Identification of genes associated with melanoma metastasis

Tao Qiu, Hongyi Wang, Yang Wang, Yu Zhang, Qiang Hui, Kai Tao*

Department of Plastic Surgery, General Hospital of Shenyang Military Area Command, Shenyang, Liaoning, China

Received 16 July 2015; accepted 14 September 2015
Available online 21 November 2015

Abstract The aims of the study were to identify the differentially expressed genes (DEGs) between primary melanomas and metastasis melanomas (MMs), and to investigate the mechanisms of MMs. The microarray data GSE8401 including 31 primary melanomas and 52 MMs were downloaded from Gene Expression Omnibus. DEGs were identified using the Linear Models for Microarray Data package. The functional and pathway enrichment analyses were performed for DEGs. Identification of transcription factors, tumor-associated genes (TAGs), and tumor suppressor genes (TSGs) were performed with the TRANSFAC, TAG, and TSGene databases, respectively. A protein–protein interaction network was constructed using Search Tool for the Retrieval of Interacting Genes. The modules construction and analysis was performed using Molecular Complex Detection and Gene Cluster with Literature Profiles, respectively. In total, 1004 upregulated and 1008 downregulated DEGs were identified. The upregulated DEGs, such as CDK1, BRCA1, MAD2L1, and PCNA, were significantly enriched in cell cycles, DNA replication, and mismatch repair. The downregulated DEGs, such as COL1A1, COL4A5, COL18A1, and LAMC2, were enriched in cell adhesion and extracellular matrix-receptor interaction. BRCA1 was identified as a transcription factor and TSG, and COL18A1 and LAMC2 were identified as a TSG and TAG, respectively. The upregulated DEGs had higher degrees in the protein–protein interaction network and module, such as PCNA, CDK1, and MAD2L1, and the heat map showed they were clustered in the functions of cell cycle and division. These results may demonstrate the potential roles of DEGs such as CDK1, BRCA1, COL18A1, and LAMC2 in the mechanism of MM.

Copyright © 2015, Kaohsiung Medical University. Published by Elsevier Taiwan LLC. All rights reserved.
Introduction

Melanoma is a lethal skin cancer that accounts for 4–5% of all cancers and metastatic melanoma (MM) accounts for 80% of skin cancer deaths [1–3]. Metastasis is the dispersal of cancer cells from their primary loci to distant organs and accounts for > 90% of deaths in cancer patients [4]. About 80% of melanomas are diagnosed at localized stages, whereas one-third of those patients will develop MMs [5]. As shown by both clinical and animal model studies, metastasis is a severe event [4]. Therefore, it is very important to understand the mechanisms of MM.

Regarding the causes of melanoma, there is a complex interaction of environmental (exogenous) and endogenous risk factors in developing malignant melanomas or MM [4]. UV radiation is clearly the predominant environmental risk factor of melanoma and accounts for ∼ 86% of MMs. BRAF, a component of the Ras-mitogen-activated protein kinase-ERK pathway, is commonly targeted in cancers with an overall occurrence of 7%, whereas the BRAF mutations frequency approaches 70% in MM [6–9]. BRAF activation has been shown to lead to the development of benign nevi, while the cooperation of the deficiency of the p53 pathway is required for the progression to frank melanoma. In addition, constitutive expression of AKT has been observed in nearly 60% of melanomas and inhibition of AKT in the onset and further development of melanoma altered the progression of melanoma invasion and metastasis. Recently, high-dose interferon-α therapy and interleukin-2 therapy, antibody blockade of cytotoxic T-lymphocyte-associated protein 4, inhibitors of the mitogen-activated protein kinase pathway, and adoptive cell therapy have also been used in the treatment of melanoma. In spite of this, the prognosis for patients with high-risk or advanced MM remains poor compared with the increasing incidence of melanoma.

Several studies have shown that primary tumors and their metastases have different gene expression signatures that predicts their propensity to metastasize. As reported by Xu et al [4], 150 differentially expressed genes (DEGs) showed different expression profiles between the primary melanomas (PMs) and MMs identified through gene expression microarray data, and functional and pathway enrichment analysis showed the DEGs were involved in processes such as extracellular matrix (ECM)-receptor interaction, focal adhesion, as well as the transforming growth factor (TGF)-β signaling pathway, suggesting the important roles of DEGs in the metastasis of melanoma. However, further identification of the transcription factors (TFs), tumor-associated genes (TAGs), and tumor suppressor genes (TSGs), as well as the construction and analysis of the protein–protein interaction (PPI) network and the subnetwork of the DEGs were not performed.

In order to further investigate the mechanisms of MM, the microarray data GSE8401 from Xu et al [4] were downloaded. Identification of DEGs were performed with lower cut-off criteria for amplifying the inclusion genes. In addition to the functional and pathway enrichment analysis, construction and analysis of the PPI network and module network of the DEGs were performed for better understanding the biological processes of MM. The reanalysis of the microarray data would provide more abundant information for understanding the mechanisms of MM, as well as for the strategy of diagnosis and therapeutic treatment.

Methods

Microarray data and data preprocessing

The gene expression microarray data of melanoma were obtained from the national center for biotechnology information (NCBI) Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) and were accessible through Gene Expression Omnibus series accession number GSE8401 [4]. The platform of this data was GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. A total of 83 samples, including 31 PM samples and 52 aggressive MM samples were available in the microarray data. The probe files were downloaded for further analysis.

Probe sets were mapped to NCBI Entrez genes using Gene ID converter. If there were multiple probe sets that corresponded to the same gene, the average value was calculated for further analysis. The probe was filtered if it had multiple genes. The data were normally distributed through log 2 transformation, and were normalized using preprocess Core package [10].

Identification of DEGs

For the GSE8401 dataset, the classical t test in Linear Models for Microarray Data package of R was used to identify the DEGs [11]. Benjamini and Hochberg’s [12] method was used to adjust the raw p value into false discovery rate. The cut-off criteria for DEGs were |log_2 fold change| > 0.58 and false discovery rate < 0.05.

Functional and pathway enrichment analysis

GO (Gene Ontology) analysis is a functional study method of large-scale transcriptomic data or genomic, while KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway database provides information on how molecules or genes work [13]. In order to investigate the biofunction of DEGs in tumor progression, GO functional enrichment analysis in biological process (BP), molecular function (MF), cellular component (CC) categories were performed using the GO database, while the pathway enrichment analysis was performed using the KEGG database. The threshold was set as p < 0.01.

Identification of TFs, TSGs, and TAGs

TRANSFAC database (http://www.embl-heidelberg.de/srs/srsc) is not only a database about eukaryotic transcription regulating DNA sequence elements, but also a database about the TFs binding to DNA sequences and acting through them. To determine whether the DEGs had a transcription regulation function, TF prediction of upregulated and downregulated DEGs were performed using the TRANSFAC database, respectively.
TAG database (http://www.binfo.ncku.edu.tw/TAG/), collects specific information of target genes from various resources and is a semiautomatic information retrieving engine [14]. Additionally, TSGene database (http://bioinfo.mc.vanderbilt.edu/TSGene/) provides a comprehensive resource of TSGs for the cancer research community to further experimental designs [15]. Furthermore, a comprehensive TSG catalog for advanced systems biology-based analyses can be provided by the TSGene database (http://bioinfo.mc.vanderbilt.edu/TSGene/).

To identify genes that play important roles in carcinogenesis and facilitate cancer research, TSGs and TAGs in upregulated and downregulated DEGs were identified using the TSGene database and TAG database, respectively.

Construction of PPI network and the module analysis

The STRING (Search Tool for the Retrieval of Interacting Genes) database provides uniquely comprehensive coverage and ease of access to both experimental as well as predicted interaction information [16]. To better understand DEGs from an interactive perspective, we used the STRING database to build an interaction network of encoding products of all of the DEGs, i.e., the PPI network. A STRING score of 0.4 was set as the reliability threshold. Cytoscape is a standard tool for integrated analysis and visualization of biological networks. Cytoscape software (version 2.8, http://www.cytoscape.org/) was used to visualize the PPI network. Then, using the scale-free properties of PPI networks, connectivity degree analysis was performed and hub nodes were obtained.

Molecular Complex Detection (MCODE) in Cytoscape, a novel graph theoretic clustering algorithm that detects densely connected regions in large PPI networks that may represent molecular complexes, was used to obtain modules of the PPI network with the parameters set to the default values. Gene Cluster with Literature Profiles analyzes human genes with enriched keywords and molecular interactions to construct molecular networks and modules related to the free terms [17]. Finally, the analysis and construction of the heat map of the module was performed with Gene Cluster with Literature Profiles.

Results

Identification of DEGs

After mapping to the NCBI Entrez genes and normalization, the expression spectrum matrix including 12,750 genes was obtained. Afterwards, a total of 2012 DEGs were identified, including 1004 upregulated DEGs (such as CDK1, MAD2L1, RFC4, PCNA, BRCA1, SPC25, SMC2, and SMC4) and 1008 downregulated DEGs (such as COL1A1, COL6A1, COL18A1, LAMA2, LAMC2, SEMA3G, SEMA3C, S100A2, and S100A7).

Functional and pathway enrichment analysis

Through the functional enrichment analysis in the BP category, upregulated DEGs were significantly involved in mitotic cell cycle (p < 0.01), such as CDK1, MAD2L1, RFC4, and PCNA; nuclear division (p < 0.01), such as CDK1, MAD2L1, and SMC2 (Table 1). After enrichment analysis in the CC category, the upregulated DEGs were significantly involved in membrane-enclosed lumen (p < 0.01), such as CDK1, RFC4, PCNA, and SM2; membrane-bounded organelle (p < 0.01), such as CDK1, MAD2L1, RFC4, and PCNA (Table 1). For the analysis in the MF category, the upregulated DEGs were significantly involved in DNA-dependent ATPase activity (p = 3.49E-06), such as RFC3, MCM6, MCM7, and MSH6; small molecule binding (p = 4.29E-06), such as CDK1, RFC4, SMC2, and SMC4; protein binding (p = 5.43E-06), such as CDK1, MAD2L1, RFC4, and SMC2 (Table 1). Moreover, after the annotation of KEGG pathway, the upregulated DEGs were significantly involved in cell cycle (p = 5.19E-11), such as CDK1, MAD2L1, RFC4, and PCNA; DNA replication (p = 9.22E-07), such as RFC3, RFC4, and PCNA; mismatch repair (p = 4.53E-05), such as RFC3,

<table>
<thead>
<tr>
<th>Category</th>
<th>ID</th>
<th>Name</th>
<th>Count</th>
<th>Gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>GO:0000278</td>
<td>Mitotic cell cycle</td>
<td>139</td>
<td>CDK1, MAD2L1, RFC4, SMC2, SMC4, PCNA, BRCA1</td>
</tr>
<tr>
<td></td>
<td>GO:000280</td>
<td>Nuclear division</td>
<td>67</td>
<td>CDK1, MAD2L1, MAD2L1BP, SMC2, SMC4, SPC25</td>
</tr>
<tr>
<td></td>
<td>GO:0007049</td>
<td>Cell cycle</td>
<td>189</td>
<td>CDK1, MAD2L1, SMC2, HMGA2, SPC25</td>
</tr>
<tr>
<td></td>
<td>GO:0007067</td>
<td>Mitosis</td>
<td>67</td>
<td>CDK1, MAD2L1, SMC2, SMC4, HMGA2, SPC25</td>
</tr>
<tr>
<td></td>
<td>GO:0022402</td>
<td>Cell cycle process</td>
<td>157</td>
<td>CDK1, MAD2L1, SMC2, SMC4, HMGA2, SPC25</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0031974</td>
<td>Membrane-enclosed lumen</td>
<td>245</td>
<td>CDK1, RFC4, PCNA, SMC2, COL11A1, HMGA2</td>
</tr>
<tr>
<td></td>
<td>GO:0031981</td>
<td>Nuclear lumen</td>
<td>192</td>
<td>CDK1, HMGA2, SMC2, PCNA, BRCA1, RFC4</td>
</tr>
<tr>
<td></td>
<td>GO:0043227</td>
<td>Membrane-bounded organelle</td>
<td>628</td>
<td>CDK1, MAD2L1, RFC4, PCNA, COL11A1, SMC4</td>
</tr>
<tr>
<td></td>
<td>GO:0043231</td>
<td>Intracellular membrane-bounded organelle</td>
<td>626</td>
<td>CDK1, COL11A1, SMC4, HMGA2, SPC25</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0043233</td>
<td>Organelle lumen</td>
<td>242</td>
<td>BRCA1 CDK1, COL11A1, HMGA2, SMC2, RFC4</td>
</tr>
<tr>
<td></td>
<td>GO:0008094</td>
<td>DNA-dependent ATPase activity</td>
<td>15</td>
<td>RFC3, MCM6, MCM7, MSH6, MSH2, DNA2</td>
</tr>
<tr>
<td></td>
<td>GO:0036094</td>
<td>Small molecule binding</td>
<td>179</td>
<td>CDK1, RFC4, RFC3, SMC2, SMC4, MCM7</td>
</tr>
<tr>
<td></td>
<td>GO:0005515</td>
<td>Protein binding</td>
<td>466</td>
<td>CDK1, MAD2L1, RFC4, SMC2, MAGEA3, MAGEA6</td>
</tr>
<tr>
<td></td>
<td>GO:0005524</td>
<td>ATP binding</td>
<td>117</td>
<td>CDK1, SMC4, SMC2, MCM6, RFC4, MCM7</td>
</tr>
<tr>
<td></td>
<td>GO:0000166</td>
<td>Nucleotide binding</td>
<td>169</td>
<td>CDK1, SMC4, SMC2, RFC3, MCM6, MCM7</td>
</tr>
</tbody>
</table>

BP = biological process; CC = cellular components; MF = molecular functions.
such as COL1A1, LAMC2, and LAMC3; and focal adhesion (p = 8.96E-05), such as COL1A1, LAMC2, and LAMC3 (Table 4).

Identification of TFs, TAGs, and TSGs

For the upregulated DEGs, 34 TFs (such as BRCA1, HIF1A, TAF4, and HMG2), 18 TAGs (such as HMG2, CCNA2, PIK3CA, and MYBL1), and 42 TSGs (such as BRCA1, SMAD4, MSH2, and CDC73) were identified. Meanwhile, 55 TFs (such as TP63, KLF4, KLF13, and FOXC1), 30 TAGs (such as CD24, LAMC2, CCND1, and FGFR2), and 67 TSGs (such as TP63, CDKN1C, KLF5, and COL18A1) were identified among the downregulated DEGs (Table 5).

PPI network and module analysis

In total, 1574 nodes and 13,876 edges were included in the PPI network (Figure 1). The top five DEGs with the highest degrees were PCNA (degree = 188), CDK1 (degree = 180), MAD2L1 (degree = 171), RFC4 (degree = 168), and BRCA1.
Identification of MM related genes

The top 10 enriched Kyoto Encyclopedia of Genes and Genomes pathways for the downregulated differentially expressed genes.

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Count</th>
<th>Gene symbol</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>4360</td>
<td>Axon guidance</td>
<td>25</td>
<td>EPHA1, GSK3B, SEMA2C, SEMA3C, SEMA3F, SEMA3G, SEMA4D</td>
<td>2.98E-07</td>
</tr>
<tr>
<td>4512</td>
<td>ECM-receptor interaction</td>
<td>17</td>
<td>COL1A1, COL4A5, COL6A1, LAMA2, LAMC2, LAMC3</td>
<td>1.59E-05</td>
</tr>
<tr>
<td>5200</td>
<td>Pathways in cancer</td>
<td>40</td>
<td>GSK3B, LAMA2, LAMA3, LAMB3, LAMC3, SLC2A1</td>
<td>2.60E-05</td>
</tr>
<tr>
<td>5146</td>
<td>Amoebiasis</td>
<td>18</td>
<td>COL1A1, COL4A5, LAMA2, LAMA3, LAMC2, LAMC3</td>
<td>8.96E-05</td>
</tr>
<tr>
<td>4510</td>
<td>Focal adhesion</td>
<td>27</td>
<td>COL1A1, COL4A5, COL6A1, COL6A2, LAMA2, LAMC3</td>
<td>0.001</td>
</tr>
<tr>
<td>590</td>
<td>Arachidonic acid metabolism</td>
<td>12</td>
<td>ALOX12, ALOX12B, ALOX15, ALOX15B, GPX2, GPX3</td>
<td>0.002</td>
</tr>
<tr>
<td>4640</td>
<td>Hematopoietic cell lineage</td>
<td>15</td>
<td>IL11RA, IL1R1, IL1R2, IL1R2, IL1R2, IL7R, IL4R, CD4, CD4, CD9</td>
<td>0.003</td>
</tr>
<tr>
<td>5150</td>
<td>Staphylococcus aureus infection</td>
<td>11</td>
<td>HLA-DPA1, HLA-DPB1, HLA-DQB1, CFB, CFD, PDF</td>
<td>0.006</td>
</tr>
<tr>
<td>480</td>
<td>Glutathione metabolism</td>
<td>10</td>
<td>GSTM3, GSTA3, GST21, GSTA4, GSTP1, GSTT1</td>
<td>0.010</td>
</tr>
<tr>
<td>5219</td>
<td>Bladder cancer</td>
<td>9</td>
<td>MMP1, MMP2, EFGFR, FGFR3, CCND1, CDH1, ERBB2, HRAS, TYMP</td>
<td>0.010</td>
</tr>
</tbody>
</table>

ECM = extracellular matrix.

Discussion

In our study, a total of 2012 DEGs between PM and aggressive MMs were identified, including 1004 upregulated and 1008 downregulated DEGs. From the PPI network we found the top five DEGs with higher node degrees were PCNA, CDK1, MAD2L1, RFC4, and BRCA1, and they were involved in the module with the highest MCODE score.

Through functional and pathway enrichment analysis, CDK1 and PCNA were enriched in the mitotic cell cycle, membrane-enclosed lumen, small molecule binding, and cell cycle. As previously reported, cyclin-dependent kinases (CDKs), such as CDK1, regulated the progression of cell cycle and DNA replication [18]. Additionally, proliferating cell nuclear antigen (PCNA) is essential for DNA replication and is involved in DNA repair [19]. In vertebrate cells, PCNA is crucially regulated by p21, an inhibitor of CDKs [20], and binding of p21 to PCNA inhibits DNA replication [19]. The expression of PCNA is correlated with the presence of malignancy and prognosis in numerous melanoma patients [21]. Moreover, the increased mRNA levels of melanoma differentiation associated MDA6 which encodes p21, displaying a loss of metastatic potential in MM cells. Taken together, these studies indicated that p21 functions as a negative regulator of melanoma growth, progression, and metastasis through regulating the CDKs and PCNA. Therefore, it could be speculated that CDK1 and PCNA participate in MM by affecting the cell cycle.

Moreover, RFC3, RFC4, and PCNA were enriched in the pathways of DNA replication and mismatch repair as well as the function enrichment of mitotic cell cycle and molecular binding. RFC1, RFC3, and RFC4 have the clamp loader activities when PCNA acts as a clamp to repair or replicate DNA [19]. A conserved motif of PIP box, PCNA-interacting protein that acts as PCNA partner, was found in RFCs [19]. When repair or replicate DNA, replication factor C (RFC ) binds to the primer template junction and catalyzes the loading of the ring-shaped replication factor PCNA, like a screw cap with a bottle that encircles DNA [19]. Moreover, the overexpression of RFC4 was highly significant in the MM within 4 years [22]. This suggested that RFC4 and PCNA might have crucial effects on cell cycle, mismatch repair, and DNA replication of MM cells, thus contributes to the development of MM.

In addition, BRCA1 and MAD2L1 were involved in mitotic cell cycle and nuclear lumen, and BRCA1 was identified as a
Figure 1. Protein–protein interaction network of the differentially expressed genes (DEGs). The red nodes indicate the upregulated DEGs. The green nodes indicate the downregulated DEGs. The molecular interactions between the DEGs are indicated by edges.
TF and TSG. It has been reported that \textit{BRCA1} is ubiquitously expressed and involved in cell proliferation, DNA damage response, apoptosis, and regulation of cell cycle progression, and is required for centrosome and mitotic spindle functions \cite{23}. \textit{BRCA1} participated in the recruitment or maintenance of \textit{MAD2L1} to kinetochores, and further investigation found that \textit{BRCA1} positively regulated \textit{MAD2L1} expression by interacting with its promoter \cite{24}. As previously reported, the loss or knockdown of \textit{BRCA1} induced the activation of \textit{AKT}, thus inducing the protein kinase B/mechanistic target of rapamycin signaling pathway \cite{25}, whose constitutive expression had been observed in most of melanomas and inhibition of \textit{AKT} in the onset and further development of melanoma altered the progression of MM \cite{26,27}. This suggested that \textit{BRCA1} and \textit{MAD2L1} might have a crucial role in the progression of MM by participating in the cell cycle pathway, and activation of the \textit{BRCA1} might be a potential therapeutic target to inhibit and alter the invasion and metastasis of MM.

However, the downregulated DEGs including laminin family members such as \textit{LAMC2} and \textit{LAMC3} and collagen family members such as \textit{COL1A1} and \textit{COL4A5} were significantly involved in cell adhesion, ECM, ECM-receptor interaction, and focal adhesion, and \textit{COL18A1} and \textit{LAMC2} was identified as the TSG and TAG, respectively. In addition to the hyperactivity of cell proliferation, metastasis is facilitated by cell–cell interactions between tumor cells and the endothelium in distant tissues. Laminin and collagen are ECM proteins that play important roles in maintaining the integrity of various tissues and mediating the adhesion, migration, and organization of cells into tissues. \textit{TGF-β} induced protein (\textit{TGF-β}I) is a \textit{TGF-β} inducible, secreted ECM protein impairs the adhesive contacts of melanoma cells with the ECM molecules, including type I collagen, type IV collagen, and laminin, and, therefore, promotes melanoma cell migration and metastatic progression \cite{28,29}. However, the knockdown of \textit{SPARC} with siRNA increased the expression of \textit{COL1A1} and \textit{COL3A1} and resulted in a significant decrease of proliferation in uveal melanoma cell lines. Furthermore, \textit{LAMC2} and \textit{LAMA4} were particularly overexpressed in angiotropic melanoma areas versus nonangiotropic areas, suggesting the increased \textit{LAMC2} and \textit{LAMA4} enhanced the focal adhesive ability of melanoma cells \cite{30}. This phenomenon demonstrated that

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{The module obtained from protein–protein interaction network with the highest score. The Molecular Complex Detection score is 48.875. A total of 104 differentially expressed genes are included in the module, and all of them are the upregulated differentially expressed genes, represented by the red nodes.}
\end{figure}
the laminin and collagen family members such as \textit{LAMC2}, \textit{COL1A1}, and \textit{COL18A1} might affect the cell and focal adhesion of the MM cells, and thus might play an important part in the progression of MM.

In conclusion, 2012 DEGs between the PM and MM groups were identified. The upregulated DEGs such as \textit{CDK1}, \textit{PCNA}, \textit{RFC4}, \textit{BRCA1}, and \textit{MAD2L1} were significantly involved in the cell cycle and proliferation, and downregulated DEGs, such as \textit{LAMA2}, \textit{LAMC2}, \textit{COL1A1}, and \textit{COL4A5} were involved in cell adhesion and ECM-receptor interaction, respectively. Furthermore, the upregulated DEGs had higher scores in the PPI network, such as \textit{CDK1}, \textit{PCNA}, \textit{RFC4}, and \textit{BRCA1}. These DEGs might be involved in the process of MM. However, further experimental verification is still needed to confirm the potential effects of the DEGs in MM.

References


**Figure 3.** The heat map of function enrichment of the 104 differentially expressed genes in the module. The horizontal axis indicates the DEGs in the module, and the vertical axis indicates the significantly enriched functions. In the map, the green color means the DEG is involved in the function, while the black color means the DEG is not enriched in the function.
Identification of MM related genes


