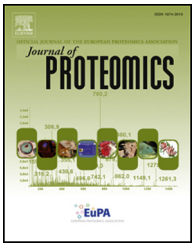
JOURNAL OF PROTEOMICS XX (2014) XXX–XXX

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot



Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature

Jordane Biarc[a,*], *Philippe Gonzalo*[b,c], *Ivan Mikaelian*[b], *Laurent Fattet*[b], *Mathieu Deygas*[b], *Germain Gillet*[b], *Jérôme Lemoine*[a,1], *Ruth Rimokh*[b,1]

[a]*UMR 5280, Institut des sciences analytiques, Université de Lyon, Lyon 1, 69622 Villeurbanne cedex, France*
[b]*CNRS UMR5286, INSERM U1052, Centre de Recherche en Cancérologie de Lyon, Université de Lyon, Centre Léon Bérard, F-69008 Lyon, France*
[c]*Laboratoire de Biochimie, Hôpitaux de Saint-Etienne, Université de Saint-Etienne, France*

ARTICLE INFO

Article history:
Received 3 March 2014
Accepted 27 May 2014

Keywords:
Epithelial–mesenchymal transition signature
Label-free quantification
Single reaction monitoring
Proteomics
Cell signaling

ABSTRACT

Disease phenotype reorganizations are the consequences of signaling pathway perturbations and protein abundance modulations. Characterizing the protein signature of a biological event allows the identification of new candidate biomarkers, new targets for treatments and selective patient therapy. The combination of discovery LC–MS/MS analyses and targeted mass spectrometry using selected reaction monitoring (SRM) mode has emerged as a powerful technology for biomarker identification and quantification owing to faster development time and multiplexing capability. The epithelial–mesenchymal transition (EMT) is a process that controls local invasion and metastasis generation by stimulating changes in adhesion and migration of cells but also in metabolic pathways. In this study, the non-transformed human breast epithelial cell line MCF10A, treated by TGFβ or overexpressing mutant K-Ras^{v12}, two EMT inducers frequently involved in cancer progression, was used to characterize protein abundance changes during an EMT event. The LC–MS/MS analysis and label-free quantification revealed that TGFβ and K-Ras^{v12} induce a similar pattern of protein regulation and that besides the expected cytoskeletal changes, a strong increase in the anabolism and energy production machinery was observed.

Biological Significance
To our knowledge, this is the first proteomic analysis combining a label-free quantification with an SRM validation of proteins regulated by TGFβ and K-Rasv12. This study reveals new insights in the characterization of the changes occurring during an epithelial–mesenchymal

Abbreviations: ABC, ammonium bicarbonate; CCT, chaperonin containing TCP-1; DTT, dithiothreitol; EMT, epithelial–mesenchymal transition; FDR, false discovery rate; IAA, iodoacetamide; K-Ras^{v12}, Kirsten rat sarcoma viral oncogene homolog mutated at valine 12; TGFβ, transforming growth factor beta.

* *Corresponding author.*
E-mail address: jordane.biarc@univ-lyon1.fr (J. Biarc).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jprot.2014.05.026
1874-3919

transition (EMT) event. Notably, a strong increase in the anabolism and energy production machinery was observed upon both EMT inducers.

1. Introduction

Local invasion can be considered as an essential step in the malignancy of carcinomas, leading to the generation of distant metastases. It appears to be controlled by a coordinated series of cellular and molecular processes that enable tumor cells to dissociate and migrate from the primary tumor. The changes in cell adhesion and migration during this event are reminiscent of an important developmental process termed epithelial–mesenchymal transition (EMT), a process that also has an active role in other stages of the metastatic cascade such as intravasation [1]. The possibility is raised that, during EMT, a crosstalk takes place between modifications of cell shape, adherence and migration, resistance to death-inducing stimuli and metabolic pathways. The precise knowledge of these modifications in cancer cells may thus lead to the characterization of new therapeutic targets.

TGFβ is known to be a potent inducer of EMT and high levels of this factor are found in a number of tumors, which is often correlated to high invasion and acquisition of metastasic properties [2] TGFβ participates in major cellular processes, such as proliferation, differentiation, migration and apoptosis, from embryogenesis to adulthood. This cytokine exerts antagonistic effects on tumor development. In the early steps of tumorigenesis, it acts as a tumor suppressor by inhibiting cell proliferation, whereas in later stages, it facilitates tumor progression by stimulating angiogenesis, immune response escape, and EMT [3]. The signaling pathways that are involved in the dual effects of TGFβ are yet to be characterized. It has been shown that mutations leading to constitutive activation of Ras signaling pathways occur in approximately 30% of cancers [4]. The mutant K-Ras^{v12} with glycine replaced by valine at codon 12, one of the most common amino acid substitutions of K-Ras found in human cancers [5], constitutes a good model of ras signaling induction. In non-transformed human breast MCF10A epithelial cells, the overexpression of this mutant gene constitutively activates ras signaling, confers characteristics of transformation, enhanced growth rate, reduced dependency on serum, and loss of contact inhibition, and induces EMT [6,7].

Despite intensive transcriptomic profiling studies, the characterization of specific signature genes remains elusive, partially because the transcriptome does not mirror the functional proteome. Some studies have compared the proteome of an epithelial and a mesenchymal breast cancer cell line [8] or the secretome of ras and ras/TGFβ mediated EMT in MDCK cells [9], but label-free quantification of proteins regulated by two EMT inducers has not been reported.

In this study, proteomic analysis was used to compare the protein profiles between the human breast epithelial cell line MCF10A, treated by TGFβ or overexpressing the mutant K-Ras^{v12}. This work has been performed by combining an LC–MS/MS analysis on a Q-exactive orbitrap and a relative quantification by MS1 filtering tool in Skyline. This label-free method allowed a fast and repeatable quantification on a large-scale pool of proteins and highlighted differences and similarities in protein regulation between these two EMT models that could be used to improve tumor characterization and selective tumor therapy.

2. Materials and methods

2.1. Cell culture

The human mammary epithelial cell line MCF10A, a spontaneous immortalized was obtained from the American Type Culture Collection. Cells were grown in DMEM/F12 medium supplemented with 5% horse serum, EGF (20 ng/mL), insulin (10 μg/mL) cholera toxin (100 ng/mL), hydrocortisone (0.5 μg/mL) and penicillin plus streptomycin. MCF10A-LXSN (control) and MCF10A-LXSN-K-Ras^{v12} cell lines, kindly provided by Ben Ho Park, have been previously described [6]. Recombinant TGFβ1 (Peprotech) was used at 5 ng/mL for 72 h. Cells were washed 2 times with phosphate-buffered saline (PBS) and lysed in 1 mL of urea 10 M before sonication for 5 min. After centrifugation (10 min, 4 °C at 14,000 × *g*), an aliquot of the supernatant was analyzed by the Bradford test to measure the concentration of protein in each sample.

2.2. F-actin staining, migration assays and immunoblotting

For immunofluorescence, cells were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS), permeabilized with 0.4% Triton X-100 in PBS for 5 min and subsequently stained with 0.25 mM tetramethyl rhodamine iso-thiocyanate-conjugated phalloidin (Sigma- Aldrich). Fluorescence was examined by confocal laser scanning microscopy (Carl Zeiss). For migration assays, cells were plated in the upper chamber of transwell plates (BD Biosciences). After 6 h (TGFb) or 2.5 h (K-Ras^{v12}) incubation, cells attached to the lower part of the filters were counted. Experiments were carried out in triplicate, and for each transwell, three microscopic fields were counted.

Lysates of cells prepared in RIPA lysis buffer were separated by SDS–polyacrylamide gel electrophoresis, electro-blotted onto polyvinylidene difluoride transfer membrane, and stained with specific primary antibodies (E-cadherin (CDH1) and N-cadherin (CDH2), BD Biosciences; vimentin, Dako; occludin, Santa Cruz; Ras, Calbiochem). Peroxidase-linked anti-mouse (P0260, Dako) and anti-rabbit (P0448, Dako) secondary antibodies and ECL detection reagents (Roche Applied Science) were used according to the manufacturer's instructions. Anti-tubulin antibody (Covalab) was used for loading controls.

2.3. Sample preparation

Each sample (1 mg) was reduced with 110 μL DTT (150 mM) for 40 min at 60 °C and alkylated with 340 μL of IAA (150 mM) in the dark for 40 min at room temperature. Proteins were diluted with 13 mL of ABC (50 mM) and digested using 200 μL trypsin (2 mg/mL, type IX-S from porcine pancreas) for 16 h at 37 °C. Digestion was stopped by adding 0.5% formic acid. After

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

a centrifugation at 13,000 × g for 10 min, peptides were separated on an Oasis® MCX 3cc (60 mg) Extraction Cartridge with 1 mL ABC 30 mM pH 8 + 20% methanol, 1 mL ammonium hydroxide 5% pH 10 + 20% methanol and finally 1 mL ammonium hydroxide 5% pH 10 + 80% methanol. The 3 MCX fractions per sample were evaporated and resuspended in 200 μL of water/0.5% formic acid for LC–MS/MS analysis (20 μL per injection, 100 μg of peptides injected for the 3 fractions).

2.4. LC–MS/MS

Mass spectrometry was performed using a Q-Exactive (Thermo Fisher Scientific, San Jose, CA). Chromatography was performed using an Agilent 1100 ultra performance liquid chromatography (Agilent, Massy, France) system (at a flow rate of 300 μL/min) on a column X-bridge™ of BEH130 C18 3.5 μm, 2.1 mm ID X 100 mm (Waters, Milford, MA), with a 120 min gradient. Solvent A was water/0.5% formic acid, and solvent B was acetonitrile/0.5% formic acid; peptides were eluted by a gradient from 5% to 30% solvent B over 113 min followed by a short wash in 95% solvent B, before returning to the starting conditions. After a precursor scan of intact peptides was measured in the orbitrap by scanning from m/z 350 –1200 (with a resolution of 35,000), the 10 most intense multiply-charged precursors were selected for HCD analysis in the C-trap (with a resolution of 17500 and an isolation window of 2.0 m/z). Normalized collision energy was set to 28.0 for HCD fragmentation. Automatic gain control (AGC) targets were 3e6 ions for orbitrap scans (max injection time 250 ms) and 2e5 for MS/MS scans (max injection time 120 ms). Dynamic exclusion for 30 s was used to reduce repeated analysis of the same components.

2.5. Peptide and protein identification

Fragmentation data were converted to peaklists using a script based on the Raw_Extract script in Xcalibur v2.4 (Thermo Fisher Scientific, San Jose, CA) ([10]) and the HCD data for each sample were searched using Protein Prospector version v 5.10.0 [11] against the Swissprot human database (downloaded March 21, 2012; 535248 entries), to which a randomized version of all entries had been concatenated. All searches used the following parameters: mass tolerances in MS and MS/MS modes were 20 ppm and 0.5 Da, respectively. Trypsin was designated as the enzyme and up to one missed cleavage was allowed. S-Carbamidomethylation of cysteine residues was designated as a fixed modification. Variable modifications considered were N-terminal acetylation, N-terminal glutamine conversion to pyroglutamate and methionine oxidation. The maximum expectation value allowed was set up at 0.01 (protein) and 0.05 (peptide). When one peptide matched multiple proteins, the results reported one member of a protein family and homologous proteins if there is at least one unique peptide matching this protein (default parameter "interesting"). At these thresholds the protein false positive rate was estimated at 0.1–1.1% for each experiment according to the concatenated database search results [12]. All spectra and searches of peptides listed in Supplementary Table 2 can be viewed using the viewer file available at http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer; (search key = MCX1_control1: **eailchn1wy**; MCX2_control1:**p7kpajp4a2**; MCX3_control1: **kfuq95rxku**; MCX1_Kras1:**fif0shqe0t**; MCX2_Kras1:**eewzr3zvio**; MCX3_Kras1:**ryttrdf2zw**; MCX1_TGFβ1:**pcrxbljvzn**; MCX2_TGFβ1: **q8r1naiymm**; MCX3_TGFβ1:**m5vwunujtr**; MCX1_control2: **sie94oaify**; MCX2_control2:**4f85tbqohp**; MCX3_control2: **8x4yuzmdkx**; MCX1_Kras2:**sl6jtcmieg**; MCX2_Kras2:**wul7nzkfah**; MCX3_Kras2:**8yyyfovfh8**; MCX1_TGFβ2:**te4bfjafto**; MCX2_TGFβ2: **nmbre5kssy**; MCX3_TGFβ2:**cblxomhaui**). They can also be directly viewed through hyperlinks provided in Supplement 4. The table displays the same columns as Supplementary Table 2.

2.6. Quantification

Label-free quantification measurements were extracted from the raw data by "MS1 filtering" tool algorithm of Skyline software (http://proteome.gs.washington.edu/software/skyline) according to [13]. Spectral libraries containing identified peptides and retention time were created from the raw data (converted into an mzXML file, MassMatrix Mass Spectrometric Data File conversion Tool version 3.9) and the peptide searches in Protein Prospector (converted into a pepXML file with Protein Prospector version v 5.10.0) with a cut-off score of 0.99. In Skyline, trypsin was designated as the enzyme, one miscleavage was allowed and the human proteome from Swissprot was used as a background proteome. The structural modification included S-carbamidomethylation of cysteine residues, N-terminal acetylation, N-terminal glutamine conversion to pyroglutamate and methionine oxidation. The filter in "transition settings" allowed the quantification of the 2,3 or 4-charged precursors (ion types = p) with 1,2 or 3 charged-ions. The "auto-select all matching transitions" box was checked. The library ion match tolerance was set up to 0.5 Th and the instrument acquired data from 50 Th to 1500 Th with a match tolerance of 0.055 Th. For MS1 filtering, the "Isotope peaks included" was set up to count and the "precursor mass analyzer" to orbitrap with a resolving power of 35,000 at 200 Th. The MS1 filtering only used scans within 1 min of MS/MS identification.

All peptides within Protein Prospector searches of the same MCX fraction in control and stimulated samples (for example, the peptides contained in MCX3 control and MCX3 stimulated) were imported and quantified together in Skyline. Only peptides with an isotope dot product > 0.95 in control and stimulated samples were considered and the maximum difference of retention time for a peptide in control and stimulated samples was set up to 2 min. The peptide ratios "control/stimulated" were calculated by dividing the area measured for each peptide in the control sample by the area measured for the same peptide in the stimulated sample. All peptide ratios in the same experiment (contained in all fractions) were gathered and the log(peptide ratios) were calculated. We assumed that the median of all log(peptide ratios) was 0 (i.e. as we assume that most of the peptides will not change upon stimulation) and all values of log(ratios) were corrected. The log(protein ratios) were measured by calculating the median of log(peptide ratios) to discard aberrant peptides or with a specific post-translational modification that would not represent protein abundance. All peptides used for protein quantification for each sample can be viewed on Supplementary Table 2.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

2.7. LC–MS/MS analysis and quantification repeatability

The MCX2 fraction of MCF10A cells overexpressing K-Ras^{v12} of experiment 1 was injected three times in the Q-exactive orbitrap, analyzed by LC–MS/MS and quantified by MS1 filtering tool in Skyline as described above in the quantification section. The coefficients of variation (CV) of each peptide area were calculated and the median CV was calculated for each protein.

2.8. SRM validation

This assay was performed by comparing the fractions MCX1, 2 and 3 of MCF10A cells expressing K-Ras^{v12} to the control sample (experiment 2). The SRM methodology was used as described in [14]. According to the Skyline spectral libraries created from raw file and Protein Prospector file, 2 intense peptides from each protein of interest were selected and the 3 best ions chosen to create 6 transitions per protein. The collision energy was optimized according to the formula: m/z*slope + intercept (with slope = 0.05 for 3+ ions or = 0.04 for 2+ ions and intercept = 4 for 3+ ions or = 5 for 2+ ions) and the retention time found in the Q-exactive experiment was used to set up the SRM scheduling. The SRM transitions were monitored and acquired at unit resolution in Q1 and Q3 at their retention time with a window of 300 s. Each MCX fraction (20 μL) was injected in a QTRAP® 5500 LC–MS/MS hybrid triple quadrupole/linear ion trap mass spectrometer (AB Sciex, Foster City, CA, USA) over the same 2 h chromatography gradient used on the Q-exactive (Thermo Fisher Scientific, San Jose, CA). Total area quantifications were performed in Skyline by adding areas of all 3 transitions per peptide and the ratios control/stimulated for proteins were determined by calculating the average of ratios measured for the 2 peptides (corrected with the same factor applied during the label-free quantification). Label-free quantification was performed as described above in the "quantification" part.

2.9. Data mining

2.9.1. GO annotations

The lists of regulated proteins were submitted to DAVID Bioinformatics Resources 6.7 [15] with their Uniprot accession number to analyze protein function regulated upon TGFβ stimulation or K-Ras^{v12} overexpression according to their membership to major categories of GO annotations.

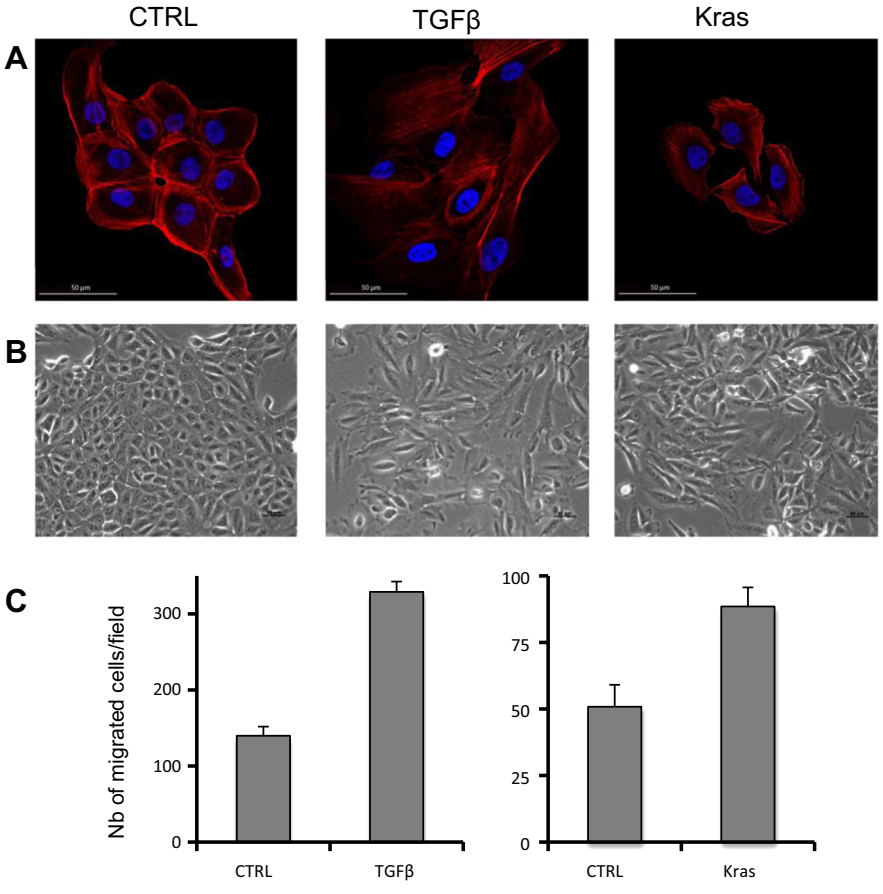


Fig. 1 – Induction of EMT in MCF10A cells treated with TGFβ or retrovirally infected by LXSN-KRas^{v12}. (A) Immunofluorescence confocal acquisitions showing actin subcellular localization detected by phalloidin. Nuclei were stained with Hoescht 33342 (blue); scale bars, 50 μm. (B) Phase contrast images of the indicated cells; scale bars, 50 μm. (C) Quantification of migrated MCF10A cells using a transwell migration assay.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

Each protein was manually assigned using the gene ontology terms found in the DAVID NIH database as previously described [16].

2.9.2. Protein network

The lists of identified proteins were submitted using their Uniprot accession numbers from HUMAN to the STRING

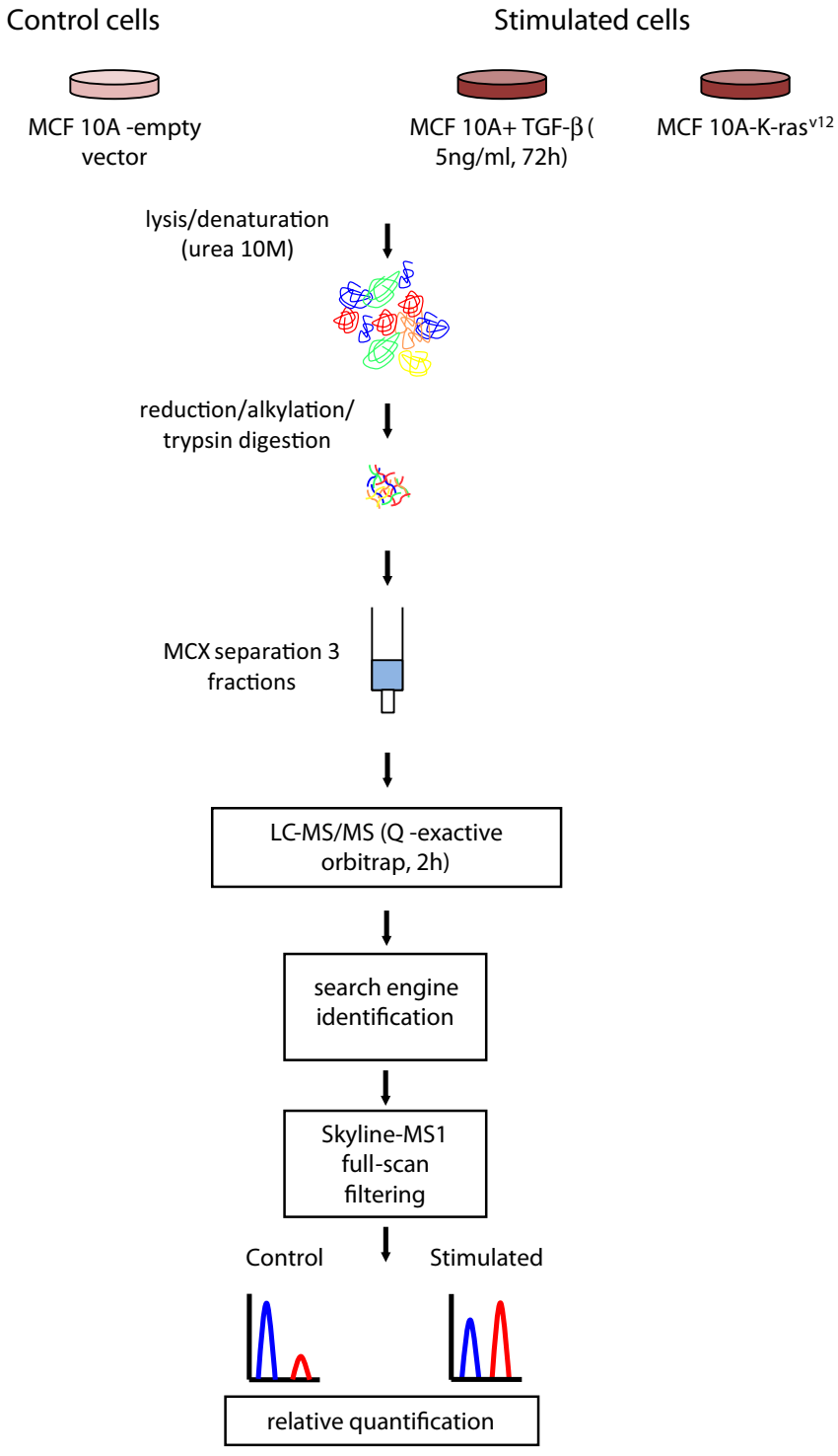


Fig. 2 – Experimental strategy. MCF10A cells infected with an empty vector were used as the "control sample" and were compared to "stimulated samples" consisting of MCF10A cells treated with TGFβ (72 h, 5 ng/mL) or overexpressing K-Ras^{v12}. The cells were lysed and 1 mg of each sample was denatured, reduced, alkylated and digested with trypsin. A separation of peptides by MCX Extraction Cartridge was performed for each sample. Samples were analyzed by LC–MS/MS for 2 h on a Q-exactive orbitrap and peptide identification was performed with Protein Prospector. Relative quantification was carried out by using MS1-filtering tool in Skyline software and log ratios (control/stimulated) proteins were reported.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

database [17] considering the combined score of experimental and database interactions. The confidence score was set to medium (0.400) for all networks and the results were visualized with Cytoscape v_2.8.2.

2.9.3. *Heatmaps*

The lists of identified proteins were hierarchically clustered according to their regulation (log [ratio protein]) upon stimulation with TGFβ or overexpressing K-Ras^{v12} by calculating the Euclidean distance using Cytoscape v_2.8.2 [18].

3. Results

3.1. *EMT models*

Human mammary epithelial MCF10A cells are commonly used to study EMT [19]. Our two models consisted of MCF10A cells treated with TGFβ (72 h, 5 ng/mL) and overexpressing mutant K-Ras^{v12} (LXSN-K-Ras^{V12}) (stimulated samples). Control cells consisted of MCF10A cells infected with the empty LXSN vector and left untreated. The characteristics of MCF10A-LXSN-K-Ras^{v12} have been previously described [6]. Along with the switch in gene expression, cells undergoing EMT are characterized by a reorganization of cortical actin into stress fibers typical of mesenchymal cells and fibroblast-like elongated morphology showing that MCF10A cells treated by TGFβ or expressing K-Ras^{v12} have undergone EMT (Fig. 1). Cell migration is an important aspect of EMT; we therefore investigated migration of the cells. We obtained evidence that MCF10A treated with TGFβ or overexpressing the mutant K-Ras^{v12} displayed increased motility compared with control (Fig. 1C).

3.2. *Quantification strategy*

All peptides identified by the LC–MS/MS analysis on the Q-exactive were quantified into skyline according to a strategy designed in Fig. 2. This analysis allowed the identification and quantification of 1415 proteins (with an average of 6 peptides/protein) in both experiments (control/TGFβ or K-Ras^{v12} experiment 1 and 2, Supplementary Tables 1 and 2). Fig. 3 presents the distribution of protein ratios measured for all samples. Proteins showing ratios over or below 2 fold (log ratio <−0.3 or >0.3) in the two experiments of each group were considered for further regulation analysis. Proteins showing ratios below 1.3 fold (log ratio <0.125 or >−0.125) in the two experiments were considered as not regulated under these conditions. Considering all proteins, the coefficient of correlation between the two biological replicates showed values of R equal to 0.5 and 0.7 for TGFβ treated cells and K-Ras^{v12} cells, respectively, indicating an average correlation between the two replicates. The median CV for protein quantification between biological replicates

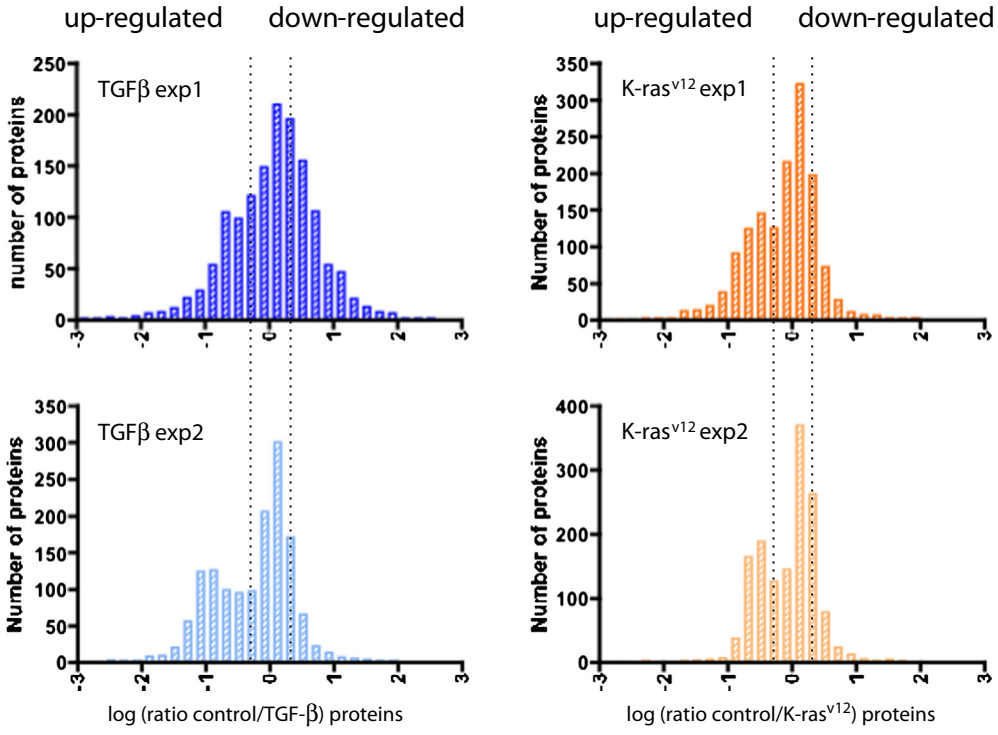


Fig. 3 – Protein abundance changes in each experiment. Distribution of log ratio (control/stimulated) proteins quantified in each experiment (MCF10A cells treated by TGFβ in two experiment replicates (TGFβ exp1 and TGFβ exp2) and MCF10A cells overexpressing K-Ras^{v12} in two experiment replicates (K-Ras^{v12} exp1 and K-Ras^{v12} exp2). Negative log(ratios) and positive log(ratios) indicate an up-regulation and a down-regulation of protein expression in stimulated samples respectively. Dotted lines (log(ratio) = 0.3 or −0.3) represent the 2 fold up or down regulation of protein abundance for each experiment.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

is 39% showing the necessity to take into account only proteins showing similar ratios in both replicates. Indeed, the correlation calculated between the two biological replicates of 2-fold or more regulated proteins showed values of R equal to 0.8 and 0.9 for TGFβ treated cells and K-Ras^{v12} cells, respectively. The biological variability can be explained by the dual and transitory effects of TGFβ. TGFβ stimulation led to the up-regulation of 300 proteins (21%), a down-regulation of 150 proteins (11%) and a constant expression of 62 proteins (4.3%) while the overexpression of K-Ras^{v12} increased the abundance of 374 proteins (26%), decreased the abundance of 143 proteins (10%) and did not regulate 113 proteins (8%). Two hundred sixty-two proteins (19%) were up-regulated and 134 proteins (10%) were down-regulated by both EMT inducers. The proteins that were not presenting a similar regulation in both replicates were not taken into account.

Another representation of the protein ratios averaged for both replicates and plotted after a hierarchical clustering on heatmaps (Supplementary Fig. 1) shows the similarities between these two EMT models. The regulation patterns of TGFβ and K-Ras^{v12} are quite similar and only 6 proteins have shown a specific regulation profile upon stimulation (>2-fold change in both replicate of one condition and <1.3-fold change in both replicate of the other condition).

3.3. Pathway analysis and protein interaction network

Functions of regulated proteins (by at least 2 fold) upon TGFβ stimulation and K-Ras^{v12} overexpression have been analyzed and compared by determining their implication in specific biological processes and cellular structures. In Fig. 4 the number of proteins categorized by major GO annotations and by their type of regulation are represented. TGFβ and K-Ras^{v12} tend to up-regulate proteins in the categories for cytosol, mitochondria, Golgi, cytoskeleton or ribonucleoprotein complexes and metabolism, glycolysis, transport, cell cycle and mitosis, apoptosis, proliferation, translation, transcription and DNA repair. However, this strong preference for up-regulation is not observed for proteins located in the nucleus, plasma membrane, endoplasmic reticulum or as part of the chromosome/chromatin structure and those involved in signaling, morphogenesis, adhesion, and RNA splicing. The

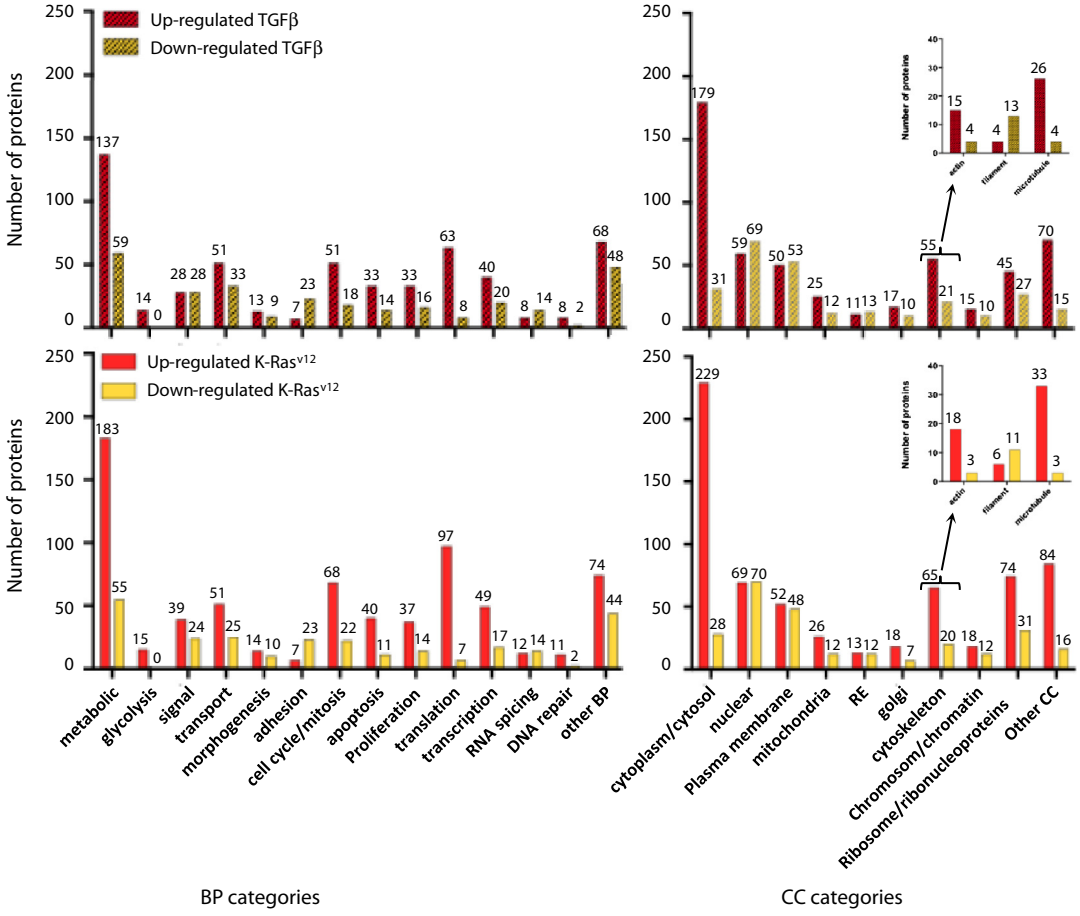


Fig. 4 – Distributions of GO annotations. Representation of the number of regulated proteins in various biological processes (BP, left panel) and cellular components (CC, right panel) annotations of the up- (red bars) and down-regulated (yellow bars) proteins identified in MCF10A cells upon stimulation with TGFβ (upper panels, dashed bars) or overexpressing K-Ras^{v12} (lower panels, plain bars). The list of identified and quantified proteins has been submitted into the DAVID Bioinformatics Resources 6.7 by manually assigning the gene ontology terms found in the database. The number of proteins is indicated above each bar. Cytoskeleton proteins are divided into 3 sections: actin, filament and microtubule and are represented in the top/right panel.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

separation of cytoskeleton proteins into three sections (actin, intermediate filaments, and microtubules including associated proteins) indicates that actin and microtubules (and associated proteins) are up-regulated by TGFβ and K-Ras^{v12} while proteins associated to intermediate filaments are mainly down-regulated.

Differences between TGFβ and K-Ras^{v12} effects were highlighted by comparing the distribution of all protein ratios averaged for both experiments and for each EMT inducer. In Supplementary Fig. 2, the distributions are presented for selected GO categories queried in the previous analysis and associated with a gaussian regression curve calculated on frequency values. These analyses were performed on GO categories with a total number of up- and down-regulated proteins for the TGFβ or K-Ras^{v12} cells equal to 30 or more. The results tend to show that TGFβ up-regulates fewer proteins but with higher ratios than K-Ras^{v12}, particularly for proteins associated with the cytoplasm, ribonucleoprotein complexes and translation (the distribution of ratios is shifted to the left compared to K-Ras^{v12}). TGFβ also appears to down-regulate more proteins and more extensively in most categories, particularly in regard to proteins involved in proliferation, apoptosis, and cell cycle/mitosis. This result is accentuated when the distributions are combined (Fig. 5A) while the ratio of down/up-regulated protein is higher for TGFβ than K-Ras^{v12} (Fig. 5B).

In order to identify common protein networks regulated by TGFβ and K-Ras^{v12} stimulation (by at least 2-fold), the STRING database, which reports several types of interactions between proteins, was used [17]. The results, visualized with Cytoscape v_2.8.2, show that several proteins forming part of the nuclear pore/spliceosome and cytoskeleton/adhesion networks are down-regulated by TGFβ and K-Ras^{v12} proteins (Fig. 6A). Proteins up-regulated by both EMT inducers are part of 4 main protein sub-networks involved in glycolysis and in protein translation, degradation and folding (ribosome structure and initiation/elongation factors, ubiquitin/proteasome complex and CCT complex) (Fig. 6B). Numerous proteins from the cytoskeleton that have not clustered together after application of the force-directed layout, which means that they are not part of a same complex, are both up- and down-regulated as already observed in the GO analysis. These biological processes and protein complexes regulated by both EMT inducers constitute the EMT signature.

3.4. *Quantification repeatability and validation*

The repeatability of LC–MS analyses and quantifications was performed by comparing the results obtained in 3 replicates of the same sample (MCX2 fraction of MCF10A cells expressing K-Ras^{v12}, experiment 2). This analysis has evaluated the variability of the chromatography separation, the MS/MS analysis and also the quantification performed by Skyline MS1 filtering tool. Proteins (1841, 1831 and 1810) were identified in samples 1, 2 and 3, respectively, and 1747 proteins in common were quantified. The median coefficient of variation for each peptide was calculated (4.6%) and the median coefficient of variation for each protein was calculated at 5.2% with 90% of the proteins presented a CV < 80%. Fig. 7 presents the number of proteins in each CV range.

In order to validate the protein quantification, an analysis by scheduled-SRM was performed, comparing the same MCF10A cells expressing K-Ras^{v12} or control cells for the label-free quantification by Skyline (experiment 2). Several proteins regulated or not by K-Ras^{v12} were selected and measured by SRM. Thus, 258 proteins were monitored in the

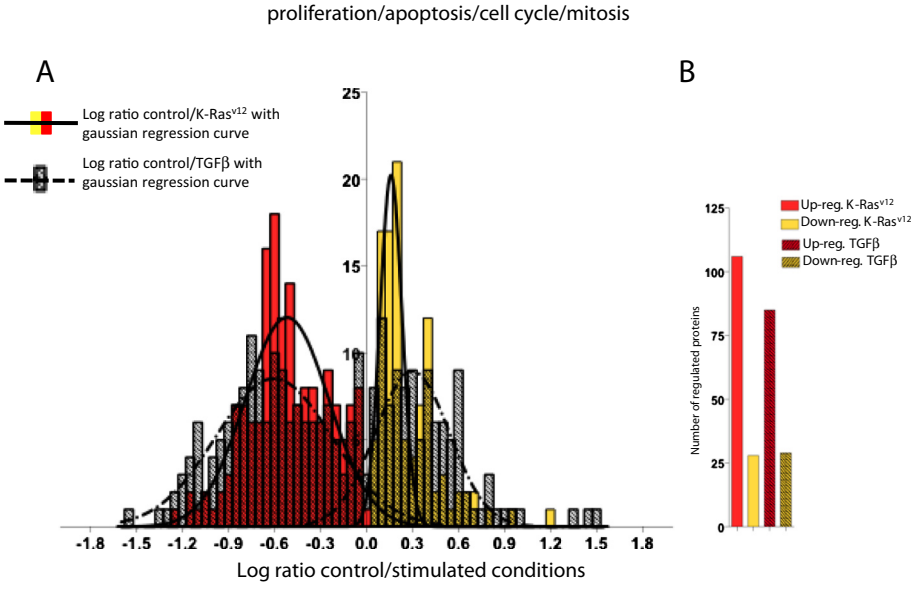


Fig. 5 – Distribution of protein ratios. (A) Protein ratio distribution (averaged for replicates) involved in biological processes proliferation/apoptosis/cell cycle and mitosis (red bars = log ratio < 0 or yellow bars = log ratio > 0) with a Gaussian regression curve calculated for TGFβ (dashed bars, dotted line) or K-Ras^{v12} (plain bars, plain line). X axis represents the ratio ranges, Y axis represents the number of proteins in each ratio range. (B) Number of proteins up-regulated (red bars) or down-regulated (yellow bars) by TGFβ (dashed bars) or K-Ras^{v12} (plain bars) involved in proliferation/apoptosis/cell cycle/mitosis.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

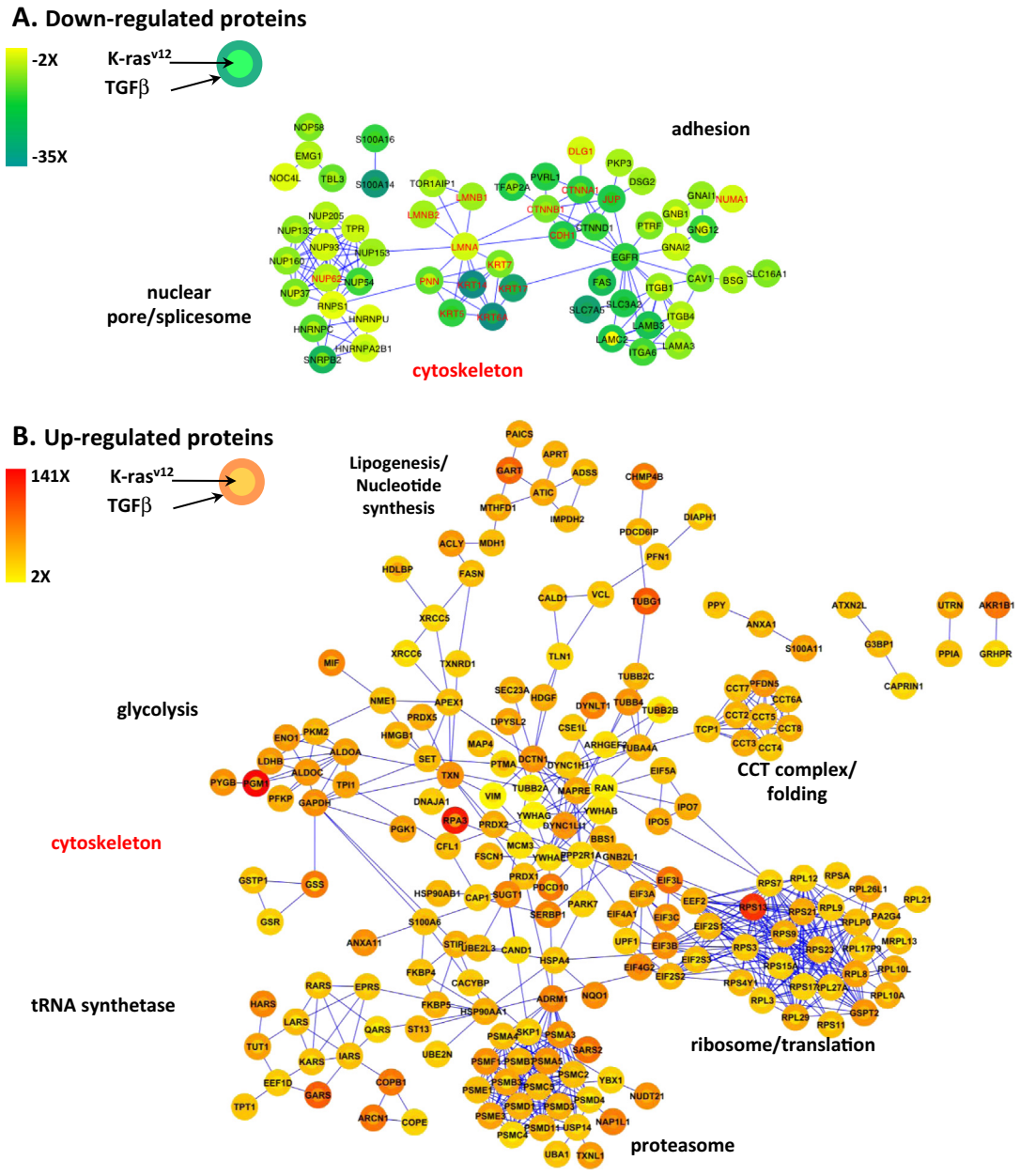


Fig. 6 – Protein sub-networks and EMT signature. Represented sub-networks of down-regulated (A) or up-regulated (B) proteins induced by both TGFβ and K-Ras^{v12}. STRING database has been used to report interaction between proteins. Only a combined score of experimental and database interactions was considered and the confidence score was set to medium (0.400) for all networks. Visualization was performed with Cytoscape v_2.8.2 (force-directed layout) and shows the interactions between these proteins, the type of regulation (yellow/blue for down-regulated proteins and yellow/red for up-regulated proteins), the strength of regulation (from 2 to 35 fold for down-regulated proteins and from 2 to 141 fold for up-regulated proteins) and the nature of the inducer (node border color for TGFβ and node color for K-Ras^{v12}). Several complexes are annotated (black complexes) and proteins that are part of the cytoskeleton are designated by red gene names.

3 MCX fractions with 2 peptides (516 peptides) and 3 transitions per peptide (1548 couples precursor ion/product ion) over the same 2 h chromatography gradient used on the Q-exactive to preserve the retention time and allow correct identification. Peptide ratios control/K-Ras^{v12} were averaged and compared to the ratios obtained by label-free quantification (Supplementary Table 3). The coefficient of correlation of the linear regression curve calculated between ratios obtained by Skyline quantification and by SRM is $r^2 = 0.81$ (Fig. 8A). Important differences usually resulted from higher ratios measured by SRM suggesting that LC–MS/MS analysis and label-free quantification underestimated some ratios (e.g. LAMA3, LAMB3, and LAMC2).

Among the proteins regulated by both EMT inducers, vimentin, a mesenchymal marker, was found in replicate label-free experiments, to be up-regulated by TGFβ and by

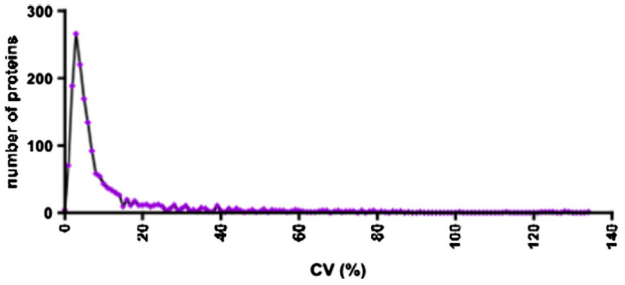


Fig. 7 – Repeatability of LC–MS analyses and quantification by MS1 filtering tool in Skyline. The same sample was analyzed three times by LC–MS/MS and quantified by MS1 filtering tool in Skyline. The median coefficient of variation (CV) was calculated for each peptide and protein, as outlined in the experimental procedures. The number of proteins in each CV range is shown. The median coefficient of variation calculated for each protein represented 5% with 83 proteins presenting a median CV of 50% and 90% of the proteins presenting a CV < 80%.

K-Ras^{v12}. Similarly, CDH1, considered as the prototypical epithelial marker, showed a decrease of abundance by TGFβ treatment and by K-Ras^{v12} expression. This result was confirmed by western blot, where we observed an increased expression of the mesenchymal markers, vimentin and N-cadherin (CDH-2) and a downregulation of the epithelial markers, E-cadherin (CDH1) and occludin in cells treated with TGFβ or expressing K-Ras^{v12} (Fig. 8B). Moreover, the measurement of CDH1 and vimentin by MRM (2 peptides per protein) in TGFβ and K-ras^{v12} samples compared to the control also confirmed this profile (Fig. 8C).

4. Discussion and conclusion

Large-scale studies of EMT usually report gene expression maps in order to understand mechanisms underlying EMT and several have revealed multiple signatures in the last decade. In this report, the protein abundance changes in two different models of EMT are given. The first one consisted of MCF10A breast cells stimulated with TGFβ, a strong EMT inducer, and the second one consisted of cells overexpressing the mutant K-Ras^{v12}, an oncogene frequently involved in cancer progression. By using mass spectrometric analysis combined with MS1 filtering tool quantification, this study allowed the comparison of the expression of a large number of proteins in a fast and repeatable manner. This method, by creating mass spectral libraries associated to retention time, allowed the development of a multiplex quantification in a SRM mode that fully validated the label-free quantification. The expression of E-cadherin (CDH1) and vimentin by both methods were also confirmed by western blot, consolidating the quantification results.

The structural features of EMT were observed at the protein level in both models by this mass spectrometric study. Indeed, in our study, 17 proteins involved in cell adhesion were down-regulated by both EMT inducers such as CDH1, catenins (CTNNB1, CTNNA1, and CTNND1), laminins (LAMB3, LAMC2, and LAMA3), and integrins (ITGB1, ITGB4, and ITGA6). It was also the case for proteins involved in the formation of the desmosome, a type of intercellular junctions (JUP, DSG2, PKP3). An EMT signature was also observed at the level of the cytoskeleton. Several cytoskeletal proteins of intermediate filaments such as keratins (KRT5/6A/7/14/17) were down-regulated, whereas actin microfilament-associated proteins such as profilin, cofilin, destrin, utrophin, vinculin, fascin, and microtubule-associated proteins such as dynein, dynactin, kynesin, septin2, and annexin were up-regulated. Several proteins from these processes have already been observed with a similar regulation in a transcriptional study performed on MCF10A cells treated with TGFβ [19].

More surprisingly, the translation apparatus was also greatly enhanced in MCF10A cells undergoing EMT, both after K-Ras^{v12} activation, which might be expected since Ras signaling leads to increased proliferation, but also upon TGFβ treatment although it drives also a proliferation arrest. Actually, more than 50 proteins that are part of, or associated with, both large and small ribosomal subunits, linked with several initiation factors (eIF3/eIF4) and tRNA synthetases were up-regulated similarly by both TGFβ and K-Ras^{v12}. It is noteworthy that this increased amount of ribosomal or translational proteins is associated with an increase in proteins involved in ribosome synthesis such as importin (IPO) or Ran.

Besides, many proteins involved in protein folding (chaperone proteins such as Hsp40 (DNAJA1), Hsp90 (Hsp90AB1), 7 subunits of the CCT (chaperonin containing TCP-1) and two additional CCT interacting proteins), but also proteins involved in the protection against oxidative stress (such as peroxiredoxin) or part of the protein degradation machinery (17 proteins belonging to the ubiquitination and proteasome-targeted degradation complex) were up-regulated by TGFβ and by K-Ras^{v12}. It is noteworthy that these functions are also requested in cells that would intently express nascent proteins [20,21]. All these structures are part of the super-complex "translasome," the eIF3 interactome, identified in a study revealing that ribosome biogenesis, translation elongation, quality control, degradation and transport are physically linked to facilitate efficient protein synthesis [22]. Taken together, all this suggests that MCF10A cells undergoing EMT not only display an increased ability to intently synthesize new proteins but also to fold, degrade and renew them.

Finally, energy is required for protein synthesis and more generally for anabolism. In this study, we observed that 15 proteins involved in glycolysis were up-regulated by TGFβ and K-Ras^{v12}. Increased glycolysis is not only necessary to produce energy but also to provide the metabolic intermediates. ATP-citrate synthase (ACLY), another enzyme up-regulated by both EMT inducers, promotes Krebs cycle activation at the mitochondria and fatty acid synthesis after export to the cytosol of the citrate excess. Of note, FASN, the multi-enzyme protein fatty acid synthase and MDH1, the malate dehydrogenase enzyme responsible for NADPH production, which is required for fatty acid synthesis, were also overexpressed by both EMT inducers. SERBP1, the major controller of cholesterol synthesis, which is an essential membrane component, was also overexpressed in both cases. It is noteworthy

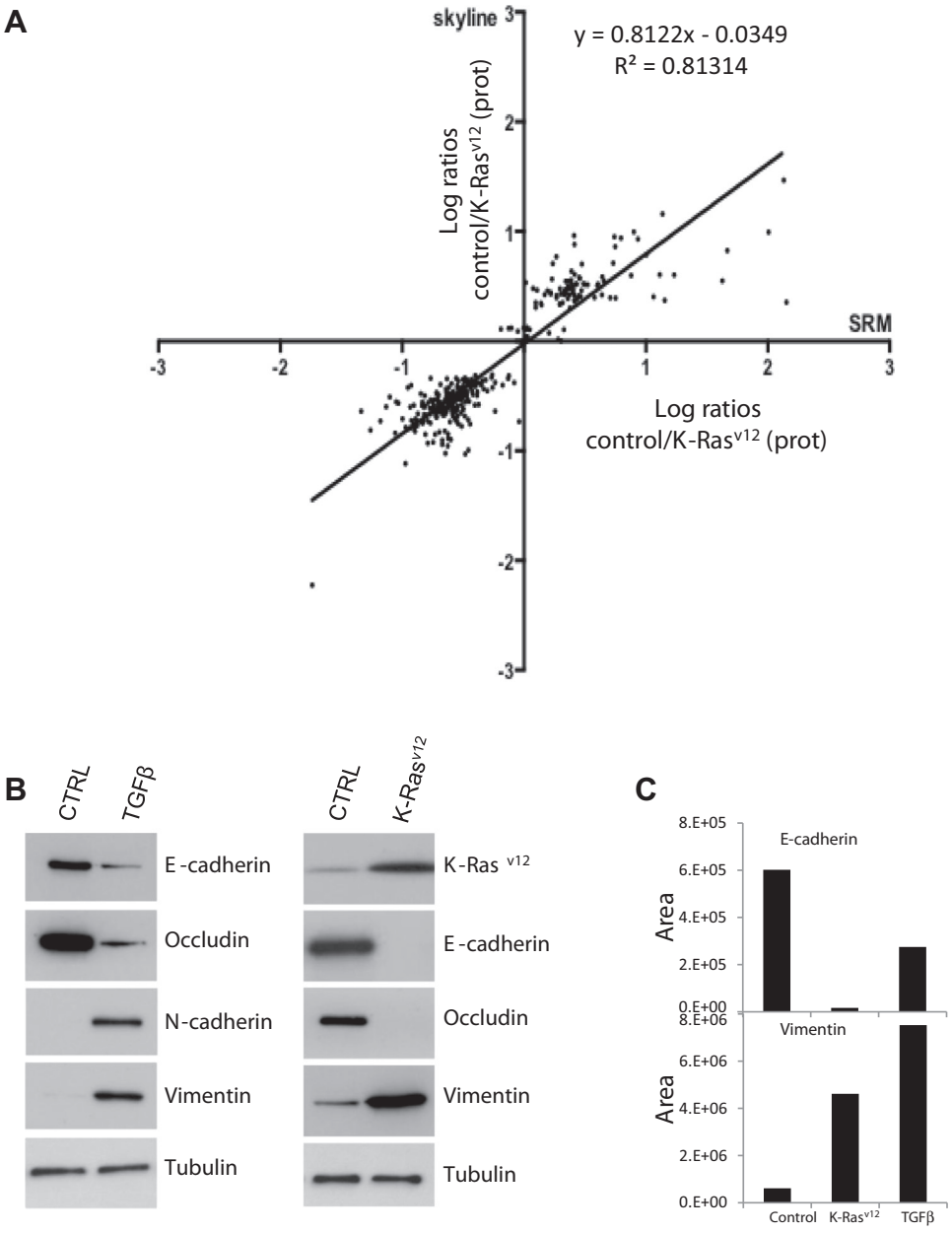


Fig. 8 – Validation of quantification performed by MS1 filtering tool in Skyline. (A) MCF10A overexpressing of K-Ras^{v12} used for the quantification by Skyline MS1 filtering tool (experiment 2) has been used to compare label-free quantification and SRM. 258 proteins were monitored in the 3 MCX fractions with 2 peptides (516 peptides) and 3 transitions per peptide (1548 couples of precursor ion/product ion) on a QTRAP® 5500 LC–MS/MS hybrid triple quadrupole/linear ion trap mass spectrometer over a 2 h chromatography gradient used on the Q-exactive. All transitions were manually integrated into Skyline software and peptide ratios control/K-Ras^{v12} were averaged for each protein and corrected with the same factor applied during the label-free quantification. The log ratios for each protein in both methods were plotted and a linear regression added. (B) E-cadherin (CDH1), occludin, N-cadherin (CDH2), vimentin and K-ras levels were determined by immunoblotting on MCF10A cells treated with TGFβ for 72 h or retrovirally infected by LXSN-K-Ras^{v12}. Tubulin was used as a loading control. Retrovirally infected MCF10A cells with empty vector (LXSN) were used as a control sample (CTRL). (C) E-cadherin and vimentin were measured by SRM in the same samples.

that 7 enzymes involved in nucleotide synthesis were also up-regulated by both EMT inducers such as MTHD1, APRT, or ATIC. All this suggests that MCF10A cells undergoing EMT exhibit high anabolic abilities, which is prominent in the K-Ras^{v12} phenotype but not so obvious upon TGFβ treatment.

In conclusion, by using mass spectrometric analysis combined with MS1 filtering tool quantification, this study allowed the comparison of the expression of a large number of proteins in two different models of EMT induced by TGFβ or by an oncogenic allele of K-Ras^{v12} in MCF10A breast cells in a fast

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026

and repeatable manner. This method creates mass spectral libraries that collect retention time and fragments of each peptide identified allowing the development of a multiplex quantification in a SRM mode. From these experiments an EMT signature was found that could be used for other cell types and at the same time revealed new cellular pathways involved in tumor progression. It is noteworthy that, besides the cytoskeletal changes that were expected in EMT, a strong increase in the anabolism and energy production machinery was observed. It is likely that these modifications are required for cells to undergo the EMT process. From these results, it would be interesting to test whether the inhibition on these metabolic changes can inhibit EMT in cell and metastatic spread in vivo.

Author contributions

JB, JL, RR, and GG designed the experiments. JB, IM, PG, MD and LF performed the experiments. JB analyzed the data and PG, IM and RR provided critical analyses of the results and their interpretations. JB, PG, IM, RR, GG and JL wrote the paper.

Competing interests

The authors declare that they have no competing interests.

Acknowledgment

The authors would like to acknowledge the Chemistry Institute of Lyon and the Physics Institute of CNRS. We thank Ben Ho Park for providing MCF10A-LXSN and MCF10A-LXSN-K-Ras^{v12} cell lines. This work was supported by INSERM, CNRS, Fondation ARC and Ligue Nationale contre le cancer. M.D. is supported by fellowships from *Ligue nationale contre le cancer* and L. F. from the *Fondation ARC*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.05.026.

REFERENCES

[1] Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial–mesenchymal transitions in development and disease. Cell 2009;139:871–90.
[2] Moustakas A, Heldin CH. Induction of epithelial–mesenchymal transition by transforming growth factor beta. Semin Cancer Biol 2012;22:446–54.
[3] Inman GJ. Switching TGFbeta from a tumor suppressor to a tumor promoter. Curr Opin Genet Dev 2011;21:93–9.
[4] Schubbert S, Shannon K, Bollag G. Hyperactive Ras in developmental disorders and cancer. Nat Rev Cancer 2007;7:295–308.
[5] Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 2004;91:355–8.
[6] Konishi H, Karakas B, Abukhdeir AM, Lauring J, Gustin JP, Garay JP, et al. Knock-in of mutant K-ras in nontumorigenic human epithelial cells as a new model for studying K-ras mediated transformation. Cancer Res 2007;67:8460–7.
[7] Vuoriluoto K, Haugen H, Kiviluoto S, Mpindi JP, Nevo J, Gjerdrum C, et al. Vimentin regulates EMT induction by Slug and oncogenic H-Ras and migration by governing Axl expression in breast cancer. Oncogene 2011;30:1436–48.
[8] Vergara D, Simeone P, del Boccio P, Toto C, Pieragostino D, Tinelli A, et al. Comparative proteome profiling of breast tumor cell lines by gel electrophoresis and mass spectrometry reveals an epithelial mesenchymal transition associated protein signature. Mol Biosyst 2013;9:1127–38.
[9] Mathias RA, Ji H, Simpson RJ. Proteomic profiling of the epithelial–mesenchymal transition using 2D DIGE. Methods Mol Biol 2012;854:269–86.
[10] Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL. A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. Mol Cell Proteomics 2011;10 [M111 010728].
[11] Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. In-depth analysis of tandem mass spectrometry data from disparate instrument types. Mol Cell Proteomics 2008;7:2386–98.
[12] Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 2007;4:207–14.
[13] Schilling B, Rardin MJ, MacLean BX, Zawadzka AM, Frewen BE, Cusack MP, et al. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol Cell Proteomics 2012;11:202–14.
[14] Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat Methods 2010;7:43–6.
[15] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
[16] Biarc J, Chalkley RJ, Burlingame AL, Bradshaw RA. The induction of serine/threonine protein phosphorylations by a PDGFR/TrkA chimera in stably transfected PC12 cells. Mol Cell Proteomics 2012;11:15–30.
[17] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011;39:D561–8.
[18] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[19] Hesling C, Fattet L, Teyre G, Jury D, Gonzalo P, Lopez J, et al. Antagonistic regulation of EMT by TIF1gamma and Smad4 in mammary epithelial cells. EMBO Rep 2011;12:665–72.
[20] Goldberg AL. Protein degradation and protection against misfolded or damaged proteins. Nature 2003;426:895–9.
[21] Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 2002;295:1852–8.
[22] Sha Z, Brill LM, Cabrera R, Kleifeld O, Scheliga JS, Glickman MH, et al. The eIF3 interactome reveals the translasome, a supercomplex linking protein synthesis and degradation machineries. Mol Cell 2009;36:141–52.

Please cite this article as: Biarc J, et al, Combination of a discovery LC–MS/MS analysis and a label-free quantification for the characterization of an epithelial–mesenchymal transition signature, J Prot (2014), http://dx.doi.org/10.1016/j.jprot.2014.05.026