HHS Public Access

Author manuscript

Transl Res. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Transl Res. 2015 December ; 166(6): 568–585. doi:10.1016/j.trsl.2015.08.001.

The Impact of the Cancer Genome Atlas on Lung Cancer

Jeremy Tzu-Huai Chang[1], **Yee-Ming Lee**[2], and **R. Stephanie Huang**[3,*]

[1]The Biological Sciences Collegiate Division, The University of Chicago, Chicago, IL 60637, USA

[2]Center for Personalized Therapeutics, The University of Chicago, Chicago, IL 60637, USA

[3]Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

Abstract

The Cancer Genome Atlas (TCGA) has profiled over 10,000 samples derived from 33 types of cancer to date, with the goal of improving our understanding of the molecular basis of cancer and advancing our ability to diagnose, treat, and prevent cancer. This review focuses on lung cancer as it is the leading cause of cancer-related mortality worldwide in both men and women. Particularly, non-small cell lung cancers (including lung adenocarcinoma and lung squamous cell carcinoma) were evaluated. Our goal is to demonstrate the impact of TCGA on lung cancer research under four themes: namely, diagnostic markers, disease progression markers, novel therapeutic targets, and novel tools. Examples were given related to DNA mutation, copy number variation, mRNA, and microRNA expression along with methylation profiling.

Keywords

TCGA; lung squamous cell carcinoma; lung adenocarcinoma

Overview of TCGA

The Cancer Genome Atlas (TCGA) project is a major collaborative effort being undertaken in the United States to advance our “understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing.” [1] Since cancer is viewed as a complex genetic disorder, there is sustained interest in finding the genetic pathways and aberrant genomic changes that precipitate the development of tumors. The ultimate goal of TCGA is to “improve our ability to diagnose, treat, and prevent cancer” through the discoveries and insights enabled by the exhaustive mapping of various forms of cancer. [1]

*Address for correspondence and reprints: R. Stephanie Huang, 900 E 57th street, KCBD room 7148, The University of Chicago, Chicago, IL 60637. Telephone: (773) 702-9363; Fax: (773) 702-0963; rhuang@medicine.bsd.uchicago.edu.

Conflict of Interest: All authors have read the journal’s policy on disclosure of potential conflicts of interest and they have none to declare.

Originally established by the National Institute of Health (NIH), TCGA is jointly run by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). The three-year pilot project of TCGA began in 2006 to develop the policies and infrastructure to handle the high-volume data generated from patient samples. In 2009, the NCI and NHGRI announced that TCGA was ready to move forward into a new five-year project. This project would comprehensively analyze the genomic landscape of over 20 cancers, ranging from glioblastomas to prostate adenocarcinomas. [2]

The cancers studied by TCGA were selected for their "poor prognosis and overall public health impact," and the "availability of human tumor and matched-normal tissue samples that meet TCGA standards for patient consent, quality, and quantity." [3] The general strategy for this unprecedented effort was to collect large samples of each tumor type along with matched normal tissue, with some tumor types receiving over 500 unique samples.

In order to generate comprehensive molecular profiles for each tumor type, different technologies were employed by TCGA, such as whole genome and exon sequencing, SNP genotyping, copy number variation profiling using microarrays (i.e. Agilent, Illumina), DNA methylation profiling, genome wide expression, functional proteomic analysis, and microRNA expression profiling through RNA sequencing. This rich amount of information generated from TCGA was then consolidated and shared with the community.

By January 2015, TCGA announced it had successfully collected "the necessary quality and quantity of samples" for all 33 selected tumor types. Table 1 shows a summary of the available TCGA data for these 33 tumor types. To date, publicly available genomic data has been shared online for 29 of the 33 tumor types. As of 2015, breast invasive carcinoma (BRCA) comprised the highest number of data with 1098 cases, while cholangiocarcinoma (CHOL) had the lowest number with 36 cases. [4] TCGA data has spawned over 2,700 papers in scientific journals, demonstrating the far-reaching effects of TCGA data on cancer research. [5] As TCGA winds down, new projects such as PanCanAtlas and Pan-Cancer Analysis of Whole Genomes (PCAWGs) are now underway to further analyze TCGA data and to gain an even clearer understanding of the molecular mechanisms of cancer. [5] This review seeks to focus on lung cancer as it is the leading cause of cancer-related mortality worldwide in both men and women. The review aims to demonstrate the impact of TCGA on lung cancer research under four themes: namely, diagnostic markers, disease progression markers, novel therapeutic targets, and novel tools.

Lung Cancers

Lung cancer is the leading cause of cancer-related death worldwide, accounting for 1.59 million deaths out of the 8.2 million total cancer deaths in 2012. [6] There are two main types of lung cancer: small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). NSCLC accounts for 85–90% of lung cancer cases and its two largest subtypes are lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD). [7] LUSC accounts for 25–30% of all total lung cancer cases while LUAD accounts for 40% of all total lung cancer cases. LUSC tends to be found in the middle of the lungs and is associated

with smoking. In contrast, LUAD forms on the periphery of the lungs and may be associated with smoking, but is the most common lung cancer type among non-smokers. [7]

The lack of effective therapeutic options to treat NSCLC patients contributes to poor outcomes and to the high number of lung cancer deaths. NSCLCs have historically been resistant to traditional platinum-based chemotherapy with response rates of only 20–50%, compared to advanced SCLCs that have chemotherapy response rates of 60–80%. [8] This resistance has proven to be a challenge as the dose response curve for advanced NSCLC plateaus at higher doses, resulting in poor prognosis and low 5-year survival rates. [9, 10] The increased focus on driver mutations in tumorigenesis provided critical insight for personalized therapeutics in the treatment of NSCLCs. Research into the mutations of the epidermal growth factor receptor (*EGFR*) and the translocation of anaplastic lymphoma kinase (*ALK*) have been fertile ground for the treatment of NSCLCs. [11–13] Targeted therapies for the active oncogenic *EGFR* mutation have significantly improved treatment outcomes, especially for LUAD since this NSCLC type is the most likely to contain *EGFR* aberrations.[11,12,14]

Despite the effective use of epidermal growth factor receptor tyrosine kinase inhibitor (*EGFR*-TKI) drugs such as gefitinib to treat *EGFR*-mutated cancers, drug resistance from mutations such as the threonine-790 to methionine (T790M) point mutation are major obstacles in the treatment of NSCLCs. [15] Pan-cancer studies have also found LUAD to be highly heterogeneous, adding to the challenge of providing effective treatment. [16] Furthermore, therapies for LUAD have been proven largely ineffective for treating LUSC. [17] A more comprehensive understanding of the genetic pathways of LUAD and LUSC is needed to develop better diagnostic tools and treatment options. The lung cancer project of TCGA has provided the opportunity to study NSCLCs in depth by amassing large amounts of genomic information from hundreds of cases.

TCGA Coverage for Lung Cancers

LUAD and LUSC were two of the three lung tumors to be studied by TCGA, with the third being mesothelioma. TCGA had three main goals for its lung cancer research: (1) to identify the specific genes that differentiate LUSC and LUAD tumors into molecular subgroups, (2) to differentiate broader gene expression patterns between LUAD and LUSC, and (3) to distinguish genomic changes between smokers and non-smokers. As of June 2015, 521 samples of LUAD and 504 samples of LUSC have been analyzed and their data submitted to TCGA's data portal. The two landmark TCGA lung cancer publications for LUAD and LUSC, however, focused on a subset of 230 and 178 tumors, respectively. [4] There were four major goals from these two landmark studies. The first was to identify the genomic differences between tumor tissue and matched normal tissue, and this involved using whole-exome sequencing, RNA sequencing, SNP array, and methylation profiling. The second goal was to relate these genomic differences to molecular pathways that drive carcinogenesis. The last two goals were to determine the impact of smoking on tumor genetics and to identify potential target genes to improve NSCLC treatment. The major findings from these two publications are summarized in Table 2. [18, 19]

Specifically, in the landmark study for LUSC, CpG transitions and transversions were the most commonly observed mutation types. Mutations in *TP53* were the most common abnormality, occurring in 81% of samples, but this figure could have been as high as 90% if it were not for suboptimal coverage or poor sample purity. This study also discovered that 44% of the samples had alterations in genes involved with squamous cell differentiation. A surprise find was that alterations in *NOTCH1*, *NOTCH2*, and *ASCL4* were mutually exclusive with *TP63* and/or *SOX2* alterations. This suggests that abnormalities in squamous cell differentiation genes have overlapping functional consequences. Given that 7% of the LUSC cases possessed amplifications of *EGFR*, the TCGA study suggested that some forms of LUSC may respond to *EGFR*-TKIs such as erlotinib and gefitinib. The study identified 114 (63%) cases with somatic alterations of a potentially targetable gene. [18]

In the landmark study for LUAD, 75% of the somatic mutations in whole-exome sequencing were present in the RNA transcriptome. Ten tumors had somatic *MET* DNA alterations with *MET* exon 14 skipping in RNA. DNA hypermethylation commonly occurred for the *CDKN2A* and *GATA2* genes in CpG island methylator phenotype- high (CIMP-H) tumors. Another notable aspect associated with CIMP-H tumors was the overexpression of *MYC* ($P = 0.003$). For important pathways, TCGA found that 76% (175/230) of LUAD tumors contained known activating mutations in driver oncogenes that precipitated the RTK/RAS/RAF pathway. The study classified the 230 TCGA LUAD samples into a transversion high and a transversion low cohort, with the transversion high group found to be strongly associated with past or present smoking as well as *KRAS* mutations. The low transversion cohort was associated with mutations in *EGFR*, as well as *PIK3CA* and *RB1*. With these new findings, increased implementation of *MET* and *ERBB2/HER2* inhibitors may improve treatment for LUADs with oncogene-negative tumors. [19]

Additional Research Based on TCGA Lung Cancer Data Sets

The datasets generated from TCGA LUAD and LUSC have served as a basis for numerous research projects ranging from identifying new diagnostics markers to developing new tools and algorithms. In addition to research that focuses primarily on these two lung cancers, there have been multiple studies comparing LUAD and LUSC to each other and with the other cancers evaluated by TCGA.

We have surveyed the literature for studies derived from TCGA LUAD/LUSC to demonstrate the diversity of research occurring with the usage of TCGA data, and organized this research into four themes with provided examples. The four themes are: molecular mechanisms leading to disease (diagnostic markers); molecular mechanisms related to prognosis (disease progression prognostic markers); identifying novel therapeutic markers; and building tools using TCGA data. We have attempted to emphasize the potential practical uses of many of the studies featured to demonstrate the salutary effects of the TCGA lung cancer project.

Molecular Mechanisms Leading to Disease (Diagnostic Markers)

With the vast amount of genomic data generated from tumor and normal tissues, many pathways, genes, and genomic phenotypes have been proposed to be potential diagnostic

markers. Table 3 shows examples of studies that have examined molecular mechanisms leading to both LUAD and LUSC.

LUAD—Here, we describe the role of two genotypes and a phenotype in LUAD: the *KRAS* and *CDKN2AB* genes [20], and the CIMP phenotype [21].

Combination of Genetic Events Leads to Increased Cancer Risk: The *KRAS* gene is a small GTPase that is mutated in a high number of LUADs. [22] Mutation in *KRAS* is not an independent mutation that singularly drives carcinogenesis, but is often associated with mutations in the *CDKN2AB* locus as well. [23–26] The *CDKN2AB* locus is important for encoding several tumor suppressors; *CDKN2A* contains $p16^{INK4A}$ and $p14^{ARF}$ while *CDKN2B* contains $p15^{INK4B}$. Both $p16^{INK4A}$ and $p15^{INK4B}$ have been shown to inhibit the *CDK4/CDK6* retinoblastoma family of tumor suppressor genes to induce G1 cell arrest while $p14^{ARF}$ has been shown to upregulate *TP53*. [27, 28] With the use of two genetically-engineered mouse models that contain *KRAS* mutations of the respiratory epithelium and human lung cancer cell lines, the featured study found that mutant *KRAS* mice conditionally deficient in the entire *CDKN2AB* locus experienced higher lung cancer mediated mortality. [20] Deficiency of the entire *CDKN2AB* gene promoted the development of high-grade tumors as *KRAS* mutant mouse displayed an increased number of tumors and increased tumor burden. *KRAS* mutant mice that were deficient only of *CDKN2A* demonstrated an intermediate number of tumors and tumor burden while the mice with intact *CDKN2AB* loci demonstrated the highest survival.

In analyzing TCGA LUAD datasets (n=323), the study found that 30% (n=96) of the samples contained oncogenic *KRAS* mutations. Around 53% of the 96 (n=51) patients had two-hit inactivation of *CDKN2A* and 33% of those 51 (n=17) also showed inactivation of *CDKN2B*. Patients with *CDKN2A* inactivation experienced poorer overall survival (OS) than the average patient with *KRAS*- mutated LUAD. While patients with the additional deficiency of *CDKN2B* tended to have even poorer OS, the results were not statistically significant. This study confirmed that the loss of function for *CDKN2AB* can drive cell proliferation and metastatic behavior in *KRAS*- mutated lung carcinomas. The study also highlighted a subset of *KRAS*- mutated lung cancers that may be sensitive to focal adhesion kinase inhibitors since the extracellular signal-regulated kinase (*ERK*)/*RHOA*/focal adhesion kinase (*FAK*) network is frequently deregulated in *KRAS*-mutated and *CDKN2A*- inactivated NSCLCs, especially for LUAD. [6] The results of this study may improve diagnostic abilities to identify at-risk groups within *KRAS*-mutated LUADs and improve clinical outcomes for these patients. [20]

Methylation Status in LUAD Risk: CIMP has been studied in LUAD as a diagnostic marker using TCGA data. [21] Using methylated CpG island amplification microarray (MCAM) analysis for a set of 41 LUAD samples received from various hospitals in Japan, the study identified six markers for CIMP in LUAD: *CCNA1*, *ACAN*, *GFRA1*, *EDARADD*, *MGC45800* and *p16* (*CDKN2A*). With these six markers, the study found that 4 (10%) of the LUAD samples from Japan were methylated in at least four of the six markers, 11 (27%) samples were methylated for two to four of the markers, and 26 (63%) samples were

methylated in only one or none of the six markers. To confirm the validity of these markers, a subset of 85 TCGA LUADs were analyzed. The methylation data were obtained by the Infinium Beadchip using 27,578 probes corresponding to 14,473 genes. The study evaluated three markers- *ACAN*, *CCNA1* and *GFRA1*- on the 85 TCGA samples and found that no single sample was methylated in all three CIMP markers, 4 (4.7%) were methylated in two CIMP markers, 14 (16.5%) were methylated for one CIMP marker, and 67 (78.8%) demonstrated no methylation for any of the three markers. The LUAD samples that were methylated for one or two of the markers had overall higher rates of methylation, indicating that the three CIMP markers could predict for higher methylated LUAD.

When combining the Japanese and TCGA LUAD cases (n=128), the subtypes of CIMP LUAD were classified as CIMP-H(igh), CIMP-L(ow), and CIMP-N(egative). CIMP-H accounted for 10 of the samples (7.8%), CIMP-L for 40 (31.3%) and CIMP-N for 78 (60.9%). CIMP-H tumors were more common among males, associated with frequent exposure to smoking, and did not possess *EGFR* mutations. CIMP-H was found to be a poor prognostic factor regardless of the presence of the *EGFR* mutation. For the CIMP-L cohort, the study did not define any distinct characteristics other than its intermediate accumulation of DNA methylation. Fourteen different adenocarcinoma cell lines (one CIMP-H, seven CIMP-L, and six CIMP-N) were treated with the DNA methylation inhibitor 5-Aza-dC to test the effectiveness of epigenetic therapy for CIMP-positive adenocarcinomas. The study found that CIMP-positive adenocarcinoma cells were sensitive to 5-Aza-dC, suggesting that DNA methylation inhibitors may be useful for treating LUAD patients diagnosed with CIMP. The study identified CIMP phenotypes within LUAD that may improve diagnostic classifications with the purpose of providing more personalized therapies for tumor subtypes such as CIMP-H LUADs [21].

LUSC—Similar to LUAD, the elucidation of individual genes, pathways, and phenotypes that drive or characterize LUSC is among the most prevalent forms of research. One study that investigated the oncogenic pathways of LUSC employed Gene Set Enrichment Analysis (GSEA) and Signaling Pathway Impact Analysis (SPIA) to TCGA datasets and ranked the pathways most associated with LUSC tumorigenesis. The top ranked pathways were the viral carcinogenesis pathway, the p53 signaling pathway, and the PI3K-Akt signaling pathway. [29] Another study investigated the regulatory effect of *SOX2* on *FOXA1* and found that *SOX2* was a negative upstream regulator of *FOXA1*. These two genes were observed to be altered in 46.4% of 179 TCGA LUSC samples, and the findings of the study may possess future promise in improving the diagnostic efficiency for LUSCs as *SOX2* alterations are common for LUSC. [30] This review explores here the co-amplification of the genes *SOX2* and *PRKCI* that drive a stem-like phenotype in LUSC tumors. [31]

Gene Coordination for LUSC Maintenance and Growth: The Cancer Stem Cell hypothesis states that a small population of cancer stem cells creates heterogeneous, hierarchical structures among tumor cells [32]. There has been research demonstrating that stem-like cells in NSCLCs play a major role in tumor initiation, relapse, and metastasis. [33–35] The *SOX2* gene has been implicated to be vital for the maintenance of cancer stem cells, and the gene *PRKC1* is a known oncogene found to be overexpressed in LUSC cells

due to the amplification of 3q26 in some LUSCs. [36, 37] The featured study first isolated stem-like cells from five human lung cancer cell lines that contain 3q26 copy number gains (H1299, H1703, ChagoK1, H520 and H1869). Three of the cell lines-H1299, H1703 and ChagoK1- grew as cell spheres (oncospheres) that exhibited numerous stem-like properties such as elevated *SOX2* expression, augmented cell proliferation, and increased soft agar growth. The oncospheres also exhibited increased *PRKCI* activity. *PRKCI* RNAi was found to severely impair soft agar growth and cell proliferation of the oncospheres, suggesting that *PRKCI* is crucial for maintaining cancer cell oncosphere stemness. The study investigated *PRKCI*-regulated pathways through RNA-seq data and found that *PRKCI* impacted the Hedgehog (Hh) pathway by regulating Hh components such as Hedgehog Acyl Transferase (*HHAT*), *ADRBK1*, *CDK19* and *GLI1*. By inhibiting *PRKCI* activity in the oncospheres, *HHAT* and *GLI1* expression was decreased in addition to the loss of stem-like properties, indicating that *PRKCI* and the Hh pathway are vital for cancer stem cells. The effects of *HHAT*, *GLI1*, and *PRKCI in vivo* were also studied with immune deficient mice. When *PRKCI*, *HHAT*, and *GLI1* RNAi oncospheres were introduced, the mice showed a decreased take-rate and failed to produce large tumors compared to controls. This confirmed the importance of *PRKCI*, *HHAT* and *GLI1* for tumor growth *in vivo*. The same procedures were also carried out for *SOX2* as the gene of interest. The conclusion was that *SOX2* also regulated the Hh pathway, especially for *HHAT*.

TCGA RNA-seq data for LUSC were then analyzed to find out if *SOX2* and *PRKCI* were coordinately amplified and overexpressed in LUSCs. The study found that the two genes were indeed frequently co-amplified and overexpressed with a strong positive correlation between the expression of the two genes and the Hh components *HHAT* and *GLI1*. To demonstrate that *SOX2* and *PRKCI* regulated *HHAT* and *GLI1*, the study used tumor cell lines from LUSC patients and performed either RNAi for the two genes or *HHAT* knock-down (KD). RNAi for either *SOX2* or *PRKCI* led to decreased *HHAT* expression, but *HHAT* KD did not generate any expression changes for *SOX2* or *PRKCI*. The importance of the study is the identification of LUSC tumors that rely on the coordination of *SOX2* and *PRKCI* to regulate the Hh pathway for cancer growth and maintenance. The discovery of *SOX2* and *PRKCI* gene coordination may prove beneficial as therapies targeting the Hh pathway, *SOX2*, or *PRKCI* may be especially effective in the future against some forms of LUSC. [31]

Diagnostic Markers for Both LUAD and LUSC—Pan-cancer analyses and studies regarding TCGA NSCLC are common, but are also diverse. For example, one pan-cancer study examined the presence of viral DNA transcripts in 3,775 neoplasms from TCGA RNA-seq data. Human papillomavirus (HPV) transcripts were detected in LUSC along with head-and-neck squamous cell carcinoma (HNSC) and uterine endometrioid carcinoma (UCEC). No DNA virus transcripts were detected in LUAD along with acute myeloid leukemia (LAML), cutaneous melanoma (SKCM), low- and high-grade gliomas of the brain (LGG and PCPG), and many of the other adenocarcinomas. An understanding of viral integration sites may provide clues for understanding the mechanisms of tumorigenesis for cancers such as LUSC. [38] As the employment of histology-based therapies increase, the accurate identification of LUSC and LUAD has increased in importance; *CLCA2* is one possible marker that may improve lung cancer diagnoses. [39]

Potential Immunohistochemical Marker for Differential Diagnosis: One of the main goals of TCGA research into NSCLC is to pinpoint genetic differences between LUAD and LUSC. The utility of TCGA data has led to the discovery of *CLCA2* as a novel immunohistochemical marker to differentiate LUAD from LUSC. [39] The study gathered RNA-seq data for 490 LUAD cases and 490 LUSC cases from TCGA. The identification of LUSC-specific genes were selected based on two conditions: (1) the median gene expression value had to be higher than 1,500 in LUSC specimens, and (2) the median gene expression value had to be less than 100 in LUAD specimens. A total of 2,693 genes (13.1%) met condition (1), 8,658 genes (42.2%) agreed to condition (2), and 24 genes satisfied both conditions, garnering the classification as LUSC-specific genes.

The gene *CLCA2* was selected because its protein expression had not been studied in lung cancers previously and the difference in the median expression value of the gene between LUSCs and LUADs was extremely large (4,860.0234 vs. 5.2242). A set of 235 LUAD and 161 LUSC paraffin-embedded tissue samples from Hamamatsu University Hospital were used for tissue microarray (TMA) blocks to confirm the difference in *CLCA2* expression. An immunohistochemical analysis using anti-*CLCA2* polyclonal antibody was performed with the TMA section to observe *CLCA2* immunoreactivity in the cytoplasm and membrane of the cancer cells. It was revealed that the immunohistochemical scores for LUSC were significantly higher than those for LUAD (median, 170 vs. 0; $P < 0.0001$, Mann-Whitney U test). When the immunohistochemical score of 100 was used as the cut-off value for defining positive *CLCA2* expression (score > 101) and negative *CLCA2* expression (score ≤ 100), LUSC samples with positive-*CLCA2* expression (104/161, 64.6%) had a much higher frequency than LUAD samples with positive-*CLCA2* expression (2/235, 0.9%). This research highlighted a novel genetic difference between LUAD and LUSC that could be a promising immunohistochemical marker for differentiating between the two cancers. [39]

Molecular Mechanisms Related to Prognosis (Disease Progression Markers)

There has been a significant effort for the discovery of prognostic markers with TCGA data. Table 4 describes examples of disease progression markers studied in LUAD and LUSC.

LUAD—The investigation of intra-tumor signaling entropy as a prognostic tool for *Stage 1* LUAD is one example of LUAD research in disease progression markers. Relying on the Cancer Stem Cell hypothesis, the study proposed that intra-tumor signaling entropy is an effective measure of intra-tumor heterogeneity and tumor stemness with predictive abilities for clinical outcomes. [40] The study found that higher signaling entropy in *Stage 1* LUAD patients was associated with worse prognosis and observed that the top third of the signaling entropy distribution conferred almost a doubling of the hazard rate for mortality, as assessed over the first three years after diagnosis. [41] In addition, studies have been focused on the use of microRNAs (miRNAs) for LUAD prognostic models. [42, 43]

MicroRNAs as Prognostic Markers for LUAD: miRNAs are believed to regulate gene expression post-transcriptionally by binding to mRNA, and have been observed to be dysregulated in cancers. [44, 45] Data of 372 LUAD patients from TCGA including miRNA expression, patient outcome, and cancer staging information were utilized in this featured

study. Common miRNA related to OS after controlling for disease stage (−N, M, and T stage) were identified through univariate survival analyses on a training set of 186 TCGA LUAD samples using a p-value of 0.1. A prognostic model of eight miRNA markers (miR-31, miR-196b, miR-766, miR-519a-1, miR-375, miR-187, miR-331 and miR-101-1) was constructed as a weighted linear combination of the detected miRNA expression levels. Prognostic scores were determined through the equation: Prognostic-score = (0.181 × expression level of miR-31) + (0.136 × expression level of miR-196b) + (−0.114 × expression level of miR-375) + (−0.148 × expression level of miR-187) + (−0.352× expression level of miR-331) + (−0.372× expression level of miR-101-1) + (0.182× expression level of miR-766) + (0.21× expression level of miR-519a-1). Prognostic scores were calculated in another set of 186 TCGA LUAD samples in order to test the predictive ability of the model. The prognostic scores were classified as high-risk or low-risk by employing the median score from the training set of 186 LUAD patients.

It was found that increased expression of four out of the eight miRNAs was associated with higher mortality and subsequently defined as high risk (hsa-miR-31, miR-196b, miR-766, miR-519a-1). Increased expression of the other four miRNAs was associated with increased survival and defined as protective (miR-375, miR-187, miR-331, miR-101-1). Tumors with high prognostic scores for poor outcomes tended to express high-risk miRNAs and tumors with low prognostic scores tended to express protective miRNAs. Patients with high prognostic scores had poorer OS compared to the OS of patients with low prognostic scores (median OS of 39.0 months *vs.* 59.3 months, HR = 1.99, *P* value < 0.001). The eight miRNA signature was found to be heavily correlated with lymph node metastasis (*P* = 0.0085) and clinical stage (*P* = 0.0252). When the LUAD cohort was separated by smoker and non-smoker, the eight miRNA signature proved to be especially accurate as a five-year prognostic model for patients who were non-smoking or had not smoked in 15 years. The eight miRNA signature has clinical significance since the signature can help clinicians identify LUAD patients who have high risk for poorer OS. [42]

A second miRNA study focused specifically on the role of miR-31, one of the high-risk miRNAs in the eight miRNA signature mentioned previously. [43] The study confirmed that miR-31 expression is highly correlated with lymph node metastasis in LUAD patients. Out of a potential 43 cases with 64 tissue specimens available from LUAD patients through the Ohio State University Comprehensive Cancer Center (OSUCCC) Tissue Procurement Shared Resource, 10 cases were analyzed by performing genome-wide miRNA-seq on tissues. Four of the 10 cases displayed lymph node metastasis (N1+) and the other six cases lacked evidence of lymph node metastasis (N0). miR-31 was found to be upregulated in the N1+ cohort through a comparison of the qPCR miR expression of the N1+ and N0 groups, which was consistent with other previous studies that observed miR-31 up-regulation in metastatic colon cancer. [46] However, these findings were inconsistent with another study that reported miR-31 up-regulation was correlated with lower frequency of breast cancer metastasis. [47]

TCGA data from 249 patients were then used to validate the correlation between increased expression of miR-31 and lymph node metastasis in LUADs. It was discovered that LUAD patients with lymph node metastasis had a 2.5 fold increase in miR-31 expression compared

to LUAD patients without lymph node metastasis. The multivariate Cox proportional hazards modeling of 233 LUAD cases with eight different factors (miR-31 expression, age, sex, lymph node stage, tumor stage, race, prior cancer diagnosis, and smoking history) revealed that the three most significant predictors of survival were miR-31 expression, nodal stage, and previous cancer diagnosis—indicating miR-31 expression has a prognostic value independent of lymph node status.

To gain a functional understanding of miR-31 expression, three LUAD cell lines (H23, H2228, and H1573) were tested *in vitro* for migration, invasion, and proliferation with the modulation of miR-31 expression level. The experiment demonstrated that the transduction of Lenti-miR containing miR-31 precursor (to cause miR-31 overexpression) significantly increased the migratory and proliferative ability of H23 and H2228 cells. Pathway analysis revealed that the *ERK1/2* pathway was one of the top predicted pathways affected by miR-31 expression. The inhibition of the *ERK1/2* pathway in the miR-31 overexpressed cells was found to suppress the gains in proliferative and migratory abilities of the cell lines, suggesting that *ERK1/2* is one of the primary pathways regulated by miR-31. The study supported previous reports that miR-31 leads to *ERK1\2* pathway activation and promotes cell migration, invasion, and proliferation. [48] This study highlighted the importance miR-31 may have as a disease progression marker in LUADs. [43]

LUSC—The exploration for prognostic biomarkers with TCGA LUSC data is just as diverse as for TCGA LUAD. For example, one study evaluated the mRNA expression of insulin receptor isoform A (*IR-A*) and insulin receptor isoform B (*IR-B*) in NSCLCs, including LUSC. The study found that decreased *IR-B* expression resulted in an elevated *IR-A/IR-B* ratio, which was associated with lower epithelial-mesenchymal transition gene signatures in NSCLCs overall and longer patient survival under standard of care for LUSC. [49] Another study evaluated a wide range of novel cancer drivers for LUSC and found that patients with high expression of the *ARHGDIB* gene had good prognoses while patients with high expression of the *HOXD3* gene had poor prognoses. [50] Another paper studied whether the addition of multiple genomic components to existing prognostic models for LUSC improved predictive powers and found no significant improvement, suggesting that measurements at the transcription level may be more predictive of clinical outcomes than measurements at the DNA\ epigenetic level for LUSC. [51]

PIAS3 Expression as a Prognostic Marker for LUSC: Signal Transducers and Activators of Transcription 3 (*STAT3*) is a cytoplasmic transcription factor that drives uncontrolled growth signaling, leading to tumorigenesis when activated. [52] Previous research has demonstrated that *STAT3* is an integration point of common cancer drivers such as the *EGFR* mutation, making the *STAT3* pathway a promising therapeutic target for cancer treatment. [53] Furthermore, the role of the *STAT3* pathway in predicting patient survival for LUSC has been suggested and is explored here. [54] The protein inhibitor of activated *STAT3* (*PIAS3*) is an endogenous inhibitor that has been observed to inhibit growth in U251 human glioblastoma cells. [55–57] *PIAS3* levels in LUSC tissue samples were investigated in this study along with an analysis of TCGA LUSC cases. [51] Western blotting of patient tumor tissues from the Lung Cancer Biospecimen Resource Network (LCBRN, University

of Virginia) confirmed that the majority of LUSC samples demonstrated minimal *PIAS3* expression when compared to a normal lung epithelial cell line and an adenocarcinoma cell line.

To explore the potential clinical impact of low *PIAS3* expression for LUSC, TCGA mRNA transcripts and patient survival data for 133 individuals were correlated. The cohort of patients with the highest level of *PIAS3* expression (<75%) demonstrated the highest OS of any group while the cohort with the lowest level of expression (<25%) demonstrated the lowest OS. A Cox regression analysis revealed a hazard of 0.57 (95% CI: 0.37–0.87), demonstrating a decrease in the risk of death by 43% for every 1 unit elevation in *PIAS3* gene expression. The same study also tested the effects of increased *PIAS3* expression on LUSC cell lines, specifically Calu-1 cells. When the Calu-1 cells were treated with varying levels of curcumin, *PIAS3* expression increased as early as 30 minutes after the treatment and remained elevated for at least 48 hours in the Calu-1 cells. In measuring the cell viability of the Calu-1 cells, there was a dose-dependent decrease in the cells after prolonged curcumin treatment. Further analysis found a higher proportion of cells in the subG1 phase for the curcumin-treated Calu-1 cells than the control cells, revealing that increased *PIAS3* levels inhibited cell growth for Calu-1 cells. This experiment demonstrated that therapeutic strategies for promoting *PIAS3* expression in LUSC may be promising, although better small molecule activators of *PIAS3* must be found to develop effective treatments. [52]

Prognostic Markers for both LUAD and LUSC—Efforts to identify prognostic markers in lung cancer relying on TCGA data have been abundant because of the large sample of NSCLC cases made accessible by TCGA. One study examined 753 lung cancer samples (521 TCGA cases) for gene fusions. The study not only observed that *NRG1, NF1* and *Hippo* pathway fusions play significant roles in oncogene-negative tumors, but that a high number of gene fusions is an independent prognostic factor for poor OS. [58] The ability of using breast cancer anti-estrogen resistance 1 (*BCAR1*) as a prognostic marker for both LUAD and LUSC will be discussed in detail. [59]

BCAR1 as a Potential Prognostic Marker for Specific Demographic Groups: Breast cancer anti-estrogen resistance 1*(BCAR1*), also known as *p130cas*, is a *CAS* protein (*Crk*-associated substrate) family member that participates in apoptosis, cell cycle, migration, differentiation, and the carcinogenic behavior of cancer cells. [60–62] While it has been previously found that elevated levels of *BCAR1* expression in tumors predict poor prognosis in a group of 151 NSCLC patients [63], TCGA data for LUAD and LUSC with additional data from the Mayo Clinic (Rochester, MN, USA) and Daping Hospital (Chongqing city, China) were analyzed for *BCAR1*-mRNA or -protein to confirm the prognostic ability of *BCAR1* expression. Matched tissues from 77 LUAD cases of never-smokers were obtained through the Mayo Clinic, and analyzed by a gene expression microarray. Tissues of 150 NSCLC cases from Daping Hospital were analyzed for *BCAR1*-protein expression using tissue microarray and immunohistochemistry. The RNA-seq datasets for 508 NSCLC cases were downloaded from the data portal of TCGA. Student's t-test was used to analyze mRNA and protein differential expression while the Cox proportional hazards model was performed to examine prognostic factors within the TCGA cohort. From the 77 matched LUAD cases

of the Mayo Clinic, increased *BCAR1*-mRNA expression was significantly correlated with advanced tumor-stage (P<0.001) and lymphatic metastasis (P<0.001).

TCGA RNA-seq data corroborated these observations for LUAD, but not for LUSC. The datasets of TCGA also showed no appreciable differences of *BCAR1*-mRNA between LUAD and LUSC (t test, P=0.707). The *BCAR1*-protein analysis revealed that elevated *BCAR1*-protein levels were significantly correlated with lymphatic metastasis in LUADs. No correlation for LUSC was found with *BCAR1*-protein levels, and there were no significant differences in *BCAR1*-protein expression between the two NSCLC subtypes. In analyzing OS, the Cox model for the 77 cases from the Mayo Clinic did not reveal a significant association between *BCAR1* expression and OS (95.0% CI for HR: 0.697–2.968). However, in TCGA data for both LUAD (HR=1.776, 95.0% CI for HR: 1.159–2.722) and LUSC cases (HR=1.566, 95.0% CI for HR: 1.082–2.266), there was a significant correlation between *BCAR1* expression and OS, demonstrating a 56–77% increased risk of death when *BCAR1* expression was doubled. The results of the study suggest that *BCAR1* expression in NSCLCs may find use as a prognostic signatures for some demographic groups, but the prognostic ability of *BCAR1* and the specific pathways associated with *BCAR1* need to be further corroborated and analyzed. [11]

Molecular Mechanisms Related to Sensitivity (Novel Therapeutic Targets)

Table 5 shows examples of recent research for the discovery of novel therapeutic targets for LUAD and LUSC.

LUAD—As described earlier, LUAD treatment experienced a great boon with targeted therapies involving *EGFR* and *ALK* aberrations, but the search for additional and more wide-ranging therapeutic targets continues. One study examined the regulatory role of the miR-200 family in tumor angiogenesis and the effects of increased miR-200 expression in tumors. Using miRNA-expression datasets from TCGA, the study first identified that low expression of several miR-200 isoforms was significantly associated with worse OS for LUAD patients. The therapeutic effects of miR-200 were tested with mice bearing 344SQ subcutaneous lung tumors. The study observed that delivery of miR-200a and -200b members led to a significant decrease in primary tumor burden compared with control miRNA expression. This research provides glimpses into the promise of miRNA-based therapies for cancer treatment. [64] This review will discuss the differential gene expression of tristetraprolin (*TTP*) in tumors and the possibility of targeted therapies for the *TTP* gene in LUADs. [65]

Potential Therapy for Tristetraprolin-Low LUADs: Tristetraprolin (TTP, ZFP36) is a newly discovered tumor suppressor that functions by transporting mRNAs with adenosine-uridine (AU)-rich elements (AREs) located in their 3' untranslated regions (3'UTRs) to cell processing centers where the mRNA undergoes deadenylation, decapping, and degradation by a series of mRNA decay enzymes. [66] Enforced *TTP* expression in *MYC*-driven mouse model of B cell lymphoma has been observed to disable maintenance of *MYC*-driven lymphoma. [67] Reduced levels of *TTP* in human breast cancer have been correlated with high tumor grade and poorer outcomes. [68] A set of 355 LUAD cases and 260 LUSC cases

were downloaded from TCGA to investigate the importance of *TTP* in LUAD, LUSC, and other cancers. RNA-sequencing by Expectation-Maximization (RSEM) normalized count was used to analyze isoform level transcription estimates for each TCGA case. Upper and lower quartile groups (*TTP*-high and *TTP*-low) were established based on *TTP* expression within each cancer. The NextBio Research platform and Ingenuity Pathway Analysis (IPA) software (Qiagen) were used to identify mRNA biosets and canonical pathways associated with the *TTP*-gene signature. A total of 50 genes were identified to be differentially expressed between the *TTP*-high and *TTP*-low tumors.

In analyzing the OS data from TCGA, LUAD patients with *TTP*-low tumors displayed lower survival rates compared to patients with *TTP*-high tumors. TCGA LUSC patients, however, did not demonstrate significant differences in survival rate between patients with *TTP*-low tumors and patients with *TTP*-high tumors. There are three molecular subtypes of LUAD: bronchioid, magnoid, and squamoid. [69] Bronchioid LUAD has been observed to have the most favorable patient outcome of the three molecular subtypes. [69, 70] Over 60% of the patients of the *TTP*-high cohort were of the bronchioid subtype. On the other hand, the magnoid and squamoid subtypes were much more prevalent in the *TTP*-low cohort of LUAD patients (36% and 44%, respectfully). It was also noted that nearly twice as many *TTP*-low LUAD patients had *Stage 3* or *Stage 4* tumors (27% combined) than *TTP*-high patients (14% combined).

The cyclic AMP response element (CRE)-binding protein (*CREB*) family of activators (*CREB1*, *CREM*, *ATF1*) was discovered to be the top upstream regulator of *TTP*. Although *CREB* expression was not significantly altered between the *TTP*-high and *TTP*-low cohorts, Activating Transcription Factor 3 (*ATF3*) was significantly reduced in *TTP*-low LUAD. Additionally, the expression of *CREM* transcript variant 2, which harbors the N-terminal domain of *CREM* but lacks the bZIP domain (CREM ZIP), is believed to compete with *CREB* family members for transcriptional cofactors. CREM ZIP was found to be significantly increased in *TTP*-low LUAD. There is a possibility that CREM ZIP may impair the transcriptional activity of *CREB* family members in some *TTP*-low cancers and lead to the repression of *CREB* target genes in these tumors. The newly discovered connection between *CREB* family members and *TTP* expression suggests that therapeutic *CREB* agonists such as colforsin, salbutamol, clenbuterol, and isoprenaline may be a potential treatments for *TTP*-low expressing LUAD tumors, particularly ones with decreased *CREB* activity. [65]

LUSC—The identification of therapeutic markers specifically for LUSC utilizing TCGA data has been difficult, but efforts to discover markers and targets for LUSC remain strong. One study investigated the *NRF2* pathway in LUSC and found a 28 gene-set called $NRF2^{ACT}$ that splits LUSC cases into two subgroups: tumors that are somatically mutated within the 28 gene set with the *NRF2* pathway activated ($NRF2^{ACT}$-high), and the wild type LUSC ($NRF2^{ACT}$-low). Retrieving clinical data from various sources, the study observed that patients with the $NRF2^{ACT}$-high subtype did not benefit from adjuvant platinum-based chemotherapy while patients with the $NRF2^{ACT}$-low subtype exhibited a trend towards chemotherapy benefit. The potential clinical application of this finding is the identification

of LUSC patients who would not benefit from adjuvant chemotherapy, sparing these patients from ineffective, toxic treatment. [71] In contrast to a 2013 review that extrapolated therapeutic possibilities from the first TCGA publication for LUSC [72], our review will examine the effect of copy number variations (CNVs) on pharmacogenes in LUSC as novel therapeutic targets. [73]

Genetic Aberrations in Pharmacogenes in LUSC Tumors: The goal of personalized medicine requires not only an understanding of the molecular subtypes of different cancers, but also an extensive knowledge of how genetic variations affect drug response and toxicity for individuals. Copy number variations are DNA segments that appear to be deleted, inserted, inverted, duplicated, or recombined when compared to a reference genome. There is recent evidence demonstrating that CNVs play a role in drug efficacy and toxicity. [74, 75] There has only been limited research on pharmacogenes with CNV data, but the datasets of TCGA provided an opportunity to analyze CNVs in pharmacogenes. [76] TCGA CNV data was downloaded for four cancers, including 371 paired LUSC tissues (the other three cancers were liver hepatocellular carcinoma (LIHC), acute myeloid leukemia (LAML), and lymphoid neoplasm diffuse large B-cell lymphoma (DLBC)). The Pharmacogenomics Knowledge Base and the FDA Table of Pharmacogenomic Biomarkers in Drug Labels were used to identify important pharmacogenomics-related genes. A total of 152 genes were studied for CNVs. The CNVs of pharmacogenes with a frequency of more than 1% were selected and differences in CNV frequencies were analyzed using the Mann–Whitney U-test.

The LUSC samples experienced 136 gene losses, 2 gene deletions, 136 gene additions, and 22 gene amplifications— all with a frequency of more than 1%. With these values, LUSC demonstrated the highest number of gene losses, additions, and amplifications among the four cancers studied. Amplification scores for CNVs were calculated with the normalized log 2 test; values of ‘1’ were interpreted as gene gains, ‘-1’ as gene losses, ‘2’ as gene amplifications, and ‘-2’ as gene deletions. The *P2RY12* and *P2RY1* genes had the highest frequency for gene gain in LUSC (26.95 and 26.68%, respectively). The *FHIT* gene ranked the highest in frequency for gene loss (27.49%) and *CDKN2A* ranked highest in frequency for gene deletion in LUSC (12.13%). Drug-metabolizing genes such as *CYP2A6*, *GSTT1*, and *SULT1A1* were found to have CNVs in all four of the cancers. CNVs in drug transporters such as *SLC19A1* and *SLC28A1* were also found in all four of the cancers. The CNVs for these pharmacogenes may potentially affect patients’ drug sensitivity, resistance and toxicity as well as act as a barrier to effective treatment for LUSC. Further research is needed to understand the effects of CNVs on pharmacogenes in LUSC and other cancers in order to advance individualized pharmacotherapy. [74]

Novel Therapeutic Targets in both LUAD and LUSC—The development of improved lung cancer treatment is the result of a combination of research that explores genetic biomarkers for drug sensitivity and specific therapeutic targets. One study pertaining to drug sensitivity analyzed *FGFR1* mRNA and protein expression for *FGFR* TKI responses in lung cancers. The study found that *FGFR1* mRNA and protein expression predict ponatinib sensitivity more accurately than *FGFR1* gene copy numbers and suggested that

perhaps as high as 20% of lung cancers may be sensitive to ponatinib, roughly equaling the combined ratio of lung cancers associated with *EGFR* and *ALK* aberrations. The findings of this study may have large implications for lung cancer treatment, but more research is required. [77] This review will also examine the potential of epigenetic therapy in NSCLCs. [78]

Potential Combination of Epigenetic and Immune Checkpoint Therapy: Azacytidine (AZA- Vidaza) is a DNA demethlyating agent that stops the activity of all active DNA methyltransferases (DNMTs) and triggers degradation of DNMTs in the nucleus. [79, 80] AZA has been noted to induce increased promoter DNA hypermethylated CT antigen expression and has been observed to up-regulate important transcription factors for immune responses such as IRF7 and IRF5 [81–86] AZA is currently being tested as a “primer” that will increase tumor sensitivity to immunotherapy by affecting the PD-1/PD-L1 pathway in NSCLCs. The specific effects of AZA on NSCLC gene expression have been tested in one particular study. [78]

NSCLC cells were treated *in vitro* with 500 nM AZA for 72 hours, harvested, and analyzed one week later for genome wide methylation and expression studies. GSEA revealed that AZA induced up-regulation of multiple pathways associated with immune checkpoints and immune evasion. Another important result was that AZA up-regulated transcripts of HLA Class I antigens, supporting previous studies on AZA. [84, 85] HLA Class I antigens are crucial for cancer immunotherapy because they assist in the recognition of tumor cells so that cytotoxic T-cells can begin degrading tumors. [87] The gene *IFI27*, which encodes a protein that triggers apoptosis in late stages of chronic viral infection, was also upregulated by AZA, indicating that AZA activates genes that can inhibit tumor growth. [79] AZA also up-regulated transcription factor IRF7, an upstream activator for recognizing the virus response element (VRE-A) to increase transcription of genes involved in type 1 IFN signaling. [80] Finally, AZA was also noted to have increased *PD-L1* expression. TCGA RNA-seq data and DNA methylation data for 129 LUAD and 224 LUSC cell samples were analyzed and revealed that a significant proportion of the NSCLC samples had low expression of AZA-induced immune genes. The potential implication is that AZA may have a broad effect in treating NSCLCs by sensitizing patients with NSCLC to PD-1/PD-L1 immune checkpoint blockade. The combination of epigenetic therapy and immune checkpoint therapy is gaining traction for the treatment of NSCLCs because this therapy may be an effective treatment that at the same time reduces the toxicity present in current treatments for NSCLCs. [79]

Novel Tools

While many mutation calling algorithms have focused on comparing the DNA from normal tissue to tumor tissue of the same individual, the introduction of vast amounts of RNA-seq data from TCGA has created a new avenue to detect mutations and to characterize the cancer genome. [88] Table 6 provides examples of novel tools developed using TCGA data.

LUAD

Mutation-Calling Algorithm RADIA Employing the Triple BAM Method: A novel computational method called RADIA (RNA and DNA Integrated Analysis) combines patient matched normal and tumor DNA along with the tumor RNA to detect somatic mutations. [89] The use of these three datasets is known as the triple BAM method (TBM) and has an advantage over the traditional DNA Only Method (DOM) because TBM can detect mutations that have weak DNA support but strong RNA support. RADIA typically employs the DOM on the pairs of tumor and matched-normal DNA and then the TBM for the DNA and RNA triplets. After the results of the DOM and TBM are each filtered, they are merged and run through a final read support filter. The filters used in RADIA assist in removing false positives, accounting for misaligned reads around INDELS, removing known pseudogenes, and accounting for redundant and highly repetitive regions. A comparison of RADIA with traditional modeling algorithms was conducted to demonstrate the potential benefits of RADIA utilization.

When RADIA was applied to 230 TCGA LUAD triplets, the algorithm detected 2,404 somatic mutations. Verification of the mutations was performed by the Broad Institute. Out of the 2,404 RADIA detected somatic mutations, 2,395 were verified (99.63%). The DOM portion of RADIA detected 2,345 somatic mutations with 2,336 of them verified (99.62%). For the TBM portion, 469 (100%) of the 469 RADIA detected mutations were verified. RADIA detected two non-synonymous *TP53* mutations that were below the detection threshold of other mutation calling algorithms used by TCGA. One of these mutations (G266E) was confirmed by another lung cancer study [90] and appears to play a role in other cancers such as prostate [91], pancreas [92], urinary tract [93], and hematopoietic and lymphoid cancer [94]. The other *TP53* mutation (G199V) is a known anti-apoptotic gain-of-function mutation through the *STAT3* pathway [95]. The study also found a nonsense mutation (W239*) that was below the detection threshold of other mutation calling algorithms in the *STK11* gene, which is a tumor suppressor gene. This mutation has been observed to produce a truncated protein by introducing an early stop codon in exon five. While only expressed alleles can be detected with this detection algorithm, RADIA presents new opportunities to outline even more cancer subtypes and to discover potential therapeutic targets. [89]

LUSC—The discovery of an effective therapeutic target for LUSC has remained elusive with no equivalent to the discovery of TKI-sensitivity for *EGFR* and *ALK* mutations for LUAD [96–101]. This may explain the greater emphasis on constructing improved mutation calling algorithms and investigating clinical applications for LUSC therapies with TCGA data. Two new approaches for finding potential therapeutic markers associated with TCGA LUSC datasets have been dGene, a new annotation tool designed to quickly identify gene mutations that belong to one of 10 druggable classes frequently targeted in cancer drug development [102], and a proposed method of integrated analysis of deep sequencing for both the genome and transcriptome to generate greater insight into LUSC [103]. The Lung Master Protocol (Lung-MAP, S1400) combines the process of discovering targetable genes and clinical trials for LUSC by utilizing next-generation DNA sequencing, followed by randomization to the pertinent targeted therapy versus standard care. [104] This review will

highlight the development of the mutation calling algorithm SigFuge that utilizes RNA-seq data. [105]

New Tool SigFuge for Cluster Analysis Using RNA-seq Data: The widespread addition of transcriptome sequencing (RNA-seq) has presented the opportunity to discover novel transcriptional events such as splicing patterns [106], alternative untranslated region usage [107], and gene fusions [108]. The importance of transcriptional events such as alternative splicing has increased in recent years as these events appear to contribute to tumorigenesis [109,110]. With RNA-seq, clustering methods provide the ability to detect alternate isoform usage of specific genes to improve our molecular understanding of cancer. SigFuge (SIGnificant Forms Using per-base Gene Expression) seeks to identify clusters within the transcriptome that express isoforms from a single gene locus at different proportions. The main advantage of SigFuge is its approach to clusters by per-base expression rather than by analysis of aggregations along whole gene boundaries, leading to greater sensitivity for discovery of important alternate splicing events. The three main steps of SigFuge are data extraction, processing, and analysis. Data extraction includes the organization of sample data into expression count matrices and exon annotation. Data processing includes the removal of low-expression samples, count normalization of sample curves, and log transformation of the data.

The new algorithm was employed on 177 TCGA LUSC RNA-seq samples to detect alternative splicing events for the cancer type. From the analysis, 27 gene loci were identified based their association with a gain or loss of a middle exon. *CDKN2A* and *KLK12* were two notable genes from the 27 identified genes whose differential isoform usages may contribute to tumor development and growth. As explicated earlier, *CDKN2A* is a known tumor suppressor that includes the two genes $p16^{INK4A}$ and $p14^{ARF}$. SigFuge identified three clusters: lack of expression for both $p16^{INK4A}$ and $p14^{ARF}$, expression of both genes, and expression of $p14^{ARF}$ only. These SigFuge clusters correlated strongly with the three main modes of *CDKN2A* inactivation, which are homozygous deletion, epigenetic silencing by methylation and inactivation by point mutation. Pearson's chi-square tests revealed highly significant p-values that confirmed the strong associations between the SigFuge clusters and previously identified alterations. Additionally, the *KLK12* locus is known to produce multiple isoforms and has been studied in epithelial ovarian [111], prostate [112], and lung cancers [114]. The SigFuge analysis identified *KLK12* expression clusters of similar differential usage among TCGA LUSC samples. The results suggest that *KLK12* and other *KLK* splice variants may be used as potential biomarkers for LUSC. The accessibility of RNA-seq data has allowed for increased research into novel transcriptional events, and new algorithms such as SigFuge are being developed to take advantage of this new source of data. [106]

Conclusions and Future Directions

The project of TCGA LUAD and LUSC provided a strong foundation for genomic research and has stimulated a diversity of lung cancer research. The consequences of this research has led to an improved understanding of the molecular mechanisms underlying LUAD and LUSC, the elucidation of the clinical implications of these molecular mechanisms, the

development of new and improved tools and algorithms, and the discovery of potential new therapies. This review of the literature pertaining to LUAD and LUSC associated with TCGA has highlighted the current and potential future research areas in LUAD and LUSC. The importance of commonly altered genes such as *CDKN2AB, TP53*, and *TP63* for tumorigenesis cannot be underestimated and will likely continue to generate an abundance of cancer research. Epigenetic aspects such as methylation and CIMP will likely garner increasing scrutiny to explain cancer cell behavior and be leveraged to produce therapeutic treatments. The popularization and accessibility of RNA-seq data has arrived, but the utilization of transcriptome data is still a new aspect of cancer research in which there is still much to be investigated. The functions of miRNA have sparked intense research efforts, and the pace of miRNA research will likely only accelerate. Burgeoning fields such as the study of proteomics, splice factors, and pharmacogenes are now gaining more traction with the large amount of TCGA data providing opportunities for these research topics to be explored more in depth. The LUAD and LUSC project of TCGA has furthered our understanding of the two cancers and will continue to do so with the ultimate goal of developing clinical measures that will improve the outcomes of patients with LUAD and LUSC.

Acknowledgments

RSH received support from National Institute of Health grant K08GM089941, National Institute of Health grant R21 CA139278, University of Chicago Support Grant (#P30 CA14599), Breast Cancer SPORE Career Development Award [CA125183], the National Center for Advancing Translational Sciences of the National Institute of Health [UL1RR024999], the National Institute of Health grant UO1GM61393, Circle of Service Foundation Early Career Investigator award, and a Conquer Cancer Foundation of ASCO Translational Research Professorship award In Memory of Merrill J. Egorin, MD (awarded to Dr. MJ Ratain). YML received support from The William F. O'Connor Foundation (MJR). The funding sources had no involvement in the research. All authors have read the authorship agreement and the manuscript has been reviewed and approved by all named authors.

Abbreviations

TCGA	The Cancer Genome Atlas
NSCLC	non-small cell lung carcinoma
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
OS	overall survival
CIMP	CpG island methylator phenotype
GSEA	Gene Set Enrichment Analysis
miRNA	MicroRNA

References

1. Mission and Goal - TCGA. [Accessed June 13, 2015] http://cancergenome.nih.gov/abouttcga/overview/missiongoal
2. Backgrounder - TCGA. [Accessed June 13, 2015] http://cancergenome.nih.gov/newsevents/forthemedia/backgrounder
3. Cancers Selected for Study - TCGA. [Accessed June 13, 2015] http://cancergenome.nih.gov/cancersselected

4. The Cancer Genome Atlas - Data Portal. [Accessed June 13, 2015] https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp
5. The Cancer Genome Atlas (TCGA): The next stage - TCGA. [Accessed June 13, 2015] http://cancergenome.nih.gov/newsevents/newsannouncements/TCGA_The_Next_Stage
6. WHO | Cancer. [Accessed June 13, 2015] http://www.who.int/mediacentre/factsheets/fs297/en/
7. What is non-small cell lung cancer? [Accessed June 13, 2015] http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-what-is-non-small-cell-lung-cancer
8. Stewart DJ. Tumor and host factors that may limit efficacy of chemotherapy in non-small cell and small cell lung cancer. Crit Rev Oncol Hematol. 2010; 75(3):173–234.10.1016/j.critrevonc.2009.11.006 [PubMed: 20047843]
9. Stewart DJ, Chiritescu G, Dahrouge S, Banerjee S, Tomiak EM. Chemotherapy dose--response relationships in non-small cell lung cancer and implied resistance mechanisms. Cancer Treat Rev. 2007; 33(2):101–137.10.1016/j.ctrv.2006.12.002 [PubMed: 17276603]
10. Stewart DJ, Tomiak E, Shamji FM, Maziak DE, MacLeod P. Phase II study of alternating chemotherapy regimens for advanced non-small cell lung cancer. Lung Cancer. 2004; 44(2):241–249.10.1016/j.lungcan.2003.10.009 [PubMed: 15084389]
11. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304(5676):1497–1500.10.1126/science.1099314 [PubMed: 15118125]
12. Antonicelli A, Cafarotti S, Indini A, et al. EGFR-targeted therapy for non-small cell lung cancer: focus on EGFR oncogenic mutation. Int J Med Sci. 2013; 10(3):320–330.10.7150/ijms.4609 [PubMed: 23423768]
13. Solomon BJ, Mok T, Kim D-W, et al. First-Line Crizotinib versus Chemotherapy in ALK -Positive Lung Cancer. N Engl J Med. 2014; 371(23):2167–2177.10.1056/NEJMoa1408440 [PubMed: 25470694]
14. Yamamoto H, Toyooka S, Mitsudomi T. Impact of EGFR mutation analysis in non-small cell lung cancer. Lung Cancer. 2009; 63(3):315–321.10.1016/j.lungcan.2008.06.021 [PubMed: 18760859]
15. Suda K, Onozato R, Yatabe Y, Mitsudomi T. EGFR T790M mutation: a double role in lung cancer cell survival? J Thorac Oncol. 2009; 4(1):1–4.10.1097/JTO.0b013e3181913c9f [PubMed: 19096299]
16. TumorPortal. [Accessed June 13, 2015] http://www.tumorportal.org/figure/S1
17. Rekhtman N, Paik PK, Arcila ME, et al. Clarifying the spectrum of driver oncogene mutations in biomarker-verified squamous carcinoma of lung: lack of EGFR/KRAS and presence of PIK3CA/AKT1 mutations. Clin Cancer Res. 2012; 18(4):1167–1176.10.1158/1078-0432.CCR-11-2109 [PubMed: 22228640]
18. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012; 489(7417):519–525.10.1038/nature11404 [PubMed: 22960745]
19. Collisson EA, Campbell JD, Brooks AN, et al. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511(7511):543–550.10.1038/nature13385 [PubMed: 25079552]
20. Schuster K, Venkateswaran N, Rabellino A, Girard L, Peña-Llopis S, Scaglioni PP. Nullifying the CDKN2AB locus promotes mutant K-ras lung tumorigenesis. Mol Cancer Res. 2014; 12(6):912–923.10.1158/1541-7786.MCR-13-0620-T [PubMed: 24618618]
21. Shinjo K, Okamoto Y, An B, et al. Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characters of lung adenocarcinoma. Carcinogenesis. 2012; 33(7):1277–1285.10.1093/carcin/bgs154 [PubMed: 22532250]
22. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. Nat Rev Cancer. 2011; 11(11):761–774.10.1038/nrc3106 [PubMed: 21993244]
23. Fisher GH, Wellen SL, Klimstra D, et al. Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. Genes Dev. 2001; 15(24):3249–3262.10.1101/gad.947701 [PubMed: 11751631]

24. Johnson L, Mercer K, Greenbaum D, et al. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature. 2001; 410(6832):1111–1116.10.1038/35074129 [PubMed: 11323676]
25. Pao W, Wang TY, Riely GJ, et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. PLoS Med. 2005; 2(1):e17.10.1371/journal.pmed.0020017 [PubMed: 15696205]
26. Konstantinidou G, Ramadori G, Torti F, et al. RHOA-FAK is a required signaling axis for the maintenance of KRAS-driven lung adenocarcinomas. Cancer Discov. 2013; 3(4):444–457.10.1158/2159-8290.CD-12-0388 [PubMed: 23358651]
27. Zhang Y, Xiong Y, Yarbrough WG. ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways. Cell. 1998; 92(6):725–734. http://www.ncbi.nlm.nih.gov/pubmed/9529249. [PubMed: 9529249]
28. Kamijo T, Weber JD, Zambetti G, Zindy F, Roussel MF, Sherr CJ. Functional and physical interactions of the ARF tumor suppressor with p53 and Mdm2. Proc Natl Acad Sci. 1998; 95(14):8292–8297.10.1073/pnas.95.14.8292 [PubMed: 9653180]
29. Cai B, Jiang X. Revealing Biological Pathways Implicated in Lung Cancer from TCGA Gene Expression Data Using Gene Set Enrichment Analysis. Cancer Inform. 2014; 13(Suppl 1):113–121.10.4137/CIN.S13882 [PubMed: 25520551]
30. Li X, Chen S, Sun T, et al. The transcriptional regulation of SOX2 on FOXA1 gene and its application in diagnosis of human breast and lung cancers. Clin Lab. 2014; 60(6):909–918. http://www.ncbi.nlm.nih.gov/pubmed/25016694. [PubMed: 25016694]
31. 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(2):139–151.10.1016/j.ccr.2014.01.008 [PubMed: 24525231]
32. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001; 414(6859):105–111.10.1038/35102167 [PubMed: 11689955]
33. Chen Y-C, Hsu H-S, Chen Y-W, et al. Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells. PLoS One. 2008; 3(7):e2637.10.1371/journal.pone.0002637 [PubMed: 18612434]
34. Eramo A, Lotti F, Sette G, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death Differ. 2008; 15(3):504–514.10.1038/sj.cdd.4402283 [PubMed: 18049477]
35. Justilien V, Regala RP, Tseng I-C, et al. Matrix metalloproteinase-10 is required for lung cancer stem cell maintenance, tumor initiation and metastatic potential. PLoS One. 2012; 7(4):e35040.10.1371/journal.pone.0035040 [PubMed: 22545096]
36. Yuan P, Kadara H, Behrens C, et al. Sex determining region Y-Box 2 (SOX2) is a potential cell-lineage gene highly expressed in the pathogenesis of squamous cell carcinomas of the lung. PLoS One. 2010; 5(2):e9112.10.1371/journal.pone.0009112 [PubMed: 20161759]
37. Regala RP, Thompson EA, Fields AP. Atypical protein kinase C iota expression and aurothiomalate sensitivity in human lung cancer cells. Cancer Res. 2008; 68(14):5888–5895.10.1158/0008-5472.CAN-08-0438 [PubMed: 18632643]
38. Khoury JD, Tannir NM, Williams MD, et al. Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. J Virol. 2013; 87(16):8916–8926.10.1128/JVI.00340-13 [PubMed: 23740984]
39. Shinmura K, Igarashi H, Kato H, et al. CLCA2 as a novel immunohistochemical marker for differential diagnosis of squamous cell carcinoma from adenocarcinoma of the lung. Dis Markers. 2014; 2014:619273. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4274868&tool=pmcentrez&rendertype=abstract. [PubMed: 25548429]
40. Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. Nat Rev Cancer. 2007; 7(10):791–799.10.1038/nrc2212 [PubMed: 17851544]
41. Banerji CRS, Severini S, Caldas C, Teschendorff AE. Intra-tumour signalling entropy determines clinical outcome in breast and lung cancer. PLoS Comput Biol. 2015; 11(3):e1004115.10.1371/journal.pcbi.1004115 [PubMed: 25793737]

42. Li X, Shi Y, Yin Z, Xue X, Zhou B. An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. J Transl Med. 2014; 12:159.10.1186/1479-5876-12-159 [PubMed: 24893932]
43. Meng W, Ye Z, Cui R, et al. MicroRNA-31 predicts the presence of lymph node metastases and survival in patients with lung adenocarcinoma. Clin Cancer Res. 2013; 19(19):5423–5433.10.1158/1078-0432.CCR-13-0320 [PubMed: 23946296]
44. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136(4):642–655.10.1016/j.cell.2009.01.035 [PubMed: 19239886]
45. Kent OA, Mendell JT. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. Oncogene. 2006; 25(46):6188–6196.10.1038/sj.onc.1209913 [PubMed: 17028598]
46. Cekaite L, Rantala JK, Bruun J, et al. MiR-9, -31, and -182 deregulation promote proliferation and tumor cell survival in colon cancer. Neoplasia. 2012; 14(9):868–879. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3459282&tool=pmcentrez&rendertype=abstract. [PubMed: 23019418]
47. Valastyan S, Reinhardt F, Benaich N, et al. A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. Cell. 2009; 137(6):1032–1046.10.1016/j.cell.2009.03.047 [PubMed: 19524507]
48. Shin S, Dimitri CA, Yoon S-O, Dowdle W, Blenis J. ERK2 but not ERK1 induces epithelial-to-mesenchymal transformation via DEF motif-dependent signaling events. Mol Cell. 2010; 38(1): 114–127.10.1016/j.molcel.2010.02.020 [PubMed: 20385094]
49. Jiang L, Zhu W, Streicher K, et al. Increased IR-A/IR-B ratio in non-small cell lung cancers associates with lower epithelial-mesenchymal transition signature and longer survival in squamous cell lung carcinoma. BMC Cancer. 2014; 14:131.10.1186/1471-2407-14-131 [PubMed: 24571613]
50. Huang T, Yang J, Cai Y-D. Novel candidate key drivers in the integrative network of genes, microRNAs, methylations, and copy number variations in squamous cell lung carcinoma. Biomed Res Int. 2015; 2015:358125.10.1155/2015/358125 [PubMed: 25802847]
51. Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. Brief Bioinform. 2015; 16(2):291–303.10.1093/bib/bbu003 [PubMed: 24632304]
52. Schindler CW. Series introduction. JAK-STAT signaling in human disease. J Clin Invest. 2002; 109(9):1133–1137.10.1172/JCI15644 [PubMed: 11994400]
53. Miklossy G, Hilliard TS, Turkson J. Therapeutic modulators of STAT signalling for human diseases. Nat Rev Drug Discov. 2013; 12(8):611–629.10.1038/nrd4088 [PubMed: 23903221]
54. Abbas R, McColl KS, Kresak A, et al. PIAS3 expression in squamous cell lung cancer is low and predicts overall survival. Cancer Med. 2015; 4(3):325–332.10.1002/cam4.372 [PubMed: 25573684]
55. Chung CD, Liao J, Liu B, et al. Specific inhibition of Stat3 signal transduction by PIAS3. Science. 1997; 278(5344):1803–1805. http://www.ncbi.nlm.nih.gov/pubmed/9388184. [PubMed: 9388184]
56. Sharrocks AD. PIAS proteins and transcriptional regulation--more than just SUMO E3 ligases? Genes Dev. 2006; 20(7):754–758.10.1101/gad.1421006 [PubMed: 16600908]
57. Brantley EC, Nabors LB, Gillespie GY, et al. Loss of protein inhibitors of activated STAT-3 expression in glioblastoma multiforme tumors: implications for STAT-3 activation and gene expression. Clin Cancer Res. 2008; 14(15):4694–4704.10.1158/1078-0432.CCR-08-0618 [PubMed: 18676737]
58. Dhanasekaran SM, Alejandro Balbin O, Chen G, et al. Transcriptome meta-analysis of lung cancer reveals recurrent aberrations in NRG1 and Hippo pathway genes. Nat Commun. 2014; 5:5893.10.1038/ncomms6893 [PubMed: 25531467]
59. Deng B, Sun Z, Jason W, Yang P. Increased BCAR1 predicts poor outcomes of non-small cell lung cancer in multiple-center patients. Ann Surg Oncol. 2013; 20 (Suppl 3):S701–S708.10.1245/s10434-013-3184-2 [PubMed: 23904007]
60. Cabodi S, Tinnirello A, Bisaro B, et al. p130Cas is an essential transducer element in ErbB2 transformation. FASEB J. 2010; 24(10):3796–3808.10.1096/fj.10-157347 [PubMed: 20505116]

61. Cabodi S, Tinnirello A, Di Stefano P, et al. p130Cas as a new regulator of mammary epithelial cell proliferation, survival, and HER2-neu oncogene-dependent breast tumorigenesis. Cancer Res. 2006; 66(9):4672–4680.10.1158/0008-5472.CAN-05-2909 [PubMed: 16651418]
62. Huang W, Deng B, Wang R-W, Tan Q-Y, Jiang Y-G. Expression of breast cancer anti-estrogen resistance 1 in relation to vascular endothelial growth factor, p53, and prognosis in esophageal squamous cell cancer. Dis Esophagus. 2013; 26(5):528–537.10.1111/j.1442-2050.2012.01376.x [PubMed: 22816673]
63. Huang W, Deng B, Wang R-W, et al. BCAR1 protein plays important roles in carcinogenesis and predicts poor prognosis in non-small-cell lung cancer. PLoS One. 2012; 7(4):e36124.10.1371/journal.pone.0036124 [PubMed: 22558353]
64. Pecot CV, Rupaimoole R, Yang D, et al. Tumour angiogenesis regulation by the miR-200 family. Nat Commun. 2013; 4:2427.10.1038/ncomms3427 [PubMed: 24018975]
65. Fallahi M, Amelio AL, Cleveland JL, Rounbehler RJ. CREB targets define the gene expression signature of malignancies having reduced levels of the tumor suppressor tristetraprolin. PLoS One. 2014; 9(12):e115517.10.1371/journal.pone.0115517 [PubMed: 25541715]
66. Brooks SA, Blackshear PJ. Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. Biochim Biophys Acta. 1829(6–7):666–679.10.1016/j.bbagrm.2013.02.003 [PubMed: 23428348]
67. Rounbehler RJ, Fallahi M, Yang C, et al. Tristetraprolin impairs myc-induced lymphoma and abolishes the malignant state. Cell. 2012; 150(3):563–574.10.1016/j.cell.2012.06.033 [PubMed: 22863009]
68. Brennan SE, Kuwano Y, Alkharouf N, Blackshear PJ, Gorospe M, Wilson GM. The mRNA-destabilizing protein tristetraprolin is suppressed in many cancers, altering tumorigenic phenotypes and patient prognosis. Cancer Res. 2009; 69(12):5168–5176.10.1158/0008-5472.CAN-08-4238 [PubMed: 19491267]
69. Hayes DN, Monti S, Parmigiani G, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol. 2006; 24(31):5079–5090.10.1200/JCO.2005.05.1748 [PubMed: 17075127]
70. Wilkerson MD, Yin X, Walter V, et al. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. PLoS One. 2012; 7(5):e36530.10.1371/journal.pone.0036530 [PubMed: 22590557]
71. Cescon DW, She D, Sakashita S, et al. NRF2 Pathway Activation and Adjuvant Chemotherapy Benefit in Lung Squamous Cell Carcinoma. Clin Cancer Res. 2015; 21(11):2499–2505.10.1158/1078-0432.CCR-14-2206 [PubMed: 25739673]
72. Devarakonda S, Morgensztern D, Govindan R. Clinical applications of The Cancer Genome Atlas project (TCGA) for squamous cell lung carcinoma. Oncology (Williston Park). 2013; 27(9):899–906. http://www.ncbi.nlm.nih.gov/pubmed/24282988. [PubMed: 24282988]
73. Kim I-W, Han N, Kim MG, Kim T, Oh JM. Copy number variability analysis of pharmacogenes in patients with lymphoma, leukemia, hepatocellular, and lung carcinoma using The Cancer Genome Atlas data. Pharmacogenet Genomics. 2015; 25(1):1–7.10.1097/FPC.0000000000000097 [PubMed: 25379720]
74. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature. 2006; 444(7118):444–454.10.1038/nature05329 [PubMed: 17122850]
75. Johansson I, Ingelman-Sundberg M. CNVs of human genes and their implication in pharmacogenetics. Cytogenet Genome Res. 2008; 123(1–4):195–204.10.1159/000184709 [PubMed: 19287156]
76. He Y, Hoskins JM, McLeod HL. Copy number variants in pharmacogenetic genes. Trends Mol Med. 2011; 17(5):244–251.10.1016/j.molmed.2011.01.007 [PubMed: 21388883]
77. Wynes MW, Hinz TK, Gao D, et al. FGFR1 mRNA and protein expression, not gene copy number, predict FGFR TKI sensitivity across all lung cancer histologies. Clin Cancer Res. 2014; 20(12):3299–3309.10.1158/1078-0432.CCR-13-3060 [PubMed: 24771645]
78. Wrangle J, Wang W, Koch A, et al. Alterations of immune response of Non-Small Cell Lung Cancer with Azacytidine. Oncotarget. 2013; 4(11):2067–2079. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3875770&tool=pmcentrez&rendertype=abstract. [PubMed: 24162015]

79. Gabbara S, Bhagwat AS. The mechanism of inhibition of DNA (cytosine-5-)-methyltransferases by 5-azacytosine is likely to involve methyl transfer to the inhibitor. Biochem J. 1995; 307(Pt 1): 87–92. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1136748&tool=pmcentrez&rendertype=abstract. [PubMed: 7536414]
80. Stresemann C, Lyko F. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer. 2008; 123(1):8–13.10.1002/ijc.23607 [PubMed: 18425818]
81. Li, Q.; Tainsky, MA. Epigenetic Silencing of IRF7 and/or IRF5 in Lung Cancer Cells Leads to Increased Sensitivity to Oncolytic Viruses. In: Katoh, M., editor. PLoS One. Vol. 6. 2011. p. e28683
82. Kulaeva OI, Draghici S, Tang L, Kraniak JM, Land SJ, Tainsky MA. Epigenetic silencing of multiple interferon pathway genes after cellular immortalization. Oncogene. 2003; 22(26):4118–4127.10.1038/sj.onc.1206594 [PubMed: 12821946]
83. Símová J, Polláková V, Indrová M, et al. Immunotherapy augments the effect of 5-azacytidine on HPV16-associated tumours with different MHC class I-expression status. Br J Cancer. 2011; 105(10):1533–1541.10.1038/bjc.2011.428 [PubMed: 22015556]
84. Fonsatti E, Nicolay HJM, Sigalotti L, et al. Functional up-regulation of human leukocyte antigen class I antigens expression by 5-aza-2'-deoxycytidine in cutaneous melanoma: immunotherapeutic implications. Clin Cancer Res. 2007; 13(11):3333–3338.10.1158/1078-0432.CCR-06-3091 [PubMed: 17545540]
85. Claus R, Almstedt M, Lübbert M. Epigenetic treatment of hematopoietic malignancies: in vivo targets of demethylating agents. Semin Oncol. 2005; 32(5):511–520.10.1053/j.seminoncol.2005.07.024 [PubMed: 16210092]
86. Karpf AR, Peterson PW, Rawlins JT, et al. Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. Proc Natl Acad Sci U S A. 1999; 96(24):14007–14012. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24181&tool=pmcentrez&rendertype=abstract. [PubMed: 10570189]
87. Raghavan M, Del Cid N, Rizvi SM, Peters LR. MHC class I assembly: out and about. Trends Immunol. 2008; 29(9):436–443.10.1016/j.it.2008.06.004 [PubMed: 18675588]
88. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63.10.1038/nrg2484 [PubMed: 19015660]
89. Radenbaugh AJ, Ma S, Ewing A, et al. RADIA: RNA and DNA integrated analysis for somatic mutation detection. PLoS One. 2014; 9(11):e111516.10.1371/journal.pone.0111516 [PubMed: 25405470]
90. Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010; 466(7308):869–873.10.1038/nature09208 [PubMed: 20668451]
91. Lindberg J, Mills IG, Klevebring D, et al. The mitochondrial and autosomal mutation landscapes of prostate cancer. Eur Urol. 2013; 63(4):702–708.10.1016/j.eururo.2012.11.053 [PubMed: 23265383]
92. Biankin AV, Waddell N, Kassahn KS, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012; 491(7424):399–405.10.1038/nature11547 [PubMed: 23103869]
93. Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet. 2011; 43(9):875–878.10.1038/ng.907 [PubMed: 21822268]
94. Abaan OD, Polley EC, Davis SR, et al. The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. Cancer Res. 2013; 73(14):4372–4382.10.1158/0008-5472.CAN-12-3342 [PubMed: 23856246]
95. Kim T-H, Lee SY, Rho JH, et al. Mutant p53 (G199V) gains antiapoptotic function through signal transducer and activator of transcription 3 in anaplastic thyroid cancer cells. Mol Cancer Res. 2009; 7(10):1645–1654.10.1158/1541-7786.MCR-09-0117 [PubMed: 19825993]
96. Kwak EL, Bang Y-J, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med. 2010; 363(18):1693–1703.10.1056/NEJMoa1006448 [PubMed: 20979469]

97. Zhou C, Wu Y-L, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol. 2011; 12(8):735–742.10.1016/S1470-2045(11)70184-X [PubMed: 21783417]

98. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012; 13(3):239–246.10.1016/S1470-2045(11)70393-X [PubMed: 22285168]

99. Mitsudomi T, Morita S, Yatabe Y, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol. 2010; 11(2):121–128.10.1016/S1470-2045(09)70364-X [PubMed: 20022809]

100. Mok TS, Wu Y-L, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009; 361(10):947–957.10.1056/NEJMoa0810699 [PubMed: 19692680]

101. Shaw AT, Camidge DR, Engelman JA, et al. Clinical activity of crizotinib in advanced non-small cell lung cancer (NSCLC) harboring ROS1 gene rearrangement. ASCO Meet Abstr. 2012; 30(15_suppl):7508. http://hwmaint.meeting.ascopubs.org/cgi/content/abstract/30/15_suppl/7508.

102. Kumar RD, Chang L-W, Ellis MJ, Bose R. Prioritizing Potentially Druggable Mutations with dGene: An Annotation Tool for Cancer Genome Sequencing Data. PLoS One. 2013; 8(6):e67980.10.1371/journal.pone.0067980 [PubMed: 23826350]

103. Stead LF, Egan P, Devery A, et al. An integrated inspection of the somatic mutations in a lung squamous cell carcinoma using next-generation sequencing. PLoS One. 2013; 8(11):e78823.10.1371/journal.pone.0078823 [PubMed: 24244370]

104. Herbst RS, Gandara DR, Hirsch FR, et al. Lung Master Protocol (Lung-MAP)-A Biomarker-Driven Protocol for Accelerating Development of Therapies for Squamous Cell Lung Cancer: SWOG S1400. Clin Cancer Res. 2015; 21(7):1514–1524.10.1158/1078-0432.CCR-13-3473 [PubMed: 25680375]

105. Kimes PK, Cabanski CR, Wilkerson MD, et al. SigFuge: single gene clustering of RNA-seq reveals differential isoform usage among cancer samples. Nucleic Acids Res. 2014; 42(14):e113.10.1093/nar/gku521 [PubMed: 25030904]

106. Sultan M, Schulz MH, Richard H, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 2008; 321(5891):956–960.10.1126/science.1160342 [PubMed: 18599741]

107. Ramsköld D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol. 2009; 5(12):e1000598.10.1371/journal.pcbi.1000598 [PubMed: 20011106]

108. Maher CA, Kumar-Sinha C, Cao X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009; 458(7234):97–101.10.1038/nature07638 [PubMed: 19136943]

109. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. Genes Dev. 2003; 17(4):419–437.10.1101/gad.1048803 [PubMed: 12600935]

110. Venables JP. Aberrant and alternative splicing in cancer. Cancer Res. 2004; 64(21):7647–7654.10.1158/0008-5472.CAN-04-1910 [PubMed: 15520162]

111. Dong Y, Kaushal A, Brattsand M, Nicklin J, Clements JA. Differential splicing of KLK5 and KLK7 in epithelial ovarian cancer produces novel variants with potential as cancer biomarkers. Clin Cancer Res. 2003; 9(5):1710–1720. http://www.ncbi.nlm.nih.gov/pubmed/12738725. [PubMed: 12738725]

112. Dong Y, Bui LT, Odorico DM, et al. Compartmentalized expression of kallikrein 4 (KLK4/hK4) isoforms in prostate cancer: nuclear, cytoplasmic and secreted forms. Endocr Relat Cancer. 2005; 12(4):875–889.10.1677/erc.1.01062 [PubMed: 16322328]

113. Planque C, Choi Y-H, Guyetant S, Heuzé-Vourc'h N, Briollais L, Courty Y. Alternative splicing variant of kallikrein-related peptidase 8 as an independent predictor of unfavorable prognosis in lung cancer. Clin Chem. 2010; 56(6):987–997.10.1373/clinchem.2009.138917 [PubMed: 20360129]

Table 1

Summary of TCGA cases with data as of June 29, 2015 in a descending order based on the data available.

Selected Cancer	# Cases with Data
Breast invasive carcinoma [BRCA]	1098
Ovarian serous cystadenocarcinoma [OV]	586
Uterine Corpus Endometrial Carcinoma [UCEC]	548
Kidney renal clear cell carcinoma [KIRC]	536
Glioblastoma multiforme [GBM]	528
Head and Neck squamous cell carcinoma [HNSC]	528
Lung adenocarcinoma [LUAD]	521
Brain Lower Grade Glioma [LGG]	516
Thyroid carcinoma [THCA]	507
Lung squamous cell carcinoma [LUSC]	504
Prostate adenocarcinoma [PRAD]	498
Skin Cutaneous Melanoma [SKCM]	470
Colon adenocarcinoma [COAD]	461
Stomach adenocarcinoma [STAD]	443
Bladder Urothelial Carcinoma [BLCA]	412
Liver hepatocellular carcinoma [LIHC]	377
Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]	308
Kidney renal papillary cell carcinoma [KIRP]	291
Sarcoma [SARC]	261
Acute Myeloid Leukemia [LAML]	200
Esophageal carcinoma [ESCA]	185
Pancreatic adenocarcinoma [PAAD]	185
Pheochromocytoma and Paraganglioma [PCPG]	179
Rectum adenocarcinoma [READ]	171
Testicular Germ Cell Tumors [TGCT]	150
Thymoma [THYM]	124
Mesothelioma [MESO]	87
Adrenocortical carcinoma [ACC]	80
Uveal Melanoma [UVM]	80
Kidney Chromophobe [KICH]	66
Uterine Carcinosarcoma [UCS]	57
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma [DLBC]	48
Cholangiocarcinoma [CHOL]	36

Data obtained from https://tcga-data.nci.nih.gov/tcga/ on June 29, 2015

Table 2

Comparison of TCGA Landmark Papers for LUAD and LUSC

	LUSC	**LUAD**
Publication Year	2012	2014
# of Samples Studied	178 TCGA	412 (230 TCGA)
% Cases of past/present smoking	96%	81%
# mutations per megabase	8.10	8.87
Most frequently mutated genes	*TP53, CDKN2A, PTEN, PIK3CA, KEAP1*	*TP53, KRAS, EGFR, BRAF*
Notable Amplifications	*SOX2, TP63*, chromosome 3q	*NKX2-1, TERT, MDM2, KRAS, EGFR, MET*
Notable Deletions/ Loss of Function	*FOXP1, NOTCH1, NOTCH2, ASCL4*	*CDKN2A, MET* exon 14 skipping
Possible Cancer Drivers for Oncogene-Negative Tumors	*FAM123B, HRAS, FBXW7, SMARCA4, NF1, SMAD4, EGFR*	*ERBB2, MET, NF1*

Summarized from Official TCGA Publications for LUSC [18] and LUAD [19]

Table 3

Examples of Molecular Mechanisms Leading to Disease (Diagnostic Markers)

Topic	Model Systems	Results
LUAD		
KRAS and *CDKN2AB*	Mouse model, TCGA (RNA-seq, OS) data	The loss of function of *CDKN2AB* in *KRAS*-mutated tumors can drive cell proliferation and metastatic behavior in *KRAS* lung cancers, especially for LUAD, revealing a high-risk LUAD subgroup. [20]
CpG island methylator phenotype (CIMP)	Tumor samples, TCGA (methylation, OS) data	CIMP-H(igh) patients experienced the poorest OS when a TCGA LUAD cohort was split into three groups: CIMP-H, CIMP-L(ow), and CIMP-N(egative), revealing a LUAD subgroup that may benefit from epigenetic therapy. [21]
LUSC		
Oncogenic Pathways	TCGA (RNA-seq) data	The Gene Set Enrichment Analysis (GSEA) and Signaling Pathway Impact Analysis (SPIA) on TCGA LUSC data revealed that the top three pathways for driving LUSC were the viral carcinogenesis pathway, the p53 pathway, and the PI3K-Akt signaling pathway. [29]
SOX2 and *FOXA1*	Cell line, TCGA (DNA-seq, RNA- seq) data	*SOX2* is a negative upstream regulator of *FOXA1*, providing insight into the subset of LUSCs where *SOX2* plays a cancer driving role. [30]
SOX2 and *PRKCI* Co-amplification and Coordination	Cell lines, TCGA (RNA-seq) data, tumor samples	For a subset of LUSC tumors, *SOX2* and *PRKCI* are co- amplified and overexpressed in cancer stem cells to regulate the Hedgehog pathway. These two genes and the Hh pathway are vital for the growth and maintenance of these cancer stem cells. [31]
Both		
Viral Transcripts	TCGA (RNA-seq) data	Human papillomavirus (HPV) transcripts were detected in LUSC while no DNA virus transcripts were observed in LUAD. The understanding of viral integration sites may provide insight into the development of LUSC tumors. [38]
CLCA2 Immunohistochemi cal Marker	Tumor samples, TCGA (RNA-seq) data	*CLCA2* expression may be an effective immunohistochemical marker for the differential diagnosis of LUAD and LUSC as *CLCA2* expression has a much higher frequency in LUSC than in LUAD. [39]

Table 4

Examples of Molecular Mechanisms Related to Prognosis (Disease Progression Markers)

Topic	Model Systems	Results
LUAD		
Intra-Tumor Signaling Entropy	TCGA (RNA-seq, OS) data	Higher signaling entropy in *Stage 1* LUAD patients implies worse prognosis. For the top third of the signaling entropy distribution, the hazard rate for mortality doubled in LUAD patients. [40]
8 miRNA Prognostic Signature	TCGA (microRNA-expression, OS) data	A weighted linear prognostic model for the miRNA markers (miR-31, miR-196b, miR-766, miR-519a-1, miR-375, miR-187, miR-331 and miR-101-1) was developed and was especially accurate as a five-year prognostic model for patients who were non-smoking or had not smoked in 15 years. [42]
miR-31	Tumor samples, TCGA (microRNA- expression, OS) data, cell lines	LUAD patients with lymph node metastasis had a 2.5 fold increase in miR-31 expression to LUAD patients without lymph node metastasis. The multivariate Cox proportional hazards test reveals that miR-31 expression has a prognostic value independent of lymph node status. [43]
LUSC		
Insulin Receptor Isoform A and B (*IR-A* and *IR-B*)	TCGA (RNA-seq, OS) data	Decreased *IR-B* expression results in an elevated *IR-A/ IR-B ratio*, which is associated with lower epithelial- mesenchymal transition gene signatures in NSCLCs and longer patient survival for LUSC. [49]
ARHGDIB and *HOXD3*	TCGA (RNA-seq, DNA-seq, microRNA-seq, CNV, methylation, OS) data	High expression of the *ARHGDIB* gene was found to be an indicator of good prognosis while high expression of the *HOXD3* gene was an indicator of poor prognosis. [50]
Components of Improving Predictive Power of Models	TCGA (RNA-seq, methylation, microRNA-expression, DNA- seq, CNV) data	There is no significant improvement in prediction when adding additional genomic measurements to gene expression and clinical covariates, suggesting that measurements at the transcription level are more predictive of clinical outcomes than at the DNA/epigenetic level for LUSC. [51]
Protein inhibitor of activated STAT3 (*PIAS3*) Expression	Tumor samples, TCGA (RNA-seq, OS) data, cell lines	The cohort of patients with the highest level of *PIAS3* expression demonstrated the highest OS of any group while the cohort of patients with the lowest level of *PIAS3* expression had the lowest OS. There is a decrease in the risk of death by 43% for every one unit of elevation in *PIAS3* expression. [54]
Both		
Gene Fusions in NSCLC	RNA-seq data of other tumor samples, TCGA (RNA-seq) data	Pathway fusions of *NRG1*, *NF1*, and *Hippo* play significant roles in tumors that do not possess conventional driver mutations. Higher numbers of gene fusions is an independent prognostic factor for poor OS. [58]
BCAR1	Tumors samples, TCGA (RNA-seq, protein expression, OS) data	Elevated *BCAR1*-protein levels were correlated with lymphatic metastasis in LUADs, but not in LUSCs. For the TCGA data, there was a 56–77% increased risk of death in both LUAD and LUSC when *BCAR1* expression was doubled. [59]

Table 5

Examples of Molecular Mechanisms Related to Sensitivity (Novel Therapeutic Targets)

Topic	Model Systems	Results
LUAD		
miR-200	TCGA (microRNA- seq, OS) data, cell lines	Low expression of several miR-200 isoforms was associated with worse OS for LUAD patients. Delivery of miR-200a and -200b members to mice bearing 344SQ subcutaneous tumors resulted in a significant decrease in primary tumor burden. [64]
Tristetraprolin (TTP)	TCGA (RNA-seq, OS) data	LUAD patients with *TTP*-low tumors displayed decreased OS compared to patients with *TTP*-high tumors. The study discovered a connection between *CREB* family members and *TTP* expression, which suggests that *CREB* agonists such as colforsin, salbutamol, or clenbuterol may be used to treat *TTP*- low expressing LUAD tumors. [65]
LUSC		
NRF2 pathway	TCGA (RNA-seq, mutation, DNA-seq OS) data	The 28 gene-set called $NRF2^{ACT}$ splits LUSC cases into $NRF2^{ACT}$-high and the wild type $NRF2^{ACT}$-low. Patients with the $NRF2^{ACT}$-high subtype did not benefit from adjuvant platinum-based chemotherapy, indicating an opportunity to spare a subset of patients from unnecessary toxicity. [71]
Copy Number Variations (CNVs) in Pharmacogenes	TCGA (CNV) data	LUSC samples experience the highest number of gene losses, additions, and amplifications among four cancers for known pharmacogenes, presenting a potential barrier for LUSC treatment. [73]
Both		
Ponatinib Sensitivity	Cell lines, tumor samples, TCGA (CNV, RNA-seq) data	*FGFR1* mRNA and protein expression predict ponatinib sensitivity more than *FGFR1* gene copy numbers. As high as 20% of lung cancers may be sensitive to ponatinib. [77]
Azacytidine (AZA)	Cell lines, TCGA (RNA-seq, methylation) data	Treatment of NSCLC cells with AZA induced up- regulation of pathways associated with immune checkpoints and immune evasion. The combination of epigenetic therapy and immune checkpoint therapy may prove to be effective while reducing patient toxicity. [78]

Table 6

Examples for Novel Tools using TCGA Data

Topic	Model Systems	Results
LUAD		
RADIA (RNA and DNA Integrated Analysis)	TCGA (DNA-seq, RNA-seq) data	RADIA is a mutation calling algorithm that combines matched normal and tumor DNA along with tumor RNA. RADIA detected two *TP53* and one *STK11* mutation below the detection threshold of other mutation calling algorithms. [89]
LUSC		
dGene	TCGA (mutation) data	dGene is an annotation tool designed to quickly identify gene mutations that belong to one of 10 druggable classes frequently targeted in cancer therapy. [102]
Integrated Deep Sequencing	Cell line, TCGA (RNA-seq, DNA- seq, mutation) data	This is a proposed method of integrated analysis of deep sequencing for both the genome and transcriptome. For LUDLU-1, a LUSC cell line, somatic mutations in *TP53* and *BRCA2* and reduced transcription of a potentially endogenous *PARP2* inhibitor were found. [103]
Lung Master Protocol	TCGA (DNA-seq, RNA-seq) data	The Lung Master Protocol combines the process of discovering targetable genes and clinical trials for LUSC by using next-generation sequencing, followed by randomization to the pertinent targeted therapy versus standard care for clinical trials. [104]
SigFuge (SIGnificant Forms Using per-base Gene Expression)	TCGA (RNA-seq, CNV) data	SigFuge is an algorithm that identifies cluster within the transcriptome that express isoforms from a single gene locus at different proportions. SigFuge identified *KLK12* expression clusters that may be used as potential biomarkers for LUSC. [105]