Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Single-cell analysis reveals multicellular immune coordination and drug targets in driver-mutated NSCLC

Congxue Hu

hucx1996@hrbmu.edu.cn

Harbin Medical University

Kuan Yang
Jiasi Wang
Wenqi Jiang
Hang Zhao
Depeng Mu
Xiao Peng
Yiming Yan
Xing Gao
Jing Bai

College of Bioinformatics Science and Technology

Xia Li

Harbin Medical University Institute of Biological Information Science and Technology
https://orcid.org/0000-0002-9794-2648

Yunpeng Zhang

Harbin Medical University https://orcid.org/0000-0002-3709-3656

Article

Keywords: Driver mutated, Immune coordination, Malignant regulatory network, Prognostic biomarkers, Single-cell transcriptomics

Posted Date: February 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8727447/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Non-small cell lung cancer (NSCLC) harboring driver mutations rely on specific oncogenic signaling pathways to promote tumor progression and remodel immune responses, yet the role of multicellular immune coordination within the tumor immune microenvironment (TIME) remains unclear. We performed single-cell transcriptomic profiling of ~ 200,000 cells from 45 untreated NSCLC patients with seven driver mutations and identified five functionally distinct multicellular modules (CM1–5). Strikingly, the invasive CM2 and CM5 were strongly associated with poor prognosis. CM2 was enriched in EGFR- and MET-driven brain metastases and featured dense astrocytic–myofibroblast interactions, with pro-invasive signals (SPP1, PTN, PSAP) and metabolic reprogramming fostering a permissive metastatic TIME. CM5 was prevalent in ROS1-, KRAS-, and EGFR-mutant tumors and comprised heterogeneous myeloid and endothelial populations enriched for immunosuppressive and pro-angiogenic signals, including MIF, GALECTIN, and RETN. We further developed a driver mutation–specific prognostic signature (DMSP.sig) based on receptor–ligand interactions and core transcriptional regulators, which robustly stratified patient survival and identified candidate therapeutic agents. Collectively, our study reveals driver-shaped TIME subtypes with implications for risk assessment and precision immunotherapy in high-risk NSCLC.

Introduction

Non-small cell lung cancer (NSCLC) is the most prevalent histological subtype of lung cancer, accounting for approximately 85% of all cases [1]. Its pronounced biological heterogeneity is manifested across multiple dimensions, including genetic alterations, cellular origins, tumor microenvironment, and immune responses [2]. Advances in high-throughput sequencing technologies, particularly next-generation sequencing (NGS), have enabled comprehensive molecular characterization of NSCLC, leading to the identification of numerous oncogenic alterations, several of which have become critical targets for therapeutic development [3]. These discoveries have substantially advanced the field of precision oncology in NSCLC, fostering a novel classification paradigm that emphasizes molecularly defined subtypes to guide the rational implementation of targeted therapies [4].

Multiple oncogenic driver mutations have been identified in NSCLC patients, including alterations in epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), and the receptor tyrosine kinase ROS1. Therapeutic agents such as monoclonal antibodies and small-molecule inhibitors targeting key signaling pathways that regulate tumor growth and progression have shown promise in improving patient survival and quality of life [1]. Notably, the distribution of driver mutations exhibits histological specificity: mutations in EGFR, KRAS, BRAF, ALK, and ROS1 are predominantly observed in lung adenocarcinoma, whereas alterations in FGFR1, DDR2, and components of the PI3K pathway are more commonly detected in squamous cell carcinoma [2]. Furthermore, NGS-based studies have revealed that individual tumors can harbor concurrent double, triple, or even multiple co-occurring mutations, which may interact to influence therapeutic response and drug sensitivity [5]. Thus, a deeper understanding of

the molecular heterogeneity of NSCLC is essential for uncovering novel pathogenic mechanisms and optimizing individualized treatment strategies.

Certain oncogenic events, such as mutations in KRAS, TP53, and MET, have been demonstrated to possess prognostic relevance and potential predictive value for neoadjuvant therapy [6–8]. In addition, differences in driver mutations can profoundly reshape the composition of the tumor immune microenvironment (TIME). Preclinical studies have shown that NSCLC driven by EGFR or KRAS mutations displays distinct immune cell infiltration patterns in terms of both cellular composition and phenotypic characteristics.

In recent years, the advent of single-cell RNA sequencing (scRNA-seq) has enabled high-resolution dissection of cellular heterogeneity in NSCLC. This technology has facilitated the construction of comprehensive cellular atlases within NSCLC tumor tissues and has highlighted the critical role of TIME in mediating responses to therapeutic interventions [9]. Specifically, immune components within TIME—including exhausted $CD8^+$ T cells, M2-like tumor-associated macrophages (TAMs), and regulatory B cells—play key roles in tumor progression and responsiveness to immunotherapy [10], and have emerged as promising biomarkers for predicting immunotherapeutic outcomes.

In this study, we integrated single-cell RNA sequencing data from 45 treatment-naïve patients with driver-mutated non-small cell lung cancer (NSCLC), classified into 10 subtypes based on primary versus metastatic status and driver mutations. Using these data, we constructed a TIME atlas and identified functionally specialized immune coordination programs and driver mutation-specific prognostic (DMSP) signatures to facilitate patient risk stratification, immune profiling, and prognostic prediction. These findings provide new insights into the molecular and immune mechanisms of NSCLC and offer a framework for precision therapy development.

Results

A single-cell atlas in NSCLC harboring driver mutations

To comprehensively dissect the heterogeneity of TIME in NSCLC harboring distinct oncogenic driver mutations, we systematically integrated scRNA-seq data from 45 treatment-naive NSCLC patients (Method). These tumors harbored specific oncogenic driver mutations, including EGFR, KRAS, ROS1, ALK, and others (Fig. 1A; Table S1). Among them, EGFR and KRAS mutations were the most common and were often accompanied by co-mutations with other genes. Given the potential of these driver mutations to promote brain metastasis [11], we further incorporated scRNA-seq data from NSCLC patients with brain metastases into our study (Figure S1B).

Following rigorous quality control and Harmony-based integration, we obtained high-quality transcriptional profiles for 196,560 cells, which were classified into seven major lineages using canonical marker genes: epithelial cells, endothelial cells, fibroblasts, astrocytes, T/NK cells, B cells, and myeloid cells (Fig. 1B, 1E, S1A Method). Concurrently, we observed that cells from different datasets and driver

subtypes consistently clustered by cell type, with each cluster comprising cells from multiple patients, indicating the conservation of major cell types across NSCLC with distinct driver mutations and the robustness of data integration (Figures S1B).

Through in-depth dissection of cellular composition patterns in NSCLC harboring distinct driver mutations, we found that immune cells comprise a significant proportion across all driver subtypes (Figs. 1C, 1D, 1F, S1C), while exhibiting distinct compositional patterns that reveal heterogeneous immunological landscapes across different mutational contexts. Notably, NSCLC with EGFR brain metastases (EGFR-BM) and ROS1-mutant exhibited significantly lower lymphocyte proportions compared with myeloid cell infiltration (Fig. 1G), clearly indicating dominance of innate immunity in their TIME (Method). In addition, we observed that endothelial cells, fibroblasts, and astrocytes were markedly enriched in malignant signaling pathways, such as angiogenesis and epithelial–mesenchymal transition (EMT) [12], whereas T/NK cells and myeloid cells predominantly exhibited enrichment in interferon-α [13] and interferon-γ pathways. These findings suggest that stromal and immune cells within the TIME may promote cancer progression and invasion through distinct oncogenic signaling axes (Fig. 1H).

To further elucidate the intrinsic heterogeneity and degree of malignancy of the TIME, we performed reclustering of its major components, including immune cells (myeloid, B, and T/NK cells) and stromal cells (astrocytes, endothelial cells, and fibroblasts), and generated a high-resolution atlas comprising 64 cellular subsets that comprehensively delineates the cellular composition of the TIME (Fig. 1I; Table S2). Notably, these TIME cell subsets exhibited driver mutation–specific distribution patterns in NSCLC (Figure S1D, Method). For example, astrocyte subsets were predominantly enriched in EGFR-BM patients, whereas distinct B cell and plasma cell subsets were more prominently enriched in patients harboring KRAS and TP53 mutations, respectively. In particular, B cells expressing HLA-DRA and HLA-DQB1, which are closely associated with MHC class II antigen presentation [14], suggest the presence of a potential B cell–mediated antitumor immune response. Collectively, we established a comprehensive single-cell atlas of NSCLC with driver mutations, revealing TIME cell subsets with subtype-specific distribution patterns.

Cellular module analyses reveal five TIME subtypes

Given the pivotal role of the heterogeneous composition of immune and stromal cells in shaping the tumor immune landscape [15], we conducted a co-occurrence analysis of TIME cells in NSCLC patients with distinct driver mutations and identified five robust multicellular modules (CM1–5) exhibiting co-occurrence patterns (Fig. 2A; Table S3; Method). Strikingly, diverse immune and stromal cell subsets associated with prognostic risk coexisted within the same module, suggesting that each module represents a distinct TIME subtype with potential impact on tumor progression (Table S4).

To further delineate the characteristics of each TIME subtype in tumor immunity, we performed a systematic analysis by integrating differentially expressed genes, functional enrichment results, and immune-related gene set scores (Figs. 2B–2D, Figures S2A–S2B, Table S5, Method). CM1, characterized

by high lymphocyte infiltration, exhibited pronounced upregulation of immunoglobulin-related genes, including IGHA1, IGHG3, and IGKC, potentially linked to antitumor immunity [16]. Moreover, enrichment of the unfolded protein response (UPR) pathway in CM1 suggests potential alterations in T cell phenotype and function [17], and its significantly elevated scores in T cell differentiation, antigen receptor–mediated signaling, and cytotoxic/proinflammatory signaling collectively indicate that CM1 represents a lymphocyte-mediated, immune-activated TIME. In contrast, CM2 exhibited a complex cellular composition, including co-enriched multiple astrocyte subtypes, myofibroblasts, and heterogeneous immune cell subsets. Upregulation of SPARC may be linked to APA-mediated mechanisms of cancer progression in tumor-associated fibroblasts [18], whereas elevated PLP1 suggests the presence of Schwann cell-mediated angiogenesis promoting tumor progression within the TIME [19]. Additionally, enrichment of cytokine-positive regulation and elevated expression of CC chemokines may act as driving forces of malignant progression. The enrichment of glial cell development, differentiation, and generation further supports astrocyte presence, collectively characterizing CM2 as a malignant TIME driven by astrocyte-myofibroblast interactions.

Interestingly, CM3 exhibited high myeloid and T cell infiltration, whereas the immunosuppressive Macrophage_04_SPP1 subset may promote metastasis via angiogenesis and ECM remodeling [20]. CM3 also exhibited specific expression of the tumor suppressors SCGB3A1 and SCGB3A2, as well as SFTPB, a gene associated with lung cancer risk, alongside NK_01_FGFBP2 enrichment linked to major pathological response [21]. Functionally, CM3 was enriched in MHC class II-mediated antigen presentation and positive regulation of leukocyte proliferation. Collectively, CM3 harbors both immune-activating signals and immunosuppressive elements, indicating a dynamic state where antitumor immunity and immune evasion coexist. CM4 characterized high infiltration of lymphocytes and alveolar fibroblasts, with marked upregulation of TTC19, involved in cell division, and EEF1G, linked to translation, alongside enrichment in ribosome biogenesis, translation-related pathways, and signals regulating lymphocyte homeostasis and T cell activation. However, its relatively low antigen presentation score suggests impaired antigen presentation or reliance on non-classical pathways for T cell activation, indicating an “activated but insufficient” immune state. In contrast, CM5 was dominated by diverse myeloid and endothelial cell subsets, with notable upregulation of macrophage-related marker genes and chemokine signaling, and enrichment in MHC class II-mediated antigen presentation and pathways regulating T cell–mediated immunity. Nonetheless, its elevated APC co-inhibition score points to mechanisms that suppress T cell–mediated antitumor responses. Therefore, CM5 represents an immunosuppressive TIME subtype that exhibits enhanced antigen presentation while concomitantly restraining T cell functionality.

To further investigate the association between the five TIME subtypes with distinct immune states and driver-mutant NSCLC, we assessed the most predominant TIME subtype in each patient (Fig. 2E; Method). The immune high-infiltration subtypes CM1 and CM3 were enriched in EGFR-driven patients, whereas CM2 was preferentially enriched in those with EGFR-BM, suggesting that brain metastasis is accompanied by remodeling of the immune microenvironment, consistent with previous findings [22]. Moreover, CM3 was not only the most prevalent TIME subtype across the cohort but may also represent

a more general immune state; CM4 was predominantly enriched in KRAS-driven tumors, suggesting it may be dominated by a specific oncogenic pathway; while CM5 encompasses patients with diverse driver alterations such as ROS1, KRAS, and others, hinting that it represents an immune subtype characterized by higher heterogeneity and more complex underlying mechanisms. Further analysis of tissue preference also confirmed these distribution patterns (Fig. 2F).

Owing to oncogenic signaling pathways, cytokines, and chemokines can profoundly influence the cellular composition and immune phenotype of the TIME [23]. To further investigate the mechanisms underlying the heterogeneous preference patterns of different TIME subtypes, we evaluated malignant pathway activity in each TIME subtype and driver-mutant NSCLC (Fig. 2G) and analyzed the expression patterns of chemokines (Fig. 2H). Distinct malignant pathway enrichment patterns were observed across TIME subtypes and driver-mutant NSCLC. In particular, CM1 and CM4 exhibited highly similar malignant pathway enrichment features, including key pathways such as PI3K and MAPK, accompanied by a pattern of upregulated chemokine expression and downregulated receptor expression. These findings suggest that chemokine expression patterns may play a critical role in shaping malignant pathways, thereby influencing the subtype-specific distribution of the TIME in driver-mutant NSCLC.

In summary, we identified five TIME subtypes with distinct immune states in driver-mutant NSCLC, characterized by multicellular co-occurrence, functional heterogeneity, and specifically malignant pathway enrichment, collectively reflecting patterns of immune microenvironment remodeling and regulatory mechanisms in the context of oncogenic drivers.

CM2 and CM5 are associated with poor prognosis

Due to cancer cells interact with the TIME through multiple oncogenic pathways, forming complex signaling networks that influence immune infiltration and patient outcomes[24–27]. To elucidate the intrinsic determinants underlying the heterogeneous distribution of different TIME subtypes, we first performed a systematic analysis focusing on the central origin of oncogenic pathways—cancer cells. Based on tumor cells identified by Copycat in the integrated dataset, we conducted multiresolution unsupervised clustering by sample and extracted upregulated marker genes for each cluster to delineate their principal molecular features. Through consensus clustering, we identified 13 gene elements (GEs) representing the key molecular characteristics of NSCLC cancer cells with distinct driver mutations (Figures S3C-S3D; Table S6; Method). Using these GEs, each cancer cell was scored and assigned a predominant cell state (Fig. 3A).

Notably, GE1 exhibited elevated ATP metabolism and oxidative phosphorylation activities, features that may be associated with a favorable prognosis [28] (Table S7). In contrast, subgroups such as GE2, GE3, and GE7 displayed pronounced cell-cycle activity and enhanced metabolic states, suggesting that they are in a highly proliferative phase and potentially possess stronger division and invasive capabilities. GE9 was characterized by high levels of phagosome activity and humoral immune responses, implying that it may reside within a microenvironment enriched with tertiary lymphoid structures (TLSs), accompanied

by substantial B cell infiltration and potentially linked to augmented anti-tumor immune responses [29]. Further dimensionality reduction and clustering analyses of cancer cells revealed pronounced intra- and inter-tumoral heterogeneity in cancer cell states across NSCLC harboring distinct driver mutations (Fig. 3B). In EGFR-driven tumors, the predominant states included GE1 with high oxidative phosphorylation, GE6 with TGF-β–associated programs, and GE13 enriched for leukocyte-mediated immune responses, indicating the coexistence of pro-invasive and antitumor immune activities. Consistent with previous studies, EGFR signaling suppressed $CD8^+$ T cell infiltration and conferred resistance to PD-1 blockade via ERK1/2–P90RSK–TGF-β activation [40], thereby fostering an immunosuppressive TIME. In EGFR brain metastases, beyond the GE1 and GE13 states observed in primary tumors, two additional states emerged: GE2, characterized by high proliferative and pyrimidine metabolism activity consistent with medulloblastoma [30], and GE3, characterized by nucleotide metabolism–driven high-proliferation and invasive behaviors observed across multiple cancer types [31]. These findings suggest substantial remodeling of cancer cell states during brain metastasis in EGFR-driven NSCLC. In KRAS-mutant patients, the cancer cells were relatively rare. However, in those with KRAS co-mutations, in addition to GE1 with high oxidative phosphorylation, GE8 with cysteine metabolism–mediated ferroptosis resistance[32], and GE9 with pronounced humoral immune activity also emerged. These findings indicate that in KRAS co-mutant patients, cancer cells may employ a dual adaptive strategy, involving metabolic reprogramming to resist oxidative stress and ferroptosis, while maintaining an immune-active state [33]. Distinct from the above subtypes, ROS1-mutant NSCLC was dominated by the GE1 subgroup with ATP metabolism and oxidative phosphorylation, showing a relatively homogeneous state with limited functional diversity.

To further investigate the potential influence of distinct GEs on TIME phenotypes, we conducted gene set scoring of 13 GE states in TIME cells. The results revealed that GE1, GE5, GE8, and GE9 exhibited significant associations with specific TIME subtypes (Fig. 3C), suggesting crucial roles in shaping the TIME. Strikingly, CM2 showed a relative paucity of GE1 cancer cells with low oxidative phosphorylation activity, suggesting the presence of a Warburg effect. Under hypoxic conditions, tumor cells may suppress mitochondrial oxidative phosphorylation and instead rely on glycolysis or amino acid metabolism to sustain energy production and rapid proliferation [34]. This metabolic profile is often associated with higher malignant potential and poorer clinical prognosis, indicating that CM2 may represent a more aggressive TIME subtype. In contrast, the GE types enriched in CM5 closely matched the predominant cancer cell features observed in KRAS co-mutant patients, consistent with its preferential distribution in this driver mutation. CM5 also exhibited significant enrichment of GE5, suggesting that this subtype may be driven by cancer cells with heterogeneous functional states, thereby generating a more complex TIME. Furthermore, GE1, GE5, and GE9 were enriched in ROS1-driven NSCLC (Figure S3E), also supporting the resemblance between the TIME of ROS1-driven tumors and CM5.

Given the critical role of copy number variations (CNVs) in driving cancer cell heterogeneity, we next characterized CNV profiles across the four key cancer cell states (Fig. 3D, Figure S3A-S3B, Table S8; Method). Each state exhibited distinct CNV patterns. In GE1, notable amplifications included PMVK, which promotes sustained activation of the aberrantly activated β-catenin signaling pathway in cancer

[35]; the proto-oncogene PBXIP1; and PYGO2, which has been implicated in immune suppression and malignant progression in prostate cancer, liver cancer, and colorectal cancer. In GE5, PFKFB2 amplification may inhibit NSCLC invasion and glycolysis via the PI3K/AKT pathway, thereby shifting the metabolic preference toward ATP production. Of particular note, MET was significantly amplified in GE8, activating the receptor tyrosine kinase c-MET to promote cancer cell migration, invasion, angiogenesis, and epithelial–mesenchymal transition (EMT). In contrast, GE9 primarily exhibited amplification of calcium-binding proteins such as S100A2/4/6, with increased S100A4 expression being strongly associated with high metastatic potential and poor prognosis [36].

To systematically evaluate the potential malignancy of the five TIME subtypes at the patient level, we conducted survival analyses (Fig. 3E, Method). Among them, CM2 and CM5 were significantly associated with worse patient outcomes, consistent with the cancer cell state profiles enriched in these subtypes. In contrast, CM3 was linked to a favorable prognosis, indicative of a TIME characterized by lower malignant potential and enhanced immune features. Collectively, cancer cell state heterogeneity is tightly associated with the functional characteristics of TIME subtypes, profoundly influencing patient outcomes and providing a theoretical basis for optimizing tumor immunotherapeutic strategies.

Potential mediators underlying the poor-prognosis CM2 and CM5

Cellular crosstalk within the TIME is pivotal for tumor initiation and progression [37]. We next focused on the intercellular communication networks between CM2 or CM5 and their corresponding cancer cells. Comparative analyses showed that CM2 had significantly more and stronger interactions than CM5, suggesting a more intricate communication network. (Fig. 4A, Figures S4A–S4B, Method). Specifically, CM2 formed a high-intensity communication network comprising Astrocyte_05_HLA-DRB5, cancer cells, myeloid cells, and multiple fibroblast subpopulations (Fig. 4B). Among these fibroblast subsets, HIGD1B, COL10A1, and MKI67 were highly expressed, indicating pronounced proliferative capacity [38] and extracellular matrix remodeling potential, which may provide a supportive microenvironment for tumor growth and immune evasion [39]. In contrast, the CM5-associated communication network was primarily composed of multiple endothelial cell subsets, alveolar fibroblasts, GE1 cells, and myeloid cells (Fig. 4C). Notably, vascular cell adhesion molecule VCAM1 and the chemokine CCL21 were predominantly expressed by endothelial cells, whereas CD68 served as a myeloid cell marker, implying that CM5 is more likely involved in vascular activation, macrophage recruitment, and chemokine-mediated NSCLC metastasis [40].

To elucidate the crucial molecular mechanisms underlying these networks, we further examined the signaling crosstalk patterns within the major communication networks of CM2 and CM5 (Figures S4C-S4D; Tables S9-S10). CM2 and CM5 exhibited distinct signaling profiles, while both were characterized by elevated SPP1 signaling mediated by different cell populations (Fig. 4D). In CM2, SPP1 signals were predominantly emitted by Astrocyte_05_HLA-DRB5 and Macrophage_06_C3, and were received by endothelial and fibroblast subclusters. Notably, the complement component C3 and its receptor C3AR1 have been implicated in the progression of human gliomas, correlating with more aggressive disease

phenotypes and shorter survival [41]. These findings suggest that CM2 may represent a specific TIME subtype in NSCLC brain metastases, in which $C3^+$ macrophages and astrocyte-like cells cooperatively mediate the SPP1 signaling pathway, potentially contributing to poorer prognosis. In contrast, in CM5, SPP1 signaling primarily originated from $CD68^+$ endothelial cells, $MS4A7^+$ fibroblasts, and $LGMN^+$ macrophages, and was largely received between macrophage subsets and $CHIT1^+$ mast cells. This intricate signaling network between myeloid and stromal cells may facilitate the establishment of an invasive and pro-metastatic tumor microenvironment, in line with previous findings that SPP1-driven signaling promotes tumor progression.

In addition, CM2-specific PTN and PSAP signaling may also serve as critical drivers of its high invasiveness. PTN signals were jointly emitted by Astrocyte_05_HLA-DRB5 and fibroblasts toward multiple GEs cancer cell subclusters, consistent with pro-tumorigenic mechanisms in bladder cancer[42]. PSAP signaling was broadly produced by nearly all key cellular populations within CM2 and ultimately targeted Astrocyte_05_HLA-DRB5, a pattern closely associated with tumor-promoting and metastatic potential. Ligand–receptor network analysis further revealed that SPP1, the SDC gene family, multiple chemokines, and angiopoietin-like protein ANGPTL formed central interaction hubs, collectively contributing to tumor progression (Fig. 4E). By comparison, CM5 exhibited activation of MIF, GALECTIN, and RESISTIN pathways, mainly originating from GE1 and multiple other cell types, which targeted $LGMN^+$ and $MARCO^+$ macrophages as well as mast cells. Previous studies have shown that in TLS-deficient microenvironments, MIF and GALECTIN pathways enhance tumor–immune cell interactions, facilitating the recruitment of immunosuppressive cells and establishing an immunosuppressive niche [43]. Furthermore, BAFF, GALECTIN, and MK signaling axes have been implicated in the induction of immunosuppressive states [44]. We also observed markedly elevated CXCL and CCL chemokine signaling from $LGMN^+$ and $MARCO^+$ macrophages to $VCAM1^+$ endothelial cells, potentially underlying the pro-angiogenic mechanisms of this subtype. Collectively, the coordinated activation of multiple chemotactic and immunosuppressive signaling pathways in CM5 may drive its immunosuppressive phenotype.

To further delineate the potential mediators of poor prognosis, we assessed the prognostic relevance of subtype-specific ligand–receptor pairs in the two aggressive TIME subtypes, CM2 and CM5. Hazard ratio (HR) analyses identified 10 CM2-specific and 6 CM5-specific ligand–receptor pairs significantly associated with patient prognosis (Fig. 4F, Method). In CM2, ANGPTL2 is known to promote lymphangiogenesis and plays an important role in lung cancer metastasis [45]. Its mediated ligand–receptor pair was predominantly expressed in fibroblasts from EGFR-BM, KRAS-mutant, and KRAS co-mutant NSCLC, indicating this molecular mechanism may enhance the metastatic potential of lung cancer. ANGPTL4 has been reported to be expressed in hypoxic NSCLC cells and is associated with the regulation of glutamine metabolism, suggesting its potential involvement in the metabolic reprogramming of cancer cells within CM2. Notably, ANGPTL4_SDC1/SDC4 was highly expressed in GE1 cancer cells from EGFR-BM, further supporting its role in cancer cell metabolic reprogramming. Additionally, PSAP_GPR37 exhibited consistently high expression across NSCLC driven by different oncogenic mutations, potentially representing a shared mediator of poor prognosis. In contrast,

CX3CL1_CX3CR1, which was associated with a favorable prognosis, showed high expression in Astrocyte_03_ROM1 from EGFR-BM. This finding suggests that multiple ligand–receptor mechanisms may cooperatively influence prognosis in this subtype.

In CM5, all six key ligand–receptor pairs were significantly linked to poor prognosis. Among them, IGF1_ITGA6_ITGB4, mediated by endothelial cells and macrophages, emerged as a critical adverse prognostic axis across NSCLC harboring KRAS, EGFR, and other driver mutations. Previous studies have also implicated ITGA6 and ITGB4 in the progression of hepatocellular carcinoma and nasopharyngeal carcinoma. RETN-mediated ligand–receptor pairs showed a distribution pattern similar to SCGB3A2_MARCO, with high expression in macrophages from ROS1- and KRAS-mutant subtypes but low expression in those from EGFR-BM NSCLC, suggesting a mutation-subtype-specific mechanism of immune evasion and tumorigenic potential.

Finally, we examined the major metabolic characteristics of the aggressive TIME subtypes and their associated cellular populations (Figs. 4H, S4E, Method). Overall metabolic activity was lower in CM2 than in CM5, whereas CM2-associated cancer cells showed higher activity. Cancer cells from both subtypes were enriched in glycosphingolipid and glycosaminoglycan metabolism, pathways previously linked to cancer cell invasiveness, survival, and metastasis, suggesting shared metabolic drivers of tumor progression. Notably, CM2-specific activity was enriched in linoleic acid and caffeine metabolism, which have been associated with enhanced T cell function and antitumor immunity, while CM5 exhibited broad metabolic activation, indicating multiple routes to sustain malignancy. Subcluster-level analysis showed that cells of the same type shared metabolic patterns but varied markedly in activity levels; among them, GE9 cancer cells displayed the highest activity, potentially enhancing humoral immune responses. Collectively, these analyses identified key mediators in the two aggressive TIME subtypes that are likely to drive poor prognosis (Fig. 4G), providing a potential biological basis for understanding the malignant evolution of the TIME.

Identifying the DMSP.sig with immune characteristics

Dysregulation of transcription factor (TF) expression or activity can disrupt cellular homeostasis, thereby driving tumor initiation and progression [46]. To investigate mechanisms of homeostatic imbalance and immune remodeling in the aggressive TIME subtypes CM2 and CM5, we assessed their key TF activities (Fig. 5A, Method). In CM2, ZNF202 and NFX1 exhibited marked activation. ZNF202 has previously been identified as a poor prognostic factor in head and neck squamous cell carcinoma [47]. It regulates cholesterol efflux by modulating ABCA1 expression through the macrophage-specific TLR2/NF-κB/ZNF202 signaling axis. NFX1 may promote immune evasion by suppressing the expression of its target MHC class II genes and modulating inflammatory responses. Thus, CM2 may be driven by macrophage-centered metabolic reprogramming and impaired antigen presentation, together contributing to its immunosuppressive phenotype and poor outcome. In contrast, the pronounced activation of epithelial-specific TES TFs ELF5 and EHF in CM5 may jointly facilitate NSCLC progression and regulate angiogenesis.

To delineate the malignant regulatory programs underlying the aggressive TIME subtypes, we constructed subtype-specific regulatory networks linking active TFs to HR-related ligand–receptor genes (Method). These networks focus on the regulation of HR-related ligand–receptor pairs within core interacting cells, indicating that these transcription factors may drive immunosuppression and tumor progression (Fig. 5B). HR analyses further identified core regulons associated with patient risk, supporting their key roles in the malignant evolution of aggressive TIME subtypes (Table S11). In parallel, we characterized the key CNV and driver gene–associated regulons in malignant cells associated with each subtype (Figures S5A–S5B). Of note, while CM5 exhibited a more intricate malignant regulatory network involving angiogenesis and metabolic reprogramming processes mediated by multiple regulons such as CMKLR1, the regulatory networks within CM2-associated cancer cells appeared even more complex than those in CM5, implying a stronger engagement of malignant driving forces.

Given the close association of these malignant regulons with patient prognosis and their potential roles in mediating immunosuppression and tumor progression, we further integrated those with significant prognostic value to construct a prognostic scoring model based on their expression features, aiming to systematically evaluate its predictive performance in NSCLC risk stratification (Method). Specifically, we utilized the TCGAplot package to access bulk RNA-seq data from 917 NSCLC patients and randomly divided the cohort into training and validation sets at a 7:3 ratio (Methods). In the training cohort, univariate Cox regression of 63 HR-related candidate regulons identified 14 high-risk regulons. We then applied LASSO regression with 10-fold cross-validation to minimize overfitting and collinearity, yielding nine TF–ligand–receptor regulons (PAX6–ITGB1, CEBPB–IGF1, CEBPB–RETN, NR1H4–RARRES2, PPARG–RARRES2, RELA–IGF1, SNAI1–ITGB4, SP1–ITGA6, and USF2–IGF1). The final Risk Score (RS) was calculated as the sum of these regulons weighted by their respective coefficients (Fig. 5C; Methods; Table S12). Furthermore, we assessed the protein expression of transcription factors and genes within these nine regulons using IHC data downloaded from the HPA (http://www.proteinatlas.org) for normal and lung cancer tissues. Notably, ITGA6, IGF1, ITGB1, ITGB4, and PPARG displayed tumor-specific upregulation, underscoring their functional relevance in the malignant context (Fig. 6A–6H; Methods).

To evaluate model performance, patients in the training and validation cohorts were stratified into high- and low-risk groups using the median RS. Model performance was assessed by Kaplan–Meier survival analysis (Methods), ROC curves, and the RS distributions. Kaplan–Meier analysis demonstrated significantly longer survival in the low-risk group in both cohorts (Fig. 5D). ROC analysis confirmed the model's discriminative capacity for overall survival (training cohort AUCs at 1, 3, and 5 years: 0.68, 0.65, 0.66; validation cohort: 0.54, 0.64, 0.69) (Figure S5C), and high-risk scores were consistently associated with increased mortality (Figure S5D). Assessment of clinical factors revealed that patients with stage II or III disease had significantly worse outcomes in both cohorts, and risk increased with tumor stage (Fig. 5E). Among the regulons, RELA–IGF1 and SNAI1–ITGB4 were robustly associated with poor prognosis in both cohorts, ranking third and fifth, respectively, by coefficient magnitude in the model (Fig. 5F). The top contributor, SP1–ITGA6, also showed strong oncogenic features; SP1 has been shown to enhance tumor resistance to programmed cell death and promote immune evasion [48], while ITGA6

can drive radio resistance in NSCLC via the ITGA6/PI3K/Akt signaling axis, suggesting a cooperative role in supporting tumor cell survival and therapy resistance in NSCLC.

To investigate the association between the risk model and immune landscape shaping, we examined the correlations between the risk scores, the nine constituent regulons, and immune cell infiltration levels (Fig. 5G, Method). Notably, high-risk scores showed reduced infiltration of $CD8^+$ T cells and helper T cells, increased infiltration of dormant memory T cells, and high infiltration of both activated and resting mast cells. These patterns suggest that high-risk patients reside in a microenvironment with impaired immune defense and sustained pro-tumor inflammatory signaling. Further analyses revealed that several key regulons exhibited consistent associations with reduced infiltration of anti-tumor immune populations, including helper T cells and activated dendritic cells (DC). Conversely, certain regulons had divergent effects: PPARG–RARRES2 and CEBPB–RETN were positively correlated with high infiltration of pro-tumor, anti-inflammatory M2 macrophages, whereas SP1–ITGA6 was associated with low infiltration of M2 macrophages. By weighting each regulon's expression with its corresponding coefficient, the model captures a composite immunological state characterized by a "desert" of immune-activating cells alongside activation of tumor-promoting cells.

Finally, we systematically examined the association between the risk model and NSCLC driver mutation subtypes (Fig. 5H). The RS varied significantly across subtypes: patients with ALK or ROS1 alterations generally showed higher scores, whereas those with EGFR-BM or HER2 mutations had lower scores. These differences suggest that the RS captures essential distinctions in tumor progression patterns and immune microenvironmental features among driver mutation subtypes, offering insights into subtype-specific mechanisms and potential therapeutic targets. On this basis, we designated the model as the Driver mutation-specific prognostic signature (DMSP.sig). We further evaluated score distributions across subtype-related cellular modules (Fig. 5I; Figure S5E). CM2, representing a brain metastasis–associated microenvironment dominated by astrocytes and diverse stromal cells, had consistently lower scores. In contrast, CM5, largely corresponding to ROS1-, KRAS-, and ALK-driven tumors, exhibited a macrophage-enriched TIME with markedly elevated scores, driven predominantly by CEBPB–RETN and CEBPB–IGF1. Taken together, the DMSP.sig not only provides robust prognostic prediction in NSCLC but also delineates striking differences in immune composition and malignant regulatory programs across driver mutation subtypes. The model captures a dysregulated immune contexture, defined by the "desertification" of immune-activated cells alongside the activation of pro-tumor cells, thereby offering a framework for molecular subtyping and precision therapeutic strategies.

Identification of potential therapeutic agents targeting prognostic biomarkers

Given the close association between the identified DMSP.sig and patient prognosis as well as immune status within the tumor microenvironment, we next sought to identify potential therapeutic agents that may target these prognostic markers. Based on a subset of tumor-specific DMSP.sig markers that were validated at the protein level, drug sensitivity analysis was performed (Methods). Using oncoPredict, we

identified multiple compounds exhibiting high predicted sensitivity across these eight prognostic biomarkers (Table 13). For each target gene, the compound with the highest predicted sensitivity was further highlighted (Fig. 7A). Overall, the predicted sensitivities of the identified compounds showed significant positive correlations with the expression levels of all eight prognostic genes. Notably, the predicted sensitivities of LY-2157299, FSC231, and JW-55 exhibited particularly strong correlations with the expression of ITGA6, PPARG, and RELA, respectively, with correlation coefficients exceeding 0.90. These results indicate that higher expression of these prognostic markers is associated with increased drug sensitivity, suggesting that these compounds may exert therapeutic effects by targeting pathways related to these genes, thereby potentially modulating immune-related phenotypes and contributing to prognostic improvement.

To further evaluate the binding capacity between candidate drugs and their corresponding target proteins, molecular docking analyses were conducted using protein structural models and small-molecule conformations (Fig. 7B, Methods). The results showed that the binding energies of all drug–target complexes were lower than − 5 kcal/mol, indicating favorable binding affinity. In particular, LY-2157299 exhibited a strong binding interaction with the protein encoded by ITGA6, with a binding energy of − 9.7 kcal/mol. This finding is consistent with the predicted drug sensitivity results, further supporting the strong interaction between LY-2157299 and ITGA6, and suggesting that this compound may effectively interfere with ITGA6-associated intercellular interactions, thereby contributing to improved patient prognosis.

Discussion

In this study, we systematically analyzed NSCLC harboring driver mutations, extending the focus from histopathology to molecular subtypes to delineate immune heterogeneity. Previous studies have shown that driver-mutant NSCLC requires distinct therapeutic strategies and exhibits variable responses to immune checkpoint inhibitors (ICIs) [87]. However, most single-cell studies have focused on LUAD and LUSC [21] or concentrated on single mutations such as EGFR, leaving the immune heterogeneity among multiple driver subtypes insufficiently characterized.

To address this gap, we integrated approximately 200,000 scRNA-seq profiles and systematically mapped the TIME landscape in NSCLC with diverse driver backgrounds. Compared with earlier studies [11], we further incorporated the contributions of stromal cells to tumor invasion and identified five CMs, representing functionally distinct TIME states that more precisely capture the immune features of driver-mutant NSCLC. Notably, CM1 closely resembled the previously reported immune-activated TIME (Figure S2B), predominantly composed of lymphocyte subsets [90]. In contrast, CM2 and CM5 exhibited immunosuppressive and tumor-promoting features, with CM5 resembling immune-excluded TIME but differing in cellular composition. Moreover, these TIME subtypes displayed subtype-specific distributions: CM1 and CM3 were enriched in EGFR-mutant NSCLC, CM2 was enriched in EGFR-mutant brain metastases, CM4 in KRAS-mutant NSCLC, and CM5 in ROS1/ALK-rearranged NSCLC. Importantly,

EGFR-mutant NSCLC showed altered TIME phenotypes upon brain metastasis, consistent with previous reports [31].

Given the pronounced differences in malignant pathway activity across cell programs, we further explored associated cancer cell states. Unlike conventional clustering analyses, our study captured both intra- and intertumoral molecular features, identifying 13 functionally distinct GEs. Among these, GE1, GE5, GE8, and GE9 were significantly associated with TIME programs, supporting prior evidence that oncogenic pathways modulate immune composition [32, 91]. Specifically, CM2 was characterized by reduced expression of oxidative phosphorylation–related GE1, reflecting a Warburg-like metabolic profile that may account for its invasiveness and immunosuppressive properties. In contrast, CM5 exhibited broad activation of EGFR, WNT, and JAK-STAT pathways and was closely linked to key malignant states, indicating complex tumor–immune interactions. Mechanistically, CM2 was defined by astrocyte–fibroblast–$C3^+$ macrophage axes (SPP1, PTN, PSAP) mediating immunosuppression and promoting EGFR- and MET-associated brain metastasis, reminiscent of immunosuppressive mechanisms observed in retinoblastoma [94]. CM5, however, relied primarily on myeloid–endothelial interactions, chemokine activation, and stromal signaling, shaping the immunosuppressive TIME in ROS1-, KRAS-, and ALK-mutant tumors. Collectively, these findings provide a comprehensive view of mutation-specific molecular mechanisms underlying invasive TIME subtypes.

Building on these observations, we identified the DMSP.sig. Unlike previous NSCLC prognostic models [95, 96], this model was derived from invasive subtypes CM2 and CM5, integrating key ligand–receptor pairs and high-risk TFs to capture dual immune states of desertification and tumor-promoting activation quantitatively. Risk scores varied significantly across driver subtypes, reflecting the mutation-specific contribution to TIME remodeling and providing a robust predictor of patient prognosis. This framework highlights the interplay among driver mutations, malignant pathways, and immunosuppressive networks, offering insights into NSCLC immune evolution and potential therapeutic targets. In addition, the stratification defined by DMSP.sig provided a practical framework for prioritizing drugs with mutation- and microenvironment-specific sensitivity, linking prognostic risk states to therapeutic vulnerability.

Several limitations should be acknowledged. First, rare driver subtypes were underrepresented due to their low incidence [4], which may limit the generalizability of our findings. Second, the study primarily focused on invasive subtypes CM2 and CM5, whereas the favorable CM3 subtype was not extensively explored; its molecular basis warrants validation in larger cohorts and functional studies. Our study provides key insights into TIME heterogeneity in driver-mutant NSCLC and lays a foundation for precision immunotherapy. Future studies integrating larger datasets and multi-omics analyses are expected to further enhance the robustness of these findings and facilitate their clinical translation.

Materials and methods

Sample characteristics

To construct a comprehensive single-cell transcriptomic atlas encompassing multiple driver mutations of NSCLC, we integrated scRNA-seq data from 45 untreated patients with driver mutation–positive NSCLC, collected from seven publicly available datasets. Based on gene mutation type (e.g., EGFR, KRAS), mutation pattern (single driver mutation or co-mutation), and tissue source (primary tumor or brain metastasis), patients were classified into ten distinct driver mutation subtypes: EGFR (n = 11), EGFR-BM (n = 10), EGFR co-mutation (n = 2), KRAS (n = 12), KRAS co-mutation (n = 3), ROS1 (n = 3), TP53 (n = 1), ALK (n = 1), MET brain metastasis (MET-BM, n = 1), and HER2 (n = 1) (Table S1).

Bulk RNA-seq data for TIME subtype survival analyses were retrieved from The Cancer Genome Atlas (TCGA), comprising 200 lung adenocarcinoma (LUAD) and 200 lung squamous cell carcinoma (LUSC) samples; after quality control, 386 cases were retained. For prognostic model training and validation, bulk RNA-seq data were obtained from the LUAD and LUSC datasets embedded in the TCGAplot package (v8.0.0), totaling 917 patients, and randomly split into a training cohort (n = 641) and a validation cohort (n = 276) in a 7:3 ratio.

Processing and annotation of scRNA-seq data

To ensure uniform gene nomenclature across datasets, gene names were converted to the "SYMBOL" format using the R package org.Mm.eg.db (v3.20.0). Given the dataset-specific distributions of UMI counts, gene features, and mitochondrial gene percentages, quality control was performed separately for each dataset. Low-quality cells were removed if they had UMI counts ≤ 50,000, ≤ 200 detected genes, or ≥ 25% mitochondrial gene content. Highly variable genes (top 2,000) were identified using the FindVariableFeatures function, followed by principal component analysis (PCA) for dimensionality reduction. Batch effects were corrected using Harmony (v1.2.3).

Based on the Harmony output, clustering was performed with Seurat (v5.3.0). a clustering tree was constructed using the Clustree package (v0.5.1), with an optimal resolution of 0.4, resulting in 23 clusters. Marker genes for each cluster were identified with FindAllMarkers.

Cell type annotation combined predictions from SingleR (v2.8.0) with top 20 cluster-specific DEGs matched to the CellMarker 2.0 database, further refined by known lineage markers. Annotation accuracy was confirmed through differential expression analysis across cell types.

Major TIME populations, including myeloid cells, B cells, T cells, and stromal cells, were subjected to secondary clustering. Optimal resolutions of 0.6, 0.4, 0.5, and 0.2 were selected based on hierarchical relationships, and subclusters were annotated by integrating their DEGs with canonical markers and functional features, e.g., Macrophage_01_LGMN within the myeloid compartment.

Immune composition analysis

To assess the immune composition characteristics of NSCLC subtypes with different driver mutations, we referenced the immune cell classification method from existing literature [49], where innate immunity is primarily mediated by myeloid cells, and adaptive immunity involves T cells and B cells. The overall

proportion of myeloid cells was calculated for each subtype, along with the overall proportion of T cells and B cells, followed by normalization. The t-test (t.test) was then used to assess the statistical significance of distribution differences between subtypes with different driver mutations.

Ro/e ratio analysis

In NSCLC samples with different driver mutations, the ratio of observed cell counts to the expected cell counts, based on the overall distribution (Ro/e), was calculated for the 64 cell subpopulations in the TIME to characterize the tissue preference of each subpopulation. A Ro/e value greater than 1 indicates that the cell subpopulation shows tissue preference in that driver mutation subtype, and statistical significance was assessed using the Chi-square test.

Similarly, to compare the distribution differences of the five TIME subtypes across NSCLC samples with different driver mutations, we calculated the Ro/e values for each TIME subtype in different driver mutation NSCLC subtypes and determined tissue preference and significance levels according to the same criteria.

Cell module identification

To dissect immune states within the TIME of driver mutation–positive NSCLC, a co-occurrence analysis was performed for the 64 TIME subpopulations. The procedure is as follows: first, the normalized frequency of each cell subpopulation in patient samples was calculated, excluding unclassified "unknown" cells. Next, the pairwise Pearson correlation coefficients between the normalized frequencies of cell subpopulations were calculated using the corr. test function from the R package psych (v2.5.6), with statistical significance assessed via a t-test. Based on the correlation matrix, hierarchical clustering of cell subpopulations (ward.D2 method) and visualization were performed using the R package pheatmap (v1.0.13). Ultimately, five stable cell modules were identified, each representing a co-existing or functionally similar TIME subtype in patients.

Differential gene expression and pathway enrichment analysis

For each TIME subtype, the top 200 differentially expressed genes (DEGs) were subjected to Gene Ontology (GO) biological process (BP) enrichment analysis using the enricherGO function of the clusterProfiler package (v4.14.6) with parameters pvalueCutoff = 0.05 and qvalueCutoff = 0.2. The top 10 enriched pathways were visualized.

Functional comparisons among TIME subtypes were further conducted via KEGG pathway enrichment using the compareCluster function (fun = "enrichKEGG"), with visualization of the top five pathways for each subtype.

Gene set scoring

Based on 24 published immune-related gene sets, we used the AddModuleScore function from the R package Seurat to score each single cell in the TIME, calculating the average score for all cells within each TIME subtype, and used the Kruskal-Wallis test to assess the statistical significance of the differences.

To assess the GEs status of cancer cells, we also used the AddModuleScore function, scoring the cancer cells identified by CopyKat using 13 GEs gene sets and defining GEs labels for each cancer cell.

To analyze the association between the 13 cancer cell GEs and the TIME subtypes, as well as NSCLC subtypes with different driver mutations, the AddModuleScore function was again applied to score the single cells, calculating the average score for all cells in each TIME subtype and each driver mutation subtype. Statistical significance was assessed using analysis of variance (ANOVA).

Determination of dominant cellular modules in patients

For each NSCLC patient, the normalized frequency of cell clusters originating from the same TIME subtype was calculated. The summed frequencies across the five TIME subtypes were subsequently normalized, and the subtype with the highest frequency was designated as the dominant TIME type for the patient.

Cancer cell state identification

Raw counts of epithelial cells were analyzed for copy number variation (CNV) using CopyKat (v1.0.8), with a minimum of five genes per chromosome and clustering based on Euclidean distance. Cells were classified into malignant and non-malignant subpopulations according to their CNV patterns, with malignant epithelial cells defined as cancer cells. The cancer cells were preprocessed using NormalizeData, FindVariableFeatures, and ScaleData functions, and patient samples with fewer than 50 cells were excluded. Principal component analysis (PCA) and dimensionality reduction were subsequently performed. To capture both intra- and inter-tumoral heterogeneity, multi-resolution unsupervised clustering was conducted with resolutions ranging from 0.01 to 2. Upregulated differentially expressed genes (DEGs) were identified for clusters at each resolution using FindAllMarkers, and DEGs across all resolutions were merged. Genes with mitochondrial genes, adjusted $P \geq 0.05$, log2FoldChange ≤ 0, or expressed in $\leq 20\%$ of cells were excluded. For each cluster, the top 200 marker genes ranked by adjusted P value and log2FoldChange were selected. A similarity matrix of cluster marker gene sets was constructed based on the Jaccard index, removing gene sets with fewer than 20 genes and eliminating redundancy with similarity > 0.95. Consensus clustering using the R package cola2.12.0 was then applied to the deduplicated Jaccard matrix to identify GEs, with ATC as the feature selection method and skmeans as the clustering algorithm. The optimal number of clusters was determined to be 13, resulting in 13 cancer cell state-related GEs.

CNV level analysis

To evaluate gene-level CNV in each GEs cancer cell population, infercnv (v1.22.0) was applied using DC and NK cells as references. CNV matrices of cancer cells identified by CopyKat and genomic annotation

were extracted. For cancer cells labeled with GEs, the mean CNV signal of each gene across GEs states was calculated, reflecting amplification (positive values) or deletion (negative values). Genes with the most significant amplification or deletion were selected based on genomic position and visualized using the R package circlize (v0.4.16).

Survival analysis

To assess the prognostic value of TIME subtypes, the top 30 marker genes identified from single-cell analysis were extracted from bulk RNA-seq of 386 NSCLC patients. The mean expression of each module per patient was calculated as the module score. Optimal cutoffs were determined using surv_cutpoint (survmine v0.5.0) to classify patients into high-expression and low-expression groups. Kaplan-Meier curves were constructed, survival differences assessed using survfit (survival v3.8-3), and curves with risk tables plotted via ggsurvplot.

For DMSP.sig, survival analyses were performed separately in the training cohort (641 patients) and validation cohort (276 patients), with median risk scores dividing patients into high- and low-risk groups, followed by Kaplan-Meier analysis to evaluate prognostic performance.

Cellular interaction analysis

CellChat (v2.2.0) was employed to analyze interactions mediated by invasive subtypes CM2 and CM5, using the CellChatDB v2 ligand-receptor database to assess intercellular communication weights. CM2 and its corresponding cancer cells were merged into a Seurat object, excluding clusters with fewer than 50 cells; the same procedure was applied for CM5. Communication probabilities between clusters were quantified based on the mean expression of ligands, receptors, and cofactors. Network centrality metrics were computed using netAnalysis_computeCentrality, and signaling input/output patterns were visualized with netAnalysis_signalingRole_heatmap. Key ligand-receptor interactions were illustrated using netVisual_chord_gene.

Metabolic analysis

The R package scMetabolism (version 0.2.1) was used to analyze the metabolic pathway activity of the invasive subtypes CM2 and CM5, as well as their corresponding cancer cell populations (CM2.epi and CM5.epi). Based on the KEGG metabolic pathway database, the AUCell algorithm was used to calculate the activity scores of 85 core metabolic pathways to assess the overall metabolic differences in the four groups. Additionally, the activity of the same set of metabolic pathways was further evaluated at the cell cluster level within these groups to reveal the metabolic heterogeneity between cell clusters. Statistical significance was assessed using the non-parametric Mann-Whitney U test, and false discovery rate (FDR) correction was performed using the Benjamini-Hochberg method.

HR analysis

To assess the prognostic risk of key cell clusters within the 64 TIME subpopulations, the top 10 marker genes for each subgroup (after removing aberrant or redundant gene sets) were used, and gene set

scores were calculated for 917 tumor samples from TCGA LUAD and LUSC cohorts using the R package GSVA (v2.0.7). Combined with patient survival data, a Cox proportional hazards regression model (with age as a covariate) was used to calculate the HR and 95% confidence intervals for each gene set in LUAD and LUSC, with statistical significance assessed via Wald tests. Ultimately, all results were classified based on the HR direction into "Worse survival" (HR > 1) and "Better survival" (HR < 1).

For the HR analysis of regulons (TFs and target genes), the same method was applied. A regulon was broken down into gene sets containing 2–3 elements, and their prognostic value was assessed in TCGA LUAD and LUSC.

TF identification

Key TFs in CM2, CM5, and their corresponding cancer cells were identified using run_ulm (decoupleR v2.12.0), selecting factors with high subtype-specific activity. Chi-square tests (chisq.test) were applied to compare TF specificity between CM2 and CM5, and factors were ranked by chi-square value to determine subtype-specific TFs.

Prognostic model construction

A prognostic model for NSCLC was built based on malignant regulatory networks in CM2 and CM5, including TFs and their target genes. GSVA (v2.0.7) computed enrichment scores of 63 regulons across 917 LUAD and LUSC patients, generating a candidate scoring matrix. Samples were randomly split 7:3 into training (n = 641) and validation (n = 276) sets. In the training cohort, univariate Cox regression identified 14 significant regulons (P < 0.05), followed by LASSO Cox regression (glmnet v4.1-8) to select nine regulons and construct the final model. Risk scores (RS) were calculated for training and validation cohorts, with Kaplan-Meier curves and time-dependent ROC curves assessing 1-, 2-, and 3-year predictive performance. Model risk scores were further correlated with survival status using ggrisk (v1.3).

HPA immunohistochemistry validation

IHC data were downloaded from the HPA (http://www.proteinatlas.org) to evaluate the expression levels (high or low) of the genes or transcription factors contained within the nine regulons in both normal lung tissue and lung cancer tissue.

Immune infiltration analysis

Immune infiltration analysis was performed using TCGAplot (v8.0.0) for 917 LUAD and LUSC samples. Immune cell proportion matrices were derived using get_immu_ratio(), matched with GSVA scores, and expression of nine regulons. Pearson correlation coefficients between regulons/RS and immune cell infiltration were calculated using corr.test(), with significance levels indicated as *P < 0.05, **P < 0.01, ***P < 0.001. Correlation heatmaps were visualized with ggplot2.

Drug sensitivity prediction and molecular docking of core targets

Drug sensitivity was predicted using the calcPhenotype function of the R package oncoPredict (v1.2) based on bulk RNA-seq data. Predictions were generated using ridge regression models trained on pharmacogenomic datasets integrating gene expression and drug response data from the GDSC2 and CTRP v2 databases, as curated by the oncoPredict development team (https://osf.io/c6tfx/).

Structural data of candidate compounds and core target proteins were obtained from the PubChem and Protein Data Bank (PDB) databases, respectively (https://www.rcsb.org/). Molecular docking was performed using CB-Dock2, which automatically preprocessed ligands and protein structures, including hydrogen addition, removal of water molecules, and energy minimization. Docking was carried out using AutoDock Vina, and the binding pose with the lowest Vina score was selected as the representative conformation [35609983]. Three-dimensional and two-dimensional representations of the ligand–receptor complexes were visualized using Discovery Studio 2025.

Statistical analysis

Statistical analysis was performed using R (v.4.4.1). Comparisons between groups were conducted using $\chi 2$ tests or Fisher's exact test for categorical variables. Student's t-tests, Wilcoxon rank-sum tests, and ANOVA were used for continuous variables. Paired t-tests were used for paired comparisons. Survival analyses were conducted using log-rank tests. $P < 0.05$ was considered to be statistically significant.

Conclusions

In conclusion, this study systematically delineates the mutation-specific molecular features of TIME, identifies CM2 and CM5 as key invasive subtypes, and proposes the DMSP.sig as a quantitative tool for prognostic prediction and individualized therapy. These findings deepen our understanding of immune evolution in driver-mutant NSCLC and provide a solid rationale for the development of precision immunotherapeutic strategies.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

Kuan Yang[1, #], Jiasi Wang[1, #], Wenqi Jiang[1], Hang Zhao[1], Depeng Mu[1], Xiao Peng[1], Yiming Yan[1], Xing Gao[1], Jing Bai[1, 2], Xia Li[1,*], Yunpeng Zhang [1,*], Congxue Hu [1,*]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China, 150081

[2]College of Biomedical Information and Engineering, Hainan Medical University, Haikou, PR China

Funding

This work was supported by grants from the National Science and Technology Major Program [2024ZD0530500]; and National Natural Science Foundation of China [62472131, U23A20166, 32570792, 62502128, 62172131], and the Key Research and Development Program of Heilongjiang Province[2024ZX12C27], and the China Postdoctoral Science Foundation [2024M760709], and Heilongjiang Postdoctoral Fund [LBH-Z24210], and Longjiang New Era Outstanding Doctoral Dissertation Project Grant [LJYXL2024-069].

Author contributions

Y.P.Z., C.X.H., and X.L. inputted conceptualization. Y.P.Z., C.X.H., and X.L. designed experiments. K.Y., J.S.W., W.Q.J., H.Z., and D.P.M. contributed to the data curation. J.S.W. K.Y., W.Q.J., and performed scRNA-seq data analysis. J.S.W., K.Y., W.Q.J., H.Z., D.P.M., X.P., Y.M.Y. and XG. performed visualization and interpreted the results. Y.P.Z. provided resources. C.X.H. and K.Y. contributed to the software. Y.P.Z., C.X.H., and X.L. supervised all work. Y.P.Z., C.X.H., and X.L. contributed to the funding acquisition. Y.P.Z., K.Y., J.S.W., W.Q.J., H.Z., and D.P.M. prepared the original manuscript. Y.P.Z., C.X.H., X.L., K.Y., W.Q.J., and J.S.W. revised the manuscript. All coauthors read, reviewed, and approved the manuscript.

Acknowledgments

We thank the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) Database for sharing a large amount of data.

Data availability

All data used in this study are publicly available from GEO (https://ncbi.nlm.nih.gov/geo/) and TCGA (https://portal.gdc.cancer.gov/). The datasets analyzed are existing publicly accessible datasets, and detailed information is provided in Table 1.

Code availability

R and other custom scripts for analyzing data are available upon request.

References

1. Wu, J., Lin, Z.: Non-Small Cell Lung Cancer Targeted Therapy: Drugs and Mechanisms of Drug Resistance. Int. J. Mol. Sci., 23 (2022)
2. Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F., Wong, K.K.: Non-small-cell lung cancers: a heterogeneous set of diseases. Nat. Rev. Cancer. **14**, 535–546 (2014)
3. N. Cancer Genome Atlas Research, Comprehensive molecular profiling of lung adenocarcinoma. Nature, **511** 543–550. (2014)
4. Tan, A.C., Tan, D.S.W.: Targeted Therapies for Lung Cancer Patients With Oncogenic Driver Molecular Alterations. J. Clin. Oncol. **40**, 611–625 (2022)
5. Schildgen, V., Schildgen, O.: The lonely driver or the orchestra of mutations? How next generation sequencing datasets contradict the concept of single driver checkpoint mutations in solid tumours - NSCLC as a scholarly example. Semin Cancer Biol. **58**, 22–28 (2019)
6. Shen, Z., Teng, M., Han, L., Bian, D., Zhang, J., Zhu, X., Qing, Y., Hu, S., Chen, Y., Yao, W., Yu, H., Zhang, L., Zhang, P.: The impact of oncogenic driver mutations on neoadjuvant immunotherapy outcomes in patients with resectable non-small cell lung cancer. Cancer Immunol. Immunother. **72**, 4235–4247 (2023)
7. Feng, J., Hueniken, K., Fan, Z.J., Zhan, L.J., Faour, E., Corke, L., Alghabban, A., Leighl, N.B., Liu, G., Bradbury, P.A., Sacher, A., Eng, L., Stockley, T.L., Tsao, M.S., Shepherd, F.A.: Prognostic and predictive effects of TP53 co-mutation in patients with non-small cell lung cancer with rare treatable driver mutations. Lung Cancer. **204**, 108452 (2025)
8. Yeung, S.F., Tong, J.H.M., Law, P.P.W., Chung, L.Y., Lung, R.W.M., Tong, C.Y.K., Chow, C., Chan, A.W.H., Wan, I.Y.P., Mok, T.S.K., To, K.F.: Profiling of Oncogenic Driver Events in Lung Adenocarcinoma Revealed MET Mutation as Independent Prognostic Factor. J. Thorac. Oncol. **10**, 1292–1300 (2015)
9. Akbay, E.A., Koyama, S., Carretero, J., Altabef, A., Tchaicha, J.H., Christensen, C.L., Mikse, O.R., Cherniack, A.D., Beauchamp, E.M., Pugh, T.J., Wilkerson, M.D., Fecci, P.E., Butaney, M., Reibel, J.B., Soucheray, M., Cohoon, T.J., Janne, P.A., Meyerson, M., Hayes, D.N., Shapiro, G.I., Shimamura, T., Sholl, L.M., Rodig, S.J., Freeman, G.J., Hammerman, P.S., Dranoff, G., Wong, K.K.: Activation of the PD-1 pathway contributes to immune escape in EGFR-driven lung tumors. Cancer Discov. **3**, 1355–1363 (2013)
10. Xu, C., Fillmore, C.M., Koyama, S., Wu, H., Zhao, Y., Chen, Z., Herter-Sprie, G.S., Akbay, E.A., Tchaicha, J.H., Altabef, A., Reibel, J.B., Walton, Z., Ji, H., Watanabe, H., Janne, P.A., Castrillon, D.H., Rustgi, A.K., Bass, A.J., Freeman, G.J., Padera, R.F., Dranoff, G., Hammerman, P.S., Kim, C.F., Wong, K.K.: Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. Cancer Cell. **25**, 590–604 (2014)

11. Liu, Z., Yang, Z., Wu, J., Zhang, W., Sun, Y., Zhang, C., Bai, G., Yang, L., Fan, H., Chen, Y., Zhang, L., Jiang, B., Liu, X., Ma, X., Tang, W., Liu, C., Qu, Y., Yan, L., Zhao, D., Wu, Y., He, S., Xu, L., Peng, L., Chen, X., Zhou, B., Zhao, L., Zhao, Z., Tan, F., Zhang, W., Yi, D., Li, X., Gao, Q., Zhang, G., Wang, Y., Yang, M., Fu, H., Guo, Y., Hu, X., Cai, Q., Qi, L., Bo, Y., Peng, H., Tian, Z., She, Y., Zou, C., Zhu, L., Cheng, S., Zhang, Y., Zhong, W., Chen, C., Gao, S., Zhang, Z.: A single-cell atlas reveals immune heterogeneity in anti-PD-1-treated non-small cell lung cancer. Cell. **188**, 3081–3096e3019 (2025)
12. Zhao, Y., Zhang, Q., Tu, K., Chen, Y., Peng, Y., Ni, Y., Zhu, G., Cheng, C., Li, Y., Xiao, X., Yu, C., Lu, K., Chen, Y., Li, C., Tang, J., Wang, G., Luo, W., Zhang, W., Che, G., Li, W., Wang, Z., Xie, D.: Single-Cell Transcriptomics of Immune Cells Reveal Diversity and Exhaustion Signatures in Non-Small-Cell Lung Cancer. Front. Immunol. **13**, 854724 (2022)
13. Galffy, G., Morocz, E., Korompay, R., Hecz, R., Bujdoso, R., Puskas, R., Lovas, T., Gaspar, E., Yahya, K., Kiraly, P., Lohinai, Z.: Targeted therapeutic options in early and metastatic NSCLC-overview. Pathol. Oncol. Res. **30**, 1611715 (2024)
14. Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., Vonderheide, R.H., Pittet, M.J., Jain, R.K., Zou, W., Howcroft, T.K., Woodhouse, E.C., Weinberg, R.A., Krummel, M.F.: Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat. Med. **24**, 541–550 (2018)
15. Tagore, S., Caprio, L., Amin, A.D., Bestak, K., Luthria, K., D'Souza, E., Barrera, I., Melms, J.C., Wu, S., Abuzaid, S., Wang, Y., Jakubikova, V., Koch, P., Brodtman, D.Z., Bawa, B., Deshmukh, S.K., Ebel, L., Ibarra-Arellano, M.A., Jaiswal, A., Gurjao, C., Biermann, J., Shaikh, N., Ramaradj, P., Georgis, Y., Lagos, G.G., Ehrlich, M.I., Ho, P., Walsh, Z.H., Rogava, M., Politis, M.G., Biswas, D., Cottarelli, A., Rizvi, N., Shu, C.A., Herzberg, B., Anandasabapathy, N., Sledge, G., Zorn, E., Canoll, P., Bruce, J.N., Rizvi, N.A., Taylor, A.M., Saqi, A., Hibshoosh, H., Schwartz, G.K., Henick, B.S., Chen, F., Schapiro, D., Shah, P., Izar, B.: Single-cell and spatial genomic landscape of non-small cell lung cancer brain metastases. Nat. Med. **31**, 1351–1363 (2025)
16. Tian, Y., Li, Q., Yang, Z., Zhang, S., Xu, J., Wang, Z., Bai, H., Duan, J., Zheng, B., Li, W., Cui, Y., Wang, X., Wan, R., Fei, K., Zhong, J., Gao, S., He, J., Gay, C.M., Zhang, J., Wang, J., Tang, F.: Single-cell transcriptomic profiling reveals the tumor heterogeneity of small-cell lung cancer. Signal. Transduct. Target. Ther. **7**, 346 (2022)
17. Watabe, T., Takahashi, K., Pietras, K., Yoshimatsu, Y.: Roles of TGF-beta signals in tumor microenvironment via regulation of the formation and plasticity of vascular system. Semin Cancer Biol. **92**, 130–138 (2023)
18. Zeng, Y., Zhao, J., Wu, Z., Huang, Y., Wang, A., Zhu, J., Xu, M., Zhang, W., Zhang, X., Li, J., Huang, J.A., Liu, Z.: Targeting TYK2 alleviates Rab27A-induced malignant progression of non-small cell lung cancer via disrupting IFNalpha-TYK2-STAT-HSPA5 axis. NPJ Precis Oncol. **8**, 74 (2024)
19. Patel, A.J., Richter, A., Drayson, M.T., Middleton, G.W.: The role of B lymphocytes in the immuno-biology of non-small-cell lung cancer. Cancer Immunol. Immunother. **69**, 325–342 (2020)

20. Zheng, S., Wang, W., Shen, L., Yao, Y., Xia, W., Ni, C.: Tumor battlefield within inflamed, excluded or desert immune phenotypes: the mechanisms and strategies. Exp. Hematol. Oncol. **13**, 80 (2024)
21. Wang, C., Yu, Q., Song, T., Wang, Z., Song, L., Yang, Y., Shao, J., Li, J., Ni, Y., Chao, N., Zhang, L., Li, W.: The heterogeneous immune landscape between lung adenocarcinoma and squamous carcinoma revealed by single-cell RNA sequencing. Signal. Transduct. Target. Ther. **7**, 289 (2022)
22. Zanetti, M., Xian, S., Dosset, M., Carter, H.: The Unfolded Protein Response at the Tumor-Immune Interface. Front. Immunol. **13**, 823157 (2022)
23. Huang, K., Zhang, Y., Shi, X., Yin, Z., Zhao, W., Huang, L., Wang, F., Zhou, X.: Cell-type-specific alternative polyadenylation promotes oncogenic gene expression in non-small cell lung cancer progression. Mol. Ther. Nucleic Acids. **33**, 816–831 (2023)
24. Rocha, B.G.S., Picoli, C.C., Goncalves, B.O.P., Silva, W.N., Costa, A.C., Moraes, M.M., Costa, P.A.C., Santos, G.S.P., Almeida, M.R., Silva, L.M., Singh, Y., Falchetti, M., Guardia, G.D.A., Guimaraes, P.P.G., Russo, R.C., Resende, R.R., Pinto, M.C.X., Amorim, J.H., Azevedo, V.A.C., Kanashiro, A., Nakaya, H.I., Rocha, E.L., Galante, P.A.F., Mintz, A., Frenette, P.S., Birbrair, A.: Tissue-resident glial cells associate with tumoral vasculature and promote cancer progression. Angiogenesis. **26**, 129–166 (2023)
25. Yang, Q., Zhang, H., Wei, T., Lin, A., Sun, Y., Luo, P., Zhang, J.: Single-Cell RNA Sequencing Reveals the Heterogeneity of Tumor-Associated Macrophage in Non-Small Cell Lung Cancer and Differences Between Sexes. Front. Immunol. **12**, 756722 (2021)
26. Kimura, S., Yokoyama, S., Pilon, A.L., Kurotani, R.: Emerging role of an immunomodulatory protein secretoglobin 3A2 in human diseases. Pharmacol. Ther. **236**, 108112 (2022)
27. Yang, Y., Xu, S., Jia, G., Yuan, F., Ping, J., Guo, X., Tao, R., Shu, X.O., Zheng, W., Long, J., Cai, Q.: Integrating genomics and proteomics data to identify candidate plasma biomarkers for lung cancer risk among European descendants. Br. J. Cancer. **129**, 1510–1515 (2023)
28. Dai, H., Yang, X., Sheng, X., Wang, Y., Zhang, S., Zhang, X., Hu, L., Zhang, Z., Dong, X., Yin, W., Yao, L., Lu, J.: XPOT Disruption Suppresses TNBC Growth through Inhibition of Specific tRNA Nuclear Exportation and TTC19 Expression to Induce Cytokinesis Failure. Int. J. Biol. Sci. **19**, 5319–5336 (2023)
29. Cristiano, L.: The pseudogenes of eukaryotic translation elongation factors (EEFs): Role in cancer and other human diseases. Genes Dis. **9**, 941–958 (2022)
30. Xie, P., Yan, H., Gao, Y., Li, X., Zhou, D.B., Liu, Z.Q.: Construction of m6A-Related lncRNA Prognostic Signature Model and Immunomodulatory Effect in Glioblastoma Multiforme. Front. Oncol. **12**, 920926 (2022)
31. Song, S.G., Kim, S., Koh, J., Yim, J., Han, B., Kim, Y.A., Jeon, Y.K., Chung, D.H.: Comparative analysis of the tumor immune-microenvironment of primary and brain metastases of non-small-cell lung cancer reveals organ-specific and EGFR mutation-dependent unique immune landscape. Cancer Immunol. Immunother. **70**, 2035–2048 (2021)
32. Nishida, N.: Role of Oncogenic Pathways on the Cancer Immunosuppressive Microenvironment and Its Clinical Implications in Hepatocellular Carcinoma. Cancers (Basel), 13 (2021)

33. Ozga, A.J., Chow, M.T., Luster, A.D.: Chemokines and the immune response to cancer. Immunity. **54**, 859–874 (2021)
34. Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., Zhang, Y., Zhao, W., Zhou, F., Li, W., Zhang, J., Zhang, X., Qiao, M., Gao, G., Chen, S., Chen, X., Li, X., Hou, L., Wu, C., Su, C., Ren, S., Odenthal, M., Buettner, R., Fang, N., Zhou, C.: Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. Nat. Commun. **12**, 2540 (2021)
35. J, J.S., Boyle, T.A.: Molecular Pathology of Lung Cancer. Cold Spring Harb Perspect. Med., **12** (2022)
36. Imyanitov, E.N., Iyevleva, A.G., Levchenko, E.V.: Molecular testing and targeted therapy for non-small cell lung cancer: Current status and perspectives. Crit. Rev. Oncol. Hematol. **157**, 103194 (2021)
37. Castellanos, E., Feld, E., Horn, L.: Driven by Mutations: The Predictive Value of Mutation Subtype in EGFR-Mutated Non-Small Cell Lung Cancer. J. Thorac. Oncol. **12**, 612–623 (2017)
38. Frederick, M., Skinner, H.D., Kazi, S.A., Sikora, A.G., Sandulache, V.C.: High expression of oxidative phosphorylation genes predicts improved survival in squamous cell carcinomas of the head and neck and lung. Sci. Rep. **10**, 6380 (2020)
39. Conejo-Garcia, J.R., Biswas, S., Chaurio, R.: Humoral immune responses: Unsung heroes of the war on cancer. Semin Immunol. **49**, 101419 (2020)
40. Huang, H., Zhu, X., Yu, Y., Li, Z., Yang, Y., Xia, L., Lu, S.: EGFR mutations induce the suppression of CD8(+) T cell and anti-PD-1 resistance via ERK1/2-p90RSK-TGF-beta axis in non-small cell lung cancer. J. Transl Med. **22**, 653 (2024)
41. Gwynne, W.D., Suk, Y., Custers, S., Mikolajewicz, N., Chan, J.K., Zador, Z., Chafe, S.C., Zhai, K., Escudero, L., Zhang, C., Zaslaver, O., Chokshi, C., Shaikh, M.V., Bakhshinyan, D., Burns, I., Chaudhry, I., Nachmani, O., Mobilio, D., Maich, W.T., Mero, P., Brown, K.R., Quaile, A.T., Venugopal, C., Moffat, J., Montenegro-Burke, J.R., Singh, S.K.: Cancer-selective metabolic vulnerabilities in MYC-amplified medulloblastoma. Cancer Cell. **40**, 1488–1502e1487 (2022)
42. Mullen, N.J., Singh, P.K.: Nucleotide metabolism: a pan-cancer metabolic dependency. Nat. Rev. Cancer. **23**, 275–294 (2023)
43. Ye, L., Wen, X., Qin, J., Zhang, X., Wang, Y., Wang, Z., Zhou, T., Di, Y., He, W.: Metabolism-regulated ferroptosis in cancer progression and therapy. Cell. Death Dis. **15**, 196 (2024)
44. Poltorack, C.D., Dixon, S.J.: Understanding the role of cysteine in ferroptosis: progress & paradoxes. FEBS J. **289**, 374–385 (2022)
45. Renner, K., Singer, K., Koehl, G.E., Geissler, E.K., Peter, K., Siska, P.J., Kreutz, M.: Metabolic Hallmarks of Tumor and Immune Cells in the Tumor Microenvironment. Front. Immunol. **8**, 248 (2017)
46. Chen, Z., Zhou, X., Zhou, X., Tang, Y., Lu, M., Zhao, J., Tian, C., Wu, M., Liu, Y., Prochownik, E.V., Wang, F., Li, Y.: Phosphomevalonate Kinase Controls beta-Catenin Signaling via the Metabolite 5-Diphosphomevalonate. Adv. Sci. (Weinh). **10**, e2204909 (2023)
47. Khumukcham, S.S., Manavathi, B.: Two decades of a protooncogene HPIP/PBXIP1: Uncovering the tale from germ cell to cancer. Biochim. Biophys. Acta Rev. Cancer. **1876**, 188576 (2021)

48. Zhu, Y., Zhao, Y., Wen, J., Liu, S., Huang, T., Hatial, I., Peng, X., Al Janabi, H., Huang, G., Mittlesteadt, J., Cheng, M., Bhardwaj, A., Ashfeld, B.L., Kao, K.R., Maeda, D.Y., Dai, X., Wiest, O., Blagg, B.S.J., Lu, X., Cheng, L., Wan, J., Lu, X.: Targeting the chromatin effector Pygo2 promotes cytotoxic T cell responses and overcomes immunotherapy resistance in prostate cancer. Sci. Immunol. **8**, eade4656 (2023)
49. Yuan, W., Zhao, H., Zhou, A., Wang, S.: Interference of EFNA4 suppresses cell proliferation, invasion and angiogenesis in hepatocellular carcinoma by downregulating PYGO2. Cancer Biol. Ther. **23**, 1–12 (2022)
50. Ling, J., Tang, Z., Yang, W., Li, Y., Dong, X.: Pygo2 activates BRPF1 via Pygo2-H3K4me2/3 interaction to maintain malignant progression in colon cancer. Exp. Cell. Res. **431**, 113696 (2023)
51. Ren, J., Zhao, S., Lai, J.: Triptolide restrains the growth, invasion, stemness, and glycolysis of non-small cell lung cancer cells by PFKFB2-mediated PI3K/AKT pathway. Chem. Biol. Drug Des. **103**, e14450 (2024)
52. Friedlaender, A., Drilon, A., Banna, G.L., Peters, S., Addeo, A.: The METeoric rise of MET in lung cancer. Cancer. **126**, 4826–4837 (2020)
53. Helfman, D.M., Kim, E.J., Lukanidin, E., Grigorian, M.: The metastasis associated protein S100A4: role in tumour progression and metastasis. Br. J. Cancer. **92**, 1955–1958 (2005)
54. Mao, X., Xu, J., Wang, W., Liang, C., Hua, J., Liu, J., Zhang, B., Meng, Q., Yu, X., Shi, S.: Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol. Cancer. **20**, 131 (2021)
55. Sultana, A., Alam, M.S., Khanam, A., Liang, H.: Unraveling the molecular landscape of non-small cell lung cancer: Integrating bioinformatics and statistical approaches to identify biomarkers and drug repurposing. Comput. Biol. Med. **187**, 109744 (2025)
56. Liang, Y., Xia, W., Zhang, T., Chen, B., Wang, H., Song, X., Zhang, Z., Xu, L., Dong, G., Jiang, F.: Upregulated Collagen COL10A1 Remodels the Extracellular Matrix and Promotes Malignant Progression in Lung Adenocarcinoma. Front. Oncol. **10**, 573534 (2020)
57. Wu, J., Ouyang, P., Huang, R., Cui, Y., Yang, Z., Xu, W., Ma, R., Xiang, G., Zeng, W., Wu, W., Li, J.: METTL16 Promotes Stability of SYNPO2L mRNA and leading to Cancer Cell Lung Metastasis by Secretion of COL10A1 and attract the Cancer-Associated Fibroblasts. Int. J. Biol. Sci. **20**, 4128–4145 (2024)
58. Wang, Y., Wang, E., Anany, M., Fullsack, S., Huo, Y.H., Dutta, S., Ji, B., Hoeppner, L.H., Kilari, S., Misra, S., Caulfield, T., Vander Kooi, C.W., Wajant, H., Mukhopadhyay, D.: The crosstalk between neuropilin-1 and tumor necrosis factor-alpha in endothelial cells. Front. Cell. Dev. Biol. **12**, 1210944 (2024)
59. Zhang, S., Wang, H., Xu, Z., Bai, Y., Xu, L.: Lymphatic Metastasis of NSCLC Involves Chemotaxis Effects of Lymphatic Endothelial Cells through the CCR7-CCL21 Axis Modulated by TNF-alpha. Genes (Basel), 11 (2020)
60. Rosberg, R., Smolag, K.I., Sjolund, J., Johansson, E., Bergelin, C., Wahlden, J., Pantazopoulou, V., Ceberg, C., Pietras, K., Blom, A.M., Pietras, A.: Hypoxia-induced complement component 3 promotes

aggressive tumor growth in the glioblastoma microenvironment. JCI Insight, 9 (2024)

61. Wei, J., Chen, Z., Hu, M., He, Z., Jiang, D., Long, J., Du, H.: Characterizing Intercellular Communication of Pan-Cancer Reveals SPP1 + Tumor-Associated Macrophage Expanded in Hypoxia and Promoting Cancer Malignancy Through Single-Cell RNA-Seq Data. Front. Cell. Dev. Biol. **9**, 749210 (2021)
62. Chen, Z., Hu, J., Ou, Y., Ye, F., Li, W., Liu, S., Jiang, H.: Integrating single-cell transcriptomics to reveal the ferroptosis regulators in the tumor microenvironment that contribute to bladder urothelial carcinoma progression and immunotherapy. Front. Immunol. **15**, 1427124 (2024)
63. Yan, L., Wen, Z., Yang, Y., Liu, A., Li, F., Zhang, Y., Yang, C., Li, Y., Zhang, Y.: Dissecting the roles of prosaposin as an emerging therapeutic target for tumors and its underlying mechanisms. Biomed. Pharmacother. **180**, 117551 (2024)
64. Xie, Y., Peng, H., Hu, Y., Jia, K., Yuan, J., Liu, D., Li, Y., Feng, X., Li, J., Zhang, X., Sun, Y., Shen, L., Chen, Y.: Immune microenvironment spatial landscapes of tertiary lymphoid structures in gastric cancer. BMC Med. **23**, 59 (2025)
65. Lu, S., Wei, X., Tao, L., Dong, D., Hu, W., Zhang, Q., Tao, Y., Yu, C., Sun, D., Cheng, H.: A novel tRNA-derived fragment tRF-3022b modulates cell apoptosis and M2 macrophage polarization via binding to cytokines in colorectal cancer. J. Hematol. Oncol. **15**, 176 (2022)
66. Liu, P.I., Jiang, Y.J., Chang, A.C., Huang, C.L., Fong, Y.C., Guo, J.H., Liu, C.L., Wang, S.W., Liu, J.F., Chang, S.L., Tang, C.H.: ANGPTL2 promotes VEGF-A synthesis in human lung cancer and facilitates lymphangiogenesis. Aging (Albany NY). **15**, 1652–1667 (2023)
67. Zhang, Y., Liu, X., Zeng, L., Zhao, X., Chen, Q., Pan, Y., Bai, Y., Shao, C., Zhang, J.: Exosomal protein angiopoietin-like 4 mediated radioresistance of lung cancer by inhibiting ferroptosis under hypoxic microenvironment. Br. J. Cancer. **127**, 1760–1772 (2022)
68. Zheng, G., Bouamar, H., Cserhati, M., Zeballos, C.R., Mehta, I., Zare, H., Broome, L., Hu, R., Lai, Z., Chen, Y., Sharkey, F.E., Rani, M., Halff, G.A., Cigarroa, F.G., Sun, L.Z.: Integrin alpha 6 is upregulated and drives hepatocellular carcinoma progression through integrin alpha6beta4 complex. Int. J. Cancer. **151**, 930–943 (2022)
69. Gong, S., Qiao, H., Wang, J.Y., Huang, S.Y., He, S.W., Zhao, Y., Tan, X.R., Ye, M.L., Li, J.Y., Liang, Y.L., Huang, S.W., Chen, J., Zhu, X.H., Liu, N., Li, Y.Q.: Ac4C modification of lncRNA SIMALR promotes nasopharyngeal carcinoma progression through activating eEF1A2 to facilitate ITGB4/ITGA6 translation. Oncogene. **43**, 2868–2884 (2024)
70. Ishihara, S., Aoki, K., Mizutani, T., Amano, M., Nishimura, S.I., Haga, H.: Glycosphingolipid GM2 Induces Invasiveness in Irradiation-tolerant Lung Cancer Cells. Cell. Struct. Funct. **43**, 177–185 (2018)
71. Han, J., Du, J., Li, X., Zhou, Q., Zeng, J., Lin, J.T., Zhou, W., Cai, J., Ye, Y., Yang, B., Wang, J., Zhou, X., Lian, R., Yang, Y., Zhu, X., Guan, H., Liu, L., Cai, J., Wu, J., Li, Y., Li, M., Tian, H.: The Glycosyltransferase XYLT1 Activates NF-kappaB Signaling to Promote Metastasis of Early-Stage Lung Adenocarcinoma. Cancer Res. **85**, 1628–1643 (2025)

72. Nava Lauson, C.B., Tiberti, S., Corsetto, P.A., Conte, F., Tyagi, P., Machwirth, M., Ebert, S., Loffreda, A., Scheller, L., Sheta, D., Mokhtari, Z., Peters, T., Raman, A.T., Greco, F., Rizzo, A.M., Beilhack, A., Signore, G., Tumino, N., Vacca, P., McDonnell, L.A., Raimondi, A., Greenberg, P.D., Huppa, J.B., Cardaci, S., Caruana, I., Rodighiero, S., Nezi, L., Manzo, T.: Linoleic acid potentiates CD8(+) T cell metabolic fitness and antitumor immunity. Cell. Metab. **35**, 633–650e639 (2023)
73. Liu, Y., Liu, Z., Hu, Y., Ling, Y., Qing, S., Liu, Y., Zhan, Y., Shen, Z., Fang, Y., Deng, H.: Caffeine enhances antitumor T-cell activity by suppressing kynurenine pathway in colorectal cancer. Nat. Commun. **16**, 5906 (2025)
74. Li, Y., Azmi, A.S., Mohammad, R.M.: Deregulated transcription factors and poor clinical outcomes in cancer patients. Semin Cancer Biol. **86**, 122–134 (2022)
75. Chai, Y., Zhang, J., Shao, W., Zhang, Z.: Single-cell insights into HNSCC tumor heterogeneity and programmed cell death pathways. Transl Oncol. **54**, 102341 (2025)
76. Peng, L., Zhang, Z., Zhang, M., Yu, X., Yao, F., Tan, Y., Liu, D., Gong, D., Chong, H., Liu, X., Zheng, X., Tian, G., Tang, C.: Macrophage-activating lipopeptide-2 downregulates the expression of ATP-binding cassette transporter A1 by activating the TLR2/NF-small ka, CyrillicB/ZNF202 pathway in THP-1 macrophages. Acta Biochim. Biophys. Sin (Shanghai). **48**, 363–370 (2016)
77. Chintala, S., Katzenellenbogen, R.A.: NFX1, Its Isoforms and Roles in Biology, Disease and Cancer, Biology (Basel), 10 (2021)
78. Luk, I.Y., Reehorst, C.M., Mariadason, J.M.: ELF3, ELF5, EHF and SPDEF Transcription Factors in Tissue Homeostasis and Cancer, p. 23. Molecules (2018)
79. Li, T., Xu, L., Wei, Z., Zhang, S., Liu, X., Yang, Y., Gu, Y., Zhang, J.: ELF5 drives angiogenesis suppression though stabilizing WDTC1 in renal cell carcinoma. Mol. Cancer. **22**, 184 (2023)
80. Kiczmer, P., Senkowska, A.P., Kula, A., Dawidowicz, M., Strzelczyk, J.K., Zajdel, E.N., Walkiewicz, K., Waniczek, D., Ostrowska, Z., Swietochowska, E.: Assessment of CMKLR1 level in colorectal cancer and its correlation with angiogenic markers. Exp. Mol. Pathol. **113**, 104377 (2020)
81. Wang, D., Mahmud, I., Thakur, V.S., Tan, S.K., Isom, D.G., Lombard, D.B., Gonzalgo, M.L., Kryvenko, O.N., Lorenzi, P.L., Tcheuyap, V.T., Brugarolas, J.: Welford, GPR1 and CMKLR1 Control Lipid Metabolism to Support the Development of Clear Cell Renal Cell Carcinoma. Cancer Res. **84**, 2141–2154 (2024)
82. Sun, X., Xiao, C., Wang, X., Wu, S., Yang, Z., Sui, B., Song, Y.: Role of post-translational modifications of Sp1 in cancer: state of the art. Front. Cell. Dev. Biol. **12**, 1412461 (2024)
83. Han, F., Chen, S., Zhang, K., Zhang, K., Wang, M., Wang, P.: Single-cell transcriptomic sequencing data reveal aberrant DNA methylation in SMAD3 promoter region in tumor-associated fibroblasts affecting molecular mechanism of radiosensitivity in non-small cell lung cancer. J. Transl Med. **22**, 288 (2024)
84. Cui, C., Wang, J., Fagerberg, E., Chen, P.M., Connolly, K.A., Damo, M., Cheung, J.F., Mao, T., Askari, A.S., Chen, S., Fitzgerald, B., Foster, G.G., Eisenbarth, S.C., Zhao, H., Craft, J., Joshi, N.S.: Neoantigen-

driven B cell and CD4 T follicular helper cell collaboration promotes anti-tumor CD8 T cell responses. Cell. **184**, 6101–6118e6113 (2021)

85. Del Prete, A., Salvi, V., Soriani, A., Laffranchi, M., Sozio, F., Bosisio, D., Sozzani, S.: Dendritic cell subsets in cancer immunity and tumor antigen sensing. Cell. Mol. Immunol. **20**, 432–447 (2023)
86. Gao, J., Liang, Y., Wang, L.: Shaping Polarization Of Tumor-Associated Macrophages In Cancer Immunotherapy. Front. Immunol. **13**, 888713 (2022)
87. A. Society of Cancer Precision Medicine of Chinese Anti-Cancer, J. Lung Cancer Expert Group of Chinese Medical, [Chinese expert consensus on immunotherapy for advanced non-small lung cancer with oncogenic driver mutations (2023 edition)]. Zhonghua Zhong Liu Za Zhi, **45** 717–740. (2023)
88. Wang, H., Xiong, A., Chen, X., Guo, J., Tang, Z., Wu, C., Ren, S., Zhou, C., Chen, J., Hou, L., Jiang, T.: CXCR1(+) neutrophil infiltration orchestrates response to third-generation EGFR-TKI in EGFR mutant non-small-cell lung cancer. Signal. Transduct. Target. Ther. **9**, 342 (2024)
89. Liu, K.J., Li, H.R., Tan, Q.Q., Jiang, T., Peng, K.C., Chen, H.J., Zhou, Q., Zhang, X.C., Zheng, Z., Chen, S.Y., Zheng, X., Zheng, H.B., Mao, B.B., Gong, L.L., Chen, X.W., Wu, W., Wu, Y.L., Jia, J., Yang, J.J.: Tumor immune microenvironment of NSCLC with EGFR exon 20 insertions may predict efficacy of first-line ICI-combined regimen. Lung Cancer. **195**, 107933 (2024)
90. Xue, R., Zhang, Q., Cao, Q., Kong, R., Xiang, X., Liu, H., Feng, M., Wang, F., Cheng, J., Li, Z., Zhan, Q., Deng, M., Zhu, J., Zhang, Z., Zhang, N.: Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. Nature. **612**, 141–147 (2022)
91. Yang, L., Li, A., Lei, Q., Zhang, Y.: Tumor-intrinsic signaling pathways: key roles in the regulation of the immunosuppressive tumor microenvironment. J. Hematol. Oncol. **12**, 125 (2019)
92. Liberti, M.V., Locasale, J.W.: The Warburg Effect: How Does it Benefit Cancer Cells? Trends Biochem. Sci. **41**, 211–218 (2016)
93. Icard, P., Shulman, S., Farhat, D., Steyaert, J.M., Alifano, M., Lincet, H.: How the Warburg effect supports aggressiveness and drug resistance of cancer cells? Drug Resist. Updat. **38**, 1–11 (2018)
94. Wu, C., Yang, J., Xiao, W., Jiang, Z., Chen, S., Guo, D., Zhang, P., Liu, C., Yang, H., Xie, Z.: Single-cell characterization of malignant phenotypes and microenvironment alteration in retinoblastoma. Cell. Death Dis. **13**, 438 (2022)
95. Han, K., Wang, J., Qian, K., Zhao, T., Zhang, Y.: Establishment of non-small-cell lung cancer risk prediction model based on prognosis-associated ADME genes. Biosci. Rep., **41** (2021)
96. Zhang, Z., Zhang, P., Xie, J., Cui, Y., Shuo, W., Yue, D.: Five-gene prognostic model based on autophagy-dependent cell death for predicting prognosis in lung adenocarcinoma. Sci. Rep. **14**, 26449 (2024)

Figures

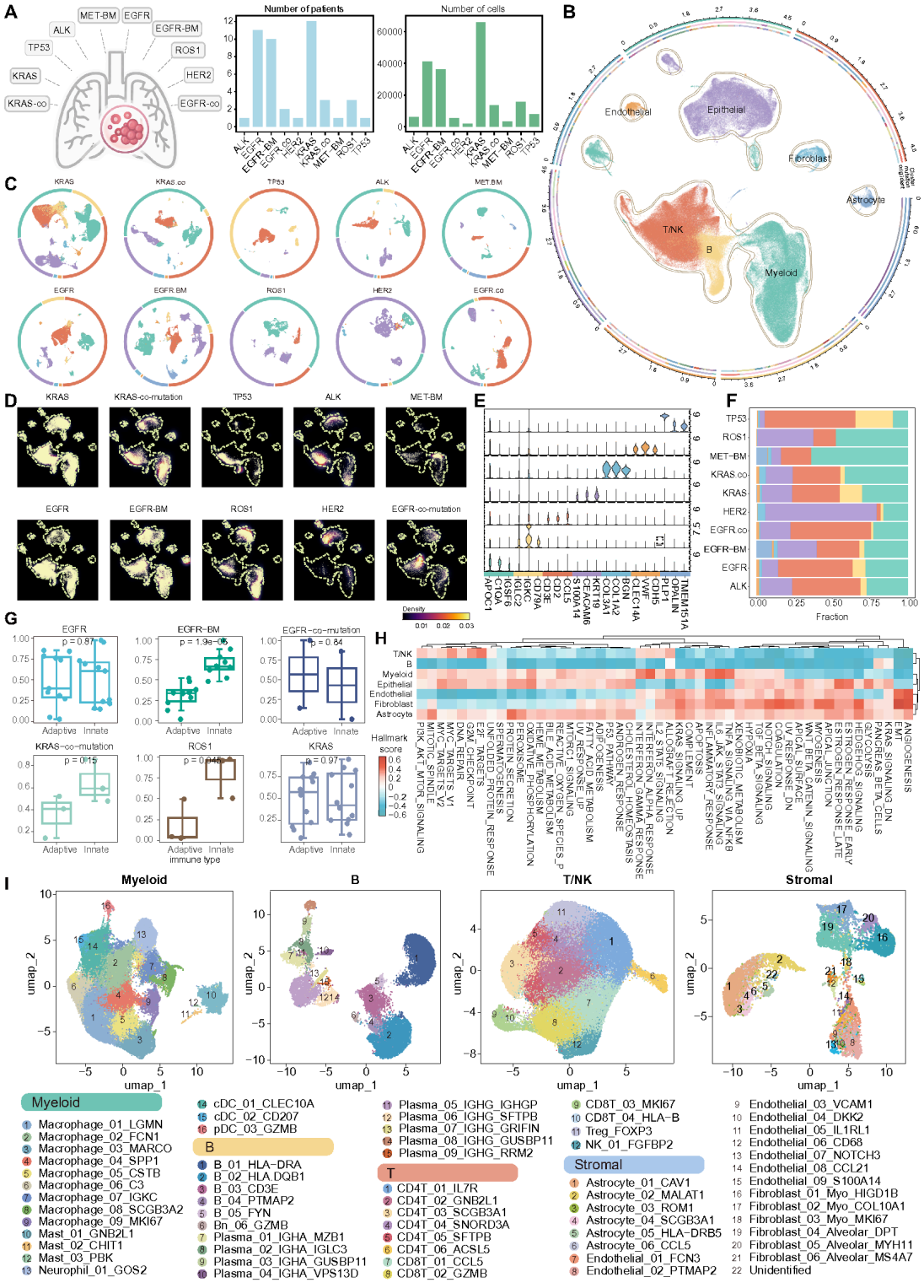


Figure 1

A comprehensive single-cell atlas of driver-mutated NSCLC

A) Data overview. Bar charts show the number of patients and the number of cells for each driver mutation subtype of NSCLC. **B)** Uniform Manifold Approximation and Projection unveiling seven major cell lineages, and colors represent different cell populations. The concentric ring plot displays, from the

innermost to the outermost layer, cell types, oncogenic driver mutations, and individual patients. **C)** UMAP plots of 10 driver-mutated NSCLC samples, colors represent different cell populations. The outer ring indicates the proportion of each cell type. **D)** UMAP density plots of 10 driver-mutated NSCLC samples, colors represent the intensity of cell density. **E)** Violin plots showing the expression of marker genes across seven major cell types. **F)**Stacked bar plot showing the distribution of major cell types in each driver-mutated NSCLC. **G)** Boxplots of innate (T/B) and adaptive (myeloid) immune cell proportions across driver mutations. **H)** GSVA heatmap of hallmark pathway enrichment across seven cell types. **I)** UMAP plots of TIME clusters from 10 driver-mutated NSCLC samples, displayed in panels according to myeloid cells, B cells, T/NK cells, and stromal cells.



Figure 2

Five distinct TIME subtypes identified in driver-mutated NSCLC

A) The five cell modules derived from correlation analysis among major cell clusters within the TIME, with key contributing cell clusters for each module shown on the right.

B) Volcano plots showing differentially expressed genes (DEGs) across the five TIME subtypes, with the top five marker genes labeled. Wilcoxon rank-sum test, adjusted P value < 0.05. **C)** Dot heatmap showing enriched pathways across TIME subtypes. Benjamini–Hochberg-adjusted hypergeometric test. **D)** Violin plots showing the scores of various immune gene sets across TIME subtypes. Chi-square test for statistical comparison. **E)** Heatmap showing the proportions of each TIME subtype across NSCLC patients. **F)**Lollipop plot showing the tissue distribution preference of each TIME subtype across different driver-mutated NSCLC samples. Dot colors represent the type of driver-mutated NSCLC samples, and the values represent the Ro/e ratio. **G)**Heatmap showing the enrichment levels of 14 oncogenic pathways in TIME subtypes (top) and driver-mutated NSCLC samples (bottom). **H)** Sankey plots illustrating the associations of chemokine (top) and chemokine receptor (bottom) expression patterns with TIME subtypes and driver-mutated NSCLC sample types. Expression levels are categorized as high or low based on the median value.

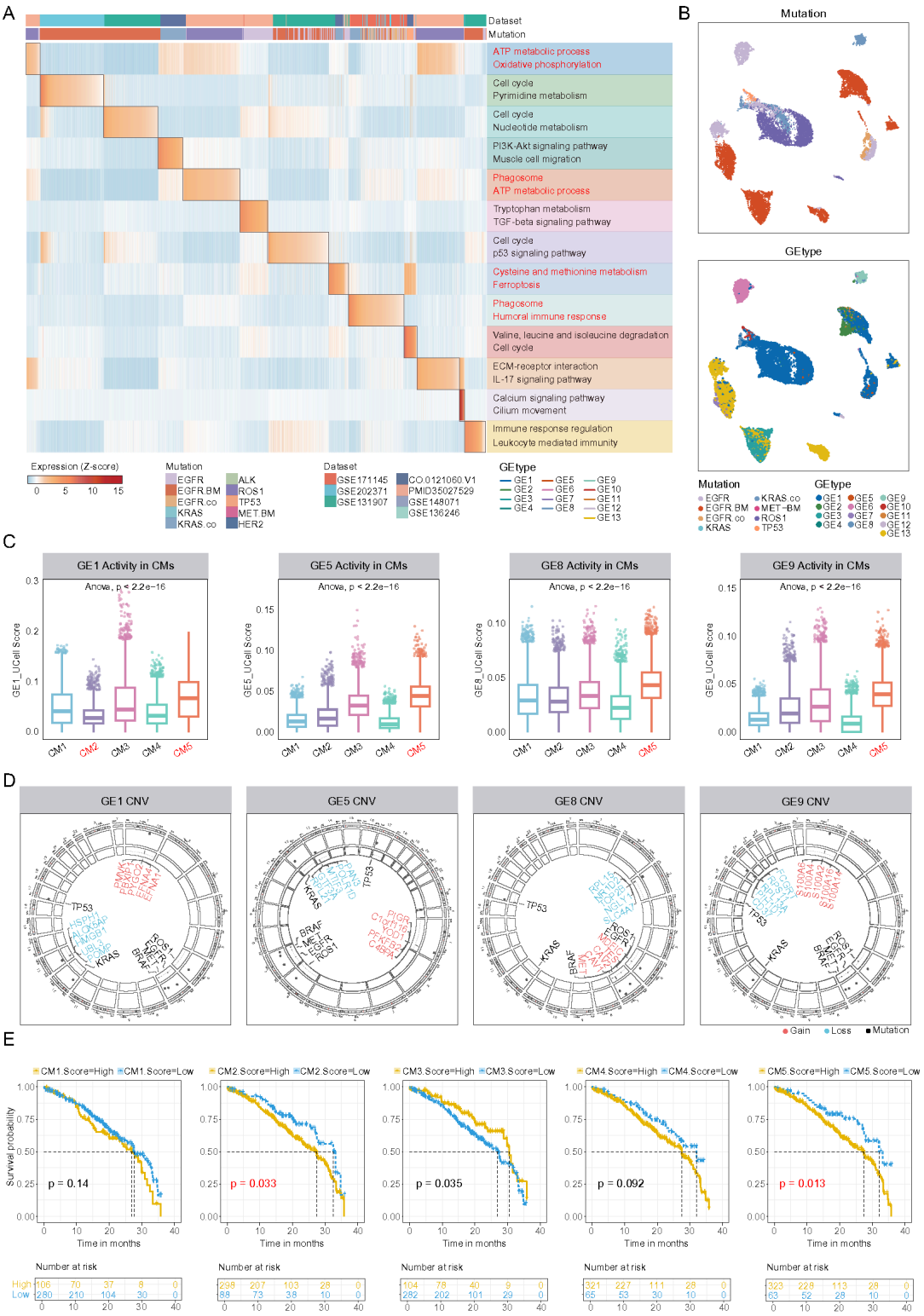


Figure 3

Characterization of malignant features of cancer epithelial cells across TIME subtypes

A) Heatmap of Z-scored signature scores of the 13 identified gene elements (GEs) across all cancer epithelial cells. Annotations indicate dataset origin and driver-mutated NSCLC subtype. The right panel shows the main functions enriched in each GE. **B)**UMAP visualization of cancer epithelial cells, colored

by driver-mutated NSCLC subtype (top) and by GE-defined classification (bottom). **C)** Boxplots showing the distribution of representative GEs signature scores across TIME subtypes. Statistical significance was assessed using ANOVA. **D)** Circo plot of the genome showing major CNV events within representative GEs. Top 5 amplified genes are shown in red, top 5 deleted genes in blue, and known driver genes in black. **E)** Progression-free survival (PFS) over three years stratified by TIME subtype, analyzed using log-rank tests.

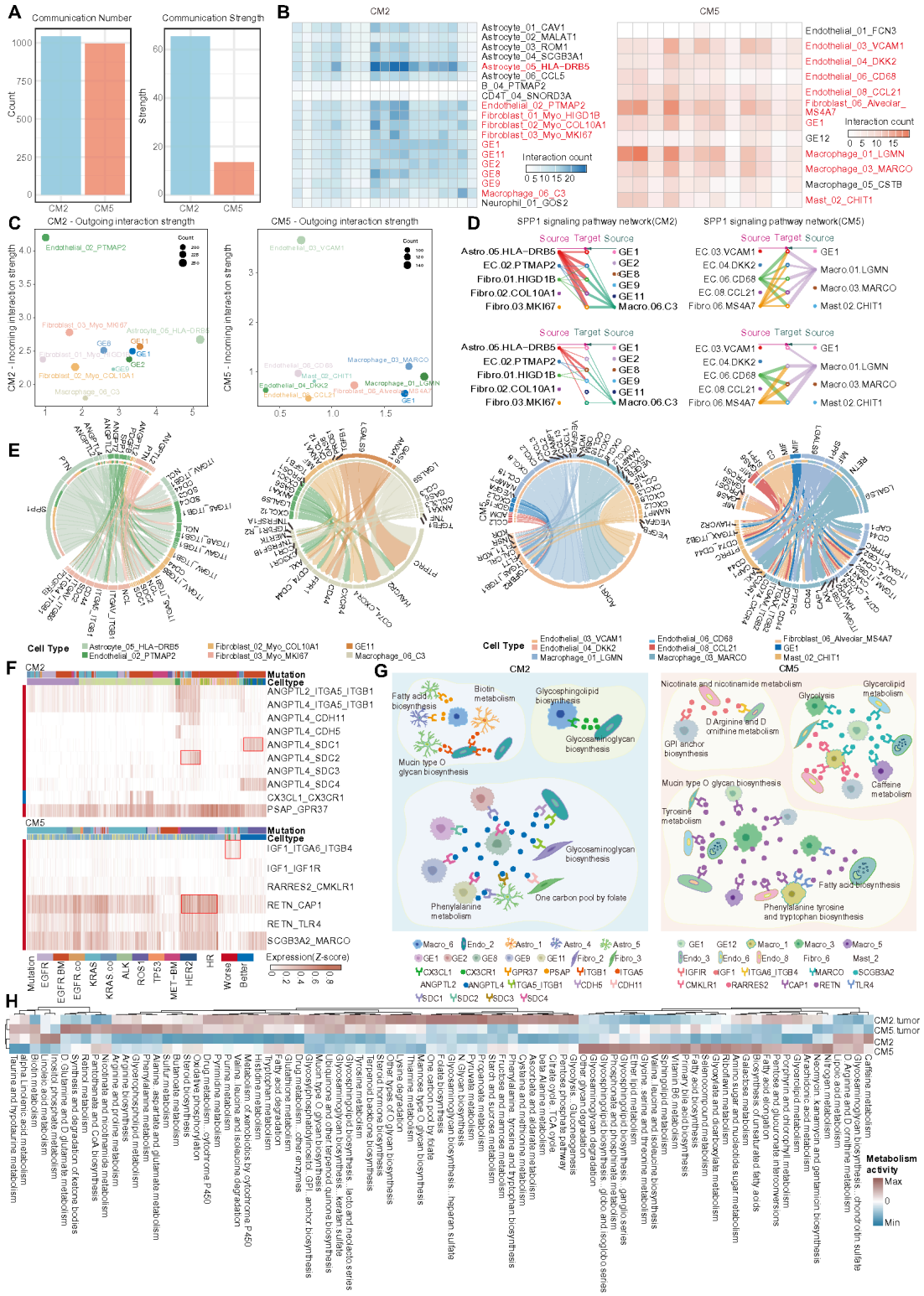


Figure 4

Dissection of potential mediators driving poor prognosis in CM2 and CM5

A) Bar plots showing the number (left) and strength (right) of cell-cell interactions mediated by the CM2 and CM5 subtypes and their associated GEs clusters. **B)** Heatmaps showing the number of cell-cell interactions for each cellular subcluster within the CM2 and CM5 and their associated GEs clusters. **C)** Scatter plots showing the outgoing and incoming interaction strength of each cell type in CM2 and CM5. Dot size reflects the number of interactions; color indicates cell identity. **D)** SPP1 signaling pathways and their corresponding communication networks within CM2- and CM5-associated GEs clusters. Arrows represent directional ligand–receptor interactions among cell types. **E)**Circo plots showing receptor-ligand pairs involved in key signaling pathways mediated by CM2 (left) and CM5 (right) subtypes. **F)** Heatmaps showing the expression of HR-related receptor-ligand pairs between CM2 (top) and CM5 (bottom) subtypes and their associated GE clusters. Annotation panels indicate the driver mutation types and cell types. **G)** Schematic diagram illustrating the mechanisms of potential mediators influencing prognosis in the CM2 (left) and CM (right). **H)** Heatmap showing metabolic pathway activity across major cell groups within CM2, CM5, and their associated GEs clusters. An overview of cell–cell communication in CM5 and their associated GEs clusters..

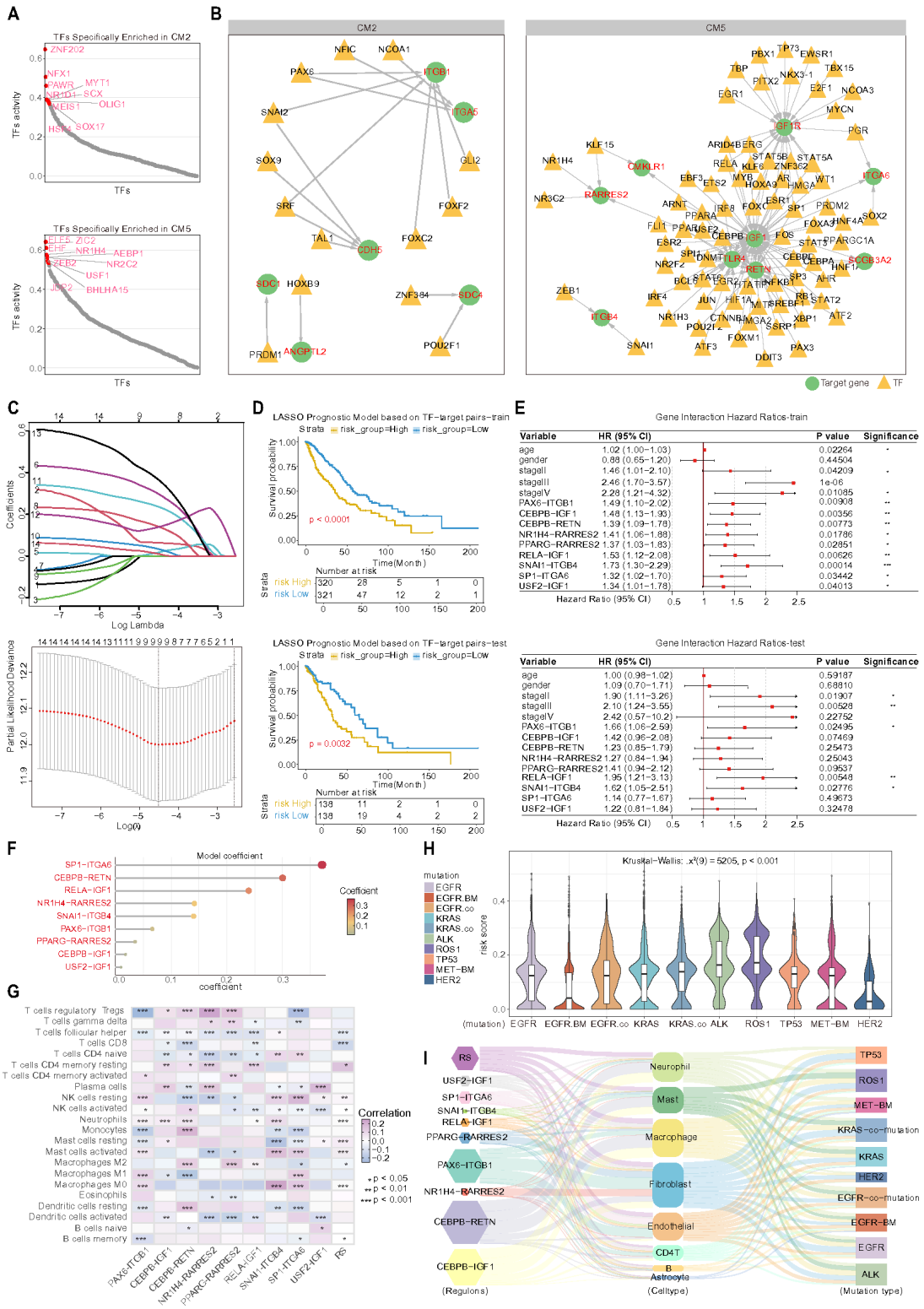


Figure 5

Discovery and functional profiling of the DMSP signature

A) Scatter plots showing high active TFs in CM2 (top) and CM5 (bottom). The top 10 TFs are highlighted in red. **B)** Malignant regulatory networks within CM2 (left) and CM5 (right). Circular nodes indicate HR-related ligand/receptor genes, and triangular nodes denote TIME-specific highly active TFs. **C)** Illustration

of the LASSO coefficient spectrum for the overlapping genes (left) and the adjusted parameters of the LASSO regression model (right). **D)**Kaplan–Meier analysis showing the association between model risk score and OS of patients in the training set (top) and validation set (bottom).

E) Multivariable analysis of clinical factors (age, sex, etc.) and the nine constituent regulators as risk factors for overall survival in the training cohort (top) and validation cohort (bottom). **F)** Lollipop plot showing the coefficient values of the 9 regulators in the model. **G)** The heatmap illustrating the correlations between the model risk score and the nine included regulators with immune cell infiltration, calculated using Pearson's correlation. **H)**Violin plots illustrating the distribution of model-derived risk scores across driver-mutated NSCLC samples, evaluated using the Kruskal–Wallis test.

I) Sankey diagram illustrating the associations of the model-derived risk score and its nine constituent regulators with cell types and driver mutation subtypes of NSCLC. The width of each flow represents the strength of the association.

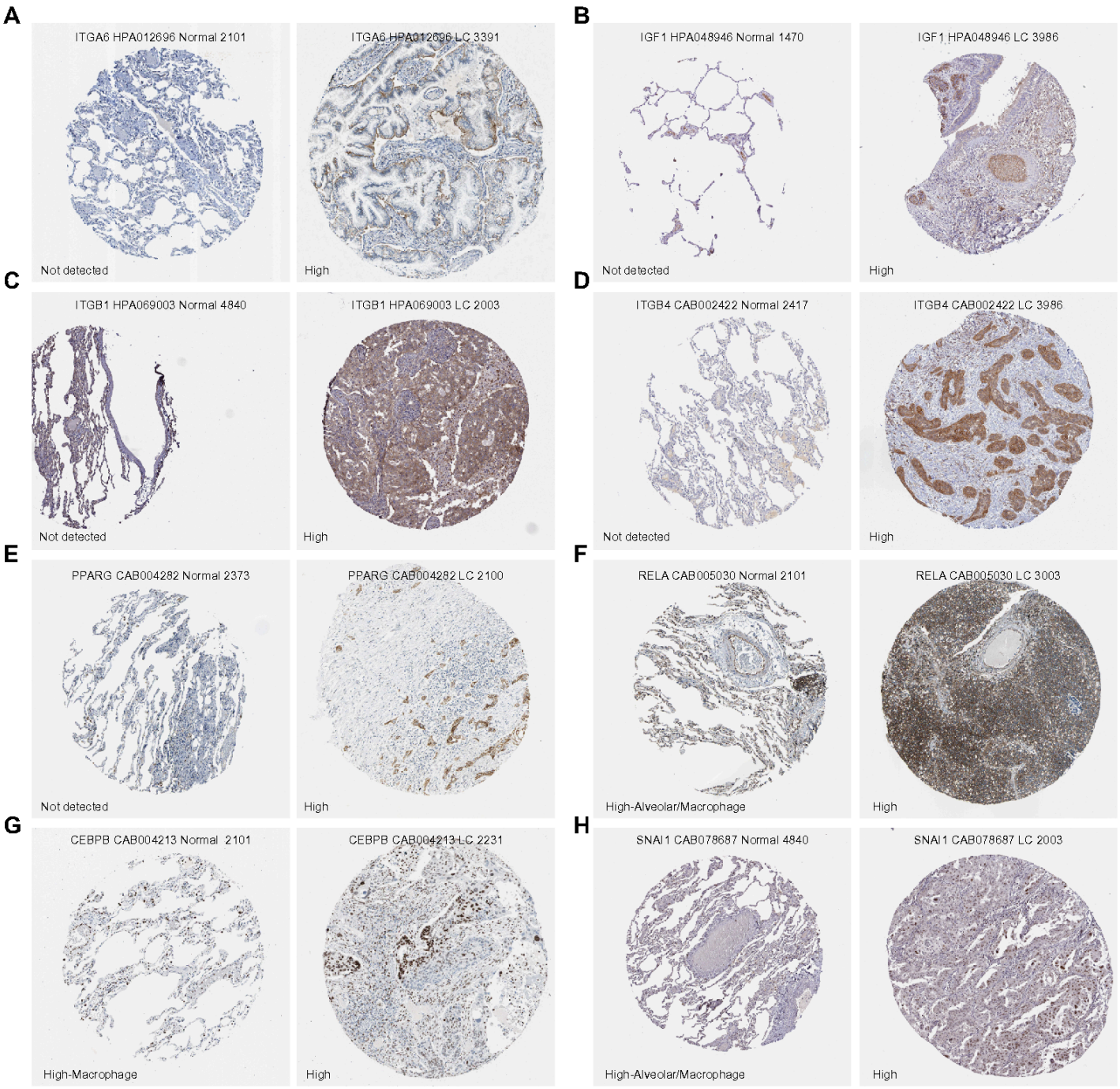


Figure 6

Immunohistochemical staining for modulators of the prognostic model.

A-H) IHC staining of ITGA6, IGF1, ITGB1, ITGB4, PPARG, RELA, CEBPB, and SNAI1 in normal lung tissue and lung cancer tissue, respectively. The high expression of some genes or transcription factors included in the model modulators at the protein level in tumor samples was validated. Each tissue section is annotated at the top with the gene/transcription factor name, antigen type, tissue type, and patient ID. The expression level of the corresponding protein is indicated at the bottom left. Scale bar = 200 μm. Each tissue core has a diameter of approximately 1 mm.

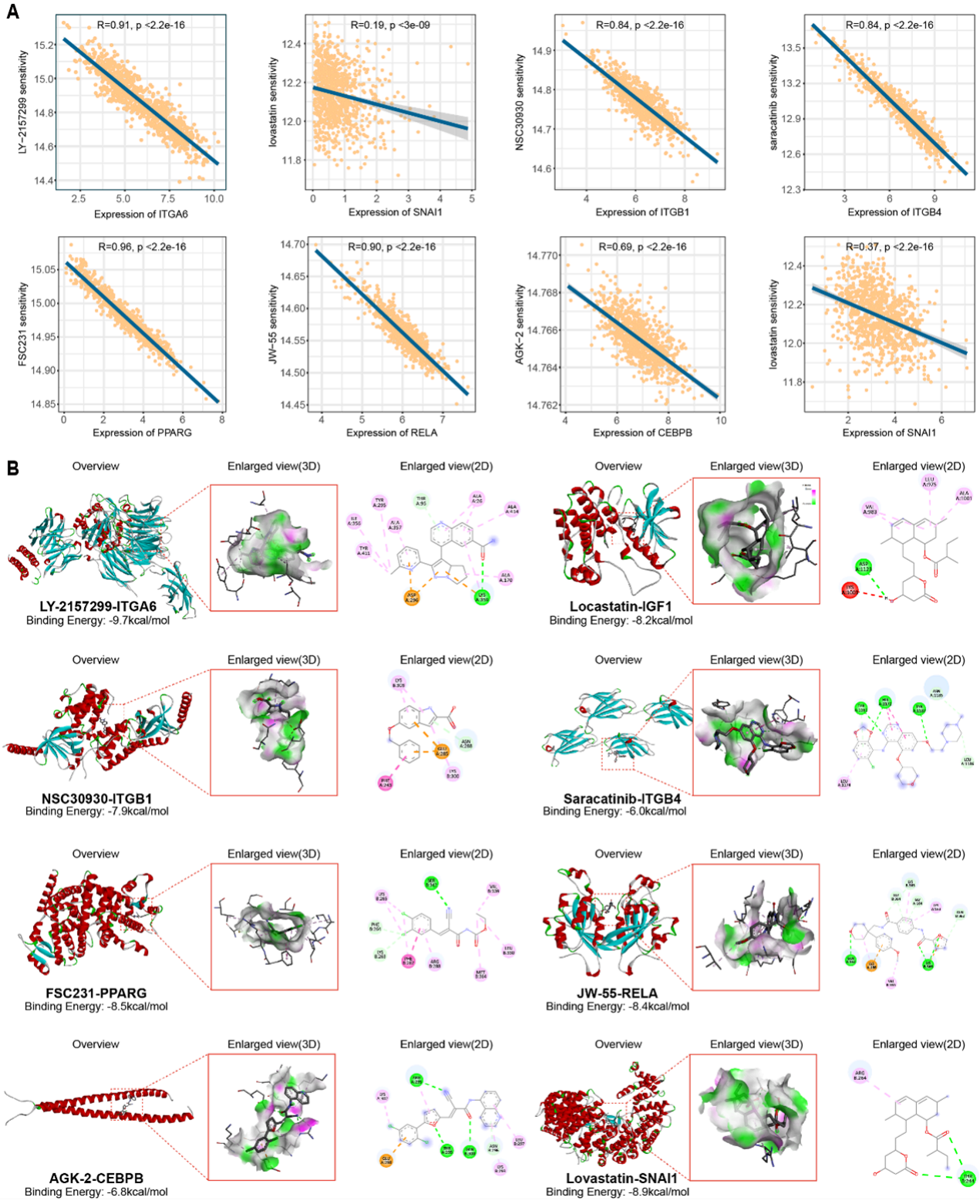


Figure 7

Drug sensitivity and molecular docking analyses

A) Scatter plots illustrating the correlations between target gene expression levels and predicted drug sensitivities, with trend lines indicating the overall strength of the associations. **B)** The molecular docking results between target drugs and core targets.

Supplementary Files

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

- TableS1S13.xlsx
- SupplementaryFile.docx