Published in final edited form as:

J Proteome Res. 2010 July 2; 9(7): 3394–3402. doi:10.1021/pr100231m.

Quantitative Proteomic Profiling Studies of Pancreatic Cancer Stem Cells

Lan Dai¹, Chen Li², Kerby A. Shedden³, Cheong.J. Lee⁴, Chenwei Li, HuyVuong Quoc¹, Diane M. Simeone⁴, and David M. Lubman^{1,2,4,5}

¹ Bioinformatics Program, University of Michigan Medical Center, Ann Arbor, MI

² Department of Chemistry, University of Michigan, Ann Arbor, MI

³ Department of Statistics, University of Michigan, Ann Arbor, MI

⁴ Department of Surgery, University of Michigan Medical Center, Ann Arbor, MI

⁵ Department of Pathology, University of Michigan Medical Center, Ann Arbor, MI

Abstract

Analyzing subpopulations of tumor cells in tissue is a challenging subject in proteomic studies. Pancreatic cancer stem cells (CSCs) are such a group of cells that only constitute 0.2-0.8% of the total tumor cells but have been found to be the origin of pancreatic cancer carcinogenesis and metastasis. Global proteome profiling of pancreatic CSCs from xenograft tumors in mice is a promising way to unveil the molecular machinery underlying the signaling pathways. However, the extremely low availability of pancreatic tissue CSCs (around 10,000 cells per xenograft tumor or patient sample) has limited the utilization of currently standard proteomic approaches which do not work effectively with such a small amount of material. Herein, we describe the profiling of the proteome of pancreatic CSCs using a capillary scale shotgun technique by coupling offline capillary isoelectric focusing(cIEF) with nano reversed phase liquid chromatography(RPLC) followed by spectral counting peptide quantification. A whole cell lysate from 10,000 cells which corresponds to ~1ug protein material is equally divided for three repeated cIEF separations where around 300ng peptide material is used in each run. In comparison with a non-tumorigenic tumor cell sample, among 1159 distinct proteins identified with FDR less than 0.2%, 169 differentially expressed proteins are identified after multiple testing corrections where 24% of the proteins are upregulated in the CSCs group. Ingenuity Pathway analysis of these differential expression signatures further suggests significant involvement of signaling pathways related to apoptosis, cell proliferation, inflammation and metastasis.

Keywords

cancer stem cells; capillary isoelectric focusing; two-dimensional separation; proteome profiling; differential expression; spectral count; pathway analysis

Supporting Information **Available:** Correlation matrix before and after transformation for all 7 datasets (S1); The detailed information about all the protein identifications (S2); The detailed information about all the differentially expressed proteins detected between the CSCs group and the tumor group (S3); The detailed information about how the cIEF recovery was determined (S4). This information is available free of charge via the Internet at <http://pubs.acs.org>.

Introduction

Pancreatic cancer has the worst prognosis of any major malignancy and is currently ranked the fourth leading cause of cancer-related mortality with a five-year survival rate less than 5%. Delayed diagnosis, relative chemotherapy and radiation resistance and an intrinsic biological aggressiveness all contribute to the abysmal prognosis[1]. Attempts to better understand the molecular characteristics of pancreatic adenocarcinoma have focused on studying the gene and protein expression profiling of pancreatic cancer compared to either normal pancreas or pancreatitis. However, these studies have not accounted for the heterogeneity of the tumor cells, in particular, the existence of a small set of distinct cells termed cancer stem cells which are responsible for tumor initiation and propagation.

Cancer stem cells have been identified in pancreatic cancer and several other tumor types including colon, prostate, and brain. In Li *et al* [2], a subpopulation of pancreatic tumor cells with cell surface markers CD44⁺CD24⁺ESA⁺ was isolated and functional studies were conducted to verify that this subpopulation possessed the ability of self-renewal and producing differentiated progeny. New strategies addressing this disease with a paradigm shift in the mechanism of the therapeutic resistance and recurrence of pancreatic tumor can be developed with an improved understanding of the cellular signaling pathways in CSCs at the protein level. The xenograft model of primary human pancreatic cancer represents a significant advance for the study of pancreatic cancer. Animal models using cancer cell lines often do not recapitulate human diseases accurately, where the biological characteristics and histology of tumor-derived human pancreatic cancer tissue are preserved in the xenograft.

The major obstacle to the study of global proteome expression profiling of the CSCs is the extremely small number of cancer stem cells available per study per tumor sample. Due to the unique features of pancreatic tissue and low percentage of pancreatic cancer stem cells (0.2% to 0.8%), from a single human tumor xenografted in a mouse, we can typically obtain 10k antibody labeled cancer stem cells using flow cytometry, which corresponds to around 1ug of total protein. Current publications of proteomic studies using capillary scale shotgun approaches are based on the analysis of entire tissue sections or cell lines instead of a subpopulation of primary human cells[3]. The amount of material consumed in a proteomic study using shotgun approaches such as MudPIT or offline 2D-LC/MS/MS is higher than 20ug[4-6]. Some studies targeting a certain subpopulation of cells have restricted the analysis to one dimensional separation before mass spectrometry[7,8] to avoid the sample loss in a 2nd dimension of separation, though limiting the ability of identifying proteins present in lower abundance.

In the present study, we employed a PPS facilitated lysis procedure combined with a high resolution two-dimensional separation to accommodate the small sample size of this study. After cell lysis, protein extracts were digested and equally divided into three aliquots. Each aliquot of around 300ng total material was then introduced into a two-dimensional separation by cIEF and nano-RPLC followed by tandem mass spectrometry analysis to identify the proteins present. A label-free protein quantification using spectral counts was employed to measure the protein fold changes between the CSCs group and the bulk tumor group. Ultimately, the signature proteins detected by the method were then uploaded to Ingenuity Pathway Analysis (IPA) for functional analysis to identify signaling pathways and protein-protein interaction networks that were significant in CSCs compared to bulk tumor cells.

Material and Experimental Procedure

1. Starting Material Preparation

a. Primary tumor specimen implantation—Samples of human pancreatic adenocarcinomas were obtained within 30 min following surgical resection according to Institutional Review Board–approved guidelines. Tumors were suspended in sterile RPMI 1640 and mechanically dissociated using scissors and then minced with a sterile scalpel blade over ice to yield two 2-mm pieces. The tumor pieces were washed with serum-free PBS before implantation. Eight-week-old male NOD/SCID mice were anesthetized using an i.p. injection of 100 mg/kg ketamine and 5 mg/kg xylazine. A 5-mm incision was then made in the skin overlying the mid-abdomen, and three pieces of tumor were implanted subcutaneously. The skin incision was closed with absorbable suture. The mice were monitored weekly for tumor growth for 16 weeks.

b. Preparation of single-cell suspensions of tumor cells—Before digestion with collagenase, low passage primary human pancreatic xenograft tumors from multiple mice were cut up into small pieces with scissors and then minced completely using sterile scalpel blades. To obtain single-cell suspensions, the resultant minced tumor pieces were mixed with ultrapure collagenase IV (Worthington Biochemicals, Freehold, NJ) in medium 199 (200 units of collagenase per ml) and allowed to incubate at 37 C for 1.5 to 2 hrs for enzymatic dissociation. The specimens were further mechanically dissociated every 15 to 20 min by pipetting with a 10-ml pipette. At the end of the incubation, cells were filtered through a 40- μ m nylon mesh and washed with HBSS/20% fetal bovine serum (FBS) and then washed twice with HBSS.

c. Flow cytometry—Dissociated cells were counted and transferred to a 5-mL tube, washed twice with HBSS containing 2% heat-inactivated FBS, and resuspended in HBSS with 2% FBS at a concentration of 10^6 per 100 μ L. Sandoglobin solution (1 mg/ml) was then added to the sample at a dilution of 1:20 and the sample was incubated on ice for 20 min. The sample was then washed twice with HBSS/2% FBS and resuspended in HBSS/2% FBS. Antibodies were added and incubated for 20 min on ice, and the sample was washed twice with HBSS/2% FBS. When needed, a secondary antibody was added by resuspending the cells in HBSS/2% FBS followed by a 20-min incubation. After another washing, cells were resuspended in HBSS/2% FBS containing 4,6-diamidino-2-phenylindole (DAPI; 1 μ g/mL final concentration). The antibodies used were anti-CD44 allophycocyanin, anti-CD24 (phycoerythrin), and anti-H2K (PharMingen, Franklin Lakes, NJ) as well as anti-ESA-FITC (Biomedex, Foster City, CA), each at a dilution of 1:40. In all experiments using human xenograft tissue, infiltrating mouse cells were eliminated by discarding H2K (mouse histocompatibility class I) cells during flow cytometry. Dead cells were eliminated by using the viability dye DAPI. Flow cytometry was done using a FACSAria (BD Immunocytometry Systems, Franklin Lakes, NJ). Side scatter and forward scatter profiles were used to eliminate cell doublets. Cells were reanalyzed for purity, which typically was >97%.

d. Material—10k Pancreatic cancer stem cells and 100k bulk tumor cells were obtained from mice xenografts after sorting by flow cytometry and gently washed three times with cold PBS (pH 7.4) by repetitive pipetting, followed each time by centrifugation at 1000g for 5min at 4°C. In the third time of washing, excessive PBS was gently sucked off with extra caution when cell pellets were observed at the bottom of the tube.

2. Cell lysis and Trypsin digestion

PPS (Protein Discovery, Knoxville, TN) powder was dissolved in 50 mM Ammonia Bicarbonate and was added to each tube at a final concentration of 0.2%(m/v). Around a 100 μ l cell suspension was then vortexed and incubated at 60°C for 10min, followed by sonication in an ice-water bath for 2 hrs. An aliquot of 5 mM DTT was added and the mixture was incubated at 60°C for 30min. After cooling, 15 mM iodoacetamide was added and the mixture was placed in the dark at room temperature for 30 min in order to allow the carboxymethylation reaction of cysteine residues. 50 mM ammonia bicarbonate was then added at a dilution ratio of 1:5 and 1:50(w/v) L-1-tosylamido-2-phenylethyl chloromethylketone modified sequencing-grade porcine trypsin (Promega, Madison, WI) was added. The mixture was incubated at 37°C in a water bath with agitation. Formic acid (FA) was then added to make a final concentration of 2% to stop the proteolysis. Following termination, the acidified mixture was placed in a 37°C water bath again for 4 hrs to facilitate the hydrolysis and allow the cleavage of PPS. The acidified tryptic peptide mixture was then desalted by a peptide micro-trap (Michrom, Auburn, CA) and eluted with 98% acetonitrile (ACN) and 0.3% FA, followed by spinning to dryness using a SpeedVac concentrator (Labconco, Kansas City) and stored in the -80°C freezer for future use. All chemicals were purchased from Sigma unless mentioned otherwise.

3. First Dimensional Separation: cIEF

A Beckman CE instrument was modified for cIEF with fraction collection. A 80 cm cIEF (100 mm id, 365 mm od) capillary was coated as previously described[10]. Lyophilized peptides were first reconstituted in gel buffer containing 2% ampholyte (pH 3-10) and were injected hydrodynamically to fill the capillary. Peptide focusing was performed by applying 21kv voltage to the two ends of the capillary using 0.1M phosphoric acid and 1mM sodium hydroxide as the anolyte and catholyte, respectively. The cathodic end of the capillary was kept in a stainless steel coaxial device. As the current reached its plateau, the focusing was complete and the focused peptides were mobilized under pressure and eluted into a 96-well auto-sampler plate by a 2 μ L/min flow of catholyte solution delivered by a syringe pump. The auto-sampler plate was moved from well to well automatically every 2 minutes by a Beckman fraction collector.

4. Second Dimensional Separation and Mass Spectrometry: nanoRPLC+nanoESI-MS/MS

When cIEF separation was completed, all cIEF runs with each containing ~30 pL fractions were injected in randomized order via Paradigm auto-sampler (Michrom Biosciences, Auburn, CA) and loaded onto a desalting nano-trap (300 $\times$ 50mm) (Michrom) connected to a nano-RP column (C18AQ, 5 μ m 200A, 100 $\times$ 150 mm)(Michrom) by a Paradigm AS1 micro-pump (Michrom). The mobile phases A and B were composed of 0.3% FA in water and 0.3% FA in ACN, respectively. Peptides were first desalted and enriched starting at 100%A with a flow rate of 50 μ L/min for 5 min. The sample was subsequently separated by a nano-RP column with a flow rate of 0.3 μ L/min after splitting. The linear gradient for separation was as follows: from 3% ACN to 12% ACN in 5 min, from 12% ACN to 40% ACN in 30min, from 40% ACN to 80% ACN in 15 min and finally decreased from 80% ACN to 3% ACN in 10min. The resolved peptides were then introduced into a ThermoFinnigan linear ion trap mass spectrometer (LTQ) (Thermo Electron, San Jose, CA) equipped with a nano-spray ion source (Thermo Electron). The LTQ was operated in data dependent mode in which one cycle of experiments consisting of one full MS scan was followed by five pairs of zoom scans and MS/MS scans with dynamic exclusion set to 30 s. The capillary temperature was set at 175°C, spray voltage was 2.8 kV, capillary voltage was 30 V and the normalized collision energy was 35% for the fragmentation.

5. Database Search and Protein Identification

MS/MS spectra were then searched against the human UniProt FASTA database by TurboSEQUEST provided by Bioworks ver3.1 SR1 (Thermo-Finnigan). The following modification was allowed in the search: 15.99 Da shift for oxidized Met residues; 58.1 Da shift for carboxymethylated Cys residues. The identified peptides were subsequently processed through PeptideProphet and ProteinProphet incorporated in the trans-proteomic pipeline (TPP: <http://proteinprophet.sourceforge.net/prot-software.html>)[11] where each protein was assigned with a probability indicating the significance level of the protein appearing in the original sample. In this study, we used a protein probability score of 0.99 as the threshold for protein identification and the FDR is below 0.2%.

6. Label-free Protein Quantitation and Data Transformation

Spectral counts were parsed out of TPP xml files after processing the SEQUEST data and used as a surrogate measure of protein abundance in our analysis. Global normalization was used to reduce technical bias when acquiring spectral count data from different runs between and across samples. The bias may come from instrument error or the inherent random sampling nature of the LTQ. Three replicated datasets for the CSCs group and four replicated datasets for the tumor group (denoted as CSC1, CSC2, CSC3 and tumor1, tumor2, tumor3, tumor4), containing proteins with 99% confidence or above, together with their spectral counts, were generated. The data were consolidated to form a matrix with seven columns; missing values were replaced with zero. To eliminate the discontinuity observed in simple count ratios when a protein shows spectral count 0, raw data were transformed according to Old *et al*[12], as originally proposed by Beissbarth *et al*[13] for serial analysis of gene expression (SAGE). The transformation uses the log₂ scale quantity

$$N = \log_2 [(n+f) / (t - n+f)] \quad (\text{Eq. 1})$$

for each protein, where *n* is the raw (globally normalized) spectral count value; *t* is the total number of spectra over all proteins in each dataset; and *f* is a correction factor. Larger values of *f* shrink the results for low spectral count proteins toward zero, thereby eliminating the discontinuity at zero and down weighting the results with greatest measurement error (i.e. the proteins with low spectral counts). Several procedures for setting the constant term *f* have been proposed; we devised a new approach that is suitable for experiments with technical replicates. We considered the following criterion:

$$R(f) = \sum \text{cor}(\text{between replicates}) \quad (\text{Eq. 2})$$

Correction factor *f* is defined to be the value that maximizes the correlation given in (Eq. 2). This maximizing effect yielded the value *f*=3. A schematic view of this computational process is depicted in Figure1(b).

After transformation, statistical significance levels between the CSCs group and the tumor group were then determined by student's t-test followed by multiple testing adjustment using FDR test. Differentially expressed proteins used for subsequent pathway analysis were declared at the level of *q*-value < 0.1. This FDR test was performed using R-package (<http://cran.r-project.org/web/packages/fdrtool/index.html>). Fold change (FC) is computed from transformed data using the mean of spectral counts from all replicates within a group: FC = (mean CSCs group) – (mean tumor group), where a positive sign indicates over

expression in the CSCs group and a negative sign indicates over expression in the tumor group.

7. Ingenuity Pathway Analysis (IPA)

To obtain detailed molecular information and infer significant signaling pathways from our global profiling results, differentially expressed proteins from the CSCs group and tumor group were uploaded to IPA. The uploaded Excel spreadsheet file contains the relevant proteins with their fold change, q-value and corresponding primary accession number. The significance values for canonical pathways were calculated using the right-tailed Fisher's Exact Test by comparing the number of proteins that participate in a given function or pathway relative to the total number of occurrences of these proteins in all functional/ pathway annotations stored in the ingenuity pathway knowledge base (IPKB).

Results and Discussion

1. Evaluation of cIEF+RPLC platform

Capillary isoelectric focusing is a powerful 1st dimensional separation for protein/peptides because of its high resolution and orthogonal separation mechanism versus RP-HPLC. This pI based separation provides an optimal resolution of 0.01 pH unit, which indicates a peak capacity of 700 in a pH range from 3 to 10, while strong cation exchange only has a peak capacity of around 50. SCX also presents undesired retention of peptides with strong interaction with the chromatographic resin and results in poor sample recovery rate. In contrast, cIEF is performed in an open capillary which is usually neutrally coated to prevent electroosmotic flow (EOF) and absorption of samples. Thus cIEF usually provides a sample recovery rate of higher than 90% (see supporting material⁴ for details) which is critical in analyzing the extremely small amount of sample in our pancreatic CSCs study.

The quality of the cIEF separation in terms of the resolution and reproducibility is essential to the accuracy of the comparative proteomic study. The theoretical pI value for each identified peptide within each fraction was calculated after database searching. The pI distribution plot from the first replicate of the CSCs group is shown in Figure 2(a). As expected, the pIs of the fractions decrease from 10 to 4 following a linear trend except the first 7 fractions where only a few proteins were identified. Peptides beyond the pH range of the ampholytes (pH 3–10) used in these experiments were not expected to be resolved. Overall, cIEF exhibited high separation resolution where more than 70% of the peptides were identified in no more than a single fraction. In addition, the off line collection method has the advantage of maintaining the separated peptide bands, whereas on line collection directly coupled to RPLC reduces the workload at the cost of sacrificing separation resolution and increased sample loss. The number of peptides identified from the four replicate runs of the same tryptic digest from the tumor sample is plotted against their pI values shown in Figure 2(b). The distribution of the peptides demonstrates excellent technical reproducibility of the experiment.

Reproducibility is also accessed by pairwise comparison of two selected replicates (CSCs replicate run 1 and 2, Tumor replicate run 1 and 2) using the Pearson correlation coefficient. No transformation was performed at this point. Common proteins identified in both replicate runs were used to calculate the correlation values (R) which are 0.87 with 95% confidence level between [0.84, 0.9] for CSC1 vs CSC2 shown in Figure2(c) and 0.91 for tumor1 vs tumor2 shown in Figure2(d) with 95% confidence level between [0.89, 0.93] shown in Figure2(d). A complete correlation matrix can be found in the supporting material¹.

2. Spectral Counting Results and Transformation

Detecting relative protein quantity change between various disease classes or different samples is central to understanding the molecular processes of the cell. Currently there are two widely used but fundamentally different label-free protein quantification strategies: spectral counting and peak-area measurement. The latter method requires aligning the retention time of the chromatogram peaks to accurately locate the same peptide and is preferred in experiments where peptides are subject to only one-dimensional separation before being analyzed by mass spectrometry for its accuracy, while the peak-area method is difficult to use with approaches containing two-dimensional separations where the same peptide might appear in more than one LC/MS/MS run. Spectral counting is relatively easy to apply to two-dimensional separation data. We have previously reported a study using spectral counting to analyze the protein expression levels in two ovarian cancer cell line samples and spectral counting results of selected proteins were all consistent with western blot experiments[9].

Different types of computational algorithms regarding the processing strategies of spectral counting data have been proposed by a number of groups[10-13]. Boris et al.[13] proposed a NSAF method to normalize and transform spectral count data based on protein length whereas Old et al.[10] adopted a transformation method to avoid the discontinuity problem which was originally used for serial analysis of gene expression. In the present study, spectral counts were assigned to each identified protein followed by global normalization by adjusting the mean of each dataset to be equal. After consolidating all 7 datasets into a matrix and replacing missing values with zero, we proceeded to identify the correction parameter f that produces the most meaningful abundance measurements for our experimental data. The transformation (Eq. 1) that we applied has the effect of shrinking the expression scores for low spectral count proteins toward zero, compensating for their relatively greater uncertainty. By maximizing $R(f)$ (Eq. 2), the data was transformed in such a way that low-abundance proteins were appropriately, but not excessively down weighted in the analysis. The approach to defining f by maximizing $R(f)$ is based on the fact that technical replicates from the same biological sample should be intrinsically the same. The correction factor was calculated to be 3 and then the whole dataset was transformed accordingly. This is similar to Old's method[10] that f is adjusted according to higher correlation between expected and observed. The fitted transformation, shown in Figure 3, indicated that above 300 expression units on the original scale, the measured data were sufficiently reliable to be used without adjustment; between 100 and 300 expression units on the original scale the data were substantially shrunk toward zero, but still contributed to the analysis; and below 100 units, the data were shrunk to the degree that they have little influence in the analysis.

An important issue with any data transformation scheme that uses the class labels or clinical outcome information is the possibility that it may induce artifactual correlations between protein abundance and outcomes. The transformation function (Eq. 1) was monotonic, meaning that if the spectral counts of protein A were larger than those of protein B, this relationship would continue to hold after the transformation is applied. A monotonic transformation is limited in its ability to induce spurious correlations. To further explore whether the transformation induced spurious correlations, we conducted a simulation study. We simulated spectral counts for 1000 proteins measured in two groups of samples, each with three replicates. All 6,000 data values were simulated independently from a standard exponential distribution. We applied the procedure described above for a range of f values from 0 to 100, and considered the number of Z-statistics (comparing the two groups of three samples for a given protein) that exceeded various thresholds (2, 2.5, and 3). We did not observe any inflation in the number of significant associations for $f > 0$ compared to $f = 0$. Furthermore, we observed in our experimental data that the results were not very sensitive to

the specific value of f , as long as it was not too close to zero. Figure 4(a) shows the clustering results after transforming by $f=3$. The CSCs group and the tumor group are well separated without introducing artificial interference.

3. Protein Profiling Results

Accepted protein identifications were obtained after applying a cutoff protein probability of 0.99 by TPP which ensures the FDR is below 0.2%. In addition to this bayes approach based FDR method, we also tested the identification confidence by applying a target decoy database search on one of the tumor datasets. A 0.9 protein probability filter resulted in a FDR of 1.4% and a 0.99 protein probability filter further decreased the decoy hits to zero. A total of 763 and 1031 distinct proteins were identified from three CSCs replicates and four tumor replicates, respectively. The detailed information about these identifications can be found in supporting material2.

Each identified protein was assigned a cellular location based on information from IPKB. Figure5(a) and Figure5(b) show the cellular distribution of 763 and 1031 identified proteins from the CSCs group and the tumor group, respectively. The cellular distribution is consistent for both of these two groups: the majorities are cytoplasmic and nuclear proteins; plasma membrane proteins occupy 10% of each total proteome, suggesting PPS has the ability to extract hydrophobic proteins. Multiple correcting testing (FDR) was then performed to capture the differentially expressed molecules and 161 out of 1159 proteins were identified by using a threshold of $q\text{-value} < 0.1$. 24% of these differentially expressed proteins show up-regulation in the CSCs group and a few of them are related to the key signaling pathways of CSCs. For example, inter-alpha trypsin inhibitor H3 (ITIH3) was identified in all 3 runs of the CSCs group with two unique peptides and an average spectral count larger than 10, while it was only identified in one of the tumor runs with low spectral counts. This protein associated with inflammatory response in local tissue[14]was previously reported to be one of the downstream target genes of Sonic Hedgehog (Shh), which plays a key role in signaling pathways that directly activate the genes involved in the self-renewal and apoptosis-inhibition functions of CSCs[15]. The over-expression of ITIH3 is consistent with previous finding of the up-regulated Shh at the mRNA level in pancreatic CSCs[2]. In contrast, a mitochondrial apoptosis-inducing factor (AIFM1) was down-regulated in the CSCs group. This protein was identified in all 4 runs of the tumor group with 5 unique peptides, whereas it was only identified in one of the CSCs runs. The decreased protein level detected in the stem cell group from our study agrees with previous reports that inactivation of AIFM1 renders embryonic stem cells resistant to cell death [16]. Beside these significantly differentially expressed proteins, some important low-abundant proteins were also detected although they were not found to be significantly different based on the FDR test. For example, NF- κ B was identified in one of the CSCs runs with two unique peptides but not in any of the tumor runs, as well as c-MET and CXCL5. Their absence in the tumor group is mainly due to their low abundance level which was below the detection limit, however their relative over expression in the CSCs group agrees with previous findings [17-21] that the elevated expressions of these proteins are related to the properties of CSCs.

4. Signaling Pathway and Connectivity Network Analysis

A list of 169 differentially expressed proteins was uploaded into IPA for functional annotation and pathway analysis. The detailed information about these molecules can be found in supporting material3. The most variant and relevant canonical signaling pathways enriched with differentially expressed molecules between the CSCs group and the tumor group were generated by IPA and are ranked by significance shown in figure6 with a

threshold of p -value < 0.1 . The length of the bars indicate the significance of the signaling pathways to which the differentially expressed proteins are related.

These enriched pathways can be grouped into four main categories related to the characteristics of cancer stem cells: resistance to apoptosis, dysregulation of cell proliferation, association with inflammation and metastasis. The top pathway, Mitochondrial Dysfunction, is related to both apoptosis and tumorigenesis[22,23] and a recent report has linked mitochondrial dysfunction to ovarian cancer stem cells[24]. Pathways mapped to cellular growth, proliferation and development by IKGB include ILK Signaling, RhoA Signaling and Integrin Signaling. CXCR4 signaling and Acute Phase Response signaling pathways categorized under cellular immune/inflammatory response are also shown to be significantly involved. The connection between inflammation and tumorigenesis has been recognized in many pathologic conditions including pancreatic cancer where some clues are suggesting that inflammation might induce an accelerated process of mutagenesis and mutation accumulation[1,25]. VEGF signaling associated with angiogenesis is also shown to be significant.

To further infer the functional relevance between these differentially expressed molecules, we have constructed connectivity networks by IPA. The top network which has the highest score of 54 (p -value $< 1 \times 10^{-54}$ from Fisher Exact Test) consists of 27 signature proteins shown in figure 7. Interestingly, NF- κ B is imported as the central node to generate this interaction network by IPA, suggesting a potential involvement of this transcription factor although this molecule is not in the experimental signature protein list.

Conclusion

This work represents the first proteome profiling study on pancreatic cancer stem cells from xenografted tumors in mice. We overcome the difficulty of analyzing the extremely small number of cancer stem cells from the xenografted tumor by using an ultrasensitive sample preparation procedure and cIEF as the 1st dimensional separation before LC/MS/MS to minimize sample loss and obtain high resolution of peptides. A modified transformation algorithm was also devised to handle spectral counting data with technical replicates in order to weight proteins with a large range of spectral counts appropriately in the subsequent statistical analysis. 169 proteins have been captured as differentially expressed signatures between the CSCs group and the bulk tumor group. Pathway analysis and network modeling by IPA has further revealed significantly involved signaling cascades relevant to the characteristics of CSCs.

It will be important to validate proteome profiling results using alternative technologies such as the Western blotting or RT-PCR. However, currently, no other techniques work effectively with less than 1 μ g of protein material. To compensate for the lack of validation, we increased the confidence level of protein identifications by adopting a more stringent threshold. We set the cutoff to a protein probability score of 0.99 which has a FDR less than 0.2% compared to the commonly used score of 0.9 which gives FDR less than 1%. Moreover, multiple testing corrections are employed to control false positives induced by performing significance tests on a large number of proteins. Thus, differentially expressed proteins identified in this way are more likely to be true positives.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the National Cancer Institute under grant R21CA134623 and the National Institutes of Health under R01GM49500.

References

1. Lee CJ, D J, Simeone DM. Pancreatic cancer stem cells. *J Clin Oncol*. 2008; 26(17):2806–12. [PubMed: 18539958]
2. Li C, H D, Dalerba P, Burant CF, Zhang L, Adsay V, Wicha M, Clarke MF, Simeone DM. Identification of Pancreatic Cancer Stem Cells. *Cancer Research*. 2007; 67(3):1030–1037. [PubMed: 17283135]
3. Fang X, B B, Wang W, Park DM, Lee CS. Comparison of multidimensional shotgun technologies targeting tissue proteomics. *Electrophoresis*. 2009; 30(23):4063–70. [PubMed: 19960471]
4. Sodek KL, E A, Ignatchenko A, Agochiya M, Brown TJ, Ringuette MJ, Jurisica I, Kislinger T. Identification of pathways associated with invasive behavior by ovarian cancer cells using multidimensional protein identification technology (MudPIT). *Mol Biosyst*. 2008; 4(7):762–73. [PubMed: 18563251]
5. Elschnerbroich S, I V, Sharma P, Schmitt-Ulms G, Gramolini AO, Kislinger T. Peptide separations by on-line MudPIT compared to isoelectric focusing in an off-gel format: application to a membrane-enriched fraction from C2C12 mouse skeletal muscle cells. *J Proteome Res*. 2009; 8(10):4860–9. [PubMed: 19670906]
6. Gortzak-Uzan L, I A, Evangelou AI, Agochiya M, Brown KA, St Onge P, Kireeva I, Schmitt-Ulms G, Brown TJ, Murphy J, Rosen B, Shaw P, Jurisica I, Kislinger T. A proteome resource of ovarian cancer ascites: integrated proteomic and bioinformatic analyses to identify putative biomarkers. *J Proteome Res*. 2008; 7(1):339–51. [PubMed: 18076136]
7. Hwang SI, T J, Lundgren DH, Rezaul K, Mayya V, Wu L, Eng J, Wright ME, Han DK. Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene*. 2007; 26(1):65–76. [PubMed: 16799640]
8. Bagnato C, T J, Mayya V, Hwang SI, Zebroski H, Claffey KP, Haudenschild C, Eng JK, Lundgren DH, Han DK. Proteomics analysis of human coronary atherosclerotic plaque: a feasibility study of direct tissue proteomics by liquid chromatography and tandem mass spectrometry. *Mol Cell Proteomics*. 2007; 6(6):1088–1102. [PubMed: 17339633]
9. Dai L, L C, Shedden KA, Misek DE, Lubman DM. Comparative proteomic study of two closely related ovarian endometrioid adenocarcinoma cell lines using cIEF fractionation and pathway analysis. *Electrophoresis*. 2009; 30(7):1119–31. [PubMed: 19288585]
10. Old WM, M-A K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinisky JR, Resing KA, Ahn NG. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics*. 2005; 4(10):1487–502. [PubMed: 15979981]
11. Zybailov B, M A, Sardi ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res*. 2006; 5(9):2339–47. [PubMed: 16944946]
12. Choi H, F D, Nesvizhskii AI. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics*. 2008; 7(12):2373–85. [PubMed: 18644780]
13. Zybailov, Boris; M, AL.; Sardi, Mihaela E.; Coleman, Michael K.; Florens, Laurence; Washburn, Michael P. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research*. 2006; 5(9):2339–2347. [PubMed: 16944946]
14. Zhuo L, K K. Structure and function of inter-alpha-trypsin inhibitor heavy chains. *Connect Tissue Res*. 2008; 49(5):311–20. [PubMed: 18991084]
15. Kato M, S N, Sugano S, Hashimoto K, Masuho Y, Muramatsu M, Kaibuchi K, Nakafuku M. Identification of sonic hedgehog-responsive genes using cDNA microarray. *Biochem Biophys Res Commun*. 2001; 289(2):472–8. [PubMed: 11716497]
16. Joza N, S S, Daugas E. Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death. *Nature*. 2002; 410(6828):549–54. [PubMed: 11279485]

17. Donahue TR, H O. CXCR2 and RET single nucleotide polymorphisms in pancreatic cancer. *World J Surg.* 2009; 33(4):710–5. [PubMed: 19057948]
18. Wente MN, K M, Burdick MD, Friess H, Büchler MW, Ceyhan GO, Reber HA, Strieter RM, Hines OJ. Blockade of the chemokine receptor CXCR2 inhibits pancreatic cancer cell-induced angiogenesis. *Cancer Lett.* 2006; 241(2):221–7. [PubMed: 16458421]
19. Hansel DE, R A, House M, Ashfaq R, Berg K, Yeo CJ, Maitra A. Met proto-oncogene and insulin-like growth factor binding protein 3 overexpression correlates with metastatic ability in well-differentiated pancreatic endocrine neoplasms. *Clin Cancer Res.* 2004; 10(18):6152–8. [PubMed: 15448002]
20. Alvero AB, C R, Fu HH, Montagna M, Schwartz PE, Rutherford T, Silasi DA, Steffensen KD, Waldstrom M, Visintin I, Mor G. Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. *Cell Cycle.* 2009; 8(1):158–66. [PubMed: 19158483]
21. Birnie R, B S, Roome C, Dussupt V, Droop A, Lang SH, Berry PA, Hyde CF, Lewis JL, Stower MJ, Maitland NJ, Collins AT. Gene expression profiling of human prostate cancer stem cells reveals a pro-inflammatory phenotype and the importance of extracellular matrix interactions. *Genome Biol.* 2008; 9(5):R83. [PubMed: 18492237]
22. Newmeyer DD, F-M S. Mitochondria: releasing power for life and unleashing the machineries of death. *Cell.* 2003; 112(4):481–90. [PubMed: 12600312]
23. Lu J, S L, Bai Y. Implications of mitochondrial DNA mutations and mitochondrial dysfunction in tumorigenesis. *Cell Res.* 2009; 19(7):802–15. [PubMed: 19532122]
24. Bapat S. Evolution of cancer stem cells. *Semin Cancer Bio.* 2007; 17(3):204–13. [PubMed: 16787749]
25. Guerra C, S A, Cañamero M, Grippo PJ, Verdaguer L, Pérez-Gallego L, Dubus P, Sandgren EP, Barbacid M. Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. *Cancer Cell.* 2007; 11(3):291–302. [PubMed: 17349585]

Figure1.
Figure1(a): Experimental flow chart

Figure1(b): Data Processing Strategy. Upper left matrix is the consolidated dataset. Lower left flowchart is the correction factor searching scheme and transformation algorithm.

Figure2.

Figure2(a): Theoretical pI distribution plot of the first run of CSC group. Fraction number shown in the X -axis is plotted against the average of peptides' pI value within each fraction shown in the Y -axis.

Figure2(b): Distribution of number of identified peptides from each run of tumor group across pI range between 3.5 to 10. X -axis shows their pI value and Y -axis shows the number of identified peptides. Different tumor replicate runs are represented by different colors.

Figure2(c): Pearson correlation plot of all proteins detected with single or more spectral counts in the first and the second run of CSC group.

Figure2(d): Pearson correlation plot of all proteins detected with single or more spectral counts in the first and second replicate of tumor group.

Figure3. Monotonic plot of original data Vs transformed data. Different color and different shapes represent csc1, csc2, csc3, tumor1, tumor2, tumor3, tumor4, respectively. Y-axis represent original data and X-axis represent transformed data on log2 scale.

Figure4.
Clustering results of the CSC group and tumor group after transforming by f=3.

Figure 5.

Figure 5(a): Cellular Distribution of identified proteins from pooled CSC group.

Figure 5(b): Cellular Distribution of identified proteins from pooled tumor group.

Figure6.
Canonical signaling pathways enriched with differentially expressed proteins ranked by significance. A threshold $p\text{-value} < 0.1$ is applied.

Figure7.
The top1 connectivity network constructed by IPA. This network only consists of differentially expressed proteins from experimental data. Red and green circles indicate overexpression and underexpression in the CSC group versus the bulk tumor group, respectively.

Figure8.

Global proteome profiling of pancreatic cancer stem cells(CSCs) is a promising way to unveil the molecular machinery underlying the signaling pathways. However, the extremely low availability of pancreatic tissue CSCs has limited the utilization of current standard proteomic approaches. We describe the profiling of the proteome of pancreatic CSCs using a capillary scale shotgun technique by coupling off-line capillary isoelectric focusing(cIEF) with nano reversed phase liquid chromatography(RPLC) to profile the proteome from < 1ug total protein.