Proteomics

ECM remodelling in breast cancer with different grade: contribution of 2D-DIGE proteomics.

Journal:	*PROTEOMICS*
Manuscript ID	pmic.201800278.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Moriggi, Manuela; University of Milan, Department of Biomedical Sciences for Health Giussani, Marta; Fondazione IRCCS Istituto Nazionale dei Tumori, Molecular Targeting Unit, Department of Research Torretta, Enrica; University of Milan, Department of Biomedical Sciences for Health Capitanio, Daniele; University of Milan, Department of Biomedical Sciences for Health Sandri, Marco; Fondazione IRCCS Istituto Nazionale dei Tumori, Molecular Targeting Unit, Department of Research Leone, Roberta; University of Milan, Department of Biomedical Sciences for Health Depalma, Sara; National Research Council (CNR), Institute of Bioimaging and Molecular Physiology Vasso, Michele; National Research Council (CNR), Institute of Bioimaging and Molecular Physiology Vozzi, Giovanni; University of Pisa, Research Center BE. Piaggio; University of Pisa, Dipartimento di Ingegneria dell'Informazione (DII) Tagliabue, Elda; Fondazione IRCCS Istituto Nazionale dei Tumori, Molecular Targeting Unit, Department of Research Gelfi, Cecilia; University of Milan, Department of Biomedical Sciences for Health; IRCCS Istituto Ortopedico Galeazzi
Keywords:	2D-DIGE, breast cancer, ECM, mass spectrometry, TN tumors

ECM remodelling in breast cancer with different grade: contribution of 2D-DIGE proteomics.

Manuela Moriggi[1], Marta Giussani[2], Enrica Torretta[1], Daniele Capitanio[1,3], Marco Sandri[2], Roberta Leone[1], Sara De Palma[4], Michele Vasso[4], Giovanni Vozzi[5,6], Elda Tagliabue[2], Cecilia Gelfi[1,3]

[1]Department of Biomedical Sciences for Health, University of Milan, Milan, Italy

[2]Molecular Targeting Unit, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

[3]IRCCS Istituto Ortopedico Galeazzi, Milano, Italy;

[4]Institute of Bioimaging and Molecular Physiology, National Research Council (CNR), Segrate-Cefalù, Italy

[5]Research Center BE. Piaggio, University of Pisa, Pisa, Italy

[6]Dipartimento di Ingegneria dell'Informazione (DII), University of Pisa, Pisa, Italy

Corresponding author: Professor Cecilia Gelfi,
Department of Biomedical Sciences for Health, University of Milan, Via F.lli Cervi 93, 20090, Segrate (MI), Italy
E-mail: cecilia.gelfi@unimi.it
Fax: +39(02)21717558

Abbreviations: A1BG, alpha-1B-glycoprotein; ARHGDIB, Rho GDP-dissociation inhibitor 2; BC, breast cancer; BVA, biological variation analysis; CCT3, T-complex protein 1 subunit gamma; CCT5, T-complex protein 1 subunit epsilon; COL6A1, collagen alpha-1(VI) chain; DDX1, ATP-dependent RNA helicase DDX1;de Tyr α-Tubulin, detyrosinated α-Tubulin; DIA, differential in-gel analysis; DN, double negative; E0G0, non-ECM3 of grade I-II; E0G3, non-ECM3 of grade III; E3G0, ECM3 of grade I-II; E3G3, ECM3 of grade III; ECM, extracellular matrix; ECM3, ECM gene expression signature; EDA, extended data analysis; ER, estrogen receptor; FBP1, fructose-1,6-bisphosphatase 1; FDR, false discovery rate; FGB, fibrinogen beta chain; FGG, fibrinogen gamma

chain; GDI2, Rab GDP dissociation inhibitor beta; HNRNPA2B1, heterogeneous nuclear ribonucleoproteins A2/B1; HSPA5, 78 kDa glucose-regulated protein; MIAPE, minimum information about a proteomics experiment; MSN, moesin; OGN, mimecan; PCA, principal component analysis; PDCD6IP, programmed cell death 6-interacting protein; PgR, progesterone receptor; PPIA, peptidyl-prolyl cis-trans isomerase A; RACK1, receptor of activated protein C kinase 1; Rho, ras homolog gene family; SERPINA1, alpha-1-antitrypsin; SET, protein SET; TN, triple negative; VCP, transitional endoplasmic reticulum ATPase; vs, versus; α5β1/αVβ3, integrin α5-integrin β1/integrin αV-integrin β3.

Keywords: 2D-DIGE, breast cancer, ECM, Mass Spectrometry, TN tumors

Total number of words: 7390

Abstract

Tumor extracellular matrix (ECM) plays a pivotal role in outcome of breast cancer (BC) patients. Overespression of 58 genes, encoding 43 structural ECM proteins, has been identified to determine a specific cluster of BC with accelerated metastatic potential only in the undifferentiated (grade III) phenotype. The scope of this study was to characterize protein repertoire able to predict patient outcome in BC according to ECM gene expression pattern and histological grade. The differential proteomic analysis has been based on 2D-DIGE, MALDI-MS, bioinformatics and immunoblotting. Results suggested a relationship among ECM remodeling, signal mechanotransduction and metabolic rewiring in BCs characterized by a specific mRNA ECM signature and identified a set of dysregulated proteins characteristic of hormone receptors expression as fibrinogen beta chain (FGB), collagen alpha-1 (VI) chain (COL6A1) and alpha-1B-glycoprotein (A1BG). Furthermore, in triple negative tumors (TN) with ECM signature, the FGG and α5β1/αvβ3 integrins increased whereas detyrosinated alpha-tubulin, and mimecan (OGN) decreased leading to unorganized integrin presentation involving focal adhesion kinase (FAK), activation of Rho GTPases associated to epithelial mesenchymal transition.

In hormone receptors negative BCs characterized by a specific ECM gene cluster, the differentially regulated proteins, identified in the present study,can be potentially relevant to predict patient's outcome.

Statement of significance of the study

Tumors can modify their microenvironment to form a tumor-associated mileu promoting cell spreading and invasion. However, key processes and signaling interactions fostered by epitelial cells and stroma remain partially unraveled and not considered in predicting patient outcome. The aim of this study is to analyze the protein dysregualtion in BCs characterized by mRNA ECM gene cluster in order to identify molecules potentially useful for patient monitoring and treatment. The investigation, based on 2D-DIGE, MALDI-MS, bioinformatics and immunoblotting of 26 BC's, confirmed the role of ECM remodelling in BCs. Results indicate that 58 overexpressed genes, encoding for 43 ECM proteins, are associated to structural and metabolic changes and provide better insight into the role of specific protein dysregulation within tumors lacking not only hormone receptors but also tyrosine protein kinase erbB-2 receptor. Specifically, results indicate that FBP1 is a trait of ECM gene signature irrespectively of tumor grade, collagen VI, fibrinogen, moesin and serpin are trait of undifferentiated grade III hormone receptor negative tumors. In TN tumors, cytoskeletal proteins are increased, high levels of COLVI are retained, fibrinogen, detyrosinated alpha-tubulin and mimecan decreased suggesting these last as a possible marker of epithelial mesenchymal transition occurring in TN with ECM gene signature.

1 Introduction

Breast cancer (BC) is a heterogeneous disease characterized by many subtypes according to molecular features of tumor cells[1, 2]. BCs are composed of malignant epithelial and stromal cells located in the tumor microenvironment which correlate with disease outcome[3-5]. Communication between epithelium and stroma is mediated through physical interactions among epithelial and stromal cells and the extracellular matrix (ECM), through the expression of signaling molecules[6]. In this context, ECM acts as a key component of microenvironment containing a number of biochemically distinct regulatory elements including proteins, glycoproteins, proteoglycans, and polysaccharides with different physical and biochemical properties[7]. Analysis of genes encoding ECM molecules showed that stromal expression patterns can vary among BC independently from neoplastic cell types, suggesting that distinct molecular patterns of stromal components can be associated to clinical outcome[8]. On this line proteomic studies, adopting different technological approaches, has been performed with the aim to identify additional proteins to subtype tumors within the current classification scale[2, 9][10][11][12]. The combination of proteomic analyses and a specific gene signature involving microenvironment, indicate, in animal models and cell cultures, that stromal proteomes play an essential role in tumor metastatic potential[13][14][15]. It has been also described that tumors colonize their microenvironment through a regulation of ECM proteome in which, collagens and cytoskeletal proteins together with complement and macrophages, play an essential role[16].

In a previous study, an ECM gene expression signature (ECM3) has been identified in ~40 % of BC introducing a new independent group of tumors[4, 5]. The analysis of an objective and reproducible set of overexpressing genes defining ECM3 tumors across different datasets, identified 58 genes relevant in determining the ECM3 cluster, of which 43 encode structural ECM proteins (e.g., different collagen chains, fibronectin, laminin, SPARC) coordinately overexpressed by both stromal and BC cells. Multivariate analysis of distant metastasis-free survival in untreated breast tumor

patients revealed a significant interaction between ECM3 and histological grade, indicating that ECM3 was significantly associated with high risk of relapse in patients with undifferentiated (grade III) tumors. Recent independent studies by our group have highlighted the function of ECM in different disorders introducing a new view of ECM as active player, pinpointing its role as signal "mechanotransducer" from outside to inside cells influencing structural organization, metabolism and nuclear signaling[17][18-20]. Furthermore, it has been found that tissue mechanics, assisted by ECM remodeling, is a key regulator in cancer progression in which structural forces and biochemical signals modulate cell and tissue behavior eliciting a specific cellular response[21][22][23][24][25].

The present study, is focused on the proteomic analysis of the molecular machinery (or protein repertoire) of BC according to ECM gene expression pattern and histological grade with the aim to contribute to a subtyping based on ECM protein signature and to define a protein repertoire that can possibly predict patient's outcome. The differential proteomic analysis was based on 2D-DIGE, MALDI mass spectrometry, bioinformatics and immunoblotting to provide quantitative data characteristic of BCs with different grade and quantitative overexpression of messengers encoding for ECM proteins. Results here described suggest a clear relationship among ECM remodelling, signal mechanotransduction and metabolic rewiring in BCs characterized by a specific mRNA ECM signature and, in particular, identifies a molecular repertoire characteristic of triple negative BCs.

2 Materials and methods

2.1 Human BC samples

Human BC specimens were obtained from women surgically treated at Fondazione IRCCS Istituto Nazionale dei Tumori of Milan. Samples were characterized for clinico-pathological characteristics (ER, PgR and HER2 expression, size, grade, node status and ECM3 feature) (Supporting Information Table 1S) and were representative of the main molecular subtypes (luminal, HER2+ and triple negative (TN)). Tissue samples were collected in collaboration with the institutional

biobank and stored frozen at -80 °C. Institutional approval from independent ethic committee (Comitato Etico Indipendente, Fondazione IRCCS, Istituto Nazionale dei Tumori) was obtained for this study (INT 144-14). Patients gave informed consent for the use of their own biological materials for future investigations and research purposes. All data were analyzed anonymously. All procedures were in accordance with the Helsinki Declaration.

Before proteomic analyses, BC samples were profiled for gene expression by Illumina platform (GSE59595)[26] and by Affymetrix Human Transcriptome Array 2.0 (GSE86945). To select ECM3 tumors, data were analyzed using the large average submatrix (LAS) biclustering method as previously reported[4].

BC specimens were then divided according to ECM3/grade features in 4 groups: ECM3 of grade I-II (E3G0), ECM3 of grade III (E3G3), non-ECM3 of grade I-II (E0G0) and non-ECM3 of grade III (E0G3).

2.2 Proteomic analysis

Soluble extracts from frozen BC specimens were analysed by 2D-DIGE and immunoblot assay to evaluate proteome changes. These analyses were conducted in two groups of patients, the first included: E3G0 (n = 5), E3G3 (n = 4), E0G0 (n= 5) and E0G3 (n = 5); the second was focused specifically on Triple Negative, negative to estrogen, progesterone and tyrosine protein kinase erbB-2 receptors, (TN) patients of E3G3 group (TN-E3G3) (n = 3) and E0G3 group (TN-E0G3) (n = 4).

2.2.1 Protein extraction

For 2D-DIGE and immunoblot assays, each sample from each subject after homogeneyzation was suspended in lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 30 mM Tris, and 1 mM PMSF) and solubilized by sonication on ice. Proteins were selectively precipitated using PlusOne 2D-Clean up Kit (GE Healthcare, Little Chalfont, UK) in order to remove non-protein impurities, and re-

suspended in lysis buffer. The pH of the protein extracts was adjusted to pH 8.5 by addition of 1 M NaOH. Protein concentration was determined by PlusOne 2D-Quant Kit (GE Healthcare).

2.2.2 2D-DIGE

Soluble extracts from frozen breast carcinomas were analysed by quantitative 2D-DIGE, followed by mass spectrometry. The 2D-DIGE method was inserted in The Minimum Information About a Proteomics Experiment-Gel Electrophoresis (MIAPE-GE) compliant form[27] in Supporting Information Table 2S. Protein minimal labelling with cyanine dyes (Cy3 and Cy5), 2D separation and analyses were performed as described previously[28]. Briefly, the proteins extracted (50 µg) from each individual sample were labelled with Cy5, while internal standards were generated by pooling (50 µg) individual samples that were Cy3-labeled. Samples were separated on 24 cm, 3–10 nonlinear immobilized pH gradient (IPG) strips; each individual sample was run in duplicate (E0G0, E0G3, E3G0 and E3G3) and in triplicate (TN-E0G3 and TN-E3G3; analytical replicates) to minimize the inter-gel variability and increase results reliability. Image analysis was performed using DeCyder 6.5 software (GE Healthcare). All gel images were imported into individual differential in-gel analysis (DIA) workspaces. Using the Batch Processor tool, automated detection of protein spots was performed with the following filter settings: estimated number of spots=10000, exclusion slope>1.2; minimal area cutoff<200 and peak height=100000. DIA workspaces were then manually edited to eliminate gel artefacts (e.g., plate scratches and dust specks) and include any incorrectly excluded spot. The resulting spot maps (containing the spot identifiers, locations and normalized volumes for all protein spots in each channel of each gel) were further processed in the biological variation analysis (BVA) module. Individual DIA workspaces for all analytical gels were imported into the BVA module. The BVA workspace was used for inter-gel protein spot matching. Statistical analysis was performed using the DeCyder 1.0 extended data analysis (EDA) module. Protein filters were set to select only those protein spots that matched 90% of the gel images and these protein spots were included in data analysis. Statistically significant differences were

computed by independent one-way ANOVA, coupled to Tukey's multiple group comparison test ($p<0.01$) for the comparison between E0G0 *vs* E0G3, E3G0 *vs* E3G3, E0G0 *vs* E3G0, E0G3 *vs* E3G3. As regards TN-E0G3 *vs* TN-E3G3 comparison, the unpaired Student's t-test ($p<0.01$) was applied. To minimize inclusion of false-positive protein spot changes, protein abundance data were filtered using the following criteria: independent one-way ANOVA, coupled to Tukey's multiple group comparison test and unpaired Student's t-test ($p<0.01$), 1.15-fold difference in abundance, and false discovery rate (FDR). The change of 1.15-fold and above in protein abundance was adopted for the present analysis, taking into account the power of the DIGE method to detect reliable differences in protein abundance down to 15%[29]. FDR correction was applied as a multiple testing correction method to keep the overall error rate as low as possible[30]. Principal component analysis (PCA) was performed using the DeCyder 1.0 EDA module. Proteins of interest were identified by Maldi mass spectrometry.

2.2.3 Protein identification by MALDI-TOF MS

Semipreparative gels, containing 400 µg of total protein extract per strip, were loaded; electrophoretic conditions were the same as for 2D-DIGE, except that gels were stained with a protein fluorescent stain (Deep Purple Total Protein Stain, GE Healthcare). Image acquisition was performed using the Typhoon 9200 laser scanner. Proteins were identified by PMF utilizing a MALDI-ToF mass spectrometer (Ultraflex III ToF/ToF; Bruker Daltonics, Bremen, Germany), as previously described[31]. In particular, the search was performed by correlation of uninterpreted spectra to Homo sapiens entries in NCBIprot database 20180429 (152.462.470 sequences; 55.858.910.152 residues). In cases where this approach was unsuccessful, additional searches were performed using electrospray ionization-MS/MS, as previously described [18]. Protein identification methods are provided in a MIAPE-MS compliant form [27] in Supporting Information Table 2S. (For further information about peptide mass fingerprinting and LC-MS/MS data are listed in Supporting

Information Table 3S and 4S, interpreted MS spectra are supplied in Supporting Information Figure 1S). A representative proteomic map is shown in Supporting Information Figure 2S

2.3 Protein quantitation by immunoblotting

Protein extracts (50 µg) from 3 samples of TN-E3G0 and 3 samples of TN-E3G3 were loaded in duplicate and resolved on 6–14% polyacrylamide gels (SE 400 Vertical Unit (GE Healthcare)) to visualized the following molecules: integrin β7, Rho-associated protein kinase 1 (ROCK1), vinculin, detyrosinated α-tubulin (de Tyr α-Tubulin), vimentin, myosin heavy chain 9 (MYH9), major vault protein (MVP), ubiquitin-binding protein p62 (p62), pyruvate dehydrogenase kinase (PDK) and prolyl hydroxylase 3 (PHD3). Protein extracts (50 µg) from pooled TN-E3G0 and TN-E3G3 were loaded in duplicate and resolved on 6–14% polyacrylamide gels to visualized the following molecules: integrin α5, integrin β1, integrin αV, integrin β3, panRAC, ras homolog gene family member A (RhoA), ras homolog gene family member B (RhoB) and fructose-1,6-bisphosphatase 1 (FBP1). Blots were incubated with rabbit, mouse or goat polyclonal primary antibodies (Santa Cruz Biotechnology, except when differently stated) as follows: anti-integrin β7, 1:500; anti-ROCK1 (Cell Signaling Technology), 1:1000; anti-vinculin, 1:1000; anti-de Tyr α-Tubulin (Abcam, Cambridge, UK), 1:500; anti-vimentin, 1:1000; anti-MYH9, 1:500; anti-MVP, 1:2000; anti-p62 (Sigma Aldrich), 1:1000; anti-PDK, 1:500; anti-PHD3 (Novous Biologicals), 1:500; anti-integrin α5, 1:500; anti-integrin β1, 1:500; anti-integrin αV, 1:500; anti-integrin β3, 1:500; anti-panRAC, 1:500; anti-RhoA, 1:500;anti-RhoB, 1:500 and anti-FBP1 (Novous Biologicals), 1:500. After washing, membranes were incubated with anti-rabbit or anti-mouse (GE Healthcare, Milan, Italy) or anti-goat (Santa Cruz Biotechnology) secondary antibodies, conjugated with horseradish peroxidase. Signals were visualized by chemiluminescence using the ECL Prime detection kit (GE Healthcare). Image analysis (Image Quant TL, Molecular Dynamics, Ragusa, Italy) was performed, followed by statistical analysis (Student's t-test, $p < 0.05$) to evaluate

differences between TN-E0G3 vs TN-E3G3. Data were normalized against the total amount of loaded proteins stained with Sypro Ruby Blot Stain (Life Technologies Europe BV, Monza, Italy).

2.4 2D-DIGE Orthogonal Validation by Immunoblotting

The 2D-DIGE results were validated by performing a random analysis via immunoblotting of identified proteins (Supporting Information Figure 3S). Immunoblotting procedures were as described above. Blots were incubated with rabbit, mouse, or goat polyclonal primary antibodies (Santa Cruz Biotechnology, except when differently stated) as follows: anti- alcohol dehydrogenase class 3 chi chain (ADH5), 1:500; anti-vimentin (VIM), 1:500; anti- glutamate dehydrogenase 1 (GLUD1), 1:500; anti-78 kDa glucose-regulated protein (HSPA5), 1:500; anti-peroxiredoxin-1 (PRDX1), 1:500; anti-gelsolin (GSN), 1:500; anti- protein disulfide-isomerase A6 (PDIA6), 1:500.

3 Results

3.1. Comparative Proteomic Analysis (2D-DIGE)

Nineteen primary human BCs classified according to ECM molecular characteristics and grade were analyzed by 2D-DIGE and mass spectrometry. Clinico-pathological features are shown in Supporting informationTable 5S. Proteomic results from sample extracts of different groups (E3G3, E3G0, E0G3 and E0G0) were compared.

Samples were processed and analysed adopting a well consolidated technology[29] and variation were determined applying a stringent statistical analysis. To overcome the individual variability and the restricted number of available samples, only spots differentially changed with the same trend in all samples of each group were considered. A representative 2D map of tumor extracts, after separation in 3-10 non linear pH gradient IPG strips, is shown in Supporting Information Figure 2SA. The entire list of proteins differentially expressed together with statistical analyses, protein accession number, gene name, theoretical molecular mass, isoelectric points and MS data are listed in Supporting Information Table 3S. The identified proteins are grouped according to their role.

An overview of 2D-DIGE results is described by the principal component analysis (PCA) which provides spots distribution in E3G0, E3G3, E0G0 and E0G3 samples shown in Figure 1. Results indicate the separation of 4 groups with a protein distribution markedly associated to grade (technical duplicates are shown in Figure1). However, the PCA highlights the presence of two samples in E3G3 group with a different spots distribution (figure 1A), both samples are negative to estrogen (ER-) and progesterone (PgR-) receptors expression (from now defined double negative, DN). Figure 1B indicates PCA without DN samples samples. Results show a quite good spots distribution according to grade and ECM signature suggesting that DN samples, within E3G3 group have a characteristic spot distribution and should be considered apart.

3.1.1 Protein signature based on grade

To answer the question whether in BCs, with the same ECM signature, protein levels were influenced by grade, a careful spot abundance analysis was performed.

E0G3 sample extracts were compared to E0G0. Overall, 1000 spots were matched among gels and 16 spots were changed in abundance, 13 of them identified. The semiquantitative proteomic profile of E0G3 *vs* E0G0 (Figure 2A) revealed changes in a restricted number of proteins. E0G3 were characterized by CAPZB increment and decrement of two proteoforms of VIM. Both proteins are involved in cytoskeleton organization. Concerning metabolic proteins: AHCY, AKR1A1, cytosolic MDH1, PGLS and NANS increased in grade III. PSMA2, GDI2, ARHGDIB, GSTP1 and PARK7 were also increased in E0G3 *vs* E0G0.

Comparing E3G3 samples to E3G0, only 5 spots were dysregulated, among those, ADH5 decreased (Figure 2A).

Overall these data revealed that differences in protein abundance between BCs of different grade are more relevant within E0 than E3 tumors.

3.1.2 Differential protein profile based on ECM molecular feature

The influence of ECM signature at protein level in tumors characterized by same grade is shown in Figure 2B.

Comparing E3G0 *vs* E0G0, only two proteins were changed and identified (Figure 2B). AHCY and ADH5 were increased in E3G0.

Comparing E3G3 *vs* E0G3, 18 proteins were dysregulated, 10 of them identified. A decrement in E3G3 was observed for the following proteins: CAPG, FAH, TPI1, ADH5, HNRPL, TUFM and two proteoforms of GDI2. ACTN1 and TPM1 increased in E3G3 compared to E0G3.

Overall these data indicate that the presence of a specific mRNA ECM signature mainly affects grade III BC.

3.1.3 Protein signature of DN-ECM3 grade III (DN-E3G3) *vs* E3G3 and DN-E3G3 *vs* E0G3

This comparison was necessary since in our first study PCA highlighted that, within E3G3 BCs, two samples (DN) negative for estrogen receptors (ER-) and progesterone receptors (PgR-) clusterized apart from ER+/PgR+ tumors.

Comparing DN-E3G3 *vs* E3G3 and DN-E3G3 *vs* E0G3, 58 and 104 proteins differentially abundant were identified, respectively. Semi-quantitative results are shown in figure 3, in particular in figure 3A indicated 57 proteins differentially expressed that were commonly dysregulated in DN-E3G3 *vs* E3G3 and DN-E3G3 *vs* E0G3; whereas figure 3B shows 47 proteins changed in one of two analyses only. Proteins were grouped according to their function. In figure 3A, among structural proteins: ACTR3, CAPZB and EZR decreased; whereas MSN increased in DN-E3G3. Proteins from the extracellular matrix apparatus, including two proteoforms of FGB, COL6A1 and two proteoforms of A1BG were increased in abundance in DN-E3G3. Concerning metabolic proteins increment of GLUD1 and of two proteoforms of APOA1 were observed in DN-E3G3; whereas PGAM1, PGK1, ENO1, NANS, PAICS, AHCY and AKR1A1 decreased in DN-E3G3 compared to E3G3 and E0G3. Among proteins involved in apoptosis, stress response and cell signaling: ARHGDIA, ARHGDIB, PDCD6IP, VCP, HSPA5, PRDX1, CCT5, CCT3, two proteoforms of PPIA, RACK1, GDI2 and

GNB1 were decreased in DN-E3G3, whereas SERPINA1 increased in abundance. Concerning protein synthesis and degradation, two proteoforms of DDX1, two proteoforms of HNRNPA2B1, OTUB1, PCBP1, PRPF19, several proteoforms from the proteasome subunit alpha (type 2, PSMA2; two proteoforms of type 4, PSMA4; type 7, PSMA7), SET, TUFM, UPS5, YWHAB and YWHAQ, NME2 and XRCC5 decreased in DN-E3G3 compared to E3G3 and E0G3. Regarding protein transport, CLIC1 decreased, whereas HPX and three proteoforms of TF were more abundant in DN-E3G3.

Figure 3B shows 47 proteins differentially expressed in DN-E3G3 *vs* E0G3. Among structural proteins: ARPC2, CAPG and KRT10 decreased; whereas GSN and two proteoforms of VIM increased in DN-E3G3 compared to E0G3. Concerning the extracellular matrix, two proteoforms of COL6A1 and DPT were increased in abundance in DN-E3G3 *vs* E0G3. Regarding metabolic proteins, a decrement of AKR7A2, ENO1, FBP1, mitochondrial IDH2, cytoplasmic MDH1, NQO1, PGK1, PGLS, PKM, TPI1 was observed in DN-E3G3; whereas DLD increased in DN-E3G3 compared to E0G3. Among proteins involved in apoptosis, stress response and cell signaling: two proteoforms of GDI2, ADH5, ANXA5, CCT2, DNAJB11, PARK7, two proteoforms of PDIA3, PRDX1, PRDX2, PRDX6, SERPINB1 and VCP were decreased in DN-E3G3 *vs* E0G3. SERPINA1 increased in abundance. Considering protein synthesis and degradation: HNRNPH2, HNRNPL, PEBP1, PSMA7, PSMB1, RBBP4, TUFM, XRCC5 and YWHAE decreased; whereas EEF1D increased in DN-E3G3 compared to E0G3. In the protein transport compartment, ALB and TF increase in DN-E3GE *vs* E0G3. Interestingly, TPM1 was down-regulated in DN-E3G3 compared to E3G3.

These data suggest significant differences in proteomic repertoire of undifferentiated BCs with ECM3 according to the expression of hormone receptors.

3.1.4 Protein signature of TN-E3G3 *vs* TN-E0G3

Considering the role of ECM gene characteristics in grade III BCs we focused on triple negative tumors which are reported to be the most undifferentiated BCs within tumors lacking the expression of estrogen, progesteron and tyrosine protein kinase erbB-2 receptors[32]. To determine the effects of the ECM3 signature in TN samples, the differential protein expression of TN-E3G3 *vs* TN-E0G3 was performed. 53 spots were found differentially regulated of them 32 were identified.

The principal component analysis is shown in figure 4A. The protein data set of TN-E3G3 was separated from TN-E0G3 by PCA1 (55,7% global variation), highlighting the overall changes in protein abundance and that TN-E0G3 differs from TN-E3G3.

Proteins differentially abundant in TN-E3G3 *vs* TN-E0G3 were grouped according to their function and shown in figure 4B. Concerning structural proteins: ARPC2, CORO1A and cytoskeletal KRT19 were decreased in TN-E3G3; whereas three proteoforms of GSN, two proteoforms of LMNA, VIM, TPM1, TPM4, MAPRE1 and cytoskeletal KRT8 increased in TN-E3G3. Extracellular matrix and metabolic proteins were also increased in TN-E3G3 such as FGG, LDHA and UGDH**.** Cell stress response indicated a decrement of PRDX1 whereas PDCD6IP, FKBP4, PDIA6 and SERPINA1 increased in TN-E3G3. Proteins involved in transcription, synthesis, degradation and signal transduction, like HNRNPK, MPV and ARHGAP1 were increased in TN-E3G3; whereas PSMB2 and ARHGDIB decreased. ALB, AKR7A2 and ANXA2 increased; whereas APOE, OGN and ANPEP decreased in TN-E3G3.

Overall these data confirm that TN E3G3 tumors have a characteristic proteomic repertoire in terms of cell-microenvironment interaction influencing cell metabolism.

3.2. Hints from Immunoblotting

To investigate whether ECM3 signature influences signal mechanotransduction from outside to inside tumor cells in TN BCs, some relevant molecules known to be involved in ECM remodelling and autophagy were determined by immunoblotting in TN-E3G3 *vs* TN-E0G3 (Figure 5A and B). Statistically significant changes were reached for detyrosinated alpha-tubulin and p62 which

decreased in TN-E3G3 whereas vimentin increased. Concerning integrin β7, major vault protein and ROCK1 trends toward differential abundance were observed (Fig 5A) whereas significant changes were observed in Integrin β1 and panRAC which decreased in TN-E3G3 compared to TN-E0G3 (Fig. 5B). Results of FBP1immunoblotting (Fig. 5C) highlighted the role of metabolism in ECM phenotype. FBP1 was significantly decreased in E3G3 and in DN-E3G3 compared to E0G3.

4 Discussion

Recent analysis of ECM gene expression profiles in BC identified an expression pattern (ECM3) in ~40% of BC that defines an independent group of tumors. ECM3 pattern has prognostic significance in relation to tumor differentiation status, being able to stratify a subgroup of tumor with poor prognosis within the most undifferentiated grade III. Conversely, distant metastasis-free survival of non-ECM3-grade III patients was similar to that of the most differentiated grade I-II patients regardless of ECM3 features[4]. PCA from in-gel proteomic 2D-DIGE analyses, highlighted grade as the main trait characterizing tumors. Considering ECM classification, PCA revealed significant differences between ECM3 and non-ECM3 within grade III tumors only, whereas grade I-II tumors did not display major variations. These findings are in agreement with our previous published results suggesting that the ECM3 trait significantly interacts with histological tumor grade leading to poor prognosis in undifferentiated (grade III) BCs. This observation turned our attention on the analysis of grade III tumors in which the differentially expressed proteins reflected group's distribution according to PCA indicating that protein expression of ECM3 signature could be a powerful tool to discriminate tumors within grade III BCs.

The 2D-DIGE is a top down proteomic technique for the study of a fraction of metabolic, structural, cytoskeletal and cell signalling protein levels and provides a picture of proteoforms and PTMs of undigested proteins[17, 19]. Considering E3G3 characteristics according to tumor hormone receptors expression, ECM proteins, in particular COL6A1, were more abundant in DN-E3G3 than in E3G3 BCs. Collagen VI is an extracellular protein that contributes to ECM structure; it is synthetized in

the endoplasmic reticulum and assembled in the extracellular space[33]. Collagen VI, in particular the endotrophin (ETP) fragment of collagen alpha-3 (VI) chain (COL6A3), was described for its involvement in breast cancer[34]. Our results indicate that MSN, entailed in connections of major cytoskeletal structures to the plasma membrane, A1BG and two proteoforms of FGB increased, suggesting that a matrix remodelling characterized DN-E3G3[35]. Previous studies from our group in inflammatory bowel disease (IBD) patients, in COL6A1 deficient animal models and mypoatic patients with mutation in COL6A1[17, 19, 36], highlighted the role of matrix, metabolism and structural remodelling in disease onset and progression. A similar "mechanosignalling" appears active also in DN-E3G3 suggesting that matrix remodelling can be able to percerive physical and chemical signals activating focal adhesion that sense the rigidity of the extracellular matrix[37]. In this context protein folding is severely influenced as indicated by dysregulation of a number of proteins as VCP, HSPA5, CCT5, CCT3 and PPIA, known to assist and accelerate the folding of actin and tubulin[38][39]. Signals, that are sensed at the cell membrane level and transduced through various secondary messengers to activate transcription factors, appear dysregulated. Proteins involved in cell signalling, such as RACK1 and GDI2, acting as key regulators of intracellular membrane trafficking and controlling the vesicular transport between the endoplasmic reticulum and Golgi, were decreased. Protein synthesis and degradation, in particular several proteoforms from the proteasome subunit alpha, participating to the ATP-dependent degradation of ubiquitinated proteins; SET, that inhibits TGF-beta response and promotes cell migration; DDX1, able to unwind both RNA-RNA and RNA-DNA duplexes; HNRNPA2B1, that associates nascent pre-mRNAs and packages them into hnRNP particles, were also decreased. Packaging plays a role in various processes and determine the mechanical properties of the nucleus including RNA nuclear export, suggesting that signalling from microenvironment can impact on translation and stability on mature mRNAs[24]. Further verification studies are required to consolidate our data; however, they can

represent a starting point for more refined studies in a larger number of samples. Furthermore, proteomics indicates that iron levels are sustained by hemopexin and serotransferrin overexpression. This study within E3G3 tumors suggested that, from a proteomic point of view, the absence of hormone receptors in grade III tumor cells exacerbates ECM3 signature. Moreover, considering TN tumors lacking the expression of estrogen, progesterone and tyrosine protein kinase erbB-2 receptors, we observed that the number of dysregulated proteins increased in TN-E3G3 *vs* TN-E0G3 compared to E3G3 *vs* E0G3 (32 and 10 differentially expressed proteins, respectively). Specifically, in TN-E3G3 compared to TN-E0G3, proteomic analysis indicated that extracellular matrix (FGG) and structural proteins (GSN, LMNA, VIM, TPM1, TPM4, MAPRE1 and KRT8) increased, further supporting the role of structural remodelling which may impact on cell signalling [40][41][23]. Our attention focused on FGG and α5β1/αVβ3 integrins, detyrosinated alpha-tubulin and OGN. These molecules regulating cell adhesion in conjunction with Rho/ROCKsignalling modulate cell migration in BC metastasis, and their variation supports the hypothesis that FGG can influence matrix composition exposing new epitopes that alter cell attachment by modulating focal adhesion plaque formation, and/or actin cytoskeleton rearrangement[41][42]. The characteristic actin architecture of TN-E3G3 tumors, may lead to unorganized surface integrin presentation that may further enhance the deposition of disorganized matrix by more invasive cells[41]. The proteomic dysregulation observed in TN-E3G3 can support changes in cell migration involving integrin–cytoskeletal interactions, focal adhesion kinase (FAK), and activation of Rho GTPases. Immunoblotting showed that components of focal adhesions, such as integrins, vimentin, de Tyr α-Tubulin, Rho and RAC were dysregulated in TN-E3G3. The signal from outside to inside the cell environment impacts on proteins involved in cytoskeleton rearrangement and protein folding. Decrement in TN-E3G3 of ARHGDIB, involved in the reorganization of actin cytoskeleton mediated by Rho family members[43], together with the increment of PDCD6IP, interplaying with F-actin and partitioning defective 3 homolog (PARD3) and tight junction protein ZO-1 (TJP1), are

essential to secure the proper assembly and positioning of actomyosin-tight junction complex and crucial for the maintenance of the epithelial cell polarity and barrier. These results can support the hypothesis of structural alterations, which may negatively impact, not only on cytoskeletal organization, but also on cell signalling[23][24][41][44]. A relevant feature characterizing TN-ECM3 tumors was the decrement of OGN. Previous studies on the mRNA changing expression pattern indicated that this proteins can play a critical role in the regulation of tissue-specific matrix assembly and function regulating collagen fibril growth, fibril organization and extracellular matrix assembly thus mediating cell-matrix interactions. Abnormal OGN expression can result in dysfunctional ECM[45]. Dysregulation has been recognized as a part of mRNA signature involved in epithelial-mesenchymal transition in prostate cancer and in meningioma suggesting a possible involvement also in BC[46]. Moreover, SERPINA1 significantly increased in TN-E3G3. The primary target of this protein is elastase, but it irreversibly inhibits trypsin, chymotrypsin and plasminogen activator suggesting that aberrant matrix deposition couldn't be removed and eliminated by proteases, which in turn are decreased (see PRDX1).

In order to understand the role of metabolism in ECM phenotype, metabolic adaptation was observed analyzing E3 tumors with different grade. The E3 signature indicates that the metabolic capacity is changed and exacerbated in E3G3 compared to E0G3 by a significant decrement of FBP1 and that this decrement is a characteristic feature of BC characterized by a ECM phenotype and is not grade dependent. Interestingly, FBP1 in TN-E3G3 vs TN-E0G3 also tends to decrease even though it does not reach statistical significance (data not shown). FBP1 is a pivotal enzyme controlling the exosamine pathway crucial in tumor progression[20, 47]. This observation, correlated with a lower level of metabolic proteins typical of ECM3 signature and suggests that a metabolic adaptation is associated to ECM remodeling in ECM3 tumors[48][49].

Overall these analyses showed that the protein repertoire of BCs mainly changes according to differentiation status of tumor cells and that the contribution of ECM features is higher in

undifferentiated grade III BCs, hormone receptors negative and particularly in TN showing characteristics related to ECM3 gene signature.

We are aware of the limitation of these results. However, we are convinced that, even if performed in a restricted number of patients, the identified dysregulated proteins could represent new biomarkers strictly dependent on ECM/tumor interaction that may turn out to be potential novel targets to block tumor progression.

5 References

[1] A. Prat, J. Baselga, Nature clinical practice. Oncology 2008, 5, 531.

[2] A. Prat, C. M. Perou, Molecular oncology 2011, 5, 5.

[3] R. B. West, D. S. Nuyten, S. Subramanian, T. O. Nielsen, C. L. Corless, B. P. Rubin, K. Montgomery, S. Zhu, R. Patel, T. Hernandez-Boussard, J. R. Goldblum, P. O. Brown, M. van de Vijver, M. van de Rijn, PLoS biology 2005, 3, e187.

[4] T. Triulzi, P. Casalini, M. Sandri, M. Ratti, M. L. Carcangiu, M. P. Colombo, A. Balsari, S. Menard, R. Orlandi, E. Tagliabue, PloS one 2013, 8, e56761.

[5] A. Bergamaschi, E. Tagliabue, T. Sorlie, B. Naume, T. Triulzi, R. Orlandi, H. G. Russnes, J. M. Nesland, R. Tammi, P. Auvinen, V. M. Kosma, S. Menard, A. L. Borresen-Dale, The Journal of pathology 2008, 214, e357.

[6] B. S. Wiseman, M. D. Sternlicht, L. R. Lund, C. M. Alexander, J. Mott, M. J. Bissell, P. Soloway, S. Itohara, Z. Werb, The Journal of cell biology 2003, 162, 1123.

[7] P. Lu, V. M. Weaver, Z. Werb, The Journal of cell biology 2012, 196, 395.

[8] M. Giussani, G. Merlino, V. Cappelletti, E. Tagliabue, M. G. Daidone, Seminars in cancer biology 2015, 35, 3.

[9] A. Gamez-Pozo, L. Trilla-Fuertes, G. Prado-Vazquez, C. Chiva, R. Lopez-Vacas, P. Nanni, J. Berges-Soria, J. Grossmann, M. Diaz-Almiron, E. Ciruelos, E. Sabido, E. Espinosa, J. A. Fresno Vara, PloS one 2017, 12, e0178296.

[10] N. Pendharkar, A. Gajbhiye, K. Taunk, S. RoyChoudhury, S. Dhali, S. Seal, A. Mane, S. Abhang, M. K. Santra, K. Chaudhury, S. Rapole, Journal of proteomics 2016, 132, 112.

[11] M. A. Muniz Lino, Y. Palacios-Rodriguez, S. Rodriguez-Cuevas, V. Bautista-Pina, L. A. Marchat, E. Ruiz-Garcia, H. Astudillo-de la Vega, A. E. Gonzalez-Santiago, A. Flores-Perez, J. Diaz-Chavez, A. Carlos-Reyes, E. Alvarez-Sanchez, C. Lopez-Camarillo, Journal of proteomics 2014, 111, 198.

[12] S. W. Lam, C. R. Jimenez, E. Boven, Cancer treatment reviews 2014, 40, 129.

[13] R. Castello-Cros, G. Bonuccelli, A. Molchansky, F. Capozza, A. K. Witkiewicz, R. C. Birbe, A. Howell, R. G. Pestell, D. Whitaker-Menezes, F. Sotgia, M. P. Lisanti, Cell Cycle 2011, 10, 2021.

[14] P. Schedin, P. J. Keely, Cold Spring Harbor perspectives in biology 2011, 3, a003228.

[15] K. R. Levental, H. Yu, L. Kass, J. N. Lakins, M. Egeblad, J. T. Erler, S. F. Fong, K. Csiszar, A. Giaccia, W. Weninger, M. Yamauchi, D. L. Gasser, V. M. Weaver, Cell 2009, 139, 891.

[16] X. Wang, A. D. Mooradian, P. Erdmann-Gilmore, Q. Zhang, R. Viner, S. R. Davies, K. L. Huang, R. Bomgarden, B. A. Van Tine, J. Shao, L. Ding, S. Li, M. J. Ellis, J. C. Rogers, R. R. Townsend, D. Fenyo, J. M. Held, Science signaling 2017, 10.

[17] M. Moriggi, L. Pastorelli, E. Torretta, G. E. Tontini, D. Capitanio, S. F. Bogetto, M. Vecchi, C. Gelfi, Proteomics 2017, 17.

[18] D. Capitanio, M. Moriggi, S. De Palma, D. Bizzotto, S. Molon, E. Torretta, C. Fania, P. Bonaldo, C. Gelfi, P. Braghetta, Frontiers in molecular neuroscience 2017, 10, 337.

[19] S. De Palma, R. Leone, P. Grumati, M. Vasso, R. Polishchuk, D. Capitanio, P. Braghetta, P. Bernardi, P. Bonaldo, C. Gelfi, PloS one 2013, 8, e56716.

[20] R. Palorini, G. Votta, Y. Pirola, H. De Vitto, S. De Palma, C. Airoldi, M. Vasso, F. Ricciardiello, P. P. Lombardi, C. Cirulli, R. Rizzi, F. Nicotra, K. Hiller, C. Gelfi, L. Alberghina, F. Chiaradonna, PLoS genetics 2016, 12, e1005931.

[21] C. W. Li, S. O. Lim, E. M. Chung, Y. S. Kim, A. H. Park, J. Yao, J. H. Cha, W. Xia, L. C. Chan, T. Kim, S. S. Chang, H. H. Lee, C. K. Chou, Y. L. Liu, H. C. Yeh, E. P. Perillo, A. K. Dunn, C. W. Kuo, K. H. Khoo, J. L. Hsu, Y. Wu, J. M. Hsu, H. Yamaguchi, T. H. Huang, A. A. Sahin, G. N. Hortobagyi, S. S. Yoo, M. C. Hung, Cancer cell 2018, 33, 187.

[22] M. Giussani, C. De Maria, M. Vasso, F. Montemurro, T. Triulzi, E. Tagliabue, C. Gelfi, G. Vozzi, Curr Mol Bio Rep 2015, 1, 71.

[23] J. Wang, I. C. Schneider, Biomaterials 2017, 120, 81.

[24] C. Uhler, G. V. Shivashankar, Nature reviews. Molecular cell biology 2017, 18, 717.

[25] J. A. M. Nuhn, A. M. Perez, I. C. Schneider, Acta biomaterialia 2018, 66, 248.

[26] X. Huang, M. Dugo, M. Callari, M. Sandri, L. De Cecco, B. Valeri, M. L. Carcangiu, J. Xue, R. Bi, S. Veneroni, M. G. Daidone, S. Menard, E. Tagliabue, Z. Shao, J. Wu, R. Orlandi, Cancer medicine 2015, 4, 1016.

[27] C. F. Taylor, N. W. Paton, K. S. Lilley, P. A. Binz, R. K. Julian, Jr., A. R. Jones, W. Zhu, R. Apweiler, R. Aebersold, E. W. Deutsch, M. J. Dunn, A. J. Heck, A. Leitner, M. Macht, M. Mann, L. Martens, T. A. Neubert, S. D. Patterson, P. Ping, S. L. Seymour, P. Souda, A. Tsugita, J. Vandekerckhove, T. M. Vondriska, J. P. Whitelegge, M. R. Wilkins, I. Xenarios, J. R. Yates, 3rd, H. Hermjakob, Nature biotechnology 2007, 25, 887.

[28] M. Moriggi, M. Vasso, C. Fania, D. Capitanio, G. Bonifacio, M. Salanova, D. Blottner, J. Rittweger, D. Felsenberg, P. Cerretelli, C. Gelfi, Proteomics 2010, 10, 3756.

[29] R. Marouga, S. David, E. Hawkins, Analytical and bioanalytical chemistry 2005, 382, 669.

[30] Y. Benjamini, Y. Hochberg, J Educ Behav Stat 2000, 25, 60.

[31] A. Vigano, M. Vasso, A. Caretti, V. Bravata, L. Terraneo, C. Fania, D. Capitanio, M. Samaja, C. Gelfi, Proteomics 2011, 11, 4202.

[32] O. Gluz, C. Liedtke, N. Gottschalk, L. Pusztai, U. Nitz, N. Harbeck, Annals of oncology : official journal of the European Society for Medical Oncology 2009, 20, 1913.

[33] M. Cescon, F. Gattazzo, P. Chen, P. Bonaldo, Journal of cell science 2015, 128, 3525.

[34] E. Karousou, M. L. D'Angelo, K. Kouvidi, D. Vigetti, M. Viola, D. Nikitovic, G. De Luca, A. Passi, BioMed research international 2014, 2014, 606458.

[35] J. Park, P. E. Scherer, The Journal of clinical investigation 2012, 122, 4243.

[36] S. De Palma, D. Capitanio, M. Vasso, P. Braghetta, C. Scotton, P. Bonaldo, H. Lochmuller, F. Muntoni, A. Ferlini, C. Gelfi, Journal of proteome research 2014, 13, 5022.

[37] B. Stutchbury, P. Atherton, R. Tsang, D. Y. Wang, C. Ballestrem, Journal of cell science 2017, 130, 1612.

[38] S. Seo, L. M. Baye, N. P. Schulz, J. S. Beck, Q. Zhang, D. C. Slusarski, V. C. Sheffield, Proceedings of the National Academy of Sciences of the United States of America 2010, 107, 1488.

[39] K. Lang, F. X. Schmid, G. Fischer, Nature 1987, 329, 268.

[40] C. Pardo-Pastor, F. Rubio-Moscardo, M. Vogel-Gonzalez, S. A. Serra, A. Afthinos, S. Mrkonjic, O. Destaing, J. F. Abenza, J. M. Fernandez-Fernandez, X. Trepat, C. Albiges-Rizo, K. Konstantopoulos, M. A. Valverde, Proceedings of the National Academy of Sciences of the United States of America 2018, 115, 1925.

[41] T. Wang, S. Hamilla, M. Cam, H. Aranda-Espinoza, S. Mili, Nature communications 2017, 8, 896.

[42] M. Rolli, E. Fransvea, J. Pilch, A. Saven, B. Felding-Habermann, Proceedings of the National Academy of Sciences of the United States of America 2003, 100, 9482.

[43] H. Leffers, M. S. Nielsen, A. H. Andersen, B. Honore, P. Madsen, J. Vandekerckhove, J. E. Celis, Experimental cell research 1993, 209, 165.

[44] S. Pan, R. Wang, X. Zhou, J. Corvera, M. Kloc, R. Sifers, G. E. Gallick, S. H. Lin, J. Kuang, The EMBO journal 2008, 27, 2077.

[45] S. Chen, D. E. Birk, The FEBS journal 2013, 280, 2120.

[46] B. Orr, A. C. Riddick, G. D. Stewart, R. A. Anderson, O. E. Franco, S. W. Hayward, A. A. Thomson, Oncogene 2012, 31, 1130; Y. Mei, Z. Du, C. Hu, N. F. Greenwald, M. Abedalthagafi, N. Y. R. Agar, G. P. Dunn, W. L. Bi, S. Santagata, I. F. Dunn, Cell communication and signaling : CCS 2017, 15, 34.

[47] F. Chiaradonna, F. Ricciardiello, R. Palorini, Cells 2018, 7.

[48] G. Jego, A. Hazoume, R. Seigneuric, C. Garrido, Cancer letters 2013, 332, 275.

[49] G. M. DeNicola, F. A. Karreth, T. J. Humpton, A. Gopinathan, C. Wei, K. Frese, D. Mangal, K. H. Yu, C. J. Yeo, E. S. Calhoun, F. Scrimieri, J. M. Winter, R. H. Hruban, C. Iacobuzio-Donahue, S. E. Kern, I. A. Blair, D. A. Tuveson, Nature 2011, 475, 106.

Acknowledgements

This work was supported by the Italian association for cancer research (Associazione Italiana per la Ricerca sul Cancro - AIRC), (No 12162) (ET). MG was recipient of a three-year fellowship from AIRC and from the Italian foundation for cancer research (Fondazione Italiana per la Ricerca sul Cancro - FIRC) (Ref 16563).

All authors have no conflict of interest to declare.

Legends

Figure 1. Principal component analysis (PCA). **Figure 1A.** PCA provided the overview of spots distribution in E0G0, E0G3, E3G0 and E3G3. Samples in the black box are the two outliers within E3G3 tumors. **Figure 1B.** PCA provided the overview of spots distribution in E0G0, E0G3, E3G0 and E3G3 without the two outlier samples. Result showed the separation of 4 groups with the major difference in protein distribution related to grade.Samples were analyzed in duplicate.

Figure 2A. Proteomic analysis by 2D-DIGE, histograms of differentially expressed proteins observed by 2D-DIGE between E0G3 *vs* E0G0 (gray bars) and E3G3 *vs* E3G0 (black bars). **Figure 2B**. Proteomic analysis by 2D-DIGE, histograms of differentially expressed proteins observed by 2D-DIGE between E3G0 vs E0G0 (gray bars) and E3G3 vs E0G3 (black bars). Statistical analysis was performed using DeCyder EDA 1.0 module; independent one-way ANOVA coupled to Tukey's multiple group comparison test ($P<0.01$) were used to determine protein spots significantly different between the analyses. Differentially expressed proteins are indicated by their gene name and they are expressed as % of fold change. List of proteins differentially expressed in all comparisons were grouped according to their role together with statistical analysis and are presented as raw data sets in Supporting Information Table 3S. Structural proteins: CAPZB, F-actin capping protein subunit beta; VIM, vimentin; ACTN1, alpha-actinin-1; CAPG, macrophage capping protein; TPM1, tropomyosin-1 alpha chain. Metabolic proteins: AHCY, adenosylhomocysteinase; AKR1A1, alcohol dehydrogenase; MDH1, malate dehydrogenase, cytoplasmic; PGLS, 6-phosphogluconolactonase; NANS, Sialic acid synthase; TPI1, triosephosphate isomerase; FAH, fumarylacetoacetase. Apoptotic and stress response proteins: ARHGDIB, Rho GDP-dissociation inhibitor 2; ADH5, Alcohol dehydrogenase class 3 chi chain; GSTP1, Glutathione S-transferase P; PARK7, Protein/nucleic acid deglycase DJ-1. Protein degradation: PSMA2, Proteasome subunit alpha type 2. Signal transduction: GDI2, Rab GDP

dissociation inhibitor beta. mRNA splicing: HNRNPL, heterogeneous nuclear ribonucleoprotein L. Protein biosynthesis: TUFM, elongation factor Tu, mitochondrial.

Figure 3. Proteomic analysis by 2D-DIGE. Histograms of differentially expressed proteins observed by 2D-DIGE in DN-E3G3 compared to E3G3 (black bars) and DN-E3G3 *vs* E0G3 (gray bars). Statistical analysis was performed using DeCyder EDA 1.0 module; independent one-way ANOVA coupled to Tukey's multiple group comparison test ($P<0.01$) were used to determine protein spots significantly different between the analyses. Differentially expressed proteins are indicated by their gene name and they are expressed as % of fold change. List of proteins differentially expressed in all comparisons were grouped according to their role together with statistical analysis and are presented as raw data sets in Supporting Information Table 3S. **Figure 3A**. Histograms of 57 proteins differentially expressed in common between the two analysis. **Figure 3B**. Histograms of 47 proteins differentially expressed in DN-E3G3 *vs* E0G3 and only one protein differentially expressed in DN-E3G3 *vs* E3G3.

Figure 4A. Shows the PCA with an overview of spots distribution in TN-E3G3 *vs* TN-E0G3. The protein data set of TN-E3G3 was separated from TN-E0G3 by PCA1 (55,7% global variation), highlighting the overall changes in protein abundance. PCA2 (11% global variation). This analysis highlight the difference between samples clearly indicating that TN-E0G3 differ from TN-E3G3 in protein abundance. Each sample is shown as triplicate**. Figure 4B.** Proteomic analysis by 2D-DIGE. Histograms of differentially expressed proteins observed by 2D-DIGE in TN-E3G3 compared to TN-E0G3. Statistical analysis was performed using DeCyder EDA 1.0 module; data were computed by an unpaired Student's t-test ($p < 0.01$) to determine protein spots significantly different between the analyses. Differentially expressed proteins are indicated by their gene name and they are expressed as % of fold change. List of proteins differentially expressed in all comparisons were grouped according to their role together with statistical analysis and are presented as raw data sets in Supporting Information Table 4S. Structural proteins: ARPC2, actin-related protein 2/3 complex

subunit 2; GSN, gelsolin; LMNA, prelamin-A/C; CORO1A, coronin-1A; VIM, vimentin; TPM1, tropomyosin-1 alpha chain; TPM4, tropomyosin alpha-4 chain; MAPRE1, microtubule-associated protein RP/EB family member 1; KRT8, keratin, type II cytoskeletal 8; KRT19, keratin, type I cytoskeletal 19. Extracellular matrix protein: FGG, fibrinogen gamma chain. Metabolic proteins: LDHA, L-lactate dehydrogenase A chain; UGDH, UDP-glucose dehydrogenase. Chaperone, apoptotic and stress response proteins: FKBP4, peptidyl-prolyl cis-trans isomerase; PDCD6IP, programmed cell death 6-interacting protein; PRDX1, peroxiredoxin-1; PDIA6, protein disulfide-isomerase A6; SERPINA1, alpha-1-antitrypsin. Transcription regulation: HNRNPK, heterogeneous nuclear ribonucleoprotein K. Protein synthesis and degradation: PSMB2, proteasome subunit beta type 2. Signal transduction: MVP, major vault protein; ARHGAP1, Rho GTPase-activating protein 1; ARHGDIB, Rho GDP-dissociation inhibitor 2. Transport: ALB, serum albumin; APOE, apolipoprotein E. Other proteins: AKR7A2, aflatoxin B1 aldehyde reductase member 2; ANXA2, annexin A2; OGN, mimecan; ANPEP, aminopeptidase N

Figure 5A. Representative immunoblot images (cropped images; full lenght blots are included as Supporting Information Figure 4S) and histograms (means ± SD) of proteins abundance of Integrin β7, Rho-associated protein kinase 1 (ROCK1), Vinculin, detyrosinated α-Tubulin (de Tyr α-Tubulin), Vimentin, myosin heavy chain 9 (MYH9), major vault protein (MVP), ubiquitin-binding protein p62 (p62), pyruvate dehydrogenase kinase (PDK) and prolyl hydroxylase 3 (PHD3) in TN-E0G3 compared to TN-E3G3. **Figure 5B**. Representative immunoblot images (cropped images; full lenght blots are included as Supporting Information Figure 4S) and histograms (means ± SD) of proteins abundance of integrin α5, integrin β1, integrin αV, integrin β3, panRAC, ras homolog gene family member A (RhoA) and ras homolog gene family member B (RhoB) in TN-E0G3 compared to TN-E3G3. **Figure 5C.** Representative immunoblot images (cropped images; full lenght blots are included as Supporting Information Figure 4S) and histograms (means ± SD) of fructose-1,6-bisphosphatase 1 (FBP1) abundance.

The data were normalized against the total amount of loaded proteins stained with Sypro Ruby. Statistical analysis was performed by Student's t-test ($p<0.05$).

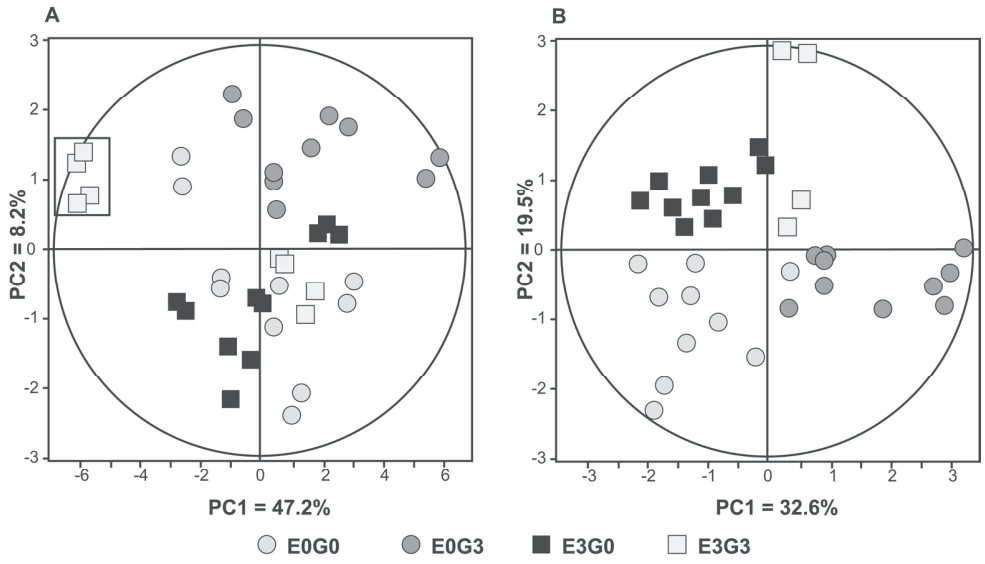

Figure 1. Principal component analysis (PCA). Figure 1A. PCA provided the overview of spots distribution in E0G0, E0G3, E3G0 and E3G3. Samples in the black box are the two outliers within E3G3 tumors. Figure 1B. PCA provided the overview of spots distribution in E0G0, E0G3, E3G0 and E3G3 without the two outlier samples. Result showed the separation of 4 groups with the major difference in protein distribution related to grade differences.Samples were analyzed in duplicate.

94x54mm (600 x 600 DPI)

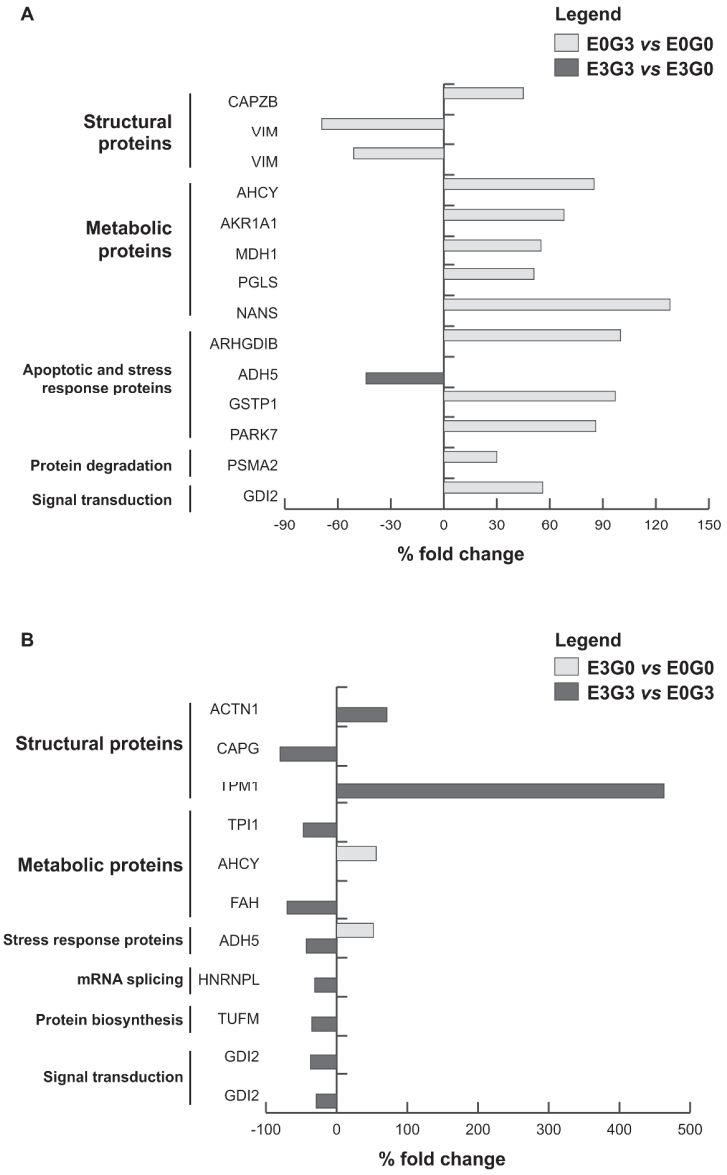

Figure 2A. Proteomic analysis by 2D-DIGE, histograms of differentially expressed proteins observed by 2D-DIGE between E0G3 vs E0G0 (gray bars) and E3G3 vs E3G0 grade I-II (black bars). Figure 2B. Proteomic analysis by 2D-DIGE, histograms of differentially expressed proteins observed by 2D-DIGE between E3G0 vs E0G0 (gray bars) and E3G3 vs E0G3 (black bars). Statistical analysis was performed using DeCyder EDA 1.0 module; independent one-way ANOVA coupled to Tukey's multiple group comparison test ($P<0.01$) were used to determine protein spots significantly different between the analyses. Differentially expressed proteins are indicated by their gene name and they are expressed as % of fold change. List of proteins differentially expressed in all comparisons were grouped according to their role together with statistical analysis and are presented as raw data sets in Supporting Information Table 3S. Structural proteins: CAPZB, F-actin capping protein subunit beta; VIM, vimentin; ACTN1, alpha-actinin-1; CAPG, macrophage capping protein; TPM1, tropomyosin-1 alpha chain. Metabolic proteins: AHCY, adenosylhomocysteinase; AKR1A1, alcohol dehydrogenase; MDH1, malate dehydrogenase, cytoplasmic; PGLS, 6-phosphogluconolactonase; NANS, Sialic acid synthase; TPI1, triosephosphate isomerase; FAH, fumarylacetoacetase. Apoptotic and

stress response proteins: ARHGDIB, Rho GDP-dissociation inhibitor 2; ADH5, Alcohol dehydrogenase class 3 chi chain; GSTP1, Glutathione S-transferase P; PARK7, Protein/nucleic acid deglycase DJ-1. Protein degradation: PSMA2, Proteasome subunit alpha type 2. Signal transduction: GDI2, Rab GDP dissociation inhibitor beta. mRNA splicing: HNRNPL, heterogeneous nuclear ribonucleoprotein L. Protein biosynthesis: TUFM, elongation factor Tu, mitochondrial.

218x354mm (600 x 600 DPI)

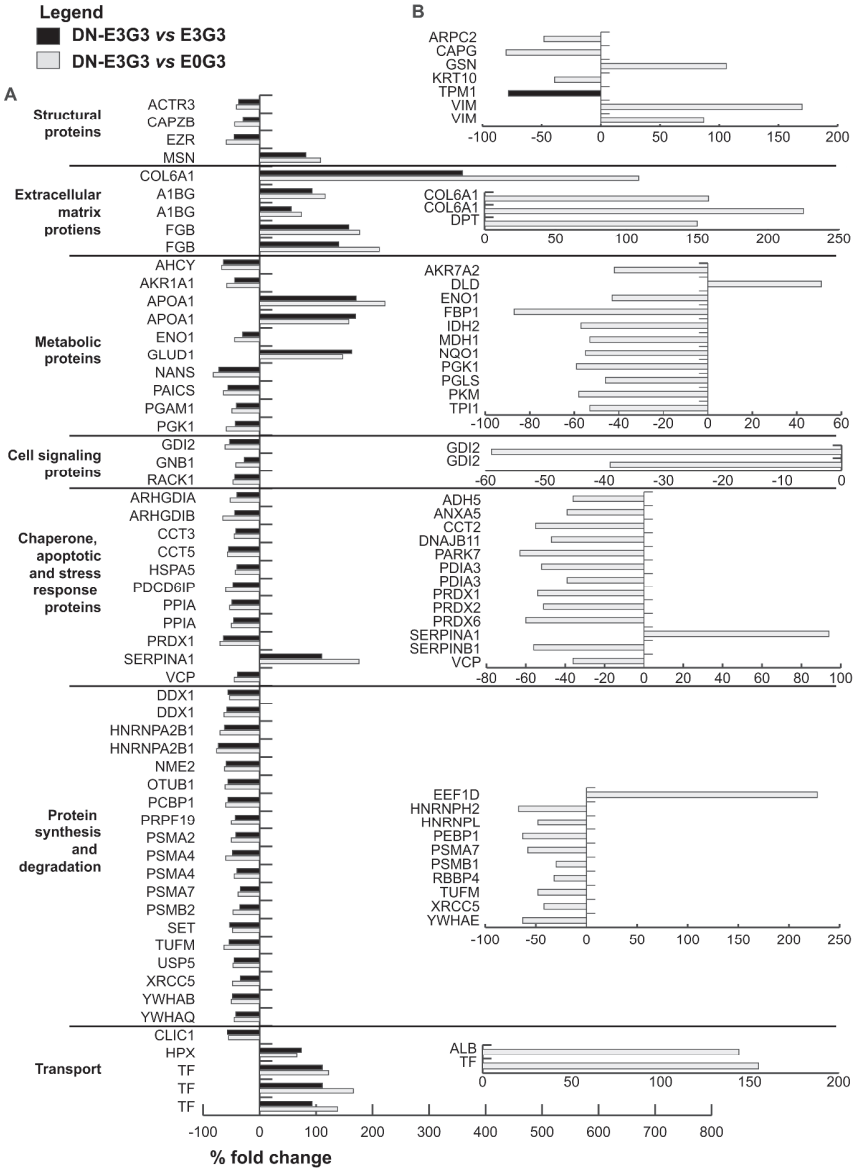

Figure 3. Proteomic analysis by 2D-DIGE. Histograms of differentially expressed proteins observed by 2D-DIGE in DN-E3G3 compared to E3G3 (black bars) and DN-E3G3 vs E0G3 (gray bars). Statistical analysis was performed using DeCyder EDA 1.0 module; independent one-way ANOVA coupled to Tukey's multiple group comparison test ($P<0.01$) were used to determine protein spots significantly different between the analyses. Differentially expressed proteins are indicated by their gene name and they are expressed as % of fold change. List of proteins differentially expressed in all comparisons were grouped according to their role together with statistical analysis and are presented as raw data sets in Supporting Information Table 3S. Figure 3A. Histograms of 57 proteins differentially expressed common between the two analysis. Figure 3B. Histograms of 47 proteins differentially expressed in DN-E3G3 vs E0G3 and only one protein differentially expressed in DN-E3G3 vs E3G3.

228x313mm (600 x 600 DPI)

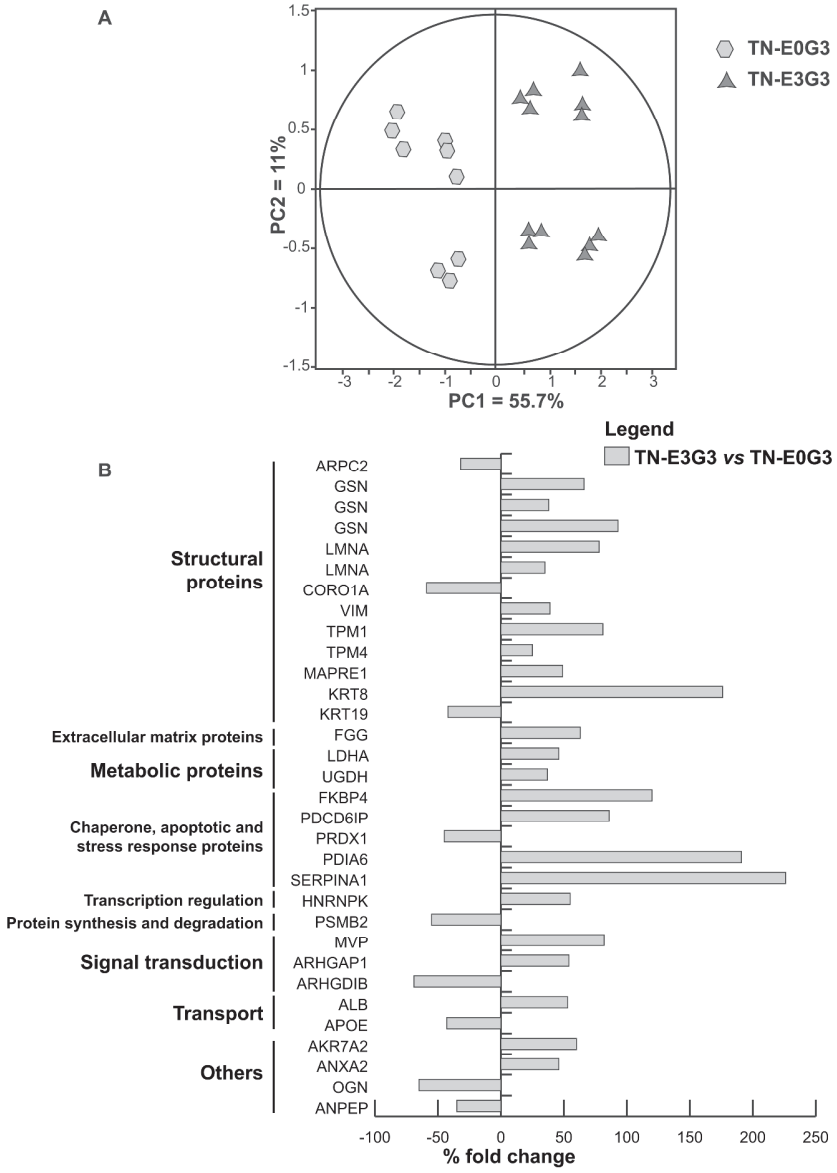

Figure 4A. Shows the PCA with an overview of spots distribution in TN-E3G3 vs TN-E0G3. The protein data set of TN-E3G3 was separated from TN-E0G3 by PCA1 (55,7% global variation), highlighting the overall changes in protein abundance. PCA2 (11% global variation). This analysis highlight the difference between samples clearly indicating that TN-E0G3 differ from TN-E3G3 in protein abundance. Each sample is shown as triplicate. Figure 4B. Proteomic analysis by 2D-DIGE. Histograms of differentially expressed proteins observed by 2D-DIGE in TN-E3G3 compared to TN-E0G3. Statistical analysis was performed using DeCyder EDA 1.0 module; data were computed by an unpaired Student's t-test ($p < 0.01$) to determine protein spots significantly different between the analyses. Differentially expressed proteins are indicated by their gene name and they are expressed as % of fold change. List of proteins differentially expressed in all comparisons were grouped according to their role together with statistical analysis and are presented as raw data sets in Supporting Information Table 4S. Structural proteins: ARPC2, actin-related protein 2/3 complex subunit 2; GSN, gelsolin; LMNA, prelamin-A/C; CORO1A, coronin-1A; VIM, vimentin; TPM1, tropomyosin-1 alpha chain; TPM4, tropomyosin alpha-4 chain; MAPRE1, microtubule-associated protein RP/EB family member 1; KRT8,

keratin, type II cytoskeletal 8; KRT19, keratin, type I cytoskeletal 19. Extracellular matrix protein: FGG, fibrinogen gamma chain. Metabolic proteins: LDHA, L-lactate dehydrogenase A chain; UGDH, UDP-glucose dehydrogenase. Chaperone, apoptotic and stress response proteins: FKBP4, peptidyl-prolyl cis-trans isomerase; PDCD6IP, programmed cell death 6-interacting protein; PRDX1, peroxiredoxin-1; PDIA6, protein disulfide-isomerase A6; SERPINA1, alpha-1-antitrypsin. Transcription regulation: HNRNPK, heterogeneous nuclear ribonucleoprotein K. Protein synthesis and degradation: PSMB2, proteasome subunit beta type 2. Signal transduction: MVP, major vault protein; ARHGAP1, Rho GTPase-activating protein 1; ARHGDIB, Rho GDP-dissociation inhibitor 2. Transport: ALB, serum albumin; APOE, apolipoprotein E. Other proteins: AKR7A2, aflatoxin B1 aldehyde reductase member 2; ANXA2, annexin A2; OGN, mimecan; ANPEP, aminopeptidase N

209x296mm (600 x 600 DPI)

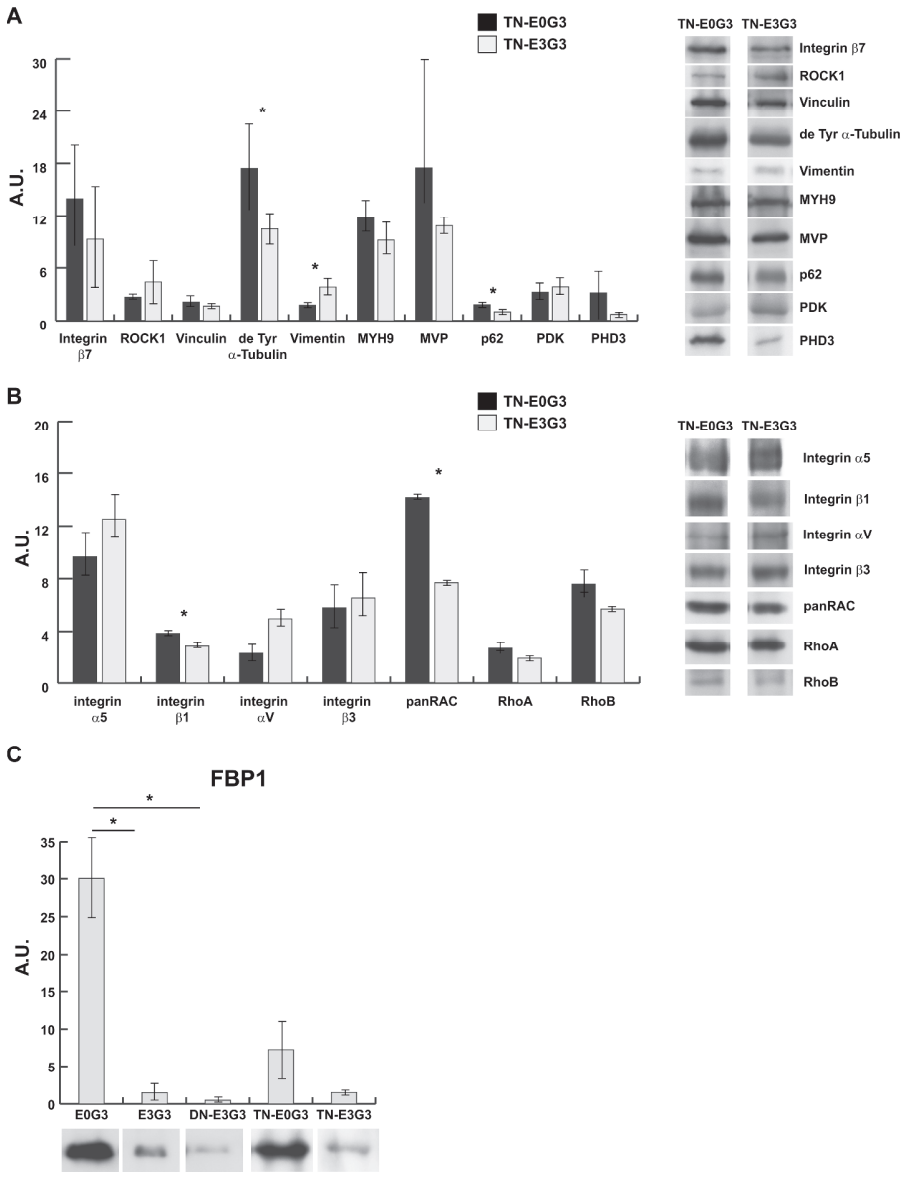

Figure 5A. Representative immunoblot images (cropped images; full lenght blots are included as Supporting Information Figure 4S) and histograms (means ± SD) of proteins abundance of Integrin β7, Rho-associated protein kinase 1 (ROCK1), Vinculin, detyrosinated α-Tubulin (de Tyr α-Tubulin), Vimentin, myosin heavy chain 9 (MYH9), major vault protein (MVP), ubiquitin-binding protein p62 (p62), pyruvate dehydrogenase kinase (PDK) and prolyl hydroxylase 3 (PHD3) in TN-E0G3 compared to TN-E3G3. Figure 5B. Representative immunoblot images (cropped images; full lenght blots are included as Supporting Information Figure 4S) and histograms (means ± SD) of proteins abundance of integrin α5, integrin β1, integrin αV, integrin β3, panRAC, ras homolog gene family member A (RhoA) and ras homolog gene family member B (RhoB) in TN-E0G3 compared to TN-E3G3. Figure 5C. Representative immunoblot images (cropped images; full lenght blots are included as Supporting Information Figure 4S) and histograms (means ± SD) of fructose-1,6-bisphosphatase 1 (FBP1) abundance.

219x286mm (600 x 600 DPI)