# Analysis of secretome of breast cancer cell line with an optimized semi-shotgun method

TANG Xiaorong<sup>1,2</sup> YAO Ling<sup>1,2</sup> CHEN Keying<sup>1,2</sup> HU Xiaofang<sup>1,2</sup> XU Lisa X.<sup>1,2,3,\*</sup> FAN Chunhai<sup>4</sup>

<sup>1</sup>Key Laboratory of Systems Biomedicine (Shanghai Jiao Tong University), China Ministry of Education, Shanghai, 200240, China <sup>2</sup>Dept. of Biomedical Engineering, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China

<sup>3</sup>Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, 200240, China

<sup>4</sup>Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, 201800, China

**Abstract** Secretome, the totality of secreted proteins, is viewed as a promising pool of candidate cancer biomarkers. Simple and reliable methods for identifying secreted proteins are highly desired. We used an optimized semi-shotgun liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) method to analyze the secretome of breast cancer cell line MDA-MB-231. A total of 464 proteins were identified. About 63% of the proteins were classified as secreted proteins, including many promising breast cancer biomarkers, which were thought to be correlated with tumorigenesis, tumor development and metastasis. These results suggest that the optimized method may be a powerful strategy for cell line secretome profiling, and can be used to find potential cancer biomarkers with great clinical significance.

Key words Secretome, Semi-shotgun, Breast cancer

## 1 Introduction

Cancer is one of leading causes of human death nowadays. Effective serum biomarkers with high sensitivity and specificity for early detection and prognostic prediction of a specialized cancer are highly desired<sup>[1]</sup>. Tumor may secrete some specific proteins, which may enter the circulation system and become useful biomarkers, such as prostate specific antigen (PSA)<sup>[2]</sup>. Therefore, secretome (the totality of secreted proteins of specific cell line, tissue or organ) of cancer cell lines provides a promising pool for identification of potential candidates for cancer biomarkers<sup>[1, 3]</sup>.

With the development of proteomics-base approaches, secretome researches of cancer cell lines have been in marked progresses in recent years. Both traditional gel electrophoresis based mass spectrometry (MS) method and the newly developed liquid chromatography(LC)-MS/MS method were applied to identify and compare the secretome of cancer cell lines<sup>[4-9]</sup>. Using "bottom-up" shotgun proteomics approach combined with 2D LC-MS/MS to separate and identify peptides on a linear ion trap, Kulasingam V et al.<sup>[10]</sup> identified 600-700 proteins in the serum-free medium (SFM) of each breast cancer cell line. However, only ~34% proteins among them were identified as secreted proteins and membrane proteins. The semi-shotgun approach divided the proteins in SFM into different fractions, followed by digestion of proteins in each fraction into peptides and identification of peptides with 1D LC-MS/MS. Mbeunkui F et  $al^{[11]}$  identified ~250 proteins in SFM of each breast cancer cell line. Percentage of secreted proteins among total identified proteins was improved, but the number of identified proteins in SFM is limited.

Supported by National Natural Science Foundation of China (No. 50725622, No. 30670495), Ministry of Science and Technology of China (No. 2006CB910100), and Special Funding from Shanghai Center for Systems Biomedicine.

\* Corresponding author. E-mail address: lisaxu@sjtu.edu.cn

Received date: 2009-03-13

To effectively identify secreted proteins in SFM, several challenges remain to be addressed<sup>[3]</sup>. First, contamination from bovine serum in the culture medium should be avoided by sufficient wash of cells with SFM when changing the culture medium. Next, as the quantity of secreted proteins is small, the intracellular proteins released from dead cells during culture process will contaminate SFM, interfering the identification and quantification of secreted proteins. Thus the cell death rate should be reduced to minimum level. Finally, effective pre-fractionation of proteins in SFM would reduce sample complexity and facilitate the MS detection. These steps are critical to gain meaningful experiment results.

In the present work, we optimized a semi-shotgun LC-MS/MS method in the detection of secretome by addressing above mentioned issues. Breast cancer cell line MDA-MB-231 (viewed as metastatic model) was chosen as an example to analyze cancer cell secretome. Through the improvement of cell culture conditions, the death rate was strictly controlled below 3%. The concentration, desalting and fractionation of proteins in SFM were completed in one single step by using the small reversed phase C2 column. Furthermore, proteins in SFM were evenly fractionated into 5 portions to facilitate the MS detection by optimizing the gradient elution method. The MS results were very satisfying. A total of 464 proteins were identified with high reliability in SFM of MDA-MB-231. Among them, 63% proteins were secreted proteins and membrane proteins. Several secreted proteins in the list of identified proteins are closely related to breast cancer tumorigenesis.

## 2 Methods

#### 2.1 Cell culture

Breast cancer cell line MDA-MB-231 was cultured in Leibovitz's L-15 media with 10% fetal bovine serum. The death rate was controlled at below 3% (measured by trypan blue staining). At approximately 90% confluence, the cells were cultured in conditioned serum-free media (SFM) for 24 h after extensive washes. Conditioned media (CM) of four 10 cm plates (about  $4 \times 10^6$  cells /plate) was collected, centrifuged at 1000 r/min for 10 min, filtrated by 0.22 µm filter, and

then added with TFA (0.1%) before storing them at  $-80^{\circ}$ C.

#### 2.2 Optimized protein separation

Secreted proteins in the CM (~50 µg) were enriched and separated by the small reverse phase C2 column. The CM samples were injected into the column at the speed of 0.25 mL/min. After being washed by 10 mL 0.1%TFA, secreted proteins were fractionated into 5 fractions by eluting solvents of the increased acetonitrile (ACN) concentration in 0.1% TFA at the speed of 0.1 mL/min(1.2 mL 40% ACN, 0.6 mL 50% ACN, 0.6 mL 60% ACN, 0.6 mL 70% ACN and 0.6mL 100% ACN). Each fraction was equally divided into two parts, one for protein quantification and the other for subsequent LC-MS/MS analysis.

### 2.3 Digestion of protein fractions

Each fraction of protein sample was lyophilized, denatured in 10  $\mu$ L 8 mol/L urea and reduced by 10mmol/L TCEP for 3 h at the room temperature. 30 $\mu$ L ammonium bicarbonate (50 mmol/L, pH 7.8) was added to give a final urea concentration of 2mol/L. Then the samples were digested by sequencing grade trypsin (Promega) with protease: protein ratio of 1:50. The digestion was lasted overnight at 37°C, and it was stopped by the addition of 1  $\mu$ L TFA.

# 2.4 LC-MS/MS

The trypstic peptide digests were analyzed using nano LC-MS/MS (QStar Elite, ABI). Each sample was re-dissolved, desalted and loaded onto nano LC (5 µm C18, 12 cm×75 µm ID). The peptides were separated using a binary solvent system with solvent A consisting of 0.1% formic acid and 99.9% water, and solvent B of 98% acetonitrile, 0.1% formic acid and 1.9% water. The peptides were eluted with linear gradient from 3% B to 25% B in 80 min, 25% B to 60% B in 15 min, 60% B to 95% B in 2 min, followed by 10min of isocratic elution at 95% B with a constant flow rate of 300 nL/min. The MS conditions were: one full MS scan (375-1600 m/z) was followed by fragmentation and MS/MS scans of the top three most intensive ions; SmartExit was enabled in TOF MS/MS with maximum accumulation time at 2 s; spray voltage, 1.8 kV. The subsequent MS data were searched by ProteinPilot against the non redundant IPI\_Human\_3.25 database with biological modification included.

#### 2.5 Bioinformatic analysis of identified proteins

The proteins output confidence cutoff was 1.3 (unused score, >95%). Total score measures the evidence of all peptides for a protein, and it could represent the relative abundance of the protein. The information of cellular location and molecular function of the identified proteins were obtained from Gene Ontology.

# **3** Results and discussion

In the present work, we optimized a semi-shotgun proteomics method to identify secreted proteins<sup>[11]</sup>. Breast cancer cell line MDA-MB-231 was used as model cell line.

First, we optimized the cell culture conditions of



MDA-MB-231 in SFM. To avoid the contamination from bovine serum in the culture medium, the cells were washed with SFM for 15 min twice and 60 min twice at 37°C before being cultured in SFM. Through extensive washing, almost no proteins from the bovine serum could be detected in SFM of MDA-MB-231. To reduce the contamination of the intracellular proteins released from the dead cells during culture process, we optimized the cell density and serum-free culture time to decrease the cell death rate. The optimal seeding density of MDA-MB-231 was  $2 \times 10^6$  in each 10 cm plate. After culture in normal bovine serum containing medium for two days, the cell confluence reached ~90%. Then the cells were changed to be cultured in SFM for 24 h before collecting the SFM. Under these conditions. the death rate of MDA-MB-231 could be controlled below 3% (Fig.1 and Table 1).



Fig. 1 Typical pictures of MDA-MB-231 cells after trypan blue staining. The MDA-MB-231 cells were collected and stained after cultured in SFM for 24 h. Dead cells were stained with blue color (arrows).

PlateNo.	Number of dead cells	Total cell number	Death rate / %	
1	5	379	1.3	
2	4	204	1.9	
3	4	517	0.7	

Second, we optimized the separation method of secretome. The semi-shortgun method was used to prepare secretome samples. Secreted proteins were firstly divided into several fractions to simplify the sample complexity, and then were digested respectively. Better fractionation could lead to better LC-MS/MS analysis results. In our work, we used reversed phase C2 column to concentrate, desalt and fractionate proteins in SFM. Before optimazation, proteins were loaded at a speed of 0.5 mL/min, and eluted with 0.6 mL 30% ACN, 0.6 mL 40% ACN, 0.6 mL 50% ACN, 0.6 ml 60% ACN, 0.6 mL 70% ACN and 0.6 mL 80% ACN. We found that proteins were concentrated in the 50% and 60% ACN fractions. A

typical result of protein fractionation was 0.5, 1.9, 5.1, 10.7, 1.4, and 1.0 µg, respectively. To obtain better fractionation result, the flow rate and flow volume of gradient elution solvents were modulated to achieve the optimal protein recovery and protein separation. The loading rate was lowered from 0.5 mL/min to 0.25 mL/min to increase the adsorption amount of proteins on the column. Low flow rate was also used to elute proteins from column to achieve better protein recovery. In addition, we modulated the volume of gradient elution solvents containing increasing amount of ACN. When proteins were eluted subsequently with 1.2 mL 40% ACN, 0.6 mL 50% ACN, 0.6 mL 60% ACN, 0.6 mL 70% ACN and 0.6 mL 100% ACN, the proteins were most evenly dispersed in 5 fractions (3.4, 3.9, 7.4, 3.4, and 2.7 µg, respectively).

After these optimizations, 5 fractions were

digested and analyzed by LC-MS/MS separately. A total of 464 proteins were identified from 5 fractions with the 95% confidence. Among them, 84 (22.8%) were located in the extracellular region and 127 (34.4%) were membrane proteins (Fig.2A). According to the classification of biological function of these proteins, the top functions were binding activity followed by catalytic activity, signal transducer activity and structural molecule activity (Fig.2B). Obviously, not only the proportion of secreted proteins and membrane proteins was significantly increased in the detected proteins after optimizations, but also a great number of low-abundent secreted proteins were identified with high reliability. These demonstrated that the optimized semi-shotgun method is powerful and applicable in cell secretome analysis.



Fig. 2 Distribution of cellular location (A) and molecular function (B) of the identified proteins in the SFM of MDA-MB-231 with the optimized cell culture and fractionation methods.

Secretome is a promising pool of candidate cancer biomarkers. Ideal biomarker should be detected easily and reliably in serum after extensive dilution by circulation system. Major of the few tumor biomarkers used routinely in clinic currently are secreted and highly expressed at specific tumor region, such as prostate specific antigen (PSA) and alpha feto protein (AFP). Thus we focused on top-50 most abundant proteins identified in this work (Table 2 and Fig.3). Among these proteins, 16 (40.0%) were located in membrane, 7 (17.5%) were located in the extracellular matrix and 18 (45.0%) were located in the extracellular region. Several proteins, such as quiescin Q6, thrombospondin-1, galectin-3-binding protein, plasminogen activator inhibitor 1 precursor and cystatin C precursor, were also identified in the secretome of other breast cancer cell lines<sup>[10, 11]</sup>, and found to be related to tumor initiation and progression<sup>[12-16]</sup>.

Plasminogen activator inhibitor 1 is the second most abundant secreted proteins indentified for breast cancer cell MDA-MB-231 in our work. This protein participates in proteolysis and fibrinolysis<sup>[15]</sup>. It has positive correlation with tumorigenesis, tumor progression and metastasis<sup>[15, 17]</sup>. It is a recommended breast cancer biomarker by European Group on Tumor Markers and American Society of Clinical Oncology<sup>[18]</sup>.

IPI number	Protein name	Function	Unused
			score
IPI00003590	Isoform 1 of Sulfhydryl oxidase 1 precursor	Catalytic activity	52.42
IPI00007118	Plasminogen activator inhibitor 1 precursor	Protein binding, hydrolase activity,	51.6
		enzyme regulator activity	
IPI00296099	Thrombospondin-1 precursor	Receptor binding, structural molecule activity, binding	50.76
IPI00455315	Annexin A2	Binding, enzyme regulator activity	41.35
IPI00011229	Cathepsin D precursor	Peptidase activity	33.7
IPI00008561	Interstitial collagenase precursor	Calcium ion binding, peptidase activity	29.96
IPI00023673	Galectin-3-binding protein precursor	Receptor activity, protein binding	28.53
IPI00219018	Glyceraldehyde-3-phosphate dehydrogenase	Protein binding, catalytic activity	27.82
IPI00218918	Annexin A1	Receptor binding, structural molecule activity, binding,	26.23
IPI00012585	Beta-hevosaminidase beta chain precursor	Binding hydrolase activity	25.7
IPI00021405	Isoform A of Lamin-A/C	Structural molecule activity protein hinding	22.54
IPI00333541	Filamin-A	Recentor hinding actin hinding	21.4
IPI00016915	Insulin-like growth factor-binding protein 7	Protein binding	20.01
	precursor	1 rown onlang	20.01
IPI00018219	Transforming growth factor-beta-induced	Receptor binding, protein binding	19.61
IPI00477950	Collagen, type IV, alpha 2		19.29
IPI00374563	Agrin precursor	Signal transducer activity, structural molecule activity,	18.94
	5 r	protein binding, transferase activity	
IPI00387168	Proprotein convertase subtilisin/kexin type 9	Signal transducer activity, binding,	17.93
	precursor	peptidase activity	
IPI00010796	Protein disulfide-isomerase precursor	Protein binding, catalytic activity	17.35
IPI00006608	Isoform APP770 of Amyloid beta A4 protein	Binding, enzyme regulator activity	17.3
	precursor (Fragment)		
IPI00419585	Peptidyl-prolyl cis-trans isomerase A	Protein binding, catalytic activity	16.69
IPI00022462	Transferrin receptor protein 1	Receptor activity	15.96
IPI00299219	Protein CYR61 precursor	Binding	15.78
IPI00002236	Lactadherin precursor	Protein binding	15.45

 Table 2
 List of membrane and extracellular proteins in Top 50 proteins identified



Fig. 3 Distribution of cellular location (A) and molecular function (B) of the top-50 abundant proteins in SFM of MDA-MB-231.

Galectin-3-binding protein (G3BP, also named Mac-2 BP) is the 12<sup>th</sup> abundant protein identified in our work. G3BP is a highly glycosylated secreted protein, while its physiological function is not very

clear by now. G3BP was identified as a tumor associated antigen of breast cancer in 1980s<sup>[19,20]</sup>. It also has been found to be positively correlated with the progression of various cancers, such as ovarian,

gastric, liver cancers etc.<sup>[21-24]</sup>. Recent secretome study identified G3BP as a potential biomarker for oral cancer<sup>[7]</sup>. G3BP was also identified with high abundance in the secretome of breast cancer cell lines in previous work <sup>[11]</sup>.

Cystatin C is the 56<sup>th</sup> abundant protein in the secretome of our invasive breast cancer cell model. Cathepsins and their inhibitor cystatin C are involved in tumor growth, invasion and metastasis<sup>[16]</sup>. Cystatin C is a non-glycosylated, low molecular weight, basic protein<sup>[25]</sup>. Cystatin C is up-regulated in serum of patients with breast cancer and hepatocellular carcinoma<sup>[26, 27]</sup>. The increased level of cystatin C is also found to be correlated with adverse outcome of cancer patients.

In the present work, we optimized a semi-shotgun method to detect the secretome of cancer cell line. The optimized culture conditions excluded the contamination from bovine serum by extensive washing and greatly reduce the contamination of intracellular proteins by minimizing the cell death rate. In semi-shotgun method, improved fractionation of secretome led to better MS results. With optimized method, we identified 464 proteins from the SFM of a single cell line. Furthermore, most of these proteins were secreted proteins and membrane proteins. These data indicate the optimized semi-shotgun method developed in this study is simple and effective way to identify cell secretome.

## References

- Kulasingam V, Diamandis E P. Int J Cancer, 2008, **123**(9): 2007-2012.
- 2 Balk S P, Ko Y J, Bubley G J. J Clin Oncol, 2003, 21(2): 383-391.
- 3 Hathout Y, Expert Rev Proteomics, 2007, 4(2): 239-248.
- 4 Huang Lingjin, Chen Shengxi, Hunag Yan, et al. Lung Cancer, 2006, 54(1): 87-94.
- 5 Volmer M W, Stuhler K, Zapatka M, et al. Proteomics, 2005, 5(10): 2587-2601.
- Lin C Y, Tsui K H, Yu C C, *et al.* Proteomics, 2006, 6(15):
   4381-4389.

- Weng L P, Wu C C, Hsu B L, *et al.* J Proteome Res, 2008, 7(9): 3765-3775.
- 8 Kreunin P, Urquidi V, Lubman D M, et al. Proteomics, 2004, 4(9): 2754-2765.
- 9 Sardana G, Marshall J, Diamandis E P. Clin Chem, 2007, 53(3): 429-437.
- 10 Kulasingam V, Diamandis E P. Mol Cell Proteomics, 2007, 6(11): 1997-2011.
- Mbeunkui F, Metge B J, Shevde L A, *et al.* J Proteome Res, 2007, 6(8): 2993-3002.
- 12 Valko M, Rhodes C J, Moncol J, *et al.* Chem-Biol Interact, 2006, **160**(1): 1-40.
- Kazerounian S, Yee K O, Lawler J, Cell Mol Life Sci, 2008, 65(5): 700-712.
- 14 Grassadonia A, Tinari N, Iurisci I, *et al.* Glycoconjugate J, 2004, **19**(7-9) 551-556.
- 15 Binder B R, Mihaly J. Immunol Lett, 2008, **118**(2): 116-124.
- 16 Keppler D. Cancer Lett, 2006, 235(2): 159-176.
- 17 Annecke K, Schmitt M, Euler U, *et al.* Adv Clin Chem, 2008, **45**: 31-45.
- Molina R, Barak V, van Dalen A, *et al.* Tumour Biol, 2005,
   26(6): 281-293.
- 19 Iacobelli S, Arno' E, D'Orazio A, et al. Cancer Res A, 1986, 46: 3005-3010.
- 20 Iacobelli S, Arno' E, Sismondi P, *et al.* Breast Cancer Res Treat, 1988, **11**(1): 19-30.
- 21 Fusco O, Querzoli P, Nenci I, *et al.* Int J Cancer, 1998, **79**(1): 23-26.
- Natoli C, Garufi C, Tinari N, *et al.* Br J Cancer, 1993, 67(3): 465-467.
- 23 Park Y P, Choi S C, Kim J H, et al. Int J Cancer, 2007, 120(4): 813-820.
- 24 Lacovazzi P A, Trisolini A, Barletta D, *et al.* Clin Chem Lab Med, 2001, **39**(10): 961-965.
- 25 Brguljan P M, Cimerman N. Turk J Biochem, 2007, **32** (3): 95-103.
- 26 Decock J, Obermajer N, Vozelj S, *et al.* Int J Biol Markers, 2008, **23**(3): 161-168.
- 27 Zinkin N T, Grall F, Bhaskar K, *et al.* Clin Cancer Res, 2008, **14**(2): 470-477.