Bioinformatical analysis identified significant genes in the PI3K-Akt signaling pathway relating with poor prognosis of skin cutaneous melanoma

Yingzi Zhang¹, Yunlan Zhang², Xuyun Hu³*

1 Department of Dermatology, Shunyi Maternal and Children’s Hospital of Beijing Children’s Hospital, Beijing, China, 101300
2 Department of Dermatology, Beijing Tongren Hospital, Capital Medical University, Beijing, China, 100730
3 Beijing Key Laboratory for Genetics of Birth Defects, Beijing Pediatric Research Institute; MOE Key Laboratory of Major Diseases in Children; Genetics and Birth Defects Control Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China, 100045

*Corresponding authors:
Xuyun Hu
hu_xuyun@126.com
Beijing Children's Hospital
56 Nanlishi Road, Beijing, China
Tel: 86-10-59616894

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Skin cutaneous melanoma (SKCM) is the most serious form of skin cancer and the reliable genetic prognostic biomarkers are not clear. The purpose of the present academic work was to identify reliable prognostic biomarkers and understand the underlying mechanism. Gene expression profiles of GSE31909, GSE104849 and GSE111766 were available from Gene Expression Omnibus (GEO). DEGs were collected from 13 melanoma samples and 12 melanocyte samples which were in the three profile datasets by GEO2R analysis. Gene ontology (GO) analysis and KEGG analysis were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID). We further established protein-protein interaction (PPI) and identify core genes with Search Tool for the Retrieval of Interacting Genes (STRING) and Molecular Complex Detection (MCODE) app in Cytoscape software. There were total of 63 consistently expressed genes in the three datasets, including eight up-regulated genes enriched in biological processes like osteoblast differentiation and extracellular matrix organization, and cell component like extracellular space, while 55 down-regulated genes enriched in positive regulation of gene expression, intracellular signal transduction, cell adhesion and apoptotic process. With prognostic information and expression data from UALCAN and Gene Expression Profiling Interactive Analysis (GEPIA), we identified three significant genes (TNC, SPP1 and KIT). After Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, all three genes were located in the PI3K-Akt Signaling Pathway. Our results provide additional genetic biomarkers for SKCM patients.

Keywords: Bioinformatical analysis, Microarray, Skin cutaneous melanoma, Differentially expressed gene
**Introduction**

Skin cutaneous melanoma (SKCM) is the most serious form of skin cancer. It is the fifth most common cancer and is increasing faster than any other potentially preventable cancer in the United States (1, 2). Five-year survival rates for people with melanoma depend on the stage of the disease (from stage 0 to stage IV) at the time of diagnosis. Most people with stage 0 or I melanoma can expect prolonged disease-free survival after treatment, whereas patients with more advanced stage II to IV melanoma are more likely to develop metastatic disease (2, 3). Therefore, identifying reliable prognostic biomarkers will help to predict the treatment effect and understand the underlying mechanism.

Gene expression profiles varied in different pathophysiological process. With the help of expression beadchips, we can identify differentially expressed genes (DEGs) in genome wide easily. Recent studies on SKCM produced public bioinformatical data which could be used to further identify prognostic biomarker genes (4-7). In this study, we chose microarrays of GSE31909, GSE104849 and GSE111766 from Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/) (8). After GEO2R analysis, 63 common DEGs from the three datasets were collected for further study. We further established protein-protein interaction (PPI) and identified core genes with Search Tool for the Retrieval of Interacting Genes (STRING, string-db.org) and MCODE (Molecular Complex Detection) app in Cytoscape software. With prognostic information and expression data from UALCANC and Gene Expression Profiling Interactive Analysis (GEPIA) (9, 10), we identified three significant genes (*TNC*, *SPP1* and *KIT*). After Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, all three genes were located in the PI3K-Akt Signaling Pathway. Our results provide additional genetic biomarkers for SKCM patients.

**Material and methods**

**Gene expression data**

Gene expression profiles data was obtained from GEO. Microarrays of GSE31909, GSE104849 and GSE111766, which contained 13 melanoma samples and 12 melanocyte samples, were used to identify DEGs. All microarrays were performed on Platform of
GPL10558 (Illumina HumanHT-12 V4.0 expression beadchip). After GEO2R analyzing, DEGs with $|\text{logFC}|>2$ and $P$ value $<0.05$ between melanoma samples and normal melanocyte samples were identified. DEGs with $\text{log FC}<0$ was considered as down-regulated genes, and DEGs with $\text{log FC}>0$ was considered as up-regulated genes. Venn software was used to detect the intersection of DEGs in the three datasets.

**DEGs function and interaction analysis**
Gene ontology (GO) analysis and KEGG analysis were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID, david.ncifcrf.gov/home.jsp), which is designed to identify genes or proteins function to understand biological meaning behind large list of genes (11, 12). STRING was used to evaluate protein-protein interaction among DEGs. MCODE app in Cytoscape software was used to reanalyze the modules of the PPI network to identify core genes.

**Survival rate analysis**
UALCANCAN, an interactive web resource containing The Cancer Genome Atlas (TCGA) and MET500 cancer OMICS data, was used to analyze patient survival information for core genes. After survival analysis, genes with $P<0.05$ were reevaluated with GEPIA to analyze the tumor/normal differential expression. Core genes with significant differential survival rate and expression were reanalyzed with DAVID to identify the key pathways related with SKCM prognosis.

**Results**

**Identification of DEGs in skin cutaneous melanoma**
After GEO2R analysis, we extracted 248, 401 and 246 DEGs from GSE31909, GSE104849 and GSE111766, respectively. Then, we used Venn diagram software to identify the commonly DEGs in the three datasets. Results showed that a total of 63 commonly DEGs were detected, including eight up-regulated genes ($\text{logFC}>0$) and 55 down-regulated genes ($\text{logFC}<0$) in 13 melanoma samples and 12 melanocyte samples in our present study. (Table 1
Fig. 1 Venn diagrams of DEGs from three datasets. Eight up-regulated genes and 55 down-regulated genes were presented in all three datasets.

Table 1 64 differentially expressed genes presented in all three datasets

<table>
<thead>
<tr>
<th>Expression</th>
<th>Genes</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>PRAME, TM4SF1, ADAM19, TMEM158, TNC, SPP1, CXCL8, VGF</td>
<td>9</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>LGI3, CDH3, MCF2L, IL17D, CAPN3, FCGRT, PIR, MLANA, ASPA, 55</td>
<td>55</td>
</tr>
</tbody>
</table>

GO analysis and KEGG pathway analysis of DEGs

GO analysis and KEGG pathway analysis were performed for eight up-regulated genes and 55 down-regulated genes by DAVID software. For up-regulated genes, GO analysis results
<table>
<thead>
<tr>
<th>Expression</th>
<th>Categ</th>
<th>Term</th>
<th>Genes</th>
<th>Count</th>
<th>%</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>BP</td>
<td>osteoblast differentiation</td>
<td>SPP1, TNC</td>
<td>2</td>
<td>25</td>
<td>3.10E-02</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>extracellular matrix organization</td>
<td>SPP1, TNC</td>
<td>2</td>
<td>25</td>
<td>5.70E-02</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>extracellular space</td>
<td>SPP1, TNC, CXCL8, VGF</td>
<td>4</td>
<td>50</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>MF</td>
<td>NONE</td>
<td>KIT, CDH3, GAS6, OLFM1, P2RX7</td>
<td>5</td>
<td>9.1</td>
<td>8.10E-03</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>positive regulation of gene expression</td>
<td>KIT, MCF2L, PAG1, PRKCB, SYK</td>
<td>5</td>
<td>9.1</td>
<td>3.40E-02</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>intracellular signal transduction</td>
<td>KIT, MCF2L, PAG1, PRKCB, SYK</td>
<td>5</td>
<td>9.1</td>
<td>5.00E-02</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>cell adhesion</td>
<td>CD36, CDH3, FEZ1, GAS6, IGFBP7</td>
<td>5</td>
<td>9.1</td>
<td>3.47E-02</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>apoptotic process</td>
<td>CAPN3, GAS6, MAL, PRKCB, PRUNE2</td>
<td>5</td>
<td>9.1</td>
<td>9.30E-02</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>cytoplasm</td>
<td>OCA2, CDH1, and other 23 genes</td>
<td>2</td>
<td>45</td>
<td>7.30E-02</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>plasma membrane</td>
<td>KIT, CDH1, and other 16 genes</td>
<td>1</td>
<td>32</td>
<td>7.30E-02</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>cytosol</td>
<td>CABLES1, DNAJA4, and other 13 genes</td>
<td>1</td>
<td>27</td>
<td>8.80E-02</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>extracellular space</td>
<td>KIT, IGFBP7, and other 9 genes</td>
<td>1</td>
<td>20</td>
<td>4.90E-03</td>
</tr>
<tr>
<td></td>
<td>MF</td>
<td>receptor binding</td>
<td>A2M, GAS6, NSG1, P2RX7</td>
<td>4</td>
<td>7.3</td>
<td>8.30E-02</td>
</tr>
<tr>
<td></td>
<td>MF</td>
<td>cysteine-type peptidase activity</td>
<td>CAPN3, CTSF, CTSK</td>
<td>3</td>
<td>5.5</td>
<td>5.40E-03</td>
</tr>
<tr>
<td></td>
<td>MF</td>
<td>protease binding</td>
<td>KIT, A2M, MBP</td>
<td>3</td>
<td>5.5</td>
<td>3.50E-02</td>
</tr>
<tr>
<td></td>
<td>MF</td>
<td>lipid binding</td>
<td>CD36, APOLD1, MAL</td>
<td>3</td>
<td>5.5</td>
<td>7.10E-02</td>
</tr>
</tbody>
</table>

BP: biological processes; CC: cell component; MF: molecular function

revealed that these genes were enriched in biological processes like osteoblast differentiation and extracellular matrix organization, and cell component like extracellular space. For down-regulated genes, in the category of biological processes, genes were major enriched in positive regulation of gene expression, intracellular signal transduction, cell adhesion and
apoptotic process. In the category of cell component, genes were major enriched in cytoplasm, plasma membrane, cytosol and extracellular space. In the category of molecular function, genes were major enriched in receptor binding, cysteine-type peptidase activity, protease binding and lipid binding (Table 2).

KEGG analysis results demonstrated that up-regulated genes were particularly enriched in ECM-receptor interaction, toll-like receptor signaling pathway, focal adhesion and PI3K-Akt signaling pathway. Down-regulated genes were enriched in melanogenesis and histidine metabolism (Table 3).

![Table 3 KEGG pathway analysis of differentially expressed genes](image)

<table>
<thead>
<tr>
<th>Expression</th>
<th>Term</th>
<th>Genes</th>
<th>Count</th>
<th>%</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>ECM-receptor interaction</td>
<td>SPP1 TNC</td>
<td>2</td>
<td>25</td>
<td>2.50E-02</td>
</tr>
<tr>
<td></td>
<td>Toll-like receptor signaling pathway</td>
<td>CXCL8 SPP1</td>
<td>2</td>
<td>25</td>
<td>3.10E-02</td>
</tr>
<tr>
<td></td>
<td>Focal adhesion</td>
<td>SPP1 TNC</td>
<td>2</td>
<td>25</td>
<td>5.90E-02</td>
</tr>
<tr>
<td></td>
<td>PI3K-Akt signaling pathway</td>
<td>SPP1 TNC</td>
<td>2</td>
<td>25</td>
<td>9.80E-02</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>Melanogenesis</td>
<td>TYRP1</td>
<td>3</td>
<td>5.5</td>
<td>3.40E-02</td>
</tr>
<tr>
<td></td>
<td>Histidine metabolism</td>
<td>ASPA HNMT</td>
<td>2</td>
<td>3.6</td>
<td>6.20E-02</td>
</tr>
</tbody>
</table>

Protein–protein interaction network (PPI) analysis

After STRING analysis of all 63 DEGs in common from three datasets, 30 of them were imported into the PPI network complex which included six up-regulated genes and 24 down-regulated genes. Cytoscape MCODE revealed two overlapped clusters (Fig. 2). 20 core genes were located in the two clusters including four up-regulated genes and 16 down-regulated genes (Fig. 2).
Fig. 2 Cluster analysis of 30 DEGs by Cytoscape MCODE. Genes in yellow were located in two overlapping clusters. Genes in green were up-regulated genes. Genes in red were down-regulated genes.

**Survival rate and expression analysis**

Patient survival information for core genes was analyzed by UALCAN. High expression in six genes and low expression in three genes had a significantly worse survival rate (P < 0.05, Fig. 3). GEPIA was used to reanalyze the nine gene expression level between cancerous and normal people. Results showed that three genes (TNC, SPP1 and KIT) had significant expression change in melanoma samples and melanocyte samples and consisted with microarrays data (P < 0.05, Fig. 4).
Fig. 3 The prognostic information of the nine core genes with a significantly worse survival rate. Green line: low expression; red line: high expression (P < 0.05)

Fig. 4 Expression level of TNC, SPP1 and KIT between cancerous and normal people. Left box means tumor tissues (n=461) and right box means normal tissues (n=558) (P < 0.05)
Reanalysis of three key genes via KEGG pathway enrichment

To understand the possible pathway of the three key DEGs, KEGG pathway enrichment was reanalyzed by DAVID. Results showed that all three genes markedly enriched in the PI3K-Akt signaling pathway ($P = 2.5E-3$, Fig. 5), while $TNC$ and $SPP1$ were also enriched in ECM-receptor interaction ($P=2.5E-2$) and Focal adhesion ($P=5.9E-2$).

![Fig. 5 KEGG pathway reanalysis of three key genes revealed enrichment in PI3K-Akt signaling pathway.](image)

Discussion

SKCM is one of the most fatal skin cancers due to its high proliferative and metastatic potential. Several biomarkers for the TNM (Topography, Lymph Node and Metastasis) classification have been identified, but these staging markers often fail to predict the patient’s prognosis accurately (13). In our study, we uncovered 63 common DEGs from microarrays of GSE31909, GSE104849 and GSE111766. Protein-protein interaction analysis identified 20 core genes in the two clusters. With prognostic information and expression information in public databases, we identified three significant genes $TNC$, $SPP1$ and $KIT$. These three genes are located in the PI3K-Akt signaling pathway, which is one of the most important signaling...
networks in cancer and the activation of this pathway plays a significant role in melanoma (14). PI3K-AKT pathway also plays important roles in melanoma therapeutic resistance. Therefore, our results provide further genetic markers for SKCM patients’ prognosis.

*TNC* encodes tenascin, an extracellular matrix protein with a spatially and temporally restricted tissue distribution. This protein contains multiple EGF-like and fibronectin type-III domains. The EGF-like domains is considered anti-adhesive and the fibronectin type-III domains are considered adhesive (15-17). Cell interaction with tenascin can promote migration in a mesenchymal state or an amoeboid state to penetrate denser matrix (18-20). Recent evidence also showed The EGF-like domains of TNC acted as an ultralow affinity ligand for the EGF receptor (21, 22) and the restriction of EGFR signaling increased cell survival protect stem cells from apoptosis (23, 24). Therefore, the function of highly expressed *TNC* in melanoma includes both promoting migration and invasion through the dermis, and colonization the distant target organ with supporting melanoma cells survival. The *TNC* gene expression is strictly regulated and it is lowly expressed in adult skin (25). *TNC* expression is dramatically elevated level in patients with higher stages melanoma than lower stages tumor (26). Thus, *TNC* was considered as a potential biomarker of melanoma aggressiveness and treatment effect. However, the mechanism behind the up-regulation of *TNC* remains to be further determined.

*SPPI* encodes secreted phosphoprotein 1 (also unknown as osteopontin, OPN). This protein is a matricellular protein that is produced by multiple tissues and mostly abundant in bone. It is involved in bone mineralization by attachment of osteoclasts to the mineralized bone matrix (27). *SPPI* protein is produced by cancer cells and plays crucial roles in cancer progression by facilitating adhesion, migration and survival of tumor cells (28-30). Several studies indicate that *SPPI* protein is overexpressed in the majority of human cancers and the up-regulated expression is related with associated with poor clinical outcome in a variety of cancers including breast cancer, lung cancer, liver cancers, prostate cancer, etc. (31-33). Similar to *TNC*, *SPPI* overexpression was significantly associated with tumor thickness, metastasis, and reduced relapse-free survival in melanoma patients (34, 35). Therefore, *SPPI* is a promising biomarker for detecting melanoma aggressiveness and a factor for the survival prediction in patients.
KIT encodes Mast/Stem Cell Growth Factor Receptor. It is the human homolog of the proto-oncogene c-kit. KIT protein is a Receptor Tyrosine Kinase containing an extracellular ligand-binding domain, a transmembrane domain, an juxta-membrane domain and two tyrosine kinase domains (36). KIT loss-of-function mutations result in piebaldism, which is characterized by the congenital absence of melanocytes in affected areas of the skin and hair (37, 38). Somatic KIT mutations that result in KIT receptor tyrosine kinase constitutive activity, were reported to cause a variety of cancers like gastrointestinal stromal tumor, acute myelogenous leukemia, testicular germ cell tumor (39, 40), as well as melanoma (mostly in mucosal and acral melanomas) (41-43). However, activating mutations in the KIT gene do not necessarily correlate with c-KIT expression (44, 45). KIT has been considered a melanoma suppressor and the expression appears to be downregulated or lost in invasive and metastatic melanomas (46, 47). Loss of KIT expression were also seen in other cancer (48, 49) and may be a required step in tumor progression due to its important roles in cell differentiation and tissue morphogenesis. Although KIT expression is down-regulated in most subsets of melanomas (including SKCM), one study showed human posterior uveal melanoma expresses KIT at high levels frequently (50). This result indicated that the KIT expression in melanomagenesis is more complex and need further investigation.

We identified three potential predictors (TNC, SPP1 and KIT) for survival rates of patients with SKCM based on three different microarray datasets. These three genes are located in the PI3K-Akt signaling pathway and were reported to have important roles in the progression of SKCM. Further researches are needed to verify these prognosis bio-markers and improve the clinical outcomes of patients with this highly aggressive cancer.

Acknowledgements

This study was partially supported by Scientific Research Cultivation Fund of Capital Medical University (PYZ19034).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the GEO repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31909),
(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104849) and

Ethics approval and consent to participate
Not applicable.

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

References


41. Gutierrez-Castaneda LD, Nova JA and Tovar-Parra JD: Frequency of mutations in BRAF,


