Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1053 (2004) 279–290
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Microscale protein expression profiling during disease evolvement

György A. Marko-Varga*, Thomas E. Fehniger

Department of Biological Sciences, AstraZeneca R&D, Respiratory and Inflammation, S-221 87 Lund, Sweden

Abstract

Advances in technology, such as laser capture microdissection (LCM), have allowed for the specific sampling of cells within their natural functional micro-environment. In model systems using LCM, we have studied the global protein expression profiles of airway epithelial cells during a response to allergen provocation. Bronchial epithelial cells were first identified and phenotyped histologically in snap frozen lung samples of experimentally sensitised mice. Consecutive thin sections of whole lung were then sampled using preparative LCM procedures. Lysates of the captured epithelium (7500 shots) or whole lung were prepared for two-dimensional gel electrophoretic separation and 1400 protein spots were annotated by image analysis. Protein identities were established by matching peptide masses detected using matrix-assisted laser desorption ionization time-of-flight MS as well as electrospray ionization MS–MS sequencing. Using the Mascot database of protein/peptide identities high significance scores in terms of sequence coverage (range 22–70%) and number of peptides (range 7–22 peptides/protein) were obtained for approximately 500 proteins, with examples listed in Table 1. In quantitative terms, the LCM procedure allows the statistical sampling of singular populations of cells distributed throughout tissues and organs. The absolute number of cells required for "entry level" measurements of protein profiles will vary over an order of magnitude depending on the physical size and frequency of the cells being studied within each biological compartment as well as the dynamic range of the proteins being measured, and the absolute limits of detection within the technologies being employed.

Keywords: Protein expression; Laser capture microscopy; Proteins

1. Introduction

A major focus of bio-medical research today are fundamental studies to unravel the complex biological and cellular interactions within tissues and organs, which occur at sites of dysfunction or pathological processes. Modern technology allows for studying the biology and biological chemistry of health and disease, by contrasting steady-state functions such as repair, growth and regulated gene and protein expression within various cellular compartments [1–3]. However, the contributions of specific cell types within specialized tissue compartments are often difficult to resolve in tissue composed of heterogeneous resident and non-resident inflammatory cells varying in both activation and differentiation states.

Proteomics expression studies of disease processes must allow access to the same sets of proteins that are occurring during changes in the temporal states of disease. The fact is though that most clinical sample studies provide only a cross-sectional glimpse into the natural progression of disease, with each sample representing its unique individual time point of sampling.

It is likely that in order to acquire samples with disease context over extended sampling periods, we will need to develop high sensitivity assays from relatively small amounts of sample. This is especially true for studies of protein expression within tissue, where samples are typically acquired by invasive biopsy during diagnostic and/or treatment procedures. Importantly, the dynamic range of expression of proteins imposes a further restriction on measurements due to the low abundance levels of many biologically important proteins. In order to solve these problems, we will need to find ways to technically sample and detect proteins at low concentration levels.

* Corresponding author. Tel.: +46 6 336000; fax: +46 6 337383.
E-mail addresses: gyorgy.marko-varga@astrazeneca.com (G.A. Marko-Varga), thomas.fehniger@astrazeneca.com (T.E. Fehniger).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.08.115

Laser capture microdissection (LCM) was introduced at the US National Institutes of Health (NIH) in 1997 as a tool for studying the contribution of individual cell types to the patterns of protein expression which promote disease development [4]. To date, this technology has most often been utilized to isolate tumour cells for use in mRNA microarray studies. Microsectioning tissue and LCM isolation of immunostained frozen sections has also proven to be a particularly powerful combination for dedicated RNA analysis [5]. Rekhter and Chen have lined out some of the more practical details of the protocols needed to obtain high quality microdissections for global analytical approaches [6]. However, only a few studies have applied LCM specifically to map protein expression within tissues of disease interest [7–16]. The more recent applications of LCM have used the technology to approach the difficult questions of characterizing and quantifying protein expression patterns within very small number of cells (∼10 000 cells). A fine example of the use of LCM in this context is the recent study by Karger's group on ductal epithelium from breast cancer patients [17]. Using isotopic labelling in combination with MS identification, the study succeeded in identifying a set of proteins which were differentially regulated in malignant ductal epithelial cells in comparison to non-cancerous controls. The cells analysed in this study represented only 1–4 μg of total starting material and thus provides a look into the possibilities of microscale analyses.

We have begun to address these issues of quantitative biology by developing model analyses systems that allow accurate determinations of proteins from defined cellular compartments which represent prototypic differences in histopathological phenotype. The specific aim of the study model presented here was to achieve a biological resolution which approaches singular cells from specific regions or compartments of tissue: for example, within the conducting airways of experimentally provoked allergen challenge, where histological disease state characterization could be performed. In order to obtain this level of resolution, we have applied LCM to obtain enriched samples of the activated mucosal epithelium. We have developed both qualitative and quantitative approaches for determining the specific patterns of proteins expressed in naïve and challenged airways with the intent of identifying as many unique proteins as possible using high resolution separation and sequence identifications by mass spectrometry. These combined approaches offer both utility and sensitivity for studying complex cellular and biological interactions at the very sites of disease development.

2. Experimental

2.1. Experimental animals

All animal studies were performed under protocols approved by the Malmö/Lund Ethical Committee for Animal Experiments (M254-99). Female BALB/c mice, weighing 20–25 g, were purchased from Bomholtgaard, Denmark and housed in plastic cages with pine chip bedding (10 mice/cage). The animal room was maintained at 22 °C with a daily light–dark cycle (06:00–18:00 light) and fed with chow and water ad lib. The mice were 8 weeks old at the time of study.

2.2. Sensitisation and provocation

We have adapted a previously described model of allergic airway inflammation in which ovalbumin is used for sensitisation and challenge [18,19].

2.3. Tissue sampling

The lungs of five animals were sampled from each of the control and allergen challenged groups. The mice were inserted with a canula into the trachea and the lungs were slowly inflated by injection with 0.7 ml of a 66% solution of Tissue-Tek O.C.T. (Sakura Finetek, Torrance, CA) in phosphate-buffered saline (PBS), pH 7.2. The trachea was tied off, and the whole lung was dissected out and the left lung lobe was carefully removed and placed immediately into a bath of isopentane on dry ice for snap freezing. The lungs were placed into storage vials and stored at −70 °C until sectioning and analysis. In some experiments, whole liver was dissected, snap frozen, and processed for analysis.

2.4. Histological staining

In some experiments, the O.C.T. sufflated lungs were placed into a vessel and fixed overnight with formalin, embedded in paraffin, and thin sections cut in the sagittal plane were placed onto glass slides. These slides were then stained using conventional methods.

2.5. Sample preparation

For studies of global protein expression, whole lung was thin sectioned in the sagittal plane on a cryostat (10 μm thick) as central transverse biopsies and kept frozen at −70 °C until use. To solubilise the whole tissue, three whole sections were placed into Eppendorf tubes containing 200 μl isoelectric focusing (IEF) lysing solution, in 7 M urea, 2 M thiourea and 4% 3-[(cholamidopropyl)dimethylamine]-1-propanesulfonate (CHAPS) and vortexed vigorously for 1 min and then centrifuged before further use. Whole liver was processed similarly.

2.6. Two-dimensional gel electrophoresis (2DE)

Immobiline dry strips (180 mm, pH 3–10 NL) were rehydrated in 350 μl of the solubilisation solution containing 7 M thiourea, 2 M urea, 4% CHAPS, 10 mM dithiothreitol (DTT), and 0.5% IPG 3–10 buffer, together with the fractionated samples (100 μl).

The IEF step was performed at 20 °C in a IPGphor (Amersham Pharmacia Biotech, Uppsala, Sweden) and run according to the following schedule: (1) 30 V for 10 h, (2) 500 V for 1 h, (3) 1000 V for 1 h, and (4) 4000 V until approximately 45 000 V h were reached. The strips were equilibrated for 10 min in a solution containing 65 mM DTT, 6 M urea, 30% (w/v) glycerol, 2% (w/v) sodium dodecyl sulfate (SDS) and 50 mM Tris–HCl, pH 8.8. A second equilibration step was also carried out for 10 min in the same solution except for DTT, which was replaced by 259 mM iodoacetamide. The strips were soaked in electrophoresis buffer (24 mM Tris base, 0.2 M glycine and 0.1% SDS) just before the second-dimensional gel electrophoresis. The strips were applied on 14% homogeneous Duracryl slabgel. The strips were overlaid with a solution of 1% agarose in electrophoresis buffer (kept at 60 °C). Electrophoresis was carried out in a Hoefer DALT gel apparatus (Amersham Pharmacia Biotech, San Francisco, CA, USA) at 20 °C and constant 100 V for 18 h. All lung samples were run in triplicate.

2.7. Gel staining

Gels were stained with silver according to Shevchenko et al. [20].

2.8. Spot analysis

Gels were scanned using a Fluor-S MultiImager (Bio-Rad Labs., Sundbyberg, Sweden). Spot analysis was performed using the PDQUEST (version 6.1.0) two-dimensional gel analysis system (Bio-Rad discovery series, Bio-Rad Labs.).

2.9. Mass spectrometry identification

Mass spectrophotometry was performed as previously described [19]. Briefly, the matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) instrument used was a Voyager DE-PRO (Perseptive Biosystems, Framingham, MA, USA) mass spectrometer. The instrument, equipped with a delayed extraction ion source, utilized a nitrogen laser at 337 nm and was operated in reflectron mode at accelerating voltages of 20 kV. The sample probes were made of polished stainless steel. Sample deposition of nanoliter fractions was made on stainless steel MALDI-target plates and on a 100-position stainless steel target plate (Perseptive Biosystems). For protein sequencing, a Q-TOF2 (Micromass, Manchester, UK) electrospray ionization (ESI) MS–MS instrument was used.

2.10. Microdissection

Hepatocytes or airway epithelial cells were microdissected by LCM (Pixcell II, Arcturus Engineering, Mountain View, CA, USA). The laser spot size used were 30 μm in diameter (pulse power: 30 mW; pulse width: 5.0 ms; threshold voltage: 250 mV).

2.11. Protein expression analysis of LCM isolated epithelium

The epithelial mucosa attached to the LCM polymer cap were lysed by IEF lysing solution containing 7 M urea, 2 M thiourea and 4% CHAPS and vortexed vigorously for 1 min. The IEF lysing solution was then reapplied to another cap holding cells from the same microdissected case, and the procedure was repeated until each tube contained material from approximately 7000 shots. The solubilised epithelial cells were then equilibrated with the Immobiline strip overnight and subsequently run on the second-dimension SDS gel and then analysed as above as earlier.

2.12. Protein identification and annotation

The proteins were annotated and identified in a two-step procedure as partly previously described [21]. First, image analysis was performed by hierarchic differential analysis where all protein spots were annotated as expressed entities on the 2D gels. These annotations were next quantified by mean intensity values from the respective animal group, where the lowest regulation factors defined in this study was set to at least 2.0. Protein spots of interest were excised by the use of an automated spot cutting instrument (Proteome Works Spot Cutter, Bio-Rad, Hercules, CA, USA), and positioned in a 96-well plate. These 96-well plates were digested and micro-extracted as described previously [19,22]. The second step protein identity was performed by applying both MALDI and ESI tandem mass spectrometry, using SwissProt and NCBI databases for MS-spectra query.

3. Results and discussion

The workflow of our dedicated analysis systems for analysing and identifying proteins within tissue is outlined in Fig. 1. The systems have the flexibility for the analyses of protein expression either in global surveys or as focused studies of targeted tissue compartments. For global protein analyses of whole tissue, we typically perform a histological characterization of the study samples in parallel to the protein identification. This allows us to standardize the cellular content of the sample in terms of cell numbers, cell phenotypes, and to assign functional biological context to the micro-environment being analysed. We can then subgroup the samples in terms of their histological similarities and heterogeneities.

Sample preparation steps process the solubilised components of the tissue for fractionation based upon physiochemical properties of charge, mass, hydrophobicity, and solvent partitioning in either of the alternative separation and fractionation platforms. The technology toolbox for assessing protein expression, is composed of dedicated platforms for either two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), or for by either multidimensional

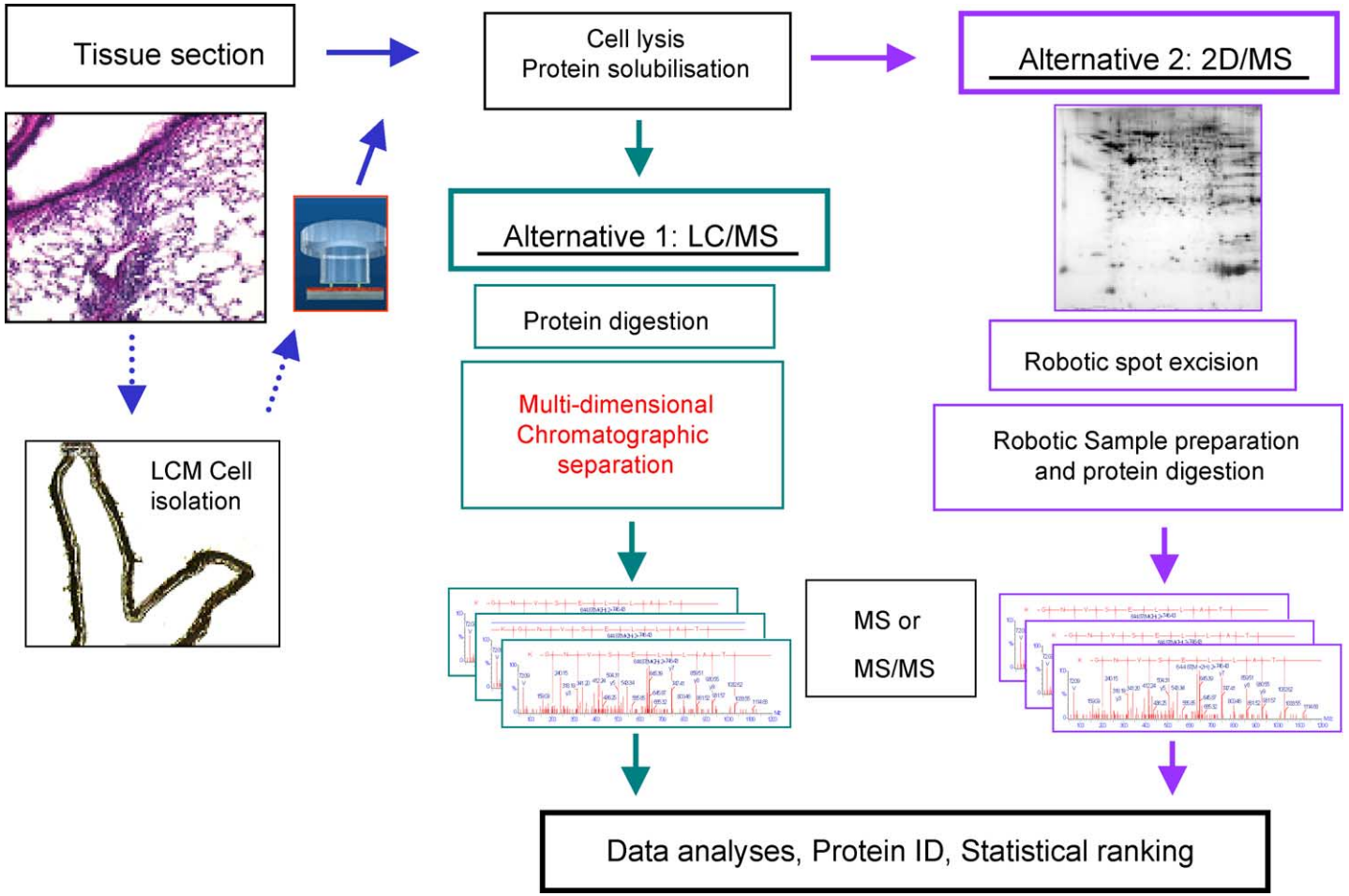

Fig. 1. The workflow process for generating protein expression maps from tissue using alternatively multi-phase chromatography or 2D-PAGE for protein separation and mass spectrometry technologies for identification.

liquid chromatography. Resulting chromatography fractions, or excised gel spots are determined by MS, and MS–MS. 2D-PAGE is performed within our research team using both large (24 cm × 20 cm), medium (13 cm × 12 cm) and small sized (8 cm × 8 cm) 2D gels. The larger the gel, the higher the resolution, and the size chosen will be determined by the actual cell number and total protein amount available from the biological sample. The practical experimental basis will always determine the tools we use from our technology toolbox. For example, we have developed protein expression maps using liquid phase separations with only one separation mode, reversed phase separation. This is applicable to samples where we have invested considerable time to make sub-cellular isolation and fractionation of any given sample type with a given sample complexity. The adaptation, choosing the right platform has two major considerations. First is the time it takes to run through a sub-set of clinical samples. The second consideration most proteomics groups face is the number of fractions that the separation platform produces linked to the dimensions of the liquid phase separation system utilized. The actual number of fractions to be analysed will generate a major work load for the mass spectrometer, that may be ESI and/or MALDI. These MS work horses will create very large data file numbers that will require server space often in the terra-byte size. The method of choice, i.e. the route taken to determine the protein expression pattern, within the tissue of interest, will be determined by the biological question being asked. This pragmatic proteomics strategy results in an unbiased approach where, in principle, any type of new development in separation methodology could find its place within our technology toolbox.

In some experiments, the specific tissue compartments of interest within the global sample, such as the epithelium, sub-mucosa, or infiltrating inflammatory cells are isolated using LCM. The LCM method may be applied to virtually any "soft" tissue, whether homogeneous or heterogeneous in cellular content (Fig. 2A). Bone, cartilage, tooth, and nail are less applicable. Fig. 2B illustrates the principles of the LCM process for obtaining cell isolations from tissue. The tissue section is positioned in the instrument adaptor under a polymer cap used for recovering the dissected tissue (see Fig. 2B, before). LCM makes use of a pulsing thermal laser source that excises the cells of interest from their tissue position using an adjustable laser pulse. The width of the beam ranges between 7.5 and 30 μm in diameter. The adjustable laser diameter has great utility allowing cell isolation from distinct regions of interest with differing sizes of cells within the very same section. For example, both small cells (i.e. diameter ≅10 μm; eosinophils, lymphocytes) as well as larger (i.e. diameter ≅25 μm; hepatocytes, monocytes) can be isolated as singular cells. The heat generated by the high frequency laser, will melt the polymer over the micro-environment that has been chosen for isolation which anneals the cells to the cap. Following the mechanical removal of the cap from the tissue surface, the resulting "captured sample" is ready for proteomic analysis. Typically, we solubilise the sample with a minimum volume of lysis buffer (5 μl).

The technology toolbox gains value when it is used to address and answer questions regarding real biological processes, and especially those processes in which structural and/or functional dysregulation promote the evolvement of disease. Ideally, a good experimental model of disease should bear the characteristics of the dynamic processes which operate during actual human disease. The differential regulation of key proteins within signal transduction pathways or proteins that can act as markers for a given disease state are examples of study areas being investigated today in many laboratory model systems. It is difficult to find models for all attributes of human disease, and this is especially true for chronic diseases and for diseases which produce morphological changes or tissue remodelling which promote pathological processes that effect function. Respiratory epithelial cells lining the airways, represent one such dynamic population of cells which can be modelled: with the dual characteristics of (1) self-replenishing structural elements with multiple functional activities ranging from mediator and mucous production to shielding from environmental insult, and (2) cells with intimate biological contact with other cells populating the laminar matrix and with migratory and inflammatory cells [23,24]. Both gene and/or protein expression alterations by these cell types can be linked to various processes which predispose or even promote disease, such as the production of growth factors, the remodelling of the mesenchymal matrix, or the production of excess mucus.

In this study, we have used an in vivo allergic airway model which produces an inflammatory response within the lungs of animals sensitised and challenged with ovalbumin (OVA). The model produces a number of changes to the airways and vasculature following an intensive recruitment and activation of inflammatory cells. In order to address the biological changes which occur following the OVA challenge to the micro-environment surrounding the epithelial lining of the major conducting airways, we differentially stained the lung sections using a variety of conventional histological staining methods. In Fig. 3A is shown a scaled micrograph of a representative thin section of the left lung lobe used in these studies. The sizes of tissue sections were in the region of 0.5–1.5 cm. Following sensitisation and challenge with OVA, the epithelial cells lining the central conducting airways become hyper-trophic and metaplastic with prominent mucus hyper-secretion (Fig. 3B). Shown here in the airway lumen are strings of mucus being expressed by the epithelium following allergen challenge. Inflammation by migratory cells trafficking into the airway compartment is also observed while the sub-mucosa within this compartment consisting of the basement membrane and the mesenchymal matrix of collagen and connective tissue elements within the elastic lamina, as well as the fibroblasts and smooth muscle cells, all become exposed, to this environment. By observing the full microanatomy of these lungs (Fig. 3C), we gain a fuller understanding of the context of the dynamic processes

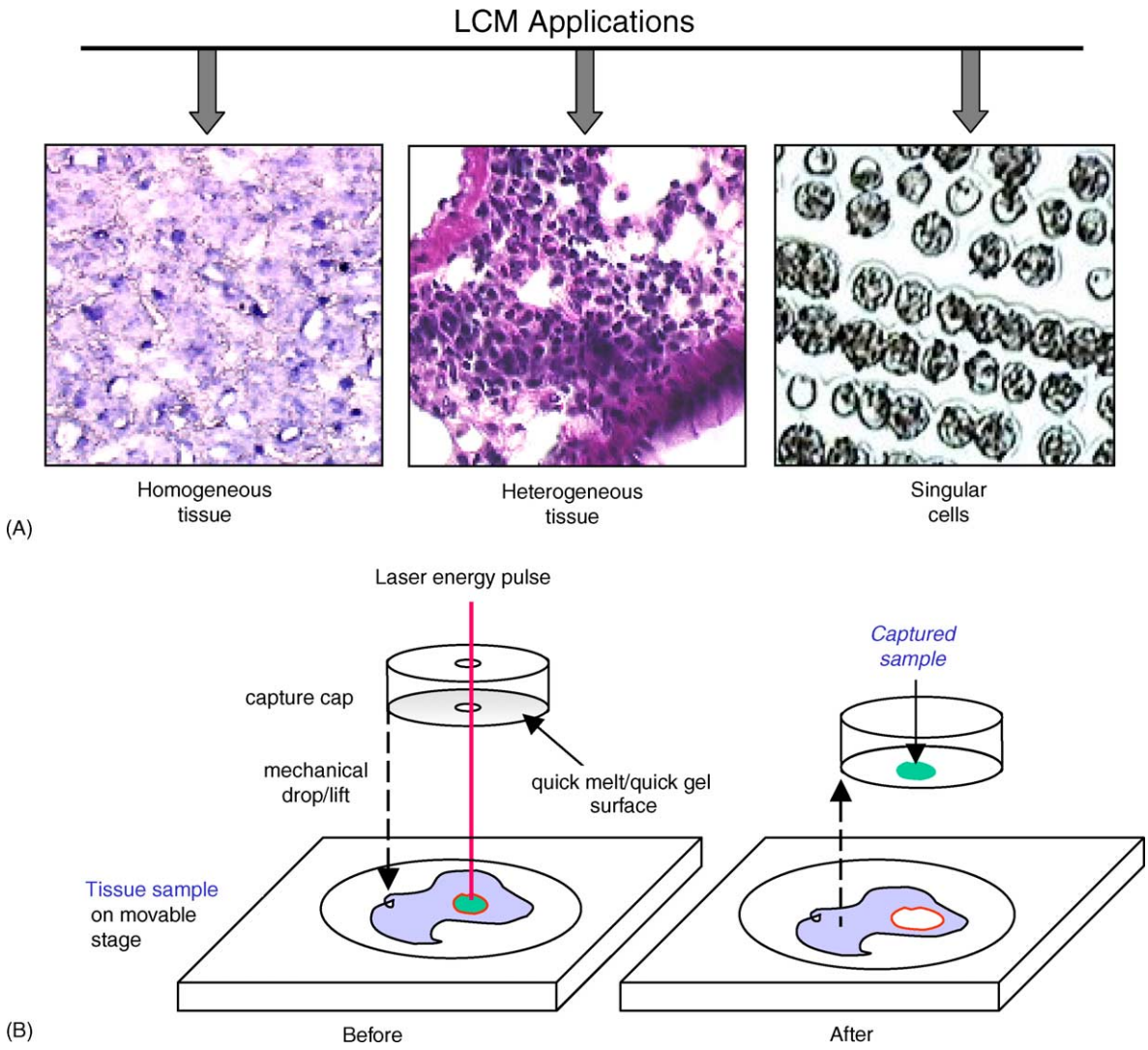

Fig. 2. Isolation of cells from by laser capture microdissection. LCM applications can be applied to a variety of tissue types (A). The principles of LCM cell isolation (B).

we are studying, even though we may only be sampling one of these compartments.

We applied the LCM at preparative scale in order to isolate sufficient quantities of airway epithelial cells for analysis. The practical aspect here is the "time" that the experiment is allowed to proceed in order to harvest sufficient quantities of cells for analysis. We have found that a sampling cycle of 20–30 min in duration can optimally isolate high quality cell preparations in terms of absolute yield and protein integrity. The tissue itself is stable under these conditions and typically not restricted in terms of cell collapse or disintegration. These are also cycle times that allow the LCM operator to be alert and less affected by physical tedium. A representative example of the preparative isolation achieved in the study are shown in Fig. 4 in which either small airways (A) or large airways (B) were isolated. These images show the intact tissue section before processing, and the captured epithelial cells "dissected" away from the remnant tissue. We use this visual validation as a control point before committing the sample for further analyses. However, as a practical experimental aspect, we found that if the sampling time gets extended, the operator will not be able to maintain high quality cell isolation more than 30 min within a session.

Can microscale isolation and protein analysis be readily attained in standard laboratory settings? Yes, we believe it can. We began by asking the question: What number of cells do we actually excise and isolate from the tissue by LCM? In Fig. 5 are shown representative micrographs of the airway epithelium used in our studies here. We performed quantitative image analysis and cell counting to estimate the actual number of cells isolated by the LCM. The green bar represents 100 μm of linear distance along the basement membrane below the epithelial cells lining the airway lumen. By comparing identical panels and fields using multi-chromatic fluorescent staining of cell nuclei, cytoplasm, and matrix components within the sub-mucosal compartment staining, we identified

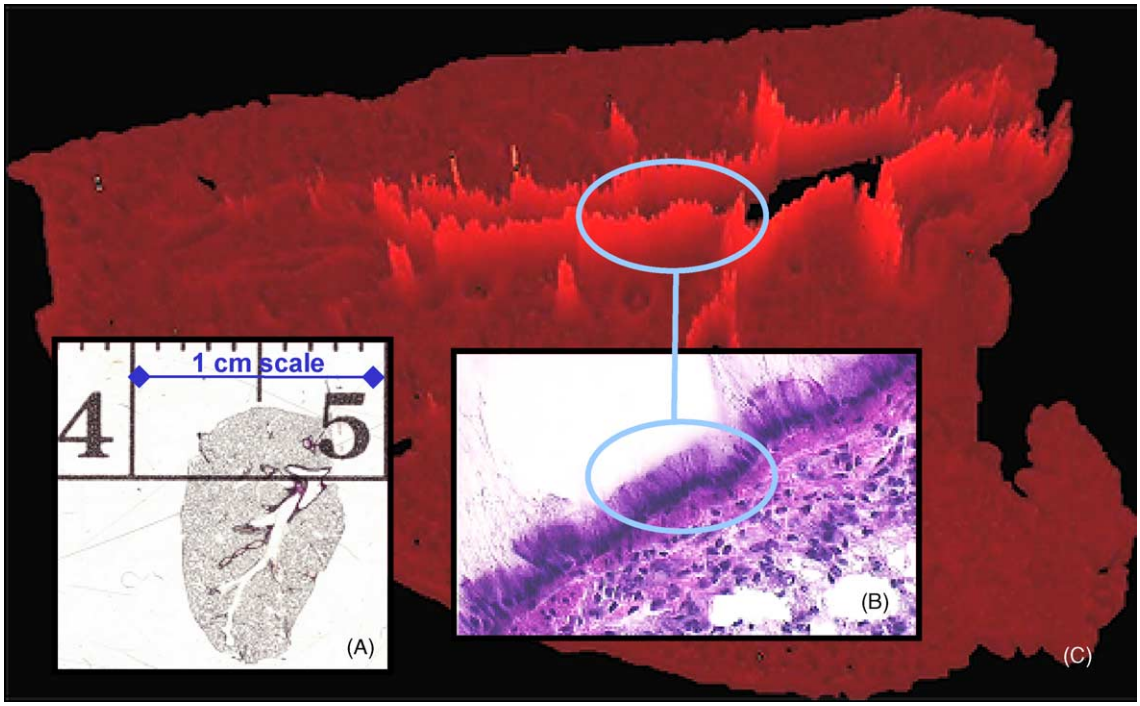

Fig. 3. Histological characterization of tissue used in protein expression studies. A cross-section of lung in scale size (A) provides overall context and identifies compartments, such as airways and bronchioles for focused cell isolations of epithelial cells by LCM (B). Seen in (C) is a three-dimensional pictogram of the entire epithelial surface of the conducting airways used in preparative scale cell isolations.

singular cells on a large set of tissues and lungs (50 tissue sections from six lungs). The dots and crosses indicate individual cells identified in cross-section within the plane of thickness of the section. We concluded from these exercises that we isolated approximately 18 000–22 000 cells/LCM cap on the preparative scale captures. The relative index of cellular density (cellularity) varied between small and large bronchioles but was fairly homogeneous within individual airways and throughout the thickness of the sections sampled. There is not a universal formula for calculating cell numbers recovered per LCM shot due to the many differences in cell sizes, distributions in tissue, and foci of abundance.

We compared the patterns of total expression between whole lung (global) and LCM captured airway epithelial cells using the 2D-PAGE, followed by robotic spot excision, elution, reduction and alkylation and enzymatic digestion prior to MS–MS for protein identification. In Fig. 6 are shown protein expression maps from three 20 μm sections of whole

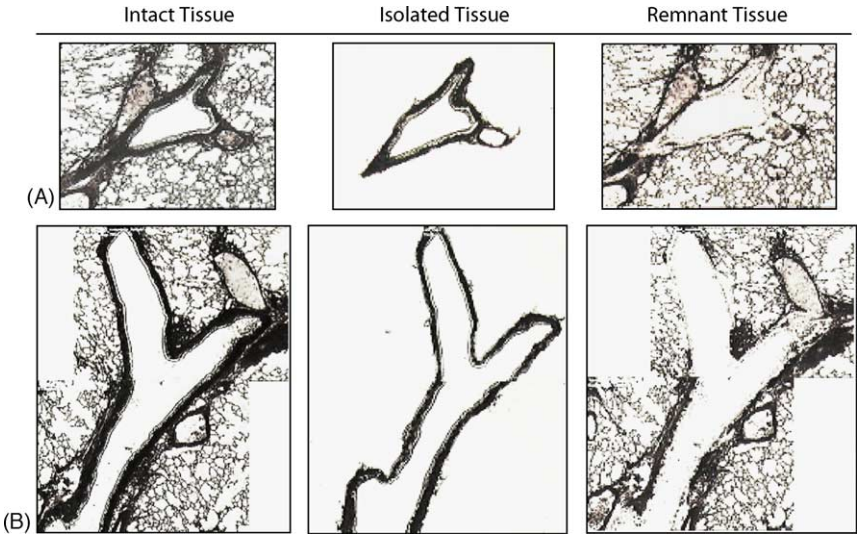

Fig. 4. Preparative scale isolation of airway epithelium isolation by laser capture microdissection (LCM), showing intact tissue, isolated epithelium and remnants after LCM isolation.

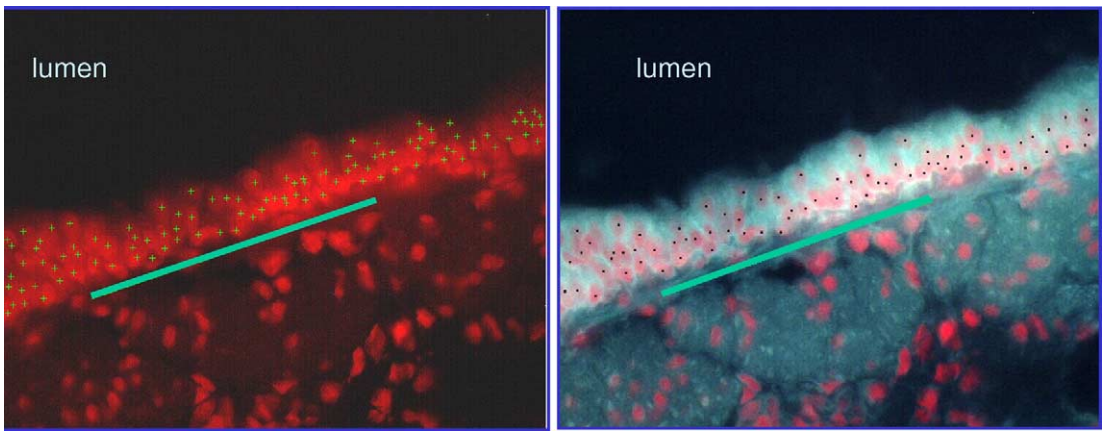

Fig. 5. Quantitative analysis of cell numbers in airway compartments. Fluorescent light micrographs of identical fields of cells using multi-chromatic staining. Plus and dot signs symbolize individual cell counts. The scale bar indicates 100 μm linear distance.

lung (A), and from 7500 LCM shots of isolated epithelial cells from mice sensitised and challenged with OVA allergen (B). The airway epithelial cells are a major cellular component of the whole lung tissue and as such there were significant overlaps between the two samples in terms of their spot profiles, while still showing distinct overall profiles, in terms of presence and absence of spots, and overall abundance. We were able to annotate 1400 protein spots with analytical quality. High resolution is obtained from both high-molecular-mass (120 000), as well as low-molecular-mass (8000) proteins, as well as good separations within the basic and acidic regions. Protein identifications were performed by peptide mass fingerprinting using MALDI-TOF-MS analysis as well as follow up electrospray ionization MS–MS sequencing. The resulting peptide map, and sequence tables were run towards data base acquisition for identification. In Table 1 is shown a list of the spots annotated in Fig. 6 gels and their identifications. The protein identities were established by matching the peptide masses identified by MS–MS against the Mascot database of protein/peptide identities [19]. The annotated proteins in Table 1 showed identities with high significance scores that matched individual peptides (range 7–22 peptides/protein) and sequence coverage (range 22–70%). As an example, in Table 2 is the annotation identity for spot 1.E10 Annexin I, which shows the identities of the 21 peptide mass identities observed by MS–MS as well as their position on

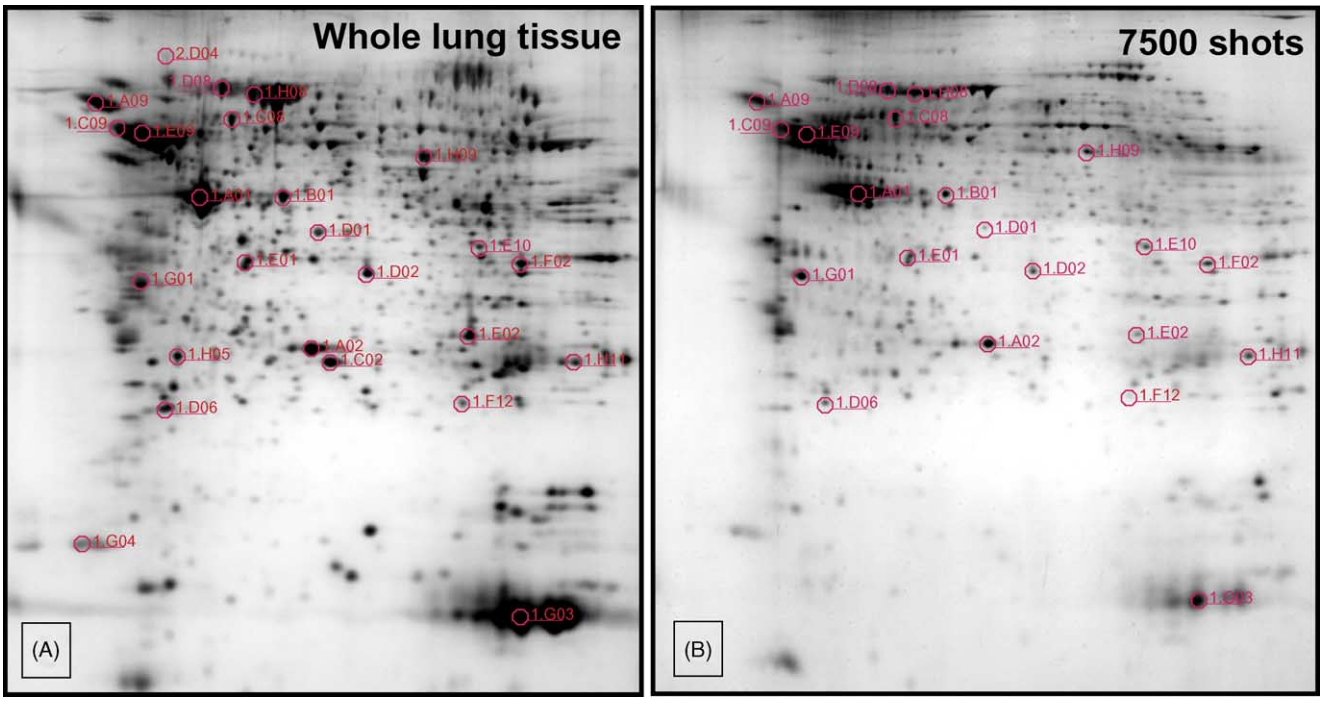

Fig. 6. Annotated 2D gel protein expression maps of whole lung tissue and LCM isolated epithelium. Annotations of proteins identified by MS–MS correspond to identities listed in Table 1.

Table 1
Selected annotation identities of proteins identified in lung by mass spectrophotometry

Spot identity	Protein name	M_r (calculated)	Peptides matched[a]	Mascot score[b]
1.A01	Actin, β (fragment)	41 432	15	171
1.A02	Thioether *S*-methyltransferase (EC 2.1.1.96)	30 068	10	113
1.A09	Contrapsin precursor	47 021	19	233
1.B01	M_r 45 000 secretory protein	46 453	16	155
1.C02	Antioxidant protein 2 (1-*cis*-peroxiredoxin) (acidic calcium-independent phospholipase A)	24 838	14	181
1.C08	M_r 60 000 heat shock protein (M_r 60 000 chaperonin)	61 088	12	158
1.C09	α-1-Antitrypsin 1–3 precursor (serine protease inhibitor 1–3)	45 996	12	122
1.D01	Novel mammalian lectin Ym12	42 637	10	93
1.D06	Peroxiredoxin 2 (thioredoxin peroxidase I)	21 936	7	110
1.D08	Heat shock cognate (M_r 71 000 protein)	70 055	13	123
1.E01	Annexin III (lipocortin III) (placental anticoagulant protein III)	36 520	13	161
1.E02	Carbonic anhydrase II (EC 4.2.1.1) (carbonate dehydratase II)	29 056	11	127
1.E10	Annexin I (lipocortin I) (calpactin II) (p35) (phospholipase A2 inhibitory protein)	38 604	21	325
1.F02	Annexin II (lipocortin II) (calpactin I heavy chain)	38 806	7	199
1.F09	β-Tubulin	50 255	16	151
1.F12	Peroxiredoxin I (thioredoxin peroxidase 2)	22 390	10	113
1.G01	Annexin V (lipocortin V) (endonexin II)	35 787	19	279
1.G03	Hemoglobin β-1 chain (B1)	15 813	9	124
1.G04	Myosin light chain alkali (smooth muscle myosin α)	16 959	8	154
1.H05	RHO GDP-dissociation inhibitor I (RHO GDI 1) (RHO-GDI ALPHA)	23 450	12	128
1.H08	Serum albumin precursor	70 700	22	126
1.H09	Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3)	57 015	15	167
1.H11	Glutathione *S*-transferase YC (EC 2.5.1.18)	25 740	19	267
2.D02	Malate dehydrogenase, cytoplasmic (EC 1.1.1.37)	36 494	10	111
2.D04	Heat shock protein 84—mouse	83 571	13	113

[a] Number of peptide mass values matched with postulated sequences derived from trypsin cleavage of identified protein.
[b] Probability based mouse score. Score is $-10\log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 66 are significant ($p < 0.05$).

the primary sequence of the protein. In this case, 58% of the sequence of the total protein could be identified.

LCM allows microscale protein expression analyses. Overall, the cells isolated by LCM were easily analysed using standard solubilisation protocols and the 2D gel separation platform. The great utility of the LCM method is the ability to selectively address specific cell populations and relate patterns of protein expression to precise biological micro-environments.

It is difficult to directly compare quantitatively the results of protein expression profiling studies from singular reports of LCM application. There are only few systems which have the precision for determining the actual cell counts of tissue sections cut as thin slices (5–20 μm thick). This is more confounded in experimental models which sample heterogeneous cell populations in proportional distributions within tissue. For these types of studies, a scale of reference is required in order to provide comparative data. In a recent report, we introduced one such comparative index, "spots per shots" [25]. We found that similar to other microscale studies utilizing 2D gel separation that an optimal amount of LCM sample was required for best separating individual proteins. In Fig. 7 is shown an example of the changes in protein profiles seen between samples prepared from 8300 and 12 400 shots of liver tissue, respectively. The changes in abundance levels of individual spots either decreased (in red) or increased (in green) at higher relative loading levels. A finer detail of one gel region is detailed in the box marked in blue. The differences in expression maps due to sample loading are of such significance that we recommend that some standardization report be included in studies demonstrating the utility of microscale measurements.

There are numerous technical considerations that need to be addressed in order to succeed in studying quantitative protein expression in clinical samples, such as the choice of tissue and the sampling strategy [23]. Both biology and technology are the drivers for choosing the overall experimental strategy. Our experience has demanded that that we closely link the expression studies to clinical phenotypes, and to measurements such as histology and functional pathophysiology, in which clinical disease states and end-points can be characterized. Our experience concludes, from working with many different tissues originating from various organs, that the levels of cellular heterogeneity/homogeneity, as well as the nature of micro-environment within various histological compartments being sampled, will determine the strategy being applied for successful protein isolation and identification (Fig. 8). The qualitative protein identity (ID) produced

Table 2
Example of peptide mass values and sequence coverage

Peptide	Start	End	Sequence	M_r (observed)	M_r (calculated)
1	12	25	FLENQEQEYVQAVK	1724.85	1723.84
2	29	52	GGPGSAVSPYPSFNVSSDVAALHK	2344.14	2343.15
3	58	70	GVDEATIIDILTK	1387.76	1386.76
4	81	96	AAYLQENGKPLDEVLR	1815.97	1814.95
5	81	97	AAYLQENGKPLDEVLRK	1944.05	1943.05
6	98	112	ALTGHLEEVVLAMLK oxidation	1639.90	1638.90
7	113	123	TPAQFDADELR	1262.61	1261.59
8	128	143	GLGTDEDTLIEILTTR	1746.92	1745.90
9	154	160	VYREELK	936.52	935.51
10	166	176	DITSDTSGDFR	1213.53	1212.53
11	166	177	DITSDTSGDFRK	1341.64	1340.62
12	188	203	CQDLSVNQDLADTDAR	1820.81	1819.80
13	204	211	ALYEAGER	908.45	907.44
14	213	227	KGTDVNVFTTILTSR	1651.91	1650.89
15	214	227	GTDVNVFTTILTSR	1523.81	1522.80
16	228	233	SFPHLR	756.41	755.41
17	234	241	RVFQNYGK	1011.54	1010.53
18	235	241	VFQNYGK	855.42	854.43
19	242	249	YSQHDMNK	1022.43	1021.43
20	242	249	YSQHDMNK oxidation	1038.43	1037.42
21	269	280	CATSTPAFFAEK	1329.62	1328.61

Residue	Primary sequence[a]
1	AMVSEFLKQAR**FLENQEQEYVQAVK**SYK**GGPGSAVSPYPSFNVSSDVAAL**
51	**HK**AIMVK**GVDEATIIDILTK**RTNAQRQQIK**AAYLQENGKPLDEVLRKALT**
101	**GHLEEVVLAMLKTPAQFDADELR**GAMK**GLGTDEDTLIEILTTR**SNEQIRE
151	INR**VYREELK**RDLAK**DITSDTSGDFRK**ALLALAKGDR**CQDLSVNQDLADT**
201	**DARALYEAGERRKGTDVNVFTTILTSRSFPHLRRVFQNYGKYSQHDMNK**A
251	LDLELKGDIEKCLTTIVK**CATSTPAFFAEK**LYEAMKGAGTRHKALIRIMV
301	SRSEIDMNEIKVFYQKKYGISLCQAILDETKGDYEKILVALCGGN

Example: Annexin I (lipocortin I) (calpactin II). Nominal mass (M_r): 38 864. Match to: SWISSNEW:ANX1_MOUSE; Score: 325. Number of mass values matched: 21.

[a] Observed peptide sequence coverage: 58% (underlined in bold).

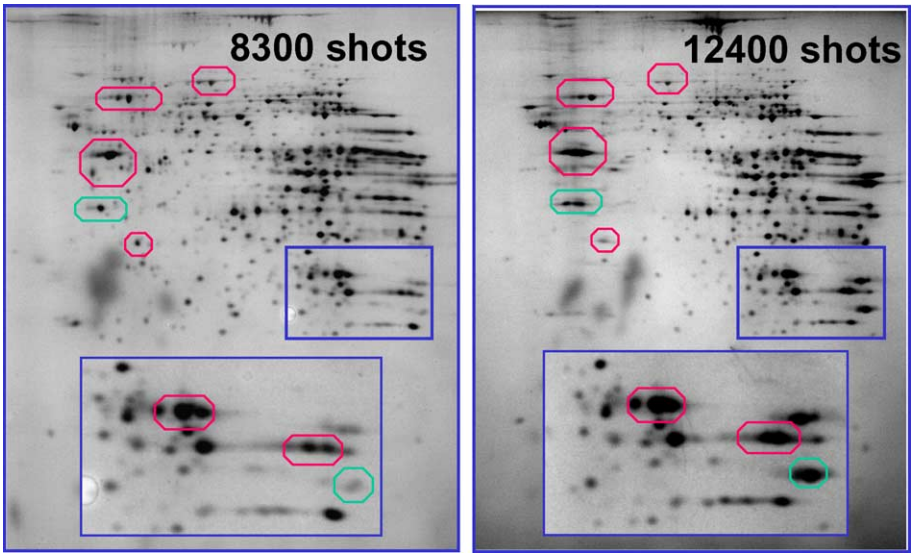

Fig. 7. Differences in protein expression maps between 8300 and 12 400 LCM shots of hepatocytes isolated from whole liver tissue.

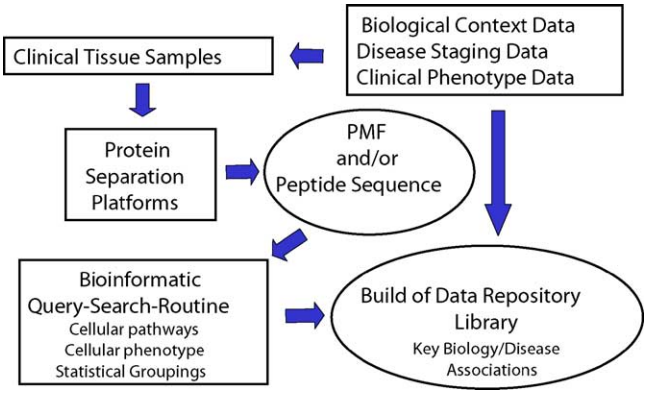

Fig. 8. Bioinformatics systems approach to relate protein expression phenotype to biological context, cellular pathways, and clinical parameters.

from scoring assures statistically that peptide sequences can be matched correctly to proteins. The list(s) of annotation identities are not the end goals of the study. These lists are the beginning of the next phase of study which analyses the relative differences in expression between individuals in experimental or clinical groupings. All data from each protein expression study are captured and stored in data repository libraries. The important linkage of proteins to pathways, pathways to cells, and cells to foci of disease requires a critical and important investment of resources for data acquisition, storage, and analyses. Even when empowered with concerted bioinformatics support these linkage analyses are difficult to interpret without strong clinical phenotype data. As such these libraries will hold additional information that has been generated from well-designed clinical studies—such as measurements of function (mechanical, physiological), imaging (CT, MRI, etc.), disease progression, and treatment history. A big effort for the future will be to build predictive biological scenarios for hypothesis testing based upon the relationships of protein expression to specific medical "factors" unique to each disease presentation.

4. Conclusions

Today, microscale protein expression profiling is easily approached using standard laboratory methods. Advances in LCM technology provide a highly attractive option for isolating specific cell types from within the micro-environment of tissue. Intact proteins can be isolated and identified from LCM isolated tissue samples. The possibility for studying protein expression levels associated with cells active in disease processes promises to deliver a significant amount of information with important potential clinical impact.

In order to match the delivery of protein expression indexes with a utility of usefulness, we need to overcome a number of technical and statistical limitations [26,27]. The important coupling of a sampling strategy with appropriate technology cannot be underestimated. The first step is to build a biological hypotheses which can be tested with the available sample. The biological questions being addressed will greatly influence the approach chosen to answer these questions. The next goal should be to establish a basis for sampling which represents the distribution of the biological question within a biological context, albeit disease evolvement or normal steady state homeostasis. This includes establishing a statistical basis for choosing one type of sample over another, or one site of sampling in time versus another [26]. Once one is satisfied that the sample is representative of the biological question, then the method of analysis requires validation, both qualitatively and quantitatively. The challenge is to generate validity in the technology as such, and to try to keep the biological window at a minimum of variance. Even though we are capable of isolating key target cells from the biological sites which most closely are associated with our study questions, we have much to learn yet about their true active contributions to disease evolvement.

The variety of resident, circulating, and/or inflammatory cells present in tissue, provides yet another level of need, for interpretive skills for relating protein expression patterns to cell phenotypes, and further to histological indications of disease evolvement. Without solid histology reporting, there is no easy way to relate one tissue bit to another. Let alone to differences between healthy and disease states.

The critical number of data points which arise from surveys of protein expression, even on the microscale, are staggering. It is not uncommon to have hundreds of thousands to millions of data points for single studies. For example: individual data points of charge, mass, time of flight, for average of 15 peptides/protein $\times$ 500 proteins $\times$ 100 samples = 2.25 million data points! Often repeated samplings over time within individual experiments are required in order to reach a level of statistical significance for baseline determinations of variance. There is a pressing need for technology developments in data analysis to be made in standardizing and de-complexing this part of the process while ensuring high quality analytical measurements. However, providing the right bioinformatic link that allows accurate predictions of clinical progression based upon the data generated in these proteomic surveys remains a clear goal for the future.

References

[1] S.W. Taylor, E. Fahy, B. Zhang, G.M. Glenn, et al., Nat. Biotechnol. 3 (2003) 281.
[2] V. Espina, K.A. Dettloff, S. Cowherd, E.F. Petricoin III, L.A. Liotta, Expert Opin. Biol. Ther. 1 (2004) 83.
[3] J.G. Paez, P.A. Janne, J.C. Lee, S. Tracy, Science 304 (2004) 1497.
[4] R.F. Bonner, M. Emmet-Buch, K. Cole, T. Pohida, R. Chuaqui, S. Goldstein, L.A. Liotta, Science 21 (1997) 1482.
[5] F. Fend, M. Emmet-Buch, R. Chuaqui, K. Cole, J. Lee, L.A. Liotta, M. Raffeld, Am. J. Pathol. 154 (1999) 61.
[6] M.D. Rekhter, J. Chen, Cell Biochem. Biophys. 35 (2001) 103.
[7] L.C. Lawrie, S. Curran, H.L. McLeod, J.E. Fothergill, G.I. Murray, J. Clin. Pathol. Mol. 54 (2001) 253.
[8] R. Banks, M. Dunn, M.A. Forbes, D. Pappin, T. Naven, M. Gough, P. Harnden, P.J. Selby, Electrophoresis 20 (1999) 689.

[9] M. Brown Jones, H. Krutzsch, H. Shu, Y. Zhao, L.A. Liotta, E.C. Kohn, E.F. Peticoin, Proteomics 2 (2002) 76.
[10] T.A. Zhukov, R.A. Johanson, A.B. Cantor, R.A. Clark, M.S. Tockman, Lung Cancer 40 (2003) 267.
[11] A. Wellmann, V. Wollscheid, H. Lu, Z.L. Ma, P. Albers, K. Schutze, V. Rohde, P. Behrens, S. Dreschers, Y. Ko, N. Wernert, Int. J. Mol. Med. 9 (2002) 341.
[12] L.H. Cazares, B.L. Adam, M.D. Ward, S. Nasim, P.F. Schellhammer, O.J. Semmes, G.L. Wright Jr., Clin. Cancer Res. 8 (2002) 2541.
[13] D.E. Palmer-Toy, D.A. Sarracino, D. Sgroi, R. LeVangie, P.E. Leopold, Clin. Chem. 46 (2000) 1513.
[14] B.J. Xu, R.M. Caprioli, M.E. Sanders, R.A. Jensen, J. Am. Soc. Mass Spectrom. 13 (2002) 1292.
[15] G. Zhou, H. Li, D. DeCamp, S. Chen, H. Shu, Y. Gong, M. Flaig, J.W. Gillespie, N. Hu, P.R. Taylor, M.R. Emmert-Buck, L.A. Liotta, E.F. Petricoin III, Y. Zhao, Mol. Cell. Proteomics 1 (2002) 117.
[16] J.D. Wulfkuhle, D.C. Sgroi, H. Krutzsch, K. McLean, K. McGarvey, M. Knowlton, S. Chen, H. Shu, A. Sahin, R. Kurek, D. Wallwiener, M.J. Merino, E.F. Petricoin III, Y. Zhao, P.S. Steeg, Cancer Res. 62 (2002) 6740.
[17] L. Zang, D. Palmer-Toy, W.S. Hancock, D.S. Sgroi, B.L. Karger, J. Proteome, Res. 3 (2004) 604.
[18] D.I. Blyth, M.S. Pedrick, T.J. Savage, E.M. Hessel, D. Fattah, Am. J. Respir. Cell Mol. Biol. 14 (1996) 425.
[19] T.E. Fehniger, J.G. Sato-Folatre, J. Malmstrom, et al., J. Proteome Res. 2 (2004) 307.
[20] A. Shevchenko, M. Wilm, o. Vorm, M. Mann, Anal. Chem. 68 (1996) 850.
[21] G. Westergren-Thorson, J. Malmström, G. Marko-Varga, Electrophoresis 24 (2003) 276.
[22] C. Bratt, C. Lindberg, C. Lindberg, G. Marko-Varga, J. Chromatogr. A 909 (2001) 279.
[23] S.T. Holgate, P. Lackie, S. Wilson, W. Roche, D. Davies, Am. J. Respir. Crit. Care Med. 3 (2000) S113.
[24] E.R. Weibel, C.R. Taylor, Fishman's Pulmonary Diseases and Disorders, third ed., McGraw-Hill, New York, 1998, p. 233.
[25] G. Marko-Varga, M. Berglund, J. Malmstrom, H. Lindberg, T.E. Fehniger, Electrophoresis 21 (2003) 3800.
[26] G. Marko-Varga, T.E. Fehniger, J. Proteome Res. 2 (2004) 167.
[27] G.L.G. Miklos, R. Maleszka, Proteomics 1 (2001) 169.