Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties

Martin J. Page*†, Bob Amess*, R. Reid Townsend*, Raj Parekh*, Athula Herath*, Luc Brusten*, Marketa J. Zvelebil‡, Robert C. Stein§, Michael D. Waterfield‡, Susan C. Davies¶, and Michael J. O'Hare¶

*Oxford GlycoSciences, 10 The Quadrant, Abingdon Science Park, Oxfordshire, OX14 3YS, United Kingdom; Ludwig Institute for Cancer Research, Departments of ‡Biochemistry and §Oncology, 91 Riding House Street, London W1P 8BT, United Kingdom; Department of Biochemistry and Molecular Biology, University College, London W1P, United Kingdom and ¶Ludwig Institute for Cancer Research/University College London Breast Cancer Laboratory, Department of Surgery, 67-73 Riding House Street, London W1P 7LD, United Kingdom

Communicated by Lloyd J. Old, Ludwig Institute for Cancer Research, New York, NY, July 16, 1999 (received for review March 31, 1999)

Normal human luminal and myoepithelial breast cells separately purified from a set of 10 reduction mammoplasties by using a double antibody magnetic affinity cell sorting and Dynabead immunomagnetic technique were used in two-dimensional gel proteome studies. A total of 43,302 proteins were detected across the 20 samples, and a master image for each cell type comprising a total of 1,738 unique proteins was derived. Differential analysis identified 170 proteins that were elevated 2-fold or more between the two breast cell types, and 51 of these were annotated by tandem mass spectrometry. Muscle-specific enzyme isoforms and contractile intermediate filaments including tropomyosin and smooth muscle (SM22) alpha protein were detected in the myoepithelial cells, and a large number of cytokeratin subclasses and isoforms characteristic of luminal cells were detected in this cell type. A further 134 nondifferentially regulated proteins were also annotated from the two breast cell types, making this the most extensive study to date of the protein expression map of the normal human breast and the basis for future studies of purified breast cancer cells.

Breast cancer continues to be a leading cause of death for women (1, 2), despite major advances in basic research over the last two decades. Response rates to therapy remain poor, with survival for women who present with metastatic disease typically only between 18 and 24 mo (3). Although many prognostic indicators have been studied (4, 5), none reliably predict response to treatment.

Prognostic uncertainty could be overcome, in part, by defining the changes that occur in a tumor at either the gene (genomic) or protein (proteomic) level. Genomics is now able to characterize gene mutations and transcript mRNAs in a high-throughput manner (6–8). Proteomics, which examines the protein repertoire of a sample and produces a unique protein expression map (PEM), including posttranslational modifications and subcellular localization, is now reaching a similar level of refinement (9–11). Each technology is capable of identifying in a single process many events at the gene or protein level that change during tumor formation and progression. Such tumor-associated changes may identify new markers and new targets for therapeutic intervention. Correlation of these new markers with drug response and ultimately patient survival should provide clinicians with new diagnostic and prognostic information that will benefit their patients.

A critical part of this process is the ability to define changes that have occurred between normal and tumor material. The human breast is a complex organ whose proliferation and differentiation are regulated by the interplay of growth factors, steroid hormones, and cell–cell interactions (12, 13). The terminal lobular-alveolar unit of the breast, which is the structure from which the majority of cancers arise, is composed of two types of epithelial cells. The inner or luminal epithelial cells which are potential milk secretory cells surrounded by an outer layer of contractile myoepithelial cells. Most breast carcinomas (95%) express phenotypic markers that are consistent with an origin from luminal cells (14).

Early attempts to resolve proteins from breast material by two-dimensional PAGE were hampered by lack of reproducibility, low sensitivity, use of biopsy material containing mixed cell populations, and an inability to obtain sequence information from proteins of interest (15–17). More recently, significant progress in key areas of proteomics (18, 19) has resulted in reports and databases with proteomes for normal and tumor-derived breast cell lines in culture (20–22) and breast tissue from benign and malignant sources (23, 24). However, these studies are limited by a number of factors: use of established cell lines that may be unrepresentative, use of normal cells of undefined phenotype, and heterogeneity of cell types where primary breast material has been studied.

Several advances now render feasible a systematic proteomic analysis of normal human luminal and myoepithelial breast cells. It is now possible, by using immunomagnetic methods (25, 26), to purify these cells from normal human breast material in sufficient quantities for proteomic characterization and in purities equivalent to those obtained by fluorescence-activated cell sorting (FACS) (27). The PEM analysis has involved high-resolution detection of protein features by using fluorescent dyes, image analysis algorithms that allow the construction of protein expression databases, high-throughput processing of peptide pools, and mass spectrometric analysis to obtain sequence data from femtomol levels of proteins. Using these systems, we describe here the first detailed PEM analysis of matched normal adult human luminal and myoepithelial breast cells. These studies demonstrate that cell type-specific proteins can be detected and quantitated by using this approach and provide the basis for future comparisons with the proteomes of purified breast tumor cells.

Materials and Methods

Preparation of Purified Cells. Tissues used in this study consisted of 10 samples of reduction mammoplasties obtained with consent from patients aged between 20 and 47 yr (mean age ≈ 25 yr). No malignancy or pathology other than minimal fibrocystic change (two cases) was detected. Purified populations of normal human breast luminal and myoepithelial cells were prepared as described by Clarke *et al.* (25), with important modifications to

Abbreviations: PEM, protein expression map; MCI, molecular cluster index; MACS, magnetic affinity cell sorting; pI, isoelectric point.

†To whom reprint requests should be addressed. E-mail: martin.page@ogs.co.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

enhance purity. Primary epithelial cultures were treated overnight with calcium-free medium, resulting in suspensions enriched in luminal cells. These were incubated for 40 min on ice with a mixture of rat monoclonal antibody (ICR-2, 10 μg/ml) against the luminal epithelial marker EMA and a mouse monoclonal antibody (Dako, clone SS2/36, 1:25) against the myoepithelial antigen CD-10. The cells were washed and labeled for 15 min with anti-rat magnetic affinity cell sorting (MACS) magnetic beads (Miltenyi Biotec, Auburn, CA) and positively labeled cells separated by using a Vario-MACS (Miltenyi Biotec) high field-intensity magnet. These were further incubated for 20 min with anti-mouse Dynabeads (Dynal, U.K.), which binds to any residual CD-10 positive myoepithelial cells, and separated by using an MPC-10 low-intensity magnet, which does not attract MACS-bead labeled luminal cells. Myoepithelial cells were purified by trypsinization and filtration (35 μm) of the luminally depleted primary cultures and incubation with a mixture of a mouse IgG2a anti CD-10 antibody [Harlan Serlab (Harlan Laboratories, Haslett, MI) clone 55; 1:50], and a mouse IgG1 monoclonal antibody (F-19, 10 μg/ml) reacting with the fibroblast activation protein (FAP; ref. 28). The labeled suspension of cells was then incubated with subclass-specific anti-mouse IgG2 MACS beads, separated by using Vario-MACS, and then incubated with anti-IgG1 mouse Dynabeads to remove any FAP-positive cells. This procedure enabled F-19 positive fibroblasts, which are the main potential contaminant of the myoepithelial preparation because of their ability to express CD-10 antigen in short-term culture (29), to be removed from the F-19 negative myoepithelial cells. Purified luminal and myoepithelial cells were obtained in yields of $5 \times 10^6 - 2 \times 10^7$ cells. Purity of the resulting cell preparations was assessed by staining for cell-type-specific filament proteins (cytokeratins and vimentin), as originally described by O'Hare *et al.* by using FACS-sorted cells (27), and by quadruple simultaneous immunofluorescence (data not shown). The purified cell populations were subsequently washed five times in serum-free medium, flash frozen, and stored at −80°C. Lysis was undertaken with a solution containing 4% wt/vol 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, 5 M urea, 2 M thiourea, 65 mM DTT, 0.8%wt/vol Resolytes 3–10 (Bio-Rad), and trace bromophenol blue, to a frozen cell pellet containing approximately 300 μg protein resulting in a final volume of 925 μl. This was vortexed, left to stand for 5 min, vortexed again, then centrifuged at 13,000 $\times$ *g* for 5 min at 15°C.

Two-Dimensional Gel Electrophoresis. Immobilized pH gradient (IPG) gels (Immobiline DryStrip 3–10 NL, Amersham Pharmacia Biotech) were rehydrated with 370 μl of solubilized sample and focused overnight (70 kVh, 20°C) according to Sanchez *et al.* (30). Immediately after being focused, IPG gels were equilibrated in 6 M urea/2% wt/vol SDS/2% wt/vol DTT/50 mM Tris, pH 6.8/30% vol/vol glycerol for 15 min before running in the second dimension on 9–16% T, 2.7% C gels, cast with the gel bound to one of the glass plates, in an electrophoresis tank similar to that described by Amess *et al.* (31) at 30 mA per gel and 20°C. Immediately after electrophoresis, gels were fixed in 40% vol/vol ethanol:10% vol/vol acetic acid and stained with the fluorescent dye OGT MP17 (molecule on the basis of ref. 32), and 16-bit monochrome fluorescence images at 200 μm resolution were obtained by scanning gels with an Apollo II linear fluorescence scanner (Oxford GlycoSciences, Oxfordshire, U.K.). Two gels were run for each sample.

Analysis of Gel Images. Primary images were processed with a custom version of MELANIE II (Bio-Rad). Individually resolved protein features were enumerated and quantified on the basis of fluorescence signal intensity. The pI and molecular weight of each feature were calculated by bilinear interpolation between landmark features on each image previously calibrated with respect to *Escherichia coli* proteins. Intensity was measured by summing pixels within each feature boundary and recorded as a percentage of the total feature intensity on the image. Replicate images were obtained for each sample, matched, and only those features detected in both replicates were retained. The index [pI, relative molecular weight (RMM), sample of origin], and percentage intensity data (mean of replicates) for each gel feature were entered into a database table. The resulting definition of protein features found in both replicates of a given sample is referred to as the PEM. For purposes of comparison between PEMs from different samples, it is necessary to assign to each individual protein a master index [referred to as molecular cluster index (MCI)], which establishes the correspondence (on the basis of pI and RMM) between equivalent proteins across the entire set of PEMs. This process is achieved by using software (MELANIE II) to match the images and is checked by human operators. The algorithms underlying this process have been previously described (33). Its purpose is to normalize individual PEMs into a single and coherent geometry, thereby allowing the precise construction of a protein expression database for all samples under investigation. Differential analysis was undertaken by using proprietary software (ROSETTA, Oxford GlycoSciences), which allows an investigator to vary the criteria by which differential expression is deemed significant, including abundance change, frequency of incidence, and statistical parameters. In this study ROSETTA was used to generate a binary proteograph listing MCIs present in 50% of either luminal or myoepithelial samples and having a positive or negative fold change with a magnitude of $\geqq 2$.

Protein Identification by Mass Spectrometry. Protein features of interest were excised from the gel by a software-driven robotic cutter, delivered into separate wells of a 96-well plate, and processed by a proteolysis workstation to yield tryptic peptide pools. A mass list of peptides from each protein was obtained by using a matrix-assisted laser desorption ionization (MALDI)—time-of-flight spectrometer (ELITE, PerSeptive Biosystems, Framingham, MA). Fragmentation spectra from 1-Da mass windows (obtained from the MALDI mass list) were recorded by using a nanospray ionization source (Z-spray) on a Q-TOF instrument (Micromass, Manchester, U.K.). The continuum fragmentation spectra were converted to centered spectra and used to search the SwissProt (version 36.0, October 1998) database with the SEQUEST computer program (34). Candidate sequences were confirmed when an ion series consistent with y-type fragmentation was observed for the complete peptide sequence (35).

Results

Derivation of PEMs. Normal breast material from 10 premenopausal women was processed by using double antibody labeling and differential magnetic purification procedures, yielding 10 separate purified populations of matched normal human luminal and myoepithelial breast cells (both of >95% purity). After lysis and two-dimensional PAGE, the separated proteins were detected by using a fluorescent dye that binds noncovalently to the SDS moiety attached to the proteins, enabling features to be visualized by a fluorescence scanner at a detection level of less than 1 ng protein. From the 10 matched duplicate luminal and myoepithelial proteomes (accounting for the presence of 43,302 total features), 1,738 unique MCIs were identified.

Identification of Differentially Expressed Proteins. The primary objective of this study was to identify those proteins uniquely or differentially expressed between the two breast cell types. This was achieved by using the ROSETTA software that analyzes each MCI individually in a qualitative and quantitative manner

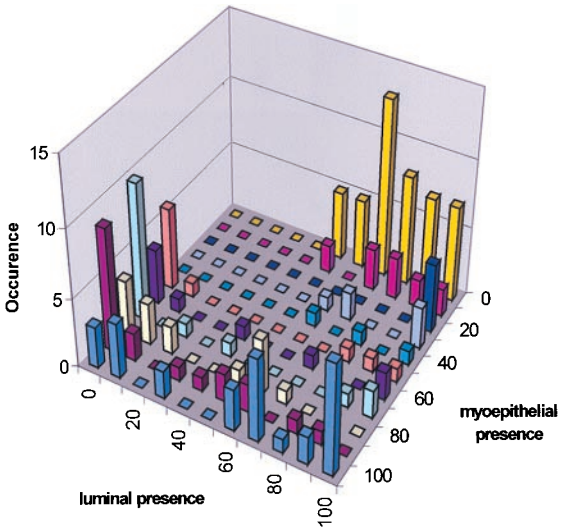

Fig. 1. Chart showing the frequency of each differentially expressed feature (in 50% or more of either cell type) across the set of 10 pairs of luminal and myoepithelial samples.

between the sets of luminal or myoepithelial PEMs. The output of this process is referred to as a proteograph. Because each of the 10 samples originated from different human donors, some natural biological variation across the proteomes was expected. The proteograph was, therefore, programmed to accept only those MCIs that were present in 50% or more of either data set (10 PEMs) and that differed by ≥2-fold. Every feature identified in the proteograph as a differential MCI was confirmed by manually checking across each complete set of PEMS. In total, 170 features met these criteria, and the frequency of their incidence across either set of 10 luminal or myoepithelial breast samples is shown in Fig. 1. This process identifies those features that are expressed either uniquely (such as cytokeratin 19, cytokeratin 18, and annexin II in luminal cells only), or rarely in each cell type, and other classes of features that are present in both cell types, but are expressed to different levels. Thus each column in Fig. 1 illustrates the frequency with which that protein is present in a particular set of samples. Such analysis, when broadened, should serve as the basis for identifying interesting subpopulations within large sample sets. Of the 170 differential features identified, it was noticeable that their abundance varied across two log orders. Of these, a series of features was chosen for analysis by tandem mass spectrometry and the annotations for 51 differentially expressed proteins (33 up-regulated in luminal cells and 18 up-regulated in myoepithelial cells) are shown in Table 1 and as MCI assignments on the individual master PEMs in Fig. 2. It is of note that some of the spectra could not be correlated with sequences in the public domain databases, suggesting the identification of novel proteins, and that some annotations reported herein are on low abundance proteins (e.g., β isoform of phosphatidylinositol transfer protein and fructose bisphosphate aldolase c), which had a percent volume presence of less than 0.05% relative to the total amount of protein detected.

Myoepithelial cells have a phenotype intermediate between epithelial cells and smooth-muscle cells. Among the 18 annotations in the myoepithelial up-regulated set, there are two muscle-specific isotypes of the enzymes fructose bisphosphate and pyruvate kinase M1. Tropomyosin was also significantly elevated in these cells, together with smooth muscle (SM22) alpha protein, a marker previously believed to be specific to adult smooth muscle (36). Myoepithelial cells were also shown to contain higher levels of actin and cytokeratin 14 than luminal cells, which is a distinguishing characteristic of these cells, as seen immunohistochemically (14).

The most striking characteristic of the luminal epithelial cell-specific proteins was the range of cytokeratin expression, including the differential expression of cytokeratins 8, 18, and 19, which are used as immunohistochemical markers of this breast cell type (14) and are the cytokeratins commonly expressed in breast carcinoma cells (14, 37). Many of the cytokeratins and some other proteins had multiple MCI values because of the presence of different isoforms of the same protein, with different electrophoretic and migration properties. This observation is probably accountable by processing and by posttranslational events that would be overlooked by genomic studies.

Creatine kinase (mitochondrial) and a particular isoform of annexin II were also expressed at higher levels in the luminal cells. This mitochondrial isoform of creatine kinase has been reported to be elevated in the serum of breast cancer patients and correlates with a significantly higher mortality rate (38). Annexins are believed to play key roles in breast cell proliferation and differentiation (39). Annexin II has been linked with multidrug resistance (40) and tumor–endothelial interactions (41) and is known to be differentially regulated between normal and malignant human breast epithelial cells (42).

Identification of Nondifferentially Expressed Proteins. Although there is an important set of differentially expressed proteins between the two breast cell types, it is nevertheless noticeable that their respective PEMS are very similar. To further advance the protein annotation of these cells, we have identified an additional 134 proteins by using tandem mass spectrometry [see supplemental Table 2 (see www.pnas.org)]. By proteograph analysis, none of these met the criteria as differential features between the two cell types, yet many of these proteins will have key roles in fundamental breast biology. These include structural, metabolic, signal transduction, processing, and nuclear proteins. By inference in the pairing and matching of the MCIs across the two cell types, an identical MCI from either the luminal or myoepithelial cell would represent the same protein. In this study, we note that there were 16 examples where an identical MCI was chosen for annotation from parallel proteomes of myoepithelial and luminal gels, and in each case the same identity was established. Moreover, no examples were found where the same MCI yielded different annotations in the two proteomes. All 185 proteins identified in this study are depicted with their MCIs on the respective master luminal or myoepithelial cell PEMs in Fig. 2.

Discussion

Although protein expression analysis can be performed on tissues and mixed cell populations, cellular heterogeneity severely limits interpretation and presents a significant obstacle in clinical studies that attempt to compare PEMs from disease with normal tissues. In this study, improved immunomagnetic separation techniques were used to produce large populations of highly purified luminal and myoepithelial breast cells from primary clinical samples as a resource for detailed proteomic analysis.

A protein expression database with a common index of MCIs was generated from 10 duplicate matched sample sets of luminal and myoepithelial cells. These two major breast cell types showed significant homology at the global protein level, which supports the proposal that they may be derived from a common stem cell (43). However, 170 proteins were identified that were expressed significantly differently between the two cell types, and mass spectrometry annotation established the identity of 51 of these differential proteins.

The myoepithelial cells contained several enzymes and structural proteins of muscle-specific origin, which were expressed at

Table 1. Annotation of 51 features that are elevated by 2-fold or more between the two breast cell types

MCI	fc	Bgnd FP	Fgnd FP	Bgnd CV	Fgnd CV	P-Val (T)	P-Val (RS)	Accession	Annotation
5252	−44.4	10	0	28.6				P08727	Keratin, type I cytoskeletal 19 (cytokeratin 19)
5372	−25.0	10	0	30.8				P05783	Keratin, type I cytoskeletal 18 (cytokeratin 18)
5164	−19.0	10	0	35.4				P08779	Keratin, type I cytoskeletal 17 (cytokeratin 17)
4614	−17.3	8	0	41.8				P02538	Keratin, type II cytoskeletal 6a (cytokeratin 6a)
4602	−17.3	10	0	44.3				P07355	Annexin II (lipocortin II)
4620	−15.6	6	0	59.4				P13647	Keratin, type II cytoskeletal 5 (cytokeratin 5)
5253	−15.0	9	0	40.8				Q04695	Keratin, type I cytoskeletal 17 (cytokeratin 17)
5280	−13.8	9	0	49.5				Q04695	Keratin, type I cytoskeletal 17 (cytokeratin 17)
5298	−13.7	10	0	50.5				P08727	Keratin, type I cytoskeletal 19 (cytokeratin 19)
5407	−13.6	8	0	13.4				P08727	Keratin, type I cytoskeletal 19 (cytokeratin 19)
4644	−13.4	5	0	53.1				P02538	Keratin, type II cytoskeletal 6a (cytokeratin 6a)
4654	−13.3	8	5	35.3	41.5	0.0001	0.0043	P02538	Keratin, type II cytoskeletal 6a (cytokeratin 6a)
5365	−12.9	8	0	42.9				Q04695	Keratin, type I cytoskeletal 17 (cytokeratin 17)
5215	−11.1	7	0	35.6				Q04695	Keratin, type I cytoskeletal 17 (cytokeratin 17)
4899	−10.7	9	0	46.9				P12532	Creatine kinase, ubiquitous mitochondrial precursor (EC 2.7.3.2)
5416	−10.1	8	0	11.6				O15509	Arp2/3 complex 20 kDA subunit
5243	−9.6	9	0	23.8				P08727	Keratin, type I cytoskeletal 19 (cytokeratin 19)
5251	−9.2	7	0	34.1				P08727	Keratin, type I cytoskeletal 19 (cytokeratin 19)
5351	−9.1	9	0	47.9				Q04695	Keratin, type I cytoskeletal 17 (cytokeratin 17)
5275	−7.4	7	0	43.1				P08727	Keratin, type I cytoskeletal 19 (cytokeratin 19)
5323	−7.3	10	1	46.1				Q04695	Keratin, type I cytoskeletal 17 (cytokeratin 17)
4938	−6.5	7	0	45.9				P09972	Fructose-bisphosphate aldolase c (EC 4.1.2.13) (brain)
4762	−6.3	7	0	38.4				P13645	Keratin, type I cytoskeletal 10 (cytokeratin 10)
5060	−5.6	9	0	45.1				P48739	Phosphatidylinositol transfer protein β isoform
5221	−4.2	7	0	49.8				P13645	Keratin, type I cytoskeletal 10 (cytokeratin 10)
4635	−3.6	10	10	51.4	23.7	0.0016	0.0024	P05787	Keratin, type II cytoskeletal 8 (cytokeratin 8)
4700	−3.1	9	7	97.1	61.1	0.0749	0.0127	P05787	Keratin, type II cytoskeletal 8 (cytokeratin 8)
4624	−2.7	10	10	36.9	31.6	0.0004	0.0007	P08729	Keratin, type II cytoskeletal 7 (cytokeratin 7)
5006	−2.4	10	3	22.0	82.7	0.0948	0.0509	P52907	F-actin capping protein α-1 subunit
4706	−2.3	10	7	50.6	37.6	0.0065	0.0046	P02533	Keratin, type I cytoskeletal 14 (cytokeratin 14)
5058	−2.0	10	4	55.1	14.8	0.0199	0.0180	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1
5129	−2.0	8	4	22.5	23.3	0.0011	0.0161	P07339	Cathepsin D precursor (EC 3.4.23.5)
4860	−2.0	7	1	33.7				Q92524	26S protease regulatory subunit S10B (proteasome subunit p42)
5764	28.7	0	7		40.8			P04075	Fructose-bisphosphate aldolase A (EC 4.1.2.13) (muscle)
5809	23.7	0	5		21.3			P07226	Tropomyosin, fibroblast nonmuscle type
5781	22.0	0	10		49.5			P02570	Actin, cytoplasmic 1
5937	19.6	0	8		51.2			P37802	Sm22-alpha homolog
5960	16.8	0	9		48.8			P09455	Retinol-binding protein I, cellular
5636	15.4	0	6		64.3			P14618	Pyruvate kinase, M1 isozyme (EC 2.7.1.40)
5889	14.7	0	9		57.6			Q06830	Thioredoxin peroxidase 2
5821	13.3	0	8		40.3			O15144	Arp2/3 complex 34 kDA subunit
5879	13.3	0	9		33.9			P04792	Heat-shock 27-kDA protein
5688	10.4	0	9		35.5			P19338	Nucleolin
5600	8.1	0	9		37.1			P07900	Heat-shock protein hsp 90-alpha
4886	3.9	2	8	30.7	67.4			P02533	Keratin, type I cytoskeletal 14 (cytokeratin 14)
5645	3.4	1	9		120.2			P50991	T-complex protein 1, Δ subunit
4461	3.1	10	10	32.9	36.2	0.0002	0.0006	P11142	Heat-shock cognate 71-kDA protein
4989	2.8	6	8	25.4	28.4	0.0002	0.0023	P52895	Probable *trans*-1,2-dihydrobenzene-1,2-diol dehydrogenase (EC 1.3.1.20)
5001	2.8	9	9	32.1	39.5	0.0010	0.0023	P02570	Actin, cytoplasmic 1
5785	2.3	3	10	30.3	29.4	0.0003	0.0329	P52907	F-actin capping protein α subunit
5820	2.0	1	9		44.4			P07339	Cathepsin D precursor (EC 3.4.23.5)

The fc value represents the fold elevation of that feature relative to the background value. In each case, the negative and positive values indicate specific elevation in the luminal or myoepithelial cells, respectively. Fgnd and Bgnd FP, foreground and background feature presence; CV, coefficient of variation; P-Val, statistical score by T-test (T) or Wilcoxon rank sum (RS) test; accession, SwissProt number.

considerably higher levels than in luminal cells. Extensive characterization of the cytokeratin profiles of each cell type revealed multiple isoforms and conformed with the known immunocytochemical characteristics of these cells (14). The roles of breast cytokeratins encompass both structural and signaling capabilities. For instance, cytokeratin 8 is known to be the major receptor for plasminogen on breast cells (44), and many of the cytokeratins are significantly down-regulated in breast tumors (45). Interestingly, several workers have now found an inverse correlation between breast cytokeratin mRNA levels and the corresponding protein levels (46, 47).

Previous reports of protein expression analysis of human breast-derived material have included breast cancer cell lines, in which extensive heterogeneity was seen (20). A limited study by using normal human breast cells obtained from human milk or human mammary epithelial (HMEC) cells (Clonetics, San Di-

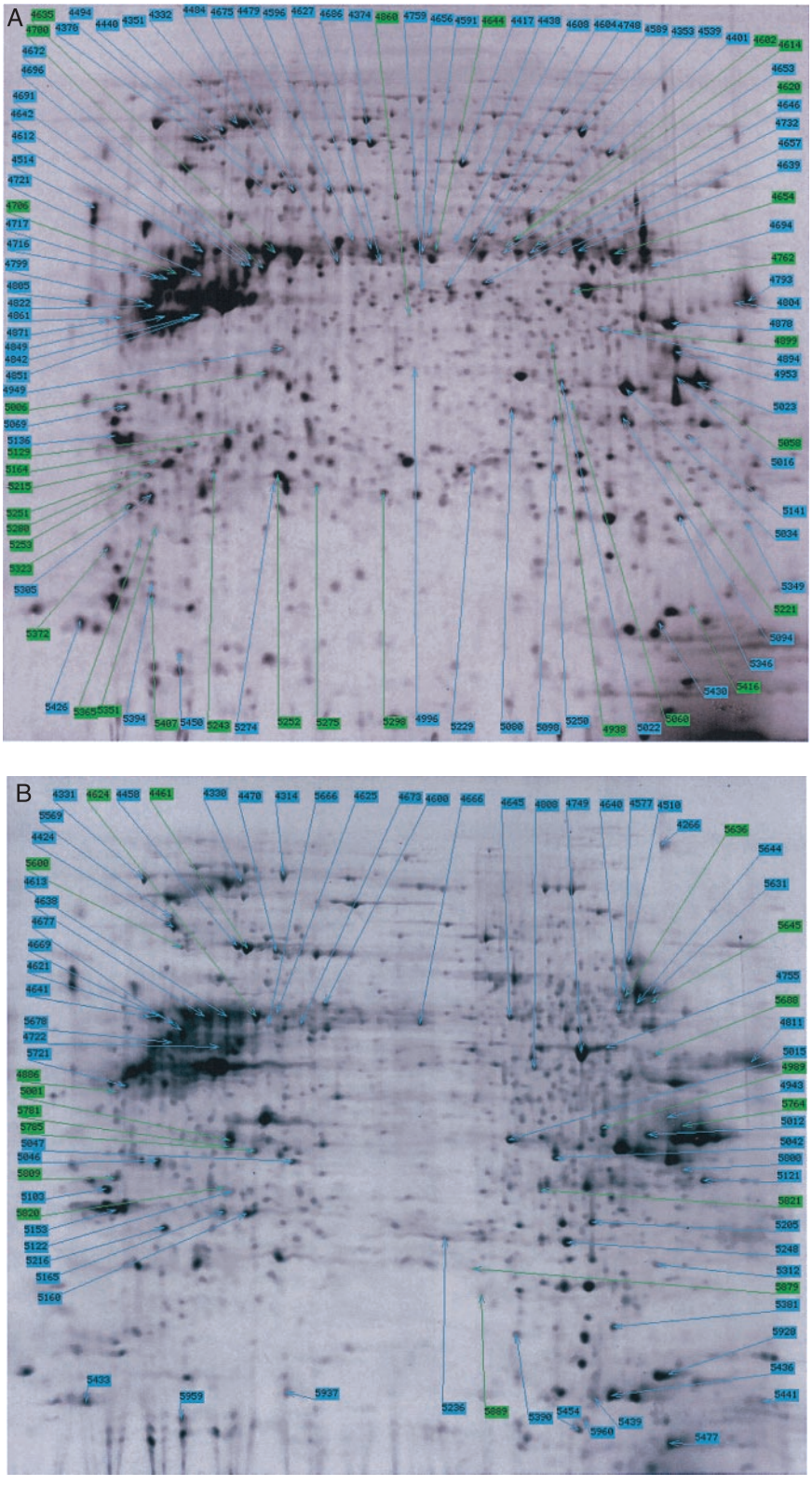

Fig. 2. The master PEMs for the luminal (*A*) and myoepithelial (*B*) cells, complete with 185 annotations. Those annotations in a green box are differentially elevated in that cell type, whereas those in a blue box are expressed at similar levels in both cells. All annotations are provided in Table 1 and supplemental Table 2 (see www.pnas.org).

ego) prepared from reduction mammoplasty has also been carried out (48). However, epithelial cells derived from human milk contain both luminal and myoepithelial cell types (M.O'H., unpublished observations), and comparison of the cytokeratin profiles of fluorescence-activated cell sorted or MACS-sorted myoepithelial cells (27, 25) shows that HMEC cells are predom-

inantly of this phenotype. In total, there are less than 60 proteins annotated on existing tumor and normal breast PEMs (20–22, 48, 49), some of which have been identified by inference.

In this report, we describe the most extensive study to date, to our knowledge, of the protein expression in breast epithelial cells utilizing highly purified populations of human breast luminal and myoepithelial cells. We have annotated 185 proteins by mass spectrometry, of which 51 are differentially expressed. On some occasions during the annotation process, we were unable to find matches in the public domain databases, implying that this approach also has the capability of identifying novel proteins. These observations demonstrate that proteomics has the refinement and sensitivity to find proteins that are either uniquely or differentially expressed between different cell types, the consequences of which could enable new strategies for drug discovery. While we continue to increase our database with additional normal breast samples, we have begun a parallel study by using immunomagnetically purified cells obtained from primary and metastatic breast tumors. From these studies, we intend to identify the global divergence between normal and tumor breast cells by using proteomics and thereby to develop new approaches to breast cancer treatment and monitoring.

We thank Prof. Munro Neville for helpful support and advice during this project. In addition, we also thank the surgeons at Queen Mary's Hospital, Roehampton, U.K., for access to the human tissues, Dr. S. Lakhani (University College London, Department of Histopathology) for pathological examination and Dr. E. Stockert (Ludwig Institute for Cancer Research, New York, NY) for the F19 antibody. M.J.P. dedicates this paper to the memory of Margarete Majer.

1. Landis, S. H., Murray, T., Bolden, S. & Wingo, P. A. (1998) *CA Cancer J. Clin.* **48,** 6–29.
2. *Cancer Research Campaign* (London, U.K.) *Factsheet* 6.1 (1996).
3. Leonard, R. C., Rodger, A. & Dixon, J. M. (1994) *BMJ* **309,** 1501–1504.
4. Ravaioli, A., Bagli, L., Zucchini, A. & Monti, F. (1998) *Cell Prolif.* **31,** 113–126.
5. Ferno, M. (1998) *Anticancer Res.* **18,** 2167–2171.
6. Crooke, S. T. (1998) *Nat. Biotechnol.* **16,** 29–30.
7. Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E. & Davis, R. W. (1998) *Trends Biotechnol.* **16,** 301–306.
8. Ramsay, G. (1998) *Nat. Biotechnol.* **16,** 40–44.
9. Anderson, N. L. & Anderson, N. G. (1998) *Electrophoresis* **19,** 1853–1861.
10. Page, M. J., Amess, B., Rohlff, C., Stubberfield, C. & Parekh, R. (1999) *Drug Discovery Today* **4,** 55–62.
11. Wang, J. H. & Hewick, R. M. (1999) *Drug Discovery Today* **4,** 129–133.
12. Dickson, R. B., Johnson, M. D., el-Ashry, D., Shi, Y. E., Bano, M., Zugmaier, G., Ziff, B., Lippman, M. E. & Chrysogelos, S. (1993) *Adv. Exp. Med. Biol.* **330,** 119–141.
13. Rudland, P. S., Barraclough, R., Fernig, D. G. & Smith, J. A. (1998) *Biochem. Soc. Symp.* **63,** 1–20.
14. Taylor-Papadimitriou, J. & Lane, E. B. (1987), *The Mammary Gland: Development, Regulation and Function*, eds. Neville, M. C. and Daniel, C. W. (Plenum, New York), pp.181–215.
15. Stastny, J., Prasad, R. & Fosslien, E. (1984) *Clin. Chem.* **30,** 1914–1918.
16. Wirth, P. J., Egilsson, V., Gudnason, V., Ingvarsson, S. & Thorgeirsson, S. S. (1987) *Breast Cancer Res. Treat.* **10,** 177–189.
17. Wirth, P. J. (1989) *Electrophoresis* **10,** 543–554.
18. Figeys, D., Gygi, S. P., Zhang, Y., Watts, J., Gu, M. & Aebersold, R. (1998) *Electrophoresis* **19,** 1811–1818.
19. Corthals, G. L., Molloy, M. P., Herbert, B. R., Williams, K. L. & Gooley, A. A. (1997) *Electrophoresis* **18,** 317–323.
20. Giometti, C. S., Williams, K. & Tollaksen (1997) *Electrophoresis* **18,** 573–581.
21. Williams, K., Chubb, C., Huberman, E. & Giometti, C. S. (1998) *Electrophoresis* **19,** 333–343.
22. Rasmussen, R. K., Ji, H., Eddes, J. S., Zugaro, L. M., Reid, G. E., Simpson, R. J. & Dorow, D. S. (1998) *Electrophoresis* **19,** 818–825.
23. Bini, L., Magi, B., Marzocchi, B., Arcuri, F., Tripodi, S., Cintorino, M., Sanchez, J. C., Frutiger, S., Hughes, G., Pallini, V., *et al.* (1997) *Electrophoresis* **18,** 2832–2841.
24. Franzen, B., Linder, S., Alaiya, A. A., Eriksson, E., Fujioka, K., Bergman, A.-C., Jornvall, H. & Auer, G. (1997) *Electrophoresis* **18,** 582–587.
25. Clarke, C., Titley, J., Davies, S. & O'Hare, M. J. (1994) *Epithelial Cell Biol.* **3,** 38–46.
26. Gomm, J. J., Browne, P. J., Coope, R. C., Liu, Q. Y., Buluwela, L. & Coombes, R. C. (1995) *Analytical Biochem.* **226,** 91–99.
27. O'Hare, M. J., Ormerod, M. G., Monaghan, P., Lane E. B. & Gusterson, B. A. (1991) *Differentiation* **46,** 209–221.
28. Rettig, WJ., Garin-Chesa, P., Healey, J. H., Su, S. L., Ozer, H. L., Schwab, M., Albino, A. P. & Old, L. J. (1993) *Cancer Res.* **53,** 3327–3335.
29. Atherton, A. J., O'Hare, M. J., Buluwela, L., Titley, J., Monaghan, P., Paterson, H. F., Warburton, M. J. & Gusterson, B. A. (1994) *J. Cell Sci.* **107,** 2931–2939.
30. Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R. & Hochstrasser, D. F. (1997) *Electrophoresis* **18,** 324–327.
31. Amess, B. & Tolkovsky, A. M. (1995) *Electrophoresis* **16,** 1255–1267.
32. Hassner, A., Birnbaum, D. & Loew, L. M. (1984) *J. Org. Chem.* **49,** 2546–2551.
33. Wilkins, M. R., Hochstrasser, D. F., Sanchez, J. C., Bairoch, A. & Appel, R. D. (1996) *Trends Biochem. Sci.* **21,** 496–497.
34. Eng, J. K., McCormack, A. L. & Yates, J. R. (1994) *J. Am. Soc. Mass Spectrom.* **3,** 976–999.
35. Biemann, K. (1990) *Methods Enzymol.* **193,** 455–479.
36. Solway, J., Seltzer, J., Samaha, F. F., Kim, S., Alger, L. E., Niu, Q., Morrisey, E. E., Ip, H. S. & Parmacek, M. S. (1995) *J. Biol. Chem.* **270,** 13460–13469.
37. Bocker, W., Bier, B., Freytag, G., Brommelkamp., B., Jarasch, E. D., Edel, G., Dockhorn-Dworniczak, B. & Schmid, K. W. (1992) *Virchows Arch. A Pathol. Anat. Histopathol.* **421,** 315–322.
38. Pratt, R., Vallis, L. M., Lim, C. W. & Chisnall, W. N. (1987) *Pathology* **19,** 162–165.
39. Ahn, S-H., Sawada, H., Ro, J-Y. & Nicolson, G. L. (1997) *Clin. Exp. Metastasis* **15,** 151–156.
40. Cole, S. P., Pinkoski, M. J., Bhardwaj, G. & Deeley, R. G. (1992) *Br. J. Cancer* **65,** 498–502.
41. Tressler, R. J., Updyke, T. V., Yeatman, T. & Nicolson, G. L. (1993) *J. Cell. Biochem.* **53,** 265–276.
42. Schwartz-Albiez, R., Koretz, K., Moller, P. & Wirl, G. (1993) *Differentiation* **52,** 229–237.
43. Stingl, J., Eaves, C. J., Kuusk, U. & Emerman, J. T. (1998) *Differentiation* **63,** 201–213.
44. Hembrough, T. A., Li, L. & Gonias, S. L. (1996) *J. Biol. Chem.* **271,** 25684–25691.
45. Franzen, B., Linder, S., Alaiya, A. A., Eriksson, E., Uruy, K., Hirano, T., Okuzawa, K. & Auer, G. (1996) *Br. J. Cancer* **74,** 1632–1638.
46. Su, L., Morgan, P. R. & Lane, E. B. (1996) *Hum. Pathol.* **27,** 800–806.
47. Brotherick, I., Robson, C. N., Browell, D. A., Shenfine, J., White, M. D., Cunliffe, W. J., Shenton, B. K., Egan, M., Webb, L. A., Lunt, L. G., *et al.* (1998) *Cytometry* **32,** 301–308.
48. Giometti, C. S., Tollaksen, S. L., Chubb, C., Williams, C. & Huberman, E. (1995) *Electrophoresis* **16,** 1215–1224.
49. Rasmussen, R. K., Ji, H., Eddes, J. S., Moritz, R. L., Reid, G. E., Simpson, R. J. & Dorrow, D. S. (1997) *Electrophoresis* **18,** 588–598.