**Collagen family proteins, thrombospondin 1 and lumican are differentially expressed across breast cancer subtypes by functional proteomics: a preliminary study with bioinformatics analysis**

Wei-Chi Ku1\*, Nam Nhut Phan2\*, Chih-Yi Liu1,3, Chi-Jung Huang4,5, Chen-Chung Liao6, Yen-Chun Huang7, Po-Hsin Kong7, Ling-Ming Tseng8,9+, Chi-Cheng Huang8,9,10+

1School of Medicine, College of Medicine, Fu-Jen Catholic University, New Taipei 242, Taiwan; 089052@mail.fju.edu.tw (WCK)

2Greehey’s Children Cancer Research Institute, University of Texas Health at San Antonio, San Antonio, Texas, USA 78229; namphanpro@gmail.com

3Division of Pathology, Cathay General Hospital SiJhih, New Taipei 221, Taiwan; cyl1124@gmail.com (CYL)

4Department of Medical Research, Cathay General Hospital, Taipei 106, Taiwan; science.man2@gmail.com

5Department of Biochemistry, National Defense Medical Center, Taipei 114, Taiwan

6Cancer and Immunology Research Center, National Yang Ming Chiao Tung University, Taipei 112, Taiwan; ccliao@nycu.edu.tw

7Marker Exploration Corporation, Taipei 112, Taiwan; leohuang@marker-x.com (YCH) and boriskong@marker-x.com (PHK)

8Division of Breast Surgery, Department of Surgery, Taipei Veterans General Hospital, Taipei 112, Taiwan; lmtseng87@gmail.com (LMT)

9School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei 112, Taiwan

10Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei 100, Taiwan; chishenh74@gmail.com

\*Equal contribution

+Correspondence: lmtseng87@gmail.com (LMT) and chishenh74@gmail.com (CCH)

**Abstract**

Most multi-gene signatures for breast cancer prognostication are mRNA (gene expression)-based, and few take protein expression as a biomarker. Candidate proteins were identified using nanoscale liquid chromatography with tandem mass spectrometry (NanoLC-MS/MS) from 61 Taiwanese breast cancers. Seventeen proteins were highly differential expression from analysis of variance (ANOVA) across four distinct immunohistochemistry subtypes and protein-protein interaction (PPI) network found the strong networks including COL2A1, COL11A1, COL6A1, COL6A2, THBS1 and LUM. Using public domain databases, we found that RNASeq expression values of *COL2A1*, *COL11A1*, *COL6A1*, *COL6A2* and *LUM* were higher in tumor than in normal tissue, and all six genes were differentially expressed across four molecular subtypes. In addition, disease-specific survival discrepancy was observed comparing breast cancer patients of the upper and lower quartile of the collage family, *THBS1* and *LUM* gene expression signature. Functional proteomics with LC-MS/MS is capable for target protein identification.

**Keywords:** breast cancer subtypes; functional proteomics; protein-protein interaction; collagen family protein; thrombospondin 1; lumican