Proteomic Analysis of CRISPR Cas 9 Mediated Mdig Deletion in Triple Negative Breast Cancer Cells

Chitra Thakur (✉ chitra.thakur@wayne.edu)

Wayne State University <https://orcid.org/0000-0002-4356-2543>

Qian Zhang

Wayne State University

Nicholas J Carruthers

Wayne State University

Liping Xu

Wayne State University

Yao Fu

Wayne State University

Zhuoyue Bi

Wayne State University

Wenxuan Zhang

Wayne State University

Priya Wadgaonkar

Wayne State University

Bandar Almutairy

Wayne State University

Paul M Stemmer

Wayne State University

Fei Chen

Wayne State University

Research article

Keywords: mdig, mass spectrometry, signaling pathways, breast cancer, biomarker

Posted Date: October 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-72208/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

We have identified an environmentally inducible gene, mdig that predicted the overall survival in breast cancer patients. We showed that mdig regulated breast cancer cell growth, motility and invasion partially through DNA and histone methylation. However, we have lacked a comprehensive analysis of the proteomic profile of mdig in triple negative breast cancer cells.

Methods

We applied mass spectrometry to acquire global proteomic and post translational modification analysis for triple negative breast cancer cells MDA-MB-231 that had mdig deleted via CRISPR Cas 9 gene editing. Using label-free bottom up quantitative proteomics, we compared wildtype control (WT) and mdig knockout (KO) MDA-MB-231 cells and identified the proteins and pathways that are significantly altered with mdig deletion. The Ingenuity Pathway Analysis (IPA) platform was further used to explore the signaling pathway networks incorporating differentially expressed proteins.

Results

904 differentially expressed ($p < 0.005$) proteins were identified in MDA-MB-231 cells that had mdig deletion. Approximately 30 pathways and networks linked to the pathogenicity of breast cancer and populated by the differentially expressed proteins were either activated or inhibited. IPA established that the differentially expressed proteins have relevant biological actions in cell growth, motility and malignancy. This analysis provides a rich source of potential candidate therapeutic targets with potential prognostic significance in triple negative breast cancer. Data are available via ProteomeXchange with identifier PXD016688.

Conclusions

These data provide the first insight into protein expression patterns in breast cancer associated with a complete disruption of the mdig gene. Differentially expressed proteins between WT and KO MDA-MB-231 triple negative breast cancer cells provide substantial information regarding key proteins, biological process and pathways that are modulated by mdig and contribute to breast cancer tumorigenicity and invasiveness. Mdig modulated signaling pathways and hub molecules provide novel targets for the development of treatment strategies and breast cancer therapies.

Background

Breast cancer is the second leading cause of cancer related deaths in women after lung cancer in the U.S. and as of year 2019, there are more than 3.1 million women with a history of breast cancer. This is an alarming situation as about 1 in 8 women in the U.S. will develop invasive breast cancer during their lifetimes (1). Breast cancer is a clinically heterogeneous and a highly complex disease composed of

different biological subtypes. Those include human epidermal growth factor receptor 2 (HER-2), luminal A, luminal B, claudin-low, and basal-like (2); in subtypes HER2, progesterone receptor (PR) and estrogen receptor (ER) the proliferation status as measured by Ki 67 remains the standard predictive and prognostic factors for developing breast cancers (3). Among these subtypes, triple negative breast cancer (TNBC) accounts for 10 to 20% of all breast cancer cases, is highly aggressive and has the worst patient outcome. Lack of a targeted therapy, aggressive metastasis and relapse remain the top factors that make TNBC treatment challenging. Virtually all metastases occur within the first five years after diagnosis giving TNBC the worst prognosis (4, 5).

Several factors pertaining to genetics, epigenetics, environment and lifestyle are involved in the etiology of breast cancer. Mutations in the BRCA1 and BRCA2 genes, age, endogenous and exogenous exposure to hormones, obesity, alcohol consumption and cigarette smoking are some of the known risk factors (6–11). Developing an understanding of gene-environment interaction in breast cancer is a promising avenue of research. By studying the environmentally modulated genes that are implicated in breast cancer, valuable information concerning the development and progression of breast cancers will be obtained. We have recently identified a gene named as mdig, whose expression status influences the survival time of the breast cancer patients. High expression of mdig predicted poor overall survival. However, for patients who are lymph node positive, mdig expression is a favorable factor for prolonged overall survival (12). Interestingly, suppression of mdig in breast cancer cells corresponded to enhanced methylation of DNA and histone suggesting that mdig's demethylase property is a factor in the pathophysiology. These data together can be interpreted to mean that mdig is likely to promote tumor growth in the early stages of cancer but act as a tumor suppressor by inhibiting migration and invasion at the later stages (13). After the initial discovery of mdig from the alveolar macrophages of coal miners exposed to mineral dust under occupational settings (14), several studies demonstrated increased expression of mdig in a variety of human cancers especially cancers of the lung and breast (15). Mdig also has a critical role in cell growth and motility (15), in pulmonary inflammation (16, 17) and in immune regulation (18, 19). Cellular assays have shown a paradoxical role of mdig in cell proliferation, motility and invasion in lung cancer (20), where mdig being an environmental induced gene is induced upon the exposures to certain environmental agents such as silica, arsenic, and tobacco smoke (21).

Development of TNBC and its related metastasis is a complex phenomenon that is poorly understood. Moreover, the role of mdig in aggressive breast cancers is still poorly understood. Very little is known about mdig except its influence on breast cancer cell proliferation, migration, invasion and on DNA/histone methylation. Therefore, identifying key proteins modulated by mdig and the biological pathways operating in development of breast cancers is pivotal. The knowledge gained will help in identifying the novel targets of therapeutic interest. The role of mdig in cancer has been studied for several years but this work presents the first data that documents changes in global proteomic profiles of mdig depleted cells in breast cancer.

For the present study, we adopted a proteomic approach to analyze the triple negative breast cancer cells MDA-MB-231 that are knocked out for mdig via the CRISPR-Cas 9 gene editing technique. Wild type and

knockout MDA-MB-231 clones were processed for high resolution mass spectrometry and the data was analyzed for the differentially expressed proteins. The underlying signaling pathways and prominent post translational modifications were then evaluated. We have demonstrated significant pathways, protein networks and the differential accumulation of critical proteins in the mdig affected cells. EIF2 signaling, the unfolded protein response, upregulation of AKT and ribosomal proteins are interesting findings. We also report some key proteins such as MAGED2, STMN1, RACK1, HYOU1, PLAUR, RIN1 and SOD2 that might have a role in predicting the overall survival in TNBC patients and are modulated by mdig. Altogether, these results provide a strong basis for a much-needed future research regarding mdig's implication in malignant breast cancers.

Methods

Cell culture

The human MDA-MB-231 cells were purchased from American Type Culture Collection (Manassas, VA). MDA-MB-231 were cultured in DMEM F-12 medium. Cells were supplemented with 10% FBS and 1% penicillin-streptomycin (Sigma, St. Louis MO) and grown in 37 °C-humidified incubators in the presence of 5% CO₂.

Construction of the CRISPR-Cas9 vector

To generate the CRISPR-Cas9 plasmid, mdig CDS sequence was supplied into the CRISPR Design tool (<http://crispr.mit.edu/>), and single guide RNA (sgRNA) sequence targeting on exon 3 of mdig was selected. The sense and antisense primer sequences are 5'-CACCGAATGTGTACATAACTCCCGC-3' and 5'-AAACGCGGGAGTTATGTACACATTC-3', respectively. Single-stranded sense and antisense primers were annealed to form double-strand oligos in 95 °C for 5 min, and then cooled down to 25 °C for 5 min. Vector pSpCas9-2A-Blast was digested with BpiI (BbsI) restriction enzymes (Thermo Fisher Scientific, Ann Arbor, MI). sgRNA pairs and linearized vector were ligated by T4 DNA ligase (Thermo Fisher Scientific) for 10 min at 22°C. Then the ligation product was transferred into DH5α competent E. coli strain (Thermo fisher scientific) according to the manufacture's protocol.

Transfection and colonies selection

MDA-MB-231 cells, 2.5×10^5 /well in 6-well plate were transfected with Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacture's protocol. Forty-eight hours after transfection, cells were sub-cultured in 10 cm dish for 24h, followed by 2mg/ml of Blasticidin (Thermo Fisher Scientific) selection for 2 weeks. Cell colonies were collected for screening of mdig expression by western blotting. Colonies without mdig knockout were used as wild type cells (WT), whereas colonies with successful mdig knockout were designated as knockout (KO) cells.

Western Blotting

Total cellular proteins were prepared by lysing cells via sonication in 1 × RIPA buffer (Millipore, Billerica, MA) supplemented with phosphatase/protease inhibitor cocktail and 1 mM PMSF. Lysed cells were then centrifuged and supernatant isolated as protein, which was quantified using the Micro BCA Protein Assay Reagent Kit (Thermo Scientific, Pittsburgh, PA). Prior to loading onto SDS–PAGE gels, samples were boiled in 4 × NuPage LDS sample buffer (Invitrogen) containing 1 mM dithiothreitol (DTT). Samples were run on SDS-PAGE gels, and separated proteins were then transferred to methanol-wetted PVDF membranes (Invitrogen). Membranes were subsequently blocked in 5% nonfat milk in TBST and probed with the indicated primary antibodies at dilutions of 1:1000 or 1:2500 overnight at 4 °C. The next day, membranes were washed with TBST and incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies at dilutions of 1:2000 or 1:5000 at room temperature for 1 h. Immunoreactive bands were visualized through SuperSignal™ West Pico Chemiluminescent Substrate detection system (Thermo Scientific, Rockford, IL). Mdig (mouse) antibody was purchased from Invitrogen. uPAR/PLAUR antibody was from cell application Inc, MAGE-D2 from Santacruz, Anti-ORP150, Anti-RIN1, Anti-SOD2, anti-Visfatin and Filamin A were from Abcam. Cathepsin D, RACK1, Stathmin, and Tubulin were from Cell Signaling Technology (Danvers, MA, USA). All presented data are representative of at least three independent experiments.

Experimental Design and Statistical Rationale

To ensure robust detection of differential expression 5 WT and 12 KO clones were analyzed, each in duplicate. To capture variability due to sample prep and analysis, each analysis was considered to be independent for statistical analysis. Moderated t-tests with q-value correction for multiple testing was used to identify differentially expressed proteins.

Preparation of samples for mass spectrometry

Cell harvesting for proteomics analysis protocol has been adopted from (22). Thereafter the samples in duplicates were submitted to the proteomics core facility of the Wayne State University. In total 34 cell pellets were submitted for proteomic analysis. Samples were weighed and volumes matched with the addition of HPLC-grade water. 1% LiDS final was added to the samples and heated at 95°C for 5 min., followed by filtering through Pierce Handee Spin Columns (Thermo Scientific) to remove non-soluble material. Protein amount was determined by BCA Protein Assay (range from 0.278 mg to 1.064 mg). 50 µg aliquots of each were buffered with 100 mM ammonium bicarbonate (AMBIC), reduced with 5 mM dithiothreitol (DTT), and alkylated with 15 mM iodoacetamide (IAA) under standard conditions. Excess IAA was quenched with an additional 5 mM DTT. Samples were diluted to decrease LiDS to 0.1% then an overnight digestion was performed with sequencing-grade trypsin (Promega, Madison, WI) in 100 mM AMBIC, 0.3 M urea, and 15% acetonitrile. The next day, detergent was removed from the samples using Pierce Detergent Removal Columns. Samples were speed vac'ed to dryness and solubilized in 0.1% FA for analysis. The peptides, 4 µg per analysis, were separated by reversed-phase chromatography (Easy Spray PepMap RSLC C18 50 cm column, Thermo Scientific), followed by ionization with the Easy Spray Ion

Source (Thermo Scientific), and introduced into a Fusion Orbitrap mass spectrometer (Thermo Scientific). Abundant species were fragmented with collision-induced dissociation (CID)

Mass spectrometry data analysis

For protein quantification and pathway analysis, mass spectrometry raw files were searched against the Uniprot human complete database downloaded 2017.07.14 (20 201 entries) using MaxQuant v1.6.2.10 with the default version of the Andromeda search engine. Match between runs was enabled and just one peptide was required for protein quantification. All other parameters were left at their default values including: tryptic cleavage with at most 1 missed cleavage was the protease, methionine oxidation and protein N-terminus acetylation were variable modifications, cysteine carbamidomethylation was a fixed modification, fragment ion tolerance was 0.5 Da, precursor tolerance was 20 ppm for the first search and 4.5 ppm for the second, peptide identifications were allowed at a 1% false discovery rate as determined by a reversed database. For PTM analyses the same raw files were searched against the same database using Proteome Discoverer v2.3.502 to take advantage of the percolator algorithm for sensitive peptide identification. Two independent searches were conducted for histidine oxidation and for lysine di- and trimethylation plus lysine acetylation. For the histidine oxidation search both histidine and methionine oxidation were set as variable modifications. For lysine acylation analysis, lysine di-methylation, lysine trimethylation, lysine acetylation and methionine oxidation were set as variable modifications. All other aspects of the Proteome Discoverer searches were the same. Sequest HT was the search engine. Trypsin with at most 1 missed cleavage was the protease. Cysteine carbamidomethylation was set as a fixed modification. MS1 mass tolerance was set to 10 ppm and MS2 mass tolerance was set to 0.6 Da. For all analyses, peptide spectra matches were accepted at a 1% false discovery rate as determined by a reversed database search. PTMRS (23) was used to assess PTM localization confidence. Peptide area under the curve was used to generate quantitative values.

Statistical analysis used R v3.4.3. Protein abundances were normalized to have the same median and differential abundance between wild type and knock-out samples was determined using a moderated t-test (24) with q-value correction for false discoveries (25). To capture variability due to sample prep and analysis, each sample was considered to be independent for statistical analysis. PTM abundance changes were assessed using the moderated t-test with q-value correction on peptide level data. Bulk changes in PTM abundance were assessed using a permutation test as follows. The mean t-statistic for all peptides bearing that PTM was calculated. Then mean t-statistics for 10000 random draws of the same number of peptides from the entire dataset were calculated. A p-value was calculated as the fraction of draws that had a mean t-statistic more extreme than the PTM mean.

Bioinformatics

Sets of proteins obtained from the MS data were processed using The Database for Annotation, Visualization and Integrated Discovery (DAVID) version 2.0 (<http://david.abcc.ncifcrf.gov/home.jsp>). Further, Protein ANalysis THrough Evolutionary Relationships (PANTHER) database v 6.1 (www.pantherdb.org) was used for gene ontology (GO) annotation. QIAGEN's Ingenuity Pathway Analysis

(IPA®, QIAGEN Redwood City, <http://www.ingenuity.com/>) software was used to investigate the functional and canonical pathways that were enriched in the differentially expressed proteins. Proteins that responded to mdig knock out (moderated t-test $p < 0.005$, $n = 8$) were submitted to IPA. All proteins identified in the study and pathways were considered significantly different with $p < 0.05$.

Kaplan–Meier survival analysis

A Kaplan–Meier survival database that contains survival information of breast cancer patients and gene expression data obtained by Affymetrix HG-U133 microarrays. The probe set for the indicated genes were used that scored to be the best among the other probe sets available by using JetSet best probe detection tool (26). Survival curves resulting in p values of < 0.05 between the gene higher (gene^{high}) and gene lower (gene^{low}) groups were considered significantly different.

Results

Generation of mdig knockout cells by CRISPR Cas 9

To create mdig knock out cells, human triple negative breast cancer cells, MDA-MB-231, were transfected with pSpCas9-2A-Blast vector containing sgRNA that targets the third exon of the mdig gene. Thereafter blasticidin selection was performed for two consecutive weeks and the colonies obtained were screened for mdig expression by western blot (Fig. 1A). Altogether we obtained 5 WT and 12 KO clones and after screening them for mdig expression at the protein level, we prepared them for proteomic analysis. Each of the WT and KO clones was cultured and analyzed in duplicate. 2 of the 34 samples were removed from further analysis for quality control reasons. 5739 proteins were detected, and 5711 were quantified in at least 1 sample. 3569 were quantified in all samples.

Principal component (Fig. 1B) and cluster analysis (not shown) indicated some within-group heterogeneity. One KO clone in particular, KO#3, appeared to be more similar to WT samples than to other KOs. The gene knock out for that clone was confirmed by western blot and by the mass spec data. The clones KO#10, KO#3 and WT#5 were removed from the dataset and not used in any further analysis. Protein data for protein quantitative analysis (MaxQuant) and peptide data supporting protein quantitative analysis (MaxQuant) has been shown in Supplementary Table S1 and Supplementary Table S2 respectively.

Identification of the differentially expressed proteins for their class and gene ontology annotation

LC-MS/MS data were analyzed to determine the fold change (FC) as a normalized ratio for KO compared to WT control cells. This first screening of the raw data identified a set of proteins for which abundances increase or decrease in the MDA-MB-231 mdig KO cells. The analysis consisted of the unique protein IDs, with their fold change, p value and t statistics as a function of KO/WT. Thereafter the differentially expressed proteins were classified based on gene ontology designations such as molecular function, cellular component, and biological process using the PANTHER classification system (Fig. 2). A total of

26 protein classes were identified at the $p < 0.05$ level. Those categories are: calcium binding, cell adhesion molecules, cell junction proteins, chaperones, cytoskeleton, immunity, enzyme modulator, hydrolase, isomerase, ligase, lyase, membrane traffic proteins, nucleic acid binding, oxidoreductase, receptor, signaling molecule, storage proteins, structural proteins, surfactants, transcription factor, carrier proteins, transferase, transmembrane receptor regulatory, transporter and viral proteins categories. Among them, proteins in the nucleic acid binding (PC00171) class were the most prevalent, encoding 340 genes for this category. According to biological process, most of the proteins belonged to the subcategories of biological adhesion, biological regulation, cell proliferation, biogenesis, cellular process, development process, immune system process, localization, metabolic process, multicellular organismal process, reproduction and response to stimulus. Among these, cellular and metabolic process were highly elevated with increased number of genes assigned to them compared to other subcategories. According to molecular functions, majority of the proteins belonged to functions pertaining to binding, catalytic activity, molecular function regulator, molecular transducer activity, structural molecule activity and transporter activities. Binding and catalytic activity were found to be the highest among the group. Finally, according to cellular components, most of the proteins were localized to the cell junction, cells, extracellular regions, membrane, organelle and protein containing complex. Among them, elevated regions were the proteins belonging to the cellular compartment, organelles and protein containing complex (Supplementary Fig. 1). These patterns of protein distribution suggest that mdig significantly affected the family of proteins that are essential for important biological and molecular processes such as binding, metabolism, immunity, and catalytic activities implicated in triple negative breast cancer. It also warrants a further detailed investigation of the individual genes and protein related to such biological functions manifested in breast cancer.

Canonical pathway analysis reveals key signaling cascades affected by mdig

We identified the ten proteins with the greatest magnitude change in abundance in KO over WT MDA-MB-231 cells (Table 1). Once the differentially expressed proteins were identified, next step was to query the role of those proteins in the pathogenesis of breast cancer.

Table 1

Top 10 proteins consisting the highest magnitude change in KO over WT MDA-MB-231 cells ($p < 0.005$), as revealed by proteomics data set obtained through mass spectrometry

Symbol	Identifier UniProt/Swiss-Prot Accession	Description	Fold change
CTSD	P07339	Cathepsin D	11.063
MAGED2	Q9UNF1	Melanoma-associated antigen D2	10.443
FLNA	P21333	Filamin-A	9.662
ABHD16A	O95870	Abhydrolase domain-containing protein 16A	9.595
STMN1	P16949	Stathmin	9.118
NPC2	P61916	Epididymal secretory protein E1	8.845
RACK1	P63244	Receptor of activated protein C kinase 1	8.820
HIST1H2BA	Q96A08	Histone H2B type 1-A	8.673
IQGAP1	P46940	Ras GTPase-activating-like protein IQGAP1	8.603
HUWE1	Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1	8.473
RIOX2	Q8IUF8	Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA	-8.521
KRI1	Q8N9T8	Protein KRI1 homolog	-8.272
CCDC51	Q96ER9	Coiled-coil domain-containing protein 51	-8.241
PLAUR	Q03405	Urokinase plasminogen activator surface receptor	-8.203
HYOU1	Q9Y4L1	Hypoxia up-regulated protein 1	-7.368
SOD2	P04179	Superoxide dismutase [Mn], mitochondrial	-7.173
RIN1	Q13671	Ras and Rab interactor 1	-7.096
NOP58	Q9Y2 × 3	Nucleolar protein 58	7.083
NAMPT	P43490	Nicotinamide phosphoribosyltransferase	-7.006
MCCC2	Q9HCC0	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	-6.863

Characteristic alterations in signaling pathways and regulatory networks are expected between disease vs healthy cells. We used the Ingenuity Pathway Analysis (IPA) Software (IPA; Ingenuity® Systems, Qiagen) to identify the major biological pathways perturbed in mdig KO cells. IPA was used to interpret the differentially expressed proteins in terms of predominant canonical pathways and derivation of mechanistic networks. Canonical pathways are well defined biochemical cascades resulting in unique

functional biological consequence. Performing the canonical pathway analysis of our dataset via IPA revealed 501 canonical pathways. The top 5 canonical pathways ($p < 0.05$) according to the number of identified proteins were EIF2 Signaling (54), Isoleucine Degradation I (8), Unfolded Protein Response (12) Regulation of eIF4 and p70S6K Signaling (31) and Caveolar-mediated Endocytosis Signaling (14) (Fig. 3). The Regulation of eIF4 and p70S6K Signaling and Unfolded Protein Response pathways have been elaborated in Fig. 4 showing the upregulated and downregulated proteins and their cellular localization.

Among the pathways that are overrepresented in mdig KO cells, EIF2 signaling was the topmost canonical pathway found in our analysis. Interestingly, PI3K, and AKT were upregulated with mdig silencing while RAS and eIF4a were downregulated. Previous reports identified the PI3K-Akt pathway as an enhancer of the expression of EMT resultant transcription factors such as Snail, Slug, ZEB1 and ZEB2 that promoted the EMT and resulted in an elevation of the cancer cell motility (27, 28). This suggests an increased motility potential of breast cancer cells upon the loss of mdig protein. Among the Unfolded Protein Response family of proteins, several heat shock proteins such as Hsp70 and Hsp40 were upregulated while TNF receptor associated factor 2 was downregulated in mdig KO cells. Analyzing the protein profiles belonging to the canonical pathway, Regulation of eIF4 and p70S6K Signaling, revealed a plethora of ribosomal proteins that were upregulated in the KO cells, such as ribosomal protein S16, S8, S9, S26, S2, S15a, S3, S6, S7, S21, S24, S3a, S27a, S10, S17, S20, S23 and S4X-linked. Since ribosome biogenesis is important for cancers; upregulation of ribosomal proteins in response to mdig deletion is a striking observation that needs further investigation. The filamin family of proteins such as filamin A, filamin B, and filamin C were upregulated in the KO cells. Notably, another interesting protein, flotillin 1 was found to be upregulated (Supplementary Fig. 2). Filamin proteins have been implicated in cancer progression while increased levels of flotillin 1 promoted cell proliferation, migration, tumorigenicity and lymph metastasis in breast cancer studies (29, 30).

These data indicate the important signaling pathways implicated in breast cancer upon mdig knockdown. The individual differentially regulated proteins in the top five canonical pathways certainly are attractive targets for further investigation where mdig is directly involved in the ribosome biogenesis, and the metastasis of triple negative breast cancers.

Cellular and molecular function gives insight into the differential biology of breast cancer cells affected by mdig

IPA-based protein network analysis was performed using all identified proteins upon mdig knockdown in TNBC cells. We identified 500 molecular and cellular functions associated with mdig deletion. The top five scoring function categories were evaluated for the predicted effect of mdig deletion on the activation status. Processes that are integral to cell growth and tumorigenesis were found, including: Protein Synthesis (177 associated proteins), RNA Damage and Repair (45 associated proteins), RNA Post-Transcriptional Modification (104 associated proteins), Cell Death and Survival (362 associated proteins) and Nucleic Acid Metabolism (74 associated proteins). These processes orchestrate the vital molecular

functions such as protein expression, decay of mRNA, processing of rRNA, necrosis and metabolism of nucleic acid component or derivative respectively (Fig. 5A). Among them, an overall increase in protein synthesis and an overall decrease in the RNA post-transcriptional modification and cell death & survival were found in the KO category (Fig. 5B). Individual proteins belonging to these molecular and cellular functions with their upregulation and downregulation status have been depicted. Additionally, we found an overall decrease in inflammation with mdig loss (supplementary Fig. 3A). This is interesting as our *in vivo* studies on mdig knockout mice suggested a decreased inflammatory status of the mice upon silica exposure (17) further corroborating the current results.

The top enriched proteins associated with diseases and the disorders with the most proteins involved belonged to the categories of Cancer, Organismal Injury and Abnormalities, Tumor Morphology, Cardiovascular Disease and Developmental Disorder. The top network identified was associated with cancer. This network consists of 458 proteins in our proteomic data set (Fig. 6A). These results suggest the involvement of mdig in regulating the process of transformation in breast cancer. The IPA also predicted the upstream regulatory molecules that are either activated or inhibited on the basis of the observed protein expression changes allowing us to understand the underlying causal network. In our analysis we found the top 5 upstream regulators to be: MYCN, NFE2L2, MYC and TCR. Moreover, MYCN was activated upon mdig knockdown (Fig. 6B). Also, Myc is known as a classical upstream regulator of mdig (31).

Post translational modification and disease-based protein network analysis reveal the catalytic activity of mdig in the oxidation and demethylation process

PTMs can change the dynamics and affinity aspects of protein-protein interactions and often serve as the basis for modulation of signaling pathways implicated in breast cancer. Epigenetically relevant PTMs such as acetylation and methylation contribute to transcription regulation and have well established roles in cancer.

Mdig catalyzes both histidine oxidation (32) and tri-methyl lysine demethylation (33). Therefore, spectra were searched for histidine oxidation and lysine acylation to quantify their changes in response to mdig knockout (Table 2). Changes in PTM abundance were assessed using the number of peptides that were significant ($q < 0.1$) and whether the mean t-statistic was different from 0. Mdig catalyzes histidine oxidation at His39 of the 60S ribosomal protein L27a (Uniprot accession: P46776) (32). The tryptic peptide containing His39 from the 60S ribosomal protein L27a was detected in both the oxidized and native forms (Fig. 7A, Supplementary Fig. 3B). The peptide sequence (GNAGGLHHHR) has no residues that can be non-enzymatically oxidized so the oxidized form must be the product of an enzymatic reaction. The abundance of the oxidized form was decreased in mdig KO samples ($q = 0.00030$,

moderated t-test, $n = 8$). The native, non-oxidized, form was detected only in mdig KO samples demonstrating that the knockout removed a specific enzymatic activity from the cells.

Table 2

Evaluation of changes in PTM abundance between MDA-MB-231 KO and WT samples. * p-value of a permutation test for the mean t-statistic being different from 0

PTM	peptides quantified	peptides increased ($q < 0.1$)	peptides decreased ($q < 0.1$)	modification mean t-statistic	modification p-value
oxidized histidine	98	7	4	0.01	0.571
di-methyl lysine	163	28	14	0.37	0.021
tri-methyl lysine	84	11	9	0.33	0.104
acetyl lysine	104	10	9	0.35	0.067
All peptides	65 281	6725	8513	0.07	n/a

In total, 98 peptides with candidate histidine oxidation sites were quantified and 11 were found to be significantly different between KO and WT samples ($q < 0.1$, moderated t-test, $n = 8$). The mean t-statistic for histidine oxidized peptides was near 0 (Table 2) indicating that they weren't changed in a uniform direction by mdig knockout. To ensure that methionine oxidation didn't interfere with our analysis we limited the set of histidine oxidized peptides to those that had no methionine residues or that had confident localization of the oxidation site to histidine by PTMRS (23). The mean t-statistic for that selected group was still approximately 0 (not shown). These data confirm the activity of mdig to catalyze the oxidation of His39. However, they don't provide evidence that mdig oxidizes other His residues outside of 60S ribosomal protein L27a His39.

In addition to catalyzing His oxidation, mdig catalyzes the demethylation of tri-methylated lysine 9 of Histone H3 (33). Hence Lysine di- and tri-methylated and acetylated peptides were evaluated for changes in abundance in response to mdig deletion. Dimethylated lysine containing peptides had an overall increase in abundance in mdig KO samples relative to WT (Table 2). This is supported by the number of dimethyl-lysine peptides that were increased in abundance, 28 vs 14 decreased ($q < 0.1$, moderated t-test, $n = 8$), and also by the mean t-statistic for lysine di-methylated peptides that was positive (0.37, $p = 0.021$, permutation test for difference from 0). The change in abundance was confirmed in a smaller set of 63 very high confidence peptides (percolator posterior error probability, (PEP) < 0.001). Those 63 very high-confidence lysine di-methylated peptides had a greater increase in abundance than the larger set (mean t-statistic of 0.83, $p = 0.0011$) demonstrating that the increase in abundance was not just limited to low quality peptide identifications. Tri-methyl lysine and acetylated lysine also had positive mean t-statistics

but did not meet our statistical threshold. These results suggest an important regulatory role of mdig on the 60S ribosomal protein L27a and on the methylation of lysine residues on histone proteins; which are likely to affect the transcription of critical genes implicated in TNBC.

PTM peptides that were differentially abundant between KO and WT has been shown in supplementary data Table S3. Peptide data for Histone oxidation analysis (Proteome Discoverer) is shown in Supplementary Table S4 and Peptide data for Lysine acylation analysis (Proteome Discoverer) is shown in Supplementary Table S5.

Individual peptides with their candidate PTM sites have also been shown for Acetylated Peptides (Supplementary Table S6), Dimethylated Peptides (Supplementary Table S7), HisHydroxylated Peptides (Supplementary Table S8) and Trimethylated Peptides (Supplementary Table S9).

The next level of regulation is the interaction of signaling networks and regulatory pathways. IPA identified 25 interaction networks built with 35 focus molecules that were affected by mdig knockdown. The five most affected gene networks as determined by IPA and a detailed interaction in the most significant networks has been shown in Fig. 7B and C. Genes with different expression patterns predominantly mapped to the networks associated with protein synthesis, RNA post transcriptional modification, DNA replication, recombination and repair. This shows an important role of mdig in regulating the genes associated with genomic stability and cancer, further indicating its influence on the pathogenicity of TNBC.

Validation of top identified differentially regulated proteins and their relevance in breast cancer growth, motility and metastasis

After having established the changes in protein abundance in mdig KO cells we selected the top five upregulated and downregulated proteins as determined by proteomic profiling and IPA. The rationale for selecting these proteins comes from the top ready molecule list as provided by the IPA and their specific relevance in breast cancer metastasis upon literature survey. Upregulated proteins consisted of CTSD, MAGED2, FLNA, STMN1 and RACK1, while downregulated proteins consisted of PLAUR, HYOU1, SOD2, RIN1 and NAMPT. To determine if an association exists between these proteins and TNBC, we performed western blotting analysis of these specific protein groups in MDA-MB-231 cells expressing mdig (WT) and deleted mdig (KO) (Fig. 8A). Five WT clones and twelve KO clones were tested using Tubulin as a loading control. We found a consistent pattern of altered protein expression in the TNBC cells, where CTSD, MAGED2, STMN1 and RACK1 were upregulated in the KO cells and PLAUR, HYOU1, SOD2, RIN1 and NAMPT were downregulated in the KO cells. Except for FLNA, all proteins at western blot assay corroborated with the IPA findings and hence validated our results. This suggests that these proteins are potential candidate biomarkers for TNBC and are strongly associated with breast cancer. This analysis also demonstrates mdig's regulation on the abundance of these proteins which are implicated in motility, EMT, genomic stability, thereby governing the overall malignant phenotype of aggressive breast cancers.

Evaluation of the identified proteins in predicting disease prognosis for the survival of breast cancer patients

To explore whether the above proteomics findings are clinically relevant for breast cancer patients, the top proteins identified as differentially abundant in *mdig* KO cells were evaluated for their performance in predicting disease prognosis and overall survival of breast cancer and TNBC. Survival data from 3951 breast cancer patients and 618 TNBC patients were obtained from an online gene profiling database (Kaplan Meir plotter, (34)). Abundance of the top identified and validated proteins in the *mdig* KO cells was evaluated for correlation to patient stratification based on the high expression of the proteins under study (Fig. 8B). In breast cancer patients, high expression of STMN1, NAMPT, PLAUR, and SOD2 predicted poor overall survival, whereas FLNA, MAGED 2, RACK1, HYOU1, and RIN 1 predicted better overall survival. However, in TNBC patients, high expression of MAGED2, and STMN1 predicted poor overall survival, whereas, RACK1, HYOU1, PLAUR, RIN1 and SOD2 predicted better overall survival. The differential regulation of such proteins by *mdig* is an important finding. Involvement of these proteins in the regulation of cell proliferation, motility, invasiveness, cancer metabolism and ER stress make them ideal candidates that can be exploited in breast cancer for therapeutic efficacy.

Discussion

Triple negative breast cancer is a malignant form of breast cancer with aggressive clinical characteristics. It has the worst patient prognosis and currently lacks targeted therapy (35, 36). Therefore, there is an urgent need to understand the molecular and biological mechanisms governing the malignant behavior of TNBC and its pathogenicity.

We have recently identified a gene named as *mdig* which predicts poor prognosis in breast cancer (12). Initially it was identified as an oncogene for lung cancer (33) and was also expressed in other cancer type with roles in cell growth and motility (15). In breast cancer we found that high *mdig* expression predicts poor overall survival of patients, however, predicted a better survival of the patients who had lymph node or distal organ metastasis, further suggesting that *mdig* is favorable for metastatic patients (12). In breast cancer cells MDA-MB-231, silencing *mdig* using an siRNA approach, enhanced the DNA and histone methylation and the migration of the cells (13), an important attribute of *mdig*. These studies indicated that *mdig* is important for the tumor growth of the early stage breast cancers but at the later advanced stage, *mdig* expression is likely to benefit the patient as it inhibits the migration and invasion of breast cancer cells.

Cancer is indeed a “disease of pathways” (37) and proteins function through complex biological pathways that involve many proteins working together. The pathways and functions that are over represented in *mdig* overexpressing cells remain the most likely cause of cancer (38) and hence determination of critical pathways and proteins enriched and altered in healthy vs cancerous cells is essential for our understanding the biological and molecular mechanisms driving the process of carcinogenesis. It is in such scenarios where the state-of-the-art technology like proteomics in conjunction

with integrated bioinformatics is essential in identifying the cancer associated signaling pathways and networks and, thereby, assisting in cancer biomarker discovery. The CRISPR/Cas 9 system is an indispensable tool applied in various kinds of human cancers and has been accelerating cancer research (39). In breast cancer, CRISPR technology has enabled breakthroughs in diagnosis, treatment and drug resistance related research (40).

Our previous studies on TNBC cells silenced for mdig via short interfering RNAs have yielded some important information about the regulatory effects of mdig on cell motility and invasion as well as on DNA and histone methylation (13). Though that model is a transient knockdown for mdig, the data generated suggest that mdig negatively regulates breast cancer cell's migration and invasion potential. The data also show that mdig expression is inversely proportional to the extent of DNA methylation. Still, the mechanisms underlying the influence of mdig on breast cancer cells are poorly understood and have not previously been explored at the system level that proteomics technology allows.

To gain a better understanding about the function of mdig in cell growth, motility and invasion in breast cancer we applied CRISPR-Cas9 gene editing technique to knockout mdig in human triple negative breast cancer cells MDA-MB-231. Those wild type and mdig deleted cells were subjected to global proteomic analysis. Analyses of these data produced significant findings related to differentially expressed proteins as well as signaling pathways and regulators modulated by mdig.

Proteomic profiling identified important proteins that are differentially expressed in TNBC cells upon mdig deletion. Loss of mdig resulted in an increase in abundance of proteins that are implicated in cell proliferation, angiogenesis and metastasis of breast cancer. Among them are Cathepsin D, MAGED2, Filamin A, Stathmin 1 and RACK 1. The cathepsin family of proteins are known to provide a necessary activity for the metastasis of cancer cells and to degrade the extracellular matrix and collagen. Cathepsin D is also overexpressed in breast cancer (41–43) and predicts a poor prognosis (44–46). It represents as a marker for invasive potential and aggressive behavior in high grade carcinomas (47) and stimulates the cell growth, angiogenesis and metastasis (48–50). MAGED2 is found to be elevated in primary tumors and to be upregulated in metastasis (51). It is noteworthy that in this present report we identify MAGED 2 as a novel protein that is increased in response to mdig knockdown. Stathmin 1 was also upregulated in KO cells. STMN1 is a microtubule destabilizing protein whose expression is associated with the breast cancer proliferation (52, 53). In breast cancer patients, high STMN1 correlates with poor prognosis (54, 55). Moreover, in breast cancer patients, elevated STMN1 is linked with high histological grade and low ER, PR expression status (52) and related with aggressive phenotypes accompanied with cancer stem cell marker expression (56). Interestingly, another protein that is associated with cell growth, adhesion invasion and metastasis is the RACK 1 protein which was upregulated in response to mdig knockout. In fact both *in vitro* and *in vivo* studies have shown that RACK 1 promotes the proliferation, invasion and metastasis of breast cancer (57) and remains one of the independent predictors for poor clinical outcome in breast cancer (58). Perhaps RACK1 was not only associated with breast cancer malignancy, but its overexpression is implicated in the growth and metastasis of several other cancer types such as lung cancer, gliomas, colon cancer, prostate cancer, liver cancer, epithelial ovarian cancer and squamous cell

carcinoma of the esophagus (59). Finally, the top five proteins found to be upregulated in our analysis included Filamin A. Several studies have reported that the overexpression of Filamin A is associated with highly metastatic cancers of the prostate (60), skin (61) and brain (62) and that FLN A is involved in the progression of neoplasia (63).

Western blot validation of all the upregulated proteins showed a similar trend of increase in the KO cells, with the exception of Filamin A for which we found a decrease in the KO cells. This is quite interesting and is relevant to the metastasis of TNBC cells as reported in the current study. FLNA has dual functions and can promote opposite outcomes depending upon its subcellular localization. In the cytoplasm, FLNA is able to facilitate cell growth and metastasis, however, its presence in the nucleus causes an inhibition of cell growth and metastasis (64). These data indicate that cells with an appropriate amount of FLNA are likely to leverage some benefits during metastasis and that the abundance of FLNA will influence the metastasis of cancer cells based on its subcellular localization. In this regard, FLNA inhibition was also found to reduce the metastatic potential of cancer (65) and silencing of FLNA in MDA-MB-231 cells was sufficient to inhibit cellular migration and invasion (66).

Among the proteins that were downregulated upon mdig knockout are the families of proteins implicated in cancer metabolism and reprogramming, DNA repair pathways, cell motility, tumor suppressor functions and stress related cellular response. These downregulated proteins were PLAUR, SOD2, RIN 1, HYOU1, and NAMPT.

Increased PLAUR expression has been found in aggressive breast cancers such as in TNBC, a subset of Her 2 + breast cancer, and in tamoxifen refractory breast cancer (67–69). PLAUR, is known to regulate the ubiquitin proteasome system during DNA damage response and silencing PLAUR impairs the DNA repair process (70). Interestingly in MDA-MB-231 cells and HeLa cells, PLAUR plays a significant role in regulating the homologous recombination (HR) DNA repair pathway (71). PLAUR is a potential molecular target for breast cancer owing to its accessibility on the surface of cancer cells (72). Strikingly, we also observed decreased enzyme manganese superoxide dismutase 2 (SOD2) in the KO cells. This is a very significant finding since loss of SOD2 represents a phenotype of tumor initiation and therefore an indicative of the tumor suppressor role of SOD2 particularly due its $O_2^\bullet$ scavenging role during the process of tumorigenesis (73). Apparently, decreased SOD2 activity and hiked up ROS are the prerequisite for the metabolic reprogramming of cancer cells (74). Additionally, forced SOD2 overexpression in cancer cells is able to decrease the metastatic potential and undermine the malignant phenotype of the cancers (75, 76). In breast cancer, SOD2 is epigenetically regulated where SOD2 expression is repressed primarily due to the hypo acetylation and hypo methylation of histone proteins thereby inhibiting the functions of transcription factor (77). More-over there is a switch from SOD2 to SOD1 during the transformation process in breast cancers (78) and SOD2 is downregulated in malignant breast cancer cells compared to their normal cell counterparts (79). Mdig is a histone demethylase and hence it is likely to exert its demethylation or hypo methylation activity on the transcriptional status of the SOD2 gene. However, this needs to be further investigated. Mdig KO cells also exhibited a decreased protein RIN1. Notably, RIN 1

downregulation has been associated with invasion and poor overall survival in liver cancer (80) and RIN 1 silencing resulted in increased motility of epithelial cells (81).

In breast cancer, RIN1 expression is decreased in neoplastic tissues as compared to normal breast tissues (82). Since RIN1 contributes to cell motility the decreased expression of RIN1 suggests that mdig influences the motility and malignant behavior of TNBC cells by downregulating the expression of RIN 1 and other proteins that promote metastasis. Among other downregulated proteins is the HYOU1, a novel HSP implicated in the ER stress response that mediates anti-apoptotic signals in certain cancers such as breast cancer (83), bladder cancer (84) and prostate cancer (85). Another downregulated protein in mdig KO cells is the multifunctional enzyme NAMPT. This protein is usually overexpressed in lymphoma (86) and in solid cancers of prostate, stomach and colon (87–89). In breast cancer, NAMPT downregulation brought via mir-206 resulted in decreased survival of breast cancer cells (90) and that the expression of NAMPT affects the metastasis and adhesion of breast cancer cells by inhibiting the functions of integrin proteins (91). NAMPT in conjunction with Her 2 and VEGF also serves as a biomarker for the diagnosis and prognosis of human breast malignancies (92). NAMPT downregulation in response to mdig silencing is a novel finding which needs further investigation. Altogether, our results provide evidence that downregulation of mdig initiates a destabilization of the breast cancer genome most likely via affecting the DNA repair pathways and enzymes involved in ROS metabolism. The subsequent effects on the gene transcription then increase the likelihood of tumorigenesis. It also negatively regulates the expression of proteins involved in adhesion, invasion and in aggressive malignant phenotype in breast cancer.

Carcinogenesis is a multistep process with alterations in signaling networks resulting from genetic, epigenetic, and environmental changes being accumulated at distinct stages of carcinogenesis. Our analysis showed the top signaling pathways that were enriched in the TNBC cells after mdig knockout. The higher scoring pathway is Eukaryotic Initiation Factor 2 (eIF2) signaling which is responsible for initiating translation in eukaryotes. In fact, dysregulated mRNA translation is critical in the etiology and pathogenesis of human malignancies. Hence it is widely reported that aberrant translation of oncogenes, tumor suppressors, and eukaryotic translation initiation factors are of paramount importance in the proliferation of cancer cells (93). IPA of the mdig KO proteome highlighted elevated level of PI3K and AKT. Increases in abundance of these proteins suggests that the PI3K-Akt pathway increases in activity. That increased activity would facilitate expression of EMT related transcription factors Snail, Slug, ZEB1 and ZEB2 thereby promoting EMT and enhanced motility of cancer cells (27, 28).

Metabolic processes in tumors change as the cancer cells become dependent on amino acids as their source of energy and metabolites. Branched chain amino acids (BCAA) such as leucine, isoleucine, and valine are preferentially up taken by the tumors (94). The degradation of branched chain amino acids then serve as an energy supply for the cancer cells (95). The Isoleucine Degradation I pathway appeared as the top signaling pathways affected by mdig deletion in TNBC cells. In breast cancer cells, there has been an increase in the downstream branched chain amino acid catabolic enzymes (96) also suggesting the roles of reprogrammed metabolism as an early event in the BRCA-1 tumorigenesis (97).

Activation of the unfolded protein response (UPR), confers a resistance to therapy on breast cancer cells and increases the likelihood of recurrence (98). Mdig loss resulted in the enriched UPR pathway suggesting the accumulation of the misfolded proteins due to the impairment of protein folding occurs in the KO cells. Alternatively, this indicates the positive influence of mdig in the development of endoplasmic reticulum stress and the UPR signaling. The increase in the heat shock proteins with mdig loss suggests the upregulation of the coping mechanisms in the KO cells in response to the EnR stress.

The increased abundance of ribosomal proteins in the canonical pathway pertaining to the regulation of eIF4 and p70S6K signaling in KO cells is a striking observation. Mdig is involved in the ribosomal biogenesis (99) and is implicated in ribosomal RNA transcription (33). Mdig belongs to the family of 2-Oxoglutarate (2OG)-dependent oxygenases' (2OG-oxygenases) that catalyzes the ribosomal protein histidyl hydroxylation where mdig targets the His-39 of Rpl27a within the large (60S) subunit (100). The presence of mdig in the nucleolus is indicative of its critical role in the ribosome biogenesis. Accumulation of several ribosomal proteins in the KO cells, especially S6, is an indication that mdig modulates expression of ribosomal proteins, a role that might have implications for neoplastic transformation. Also the upregulation of AKT in the KO cells reflects the role of mdig in tumor survival, EMT and metastasis, as AKT activation is a hallmark of several cancers (101).

Mdig has two catalytic activities that change PTMs, histidine oxidation and tri-methyl lysine demethylation (32, 33). PTM analysis of the mass spectrometry data identified the known target of mdig oxidation activity, histidine 39 of 60S ribosomal protein L27a. The decreased abundance of the oxidized form and increased abundance of the native form in the mdig knock out samples (Fig. 7A) are consistent with the known activity of mdig. No other peptides with oxidized His had a change in abundance of the magnitude observed for histidine 39 of 60S ribosomal protein L27a and there was no apparent global increase in abundance for histidine-oxidized peptides. This evidence does not support a role of mdig as a general histidine oxidatase and reinforces the selective action on the L27a protein. The lysine PTMs di-methylation, tri-methylation and acetylation were also tested for quantitative differences in mdig KO cells compared to WT. Because mdig is a demethylase, increased tri-methylation and decreased di-methylation at mdig target lysine residues could be expected. Our data indicate that di-methyl lysine was more abundant in mdig knockouts compared to wild-type. In addition, our results were suggestive of global changes in abundance for lysine tri-methylated and acetylated peptides ($p = 0.104$ and $p = 0.67$ respectively). These results suggest that mdig has a global impact on lysine acetylation in addition to its specific de-methylase activity.

Metabolic pathways known to be modulated in cancer were also implicated in our analysis. These include the glycolysis, TCA cycle and the Pentose phosphate pathway (PPP) (data not shown). Mdig loss affected critical enzymes involved in glycolysis where its loss resulted in the upregulation of glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase, phosphopyruvate hydratase, and pyruvate kinase, while 6-phosphofructokinase was found to be downregulated in the KO cells. Within the TCA cycle, downregulation of aconitate hydratase, 2-oxoglutarate dehydrogenase E1 component like, succinate-CoA ligase and succinate dehydrogenase were observed in the KO cells. In the PPP,

upregulation of glucose-6-phosphate 1-dehydrogenase, phosphogluconate dehydrogenase and transaldolase were found whereas ribose-5-phosphate isomerase was upregulated. These findings suggest an important role of mdig in regulating the breast cancer cell growth and survival most likely by interfering with the prominent signaling cascade related to energy supply, anabolism and catabolism.

The heterogeneity of the TNBC and lack of effective therapeutic targets along with insufficient predictive biomarkers are reasons for the challenges associated with TNBC therapy. Malignant transformation includes changes in protein abundance. Monitoring these changes at the protein level provides unique protein signatures that might facilitate effective diagnosis and prognosis. The high throughput proteomics study of the TNBC cells has provided a large and rich dataset that has allowed us to stratify systemic differences between the MDA-MB-231 cells with and without mdig. The top differentially regulated proteins have been validated at the protein level and have been found to predict disease prognosis both in breast cancer and in TNBC. Among them, high expression of STMN1, NAMPT, PLAUR and SOD2 predict poor overall survival in breast cancer patients whereas, high expression of FLNA, MAGED2, RACK1, HYOU1 and RIN1 predict better OS. Within the TNBC patient category, high expression of MAGED2 and STMN1 predicted poor OS, however elevated RACK1, HYOU1, PLAUR, RIN1 and SOD2 predicted better OS. Hence these proteins may serve as additional biomarkers in the prognosis of the TNBCs. Mechanistic regulation of these proteins by mdig needs further investigation. Nevertheless, we can see using the current data set that the TNBC protein repertoire displayed in the mdig KO cells indicates that the signaling pathways and metabolic alterations induced in the MDA-MB-231 cell line recapitulates the physiological changes *in vivo* as influenced by mdig on the mammary cells. This study provides a bioinformatical insight into the TNBC associated protein profiles in context of mdig deletion that has laid the foundation for identifying additional pathway specific biomarkers and their functional implications towards a better understanding of the development of breast cancers. The heterogeneity of the TNBC and lack of effective therapeutic targets along with insufficient predictive biomarkers are reasons for the challenges associated with TNBC therapy. Malignant transformation includes changes in protein abundance. Monitoring these changes at the protein level provides unique protein signatures that might facilitate effective diagnosis and prognosis. The high throughput proteomics study of the TNBC cells has provided a large and rich dataset that has allowed us to stratify systemic differences between the MDA-MB-231 cells with and without mdig. The top differentially regulated proteins have been validated at the protein level and have been found to predict disease prognosis both in breast cancer and in TNBC. Among them, high expression of STMN1, NAMPT, PLAUR and SOD2 predict poor overall survival in breast cancer patients whereas, high expression of FLNA, MAGED2, RACK1, HYOU1 and RIN1 predict better OS. Within the TNBC patient category, high expression of MAGED2 and STMN1 predicted poor OS, however elevated RACK1, HYOU1, PLAUR, RIN1 and SOD2 predicted better OS. Hence these proteins may serve as additional biomarkers in the prognosis of the TNBCs. Mechanistic regulation of these proteins by mdig needs further investigation. Nevertheless, we can see using the current data set that the TNBC protein repertoire displayed in the mdig KO cells indicates that the signaling pathways and metabolic alterations induced in the MDA-MB-231 cell line recapitulates the physiological changes *in vivo* as influenced by mdig on the mammary cells. This study provides a bioinformatical insight into the TNBC

associated protein profiles in context of mdig deletion that has laid the foundation for identifying additional pathway specific biomarkers and their functional implications towards a better understanding of the development of breast cancers.

Conclusions

Current data regarding proteomic changes, post translation modification profiles as well as differentially expressed proteins identified in the mdig deleted breast cancer cells revealed mdig's regulation on the abundance of crucial proteins which are implicated in EMT, genomic stability and metastasis. Our results on mdig modulated signaling pathways and hub molecules have provided novel targets that can be utilized for the development of treatment strategies and breast cancer therapies. Most importantly, this study provides the first insight into the molecular effects of mdig in governing the overall malignant phenotype of aggressive breast cancers thereby offering a new realm where mdig can be exploited in breast cancer therapies

Abbreviations

TNBC, triple negative breast cancer

WT, wild type

KO, knockout

PTM, post translation modification

EMT, epithelial mesenchymal transition

OS, overall survival

Declarations

Ethics approval and consent to participate Not applicable

Consent for publication Not applicable

Availability of data and materials

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (102) partner repository with the dataset identifier PXD016688 and 10.6019/PXD016688.

Western blots have been deposited to the Biostudies database (<https://www.ebi.ac.uk/biostudies/>) with the accession number S-BSST333.

Competing interests, the authors declare that they have no competing interests

Funding

National Institutes of Health grants R01 ES028263, R01 ES028335, and P30 ES020957 supported the design of this study, collection analysis and assisted in drafting the manuscript and overall needs of this project. National Institutes of Health grants P30 ES020957, P30 CA 022453 and S10 OD010700 supported the mass spectroscopy experiments and proteomics data analysis.

Authors' contributions

CT and FC conceived and designed the experiments and drafted the manuscript. CT and QZ carried out the CRISPR Cas-9 knockout assays. NJC carried out the mass spectrometry assay and assisted CT in conducting the bioinformatics and statistical analysis. LX, YF, ZB, WZ, PW, BA participated in the analysis. PMS supervised the proteomics assay and participated in its design and coordination and helped to review the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge the assistance of the Wayne State University Proteomics Core where NJC and PS are supported through NIH grants P30 ES020957, P30 CA 022453 and S10 OD010700. This research project is supported by the NIH grants R01 ES028263, R01 ES028335, and P30 ES020957 to FC.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
2. Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. *Breast cancer research treatment*. 1998;52(1–3):261–88.
3. Karlsson E, Appelgren J, Solterbeck A, Bergenheim M, Alvariza V, Bergh J. Breast cancer during follow-up and progression - A population based cohort on new cancers and changed biology. *European journal of cancer (Oxford England: 1990)*. 2014;50(17):2916–24.
4. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2010;28(20):3271–7.
5. Dent R, Hanna WM, Trudeau M, Rawlinson E, Sun P, Narod SA. Pattern of metastatic spread in triple-negative breast cancer. *Breast cancer research treatment*. 2009;115(2):423–8.
6. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. *J Natl Cancer Inst*. 1995;87(22):1681–5.
7. Steiner E, Klubert D, Knutson D. Assessing breast cancer risk in women. *Am Family Phys*. 2008;78(12):1361–6.
8. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. *N Engl J Med*. 2006;354(3):270–82.

9. Palmer JR, Rosenberg L. Cigarette smoking and the risk of breast cancer. *Epidemiol Rev.* 1993;15(1):145–56.
10. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. *Oncogene.* 2006;25(43):5898–905.
11. Greenwald P. Role of dietary fat in the causation of breast cancer: point. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research.* cosponsored by the American Society of Preventive Oncology. 1999;8(1):3–7.
12. Thakur C, Lu Y, Sun J, Yu M, Chen B, Chen F. Increased expression of mdig predicts poorer survival of the breast cancer patients. *Gene.* 2014;535(2):218–24.
13. Thakur C, Chen B, Li L, Zhang Q, Yang ZQ, Chen F. Loss of mdig expression enhances DNA and histone methylation and metastasis of aggressive breast cancer. *Signal transduction targeted therapy.* 2018;3:25.
14. Zhang Y, Lu Y, Yuan BZ, Castranova V, Shi X, Stauffer JL, et al. The Human mineral dust-induced gene, mdig, is a cell growth regulating gene associated with lung cancer. *Oncogene.* 2005;24(31):4873–82.
15. Thakur C, Chen F. Current understanding of mdig/MINA in human cancers. *Genes cancer.* 2015;6(7–8):288–302.
16. Mori T, Okamoto K, Tanaka Y, Teye K, Umata T, Ohneda K, et al. Ablation of Mina53 in mice reduces allergic response in the airways. *Cell Struct Funct.* 2013;38(2):155–67.
17. Thakur C, Wolfarth M, Sun J, Zhang Y, Lu Y, Battelli L, et al. Oncoprotein mdig contributes to silica-induced pulmonary fibrosis by altering balance between Th17 and Treg T cells. *Oncotarget.* 2015;6(6):3722–36.
18. Hemmers S, Mowen KA. T(H)2 bias: Mina tips the balance. *Nature immunology.* 2009;10(8):806–8.
19. Yosef N, Shalek AK, Gaublomme JT, Jin H, Lee Y, Awasthi A, et al. Dynamic regulatory network controlling TH17 cell differentiation. *Nature.* 2013;496(7446):461–8.
20. Yu M, Sun J, Thakur C, Chen B, Lu Y, Zhao H, et al. Paradoxical roles of mineral dust induced gene on cell proliferation and migration/invasion. *PloS one.* 2014;9(2):e87998.
21. Sun J, Yu M, Lu Y, Thakur C, Chen B, Qiu P, et al. Carcinogenic metalloid arsenic induces expression of mdig oncogene through JNK and STAT3 activation. *Cancer letters.* 2014;346(2):257–63.
22. Rogers LD, Fang Y, Foster LJ. An integrated global strategy for cell lysis, fractionation, enrichment and mass spectrometric analysis of phosphorylated peptides. *Molecular bioSystems.* 2010;6(5):822–9.
23. Taus T, Kocher T, Pichler P, Paschke C, Schmidt A, Henrich C, et al. Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res.* 2011;10(12):5354–62.
24. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology.* 2004;3:Article3.

25. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*. 2003;100(16):9440.
26. Li Q, Birkbak NJ, Györfy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*. 2011;12(1):474.
27. Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nature reviews Cancer*. 2007;7(6):415–28.
28. Mallini P, Lennard T, Kirby J, Meeson A. Epithelial-to-mesenchymal transition: what is the impact on breast cancer stem cells and drug resistance. *Cancer treatment reviews*. 2014;40(3):341–8.
29. Li L, Luo J, Wang B, Wang D, Xie X, Yuan L, et al. MicroRNA-124 targets flotillin-1 to regulate proliferation and migration in breast cancer. *Mol Cancer*. 2013;12:163.
30. Lin C, Wu Z, Lin X, Yu C, Shi T, Zeng Y, et al. Knockdown of FLOT1 impairs cell proliferation and tumorigenicity in breast cancer through upregulation of FOXO3a. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011;17(10):3089–99.
31. Tsuneoka M, Koda Y, Soejima M, Teye K, Kimura H. A novel myc target gene, mina53, that is involved in cell proliferation. *J Biol Chem*. 2002;277(38):35450–9.
32. Ge W, Wolf A, Feng T, Ho CH, Sekirnik R, Zayer A, et al. Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nature chemical biology*. 2012;8(12):960–2.
33. Lu Y, Chang Q, Zhang Y, Beezhold K, Rojanasakul Y, Zhao H, et al. Lung cancer-associated JmJc domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3. *Cell cycle (Georgetown Tex)*. 2009;8(13):2101–9.
34. Nagy Á, Lánczky A, Menyhárt O, Györfy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep*. 2018;8(1):9227.
35. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52.
36. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010;363(20):1938–48.
37. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57–70.
38. Sawyers CL. The cancer biomarker problem. *Nature*. 2008;452(7187):548–52.
39. Zhan T, Rindtorff N, Betge J, Ebert MP, Boutros M. CRISPR/Cas9 for cancer research and therapy. *Sem Cancer Biol*. 2019;55:106–19.
40. Yang H, Jaeger M, Walker A, Wei D, Leiker K, Weitao T. Break Breast Cancer Addiction by CRISPR/Cas9 Genome Editing. *J Cancer*. 2018;9(2):219–31.
41. Garcia M, Platet N, Liaudet E, Laurent V, Derocq D, Brouillet J-P, et al. Biological and Clinical Significance of Cathepsin D in Breast Cancer Metastasis. *STEM CELLS*. 1996;14(6):642–50.
42. Capony F, Rougeot C, Montcourrier P, Cavailles V, Salazar G, Rochefort H. Increased secretion, altered processing, and glycosylation of pro-cathepsin D in human mammary cancer cells. *Cancer research*. 1989;49(14):3904–9.

43. Rochefort H, Cavailles V, Augereau P, Capony F, Maudelonde T, Touitou I, et al. Overexpression and hormonal regulation of pro-cathepsin D in mammary and endometrial cancer. *Journal of steroid biochemistry*. 1989;34(1–6):177–82.
44. Rochefort H. Cathepsin D in breast cancer: a tissue marker associated with metastasis. *European journal of cancer (Oxford, England: 1990)*. 1992;28a(11):1780-3.
45. Ferrandina G, Scambia G, Bardelli F, Benedetti Panici P, Mancuso S, Messori A. Relationship between cathepsin-D content and disease-free survival in node-negative breast cancer patients: a meta-analysis. *Br J Cancer*. 1997;76(5):661–6.
46. Westley BR, May FE. Prognostic value of cathepsin D in breast cancer. *British journal of cancer*. 1999;79(2):189–90.
47. Sivaranjini Y, Angadi PV, Kumar GS. Immuno-histochemical evaluation of Cathepsin D in malignant salivary gland carcinomas. *The Gulf journal of oncology*. 2012(11):38–44.
48. Ohri SS, Vashishta A, Proctor M, Fusek M, Vetvicka V. The propeptide of cathepsin D increases proliferation, invasion and metastasis of breast cancer cells. *Int J Oncol*. 2008;32(2):491–8.
49. Vashishta A, Ohri SS, Proctor M, Fusek M, Vetvicka V. Ribozyme-targeting procathepsin D and its effect on invasion and growth of breast cancer cells: an implication in breast cancer therapy. *Int J Oncol*. 2007;30(5):1223–30.
50. Berchem G, Glondu M, Gleizes M, Brouillet JP, Vignon F, Garcia M, et al. Cathepsin-D affects multiple tumor progression steps in vivo: proliferation, angiogenesis and apoptosis. *Oncogene*. 2002;21(38):5951–5.
51. Kidd M, Modlin IM, Mane SM, Camp RL, Eick G, Latich I. The role of genetic markers–NAP1L1, MAGE-D2, and MTA1–in defining small-intestinal carcinoid neoplasia. *Ann Surg Oncol*. 2006;13(2):253–62.
52. Curmi PA, Nogues C, Lachkar S, Carelle N, Gonthier MP, Sobel A, et al. Overexpression of stathmin in breast carcinomas points out to highly proliferative tumours. *Br J Cancer*. 2000;82(1):142–50.
53. Arnedos M, Drury S, Afentakis M, A'Hern R, Hills M, Salter J, et al. Biomarker changes associated with the development of resistance to aromatase inhibitors (AIs) in estrogen receptor-positive breast cancer. *Annals of oncology: official journal of the European Society for Medical Oncology*. 2014;25(3):605–10.
54. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci USA*. 2007;104(18):7564–9.
55. Baquero MT, Hanna JA, Neumeister V, Cheng H, Molinaro AM, Harris LN, et al. Stathmin expression and its relationship to microtubule-associated protein tau and outcome in breast cancer. *Cancer*. 2012;118(19):4660–9.
56. Obayashi S, Horiguchi J, Higuchi T, Katayama A, Handa T, Altan B, et al. Stathmin1 expression is associated with aggressive phenotypes and cancer stem cell marker expression in breast cancer patients. *Int J Oncol*. 2017;51(3):781–90.

57. Cao X-X, Xu J-D, Xu J-W, Liu X-L, Cheng Y-Y, Wang W-J, et al. RACK1 promotes breast carcinoma proliferation and invasion/metastasis in vitro and in vivo. *Breast cancer research treatment*. 2010;123(2):375–86.
58. Cao XX, Xu JD, Liu XL, Xu JW, Wang WJ, Li QQ, et al. RACK1: A superior independent predictor for poor clinical outcome in breast cancer. *International journal of cancer*. 2010;127(5):1172–9.
59. Lv Q-L, Huang Y-T, Wang G-H, Liu Y-L, Huang J, Qu Q, et al. Overexpression of RACK1 Promotes Metastasis by Enhancing Epithelial-Mesenchymal Transition and Predicts Poor Prognosis in Human Glioma. *Int J Environ Res Public Health*. 2016;13(10):1021.
60. Bedolla R, Wang Y, Asuncion A, Chamie K, Siddiqui S, Mudryj M, et al. Nuclear versus cytoplasmic localization of filamin A in prostate cancer: immunohistochemical correlation with metastases. *Clin Cancer Res*. 2009;15:788–96.
61. Coughlin M, Puig-de-Morales M, Bursac P, Mellema M, Millet E, Fredberg J. Filamin-A and rheological properties of cultured melanoma cells. *Biophys J*. 2006;90:2199–205.
62. Bachmann A, Howard J, Vogel C. Actin-binding protein filamin A is displayed on the surface of human neuroblastoma cells. *Cancer Sci*. 2006;97:1359–65.
63. Yue J, Huhn S, Shen Z. Complex roles of filamin-A mediated cytoskeleton network in cancer progression. *Cell Bioscience*. 2013;3(1):7.
64. Savoy RM, Ghosh PM. The dual role of filamin A in cancer: can't live with (too much of) it, can't live without it. *Endocrine-related Cancer*. 2013;20(6):R341-56.
65. Jiang X, Yue J, Lu H, Campbell N, Yang Q, Lan S, et al. Inhibition of filamin-A reduces cancer metastatic potential. *Int J Biol Sci*. 2013;9(1):67–77.
66. Ji Z-m, Yang L-l, Ni J, Xu S-p, Yang C, Duan P, et al. Silencing Filamin A Inhibits the Invasion and Migration of Breast Cancer Cells by Up-regulating 14-3-3 σ . *Current Medical Science*. 2018;38(3):461–6.
67. Gelder MEM-v, Look MP, Peters HA, Schmitt M, Brnner N, Harbeck N, et al. Urokinase-Type Plasminogen Activator System in Breast Cancer. Association with Tamoxifen Therapy in Recurrent Disease. 2004;64(13):4563–8.
68. Bianchi E, Cohen RL, Thor AT, Todd RF, Mizukami IF, Lawrence DA, et al. The Urokinase Receptor Is Expressed in Invasive Breast Cancer but not in Normal Breast Tissue. *Cancer research*. 1994;54(4):861–6.
69. Konecny G, Untch M, Arboleda J, Wilson C, Kahlert S, Boettcher B, et al. HER-2/neu and Urokinase-Type Plasminogen Activator and Its Inhibitor in Breast Cancer. *Clin Cancer Res*. 2001;7(8):2448–57.
70. Narayanaswamy PB, Hodjat M, Haller H, Dumler I, Kiyan Y. Loss of urokinase receptor sensitizes cells to DNA damage and delays DNA repair. *PloS one*. 2014;9(7):e101529.
71. Narayanaswamy PB, Baral TK, Haller H, Dumler I, Acharya K, Kiyan Y. Transcriptomic pathway analysis of urokinase receptor silenced breast cancer cells: a microarray study. *Oncotarget*. 2017;8(60):101572–90.

72. LeBeau AM, Duriseti S, Murphy ST, Pepin F, Hann B, Gray JW, et al. Targeting uPAR with Antagonistic Recombinant Human Antibodies in Aggressive Breast Cancer. *Cancer research*. 2013;73(7):2070–81.
73. Kim YS, Gupta Vallur P, Phaeton R, Mythreye K, Hempel N. Insights into the Dichotomous Regulation of SOD2 in Cancer. *Antioxidants* (Basel, Switzerland). 2017;6(4).
74. Finley LW, Carracedo A, Lee J, Souza A, Egia A, Zhang J, et al. SIRT3 opposes reprogramming of cancer cell metabolism through HIF1alpha destabilization. *Cancer cell*. 2011;19(3):416–28.
75. Li JJ, Oberley LW, St Clair DK, Ridnour LA, Oberley TD. Phenotypic changes induced in human breast cancer cells by overexpression of manganese-containing superoxide dismutase. *Oncogene*. 1995;10(10):1989–2000.
76. Weydert C, Roling B, Liu J, Hinkhouse MM, Ritchie JM, Oberley LW, et al. Suppression of the malignant phenotype in human pancreatic cancer cells by the overexpression of manganese superoxide dismutase. *Mol Cancer Ther*. 2003;2(4):361–9.
77. Hitchler MJ, Oberley LW, Domann FE. Epigenetic silencing of SOD2 by histone modifications in human breast cancer cells. *Free Radic Biol Med*. 2008;45(11):1573–80.
78. Papa L, Hahn M, Marsh EL, Evans BS, Germain D. SOD2 to SOD1 switch in breast cancer. *J Biol Chem*. 2014;289(9):5412–6.
79. Hitchler M, Wikainapakul K, Yu L, Powers K, Attatippaholkun W, Domann F. Epigenetic Regulation of Manganese Superoxide Dismutase Expression in Human Breast Cancer Cells. *Epigenetics*. 2006;1(4):163–71.
80. He H, Wu G, Liu H, Cheng Y, Yu Y, Wang Y, et al. Low RIN1 Expression in HCC Is Associated With Tumor Invasion and Unfavorable Prognosis. *Am J Clin Pathol*. 2013;140(1):73–81.
81. Hu H, Milstein M, Bliss JM, Thai M, Malhotra G, Huynh LC, et al. Integration of transforming growth factor beta and RAS signaling silences a RAB5 guanine nucleotide exchange factor and enhances growth factor-directed cell migration. *Mol Cell Biol*. 2008;28(5):1573–83.
82. Milstein M, Mooser CK, Hu H, Fejzo M, Slamon D, Goodglick L, et al. RIN1 is a breast tumor suppressor gene. *Cancer research*. 2007;67(24):11510–6.
83. Tsukamoto Y, Kuwabara K, Hirota S, Kawano K, Yoshikawa K, Ozawa K, et al. Expression of the 150-kd oxygen-regulated protein in human breast cancer. *Lab Invest*. 1998;78(6):699–706.
84. Asahi H, Koshida K, Hori O, Ogawa S, Namiki M. Immunohistochemical detection of the 150-kDa oxygen-regulated protein in bladder cancer. *BJU Int*. 2002;90(4):462–6.
85. Miyagi T, Hori O, Koshida K, Egawa M, Kato H, Kitagawa Y, et al. Antitumor effect of reduction of 150-kDa oxygen-regulated protein expression on human prostate cancer cells. *International journal of urology: official journal of the Japanese Urological Association*. 2002;9(10):577–85.
86. Olesen UH, Hastrup N, Sehested M. Expression patterns of nicotinamide phosphoribosyltransferase and nicotinic acid phosphoribosyltransferase in human malignant lymphomas. *APMIS: acta pathologica, microbiologica, et immunologica Scandinavica*. 2011;119(4–5):296–303.

87. Nakajima TE, Yamada Y, Hamano T, Furuta K, Gotoda T, Katai H, et al. Adipocytokine levels in gastric cancer patients: resistin and visfatin as biomarkers of gastric cancer. *Journal of gastroenterology*. 2009;44(7):685–90.
88. Nakajima TE, Yamada Y, Hamano T, Furuta K, Matsuda T, Fujita S, et al. Adipocytokines as new promising markers of colorectal tumors: adiponectin for colorectal adenoma, and resistin and visfatin for colorectal cancer. *Cancer Sci*. 2010;101(5):1286–91.
89. Wang B, Hasan MK, Alvarado E, Yuan H, Wu H, Chen WY. NAMPT overexpression in prostate cancer and its contribution to tumor cell survival and stress response. *Oncogene*. 2011;30(8):907–21.
90. Hesari Z, Nourbakhsh M, Hosseinkhani S, Abdolvahabi Z, Alipour M, Tavakoli-Yaraki M, et al. Down-regulation of NAMPT expression by mir-206 reduces cell survival of breast cancer cells. *Gene*. 2018;673:149–58.
91. Santidrian AF, LeBoeuf SE, Wold ED, Ritland M, Forsyth JS, Felding BH. Nicotinamide phosphoribosyltransferase can affect metastatic activity and cell adhesive functions by regulating integrins in breast cancer. *DNA Repair*. 2014;23:79–87.
92. Zhu Y, Guo M, Zhang L, Xu T, Wang L, Xu G. Biomarker triplet NAMPT/VEGF/HER2 as a de novo detection panel for the diagnosis and prognosis of human breast cancer. *Oncol Rep*. 2016;35(1):454–62.
93. Sharma DK, Bressler K, Patel H, Balasingam N, Thakor N. Role of Eukaryotic Initiation Factors during Cellular Stress and Cancer Progression. *Journal of Nucleic Acids*. 2016;2016:19.
94. Ananieva E. Targeting amino acid metabolism in cancer growth and anti-tumor immune response. *World journal of biological chemistry*. 2015;6(4):281–9.
95. Antanavičiūtė I, Mikalayeva V, Ceslevičienė I, Milašiūtė G, Skeberdis VA, Bordel S. Transcriptional hallmarks of cancer cell lines reveal an emerging role of branched chain amino acid catabolism. *Sci Rep*. 2017;7(1):7820.
96. Zhang L, Han J. Branched-chain amino acid transaminase 1 (BCAT1) promotes the growth of breast cancer cells through improving mTOR-mediated mitochondrial biogenesis and function. *Biochem Biophys Res Commun*. 2017;486(2):224–31.
97. Cuyas E, Fernandez-Arroyo S, Alarcon T, Lupu R, Joven J, Menendez JA. Germline BRCA1 mutation reprograms breast epithelial cell metabolism towards mitochondrial-dependent biosynthesis: evidence for metformin-based "starvation" strategies in BRCA1 carriers. *Oncotarget*. 2016;7(33):52974–92.
98. Rajapaksa G, Thomas C, Gustafsson J-Å. Estrogen signaling and unfolded protein response in breast cancer. *The Journal of Steroid Biochemistry Molecular Biology*. 2016;163:45–50.
99. Eilbracht J, Kneissel S, Hofmann A, Schmidt-Zachmann MS. Protein NO52—a constitutive nucleolar component sharing high sequence homologies to protein NO66. *Eur J Cell Biol*. 2005;84(2–3):279–94.
100. Ge W, Wolf A, Feng T, Ho C-h, Sekirnik R, Zayer A, et al. Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nature chemical biology*. 2012;8:960.

101. Sheng S, Qiao M, Pardee AB. Metastasis and AKT activation. *J Cell Physiol.* 2009;218(3):451–4.

102. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic acids research.* 2019;47(D1):D442-d50.

Figures

Figure 1

Mdig deletion by CRISPR Cas 9 approach and validation in MDA-MB-231 cells MDA-MB-231 cells were subjected to CRISPR Cas 9 based gene editing for mdig. (A) After transfection and selection, the obtained colonies were screened for mdig protein expression by western blot. WT cells represent failed mdig targeting by CRISPR Cas 9, whereas KO cells represent successful mdig knockout by the CRISPR Cas 9 system. N=no. of colonies, 5 WT and 13 KO. GAPDH was used as a loading control. Each time the WT proteins were loaded along with the KO in a series of there independent gels. Image is a representative of three independent western blots. Full-length blots/gels are presented in Supplementary Figure S10 (B) All the WT and KO colonies in duplicates were subjected to proteomics assay. Principal component analysis (PCA) of the colonies indicated some within group heterogeneity. PCA of with samples labelled by sample name and are colored by WT/KO ratio. The colonies were validated and were proceeded for downstream data analysis. KO #10, KO#3 and WT#5 were excluded from further analysis.

Figure 2

PANTHER functional classification of differentially expressed proteins in MDA-MB-231 KO over WT cells. Histograms showing classification according to the class of proteins and biological process. Proteins consisting of nucleic acid binding (PC00171) class was the highest, encoding 340 genes for the Protein Class category. In biological process, cellular and metabolic process showed increased number of genes assigned to them compared to other subcategories.

A

Name	p-value	Overlap
EIF2 Signaling	5.61E-09	34.6 % 54/156
Isoleucine Degradation I	4.35E-05	72.7 % 8/11
Unfolded protein response	9.51E-04	41.4 % 12/29
Regulation of eIF4 and p70S6K Signaling	1.46E-03	27.2 % 31/114
Caveolar-mediated Endocytosis Signaling	4.19E-03	33.3 % 14/42

B

Figure 3

Canonical pathways identified by Ingenuity Pathway Analysis, in MDA-MB-231 KO over WT cells (A) Illustrated are the top five canonical pathways ($p < 0.05$) based on 904 differentially expressed proteins. (B) Canonical pathway analysis revealed the eukaryotic initiation factor 2 (EIF2) signaling pathway among the topmost canonical pathways enriched upon mdig knock down. The EIF2 Signaling has been elaborated showing the upregulated and downregulated proteins and their cellular localization. Nodes represent molecules in a pathway, whereas the biological relationship between nodes is represented by a line (edge). Edges are supported by at least one reference in the Ingenuity Knowledge Base. The intensity of color in a node indicates the degree of up (red) or down (green) regulation.

Regulation of eIF4 and p70S6K Signaling

Unfolded Protein Response

Figure 4

IPA of top canonical pathways of differentially altered proteins in MDA-MB-231 KO over WT cells Canonical pathway analysis revealed the Regulation of eIF4 and p70S6K Signaling pathway and Unfolded Protein Response among the topmost canonical pathways enriched upon mdig knock down. The Regulation of eIF4 & p70S6K Signaling and Unfolded Protein Response pathways have been elaborated showing the upregulated and downregulated proteins and their cellular localization. Nodes represent molecules in a pathway, whereas the biological relationship between nodes is represented by a line (edge). Edges are supported by at least one reference in the Ingenuity Knowledge Base. The intensity of color in a node indicates the degree of up (red) or down (green) regulation.

Figure 5

Activation status of molecular and cellular functions in the MDA-MB-231 KO over WT cells (A) List of top molecular and cellular functions with their respective scores obtained from the IPA. (B) Molecular and Cellular Functional Analysis predicts the activation of protein synthesis in KO cells, whereas processing of rRNA and cancer cell death are decreased.

B Upstream Regulators

Name	p-value	Predicted activation
MYCN	6.57E-13	Activated
sirolimus	2.07E-09	Inhibited
NFE2L2	7.03E-09	
MYC	2.72E-08	
TCR	1.50E-07

Figure 6

IPA of Top Diseases & Disorders and Upstream Regulators in MDA-MB-231 KO over WT cells (A) Hierarchical heat map depicting affected functional categories based on differentially expressed proteins where the major boxes represent a category of diseases and functions showing Cancer, Organismal Injury and Abnormalities, Tumor Morphology, Cardiovascular Disease and Developmental Disorder as the top diseases in the proteomics data set. (B) Upstream regulator analysis predicted the upstream molecules which are likely to cause the observed gene expression changes in the current analysis. Myc was found to be the top molecule which showed activated status upon mdig knockdown in the breast cancer cells. The figure legend describes predicted relationships of all the genes to the upstream regulator.

Figure 7

Post Translation Modification and Network Analysis in the MDA-MB-231 cells. (A) Relative abundance of modifications of sites on 60S ribosomal protein L27a peptide in MDA-MB-231 cells. Abundance of 60S ribosomal protein L27a peptide [R32] GNAGGLHHHR [I43] in the native or His-oxidized form. The peptide includes His39, the known target for Mdig catalyzed oxidation. The abundance of the oxidized form was decreased in the KO category. Black bars indicate abundance in WT samples and grey bars indicate abundance in mdig KO samples. $q = 0.00030$, moderated t-test, $n = 8$. (B) Ingenuity Pathway Analysis of protein networks in the MDA-MB-231 KO over WT cells. The top five significant networks as determined by the IPA with their scores and associated functions has been shown. (C) Network ID 1 has been elaborated showing the relationships and connectivity between the focus molecules of the network related to protein synthesis, RNA post transcriptional modification, DNA replication, Recombination and Repair. Red indicated increased and green indicates decreased expression.

Figure 8

Proteomics identified key proteins that are differentially regulated in TNBC and predicts patient's overall survival. Top differentially expressed proteins ($p < 0.05$, $n = 8$, moderated t-test) as revealed by the proteomics study were validated in MDA-MB-231 cells corroborating the proteomics findings. These proteins have prognostic significance in breast cancer patients. (A) Validation of proteomics results using western blot analysis of proteins isolated from MDA-MB-231 WT and KO cells. Red indicates upregulated and green indicates downregulated proteins in KO over WT cells. Representative image from three independent assays. Tubulin was used as a loading control. Full-length blots/gels are presented in Supplementary Figure S11 to S14. WT and KO numbers indicates individual colony number generated after CRISPR Cas 9 mediated knockdown for mdig. (B) Survival curves were plotted for breast cancer patients ($n = 3951$) and TNBC patients ($n = 618$) for the indicated proteins using Kaplan-Meier Plotter. X axis denotes time in months and Y axis denotes survival probability. P value = ** $P \leq 0.01$ *** $P \leq 0.001$

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure3.tif](#)
- [S14.tif](#)
- [SupplementaryFigure2.tif](#)
- [SupplementaryFigure1.tif](#)
- [S13.tif](#)

- [S12.tif](#)
- [S11.tif](#)
- [S10.tif](#)
- [SupplementaryTableS9forTrimethylatedPeptides.csv](#)
- [SupplementaryTableS8forHisHydroxylatedPeptides.csv](#)
- [SupplementaryTableS7forDimethylatedPeptides.csv](#)
- [SupplementaryTableS6forAcetylatedPeptides.csv](#)
- [SupplementaryTableS5.csv](#)
- [SupplementaryTableS4.csv](#)
- [SupplementaryTableS3.xlsx](#)
- [SupplementaryTableS2.csv](#)
- [SupplementaryTableS1.csv](#)