Comprehensive comparison of immune features between ultra-low and ultra-high mutation burden non-small cell lung cancers

Zhigang Wang

 Chinese Academy of Medical Sciences and Peking Union Medical College

Zhaojun Duan

 Chinese Academy of Medical Sciences and Peking Union Medical College Institute of Basic Medical Sciences

Chong Chen

 Chinese Academy of Medical Sciences and Peking Union Medical College Institute of Basic Medical Sciences

Xiaoyin Bai

 Peking Union Medical College Hospital

Wei Liu

 Institute of Basic Medical Sciences

Yongqun He

 University of Michigan Medical School

Xiaolin Yang

 Peking Union Medical College and Chinese Academy of Medical Sciences

Bo Yang

 Department of Oncology, Chinese PLA General Hospital

Research article

Keywords: Non-small cell lung cancer, Tumor mutation burden, Tumor microenvironment

Abstract

Background: Tumor mutation burden (TMB) is associated with the formation of tumors and the outcomes of immunotherapy, especially in non-small cell lung cancers (NSCLC). The aim of this study was to study immune features of NSCLC with different mutation burdens. We performed a comprehensive comparison of immune features, including driver mutations, up- or down-regulated expression of key biological pathways, tumor cell evolution, tumor cell stemness, etiology, and immune cell infiltration, between ultra-low (UL) and ultra-high (UH) TMB NSCLC.

Results: Generally, we found that UL-TMB tumors contained a higher proportion of driver mutations and maintained an "immune-desert" tumor microenvironment with reduced effector cell infiltration and increased suppressor cell infiltration. In UH-TMB tumors, the expression levels of cell cycles, DNA replication and mismatch repair pathways were up-regulated. We discovered that UL-TMB tumors had different up-regulated pathways between a denocarcinoma (LUAD) and squamous cell carcinoma (LUSC). Basically, the expression levels of primary bile acid biosynthesis, arachidonic acid metabolism, and drug metabolism pathways were up-regulated in LUAD; however, the expression level of CaM pathway and the PLC-gamma1 signaling pathway were up-regulated in LUSC. Most mutations in UH-TMB tumors occurred in the early stages, while UL-TMB tumors showed the opposite trend. UH-TMB tumors had a relatively high stemness score. Despite the mutations and neo-antigens increasing in UH-TMB tumors, the overall antigen presentation capacity was weak.

Conclusions: These results support rational design of therapeutic strategies for NSCLC with different mutation burdens.

Background

NSLSC is a high-prevalence tumor worldwide. Immunotherapy might turn the malignant cancer into a controllable chronic disease by normalizing the immune system of patients. Different mutational burdens and mutational content can lead to different biological and immunological characteristics [1] of tumor microenvironments. This may also cause different responses to immunotherapy. Tumor mutational burden (TMB) can partly predict the efficacy of immunotherapy [2, 3], but not all tumors with high TMB will give an effective clinical response. Therefore, it is of great importance to study the correlations between TMB and immune status for NSCLC. By comparing the immune status of ultra-low (UL) and ultra-high (UH)-TMB tumor samples, we may have a deepened understanding of the potential mechanisms of how high-TMB tumors respond to immunotherapy and offer new treatment options for patients with UL-low-TMB tumors. Yet, only limited studies are available in this respect.

The development of tumors involves a variety of important changes, including driver gene mutations, up- or down-regulated expression of key biological pathways, tumor cell evolution, tumor cell stemness, and immune cell infiltration. Driver events occur when a tumor is developed, such as deactivation and

mutations of tumor suppressor genes (e.g., TP53), activation and mutations of oncogenes (e.g., PIK3CA), and absent expression of DNA repair genes. In this case, if the immune system fails to eliminate the deteriorated cells, the cell growth will become out of control and result in the formation and production of tumors. It is widely known that tumor cells share some similarities with stem cells, and tumor growth is driven by tumor stem cells. Tathiane M et al. noted greater similarities between metastatic tumors and stem cells compared to primary lesions and argued that tumor cell stemness is the root cause of metastasis [4]. Normal immune cells identify and eliminate tumor cells. A previous study showed that mononuclear cell infiltration could be used as a biomarker in endometrial and breast cancers [5]. L. Cassetta et al. found that tumor-associated macrophages enhanced metastasis and invasion of tumor cells and inhibited anti-tumor immunity by promoting vascularization, which accelerated the development and progression of tumors [6]. Tumor-associated signaling pathways also play an important role in the development and progression of tumors. For example, gain-of-function and loss-of-function mutations caused by abnormal expression of relevant genes and molecules in the phosphatidylinositol 3-kinase/Protein Kinase B, PI3K/Akt pathways can result in abnormal proliferation, apoptosis, and invasion of tumor cells [7].

However, little has been done on discovering and differentiating these immune features of UL- and UH-TMB NSCLCs. In this study, comparisons were drawn between the immune features of UL- and UH-TMB LUAD and LUSC cases while an elaborate analysis was performed to investigate relevant factors related to the development and progression of tumors, including driver mutations, driver genes, key biological pathways, mutation timing, tumor cell stemness, and immune cell infiltration. These results may suggest therapeutic strategies for different types of NSCLC with ultra-low or -high mutational burdens and provide further insights for rational biomarker development for immunotherapies.

Results

1. **Top mutated genes.**

Through segmented linear regression analysis, the low and high-TMB thresholds in LUAD were 25 mutations and 667 mutations. In other words, there were less than 25 mutations in the UL-TMB samples while the UH-TMB samples had over 667 mutations in their whole exome sequencing regions. For LUSC, the break points were 60 and 478 mutations, respectively (Supp. Fig. S1). This study presented a comprehensive comparison of the immune features of UL- and UH-TMB lung cancers. First of all, we analyzed the distribution of mutations on genes in different groups. It was easy to understand that the important cancer-related genes such as TP53 and KRAS were both highly mutated in the UL and UH samples. However, there was a difference in the ranking of top mutated genes. Besides, the spectrum of mutations in LUAD also differed from that in LUSC. As shown in the Figure 1, for LUAD, in the UL-TMB group, mutations occurred in no more than 20% of the samples, with the median as merely 8%. As to the UH group, the top 10 mutated genes were found in over 73% of the samples and there was a higher

degree of consistency in the patterns of mutation spectra. Among the top 10 mutated genes in the UL group, the known proto-oncogenes included EGFR and KRAS and the known tumor suppressor genes included TP53 and KMT2D. The SCN8A, GRIK5, and FAT3 genes in the UL-TMB LUAD group were highly mutated compared to the corresponding mutation rates in the UH group, and the differences were statistically significant.

Likewise, mutations were found in no more than 30% of the UL-TMB LUSC samples (see Figure S2) whereas about 90% of the UH group underwent mutations. However, compared to the UH group, no gene had a significant increased mutation rate in the UL group.

2. Driver mutations and driver genes.

There are studies showing that about 10 mutations can lead to the formation of tumors. So, is there a higher proportion of driver mutations in low-TMB tumors? To find out the answer, the algorithm Oncodrive-MUT was used in this study to predict the driving force of carcinogenic mutations. With extensive tumor genome data taken into consideration, the algorithm predicted the carcinogenicity of the mutations based on 6,792 samples of 28 cancer types and massive samples from healthy donors. This is combined with the features that describe the functional consequences of genetic mutations. As a result, the mutations were classified as driver and passenger mutations. In this study, for LUAD, driver mutations accounted for 4.8% of all 454 mutations in the UL-TMB group. These driver mutations were derived from the following genes: EGFR, GNAS, TP53, KRAS, RPS6KA3, HCFC1, TSC1, BRAF, SYNCRIP, MED23, FAT1, SRGAP3, SETD2, ZFP36L2, CAD, COL1A1, TP53, and ATR. In the UH-TMB LUAD group, 2.1% of the 47,908 mutations were driver mutations, which originated from a total of 422 genes. The UL group had a significantly higher proportion of driver mutations. In terms of the LUAD samples, the following genes were associated with the driver mutations that occurred in the UL group, including lRPS6KA3, TSC1, SYNCRIP, ZFP36L2 and ATR. TSC1 is a tumor suppressor in the mTOR signaling pathway, which is inactivated by mutation or deletion in a diverse range of cancers. Germline and somatic TSC1 mutation is ais feature of the disease tuberous sclerosis complex. ATR is a tumor suppressor gene involved in DNA damage repair, which is mutated in various cancer types. These UL specific driver genes interfered with the pathway such as TOR signaling pathway which control cell growth and proliferation, cell aging, and DNA replication signaling pathways (see the Figure 2A and Table S1).

For LUSC, driver mutations in the UL-TMB samples were derived from the TP53, BRCA1, CDC73, HERC2, HGF, MAP2K1, MGA, MYD88, NF1, NFE2L2, PIK3C2B, POT1, RB1, SPTAN1, TAOK2, TJP1, TNPO1, and TRAF3, among which MAP2K1, MYD88, PIK3C2B, SPTAN1, TNPO1, and TRAF3 were driver genes only found in the UL group. Particularly, MAP2K1 is an oncogene and intracellular kinase that is mutated at

low frequencies in a variety of cancer types, including melanoma, colorectal and lung cancers; MYD88, an oncogene and adaptor protein, is frequently altered in hematologic malignancies including Waldenström's macroglobulinemia; TRAF3 is a tumor suppressor, signaling molecule, and E3 ligase, which is recurrently mutated and deleted in B cell lymphoma and multiple myelomas. These genes are involved in the pattern recognition receptor signaling pathway, innate immune response-activating signal transduction, toll-like receptor signaling pathway, and regulation of innate immune response, as shown in the Figure 2B and Table S1.

Mutations were classified as either early or late based on their clonal status. It was found that in the UL TMB LUAD and LUSC groups, mutations largely occurred following subcloning or changes in chromosome number. In other words, most mutations occurred at the later stages of tumors; as to the UH groups, the opposite was the case (Figure S6).

3. **Pathways dis-regulation related to mutation burden.**

Generally, in addition to driver genes, there are plenty of infrequently mutated genes that may contribute to tumor biology by affecting the expression levels of relevant genes. Hence, it is interesting to analyze pathways and networks involved and illuminate the biological processes involved. In this study, Gene Set Variation analysis (GSVA) was employed in the comparison between the pathways regulation of the UL- and UH-TMB groups. See the figure below.

As to LUAD, it was found that in the UH-TMB group, the biological pathways with upregulated expression were mainly associated with the cell cycle, base excision repair, homologous recombination, DNA replication, and mismatch repair pathways. Interestingly, in the UL-TMB group, upregulation of expression occurred in the primary bile acid biosynthesis, arachidonic acid metabolism, drug metabolism cytochrome 50, complement and coagulation cascades, and calcium signaling pathways (see Figure 3). For the LUSC samples, expression of the Eicosanoid ligand-binding receptors, the CaM pathway, and the PLC-gamma1 signaling pathway were upregulated in the UL group (Figure S3).

The tumor map was drawn using an interactive plotting system to visualize, the expression data and explore the molecular similarities of the cancer samples. By labeling TMB groups' information of each sample, it was noted that the UL-TMB samples frequently fell between normal tissue and the major tumor subtypes. The tumor map of the LUAD and LUSC samples is shown shown as Figure S4.

4. Stemness analysis.

Cancer progression involves the gradual loss of a differentiated phenotype and acquisition of progenitor and stem-cell-like features. This study used a one-class logistic regression (OCLR) machine-learning algorithm to analyze the stemness features of the samples, which were associated with oncogenic dedifferentiation. The results showed that the stemness of the UH group was significantly higher than that of the UL group (Wilcoxon rank-sum test $P < 0.05$) for both LUAD and LUSC (Figure 4 A is for LUAD, B is for LUSC).

5. Immunophenotype analysis.

The immunophenoscore (IPS) was calculated on an arbitrary 0–10 scale based on the sum of the weighted average Z score of the four categories. In terms of the overall IPS, there was no difference between the two groups. However, when the IPS was decomposed into 4 categories, the differences between the UL- and UH-TMB groups were statistically significant in antigen processing by major histocompatibility complex (MHC) proteins, effector cells (ECs), and suppressor cells (SCs) for LUAD. As to checkpoints (CPs) or immunomodulators, the UL group's score appeared to be higher than the UH group's. See the Figure 5B, 5C and 5D.

Although the differences between the UL- and UH-TMB groups lacked statistical significance for LUSC; they exhibited similar trends as the differences between the UL- and UH-TMB LUAD groups. Specifically, the UL group had higher scores for antigen presentation mediated by MHC and immunomodulators, indicating that the immune status remained normal. Also, it was found that the UL group's scores for ECs and SCs were significantly lower than those of the UH group's, indicating immunosuppression and inactivation of all immune cells in the samples.

The findings above were proved in the analysis of 22 immune cell type fractions using the CIBERSORT method. However, some special cells showed the opposite. For example, the proportion of dendritic cells was significantly higher compared to other activated cells while there were much less T cells follicular helpers in the UL group. T cells CD4 memory resting, dendritic cells resting, and mast cells resting were also highly infiltrated in the UL group (Figure S5).

6. Mutation signature analysis.

Results from the mutation signature analysis were rather interesting (see the Figure 6). Comparing the UL groups of the LUAD and LUSC samples, their mutation spectra shared great similarities, which were

dominated by signature 1 and 3. As to the UH groups, the mutation spectra were basically covered by signature 4. Signature 1 was present in all cancer types and in most cancer samples as a result of the endogenous mutational process initiated by the spontaneous deamination of 5-methylcytosine. Signature 3 was associated with failure of DNA double-strand break-repair by homologous recombination. Signature 4 was found in head and neck cancer, liver cancer, lung adenocarcinoma, lung squamous carcinoma, small cell lung carcinoma, and esophageal cancer, which was associated with smoking, with its profile similar to the mutational pattern present in the experimental systems exposed to tobacco carcinogens. It was likely to be caused by tobacco mutagens. Signature 29 was seen in cancers associated with tobacco chewing and appeared to be different from Signature 4.

Discussion

Through analysis of the mutation spectra of the sample genes, it was noted that the UL-TMB samples showed a higher degree of heterogeneity and greater differences in their mutation patterns. In low-TMB tumors, mutation might be a less important driving factor. Since there are fewer mutations and mutation-associated genes and a limited number of driver genes in a single sample, targeted therapy designed for low-TMB tumors may produce better outcomes; yet such targeted therapy must be highly individualized. Considering the acquired immune privilege brought by a low-TMB, the immune cell recruitment becomes less effective; immune escape occurs more frequently; the prognosis is no better than that in high-TMB cases. In other words, TMB can be used to predict the response to immunotherapy but never serves as an effective predictive factor of treatment outcomes and prognosis.

Unlike the low-TMB samples, the high-TMB ones revealed a higher degree of similarity, with the mutation status of many genes consistent with each other in over 90% of all samples. This indicates that relevant tumors are driven by the same force, in which case, universal targeted therapy may outweigh individualized therapy. This finding is agreed upon by Brittany B. Campbell et al. who argued that high-TMB tumor samples had a highly similar mutation spectrum, which suggested that the samples had the same driving force [8]. Based on the etiological analysis of mutational signatures on the mutation spectrum using the Catalogue of Somatic Mutations in Cancer (COSMIC) database, it was found that hypermutation was largely caused by environmental factors (e.g., smoking), which is in agreement with the conclusions given in Steven A. Roberts et al. [9] Erin D Pleasance et al. also proved that over 60 carcinogens in tobacco smoke were the primary cause of hypermutation in tumors in the lung [10]. This indicates that high-TMB lung cancers are probably caused by exogenous factors and the low-TMB ones by endogenous factors as the mutation spectra reveal common features present in multiple cancer types. Further, the LUAD and LUSC groups exhibited the same patterns, which provides solid evidence for the hypothesis.

The analysis of ploidy showed that the UH-TMB samples had higher ploidy levels. Konrad H Stopsack et al. found that a greater degree of aneuploidy in the chromosomes of prostate cancer cells indicated a higher mortality rate [11]. On this basis, the degree of aneuploidy can be used as a prognostic indicator to facilitate the determination of a radical treatment plan. However, the timing of mutation suggested that compared to changes in chromosome number, most mutations occurred during the earlier stages of cancer progression, whereas aneuploidy of the UH-TMB samples seemed to be a phenomenon in the later stages. In other words, aneuploidy is more of the result of cancer progression than a driver event. A higher ploidy level indicates that the mitosis mechanism of tumor cells was thrown into disorder, which entails a higher risk of errors or rearrangement and enhances cancer cell proliferation. The presence of aneuploids strengthens the invasiveness of cancer cells. These findings are consistent with the significant increase in stemness. Aavoli T et al. found that aneuploidy might attenuate the antigen presentation mediated by MHC [12], an essential part of the immune system to identify the presence of tumors. The immunophenotype score proved that high-TMB tumors, namely the tumors at higher ploidy levels, scored higher for MHC compared to the low-TMB tumors in the LUAD group. Additionally, the LUSC samples also followed similar trends.

This study found that mutations were less frequent in the UL-TMB samples and yet, the driver mutation percentages in the UL-TMB groups were significantly higher than those in the UH-TMB groups. This indicates that driver mutations might promote carcinogenesis in the UL-TMB samples. In contrast, the UH-TMB groups had more non-driver mutations that might interfere with cell functions, affect cell proliferation and enhance immune identification, thereby inhibiting tumor growth. Mutations of the TP53 and KRAS genes occurred frequently in both the UH and UL-TMB groups. This finding is agreed upon by Bailey MH et al. [13]. For the LUAD samples, driver mutations in the UL-TMB group were derived from RPS6KA3, TSC1, SYNCRIP, ZFP36L2, and ATR. Particularly, RPS6KA3 and LIHC were considered as driver genes; TSC was a driver gene in the pan-cancer and BLCA analysis; ATR was a driver gene in the pan-cancer and UCEC analysis [6] – All these are tumor suppressor genes. In the pan-cancer and COADREAD BLCA analysis, ZFP36L2 is defined as a driver gene. The RPS6KA3 gene is involved in the MAP kinase signaling pathway and specifically, is responsible for upregulation of cell proliferation and differentiation in a cell cycle. TSC1 is involved in the mTOR signaling pathway, which appears to be highly sensitive to Rapamune, a typical inhibitor of mTOR in tumors driven by the gene. SYNCRIP is a negative regulator of translation and makes cellular responses to interferon-gamma; in the immune response system, the SYNCRIP gene is responsible for antigen presentation to achieve autoimmunity. ZFP36L2 (ZFP36 ring finger protein like 2) is mainly associated with transcription factors, T cell differentiation in thymus, and B cell differentiation; ZFP36L2 also plays a role in cell growth and damage response. ATR serine/threonine kinase (ATR) is important enhancer of DNA damage response in the cell cycle. As for the genes that only existed in the UL-TMB LUSC samples, driver mutations originated from MAP2K1, MYD88, PIK3C2B, SPTAN1, TNPO1, and TRAF3, and the presence of MAPK1, MYD88, SPTAN1, and TRAF3 was in line with the conclusions drawn by Mathew H. Balley et al. Identification of driver mutations may facilitate the development of molecular targeted therapeutics for UL-TMB tumors.

It has been suggested that all cancers display defects in DNA repair. In some tumors like melanoma, cell cycle checkpoint/DNA repair gene upregulation is associated with tumor metastasis, whereas in many other cancers the opposite has been observed [14]. Upregulation of DNA damage response and repair genes can also cause cancer and increase resistance of cancer cells to DNA damaging therapy [15]. In this study, it was noted that the cell cycle, homologous recombination, DNA replication, DNA deletion and repair, and mismatch repair pathways were upregulated in the UH-TMB group for LUAD, indicating insignificant therapeutic effect on DNA damage caused by the tumors. As to the UL-TMB tumors, upregulation was observed in the primary bile acid biosynthesis, arachidonic acid metabolism, drug metabolism cytochrome p450, complement and coagulation cascades, and calcium signaling pathways. Compared to the UH group, the UL-TMB samples often had metabolic and coagulation abnormalities. Bile acids serve as a regulator of cell growth and proliferation. In case of liver injury and tumorigenesis, changes in the level of bile acids can be observed [16]. This is probably a symptom accompanying low-TMB lung cancer. Studies have shown that subclinical activation of blood coagulation may occur during the earlier stages of lung cancer. In addition, clotting activation is confirmed as a predictor of survival [17]. Altered expression of specific Ca2+ channels and Ca2+-binding proteins is a characterizing feature of lung cancer, which regulates cell signaling pathways leading to cell proliferation or apoptosis. Hypoxia has a vital role in tumor angiogenesis, metastasis, and apoptosis; also, it induces Ca2+ channels to open with an increase in the Ca2+ influx that causes tumor growth [18]. For the LUSC samples, the pathways in the UH group that underwent upregulation were basically the same as those in the UH-TMB LUAD group, which were mainly associated with the DNA cycle; the Eicosanoid ligand-binding receptors associated with inflammation and the CaM pathway related to autophagy were upregulated in the UL group.

Generally, it is believed that cancer stem cell subpopulations induce cancer cell reproduction and growth in a disorderly style. Breast cancer stem cells (BCSCs) are characterized in their resistance to chemotherapy, radiotherapy, and hypoxia, as well as the high oncogenicity, invasiveness, and metastasis. These features play a critical role in the development, progression, recurrence, and metastasis of breast cancer as significant causes of recurrent breast cancer and important factors closely associated with prognosis in breast cancer. The similarity between cancer cells and stem cells can be assessed using the stemness indices. Cell similarity indices were analyzed on the basis of gene expression features, with a higher value indicating closer correlations with cancer progression of multiple cancer types. A previous study revealed strong similarities between metastatic tumor cells and stem cells [19]. Tumor cell stemness is a useful tool to identify new targets of anticancer drugs, thereby promoting the development of novel therapy to inhibit cancer progression. In this study, the UH groups showed a higher degree of stemness compared to the UL groups of the LUAD and LUSC samples. Yet, there is no significantly positive correlation between TMB and new tumor events. This is probably because the highly activated immune cells are subject to large-scale infiltration, which inhibits distant metastasis of tumor cells.

Components of immune status of the samples were analyzed using IPS immunophenoscore that involved MHC, ECs, SCs, and CP (checkpoint) molecules. It was found that in the UL groups, the SC and EC scores were significantly lower compared to the UH groups, indicating that the UL groups were under immunosuppression where immune-desert tumors were observed. Despite the high scores for immune cell infiltration, the UH groups scored relatively low in MHC, indicating the poor antigen presentation capacity of the UH-TMB samples. Joan Font-Burgada et al. found that in TCGA patients, a higher mutation frequency meant poorer antigen presentation of mutant peptides by HLA [20], which is in agreement with our findings. From the perspective of tumor evolution, the mutant clones strongly presented by HLAs in TCGA patients have been identified and largely eliminated by immune cells during early immune surveillance. Only the mutant clones with poor presentation by HLAs are likely to escape from immune surveillance, thereby leading to a high mutation frequency after formation and production of tumors. Hence, the TMB is significantly higher in a patient with a lower MHC score than in those with normal MHC antigen presentation. If the immune surveillance mediated by new antigens that are identified by HLAs is impaired, there is a higher risk of immune escape in the tumor clones carrying mutations, which at the macroscopic level leads to a high-TMB. HLA alleles have innate differences in antigen presentation. Timothy A. Chan et al. revealed that a greater variety of HLA alleles indicated a higher possibility of a patient benefiting from immunotherapy [21]. It was found in this study that the UH groups had relatively low scores for MHC while the sample dispersion remained very high. In this case, the MHC status should be taken into full account when developing individualized immunotherapy.

Looking into effector cells, dendritic cells (DCs) in the UL groups were highly infiltrated. As one of the most powerful dedicated antigen presentation cells (APCs) in the immune system, DCs play a crucial role in the activation and regulation of immune responses. DC-based immunotherapy is safe and can induce antitumor immunity, even in patients with advanced disease. However, clinical responses have been disappointing, with classic objective tumor response rates rarely exceeding 15% [22]. In high-TMB cases, patients with low DC infiltration may benefit better from DC-based immunotherapy. These findings also apply to LUAD and LUSC patients.

It is important to note that our study is not without limitations. Only UL- and UH-TMB samples are included in the present study and thus the findings have maximized the differences in immune features. Further, the conclusions in relation to LUAD and LUSC do not necessarily apply to other cancers.

Conclusion

In the present study we performed a comprehensive comparison of immune features, including driver mutations, up- or down-regulated expression of key biological pathways, tumor cell evolution, tumor cell

stemness, etiology, and immune cell infiltration, between ultra-low (UL) and ultra-high (UH) TMB NSCLC. These findings support rational design of therapeutic strategies for NSCLC with different mutation burdens.

Methods

In this study, exome sequencing, RNA-seq and clinical data of the NSCLC samples were acquired from the public TCGA project. The project molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. No cell line was used in this study. The database has been extensively applied to studies on pan-cancer tumor classification at the molecular level, cancer progression [23], and signaling pathways[24] in tumors. A systematic comparison was drawn between UL- and UH-TMB tumors in patients diagnosed with non-small cell lung cancer (NSCLC), including driver mutations, driver genes, mutation timing, immune cell infiltration, changes in expression of biological pathways, and tumor cell stemness. To guarantee the applicability of the conclusions in this study, the LUAD and LUSC samples were analyzed respectively.

To determine the UL and UH cutoffs, somatic mutation data of 565 LUAD samples and 491 LUSC samples using the mutect2 algorithm were obtained from the database. The synonymous variants and variants in the intergenic or non-coding regions were filtered out for mutation burden analysis. The somatic mutation data in relation to the LUAD and LUSC samples were subject to segmented linear regression analysis [25]using the R package segmented [3]. If a sample had a TMB below the lower cutoff, the sample would be categorized into the UL-TMB group; in contrast, if a sample had a TMB higher than the upper cutoff, it would be classified as an UH-TMB sample.

The mutation spectrum was visualized using the R package MAFtools [26], through which the top mutated genes, mutation rates, and mutation classes were presented vividly. The most likely driver mutations of a tumor were identified by the OncodriveMUT bioinformatics method [27]. This method incorporates features characterizing the genes (or regions within genes) where mutations occur, derived from the analysis of cohorts of tumors (6,792 samples across 28 cancer types) and samples from healthy donors (60,706 unrelated individuals). The GeneMANIA plugin [6] in the Cytoscape [29] software (version 3.6.0; National Institute of General Medical Sciences, Seattle, WA, USA) based on a large set of functional association data, including protein and genetic interactions, co-expression, co-localization pathways, and protein domain similarity, was employed in investigating the functional similarities of the driver genes.

The analysis concerning the timing of mutational processes in cancer evolution can facilitate the identification of driver events that occur during evolution of tumor clones. All this is useful information for the selection of targeted therapy. Information in relation to mutation timing was obtained from McGranahan N et al [30]. The chi-squared test was utilized in the comparison between the percentages of mutations in the UL- and UH-TMB LUAD and LUSC groups during the early and late stages.

Somatic mutations might be caused by endogenous factors (e.g., DNA repair deficiency) or exogenous factors (e.g., exposure to carcinogens). Mutations caused by different factors have different mutation spectra. Through analysis of mutation spectra, we may gain a better understanding of the underlying biological processes regarding the formation and production of tumors and select optimal treatment plans for patients. In this study, deconstructSigs [31] was used to calculate the weighted contribution of the 30 published COSMIC signatures and one additional unknown signature to the collection of somatic variants for each group. As the UL groups had very limited mutations, all mutations in the UL and UH groups were divided into two classes, respectively.

The expression of transcriptome was analyzed following the mutation data. The differences in regulation of pathway expression in the UL and UH groups were assessed using the GSVA (gene-set variance analysis) package from R bioconductor [32], a non-parametric, unsupervised method for estimating variation of gene set enrichment through the samples of an expression data set. In the present study, the gene sets, KEGG, Reactome, and BioCarta, were selected. To calculate the GSVA enrichment scores of these gene sets using count data, the kernel parameter was set to "Poisson". For the LUAD samples, there were plenty of pathways having statistically significant differences in their expression levels when FDR < 0.05 was considered as the standard for statistical significance; hence, FDR < 0.00001 was set as the standard for statistical significance in the LUAD cases. As to the LUSC samples, FDR < 0.05 indicated the level of statistical significance.

The UCSC TumorMap [33] analysis was used to visualize clusters of expression profiles across all LUAD or LUSC samples. The expression RPKM values were employed in generating genes from the sample matrix containing 56,716 genes and 574 samples for LUAD. The LUSC sample matrix included 552 samples. The expression matrix above and the sample annotations including the TMB groups were uploaded to the TumorMap portal for analysis.

Tumor cell stemness was analyzed using a machine-learning algorithm. NSCLC data from the TCGA database were directly obtained from the work of Tathiane M.Malta et al [19]. Through sample ID

mapping, the samples were divided into different groups to compare the stemness of the UL and UH groups, which was examined using the Wilcoxon rank-sum test.

IPS immunophenoscore [34] is an aggregated score based on the expression of the representative genes or gene sets comprising four categories: MHC molecules, immunomodulators, effector cells (activated CD8+ T cells and CD4+ T cells, Tem CD8+, and Tem CD4+ cells), and suppressor cells (Tregs and MDSCs). IPS can be used for predicting the response of melanoma to checkpoint blockade while it takes into consideration the tumor-intrinsic factors and extrinsic factors present in the tumor microenvironment. Downloaded FPKM data were converted into log2 (TPM+1) to calculate each sample's IPS score and the differences in the four categories between the UL and UH groups.

Based on the log2 (TPM+1) data, the CIBERSORT algorithm was used to infer the TIL proportions of the tumor microenvironment. The LM22 dataset that was comprised of 22 distinct immune cell types and constructed from the gene expression profiles of these cell types was downloaded from the CIBERSORT [35] website (https://cibersort.stanford.edu/download.php).

Abbreviations

TMB : Tumor mutation burden; NSCLC: Non-small cell lung cancer; UH-TMB: Ultra high tumor mutation burden; UL-TMB: Ultra low tumor mutation burden

Declarations

-Ethics approval and consent to participate

Not applicable

-Consent for publication

Not applicable.

-Availability of data and materials

The datasets used and/or analysed during the current study are available from TCGA project.

-Competing interests

The authors declare that they have no competing interest.

-Funding

This research is supported by CAMS Innovation Fund for Medical Sciences, 2018-I2M-AI-009; National Key R&D Program of China, 2017YFC0907505; Peking Union Medical College Youth Fund 2014, No. 06061; Science, Technology and Innovation Commission of Shenzhen Municipality, JCYJ20160428164548896. The funding bodies had no role in the design of the study or in the collection,analysis and interpretation of data or in writing the manuscript.

-Authors contribution

X.Y. and B.Y. designed the research study. Z.W. performed the analysis. W.L. and X.B. gave suggestion on the analysis pipeline. X.Y., Z.D. and C.C. assisted in the analysis. Z.W., C.C. and Z.D. wrote the paper. D.B. and J.L. asked many key questions about this study and helped to check results. Y.H. helped to proofread part of the paper in English, and gave some research plan.

-Acknowledgements

References

1. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354:618–22.

2. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N Engl J Med. 2017;377:2500–1.

3. Samstein RM, Lee C-H, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019;51:202–6.

4. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell. 2018;173:338-354.e15.

5. Cassetta L, Fragkogianni S, Sims AH, Swierczak A, Forrester LM, Zhang H, et al. Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. Cancer Cell. 2019;35:588-602.e10.

6. Cassetta L, Pollard JW. Targeting macrophages: therapeutic approaches in cancer. Nat Rev Drug Discov. 2018;17:887.

7. Millis SZ, Ikeda S, Reddy S, Gatalica Z, Kurzrock R. Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19784 Diverse Solid Tumors. JAMA Oncol. 2016;2:1565–73.

8. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell. 2017;171:1042-1056.e10.

9. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer. 2014;14:786–800.

10. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, et al. A small cell lung cancer genome reports complex tobacco exposure signatures. Nature. 2010;463:184–90.

11. Stopsack KH, Whittaker CA, Gerke TA, Loda M, Kantoff PW, Mucci LA, et al. Aneuploidy drives lethal progression in prostate cancer. Proc Natl Acad Sci. 2019;116:11390–5.

12. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science. 2017;355.

13. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;174:1034–5.

14. Gavande NS, VanderVere-Carozza PS, Hinshaw HD, Jalal SI, Sears CR, Pawelczak KS, et al. DNA repair targeted therapy: The past or future of cancer treatment? Pharmacol Ther. 2016;160:65–83.

15. Broustas CG, Lieberman HB. DNA Damage Response Genes and the Development of Cancer Metastasis. Radiat Res. 2014;181:111–30.

16. Li T, Apte U. Bile acid metabolism and signaling in cholestasis, inflammation and cancer. Adv Pharmacol San Diego Calif. 2015;74:263–302.

17. Ferrigno D, Buccheri G, Ricca I. Prognostic significance of blood coagulation tests in lung cancer. Eur Respir J. 2001;17:667–73.

18. Yang H, Zhang Q, He J, Lu W. Regulation of calcium signaling in lung cancer. J Thorac Dis. 2010;2:52–6.

19. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell. 2018;173:338-354.e15.

20. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-I Genotype Restricts the Oncogenic Mutational Landscape. Cell. 2017;171:1272-1283.e15.

21. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell. 2017;171:1259-1271.e11.

22. Anguille S, Smits EL, Lion E, van Tendeloo VF, Berneman ZN. Clinical use of dendritic cells for cancer therapy. Lancet Oncol. 2014;15:e257-267.

23. Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, et al. Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. Cell. 2018;173:305-320.e10.

24. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173:321-337.e10.

25. Muggeo VMR. Estimating regression models with unknown break-points. Stat Med. 2003;22:3055–71.

26. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747–56.

27. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10:25.

28. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinforma Oxf Engl. 2010;26:2927–8.

29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

30. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med. 2015;7:283ra54.

31. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17. doi:10.1186/s13059-016-0893-4.

32. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013;14:7.

33. Newton Y, Novak AM, Swatloski T, McColl DC, Chopra S, Graim K, et al. TumorMap: Exploring the Molecular Similarities of Cancer Samples in an Interactive Portal. Cancer Res. 2017;77:e111–4.

34. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18:248–62.

35. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7.

Supplementary Information

Table S1. GO enrichment for UL specific dirver genes in GeneMANIA network.

Figure S1. Segmented linear regression analysis to determine UL and UH cutoff for LUAD and LUSC.

Figure S2. Mutation waterfall plot for LUSC. (A) UL samples. (B) UH samples. The block color represent mutation type. The right bar-plot means percentage of samples mutated on a specific gene. The top bar-plot means mutation burden of each sample.

Figure S3. Pathway regulation heatmap for LUSC by Gene Set Variation analysis. The column represents samples, The UL samples was marked by blue color. The UH samples was colored red.

Figure S4. Tumor map by gene expression. (A) LUAD. (B) LUSC. The samples groups were colored as legend. The gray color represent normal sample.

Figure S5. TILs by CIBERSORT boxplot.

Figures

Figure S6. Mutation timing comparison.
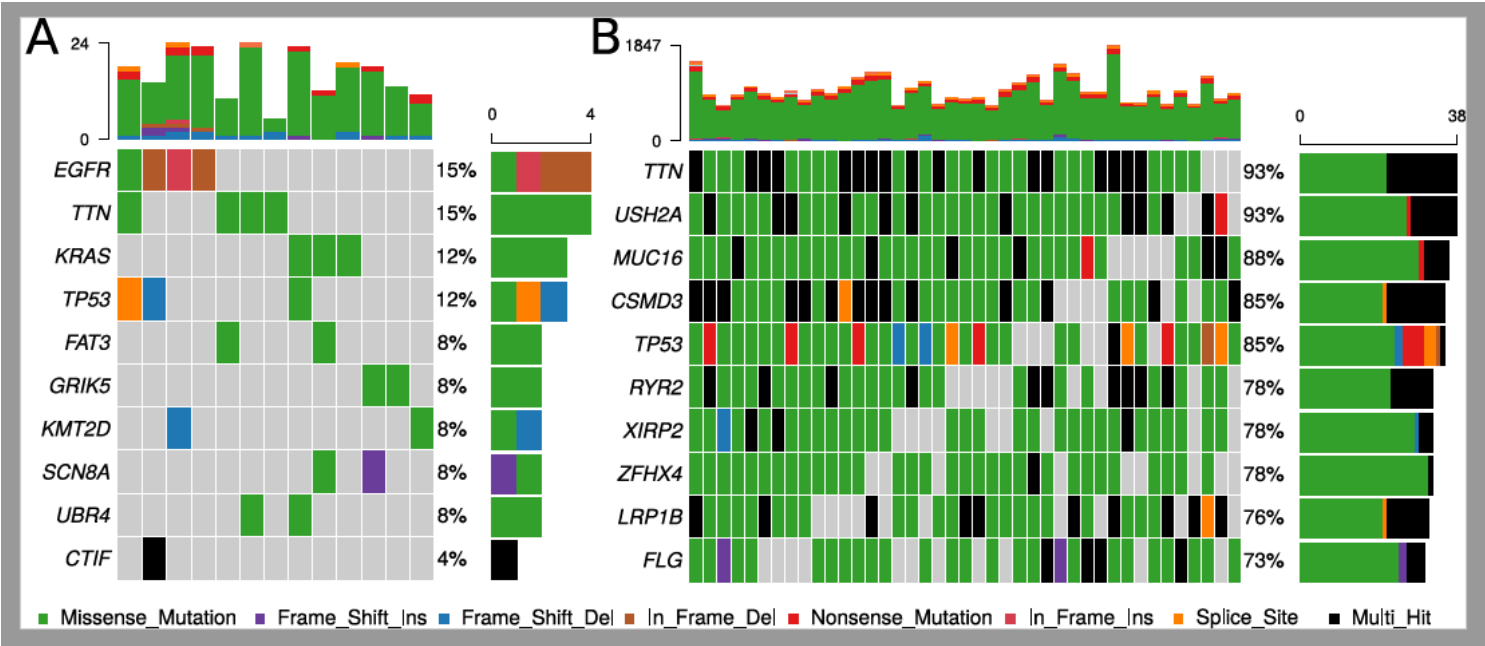
Figure 1

Mutation waterfall plot for LUAD. (A) UL samples. (B) UH samples. The block color represent mutation type. The right bar-plot means percentage of samples mutated on a specific gene. The top bar-plot means mutation burden of each sample.
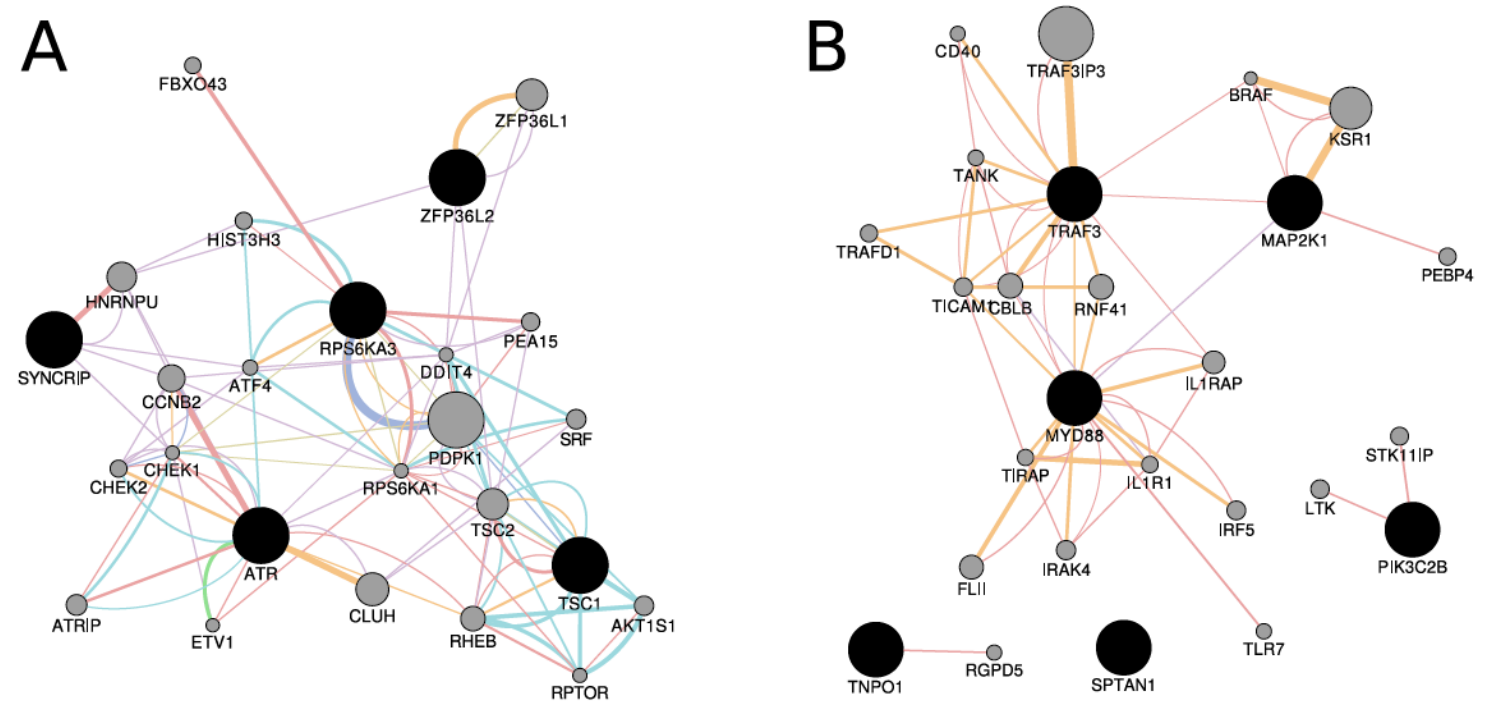
Figure 2

GeneMANIA network for UL specific dirver genes. (A) for LUAD samples. (B) for LUSC samples.
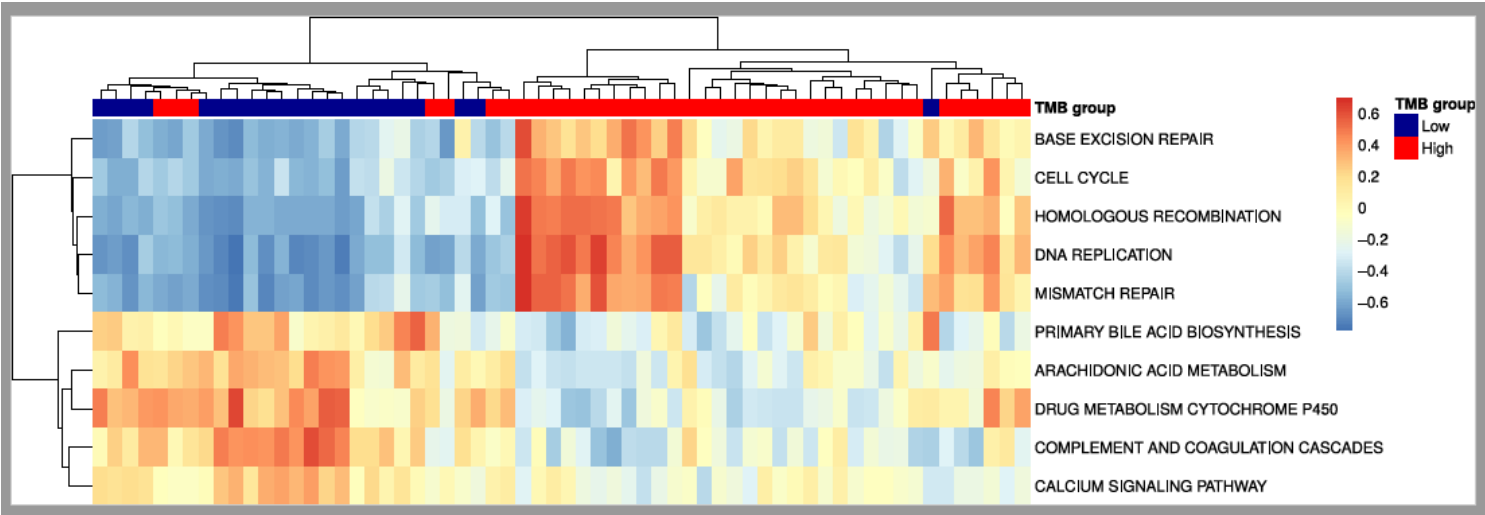
Figure 3

Pathway regulation heatmap for LUAD by Gene Set Variation analysis. The column represents samples, The UL samples was marked by blue color. The UH samples was colored red.
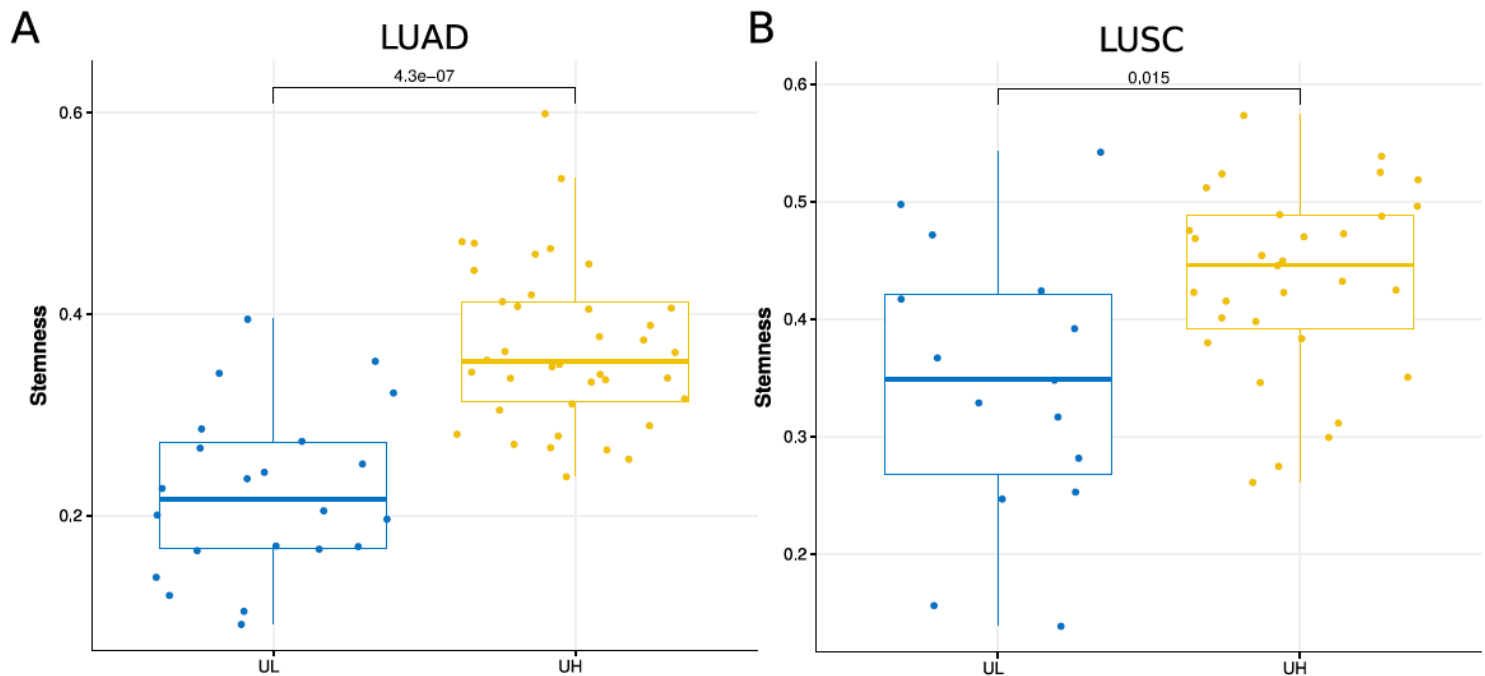
Figure 4

Stemness comparison between UL and UH samples. (A) for LUAD samples. (B) for LUSC samples. The stemness of the UH group was significantly higher than that of the UL group.
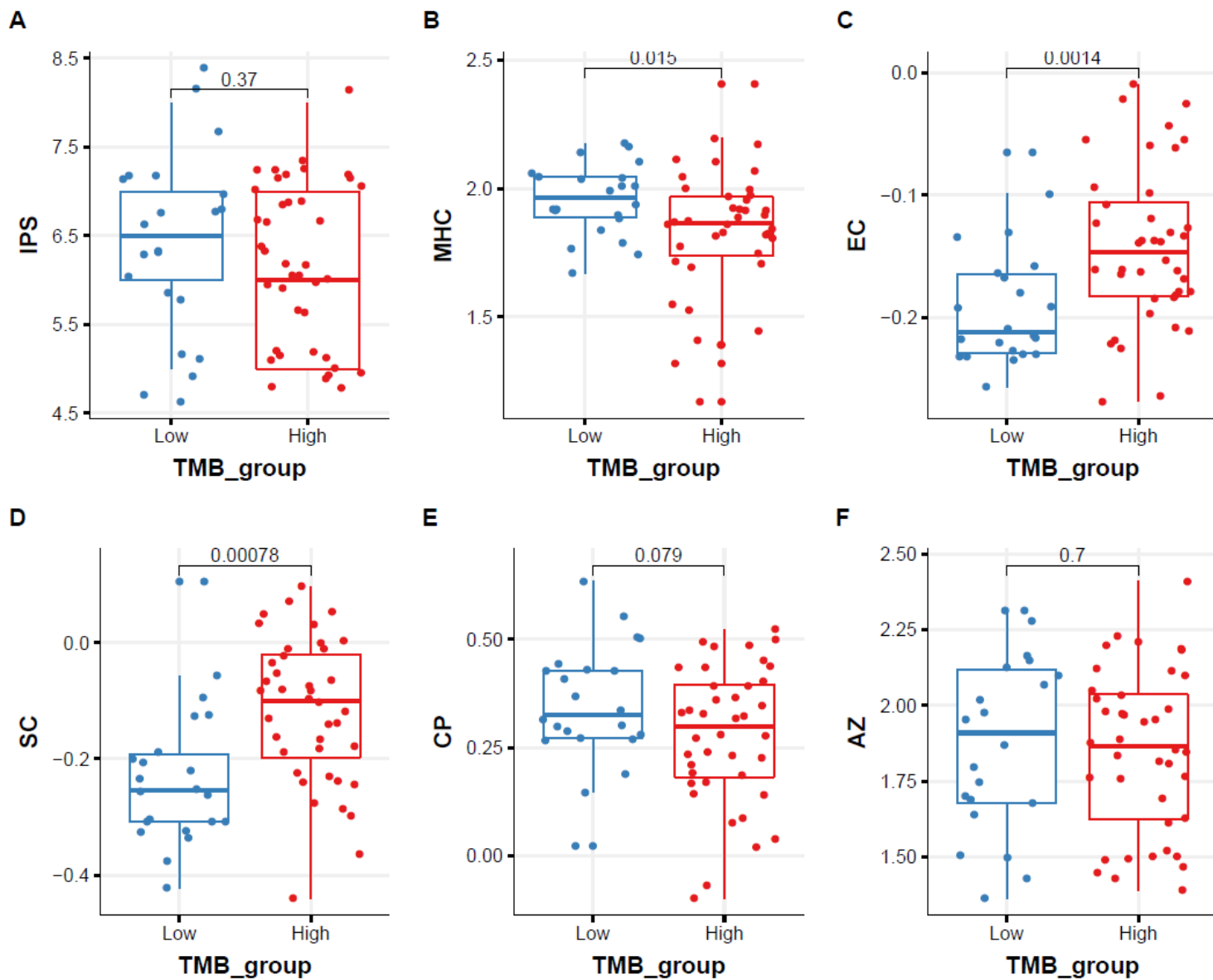
Figure 5

Immunophenotype comparison between UL and UH samples for LUAD. The Immunophenotype was decomposed into 4 categories, major histocompatibility complex (MHC) proteins, effector cells (ECs), checkpoints (CPs) and suppressor cells (SCs).
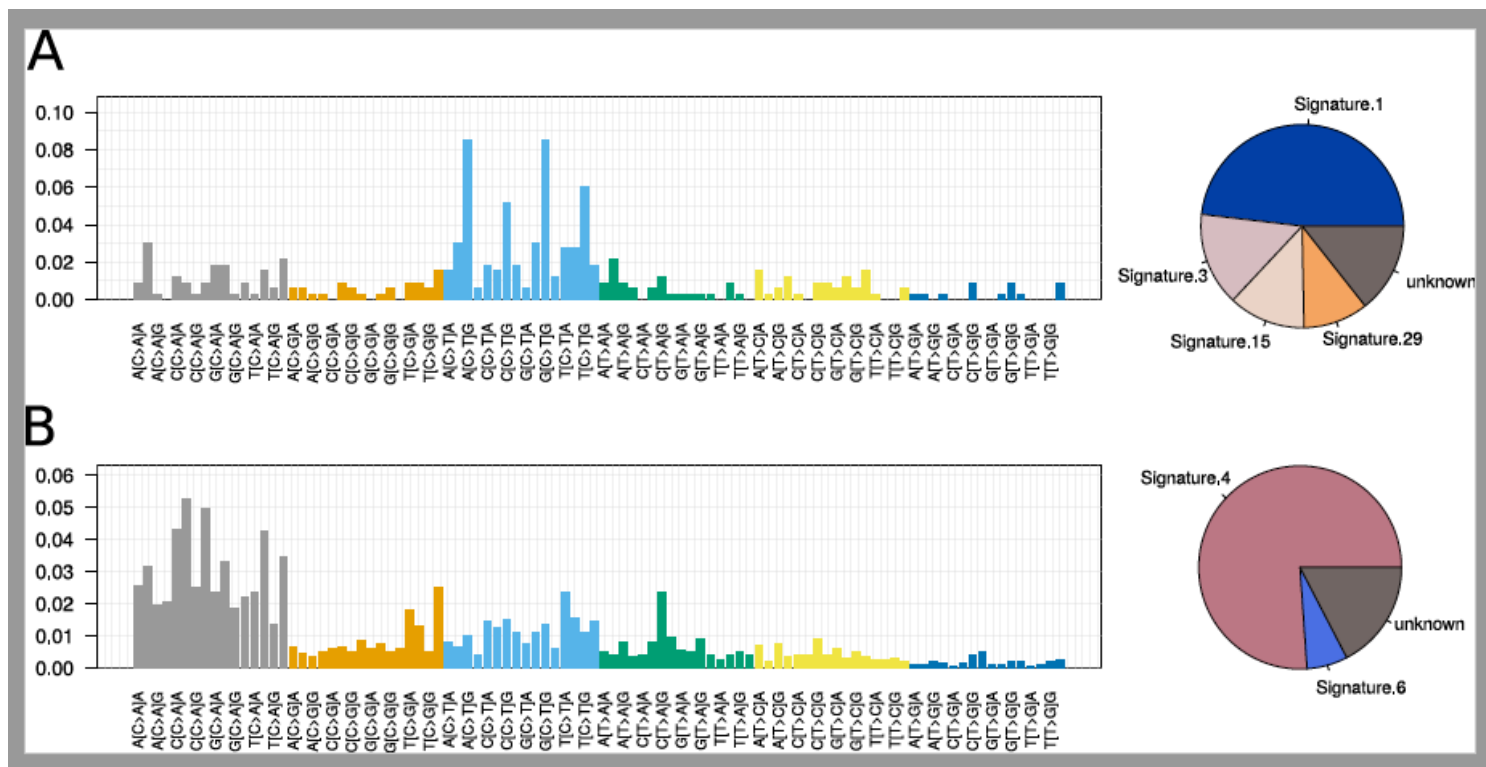
Figure 6

Mutation signature analysis based on COSMIC 30 signatures. As the UL groups had very limited mutations, all mutations in the UL and UH groups were divided into two classes, respectively