RESEARCH

Open Access

High expression of MKK3 is associated with worse clinical outcomes in African American breast cancer patients

Xuan Yang[1,2], Mohamed Amgad[3], Lee A. D. Cooper[4], Yuhong Du[1,2,5], Haian Fu[1,2,5,6*] and Andrey A. Ivanov[1,2,5*]

Abstract

Background: African American women experience a twofold higher incidence of triple-negative breast cancer (TNBC) and are 40% more likely to die from breast cancer than women of other ethnicities. However, the molecular bases for the survival disparity in breast cancer remain unclear, and no race-specific therapeutic targets have been proposed. To address this knowledge gap, we performed a systematic analysis of the relationship between gene mRNA expression and clinical outcomes determined for The Cancer Genome Atlas (TCGA) breast cancer patient cohort.

Methods: The systematic differential analysis of mRNA expression integrated with the analysis of clinical outcomes was performed for 1055 samples from the breast invasive carcinoma TCGA PanCancer cohorts. A deep learning fully-convolutional model was used to determine the association between gene expression and tumor features based on breast cancer patient histopathological images.

Results: We found that more than 30% of all protein-coding genes are differentially expressed in White and African American breast cancer patients. We have determined a set of 32 genes whose overexpression in African American patients strongly correlates with decreased survival of African American but not White breast cancer patients. Among those genes, the overexpression of mitogen-activated protein kinase kinase 3 (MKK3) has one of the most dramatic and race-specific negative impacts on the survival of African American patients, specifically with triple-negative breast cancer. We found that MKK3 can promote the TNBC tumorigenesis in African American patients in part by activating of the epithelial-to-mesenchymal transition induced by master regulator MYC.

Conclusions: The poor clinical outcomes in African American women with breast cancer can be associated with the abnormal elevation of individual gene expression. Such genes, including those identified and prioritized in this study, could represent new targets for therapeutic intervention. A strong correlation between MKK3 overexpression, activation of its binding partner and major oncogene MYC, and worsened clinical outcomes suggests the MKK3-MYC protein–protein interaction as a new promising target to reduce racial disparity in breast cancer survival.

Keywords: Triple-negative breast cancer, Racial disparity, Differential expression, MKK3

Background

Breast cancer is the most common cancer and the leading cause of cancer-related death in women [1]. Recent studies have shown up to a twofold higher incidence of triple-negative breast cancer (TNBC) among African American women as compared to White women [2–4]. Moreover, African Americans die from breast cancer at

*Correspondence: hfu@emory.edu; andrey.ivanov@emory.edu
[1] Department of Pharmacology and Chemical Biology, Emory University School of Medicine, Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA
Full list of author information is available at the end of the article

Yang *et al. J Transl Med*   (2020) 18:334

Page 2 of 19

up to 40% higher rate than White and Hispanic women [5–7]. The American College of Radiology (ACR) has assigned a special status for African American women at higher-than-average risk for breast cancer [8].

Previous studies have revealed significant differences in the mutation rates of several cancer driver genes in African American and White breast cancer patients (Table 1) [9–13].

For example, it was shown that African American women with at least 50% African ancestry have a higher rate of mutations in the major tumor suppressor gene TP53 (43%) as compared to White women with at least 90% European ancestry (28%) [11, 12]. Huo et al. [12] also demonstrated that the mutation frequency in the ubiquitin ligase FBXW7 is almost four times higher in African American breast cancer patients (4.2%) than in White patients (1.2%). Furthermore, African American patients show a higher mutation frequency of BRCA1 (10.2%) and BRCA2 (5.7%) tumor suppressor genes comparing to European non-Ashkenazi Jews White patients (BRCA1: 6.9%, BRCA2: 5.2%) [9, 10]. In contrast, mutations in the catalytic subunit of the Alpha isoform of the Phosphatidylinositol 4,5-Bisphosphate 3-Kinase (PIK3CA) were rarer in African American patients than in White breast cancer patients (20% vs 34%). This difference was even more significant between European White patients (36%) and a cohort of Nigerian breast cancer patients (17%) [13]. In the same study [13], Pitt et al. also determined a significantly lower mutation rate of Cadherin 1 (CDH1) in Nigerian patients (0.8%) and TCGA African American patients (6.4%) as compared to White patients (16.2%).

Besides the mutation rates, the frequency of the DNA copy number alterations has been recently analyzed [12]. It was shown that retinoblastoma protein 1 (RB1), a cell cycle suppressor and the CUB And Sushi Multiple Domains 1 (CSMD1), a tumor suppressor that control cell proliferation, invasion, and migration, are more frequently deleted in Black/African American breast cancer patients (14.5% and 8.6%, respectively) as compared to White patients (8.7% and 4.1%, respectively). Conversely, MYC and Cyclin E1, critical activators of the cell cycle, are more frequently amplified in Black/African American breast cancer patients (30.9% and 9.2%, respectively) than in White patients (20.4% and 3.6%). Together, accumulating clinical and genomics data reveal unique molecular features that may contribute to survival disparity in breast cancer. As summarized in Table 1, the majority of genes that are differentially altered in White and African American breast cancer patients play critical functions in cell proliferation and survival. Meanwhile, most of those genes, including TP53, BRCA1/2, FBXW7, RB1, CDH1, and CSMD1, are tumor suppressors lost due to the inactivating mutations or deletions. The discovery of race-specific and therapeutically actionable targets to decrease the mortality in African American breast cancer patients remains a challenge.

To address this unmet medical need, we performed a systematic analysis of clinical outcomes and gene expression determined for the TCGA PanCancer cohorts of White and African American breast cancer patients. We have identified 32 genes as potential targets to decrease the mortality of African American breast cancer patients. The mitogen-activated protein kinase 3

Table 1 Frequency of tumor driver gene alterations in Black/African American and White breast cancer patients

	Black or African American	White	Oncogenic function	Regulated pathways	References
Mutation, %					
TP53	43	28	TSG	Apoptosis, senescence, DNA repair	[11]
BRCA1	10.2	6.9	TSG	DNA repair Checkpoint control	[9]
BRCA2	5.7	5.2	TSG	DNA repair, checkpoint control	[9]
PIK3CA	20	34	OG	Cell survival, proliferation	[11]
FBXW7	4.2	1.2	TSG	Cell cycle, apoptosis, differentiation	[12]
CDH1	6.4	16.2	TSG	Proliferation, adhesion, polarity, EMT	[13]
Deletion, %					
CSMD1	14.5	8.7	TSG	Proliferation, migration and invasion	[12]
RB1	8.6	4.1	TSG	Cell cycle, apoptosis	[12]
Amplification, %					
MYC	30.9	20.4	OG	Cell growth, survival, immune response, other	[12]
CCNE1	9.2	3.6	OG	Cell cycle	[12]

TSG tumor suppressor gene, *OG* oncogene, *EMT* epithelial-to-mesenchymal transition

Yang *et al. J Transl Med* (2020) 18:334

Page 3 of 19

(MKK3) appeared among the proteins with the most dramatic impact on the survival of African American TNBC patients. We determined that MKK3 promotes TNBC tumorigenesis in African American but not White or Asian patients, and its overexpression leads to the activation of the transcriptional program of major tumor driver MYC.

Together, our data revealed multiple proteins as new promising targets for therapeutic intervention in breast cancer African American patients. As one example, we showed that MKK3 has critical oncogenic functions and promotes TNBC tumorigenesis in African Americans through the activation of the MYC program. The discovery of small-molecule inhibitors to control MKK3 signaling may provide a new therapeutic strategy to decrease mortality in African American TNBC patients.

Methods

Breast cancer patient cohort

In this study, the clinical and genomics data from the Breast Invasive Carcinoma TCGA PanCancer cohorts [14] that consists of a total of 1055 female patients with determined DNA copy-number and mRNA expression were analyzed (Fig. 1). The gene RNA expression, DNA copy number, and breast cancer patient survival data were obtained from the NCI Genomics Data Commons (GDC) [15]. The dataset included samples from 729 White patients (69% of all samples) and 178 samples from Black or African American (BAA) patients (17%), as well as 60 Asian patient samples (6%), and 88 samples (8%) from patients with unspecified race. The breast cancer subtype annotations were added based on the original publication [14].

Differential expression

The subset of 17,211 protein-coding genes was identified based on the HUGO Gene Nomenclature Committee (HGNC) annotations [16]. The DNA amplifications or deletions were determined based on the GISTIC 2.0 scores (2-amplification, -2-homozygous deletion) [17]. For the differential expression analysis, the TCGA RNA-seqV2 expression data (EBPlusPlusAdjustPAN-CAN_IlluminaHiSeq_RNASeqV2.geneExp.tsv; http://api.gdc.cancer.gov/data/3586c0da-64d0-4b74-a449-5ff4d9136611) were used. For each gene, the log2 fold change was calculated as $log2\,fold\,change = \mu_{BAA} - \mu_{WT}$, where μ_{BAA} and μ_{WT} are the mean values of the log2 (x+1)-transformed gene expression obtained for the Black/African American and White patient cohorts, respectively. The p-values were calculated with the Wilcoxon test. The false discovery rate adjusted q-values were calculated with the Benjamini–Hochberg procedure [18].
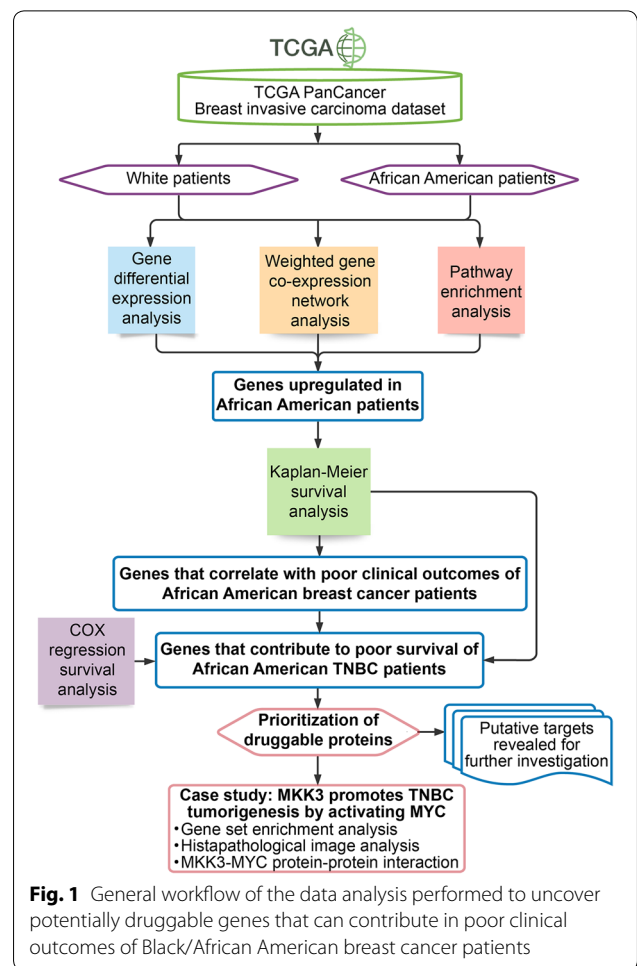

Fig. 1 General workflow of the data analysis performed to uncover potentially druggable genes that can contribute in poor clinical outcomes of Black/African American breast cancer patients

The mRNA overexpression was determined based on the z-scores. First, the average (μ) and standard deviation (σ) values were calculated for the samples in which gene is diploid. Then, the z-score was calculated as (τ − μ)/σ, were τ is the gene mRNA expression in the sample. Z-score > 2 and z-score < -2 indicates gene overexpression or underexpression, respectively.

The signed weighted gene co-expression network was constructed for 5256 genes differentially expressed in Black/African American and White breast cancer patients using the WGCNA R package [19]. Pearson correlation coefficients between the expression of all gene pairs were also calculated and used to construct the adjacency correlation matrix and the topological overlap matrix (TOM). The optimal value of the soft threshold power β = 11 was selected using the pickSoftThreshold function to maintain the scale-free topology and sufficient node connectivity [20]. The hierarchical clustering of genes was performed based on the TOP matrix using the average agglomeration method implemented in the flashClust function [21]. The gene modules were

Yang *et al. J Transl Med*   (2020) 18:334

Page 4 of 19

identified using the dynamic tree cut method [22]. Specifically, the cutreeDynamic R function was used with the minModuleSize = 100 and method = "tree" options.

Survival analysis

The Kaplan–Meier survival curves and the logrank p-values have been calculated using the Lifelines python package. The mean survival time (MST) values were calculated with the Lifelines package based on the area under the survival curve. The COX regression analysis has been performed using the fit proportional hazards regression model function coxph from the Survival R package.

Enrichment analysis

The disease association and pathway enrichment analysis were performed using the DisGENet [23], KEGG [24], and Reactome [25] datasets. The p-values were calculated with the Fisher Exact test using 17,211 protein-coding genes as the reference set. The false discovery rate adjusted q-values were calculated with the Benjamini–Hochberg procedure [18]. The gene set enrichment analysis (GSEA) was performed using the GSEA program [26]. The High and Low phenotypes were defined as the 10% of samples with the highest and the lowest gene expression, respectively. The GSEA curves were rebuilt using the GseaPy python package.

The breast cancer histological image analysis
Fully-convolutional model training

To extract tumor features we used our established standard 16-layer VGG fully-convolutional neural network (VGG16-FCN8) constructed using ImageNet [27] pre-trained weights as described previously [28]. We have previously shown that for this particular dataset, the VGG-16 FCN-8 architecture shows more favorable model convergence and fitting properties than the deeper and more complex DenseNet architecture [29]. Using this particular architecture and number of layers enabled us to leverage the publicly available pre-trained weights, hence improving accuracy [28, 29].

The model is trained to classify pixels into one of five classes: tumor (including DCIS), stroma, tumor-infiltrating lymphocytes (including plasma cells and mixed inflammatory infiltrates), necrosis or debris, and others. Regions of interest were divided into 800×800 pixel tiles that are overlapping, where the amount of overlap increased for smaller regions of interest to create a balanced training dataset. Random cropping of 768×768 pixel regions was used as a data augmentation strategy to improve robustness during training. The model was trained on 4 GPUs with a per-GPU batch size of 4 tiles (16 tiles per batch) using data parallelization and gradient averaging. Adap optimizer was used with a starting learning rate of 1e−5. The loss function used is weighted categorical cross-entropy, where the weight associated with each region class, Wc, is calculated using the equation:

$$W_c = \begin{cases} 0 : & \text{if } c = 0 \\ 1 - \dfrac{N_c}{N} : & \text{if } c > 0 \end{cases}$$

where N is the total number of pixels and Nc is the total number of pixels belonging to region class c.

Fully-convolutional model inference

We used whole-slide images (WSI) formalin-fixed paraffin-embedded hematoxylin and eosin-stained slides from the TCGA cohort. The analysis was focused on WSIs from African-American patients with triple-negative breast cancer, and limited to infiltrating ductal histologic subtype (determined using TCGA clinical records). The focus on infiltrating ductal subtype is for pragmatic reasons since the fully-convolutional model has been trained and optimized on this histologic subset. Only one diagnostic slide was used per patient ("-DX" designation in TCGA) and only WSIs scanned at $40\times$ were used in the analysis. The analysis was performed at scan magnification.

Analysis regions were chosen semi-automatically and constituted the main tumor bulk within a WSI. A low-resolution RGB image of the slide (at 0.3–0.5 x) was loaded and converted to the Hue-Saturation-Intensity (HSI) space. Default thresholds for each of the HSI channels were manually adjusted for each slide to capture the majority or entirety of the tumor within the slide. This region of interest was divided into non-overlapping 1024×1024 pixel tiles and fed into the trained FCN-8 model after color normalization using the Reinhard method [30]. The Reinhard normalization used target statistics derived from the RGB image corresponding to the mask called *"TCGA-A2-A3XS-DX1_xmin21421_ymin37486_.png"* [28].

Feature extraction of tumor nests

A total of nine features (four global and five local) were derived from the slides. "Local" features are those features derived from each individual tumor nest (defined as a coherent collection of carcinoma cells) and are averaged to get slide-level features. The global features were: tumor-to-stroma ratio, stromal tumor-infiltrating lymphocyte score, necrosis-to-tumor ratio, and the number of tumor nests, normalized for the area of the region of interest (i.e. "per pixel"). Local features included area and shape descriptors for each tumor nest.

Yang *et al. J Transl Med*   (2020) 18:334

Page 5 of 19

Histologic-genomic correlation

Histological descriptors were compared against gene expression data derived from the same patients in the TCGA cohort. Spearman correlation coefficient was used and the Benjamini–Hochberg adjustment was used for multiple hypothesis testing.

Results

Differential gene expression in African American and White breast cancer patients

To determine the differences between gene expression in White and Black/African American breast cancers, we have performed the differential expression (DE) analysis for a total of 17,211 protein-coding genes. We found that 7195 genes showed statistically significant differences in expression between White and Black/African American cohorts, as determined with the Wilcoxon test p-values adjusted for the false discovery rate (q-value < 0.001, Additional file 1: Table S1). To increase the stringency of the analysis, we further prioritized 5268 genes with q-values < 0.001 and at least 20% difference in the mRNA expression in White and Black/African American patients (Fig. 2a). Among those genes, expression of 2501 genes was decreased in Black/African American patients, as compared to White breast cancer patients (BAA_{low} gene set). In contrast, the expression of 2767 genes was significantly higher in the Black/African Americans cohort than in White patients (BAA_{high} gene set). These data indicate that White and Black/African American breast cancer patients have very different genomic backgrounds with over 30% of the protein-coding genes expressed differently in these two patient cohorts.

Cancer association and pathway enrichment analysis

To determine whether the differentially expressed genes are associated with the regulation of oncogenic processes, we performed the enrichment analysis. First, using the carcinoma-associated gene sets defined in DisGeNET database [23], we found that both, BAA_{high} and BAA_{low} gene sets are significantly enriched (p-value < 0.01, q-value < 0.1) in genes associated with different cancer types, including colon (q-value = 0.006), liver (q-value = 0.036 for BAA_{high} and q-value = 0.087 for BAA_{low}), pancreatic (q-value = 0.068), and lung cancers (q-value = 0.091) (Fig. 2b). Moreover, genes associated with breast carcinoma were among the most significantly overrepresented genes in both, BAA_{high} (p-value < 0.001, q-value < 0.031) and BAA_{low} sets (p-value < 0.001, q-value = 0.004, Fig. 2b). 2567 out of 2767 BAA_{high} genes (92%) are overexpressed in at least 5% of Black/African

American breast cancer patients, supporting their potential roles in breast carcinogenesis.

Then, we sought to determine specific biological programs associated with identified differentially expressed genes. Through the enrichment analysis of signaling and metabolic pathways defined in the KEGG database [24], we found that overexpressed BAA_{high} genes (BAA_{OVR} genes) showed the enrichment in genes associated with several major oncogenic pathways. The most significant enrichment (p-value ≪ 0.001, q-value ≪ 0.001) was observed for the cell cycle-associated genes. Furthermore, genes involved in senescence, phagosome maturation, JAK-STAT, cAMP, cGMP-PKG, and PIK3-AKT signaling pathways, retinol and pyrimidine metabolism, and oxidative phosphorylation (Fig. 2c) were significantly overrepresented in the BAA_{OVR} genes (p-value < 0.001, q-value < 0.05). Thus, overexpression of genes upregulated in Black/African American patients may promote breast cancer development and progression through the dysregulation of multiple oncogenic processes.

To identify functional modules of co-regulated genes we applied the weighted gene co-expression network analysis (WGCNA) [19, 20]. The WGCNA performed for all 5268 BAA_{high} and BAA_{low} genes revealed 12 distinct modules of significantly co-expressed genes (Additional file 1: Table S2, Additional file 2: Figure S1). The "pink" (332 genes), "black" (355 genes), "cyan" (846 genes), "red" (368 genes), "green" (431 genes), and "magenta" (292 genes) modules were comprised almost completely by BAA_{high} genes. The "yellow" (521 genes), "blue" (798 genes), "brown" (648 genes), "purple" (174 genes), "greenyellow" (164 genes), and "tan" (143 genes) modules included mostly BAA_{low} genes. To uncover the biological pathways associated with individual modules, we performed the enrichment analysis using the gene sets defined in the KEGG database [24] (Additional file 1: Table S3; Additional file 2: Figure S2). We found that five modules were more than tenfold overrepresented by genes involved in pathways defined in the KEGG database as compared to the reference human genome. Among all modules, the most significant enrichment was determined for the "pink" module, which appeared to be overrepresented (overrepresentation fold, OVF = 21.08, q-value = 8.78×10^{-26}) in the cell cycle regulating genes (Additional file 1: Table S3). An equally high overrepresentation (OVF = 20.83, q-value = 2.38×10^{-6}) was determined for the "magenta" module that was enriched in genes that control primary immunodeficiency, including ADA, CD19, CD79A, IGLL1, TAP1, TAP2, TNFRSF13C, and ZAP70. The "green" module appeared to be enriched in genes involved in oxidative phosphorylation (OVF = 12.5, q-value = 8.50×10^{-14}), ribosome (OVF = 12.43, q-value = 4.83×10^{-12}), and genes involved

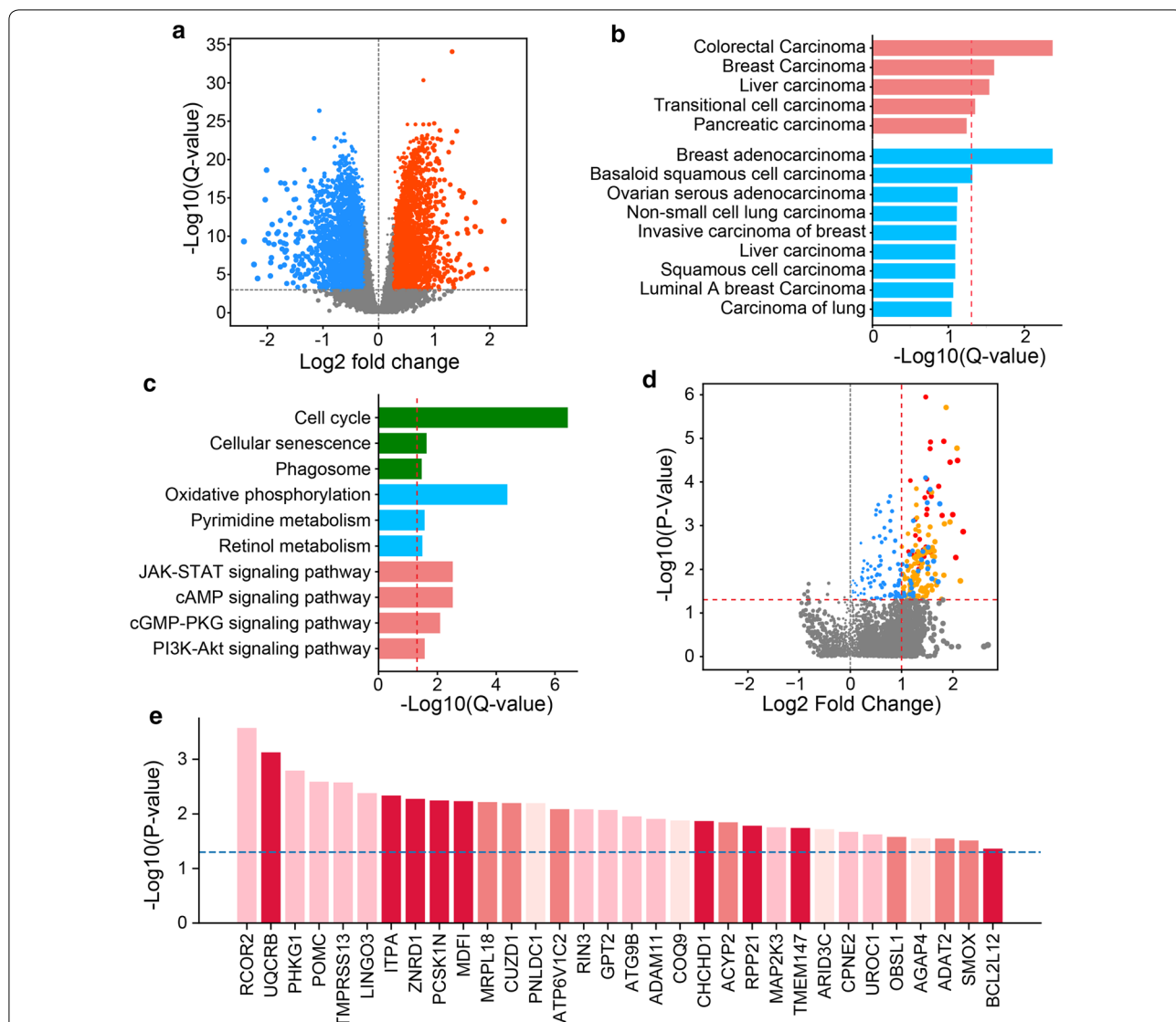Yang *et al. J Transl Med*   (2020) 18:334

Page 6 of 19

Fig. 2 Differential expression and breast cancer patient survival analysis. **a** Differential expression analysis revealed 2501 genes downregulated (blue) and 2767 genes upregulated (orange) in Black/African American (BAA) patients as compared to White patients. **b** Enrichment analysis of the disease-associated datasets revealed strong association (p-value < 0.01 and q-value < 0.1) between different cancer types and genes either upregulated (red) or downregulated (blue) in BAA breast cancer patients as compared to White patients. The dash-line indicates the q-value < 0.05 cut-off for the most statistically significant enrichment. **c** Enrichment analysis revealed the biological processes (green), metabolic (blue), and signaling (red) pathways associated with genes overexpressed in at least 5% of BAA breast cancer patients and upregulated as compared to White patients. **d** The volcano plot shows the association of high gene expression with decreased survival of BAA breast cancer patients. The fold change was calculated as the ratio of mean survival time of patients with normal gene expression to mean survival time of patients with gene overexpression. Overexpression of 257 genes (blue) correlates with decreased BAA patient survival (log-rank p_{BAA}-values < 0.05). Overexpression of 147 genes was not correlated with reduced survival of White patients with the p_{WT}-value > 0.1 (orange). The most significant difference between BAA and White breast cancer patient survival was observed for overexpression of 32 genes highlighted in red (p_{BAA}-values < 0.01, p_{WT}-value > 0.1, $P_{WT/BAA}$-value < 0.05, MST_{WT}/MST_{BAA} > 2). **e** The top-32 genes with the most BBA patient-specific contribution in decreased survival are shown. Y-axis indicates the log-rank p-values calculated for BAA and White patients with gene overexpression. The blue line indicates the p-value cut-off of 0.05. The bars are colored based on the frequency of gene overexpression in BAA breast cancer patients. Light-pink: 5–6%, pink: 6–10%, light-red: 10–15%, dark-red: > 15%

in neurodegenerative disorders, such as the Parkinson's disease (PD) (OVF = 11.31, q-value = 4.83 × 10^{-12}). Notably, multiple studies have suggested that the development of PD and cancer, including breast cancer, can progress through the same genes and molecular mechanisms [31–33]. In contrast to "pink", "magenta", and "green"

Yang *et al. J Transl Med* (2020) 18:334

Page 7 of 19

modules, the "tan" and "greenyellow" modules are comprised of genes with higher expression in White breast cancer patients as compared to Black/African American patients. We found that the "tan" module is enriched in genes involved in extracellular matrix receptor interactions (OVF = 15.31, q-value = 2.39×10^{-6}). The "greenyellow" module appeared to be overrepresented in the ATP-binding cassette (ABC) transporters (OVF = 19.77, q-value = 4.85×10^{-5}), and genes that control tyrosine metabolism (OVF = 10.15, q-value = 2.98×10^{-2}) and the complement and coagulation cascades (OVF = 14.37, q-value = 4.85×10^{-5}). The role of the ATP-binding cassette (ABC) transporters in tumorigenesis of White breast cancer patients is further supported by the enrichment of the "brown" module in basal transcription factors, including the ATP-binding cassette subfamily members ABCA9, ABCC9, ABCG2, ABCB1, ABCA6, and ABCA8. Previous studies have demonstrated the association of ATP-binding cassette transporters with breast cancer aggressiveness and reduced survival of breast cancer patients [34, 35]. We also noticed that the "blue" module is enriched (OVF = 6.21, q-value = 0.036) in genes that control sphingolipid metabolism that play critical functions in cancer growth and progression [36]. Furthermore, the "purple" module appeared to be enriched in genes that are associated with the dilated cardiomyopathy, a known side effect of breast cancer radiotherapy and chemotherapy [37, 38].

Differential expression and survival disparity

The genes with abnormally high expression may represent putative targets for therapeutic intervention. We used the set of 2567 BAA$_{OVR}$ genes to determine the impact of their overexpression on breast cancer patient survival. We found that overexpression of 257 BAA$_{OVR}$ genes (Additional file 1: Table S4, Group I) correlates with decreased survival of Black/African American patients (p_{BAA}-value < 0.05) (Fig. 2d). Furthermore, the overexpression of 174 out of 257 genes (Additional file 1: Table S4, Group II) correlated with more than twofold decreased survival. Among the 174 genes, overexpression of 147 genes (Additional file 1: Table S4, Group III) was associated with the reduced survival of Black/African American patients, but not White patients (p_{WT}-value > 0.1, Fig. 2d). This group of genes includes several genes previously linked with breast cancer

development and progression. For example, overexpression of protein arginine methyltransferase 1 (PRMT1) has been associated with the methylation of the transcription factor C/EBPα and inhibition of its tumor suppressor function in breast cancer [39]. Interestingly, PRMT1 knockdown was also correlated with decreased EGFR activity and suppressed proliferation of in MDA-MB-468 breast cancer cells that are derived from an African American breast cancer patient [40]. Kinesins KIF1C and KIFC3 promotes breast cancer cell growth and survival and mediate taxane resistance [41, 42]. Syndecan-1 (SDC1) has been linked with the accelerated metastasis of breast cancer to the brain [43]. Meanwhile, our data revealed genes previously not associated with the increased breast cancer progression, providing new opportunities for therapeutic interventions in breast cancer.
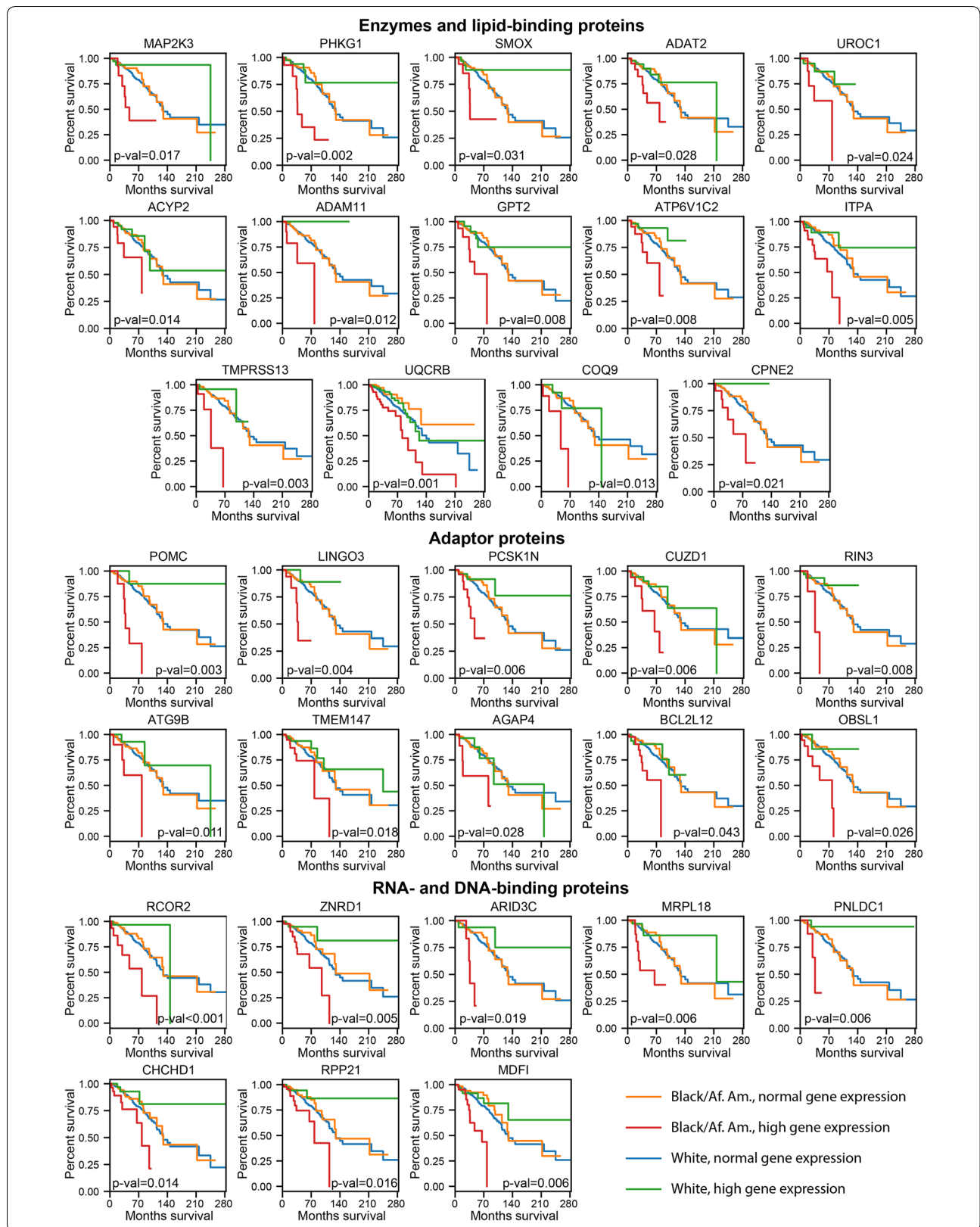
To further prioritize genes with the most significant contribution to the survival disparity between African American and White patients, we applied more stringent statistical cut-offs: p_{BAA}-value < 0.01, p_{WT}-value > 0.1, a significant difference between survival time of White and Balck/African American patients with the overexpressed gene ($p_{WT/BAA}$-value < 0.05), and at least twofold decreased mean survival time of Black/African American patients (MST$_{BAA}$) comparing to the MST of White patients (MST$_{WT}$) determined for the samples with the overexpressed gene. Using these parameters, a total of 32 genes with a most significant and race-specific impact on the breast cancer tumorigenesis in Black/African American patients have been prioritized (Additional file 1: Table S4, Group IV, Fig. 2e, and Fig. 3).

Evaluation of protein druggability for therapeutic discovery

To assess the potential druggability of the top-32 prioritized proteins, we have classified them into three groups based on the protein types (Fig. 3). RCOR2, ZNRD1, ARID3C, MRPL18, PNLDC1, CHCHD1, RPP21, MDFI are either DNA- or RNA-binding proteins. These proteins may represent the most challenging targets for direct interrogation with small molecules or specific antibodies due to their nuclear localization the lack of a defined pocket for a small-molecule binding. POMC, LINGO3, PCSK1N, CUZD1, RIN3, ATG9B, TMEM147, AGAP4, BCL2L12, OBSL1 also lack an enzymatic activity

(See figure on next page.)

Fig. 3 Survival curves for the top-32 genes that contribute to survival disparity between Black/African American and White patients. Orange and red lines indicate the survival of Black/African American breast cancer patients with normal and overexpressed gene levels, respectively. Blue and green lines indicate the survival of White breast cancer patients with normal and overexpressed gene levels, respectively. The log-rank p-values calculated for the survival rates of White and Black/African American patients with gene overexpression are indicated

Yang *et al. J Transl Med*   (2020) 18:334

Page 8 of 19

Enzymes and lipid-binding proteins

Adaptor proteins

RNA- and DNA-binding proteins

— Black/Af. Am., normal gene expression
— Black/Af. Am., high gene expression
— White, normal gene expression
— White, high gene expression

Yang *et al. J Transl Med*   (2020) 18:334

Page 9 of 19

and contribute in breast cancer tumorigenesis acting as adaptors for other proteins. A large area, hydrophobicity, and relatively flat configuration of the protein–protein interaction (PPI) interface surfaces are among the limiting factors for the design and discovery of low molecular weight PPI inhibitors [44]. On the other hand, the growing number of potent cell-permeable inhibitors for PPI discovered over the past decades, including the FDA-approved BCL2 inhibitor venetoclax [45], indicates the PPI druggability for therapeutic discovery [46]. Meanwhile, enzymes and receptors represent the largest class of therapeutic targets [47]. We found that 14 out of 32 proteins belong to protein families known to be druggable by low molecular weight compounds. Specifically, COQ9 and CPNE2 are the lipid-binding proteins with a defined binding site for a lipid molecule that can be targeted by small molecules [48]. Furthermore, 12 proteins belong to different types of enzymes, including a subunit of the ubiquinol-cytochrome c oxidoreductase UQCRB, serine protease TMPRSS13, inosine triphosphate pyrophosphatase ITPA, proton ATPase ATP6V1C2, Alanine aminotransferase GPT2, metalloproteinase ADAM11, acylphosphatase ACYP2, urocanate hydratase UROC1, tRNA-specific adenosine deaminase ADAT2, spermine oxidase SMOX, and two kinases: PHKG1 and MKK3 also known as MAP2K3. The discovery of potent inhibitors for these enzymes may lead to new therapeutic strategies for African American breast cancer patients.

COX regression survival analysis for TNBC Black/African American patients

The COX regression analysis is a widely used approach to identify predictive biomarkers of poor clinical outcomes [49, 50]. We applied the COX regression analysis to determine the overall impact of the prioritized genes on clinical outcomes of Black/African American patients specifically with the triple-negative breast cancer subtype. First, we built the univariate COX regression models to determine the hazard ratios and significance for each of the 32 prioritized genes. We found that for each gene the Hazard ratio values (HR) were higher than 1 indicating a positive correlation between gene expression and decreased patient survival (Additional file 1: Table S5). This result is consistent with the Kaplan–Meier analysis performed for all breast cancer subtypes (Fig. 3). Eight out of 32 genes demonstrated highly significant correlation with poor clinical outcomes with the Hazard ratio (HR) > 2 and the p-values ≤ 0.05, including ACYP2, ADAT2, AGAP4, CHCHD1, MKK3, MRPL18, RPP21, and ZNRD1 (Additional file 1: Table S5; Additional file 2: Figure S3).
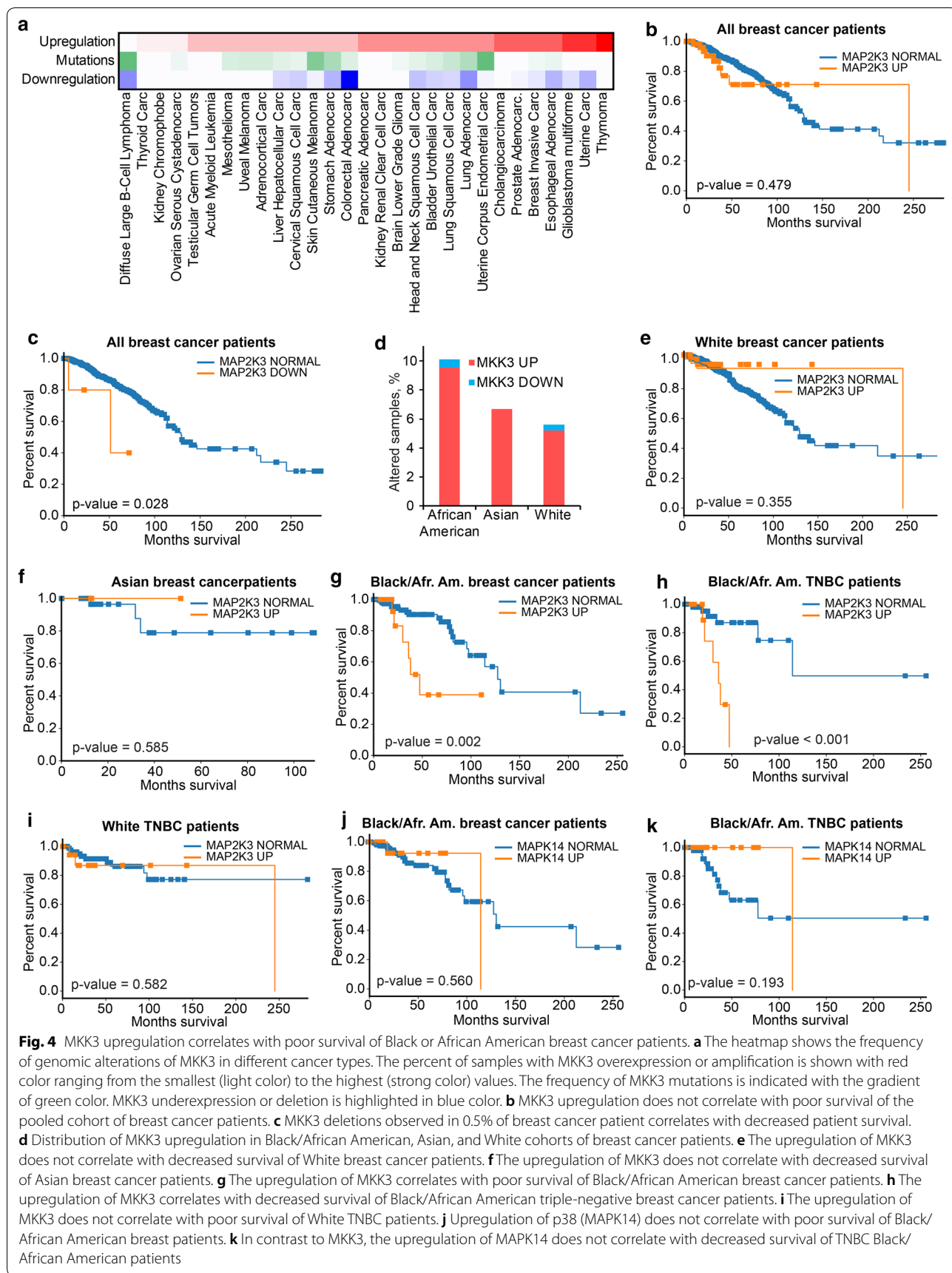
To evaluate the combined effect of these 8 genes on the clinical outcomes of TNBC Black/African American patients, we built a multivariate COX regression model. The resulting Model 1 demonstrated a high concordance index (c-index = 0.93) and statistical significance (p-value = 1×10^{-4}), indicating the satisfactory prognostic ability of the model (Additional file 1: Table S6). The detailed evaluation of the model revealed that expression of MKK3 (HR = 27.98, p-value = 0.002), AGAP4 (HR = 1.73, p-value = 0.017), and ACYP2 (HR = 1.30, p-value = 0.04) made the most significant contribution to the model. To determine if MKK3, AGAP4, and ACYP2 can be used as markers for poor clinical outcomes, we built another model based on these three genes only. The resulting Model 2 (Additional file 1: Table S6) was characterized by an equally high c-index of 0.91 and improved statistical significance (p-value = 2×10^{-5}) as compared to the 8-parameter model. We noticed that in both Model 1 and Model 2 the highest HR value was obtained for MKK3, suggesting its significance for clinical outcomes of Black/African American TNBC patients.

MKK3 overexpression promotes triple-negative breast cancer in African American patients

MKK3 is frequently altered in different cancers and recent studies have suggested that MKK3 may contribute in tumorigenesis in multiple cancer types [51–55]. Analysis of the TCGA PanCancer datasets indicates that MKK3 is mutated in 5% of uterine carcinoma, 5% of B-cell lymphoma, and 4% of skin melanoma patients. MKK3 is homozygously deleted in 6% of colon cancer patients. On the other hand, MKK3 is either overexpressed or amplified in 3 to 8% of patients in the vast majority of cancers, including thymoma (8%), glioblastoma multiform (7%), and breast invasive carcinoma (6%) (Fig. 4a, Additional file 1: Table S7). Furthermore, MKK3 overexpression can be triggered by TP53 mutations [56], that can link MKK3 to TP53-dependent cancers, such as breast cancer, particularly in African American patients.

The evaluation of the overall survival data for the pooled dataset of 1055 samples revealed no correlation between the MKK3 overexpression and patient survival (p = 0.479, Fig. 4b, Additional file 1: Table S8). Instead, five patients with lost MKK3 demonstrated decreased survival comparing to patients with normal MKK3 (p = 0.028, Fig. 4c). This observation is consistent with a previous report that MKK3 may play a tumor-suppressive role in breast cancer [57]. Meanwhile, we found that MKK3 is the most frequently overexpressed in the Black/African American cohort (9.6%) (Fig. 4d). In the Asian and White breast cancer patients, MKK3 is overexpressed in 6.7% and 5.2%, respectively. Conversely, MKK3 downregulation is not frequent in breast cancer patients. MKK3 is not underexpressed or deleted in the Asian cohort, and it

Yang *et al. J Transl Med*   (2020) 18:334

Page 10 of 19

Fig. 4 MKK3 upregulation correlates with poor survival of Black or African American breast cancer patients. **a** The heatmap shows the frequency of genomic alterations of MKK3 in different cancer types. The percent of samples with MKK3 overexpression or amplification is shown with red color ranging from the smallest (light color) to the highest (strong color) values. The frequency of MKK3 mutations is indicated with the gradient of green color. MKK3 underexpression or deletion is highlighted in blue color. **b** MKK3 upregulation does not correlate with poor survival of the pooled cohort of breast cancer patients. **c** MKK3 deletions observed in 0.5% of breast cancer patient correlates with decreased patient survival. **d** Distribution of MKK3 upregulation in Black/African American, Asian, and White cohorts of breast cancer patients. **e** The upregulation of MKK3 does not correlate with decreased survival of White breast cancer patients. **f** The upregulation of MKK3 does not correlate with decreased survival of Asian breast cancer patients. **g** The upregulation of MKK3 correlates with poor survival of Black/African American breast cancer patients. **h** The upregulation of MKK3 correlates with decreased survival of Black/African American triple-negative breast cancer patients. **i** The upregulation of MKK3 does not correlate with poor survival of White TNBC patients. **j** Upregulation of p38 (MAPK14) does not correlate with poor survival of Black/African American breast patients. **k** In contrast to MKK3, the upregulation of MAPK14 does not correlate with decreased survival of TNBC Black/African American patients

Yang *et al. J Transl Med*   (2020) 18:334

Page 11 of 19

was deleted in 3 White patients (0.4%), 1 Black patient (0.6%), and 1 patient with the unspecified race (1.1%). In agreement with the genomic status of MKK3, its upregulation does not correlate with poor survival of White (p = 0.355, Fig. 4e) nor Asian (p = 0.585, Fig. 4f) patients. In contrast, a strong decrease in patient survival (p = 0.002) was observed for the Black/African American cohort (Fig. 4g).

The analysis of the histological subtypes of breast patients indicates, that the majority of Black/African American patients in the breast cancer TCGA PanCancer cohort (N = 178) have either basal-like/triple-negative (TNBC) (63 patients) or Luminal A (61 patients) breast cancer. The number of patients with Luminal B, HER2, and Normal breast cancers was 28, 16, and 10, respectively. Surprisingly, MKK3 was upregulated in only one patient with the Luminal A breast cancer. In contrast, MKK3 was overexpressed in 19% of Black/African American TNBC patients.

Similar to the combined set of breast cancer samples of all subtypes, the MKK3 overexpression correlates with poor survival of TNBC Black/African American patients (p < 0.001, Fig. 4h), but not White patients (p = 0.582, Fig. 4i). Moreover, through a systematic analysis of all breast cancer subtypes in all racial groups of patients, we have determined that MKK3 upregulation correlates uniquely with the poor survival of Black/African American patients specifically with the TNBC, and not with any other race or other breast cancer subtypes (Additional file 1: Table S9).

MKK3 promotes TNBC through a p38-distinct mechanism

MKK3 is the main activator of its only known substrate p38 which plays a key role in the induction of apoptosis and regulation of inflammation in response to extracellular stress [58, 59]. It can be expected that the poor survival of Black/African American patients is also associated with p38 activation. p38 (encoded by the MAPK14 gene) is amplified or overexpressed in 9.5% of the Black/African American breast cancer patients. However, in contrast to MKK3, p38 upregulation does not correlate with decreases survival of Black/African American patients neither for all breast cancer subtypes (p = 0.986) (Fig. 4j) nor specifically for the TNBC (p = 0.193, Fig. 4k). These data suggest a p38-distinct role for MKK3 in TNBC tumorigenesis. These results are further supported by the recent discovery of MKK3 as a hub protein in the PPI network determined for cancer-associated proteins [60, 61]. It was shown that besides p38, MKK3 can bind to multiple other proteins, including several drivers of breast cancer, such as CDK4, AURKA, FGFR4, EPHA2, and MYC [60].
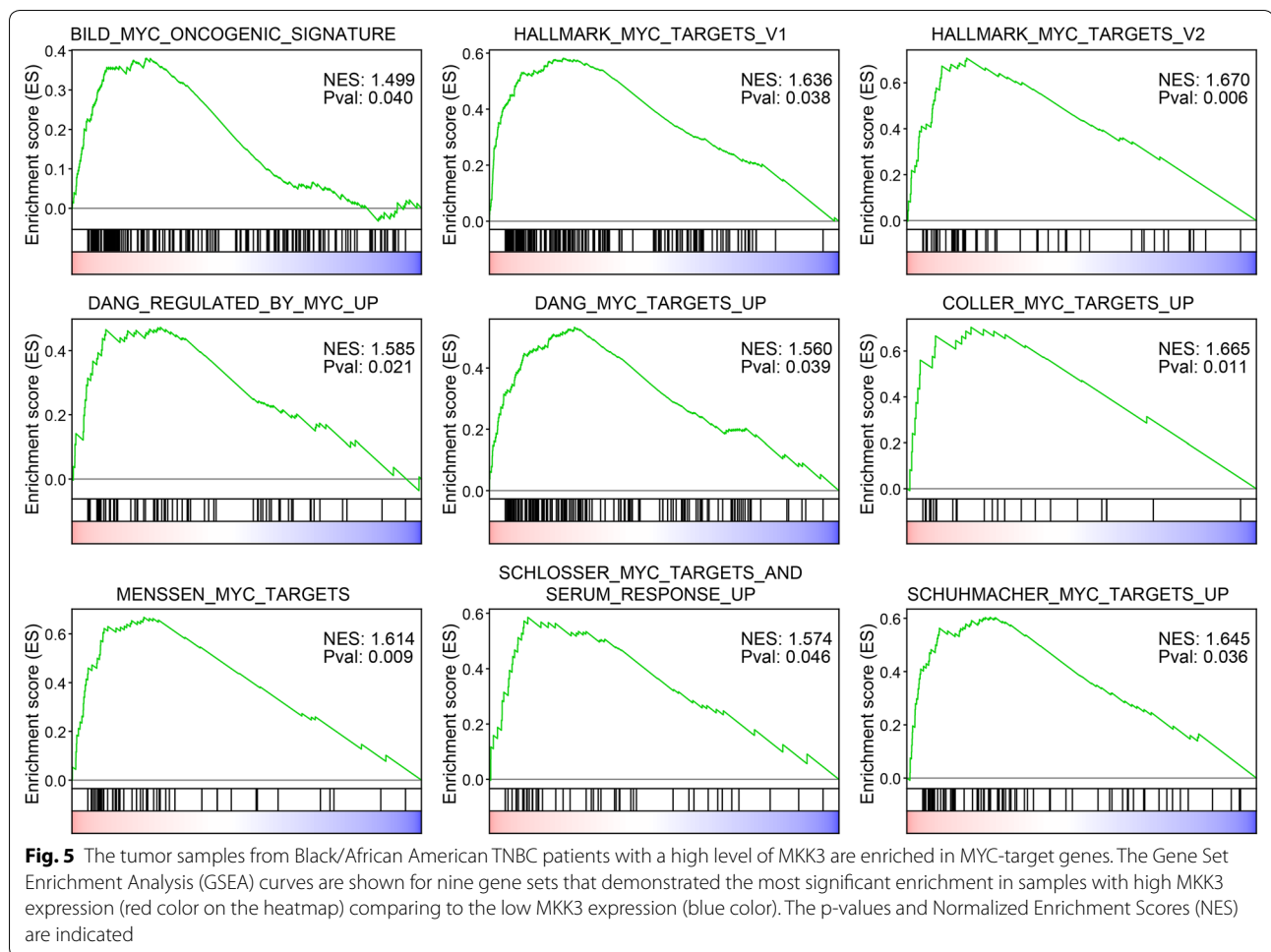
MKK3 activates MYC transcriptional program in TNBC African American patients

To uncover the molecular bases underlying the decreased survival of TNBC Black/African American patients, we performed the Gene Set Enrichment Analysis (GSEA) [62] against 50 hallmark sets of genes that define signatures of specific biological state or process [26, 63, 64] (Additional file 1: Table S10). Only five gene sets showed significant enrichment in samples with upregulated MKK3 expression (p-value < 0.05 and FDR < 25%), including genes involved in unfolded protein response, mTORC1 signaling, response to the UV irradiation, and two sets of MYC target genes [64, 65] (Additional file 1: Table S10).

To further increase the confidence in the MKK3-MYC functional association, we have expanded the GSEA analysis using 16 more sets of MYC-upregulated genes independently defined in different studies (Additional file 1: Table S11). We found that 17 out of 18 tested MYC-target gene sets demonstrate the enrichment in Black/African American TNBC samples with a high level of MKK3. Furthermore, 9 out of 18 sets showed a statistically significant enrichment with the p-value < 0.05, including the MYC oncogenic signature genes derived from the DNA microarray analysis of the breast cancer cells (p = 0.040, FDR = 5.6%, normalized enrichment score, NES = 1.5) (Fig. 5). Meanwhile, no enrichment in MYC-target genes was found for the samples with upregulated p38, further supporting p38-distinct functions of MKK3 in Black/African American TNBC patients.

As a master regulator, MYC controls multiple oncogenic programs. We sought to determine biological pathways that could be dysregulated specifically in response to MKK3-mediated MYC MYC-activation. Based on the GSEA analysis for each MYC-dependent gene set we determined a total of 323 core genes that contribute the most to the enrichment. The pathway overrepresentation analysis revealed a strong association of 222 of MYC-regulated genes enriched in patients with overexpressed MKK3 with 117 signaling and metabolomic pathways defined in the Reactome database (q-value < 0.01, at least twofold overrepresentation as compared to the reference human genome). We found that the cell cycle and RNA metabolism and processing appeared among the pathway with the most significant overrepresentation in MKK3-MYC core enrichment genes (Additional file 1: Table S12). Interestingly, the REACTOME_DEASES gene set also appeared within the top-10 the most overrepresented pathways suggesting the pathological functions for the genes upregulated through the MKK3-MYC interaction.
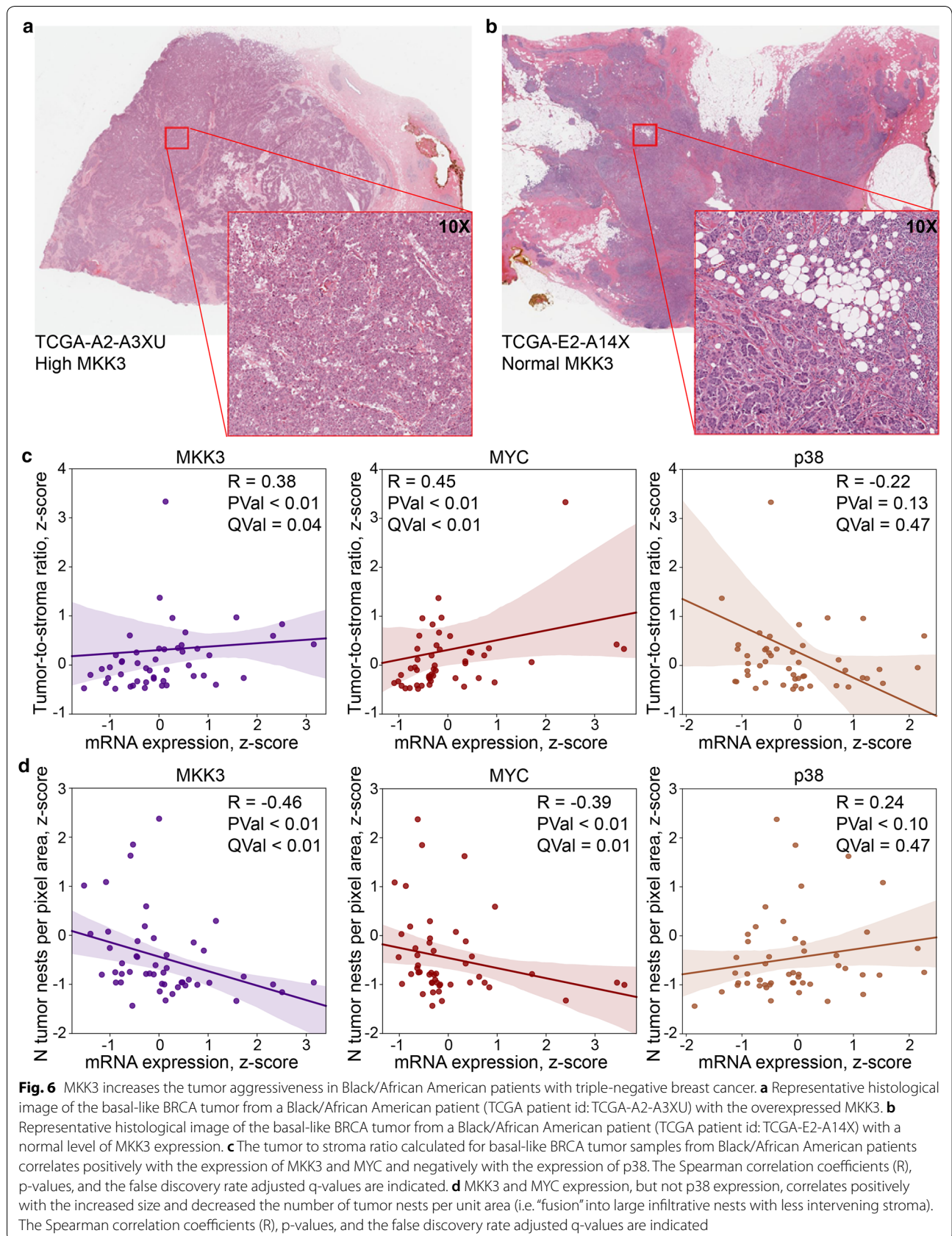
To support the clinical significance of MKK3 as a mediator of TNBC pathology, we performed a quantitative

Yang *et al. J Transl Med*   (2020) 18:334

Page 12 of 19

Fig. 5 The tumor samples from Black/African American TNBC patients with a high level of MKK3 are enriched in MYC-target genes. The Gene Set Enrichment Analysis (GSEA) curves are shown for nine gene sets that demonstrated the most significant enrichment in samples with high MKK3 expression (red color on the heatmap) comparing to the low MKK3 expression (blue color). The p-values and Normalized Enrichment Scores (NES) are indicated

analysis of the histopathological images from Black/African American TNBC patients (Fig. 6a, b). We found that the high level of MKK3 expression is associated with the increased overall tumor to stroma ratio (Spearman R = 0.38, p-value < 0.01, q-value = 0.04, Fig. 6c), and fewer discrete tumor "nests" (Spearman R = −0.46, p-value < 0.01, q-value < 0.01, Fig. 6d). Note that the smaller number of discrete nests is, in this phenotype, a consequence of their larger size, causing less intervening stroma and apparent "fusion" into large invasive tumors (Fig. 6a versus b). Similar trends have been observed for MYC. The elevation of MYC expression leads to increased overall tumor-to-stroma ratio (Spearman R = 0.45, p-value < 0.01, q-value < 0.01, Fig. 6c), as well as fewer discrete tumor nests (Spearman R = −0.39, p-value < 0.01, q-value = 0.01, Fig. 6c), which are, individually, significantly larger in size (Spearman R = 0.42, p-value < 0.01, q-value < 0.01, not shown). Unlike MKK3 and MYC, p38 upregulation does not correlate with either the overall tumor-to-stroma ratio (Spearman R = −0.22, p-value = 0.13, q-value = 0.47, Fig. 6b) or the

number of discrete tumor nests (Spearman R = 0.24, p-value = 0.10, q-value = 0.47, Fig. 6c). Moreover, the observed trends, although not statistically significant, were opposite compared to trends determined for MKK3 and MYC. Together, these data suggest a critical role of MKK3 in promoting the TNBC tumorigenesis in African American patients and its strong association with the activation of the MYC program.

To identify which of MKK3-activated MYC-regulated genes can contribute most in poor clinical outcomes of African American patients, we evaluated the correlations between overexpression of MKK3-MYC core enrichment genes and TNBC patient survival. We prioritized 8 MKK3-MYC core enrichment genes whose overexpression correlates with decreased survival of TNBC African American patients, including EIF5AL1 (log-rank test p-value = 0.029), EIF5A (p-value = 0.015), SNAI1 (p-value = 0.050), TAF12 (p-value = 0.004) as well APEX1 (p-value = 0.001), FASN (p-value = 0.033), HNRNPA2B1 (p-value = 0.036), and GRSF1 (p-value < 0.001). Notably, overexpression of these genes does not worsen clinical

Yang *et al. J Transl Med*   (2020) 18:334

Page 13 of 19

Fig. 6 MKK3 increases the tumor aggressiveness in Black/African American patients with triple-negative breast cancer. **a** Representative histological image of the basal-like BRCA tumor from a Black/African American patient (TCGA patient id: TCGA-A2-A3XU) with the overexpressed MKK3. **b** Representative histological image of the basal-like BRCA tumor from a Black/African American patient (TCGA patient id: TCGA-E2-A14X) with a normal level of MKK3 expression. **c** The tumor to stroma ratio calculated for basal-like BRCA tumor samples from Black/African American patients correlates positively with the expression of MKK3 and MYC and negatively with the expression of p38. The Spearman correlation coefficients (R), p-values, and the false discovery rate adjusted q-values are indicated. **d** MKK3 and MYC expression, but not p38 expression, correlates positively with the increased size and decreased the number of tumor nests per unit area (i.e. "fusion" into large infiltrative nests with less intervening stroma). The Spearman correlation coefficients (R), p-values, and the false discovery rate adjusted q-values are indicated

Yang *et al. J Transl Med*   (2020) 18:334

Page 14 of 19

outcomes in Caucasian TNBC patients (p-values > 0.1), suggesting their unique functions in African American patients. We found that EIF5A, EIF5AL1, and SNAI1 are the most frequently overexpressed genes (> 20%) in African American TNBC patients. These genes also demonstrate the highest correlation with both MKK3 and MYC expression (Pearson correlation p < 0.01, Fig. 7a) and decreased survival of African American patients (Fig. 7b). Importantly, both Snail Family Transcriptional Repressor 1 (SNAI1) and Eukaryotic Translation Initiation Factor 5A (EIF5A) have been associated with the induction of the epithelial-to-mesenchymal transition (EMT) in breast cancer, promotion of breast cancer metastasis, and chemoresistance [66–68]. Together, these findings suggest a new function for MKK3 as an inducer of MYC-dependent epithelial-to-mesenchymal transition in African American TNBC patients (Fig. 7c).

Discussion

Breast invasive carcinoma is the most common cancer type in women. It is especially aggressive in African American patients. The discovery of new therapeutic targets is urgently needed to decrease breast cancer mortality and reduce the racial disparity in breast cancer outcomes. In contrast to the tumor suppressor genes, such as TP53 or BRCA1/2, that are lost due to deletions or mutations, the mRNA overexpression represents an actionable alteration that can be reached therapeutically. The identification of therapeutically actionable upregulated genes that contribute in poor clinical outcomes may facilitate the development of new clinical strategies in breast cancer. Toward this goal, we have performed a systematic analysis of clinical outcomes and differential gene expression in White and African American breast cancer patients.

We found that more than 2500 genes overexpressed in African American patients are also significantly upregulated in African Americans as compared to the White breast cancer patients. Our analysis has also confirmed 117 out of 142 (82%) genes previously reported as differentially expressed in African American and White/European cohorts of breast cancer patients [12]. The enrichment analysis revealed a strong functional association of these genes with breast cancer as well as several other cancer types, and multiple key oncogenic pathways including cell cycle, PI3K-AKT, and JAK-STAT pathways. Through the gene co-expression analysis integrated with the analysis of pathway overrepresentation, we determined specific modules of co-regulated genes. Notably, three distinct gene modules of genes with higher expression in African American patients as compared to White patients were significantly enriched in genes that control cell cycle progression, immunodeficiency, and oxidative
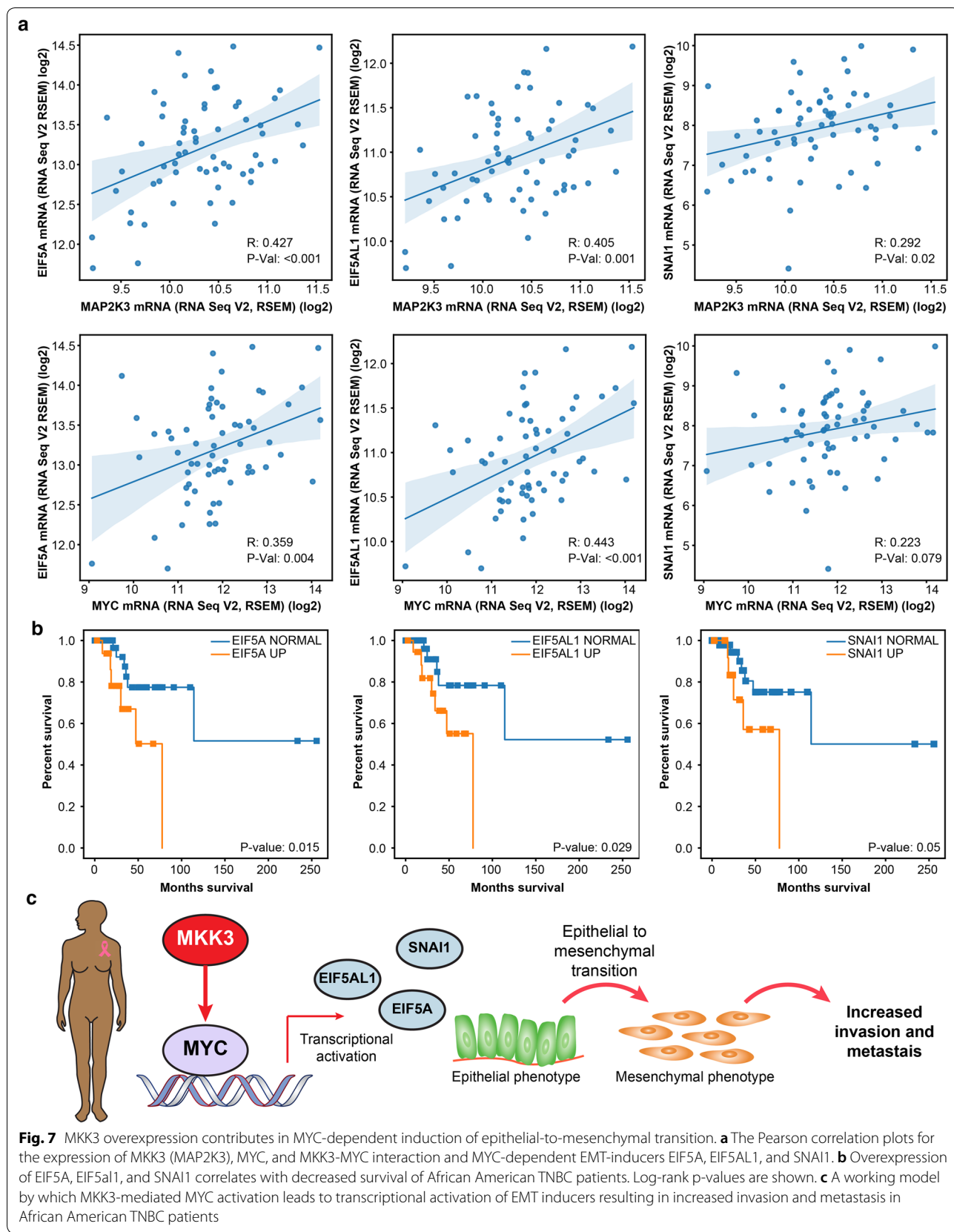
phosphorylation. The identification and targeting of the key druggable regulators of these fundamental oncogenic processes may facilitate the development of new clinical strategies to reduce survival disparity in breast cancer. Meanwhile, the diversity of differentially expressed genes and dysregulated pathways (summarized in Additional file 1: Table S3) indicate the heterogeneity and complexity of the molecular mechanisms underlying survival disparity in breast cancer. Thus, the discovery and prioritization of the most biologically clinically essential genes is critical to facilitate the translation of breast cancer patient genomics data into the clinic.

Through the rigorous statistical analysis, we prioritized 32 proteins that demonstrate the most prominent and race-specific association with decreased survival of African American women. We found that 14 of these prioritized proteins belong to proteins classes known to be druggable and thus represent promising targets for therapeutic discovery. Indeed, at least two proteins, ITPA and GPT2, are well-established therapeutic targets for rheumatoid arthritis and anxiety disorders with the FDA-approved inhibitors, azathioprine and phenelzine, respectively. Currently, phenelzine is also in clinical trials in patients with different cancer types, including patients with advanced or metastatic breast cancer. Our data may open new opportunities for the repurposing of these approved drugs and other reported inhibitors as the anticancer agents for African American breast cancer patients.

Through the COX regression analysis, we have identified several proteins as new promising targets for the therapeutic discovery in TNBC. The multivariate COX regression model suggests that expression of MKK3, AGAP2, and ACYP2 has a significant negative impact on clinical outcomes of TNBC Black/African American patients, and these genes may serve as putative biomarkers for decreased patient survival. Among them, MKK3 showed the most dramatic impact on the survival of African American patients, specifically with triple-negative breast cancer.

The integration of the survival data analysis, gene set enrichment analysis, and the analysis of the breast cancer histopathological images revealed that MKK3 can promote TNBC tumorigenesis through the activation of the MYC transcriptional program.

MKK3 is a well-established activator of the p38 pro-inflammatory and pro-apoptotic pathway [58], and MKK3 functions have been associated primarily with the regulation of p38 signaling [69–74]. The loss of p38-activation may promote tumor growth in cancers with a decreased level of MKK3 [57, 75], suggesting its tumor-suppressive function. On the other hand, the oncogenic role for MKK3 has been reported in

Yang *et al. J Transl Med*   (2020) 18:334

Page 15 of 19

Fig. 7 MKK3 overexpression contributes in MYC-dependent induction of epithelial-to-mesenchymal transition. **a** The Pearson correlation plots for the expression of MKK3 (MAP2K3), MYC, and MKK3-MYC interaction and MYC-dependent EMT-inducers EIF5A, EIF5AL1, and SNAI1. **b** Overexpression of EIF5A, EIF5al1, and SNAI1 correlates with decreased survival of African American TNBC patients. Log-rank p-values are shown. **c** A working model by which MKK3-mediated MYC activation leads to transcriptional activation of EMT inducers resulting in increased invasion and metastasis in African American TNBC patients

Yang *et al. J Transl Med*   (2020) 18:334

Page 16 of 19

multiple tumor types, including melanoma, colorectal, liver, esophageal, cervical, and breast cancers [51, 52, 76–80].

Analysis of genomics data shows that MKK3 is either up- or downregulated in different cancer types and different groups of cancer patients. Thus, MKK3 can play a dual role in cancer: one as a lost tumor suppressor acting through the p38-pathway [81, 82], and another as an oncogene through upregulation of different oncogenic programs, such as the MYC transcription. MYC is a major tumor driver, and the master regulator of multiple key cellular processes, including cell growth and proliferation, immune response, and metabolism. Over the past decades, MYC became a well-established and highly-appealing therapeutic target in breast cancer [83]. Therapeutic regulation of MYC activation may provide new clinical strategies to suppress different oncogenic mechanisms in African American breast cancer patients [84–86].

Recent studies have established strong functional connectivity between TP53 mutations in breast cancer patients and MYC activation [87]. The frequency of TP53 mutations is more than 40% higher in African American patients than in White patients [11, 12]. Furthermore, MKK3 overexpression was linked to TP53 mutations in colon and breast cancer cells [56]. These data suggest that MKK3 may cooperate with TP53 to activate MYC and promote TNBC progression in the Black/African American cohort. This model is further supported by a synthetic lethal relationship [88] and a physical protein–protein interaction observed between MKK3 and MYC in cancer cells [60, 61, 89, 90]. Thus, the MKK3-MYC oncogenic axis may represent a new promising target for therapeutic discovery for African American TNBC patients.

Through a systematic gene set enrichment analysis and clinical outcome profiling, we have discovered a new oncogenic function for MKK3 in African American TNBC patients as an activator of MYC-dependent epithelial-to-mesenchymal transition, specifically through EIF5A, EIF5AL1, and SNAI1 genes. Overexpression of these MKK3-MYC signature genes has been linked with the induction of epithelial-to-mesenchymal transition in breast cancer and strongly correlates with worsened clinical outcomes in African American patients. These findings suggest that the inhibition of MKK3-MYC interaction itself and its downstream-activated genes EIF5A, EIF5AL1, and SNAI1 may provide new therapeutic options for African American patients with triple-negative breast cancer.

Conclusions

In this study, the relationship between gene expression and survival disparity in breast cancer patients has been investigated. Through the integrative statistical analyses of clinical and genomics data, we identified 32 genes as putative targets for therapeutic intervention in Black or African American breast cancer patients. The success of the translation of these findings into the clinic would certainly rely on the further rigorous experimental validation and can be complicated by diverse molecular mechanisms underlying survival disparity. To facilitate this process, the identification and prioritization of the most biologically relevant, clinically significant, and druggable targets is crucial. Toward this goal, we performed a systematic analysis of the genomics and clinical data available for MKK3 gene that demonstrated one of the most significant negative impacts on the survival of African Americans with triple-negative breast cancer. Through a comprehensive systems biology approach, we have linked MKK3-mediated worsened clinical outcomes in African American TNBC patients with the activation of MYC transcriptional program. We have determined that besides its well-defined function in the p38-inflammatory pathway, MKK3 can induce MYC-dependent epithelial-to-mesenchymal transition in breast cancer patients in part through upregulation of EIF5A, EIF5AL1, and SNAI1 genes. These findings suggest new oncogenic functions for MKK3 in breast cancer and define MKK3-MYC interaction as a promising target to reduce survival disparity in African American TNBC patients.

Supplementary information

Additional file 1: Table S1. Differential gene expression analysis. **Table S2.** Modules of co-expressed genes identified with the Weighted Gene Co-expression Network Analysis (WGCNA). **Table S3.** Gene module pathway overrepresentation analysis. **Table S4.** Breast cancer patient survival analysis for genes upregulated in African American patients as compared to White patients. **Table S5.** Univariate COX regression survival analysis for TNBC Balck/African American patients. **Table S6.** Multivariate COX regression survival analysis for TNBC Balck/African American patients. **Table S7.** Analysis of genomic alterations of MKK3. **Table S8.** Genomic status of MKK3 (MAP2K3 gene) and p38 (MAPK14 gene) in BRCA patients and associated clinical data. **Table S9.** Correlation between MKK3 overexpression and breast cancer patient survival. **Table S10.** MKK3 gene set enrichment analysis using the MSigDB cancer Hallmark sets. **Table S11.** MKK3 gene set enrichment analysis for MYC-regulated gene sets. **Table S12.** Pathway overrepresentation analysis for MYC-dependent genes enriched in TNBC patients with overexpressed MKK3.

Additional file 2. Additional figures.

Abbreviations

ACR: American College of Radiology; ACYP2: Acylphosphatase 2; ADAT2: Adenosine deaminase TRNA specific 2; AGAP2: ArfGAP with GTPase domain, Ankyrin repeat and PH domain 2; BAA: Black or African American; CHCHD1: Coiled-coil-helix-coiled-coil-helix domain containing 1; CDH1: Cadherin 1;

Yang *et al. J Transl Med*   (2020) 18:334

Page 17 of 19

CCNE1: Cyclin E1; EIF5A: Eukaryotic translation initiation factor 5A; EIF5AL1: Eukaryotic translation initiation factor 5A like 1; EMT: Epithelial-to-mesenchymal transition; FDA: The United States Food and Drug Administration; FBXW7: F-Box and WD repeat domain containing 7; GDC: Genomics Data Commons; GSEA: Gene set enrichment analysis; HGNC: HUGO Gene Nomenclature Committee; HSI: Hue-Saturation-Intensity; HUGO: Human Genome Organization; MKK3: Mitogen-activated protein kinase kinase 3; MAP2K3: Mitogen-activated protein kinase kinase 3; MRPL18: Mitochondrial ribosomal protein L18; MST: Mean survival time; MYC: MYC proto-oncogene, basic helix-loop-helix transcription factor; NCI: National Cancer Institute; PIK3CA: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; RPP21: Ribonuclease P/MRP subunit P21; RB1: Retinoblastoma protein 1; SNAI1: Snail family transcriptional repressor 1; TCGA: The Cancer Genome Atlas; TNBC: Triple-negative breast cancer; TP53: Tumor protein 53; WGCNA: Weighted gene co-expression network analysis; WSI: Whole-slide images; ZNRD1: RNA Polymerase I Subunit H, also known as POLR1H.

Authors' contributions
XY and AAI designed the studies. XY, MA, and AAI conducted expression, clinical, and histological image analysis and calculations. XY, MA, LADC, YD, HF, AAI participated in data analysis, discussion, manuscript preparation and editing, XY, HF, and AAI wrote the paper. All authors read and approved the final manuscript.

Availability of data and materials
Training images and annotations are publicly available from TCGA sources.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Pharmacology and Chemical Biology, Emory University School of Medicine, Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA. [2] Emory Chemical Biology Discovery Center, Emory University School of Medicine, Emory University, Atlanta, GA, USA. [3] Department of Biomedical Informatics, Emory University School of Medicine, Emory University, Atlanta, GA, USA. [4] Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [5] Winship Cancer Institute, Emory University, Atlanta, GA, USA. [6] Department of Hematology & Medical Oncology, Emory University, Atlanta, GA, USA.

References
1. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global cancer in women: burden and trends. Cancer Epidemiol Biomarkers Prev. 2017;26(4):444–57.
2. Newman LA, Kaljee LM. Health disparities and triple-negative breast cancer in African American women: a review. JAMA Surg. 2017;152(5):485–93.
3. Yedjou CG, Sims JN, Miele L, Noubissi F, Lowe L, Fonseca DD, et al. Health and racial disparity in breast cancer. Adv Exp Med Biol. 2019;1152:31–49.
4. Amirikia KC, Mills P, Bush J, Newman LA. Higher population-based incidence rates of triple-negative breast cancer among young African-American women: implications for breast cancer screening recommendations. Cancer. 2011;117(12):2747–53.
5. Nolan TS, Ivankova N, Carson TL, Spaulding AM, Dunovan S, Davies S, et al. Life after breast cancer: 'Being' a young African American survivor. Ethn Health. 2019. https://doi.org/10.1080/13557858.2019.1682524.
6. Yedjou CG, Tchounwou PB, Payton M, Miele L, Fonseca DD, Lowe L, et al. Assessing the racial and ethnic disparies in breast cancer mortality in the United States. Int J Environ Res Public Health. 2017;14(5):486.
7. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. CA Cancer J Clin. 2019;69(6):438–51.
8. Monticciolo DL, Newell MS, Moy L, Niell B, Monsees B, Sickles EA. Breast cancer screening in women at higher-than-average risk: recommendations from the ACR. J Am Coll Radiol. 2018;15(3 Pt A):408–14.
9. Friebel TM, Andrulis IL, Balmana J, Blanco AM, Couch FJ, Daly MB, et al. BRCA1 and BRCA2 pathogenic sequence variants in women of African origin or ancestry. Hum Mutat. 2019;40(10):1781–96.
10. Hall MJ, Reid JE, Burbidge LA, Pruss D, Deffenbaugh AM, Frye C, et al. BRCA1 and BRCA2 mutations in women of different ethnicities undergoing testing for hereditary breast-ovarian cancer. Cancer. 2009;115(10):2222–33.
11. Keenan T, Moy B, Mroz EA, Ross K, Niemierko A, Rocco JW, et al. Comparison of the genomic landscape between primary breast cancer in African American versus White women and the Association of racial differences with tumor recurrence. J Clin Oncol. 2015;33(31):3621–7.
12. Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, et al. Comparison of breast cancer molecular features and survival by African and European Ancestry in The Cancer Genome Atlas. JAMA Oncol. 2017;3(12):1654–62.
13. Pitt JJ, Riester M, Zheng Y, Yoshimatsu TF, Sanni A, Oluwasola O, et al. Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. Nat Commun. 2018;9(1):4181.
14. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A comprehensive pan-cancer molecular study of gynecologic and breast cancers. Cancer Cell. 2018;33(4):690–705.
15. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375(12):1109–12.
16. Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2017. Nucleic Acids Res. 2017;45(D1):D619–25.
17. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12(4):R41.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B. 1995;57(1):289–300.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559.
20. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005. https://doi.org/10.2202/1544-6115.1128.
21. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. J Stat Softw. 2012;46(11):i11.
22. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics. 2008;24(5):719–20.
23. Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1):D845–55.
24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
25. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2018;46(D1):D649–55.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.

Yang *et al. J Transl Med*   (2020) 18:334

Page 18 of 19

27. Deng J, Dong W, Socher R, Li L, Kai L, Li F-F, editors. ImageNet: A large-scale image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009 20–25 June 2009.

28. Amgad M, Elfandy H, Hussein H, Atteya LA, Elsebaie MAT, Abo Elnasr LS, et al. Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics. 2019;35(18):3461–7.

29. Amgad M, Sarkar A, Srinivas C, Redman R, Ratra S, Bechert CJ, et al. Joint region and nucleus segmentation for characterization of tumor infiltrating lymphocytes in breast cancer. Proc SPIE Int Soc Opt Eng. 2019. https://doi.org/10.1117/12.2512892.

30. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graphics Appl. 2001;21(5):34–41.

31. Li Z, Zheng Z, Ruan J, Li Z, Tzeng CM. Chronic inflammation links cancer and Parkinson's disease. Front Aging Neurosci. 2016;8:126.

32. Li L. Secondary Parkinson disease caused by breast cancer during pregnancy: a case report. World J Clin Cases. 2019;7(23):4052–6.

33. Feng DD, Cai W, Chen X. The associations between Parkinson's disease and cancer: the plot thickens. Transl Neurodegener. 2015;4:20.

34. Park S, Shimizu C, Shimoyama T, Takeda M, Ando M, Kohno T, et al. Gene expression profiling of ATP-binding cassette (ABC) transporters as a predictor of the pathologic response to neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Res Treat. 2006;99(1):9–17.

35. Yamada A, Ishikawa T, Ota I, Kimura M, Shimizu D, Tanabe M, et al. High expression of ATP-binding cassette transporter ABCC11 in breast tumors is associated with aggressive subtypes and low disease-free survival. Breast Cancer Res Treat. 2013;137(3):773–82.

36. Ogretmen B. Sphingolipid metabolism in cancer signalling and therapy. Nat Rev Cancer. 2018;18(1):33–50.

37. Serie DJ, Crook JE, Necela BM, Axenfeld BC, Dockter TJ, Colon-Otero G, et al. Breast cancer clinical trial of chemotherapy and Trastuzumab: potential tool to identify cardiac modifying variants of dilated cardiomyopathy. J Cardiovasc Dev Dis. 2017;4(2):6.

38. Bellmann B, Alushi B, Bigalke B, Landmesser U, Morguet AJ. Restrictive cardiomyopathy: delayed occurrence after radiotherapy of breast cancer. Wien Klin Wochenschr. 2017;129(7–8):278–83.

39. Liu LM, Sun WZ, Fan XZ, Xu YL, Cheng MB, Zhang Y. Methylation of C/EBPalpha by PRMT1 inhibits its tumor-suppressive function in breast cancer. Cancer Res. 2019;79(11):2865–77.

40. Nakai K, Xia W, Liao HW, Saito M, Hung MC, Yamaguchi H. The role of PRMT1 in EGFR methylation and signaling in MDA-MB-468 triple-negative breast cancer cells. Breast Cancer. 2018;25(1):74–80.

41. Zou JX, Duan Z, Wang J, Sokolov A, Xu J, Chen CZ, et al. Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. Mol Cancer Res. 2014;12(4):539–49.

42. Tan MH, De S, Bebek G, Orloff MS, Wesolowski R, Downs-Kelly E, et al. Specific kinesin expression profiles associated with taxane resistance in basal-like breast cancer. Breast Cancer Res Treat. 2012;131(3):849–58.

43. Sayyad MR, Puchalapalli M, Vergara NG, Wangensteen SM, Moore M, Mu L, et al. Syndecan-1 facilitates breast cancer metastasis to the brain. Breast Cancer Res Treat. 2019;178(1):35–49.

44. Ivanov AA, Khuri FR, Fu H. Targeting protein-protein interactions as an anticancer strategy. Trends Pharmacol Sci. 2013;34(7):393–400.

45. Walensky LD. Targeting BAX to drug death directly. Nat Chem Biol. 2019;15(7):657–65.

46. Mabonga L, Kappo AP. Protein-protein interaction modulators: advances, successes and remaining challenges. Biophys Rev. 2019;11(4):559–81.

47. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov. 2017;16(1):19–34.

48. Niphakis MJ, Lum KM, Cognetta AB 3rd, Correia BE, Ichu TA, Olucha J, et al. A global map of lipid-binding proteins and their ligandability in cells. Cell. 2015;161(7):1668–80.

49. Cui L, Li H, Hui W, Chen S, Yang L, Kang Y, et al. A deep learning-based framework for lung cancer survival analysis with biomarker interpretation. BMC Bioinform. 2020;21(1):112.

50. Tang Z, Lei S, Zhang X, Yi Z, Guo B, Chen JY, et al. Gsslasso Cox: a Bayesian hierarchical model for predicting survival and detecting associated genes by incorporating pathway information. BMC Bioinform. 2019;20(1):94.

51. Baldari S, Ubertini V, Garufi A, D'Orazi G, Bossi G. Targeting MKK3 as a novel anticancer strategy: molecular mechanisms and therapeutical implications. Cell Death Dis. 2015;6:e1621.

52. Bossi G. MKK3 as oncotarget. Aging. 2016;8(1):1–2.

53. Boyle DL, Hammaker D, Edgar M, Zaiss MM, Teufel S, David JP, et al. Differential roles of MAPK kinases MKK3 and MKK6 in osteoclastogenesis and bone loss. PLoS ONE. 2014;9(1):e84818.

54. Samulin Erdem J, Skaug V, Haugen A, Zienolddiny S. Loss of MKK3 and MK2 copy numbers in non-small cell lung cancer. J Cancer. 2016;7(5):512–5.

55. Stramucci L, Pranteda A, Bossi G. Insights of crosstalk between p53 protein and the MKK3/MKK6/p38 MAPK signaling pathway in cancer. Cancers. 2018;10(5):131.

56. Gurtner A, Starace G, Norelli G, Piaggio G, Sacchi A, Bossi G. Mutant p53-induced up-regulation of mitogen-activated protein kinase kinase 3 contributes to gain of function. J Biol Chem. 2010;285(19):14160–9.

57. MacNeil AJ, Jiao SC, McEachern LA, Yang YJ, Dennis A, Yu H, et al. MAPK kinase 3 is a tumor suppressor with reduced copy number in breast cancer. Cancer Res. 2014;74(1):162–72.

58. Schieven GL. The biology of p38 kinase: a central role in inflammation. Curr Top Med Chem. 2005;5(10):921–8.

59. Cuadrado A, Nebreda AR. Mechanisms and functions of p38 MAPK signalling. Biochem J. 2010;429(3):403–17.

60. Ivanov AA, Gonzalez-Pecchi V, Khuri L, Niu T, Wang Y, Xu R, et al. OncoPPi-informed discovery of mitogen-activated protein kinase 3 as a novel binding partner of c-Myc. Oncogene. 2017;36(42):5852–60.

61. Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, et al. The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. Nat Commun. 2017;8:14356. https://doi.org/10.1038/ncomms.

62. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006;439(7074):353–7.

63. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–40.

64. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell systems. 2015;1(6):417–25.

65. Bouvard C, Lim SM, Ludka J, Yazdani N, Woods AK, Chatterjee AK, et al. Small molecule selectively suppresses MYC transcription in cancer cells. Proc Natl Acad Sci USA. 2017;114(13):3497–502.

66. Wang Y, Shi J, Chai K, Ying X, Zhou BP. The Role of Snail in EMT and Tumorigenesis. Curr Cancer Drug Targets. 2013;13(9):963–72.

67. Liu Y, Du F, Chen W, Yao M, Lv K, Fu P. EIF5A2 is a novel chemoresistance gene in breast cancer. Breast Cancer. 2015;22(6):602–7.

68. Ma SY, Park JH, Jung H, Ha SM, Kim Y, Park DH, et al. Snail maintains metastatic potential, cancer stem-like properties, and chemoresistance in mesenchymal mouse breast cancer TUBOP2J cells. Oncol Rep. 2017;38(3):1867–76.

69. Holand T, Riffo-Vasquez Y, Spina D, O'Connor B, Woisin F, Sand C, et al. A role for mitogen kinase kinase 3 in pulmonary inflammation validated from a proteomic approach. Pulm Pharmacol Ther. 2014;27(2):156–63.

70. Inoue T, Boyle DL, Corr M, Hammaker D, Davis RJ, Flavell RA, et al. Mitogen-activated protein kinase kinase 3 is a pivotal pathway regulating p38 activation in inflammatory arthritis. Proc Natl Acad Sci USA. 2006;103(14):5484–9.

71. Kang Y, Wang F, Lu Z, Ying H, Zhang H, Ding W, et al. MAPK kinase 3 potentiates Chlamydia HSP60-induced inflammatory response through distinct activation of NF-kappaB. J Immunol. 2013;191(1):386–94.

72. Kim EK, Choi EJ. Pathological roles of MAPK signaling pathways in human diseases. Biochem Biophys Acta. 2010;1802(4):396–405.

73. Dong C, Davis RJ, Flavell RA. MAP kinases in the immune response. Annu Rev Immunol. 2002;20:55–72.

74. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res. 2002;12(1):9–18.

75. Wang L, Chen C, Feng S, Lei P, Tian J. Mitogen-activated protein kinase kinase 3 induces cell cycle arrest via p38 activation mediated Bmi-1 downregulation in hepatocellular carcinoma. Mol Med Rep. 2016;13(1):243–8.

Yang *et al. J Transl Med*   (2020) 18:334

Page 19 of 19

76. Peng R, Cheng X, Zhang Y, Lu X, Hu Z. miR-214 down-regulates MKK3 and suppresses malignant phenotypes of cervical cancer cells. Gene. 2020;724:144146.

77. Zhou M, Yu X, Jing Z, Wu W, Lu C. Overexpression of microRNA21 inhibits the growth and metastasis of melanoma cells by targeting MKK3. Mol Med Rep. 2019;20(2):1797–807.

78. Stramucci L, Pranteda A, Stravato A, Amoreo CA, Pennetti A, Diodoro MG, et al. MKK3 sustains cell proliferation and survival through p38DELTA MAPK activation in colorectal cancer. Cell Death Dis. 2019;10(11):842.

79. Luo S, Ren B, Zou G, Liu J, Chen W, Huang Y, et al. SPAG9/MKK3/p38 axis is a novel therapeutic target for liver cancer. Oncol Rep. 2019;41(4):2329–36.

80. Xie X, Liu K, Liu F, Chen H, Wang X, Zu X, et al. Gossypetin is a novel MKK3 and MKK6 inhibitor that suppresses esophageal cancer growth in vitro and in vivo. Cancer Lett. 2019;442:126–36.

81. Gupta J, del Barco Barrantes I, Igea A, Sakellariou S, Pateras IS, Gorgoulis VG, et al. Dual function of p38alpha MAPK in colon cancer: suppression of colitis-associated tumor initiation but requirement for cancer cell survival. Cancer Cell. 2014;25(4):484–500.

82. Wakeman D, Schneider JE, Liu J, Wandu WS, Erwin CR, Guo J, et al. Deletion of p38-alpha mitogen-activated protein kinase within the intestinal epithelium promotes colon tumorigenesis. Surgery. 2012;152(2):286–93.

83. Xu J, Chen Y, Olopade OI. MYC and breast cancer. Genes Cancer. 2010;1(6):629–40.

84. Siddharth S, Sharma D. Racial disparity and triple-negative breast cancer in African-American Women: a multifaceted affair between obesity, biology, and socioeconomic determinants. Cancers. 2018;10(12):514.

85. Naab TJ, Gautam A, Ricks-Santi L, Esnakula AK, Kanaan YM, DeWitty RL, et al. MYC amplification in subtypes of breast cancers in African American women. BMC cancer. 2018;18(1):274.

86. Khan F, Ricks-Santi LJ, Zafar R, Kanaan Y, Naab T. Expression of p27 and c-Myc by immunohistochemistry in breast ductal cancers in African American women. Ann Diag Pathol. 2018;34:170–4.

87. Santoro A, Vlachou T, Luzi L, Melloni G, Mazzarella L, D'Elia E, et al. p53 loss in breast cancer leads to Myc activation, increased cell plasticity, and expression of a mitotic signature with prognostic value. Cell Rep. 2019;26(3):624–38.

88. Toyoshima M, Howie HL, Imakura M, Walsh RM, Annis JE, Chang AN, et al. Functional genomics identifies therapeutic targets for MYC-driven cancer. Proc Natl Acad Sci USA. 2012;109(24):9545–50.

89. Mo XL, Qi Q, Ivanov AA, Niu Q, Luo Y, Havel J, et al. AKT1, LKB1, and YAP1 revealed as MYC interactors with NanoLuc-based protein-fragment complementation assay. Mol Pharmacol. 2017;91(4):339–47.

90. Heidelberger JB, Voigt A, Borisova ME, Petrosino G, Ruf S, Wagner SA, et al. Proteomic profiling of VCP substrates links VCP to K6-linked ubiquitylation and c-Myc function. EMBO Rep. 2018;19(4):e44754.

Publisher's Note