

Integrative analysis of multi-omics data reveals the heterogeneity and signatures of immune therapy for small cell lung cancer

Yabin Chen

Shanghai Tech University

Zhaoyuan Fang

Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0002-0393-2052>

Ying Tang

Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences

Yujuan Jin

State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences

Chenchen Guo

Institute of Biochemistry and CState Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Biology, Shanghai Institutes for Biological Sciences, C

Liang Hu

State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences

Yang Xu

Chinese PLA General Hospital

Jie Gao

Chinese PLA General Hospital

Mei Xie

Affiliated Zhongshan Hospital of Dalian University

Xuelai Zang

Chinese PLA General Hospital

Xi Ma

Chinese PLA General Hospital

Sanhong Liu

Shanghai University of Traditional Chinese Medicine

Haiquan Chen

Fudan University Shanghai Cancer Center

Xinying Xue

Chinese PLA General Hospital

Hongbin Ji

State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences <https://orcid.org/0000-0003-0891-6390>

Luonan Chen (✉ Inchen@sibs.ac.cn)

Institute of Biochemistry and CState Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Biology, Shanghai Institutes for Biological Sciences, C

Article

Keywords: Small cell lung cancer, network modules, tumor heterogeneity, immune subtype, immunotherapy

Posted Date: January 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-148857/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Clinical and Translational Medicine on December 1st, 2021. See the published version at <https://doi.org/10.1002/ctm2.620>.

Abstract

Small cell lung cancer (SCLC) is highly invasive and lethal. Genomic studies are beginning to characterize its genetic determinants and heterogeneity. Here we performed RNA-Seq and whole exome sequencing (WES) in 19 Chinese SCLC clinical tumor specimens. Integration with other two public cohorts (n = 129 for RNA-seq, n = 171 for WES), we carried out gene co-expression network analysis and classified into four subgroups: ASCL1-high, NEUROD1-high, a novel variant subtype with CCSP/CC10-high, and a fourth subtype with high expression of NOTCH family and inflammatory genes. We further found that this fourth subtype was characterized by overexpression of immune-related pathways and immunosuppressive factors, referred to as immune subtype. In addition to transcriptomic signatures, specific genomic alterations were also significantly enriched in each subtype. We successfully built a machine learning model to predict the subtypes of newly coming specimens based on transcriptomic data. In particular, we found that POU2F3 protein was effective in distinguishing immunotherapy patients, i.e. the immune subtype, which was further validated by an independent data of chemo-resistant patients received 2nd -line immunotherapy. Collectively, our work systematically studied the transcriptomic and genomic heterogeneity of SCLC, and characterized an immune subtype that displayed features of sensitivity to immune checkpoint-based therapies.

Introduction

Small cell lung cancer (SCLC) is one of the deadliest cancers marked with fast proliferation and early metastasis. It is estimated that every year ~250,000 people would die from SCLC in the world¹. Over the past decades, the mainstay treatment for SCLC remains combination chemotherapy, with only minor improvements². Nonetheless, recent breakthroughs in cancer immunotherapy have created new opportunities for effective treatment of SCLC³.

As revealed by genomic analyses of recent cohort studies, SCLC is characterized by extremely prevalent TP53 and RB1 inactivating alterations and frequent SOX2 amplifications^{4,5,6}. Other alterations are less frequent and possibly related to heterogeneity among SCLC tumors. Two subtypes of SCLC have been previously recognized for decades: the classic subtype and the variant subtype^{1,7,8,9}. The classic subtype has a typical neuroendocrine (NE) phenotype, which has been partially or completely lost in the variant subtype. The lineage-specific transcription factor, ASCL1, is associated with the NE property and overexpressed in the classical subtype¹⁰. Down-regulation of ASCL1 expression in the variant subtype may contribute to its loss of NE property. Another transcription factor, NEUROD1, is also reported to be important in maintenance of the variant subtype^{10,11}. In addition, cell lines of the classical subtype grow in the non-adherent pattern, whereas the variant subtype is partially adherent *in vitro*. Recently, distinct four SCLC molecular subtypes have been proposed based on the expression of ASCL1, NEUROD1, POU2F3 and YAP1¹². But how the immune milieu and microenvironment differ between those subtypes remains unclear. On the other hand, network-based approaches have attracted increasing attentions, since it could capture the systemic nature of complex diseases from a network perspective by integrating omics data^{13,14}. Thus, the network-based subtyping with a larger dataset is expected to provide reliable subtypes of SCLC patients, in contrast to the four subtypes obtained by a traditional expression-based subtyping with a relatively small dataset¹².

To further dissect the genomic and transcriptomic signatures of SCLC, we sequenced the exomes (WES) of 19 Chinese SCLC specimens with matched control samples, and transcriptomes (RNA-seq) of the same 19 specimens with 5 normal adjacent lung samples. Through the integration with two recent SCLC cohorts^{4,6}, we have obtained the largest dataset to date (n=129 for RNA-seq, n=171 for WES), performed network-based gene module analyses and clustering, and revealed four dominant subgroups/subtypes, including one novel SCLC subtype with strong immune suppression and high tumor mutation burden (TMB), i.e. immune/immunogenic subtype, which can be effectively distinguished by the POU2F3 protein levels. Both specific transcriptomic signatures and genomic alterations are significantly enriched in each subtype, respectively.

Importantly, this immune subtype of SCLC benefits from new therapies targeting immune checkpoints, which was actually validated further by independent data of our immunotherapy patients relapsed from chemotherapies.

Results

Somatic genomic alterations in SCLC patients from eastern and western populations

We then extracted mutational signatures of these 171 SCLC patients by decomposing matrix of nucleotide substitutions according to the previous study using pan-cancer analysis¹⁸. Among the three signatures enriched in SCLC, signature 4 showed transcriptional strand bias for C>A mutations (Fig. 1c), which is known to be caused by tobacco smoking¹⁸. This signature is probably an imprint of the bulky DNA adducts generated by polycyclic hydrocarbons found in tobacco smoke and their removal by transcription-coupled NER. Signature 13 has been proved to be attributed to the APOBEC family of cytidine deaminases. As APOBEC activation constitutes part of the innate immune response to viruses and retrotransposons, this signature may represent collateral damage on the human genome from a response originally directed at retrotransposing DNA elements or exogenous viruses¹⁸.

In general, mean copy number of each gene in three datasets showed similar patterns (Fig. 1d), indicating availability of integration of CNV data. We next combined 158 samples with both mutational data and CNV data to calculate gene alteration ratios in the four pathways (Fig. 1e). Alteration ratios of each pathway in three datasets were similar based on at least one gene change in mutation or copy number. Besides the high mutation in tumors suppressor genes, such as TP53, RB1 and NOTCH family members, many oncogenic genes also showed dramatic amplification in SCLC (Fig. 1e).

Gene co-expression network modules revealed four SCLC subtypes

Although much attention has been put on the heterogeneity of SCLC, most of these studies employed cell lines and mouse models. To take advantage of transcriptomic data of clinical specimens, we collected another two SCLC cohorts, one with 81 specimens⁶ and the other with 31 specimens⁴, and integrated them with 19 Chinese SCLC specimens after batch effect removal (see Methods), with the exclusion of two outliers. We carried out weighted gene co-expression network analysis (WGCNA) to reconstruct a topological similarity network, from which 12 coherent gene modules were extracted. Based on module eigenvectors, the 129 samples formed an unsupervised clustering tree with four major clusters (Fig. 2a). Samples from different batches were evenly distributed in four clusters, suggesting the remaining batch differences are minor after correction for the batch effect.

We then carried out Kaplan-Meier survival analysis between clusters, and found patients of cluster 4 had a significantly poorer prognosis than other three subgroups, with no significant difference among cluster 1, cluster 2 and cluster 3, indicative of cluster 4 as a more malignant subtype (Fig. 2b). The previous classification of SCLC was based on molecular makers, including ASCL1, NEUROD1, POU2F3 and YAP1¹². Borromeo et al. claimed that neuroendocrine genes ASCL1 and NEUROD1 can reveal tumor heterogeneity and determine the distinct subtypes of SCLC¹¹. The transition from ASCL1-positive SCLC to NEUROD1-positive SCLC was considered as a transition from the classic subtype to a variant subtype. To align our subtyping system with previous work, we evaluated the expression of ASCL1, NEUROD1 and their inhibitor NOTCH. Samples in cluster 2 had a higher level of ASCL1 expression, whereas cluster 3 had a higher expression of NEUROD1, suggesting that cluster 2 and cluster 3 correspond to the classic and variant subtypes of SCLC, respectively (Fig. 2g). Cluster 1 showed a lower expression of both ASCL1 and NEUROD1, and higher expression of NOTCH genes¹⁹, representative of a novel non-classic, or variant, SCLC subtype. The remaining cluster 4 had significantly higher expression of CCSP than any other cluster. CCSP is a main product of club cells, which are especially abundant in the peripheral airways where they contribute to maintenance of airway integrity and repair²⁰. It is suggested that cluster 4 may have different cells of origin. Moreover, the cluster 4 had a poorer survival.

To further explore the characteristics of the subgroups, we correlated the WGCNA modules with subgroups by calculating the Pearson correlation coefficient (Supplementary Fig. 2a). Then, we selected the most relevant module for each subgroup and analyzed the subgroup-specific gene functions (Supplementary Fig. 2b). The cluster 1 specific module (ME5) was enriched in the LPS/IL-1-mediated inhibition of RXR function, IL-10 signaling, and T cell receptor signaling pathways (IPA canonical pathways, $p < 0.01$), suggestive of a strong connection of cluster 1 with immune responses. Gene set enrichment analysis (GSEA) showed the unique up/down-regulated pathways in the four SCLC subtypes (Fig. 2c). Multiple neuronal-related pathways were enriched in the neuroendocrine SCLC subtypes (cluster 2 and cluster 3) (Fig. 2d, e). Interestingly, the pathway of error-prone translesional synthesis (TLS) was specifically up-regulated in cluster 2 (Fig. 2d), which is known to underlie the mutagenic effects of numerous anticancer agents and relevant to drug resistance²¹. In contrast to cluster 1, down-regulated pathways of cluster 4 were mainly related to immunity (Fig. 2c, f). We named cluster 2 as the classical subtype, cluster 3 as variant 1 subtype, and cluster 4 as variant 2 subtype

Application of the Immune Checkpoint Blockade in the Immune subtype (cluster 1)

Cluster 1 was characterized by elevated immune cell signaling based on the analysis of the gene expression data. We then named cluster 1 as “Immune subtype”. GSEA demonstrated several immune cell receptor signaling pathways (Fig. 2c), up-regulated genes in comparison of Th1 cells versus Th17 cells (GSE11924 Gene set), and IL-12 signaling in Immune subtype versus other subtypes (Fig. 3b). Differential expression profiling revealed that the Immune subtype was enriched for both activated immune cells and immuno-stimulators (Fig. 3c). Genes mediating immunosuppressive functions (PD1, IL10RA, LAG3) were up-regulated (Fig. 3a). Notably, PD-L1 is potential biomarkers for sensitivity to immune-checkpoint-based therapies. We also assessed the helper T cell composition by comparing gene products of Th1 and Th2 (Supplementary Fig. 3). The ratios of IFNGR/IL4R and IFNGR/IL6R were lower in cluster 1, suggestive of a potential deviation from Th1 to Th2, which may further confer immune escape²². FOXP3, which is the marker gene of regulatory T cells (Tregs)²³, was also overexpressed in cluster 1 (Fig. 3a). Furthermore, estimation of abundance of immune cells using ImmuCellAI²⁴ showed that B cell, iTreg, dendritic cells, gamma delta T cells, natural killer T cells and Tfh cells were significantly higher in the Immune subtype, whereas Th1/Th2 ratio was significantly reduced (Fig. 3d). Tumor-infiltrating immune cells are considered to be primary immune signatures and strongly associated with the clinical outcomes of immunotherapies²⁵. Thus, the mechanisms by which these tumors achieved immune escape likely involve the recruitment of immune suppressive cells or the activation of immune checkpoint molecules. In addition, we found that POU2F3, ANXA1, MYC and ASCL2 were expressed extremely high in the immune subtype versus other samples (Fig. 6a, b), and thus could be used as molecular biomarkers to distinguish this immune subtype.

Significant alterations of four subtypes on WES level

To further explore specific alterations between four SCLC subtypes on WES level, we collected 115 WES mutational samples and 98 CNA samples corresponding to the four clusters at the RNA-seq level^{4, 6}. Significantly mutated genes based on prop.test with certain functions in each subtype were shown in waterfall plot (Fig. 4a). Lawson et al. identified a core set of 182 genes across mouse cancer models related to cancer-intrinsic cytotoxic T lymphocytes (CTLs) evasion²⁶, but these genes were not on our SCLC mutation list (Supplementary Data 2). On average, samples in classic subtype had more mutated genes (Fig. 4b). In contrast, variant 2 subtype had the least average mutated genes and lowest TMB (Fig. 4b, c). Interestingly, NEUROD1 was only mutated in Immune subtype, whereas NOTCH1 had higher mutation ratio in other subtypes than Immune subtype (Fig. 4a). The RASGRF1/DENND1B/ARHGEF5/ITSN2/SYTL5 as a gene set, participated in several pathways correlated with guanyl-nucleotide exchange factor activity (Fig. 4a). CD163 had 18.5% mutation frequency in Immune subtype, which was considered as the most specific marker of tumor-associated macrophages (TAMs) in tumor tissues^{27, 28}. Classical and Variant 1 subpopulations had more mutations in neuroendocrine related genes. TIE1 was observed mutated in Variant 2 subtype (Fig. 4a), which is an essential component of the angiotensin signaling system that has potential for therapeutic targeting in disease²⁹. The endothelial angiotensin/Tie (ANG/Tie) system regulates

angiogenesis during development and tumor growth, contributes to capillary-to-venous remodeling in inflammation, and maintains vascular integrity³⁰.

In general, we found differential functional features of four subtypes at CNA level. Circos heatmap depicted median gene copy number in each subtype. Immune, Classic, Variant 1 and Variant 2 subtype were inwards apart (Fig. 4d). Significantly amplified genes with high alteration ratio (>20%) in each subtype were used for pathway enrichment analysis (Supplementary Data 3) (Fig. 4e). It is showed that response to oxidative stress was observed in Classic subtype. Besides the well-recognized effect of ROS in mutation and cancer cell growth, growing evidence suggests its involvement in drug resistance³¹. Of note, significant amplified genes in Immune subtype were enriched in several immune-related pathways (Fig. 4e). The PI3K events in ERBB4 signaling were observed in Variant 2 subgroup (Fig. 4e). Moreover, in addition to the transcriptomic signatures, specific genomic alterations were also significantly enriched in each subtype, particular in the immune subtype (Fig. 4).

Identification of four subtypes of SCLC using a machine learning model

We next build a random forest model to classify four subtypes based on transcriptomic data. Random forest algorithm is a kind of classical machine learning algorithm based on statistical learning theory. It combines bootstrap resampling method with decision tree algorithm, and can evaluate the importance of features while building models. We used 460 differential expressed genes in 80% samples to construct Pearson correlation coefficient network which were clustered into 20 modules (Fig. 5b) (Supplementary Data 4). Mean gene expression value of each module was used as input features for the training (Fig. 5a, Methods). Fourteen features were chosen as the final classifier since the error of cross validation was relatively stable (Fig. 5c). High positive margin indicates correct classification of four SCLC subtypes (Fig. 5d). The prediction results with cross validation showed that this model achieved 99% accuracy to distinguish testing samples.

We further tested our model on an independent data set consisted of 79 SCLC patients with their gene expression data³². We found that 6 patients were predicted as Immune subtype, 15 patients were predicted as classic subtype, 57 patients were predicted as variant 1 subtype, and 1 patient was predicted as variant 2 subtype. The expressions of the marker genes were consistent with each predicted subtype (Fig. 5e), indicative of the overall effectiveness of our classification model.

Validation of Immune subtype using POU2F3 antibody in the immunotherapy patient cohort

In order to find the unique marker for Immune subtype, we performed the analysis of the differential expression genes between Immune subtype and other groups. POU2F3 stood out as the most significantly up-regulated gene in the Immune subtype (Fig. 6a, b). To identify the Immune subtype of SCLC clinically, we collected 28 SCLC specimens and performed immunohistochemical staining of the candidate. All the 28 SCLC patients were resistant to chemotherapy and given immunotherapy or chemo-immunotherapy for more than two cycles (every 21 days for one cycle). Therefore, we can evaluate properly the effectiveness of immunotherapy.

The immunohistochemical staining of POU2F3 showed that the patients with high expression level (score \geq 100) are sensitive to immunotherapy (Fig. 6c; Supplementary Data 5). According to Marina et al.³³, ASCL1, NEUROD1, YAP1, POU2F3 expression have different distribution in 174 SCLC patient samples by immunohistochemistry, and POU2F3 was expressed in 7% of SCLC. It appeared that POU2F3 was expressed in 43% SCLC chemotherapy-resistant patients. Interestingly, 100% (12/12) of the patients with POU2F3-high tumors had an significantly improved response following immunotherapy (Fig. 6d), and p value was 0.013 when compared with POU2F3-low tumors. High F1 score (75%) and AUC (0.872) were achieved (Fig. 6e), indicating the good ability of this classification. In addition, 2 out of the 12 cases with POU2F3-high level showed dramatical regression of the intrapulmonary tumors (Fig. 6f and Supplementary Fig. 4a. These results support that the Immune subtype identified here is indeed sensitive to immunotherapy and the POU2F3 expression could serve as an biomarker for effective immunotherapy.

Discussion

To explore the tumor heterogeneity and potential clinical implications, we performed an integrative transcriptomic classification by combining data from three independent cohorts, including 19 Chinese SCLC specimens and two previous cohorts from both Western and Eastern populations. Here, we identified four dominant subtypes with a robust clustering approach based on gene modules, two of which correspond well with the classic (ASCL1-high, NEUROD1-low) and variant (ASCL1-low, NEUROD1-high) subtypes. More importantly, two newly identified subgroups also have been identified. One subtype expressed ASCL1-high and CCSP-high. The remaining subtype represents a novel variant subtype characterized with lower expression of both ASCL1 and NEUROD1, higher expression of NOTCH1/2, showing increased immunogenic and consequent immunosuppressive features, which is defined as an immune subtype.

Although four subtypes have also been proposed in a recent study using mixed primary tumors and cell lines¹², they were obtained by a traditional expression-based subtyping with a relative small dataset. Here, by adopting a network-based subtyping fully based on primary tumors and further by integrating previous two datasets with also our data into a larger dataset, we have identified a partially different four-subtype clustering in a reliable manner. Three of the subtypes discovered are largely shared between that study and our study, including the immune subtype defined by us (named “SCLC-P” in that study). Traditionally, SCLC lack alternative therapeutic choices after rapid resistance to chemotherapy, resulting in poor prognosis and high mortality. Recently, several clinical trials have been actively evaluating the efficacy of checkpoint inhibitors targeting PD-1/PD-L1 or CTLA-4. For instance, CheckMate032 (ClinicalTrials.gov ID NCT01928394) was designed as a multicentre, open-label, phase 1/2 trial for limited-stage (LS) or extensive-stage (ES) SCLC, an objective response rate (ORR) was achieved in ten (~10%) of 98 patients receiving nivolumab 3 mg/kg, one (33%) of three patients receiving nivolumab 1 mg/kg plus ipilimumab 1 mg/kg, 14 (23%) of 61 receiving nivolumab 1 mg/kg plus ipilimumab 3 mg/kg, and ten (19%) of 54 receiving nivolumab 3 mg/kg plus ipilimumab 1 mg/kg³. Another clinical trial with ES SCLC patients showed that only phased ipilimumab, but not concurrent ipilimumab, improved immune-related progression-free survival (ir-PFS) versus control. No improvement in progression free survival (PFS) or overall survival (OS) occurred³⁴. The rudimentary results of these studies suggested further investigation of nivolumab and ipilimumab in SCLC patients. In this study, we identified a specific type of SCLC with significant features of immune suppression, i.e. immune subtype. Besides transcriptomic signatures, there are also specific genomic alterations significantly enriched in this immune subtype. We emphasized the immunogenic feature and potential clinical value of this immune subtype. Importantly, we found that POU2F3 is highly expressed in this immune subtype and the validation in SCLC specimens showed that POU2F3 protein is effective in distinguishing immunotherapy patients. We also find that 13 SCLC-POU2F3 primary tumor samples identified in previous study¹² are all included in our immune subtype. Consistently, the immune-related pathway was also up-regulated in these 13 samples versus other 97 samples (Supplementary Fig. 4e).

Recently, Tlemsani et al. built an integrated platform to examine drug sensitivities across 116 SCLC cell lines and found that immune pathways were selectively expressed in SCLC-YAP³⁵, while our results based on primary SCLC samples indicated that SCLC-POU2F3 tumors are more sensitive to immunotherapy. Besides, the only two primary tumor samples in the SCLC-YAP are not involved in immune subtype according to our classification strategy. Although our work identified an immunogenic subtype that displayed features of sensitivity to immune checkpoint-based therapies, further efforts are required to validate the effectiveness of immunotherapy in this Immune subtype.

Materials And Methods

Classification of Human SCLC specimens

The 19 SCLC surgical specimens with paired pathological normal lungs were collected from July 2007 to June 2011. This study was approved by the institutional review board of Fudan University Shanghai Cancer Hospital, Shanghai, China. All participants gave written informed consent. Fresh surgical specimens were snap-frozen and stored in liquid nitrogen upon

resection until use. The pathology of each tumor sample was determined by pathologists. All these specimens were with a minimum of 70% of tumor cellularity. DNA was extracted from fresh-frozen tissues using the Genra Puregene DNA extraction kit (Qiagen) following the protocol of the manufacturer. Total RNA was extracted from tissues with Trizol Reagent (Invitrogen) according to the manufacturer's instruction. The whole-exome sequencing was done in all of these specimens and their matched normal tissues, and RNA-seq was carried out in 19 tumor specimens and 5 normal specimens. All sequencing reactions were performed on an IlluminaHiSeq 2500 platform (Berry Genomics Corporation, China).

In addition, we downloaded other two published SCLC datasets with RNA-seq, gene mutation and CNA information. In George's material, we obtained 81 samples for RNA-seq, 110 samples for gene mutational data and 109 samples for CNA data⁶. In Rudin's material, we obtained 31 samples for RNA-seq, 42 samples for gene mutational data and 30 samples for CNA data⁴.

Processing of whole-exome sequencing (WES) data

Raw sequencing data was aligned to the human genome (GRCh 38/hg38) using Burrows-Wheeler Aligner (BWA mem, version: 0.7.15). Then, the data was processed following the GATK best practices workflow (version: 3.5) to process the data. We marked duplicates with Picard (version: 2.0.1), performed local realignment around the indels (GATK RealignerTargetCreator) and recalibrated the base quality score (GATK BaseRecalibrator). Mutation calling was carried out by Mutect2 (DbSNP version 144 and COSMIC version v76). ANNOVAR was applied to the annotation of somatic mutations. Mutational landscape of 171 samples from three datasets was completed by R package 'maftools'.

Genome-wide copy number analysis was performed using CNVKIT (version: 0.9.6). CNA heatmap was completed by R package 'RCircos'. Mean value was used for the three datasets. Median value was used for each subgroup.

Processing of RNA-seq data

Raw sequencing data was aligned to the human genome (GRCh 38/hg38) using STAR (version: 2.5.2b) 2-pass mode. Read counts were calculated by HTSeq (version: 0.6.0). FPKM was calculated by Cuffdiff (version: 2.2.1) for co-expression network construction.

Weighted gene co-expression network analysis (WGCNA)

The weighted gene co-expression network analysis was carried out by R package 'WGCNA'. The expression data for network construction from 129 samples, including 19 patients, and another 81 patients⁶ and 31 patient⁴ collected from public databases. We combined three datasets after removing the batch effect by ComBat from 'sva' package in R. Genes with a low signal value (FPKM<0.5 in more than half samples) or a low coefficient of variation (CV<0.7) were filtered out. We divided the network into modules according to the correlations between the genes. The minimum size of the module was 50. If the distance between two modules was less than 0.3, we merged the modules. The co-expression network was finally partitioned into 12 modules. We clustered the samples according to the module eigengenes by hierarchical clustering analysis (Euclidean distance, ward.D2 linkage). The 129 samples were finally clustered into 4 subgroups. The correlations between modules and subgroups were measured by Pearson correlation. Survival analysis was finished with R package 'survminer'.

Gene set enrichment analysis (GSEA)

GSEA was performed using the GSEA software (v4.0) available at <http://software.broadinstitute.org/gsea/downloads.jsp> with the option pre-ranked and default parameters^{36,37}. Reactome gene sets, KEGG gene sets, Gene Ontology gene sets, BioCarta gene sets, immunologic signatures, oncogenic signatures and hallmark gene sets were used in the molecular signature analysis (MSigDB available at <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). We used NES and NOM p-value to present up-regulated and down-regulated pathway in each subtype.

Classification Model

We used random forest model to select features to identify the subgroup of SCLC patients. Considering the fluctuation of single gene expression level, we chose the mean expression value of a group of genes instead of single genes as the candidate features. We firstly computed differential expression genes in 80% randomly selected samples between any pairwise subgroups. Each top 100 differential expression genes were chosen with the following parameters: $p\text{-value} < 0.05$, $\log_2(\text{fold-change}) > 1$, $\log_2(\text{fold-change}) < -1$, sorted by p values. In total 460 unique genes were finally used to build Pearson correlation coefficient network. We used (1-correlation) as distance for hierarchical clustering (ward.D2 linkage) and divided the 460 genes into 20 modules. Scaling of mean gene expression value of those modules were used as input to build decision tree. We used 10-fold cross-validation and repeated 5 times. Fourteen features were chosen for the final classifier model. Margin of prediction plot was completed by R package 'randomForest'. We also used another independent dataset consisting of 79 SCLC patients³² to test the model.

Validation of POU2F3 in SCLC specimens

For the validation of POU2F3 in the determination of the immune response in SCLC patients, we collected 28 SCLC patients from January 2016 to January 2020, in the Chinese PLA General Hospital and Beijing Shijitan Hospital of Capital Medical University. All the patients had disease progression after at least one previous cycle of chemotherapy and then received immunotherapy or chemo-immunotherapy for no less than two cycles (every 21 days for one cycle). Assessments of tumor response were performed by the Response Evaluation Criteria in Solid Tumors (RECIST), version 1.1 (RECIST 1.1). The study was conducted in accordance with ethical guidelines of U.S. Common Rule and was approved by the Ethics Committee of Beijing Shijitan Hospital of Capital Medical University. Written informed consent was obtained from all patients.

For IHC staining, slides were deparaffinized in xylene and ethanol, and rehydrated in water. Antigen retrieval was performed by microwave in sodium citrate (pH 6.0) buffer. Slides were quenched in hydrogen peroxide (3%) to block endogenous peroxidase activity and washed in PBS buffer. Then incubated with primary antibodies at 4°C overnight. The antibodies used were BRG1 (ab110641, dilution 1:100). Immunohistochemistry kit (Zhongshanjinqiao) was used for the IHC staining experiment. The IHC staining was blindly scored as high or low according to staining density and subjected to analyses for clinical relevance described previously³⁸.

Declarations

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

L.C., H.J. and Z.F. conceived the idea and designed the experiments. Y.C., Z.F., Y.T., C.G. and Y.J. performed all experiments and analyzed the data. Y.X., J.G., M.X., X.Z., X.M. and L.H. provided technical assistance and contributed intellectually. H.C., S.L., and X.X. kindly provided human SCLC specimens. Y.C., Z.F., Y.T., performed the bioinformatics analyses. Y.C., Y.T. and Z.F. wrote the original draft of the manuscript. L.C., H.J., X.X., Y.J., L.H. edited the manuscript.

Acknowledgments

This work was supported by the National Basic Research Program of China (No. 2017YFA0505500 to L.C and H.J., No. 2017YFA0505501 to H.J., 2020YFA0803300 to H.J.), the Strategic Priority Research Program of the Chinese Academy of Sciences (grants. XDB19020201 to H.J., XDB38040400 to L.C.), the National Natural Science Foundation of China: 31930022 (L.C.), 31771476 (L.C.), 82030083 (H.J.), 31621003 (H.J.), 81872312 (H.J.), 82011540007 (H.J.), 81871875

(L.H.), 81802279 (H.H.), 81902326 (X.W.), Basic Frontier Scientific Research Program of Chinese Academy of Science (grant ZDBS-LY-SM006 to H.J.), International Cooperation Project of Chinese Academy of Sciences (grant 153D31KYSB20190035 to H.J.), and Science and Technology Commission of Shanghai Municipality 15XD1504000 (H.J.), 2017SHZDZX01 (L.C.).

Data Availability

The datasets used for this study are all accessible either through a reasonable contact of the corresponding authors or elsewhere from original deposits following descriptions in the Materials and Methods.

References

1. Gazdar AF, Bunn PA, Minna JD. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nat Rev Cancer* **17**, 725–737 (2017).
2. Govindan R, *et al.* Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. *J Clin Oncol* **24**, 4539–4544 (2006).
3. Antonia SJ, *et al.* Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): a multicentre, open-label, phase 1/2 trial. *The Lancet Oncology* **17**, 883–895 (2016).
4. Rudin CM, *et al.* Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat Genet* **44**, 1111–1116 (2012).
5. Peifer M, *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet* **44**, 1104–1110 (2012).
6. George J, *et al.* Comprehensive genomic profiles of small cell lung cancer. *Nature* **524**, 47–53 (2015).
7. Poirier JT, Dobromilskaya I, Moriarty WF, Peacock CD, Hann CL, Rudin CM. Selective tropism of Seneca Valley virus for variant subtype small cell lung cancer. *J Natl Cancer Inst* **105**, 1059–1065 (2013).
8. Pedersen N, *et al.* Transcriptional gene expression profiling of small cell lung cancer cells. *Cancer Res* **63**, 1943–1953 (2003).
9. Carney DN, *et al.* Establishment and identification of small cell lung cancer cell lines having classic and variant features. *Cancer Res* **45**, 2913–2923 (1985).
10. Mollaoglu G, *et al.* MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer Cell* **31**, 270–285 (2017).
11. Borromeo MD, *et al.* ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. *Cell Rep* **16**, 1259–1272 (2016).
12. Rudin CM, *et al.* Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat Rev Cancer* **19**, 289–297 (2019).
13. Chen L, Wang R-S, Zhang X-S. *Biomolecular networks: methods and applications in systems biology*. John Wiley & Sons (2009).
14. Chen L, Wang R, Li C, Aihara K. *Modelling Biomolecular Networks in Cells: Structures and Dynamics*. 2010. NY: Springer-Verlag London Limited[Google Scholar].
15. Ding L, *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
16. Chen H, Chong W, Wu Q, Yao Y, Mao M, Wang X. Association of LRP1B Mutation With Tumor Mutation Burden and Outcomes in Melanoma and Non-small Cell Lung Cancer Patients Treated With Immune Check-Point Blockades. *Front Immunol* **10**, 1113 (2019).
17. Cowin PA, *et al.* LRP1B deletion in high-grade serous ovarian cancers is associated with acquired chemotherapy resistance to liposomal doxorubicin. *Cancer Res* **72**, 4060–4073 (2012).
18. Alexandrov LB, *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

19. Lim JS, *et al.* Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. *Nature* **545**, 360–364 (2017).
20. Wong AP, Keating A, Waddell TK. Airway regeneration: the role of the Clara cell secretory protein and the cells that express it. *Cytotherapy* **11**, 676–687 (2009).
21. Xie K, Doles J, Hemann MT, Walker GC. Error-prone translesion synthesis mediates acquired chemoresistance. *Proc Natl Acad Sci U S A* **107**, 20792–20797 (2010).
22. Yamazaki K, Yano T, Kameyama T, Suemitsu R, Yoshino I, Sugio K. Clinical significance of serum TH1/TH2 cytokines in patients with pulmonary adenocarcinoma. *Surgery* **131**, S236-241 (2002).
23. Sakaguchi S, Miyara M, Costantino CM, Hafler DA. FOXP3 + regulatory T cells in the human immune system. *Nat Rev Immunol* **10**, 490–500 (2010).
24. Miao YR, *et al.* ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. *Adv Sci (Weinh)* **7**, 1902880 (2020).
25. Pandya PH, Murray ME, Pollok KE, Renbarger JL. The Immune System in Cancer Pathogenesis: Potential Therapeutic Approaches. *J Immunol Res* 2016, 4273943 (2016).
26. Lawson KA, *et al.* Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature* **586**, 120–126 (2020).
27. Franco R, Fernández-Suárez D. Alternatively activated microglia and macrophages in the central nervous system. *Prog Neurobiol* **131**, 65–86 (2015).
28. Colin S, Chinetti-Gbaguidi G, Staels B. Macrophage phenotypes in atherosclerosis. *Immunol Rev* **262**, 153–166 (2014).
29. Korhonen EA, *et al.* Tie1 controls angiopoietin function in vascular remodeling and inflammation. *J Clin Invest* **126**, 3495–3510 (2016).
30. Augustin HG, Koh GY, Thurston G, Alitalo K. Control of vascular morphogenesis and homeostasis through the angiopoietin-Tie system. *Nat Rev Mol Cell Bio* **10**, 165–177 (2009).
31. Trachootham D, Alexandre J, Huang P. Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach? *Nat Rev Drug Discov* **8**, 579–591 (2009).
32. Jiang L, *et al.* Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. *PLoS Genet* **12**, e1005895 (2016).
33. Baine MK, *et al.* SCLC Subtypes Defined by ASCL1, NEUROD1, POU2F3, and YAP1: A Comprehensive Immunohistochemical and Histopathologic Characterization. *J Thorac Oncol*, (2020).
34. Reck M, *et al.* Ipilimumab in combination with paclitaxel and carboplatin as first-line therapy in extensive-disease-small-cell lung cancer: results from a randomized, double-blind, multicenter phase 2 trial. *Ann Oncol* **24**, 75–83 (2013).
35. Tlemsani C, *et al.* SCLC-CellMiner: A Resource for Small Cell Lung Cancer Cell Line Genomics and Pharmacology Based on Genomic Signatures. *Cell Rep* **33**, 108296 (2020).
36. Subramanian A, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).
37. Mootha VK, *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267–273 (2003).
38. Li F, *et al.* LKB1 Inactivation Elicits a Redox Imbalance to Modulate Non-small Cell Lung Cancer Plasticity and Therapeutic Response. *Cancer Cell* **27**, 698–711 (2015).

Figures

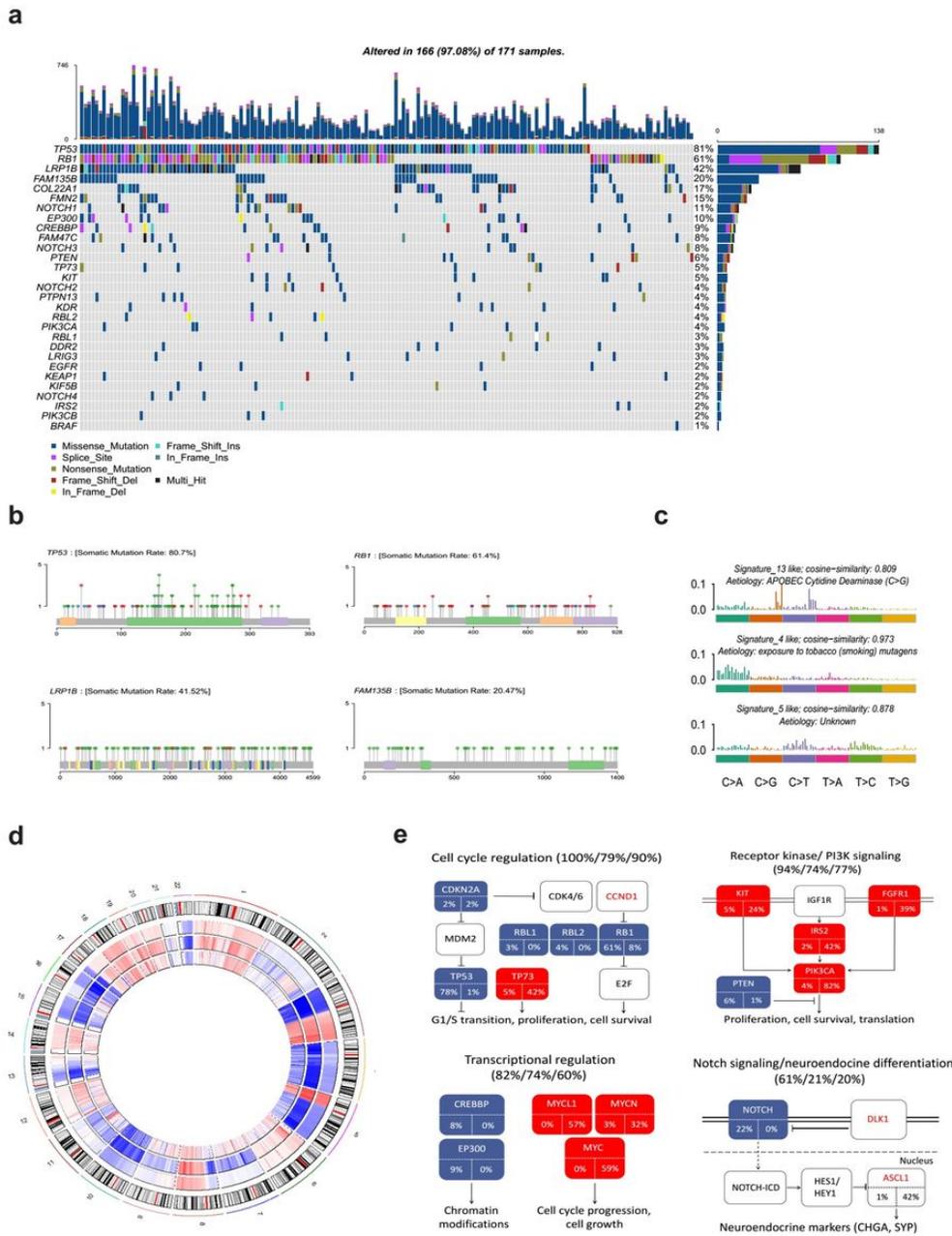


Figure 1

The Genomic landscape of small cell lung cancer merging three datasets. (A) Known cancer-related genes that were mutated in at least 2% of the cases in 171 samples. (B) Three mutational signatures of specific pattern of nucleotide substitutions in SCLC. (C) Schematic representation of 4 candidate genes with significant clustering of mutations in respective protein domains. Somatic mutations and genomic translocations are mapped to the respective protein regions. Hotspot mutation types are highlighted in different color. (D) CNV heatmap using mean copy number in each dataset. From outside to inside, S dataset, t dataset and our dataset. (E) Four signaling pathways recurrently affected in SCLC. 158 samples with both CNV and mutation data are used to calculate alteration proportions (S/X/t dataset). Mutation ratios of each pathway in three datasets are shown in brackets. Red and blue boxes denote genes with activating and inactivating alterations, respectively. Gene mutation ratios are on the bottom left corner of each colored box. Gene amplification ratios are on the right bottom right corner of red boxes, gene loss ratios are on blue boxes. Genes found expressed at high levels are shown in red font.

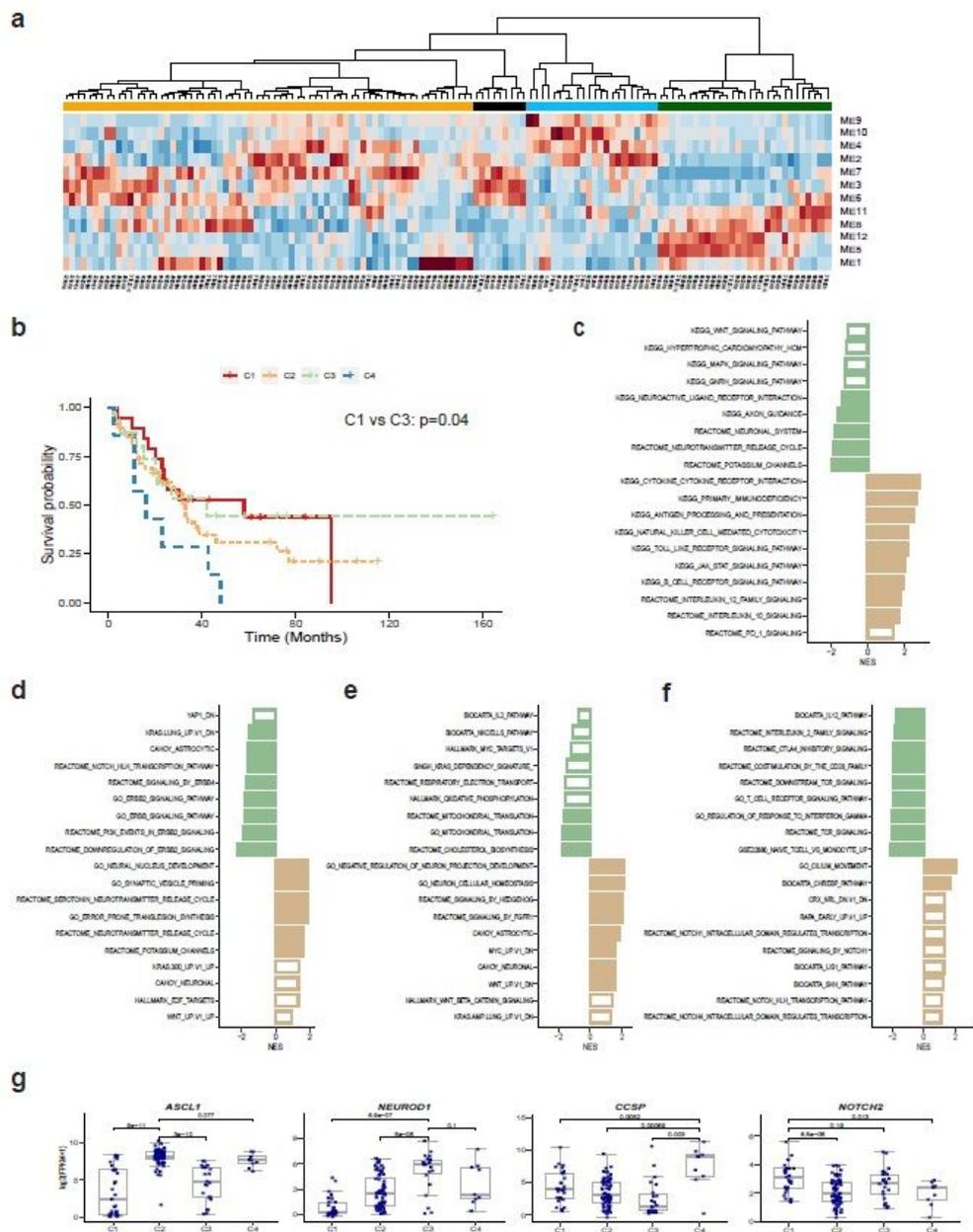


Figure 2

Subtyping 129 SCLC mRNA samples with weighted gene co-expression network. (A) Heatmap of eigenvectors of twelve network modules. The clustering method is ward.D2. (B) Survival analysis by four clusters. (C) Top up-regulating and down-regulated pathways enriched in C1 subtype. Significantly enriched pathways with GSEA p-value < 0.05 are highlighted in orange (higher in C1 subtype) or green (lower in C1 subtype). (D) Same as in (C) but for the C2 subtype. (E) Same as in (C) but for the C3 subtype. (F) Same as in (C) but for the C4 subtype. (G) Boxplots of 4 marker genes in each cluster.

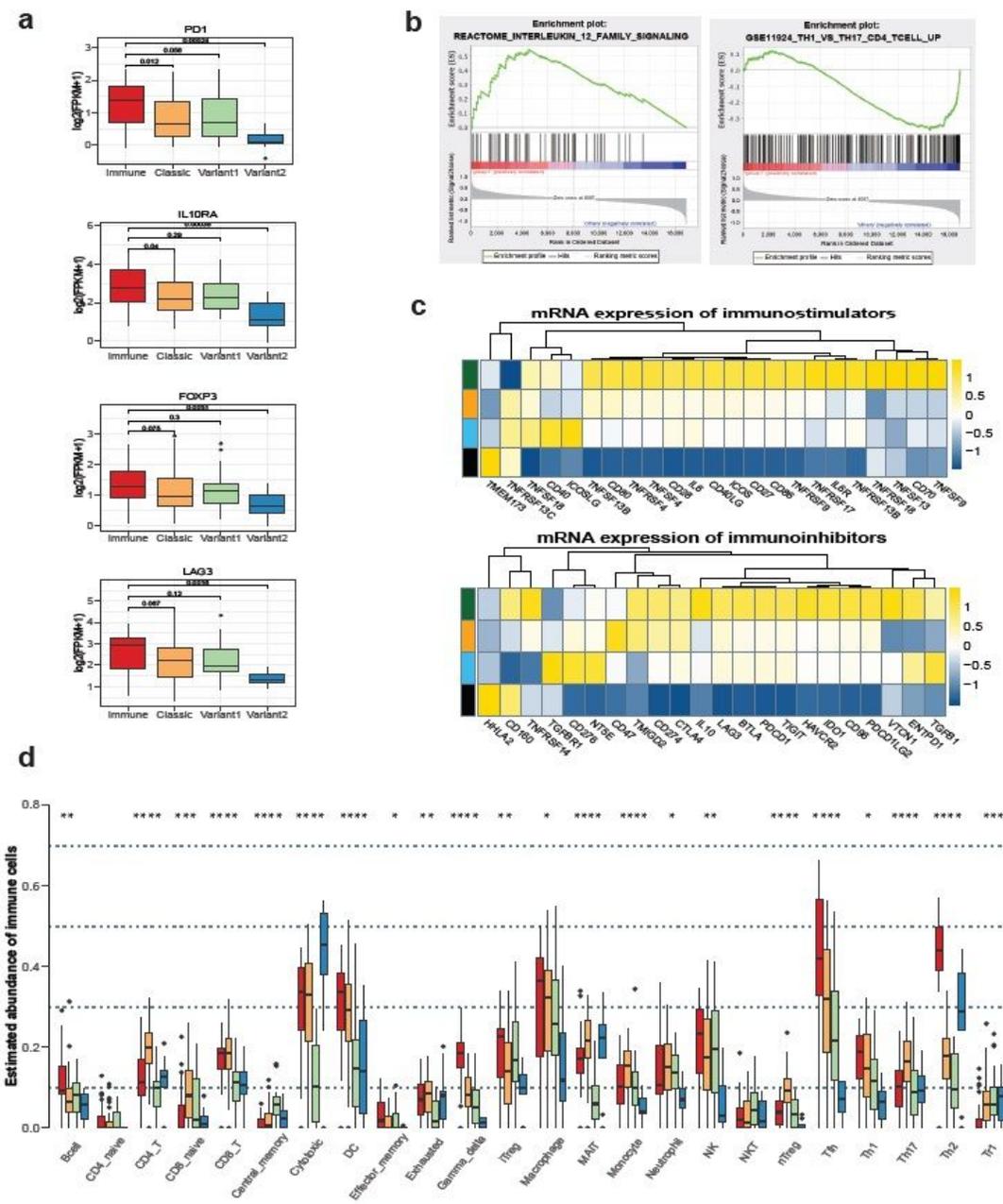


Figure 3

Immune signatures in four subtypes. (A) Boxplots of three immunoinhibitors (PD1, IL10RA, LAG3) and Tregs specific marker gene (FOXP3). C1: immune subtype; C2: classic subtype; C3: variant 1; C4: variant 2. (B) Representative gene set enrichment analysis plot showing Interleukin 12 family signaling (left) and Th1 vs Th17 CD4 T cell up (right) in the Immune subtype versus the other subtypes. (C) The mean expression of immunostimulatory molecules in four mRNA subtypes (up) and the mean expression of immune checkpoint genes (bottom) in four mRNA subtypes. Annotated color represent four subtypes (green = Immune, orange = Classic, blue = Variant1, black = Variant2). (D) Relative abundance of immune cells in the four subtypes (calculated using the ImmuCellAI algorithm). The Kruskal.test was conducted and annotated as follows: ***p < 0.001, **p < 0.01, *p < 0.05.

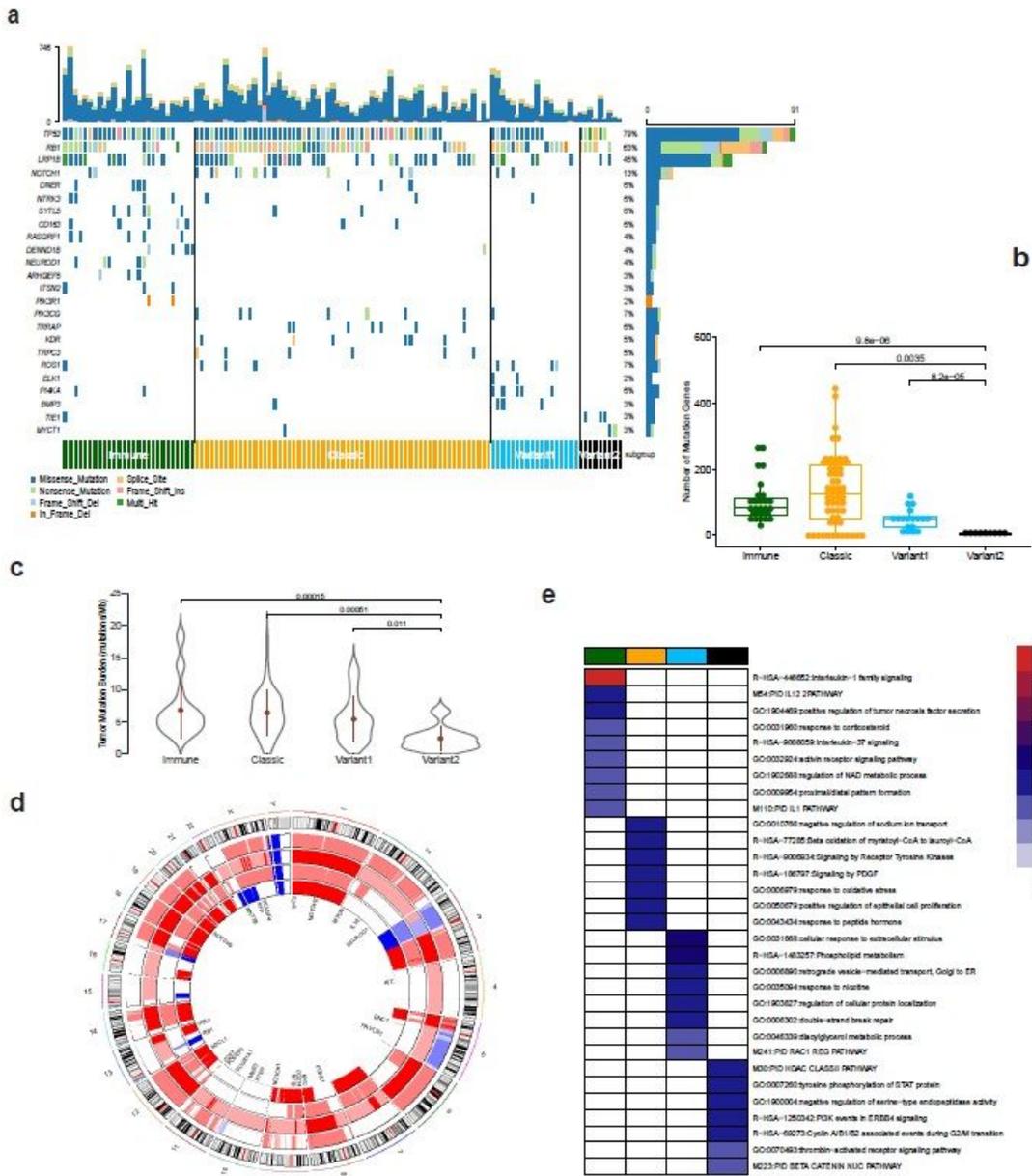


Figure 4

Genomic mutation and CNA in four subtypes. (A) Tumour samples are arranged from left to right. Genes with high mutation rate or significantly mutated in each subtype are shown in waterfall plot. Annotated colors represent 4 subtypes (green = Immune, orange = Classic, blue = Variant1, black = Variant2). (B) Number of mutant genes in each sample of subtypes. (C) Tumor mutational burden in four subtypes. (D) CNA heatmap using median copy number in each subtype. From outside to inside, Immune, Classic, Variant1, Variant2. Some significant alteration genes in each subtype are highlighted inner ring. (E) Enriched pathways using significant amplification genes (freq>20%, p value < 0.01) in each subtype (p-value < 0.001). Heatmap is colored by “ $-\log_{10}(P)$ ”. Clearly there are specific genomic alterations in each subtype, in particular, in immune subtype.

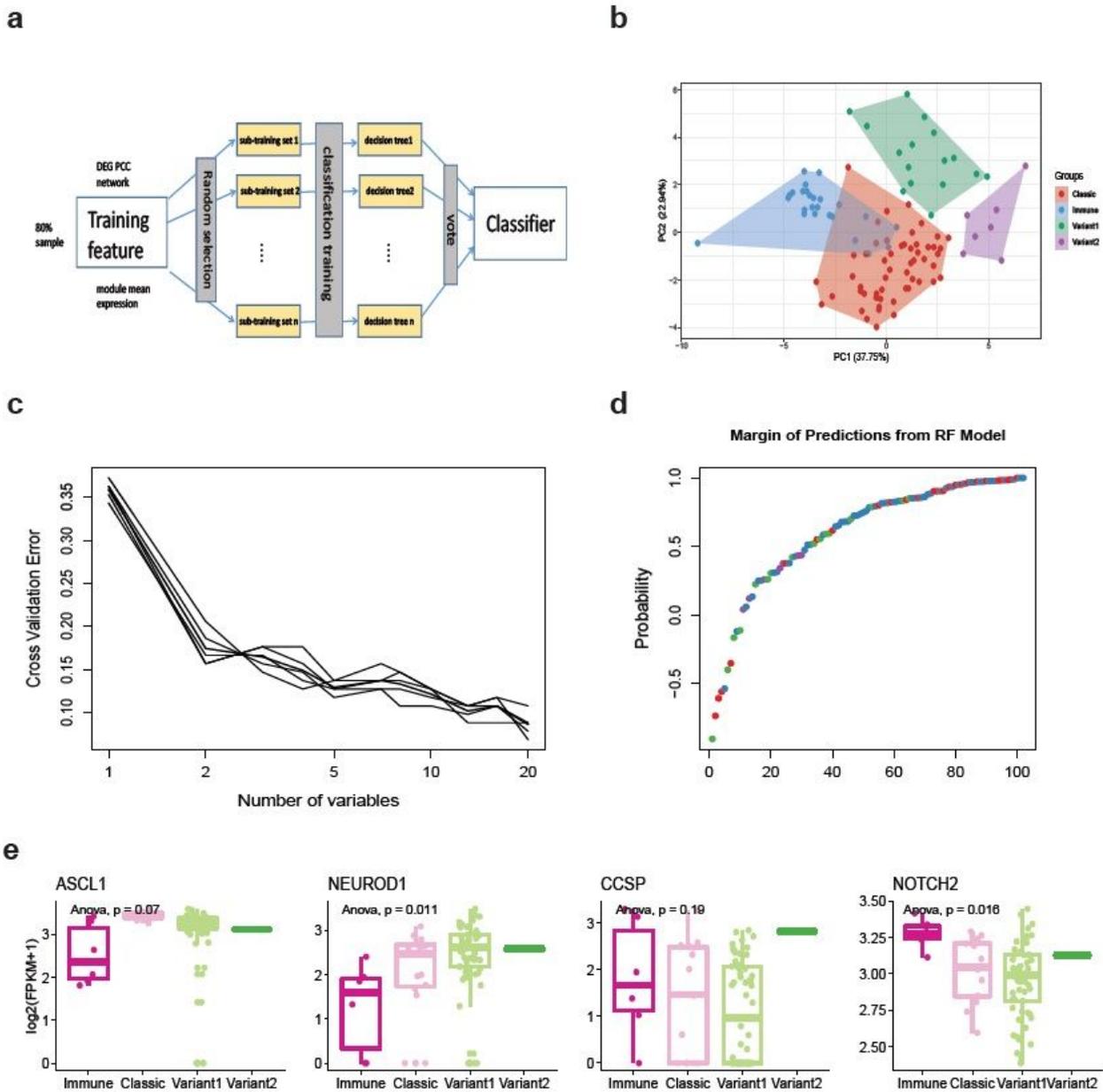


Figure 5

Classifier for distinguishing four subtypes by gene expression data. (A) Schematic representation of our random forest model. (B) PCA plot of 80% training samples using 461 differential expression genes. (C) 10 fold Cross validation with 5 repeats to select minimal feature. (D) Margin of predictions of our model. Samples in four subtypes are dotted in different colors. (E) 79 independent samples are used to distinguish subtypes. Boxplots of 4 marker genes are shown to examine classification prediction.

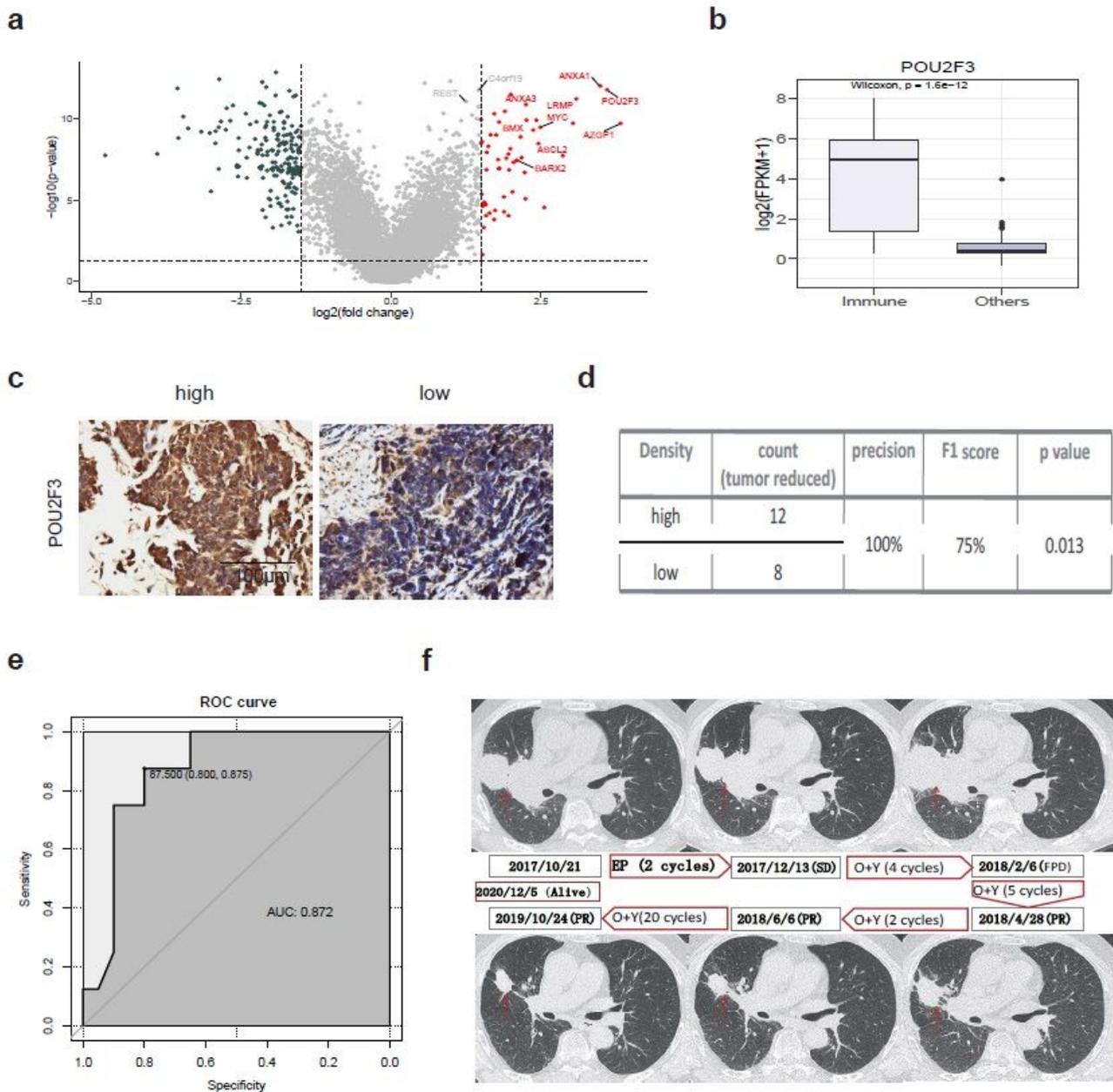


Figure 6

Further validation of Immune subtype according to POU2F3 of immunotherapy patients. (A) Volcano plot of differential expression genes between Immune subtype and others. (B) Expression of POU2F3 in immune subtype versus others. (C) Immunohistochemical staining of tumor sections using POU2F3 antibodies. Sections with high/low expression of POU2F3 are shown separately. (D) Statistics of immunohistochemical staining results, in which “high” indicates Immune subtype. (E) ROC curve of POU2F3 protein score in treatment diagnosis. (F) CT image of case 1 at different time points during the whole treatment. In this process, 2 cycles chemotherapy were used first, and the tumor remained SD. Then, after 4 cycles of immunotherapy, tumor appeared pseudo progression. We continued to give drugs for 5 cycles, tumor showed PR. After that, tumor became significantly reduced under subsequent immunotherapy.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.pdf](#)
- [FigureS2.pdf](#)
- [FigureS3.pdf](#)
- [FigureS4.pdf](#)
- [TableS1.txt](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)