Bronchoalveolar Lavage Proteomics in Patients with Suspected Lung Cancer

Authors and Affiliations:

Ana Sofia Carvalho, Célia Marina Cuco, Carla Lavareda, Francisco Miguel, Mafalda Ventura, Sónia Almeida, Paula Pinto, Tiago Tavares de Abreu, Luís Vaz Rodrigues, Susana Seixas, Cristina Bárbara, Mikel Azkargorta, Felix Elortza, Júlio Semedo, John K Field, Leonor Mota, Rune Matthiesen.

1. Computational and Experimental Biology Group, Health Promotion and Chronic Diseases Department, National Institute of Health Dr Ricardo Jorge, Lisbon, Portugal.
2. CEDOC, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal
4. Instituto de Saúde Ambiental, Faculdade de Medicina da Universidade de Lisboa
5. Department of Pulmonology, Unidade Local de Saúde da Guarda, Faculty of Health Sciences, University of Beira Interior, Portugal
6. Instituto de Investigação e Inovação em Saúde, Universidade do Porto (I3S), Porto, Portugal;
7. Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal
9. Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, University of Liverpool, William Duncan Building, 6 West Derby Street, Liverpool, L7 8TX, UK.

Running Title

BAL proteomics in lung cancer diagnostics

Keywords

Lung Cancer, Bronchoalveolar Lavage, Diagnosis, Proteomics, Biomarkers
**Supplementary figure legends**

Figure S1. Distribution of Jaccard indexes comparing protein identifications from replica and non-replica samples.

Figure S2. Increase in the number of proteins (black bars) and protein encoding genes (grey bars) identified as sample size increase.

Figure S3. Principal component analysis using spectral count values from all abundant proteins. A) Annotated based on patients’ lung cancer status as indicated in the legend. Black symbols indicate patients who died within the two years of follow-up. B) Annotated based on patients’ lung cancer staging. Black symbols indicate patients who died within the two years of follow-up. C) Annotated based on patients’ gender. D) Annotated based on patients’ smoking status.

Figure S4. Comparison of ratios obtained from VEMS and MaxQuant analysis for biomarkers found significant regulated for non-lung cancer versus lung cancer by both software.

Figure S5. Heatmap of significant regulated proteins based on iBAQ values obtained from VEMS and MaxQuant. Left bar depicts lung cancer cases (red) versus non lung cancer controls (green).

Figure S6. Boxplot of iBAQ expression values for all significant regulated proteins between cases and controls in the consensus list. Barplot titles includes gene name, P value correct for multiple testing, M log2 fold ratio and N the sample size.

Figure S7. Boxplot of iBAQ expression values for all significant regulated proteins between cases and controls factored on histology groups. Barplot titles includes gene name, P value correct for multiple testing, M log2 fold ratio and N sample size. The indicated P values results from comparing cases and controls.

Figure S8. Heatmap of protein biomarkers identified exclusively in lung cancer samples. The ten most significantly functional enriched KEGG pathways are indicated in the top of the heatmap where a black bars depicts which functional categories the protein belong.

Figure S9. Heatmap of peptide mziXIC values for most abundant peptides. "Yes" and "No" indicates the patients status with respect to lung cancer. Two sub clusters are highlighted with only lung cancer (light red) and non-lung cancer (light green).

Figure S10. Significantly enriched GO terms within the cellular component category (CC) for all identified proteins across all patient samples. The heatmap depicts the log2 number of proteins identified in different cellular component categories across patient samples.
Figure S11. Significantly enriched GO terms within the cellular component (CC) category for proteins in the consensus list across all patient samples. The heatmap depicts the log2 number of proteins in consensus list and in different cellular component categories across patient samples.

Figure S12. Significantly enriched GO terms within the molecular function (MF) category for all identified proteins across all patient samples. The heatmap depicts the log2 number of proteins identified in different molecular function categories across patient samples.

Figure S13. Significantly enriched GO terms within the molecular function (MF) category for proteins in the consensus list across all patient samples. The heatmap depicts the log2 number of proteins in consensus list and in different functional molecular function categories across patient samples.

Figure S14. Significantly enriched GO terms within the biological process (BP) category for all identified proteins across all patient samples. The heatmap depicts the log2 number of proteins identified in different biological process categories across patient samples.

Figure S15. Significantly enriched GO terms within the biological process category (BP) for proteins in the consensus list across all patient samples. The heatmap depicts the log2 number of proteins in consensus list and in different biological process categories across patient samples.

Figure S16. Boxplot factored on lung cancer versus non-lung cancer cases of total spectral counts for all identified proteins annotated as “extracellular vesicular exosome”.

Figure S17. Venn diagram comparing identified proteins in five major cellular component categories.

Figure S18. Barplot of mean log2(iBAQ+1) for proteins specific for one of the major cellular component category in Figure S15. The numbers of the bars indicates the number of specific proteins for each of the major cellular component categories that were utilized for the calculation of mean log2(iBAQ+1) values.

Figure S19. GO enrichment analysis (biological process) of all identified proteins (A) and significant regulated proteins between cases and controls (B). The y-axis indicates the number of proteins in each biological process category for all identified proteins and for the significantly regulated proteins. The numbers on top of the columns indicates the number of proteins in each category in total.

Figure S20. Reproducibility of VEMS/MaxQuant consensus lung cancer biomarkers in other OMICs lung cancer studies published in the literature. The proposed genes are
order with the genes reproduced most times in the top. The heatmap depicts down-regulated in green, up-regulated in red and non-significant or not detected in blue.

Figure S21. Boxplot comparing iBAQ expression of CD molecules from human leucocyte in non-lung cancer versus lung cancer (http://www.hcdm.org). The iBAQ expression values were compared by a Wilcox test. N indicates the number of markers that were identified by MS for the specific Leucocyte.
Distribution of Jaccard indexes

p-value < 2.2e-16
Increase in identified proteins as sample size increase

Figure S2
A. Lung cancer $N=90$

- Suspicious and non lung cancer that two years later are diagnosed as lung cancer
- Non
- Yes

B. Stage $N=90$

- 1
- 2
- 3
- 4
- na
- No

C. Gender $N=90$

- F
- M

D. Smoking $N=90$

- Current smoker
- Former smoker
- Nonsmoker
- Unknown

Figure S3
Comparisons of ratios obtained from VEMS and MaxQuant

Figure S4
Cactin
CACTIN, P= 0.000679, M= 3.06, N= 83

Adenylosuccinate lyase
ADSL, P= 0.00352, M= 2.38, N= 83

40S ribosomal protein S20
RPS20, P= 0.00392, M= 1.86, N= 83

Ubiquitin carboxyl−terminal hydrolase 5
USP5, P= 0.00599, M= 2.75, N= 83

Acylamino−acid−releasing enzyme
APEH, P= 0.00631, M= 1.68, N= 83

Staphylococcal nuclease domain−containing protein 1
SND1, P= 0.00759, M= 1.56, N= 83

Cullin−associated NEDD8−dissociated protein 1
CAND1, P= 0.00855, M= 1.52, N= 83

Haptoglobin
HP, P= 0.00867, M= 2.04, N= 83

Alpha−synuclein
SNCA, P= 0.00883, M= 3.01, N= 83

Figure S6
<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Description</th>
<th>P-value</th>
<th>Median</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4b-binding protein beta chain</td>
<td>C4BPB</td>
<td>0.0168</td>
<td>2.43</td>
<td>83</td>
</tr>
<tr>
<td>DNA-directed RNA polymerases I, II, and III subunit RPA</td>
<td>POLR2H</td>
<td>0.0187</td>
<td>1.19</td>
<td>83</td>
</tr>
<tr>
<td>Protein transport protein Sec23A</td>
<td>SEC23A</td>
<td>0.0205</td>
<td>1.26</td>
<td>83</td>
</tr>
<tr>
<td>Protein AMBP</td>
<td>AMBP</td>
<td>0.0209</td>
<td>1.1</td>
<td>83</td>
</tr>
<tr>
<td>Ethylmalonyl-CoA decarboxylase</td>
<td>ECHDC1</td>
<td>0.021</td>
<td>1.15</td>
<td>83</td>
</tr>
<tr>
<td>Acyl-coenzyme A thioesterase 2, mitochondrial</td>
<td>ACOT2</td>
<td>0.0232</td>
<td>1.52</td>
<td>83</td>
</tr>
<tr>
<td>Ankyrin-1</td>
<td>ANK1</td>
<td>0.0232</td>
<td>1.1</td>
<td>83</td>
</tr>
<tr>
<td>3-hydroxyisobutyryl-CoA hydrolase, mitochondrial</td>
<td>HIBCH</td>
<td>0.0239</td>
<td>1.13</td>
<td>83</td>
</tr>
<tr>
<td>Carbonic anhydrase 2</td>
<td>CA2</td>
<td>0.0243</td>
<td>2.12</td>
<td>83</td>
</tr>
</tbody>
</table>
Farnesyltransferase, CAAX box, alpha, isoform CRA
FNTA, P= 0.0249, M= 1.13, N= 83

Secretoglobin family 3A member 2
SCGB3A2, P= 0.0254, M= 1.39, N= 83

Alpha–
FUT6, P= 0.0254, M= 1.09, N= 83

Platelet glycoprotein Ib alpha chain
GP1BA, P= 0.0254, M= 1.15, N= 83

Pregnancy zone protein
PZP, P= 0.0256, M= 1.12, N= 83

26S proteasome non–ATPase regulatory subunit 12
PSMD12, P= 0.0262, M= 1.25, N= 83

Protein THEMIS2
THEMIS2, P= 0.0264, M= 1.14, N= 83

Uncharacterized protein C1orf87
C1orf87, P= 0.0265, M= 1.29, N= 83

Calcium–activated chloride channel regulator 1
CLCA1, P= 0.0265, M= 1.4, N= 83
Hemoglobin subunit delta
HBD, $P = 0.0265$, $M = 1.21$, $N = 83$

Triokinase/FMN cyclase
TKFC, $P = 0.0265$, $M = 1.31$, $N = 83$

Carbonic anhydrase 1
CA1, $P = 0.0267$, $M = 1.27$, $N = 83$

UPF0764 protein C16orf89
C16orf89, $P = 0.0267$, $M = 1.29$, $N = 83$

26S proteasome non–ATPase regulatory subunit 1
PSMD1, $P = 0.0284$, $M = 1.32$, $N = 83$

Protein DDI1 homolog 2
DDI2, $P = 0.0294$, $M = 1.79$, $N = 83$

Platelet basic protein
PPBP, $P = 0.0295$, $M = 1.7$, $N = 83$

Eukaryotic translation initiation factor 5
EIF5, $P = 0.0304$, $M = 1.15$, $N = 83$

Tyrosine–protein kinase BTK
BTK, $P = 0.0304$, $M = 1.06$, $N = 83$
Exportin-7
XPO7, P= 0.0349, M= 1.16, N= 83

SUMO-activating enzyme subunit 1
SAE1, P= 0.0349, M= 1.07, N= 83

Ras GTPase-activating-like protein IQGAP2
IQGAP2, P= 0.0349, M= 1.55, N= 83

Nuclear protein localization protein 4 homolog
NPLOC4, P= 0.0349, M= 1.06, N= 83

26S proteasome non-ATPase regulatory subunit 5
PSMD5, P= 0.0349, M= 1.33, N= 83

Dynamin-1-like protein
DNM1L, P= 0.0349, M= 1.06, N= 83

Serine/threonine-protein phosphatase
PPP1CC, P= 0.0349, M= 1.13, N= 83

Cytochrome b-c1 complex subunit 2, mitochondrial
UQCRC2, P= 0.0349, M= 1.04, N= 83

Exportin-1
XPO1, P= 0.0349, M= 1.1, N= 83
Dipeptidyl peptidase 1
CTSC, \( P = 0.000679, M = -2.87, N = 83 \)

Protein–glutamine gamma–glutamyltransferase 2
TGM2, \( P = 0.00138, M = -2.17, N = 83 \)

Pulmonary surfactant–associated protein A2
SFTP A2, \( P = 0.0014, M = -3.21, N = 83 \)

UMP–CMP kinase
CMPK1, \( P = 0.00177, M = -2.43, N = 83 \)

Annexin
ANXA2, \( P = 0.00177, M = -2.66, N = 83 \)

WAP four–disulfide core domain protein 2
WFDC2, \( P = 0.00177, M = -3.11, N = 83 \)

Isocitrate dehydrogenase [NADP] cytoplasmic
IDH1, \( P = 0.00277, M = -2.34, N = 83 \)

Rho GDP–dissociation inhibitor 1
ARHGDA, \( P = 0.00277, M = -2.55, N = 83 \)

Tryptophan—tRNA ligase, cytoplasmic
WARS, \( P = 0.00277, M = -2.29, N = 83 \)
Heterogeneous nuclear ribonucleoprotein K (HNRNPK), P= 0.00288, M= −2.23, N= 83

Glutathione S–transferase P (GSTP1), P= 0.00333, M= −1.57, N= 83

Plastin-2 (LCP1), P= 0.00442, M= −2.58, N= 83

PSAP protein (PSAP), P= 0.0047, M= −2.18, N= 83

Cathepsin G (CTSG), P= 0.00498, M= −1.61, N= 83

Neutrophil gelatinase–associated lipocalin (LCN2), P= 0.00599, M= −1.88, N= 83

Vimentin (VIM), P= 0.00599, M= −1.96, N= 83

Beta–microsemionoprotein (MSMB), P= 0.00651, M= −2.95, N= 83

Omega–amidase NIT2 (NIT2), P= 0.00804, M= −1.76, N= 83
<table>
<thead>
<tr>
<th>Protein</th>
<th>CES1</th>
<th>LAP3</th>
<th>CES1</th>
<th>CES1</th>
<th>CES1</th>
<th>CES1</th>
</tr>
</thead>
<tbody>
<tr>
<td>S100A11, P= 0.0207, M= −2.12, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Liver carboxylesterase 1</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0207, M= −1.94, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Lactotransferrin</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0245, M= −1.4, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cytosol aminopeptidase</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.025, M= −1.04, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LTF, P= 0.025, M= −1.04, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0254, M= −1.51, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Myosin light polypeptide 6</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0256, M= −1.47, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0262, M= −1.86, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0267, M= −1.68, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0267, M= −1.87, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0267, M= −1.87, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CES1, P= 0.0267, M= −1.87, N= 83</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Actin-related protein 2/3 complex subunit 5 (ARPC5), $P = 0.0302$, $M = -1.78$, $N = 83$

Synaptic vesicle membrane protein VAT-1 homolog (VAT1), $P = 0.0349$, $M = -1.7$, $N = 83$

Erythrocyte band 7 integral membrane protein (STOM), $P = 0.0349$, $M = -1.53$, $N = 83$

Ras-related C3 botulinum toxin substrate 2 (RAC2), $P = 0.0349$, $M = -1.19$, $N = 83$

Macrophage-capping protein (CAPG), $P = 0.0349$, $M = -2.04$, $N = 83$

Leukocyte elastase inhibitor (SERPINB1), $P = 0.0362$, $M = -1.07$, $N = 83$

Aldehyde dehydrogenase, dimeric NADP-prefering (ALDH3A1), $P = 0.0372$, $M = -1.57$, $N = 83$

Ras-related protein Rab-5C (RAB5C), $P = 0.0427$, $M = -1.28$, $N = 83$

Annexin (ANXA3), $P = 0.0479$, $M = -1.15$, $N = 83$
Figure S7
Dipeptidyl peptidase 1
CTSC, P = 0.000679, M = −2.87, N = 83

Protein–glutamine gamma–glutamyltransferase 2
TGM2, P = 0.00138, M = −2.17, N = 83

Pulmonary surfactant–associated protein A2
SFTP A2, P = 0.0014, M = −3.21, N = 83

UMP–CMP kinase
CMPK1, P = 0.00177, M = −2.43, N = 83

Annexin
ANXA2, P = 0.00177, M = −2.66, N = 83

WAP four–disulfide core domain protein 2
WFDC2, P = 0.00177, M = −3.11, N = 83

Isocitrate dehydrogenase [NADP] cytoplasmic
IDH1, P = 0.00277, M = −2.34, N = 83

Rho GDP–dissociation inhibitor 1
ARHGDIA, P = 0.00277, M = −2.55, N = 83

Tryptophan–tRNA ligase, cytoplasmic
WARS, P = 0.00277, M = −2.29, N = 83
catalytic activity
multicellular organismal development
vesicle-mediated transport
positive regulation of ERK1 and ERK2 cascade
interferon-gamma-mediated signaling pathway
cytokine-mediated signaling pathway
in utero embryonic development
nucleobase-containing small molecule metabolic process
T cell receptor signaling pathway
protein heterooligomerization
negative regulation of transcription, DNA-templated
translation
signal transducer activity
cell adhesion
GTP catabolic process
cell proliferation
cell redox homeostasis
proteolysis
oxidoreductase activity
phospholipase inhibitor activity
cellular nitrogen compound metabolic process
threonine-type endopeptidase activity
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent
antigen processing and presentation of exogenous peptide antigen via MHC class I
GTPase activity
small molecule metabolic process
viral process
protein transport
small GTPase-mediated signal transduction
RNA metabolic process
mRNA metabolic process
gene expression
apoptotic process
Spectral counts for “extracellular vesicular exosome”
Extracellular vesicular exosome

Membrane

Cytoplasm

Extracellular space

Nucleus

Figure S17
iBAQ for proteins specific for a subcellular localization

Mean log2(iBAQ+1)

<table>
<thead>
<tr>
<th>Subcellular Localizations</th>
<th>iBAQ+1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular vesicular exosome</td>
<td>134</td>
</tr>
<tr>
<td>Cytoplasm</td>
<td>65</td>
</tr>
<tr>
<td>Nucleus</td>
<td>55</td>
</tr>
<tr>
<td>Extracellular space</td>
<td>30</td>
</tr>
<tr>
<td>Membrane</td>
<td>21</td>
</tr>
</tbody>
</table>
Enriched biological process for all identifications

Number of protein identifications

- small molecule metabolic process
- innate immune response
- cellular response to virus
- viral process
- apoptotic process
- carbohydrate metabolic process
- RNA metabolic process
- cellular nitrogen compound metabolism
- cellular amino acid metabolic process
- regulation of transcription
- negative regulation of transcription
- positive regulation of transcription
- G1/S transition of mitotic cell cycle
- mitotic cell cycle
- translational elongation
- cell division
- cell cycle
- cell assembly
- protein metabolic process
- protein binding
- protein modification
- cis-acting regulator binding
- sequence-specific DNA binding
- transcription factor activity
- enzyme activity
- biological regulation
- catalytic activity
- structural molecule activity
- localization
- small molecule metabolic process

Enriched biological process for 133 significant regulated proteins

- gene expression
- viral process
- small molecule metabolic process

Figure S19
T.cell N= 6
P(Wilcox) = 0.214

Figure S21
B.cell N= 5
P(Wilcox)= 0.519
Dendritic.cell N= 3
P(Wilcox) = 0.191
NK.cell N = 5
P(Wilcox) = 0.0433
Stem.cell.Precursor N= 2
P(Wilcox)= 0.113
Macrophage.Monocyte N= 11
P(Wilcox)= 0.718
Granulocyte N= 7
P(Wilcox)= 0.235
Platelet N = 5
P(Wilcox) = 0.531
Erythrocyte N= 5
P(Wilcoxon)= 0.0955
Endothelial cell N = 7
P(Wilcoxon) = 0.711
Epithelial.cell N = 4
P(Wilcox) = 0.387
References for table S1


