

RESEARCH ARTICLE

Functional analysis of the nuclear proteome of human A549 alveolar epithelial cells by HPLC-high resolution 2-D gel electrophoresis

Jeffery Forbus¹, Heidi Spratt^{2, 3, 4}, John Wiktorowicz^{2, 3}, Zheng Wu^{2, 3}, Istvan Boldogh⁵, Larry Denner¹, Alexander Kurosky^{2, 3}, Robert C. Brasier^{2, 5}, Bruce Luxon^{2, 3, 4} and Allan R. Brasier^{1, 6}

¹ Departments of Medicine, The University of Texas Medical Branch, Galveston, TX, USA

² Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, TX, USA

³ NHLBI Proteomics Center, The University of Texas Medical Branch, Galveston, TX, USA

⁴ Bioinformatics Program, The University of Texas Medical Branch, Galveston, TX, USA

⁵ Department of Microbiology and Immunology, The University of Texas Medical Branch, Galveston, TX, USA

⁶ Sealy Center for Molecular Science, The University of Texas Medical Branch, Galveston, TX, USA

The airway epithelial cell plays a central role in coordinating airway inflammatory responses, where significant changes in the proteome occur in response to infectious stimuli. To further understand the spectrum of proteins and the biological processes they control, we have initially determined the nuclear proteome of human type II-like alveolar epithelial cells (A549) using a sequential method of organellar enrichment followed by HPLC prefractionation prior to 2-DE-based protein identification using MALDI-TOF MS. This approach yielded 719 high-confidence identifications, 433 mapping to unique gene identifiers. Expert classification showed that these proteins controlled chromatin remodeling, protein refolding, cytoskeletal structure, membrane function, metabolic processes, mitochondrial function, RNA binding, protein synthesis, signaling, and transcription factor activities. The proteins were mapped to gene ontology classifications, where metabolism and catalytic activity functions were significantly enriched, representing 43 and 32% of the protein set, respectively. Pathways analysis indicated a protein network affecting tumor necrosis factor-nuclear factor- κ B signaling pathway interacting with intermediate cytoskeletal filaments. Forty-five proteins of unknown function were subjected to domain analysis and inferred to have additional nuclear functions controlling purine nucleotide metabolism and protein-protein interactions. This database represents the most comprehensive data set of mammalian nuclear proteins and will serve as a foundation for further discovery.

Received: September 8, 2005

Revised: November 30, 2005

Accepted: December 17, 2005



Keywords:

Airway epithelial cell / 2-D gel electrophoresis / Cell nucleus / Protein networks

Correspondence: Dr. Allan R. Brasier, Division of Endocrinology, MRB 8.138, The University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-1060, USA

E-mail: arbrasie@utmb.edu

Fax: +1-409-772-8709

Abbreviations: DDX3, DEAD box RNA helicase; GO, gene ontology; IPA, Ingenuity pathway analysis; KRT, keratin; NCBI, national center for biotechnology; NF- κ B, nuclear factor- κ B; TNF, tumor necrosis factor; TPR, tetratricopeptide repeat; Trx, thioredoxin reductase

1 Introduction

The human airway is one of the largest uninterrupted epithelial surfaces in humans, covering over 150 m² in surface area that lines the nasopharynx to the terminal alveoli [1]. This mucosal surface is exposed to over 10 000 L of air daily, and is therefore highly specialized to protect against insensible fluid loss while allowing gas exchange, mediate particulate clearance, and protecting against the introduction of noxious agents, chemicals, or microorganisms. To accom-

plish these diverse functions, epithelial cells are highly differentiated to perform specialized functions depending on their particular location in the respiratory tract (reviewed in [2]). Type I and -II pneumocytes are those lining the gas-exchanging alveoli; although the type I pneumocyte is numerically the most abundant [1], the type II alveolar epithelial cell appears to play the major functional, regenerative, and inflammatory role in lung physiology. For example, type II cells secrete pulmonary surfactant, a phospholipid important for reducing alveolar surface tension and maintaining alveolar patency, and also inducibly produce antimicrobial peptides, cytokines, chemokines, and other metabolites that initiate the complex processes of vascular permeability, cellular recruitment, mucous hyper-secretion, and airway remodeling that characterize airway inflammation [3–5]. Understanding the coordinated phenotypic responses of the type II alveolar cell, therefore, will be important to the pathogenesis of many types of human respiratory disease characterized by inflammation, such as viral induced airway inflammation.

The eukaryotic nucleus is a highly organized organelle containing specific functional domains essential for normal cellular homeostasis and mediating genomic response to environmental manipulations. Specifically because the nucleus is essential for initiating expression of reparative and inflammatory responses, we previously applied focused organellar proteomics to characterize differential nuclear protein expression in uninfected and respiratory syncytial virus (RSV)-infected alveolar A549 cells. We identified 24 differentially expressed high-abundance proteins by PMF using MALDI-TOF MS, and validated them by LC-tandem MS/MS and immunofluorescence [6]. These studies identified proteins important in the nuclear cytoskeleton, chromatin remodeling, oxidant/antioxidant response, and heat shock response. Immunofluorescence studies further showed that cytoplasmic RSV replication induces a nuclear heat shock response and redistribution of nuclear dot ten proteins into the cytoplasm. Although these studies provided important new information about the biology of the airway epithelial cells, the proteins identified were highly abundant.

To further extend the depth of proteome interrogation, we adapted a recently described method of HPLC pre-fractionation prior to 2-DE [7, 8]. Using this method, we observed that HPLC pre-fractionation increased the number of proteins that can be identified in a complex mixture by ~10 times more than could be seen in the original starting material [7]. We therefore sought to further increase proteome coverage by combining subcellular organellar fractionation and HPLC pre-fractionation prior to 2-DE to determine a comprehensive database of nuclear proteins and determine their functional activities. Here we report our identification of 719 proteins, 433 unique, with a high degree of statistical confidence. These proteins were analyzed by mapping to gene ontology (GO) biological process and molecular functions, and pathway analysis against a

curated database of protein interactions. To our knowledge, this represents the largest experimentally defined nuclear proteome data set in mammalian cells.

2 Materials and methods

2.1 Cell culture

Human A549 pulmonary type II epithelial cells (American Type Culture Collection) were grown in F-12K Nutrient Mixture (Kaighn's modification) with 10% fetal bovine serum, 5% L-glutamine with penicillin (100 U/mL), and streptomycin (100 µg/mL) at 37°C in a 5% CO₂ incubator.

2.2 Preparation of nuclear extracts

A549 cells were removed from the culture plate by trypsinization, washed in cold PBS twice, then transferred to hypotonic buffer A (50 mM HEPES, pH 7.9, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.1 µg/mL of PMSF, 1 µg/mL of pepstatin A, 1 µg/mL leupeptin, 10 µg/mL soybean trypsin inhibitor, 10 µg/mL aprotinin, and 0.1% IGEPAL CA-630) and incubated on ice for 10 min [9]. The lysates were centrifuged at 6000 rpm (4000 × g) for 30 s at 4°C. The pellet (containing the nuclei) was resuspended in Buffer B (buffer A containing 1.0 M sucrose) and centrifuged at 12 000 rpm (15 000 × g) for 10 min at 4°C [9]. The pellet was resuspended in Reagent 2 (catalog no. 163-2103; BioRad)/TBP (catalog no. 163-2101; BioRad) and vortexed for 5 min at room temperature. Nucleic acid was removed by adding 300 U/mL of endonuclease (Sigma Aldrich), if needed briefly sonicated, and then incubated for 30 min at room temperature. After incubation the sample was centrifuged at 13 000 rpm for 30 min at 4°C. The supernatant was desalted using a MicroSpin™ G-25 Column (Amersham Biosciences). The extract was then normalized for protein amounts determined by protein assay (BioRad).

2.3 RP-HPLC

LC was performed with the Agilent 1100 series system using a Vydac C4 RP column (250 mm × 10 mm, 10–15 µm, 300 Å) connected to a guard column. Before injection the sample was centrifuged at 14 000 rpm for 10 min. Approximately 5 mg of protein was loaded on the column, washed in Buffer A (0.1% TFA in H₂O) with 5% buffer B (0.1% TFA in ACN), and eluted by a linear gradient from 5% elution buffer B to 55% buffer B within 80 min. The protein elution profile was monitored at 214 nm. The flow rate was set at 2 mL/min with automatic collection of 2 min *per* tube. A small fraction of each collection was dried using a SpeedVac concentrator and then assayed to determine the protein concentration within each fraction. Upon concentration determination the fractions were dried, washed using water, and then dissolved in DeStreak™ Rehydration buffer (Amersham Biosciences) for 2-DE.

2.4 2-DE

IEF was performed with 11-cm-long precast IPG strips (pH 5–8; BioRad) [6]. Not to exceed 200 µg, 200 µL of 1-mg/mL protein samples were loaded onto an IPG strip and rehydrated overnight. IEF was performed at 20°C with the parameters: 50 V for 11 h, 250 V for 1 h, 500 V for 1 h, 1000 V for 1 h, 8000 V for 2 h, and 8000 for 48 000 Vh. Upon completion of IEF the IPG strips were stored at –80°C until 2-D SDS-PAGE was performed. To equilibrate the IPG strips for 2-D SDS-PAGE the strips were incubated in 4 mL of equilibration buffer (6 M urea, 2% SDS, 50 mM Tris-HCl, pH 8.8, 20% glycerol) containing 10 µL/mL of tri-2(2-carboxyethyl)phosphine (Geno Technology, St. Louis, MO) for 15 min at 22°C with shaking. Afterward the samples were then incubated in another 4 mL of equilibration buffer with 25 mg/mL of iodoacetamide/mL for 15 min at 22°C with shaking. Upon equilibration completion electrophoresis was performed at 150 V for 2.25 h at 4°C with precast 8–16% polyacrylamide gels in Tris-glycine buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% SDS, pH 8.3). Upon completion of 2-DE, the gels were fixed in fix buffer (10% methanol, 7% acetic acid in double distilled H₂O), then stained with SYPRO-Ruby (BioRad), and destained in fix buffer.

2.5 Gel imaging

The destained gels were scanned at a 100-µm resolution using the Perkin-Elmer (Boston, Mass.) ProXPRESS Proteomic Imaging System with 480-nm excitation and 620-nm emission filters. The exposure time was adjusted to achieve a value of 55 000- to 63 000-pixel intensity on the most intense protein spots on the gel. The 2-D gel images were subsequently analyzed using Progenesis Discovery software version 2003.03 (Nonlinear Dynamics, Ltd., Newcastle Upon Tyne, UK).

2.6 Protein identification

Protein gel spots were excised and prepared for MALDI-TOF-MS analysis using Genomic Solutions' ProPic and ProPrep robotic instruments following the manufacturer (Ann Arbor, MI). Briefly, gel pieces were incubated with trypsin (20 mg/mL in 25 mM ammonium bicarbonate, pH 8.0; Promega) at 37°C for 4 h. MALDI-TOF-MS was performed using an Applied Biosystems Voyager model DE STR for PMF [6]. Tryptic digests of some gel spots selected for validation were subsequently subjected to analysis by LC nanospray TOF MS using a Micromass Q-TOF II mass spectrometer. These identifications were Hsp70, keratin (KRT) 18, and 3'-5' exonuclease (Table 1 and [6]). LC was performed on a C18 New Objective PicoFrit column that allowed the eluate to be sprayed directly into the source of the mass spectrometer. MS/MS was performed in a data-

dependent mode and the data were processed using a Proteometrics Sonar MS/MS search engine (Genomic Solutions).

Mass spectral peaks were extracted from the raw data using Knexus automation client. A 3:1 S/N criteria was used and 3-point smoothing applied. Resulting peaks were deisotoped and the monoisotopic masses were used for database searching. Proteins were identified by searching the nonredundant National Center for Biotechnology (NCBI) database (including a finalized May 2004 human reference sequence based on NCBI Build 35) using the Profound search engine (version 2004.04.15). The search engine parameters included searching all entries of the taxonomy "primates", carbamidomethylation of cysteine (fixed), oxidation of methionine (variable), and one possible missed cleavage for trypsin. A 0.1 Da window for the masses was used. Spectra that matched human proteins using a matching algorithm were based on Bayesian probabilities [10]. Proteins matching the tryptic digestion pattern are assigned expectation scores; the expectation score is the number of matches expected if the matches were completely random. For example, an expectation score of 1 means that at least one match would be expected when searching a database that did not contain the actual protein sequence of interest. An expectation value of 1×10^{-3} means that the chance was 1 in 1000 that the match was random. Proteins with significant expectation scores were extracted and, for each, the calculated *pI*, percent protein coverage, and molecular weight were tabulated [6].

2.7 Western immunoblot analysis

For Western blots, nuclear proteins were fractionated on 10% SDS-PAGE and transferred to PVDF membranes (Millipore, Bedford, MA) [11]. Membranes were treated with 5% nonfat dried milk/Tris buffered saline (TBS, 100 mM Tris-HCl, 150 mM NaCl, pH 7.5) containing 0.1% vol/vol Tween for 1 h and then incubated with indicated antibody overnight at 4°C. Anti-β-tubulin, β-actin, p300, Pol II, and IκBα antibodies were from Santa Cruz Biotech (Santa Cruz, CA). Anti lamin B was from Calbiochem (San Diego, CA). Membranes were washed four times in TBS with 0.1% v/v Tween 20, and then incubated with HRP-conjugated secondary antibody. After washing, immune complexes were detected by reaction in the ECL assay (ECL, Amersham) according to the manufacturer.

2.8 Immunofluorescence microscopy

Smears of purified nuclei preparations on microscopic slides were stained for 15 min with 10 ng/mL 4'-6-diamidino-2-phenylindole dihydrochloride (excitation 345 nm, emission 455 nm; Molecular Probes), and/or with 50 ng/mL MitoTracker[®] Red (excitation 578 nm, emission 599 nm; Molecular Probes) for 5 min. Nuclei were washed and were then mounted in antifade medium (Dako Carpinteria, CA) on the

Table 1. Classification of unique proteins in A549 nuclear proteome. Shown are high-probability identifications from PMF using masses measured by MALDI-TOF in a Bayesian algorithm (ProFound). Identifications are filtered for expectation score of <0.05 , and grouped by functional activity. For each identification, the gene identifier (GeneID), common name (Name) percent coverage (Cov), expectation score (Expect), calculated p (p), and calculated molecular weight (MW, in kDa) are given. Expectation score is the number of matches expected if the matches were completely random. An expectation value of 1 means that at least one match would be expected when searching a database that did not contain the actual protein sequence of interest. An expectation value of 1×10^{-3} means that the chance was 1 in 1000 that the match was random. A complete list of all protein identifications is included in Suppl. Table I

GeneID	Common name	Expect.	% Cov.	p / p	MW, kDa
13279152	Polymerase I and transcript release factor	5.50E-07	8	6.6	22.96
3023628	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked	4.90E-05	18	6.7	73.62
4502389	Barrier to autointegration factor 1	2.30E-04	45	5.8	10.27
5031753	Heterogeneous nuclear ribonucleoprotein H1 (H)	6.10E-03	14	5.9	49.5
7661920	DEAD (Asp-Glu-Ala-Asp) box polypeptide 48	9.10E-03	15	6.3	47.14
3002951	Barrier to autointegration factor 1	0.015	26	6.1	15.9
26996739	mutS homolog 5 (<i>Escherichia coli</i>)	0.032	9	5.8	92.64
16507237	Heat shock 70 kDa protein 5 (glucose-regulated protein, 78 kDa)	1.80E-11	40	5.1	72.43
21040386	Heat shock 70 kDa protein 9B (mortalin-2)	2.20E-07	32	6	74.12
1633054	Cyclophilin	8.00E-06	48	8.1	18.09
2136253	TCP1 ring complex TRiC5	3.80E-05	16	6.2	60.88
24234688	Heat shock 70 kDa protein 9B (mortalin-2)	1.50E-05	21	5.9	73.95
4204880	Heat shock 70 kDa protein 2	1.90E-05	25	5.6	70.26
87528	Dnak-type molecular chaperone	3.70E-05	15	5	72.21
31542947	Heat shock 60 kDa protein 1 (chaperonin)	1.10E-04	28	5.7	61.21
5123454	Hsp70	1.50E-04	14	5.5	70.29
306890	Heat shock 60 kDa protein 1 (chaperonin)	8.40E-04	18	5.7	61.18
7672663	Heat shock cognate 71 kDa protein	1.60E-03	18	5.4	71.11
31542292	Chaperonin containing TCP1, subunit 2 (beta)	2.00E-03	13	6.1	60.96
2119712	Dnak-type molecular chaperone	2.20E-03	21	5.4	70.14
5729877	Heat shock 70 kDa protein 8	5.80E-03	13	5.6	53.61
11907570	Desmin	4.10E-12	10	5.2	53.6
22415740	Mutant desmin	9.40E-12	10	5.2	53.58
553163	KRT 8	7.10E-11	28	5.8	26.16
7161776	Cytokeratin type II	2.20E-10	3	7.7	59.8
23194177	Mutant desmin	3.30E-10	10	5.2	53.46
340234	Vimentin	5.80E-10	43	4.7	35.09
181573	KRT 8	1.80E-09	33	5.5	53.54
5031875	Lamin A/C	1.40E-07	27	6.4	65.17
34035	KRT 18	5.80E-07	35	4.9	26.43
15126742	Lamin B1	2.50E-06	20	5.1	66.67
5030431	Vimentin	2.80E-06	24	4.8	41.66
16924319	Actin, gamma 1	3.10E-06	31	5.8	40.83
4501885	Actin, beta	2.10E-05	24	5.3	42.06
3157976	Actinin, alpha 4	2.30E-05	15	5.5	105.6
14550542	KRT 7	3.10E-05	19	5.7	22.51
547749	KRT 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)	5.50E-05	19	5.1	59.73
88045	KRT K7, type II	1.00E-04	8	5.4	51.46
105815	KRT 8, type II	1.40E-04	30	5.4	53.73
4826659	Capping protein (actin filament) muscle Z-line, beta	2.10E-04	19	5.7	30.95
12025678	Actinin, alpha 4	2.50E-04	20	5.3	105.3
8569616	Moesin	3.60E-04	19	9	34
27436951	Lamin B2	3.70E-04	18	5.3	67.79
16306948	Actin, gamma 1	7.60E-04	31	5.2	17.99
547754	KRT 2A (epidermal ichthyosis bullosa of Siemens)	1.50E-03	17	8.3	66.13
4501889	Actin, gamma 2, smooth muscle, enteric	2.40E-03	13	5.3	42.26
10436392	Tubulin, delta 1	2.40E-03	21	5.2	22.59
30089956	KRT 7	2.70E-03	22	5.4	51.46
190028	Plastin 3 (T isoform)	7.60E-03	22	5.7	64.3

Table 1. Continued

GeneID	Common name	Expect.	% Cov.	pI	MW, kDa
15277503	Actin, beta	7.80E-03	23	5.6	40.54
4557705	KRT 9 (epidermolytic palmoplantar keratoderma)	7.90E-03	10	5.1	62.2
6912494	Microtubule-associated protein, RP/EB family, member 1	8.50E-03	26	5	30.15
14249342	Internexin neuronal intermediate filament protein, alpha	8.90E-03	19	5.3	55.54
4501891	Actinin, alpha 1	0.012	7	5.2	103.6
1070613	Actin alpha 2	0.013	14	5.2	42.49
4507877	Vinculin	0.016	9	5.8	117.3
6636344	Beta-actin	0.017	19	5.2	29.55
11493522	PRO1512	0.017	18	5.8	34.28
24234699	KRT 19	0.017	24	5	44.09
7161769	KRT, hair, basic, 4	0.025	14	8.6	65.96
7657381	PRP19/PSO4 homolog (<i>Saccharomyces cerevisiae</i>)	0.028	14	6.1	55.62
4557775	Myosin, light polypeptide 2, regulatory, cardiac, slow	0.029	34	4.9	18.76
28336	Actin, beta	0.032	31	5.2	42.14
2116655	Actin, cytoplasmic 1	0.04	18	5.3	42.06
88049	KRT 55k type II	0.043	15	6.4	53.9
5114261	Voltage-dependent anion channel 2	4.80E-03	18	7.7	32.06
12803591	Integrin, beta 8	2.10E-03	7	5.8	49.37
2073428	Immunoglobulin heavy chain	0.018	33	8.2	14.09
14251209	Chloride intracellular channel 1	0.025	19	5.1	27.25
14625919	Immunoglobulin kappa	0.026	40	9.1	11.89
20269752	Leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 3	0.027	9	9.1	46.27
28373129	Contactin 5	0.039	9	5.6	113.5
3318841	Peroxidase enzyme	6.00E-11	29	6	24.9
999892	Triosephosphoate isomerase	1.60E-09	58	6.5	26.81
26224790	Glucose-6-phosphate dehydrogenase	3.10E-09	38	6.9	55.21
68420	Triosephosphoate isomerase	5.60E-09	58	6.4	26.94
13786847	L-Lactate dehydrogenase	3.90E-07	34	5.7	36.77
4507813	UDP-glucose dehydrogenase	1.10E-06	29	6.8	55.69
577295	Glucosidase, alpha; neutral AB	1.60E-06	16	5.7	107.2
5032057	S100 calcium-binding protein A11 (calgizzarin)	3.50E-06	56	6.6	11.84
736677	Dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)	4.60E-06	25	9.4	49
22907049	Aldehyde dehydrogenase 3 family, member A1	9.30E-06	25	6.1	50.78
4758638	Peroxiredoxin 6	2.50E-05	30	6	25.13
26224870	Glucose-6-phosphate dehydrogenase	2.50E-05	38	6.7	55.21
2832346	Trx 1	5.10E-05	16	6.1	55.27
21614520	Glucose-6-phosphate dehydrogenase	5.20E-05	22	6.3	59.7
4557032	Lactate dehydrogenase B	1.00E-04	31	5.7	36.91
1709442	Dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)	1.40E-04	13	9.4	48.96
21361176	Aldehyde dehydrogenase 1 family, member A1	1.50E-04	28	6.3	55.47
1465733	Malic enzyme 1, NADP(+)-dependent, cytosolic	2.40E-04	11	5.9	63.88
12644008	UMP-CMP kinase	2.40E-04	21	5.4	22.43
5802974	Peroxiredoxin 3	2.50E-04	14	8	28.02
29468184	Nonmetastatic cells 1, protein (NM23A) expressed in	2.60E-04	27	5.4	19.86
2554831	Chain A, crystal structure of human glutathione-S-transferase	3.20E-04	40	5.4	23.55
3123721	24-kDa subunit of complex I	5.10E-04	29	7.1	25.7
3820535	Trx 1	4.80E-04	21	6.4	55.34
14250063	Peroxiredoxin 3	5.60E-04	24	7.1	28.05
87018	Carbamoyl-phosphate synthase (ammonia)	7.40E-04	10	6.3	166
4758788	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30 kDa (NADH-coenzyme Q reductase)	7.50E-04	30	7	30.34
494066	Glutathione-S-transferase	8.50E-04	23	5.4	23.43

Table 1. Continued

GeneID	Common name	Expect.	% Cov.	p/ Cov.	MW, kDa
6729803	HSP 70 kDa ATPase N-terminal domain	8.80E-04	23	6.7	41.98
1071834	Dihydrolipoamid S-succinyltransferase	8.90E-04	15	9.4	49.01
5174743	Ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	1.00E-03	15	8.9	29.92
4758786	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49 kDa (NADH-coenzyme Q reductase)	1.60E-03	18	7.2	52.93
20336761	Heme binding protein 1	3.80E-03	38	5.7	21.19
23308751	3-hydroxyisobutyrate dehydrogenase	4.50E-03	19	8.8	35.71
4505701	Pyridoxal (pyridoxine, vitamin B6) kinase	6.00E-03	21	5.7	35.31
13325114	Aldehyde dehydrogenase 3 family, member A1	6.90E-03	18	6.1	50.76
4557525	Dihydrolipoamide dehydrogenase (E3 component of pyruvate dehydrogenase complex, 2-oxo-glutarate complex, branched chain keto acid dehydrogenase complex)	7.00E-03	16	7.8	54.7
4758158	Neural precursor cell expressed, developmentally down-regulated 5	9.40E-03	16	6.1	41.7
5803013	Chromosome 12 ORF 8	8.70E-03	16	6.8	29.03
6538774	Trx 1	0.013	13	6.1	55.25
7431153	Malate dehydrogenase	0.013	19	5.9	36.63
15080016	Enoyl Coenzyme A hydratase 1, peroxisomal	0.013	18	8.7	36.14
26655526	Aldehyde dehydrogenase, cytosolic	0.013	17	6.3	55.35
387033	Nucleoside phosphorylase	0.015	28	7.1	32.39
182516	Ferritin, light polypeptide	0.016	40	5.6	16.43
4758484	Glutathione-S-transferase omega 1	0.019	13	6.2	27.83
18088311	Nit protein 2	0.02	16	6.9	30.96
4505753	Phosphoglycerate mutase 1 (brain)	0.024	26	6.7	28.9
7546523	Chain A, X-ray structure of human glucose 6-phosphate dehydrogenase	0.026	14	6.3	59.53
5453549	Peroxiredoxin 4	0.028	8	5.9	30.75
2118088	Haptoglobin	0.035	21	6.6	38.72
283971	Aldehyde dehydrogenase	0.036	18	6	50.72
18677088	Chondroitin sulfate GalNAcT-2	0.039	18	6	21.13
227920	Beta galactosidase	0.045	44	5.3	14.91
4502171	Adenine phosphoribosyltransferase	0.048	28	5.8	19.76
14249348	Thioredoxin-like 5	0.048	31	5.4	14.21
28940	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, beta polypeptide	4.60E-11	41	5.8	57.99
31873242	Inner membrane protein, mitochondrial (mitofilin)	6.60E-09	29	6.2	83
28931	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, beta polypeptide	5.30E-07	42	4.9	34.03
516766	Inner membrane protein, mitochondrial (mitofilin)	4.50E-06	41	5.2	18.3
14286220	Enoyl Coenzyme A hydratase, short chain, 1, mitochondrial	2.60E-04	40	8.9	31.84
14290586	Translocase of inner mitochondrial membrane 50 homolog (yeast)	8.40E-04	25	8.9	39.02
15451305	Electron transporter activity	8.90E-03	15	6.8	61.01
27498544	Similar to cytochrome c oxidase subunit VIa polypeptide 1 precursor; cytochrome C oxidase subunit VIa homolog	0.021	29	9.3	12.17
14789599	Translocase of inner mitochondrial membrane 50 homolog (yeast)	0.027	16	8.9	39.65
21361331	Carbamoyl-phosphate synthetase 1, mitochondrial	0.028	7	6.3	166.1
18999392	Cytochrome c oxidase subunit Va	0.044	28	6.3	16.92
6841066	Solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 24	0.047	18	5.3	46.09
9961244	ATP-binding cassette, sub-family B (MDR/TAP), member 10	0.2	10	10.1	79.48
8131894	Inner membrane protein, mitochondrial (mitofilin)	0.22	8	5.6	68.34
4826992	SCO cytochrome oxidase-deficient homolog 2 (yeast)	0.28	15	9	29.9
14165437	Heterogeneous nuclear ribonucleoprotein K	1.30E-09	28	5.2	51.3
4557645	Heterogeneous nuclear ribonucleoprotein L	4.80E-06	23	6.7	60.74
9910244	Mitochondrial ribosomal protein S22	5.50E-06	26	7.8	41.43
7739437	Heterogeneous nuclear ribonucleoprotein H3 (2H9)	6.00E-05	35	6.8	31.51
14277700	Ribosomal protein S12	7.50E-05	49	6.8	14.9
18104577	Paraspeckle component 1	8.40E-05	24	6.3	58.84

Table 1. Continued

GeneID	Common name	Expect.	% Cov.	p/	MW, kDa
16209579	tRNA isopentenyltransferase 1	0.021	17	6.4	49.27
9624998	Heterogeneous nuclear ribonucleoprotein H2 (H')	1.50E-04	11	5.9	49.53
14141159	Heterogeneous nuclear ribonucleoprotein H3 (2H9)	0.011	20	6.4	35.28
12654583	Ribosomal protein, large, P0	0.013	21	5.4	34.43
21619877	Cleavage stimulation factor, 3' preRNA, subunit 2, 64 kDa	0.015	19	6.3	59.22
21396489	Lon protease-like	1.50E-09	14	6	107
631472	Translation initiation eIF-4A2 homolog	9.30E-07	28	6.1	47.1
32425705	Tu translation elongation factor, mitochondrial	9.70E-07	19	7.3	50.2
631285	Endopeptidase La homolog	1.10E-06	15	5.8	104.4
5174419	ClpP caseinolytic protease, ATP-dependent, proteolytic subunit homolog (<i>E. coli</i>)	6.70E-06	26	8.6	30.45
2136315	Translation elongation EF-Tu	1.00E-05	17	7.9	49.86
13489087	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 1	2.70E-05	31	5.9	42.84
28872725	Proteasome (prosome, macropain) 26S subunit, nonATPase, 11	3.00E-04	16	6.1	47.73
17402904	Exosome component 6	3.20E-04	24	6.1	28.5
14165435	Heterogeneous nuclear ribonucleoprotein K	4.00E-04	22	5.4	51.24
5729991	Proteasome (prosome, macropain) 26S subunit, ATPase, 4	4.50E-04	23	5.1	47.46
5453854	Poly(rC) binding protein 1	7.50E-04	18	6.7	38.02
6707736	Poly(rC) binding protein 2	1.20E-03	23	6.3	38.96
21361657	Glucose-regulated protein, 58 kDa	1.20E-03	21	6	57.16
4099506	Proliferation-associated 2G4, 38 kDa	1.60E-03	9	7.2	38.33
15530265	Eukaryotic translation elongation factor 1 gamma	2.20E-03	18	6.3	50.47
6005942	Valosin-containing protein	2.90E-03	13	5.1	89.99
5453990	Proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	3.50E-03	28	5.8	28.88
22538465	Proteasome (prosome, macropain) subunit, beta type, 3	4.70E-03	36	6.1	23.22
414046	Protease, serine, 15	9.50E-03	10	5.9	95.58
7443384	EF-Tu-like protein	0.018	13	7.9	49.86
1706611	Tu translation elongation factor, mitochondrial	0.02	10	7.3	49.87
5453842	Proliferation-associated 2G4, 38 kDa	0.024	10	6.13	43.81
2245365	Glucose-regulated protein, 58 kDa	0.025	17	5.9	57.16
4502027	Albumin	4.50E-09	25	5.9	71.35
4826760	Heterogeneous nuclear ribonucleoprotein F	7.80E-04	26	5.4	46
285975	GDP dissociation inhibitor 2	1.40E-07	31	5.9	51.1
4502101	Annexin A1	1.50E-07	29	6.6	38.92
809185	Annexin V	9.70E-07	46	4.9	35.84
189617	Annexin A4	9.80E-07	44	5.6	36.27
999926	Annexin V	1.80E-06	29	5	35.84
4757768	Rho GDP dissociation inhibitor (GDI) alpha	3.00E-06	35	5	23.25
4757756	Annexin A2	5.80E-05	30	7.7	38.81
7706485	Heat shock protein 75	2.20E-04	17	8.2	80.28
18645167	Annexin A2	4.40E-04	28	7.7	38.79
18089095	PreB-cell colony enhancing factor 1	2.30E-03	21	6	42.07
17511976	Heat shock protein 75	2.90E-03	14	8.5	80.38
11496277	Mitogen-activated protein kinase kinase 1 interacting protein 1	3.10E-03	56	6.7	13.66
5031977	PreB-cell colony enhancing factor 1	4.80E-03	11	6.7	55.79
18448452	Luteinizing hormone beta polypeptide	7.50E-03	36	10	15.38
1082886	TNF type 1 receptor-associated protein	0.011	11	8.7	75.72
223127	Lutrophin beta	0.02	36	10	12.89
442631	Annexin I	0.021	13	7.9	35.25
1256422	p21/Cdc42/Rac1-activated kinase 1 (STE20 homolog, yeast)	0.022	13	5.5	60.81
4504989	Luteinizing hormone beta polypeptide	0.036	18	9.5	16.01
12652977	Phosphodiesterase 6B, cGMP-specific, rod, beta (congenital stationary night blindness 3, autosomal dominant)	0.044	10	5.1	99.42

Table 1. Continued

GeneID	Common name	Expect.	% Cov.	p/	MW, kDa
4506753	RuvB-like 1 (<i>E. coli</i>)	1.40E-07	42	6	50.55
4505773	Prohibitin	1.90E-05	29	5.6	29.84
1730078	Leucine-rich PPR-motif containing	2.30E-05	9	5.5	146.4
4503571	Enolase 1, (alpha)	1.60E-04	35	7	47.49
6678271	TAR DNA binding protein	3.90E-03	11	5.8	45.06
460789	Heterogeneous nuclear ribonucleoprotein K	0.019	14	5.1	51.34
3478669	Mesenchyme homeo box 2 (growth arrest-specific homeo box)	0.034	15	7.2	25.74
32450454	Zinc finger protein 364	8.80E-03	12	5.4	34.26

microscope slides and the phase-contrast as well as the fluorescent images were photographed. To determine subcellular distribution of mitochondria, cover slip cultures of A549 cells were placed in a thermo-controlled microscopic chamber, loaded with 1 μ M Rhodamine 123 (excitation: 507 nm and emission 529 nm) for 5 min [12]. Microscopy was performed on a Zeiss LSM510 META System.

2.9 Data analysis

Functional genome ontology annotation mapping was performed using the NIAID Database for annotation, visualization, and integrated discovery (DAVID, <http://david.niaid.nih.gov/david/version2/index.html>, [13]). Pathways analysis was performed using individual clusters as input into the Ingenuity Knowledge Base database (<http://www.ingenuity.com/>). The three highest scoring networks identified were merged. Protein domain analysis was by InterPro release 11 hosted by the European Bioinformatics Institute (<http://www.ebi.ac.uk/interpro/>). InterPro is an integrated documentation resource of UniProt, PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, PANTHER, and Gene3D [14].

3 Results

3.1 Nuclear fractionation

A549 is a transformed human cell line that retains features of type II alveolar cells including surfactant secretion [15], and has been widely used as a model to investigate the response of epithelial cells to respiratory virus infection *in vitro* [6, 16–18]. In this study, we isolated highly purified A549 nuclei by sucrose step gradient centrifugation [6, 19]. To demonstrate the quality of the nuclear preparations, the extracts were visualized by phase-contrast microscopy and stained with DAPI, a stain selective for DNA (Suppl. Material). Here, nuclear preparations consisted of homogenous membrane-bound forms with intact nucleoli (identified by phase-contrast microscopy) that intensely stained with DAPI. As biochemical markers for enrichment of nuclear

protein and to assess the degree of cytoplasmic contamination, Western immunoblot analysis was performed on A549 nuclear and cytoplasmic fractions. Nuclear and cytoplasmic proteins corresponding to the same number (1×10^6) of cells were stained with antibodies to either nuclear proteins (the nuclear envelope protein, lamin B, or nucleoplasmic RNA polymerase II, and the p300 coactivator), cytoplasmic proteins (I κ B α or β -tubulin), or both compartments (β -actin). We found a strongly reactive 80 kDa lamin B band was detected specifically in the nuclear fractions (Suppl. Material). Similarly, Pol II and p300 were greatly enriched relative to cytoplasmic signals. In contrast, 37 kDa I κ B α was strongly cytoplasmic, whereas the recycling phospho-I κ B α (a protein that runs \sim 8 kDa larger [19]) was predominately nuclear. The anti- β -tubulin antibody strongly stained an 80 kDa band significantly enriched in the cytoplasmic fraction. Together these observations gave strong evidence that the nuclear fractions were homogenous and enriched in nuclear proteins.

To represent the data, a reference template gel of the nuclear proteins was prepared, with the mapped locations of proteins of interest to this study (Fig. 1). On this template gel, we were able to resolve 1416 spots. To identify the proteins in the nucleus, the nuclear proteins were pre-fractionated on HPLC RP chromatography, where nine distinct fractions were collected and individually run on high-resolution criterion 13.3×8.7 cm² 2-DE gels for protein identification. In aggregate, a total of 6648 spots were identified by Sypro-Ruby staining, increasing the number of proteins for identification by \sim 4.5-fold. Protein identification was performed by PMF using MALDI-TOF MS where the resulting peptide masses were then used to match proteins in the NCBI human database using a Bayesian matching algorithm [10]. Proteins whose statistical indicators gave a high probability of identification were collated.

3.2 Functional analysis of the A549 nuclear proteome

The 719 proteins were compared to eliminate duplicate identifications, yielding 433 proteins with unique NCBI gene identifiers (GI). These proteins were tabulated and classified into major biochemical pathways. (The full data set



Figure 1. 2-DE template of A549 nuclear proteins. Shown is a master gel representation of the nuclear proteins subjected to HPLC prefractionation and identification by PMF. Horizontal dimension, IEF was conducted over a pH range from 5 to 8. Vertical dimension, fractionation by SDS-PAGE. Migration of MW standards (in kDa) are shown at left. Identified proteins of interest are labeled as: (1) β -actin; (2) heat shock protein 70; (3) moesin; (4) TNF receptor associate protein-1; (5) proteasome 26S subunit ATPase subunit 4; (6) cytokeratin 8; (7) thioredoxin R1; (8) DDX3X; (9) proteasome 26S subunit, non-ATPase; (10) KRT 18; (11) glutathione-S-transferase; (12) peroxiredoxin 3; (13) proteasome α 6 subunit; (14) proteasome α 2 subunit. A more fully annotated 2DE is shown in the Suppl. Material.

is found in the Suppl. Material. A filtered list of proteins with identified function, representing those with a highly significant expectation score is shown in Table 1.) The major biochemical pathways identified included cytoskeletal/structural (98 entries), metabolism (82 entries), and signaling (32 entries). Other identified categories, numerically smaller, but specifically indicative of nuclear function, included chromatin remodeling (15 entries), protein synthesis/degradation (39 entries), RNA binding/ribosomal (14 entries), and transcription/DNA binding (21 entries). Proteins whose function could not be classified, listed as “unknown”, constituted 46 proteins in the group.

Although our Western blotting did not reveal significant cytoplasmic contamination, we found 20 proteins that were known to be mitochondrial, including multiple identifications of the inner mitochondrial membrane protein, mitofilin, and mitochondrial ATP synthase (Table 1). This identification could be interpreted several ways – either these mitochondrial proteins were localized to both nucleus and mitochondria, or mitochondria were present in our nuclear preparations. To assess the latter possibility, the presence of intact mitochondria was assayed using two mitochondrial-selective dyes, MitoTracker Red CMXRos [20] and rhodamine-123 (RD-123) [12]. MitoTracker Red accumulated in active mitochondria and stained small organelles peripherally associated with the nuclear membrane of our nuclear preparation (Fig. 2A). Similar results were found using

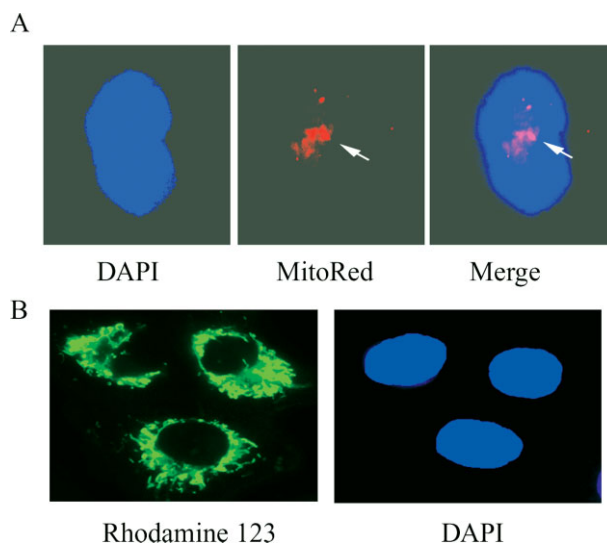


Figure 2. Peripheral mitochondrial association with nuclear membranes. (A) MitoTracker Red staining. Sucrose cushion-purified nuclei were stained with MitoTracker Red [20] and DAPI and imaged. Right panel, merged image. Faint staining of mitochondria directly associated with nuclear membranes was observed. (B) Rhodamine 123 staining. Sucrose-cushion purified nuclei were stained with rhodamine 123, and the oxidized form was visualized. Green (oxidized) rhodamine stain indicating mitochondrial uptake and metabolism is clearly seen in the periphery of the nuclear preparations.

RD-123 (Fig. 2B). Together these indicate that mitochondria are tightly associated with nuclear membranes in our sucrose cushion preparations. We therefore conclude that the mitochondrial proteins identified were due to organellar contamination. These proteins were excluded in further functional analysis.

To eliminate bias in the functional classifications, the proteins were subjected to GO analysis to identify the major biological processes and molecular functions. The GO is a structured, controlled vocabulary that describes gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner [21]. Two hundred seventy-eight of the 433 proteins were annotated to map to GO biological process terms and are graphically shown in Fig. 3. The biological processes containing numerically the greatest number of proteins in the nuclear data set include “metabolism” (43%), “cellular physiological process” (25%), and “macromolecule metabolism” (21%). A similar analysis for molecular function was conducted; here, 316 of the 433 input proteins were successfully classified (Fig. 4). The major molecular functions included “catalytic

activity” (33% of entries), “structural molecule activity” (16% entries), and “purine nucleotide binding” (15% of entries). To determine whether these biological processes represent function(s) enriched in the nuclear data set relative to the human proteome, the probability of over-representation was calculated [13, 22]. In brief, the probability of over-representation assigns a statistical significance to protein functions in the nuclear proteome relative to those functions in the annotated human proteome to exclude common protein functions from skewing functional analysis of the nuclear subcellular fractions [22]. The biological processes and the molecular functions with the greatest enrichment were calculated and tabulated in Table 2. Here it can be seen for the GO biological process group, metabolic processes were functions greatly enriched relative to the annotated human proteome and in the GO molecular function, nucleotide binding, RNA binding, and oxidoreductase activities were those with the greatest enrichment probabilities. Together the GO classification and enrichment probabilistic analysis supported the major biochemical pathways suggested by expert classification.

Biological Process

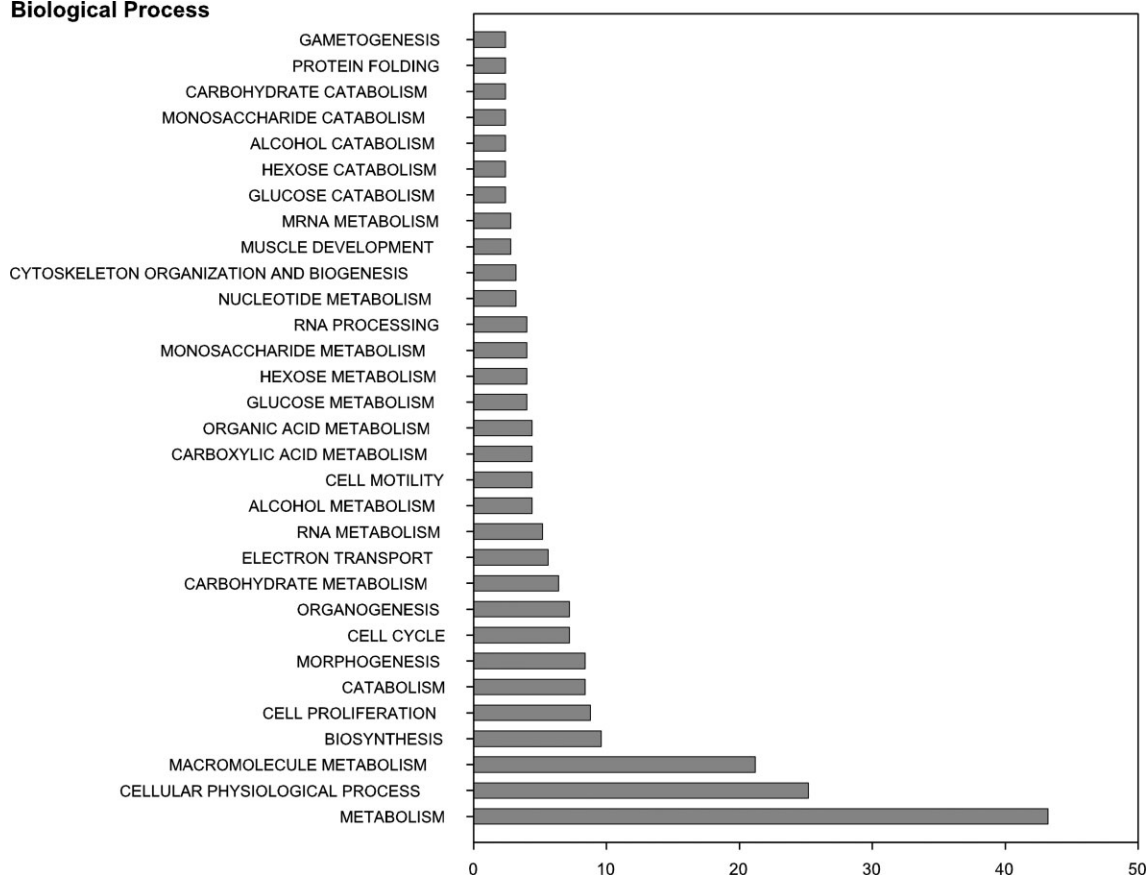


Figure 3. GO analysis of nuclear proteome: biological process. Two hundred and seventy-eight of the 433 unique proteins were classified into a biological process using GO. Shown is a bar histogram of the percentage of the group that were classified.

Molecular Function

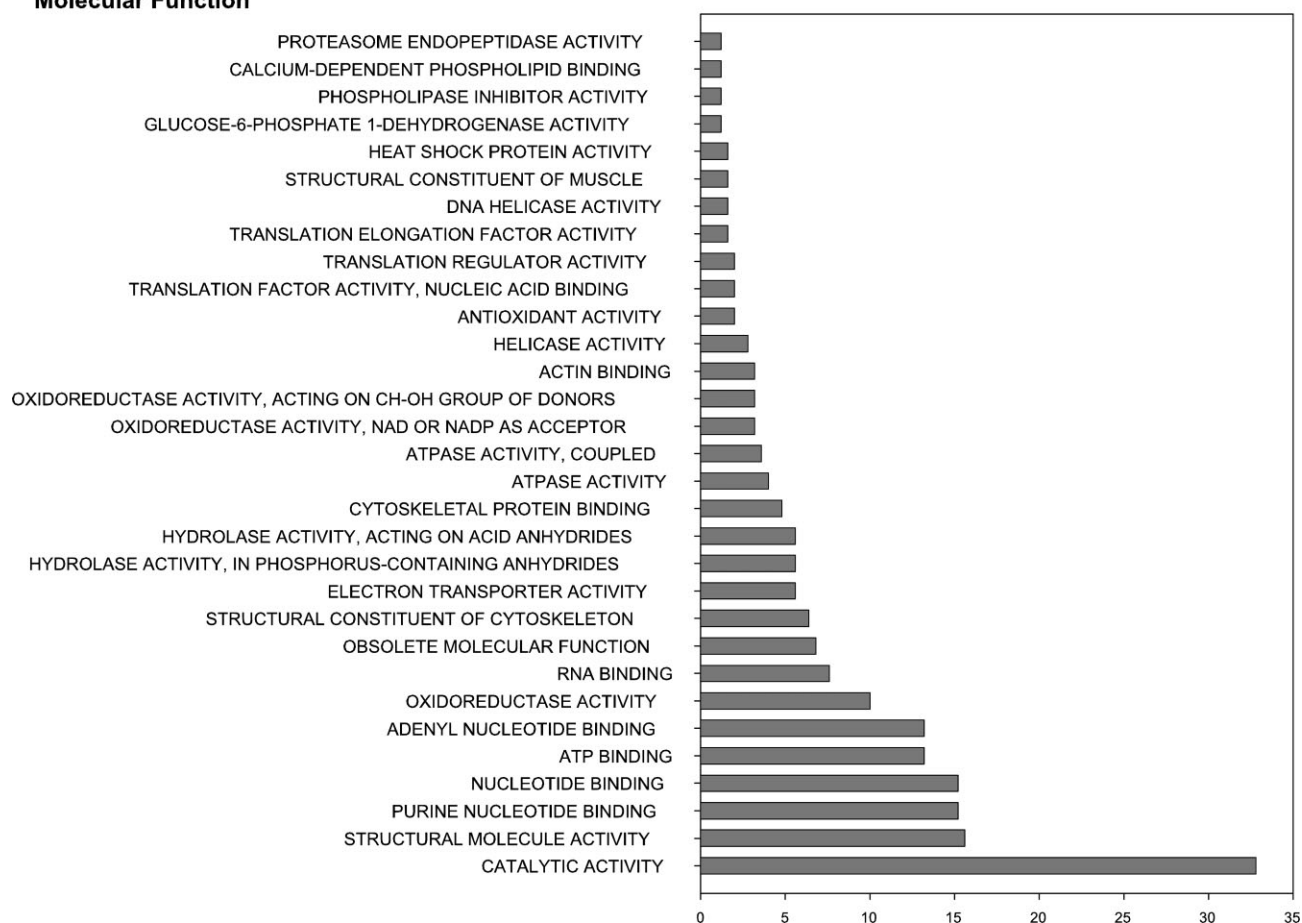


Figure 4. GO analysis of nuclear proteome: molecular function. Three hundred and sixteen of the 433 unique proteins were classified by molecular function. Shown is a histogram of the relevant molecular functions identified.

Next we used Ingenuity pathways analysis (IPA) as an additional method for analysis of the biological networks contained in the nuclear proteome data set. In this method, the 433 unique proteins from the nuclear proteome database were analyzed for network association by the Ingenuity Knowledge Base, a curated database of publicly available pathways and relationships. IPA compares proteins in the input group and displays a rank-ordered list of networks whose activities are most likely to be affected. For each network, its members and their relationships (functional and physical) are displayed graphically as a network diagram with each node representing a protein and its association represented as a line (edge). From the nuclear proteome data set, six high-ranking networks were identified. Figure 5 shows one of the top networks controlling cytoskeletal organization, and tumor necrosis factor (TNF) signaling is identified by the Knowledge Base from proteins in the 2-DE data set. In this network, the DDX3X DEAD box RNA helicase is known to bind to NF- κ B2, an interaction identified in tandem affinity purification studies in HEK 293 cells [23]. Similarly, coimmunoprecipitation studies show that p38 mito-

gen-activated protein kinase (MAPK)14, KRT isoforms-8 and -18, and 14-3-3z (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, YWHAZ) proteins form a complex, where KRT-8 is a substrate for phosphorylation by MAPK14. This process induces KRT filament reorganization [24]. Also, KRT-8 and -18 form a complex with TNF receptor associated death domain (TRADD) [25]. This pathway analysis suggests an intimate link between experimentally identified nuclear intermediate filaments and TNF-nuclear factor- κ B (NF- κ B) signaling pathway in alveolar epithelial cells.

Forty-five proteins of the experimentally determined nuclear proteome data set could not be classified into a primary biological pathway (Table 1). To determine putative functions of these unknown proteins, we submitted this set of proteins for functional domain analysis. As seen in Table 3, the function for the majority of proteins could be inferred. The two largest groups were those proteins containing domains homologous to those involved in purine nucleotide metabolism (IMP dehydrogenase/GMP reductase), and protein interaction (tetra- α peptide repeat (TPR)). GMP reductase catalyzes the irreversible and

Table 2. Over-represented GO biological processes and molecular function in the nuclear proteome data set. Tabulated are the top ten biological processes and molecular functions with the greatest statistical significance for enrichment in the nuclear proteome data set

	%	P value
Biological process		
Glucose metabolism	4	2E-06
Hexose metabolism	4	3E-05
Monosaccharide metabolism	4	4E-05
Translational elongation	2	0.0002
Carbohydrate metabolism	6.4	0.0003
Alcohol metabolism	4.4	0.0007
Glucose catabolism	2.4	0.0013
Metabolism	43.2	0.0018
RNA metabolism	5.2	0.0018
Molecular function		
Structural molecule activity	15.6	1E-13
Structural constituent of cytoskeleton	6.4	1E-12
ATP binding	13.2	2E-05
Adenyl nucleotide binding	13.2	3E-05
Purine nucleotide binding	15.2	5E-05
Nucleotide binding	15.2	8E-05
RNA binding	7.6	8E-05
Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	3.2	0.0003
Oxidoreductase activity, acting on CH-OH group of donors	3.2	0.0004

NADPH-dependent reductive deamination of GMP into IMP; IMP dehydrogenase catalyzes the rate-limiting reaction of *de novo* GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [26]. This enzyme is responsible for maintaining an appropriate balance of A and G nucleotides. The TPR is a repeated structural motif that mediates the assembly of multiprotein complexes, present in approximately 360 annotated human proteins [27]. Together, this analysis indicates that domain analysis can be used to infer a biological process from the majority of unknown or unclassified proteins in the nuclear proteome data set.

4 Discussion

The eukaryotic nucleus is highly structured organelle important for chromosomal maintenance/segregation, gene expression, RNA processing/export, antiviral defense, and others [28]. The airway epithelial cell nucleus mediates the phenotypic responses to exogenous inflammatory stimuli, such as those induced by viral infection [3, 6]. Previously we identified 24 differentially expressed proteins in the nucleus by an organelle-focused differential expression study that identified proteins important in the nuclear cytoskeleton, chromatin remodeling, oxidant/antioxidant response, and

heat shock response [6]. Recently we found that HPLC pre-fractionation can increase the depth of interrogation of the proteome, enabling the identification of up to ten times the number of proteins in whole cell extracts. Here we have applied organellar isolation coupled with HPLC fractionation to more comprehensively study the spectrum of proteins in the mammalian nucleus. We report our experience using a combined informatics approach to systematically identify the biological processes and molecular functions affected by these proteins that indicate the nucleus is highly metabolically active, in addition to the complement of structural, DNA and RNA binding proteins. Preliminary systems analysis identifies networks of protein interactions including those at the interface between cytoskeletal and metabolic molecules and those of the TNF-NF- κ B signaling pathway.

To our knowledge, this work represents the most comprehensive experimental analysis of the nuclear proteome. Previous gene trap approaches have been used to identify proteins that are preferentially nuclear [29]. Although this approach revealed a number of proteins involved in chromatin maintenance (*ca.* ~100 proteins were identified), the approach was highly biased toward proteins that were only nuclear and, by nature of the technique, identified a number of ESTs for which the complete protein remains unknown. It is important to emphasize the point that components of the eukaryotic nucleus are dynamic organelles, with some proteins undergoing cytoplasmic–nuclear shuttling, including those important in cytoskeletal structure and inducible gene expression (transcription factors). For example, the cytoskeletal protein, actin, is under dynamic nuclear import/export control and is partitioned in both the cytoplasm and nucleus (reviewed in [30]). These types of proteins would not be identified in the screen used in this prior study. A caveat to our study is that mitochondria are copurifying with our nuclear preparations. Mitochondria have similar densities as that of the 1 M sucrose cushion used in our preparation, and therefore could preferentially be enriched with nuclei during centrifugation, or, alternatively, the mitochondria could be contiguous with the nuclear membranes. Normally, mitochondria are closely associated with the smooth ER, a membrane structure contiguous with the nuclear membrane, and could remain attached to the nuclear membrane after cellular disruption in nonionic detergent. Because of their staining characteristics, being adherent to the nuclear membranes, rather than as free organelles in our preparations (Fig. 2A and B), we think the latter explanation is more likely. Mitochondria in particular are associated closely with the ER, a membrane contiguous with the nuclear membrane. More work will be required to determine if these proteins are also part of the nuclear membrane or only the associated mitochondrial structures. Recently a subtractive proteomics strategy has been used to identify proteins enriched in the rat liver nuclear membrane [31]. A comparison of this data set with ours indicates that with the exception of lamin A/C isoforms, the membrane proteins are not well represented in

Table 3. Domain analysis of proteins with unknown function. For each of the proteins identified as “unknown” in Table 1, the InterPro site was queried for domains in SMART, Panther, and Pfam databases to identify functional domains of each protein. Domains are individually structurally folded functional units of proteins which mediate a biochemical function. Shown are the unknown proteins grouped by primary biological function (redundant domains were eliminated). Abbreviations: GI, GenBank identifier; Expect, expectation score; InterPro, InterPro domain accession number

GI	Name	Expect	InterPro	Domain name
NADPH Dehydrogenase				
6624942	Hypothetical protein DKFZ	6.30E-05	IPR001268	NADH dehydrogenase, 30 kDa
–	–	–	IPR008992	Bacterial enterotoxin
–	–	–	IPR010218	NADH (or F420H2) dehydrogenase
IMP Dehydrogenase/GMP Reductase				
20072753	Mitofilin	1.80E-06	IPR001093	IMP dehydrogenase/GMP reductase
2134737	Alpha-complex protein 1	7.10E-06	IPR001093	IMP dehydrogenase/GMP reductase
–	–	–	IPR004087	Khomology, RNA binding domain
5729937	Metaxin 2	2.00E-05	IPR001093	IMP dehydrogenase/GMP reductase
–	–	–	IPR010987	Glutathione-S-transferase, C-terminal
–	–	–	IPR012336	Thioredoxin-like fold
284289	Leucine-rich protein	1.70E-04	IPR001093	IMP dehydrogenase/GMP reductase
–	–	–	IPR002885	Pentatricopeptide repeat
–	–	–	IPR008940	Protein prenyltransferase
21362024	Hypothetical protein	0.012	IPR001093	IMP dehydrogenase/GMP reductase
31753196	DKFZP434I116 protein	0.023	IPR001093	IMP dehydrogenase/GMP reductase
2204215	Stromal antigen 2	0.038	IPR001093	IMP dehydrogenase/GMP reductase
10434568	DKFZP434I116 protein	0.047	IPR001093	IMP dehydrogenase/GMP reductase
13623521	Hypothetical	0.047	IPR001093	IMP dehydrogenase/GMP reductase
29789373	Hypothetical	0.092	IPR001093	IMP dehydrogenase/GMP reductase
–	–	–	IPR001440	TPR repeat
16041090	Hypothetical protein	0.16	IPR001093	IMP dehydrogenase/GMP reductase
13375569	PCD 6 interacting protein	0.26	IPR001093	IMP dehydrogenase/GMP reductase
–	–	–	IPR004328	BRO1
7242967	CASK interacting protein 1	0.21	IPR001093	IMP dehydrogenase/GMP reductase
–	–	–	IPR001452	SH3
–	–	–	IPR001660	Sterile alpha motif SAM
Heat Shock/Chaperonin				
16041102	Hypothetical protein	8.90E-07	IPR001023	Heat shock protein Hsp70
Intermediate Filament				
21752884	Hypothetical protein	3.00E-10	IPR001664	Intermediate filament protein
–	–	–	IPR011000	Apolipoprotein III-like
Immunoglobulin				
16923185	Ig heavy chain	2.50E-04	IPR007110	Immunoglobulin-like
17224464	Hypothetical protein	0.12	IPR003006	Immunoglobulin/MHC
Peptidase				
21756162	Protease, serine, 15	8.80E-08	IPR001270	Chaperonin clpA/B
–	–	–	IPR001984	Peptidase S16, ion protease
–	–	–	IPR003593	AAA ATPase
PDZ domain				
8922964	Synaptojanin 2 binding protein	1.40E-05	IPR001478	PDZ/DHR/GLGF
22261816	Synaptojanin 2 binding protein	6.00E-05	IPR001478	PDZ/DHR/GLGF
Redox				
23271051	Metaxin 2	2.70E-03	IPR010987	Glutathione-S-transferase, C-terminal
–	–	–	IPR012336	Thioredoxin-like fold
RNA binding				
4505087	Mago-nashi homolog	1.40E-04	IPR004023	Mago nashi protein

Table 3. Continued

GI	Name	Expect	InterPro	Domain name
8922331	Mago-nashi homolog	0.099	IPR004023	Mago nashi protein
8439384	Similar to ribosomal L12	0.15	IPR000911	Ribosomal protein L11
TPR repeat				
12698001	KIAA1728	9.60E-03	IPR001440	TPR repeat
–	–	–	IPR002110	Ankyrin
18088455	KIAA0103	0.027	IPR001440	TPR repeat
7661910	KIAA0103	0.033	IPR001440	TPR repeat
4240273	KIAA0892	0.083	IPR001440	TPR repeat
Miscellaneous				
21264404	Bystin-like	0.022	IPR007955	Bystin
4502395	Beclin 1	0.028	IPR007243	Autophagy protein Apg6
14388456	Hypothetical protein	0.035	IPR000504	RNA-binding region RNP-1
21740144	Hypothetical protein	0.042	IPR011046	WD40-like
4680381	Beclin 1	0.056	IPR007243	Autophagy protein Apg6
19584401	Huntingtin interacting protein B	0.061	IPR001202	WW/Rsp5/WWP
14017883	KIAA1833	0.072	IPR000357	Heat/intracellular transport
31542242	Chromosome 14 ORF	0.22	IPR002048	Calcium-binding EF-hand
–	–	–	IPR008946	Steroid nuclear receptor, ligand-binding
4758302	Enhancer of rudimentary	0.27	IPR000781	Enhancer of rudimentary
16550881	Hypothetical	0.27	IPR001909	KRAB box
–	–	–	IPR007086	Zinc finger, C2H2-subtype
–	–	–	IPR007087	Zinc finger, C2H2-type
20521820	Paladin	0.18	IPR003595	Protein tyrosine phosphatase, catalytic
Not found				
21739618	KIAA1833	0.1	IPR000357	Heat
12804041	Nuclear protein E3-3	4.20E-06	IPR006729	Protein of unknown function DUF598
13376290	Hypothetical protein	0.08	N/F	N/A
4106983	R29828_1	0.076	IPR007998	Protein of unknown function DUF719
7020296	Hypothetical protein	0.12	N/F	N/A
22761659	<i>Homo sapiens</i> cDNA	0.12	N/F	N/A

our 2-DE data set. The approach of subcellular fractionation is therefore complementary to other approaches used to identify proteins in the eukaryotic nucleus.

Our findings suggest a significant enrichment of proteins involved in metabolic processes, which are contained in the A549 nuclear data set (Tables 1, 3). Although the identified proteins encode a spectrum of metabolic activities, a significant fraction of the identified proteins are those involved in protecting against oxidative damage. These proteins include peroxiredoxin, thioredoxin reductase (Trx), and glutathione-S-transferase (GST). The peroxiredoxin family contains six known isoforms (Prx1 to 6) that play an important role in protecting lipids, enzymes, and DNA against peroxides, such as hydrogen peroxide [32]. Trx is a low-molecular weight flavoprotein protein that catalyzes the reduction of protein disulfides and other oxidized biomolecules through its COOH terminal catalytic site to maintain essential Cys residues in their active thiol forms [33]. GST is a family of detoxifying enzymes that catalyzes the reaction of glutathione with an oxidized acceptor molecule to form an

S-substituted glutathione. In the airway, GST is important for reducing oxidized glutathione, the major antioxidant molecule secreted in the alveolar lining fluid to protect against cellular damage induced by inhalation of inhaled oxidants, such as ozone.

Another major cellular function of the proteins in the nuclear proteome data set are intermediate filaments in nuclear cytoskeletal function. Lamins are proteins with well-established roles in forming subnuclear membrane scaffolds [34]. Similarly, increasing evidence shows the role of actin and actin-binding proteins in the formation of nuclear matrix, control of RNA splicing, and organization of chromatin-remodeling complexes (reviewed in [30]). In addition, KRT 8 and 18 isoforms have been identified in this study and our previous work [6]. KRT are important cytoskeletal intermediate filaments that are expressed in cell type-restricted patterns in the respiratory tract. Previous work has shown that KRT-7, -18, and -19 are normally expressed in bronchial epithelial cells [35], and their expression can be induced in response to toxicant exposure [36], and viral infections [6, 37].

Our pathways analysis indicates that KRTs further participate in networks linked to TNF-NF- κ B signaling pathway (Fig. 5), suggesting that cytoskeletal elements may also be important in mediating inflammatory toxicant and viral responses in the alveolar nucleus.

Perhaps surprisingly, we identified a significant fraction of nuclear proteins as those involved in the ubiquitin-proteasome pathway. Although the cytoplasmic proteasome has been best demonstrated to have a role in ubiquitin-mediated turnover of misfolded proteins and in signal transduction [38], recent studies have indicated the important role of nuclear proteasomal components in chromatin-associated protein degradation [39] and gene expression control [40]. For example, coimmunoprecipitation experiments have

shown that members of the 26S proteasome associates with RNA polymerase II complexes at the 3' end of genes and in stalled complexes associated with UV damage [40].

Our study further identifies proteins important in signaling through the TNF-NF- κ B pathway. TNF α is a central mediator of the pulmonary inflammatory response by its ability to activate adhesion molecule expression, enhance leukocyte trafficking, and affect the expression of secondary cytokine cascades controlling leukocyte recruitment and activation [41–43]. TNF signaling activates a major downstream signaling network known as the I κ B Kinase (IKK)-NF- κ B pathway that mediates airway inflammation [44–47]. NF- κ B is a latent cytoplasmic transcription factor maintained in a cytoplasmic location by binding the I κ B inhibitors, pro-

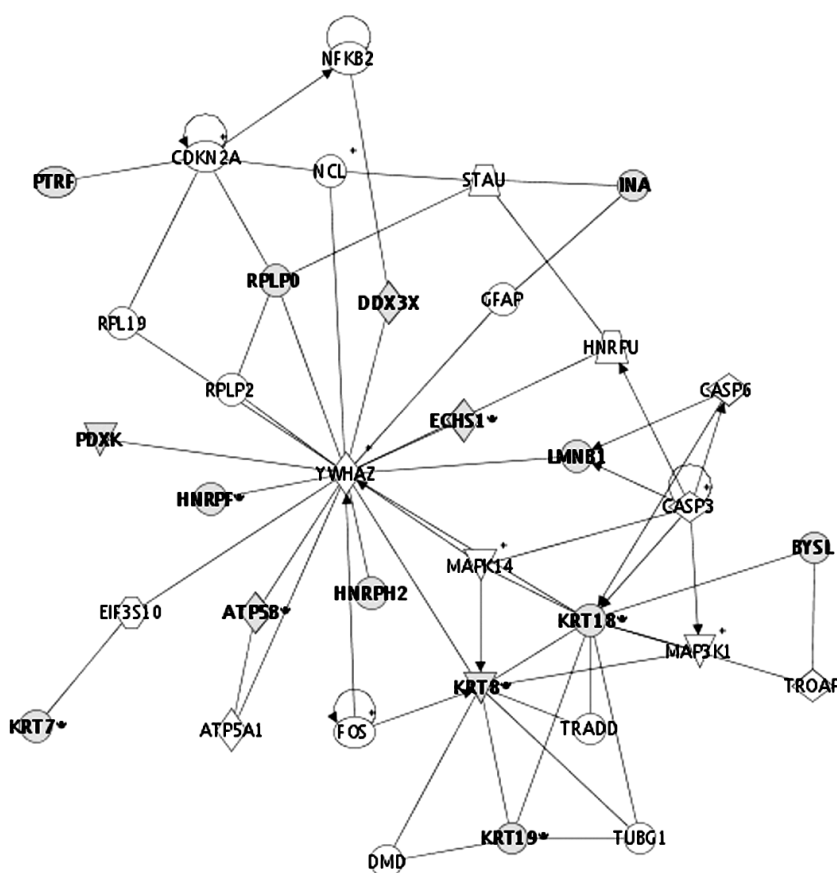


Figure 5. Pathway analysis of nuclear proteins. Pathway analysis of proteins by the Ingenuity Knowledge Base. Shown is a graphical representation of the highest scoring of 15 networks containing the largest number of focus genes. Protein protein associations are indicated by edges containing single lines, whereas proteins that act upon another protein (controlling their expression) are indicated by arrows. Nodes are represented by highlighting and shapes. The shaded nodes in bold are those proteins identified with high confidence in the nuclear proteome data set. For the shapes, circles indicate structural protein, ovals indicate transcription factor, and diamonds indicate proteases. Relationships between nodes are represented as edges. For the edges, a simple line indicates "binds to" and an arrow indicates "acts on" (or controls expression of). For example, in this network, KRT18 binds to KRT8, and KRT8 is acted upon (phosphorylated) by MAPK14. Similarly FOS acts on (induces the expression of) KRT8 [55]. Abbreviations used are: ACT, actin; ACTG, actin gamma; annexin, ANX; ATP5B, ATP synthase; BYSL, bystin-like; CASP3, caspase 3; DDX3X, DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked; ECHS, enoyl coenzyme A hydratase; HNRP, heterogeneous nuclear ribonucleoprotein; INA, internexin neuronal intermediate filament protein; KRT; MAP mitogen-activated protein kinase; LMNB1, lamin B1; RPLP0, ribosomal protein, large P0; PDXX, pyridoxal kinase; TRADD, TNF receptor-associated death domain; YWHAZ.

teins that bind and specifically inactivate it by masking its nuclear localization sequence, thereby preventing its nuclear entry [48]. NF- κ B is activated TNF signaling indirectly by targeted proteolysis of I κ B (reviewed in [49]). As a result, liberated NF- κ B rapidly enters the nucleus to activate target gene expression by forming a nucleoprotein complex with chromatin-remodeling proteins, kinases, and other transcription factors [50–52].

Our data has identified a number of nuclear proteins that modify the function of NF- κ B signaling pathway. The antioxidant flavoprotein, Trx-1, has been found to enhance DNA binding and transcriptional activity of NF- κ B upon its nuclear translocation [33, 53]. Interestingly, inhibitors of Trx-1 suppress TNF-induced expression of inflammatory genes, such as cyclo-oxygenase, in endothelial cells [33]; our findings would suggest Trx-1 could also be a target for anti-inflammatory therapy in airway inflammation. In addition, IPA network analysis identifies an interface between KRT-8 and 18 with signaling molecules in the TNF signaling pathway. As seen in Fig. 5, KRT-8 and -18 form a complex with TRADD. Earlier studies have found that cells deficient in KRT expression are 100 times more sensitive to TNF-induced cytotoxicity than those with normal KRT expression [25]. These findings would suggest that intermediate filaments of the cytoskeleton play important roles in TNF signaling. Finally, the DDX3 identified in our studies has been found to complex with the inducible NF- κ B2 DNA binding isoform in TAP-affinity tagging studies [23]. DDX3 is a protein known to be involved in RNA splicing and nucleo-cytoplasmic shuttling by binding the CRM1 export factor and localizing to nuclear membrane pores [54]. DDX3 may be important for the oscillatory control of NF- κ B dependent gene expression [46].

In summary, a combined approach of subcellular isolation followed by HPLC prefractionation can significantly enhance the systematic identification of nuclear proteins in the alveolar A549 epithelial cell. This information coupled with analysis of biological processes, molecular functions, regulatory networks, and domain identification can be used to gain insights into the complexity of functions controlled by the protein machines in the epithelial nucleus.

This work was supported by NIAID grant AI40218 (A.R.B.), NHLBI contract N01-HV-28184 "Proteomic Technologies in Airway Inflammation" (A.K.), NCI grant 1R24CA88317 (A.K.), and NIEHS P30 ES06676 "Center in Environmental Toxicology" (J. Halpert, UTMB).

5 References

- [1] Dobbs, L., *Am. J. Respir. Crit. Care Med.* 1994, 150, 531–532.
- [2] Rennard, S. I., Romberger, D. J., Sisson, J. H., von Essen, S. G. *et al.*, *Am. J. Respir. Crit. Care Med.* 1994, 150, S27–S30.
- [3] Zhang, Y., Luxon, B. A., Casola, A., Garofalo, R. P. *et al.*, *J. Virol.* 2001, 75, 9044–9058.
- [4] Martin, L. D., Rochelle, L. G., Fischer, B. M., Krunkosky, T. M., Adler, K. B., *Eur. Respir. J.* 1997, 10, 2139–2146.
- [5] Polito, A. J., Proud, D., *J. Allergy Clin. Immunol.* 1998, 102, 714–718.
- [6] Brasier, A. R., Spratt, H., Wu, Z., Boldogh, I. *et al.*, *J. Virol.* 2004, 78, 11461–11476.
- [7] Martosella, J., Zolotarjova, N., Liu, H., Nicol, G. *et al.*, *J. Proteome Res.* 2005, 4, 1522–1537.
- [8] Lefler, D. M., Pafford, R. G., Black, N. A., Raymond, J. R., Arthur, J. M., *J. Proteome Res.* 2004, 3, 1254–1260.
- [9] Brasier, A. R., Lu, M., Hai, T., Lu, Y., Boldogh, I., *J. Biol. Chem.* 2001, 276, 32080–32093.
- [10] Zhang, W., Chait, B. T., *Anal. Chem.* 2000, 72, 2482–2489.
- [11] Jamaluddin, M., Meng, T., Sun, J., Boldogh, I. *et al.*, *Mol. Endocrinol.* 2000, 14, 99–113.
- [12] Benel, L., Ronot, X., Mounolou, J. C., Gaudemer, F., Adolphe, M., *Basic Appl. Histochem.* 1989, 33, 71–80.
- [13] Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J. *et al.*, *Genome Biol.* 2003, 4, R60.
- [14] Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A. *et al.*, *Nucleic Acids Res.* 2005, 33, D201–D205.
- [15] Bennett, C. R., Jr., Hamre, D., *J. Infect. Dis.* 1962, 110, 8–16.
- [16] Garofalo, R., Sabry, M., Jamaluddin, M., Yu, R. K. *et al.*, *J. Virol.* 1996, 70, 8773–8781.
- [17] Jamaluddin, M., Garofalo, R., Ogra, P. L., Brasier, A. R., *J. Virol.* 1996, 70, 1554–1563.
- [18] Casola, A., Garofalo, R. P., Haeberle, H., Elliott, T. F. *et al.*, *J. Virol.* 2000, 75, 6428–6439.
- [19] Han, Y., Brasier, A. R., *J. Biol. Chem.* 1997, 272, 9823–9830.
- [20] Poot, M., Zhang, Y. Z., Kramer, J. A., Wells, K. S. *et al.*, *J. Histochem. Cytochem.* 1996, 44, 1363–1372.
- [21] Blake, J. A., Harris, A. M., in: Baxevanis, A. D., Davison, D. B., Page, R., Stormo, G., Stein, L. (Eds.), *Current Protocols in Bioinformatics*, Wiley and Sons, New York 2003.
- [22] Hosack, D. A., Dennis, G., Sherman, B. T., Lane, H. C., Lempicki, R. A., *Genome Biol.* 2003, 4, R70.
- [23] Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O. *et al.*, *Nat. Cell Biol.* 2004, 6, 97–105.
- [24] Ku, N. O., Azhar, S., Omary, M. B., *J. Biol. Chem.* 2002, 277, 10775–10782.
- [25] Inada, H., Izawa, I., Nishizawa, M., Fujita, E. *et al.*, *J. Cell Biol.* 2001, 155, 415–426.
- [26] Collart, F. R., Huberman, E., *J. Biol. Chem.* 1988, 263, 15769–15772.
- [27] D'Andrea, L. D., Regan, L., *Trends Biochem. Sci.* 2003, 28, 655–662.
- [28] Dunder, M., Misteli, T., *Biochem. J.* 2001, 356, 297–310.
- [29] Sutherland, H. G. E., Mumford, G. K., Newton, K., Ford, L. V. *et al.*, *Hum. Mol. Genet.* 2001, 10, 1995–2011.
- [30] Rando, O. J., Zhao, K., Crabtree, G. R., *Trends Cell Biol.* 2000, 10, 92–97.
- [31] Schirmer, E. C., Florens, L., Guan, T., Yates, J. R. III, Gerace, L., *Science* 2003, 301, 1380–1382.
- [32] Shen, C., Nathan, C., *Mol. Med.* 2002, 8, 95–102.

- [33] Sakurai, A., Yuasa, K., Shoji, Y., Himeno, S. *et al.*, *J. Cell. Physiol.* 2004, *198*, 22–30.
- [34] Gruenbaum, Y., Margalit, A., Goldman, R. D., Shumaker, D. K., Wilson, K. L., *Nat. Rev. Mol. Cell Biol.* 2005, *6*, 21–31.
- [35] Schlage, W. K., Bulles, H., Friedrichs, D., Kuhn, M., Teredesai, A., *Toxicol. Pathol.* 1998, *26*, 324–3343.
- [36] Stosiek, P., Kasper, M., Moll, R., *Virchows Archiv.* 1992, *421*, 133–141.
- [37] Domachowske, J. B., Bonville, C. A., Rosenberg, H. F., *J. Infect. Dis.* 2000, *182*, 1022–1028.
- [38] DiDonato, J., Mercurio, F., Rosette, C., Wu-Li, J. *et al.*, *Mol. Cell Biol.* 1996, *16*, 1295–1304.
- [39] Gardner, R. G., Nelson, Z. W., Gottschling, D. E., *Cell* 2005, *120*, 803–815.
- [40] Gillette, T. G., Gonzalez, F., Delahodde, A., Johnston, S. A., Kodadek, T., *Proc. Natl. Acad. Sci. USA* 2004, *101*, 5904–5909.
- [41] Beutler, B., *J. Invest. Med.* 1995, *43*, 227–235.
- [42] Brasier, A. R., Li, J., *J. Hypertens.* 1996, *27*, 465–475.
- [43] Gabay, C., Kushner, I., *N. Engl. J. Med.* 1999, *340*, 448–454.
- [44] Raingeaud, J., Gupta, S., Rogers, J. S., Dickens, M. *et al.*, *J. Biol. Chem.* 1995, *270*, 7420–7426.
- [45] Hsu, H., Shu, H.-B., Pan, M.-G., Goeddel, D. V., *Cell* 1996, *84*, 299–308.
- [46] Tian, B., Nowak, D., Brasier, A. R., *BMC Genom.* 2005, *6*, 137.
- [47] Tian, B., Nowak, D. E., Jamaluddin, M., Wang, S., Brasier, A. R., *J. Biol. Chem.* 2005, *280*, 17435–17448.
- [48] Beg, A. A., Baldwin, A. S. J., *Genes Dev.* 1993, *7*, 2064–2070.
- [49] Karin, M., *J. Biol. Chem.* 1999, *274*, 27339–27342.
- [50] Yie, J., Senger, K., Thanos, D., *Proc. Natl. Acad. Sci. USA* 1999, *96*, 13108–13113.
- [51] Yamamoto, Y., Verma, U. N., Prajapati, S., Kwak, Y. T., Gaynor, R. B., *Nature* 2003, *423*, 655–659.
- [52] Anest, V., Hanson, J. L., Cogswell, P. C., Steinbrecher, K. A. *et al.*, *Nature* 2003, *423*, 659–663.
- [53] Jin, D.-Y., Chae, H. Z., Rhee, S. G., Jeang, K.-T., *J. Biol. Chem.* 1997, *272*, 30952–30961.
- [54] Yedavalli, V. S., Neuveut, C., Chi, Y. H., Kleiman, L., Jeang, K. T., *Cell* 2004, *119*, 381–392.
- [55] Johnston, I. M. P., Spence, H. J., Winnie, J. N., McGarry, L. *et al.*, *Oncogene* 2000, *19*, 5348–5358.