Identification of Potential Core Genes Between Primary and Metastatic Malignant Melanoma and Analysis of Their Immune Correlation

Cong-Li Jia
Fu Yang
Rui-Ning Li

1Weifang Medical College, Weifang, Shandong, People’s Republic of China; 2Department of Hepatobiliary Surgery, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, People’s Republic of China; 3Second Clinical Medical College, Lanzhou University, Lanzhou, Gansu, People’s Republic of China

Purpose: To identify the potential differential genes between primary and metastatic melanoma, screen out immune-related genes in core genes and analyze their immune correlation, thus searching for the early diagnostic biomarkers of cutaneous malignant melanoma (CMM) and the targets of curbing metastasis.

Materials and Methods: We analyzed two microarray datasets (GSE8401 and GSE46517) derived from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) between primary and metastatic melanoma were screened out using the GEO2R tool. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed to identify the functions and pathways of DEGs. We analyzed protein–protein interaction of these DEGs based on the Search Tool for the Retrieval of Interacting Genes database and showed by Cytoscape software. In addition, the online Gene Expression Profiling Interactive Analysis tool (GEPIA) was used to analyze the prognostic value of hub genes expressed in metastatic melanoma patients. Immune-related genes in hub genes were screened and further analyzed.

Results: A total of 178 upregulated DEGs and 4 downregulated DEGs were identified. 23 terms and 4 pathways were confirmed related to metastatic melanoma. Ten hub genes with a high degree of connectivity were found. Overexpression of three hub genes (DSG1, FLG, PKP1) (P<0.01) was associated with metastasis and poor prognosis of CMM. Among them, the patients with overexpression of PKP1 suffered shorter survival. In addition, 2 immune-related genes (EGFR and CDH1) in hub genes were screened out and both of them were related to anti-tumor immunity, although their expression level did not affect the overall survival of CMM patients significantly.

Conclusion: Our study suggests that DSG1, FLG and PKP1 were overexpressed in metastatic melanoma compared with primary melanoma, and overexpression of these three genes was an unfavorable prognostic factor for CMM patients, which may indicate that they are associated with promoting metastasis of malignant melanoma. EGFR and CDH1 play a crucial role in anti-tumor immunity for CMM. Further research is needed to explore the value of these genes in the inhibition of metastasis and treatment of CMM.

Keywords: cutaneous malignant melanoma, metastasis, genes, PKP1, DSG1, FLG, CDH1, EGFR

Introduction
Cutaneous malignant melanoma is a highly malignant tumor originating from skin melanocytes, which is usually ignored as melanin nevus in the early stage.1 Once metastasis occurs, the cure rate is low. The primary lesion is often seen in the lower
extremities, and a few can occur in tissues and organs, such as digestive tract and eyes. CMM with insidious onset is prone to metastasis and diffusion. Currently, surgery combined with radiotherapy, chemotherapy and high-dose interferon are the main treatment methods for CMM, but the prognosis is unfavorable. Despite the rise of immunotherapy targeting certain mutated genes in recent years, the survival rate is still low.\(^2\) Metastatic potential for CMM is very high, and the prognosis of patients with metastatic tumor is very poor. Molecular targeted therapy is the main strategy for the treatment of metastatic melanoma, however, lacking targets to inhibit tumor metastasis results that only a small number of patients can benefit.\(^3\) Recent studies have focused on finding diagnostic biomarkers and therapeutic targets for malignant melanoma, but their focus is mostly on identifying gene expression differences between normal tissues and tumors, while few studies have explored genes differentially expressed between primary cancer and metastatic cancer. However, early melanoma treatment is often delayed due to misdiagnosis, and intervention before metastasis may be a promising strategy for improving survival. Due to this, it is urgent to find novel biomarkers in the genetic differences between primary and metastatic melanoma to guide early diagnosis and treatment.

In this study, we sought to explore new indicators of metastasis in CMM and to provide new therapeutic targets for this disease. To detect the DEGs between metastatic and primary tumor, bioinformatics methods were used to analyze the gene expression profiling data downloaded from the GEO database. We performed GO function and KEGG pathway enrichment to analyze DEGs. In addition, a protein–protein interaction (PPI) network was established to confirm hub genes associated with metastatic melanoma. The survival analysis of these hub genes was performed to screen out genes associated with poor prognosis as biomarkers for further research. More importantly, we screened out immune-related genes for anti-tumor immunity analysis.

**Materials and Methods**

**Data Acquisition**

The gene expression datasets analyzed in our study were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). A total of 1154 series for human cutaneous malignant melanoma were retrieved from the database. After careful screening, two gene expression profiles (GSE8401 and GSE46517) were picked out. Both gene expression profiles GSE8401 and GSE46517 were based on platform GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array). All the data were freely available online, and the study did not refer to any experiments on humans or animals by any of the authors.

**Differential Genes Screening**

We used GEO2R online analysis tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) to screen the DEGs between metastatic malignant melanoma and primary tumor samples. Genes that matched the cutoff criteria, adjusted \(P<0.05\) and \(|\log FC|\geq 1.5\), were viewed as DEGs. For each dataset, a statistical analysis was performed, and the Venn diagram online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to identify the intersected portion.

**GO Functions and KEGG Pathways**

**Analysis of DEGs**

GO analysis is mainly used to study functional enrichment of genes. Gene functions include biological process (BP), molecular function (MF), and cellular component (CC). KEGG, as a database, can be used to analyze the pathway enrichment of genes. In this study, the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tools (https://david.ncifcrf.gov/) were designed to carry out the GO function analysis and KEGG pathway enrichment analysis of DEGs. \(P<0.01\) and gene counts\(\geq10\) were considered statistically significant.

**PPI Network Establishment and Hub Gene Confirming**

PPI is usually searched by The Search Tool for the Retrieval of Interacting Genes (STRING) database (http://string-db.org/), which can construct PPI network. Mapping the DEGs selected previously to the STRING database to evaluate the latent PPI correlation. The PPI pairs were picked out with a combined score\(>0.4\). Afterwards, Cytoscape software (www.cytoscape.org/) was used to plot the PPI network. The higher the connection degree of the nodes are, the more important the nodes are to maintain the stability of the protein network. We use CytoHubba, a plugin in cytoscape to calculate the degree of each protein node. In this study, ten genes with the highest connectivity were confirmed as hub genes.
Survival Analysis of Hub Genes
An online tool, Gene Expression Profiling Interactive Analysis (http://gepia.cancer-pku.cn/) is an interactive web server for analyzing the RNA sequencing expression data of tumors and normal samples, using a standard processing pipeline. The prognostic value of hub gene in metastatic melanoma was evaluated using this tool. For each gene, patients were divided into two groups according to the median values of mRNA expression. P<0.01 was considered statistically significant.

Immune-Related Gene Analysis
The immune-related gene sets were then downloaded via the website (http://www.immport.org/home). We used an online tool (http://bioinformatics.psb.ugent.be/webtools/Venn) to select the intersection of hub genes and immune-related gene sets, immune-related genes in hub genes were screened out and their immune correlation for CMM was further analyzed by TIMER online tool (http://timer.comp-genomics.org/).

Results
Identification of DEGs
Two expression profiles (GSE8401 and GSE46517) were selected in this study. Among them, GSE8401 contained 52 metastatic melanoma samples and 31 primary melanoma samples, and GSE46517 included 73 metastatic melanoma samples and 31 primary melanoma samples (Table 1). Based on the standard of P<0.05 and |logFC|≥1.5, in gene chip GSE8401, 324 DEGs were selected, containing 281 upregulated genes and 43 downregulated genes, while a total of 206 DEGs were picked out from the gene chip GSE46517, including 200 upregulated genes and 6 downregulated genes. All DEGs were identified by comparing metastatic and primary melanoma samples. The intersection of DEG profiles was showed by Venn Analysis (Figure 1). Subsequently, we identified 182 significant DEGs. Including 178 upregulated genes and 4 downregulated genes.

Functional Enrichment Analyses of DEGs
We obtained the result of GO function and KEGG pathway enrichment for DEGs using the DAVID. The results of GO analysis showed that the enriched GO terms were classified into three types containing CC, BP, and MF, DEGs were mainly enriched in BPs, containing epidermis development, keratinocyte differentiation, keratinization, peptide cross-linking, single organismal cell–cell adhesion, cell adhesion, negative regulation of endopeptidase activity, extracellular matrix organization and cytoskeleton organization. CC analysis indicated that the DEGs were enriched in extracellular exosome, desmosome, cornified envelope, intermediate filament, extracellular space, cell–cell adherens junction, extracellular region and extracellular matrix. For the molecular function, the DEGs were enriched in structural molecule activity, structural constituent of cytoskeleton, calcium ion binding, serine-type peptidase activity, serine-type endopeptidase inhibitor activity and serine-type endopeptidase activity (Figure 2). Besides, KEGG pathway analysis figured out that DEGs were principally enriched in the pathway of amoebiasis, ECM–receptor interaction, protein digestion and absorption, and arrhythmogenic right ventricular cardiomyopathy (ARVC) (Figure 3).

Table 1 Statistics of Two Gene Chips from GEO Database

<table>
<thead>
<tr>
<th>Dataset ID</th>
<th>Primary Melanoma</th>
<th>Metastatic Melanoma</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE8401</td>
<td>31</td>
<td>52</td>
<td>83</td>
</tr>
<tr>
<td>GSE46517</td>
<td>31</td>
<td>73</td>
<td>104</td>
</tr>
</tbody>
</table>

Abbreviation: GEO, Gene Expression Omnibus.
PPI Network Establishment and Hub Gene Identification

The protein interactions among the DEGs indicated that 182 nodes and 773 edges were detected in the PPI network (Figure 4). The average node degree is 8.49. We screened the top ten genes as hub genes according to connectivity degree in the PPI network (Table 2) and showed their interactions (Figure 5). The results suggested that cadherin 1, type 1, E-cadherin (epithelial) (CDH1; degree=43) was the most significant gene with connectivity, subsequent ones were keratin 5 (KRT5; degree=40), keratin 14 (KRT14; degree=39), epidermal growth factor receptor (EGFR, degree=38), filaggrin (FLG; degree=36), loricrin (LOR; degree=34), desmoglein 1 (DSG1; degree=33), cornodesmosin (CDSN; degree=32), desmoplakin (DSP; degree=32), and plakophilin 1 (PKP1; degree=30). All of these hub genes were upregulated in metastatic melanoma.

Survival Analysis of Core Genes

The prognostic values of the ten potential hub genes were investigated using Gene Expression Profiling Interactive Analysis (GEPIA) online tool. The overall survival of 8 potential hub genes was obtained except LOR and CDSN, due to the deficiency of the sample size at the custom thresholds for these two genes. The results showed that overexpression of three hub genes known as DSG1 (p=0.0089; HR (high)=1.5; n=403), FLG (p=0.0046; HR (high)=1.5; n=452) and PKP1 (p=0.00031; HR (high)=1.6; n=455) were significantly related to poor overall survival of malignant melanoma patients (P<0.01); among these, overexpression of PKP1 (p=0.00031; HR (high)=1.6; n=455) was more unfavorable prognostic factor for overall survival in CMM patients (Figure 6). There was no statistically significant difference in disease-free survival among the ten hub genes (P>0.01).
Immune-Related Gene Analysis

We intersected the hub genes with the immune-related genes and screened out two immune-related genes (EGFR and CDH1) (Figure 7). The relationship between EGFR or CDH1 and immune cells infiltration in metastatic melanoma was showed in Figures 8 and 9, respectively. Rho<0 represented negative correlation, while Rho>0 mean positive correlation, P<0.05 indicated the difference was statistically significant. For EGFR, CD8+T cell (Rho=−0.184) and NK cell (Rho=−0.172) were negative correlation with its expression level, meanwhile, the expression of EGFR positively affected the infiltration of Tregs (Rho=0.192). On the other hand, CDH1 had a positive correlation with CD8+T cell (Rho=0.164) and NK cell (Rho=0.174). However, the expression of CDH1 plays a negative role on Tregs (Rho=−0.152). We also analyzed the effect of several immune cells...

Table 2 Top Ten Hub Genes with the Highest Degree of Connectivity

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Description</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH1</td>
<td>Cadherin 1</td>
<td>43</td>
</tr>
<tr>
<td>KRT5</td>
<td>Keratin 5</td>
<td>40</td>
</tr>
<tr>
<td>KRT14</td>
<td>Keratin 14</td>
<td>39</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
<td>38</td>
</tr>
<tr>
<td>FLG</td>
<td>Filaggrin</td>
<td>36</td>
</tr>
<tr>
<td>LOR</td>
<td>Loricrin</td>
<td>34</td>
</tr>
<tr>
<td>DSG1</td>
<td>Desmoglein 1</td>
<td>33</td>
</tr>
<tr>
<td>CDSN</td>
<td>Corneodesmosin</td>
<td>32</td>
</tr>
<tr>
<td>DSP</td>
<td>Desmoplakin</td>
<td>32</td>
</tr>
<tr>
<td>PKP1</td>
<td>Plakophilin 1</td>
<td>30</td>
</tr>
</tbody>
</table>

Figure 3 KEGG analysis for differentially expressed genes.

Note: Barplot graph for KEGG (The longer bar means the more genes enriched).

Abbreviation: KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 4 Protein–protein interaction network established with the DEGs.

Note: Red nodes represent upregulated genes, and blue nodes represent downregulated genes.

Abbreviation: DEG, differentially expressed gene.
on the survival of CMM, which showed that CD8+T cell, NK cell and Tregs were associated with good prognosis (Figure 10).

Discussion

Cutaneous malignant melanoma is a highly malignant tumor with insidious onset, strong invasion and high mortality deriving from melanocytes. Primary melanoma is easily misdiagnosed as melanin nevus and ignored. Once metastasis occurs, the 5-year survival rate is low, and the prognosis is unfavorable. Due to the lack of effective targets and biomarkers that predict metastasis, it is difficult for CMM patients to benefit from inhibition of metastasis. Meanwhile, immunotherapy, radiotherapy and chemotherapy remain the main adjuvant therapies. Nevertheless, radiation and chemotherapy are not very effective. Hence, it is urgent to confirm novel target.

In this study, to identify the underlying core genes involved in metastatic melanoma, we performed gene expression and protein interaction analyses based on public databases. DEGs between metastatic and primary melanoma were identified according to gene expression profiling data from the GEO database. The DEGs picked out were mainly correlated with the GO BP terms, such as keratinocyte differentiation, keratinization, peptide crosslinking, cell adhesion and extracellular matrix. Two of the detected KEGG pathways, known as ECM-receptor interaction and protein digestion, were associated with cancer progression. Ten core genes were screened out from the PPI network established based on DEGs, containing KRT5, FLG, KRT14, LOR, CDH1, DSP, DSG1, EGFR, PKP1, CDSN. All of these genes were upregulated in metastatic melanoma. However, among them, only overexpression of three genes including DSG1, FLG and PKP1 had a statistically significant effect on overall survival, which indicated that these three genes may play a crucial role in the metastasis and prognosis of CMM. PKP1 is considered to be an important component of desmosomes, and desmosomes are active cellular junctions that participate in cellular cohesion and are essential for maintaining tissue integrity. Thus, PKP1 mutations lead to fewer and poorly formed desmosomes and loss of epidermal integrity. Disruption of the integrity and stability of the epidermal barrier is a hallmark of cancer, which suggests that PKP1 is a cancer suppressor. Nevertheless, PKP1 overexpression has been demonstrated in several cancers, such as oral/pharyngeal squamous cell cancer and squamous cell lung cancer, and its expression level was negatively correlated with the degree of malignancy in oral/pharyngeal squamous cell cancer. In our study, PKP1 overexpressed in metastatic melanoma compared with primary melanoma. Furthermore, high expression of PKP1 was associated with shortened survival. We speculated that this might due to the fact that PKP1 exists not only in the desmosomes but also in the nucleus and cytoplasm, and it performs different functions in different cellular components. Another possibility is that PKP1 is modified during carcinogenesis and the tumor may progress due to uncontrolled adhesion and migration.

In addition to PKP1, we also detected 9 potential core genes for metastatic melanoma, but only DSG1 and FLG were the main genes that had a significant impact on the survival of CMM patients, suggesting that DSG1 and FLG may play a role in the progression and metastasis of CMM. FLG encodes proteins that are also part of the desmosome and are involved in cell adhesion and regulation of skin barrier function. A study confirmed that disruption of cell adhesion might lead to uncontrolled cell proliferation and played a key role in epithelial-to-mesenchymal transition (EMT). EMT is a critical developmental procedure in which abnormal activation enables tumors to gain the ability to invade and metastasize far away. This process may also occur in melanoma metastasis. FLG is located on chromosome 1q21.3, which is the
reported CMM susceptibility site. In our study, FLG expression was significantly higher in metastatic melanoma compared with primary melanoma. Kezie S. has reported that FLG may serve as a potential biomarker for reducing melanoma risk. Some studies have confirmed that DSG1 expression is decreased in lung cancer, head and neck cancer and gastric cancer, and its low expression is associated with longer survival or tumor inhibition. However, some people think DSG1 is carcinogenic. Li et al found that DSG1 expression decreased when melanoma cells proliferated, but in our study, DSG1 overexpressed in metastatic...
melanoma. These genes play different or even opposite roles in the occurrence and progression of different tumors, which may be related to the different locations of genes or stage of tumor. For instance, the expression of genes in desmosomes inhibits tumor invasion, while the expression in cytoplasm or nucleus may promote tumor metastasis. Gene expression may increase in the middle stage of the tumor to initiate metastasis and decrease in the later stage. However, the role and mechanism of these genes in CMM are still unclear.

We identified two immune-related core genes for CMM, namely EGFR and CDH1. On the one hand, Epidermal Growth Factor Receptor (EGFR) is one of the four members of the HER family receptors, which comprise EGFR/HER1/erbB1, HER2/erbB2, HER3/...

Figure 6 Overall survival analyses for the top ten hub genes expressed in metastatic melanoma by Gene Expression Profiling Interactive Analysis tool.
Notes: Gene description can be seen in Table 2 the overall survival of CDSN and LOR cannot be obtained due to the deficiency of the sample size at the custom thresholds for these two genes.
Figure 7 Intersection graph of hub genes and immune-related genes.

ErbB3, and HER4/erbB4.\textsuperscript{17,18} Previous research reported that 10–30% of all lung adenocarcinomas\textsuperscript{19,20} included an EGFR-activating mutation. EGFR mutations occur mostly in adenocarcinoma, younger women and girls,\textsuperscript{17} and never-smokers.\textsuperscript{17} ErbB1 (also known as EGFR) receptor signaling is involved in cell proliferation, migration, differentiation and tissue morphogenesis.\textsuperscript{21} ErbB signaling is often found to be overactive in epithelial malignancies, accounting for 80–90% of all cancer cases,\textsuperscript{22} indicating that this pathway is a very promising target in oncology. Tumor antigens are expressed specifically or highly in tumor tissues. In addition, tumor antigens can result from mutated gene products, which are not produced by non-cancer cells. Mutated epidermal growth factor receptor known as EGFRvIII is a tumor-specific product that is expressed by 30% Glioblastoma multiforme (GBM) cases.\textsuperscript{23} EGFRvIII has been extensively studied as a tumor antigen. It has been reported that the frequency of oncogenic mutations in the EGFR gene is closely associated with the occurrence of melanoma.\textsuperscript{24,25} Our study found that high expression of EGFR promoted the infiltration of Tregs in CMM but inhibited the expression of CD8+T cells and NK cells. Meanwhile, high expression of CD8+T cell, NK cell, and Tregs was positively associated with longer survival and anti-tumor immunity. The results above indicated EGFR as a pathogenic gene of CMM, which is consistent with previous reports. On the other hand, our study suggested that high expression of CDH1 resulted in the increased infiltration of CD8+T cell and NK cell, while decreased infiltration of Tregs. The CDH1 gene encodes a transmembrane glycoprotein E-cadherin that maintains Ca 2+ -dependent cell–cell adhesion in epithelial tissues and therefore plays an important role in tumorigenesis as a tumor suppressor gene.\textsuperscript{26} Reduced E-cadherin expression is reported to be a feature of colorectal cancer.\textsuperscript{27} Recent evidence suggested that reduced E-cadherin expression was associated with pathological features, such as poor differentiation, infiltrative growth, lymph node metastasis, and poor prognosis.\textsuperscript{28} Altered expression of e-cadherin has been observed in a variety of malignancies, including ovarian,\textsuperscript{29} prostate,\textsuperscript{30} liver,\textsuperscript{31} and breast cancers.\textsuperscript{32} CDH1 hypermethylation is one of the mechanisms causing E-cadherin expression reduction. Lack of E-cadherin results in the dysfunction of the cell–cell adhesion system, causing a loss of cell–cell adhesion, release of cytoplasmic β-catenin, activating Wnt signaling and increasing tumor invasiveness.\textsuperscript{33} Hypermethylation and reduced expression of CDH1 have been observed in a variety of malignancies, including gastric cancer, ovarian cancer and breast cancer.\textsuperscript{34,35} Interestingly, we found overexpression of CDH1 in metastatic melanoma in our study, which was not consistent with the results of previous studies on other tumors, thus further exploring is needed.

The limitation of our study lies in the fact that hub gene is identified through calculation, and there is still a lack of clinical and basic experimental verification. Therefore, further research is needed. However, our study is scientifically supported and credible, and may provide new ideas for the treatment of CMM.

Conclusion
In summary, this study identified 206 DEGs between metastatic melanoma and primary melanoma, of which, ten hub genes with the most connection degree were picked out. These genes are related to the cell functions of tumor-causing and are involved in some of the signaling pathways of tumorigenesis and development. Among them, overexpression of PKP1, FLG and DSG1 are significantly associated with overall survival of CMM patients. CDH1 and EGFR were related to tumor immunity. Due to strong invasion of CMM, only the early diagnosis and treatment can improve prognosis, these genes may become biomarkers of diagnosis for primary melanoma and therapeutic targets to inhibit metastasis, opening up a novel way for the treatment of CMM. Nevertheless, further studies for the biological function
Figure 8 The relationship between EGFR and immune cells infiltration in metastatic melanoma.

Notes: Black dots represent the number of immune cell infiltration, the more black dots, the more infiltration. The blue line represents the relationship between immune cell infiltration and gene expression level, the blue line upward sloping represents the positive relationship, the blue line downward sloping represents the negative relationship, the greater the gradient, the closer the relationship.

Abbreviation: EGFR, epidermal growth factor receptor.
Figure 9 The relationship between CDH1 and immune cells infiltration in metastatic melanoma.

Notes: Black dots represent the number of immune cell infiltration, the more black dots, the more infiltration. The blue line represents the relationship between immune cell infiltration and gene expression level, the blue line upward sloping represents the positive relationship, the blue line downward sloping represents the negative relationship, the greater the gradient, the closer the relationship.

Abbreviation: CDH1, cadherin 1.
of these genes, laboratory and clinical studies are still needed.

**Ethics and Consent Statements**

The data in this study came from the GEO public database, and the original data had passed the ethical review and obtained the written informed consent of the patients when provided. Our use of the data from the GEO public database was reviewed by the Ethical Review Committee of Weifang Medical College.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


