



OPEN ACCESS

EDITED BY

Kang-Seo Park,
University of Ulsan, Republic of Korea

REVIEWED BY

Hye Jin You,
National Cancer Center, Republic of Korea
Fei Fang,
Michigan State University, United States
Remigiusz Marcin Bachor,
University of Wrocław, Poland
Oscar Medina-Contreras,
Mexico Children's Hospital, Mexico

*CORRESPONDENCE

Toshihide Nishimura
✉ t-nishimura@marianna-u.ac.jp

[†]These authors have contributed equally to this work

RECEIVED 07 August 2023

ACCEPTED 30 January 2024

PUBLISHED 21 February 2024

CITATION

Nishimura T, Végvári Á, Nakamura H, Fujii K, Sakai H, Naruki S, Furuya N and Saji H (2024) Cancer cell immunity-related protein co-expression networks are associated with early-stage solid-predominant lung adenocarcinoma. *Front. Oncol.* 14:1273780. doi: 10.3389/fonc.2024.1273780

COPYRIGHT

© 2024 Nishimura, Végvári, Nakamura, Fujii, Sakai, Naruki, Furuya and Saji. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cancer cell immunity-related protein co-expression networks are associated with early-stage solid-predominant lung adenocarcinoma

Toshihide Nishimura^{1*†}, Ákos Végvári^{2†}, Haruhiko Nakamura^{1†}, Kiyonaga Fujii^{1,3†}, Hiroki Sakai^{1†}, Saeko Naruki^{4†}, Naoki Furuya⁵ and Hisashi Saji¹

¹Department of Chest Surgery, St. Marianna University School of Medicine, Kawasaki, Japan, ²Division of Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Biomedicum, Stockholm, Sweden, ³Laboratory of Analytical Chemistry, Daiichi University of Pharmacy, Fukuoka, Japan, ⁴Department of Pathology, St. Marianna University Hospital, Kawasaki, Japan, ⁵Division of Respiratory Medicine, Department of Internal Medicine, St. Marianna University School of Medicine, Kawasaki, Japan

Background: Solid-predominant lung adenocarcinoma (SPA), which is one of the high-risk subtypes with poor prognosis and unsatisfactory response to chemotherapy and targeted therapy in lung adenocarcinoma, remains molecular profile unclarified. Weighted correlation network analysis (WGCNA) was used for data mining, especially for studying biological networks based on pairwise correlations between variables. This study aimed to identify disease-related protein co-expression networks associated with early-stage SPA.

Methods: We assessed cancerous cells laser-microdissected from formalin-fixed paraffin-embedded (FFPE) tissues of a SPA group ($n = 5$), referencing a low-risk subtype, a lepidic predominant subtype group (LPA) ($n = 4$), and another high-risk subtype, micropapillary predominant subtype (MPA) group ($n = 3$) and performed mass spectrometry-based proteomic analysis. Disease-related co-expression networks associated with the SPA subtype were identified by WGCNA and their upstream regulators and causal networks were predicted by Ingenuity Pathway Analysis.

Results: Among the forty WGCNA network modules identified, two network modules were found to be associated significantly with the SPA subtype. Canonical enriched pathways were highly associated with cellular growth, proliferation, and immune response. Upregulated HLA class I molecules HLA-G and HLA-B implicated high mutation burden and T cell activation in the SPA subtype. Upstream analysis implicated the involvement of highly activated oncogenic regulators, MYC, MLXIPL, MYCN, the redox master regulator NFE2L2, and the highly inhibited LARP1, leading to oncogenic IRES-dependent translation, and also regulators of the adaptive immune response, including highly activated IFNG, TCRD, CD3-TCR, CD8A, CD8B, CD3, CD80/CD86, and highly inhibited LILRB2. Interestingly, the immune checkpoint molecule HLA-G, which is the counterpart of LILRB2, was highly expressed characteristically in the SPA subtype and might be associated with antitumor immunity.

Conclusion: Our findings provide a disease molecular profile based on protein co-expression networks identified for the high-risk solid predominant adenocarcinoma, which will help develop future therapeutic strategies.

KEYWORDS

solid predominant adenocarcinomas of the lung, WGCNA, data-driven co-expression protein networks, upstream regulator analysis, laser microdissection, proteomics, mass spectrometry

1 Introduction

Lung cancer is the leading cause of death globally, of which lung adenocarcinoma is the most common pathological subtype. Invasive nonmucinous adenocarcinoma is primarily categorized into five histopathological subtypes: lepidic, acinar, papillary, micropapillary, and solid, based on the 2021 World Health Organization Classification of Lung Tumors (1). Micropapillary- and solid-predominant lung adenocarcinoma (MPA and SPA, respectively) are high-risk subtypes with high metastatic potential and the worst prognosis. In contrast, low-risk subtypes characterized by well- and moderately-differentiated morphologies, such as lepidic predominant adenocarcinoma (LPA) have a favorable prognosis (2, 3). Lung adenocarcinomas usually contain complex mixtures of different subtypes. MPA exhibits a micropapillary pattern, which is the primary histological pattern assessed semi-quantitatively in 5% increments in resected specimens. It is associated with lymphatic invasion, pleural invasion, and lymph node metastases (4).

Caso et al. conducted a genomic characterization of prognostically important predominant histologic subtypes of lung adenocarcinoma and found that MPA and SPA exhibited a higher tumor mutational burden, increased chromosomal instability (CIN), higher APOBEC (the enzyme with DNA mutagenesis function) mutational signatures, more oncogenic pathway alterations, and the lowest frequency of targetable genomic alterations among the subtypes (5). Micropapillary or solid patterns are risk factors for predicting poor recurrence-free survival in early-stage IA lung adenocarcinoma (6). Recently, Jeon et al. compared the clinicopathological features and clinical course of patients with the MPA and SPA subtypes, including predominant and non-predominant subtypes after curative resection of stage I lung adenocarcinoma and analyzed the prognostic factors. The clinical results were different for stage I high-grade adenocarcinoma and the predominant micropapillary subtype was an independent prognostic factor for recurrence, whereas the solid subtype was a significant factor for overall survival (7).

A pivotal challenge is to unravel the underlying cancer biology of those high-risk lung adenocarcinomas and how tumor morphologic appearances are linked to malignant clinical outcomes. Solid predominant adenocarcinoma (SPA) shows a

major component of polygonal tumor cells forming sheets that lack recognizable patterns of adenocarcinoma, such as acinar, papillary, micropapillary, or lepidic growth. SPA subtype and the presence of solid pattern are associated with numerous poor prognostic factors, including higher mitotic count, high risk of occult lymph node metastases, thyroid transcription factor-1 (TTF-1) negativity, and less frequent epidermal growth factor receptor (EGFR) mutations (8). However, molecular profiles characterizing the SPA subtype remain unclear. In this study, we focus on identifying disease-related co-expression protein networks and their upstream regulators associated with the SPA subtype.

Mass spectrometry (MS)-based proteomics has proven feasible in the identification and quantification of proteins expressed in clinical specimens. Quantitative proteome data can be used to identify key disease-related proteins and therapeutic targets (9). We have adopted label-free spectral counting-based semiquantitative MS-based proteomics, following the collection of target cancerous cells from formalin-fixed paraffin-embedded (FFPE) tumor specimens by laser microdissection (LMD). This study aimed at identifying the co-expression protein networks associated with the early-stage high-risk lung adenocarcinoma SPA, by comparing the early-stage low-risk subtype LPA and another high-risk subtype MPA. Upstream regulator and causal network analysis (10) were performed for data-driven protein co-expression networks significant to SPA, obtained by the weighted gene co-expression network analysis (WGCNA), an unsupervised clustering method based on the correlation network expression (11), applied to quantitative proteome datasets.

2 Materials and methods

2.1 FFPE tissue specimens and sample preparation

Of 1,293 patients who underwent surgery for lung cancer at St. Marianna University Hospital between 2000 and 2020, 186 (14%), 12 (1%), and 49 (4%) had tumors that were histologically confirmed to represent the LPA, MPA, and SPA subtypes, respectively (Figure 1A). SPA, a high-risk subtype, was contrasted to LPA, a low-risk subtype. The pathological specimens were independently

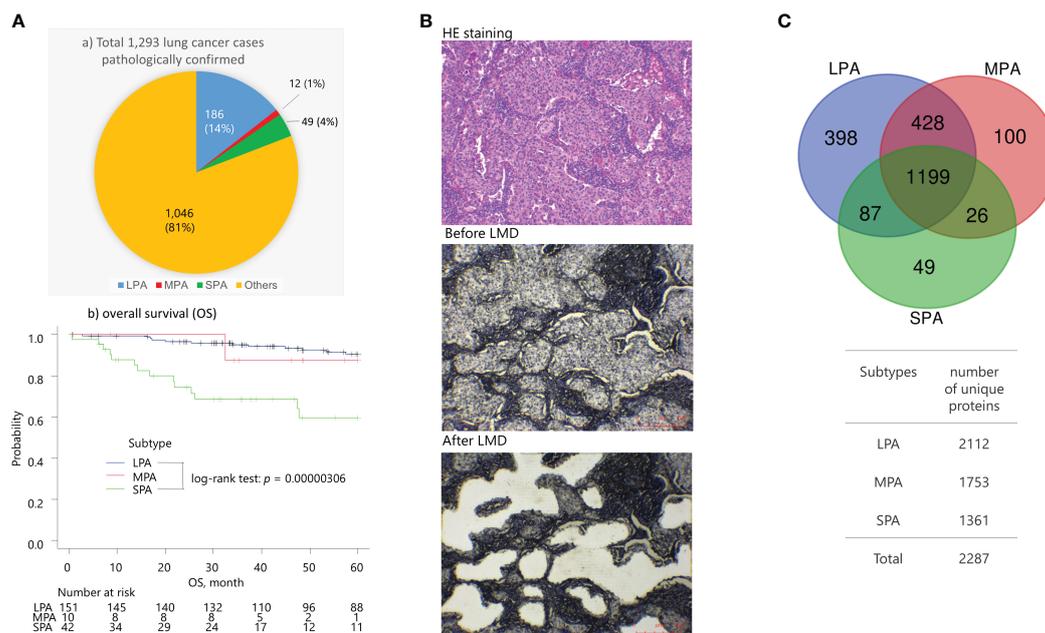


FIGURE 1

(A) 1,293 lung cancer patients were pathologically confirmed and received surgical operations between January 2008 to April 2022, at the St. Marianna University School of Medicine Hospital (Kawasaki, Kanagawa, Japan). The proportions of subtypes were LPA (14%), MPA (1%), SPA (4%), and others (81%). The SPA cases showed a worse overall survival (OS) compared with the LPA cases (log-rank test: $p = 3.06 \times 10^{-5}$), whereas the MPA cases were not statistically significant due to the limited number of cases against the LPA and SPA cases. (B) Representative images of tumor tissues from solid predominant adenocarcinomas (SPA) (sample SPA_T21) stained with hematoxylin & eosin (HE) using standard histological methods, and only with hematoxylin prior to and after laser microdissection (LMD). Scale bars in red color are indicated at the bottom right. (C) Venn map of the identified proteins.

reviewed by two pathologists (H. N. and S. N.) to confirm that they fulfilled the 2015 World Health Organization classification criteria for lung tumors (histological criteria) (12). FFPE tumor tissue blocks from 12 surgical specimens histologically confirmed as lung LPA, MPA, and SPA were obtained without patient identifiers from the St. Marianna University School of Medicine Hospital. Informed consent was obtained from all participating subjects. The protocol was approved by the Institutional Review Board of St. Marianna University School of Medicine (approval no. 1461) and the study adhered to the Helsinki Declaration.

For tissue microdissection, 10- μm -thick sections from the FFPE tumor blocks were cut and placed on DIRECTOR slides (OncoPlex Diagnostics Inc., Rockville, MD, USA). The sections were deparaffinized and stained with hematoxylin using standard protocols before dissection. Microdissection was performed using a Leica LMD7 microdissection microscope (Leica, Wetzlar, Germany) (Figure 1B). A total area of 4 mm^2 with approximately 15,000 tumor cells was directly transferred from the FFPE sections via laser dissection into the cap of a 200 μL low-binding tube. Proteins were extracted and digested with trypsin using the Liquid TissueTM MS Protein Prep kit (OncoPlex Diagnostics Inc.) according to the manufacturer's instructions (13). Briefly, dried microdissected pellets were suspended in 20 μL of Liquid Tissue buffer and heated at 95°C for 90 min and then cooled on ice for 3 min before the addition of 0.1 μg of trypsin. The tubes were then incubated at 37°C overnight. Dithiothreitol was added to a final concentration of 10 mM and the samples were heated for 5 min at

95°C. The digested samples were dried, resuspended in 50 μL of a 2% acetonitrile aqueous solution containing 0.1% trifluoroacetic acid, and stored at -20°C until analysis.

2.2 Proteomic analysis by LC-MS/MS

A label-free quantitation approach using spectral counting by LC-MS/MS was adopted for global proteomic analysis. The digested samples (5 μL for a single run) were analyzed in triplicate by LC-MS/MS using a reverse-phase LC system interfaced with a Q Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) via a nano-electrospray ionization device (AMR Inc., Tokyo, Japan). The mass spectrometer was operated in data-dependent acquisition mode. Full-scan MS spectra were acquired in the range of m/z 350–1600 at a resolution of 70,000. The top ten most intense peaks from the survey scan were selected for fragmentation with higher-energy collisional dissociation with a normalized collision energy of 27% and isolation window of m/z 1.6. The dynamic exclusion time for precursor ions selected for MS/MS fragmentation was 15 s, and the automatic gain control target values for MS and MS/MS were 1×10^6 and 1×10^5 , respectively.

The LC system consisted of an Ultimate3000 HPLC System (Thermo Fisher Scientific), a trap cartridge (0.3 $\text{mm} \times 5.0 \text{ mm}$, CERI, Tokyo, Japan), and a capillary separation column (Zaplous column alpha-PepC18, 3 μm , 12 nm , 0.1 $\text{mm} \times 150 \text{ mm}$, AMR Inc.) fitted with an emitter tip (FortisTip, OmniSeparo-TJ), Hyogo,

Japan). An auto-sampler (HTC-PAL, CTC Analytics, Zwingen, Switzerland) was used to load the samples into the trap, which was then washed with solvent A (2% acetonitrile aqueous solution containing 0.1% formic acid) to concentrate and desalt the peptides in the trap. Subsequently, the trap was connected in series to the separation column, and the peptides were eluted from the whole column with 0.1% formic acid aqueous solution and acetonitrile using a linear 5%–40% acetonitrile concentration gradient over 90 min at a flow-rate of 500 nL min⁻¹.

2.3 Protein identification

The raw data were processed using PatternLab for Proteomics software v4.0. Peptide sequence matching (14) was done using the Comet algorithm against the UniProt *Homo sapiens* database. A target-reverse strategy was employed for increased confidence in protein identification. This search considered tryptic peptide candidates as well as the formylation of lysine and oxidation of methionine as variable modifications. The Comet search engine considered a precursor mass tolerance of 40 ppm and a fragment bin tolerance of 0.02 Da. The validity of the peptide spectrum matches was assessed using the Search Engine Processor (SEPro) module of PatternLab. The acceptable FDR for spectra, peptide, and protein were 3%, 2%, and 1%, respectively. The expressions of the identified proteins were assessed using spectral count-based protein quantification. The spectral count represents the number of MS/MS spectra assigned to each protein.

2.4 Weighted correlation network analysis

The similarity in protein expression patterns for all protein pairs was calculated according to their pairwise Pearson's correlation coefficient [i.e., the similarity between proteins *i* and *j* was defined as $(1-r_{ij})/2$, where r_{ij} is the Pearson's correlation coefficient of the protein expression pattern between the two proteins]. We performed a network topology analysis for various soft-thresholding powers ranging from 1 to 2 to choose an optimal value of balance between independence and mean connectivity. A topological overlap matrix (TOM) that considers topological similarities between a pair of proteins in the network was then generated from the resultant scale-free co-expression network. We generated a tree by hierarchical clustering using dissimilarity according to TOM (1–TOM), and protein modules were determined using dynamic tree-cutting to trim the branches (11).

The modules were summarized by the first principal component, which is referred to as eigen proteins in the text, as they express the highest connectivity in the module. Module membership, defined as the correlation between the protein expression profile and the module eigen-protein, was measured with values ranging from 0 to 1, with “0” representing a gene that is not part of the module and “1” representing high connectivity with the module. Subsequently, the module-trait association was

determined using the correlation between the module eigen-protein and the three subtypes: MPA, SPA, and LPA. A protein module was summarized by the top hub protein (referred to as “eigen-protein”) with the highest connectivity in the module. The WGCNA analysis was performed using the WGCNA R-package (11) implemented in RStudio.

2.5 Protein-PPI network construction

We used the STRING database (version 11.5) (<https://string-db.org/>) to construct a PPI network for a protein module (15). STRING networks were calculated under the criteria for linkage with experiments, databases, text mining, and co-expression using the default settings (medium confidence score: 0.400; network depth: 0 interactions). Functional enrichment results were obtained for canonical pathways with a *p*-value <0.05. Proteins in a module were mapped in the PPI network from the STRING database to produce the results of the enrichment analysis regarding the biological process (GO) and Reactome pathways (HAS). Protein networks were subsequently exported to Cytoscape (version 3.9.1) (Institute for Systems Biology, Seattle, WA, USA: <https://cytoscape.org/>) (16) from the STRING database. The hub proteins in each module were identified according to their intramodular connectivity and their correlation with module eigen-proteins. The proteins inside the co-expression modules exhibit high connectivity and the proteins within the same module may play similar roles. The top 10 high-degree proteins were identified using the cytoHubba plugin (17). The top-ranked proteins in each module were considered hub proteins and designated “highly connected proteins.” Functional enrichment results were obtained for canonical pathways by considering a network bias-corrected *p*-value of <0.05 for statistical significance.

The multivariate correlation analysis (MVA) of semiquantitative key protein expressions was performed using the JMP software (SAS Institute, Cary, NC, USA), and which result was visualized using the Intervene Shiny App (<https://intervene.shinyapps.io/intervene/>) (18).

2.6 Upstream regulator and causal network analysis by IPA

Upstream regulators, causal networks, and canonical pathways were predicted using IPA software (<http://www.ingenuity.com>) (10). Quantile-normalized protein expression data of the selected modules were used as input datasets. Both the upstream regulators and causal networks (*p* < 0.05) predicted from the WGCNA network modules were significantly associated with the three subtypes (LPA, MPA, and SPA), in which the activation and the inhibition of a predicted network were defined by *z*-values that were >2.0 and <–2.0, respectively. The upregulation was defined by *z*-values > 1.5 and < 2.0, whereas downregulation was defined by *z*-values > –2.0 and < –1.5.

3 Results

3.1 Proteome datasets

At the St. Marianna University School of Medicine Hospital (Kawasaki, Kanagawa, Japan), 1,293 lung cancer patients were pathologically confirmed and underwent surgery between January 2008 to April 2022. These cases consisted of the following subtypes: LPA (14%), MPA (1%), SPA (4%), and others (81%) (Figure 1Aa). The SPA cases exhibited worse overall survival (OS) compared with the LPA cases (log-rank test: $p = 3.06 \times 10^{-5}$) (Figure 1Ab), whereas the MPA cases were not statistically significant because of the limited number of samples.

FFPE tissue specimens were obtained by surgical resection of early-stage (IA-IB) lung adenocarcinoma patients who were pathologically confirmed as having solid predominant (SPA, $n = 5$), micropapillary (MPA, $n = 3$), and lepidic (LPA, $n = 4$) subtypes (Table 1). Presurgical treatment was not performed for any of the lung adenocarcinoma patients. MS-based proteomic analysis was carried out on approximately 15,000 cancerous cells collected from the FFPE specimens by laser microdissection (LMD) (Figure 1B). A total of 2,287 proteins were identified, of which 1,199 (52.4%) were commonly expressed in the cancerous cells. LPA, MPA, and SPA contained 398 (17.4%), 100 (4.4%), and 49 (2.1%) unique proteins, respectively (Figure 1C).

Overall, 2,112, 1,753, and 1,361 proteins were identified from LPA, MPA, and SPA, respectively. A gene ontology (GO) analysis was performed using the Protein Analysis Through Evolutionary Relationships (PANTHER, version 17.0) program (Paul D. Thomas, University of Southern California, Los Angeles, CA, USA) (19) and revealed similar results among the three subtypes (Supplementary Figure S1). Volcano plots were generated from the protein expression data obtained by SimpliFi™ software (PROTIFI, Farmingdale, NY, USA; <https://simplifi.protifi.com>). Upregulated proteins significant to SPA included large ribosomal subunit protein uL29 (RPL35), human leukocyte antigen G (HLA-G), GTP-binding nuclear protein Ran (RAN), human leukocyte antigen B (HLA-B), and putative heat shock protein HSP90-beta-3 (HSP90AB3P). The expression of HLA-G protein was significantly associated with and highly upregulated in the SPA subtype (Figure 2A), which was confirmed by its ANOVA test with a high significance to the SPA subtype with p -ANOVA = 0.0026 (Figure 2C).

3.2 Identification of protein co-expression networks by WGCNA

Weighted correlation network analysis, also known as weighted gene co-expression network analysis (WGCNA) incorporates traditional data exploration techniques, but its intuitive network language and analysis framework go beyond standard analysis techniques. It uses a network approach and is suitable for the integration of complementary genomic/proteomic datasets so that

it can be interpreted as a data analysis technique for systems biology (11).

Following hierarchical clustering of the samples based on protein abundance (Figure 3A), a WGCNA analysis (11) was performed with a soft threshold power of 10, which was selected to approximate a scale-free topology, a minimum module size of 10, and a module detection sensitivity (*deepSplit*) of 4. Correlations between the resultant modules and the traits were obtained to identify protein modules that were significantly associated with the respective traits. Forty protein modules were identified by clustering all of the proteins and constructing weighted protein co-expression networks. The protein cluster dendrogram is presented in Figure 3B. A heatmap of eigen-protein expression for pairwise correlations between the modules in the connectivity measure (kME) of the module eigen-protein are presented in Figure 3C, together with lists of module IDs, module colors, and their eigen-proteins representing protein expressions in a module.

Six WGCNA modules were identified with high correlations ($r > 0.5$) and statistical significance (multiple testing correction using the Benjamini-Hochberg method: $q < 0.05$) with clinical traits (Figure 4). The WM25 (cyan: $r = 0.87$, $q = 0.0102$) and WM33 (green-yellow: $r = 0.81$, $q = 0.0286$) modules were significantly correlated with the LPA subtype. The WM23 (dark-grey: $r = 0.91$, $q = 0.014$) and WM24 (dark-orange: $r = 0.94$, $q = 0.0478$) modules were significantly correlated with the MPA subtype, and WM1 (pink: $r = 0.82$, $q = 0.0216$) was well correlated with the SPA subtype.

Several other WGCNA modules were also correlated ($r > 0.5$) with one of the three traits. WM8 (dark-magenta: $r = 0.76$), WM19 (light-cyan: $r = 0.61$), and WM32 (violet: $r = 0.61$) correlated with the LPA subtype, WM11 (dark-red: $r = 0.72$) and WM18 (orange: $r = 0.59$) with the MPA subtype, and WM2 (turquoise: $r = 0.57$) with the SPA subtype. However, none of these correlations were significant ($q > 0.05$). Trait correlation analysis often tends to overlook important modules. Statistical over-representative analysis (ORA) may help evaluate potential WGCNA modules concerning overlap with uniquely upregulated proteins for each trait. The numbers of proteins in each subtype more than twice as highly expressed compared with other subtypes were 407 in LPA, 321 in MPA, and 482 in SPA, respectively. Overlap of the WGCNA modules with that of the protein groups was assessed using ORA (Figure 4). In this study, we focused on the WGCNA modules, WM1 (pink) and WM2 (turquoise), significantly associated with the high-risk subtype, SPA (Figure 4).

3.3 Protein-protein interaction networks and functional enrichment

Using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database version 11.5 (<https://string-db.org/>) (15), human PPI networks were obtained for the WGCNA modules. The PPI networks for the WM1 and WM2 modules associated with the SPA subtype were reconstructed using Cytoscape (version 3.9.1)

TABLE 1 Clinicopathological information of the recruited patients.

Variable	Category	No. patients	%
Gender			
	Female	5	41.7
	Male	7	58.3
Age(y)			
	Average ± SD	64.4 ± 5.5	
Smoking index (Brinkmann Index, BI)			
	Female		
	BI=0	4	80.0
	0 < BI ≤ 400	1	20.0
	400 < BI ≤ 600	0	0.0
	600 < BI ≤ 1200	0	0.0
	BI > 1200	0	0.0
	Male		
	BI=0	2	28.6
	0 < BI ≤ 400	1	14.3
	400 < BI ≤ 600	1	14.3
	600 < BI ≤ 1200	1	14.3
	BI > 1200	2	28.6
Histologic type			
	Adenocarcinoma	12	100.0
Subtype			
	Lepidic predominant adenocarcinoma (LPA)	4	33.3
	Micropapillary adenocarcinoma (MPA)	3	25.0
	Solid adenocarcinoma (SPA)	5	41.7
Surgical method			
	Radical lobectomy	11	66.7
	Limited resection	1	33.3
Tumor size on CT			
	T1a (≤1cm)	0	0.0
	T1b (1-2cm)	2	16.7
	T1c (2-3cm)	4	33.3
	T2a (3-4cm)	4	33.3
	T2b (4-5cm)	2	16.7
	T3 (5-7cm)	0	0.0
	T4 (>7cm)	0	0.0
Clinical stage			
	cIA	6	50.0

(Continued)

TABLE 1 Continued

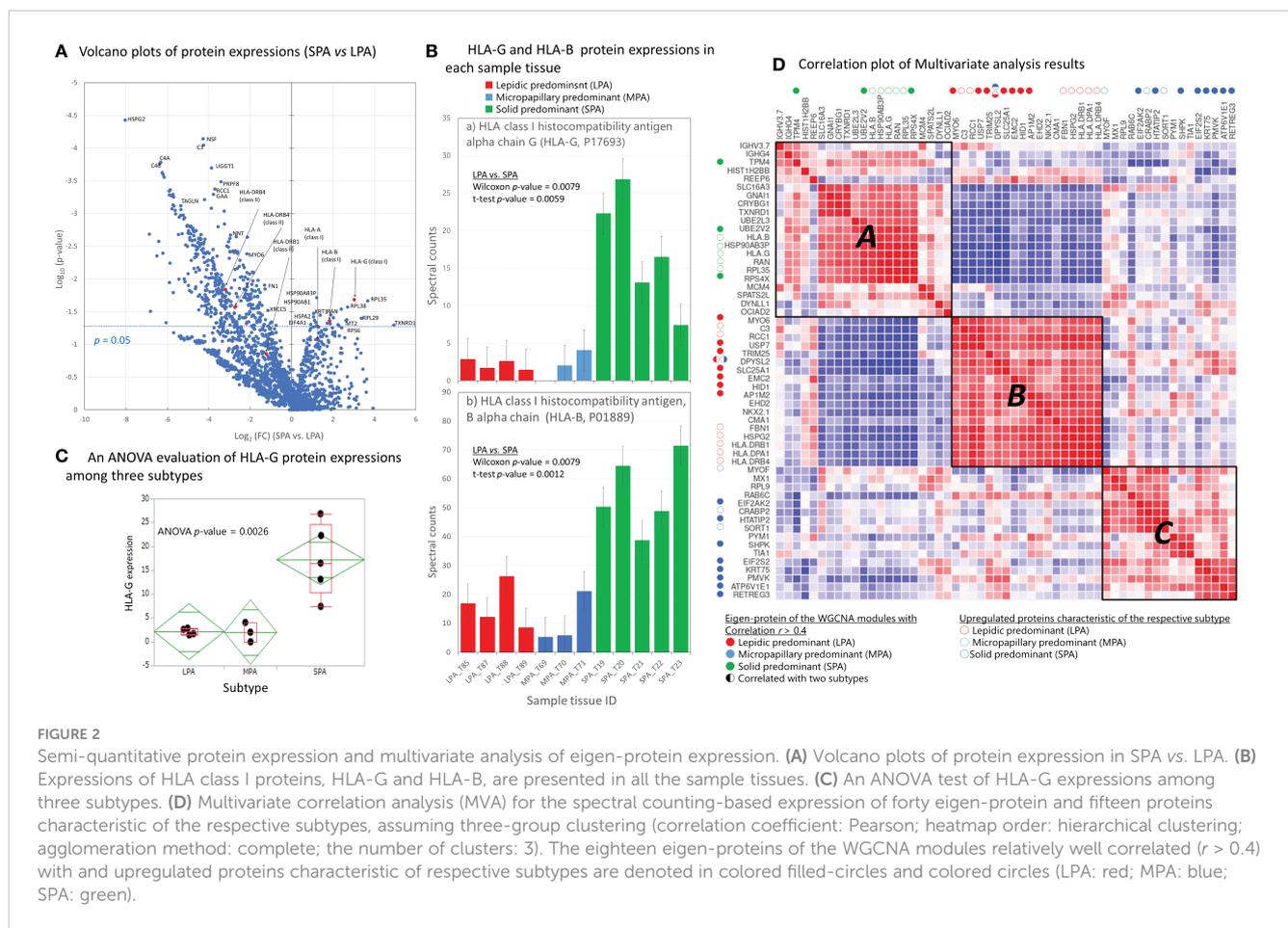
Variable	Category	No. patients	%
Clinical stage			
	cIIA	0	0.0
	cIB	6	50.0
	cIIB	0	0.0
	cIV	0	0.0
EGFR mutation status			
	Positive		
	L858R	1	8.3
	Ex19 E746-A750 del	3	25.0
	Negative		
	Neither L858R nor Ex19del	8	66.7

LPA, lepidic predominant adenocarcinoma; MPA, micropapillary adenocarcinoma; SPA, solid adenocarcinoma; BI, Brinkmann Index.

software (Institute for Systems Biology, Seattle, WA, USA: <https://cytoscape.org/>) (16) (Supplementary Figure S2). Top hub proteins were calculated using the *cytoHubba* plugin with maximal clique centrality (MCC) (17). In these data-driven protein co-expression networks, eigen-proteins are indicated in red letters, hub proteins in red to orange fill colors, and some key proteins are denoted with red borders. Top STRING enrichment results for the WGCNA modules, WM1 and WM2, significant to the SPA subtype are shown in Supplementary Table S1.

The functional enrichment obtained for the WM1 module included the following: (i) SRP-dependent co-translational protein targeting to membrane, translational initiation, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, and protein targeting to membrane as biological processes (GO); and (ii) L13a-mediated translational silencing of ceruloplasmin expression, GTP hydrolysis and joining of the 60S ribosomal subunit, eukaryotic translation termination, and NMD independent of the exon junction complex (EJC) (Supplementary Table S1). The eigen-protein RPS4X (small ribosomal subunit protein eS4, also known as SCR10) is a component of the 40S subunit and is involved in L13a-mediated translational silencing of ceruloplasmin expression, eukaryotic translation termination, and NMD independent of the exon junction complex (EJC).

The pathways enriched for the WM2 (turquoise) module included the following: (i) antigen processing and presentation of peptide antigen via MHC class I, NIK/NF-κB signaling, and regulation of transcription from RNA polymerase II promoter in response to hypoxia as biological processes (GO); (ii) regulation of expression of SLITs and ROBOs, cross-presentation of soluble exogenous antigens (endosomes), and NIK to noncanonical NF-κB signaling (Supplementary Table S1). The eigen-protein tropomyosin 4 (TPM4) is a member of the tropomyosin family of



actin-binding proteins, which is involved in stabilizing cytoskeleton actin filaments.

3.4 Semi-quantitative protein expression and multivariate correlation analysis

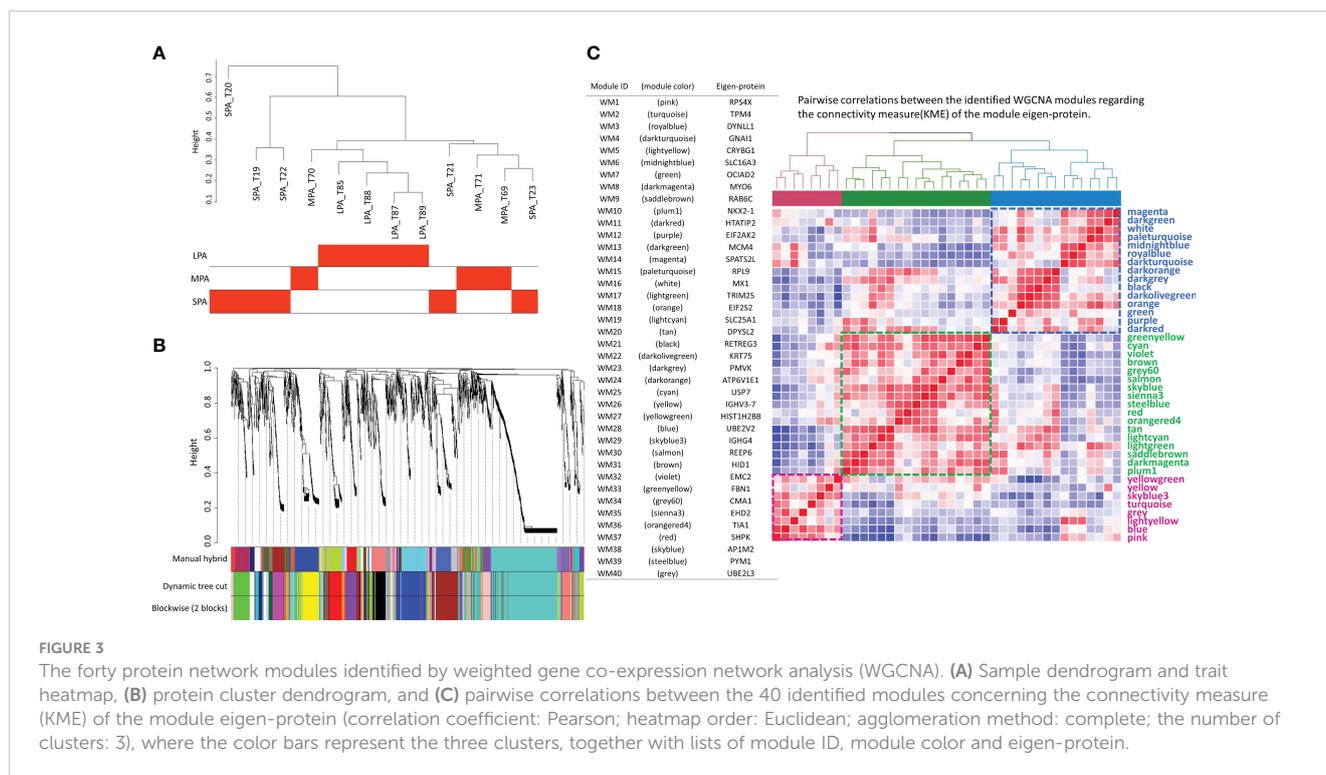
Forty eigen-proteins together with fifteen key proteins, which were representatively expressed throughout all 40 modules, were subjected to multivariate correlation analysis (MVA), assuming three-group clustering (Figure 2D). The eighteen eigen-proteins of the WGCNA modules relatively well correlated ($r > 0.4$) with respective subtypes are denoted in colored filled circles (LPA: red; MPA: blue; SPA: green). Those eighteen eigen-protein expressions were clustered into Cluster A, Cluster B, and Cluster C, which well corresponded to the SPA, LPA, and MPA subtypes, respectively. (Figure 2D). Upregulated proteins characteristic of the respective subtypes were also well clustered consistently: HLA-G and HLA-B in SPA; C3, RCC1, and HSPG2 in LPA; MYOF and CRABP2 in MPA.

Zhou et al. (20) performed proteomic analyses between low-risk and high-risk subtypes of early-stage lung adenocarcinomas, where the low-risk LPA subtype group ($n = 31$) and the high-risk subtype group ($n = 28$) consisting of both MPA and SPA and reported several proteins differentially expressed more to the high-risk group.

Their results were not well linked to our proteomic observations. Some proteins differentially expressed in their high-risk subtype group, such as DNA replicating licensing factors MCMs, were oppositely upregulated in the low-risk subtype LPA subtype in our study, except that the expression of upregulated P4HA2 (Prolyl-4 hydroxylase subunit alpha-2) was observed in the high-risk subtype MPA ($p = 0.03$). There are intrinsic differences between their approach and ours. They directly used sections of tissues collected immediately after resection for proteomic analysis while we collected only cancerous cells directly from the FFPE tissue sections via laser dissection, and they grouped both MPA and SPA as one high-risk group while we did not group MPA and SPA into one.

3.5 Canonical pathways enriched by IPA

The canonical pathways and their pathway categories enriched were obtained by the IPA software (<http://www.ingenuity.com>) (10) for the WM1 and WM2 modules significant to the SPA subtype (Figure 5). The WM1 network module was associated predominantly with activated pathways of EIF2 signaling, which are involved in the following pathway categories: Cellular Growth, Proliferation and Development, Cellular Stress and Injury, and Intercellular and Second Messenger Signaling (Figure 5A). The



WM2 network module was associated with various pathways. The top pathway category was Cellular Immune Response, which included the FAT10 signaling pathway, caveolar-mediated endocytosis signaling, antigen presentation pathway, upregulated leukocyte extravasation signaling, and activated Fcγ receptor-mediated phagocytosis in macrophages and monocytes. The pathway category Cellular Stress and Injury included the FAT10 signaling pathway, inhibition of ARE-mediated mRNA degradation pathway, regulation of eIF4 and p70S6K Signaling, activated pathways of EIF2 signaling, activated NRF2-mediated oxidative stress response, senescence, leukocyte extravasation signaling, hepatic fibrosis signaling, HIF1A signaling, and necroptosis signaling (Figure 5B).

3.6 Upstream regulators and causal networks enriched by IPA

Upstream regulator and causal network analysis was performed for the WGCNA modules using IPA software (<http://www.ingenuity.com>) (10). Table 2 summarizes the top upstream regulators, causal networks, canonical pathways, and diseases or functions predicted for the four WGCNA modules.

Highly activated upstream regulators in the WM1 protein networks included MLXIPL ($z = 4.80$, p -value of overlap = 4.16×10^{-31}), MYCN ($z = 4.12$, p -value = 6.30×10^{-17}), MYC ($z = 4.87$, p -value = 2.60×10^{-24}), Lh ($z = 4.87$, p -value = 6.30×10^{-17}), DDX3X ($z = 3.13$, p -value = 9.48×10^{-17}), TCR ($z = 2.91$, p -value = 1.94×10^{-10}), CD3 ($z = 3.16$, p -value = 1.40×10^{-7}), NFE2L2 ($z = 2.97$, p -value = 0.0001), and both CD8A and CD8B ($z = 4.15$, p -value = 0.0001), whereas highly inhibited regulators included LARP1 ($z =$

-5.10 , p -value = 7.19×10^{-44}) and RICTOR ($z = -4.24$, p -value = 3.80×10^{-17}) (Table 2). Molecules in the WM1 module targeted by LARP1, MYCN, and TCR included the eigen-protein RPS4X. The carbohydrate-responsive element-binding protein (ChREBP) encoded by *MLXIPL* is a basic helix-loop-helix leucine zipper transcription factor of the MYC/MAX/MAD superfamily. Its activation promotes cancer cell proliferation and inhibits cancer cell apoptosis by promoting aerobic glycolysis in various cancers (21).

CD3, CD8A, CD8B, and TCR (T-cell receptor) are drivers of the immune system and the immune defense against intracellular pathogens, such as viruses and bacteria, as well as tumors. CD8 is expressed on the surface of cytotoxic T lymphocytes (CTLs), dendritic cells, macrophages, monocytes, and NK cells and exists as either a homodimer (two CD8A chains) or heterodimer (CD8A and CD8B chains), which are co-receptors of the TCR-CD3 complex and binds to MHC-I to activate T cells. CD8A recruits T-cell-specific protein tyrosine kinase (LCK) or linker for activation of T-cells family member 1 (LAT) to activate downstream signaling pathways, whereas CD8B increases the avidity of the CD8/MHC/TCR complex to increase IL-2 production.

Highly inhibited La-related protein 1 (LARP1), the master regulator of the cap-dependent top mRNA translation, strongly suggests the inhibition of protein synthesis via cap-dependent mRNA translation or the activation of the cap-independent, IRES-mediated translation of mRNA subsets encoding oncogenic proteins, such as HIF1A, and MYC (22).

Regarding the WM2 module, highly activated upstream regulators included MYC, NFE2L2, MLXIPL, IFNG ($z = 4.51$, p -value of overlap = 1.39×10^{-8}), NFκB (complex) ($z = 3.87$, $p = 8.69 \times 10^{-5}$), FGFR1 ($z = 7.46$, network bias-corrected $p = 0.0005$), TCRD

Module ID.	Module color	#Module proteins	Correlation (r)			p-value			BH-corrected q-value			Max q-value	# of overlaps between module proteins and upregulated proteins in each subtype*			Hypergeometric overlap BH-corrected q-value			Max q-value
			LPA	MPA	SPA	LPA	MPA	SPA	LPA	MPA	SPA		LPA	MPA	SPA	LPA	MPA	SPA	
WM1	pink	91	-0.7	-0.18	0.82	0.0120	0.5830	0.0011	0.1200	0.8040	0.0216	0.0216	6	0	7	0.0069	1.0000	0.0035	0.0035
WM2	turquoise	516	-0.34	-0.28	0.57	0.2750	0.3790	0.0512	0.6110	0.8920	0.2280	0.2276	3	0	316	8.52E-43	1.0000	2.89E-126	2.89E-126
WM3	royalblue	36	-0.33	0.063	0.26	0.2890	0.8450	0.4080	0.6080	0.8890	0.6530	0.6084	0	0	5	1.0000	1.0000	0.6529	0.6529
WM4	darkturquoise	31	-0.15	-0.2	0.32	0.6360	0.5420	0.3140	0.7710	0.8030	0.5980	0.5981	0	0	8	1.0000	1.0000	0.8119	0.8119
WM5	lightyellow	36	-0.16	-0.14	0.28	0.6130	0.6590	0.3770	0.7660	0.8240	0.6280	0.6283	0	1	14	1.0000	0.0987	0.0334	0.0334
WM6	midnightblue	47	-0.36	0.38	0.008	0.2480	0.2170	0.9800	0.6610	0.8680	0.9800	0.6613	0	10	2	1.0000	0.2865	0.0065	0.0065
WM7	green	115	-0.074	-0.033	0.1	0.8200	0.9180	0.7580	0.8410	0.9420	0.8420	0.8410	2	3	21	3.24E-07	0.0002	0.8454	0.0000
WM8	darkmagenta	11	0.76	-0.016	-0.71	0.0045	0.9600	0.0099	0.0590	0.9600	0.0790	0.0593	6	0	0	0.0159	1.0000	1.0000	0.0159
WM9	saddlebrown	20	0.33	0.33	-0.61	0.2910	0.2970	0.0365	0.5820	0.9140	0.1830	0.1825	0	2	0	0.0171	1.0000	1.0000	0.0171
WM10	plum1	12	0.38	0.2	-0.54	0.2270	0.5250	0.0703	0.7570	0.8080	0.2560	0.2556	0	0	0	0.2373	1.0000	1.0000	0.2373
WM11	darkred	35	-0.17	0.72	-0.47	0.6000	0.0079	0.1200	0.8000	0.1060	0.3430	0.1056	0	25	0	0.0043	3.89E-14	1.0000	3.89E-14
WM12	purple	78	-0.16	0.52	-0.3	0.6090	0.0827	0.3440	0.7860	0.5510	0.6250	0.5513	0	52	1	1.0000	5.61E-27	2.83429E-06	5.61E-27
WM13	darkgreen	34	0.21	-0.28	0.043	0.5070	0.3770	0.8950	0.8110	0.9430	0.9680	0.8112	9	0	1	0.2536	1.0000	0.0140	0.0140
WM14	magenta	86	-0.18	-0.19	0.34	0.5710	0.5490	0.2750	0.8160	0.7840	0.5790	0.5789	0	0	50	1.0000	1.0000	1.96E-13	1.96E-13
WM15	paleturquoise	20	-0.018	0.25	-0.21	0.9570	0.4260	0.5210	0.9570	0.7750	0.7190	0.7186	1	3	0	0.2527	1.0000	1.0000	0.2527
WM16	white	20	-0.13	0.3	-0.14	0.6950	0.3480	0.6640	0.7510	0.9940	0.8300	0.7514	0	4	0	1.0000	0.7272	1.0000	0.7272
WM17	lightgreen	41	0.42	0.36	-0.73	0.1700	0.2440	0.0076	0.6800	0.8130	0.0759	0.0759	2	1	1	0.0472	0.0617	0.0045	0.0045
WM18	orange	26	-0.2	0.59	-0.32	0.5240	0.0425	0.3030	0.8060	0.4250	0.6060	0.4250	0	20	0	1.0000	1.98E-12	1.0000	1.98E-12
WM19	lightcyan	48	0.61	0.13	-0.7	0.0340	0.6980	0.0119	0.2720	0.8460	0.0793	0.0793	14	2	0	0.0171	0.0993	1.0000	0.0171
WM20	tan	57	0.46	0.53	-0.91	0.1280	0.0784	4.73E-05	0.6400	0.6270	0.0019	0.0019	0	2	3	1.0000	0.0418	0.0052	0.0052
WM21	black	101	-0.12	0.46	-0.29	0.7010	0.1290	0.3640	0.7380	0.5730	0.6330	0.5733	1	65	0	3.14E-07	1.57E-32	1.0000	1.57E-32
WM22	darkolivegreen	18	-0.25	0.47	-0.18	0.4380	0.1210	0.5780	0.8340	0.6050	0.7460	0.6050	0	0	0	1.0000	1.0000	1.0000	1.0000
WM23	darkgrey	29	-0.15	0.91	-0.66	0.6510	0.0000	0.0189	0.7660	0.0014	0.1080	0.0014	0	25	0	1.0000	1.66E-17	1.0000	1.66E-17
WM24	darkorange	23	-0.22	0.79	-0.48	0.5000	0.0024	0.1110	0.8330	0.0478	0.3420	0.0478	0	20	0	1.0000	3.17E-14	1.0000	3.17E-14
WM25	cyan	58	0.87	-0.11	-0.73	0.0003	0.7290	0.0069	0.0100	0.7880	0.0917	0.0102	27	0	1	6.98E-07	1.0000	1.94E-04	6.98E-07
WM26	yellow	175	0.37	-0.26	-0.12	0.2410	0.4100	0.7100	0.6890	0.8630	0.8350	0.6886	84	0	0	1.25E-21	1.0000	1.0000	1.25E-21
WM27	yellowgreen	15	0.35	-0.17	-0.18	0.2640	0.5930	0.5670	0.6210	0.7910	0.7560	0.6212	0	0	0	1.0000	1.0000	1.0000	1.0000
WM28	blue	200	-0.19	-0.26	0.41	0.5520	0.4110	0.1830	0.8180	0.8220	0.4310	0.4306	0	1	48	1.0000	6.17E-12	0.4509	6.17E-12
WM29	skyblue3	12	0.13	-0.16	0.015	0.6940	0.6300	0.9640	0.7710	0.8130	0.9890	0.7711	0	0	0	1.0000	1.0000	1.0000	1.0000
WM30	salmon	63	0.38	-0.25	-0.14	0.2290	0.4340	0.6640	0.7050	0.7550	0.8050	0.7046	29	0	0	3.61E-07	1.0000	1.0000	3.61E-07
WM31	brown	177	0.51	-0.28	-0.24	0.0904	0.3810	0.4460	0.5170	0.8470	0.6610	0.5166	121	0	0	3.52E-55	1.0000	1.0000	3.52E-55
WM32	violet	19	0.61	-0.37	-0.26	0.0350	0.2330	0.4210	0.2330	0.8470	0.6480	0.2333	18	0	0	1.90E-12	1.0000	1.0000	1.90E-12
WM33	greenyellow	67	0.81	-0.29	-0.52	0.0014	0.3570	0.0851	0.0290	0.9520	0.2840	0.0286	44	0	0	4.94E-18	1.0000	1.0000	4.94E-18
WM34	grey60	47	0.36	-0.26	-0.12	0.2530	0.4210	0.7180	0.6330	0.8020	0.8210	0.6325	25	0	1	1.14E-07	1.0000	0.0017	1.14E-07
WM35	sienna3	15	0.39	0.12	-0.47	0.2160	0.7210	0.1230	0.7850	0.8010	0.3280	0.3280	2	0	0	1.0053	1.0000	1.0000	1.0000
WM36	orangered4	10	0.24	0.22	-0.43	0.4460	0.4880	0.1650	0.8110	0.7810	0.4130	0.4125	0	1	0	1.0000	1.0000	1.0000	1.0000
WM37	red	105	-0.23	0.48	-0.21	0.4790	0.1110	0.5170	0.8330	0.6340	0.7390	0.6343	0	77	0	1.0000	2.29E-45	1.0000	2.29E-45
WM38	skyblue	21	0.46	0.12	-0.54	0.1350	0.7080	0.0674	0.6000	0.8330	0.2700	0.2696	4	0	0	0.9674	1.0000	1.0000	0.9674
WM39	steelblue	20	0.18	0.23	-0.37	0.5800	0.4700	0.2320	0.8000	0.7830	0.5160	0.5156	0	1	0	1.0000	0.5782	1.0000	0.5782
WM40	grey	13	-0.14	0.12	0.03	0.6650	0.7150	0.9270	0.7600	0.8170	0.9760	0.7600	0	1	0	1.0000	1.0000	1.0000	1.0000

FIGURE 4 Relationship between module eigen-proteins and the clinical traits of three subtypes LPA, MPA, and SPA. Each row in the embedded table represents weighted gene co-expression network analysis results for each module. The first and second columns in the table represent the module identification and color name of a module, respectively. The third column represents the number of proteins in each module. The table is color-coded by the correlation coefficient, in which the intensity and direction of the correlations are indicated (red, positive correlation; blue, negative correlation). The seventh to ninth columns present the p-values of the correlation coefficients, and the tenth to thirteenth columns are q-values by multiple testing correction using the Benjamini–Hochberg method, where significant p-value (<0.05) and q-values (< 0.1) are highlighted in bright red background. p-values between 0.05 and 0.1 are highlighted in yellow background. The fourteenth to sixteenth columns present a number of overlaps between module proteins and proteins upregulated in each subtype. The eighteenth to twenty-one columns present results of the statistical over-representative analysis (ORA) using the hypergeometric overlapping test of all the WGCNA modules with proteins upregulated (more than twice) uniquely for the individual subtypes: 397 proteins in LPA, 314 in MPA, and 474 in SPA, where the Benjamini–Hochberg corrected significant q-values (< 0.05) are highlighted in bright red background.

(T-cell receptor delta chain) ($z = 4.57, p = 0.0001$), CTNN β -LEF1 ($z = 4.52, p = 0.0001$), CTNN β 1 ($z = 4.40, p = 0.0001$), CD80 ($z = 7.31, p = 0.019$), CD86 ($z = 6.06, p = 0.019$), hypoxia-induced factors *Hif* ($z = 4.55, p = 0.0001$), and LCK ($z = 3.76, p = 0.0219$), whereas highly inhibited proteins included LARP1 and RICTOR, which were common with the WM1 module, and CLPP ($z = -4.47, p = 0.0001$) and LILRB2 ($z = -4.24, p = 0.0007$) (Table 2).

FGFR1 is a cell-surface receptor for fibroblast growth factors that regulate embryonic development, cell proliferation, differentiation, and migration. Its amplification has been frequently reported in NSCLC (23). Activation of CTNN β 1 and CTNN β -LEF1 implied activation of Wnt/ β -catenin signaling in which β -catenin is stabilized and accumulated as free β -catenin in the cytosol and is subsequently translocated into the nucleus and

activates the TCF (T-cell factor)/LEF (lymphoid enhancer factor)-dependent transcription of Wnt target genes, key factors in cell proliferation and invasion (24, 25). Proto-oncogene *LCK* encodes lymphocyte-specific protein tyrosine kinase LCK, an Src family tyrosine kinase, which plays an important role in TCR-linked signal transduction. CD28-induced LCK activation is an important mediator of T-cell activation. Recently, mathematical models for LCK autophosphorylation suggest that LCK is involved in the early stages of T-cell activation and response derived from PD-1 signaling suppression (26). It also potentially mediates PD-1-induced inhibition of early TCR signaling (27), which is important to cancer immune checkpoint therapy. LILRB2 is Leukocyte immunoglobulin-like receptor subfamily B member 2 (also known as ILT-4, immunoglobulin-like transcript 4) is a tumor

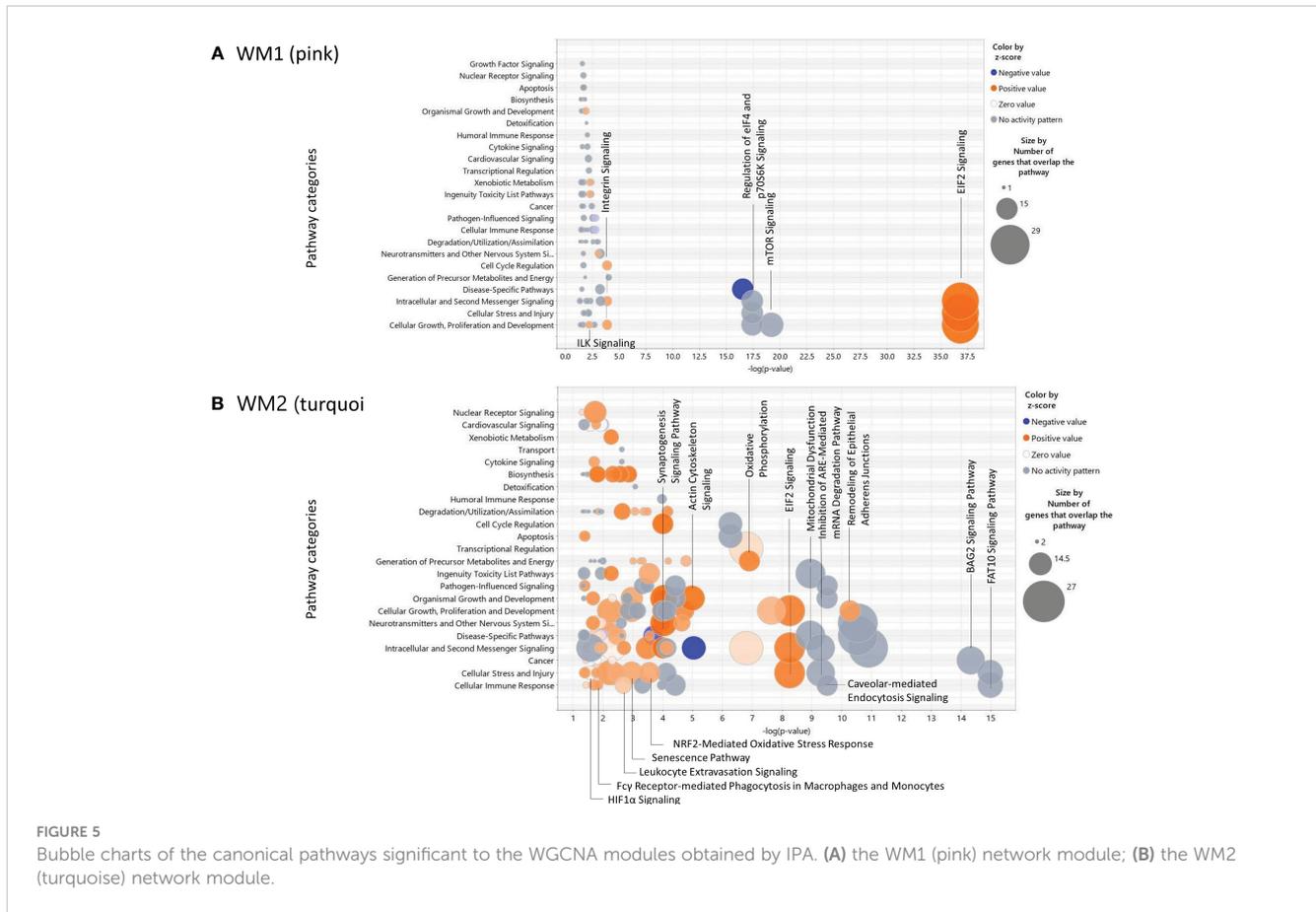


FIGURE 5

Bubble charts of the canonical pathways significant to the WGCNA modules obtained by IPA. (A) the WM1 (pink) network module; (B) the WM2 (turquoise) network module.

immune checkpoint molecule. Caseinolytic protease P (CLPP) is involved in the mitochondrial unfolded protein response and cellular bioenergetics (28).

Interestingly, all of *Hif* (complex), TRCD, FGFR1, IFNG, CD80, CD86, LCK, and LILRB2 target molecules including HLA-G in the WM2 module, whereas LILRB2 targets module molecules including both HLA-G and eigen-protein TPM4 (Table 2). Schreiber et al. demonstrated that the effects of IFNG are generally pro-tumorigenic during the immune escape stage of cancer immunoediting, in which IFNG increases inhibitory immune checkpoint molecules, including HLA-G, to promote the formation of a tolerant immune microenvironment (29).

HLA-G exerts inhibitory effects on both innate and adaptive effectors through direct binding to inhibitory receptors, such as immunoglobulin-like transcripts, LILRB1 (also known as ILT-2), and LILRB2, with high affinity compared with other HLA class I molecules. HLA-G can inhibit many players associated with the anti-tumor response throughout early and late tumor stages, although CTLA-4 and PD-1 are predominantly expressed in higher tumor grades (30, 31). The networks associated with LILRB2 inhibition include the activated hub regulator, toll-like receptor 4 (TLR4, also known as CD284). TLR signaling activates various signaling molecules, including nuclear factor κ B (NF κ B), extracellular signal-regulated kinase (ERK/JNK/p38), and induces the synthesis of immunologic factors including IL-6, IL-12, PD-L1, and HLA-G. This results in the resistance of tumor cells to CTL

attack and tumor cell immune evasion (32). Highly expressed HLA-G and highly inhibited LILRB2 implicated involvement of immune tolerance or anti-tumor immune escape. The volcano plots revealed high expression of HLA-G associated with the SPA subtype (Figures 2A–C). A web-based survival analysis (KMplot) for mRNA data of lung adenocarcinoma ($n = 719$) indicated that high HLA-G expression was significantly associated with poor OS (log-rank test $p = 1.0 \times 10^{-11}$; hazard ratio: 3.1 (2.2–4.36)) (<https://kmplot.com/analysis/>) (Supplementary Figure S3) (33). Yan et al. interestingly showed a predominant expression of soluble HLA-G (sHLA-G) in tumor cells not from squamous cell carcinoma but adenocarcinoma of lung (34).

3.7 Genomic alterations of LPA, MPA, and SPA based on the MSKCC database

Lung adenocarcinoma genomic data from the MSKCC database (Stage I to III: $n = 604$) was analyzed for prognostic markers associated with the predominant histologic subtypes in lung adenocarcinoma using the Memorial Sloan Kettering–Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) platform (5). We extracted data for the LPA ($n = 88$), MPA ($n = 37$), and SPA ($n = 68$) subtypes, and the genomic alteration profiles were visualized using the cBioPortal for Cancer Genomics (<https://www.cbioportal.org/>) (Supplementary Figure S4). The oncprints

TABLE 2 Representative upstream and predicted master regulators (activated or inhibited: $z\text{-value} \geq 2.0$, and upregulated: $1.5 < z\text{-value} < 2.0$) are summarized for the WGCNA modules, WM1 (pink) and WM2 (turquoise), significantly associated with the SPA subtype.

Module ID (color)	Upstream regulators			Causal networks			Canonical pathways			Diseases or functions					
	Top upstream regulators	z-score	p-value of overlap	Top master regulators*	z-score	Network bias-corrected p-value	Top 5 annotations	z-score	p-value	Top annotations	z-score	p-value			
WM1 (pink)	<i>LARPI</i> ^a	-5.10	7.19E-44	<i>PCGEMI</i>	4.87	1.00E-04	EIF2 Signaling	3.74	1.58E-37	Viral Infection	3.57	3.17E-08			
	<i>MLXIPL</i>	4.80	4.16E-31	<i>MLXIPL</i>	4.80	0.0001	mTOR Signaling	–	6.31E-20	Infection of cervical cancer cell lines	2.97	0.000324			
	<i>MYCN</i> ^a	4.31	1.49E-28	<i>MYC</i>	4.77	0.0001	Regulation of eIF4 and p70S6K Signaling	–	3.98E-18	Development of cytoplasm	2.78	0.00000629			
	<i>MYC</i>	4.87	2.60E-24	<i>HMGCR</i>	4.53	0.0001	TCA Cycle II (Eukaryotic)	–	8.91E-05	Fibrogenesis	2.78	0.000388			
	<i>RICTOR</i>	-4.24	3.80E-17	<i>CXCL14</i>	4.36	0.0001	Integrin Signaling	2.45	1.32E-04	Infection of cells	2.78	0.0000202			
	<i>Lh</i>	4.12	6.30E-17	<i>MYCN</i> ^a	4.32	0.0001				Necrosis	-2.93	5.33E-09			
	<i>DDX3X</i>	3.13	9.48E-17	<i>CD8B</i> ^a	4.15	0.0001				Cell death of cancer cells	-3.77	6.12E-14			
	<i>TCR</i> ^a	2.91	0.0000	<i>CD8A</i> ^a	4.15	0.0001				Cell death of tumor cells	-3.90	1.6E-13			
	<i>RBM20</i>	-2.12	1.08E-08	<i>Lh</i>	4.12	0.0001				Cell death of osteosarcoma cells	-4.00	4.22E-20			
	<i>CD3</i>	3.16	1.40E-07	<i>CD3-TCR</i> ^a	3.41	0.0001									
	<i>NFE2L2</i>	2.97	0.0001	<i>DDX3X</i>	3.16	0.0001									
	<i>HRAS</i>	2.21	0.0015	<i>RICTOR</i>	-4.24	0.0001									
	<i>NKX2-3</i>	-2.24	0.0016	<i>LARPI</i> ^a	-5.10	0.0001									
	<i>TGFB1</i>	2.44	0.0041	<i>CD3</i>	3.16	0.0026									
				<i>CD37</i> ^a	-3.40	0.0041									
			<i>UNC119</i> ^a	4.84	0.0045										
WM2 (turquoise)	<i>MYC</i>	3.90	5.72E-27	<i>LARPI</i>	-4.36	0.0001				Synaptogenesis Signaling Pathway	3.77	9.33E-05	Viral Infection	8.48	2.28E-19
	<i>RICTOR</i>	-5.67	9.94E-23	<i>CLPP</i>	-4.47	0.0001				Integrin Signaling	3.74	9.77E-05	Infection by RNA virus	7.16	3.29E-16
	<i>CLPP</i>	-4.45	2.83E-17	<i>RICTOR</i>	-5.78	0.0001					3.61	1.05E-05	Infection of cells	7.12	1.13E-14

(Continued)

TABLE 2 Continued

Module ID (color)	Upstream regulators			Causal networks			Canonical pathways			Diseases or functions		
	Top upstream regulators	z-score	p-value of overlap	Top master regulators*	z-score	Network bias-corrected p-value	Top 5 annotations	z-score	p-value	Top annotations	z-score	p-value
							Actin Cytoskeleton Signaling					
	<i>IL4</i>	5.21	2.86E-17	<i>Cbp/p300</i> ^c	5.46	0.0001	EIF2 Signaling	3.32	5.62E-09	Cell viability	6.72	1.28E-11
	<i>LARP1</i>	-4.36	4.93E-15	<i>Hif</i> ^c	4.55	0.0001	SNARE Signaling Pathway	3.32	1.10E-04	Cell survival	6.54	7.39E-14
	<i>TGFB1</i>	5.45	1.19E-12	<i>CTNNA1</i> ^c	4.59	0.0001	Oxidative Phosphorylation	3.21	1.26E-07	Cell viability of tumor cell lines	6.15	1.81E-11
	<i>TEAD1</i>	4.57	1.36E-11	<i>Tcrd</i> ^c	4.57	0.0002				HIV infection	5.68	2.93E-06
	<i>NFE2L2</i>	5.80	1.40E-11	<i>CTNNβ-LEF1</i>	4.52	0.0001				Infection by HIV-1	5.41	1.26E-06
	<i>INSR</i>	3.80	8.99E-10	<i>CTNNB1</i>	4.40	0.0001				Infection of tumor cell lines	5.39	6.95E-09
	<i>MLXIPL</i>	4.20	2.39E-09	<i>IL20RA</i> ^c	4.36	0.0002				Infection of cervical cancer cell lines	5.19	3.87E-07
	<i>IL1B</i>	3.52	4.60E-09	<i>NFE2L2</i>	5.92	0.0003				Cell movement	5.04	8.69E-12
	<i>ESR2</i>	3.74	8.05E-09	<i>FGFR1</i> ^c	7.46	0.0005				Organization of cytoplasm	5.04	3.35E-06
	<i>IFNG</i> ^c	4.51	1.39E-08	<i>RAB27A</i> ^c	3.94	0.0007				Cell death of osteosarcoma cells	-5.29	1.71E-20
	<i>TEAD1</i>	2.12	2.09E-08	<i>Lilrb2</i> ^{b,c}	-4.24	0.0007				Organismal death	-9.60	2.56E-12
	<i>RICTOR</i>	-2.65	2.03E-05	<i>IGF1R</i> ^c	3.88	0.0008						
	<i>NFκB (complex)</i>	3.87	8.69E-05	<i>CD19</i>	5.81	0.0009						
	<i>SMARCA4</i>	3.72	0.0001	<i>PSMB1</i> ^c	-4.07	0.0009						
	<i>IFN Beta</i> ^c	2.07	0.0006	<i>CD80</i> ^c	7.31	0.0186						
	<i>Interferon alpha</i> ^c	3.61	0.0022	<i>CD86</i> ^c	6.06	0.0187						
	<i>LCK</i> ^c	3.76	0.0219	<i>CD80/CD86</i> ^c	6.35	0.0438						

Superscripts a), b), and c) denote master and upstream regulators targeting module-member proteins, including the eigen-proteins, PRS4X and TPM4, and HLA-G, respectively.

showed genomic alterations characteristic of the three subtypes as follows: LPA, TP53 (17%), EGFR (43%), and KRAS (33%), whereas EGFR and KRAS were mutually exclusive ($q < 0.001$); MPA, TP53 (41%), EGFR (22%), and KRAS (32%); SPA, TP53 (75%), which was the highest among the three subtypes, EGFR (22%), and KRAS (38%).

In connection with the above genomic alterations, upregulated expressions of HLA I class molecules such as HLA-G and HLA-B were observed in proteomic analysis. (Figures 2A, B). It has been reported that there is a positive correlation between mutational burden and the expression of most HLA class I molecules (35, 36), and HLA class I molecules are associated with the innate immune response as ligands of inhibitory killer cell immunoglobulin-like receptors (KIRs) of Natural Killer (NK) cells (37).

4 Discussion

The WGCNA analysis was successfully applied to MS-based proteomic data. Forty co-expression network modules were identified, among which two network modules were significantly associated with solid predominant adenocarcinoma, SPA (Figure 4). Volcano plots of the proteins revealed high expression levels of HLA-G (human leukocyte antigen G) in the SPA subtype (Figures 2A, B). A multivariate correlation analysis (MVA) of eigen-proteins mostly grouped into Cluster A, B, and D, which correspond well to the SPA, MPA, and LPA subtypes, respectively (Figure 2C). Oncoprint analysis of three subtypes exhibited that SPA interestingly exhibited a highly frequent p53 alteration of 75%

although MPA and LPA harbored a frequency of p53 alterations, 41% and 17%, respectively (Supplementary Figure S4). High TP53 alteration seems to reflect a high malignancy of SPA.

Upstream analysis using IPA predicted that the highly activated oncogenic regulators, MYC, MLXIPL, and MYCN may reflect an aberrant feature of the SPA subtype, that activation of the redox master regulator NFE2L2 indicates an occurrence of oncogenic signaling in response to oxidative stresses and that both the highly inhibited LARP1 and activated pathways of EIF2 signaling indicated a switch to IRES-mediated mRNA translation, resulting in the generation of oncogenic proteins, such as hypoxia-inducible factors (*Hif*). Activation of Wnt/ β -catenin signaling also might participate in the disease-related interaction networks of SPA, which results in Wnt target gene products by the TCF/LEF-dependent transcription.

Interestingly, SPA-significant co-expression network modules, WM1 and WM2, also involved highly activated IFNG, TCRD, CD3-TCR, CD8A, CD8B, CD3, CD80/CD86, and highly inhibited LILRB2, all of which are associated with the adaptive immune system and defense, and importantly which all target module molecules including the HLA class I molecule HLA-G. Representative immune system-related causal networks associated with the SPA subtype are presented in Figure 6. Moreover, our observation of the immune checkpoint molecule, HLA-G, significantly upregulated in the SPA subtype and its counterpart LILRB2 predicted to be significant and highly inhibited implicate the involvement of the immune tolerance or anti-tumor immune escape processes in early-stage SPA, although CTLA-4 and PD-1 may not yet be expressed at early-stage.

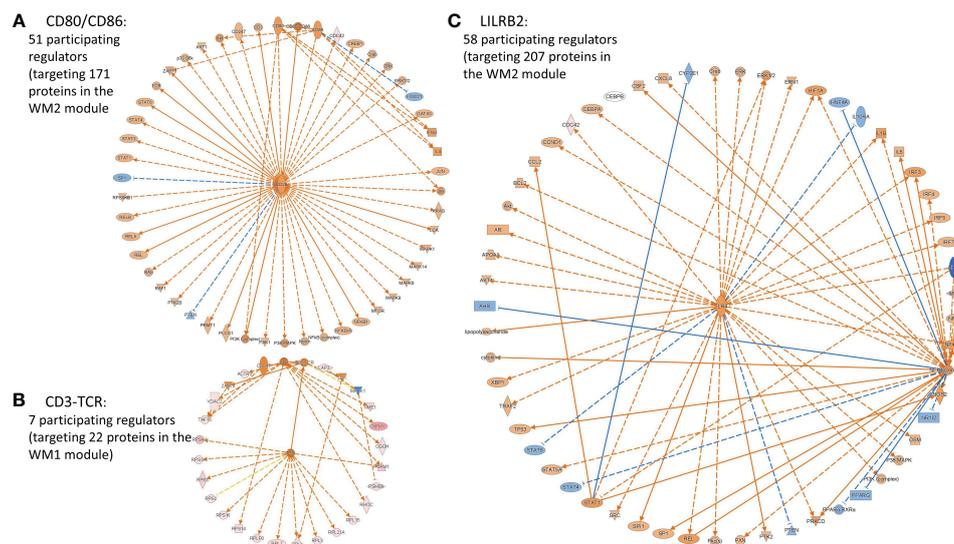


FIGURE 6

Representative cancer cell immunity-related causal networks of (A) CD80/CD86, (B) CD-TCR, and (C) LILRB2, associated with the SPA subtype, predicted by IPA. Node shapes indicate molecular types: triangle, kinase; square (dashed), growth factor; rectangle (horizontal), ligand-dependent nuclear receptor; rectangle (vertical), ion channel; diamond (vertical), enzyme; diamond (horizontal), peptidase; trapezoid, transporter; oval (horizontal), transcription regulator; oval (vertical), transmembrane receptor; double circle, complex; and circle, other. Orange/light orange and blue/light blue colors indicate the extent of confidence for predicted activation and inhibition, respectively. Lines denote predicted relationships. A solid or dashed line indicates direct or indirect interaction, respectively. Orange, leading to activation; blue, leading to inhibition; yellow, findings inconsistent with the state of a downstream molecule; grey, not predicted effect. CD80/CD86, T-lymphocyte activation antigens CD80/CD86 (Activation B7-1/B7-2 antigens); CD3-TCR, CD3, TCR (T cell receptor complex); LILRB2, Leukocyte immunoglobulin-like receptor subfamily B member 2, also known as Immunoglobulin-like transcript 4 (ILT-4).

The Applied Proteogenomic Organizational Learning and Outcomes (APOLLO) research network (38) conducted the deep proteogenomic profiling of 87 lung adenocarcinoma tumors from a cohort of individuals in the United States, which utilizes the integrative approaches, including whole-genome, transcriptome, and MS-based proteomic and phosphoproteomic sequencing, and in which adenocarcinoma tumor subtypes were stratified according to RNA expression subtypes, such as terminal respiratory unit (TRU), proximal proliferative (PP), and proximal inflammatory (PI). They reported that PI enriched with solid predominant histological subtype (SPA) overexpressed the immune cell marker clusters, and its network is characteristically associated with enhanced IFNG signaling and inflammation, high tumor mutational burden (TMB), coupled with PD-L1 protein and CTLA4 RNA expression, and suggested that the PI subtype may have the subset of tumors most likely to respond to immune checkpoint inhibitors (38).

Recently, Li et al. (39) conducted a multi-omics analysis of 1,078 untreated lung adenocarcinoma patients with clinicopathologic, genomic, transcriptomic, and proteomic data from public and internal cohorts. SPA had molecular features including significantly higher tumor mutation burden (TMB), the higher frequency of TP53 mutation together with EGFR/TP53 co-mutation, and higher immuno-resistant microenvironment, that indicates a poor response to chemotherapy, such as tyrosine kinase inhibitors (TKIs). Indeed, SPA showed upregulated expression of immunotherapy-related genes relevant to strong immunogenicity. Moreover, it was shown in the cohort of lung adenocarcinoma patients who received neoadjuvant immunotherapy that SPA relatively well responded to immunotherapy. They concluded that SPA would be more suitable for immunotherapy while less suitable for chemo- and targeted therapy (39).

Both studies (38, 39) seem to consistently reflect the molecular features of SPA and give corroboration to our findings in this study. One limitation of this study was the number of patients examined because of the limited number of early-stage cases pathologically confirmed and available for proteomic analysis in our hospital.

5 Conclusion

In summary, we successfully identified disease-related co-expression protein networks by WGCNA analysis applied to the proteomic datasets, following MS-based proteomic analysis of cancerous cells laser microdissected from FFPE tissue specimens of three lung adenocarcinoma subtypes, LPA, MPA, and SPA. Upstream regulator and causal network analysis performed for protein co-expression networks significantly associated with the SPA subtype implicated not only oncogenic signalings including cap-independent IRES-dependent mRNA translation and NRF2-mediated oxidative stress response but also highly activated molecular networks relating to the adaptive immune system and immune tolerance more likely associated with malignancies characterizing the SPA subtype, together with its high frequency of TP53 mutation. Although the number of patient samples examined was limited in this study, we are planning a larger cohort study of patient-derived samples, including genomic alteration analysis to investigate core data-driven proteogenomic

networks. This approach will provide clinically important information on proteogenomic landscapes of the high-risk adenocarcinoma subtype SPA of the lung.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://www.proteomexchangeek.org/>, PXD037831.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of St. Marianna University School of Medicine (approval no. 1461). Written informed consent to participate in this study was provided by the patient/participants or the patient/participants' legal guardian/next of kin. FFPE tumor tissue blocks from 12 surgical specimens histologically confirmed as lung LPA, MPA, and SPA were obtained without patient identifiers from the St. Marianna University School of Medicine Hospital.

Author contributions

TN: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing. ÁV: Conceptualization, Data curation, Project administration, Writing – review & editing. HN: Conceptualization, Project administration, Writing – review & editing. KF: Formal analysis, Writing – review & editing. SN: Resources, Writing – review & editing. NF: Supervision, Writing – review & editing. HSak: Writing – review & editing, Resources. HSaj: Conceptualization, Project administration, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. WHO classification of tumours editorial board. In: *Thoracic Tumours, 5th ed.* International Agency for Research on Cancer, Lyon, France.
2. Yuan Y, Ma G, Zhang Y, Chen H. Presence of micropapillary and solid patterns are associated with nodal upstaging and unfavorable prognosis among patient with cT1N0M0 lung adenocarcinoma: a large-scale analysis. *J Cancer Res Clin Oncol* (2018) 144:743–9. doi: 10.1007/s00432-017-2571-7
3. Yoshiya T, Mimae T, Tsutani Y, Tsubokawa N, Sasada S, Miyata Y, et al. Prognostic role of subtype classification in small-sized pathologic N0 invasive lung adenocarcinoma. *Ann Thorac Surg* (2016) 102:1668–73. doi: 10.1016/j.athoracsur.2016.04.087
4. Cao Y, Zhu LZ, Jiang MJ, Yuan Y. Clinical impacts of a micropapillary pattern in lung adenocarcinoma: a review. *Oncol Targets Ther* (2015) 9:149–58. doi: 10.2147/OTT.S94747
5. Caso R, Sanchez-Vega F, Tan KS, Mastrogiacomo B, Zhou J, Jones GD, et al. The underlying tumor genomics of predominant histologic subtypes in lung adenocarcinoma. *J Thorac Oncol* (2020) 15:1844–56. doi: 10.1016/j.jtho.2020.08.005
6. Zhao Y, Wang R, Shen X, Pan Y, Cheng C, Li Y, et al. Minor components of micropapillary and solid subtypes in lung adenocarcinoma are predictors of lymph node metastasis and poor prognosis. *Ann Surg Oncol* (2016) 23:2099–105. doi: 10.1245/s10434-015-5043-9
7. Jeon HW, Kim YD, Sim SB, Moon MH. Comparison of clinical results between high grade patterns in stage I lung adenocarcinoma. *Thorac Cancer* (2022) 13:2473–9. doi: 10.1111/1759-7714.14578
8. Ujiié H, Kadota K, Chaff JE, Buitrago D, Sima CS, Lee MC, et al. Solid predominant histologic subtype in resected stage I lung adenocarcinoma is an independent predictor of early, extrathoracic, multisite recurrence and of poor postrecurrence survival. *J Clin Oncol* (2015) 33:2877–84. doi: 10.1200/JCO.2015.60.9818
9. Nishimura T, Nakamura H, Végvári Á, Marko-Varga G, Furuya N, Saji H. Current status of clinical proteogenomics in lung cancer. *Expert Rev Proteomics* (2019) 16:761–72. doi: 10.1080/14789450.2019.1654861
10. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* (2014) 30:523–30. doi: 10.1093/bioinformatics/btt703
11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
12. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10:1243–60. doi: 10.1097/JTO.0000000000000630
13. Prieto DA, Hood BL, Darfler MM, Guiel TG, Lucas DA, Conrads TP, et al. Liquid Tissue: proteomic profiling of formalin-fixed tissues. *Biotechniques* (2005) Suppl:32–5. doi: 10.2144/05386su06
14. Carvalho PC, Lima DB, Leprevost FV, Santos MD, Fischer JS, Aquino PF, et al. Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. *Nat Protoc* (2016) 11:102–17. doi: 10.1038/nprot.2015.133
15. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* (2021) 49:D605–12. doi: 10.1093/nar/gkaa1074
16. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape stringApp: network analysis and visualization of proteomics data. *J Proteome Res* (2019) 18:623–32. doi: 10.1021/acs.jproteome.8b00702
17. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* (2014) 8 Suppl 4:S11. doi: 10.1186/1752-0509-8-S4-S11
18. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinf* (2017) 18:287. doi: 10.1186/s12859-017-1708-7
19. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Alhou LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* (2022) 31:8–22. doi: 10.1002/pro.4218
20. Zhou J, Liu B, Li Z, Li Y, Chen X, Ma Y, et al. Proteomic analyses identify differentially expressed proteins and pathways between low-risk and high-risk subtypes of early-stage lung adenocarcinoma and their prognostic impacts. *Mol Cell Proteomic* (2021) 20:100015. doi: 10.1074/mcp.RA120.002384
21. Iizuka K, Takao K, Yabe D. ChREBP-mediated regulation of lipid metabolism: involvement of the gut microbiota, liver, and adipose tissue. *Front Endocrinol (Lausanne)* (2020) 11:587189. doi: 10.3389/fendo.2020.587189
22. Silvera D, Formenti SC, Schneider RJ. Translational control in cancer. *Nat Rev Cancer* (2010) 10:254–66. doi: 10.1038/nrc2824
23. Bogatyrova O, Mattsson JSM, Ross EM, Sanderson MP, Backman M, Botling J, et al. FGFR1 overexpression in non-small cell lung cancer is mediated by genetic and epigenetic mechanisms and is a determinant of FGFR1 inhibitor response. *Eur J Cancer* (2021) 151:136–49. doi: 10.1016/j.ejca.2021.04.005
24. Marie PJ, Hay E. Cadherins and Wnt signalling: a functional link controlling bone formation. *Bonekey Rep* (2013) 2:330. doi: 10.1038/bonekey.2013.64
25. Osuka S, Zhu D, Zhang Z, Li C, Stackhouse CT, Sampetean O, et al. N-cadherin upregulation mediates adaptive radioresistance in glioblastoma. *J Clin Invest* (2021) 131:e136098. doi: 10.1172/JCI136098
26. Kreusser LM, Rendall AD. Autophosphorylation and the dynamics of the activation of lck. *Bull Math Biol* (2021) 83:64. doi: 10.1007/s11538-021-00900-9
27. Arulraj T, Barik D. Mathematical modeling identifies Lck as a potential mediator for PD-1 induced inhibition of early TCR signaling. *PLoS One* (2018) 13:e0206232. doi: 10.1371/journal.pone.0206232
28. Cormio A, Sanguedolce F, Pesce V, Musicco C. Mitochondrial caseinolytic protease P: A possible novel prognostic marker and therapeutic target in cancer. *Int J Mol Sci* (2021) 22:6228. doi: 10.3390/ijms22126228
29. Alspach E, Lussier DM, Schreiber RD. Interferon γ and its important roles in promoting and inhibiting spontaneous and therapeutic cancer immunity. *Cold Spring Harb Perspect Biol* (2019) 11:a028480. doi: 10.1101/cshperspect.a028480
30. Carosella ED, Ploussard G, LeMaout J, Desgrandchamps F. A systematic review of immunotherapy in urologic cancer: evolving roles for targeting of CTLA-4, PD-1/PD-L1, and HLA-G. *Eur Urol* (2015) 68:267–79. doi: 10.1016/j.eururo.2015.02.032
31. Carosella ED, Gregori S, Tronik-Le Roux D. HLA-G/LILRBs: A cancer immunotherapy challenge. *Trends Cancer* (2021) 7:389–92. doi: 10.1016/j.trecan.2021.01.004
32. Huang B, Zhao J, Li H, He KL, Chen Y, Chen SH, et al. Toll-like receptors on tumor cells facilitate evasion of immune surveillance. *Cancer Res* (2005) 65:5009–14. doi: 10.1158/0008-5472.CAN-05-0784
33. Lánckzy A, Györfy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. *J Med Internet Res* (2021) 23:e27633. doi: 10.2196/27633
34. Yan WH, Lu HY, Li YY, Zhang X, Lin A. Significance of tumour cell HLA-G5/G6 isoform expression in discrimination for adenocarcinoma from squamous cell carcinoma in lung cancer patients. *J Cell Mol Med* (2015) 19:778–85. doi: 10.1111/jcmm.12400
35. Noblejas-López MDM, Nieto-Jiménez C, Morcillo García S, Pérez-Peña J, Nuncia-Cantarero M, Andrés-Pretel F, et al. Expression of MHC class I, HLA-A, and HLA-B identifies immune-activated breast tumors with favorable outcome. *Oncimmunology* (2019) 8:e1629780. doi: 10.1080/2162402X.2019.1629780
36. Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* (2018) 359:582–7. doi: 10.1126/science.aao4572
37. Ajitkumar P, Geier SS, Kesari KV, Borriello F, Nakagawa M, Bluestone JA, et al. Evidence that multiple residues on both the alpha-helices of the class I MHC molecule are simultaneously recognized by the T cell receptor. *Cell* (1988) 54:47–56. doi: 10.1016/0092-8674(88)90178-X
38. Soltis AR, Bateman NW, Liu J, Nguyen T, Franks TJ, Zhang X, et al. Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. *Cell Rep Med* (2022) 3:100819. doi: 10.1016/j.xcrm.2022.100819
39. Li F, Wang S, Wang Y, Lv Z, Jin D, Yi H, et al. Multi-omics analysis unravels the underlying mechanisms of poor prognosis and differential therapeutic responses of solid predominant lung adenocarcinoma. *Front Immunol* (2023) 14:1101649. doi: 10.3389/fimmu.2023.1101649

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1273780/full#supplementary-material>