BMC Cancer

RESEARCH

Open Access

Long-read sequencing and proteomics reveal blood transcriptome and protein expression profiles in multiple primary lung cancers

Yingkui Huang[1†], Guangzhi Zhu[1†], Huanjuan Yang[2†], Yanfei Zhang[1], Hongwu Tian[1], Jing Cao[1], Shengmei Li[3], Guowei Liu[4*] and Husai Ma[1*]

Abstract

Multiple primary cancers (MPC) refer to the occurrence of two or more independent primary malignant tumors in the same patient, either simultaneously or metachronously. Their clinical diagnosis and differential diagnosis are challenging, treatment strategies are complex, and the incidence has been increasing in recent years. However, there is still a lack of potential biomarkers that can be used for early identification and prognosis assessment, which has become a key bottleneck in current research. This study employed Oxford Nanopore Technologies (ONT) long-read transcriptome sequencing and data-independent acquisition (DIA) proteomics to perform integrated multi-omics analysis of blood samples from patients with only primary lung cancer (OPLC), lung cancer with MPC, and healthy controls (HC), systematically characterizing the molecular features of MPC at both transcript and protein levels. The results showed that MPC patients exhibited significantly increased transcript complexity, with the numbers of differentially expressed genes (DEGs), differentially expressed transcripts (DETs), and differential transcript usage (DTU) events being substantially higher than those in OPLC and HC groups. Multiple genes displayed rich isoform diversity. Functional enrichment analysis indicated that MPC-specific molecules were significantly enriched in immune- and inflammation-related pathways such as NF-κB, NOD-like receptor, Toll-like receptor, and TNF signaling. By integrating gene, transcript, and protein expression profiles, several core molecules (e.g., ATP6AP2, APMAP, CIAO2A, and ITGB2) with consistent expression across multiple regulatory levels were identified, all showing significant and consistent alterations in MPC. This study reveals the molecular characteristics of transcript isoform complexity in the blood samples of MPC patients through long-read sequencing combined with proteomics, providing important theoretical insights for understanding the mechanisms of MPC and identifying potential targets.

Keywords Multiple primary cancers, Primary lung cancer, ONT, DIA

[†]Yingkui Huang, Guangzhi Zhu and Huanjuan Yang contributed equally to this work.

*Correspondence:
Guowei Liu
94528086@qq.com
Husai Ma
mahusai197981@163.com

[1]Department of Thoracic Surgery, Qinghai Red Cross Hospital, Xining, Qinghai, China
[2]Department of Medical Service, Qinghai Provincial People's Hospital, Xining, Qinghai, China
[3]Department of Laboratory Medicine, Qinghai Red Cross Hospital, Xining, Qinghai, China
[4]Department of Thoracic Surgery, Qinghai Provincial People's Hospital, Xining 810000, Qinghai, China

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

Multiple Primary Malignancies (MPM), also known as multiple cancers or repeated cancers, refer to the development of two or more independent primary malignant tumors in the same individual, either simultaneously or sequentially. These tumors may arise in different organs or tissues, or within the same organ or tissue [1]. In 1932, Warren and Gates conducted the earliest systematic research on MPM [2]. Based on the number of primary tumors, MPM can be classified into double primary cancer, triple primary cancer, quadruple primary cancer, etc., with double primary cancer being the most common [3]. Cancer patients often possess a susceptible constitution. The overall reported frequency of MPM ranges from 2.4% to 20.0%, and its incidence has been increasing in recent years [4, 5]. Several hypotheses have been proposed for the etiology of MPM, including genetic background, hormonal factors, dietary factors, prior treatments, smoking and environmental exposures, among others [6]. It is noteworthy that the relationship between immunological status, as one of the potential influencing factors, and the development of MPM has not been fully explored. As a key component of the systemic immune microenvironment, blood not only carries various immune cells but also circulates numerous immune-related molecules. These elements collectively form a dynamic response network that may participate in tumor initiation and progression [7]. Therefore, exploring potential biomarkers of MPM from the perspective of the blood immune system holds significant theoretical value. Lung cancer is one of the most frequently diagnosed cancers and accounts for the highest number of cancer-related deaths [8]. The extended survival of cancer patients has increased the likelihood of developing a second primary lung cancer.

Currently, mainstream transcriptomic research still primarily relies on short-read sequencing technologies (such as the Illumina platform) to generate RNA-seq data [9]. These studies typically focus on differential gene expression analysis and further employ Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis to elucidate potential biological functions. While it is undeniable that short-read RNA-seq technology demonstrates high reliability and broad applicability in detecting differential gene expression, its read length limitations also lead to significant shortcomings in capturing key information related to post-transcriptional regulation, particularly in the precise resolution of full-length transcript structures [10]. In contrast, emerging third-generation long-read sequencing technologies (such as PacBio SMRT and Oxford Nanopore sequencing) can directly obtain full-length sequences covering complete transcripts, thereby providing a more comprehensive view of the transcriptome. This technology not only accurately identifies various important molecular features, including poly(A) tail length, alternative splicing events, and differential isoform usage, but also demonstrates unique advantages in discovering novel isoforms and refining existing genome annotations. By performing single-molecule-level full-length sequencing of transcripts, researchers can systematically uncover complex splicing variations, gene fusion events, and low-abundance isoforms that were previously overlooked by short-read technologies, significantly expanding our understanding of the complexity of the transcriptome [11]. This information is crucial for gaining a deeper understanding of gene regulatory mechanisms and functional diversity.

Previous studies on multiple primary cancers have primarily employed exon or first-generation sequencing technologies, with relatively few overall sequencing-related investigations [12, 13]. This project utilizes the internationally recognized third-generation sequencing technology platform—ONT long-read RNA sequencing (ONT lrRNA-seq)—to obtain quantitative results of gene expression and transcript expression levels in samples. Additionally, data-independent acquisition (DIA) mass spectrometry-based proteomics is used to quantify protein expression. By comparing healthy controls (HC), patients with only primary lung cancer (OPLC), and those with multiple primary cancers (MPC), we identified differentially expressed genes (DEGs), differentially expressed transcripts (DETs), and differentially expressed proteins (DEPs). Through in-depth integrated analysis of ONT and DIA proteomics data, it is anticipated that potential biomarkers for multiple primary lung cancer can be systematically uncovered, and more precise therapeutic targets can be identified.

Materials and methods

Patients and specimens

Whole blood samples were collected from cohorts of four patients with HC, four with OPLC, and four with MPC. Serum samples, however, were obtained from a distinct set of subjects, which likewise included four patients per diagnostic group (HC, OPLC, and MPC). The study protocol was reviewed and approved by the Medical Ethics Committee of the QINGHAI RED CROSS HOSPITAL (approval number: KY-2025-18), and all patients who participated in this study or their families understood its approach. Informed consent for participation in the study has been obtained from the patients.

Sample Preparation

The samples were enriched using superparamagnetic iron oxide nanoparticles. Twenty μl of the sample was diluted with loading buffer (10 mM Tris-Cl, 1 mM EDTA, 150 mM KCl, 0.05% CHAPS) and mixed with 1 mg of magnetic beads. The mixture was incubated at

37 °C for 1 h. The magnetic beads were washed twice with the loading buffer, followed by one time wash with a CHAPS-free buffer (10 mM Tris-Cl, 1 mM EDTA, 150 mM KCl). The magnetic beads were captured on a magnetic rack, and the supernatant was discarded to obtain the protein-enriched magnetic beads. The sample was then added with lysis buffer (1% SDC/100 mM Tris-HCl, pH = 8.5/10 mM TCEP/40 mM CAA) and incubated at 60 °C for 30 min for protein reduction and alkylation. An equal volume of ddH2O was added to dilute the SDC to a concentration below 0.5%, and 1 μg of trypsin was added. The mixture was incubated overnight at 37 °C for enzymatic digestion. The next day, TFA was used to bring the pH down to 6.0 to end the digestion. After centrifugation, the supernatant was subjected to peptide purification using self-made SDB-RPS desalting column. The peptide eluate was vacuum dried and stored at -20℃ for later use.

LC-MS/MS detection

Mass spectrometry was conducted using a timsTOF Pro instrument (Bruker Daltonics), a trapped ion mobility quadrupole time-of-flight mass spectrometer. The system was equipped with an online coupled UltiMate 3000 RSLCnano chromatography system (Thermo) and a CaptiveSpray nano ion source (Bruker Daltonics). For separation, peptide samples were first loaded onto a Trap column (75 μm × 20 mm, 2 μm particle size, 100 Å pore size; Thermo) and then separated on a reversed-phase C18 analytical column (75 μm × 250 mm, 1.6 μm particle size, 100 Å pore size; ionopticks). The mobile phases consisted of 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in acetonitrile (mobile phase B), which were used to generate the separation gradient. Chromatography was carried out at a constant flow rate of 300 nL/min. The mass spectrometer was operated in diaPASEF (parallel accumulation–serial fragmentation) mode with trapped ion mobility spectrometry (TIMS) enabled. Each acquisition cycle included one full MS1 scan followed by subsequent MS/MS scans.

Data analysis

DIA raw data were analyzed with DIA-NN (DIA-NN 1.9.2) in library-free mode. Spectra files were searched against against the Human protein sequence database(2025-01-13, 20421 entries) downloaded from Uniprot. Search parameters were mainly default and set as follows: Precursor ion generation options were enabled for in silico-predicted spectral library generation; Trypsin/P with maximum 1 missed cleavage; Carbamidomethyl on C as fixed modification; Oxidation on M and protein N-terminal acetylation as variable modifications; the mass and MS1 accuracy was set to 15 ppm; MBR was enabled; Heuristic protein inference was enabled; Precursor FDR was set to 1% for reliable identifications. Protein intensities were normalized with MaxLFQ algorithm.

Further bioinformatics analysis was conducted in R statistical programming environment. The missing values was imputed from a random Gaussian-shaped distribution around the detection limit of the mass spectrometer, that applied a downshift of 1.8 times the standard deviations and a width of 0.25 times the standard deviations. Valid value filtering was performed with at least 50% of valid values in at least one group. Significantly differentially expressed proteins (DEPs) were determined by the P value from Student's t-test and fold change thresholds. To define the final DEP set, we applied the following criteria: an absolute fold change > 1.5 and P-value < 0.05.

RNA isolation and purity identification

Total RNA was extracted with the RNA Easy Fast Tissue/Cell Kit (DP451; TIANGEN, China). RNA purity and concentration were assessed by measuring the absorbance ratio at 260 nm and 280 nm (A260/A280) using an Ultrafine spectrophotometer (N50 touch; IMPLEN, Germany). RNA integrity was subsequently confirmed by 1.0% agarose gel electrophoresis.

RNA extraction, library construction, and sequencing

Oxford Nanopore Technologies (ONT) sequencing was carried out by Wuhan Ruixing Biotechnology Co., Ltd (http://www.rxbio.cc). Total RNA was isolated using the Total RNA Extraction Kit (DP431; TIANGEN, China). RNA purity and concentration were assessed by measuring the A260/A280 ratio with an Ultrafine spectrophotometer (N50 touch; IMPLEN, Germany), and integrity was confirmed by 1.0% agarose gel electrophoresis.

For cDNA synthesis, 500 ng of total RNA per sample was placed in a 0.2 ml PCR tube and brought to a volume of 9 μl with nuclease-free water. The RNA was mixed with 1 μl of 10 μM VNP primer and 1 μl of 10 mM dNTPs, incubated at 65 °C for 5 min, and then rapidly cooled on a pre-chilled block. Subsequently, a strand-switching mix containing 4 μl of 5× RT buffer, 1 μl of RNaseOUT, 1 μl of nuclease-free water, and 2 μl of 10 μM strand-switching primer was added. After a 2-min incubation at 42 °C, 1 μl of Maxima H Minus Reverse Transcriptase was introduced, and the reaction was performed at 42 °C for 90 min, followed by 85 °C for 5 min and a hold at 4 °C.

The resulting first-strand cDNA (20 μl) was used for PCR library construction with 2 μl of PR1 Primer, 2 μl of PR2 Primer, 25 μl of LongAmp Taq 2× master mix, and 1 μl of nuclease-free water. Thermal cycling conditions were: 94 °C for 3 min; 12 cycles of 94 °C for 15 s, 50 °C for 15 s, and 65 °C for 2 min; a final extension at 65 °C for 10 min; and a hold at 4 °C.

The cDNA libraries were barcoded using the Native Barcode kit (SQK-NBD114.96, NB01-96) according to the manufacturer's instructions. Each barcoded product was purified with 1× AMPure XP Beads, eluted in 20 µl of nuclease-free water, and quantified using a Qubit system. Finally, 300 fmol of the adapter-ligated library was loaded onto a PromethION flow cell (vR10.4.4) for sequencing.

Long-read sequencing data quality control, alignment, correction and identification of transcripts

Raw Nanopore sequencing data were obtained in POD5 format. Base calling was performed using dorado (version 0.8.2) to convert POD5 files into FASTQ format, which contains nucleotide sequences (reads) and their corresponding quality scores. Reads were then processed with Pychopper (version 2.7.9; github.com/nanoporetech/pychopper) to trim adapters and identify full-length transcripts. Quality filtering was applied using chopper (version 0.7.0; https://github.com/wdecoster/chopper) to retain only reads with a mean Phred quality score ≥ 10 and a length ≥ 50 bp for downstream analysis. Filtered reads were aligned to the GENCODE human reference genome (hg38) using minimap2 (version 2.27-r1193) [14] in spliced alignment mode with the command: minimap2 -ax splice -uf -k14. Alternative splicing isoforms were detected using the Pinfish package (github.com/nanoporetech/pinfish), which outputs polished transcripts in GTF format. Transcripts assembled from long reads were merged across all samples using StringTie (version 2.1.6) [15] to generate a non-redundant transcript set. These transcripts were compared against known genomic annotations using gffcompare (version 0.12.6; parameter: -G) [16] to identify novel transcripts and genes. Transcripts were classified as novel or known based on gffcompare class codes (e.g., 'u' for novel intergenic, '=' for exact match) and subsequently merged for further analysis.

Conversion of abundance estimates to transcripts per million (TPM)

To assess quantification accuracy at both transcript and gene levels, transcript-per-million (TPM) values for each transcript and gene in every sample were calculated using Salmon (v1.7.0) [17] in quasi-mapping mode.

Differential gene and transcript expression, and transcript usage analysis

Differential expression analysis of transcripts and genes between the two groups was conducted using DESeq2 (version 1.30.1) [18]. For the purpose of identifying differentially expressed transcripts (DTEs) and genes (DEGs), we applied a dual threshold: an absolute fold change > 1.5 and P-value < 0.05. The corresponding adjusted P-values (FDR) for all results are reported in Supplementary Data to provide a measure of confidence after multiple testing correction.

Differential transcript usage (DTU) was analyzed with the DEXSeq [19] method implemented in the IsoformSwitchAnalyzeR package (version 2.2.0) [20]. Using transcripts identified by gffcompare (version 0.12.6) [16] as input, a change in isoform fraction (dIF) and an adjusted p-value (FDR) were calculated for each isoform. Isoforms with an adjusted p-value < 0.05 were considered to exhibit DTU. At the gene level, the most significant adjusted p-value among its isoforms was assigned to the gene, and genes containing at least one significant isoform were deemed statistically significant.

Pathway enrichment analysis

Gene Ontology (GO) and KEGG pathway analyses were conducted with KOBAS (version 2.0) [21]. Term enrichment was assessed using the hypergeometric test, and significance was adjusted via the Benjamini-Hochberg false discovery rate (FDR) procedure.

Fusion transcripts analyses

Fusion transcripts are always formed by having the coding regions of two or more transcripts joined end to end. It is worth noting that the fusion transcripts are uniformly regulated. Tofu software (version 13.0.0) [22] was used to find fusion transcripts and the identification criteria as follows: (1) a fusion transcript must be mapped to two or more gene locus in the R. typus genome; (2) each gene loci must be aligned to 10% region of a fusion transcript; (3) the fusion transcript must be more than 99% of coverage on the R. typus genome; (4) distance between two mapped locus must be at least 100 kb.

Cell-type quantification

The CIBERSORT algorithm (v1.03) was used with the default parameter for estimating immune cell fractions using TPM values of each expressed gene. A total of 22 human immune cell phenotypes can be deconstructed by CIBERSORT, including seven T cell types [CD8 T cells, naïve CD4 T cells, memory CD4 resting T cells, memory CD4 activated T cells, T follicular helper cells, and regulatory T cells (Tregs)]; naïve and memory B cells; plasma cells; resting and activated NK cells; monocytes; macrophages M0, M1, and M2; resting and activated dendritic cells; resting and activated mast cells; eosinophils; and neutrophils. An R package, Immunedeconv (v2.0.2) [23], that provided a unified interface to seven deconvolution methods was used for estimating immune cell fractions, while CIBERSORT algorithm was applied for estimating immune cell fractions.

Statistical analysis

All plots were done in R (v 4.2.3) as implemented in RStudio in addition to pattern diagrams and stacked bar chart. Data are presented as means ± standard error of the mean (SEM).

Result

Revealing the complexity of blood transcript in patients with multiple primary lung cancer through ONT lrRNA-seq

To characterize the transcriptomic profiles of patients with only primary lung cancer (OPLC), lung cancer with multiple primary cancers (MPC), and healthy controls (HC), this study collected whole blood samples from 4 patients with healthy controls (HC), 4 patients with single primary lung cancer (OPLC), and 4 patients with lung cancer and multiple primary cancers (MPC). These samples were analyzed using ONT lrRNA-seq. A total of 63,191 genes and 254,184 transcripts were identified. Sample-level analysis revealed that the number of transcripts in each sample was much greater than the number of genes, and the numbers of detected genes and transcripts were relatively consistent across samples (Fig. 1A). Further analysis of the types of identified genes showed that protein-coding genes were the predominant category (Fig. 1B). In terms of transcript length, most newly predicted transcripts were significantly shorter than known transcripts (Fig. 1C). Additionally, analysis of transcript count per gene indicated that most genes corresponded to only one transcript (Fig. 1D). In conclusion, this study demonstrates significant complexity in the transcriptomes of patients with multiple primary cancers through the use of ONT lrRNA-seq.

Identification and functional analysis of Blood-Borne gene expression signatures specific to lung cancer with multiple primary cancers

To investigate the dynamic differences in gene expression among patients with OPLC, patients with MPC, and HC, this study performed systematic differential expression analysis using DESeq2. The results revealed that in the comparison between MPC and HC, a total of 2474 differentially expressed genes (DEGs) were identified; 1866

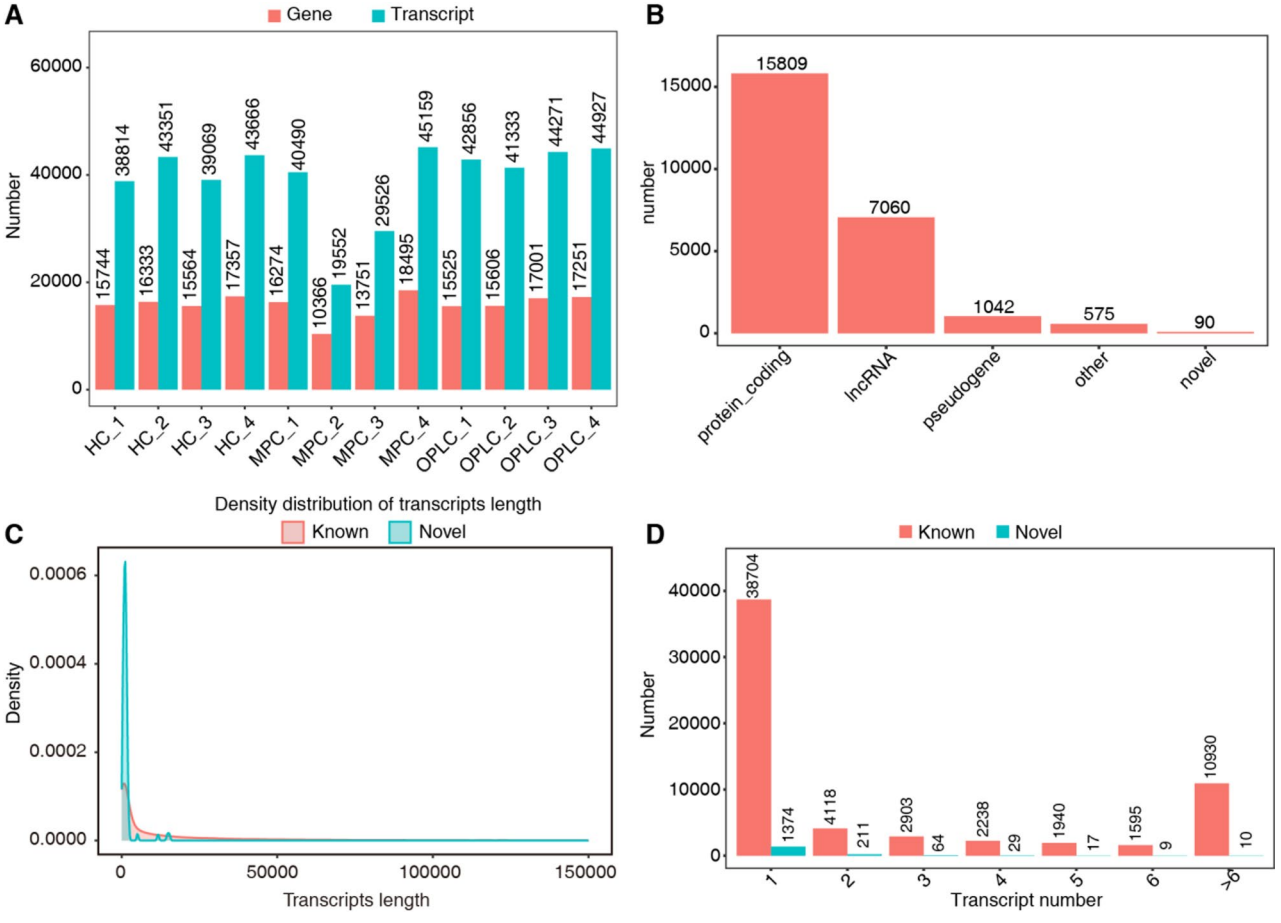

Fig. 1 Transcriptome Analysis of Whole Blood Samples in OPLC, MPC, and HC Based on Full-Length Sequencing. **A** Bar plot showing the number of detected genes and transcripts. **B** Bar plot showing the number of different types of detected genes. **C** Density distribution plot illustrating the length distribution of different types of detected genes. **D** Bar plot showing the number of each gene transcripts

DEGs were found between MPC and OPLC groups; and 345 DEGs were detected between OPLC and HC groups (Fig. 2A). The DEGs are primarily enriched in immune-related pathways (Fig. 2B). Further dynamic clustering analysis of all differential genes across the three groups using clusterGVis showed that clusters C1–C3 exhibited specific high-expression or low-expression patterns in MPC patients. Among them, genes in cluster C1 were mainly enriched in pathways such as viral defense response and oxygen transport; genes in cluster C2 were significantly enriched in functions related to the apoptotic process and positive regulation of interleukin-6 production; and genes in cluster C3 were primarily involved in processes such as protein sulfidation modification and signal transduction (Fig. 2C, Figure S1). In summary, this study reveals that both OPLC and MPC induce extensive dynamic changes in gene expression in the blood at the transcriptomic level, with MPC patients exhibiting more pronounced complexity in transcriptional regulation. This finding provides a new perspective and a basis for the exploration of potential biomarkers in multiple primary lung cancer.

Identification and functional analysis of Blood-Borne transcript expression signatures specific to lung cancer with multiple primary cancers

To investigate the dynamic differences in transcript expression among OPLC, patients with MPC, and HC, this study conducted systematic differential expression analysis using DESeq2. The results showed that a total of 6359 differentially expressed transcripts (DETs) were detected between the MPC and HC groups, 4799 DETs were identified between the MPC and OPLC groups, and 965 DETs were found between the OPLC and HC

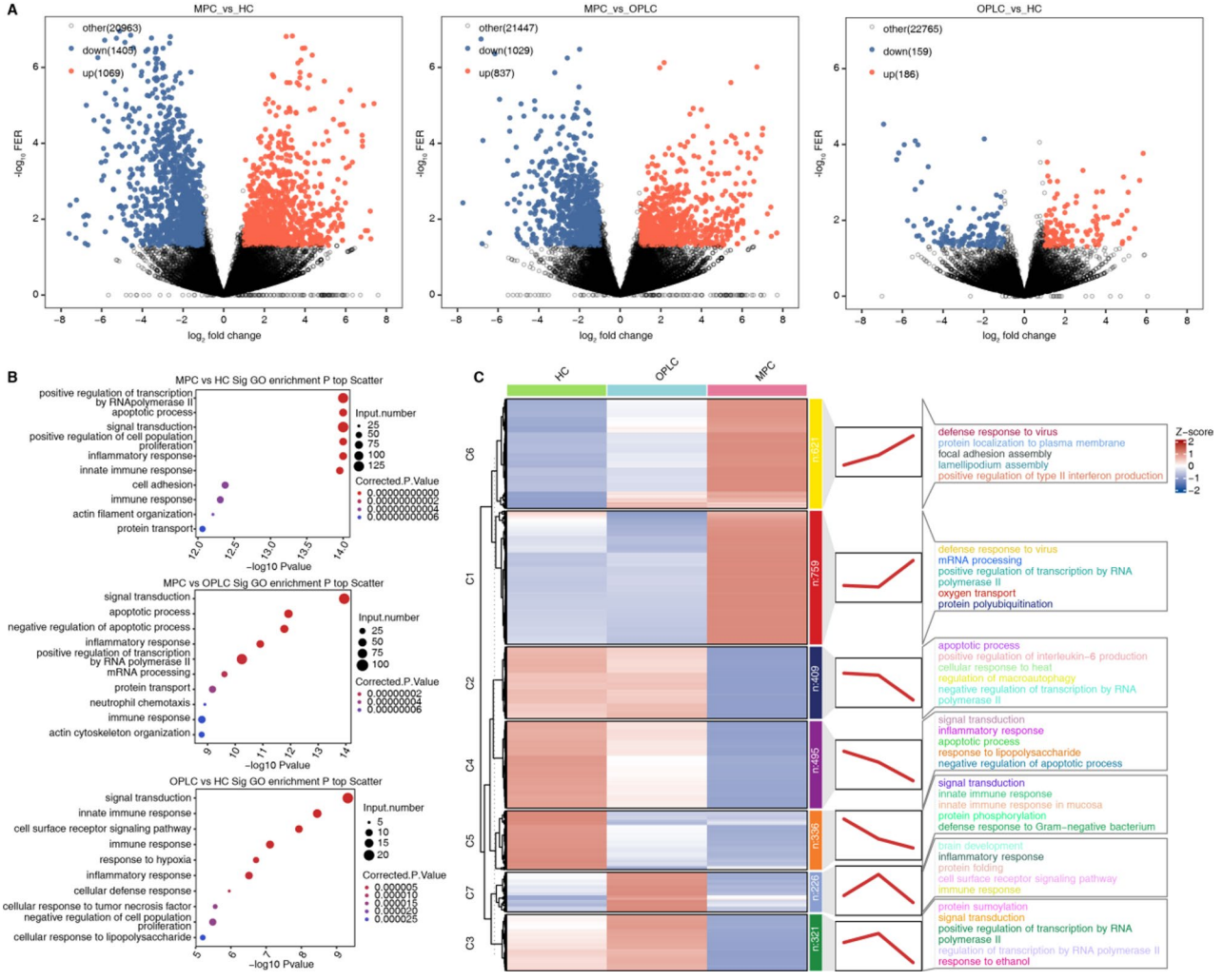

Fig. 2 Analysis of Differentially Expressed Genes in Whole Blood Samples of OPLC, MPC, and HC Based on Full-Length Sequencing. **A** The volcano map shows all the differentially expressed genes (DEGs) in the three comparison groups(FC (fold change) ≥ 1.5 or ≤ 0.67, Pvalue < 0.05). **B** The bubble chart shows the GO BP enrichment analysis results of three groups of differentially expressed genes. **C** The heatmap displays the expression of genes and the corresponding GO enrichment results for each cluster, respectively

groups (Fig. 3A). The genes associated with differentially expressed transcripts were predominantly enriched in immune-related pathways (Fig. 3B). Further dynamic clustering analysis of all differential transcripts across the three groups using clusterGVis revealed that clusters C2–C4 and C7 exhibited specific high-expression or low-expression patterns in MPC patients. Among them, genes associated with clusters C2 and C3 were mainly enriched in pathways such as viral defense response and innate immune response; genes in cluster C4 were significantly enriched in mitochondrial respiratory chain complex I assembly and proton gradient-driven ATP synthesis processes; and genes in cluster C7 were primarily involved in biological processes such as apoptosis and protein folding (Fig. 3C, Figure S2). In summary, this study demonstrates at the transcriptomic level that both OPLC and MPC can induce extensive alterations in transcript expression in whole blood samples, with MPC patients exhibiting more complex transcriptional regulation features. This finding provides a new perspective and a basis for the exploration of potential biomarkers in multiple primary lung cancer.

Screening and characterization of key genes and transcripts in lung cancer with multiple primary cancers

Given the extensive enrichment of immune and inflammation-related pathways identified in the differential expression analyses,. We employed the computational deconvolution algorithm CIBERSORT to estimate the relative abundance of immune cell subsets based on the whole-blood transcriptomic data (Figure S3). The analysis revealed a distinct remodeling of the immune cell

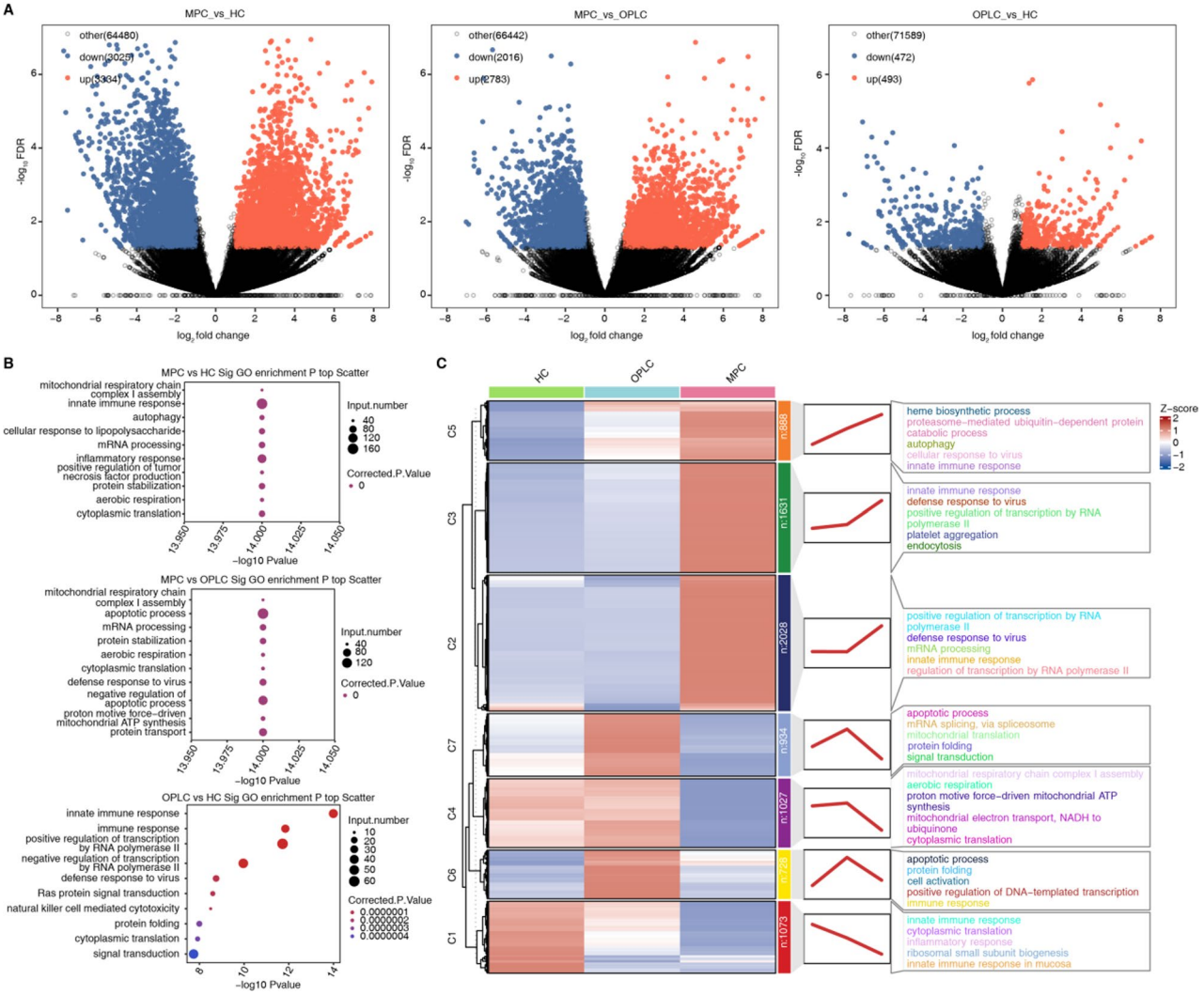

Fig. 3 Analysis of Differentially Expressed Transcripts in Whole Blood Samples of OPLC, MPC, and HC Based on Full-Length Sequencing. **A** The volcano map shows all the differentially expressed transcripts (DETs) in the three comparison groups(FC (fold change) ≥ 1.5 or ≤ 0.67, Pvalue < 0.05). **B** The bubble chart shows the GO BP enrichment analysis results of the three groups of differentially expressed transcript genes. **C**The heatmap displays the expression of transcripts and the corresponding GO enrichment results for each cluster, respectively

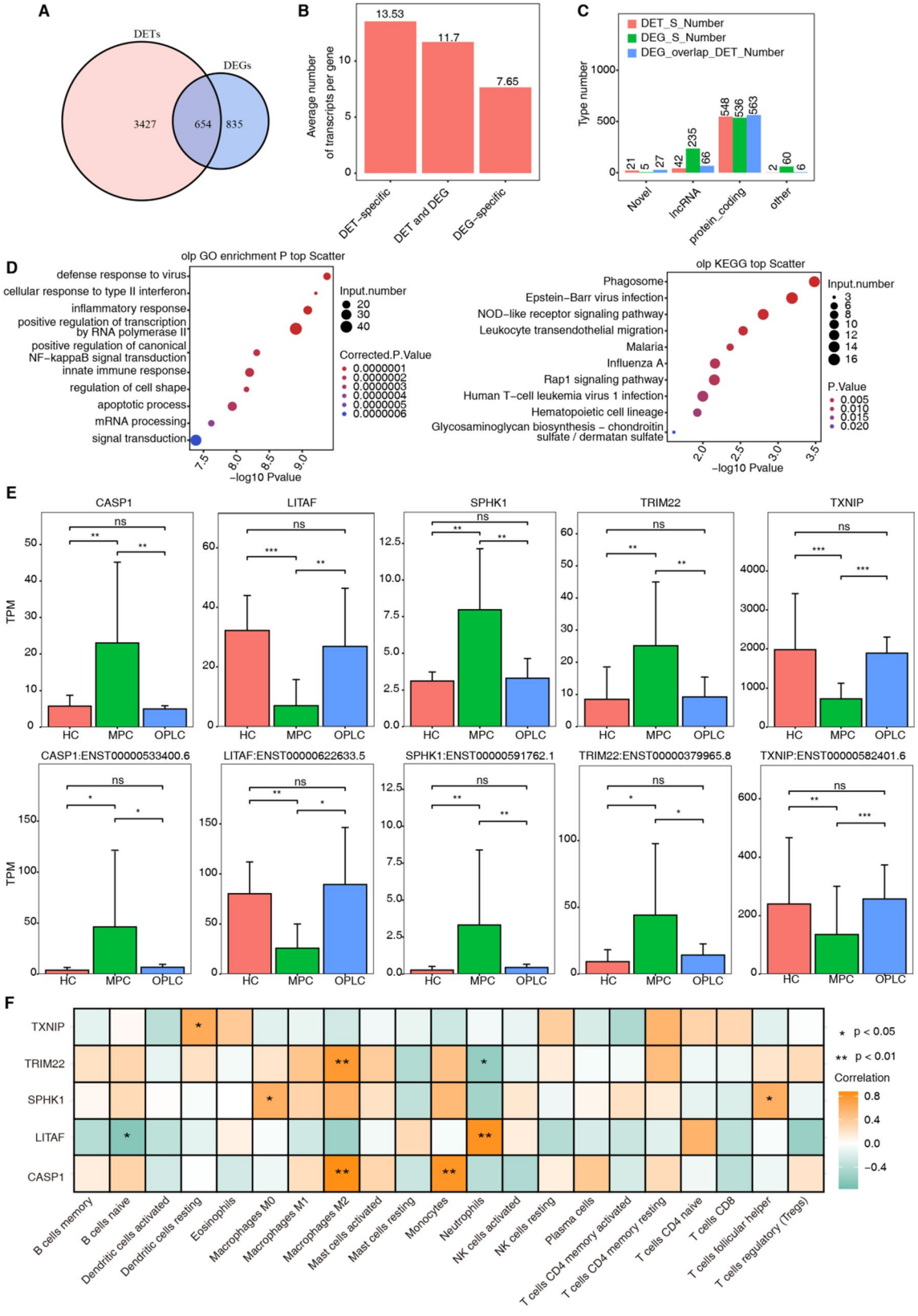

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Integrated analysis of differentially expressed genes and transcripts with key gene expression. **A** Venn diagram illustrating the overlap of gens between differentially expressed genes (DEGs) and differentially expressed transcripts (DETs). **B** Bar plot showing the number of average number of transcripts for each gene in each part of overlap. **C** Bar plot showing the number of different types of detected genes in each part of overlap. **D** The scatter plot exhibiting the most enriched GO (Biological Process) and KEGG pathway enrichment results of the overlap genes. **E** Bar plot showing the expression pattern and statistical difference of DEG sand DETsfrom overlap genes. Error bars represent mean ± SEM(*: P value ≤ 0.05, **: P value ≤ 0.01,***: P value ≤ 0.001)

landscape in MPC patients compared to the HC and OPLC groups. Specifically, we observed a significant decrease in the estimated proportion of neutrophils, concurrently, the monocytes fraction exhibited an upward trend in the MPC group. To explore the molecular mechanisms of MPC from the perspectives of gene expression and transcriptional regulation, this study performed an integrated analysis of DEGs and DETs screened through the aforementioned clustering analysis. The results showed that a total of 654 genes exhibited significant differential expression at both the gene and transcript levels (Fig. 4A). Further analysis revealed that these overlapping genes corresponded to an average of approximately 12 transcripts per gene (Fig. 4B), and the vast majority of them were protein-coding genes (Fig. 4C). Gene Ontology biological process (GO-BP) enrichment analysis indicated that these genes were significantly enriched in biological processes such as inflammatory response and positive regulation of the NF-κB signaling pathway (Fig. 4D). It is noteworthy that several genes, including CASP1, LITAF, TRIM22, SPHK1, and TXNIP, exhibited characteristic expression patterns in the MPC group (Fig. 4E) and were closely associated with immune cell infiltration (Fig. 4F), suggesting that they may play pivotal roles in the development and progression of MPC. In conclusion, this study has identified potential biomarkers associated with immune and inflammatory responses in MPC at the genetic level, providing critical molecular targets for further elucidating its pathogenic mechanism.

Lung cancer with multiple primary cancers affect differential transcript usage in blood

To analyze differential transcript usage (DTU) patterns (i.e., the abundance of specific transcripts relative to other transcripts of the same gene), this study further investigated the potential role of transcript isoform switching in increasing proteomic complexity and functional diversity [24]. Using the DEXSeq tool from IsoformSwitchAnalyzeR, differential transcript usage analysis was performed by calculating the differential isoform fraction (dIF) and false discovery rate (FDR)-adjusted p-value for each isoform under different conditions. Isoforms with an adjusted p-value < 0.05 were considered to exhibit DTU (regardless of dIF magnitude). At the gene level, the smallest adjusted p-value among all isoforms of a gene was taken as the p-value for that gene. Therefore, if a gene had at least one significantly differential isoform, it was considered to exhibit DTU. The analysis results showed that 465 DTU transcripts were identified in the comparison between the MPC and HC groups, and 172 were identified in the comparison between the MPC and OPLC groups, while no significant DTU transcripts were found between the OPLC and HC groups (Fig. 5A). To screen for core regulatory genes, this study integrated the previously highlighted differentially DEGs and DETs from clustering analysis with the DTU results. A total of 41 genes showed significant differences across all three levels (Fig. 5B). Enrichment analysis of these 41 overlapping genes revealed significant enrichment in KEGG pathways closely related to inflammatory responses, including the NOD-like receptor signaling pathway, NF-κB signaling pathway, Toll-like receptor signaling pathway, and TNF signaling pathway (Fig. 5C). It is noteworthy that the transcript ENST00000432176.7 of the HYCC1 gene not only exhibited a significant upregulation in expression abundance but also a marked increase in its relative expression proportion in the MPC group (Fig. 5D, Figure S4). More importantly, this gene was closely associated with the level of immune cell infiltration (Fig. 5E, Figure S5), suggesting that it may be a central driver of the overall upregulation of the HYCC1 gene and plays a pivotal role in regulating the immune microenvironment of MPC.

Lung cancer with multiple primary cancers affect fusion gene in blood

Fusion genes typically incorporate functional domains from different genes, potentially generating chimeric proteins with novel functions. They are frequently observed in cancer and play significant roles. Further analysis of fusion genes revealed a total of 3105 in the HC group, 2522 in the MPC group, and 3822 in the OPLC group (Fig. 6A). Circos plots were used to visualize the fusion events (Fig. 6B-D), displaying the top 20 chromosomes with the highest number of fusion genes in each of the three groups. Subsequently, GO enrichment analysis was performed on coding genes specifically involved in fusions within the MPC group. The results indicated significant enrichment in pathways related to apoptosis, among others (Fig. 6E).

Lung cancer with multiple primary cancers affect protein expression in serum

To characterize the proteomic profiles of patients with only primary lung cancer (OPLC), lung cancer with multiple primary cancers (MPC), and healthy controls (HC),

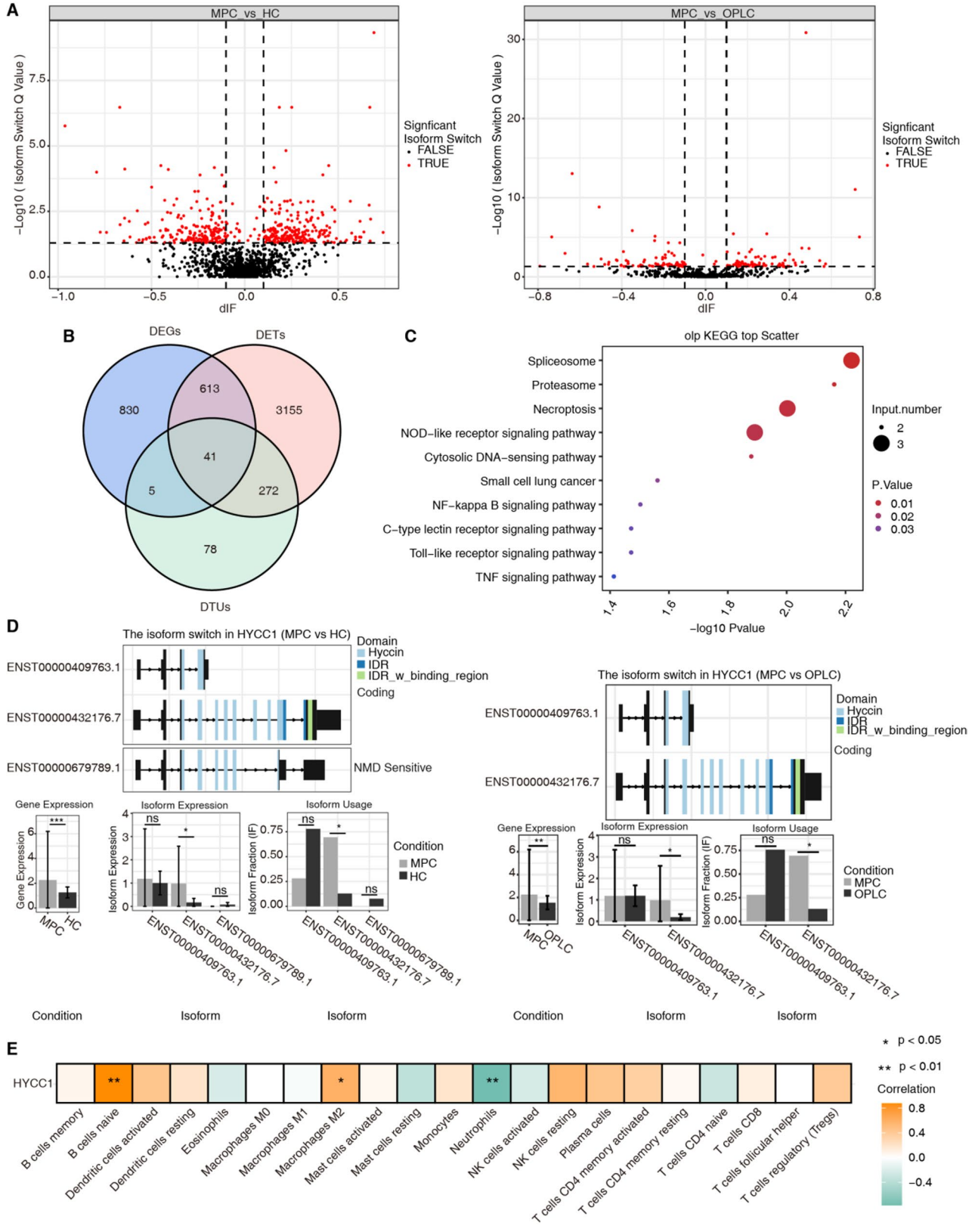

Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Analysis of differential transcript usage (DTU) and identification of core regulatory genes. **A** The volcano plot illustrates isoform switching between groups. The x-axis shows the difference in isoform fraction (dIF) for the MPC-vs-OPLC and MPC-vs-HC comparisons, while the y-axis shows the -log10 (p-value), with higher values indicating greater statistical significance. **B** Venn diagram depicting the overlap of genes with significant DGEs, DETs, and DTUs. **C** The Scatter plot exhibiting the most enriched KEGG pathway results of the overlap genes. **D** Isoform switching observed within the HYCC1 is presented in this figure generated by IsoformSwitchAnalyzeR. The upper part shows the isoform switch with colors denoting different functional domains. The lower part has three sub - plots for gene expression, isoform expression, and isoform fraction (IF)(*: P value < 0.05, **: P value < 0.01,***: P value < 0.001)

this study utilized data-independent acquisition (DIA) proteomics technology to perform relative quantitative analysis of proteins in serum samples from the three groups. A total of 14,830 peptides were identified, corresponding to 2,292 unique proteins (Fig. 7A). Statistical analysis and differential screening of the protein quantitative results revealed 48 differentially expressed proteins in the MPC_vs_HC group, 41 in the MPC_vs_OPLC group, and 80 in the OPLC_vs_HC group (Fig. 7B). Functional enrichment analysis showed that the differential proteins in the MPC_vs_HC group were primarily involved in pathways such as cellular redox status and endocytosis; the differential proteins in the MPC_vs_OPLC group were significantly enriched in pathways related to cell motility and cell adhesion; while the differential proteins in the OPLC_vs_HC group were associated with pathways such as cellular homeostasis and inflammatory response (Fig. 7C). Furthermore, protein-protein interaction networks constructed based on the STRING database further revealed functional associations and potential synergistic effects among the differentially expressed proteins in the three comparison groups (Fig. 7D).

Screening and characterization of key genes, transcripts, and proteins in lung cancer with multiple primary cancers

To identify core regulatory molecules in lung cancer with multiple primary cancers (MPC) at both the transcriptional and translational levels, this study conducted a systematic integrated analysis of differentially expressed genes (DEGs), differentially expressed transcripts (DETs), and differentially expressed proteins (DEPs). Initially, 19 overlapping genes were identified among the three categories (Fig. 8A). However, due to inconsistent expression trends of these genes across the gene, transcript, and protein levels, they could not be directly regarded as reliable multi-omics candidate molecules. Therefore, the focus was extended to pairwise intersections: specifically, 26 genes overlapping between DEGs and DEPs, and 49 genes overlapping between DETs and DEPs. Enrichment analysis of the DEG-DEP overlapping genes showed significant enrichment in pathways such as leukocyte transendothelial migration and natural killer cell-mediated cytotoxicity; while the DET-DEP overlapping genes were significantly enriched in pathways including Rap1 signaling pathway and regulation of actin cytoskeleton (Fig. 8B). Notably, ATP6AP2 exhibited significant differences and consistent expression trends at both the gene and protein levels (Fig. 8C). Meanwhile, APMAP (ENST00000217456.3), CIAO2A (ENST00000558779.1), and ITGB2 (ENST00000475170.5) also showed consistent expression trends at both the transcript and protein levels (Fig. 8D). The consistent alterations of these molecules across multiple regulatory levels suggest that they may play key roles in the development and progression of MPC, providing high-confidence candidate targets for subsequent functional validation and research.

Discussions

This study utilized Oxford Nanopore Technologies (ONT) long-read transcriptome sequencing and DIA proteomics to perform integrated multi-omics analysis of blood samples from MPC, OPLC, and HC. It systematically revealed the dynamic regulatory features of MPC across multiple levels, including transcript complexity, isoform usage, and protein expression, providing new perspectives and data support for a deeper understanding of the molecular mechanisms of MPC.

Transcript isoform switching plays an important role in increasing proteomic complexity and functional diversity [24]. Therefore, through in-depth analysis of genes and transcripts, we found that MPC patients exhibited significantly increased transcript complexity. Compared with the OPLC and HC groups, the MPC group not only had a significantly greater number of DEGs and DETs, but also multiple genes showed rich isoform diversity. This finding suggests that the occurrence of MPC may be closely related to alterations at the transcript level and isoform switching, and these changes likely promote tumor heterogeneity and malignant progression by increasing the functional diversity of the proteome [25]. Notably, a specific transcript of the HYCC1 gene (ENST00000432176.7) showed significantly increased expression level and relative proportion in MPC. Studies have reported that HYCC1 overexpression can accelerate the proliferation, invasion, and migration of pancreatic cancer cells in vitro and promote liver metastasis in vivo [26]. Collectively, these findings indicate that HYCC1 is a potential key regulator of MPC.

On the other hand, MPC-specific expressed genes, transcripts, and differentially expressed proteins were significantly enriched in immune- and inflammation-related pathways, such as NF-κB, NOD-like receptor, Toll-like receptor, and TNF signaling pathways. These

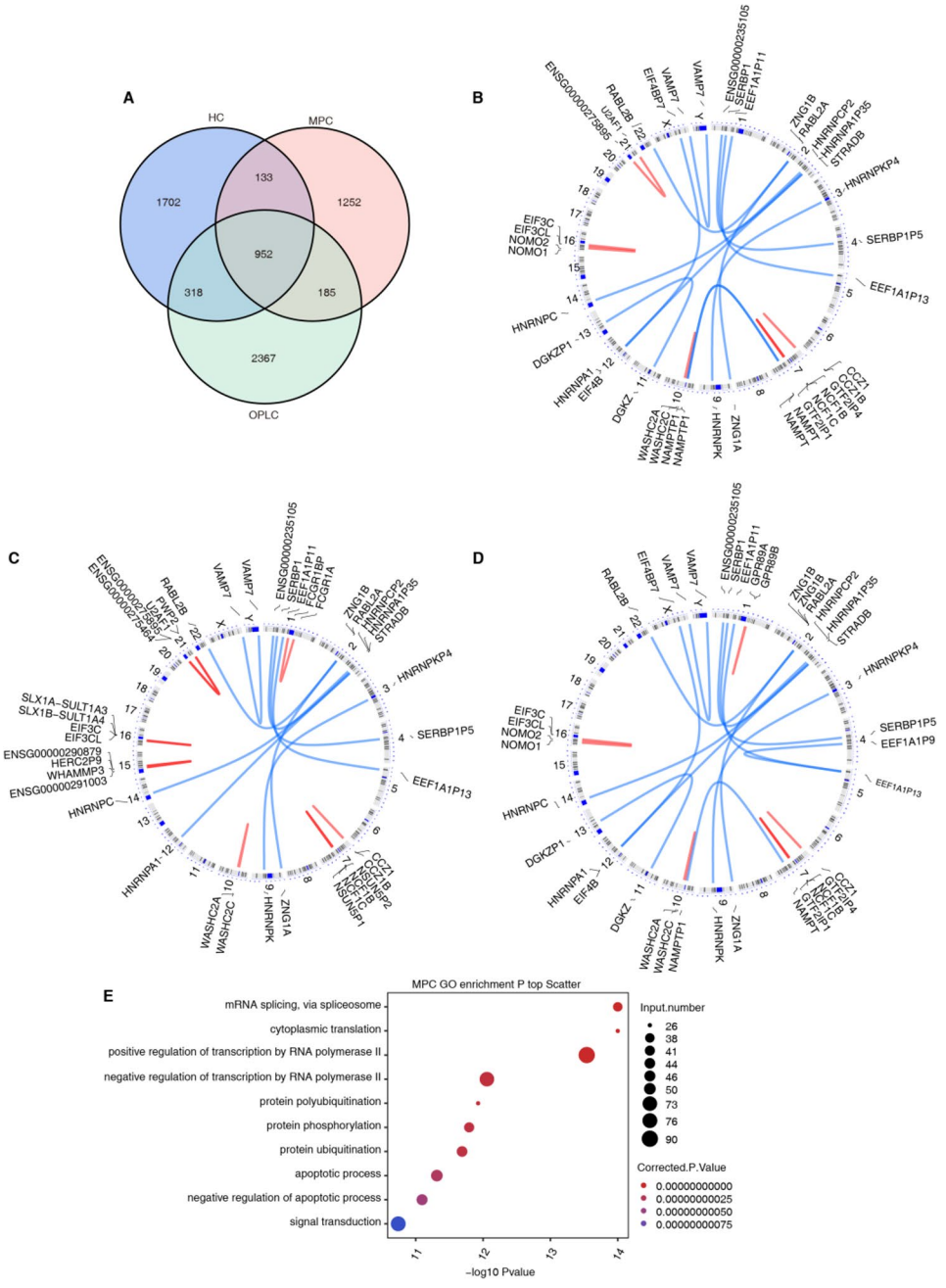

Fig. 6 Lung Cancer with Multiple Primary Cancers Affect Fusion Gene in Blood. **A** Venn diagram of fusion genes in OPLC, MPC, and HC. **B** Circos plot showing the top 20 fusion genes in the HC group. **C** Circos plot showing the top 20 fusion genes in the MPC group. **D** Circos plot showing the top 20 fusion genes in the OPLC group. **E** The scatter plot exhibiting the most enriched GO (Biological Process) pathway enrichment results of the specific fusion genes in the MPC group

pathways play core roles in shaping the tumor microenvironment [27, 28] and the release of pro-inflammatory factors [29, 30]. It is noteworthy that the activation of immune-inflammatory signals was observed across multiple regulatory levels from genes and transcripts to proteins, indicating that the occurrence of MPC may be closely related to dysregulated immune-inflammatory responses [31]. Furthermore, whole blood samples is likely an important window reflecting the body's immune status and tumor burden [32], Whole-blood transcriptomics are often influenced by leukocyte proportions [33]. Our deconvolution analysis highlighted a specific immune alteration in MPC patients, characterized by significantly reduced neutrophils and a tendency towards increased macrophage proportions. The upregulation of inflammatory pathways could stem from the altered

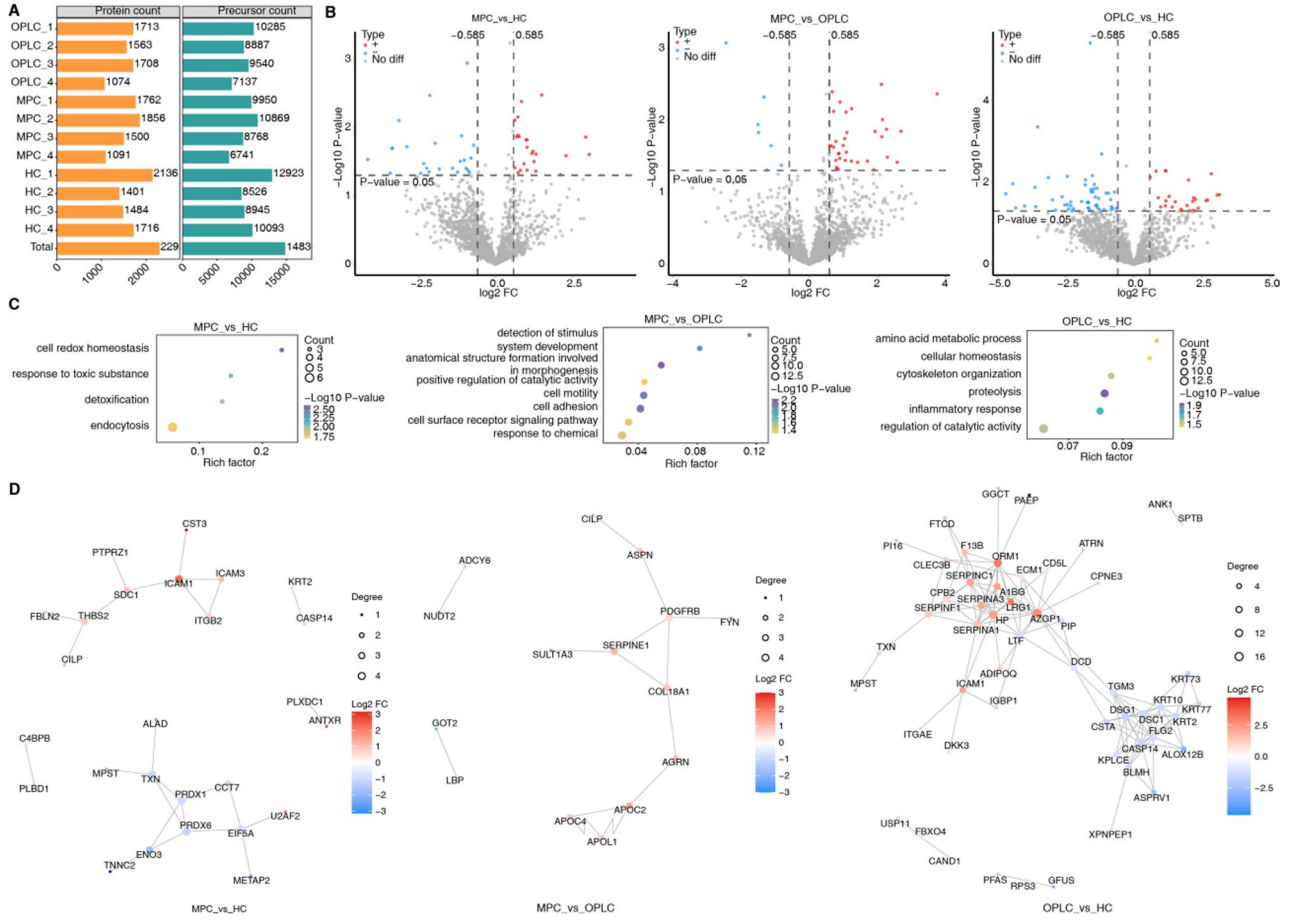

Fig. 7 Proteomic Analysis of Serum Samples in OPLC, MPC, and HC. **A** Bar graphs showing the number of peptide precursors and proteins detected in all samples. **B** Volcano plot showing differential proteins analyzed by DESeq2. **C** The bubble chart shows the Go enrichment results of the three groups of differentially expressed proteins. **D** PPI network diagram, each dot in the network represented a protein, and the more segments connected with the dot, the more important this protein was in the network structure

activation state of the remaining leukocytes or the relative enrichment of specific mononuclear subsets. Future studies employing single-cell RNA sequencing (scRNA-seq) are expected to further elucidate the cell-specific contributions underlying these signals.

By integrating expression data from the three levels of genes, transcripts, and proteins, we screened multiple core molecules with consistent differential expression in MPC. For example, ATP6AP2 was significantly downregulated at both the gene and protein levels, while molecules such as APMAP, CIAO2A, and ITGB2 showed consistent changes at the transcript and protein levels. ATP6AP2 is involved in V-ATPase assembly and Wnt signaling pathway regulation, and its functional disruption is associated with various cancers [34, 35]. We speculate that in MPC, it may promote tumor progression by affecting lysosomal acidification and canonical signal transduction. ITGB2 is an important immune cell adhesion molecule [36] and plays a key role in leukocyte recruitment and activation [37]. Its consistent upregulation in MPC suggests that this molecule may be involved in regulating immune cell infiltration and inflammatory responses in the tumor microenvironment, thereby affecting the immune escape capability of MPC. The consistent changes of these molecules across multiple regulatory levels not only indicate that they may play key roles in the development and progression of MPC but also provide high-confidence candidate targets and translational research directions for subsequent functional validation and mechanistic studies.

In summary, this study used ONT long-read sequencing combined with DIA proteomics to systematically reveal extensive transcript isoform variations and dysregulated protein expression in the blood samples of MPC patients from a multi-omics perspective, highlighting the central role of immune-inflammatory signaling pathways in the pathogenesis of MPC. It provides an important basis for mechanistic research and target discovery for this disease.

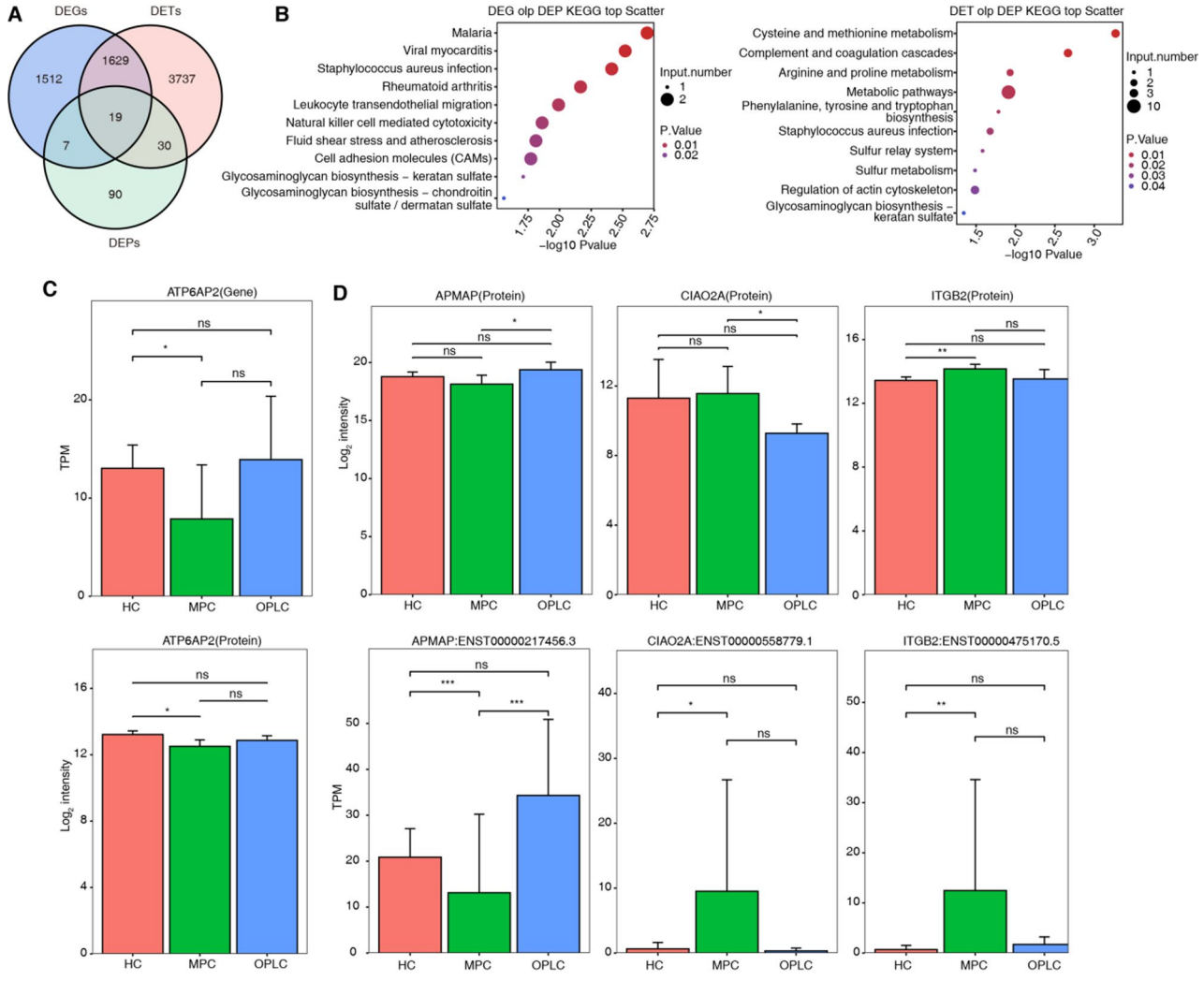

Fig. 8 Integrated analysis of differentially expressed genes (DEG), differentially expressed transcripts (DET) and differentially expressed proteins (DEP) with key molecule expression. **A** Venn diagram depicting the overlap of genes with significant DGEs, DETs, and DEPs. **B** The bubble chart shows the KEGG enrichment results for two distinct gene sets: the intersection of DEGs and DEPs, and the intersection of DETs and DEPs. **C** The bar chart simultaneously displays the gene and protein expression levels of ATP6AP2. **D** The bar chart shows the expression of APMAP, CIAO2A, and ITGB2 proteins and their corresponding transcripts

Shortcomings and outlook

This study has several limitations. First, while long-read sequencing enables direct analysis of full-length transcripts, the absence of orthogonal validation using short-read RNA-seq data somewhat restricts the confirmation of differential expression findings and the detection sensitivity for low abundance transcripts. Second, due to sample availability constraints, transcriptomic and proteomic analyses were performed using non-matched cohorts, which prevents correlation analysis at the individual level across omics layers. Furthermore, the small sample size ($n = 4$ per group) limits statistical power, and the exhaustion of clinical specimens precluded experimental validation such as RT-PCR. Therefore, the molecular signatures identified should be regarded as preliminary and hypothesis-generating, requiring further validation in larger independent cohorts.

Abbreviations

MPM	Multiple Primary Malignancies
ONT	Oxford Nanopore Technologies
DIA	Data-independent acquisition
OPLC	Only primary lung cancer
DEGs	Differentially expressed genes
DETs	Differentially expressed transcripts
DTU	Differential transcript usage
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-026-15709-9.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Acknowledgements
Not applicable.

Authors' contributions
Yingkui Huang, Guowei Liu and Husai Ma : Writing-review & editing, Supervision, Resources, Project administration, Conceptualization. **Yanfei Zhang and Hongwu Tian: ** Visualization, Validation. **Jing Cao and Shengmei Li: ** Software, Methodology, Investigation, Formal analysis. **Yingkui Huang, Guangzhi Zhu and Huanjuan Yang: ** Writing–original draft, Methodology, Investigation, Data curation, Conceptualization.

Funding
The Fundamental Research Project of Qinghai Provincial Science and Technology Department (No. 2023-ZJ-782).

Data availability
The dataset supporting the conclusions of this article is available in the GEO repository, GSE311391 and hyperlink to dataset in https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi? acc=GSE311391.

Declarations

Ethics approval and consent to participate
This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was reviewed and approved by the Medical Ethics Committee of the QINGHAI RED CROSS HOSPITAL (approval number: KY-2025-18). Written informed consent was obtained from all individual participants or their legal guardians.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 1 December 2025 / Accepted: 4 February 2026
Published online: 27 February 2026

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71(3):209–49.
2. Warren S. Multiple primary malignant tumors: a survey of the literature and a statistical study. Gastroenterology. 1932;93(4):779.
3. Etiz D, Metcalfe E, Akcay M. Multiple primary malignant neoplasms: A 10-year experience at a single institution from Turkey. J Cancer Res Ther. 2017;13(1):16.
4. Mehmet S. Copur, Suresh, manapuram: multiple primary tumors over a lifetime. Oncol (Williston Park NY). 2019;33(7):629384.
5. Halamkova J, Kazda T, Pehalova L, Gonec R, Kozakova S, Bohovicova L, et al. Second primary malignancies in colorectal cancer patients. *Frontiers in Oncology* 2021(1).
6. Ventura L, Carbognani P, Gnetti L, Rossi M, Tiseo M, Silini EM, et al. Multiple primary malignancies involving lung cancer: a single-center experience. Tumori J. 2021;107(3):196–203.
7. Hiam-Galvez KJ, Allen BM, Spitzer MH. Systemic immunity in cancer. Nat Rev Cancer. 2021;21(6):345–59.
8. Kratzer TB, Bandi P, Freedman ND, Smith RA, Travis WD, Jemal A, et al. Lung cancer statistics, 2023. Cancer. 2024;130(8):1330–48.
9. Rukhsar L, Bangyal WH, Ali Khan MS, Ag Ibrahim AA, Nisar K, Rawat DB. Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification. Appl Sci. 2022;12(4):1850.
10. Monzó C, Liu T, Conesa A. Transcriptomics in the era of long-read sequencing. Nat Rev Genet. 2025;26(10):681-701.
11. Jain M, Olsen HE, Paten B, Akeson M. The Oxford nanopore minion: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17(1):239.
12. Cavazos TB, Kachuri L, Graff RE, Nierenberg JL, Thai KK, Alexeeff S, et al. Assessment of genetic susceptibility to multiple primary cancers through whole-exome sequencing in two large multi-ancestry studies. BMC Med. 2022;20(1):332.
13. Hu C, Zhao L, Liu W, Fan S, Liu J, Liu Y, et al. Genomic profiles and their associations with TMB, PD-L1 expression, and immune cell infiltration landscapes in synchronous multiple primary lung cancers. J Immunother Cancer. 2021;9(12):e003773.
14. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
15. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.
16. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Research* 2020, 9.
17. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9.
18. Love MI, Huber W, Anders S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.
19. Li Y, Rao X, Mattox WW, Amos CI, Liu B. RNA-seq analysis of differential splice junction usage and intron retentions by dexseq. PLoS ONE. 2015;10(9):e0136653.
20. Vitting-Seerup K, Sandelin A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. Bioinformatics. 2019;35(21):4469–71.
21. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39(suppl2):W316–22.
22. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat Commun. 2016;7(1):11708.
23. Sturm G, Finotello F, List M. Immunedeconv: an R package for unified access to computational methods for estimating immune cell fractions from bulk RNA-sequencing data. Bioinformatics for cancer immunotherapy: methods and protocols. edn.: Springer; 2020. pp. 223–32.
24. Wright DJ, Hall NA, Irish N, Man AL, Glynn W, Mould A, et al. Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. BMC Genomics. 2022;23(1):42.
25. Vitting-Seerup K, Sandelin A. The landscape of isoform switches in human cancers. Mol Cancer Res. 2017;15(9):1206–20.
26. Li Y, Li Y, Luo J, Fu X, Liu P, Liu S, et al. FAM126A interacted with ENO1 mediates proliferation and metastasis in pancreatic cancer via PI3K/AKT signaling pathway. Cell Death Discovery. 2022;8(1):248.
27. Cao Y, Yi Y, Han C, Shi B. NF-κB signaling pathway in tumor microenvironment. Front Immunol. 2024;15:1476030.
28. Yanai H, Hangai S, Taniguchi T. Damage-associated molecular patterns and Toll-like receptors in the tumor immune microenvironment. Int Immunol. 2021;33(12):841–6.
29. Mukherjee S, Patra R, Behzadi P, Masotti A, Paolini A, Sarshar M. Toll-like receptor-guided therapeutic intervention of human cancers: molecular and immunological perspectives. Front Immunol. 2023;14:1244345.
30. Madahar SS, Gideon A, Abdul-Sater AA. Nod-like receptors in inflammatory arthritis. Biomedical J. 2024;47(1):100655.
31. Ye X, Liu X, Yin N, Song W, Lu J, Yang Y, et al. Successful first-line treatment of simultaneous multiple primary malignancies of lung adenocarcinoma and renal clear cell carcinoma: A case report. Front Immunol. 2022;13:956519.
32. Mackenzie NJ, Nicholls C, Templeton AR, Perera MP, Jeffery PL, Zimmermann K, et al. Modelling the tumor immune microenvironment for precision immunotherapy. Clin Translational Immunol. 2022;11(6):e1400.

33. O'Connell GC, Treadway MB, Tennant CS, Lucke-Wold N, Chantler PD, Barr TL. Shifts in leukocyte counts drive the differential expression of transcriptional stroke biomarkers in whole blood. Translational Stroke Res. 2019;10(1):26–35.
34. Qi C, Lei L, Hu J, Wang G, Liu J, Ou S. Identification of a five-gene signature deriving from the vacuolar ATPase (V-ATPase) sub-classifies gliomas and decides prognoses and immune microenvironment alterations. Cell Cycle. 2022;21(12):1294–315.
35. Kong X-M, Zhang G-H, Huo Y-K, Zhao X-H, Cao D-W, Guo S-F, et al. MicroRNA-140-3p inhibits proliferation, migration and invasion of lung cancer cells by targeting ATP6AP2. Int J Clin Exp Pathol. 2015;8(10):12845.
36. Wei J, Huang X-J, Huang Y, Xiong M-Y, Yao X-Y, Huang Z-N, et al. Key immune-related gene ITGB2 as a prognostic signature for acute myeloid leukemia. Annals Translational Med. 2021;9(17):1386.
37. Celiksoy MH, Köker MY, Gezdirici A, Ozsoy S, Malbora B, Gungor S. A novel ITGB2 variant with long survival in patients with leukocyte adhesion defect type-I. Immunol Res. 2021;69(5):461–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.