compared mRNA expression between ADCs and SCCs in the esophagus and lung. We identified 3443 differentially expressed genes between EAC and ESCC, and 4198 differentially expressed genes between LUAD and LUSC, with 1733 genes that were common in both comparisons (S2 Fig and S1 File). Among these were known markers of squamous cell-fate determination, *TP63* and *SOX2*, which have a role in both the development of normal squamous tissue in the esophagus as well as tumorigenesis along the squamous lineage in the lung [12,13] and esophagus [14,15]. Other notable genes relatively upregulated in SCCs were keratins *KRT14* and *KRT17*; *DSC3*, a member of the desmocollin family (a component of intercellular desmosomes); *FAT2*, an atypical cadherin found to be induced by ΔNp63α (isoform of TP63) and promote invasion in LUSC [16]; *EGFR*, a key growth factor receptor in many cancers; *CCNA1*, which encodes for cyclin A1, a cell cycle regulator; and *WNT3A*, a member of the Wnt family of signaling proteins, which also may have a role in SCCs [17]. Upregulated in ADCs were glandular markers such as mucins *MUC3A* and *MUC13*; claudins (which modulate tight junction permeability) *CLDN2* and *CLDN18*, which has recently been targeted therapeutically in advanced gastric and gastroesophageal junction adenocarcinomas [18]; drivers of cell differentiation such as *GATA6*; hepatic nuclear factors such as *HNF1B*, *FOXA2*, *FOXA3*; and *SPINK1* (serine protease inhibitor Kazal-type 1), which has been associated with a number of gastrointestinal and genitourinary cancers [19].

Next, we evaluated if global patterns of histology-driven gene expression were similar across organ sites by constructing heatmaps with clustering using the differentially expressed genes (DEGs) identified in the esophagus and lung, respectively (Fig 2A and 2B). We found that histology drove a clear and consistent pattern of expression in both the esophagus and lung. Importantly, these DEGs were able to distinguish ADCs from SCCs across both organs after applying hierarchical clustering. These data indicate gene expression profiles determined by histology are largely consistent across different organ sites.

Histology-driven epigenetic patterns are similar across organs

To determine if the patterns observed in differential gene expression in ADCs versus SCCs were associated with epigenetic changes, we compared DNA methylation in each histological subtype. We identified 1734 differentially methylated CpG sites between EAC and ESCC, 1650 differentially methylated CpG sites between LUAD and LUSC, with 346 CpG sites in common between the comparisons (S2B Fig and S2 File). When we observed patterns of DNA methylation and applied hierarchical clustering, we again found that the cancers grouped by histology and not by organ site (Fig 3A and 3B). Interestingly, while EAC and LUAD appeared to form distinct subclusters within the ADC cluster, ESCC and LUSC were more homogeneous in DNA methylation profile and thus did not form separate subclusters.

We then sought to identify particularly important genes by intersecting differentially expressed genes with those that were differentially methylated. We identified 174 such genes in the esophagus, 193 genes in the lung, and 33 common genes between them. Genes that were downregulated and hypermethylated in ADCs were squamous markers such as *DSC3* and *KRT5*, as well as transcription factors such as *FOXE1* and *TP73*, whose isoforms have demonstrated both tumor suppressor and oncogenic properties [20]. Interestingly, *TP73* overexpression in SCCs coincided with *TP63* overexpression in SCCs relative to ADCs (S2C Fig). The
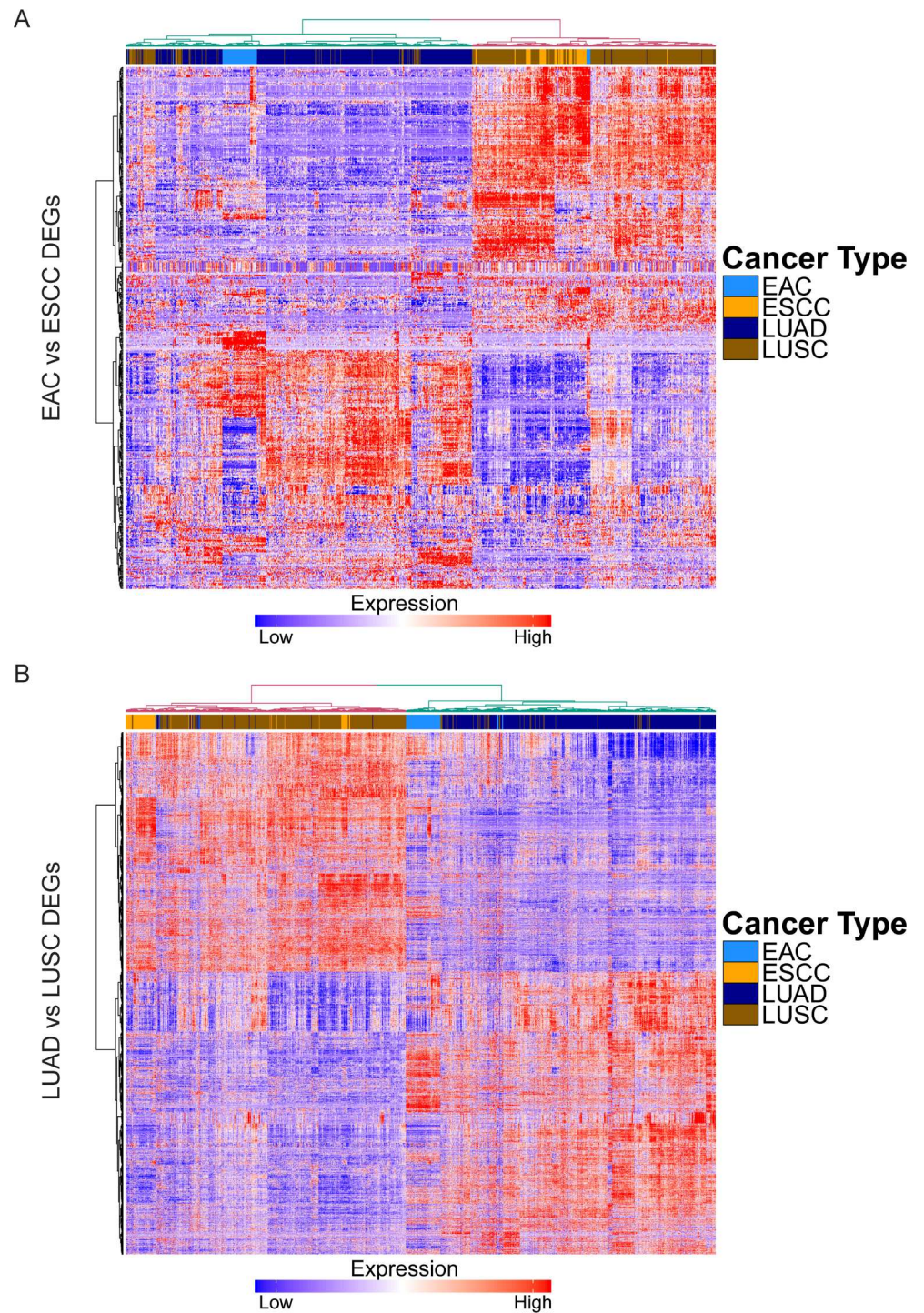
Fig 2. Global molecular patterns defined by histology are consistent across both esophagus and lung. (A) Heatmap depicting mRNA expression of DEGs between EAC and ESCC in ADCs and SCCs of esophagus and lung, with hierarchical clustering. (B) Heatmap depicting mRNA expression of DEGs between LUAD and LUSC in ADCs and SCCs of esophagus and lung, with hierarchical clustering.

predominant isoform of *TP63* in squamous epithelia and SCCs is ΔNp63 [21], which has been demonstrated previously to reversibly inhibit *TP73*-dependent transcription by direct
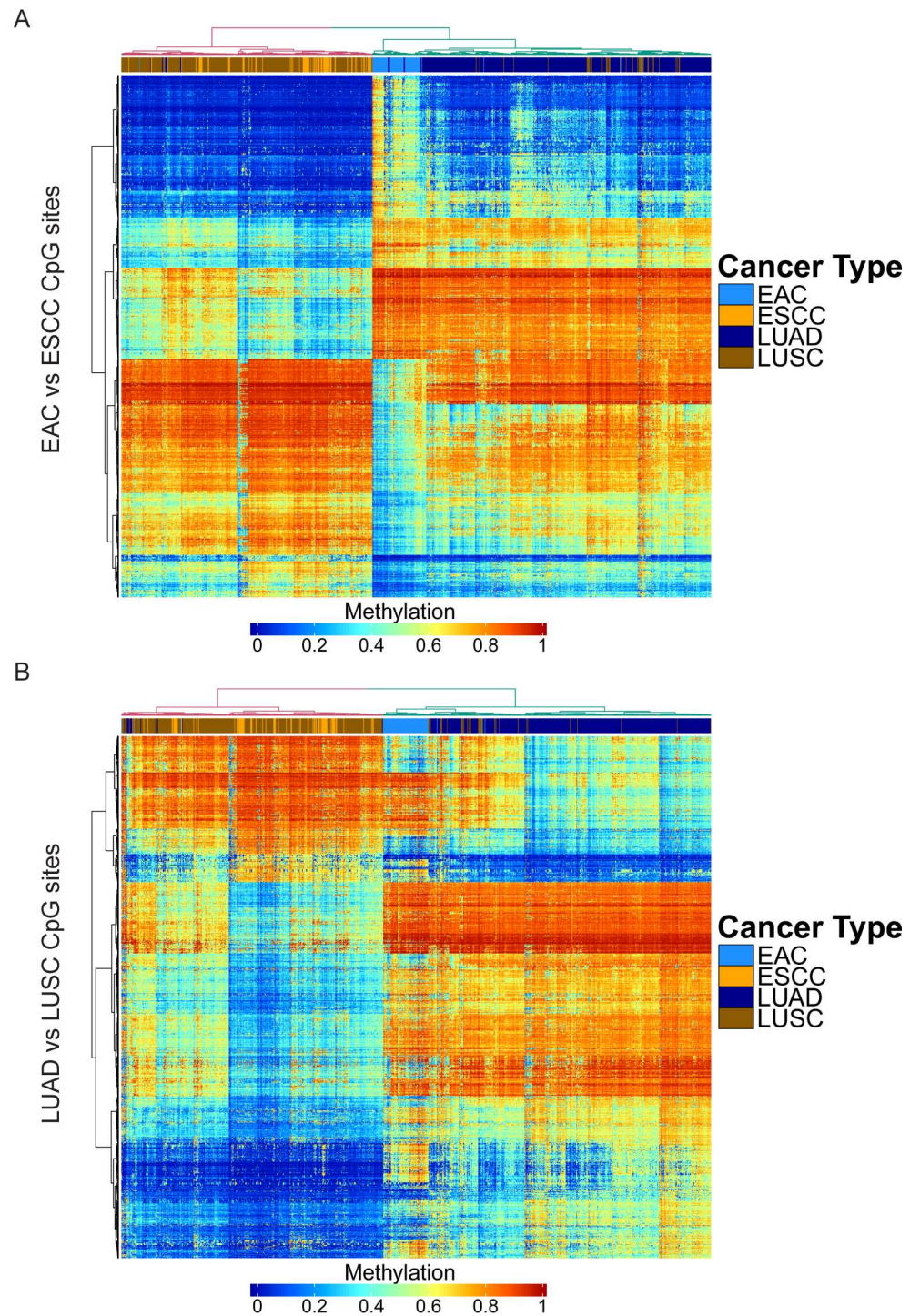
Fig 3. Overall DNA methylation patterns defined by histology are consistent across esophagus and lung. (A) Heatmap depicting DNA methylation of differentially methylated CpG sites between EAC and ESCC in ADCs and SCCs of esophagus and lung, with hierarchical clustering. (B) Heatmap depicting DNA methylation of differentially methylated CpG sites between LUAD and LUSC in ADCs and SCCs of esophagus and lung, with hierarchical clustering.

https://doi.org/10.1371/journal.pgen.1006938.g003

promoter binding or physical interaction with the p73 protein, leading to decreased apoptosis [22]. We did not, however, detect a consistent downregulation of p73 target genes [23],

including those involved in apoptosis (*PMAIP1* and *BBC3*), which may be due to selective inhibition of specific downstream targets or activation by alternative regulators (S2C Fig). By contrast, genes that were upregulated and hypomethylated included mucins like *MUC1*; *SERPINA1*, a serine protease inhibitor that acts in the liver and lung; *HNF1B*, a transcription factor that regulates in renal and pancreatic development; and *ME3*, a mitochondrial malic enzyme whose deletion was shown recently to confer lethality in *SMAD4*-deleted pancreatic adenocarcinoma [24]. These data demonstrate that histological subtype-specific patterns of gene expression correspond to changes in epigenetic profiles.

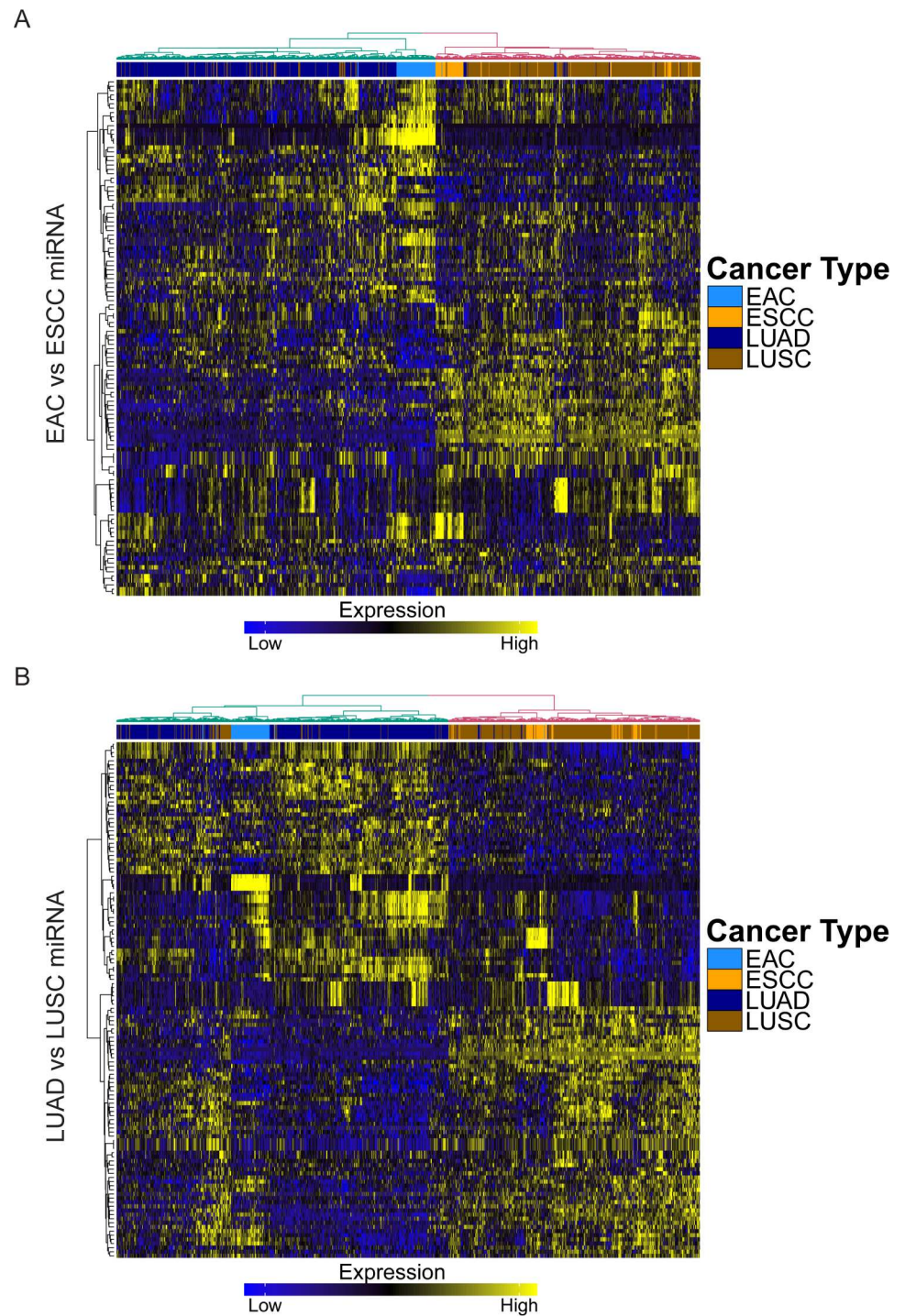## Histology-defined microRNA expression

We then analyzed other potential influences on gene expression. MiRNAs have gained significant attention due to their post-transcriptional regulation of gene expression by destabilizing mRNA and silencing translation. Using miRNA expression data acquired from TCGA, we compared the levels of specific miRNAs in ADCs and SCCs of the esophagus and lung. We identified 118 differentially expressed miRNAs in the esophageal comparison, and 125 in the lung, with 42 overlapping miRNAs (S2D Fig and S3 File). The miRNAs relatively overexpressed in SCCs included miR-994, which has been found to be part of intronic sequence of ΔNp63 and is highly expressed keratinocytes [25], as well as miR-224, which has been implicated in the progression of colorectal cancer and non-small cell lung cancer (NSCLC) [26,27]. On the contrary, miRNAs upregulated in ADCs included miR-215, which has been shown to modulate gastric tumor cell proliferation by targeting *RB1* [28], and miR-375, which has been observed to be upregulated in lung adenocarcinoma but downregulated in lung squamous cell carcinoma, and promotes cell proliferation by decreasing levels of *ITPKB*, a putative tumor suppressor [29,30]. Interestingly, also downregulated in ADCs was miR-149, which was demonstrated previously to have tumor suppressor capacity in breast cancer migration and invasion by targeting small GTPases Rap1a and Rap1b [31]. Furthermore, similar to mRNA expression, patterns of miRNA expression were largely conserved for ADC and SCC across both esophagus and lung (Fig 4A and 4B). These findings suggest that similar patterns of miRNA expression also govern differences in histology, which can be observed across different primary organ sites.

Distinct functional pathways are upregulated in ADC versus SCC

Given the histologically-driven patterns observed in transcriptome, epigenome and post-transcriptional profile, we next performed pathway analysis based on differentially expressed genes to identify functional pathways specific to ADCs or SCCs. Using DEGs from our previous ADC vs. SCC comparisons (i.e. EAC vs. ESCC), we performed pathway analysis with Ingenuity software. As a reference, we also performed pathway analysis using DEGs from cancer vs. normal comparisons (i.e. EAC vs. normal esophagus) (Fig 5A). This reference comparison allowed us to determine whether differential pathways identified in ADC vs. SCC were significantly altered in cancer biology (relative to normal), and whether these differences were due to changes in ADC, SCC, or both. Importantly, the LUAD vs. normal lung (both glandular) and ESCC vs. normal esophagus (both squamous) comparisons provided insight into whether differential pathways were simply markers of a specific histological state, since pure histological markers would be unlikely to result in differences observed in these comparisons.

In our analysis, there were several pathways relatively upregulated in ADCs compared to SCCs (Fig 5B). Intriguingly, the only pathway that was also consistently upregulated in ADC versus normal tissue was liver X receptor / retinoid X receptor (LXR/RXR) activation (S3A Fig). LXR/RXR activation was also downregulated in SCCs relative to normal tissue (S3A Fig),

Fig 4. Global miRNA expression patterns defined by histology are consistent across both esophagus and lung. (A) Heatmap depicting miRNA expression of differentially expressed miRNAs between EAC and ESCC in ADCs and SCCs of esophagus and lung, with hierarchical clustering. (B) Heatmap depicting miRNA expression of differentially expressed miRNAs between LUAD and LUSC in ADCs and SCCs of esophagus and lung, with hierarchical clustering.

suggesting a dichotomous function in these different histological subtypes. Furthermore, differences in LXR/RXR activation were maintained in the LUAD vs. normal lung and ESCC vs.
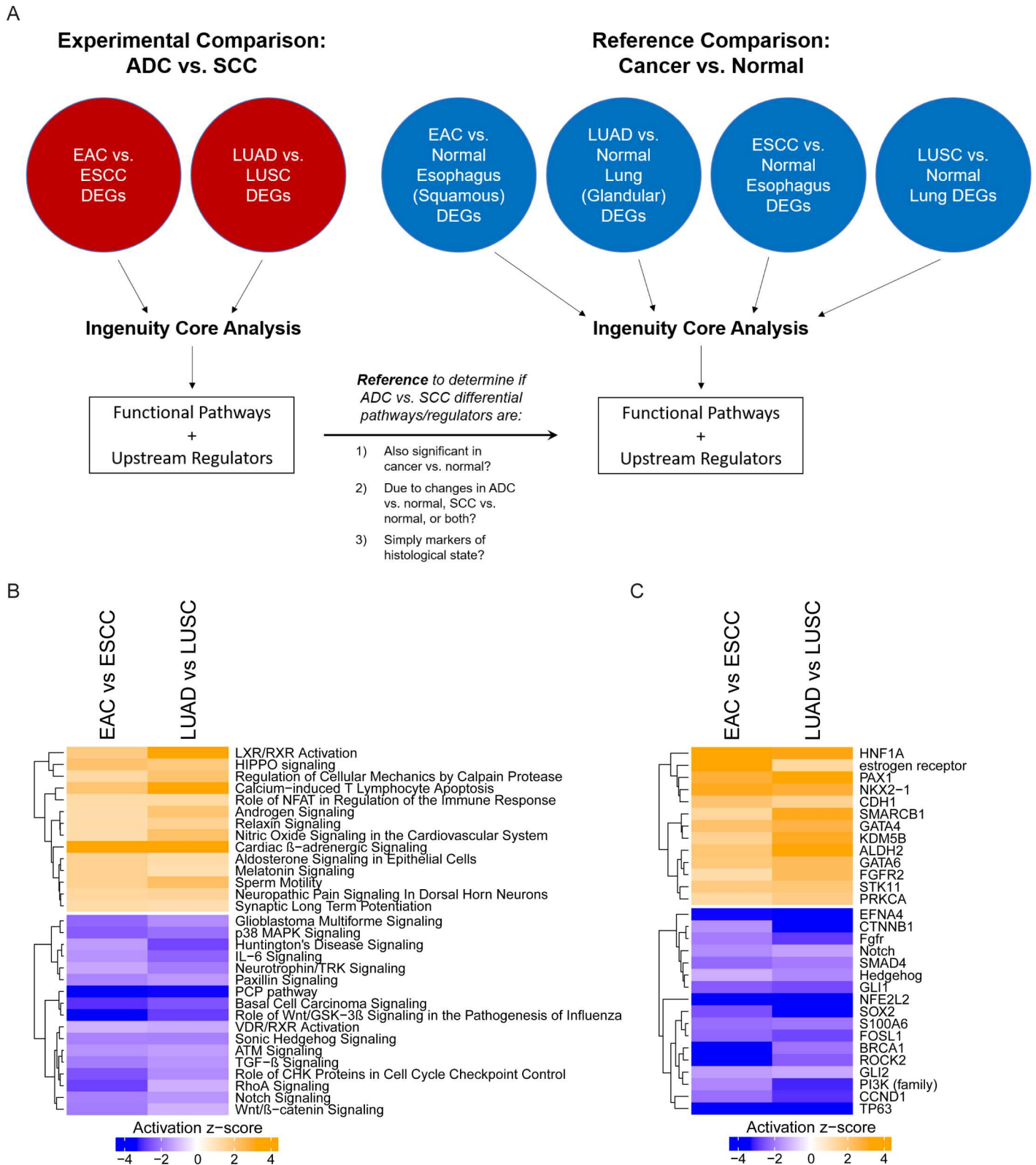
Fig 5. Pathway and upstream regulator analysis using DEGs between ADC and SCC. (A) Schematic illustrating workflow from differential gene expression analysis to pathway and upstream regulator analysis, along with use of reference comparisons (cancer vs. normal). (B) Putative functional pathways that may play a role in the development of ADC and SCC, by activation z-score. (C) Potential upstream regulators with a functional impact on ADC versus SCC, by activation z-score.

normal esophagus comparisons, indicating that it is likely more than a simple marker for glandular histology (S3A Fig). Other pathways relatively upregulated in ADCs included HIPPO signaling, Relaxin signaling, and Androgen Signaling; however, these pathways were mostly downregulated in all cancers relative to normal tissue (S3A Fig). The differentially expressed genes corresponding to each pathway are listed in a supplementary table (S1 Table).

There were several pathways that were relatively downregulated in ADCs compared to SCCs. Some of these pathways were also downregulated in ADCs relative to normal tissue, including p38 MAPK Signaling and Paxillin Signaling. The remaining pathways were upregulated in SCCs relative to both ADCs and normal tissue, suggesting an important functional role in the carcinogenesis of SCCs. This included Basal Cell Carcinoma Signaling, the planar cell polarity (PCP) pathway, and Wnt signaling. Of note, Basal Cell Carcinoma Signaling is closely related to Sonic Hedgehog (SHH) Signaling, which has been associated with the carcinogenesis of SCCs of the oral mucosa, uterine cervix, esophagus and lung [32–35], where it may have a role in mediating stemness via transcription factors like *SOX2* [34,35]. Furthermore, aberrant Wnt signaling and non-canonical Wnt/PCP signaling, which normally regulates cell shape via the cytoskeleton [36], have been hypothesized to have a unique role in SCCs [17], and may represent a functional distinction between SCCs and ADCs [37].

Expanding on this pathway analysis, we investigated potential upstream regulators driving changes in gene expression (Fig 5B). Notably upregulated in ADCs relative to SCCs was *STK11* (or LKB1), loss of which has been previously shown to induce adeno-to-squamous differentiation of lung tumors in mice [38]; *GATA4* and *GATA6*, important for cell differentiation and commonly amplified in EAC and gastric cancers [4]; and *NKX2-1*, a commonly used marker for lung adenocarcinomas. Furthermore, the estrogen receptor was found to be relatively upregulated in ADCs versus SCCs and normal tissue. Several genes were also found to be likely upstream regulators of SCC formation. For instance, *CDH1*, which codes for E-cadherin, a membrane protein important for cellular adhesion, was relatively downregulated in SCCs, which aligns with reports of its methylated and suppressed expression in oral SCCs [39]. Other notable genes relatively upregulated in SCC were *GLI1*, a downstream mediator of hedgehog signaling; *SMAD4*, which is involved in TGF-β signaling and is commonly is deleted in EAC and LUAD [4,5]; *CCND1*, a cyclin that is commonly amplified in ESCC [4]; and *TWIST1*, thought to be important in epithelial-to-mesenchymal transition [40]. The activation of these upstream regulators in each cancer type relative to corresponding normal tissue is shown in a supporting figure (S3B Fig).

Identifying genes important for survival in ADC

We next sought to highlight additional gene candidates that may serve as prognostic markers among ADCs, given the lack of unifying features previously described in the literature. Using the 500 genes showing the greatest variance in expression, we performed Cox proportional hazards survival analysis to find genes significantly correlated with survival outcomes in both EAC and LUAD. We discovered 109 such genes in EAC, and 691 genes in LUAD, with 32 genes in common between them (Fig 6A and 6B). Genes whose increased expression was correlated with worse survival outcomes included *TPX2*, *KIF4A*, *IGF2BP1*, and *HSPA6*. On the contrary, genes whose expression was correlated with a better prognosis included *MS4A1* (CD20), *SUSD2*, and *CX3CL1* (Fractalkine). As an example, we constructed Kaplan-Meier curves for *IGF2BP1*, or IMP-1, which we described previously as a modulator of tumor growth in colorectal cancer [41]. Not only did higher expression of *IGF2BP1* correlate with poorer outcomes in EAC and LUAD (Fig 6C and 6D), but the same trend was seen when we pooled several other ADCs from the TCGA Pan-Cancer dataset (including cancers of the breast,
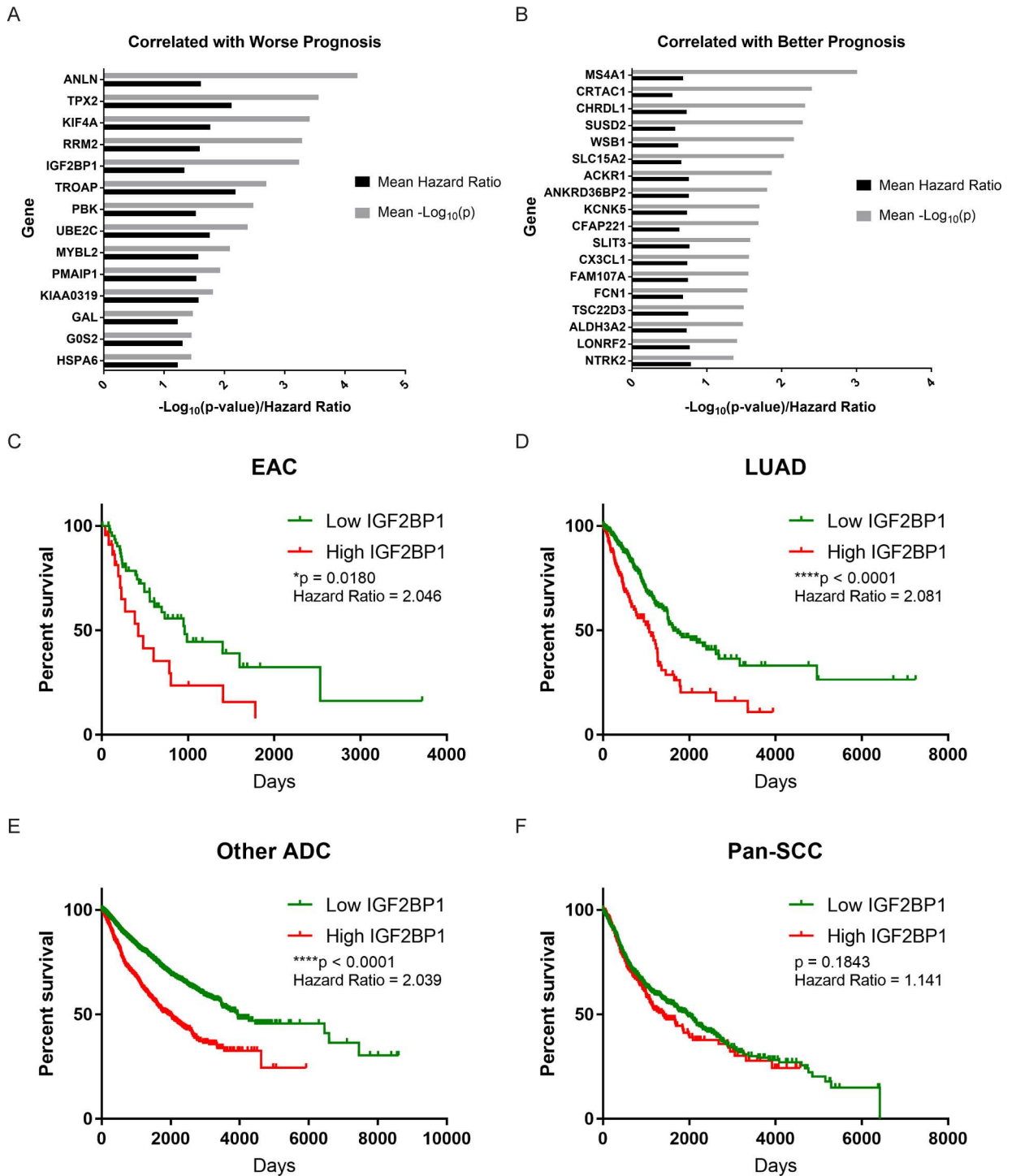
Fig 6. Genes correlated with survival in ADCs. (A) Genes whose expression is correlated with worse survival in ADCs. (B) Genes whose expression is correlated with better survival in ADCs. (C) Kaplan-Meier curve depicting survival in high *IGF2BP1* and low *IGF2BP1* expression groups in EAC (log-rank test). (D) Kaplan-Meier curve depicting survival in high *IGF2BP1* and low *IGF2BP1* expression groups in LUAD (log-rank test). (E) Kaplan-Meier curve depicting survival in high *IGF2BP1* and low *IGF2BP1* expression groups in pooled ADCs, excluding EAC and LUAD (log-rank test). (F) Kaplan-Meier curve depicting survival in high *IGF2BP1* and low *IGF2BP1* expression groups in pooled SCCs (including ESCC and LUSC) (log-rank test).

https://doi.org/10.1371/journal.pgen.1006938.g006

prostate, endocervix, endometrium, ovary, pancreas, stomach, kidney, colon, rectum, and thyroid) (Fig 6E). Notably, the same trend was not observed in a pooled SCC dataset (including cancers of head and neck, esophagus, lung, and ectocervix) (Fig 6F). Intriguingly, while ESCC and LUSC had 88 and 206 genes associated with survival, respectively, they shared no genes that had the same directional association with outcomes.

## Validation: Expression patterns in the uterine cervix

As a validation, we attempted to reproduce our results in a third organ site in which both ADCs and SCCs arise: the uterine cervix. While the predominant subtype of cancer in the cervix is squamous cell carcinoma (CESC) associated with human papilloma virus (HPV) infection, several types of adenocarcinoma, specifically endocervical, endometroid, and mucinous adenocarcinoma, can also develop. Using data acquired from TCGA, we pooled 44 samples of these various types of cervical ADCs (ECA) and compared them with 254 CESC samples by using gene expression, DNA methylation, and miRNA expression data acquired from the esophagus and lung comparison. Upon clustering, we found that the same genes, methylated CpG sites, and miRNAs distinguishing histology in the esophagus and lung were also largely able to distinguish histology in the cervix (Fig 7A–7C).

The pathways we had identified previously to be important to ADC and SCC, respectively, also maintained similar patterns in the cervix, with deviations only in Regulation of Cellular Mechanics by Calpain Protease, Melatonin Signaling, and Calcium-induced T Lymphocyte Apoptosis in pathways favoring ADCs (Fig 7D). LXR/RXR Activation was once again upregulated in ECA relative to CESC. In pathways that were relatively downregulated in ADCs, Notch Signaling and the PCP pathway maintained strong associations with SCC; however, there was a weaker relationship in the cervix between squamous histology and Basal Cell Carcinoma Signaling, Sonic Hedgehog Signaling, or TGF-β Signaling. We also performed upstream regulator analysis of ECA and CESC, again observing similar patterns to the histological subtypes of the esophagus and lung (Fig 7E). Notable differences included *ALDH2*, *STK11* in genes relatively upregulated in ADCs, and *CTNNB1* and *GLI2* in genes relatively downregulated in ADCs. Altogether, histology-driven differences in transcriptome, epigenome, as well as pathways and upstream regulators, were largely consistent across all three organ sites studied.

Discussion

Our analyses have revealed major differences in the transcriptomic and epigenomic profile of ADCs and SCCs, but major similarities among ADCs and SCCs spanning several anatomic sites. Importantly, histology was the greater determinant of molecular variation than organ site, which further emphasizes the need to re-classify tumors and re-structure current treatment paradigms. While genomic subtyping studies have demonstrated the need to consider molecular characteristics (and biological relevance) when managing tumors arising within the same organ, we also highlight the premise that some molecular features may be shared by tumors arising in different organs. Capitalizing on this property of cancer could lead to the development of widely applicable targeted therapies.

Importantly, we highlighted several specific properties that may unify ADCs, including a large collection of genes, CpG methylation sites, and miRNAs. Specifically, we identified LXR/RXR activation as having a putative role in the development of ADC. Indeed, several recent reports have highlighted liver X receptors as potential therapeutic targets, as they appear to regulate aerobic glycolysis (the Warburg effect) and lipogenesis [42], which malignant cells depend upon to sustain rapid proliferation, while also modulating antitumor immunity [43].
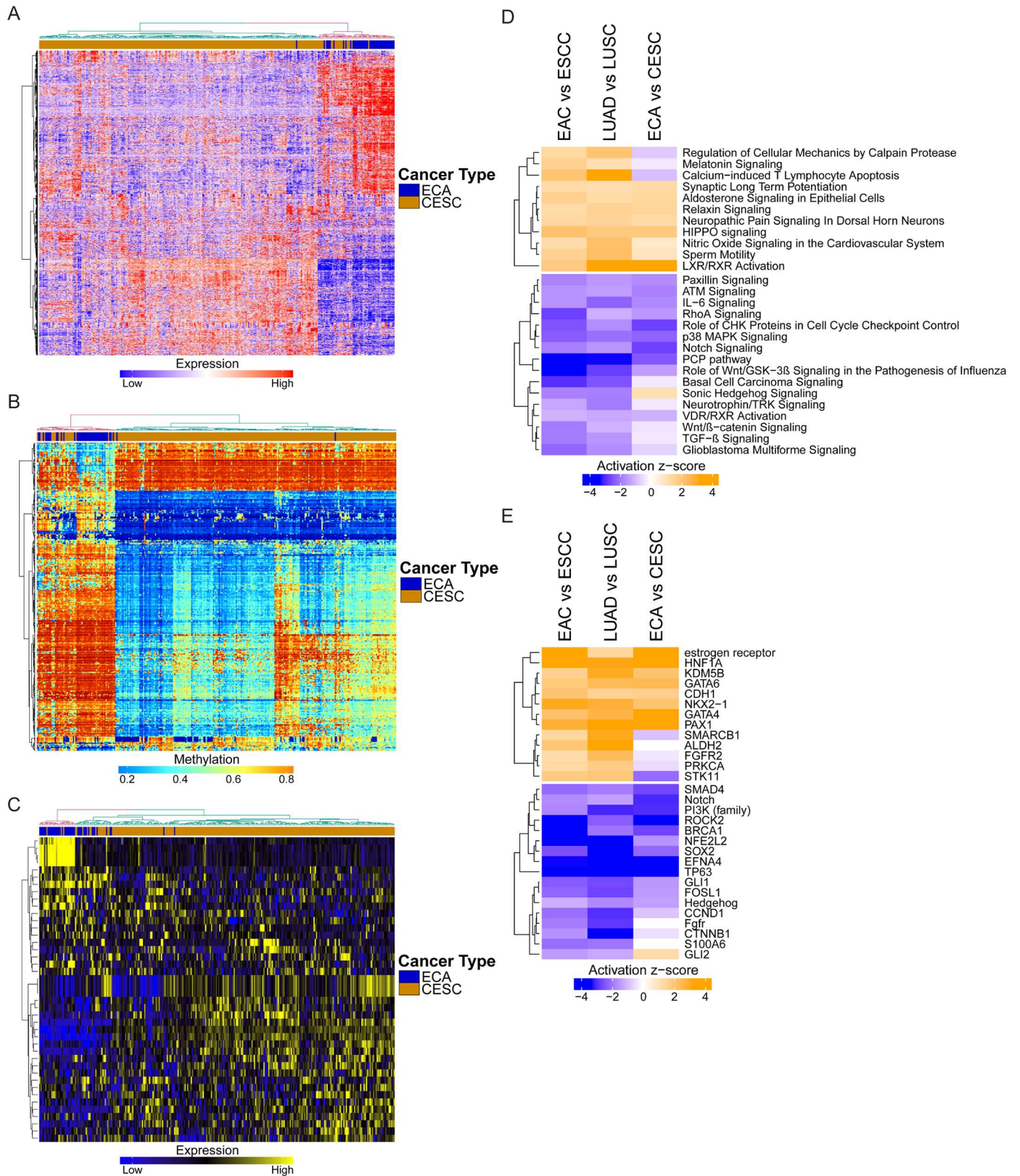
Fig 7. Validation in endocervical adenocarcinoma and cervical squamous cell carcinoma. (A) Heatmap depicting mRNA expression of common DEGs between ADC and SCC (esophagus and lung) in the cervix, with hierarchical clustering. (B) Heatmap depicting DNA methylation of common

differentially methylated CpG sites between ADC and SCC (esophagus and lung) in the cervix, with hierarchical clustering. (C) Heatmap depicting miRNA expression of common differentially expressed miRNAs between ADC and SCC (esophagus and lung) in the cervix, with hierarchical clustering. (D) Pathways identified using DEGs from ECA versus CESC, compared to results from the esophagus and lung. (E) Upstream regulators identified using DEGs from ECA versus CESC, compared to results from the esophagus and lung.

In fact, one study utilized an inverse agonist SR9243 to inhibit LXR activity, suppressing tumor growth in several colon, prostate, lung and pancreatic adenocarcinoma cell lines [42]. Curiously, our analyses also indicated that LXR/RXR activation was downregulated in SCCs relative to normal tissue. This is in line with a previous report that described the spontaneous development of peripheral squamous cell lung cancer in mice with ablated LXRα,β (*NR1H3-/-* and *NR1H2-/-*) [44]. Taken together, this evidence points to a possible dichotomous role of LXRs in ADCs versus SCCs and may indicate key differences in cell metabolism between these histological subtypes of solid malignancies. Despite these findings, however, other studies on LXRs have reported a downstream antiproliferative effect that can be harnessed by using LXR agonists (oxysterols) in several cancer types, including multiple ADCs [45]. It is clear, therefore, that further investigation on the precise role of LXRs in different contexts is required.

In addition, we identified candidate genes that correlate with patient outcomes and may have a functional role in ADC carcinogenesis, such as *IGF2BP1*. Intriguingly, increased *IGF2BP1* expression was found to be associated with worse survival in EAC, LUAD, and a pooled dataset of other ADCs spanning breast to colon (Fig 6C–6F). The same pattern was not observed in a pooled SCC dataset. *IGF2BP1* is an mRNA binding oncofetal protein, not normally expressed in adult tissues, that we previously demonstrated to have a role in modulating colorectal tumor growth *in vitro* and *in vivo*, and may also play a role in early metastasis [41]. Similar findings were reported previously in ovarian and breast adenocarcinomas [46,47]. This association further emphasizes the importance of developmental pathways in tumorigenesis, and specifically indicates a possible link between embryonic drivers of differentiation and histological subtypes in cancer. Notably, there have been efforts to develop inhibitors of *IGF2BP1*, with some recent success in the identification of a compound that selectively kills *IGF2BP1* expressing cells [48].

Liver X receptors and *IGF2BP1* are just two examples of therapeutic targets that may be particularly effective in ADCs. Our analysis has revealed an entire catalog of gene candidates for further study, including those that correlate with survival outcomes, though an association with survival is not a prerequisite for being an effective target (i.e. estrogen receptor positivity in breast cancer). Notably, our findings also have implications for the use of currently available therapeutics in ADCs versus SCCs. For example, our upstream regulator analysis highlighted several putative regulators that have greater activity in SCCs than ADCs, such as GLI1 and the PI3K family. These results suggest that SCCs may have greater sensitivity to GLI1 or Hedgehog inhibitors, as well as PI3K inhibitors, than ADCs. Another finding with potential therapeutic relevance is *TP73* methylation and low expression in ADCs. *TP73* is a known tumor suppressor [20], and its methylation could contribute to reduced apoptosis in response to chemo- or radiotherapy relative to SCCs [3]. One could hypothesize that the use of DNA methylation inhibitors, such as 5-azacitidine or decitabine, could restore sensitivity to cytotoxic therapy in ADCs, although this requires experimental validation.

We then noted that although global histology-driven patterns were conserved in esophagus and lung, some of these similarities may have been due to a common embryonic precursor, as both the esophagus and lung are derived from the foregut endoderm [49]. To address this, we tested our findings in a third organ, the uterine cervix, which develops from the paramesonephric ducts and is mesodermal in origin. We found that histology-driven differences were

largely maintained even in the cervix (Fig 7A–7C). There were some notable exceptions, however, which could be related to different embryonic tissue origin or differences in other factors such as the local tumor microenvironment. For example, *STK11* was discovered to be relatively downregulated in ESCC and LUSC, but relatively upregulated in CESC (Fig 7E). These findings are consistent with other reports: while loss of *STK11* has been associated with lung, skin, and head and neck SCCs [50], inherited loss of *STK11*, as in Peutz-Jeghers syndrome, has been associated with endometrial adenocarcinoma and a rare variant of endocervical carcinoma, minimal deviation adenocarcinoma of the endocervix [51]. Differences observed between the cervix and the esophagus and lung could also be due to distinct risk exposures. For instance, *ALDH2*, which is important for acetaldehyde (a metabolite of ethanol and an ingredient in tobacco smoke) metabolism and whose deficiency increases the risk for esophageal and head and neck SCCs, was found to be relatively downregulated in ESCC and LUSC, but was not significantly different when comparing ECA and CESC (Fig 7E). This could be due to direct exposure of the aerodigestive tract and respiratory mucosa to acetaldehyde via alcohol consumption and cigarette smoking, respectively, in *ALDH2-/-* individuals.

Lastly, there were some technical limitations to our *in silico* analysis. First, TCGA datasets can be susceptible to batch effects, although when we assessed the data, variation due to batch was determined to be of relatively insignificant impact (S1 Fig). Similarly, although TCGA datasets are generally large, they may not be representative of the general population, and critically, samples sent for sequencing are often not purely tumor cells and may contain associated stromal and immune components, as well as normal tissue. Additionally, because we were primarily interested in global molecular patterns, our analysis was not sensitive to relatively rare variants such as those described in genomic and mutational analyses. It should be noted that highlighted candidate pathways and genes need to be verified further for functional significance with additional *in vitro* or *in vivo* analyses.

Conclusion

These analyses further emphasize the need to re-categorize tumors in a biologically relevant manner—not only as a practice in pathology and research, but also in the clinic. Over the last decade, molecular subtyping efforts have already called for a shift in approach to viewing cancers—such that esophageal carcinoma, for instance, would not be viewed as a single entity [4–6], but as multiple diseases with distinct characteristics and patterns of behavior. While our findings distinguishing between histological subtypes support this premise, we also propose that broader strategies of classification to direct research efforts and clinical management may be more aligned to the pragmatic concerns of drug development while still being biologically relevant. Consider the example of immune checkpoint inhibitors, which are being widely applied across cancers.

In this report, we emphasize the call for a unified approach to the study, prevention, and treatment of SCCs, and on possible commonalities underlying ADCs. Despite apparent heterogeneity among ADCs, we have identified patterns of gene expression, functional pathways, and prognostic markers that may similarly unify our approach to ADCs. Furthermore, our analyses have provided a catalog of additional candidate genes and markers that can be further investigated. To our knowledge, this is the first study identifying common features among adenocarcinomas arising from different organ sites.

Methods

Data acquisition

Publicly available gene expression, DNA methylation, and microRNA (miRNA) data from TCGA were downloaded from the Genomic Data Commons (GDC) Data Portal and pre-

processed via the *TCGAbiolinks* R package [52]. Details regarding specific data type and subsequent analysis are described below. Survival outcomes data were acquired from the University of California Santa Cruz (UCSC) Xena at http://xena.ucsc.edu.

Differential gene and microRNA expression

Harmonized gene expression data (*HTSeq* counts)—to allow for easy comparison between cancer types—from GDC was downloaded for esophageal carcinoma (TCGA-ESCA), lung adenocarcinoma (TCGA-LUAD), lung squamous cell carcinoma (TCGA-LUSC), and cervical and endocervical carcinoma (TCGA-CESC), and assembled into a matrix using *TCGAbiolinks* [52]. Potential batch effects were visualized by constructing a PCA plot by batch and cancer type, using the R package *DESeq2* [53]. Considering insignificant batch effects and to avoid masking biologically significant differences, no corrections for batch were applied. Using the *edgeR* package in R [54], the counts data were filtered for low counts (less than 1 count per million in greater than 50% of each group, i.e. esophageal adenocarcinoma vs. esophageal squamous cell carcinoma) and normalized within samples using the trimmed means of M-values (TMM) method [55]. Differentially expressed genes (DEGs) were filtered using the *glmTreat* function in *edgeR*, which tests for significant differences relative to a set $\log_2$-fold change (logFC) cutoff of 0.5. Results were then filtered by a false discovery rate (FDR) of 0.05.

Harmonized microRNA expression data for ESCA, LUAD, LUSC, and CESC were acquired from the GDC Data Portal and preprocessed via *TCGAbiolinks* [52]. Low counts were filtered as above differential expression analysis was performed using *edgeR*. A logFC cutoff of 0.5 and FDR cutoff of 0.05, as above, were used.

DNA methylation analysis

Legacy DNA methylation data (Illumina Human Methylation 450, aligned to hg19) from GDC was downloaded via *TCGAbiolinks* [52]. Using the *TCGAanalyze_DMR* function, we searched for differentially methylated CpG sites using beta-values, with a cutoff difference in beta-values of 0.25 and adjusted p-value of 10^{-20}. The results from these data were then integrated with the differential gene expression data.

Principal component analysis, hierarchical clustering and visualization

To identify global patterns in gene expression, raw counts were first quantile-normalized within and across samples, using the *TCGAanalyze_Normalization* function, then transformed using the *varianceStabilizingTransformation* function in *DESeq2* to reduce heteroscedasticity (chosen over log transformation to minimize variance at low count values) [53]. Principal component analysis (PCA) was performed using the *prcomp* function in R. Hierarchical clustering was performed in R by Euclidean distance and the ward.D2 method. To determine the relative contribution of explanatory variables (Histology and Organ) to variance, we first constructed multiple linear regression models for response variables "PC1" and "Cluster", respectively, using the *lm* function in R. We then utilized the *relaimpo* package [56], which refers to the Lindeman, Merenda, Gold (1980) method of R^2 partitioning by average over orders [57], in order to determine the contribution to variance of each explanatory variable.

For visualization, the PCA were plotted by cancer type using the *plotPCA* function in *DESeq2* [53]. For heatmaps, normalized gene expression data (as described above) and miRNA expression data were further scaled to a mean of 0 and standard deviation of 1, while for DNA methylation, β-values were used directly in visualization. All heatmaps were constructed using the *ComplexHeatmap* R package [58]. Bar charts were constructed in GraphPad Prism.

Pathway and upstream regulator analysis

Pathway analysis and identification of upstream regulators, using differentially expressed genes as input, were performed using QIAGEN Ingenuity Pathway Analysis software (using the core analysis feature). The same cutoffs as described in the differential gene expression analysis were used to filter significant genes. To determine significant pathways across different organs, we used the comparison analysis function in Ingenuity. Significant pathways were filtered by activation z-score (absolute value ≥ 1.0) and adjusted p-value (≤ 0.05), in the ADC vs. SCC experimental comparisons.

Specifically, the DEGs from seven comparisons were initially used as input for Ingenuity (Fig 5A). These included three ADC vs. SCC comparisons (esophagus, lung, and cervix), as well as four reference cancer vs. normal comparisons. As discussed in the text, these reference comparisons help to determine whether differential pathways identified in ADC vs. SCC were also significantly altered in cancer biology (relative to normal), and whether these differences were due to changes in ADC, SCC, or both. Additionally, the LUAD vs. normal lung (both glandular) and ESCC vs. normal esophagus (both squamous) comparisons provided insight into whether differential pathways were simply markers of a specific histological state, since changes in these comparisons would not be expected for simple histological markers.

Survival analysis

Survival outcomes for TCGA samples were downloaded from UCSC Xena. To identify genes that have potential prognostic and functional significance, we only analyzed the top 500 genes with greatest variance in either ADCs or SCCs. Given the continuous nature of gene expression data, we used a cox proportional hazards model to identify genes correlated with survival. In order to improve sensitivity, no correction for multiple testing was applied. Survival was visualized by plotting Kaplan-Meier curves in GraphPad Prism 7.0. Importantly, TCGA pan-cancer data was acquired from UCSC Xena. Pooled ADC data contained 14 different ADC types with 4934 samples. Pooled SCC data contained 4 SCC types with 1346 samples. For survival, high expression and low expression groups were determined by cutoff at the 75th percentile.

Statistical analysis

Statistical analysis for principal component analysis, linear correlation, differential expression, and Cox proportional hazards were performed in R. Pathway statistical analysis was built-in to Ingenuity. Log-rank tests for Kaplan-Meier were performed using GraphPad Prism.

Supporting information

S1 Fig. Principal Component Analysis plot depicting batch effect versus group (cancer type) effect.
(TIF)

S2 Fig. ADC versus SCC of esophagus and lung show overlap in DEGs, DNA methylation and miRNA expression. (A) Venn diagram showing number of DEGs in EAC versus ESCC and LUAD versus LUSC, with overlap. (B) Venn diagram showing number of differentially methylated CpG sites in EAC versus ESCC and LUAD versus LUSC, with overlap. (C) Heatmap depicting relationships in mRNA expression between *TP63* and *TP73*, followed by expression of downstream target genes of p73 transcriptional regulation. (D) Venn diagram showing number of differentially expressed miRNAs in EAC versus ESCC and LUAD versus LUSC,

with overlap.
(TIF)

S3 Fig. Pathway and upstream regulator analysis of cancer versus normal tissues. (A) Significant pathways from Fig 5, with relationships between four cancer types (EAC, LUAD, ESCC, LUSC) relative to respective normal tissue. (B) Significant upstream regulators from Fig 5, with relationships between four cancer types (EAC, LUAD, ESCC, LUSC) relative to respective normal tissue.
(TIF)

S1 Table. Differentially expressed genes corresponding to each functional pathway identified by ingenuity pathway analysis.
(XLSX)

S1 File. ADC versus SCC differential gene expression data.
(XLSX)

S2 File. ADC versus SCC differential DNA methylation data.
(XLSX)

S3 File. ADC versus SCC differential miRNA expression data.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Eric W. Lin, Tatiana A. Karakasheva, Anil K. Rustgi.

Formal analysis: Eric W. Lin.

Funding acquisition: Adam J. Bass, Kwok K. Wong, Anil K. Rustgi.

Investigation: Eric W. Lin.

Methodology: Eric W. Lin, Tatiana A. Karakasheva, Dong-Jin Lee, Ju-Seog Lee, Qi Long, Adam J. Bass, Anil K. Rustgi.

Validation: Eric W. Lin.

Visualization: Eric W. Lin.

Writing – original draft: Eric W. Lin.

Writing – review & editing: Eric W. Lin, Tatiana A. Karakasheva, Adam J. Bass, Kwok K. Wong, Anil K. Rustgi.

References

1. Lieberman MD, Shriver CD, Bleckner S, Burt M. Carcinoma of the esophagus. Prognostic significance of histologic type. J Thorac Cardiovasc Surg. 1995; 109: 130–8; discussion 139. https://doi.org/10.1016/S0022-5223(95)70428-0 PMID: 7815789

2. Lee KB, Lee JM, Park CY, Lee KB, Cho HY, Ha SY. What is the difference between squamous cell carcinoma and adenocarcinoma of the cervix? A matched case-control study. Int J Gynecol Cancer. 2006; 16: 1569–73. https://doi.org/10.1111/j.1525-1438.2006.00628.x PMID: 16884367

3. Siewert JR, Ott K. Are squamous and adenocarcinomas of the esophagus the same disease? Semin Radiat Oncol. 2007; 17: 38–44. https://doi.org/10.1016/j.semradonc.2006.09.007 PMID: 17185196

4. Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017; 541: 169–175. https://doi.org/10.1038/nature20805 PMID: 28052061

5. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet. 2016; 48: 607–616. https://doi.org/10.1038/ng.3564 PMID: 27158780

6. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017; 543: 378–384. https://doi.org/10.1038/nature21386 PMID: 28112728

7. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513: 202–209. https://doi.org/10.1038/nature13480 PMID: 25079317

8. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. Cell. Elsevier; 2014; 158: 929–944. https://doi.org/10.1016/j.cell.2014.06.049 PMID: 25109877

9. Dotto GP, Rustgi AK. Squamous Cell Cancers: A Unified Perspective on Biology and Genetics. Cancer Cell. Elsevier; 2016; 29: 622–637. https://doi.org/10.1016/j.ccell.2016.04.004 PMID: 27165741

10. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ CK. SEER Cancer Statistics Review, 1975–2013 [Internet]. Bethesda; 2015. Available: https://seer.cancer.gov/csr/1975_2013/sections.html

11. Rustgi AK, El-Serag HB. Esophageal Carcinoma. N Engl J Med. 2014; 371: 2499–2509. https://doi.org/10.1056/NEJMra1314530 PMID: 25539106

12. Que J, Okubo T, Goldenring JR, Nam K-T, Kurotani R, Morrisey EE, et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development. NIH Public Access; 2007; 134: 2521–31. https://doi.org/10.1242/dev.003855 PMID: 17522155

13. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet. Nature Publishing Group; 2009; 41: 1238–1242. https://doi.org/10.1038/ng.465 PMID: 19801978

14. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet. 2009; 41: 1238–1242. https://doi.org/10.1038/ng.465 PMID: 19801978

15. Watanabe H, Ma Q, Peng S, Adelmant G, Swain D, Song W, et al. SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. J Clin Invest. 2014; 124: 1636–1645. https://doi.org/10.1172/JCI71545 PMID: 24590290

16. Dang TT, Westcott JM, Maine EA, Kanchwala M, Xing C, Pearson GW. ΔNp63α induces the expression of FAT2 and Slug to promote tumor invasion. Oncotarget. 2016; 7: 28592–611. https://doi.org/10.18632/oncotarget.8696 PMID: 27081041

17. Shiah S-G, Shieh Y-S, Chang J-Y. The Role of Wnt Signaling in Squamous Cell Carcinoma. J Dent Res. 2016; 95: 129–134. https://doi.org/10.1177/0022034515613507 PMID: 26516128

18. Schuler M, Al-Batran S-E, Zvirbule Z, Manikhas G, Lordick F, Rusyn A, et al. Final results of the FAST study, an international, multicenter, randomized, phase II trial of epirubicin, oxaliplatin, and capecitabine (EOX) with or without the anti-CLDN18.2 antibody IMAB362 as first-line therapy in patients with advanced CLDN18.2+ gast. Ann Oncol Off J Eur Soc Med Oncol. 2016; 27.

19. Rasanen K, Itkonen O, Koistinen H, Stenman U-H. Emerging Roles of SPINK1 in Cancer. Clin Chem. 2016; 62: 449–457. https://doi.org/10.1373/clinchem.2015.241513 PMID: 26656134

20. Rufini A, Agostini M, Grespi F, Tomasini R, Sayan BS, Niklison-Chirou MV, et al. p73 in Cancer. Genes Cancer. Impact Journals, LLC; 2011; 2: 491–502. https://doi.org/10.1177/1947601911408890 PMID: 21779517

21. Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dötsch V, et al. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. Mol Cell. 1998; 2: 305–16. Available: http://www.ncbi.nlm.nih.gov/pubmed/9774969 PMID: 9774969

22. Rocco JW, Leong C-O, Kuperwasser N, DeYoung MP, Ellisen LW. p63 mediates survival in squamous cell carcinoma by suppression of p73-dependent apoptosis. Cancer Cell. 2006; 9: 45–56. https://doi.org/10.1016/j.ccr.2005.12.013 PMID: 16413471

23. Fontemaggi G, Kela I, Amariglio N, Rechavi G, Krishnamurthy J, Strano S, et al. Identification of Direct p73 Target Genes Combining DNA Microarray and Chromatin Immunoprecipitation Analyses. J Biol Chem. 2002; 277: 43359–43368. https://doi.org/10.1074/jbc.M205573200 PMID: 12213815

24. Dey P, Baddour J, Muller F, Wu CC, Wang H, Liao W-T, et al. Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. Nature. 2017; 542: 119–123. https://doi.org/10.1038/nature21052 PMID: 28099419

25. Kim K-H, Cho E-G, Yu SJ, Kang H, Kim Y-J, Kim SH, et al. ΔNp63 intronic miR-944 is implicated in the ΔNp63-mediated induction of epidermal differentiation. Nucleic Acids Res. Oxford University Press; 2015; 43: 7462–7479. https://doi.org/10.1093/nar/gkv735 PMID: 26202967

26. Liao W-T, Li T-T, Wang Z-G, Wang S-Y, He M-R, Ye Y-P, et al. Human Cancer Biology microRNA-224 Promotes Cell Proliferation and Tumor Growth in Human Colorectal Cancer by Repressing PHLPP1 and PHLPP2. Clin Cancer Res. 19: 1–11. https://doi.org/10.1158/1078-0432.CCR-13-0244 PMID: 23846336

27. Cui R, Meng W, Sun H-L, Kim T, Ye Z, Fassan M, et al. MicroRNA-224 promotes tumor progression in nonsmall cell lung cancer. Proc Natl Acad Sci U S A. National Academy of Sciences; 2015; 112: E4288–97. https://doi.org/10.1073/pnas.1502068112 PMID: 26187928

28. Deng Y, Huang Z, Xu Y, Jin J, Zhuo W, Zhang C, et al. MiR-215 modulates gastric cancer cell proliferation by targeting RB1. Cancer Lett. 2014; 342: 27–35. https://doi.org/10.1016/j.canlet.2013.08.033 PMID: 23981575

29. Jin Y, Liu Y, Zhang J, Huang W, Jiang H, Hou Y, et al. The Expression of miR-375 Is Associated with Carcinogenesis in Three Subtypes of Lung Cancer. Dahiya R, editor. PLoS One. Public Library of Science; 2015; 10: e0144187. https://doi.org/10.1371/journal.pone.0144187 PMID: 26642205

30. Westernberg L, Conche C, Huang YH, Rigaud S, Deng Y, Siegemund S, et al. Non-canonical antagonism of PI3K by the kinase Itpkb delays thymocyte β-selection and renders it Notch-dependent. Elife. eLife Sciences Publications, Ltd; 2016; 5. https://doi.org/10.7554/eLife.10786 PMID: 26880557

31. Bischoff A, Huck B, Keller B, Strotbek M, Schmid S, Boerries M, et al. miR149 functions as a tumor suppressor by controlling breast epithelial cell migration and invasion. Cancer Res. American Association for Cancer Research; 2014; 74: 5256–65. https://doi.org/10.1158/0008-5472.CAN-13-3319 PMID: 25035394

32. Xuan YH, Jung HS, Choi Y-L, Shin YK, Kim HJ, Kim KH, et al. Enhanced expression of hedgehog signaling molecules in squamous cell carcinoma of uterine cervix and its precursor lesions. Mod Pathol. 2006; 19: 1139–47. https://doi.org/10.1038/modpathol.3800600 PMID: 16778829

33. Cavicchioli Buim ME, Gurgel CAS, Gonçalves Ramos EA, Lourenço SV, Soares FA. Activation of sonic hedgehog signaling in oral squamous cell carcinomas: a preliminary study. Hum Pathol. 2011; 42: 1484–1490. https://doi.org/10.1016/j.humpath.2010.12.015 PMID: 21496866

34. Yang Z, Cui Y, Ni W, Kim S, Xuan Y. Gli1, a potential regulator of esophageal cancer stem cell, is identified as an independent adverse prognostic factor in esophageal squamous cell carcinoma. J Cancer Res Clin Oncol. 2017; 143: 243–254. https://doi.org/10.1007/s00432-016-2273-6 PMID: 27680978

35. Justilien V, Walsh MP, Ali SA, Thompson EA, Murray NR, Fields AP. The PRKCI and SOX2 Oncogenes Are Coamplified and Cooperate to Activate Hedgehog Signaling in Lung Squamous Cell Carcinoma. Cancer Cell. 2014; 25: 139–151. https://doi.org/10.1016/j.ccr.2014.01.008 PMID: 24525231

36. Walck-Shannon E, Hardin J. Cell intercalation from top to bottom. Nat Rev Mol Cell Biol. NIH Public Access; 2014; 15: 34–48. https://doi.org/10.1038/nrm3723 PMID: 24355988

37. Hu Y, Galkin A V., Wu C, Reddy V, Su AI. CAFET Algorithm Reveals Wnt/PCP Signature in Lung Squamous Cell Carcinoma. Minna JD, editor. PLoS One. Public Library of Science; 2011; 6: e25807. https://doi.org/10.1371/journal.pone.0025807 PMID: 22016777

38. Xu C, Fillmore CM, Koyama S, Wu H, Zhao Y, Chen Z, et al. Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. Cancer Cell. NIH Public Access; 2014; 25: 590–604. https://doi.org/10.1016/j.ccr.2014.03.033 PMID: 24794706

39. Pannone G, Santoro A, Feola A, Bufo P, Papagerakis P, Lo Muzio L, et al. The role of E-cadherin downregulation in oral cancer: CDH1 gene expression and epigenetic blockage. Curr Cancer Drug Targets. 2014; 14: 115–27. Available: http://www.ncbi.nlm.nih.gov/pubmed/24274398 PMID: 24274398

40. Zhu Q-Q, Ma C, Wang Q, Song Y, Lv T. The role of TWIST1 in epithelial-mesenchymal transition and cancers. Tumor Biol. 2016; 37: 185–197. https://doi.org/10.1007/s13277-015-4450-7 PMID: 26602382

41. Hamilton KE, Noubissi FK, Katti PS, Hahn CM, Davey SR, Lundsmith ET, et al. IMP1 promotes tumor growth, dissemination and a tumor-initiating cell phenotype in colorectal cancer cell xenografts. Carcinogenesis. 2013; 34: 2647–2654. https://doi.org/10.1093/carcin/bgt217 PMID: 23764754

42. Flaveny CA, Griffett K, El-Gendy BE-DM, Kazantzis M, Sengupta M, Amelio AL, et al. Broad Anti-tumor Activity of a Small Molecule that Selectively Targets the Warburg Effect and Lipogenesis. Cancer Cell. 2015; 28: 42–56. https://doi.org/10.1016/j.ccell.2015.05.007 PMID: 26120082

43. Villablanca EJ, Raccosta L, Zhou D, Fontana R, Maggioni D, Negro A, et al. Tumor-mediated liver X receptor-α activation inhibits CC chemokine receptor-7 expression on dendritic cells and dampens antitumor responses. Nat Med. 2010; 16: 98–105. https://doi.org/10.1038/nm.2074 PMID: 20037595

44. Dai Y-B, Miao Y-F, Wu W-F, Li Y, D'Errico F, Su W, et al. Ablation of Liver X receptors α and β leads to spontaneous peripheral squamous cell lung cancer in mice. Proc Natl Acad Sci U S A. National Academy of Sciences; 2016; 113: 7614–9. https://doi.org/10.1073/pnas.1607590113 PMID: 27335465

45. Lin C-Y, Gustafsson J-Å. Targeting liver X receptors in cancer therapeutics. Nat Rev Cancer. 2015; 15: 216–224. https://doi.org/10.1038/nrc3912 PMID: 25786697

46. Köbel M, Weidensdorfer D, Reinke C, Lederer M, Schmitt WD, Zeng K, et al. Expression of the RNA-binding protein IMP1 correlates with poor prognosis in ovarian carcinoma. Oncogene. 2007; 26: 7584–7589. https://doi.org/10.1038/sj.onc.1210563 PMID: 17546046

47. Tessier CR, Doyle GA, Clark BA, Pitot HC, Ross J. Mammary tumor induction in transgenic mice expressing an RNA-binding protein. Cancer Res. 2004; 64: 209–14. Available: http://www.ncbi.nlm.nih.gov/pubmed/14729626 PMID: 14729626

48. Mahapatra L, Mao C, Andruska N, Zhang C, Shapiro DJ. High-Throughput Fluorescence Anisotropy Screen for Inhibitors of the Oncogenic mRNA Binding Protein, IMP-1. J Biomol Screen. 2014; 19: 427–436. https://doi.org/10.1177/1087057113499633 PMID: 24108120

49. Ioannides AS, Massa V, Ferraro E, Cecconi F, Spitz L, Henderson DJ, et al. Foregut separation and tracheo-oesophageal malformations: the role of tracheal outgrowth, dorso-ventral patterning and programmed cell death. Dev Biol. Elsevier; 2010; 337: 351–62. https://doi.org/10.1016/j.ydbio.2009.11.005 PMID: 19913007

50. Gurumurthy S, Hezel AF, Berger JH, Bosenberg MW, Bardeesy N. LKB1 Deficiency Sensitizes Mice to Carcinogen-Induced Tumorigenesis. Cancer Res. 2008; 68. Available: http://cancerres.aacrjournals.org/content/68/1/55

51. Contreras CM, Gurumurthy S, Haynie JM, Shirley LJ, Akbay EA, Wingo SN, et al. Loss of Lkb1 Provokes Highly Invasive Endometrial Adenocarcinomas. Cancer Res. 2008; 68. Available: http://cancerres.aacrjournals.org/content/68/3/759

52. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. *TCGAbiolinks*: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44: e71–e71. https://doi.org/10.1093/nar/gkv1507 PMID: 26704973

53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

54. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

55. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010; 11: R25. https://doi.org/10.1186/gb-2010-11-3-r25 PMID: 20196867

56. Groemping U. Relative Importance for Linear Regression in R: The Package relaimpo. J Stat Softw. 2006; 17: 1–27. https://doi.org/10.18637/jss.v017.i01

57. Lindeman RH, Merenda PF, Gold RZ. Introduction to bivariate and multivariate analysis. Scott, Foresman; 1980.

58. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016; 32: 2847–2849. https://doi.org/10.1093/bioinformatics/btw313 PMID: 27207943