Accelerated Discovery of Novel Protein Function in Cultured Human Cells*†

Emily Hodges, Jenny Stjernedahl Redelius, Weilin Wu, and Christer Höög‡

Experimental approaches that enable direct investigation of human protein function are necessary for comprehensive annotation of the human proteome. We introduce a cell-based platform for rapid and unbiased functional annotation of undercharacterized human proteins. Utilizing a library of antibody biomarkers, the full-length proteins are investigated by tracking phenotypic changes caused by overexpression in human cell lines. We combine reverse transfection and immunodetection by fluorescence microscopy to facilitate this procedure at high resolution. Demonstrating the advantage of this approach, new annotations are provided for two novel proteins: 1) a membrane-bound O-acyltransferase protein (C3F) that, when overexpressed, disrupts Golgi and endosome integrity due likely to an endoplasmic reticulum-Golgi transport block and 2) a tumor marker (BC-2) that prompts a redistribution of a transcriptional silencing protein (BMI1) and a mitogen-activated protein kinase mediator (Rac1) to distinct nuclear regions that undergo chromatin compaction. Our strategy is an immediate application for directly addressing those proteins whose molecular function remains unknown. Molecular & Cellular Proteomics 4: 1319–1327, 2005.

The pursuit of whole genome annotation has led to the development of a variety of high throughput (HTP) methods with the objective to study gene function en masse and the capacity to deliver a spectrum of data ranging from transcriptional information at the RNA level to identifying interaction partners at the protein level (1, 2). HTP array platforms have evolved from cDNA and protein chips to cell-based arrays adapted for high resolution microscopy (3) automating both transient protein expression (4–6) and mRNA knock-down by siRNAs (7, 8). Bioinformatic research based on comparative genome sequence analysis and the cross-referencing of information depositories provides assistance by annotating genes through gene ontology (www.geneontology.org) (9). Beyond HTP methods profiling aspects of gene and protein expression, much work has been carried out to develop and improve methods to resolve protein function at the cellular level by focusing on protein-protein interactions that occur in vitro or in different cellular systems (10) as well as to comprehensively define the cellular distribution of all proteins (11). These collective efforts although indispensable provide only a limited assessment of protein function based on indirect associations. Instead gaining more in-depth insights requires traditional, labor-intensive, and time-consuming approaches, such as overexpression or inactivation of individual genes in cell-based systems or by using in vivo models. Here we confront these limitations by describing a practical system that allows more direct functional analysis for a large set of undercharacterized human proteins. We sought to exploit protein overexpression in cultured cells as a tool for understanding function. The exogenous expression of proteins in mammalian cells may impact cellular function by generating a gain-of-function phenotype or by perturbing specific processes, also termed “dominant-negative.” These events suggest a link between the analyzed protein and a cellular activity, a feature frequently utilized in transgenic animal studies in which exogenously added genes are overexpressed. Studies in cultured cells, however, are more amenable to HTP formats; not to mention that for the study of human proteins few alternatives exist. Therefore, we incorporated the throughput and high content capacity of reverse transfection arrays, the targeting power of antibody biomarkers, and the resolution of immunofluorescence microscopy to investigate our set of undercharacterized proteins directly within their cellular context. Our approach complements current proteomic strategies by illustrating how cell-based techniques may be applied in accelerating biological discovery and enabling improved annotation of protein function.

EXPERIMENTAL PROCEDURES

Membrane Topology Predictions and Domain Assignments—Nine methods for predicting membrane topology were consulted and displayed graphically by the SPINX web server from plots generated for each protein sequence (12). Transmembrane prediction programs included TMHMM2.0 (13) and Phobius (14). Supplementary Kyte-Doolittle plots for analyzing hydrophobicity were also generated (15).

From the Center for Genomics and Bioinformatics, Karolinska Institute, SE-171 77 Stockholm, Sweden
Received, April 26, 2005, and in revised form, June 14, 2005
Published, MCP Papers in Press, June 19, 2005, DOI 10.1074/mcp.M500117-MCP200
† The abbreviations used are: HTP, high throughput; TM, transmembrane; SP, signal peptide; MBOAT, membrane-bound O-acyltransferase; PcoG, polycomb group; ER, endoplasmic reticulum; PH3, phosphorylated Histone 3; MAPK, mitogen-activated protein kinase; siRNA, small interfering RNA; HEK, human embryonic kidney; BFA, brefeldin A; Bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-propane-1,3-diol; PFAM, Protein Families Database of Alignments and HMMs (hidden Markov models); EEA1, early endosomal antigen 1; M6PR, mannose 6-phosphate receptor; COPII, coat protein II; DAPI, 4′,6-diamidino-2-phenylindole; BC-2, breast adenocarcinoma 2 gene; H3, Histone 3; DUF, DUF, domain of unknown function.

© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

Molecular & Cellular Proteomics 4.9 1319
Transmembrane domains were recognized only if predicted by at least four of the prediction methods and could be verified by the hydrophobicity curve. N-terminal signal peptides determined by Phobius and SignalP (16) were also displayed. Curated (PFAM-A) and noncurated (PFAM-B) domains were assigned by the Protein Families Database of Alignments and HMMs (hidden Markov models) (PFAM) for each protein sequence (17). However, only the higher quality PFAM-A domains were considered.

RT-PCR and Cloning—Full-length ORFs representing human genes were cloned into mammalian expression vectors using the recombination-based Gateway™ system (Invitrogen). Gene-specific primers were designed to amplify predicted coding regions according to nucleotide sequences listed by accession number in Supplemental Table 1 along with forward and reverse primer sequences. Full-length cDNAs were obtained by PCR amplification from two different sources, either from reverse transcription of mRNA (described in more detail below) or directly from a commercially available plasmid already containing the gene (Origene and GeneCopoeia). Sense and antisense primers consisted of 18–25 nucleotides flanked by the Gateway recombination sites (attB sites) that are 31 and 30 nucleotides, forward and reverse, respectively (Supplemental Table 1). cDNA was transcribed from mRNA selected in the presence of oligo(dT)20 from human total RNA and reverse transcribed using the Thermoscript RT system according to the manufacturer’s recommendations (Invitrogen). Human total RNA isolated from human liver, fetal brain, placenta, testis, skeletal muscle, and HeLa cells (Clontech) was digested (Invitrogen). Human total RNA isolated from human liver, fetal brain, placenta, testis, skeletal muscle, and HeLa cells (Clontech) was digested with BP clonase (Invitrogen). For protein expression studies, attB-flanked PCR products into the pDONRTM 201 vector in the Biosystems). PCR products were obtained under the following cycle conditions: 95 °C for 30 s, 68 °C for 1 min after which a final extension at 15–65 °C for 30 min was added. Entry clones were generated by inserting attB-flanked PCR products into the pDONR™ 201 vector in the presence of BP clonase (Invitrogen). For protein expression studies, the pcDNA-DEST40 vector (Invitrogen), a vector for mammalian expression of C-terminal V5 and His6 fusion proteins, and the pcDNA 3.1/V5-DEST™ vector for N-terminal V5 fusions were chosen. Full-length inserts were subsequently transferred by recombination from the entry clone to the expression vector to generate expression plasmids.

RNA expression profiles were generated for every gene in this collection. RT-PCR was performed with RNA prepared from three human cell lines including HeLa, HEK293, and human umbilical vein endothelial cells, a primary endothelial cell line, to determine minimal endogenous expression of the proteins. All genes were expressed in all three cell lines excluding five that exhibited the following expression pattern: AAH21119 was not detected in human umbilical vein endothelial cells; BAB13884, NP.149112, AAH16392, and NP.065702 were not detected in any of the three cell lines.

Clone Validation—Inserts from entry plasmids were sequence-verified from 5’ and 3’ directions with sense and antisense primers targeting the pDONR vector to confirm that no frameshifts or obvious point mutations had occurred during the cloning process. Sequencing reactions were performed with DYEEnamic™ ET Terminator (Amersham Biosciences) according to recommended protocols. Full-length protein expression was verified by transfecting HEK293 and HeLa cells with pcDNA-V5 expression plasmids containing each gene. Cell lysates were analyzed by standard SDS-PAGE and Western blot procedures. V5 fusions were detected by mouse anti-V5 conjugated with horseradish peroxidase (1:1000, Invitrogen) and subsequent chemiluminescence detection (SuperSignal®West Pico, Pierce).

Array Construction—50-well silicon gaskets (CultureWell™, Grace Bio-Labs) were affixed to standard poly-l-lysine microscope slides and sterilized. Under a sterile cell culture hood, transfection mixtures were generated according to methods described previously by Silva et al. (7) with some minor alterations. Briefly 0.5 μg of plasmid DNA was diluted in EC buffer (Qiagen) and 1 μl sucrose for a final concentration of 30 ng/μl DNA and 0.4 μl sucrose in a 15-μl total solution. DNA mixtures were incubated for 5 min at room temperature with 4 μl of Enhancer solution (Qiagen). 5 μl of Effectene™ transfection reagent (Qiagen) was added, and mixtures were incubated an additional 15 min at room temperature. Finally an aqueous gelatin solution (gelatin type B, Sigma) was added to the mixture to achieve a 0.22% gelatin concentration in a 45-μl total volume. We found a 10:1 ratio of lipid:DNA to be the optimal ratio for achieving the highest transfection efficiency. However, to reduce cell toxicity, solutions were diluted 1:4 in 0.22% gelatin. 2 μl were added to each well on the slide, and slides were allowed to dry overnight in the cell culture hood.

Cell Culture and Reverse Transfection—HeLa and HEK293 cells (ATCC) were maintained at 37 °C, 5% CO2 in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum and 1000 units/ml penicillin/streptomycin (Invitrogen). Cells were cultured in 75-cm2 T-flasks and were allowed to reach 90–95% confluency just prior to transfection. On the day of transfection, cells were passaged with 0.25% trypsin and were resuspended in 10 ml of media without antibiotics. 5 ml of the cell suspension were diluted in 5 ml of media. 12 μl were added to each well on the slide being careful not to disturb the surface. Slides were incubated at 37 °C at 5% CO2, for 2 h to allow cells to attach after which wells were covered with more media. These measures were taken to prevent well-to-well contamination. Cells were allowed to grow for 48 h.

Fixation and Post-transfection Processing—Cells were fixed and permeabilized for 15 min at room temperature in 2% paraformaldehyde, 1.6% sucrose, 0.5% Triton X-100 or by methanol:acetone at −20 °C. After fixation, slides were rinsed three times in PBS before incubating in blocking buffer (1% BSA in PBS) for 30 min at room temperature. Exogenously expressed proteins were detected by staining with two different fluorescein-conjugated anti-V5 antibodies derived from mouse (Invitrogen) and goat (Bethyl Laboratories). The antibodies were diluted 1:500 (mouse) and 1:800 (goat) in blocking buffer. In some instances, a third primary mouse anti-V5 antibody (Invitrogen) without any dye conjugation was applied. For this antibody a secondary detection with goat anti-mouse conjugated with Alexa 488 (Molecular Probes, 1:1000) was necessary. Diluted antibodies were applied to the samples on the slides and were incubated at 37 °C for 1 h or at 4 °C overnight (according to the manufacturer’s recommendations) before washing in PBS. Samples were costained with various organelle-specific, cytoskeletal, and phosphospecific antibodies. All antibodies are listed in Supplemental Table 2 along with details regarding dilutions and manufacturers. Antibodies derived from rabbit were detected with donkey anti-rabbit Cy3 (Jackson Laboratories, 1:1200). Those derived from mouse were detected with donkey anti-mouse Cy3 (Jackson Laboratories, 1:1200). Coverslips (22 × 50 mm) were mounted on slides with Prolong Anti-Fade (Molecular Probes) mounting medium. Slides were viewed by Leica DMRA2 and DMRXA microscopes and 100× objectives with epifluorescence, and images were captured with Hamamatsu digital charge-coupled device camera C4742-95 and Openlab™ software version 3.1.4.

siRNA Experiments—siRNAs were designed and generated according to methods described previously (18). Transfection procedures were also described previously. The following sequences were applied for targeting the C3F transcript: C3F sense siRNA strand, 5′-UUCUCUUGCCUCUGACAAAT-3′; C3F antisense siRNA strand, 3′-TTAAGGACAGAGACGUGU-5′.

Brefeldin A (BFA) Experiments—Cells were grown overnight on round coverslips in 24-well plates. The following day, standard trans-
fecteds were performed with Lipofectamine 2000 according to the manufacturer’s recommendations. Prior to fixation, cells were treated with 5 ng/ml brefeldin A (Sigma) added directly to the medium for 30 min at 37 °C at 5% CO₂. Coverslips were treated according to immunofluorescence staining methods.

C3F Peptide Antibody — Anti-C3F antibodies were raised in guinea pigs using four short peptides corresponding to amino acids 130–146, 160–177, 250–270, and 473–487 (GenBank™ accession number NP_005759) coupled to keyhole limpet hemocyanin (Peptide Specialty Laboratories GmbH). The individual antisera were affinity-purified on columns coupled with the corresponding peptide. For immunoblot analysis of the antibodies, 30 μl of the cell extracts were loaded on a 10-well 10% Nupage™ Bis-Tris gel (Invitrogen), separated, and blotted onto a PVDF membrane (Immobilon™, Millipore). Primary antibodies were used at a 1:100 dilution. Binding of the primary antibodies to the blot was detected using a donkey anti-guinea pig horseradish peroxidase-conjugated secondary antibody (1:10,000, Jackson Laboratories). In HeLa and HEK293 cell extracts, the peptide antibody from residues 473–487 detected a 50-kDa band corresponding with the predicted size of the protein (data not shown). Immunostaining was performed with a 1:50 dilution of the antibody subsequently detected with donkey anti-guinea pig conjugated with Cy3 (1:1000, Jackson Laboratories).

RESULTS

Sequence Identification and Data Mining — The focus of this study was to develop a framework for rapid and unbiased functional annotation of a large set of undercharacterized human proteins. We selected 46 human genes encoding proteins for which little or no biochemical description has been produced (Table I). The proteins listed in Table I are conserved between several eukaryotic genomes, including Homo sapiens, Caenorhabditis elegans, and Drosophila melanogaster, suggesting they share a conserved function. We began the process of functional annotation by assigning putative domain family homology and by predicting membrane topology. Analysis of the domain organization and domain sequence homology by PFAM (www.sanger.ac.uk/Software/Pfam/) (11, 17) allowed us to assign a PFAM domain family to 44 of the 46 proteins (Table I). In a few instances, more than one PFAM domain was identified within a protein sequence. Domain families are referred to by name or an abbreviation derived from a functional association. A “DUF” PFAM identification, which denotes a conserved domain of unknown function, was retrieved for five of the proteins (Table I). Utilizing the SFINX tool (sfinx.cgb.ki.se), which displays nine different methods for predicting membrane topology and two methods for signal peptide prediction (see “Experimental Procedures” for a detailed description), we were able to estimate with confidence topological features such as transmembrane (TM) helices and signal peptide (SP) cleavage sites from the amino acid sequence of the selected proteins (Table I) (12, 19). Notably 16 of the selected protein sequences contained one to nine putative transmembrane helices, whereas five contained N-terminal signal peptides.

Biomarker Selection and Array Design — To study the functions of the selected proteins at the cellular level, a full-length cDNA sequence corresponding to each of the 46 genes was inserted into mammalian expression vectors; each cDNA was fused in-frame with an N- or C-terminal short epitope tag (V5) for subsequent antibody detection. The subcellular distributions of the expressed fusion proteins were categorized in HEK293 and HeLa cells by comparison to cellular localization markers (Table I).

It has previously been observed in protein overexpression studies that some proteins may form aggregates, referred to as aggresomes (20, 21). The aggresomes accumulate at the centrosome region in interphase mammalian cell culture cells where the proteins become ubiquitinated and targeted for proteasomal degradation. As an important control, we carefully monitored the localization of our proteins relative to γ-tubulin, a centrosomal marker. We did not observe co-localization between γ-tubulin and the analyzed proteins (data not shown), strongly arguing that the localization patterns reported here are not results of artificial aggregate formation. Another potential limiting factor in overexpression studies is the appearance of artifacts arising from the expression of proteins outside of their normal cell environment. As an additional control step and to alleviate these concerns, we performed RT-PCR to establish that the investigated genes are endogenously expressed in the chosen cell lines.

Next we assembled a panel of antibody biomarkers representing assays that monitor unique subcellular structures or activities (Table II). These biomarkers serve to identify distinct cellular patterns that, if distorted, expose a potential link between the overexpressed protein and a biological process. Overexpression of a protein may trigger global consequences such as cell cycle arrest and apoptosis. These effects are not always attributed directly to the involvement of the protein in a pathway leading to this event but rather could be seen as by-products of an overloaded system. For this reason, we have chosen more targeted biomarkers that pinpoint more specific pathways rather than assaying general cell processes to distinguish broad consequences from specific ones.

To increase the number of proteins to be investigated in parallel, we customized a reverse transfection or transected cell array format by modifying and integrating protocols from techniques reported previously. Our array format facilitates screening of 50 different genes in tandem while allowing the parallel analysis of more than 500 transfected cells within the area spotted with the cDNA (depending on the cell line of interest). Briefly described, a silicon gasket containing 50 miniature wells is fitted onto a poly-L-lysine-coated microscope slide (Fig. 1). Lipid complexes containing unique plasmid constructs are suspended in a gelatin solution subsequently spotted into individual wells of the gasket, and slides are allowed to dry overnight. Afterward adherent cells are added to each well and are reverse transfected. We devised a co-observation strategy monitoring the distribution of our panel of biomarkers in either HEK293 or HeLa cells transfected in parallel with our 46 plasmid cDNAs. Cell clusters within each array coordinate were screened for changes in
biomarker patterns. The analogous expression of individual proteins on the arrayed slide with each coordinate containing a background of untransfected cells acts as a stringent internal control for nonspecific changes in biomarker appearance. Any protein eliciting phenotypic alterations in biomarker expression may be filtered through a secondary round of biomarkers targeting more specific cellular compartments or pathways to further narrow the function of the protein. Here is the table:

### TABLE I

**Summary of human proteins included in this study**

The RefSeq/DDBJ/EMBL/GenBank™ accession numbers are given in the far left column. Putative topological features such as TM helices and SP cleavage sites predicted from the amino acid sequence of the selected proteins are given. PFAM-A domains are listed and described according to the domains assigned by PFAM. Putative locations have been categorized in HEK293 and HeLa cells and described according to subcellular location results determined in this study using standard Gene Ontology terms. Specific locations are given only in cases where fusion proteins were successfully costained with a specific antibody. PH, pleckstrin homology; GAP, GTPase-activating protein; SH3, Src homology 3; CHCH, coiled coil-helix-coiled helix; zf, zinc finger; DAGAT, diacylglycerol acyltransferase; SAM, sterile alpha motif; CLPTM1, cleft lip and palate transmembrane protein 1; GDPD, glycerophosphoryl diester phosphodiesterase; NIF, NLI interacting factor-like phosphatase.

| Accession number | Topological features | PFAM identifier | Domain description | Putative cellular component with antibody costain |
|------------------|---------------------|----------------|-------------------|-----------------------------------------------|
| AAC19158         | 1 TM, SP            | PF00929        | Exonuc_X-T        | Endoplasmic reticulum, nucleus                |
| AAH02512         | 8 TM                | PF03062        | MBOAT             | Cellular component unknown                    |
| AAH04456         | 0                   | PF00929        | Exonuc_X-T        | Endoplasmic reticulum                         |
| AAH15018         | 0                   | PF00169        | PH                | Cellular component unknown                    |
| AAH16392         | 0                   | PF06920        | Ded_cytol         | Cytoplasmal vesicle, endoplasmic reticulum (partial) |
| AAH21119         | 0                   | PF00616        | RasGAP            | Perinuclear region                            |
| BAB13884         | 4 TM                | None           | None              | Endoplasmic reticulum                         |
| NP_003016        | 0                   | PF00018        | SH3, BAR          | Microtubule cytoskeleton, actin (partial) (α-tubulin) |
| NP_005799        | 9 TM                | PF03062        | MBOAT             | Endoplasmic reticulum                         |
| NP_005799        | 0                   | PF03031        | NIF               | Actin cytoskeleton, plasma membrane (Annexin II, β-actin) |
| NP_006402        | 2 TM, SP            | PF01553        | Acyltransferase   | Endoplasmic reticulum                         |
| NP_006867        | SP                  | PF01501        | Glyco tranf 8     | Endoplasmic reticulum                         |
| NP_149112        | 0                   | PF00023        | Ank               | Endoplasmic reticulum (partial), late endosome |
| NP_055268        | 0                   | PF03357        | SNF7              | Nucleoplasm, nuclear body, cytoplasm          |
| NP_055423        | 0                   | PF05994        | FragX_IP          | Cytoplasm, endoplasmic reticulum (partial)    |
| NP_055640        | 0                   | None           | None              | Endoplasmic reticulum                         |
| NP_055672        | 0                   | PF00566        | TBC               | Cellular component unknown                    |
| NP_055777        | 0                   | PF00651        | BTB               | Cytoplasm, endoplasmic reticulum (partial)    |
| NP_056158        | 1 TM, SP            | PF03031        | NIF               | Cytoplasmal vesicle                           |
| NP_057083        | 7 TM                | PF03006        | HlyIII            | Golgi apparatus (partial)                     |
| NP_057146        | 0                   | PF01883        | DUF59             | Endoplasmic reticulum (partial), nucleus      |
| NP_057163        | 0                   | PF03357        | SNF7              | Cytoplasmal vesicle                           |
| NP_057223        | 0                   | PF06747        | CHCH              | Cytoplasm                                     |
| NP_057385        | 0                   | PF03029        | ATP_bind_1        | Perinuclear region                            |
| NP_057480        | 0                   | PF03031        | NIF               | Nucleus, condensed chromosomes                |
| NP_057707        | 0                   | PF07159        | DUF1394           | Nuclear body                                  |
| NP_060175        | 5 TM                | PF03006        | HlyIII            | Perinuclear region                            |
| NP_060294        | 4 TM                | PF00097        | zf-C3HC4          | Endoplasmic reticulum                         |
| NP_060357        | 0                   | PF05907        | DUF666            | Cytoplasmal vesicle, endoplasmic reticulum (partial) |
| NP_060393        | 2 TM                | PF00097        | DUF686            | Cytoplasmal vesicle, endoplasmic reticulum (partial) |
| NP_060536        | 0                   | PF03029        | ATP_bind_1        | Cytoplasmal vesicle                           |
| NP_060579        | 0                   | PF03828        | PAP_assoc         | Endoplasmic reticulum                         |
| NP_060791        | 0                   | PF04427        | Brix              | Nucleolus                                     |
| NP_062539        | 0                   | PF03009        | GDPD              | Nucleus, cytoplasm                            |
| NP_065702        | 0                   | PF04710        | Pellino           | Cellular component unknown                    |
| NP_065946        | 1 TM                | PF02893        | GRAM              | Endoplasmic reticulum                         |
| NP_073736        | 9 TM                | PF03062        | MBOAT             | Endoplasmic reticulum                         |
| NP_079374        | 1 TM                | PF03982        | DAGAT             | Cellular component unknown                    |
| NP_079514        | 0                   | PF00651        | BTB               | Cytoplasmal vesicle                           |
| NP_110409        | 6 TM                | PF05602        | CLPTM1            | Cytoplasm                                     |
| NP_115607        | 2 TM                | PF01883        | DUF59             | Cytoplasm                                     |
| NP_115683        | 0                   | PF01472        | PUA               | Endoplasmic reticulum (partial)               |
| NP_219484        | 0                   | PF04427        | Brix              | Nucleolus, nucleoplasm, nuclear body          |
| NP_671512        | 5 TM                | PF00536        | SAM_1             | Endoplasmic reticulum, Golgi apparatus        |
| NP_775746        | 0                   | PF03556        | DUF298            | Cellular component unknown                    |
| NP_848545        | 2 TM                | PF00097        | zf-C3HC4          | Nuclear lamina, endoplasmic reticulum          |
we describe a few examples of unique changes in biomarker expression revealed during a set of initial screens, which provide new insights into important cellular roles for two novel human proteins.

The Endoplasmic Reticulum (ER) Protein C3F Impacts Golgi Assembly—We identified one candidate (NP_005759, also referred to as C3F) from our gene collection that severely disrupts the distribution of Golgi matrix biomarker GM130 when overexpressed in both HeLa and HEK293 cells (Fig. 2, top row). C3F also induces complete cytosolic dispersal of biomarkers for early (early endosomal antigen 1 (EEA1)) (Fig. 2, bottom row) and late (mannose 6-phosphate receptor (M6PR)) endosomes (data not shown) that normally form peripheral vesicles. The overexpressed C3F is found to accumulate in the ER (Fig. 3A). To substantiate the ER localization of C3F, we generated an antibody against a peptide sequence derived from C3F. The peptide antibody detected a 50-kDa protein in HeLa and HEK293 total protein extracts and distinctly labeled the ER in these cells (Fig. 3B). By comparison to cell clusters containing the other 45 cDNAs listed in Table I (including 20 overexpressed proteins found to accumulate in the ER), the C3F-induced redistributions of GM130, EEA1, and M6PR are unique events. Importantly a protein related to C3F by membrane-bound O-acyltransferase (MBOAT) domain homology (NP_073736 or MG61) exhibited no effects on Golgi structure despite localization to the ER (Fig. 4), demonstrating further that the impact on Golgi assembly is specific
for C3F and not necessarily characteristic for all MBOAT proteins. The remaining biomarkers listed in Table II were unaffected by C3F overexpression, indicating that the concomitant loss of Golgi and endosome integrity represents a specific dominant-negative phenotype. siRNA knock-down of the C3F transcript offered another line of evidence complementing these data (Fig. 5). Reduction in C3F mRNA levels resulted in severe fragmentation of the Golgi structure when monitored by GM130, a "loss-of-function" phenotype supporting the dominant-negative observation. To define more precisely how C3F regulates Golgi integrity, we monitored the distribution of a secondary set of biomarkers representing ER to Golgi transport (COPII), medial Golgi (CTR433), and Golgi to ER transport (GS28). We found that overexpression of C3F led to a loss of both anterograde and retrograde transport between ER and Golgi as all three analyzed markers collapsed into a similar diffuse cytosolic or juxtanuclear vesicle pattern (Fig. 6A). For further diagnostic purposes, we compared the C3F phenotype to wild-type cells treated with BFA, which blocks Golgi to ER transport (22, 23). Whereas COPII coatamer vesicles accumulate in the ER-Golgi intermediate compartment (24) of cells treated with BFA (Fig. 6B), C3F overexpression causes a complete cytosolic dispersal of COPII (Fig. 6A). This contrast strongly suggests that C3F affects...
an early step in COPII transport perhaps even prior to ER exit. The resulting impairment of anterograde transport will induce ER absorption of Golgi proteins as shown previously (25).

**BC-2 Prompts Rac1 Translocation, BMI1 Accumulation, and Elevated Histone 3 Phosphorylation at Nuclear Sites of Overexpression**—The overexpressed putative breast adenocarcinoma 2 gene (BC-2, GenBank™ accession number NP_055268) displays three distinct distribution patterns: cytoplasmic, diffuse nuclear, and nuclear foci (not associated with nucleoli or other known nuclear bodies). BC-2 overexpression triggered a change in the expression pattern of two different biomarkers including Rac1, a small GTPase of the Ras family, and phosphorylated Histone 3 (PH3), a marker for mitotic chromosomes (Fig. 7, A and B). In response to BC-2, a general nuclear translocation of Rac1 occurred that overlapped with nuclear BC-2 foci in HEK293 cells (Fig. 7A). Rac1 normally displays a cytoplasmic staining predominantly associated with the plasma membrane and filamentous structures extending from the perinuclear region (see also non-transfected cells in Fig. 7A). In addition, enhanced DAPI staining surrounding BC-2-positive foci coincides with Rac1 sites of enrichment. Likewise substantial accumulation of PH3 occurred in interphase cells at similar sites labeled by BC-2 (Fig. 7B), suggesting that BC-2 overexpression induces local chromatin compaction, an interpretation supported by the enhanced local DAPI staining shown in Fig. 7A. Similarly overexpression of chromatin modifying protein 1; charged multivesicular body protein (CHMP1), the closest human relative of BC-2, has been shown to induce the formation of nuclear foci that are PH3-positive and heavily H3-acetylated and to which the Polycomb group (PcG) protein BMI1 is recruited (26). When associated with chromatin, PcG proteins form repressive complexes targeting genes important for cell cycle control, cell proliferation, and apoptosis (27). For this reason, CHMP1 has been implicated in local gene-silencing events. Although we did not observe an increased acetylation affecting H3 (Fig. 8A), we did find that overexpression of BC-2 recruits BMI1 to nuclear foci (Fig. 8B), indicating that BC-2 participates in local chromatin modification events, possibly resulting in gene silencing.

**DISCUSSION**

We have introduced a simple, flexible, and objective system for functional annotation of novel human proteins. The transfected cell array format represents a robust platform for monitoring cellular events caused by protein overexpression. Although the previously reported cell microarrays permit gene and protein analysis on a much larger scale than what we present here, we believe our approach provides improvements or advantages that are immediate, practical, and readily applicable. One advantage our system has over cell microarrays involves surface area. Miniaturization becomes a disadvantage because the spots cover such a small surface area that an insufficient number of cells are transfected locally to be statistically informative. Moreover not all proteins are expressed equally well, so it becomes tedious to standardize expression for the entire array. One way to circumvent this was addressed by the spotting method developed by Silva et al. (7). Instead of spotting once with an individual lipid-DNA complex, the robotic arrayer spots the mixture nine times in close proximity to form a larger transfection area with a diameter that is 4-fold larger than the spots on the original Sabatini array, which were about 120–150 μm in diameter and contained 30–80 fluorescent cells. In our system, each well is 4 mm in diameter and may accommodate up to 1000 cells at 100% confluence (depending on the cell type, in this case HeLa cells were used), which increases the opportu-
nity for transfection and improves the efficiency of gene delivery.

Additionally our selected panel of well characterized bio-

markers offers a high resolution tool for analysis by immuno-

fluorescence microscopy. The use of biomarkers that repre-

sent a wide range of basic cell processes enables

comprehensive experimental screening with thorough cellular

coverage. Through our cell-based proteomic approach, we

have demonstrated that it is possible to "refine" the functional

annotation of novel proteins as well as to accelerate the

process of discovery. We observed three varieties of pheno-
typic changes associated with overexpression that were

pointed by our biomarkers: organelle disassembly (GM130),

protein translocation (Rac1), and post-translational modifi-
cations (PH3). Subsequently we performed several follow-up
experiments to confirm our results. As a result, we have
identified new tentative biological roles for two previously
uncharacterized human proteins.

The C3F membrane topology and sequence analysis sug-
gest a protein with a single MBOAT domain and multiple TMs
in close proximity spanning most of the protein sequence. Members of the MBOAT family exhibit similar sequence to-
polologies and are responsible for post-translational lipid mod-
fications of proteins (28). The lipid moieties are essential for
membrane tethering and secretion. In Drosophila, MBOAT
members Porcupine and Rasp have also been shown to re-
side in the ER and to be responsible for the palmitoylation of
the secreted proteins Wnt (Wingless, Wg) and Hedgehog (Hh),
respectively (29). Mutations in Porcupine and Rasp selectively
influence aspects of Wg and Hh secretion, ultimately leading
to aberrant developmental phenotypes. Importantly we have
observed that overexpression of NP_073736 (MG61), the
MBOAT-containing putative human homologue of Porcupine,
does not affect Golgi integrity in cultured cells. By contrast,
the loss of Golgi and endosome structures caused by C3F
overexpression and reiterated by siRNA experiments would
eventually lead to cell cycle arrest and cell death, implying a
more global importance for C3F in the cell. Overproduction of
the C3F protein may result in consumption of as yet uniden-
tified interacting partners necessary for COPII vesicle forma-
tion and/or the ER exit machinery. Alternatively an increase in
the putative acyltransferase activity of C3F could affect target
proteins in the ER secretory compartment, affecting the dy-
namics of ER exit and ER to Golgi transport. Collectively these
findings point to an essential cellular role for C3F in the ER.

BC-2 overexpression generated numerous intense nuclear
foci representing domains of locally condensed chromatin
identified by DAPI and PH3 while also inducing nuclear trans-
location and recruitment of Rac1 to these sites. Furthermore
BC-2 recruited PcG complex protein BMI1 to the same nu-
clear foci, strongly suggesting that BC-2 may take part in local
gene silencing. Unexpectedly Rac1 could play several roles in
nuclear foci generated by BC-2 overexpression. This GTPase
interacts with SmgGDS, a guanine nucleotide exchange fac-
tor that shuttles between the cytoplasm and nucleus, facil-
itating the nuclear import of Rac1 (30). SmgGDS has been
shown to be indirectly associated with a member of the struc-
tural maintenance of chromosomes (SMC) family of condensins
(human chromatin-associated protein) responsible for the stability and structural maintenance of mitotic chro-
mosomes (31). The condensins could participate in the local
chromatin compaction seen in foci accumulating BC-2. Rac1
also mediates upstream events in the p38/MAPK pathway
leading to the activation of downstream mitogen and stress
kinases Msx1 and Msx2, modulators of H3 phosphorylation
(32). Another downstream MAPK family kinase, MAPK-acti-

vated protein kinase 3, interacts with PcG proteins including
BMI1 and potentially regulates phosphorylation-dependent
PcG association with chromatin (33). This pathway intersec-
ting Rac1 signaling, chromatin condensation, and epi-
genetic control of gene expression provides us with new
avenues for exploring the contribution of BC-2 overexpres-
sion to tumor development.

Our cell-based approach is intended to assist the process
of functional annotation for uncharacterized proteins by ac-
celerating the discovery process through the assays we have
described. Therefore, as a final point, we must consider how
and through which forum the observations that we present
should be reported to contribute to community resources. The
Mouse Genome Informatics database (Jackson Laboratories),
a member of the Gene Ontology Consortium (34), offers an
open resource called the Mammalian Phenotype browser (35)
allowing users to browse vocabulary terms (referred to as
Mammalian Phenotype Ontology) tailored specifically to de-
scribe and compare phenotypic observations derived from ab-
normal genetic input. Although the terms were created with
the mouse model in mind, the term structure is amenable for cell-
based overexpression and knock-down data and could be
seen as a potential strategy for describing data from our
assays. The following suggestion of annotation terms (in
hierarchical order) from the Mammalian Phenotype browser
could be appropriate for describing BC-2 as an example:
Phenotype ontology; Cellular phenotype; Abnormal cell con-
tent/morphology; Abnormal nucleus morphology; Abnormal
chromosome morphology.

This controlled vocabulary is organized in a hierarchical
structure by how broad or narrow the annotation thus indi-
cating its level of completeness with each term connected to
a parent term and an accession number. By carefully selecting
the appropriate ontological mapping for a given observation,
we may establish a relationship between the gene and other
markers providing potential leads for further investigation.
Complying with the proposed Gene Ontology standards for
describing our data will expedite the functional annotation
process by unifying biological information making it search-
able, well defined, and well classified and making more obvi-
ous the relationships between genes, gene products, their
cellular components, and biological processes.
Acknowledgments—We thank Margareta Faxon and Susanne Stier for mRNA expression analysis, Boris Lenhard and Bill Wilson for bioinformatic analysis and data integration, Danielle Kemmer for providing several of the original clones, and Claes Wahlestedt for valuable collaboration.

* This work was supported by grants from Pfizer, the Swedish Graduate School for Functional Genomics and Bioinformatics, and the Karolinska Institute. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

S The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

‡ To whom correspondence should be addressed. Tel.: 46-8-5248-2488–2490; Fax 46-8-6548–6552; E-mail: christer.hoog@cmb.ki.se.

REFERENCES

1. Ashurst, J. L., and Collins, J. E. (2003) Gene annotation: prediction and testing. *Annu. Rev. Genomics Hum. Genet.* 4, 69–88

2. de Hoog, C. L., and Mann, M. (2004) Proteomics. *Annu. Rev. Genomics Hum. Genet.* 5, 267–293

3. Ziouzenik, D., and Sabatini, D. M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411, 107–110

4. Redmond, T. M., Ren, X., Kubish, G., Atkins, S., Low, S., and Uhler, M. D. (2004) Microarray transfection analysis of transcriptional regulation by cAMP-dependent protein kinase. *Mol. Cell. Proteomics* 3, 770–779

5. Conrad, C., Erfe, H., Warrant, P., Daigle, N., Lorch, T., Ellenberg, J., Pepperkok, R., and Ellis, R. (2004) Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res.* 14, 1130–1136

6. Starkviienei, V., Liebel, U., Simpson, J. C., Erfe, H., Poutakta, A., Wiemann, S., and Pepperkok, R. (2004) High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic. *Genome Res.* 14, 1948–1956

7. Silva, J. M., Mizuno, H., Brady, A., Lucito, R., and Hannon, G. J. (2004) RNA interference microarrays: high-throughput loss-of-function genetics in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6548–6552

8. Mousses, S., Caplen, N. J., Cornelison, R., Weaver, D., Basik, M., Hautaniemi, S., Elkahlon, A. G., Lotufo, R. A., Choudary, A., Dougherty, E. R., Suh, E., and Kallioniemi, O. (2003) RNA microarray analysis in cultured mammalian cells. *Genome Res.* 13, 2341–2347

9. Chalmel, F., Lardenois, A., Thompson, J. D., Muller, J., Sahel, J. A., Levieillard, T., and Poch, O. (2005) GOAnno: GO annotation based on multiple alignment. *Bioinformatics* 21, 2095–2096

10. Bork, P., Jensen, L. J., von Mering, C., Ramani, A. K., Lee, I., and Marcotte, E. M. (2004) Protein interaction networks from yeast to human. *Curr. Opin. Biol.* 14, 292–299

11. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O’Shea, E. K. (2003) Global analysis of protein localization in budding yeast. *Nature* 425, 686–691

12. Chalk, A. M., Wennerberg, M., and Sonnhammer, E. L. (2004) SfIx: an XML-based graphical sequence feature display in Java. *Bioinformatics* 20, 2488–2490

13. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580

14. Kall, L., Krogh, A., and Sonnhammer, E. L. (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027–1036

15. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophatic character of a protein. *J. Mol. Biol.* 157, 105–132

16. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795

17. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) The Pfam protein families database. *Nucleic Acids Res.* 32, D138–D141

18. Wu, W., Hodges, E., Redelius, J., and Hoog, C. (2004) A novel approach for evaluating the efficiency of siRNA on protein levels in cultured cells. *Nucleic Acids Res.* 32, e17

19. Sonnhammer, E. L., and Wootton, J. C. (2001) Integrated graphical analysis of protein sequence features predicted from sequence composition. *Proteins* 45, 262–273

20. Johnston, J. A., Ward, C. L., and Kopito, R. R. (1998) Aggresomes: a cellular response to misfolded proteins. *J. Cell Biol.* 143, 1863–1868

21. Kopito, R. R. (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* 10, 524–530

22. Shinotsuka, C., Yoshida, Y., Kawamoto, K., Takatsu, H., and Nakayama, K. (2002) Overexpression of an ADP-ribosylation factor-guanine nucleotide exchange factor, BIG2, uncouples brefeldin A-induced adaptor protein-1 coat dissociation and membrane tubulation. *J. Biol. Chem.* 277, 9468–9473

23. Jackson, C. L., and Casanova, J. E. (2000) Turning on ARF: the Sec7 family of guanine-nucleotide-exchange factors. *Trends Cell Biol.* 10, 60–67

24. Breuza, L., Halbeisen, R., Jeno, P., Otte, S., Barlowe, C., Hong, W., and Hauri, H. P. (2004) Proteomics of endoplasmic reticulum-Golgi intermediate compartment (ERGIC) membranes from brefeldin A-treated HepG2 cells identifies ERGIC-32, a new cycling protein that interacts with human ERV6. *J. Biol. Chem.* 279, 47242–47253

25. Ward, T. H., Polishchuk, R. S., Caplan, S., Hirschberg, K., and Lippincott-Schwartz, J. (2001) Maintenance of Golgi structure and function depends on the integrity of ER export. *J. Cell Biol.* 155, 557–570

26. Stauffer, D. R., Howard, T. L., Nyun, T., and Hollenberg, S. M. (2001) CHMP1 is a novel nuclear matrix protein affecting chromatin structure and cell-cycle progression. *J. Cell Sci.* 114, 2383–2393

27. Lund, A. H., and van Lohuizen, M. (2004) Polycrom complexes and silencing mechanisms. *Curr. Opin. Cell Biol.* 16, 239–246

28. Hofmann, K. (2000) A superfamily of membrane-bound O-acetyltransferases with implications for wnt signaling. *Trends Biochem. Sci.* 25, 111–112

29. Nussle, R. (2003) Wnts and Hedgehogs: lipid-modified proteins and similarities in signaling mechanisms at the cell surface. *Development* 130, 5297–5305

30. Williams, C. L. (2003) The polybasic region of Ras and Rho family small GTPases: a regulator of protein interactions and membrane association and a site of nuclear localization signal sequences. *Cell. Signal.* 15, 1071–1080

31. Jessberger, R. (2002) The many functions of SCMC proteins in chromosome dynamics. *Nat. Rev. Mol. Cell. Biol.* 3, 767–776

32. Solaoa, A., Thomson, S., Wiggin, G. R., Rampersaud, N., Dyson, M. H., Hazzalin, C. A., Mahadevan, L. C., and Arthur, J. S. (2003) MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J.* 22, 2788–2797

33. Vonczen, J. W., Niessen, H., Neufeld, B., Rennefaht, U., Dahlmans, V., Kubben, N., Holzer, B., Ludwig, S., and Rapp, U. R. (2005) MAPKAP kinase 3 (p90) phosphorylates and regulates chromatin association of the polycrom group protein Bm1. *J. Biol. Chem.* 280, 5178–5187

34. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* 25, 24–25

35. Smith, C. L., Goldsmith, C. A., and Eppig, J. T. (2005) The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol.* 6, R7