Epidermal Growth Factor-Like Repeats and Discoidin I-Like Domains 3 is a Novel Regulator of Epithelial-Mesenchymal Transition in Clear Cell Renal Cell Carcinoma: In Silico Analysis

Rahşan Ilıkçı Sağkan1, Dilara Fatma Akın Balı2

Objective: Epithelial-mesenchymal transition (EMT) contributes to cancer metastasis and recurrence, which are major obstacles in changing the course of cancer. However, studies on the mutational and gene expression profiles of epidermal growth factor-like repeats and discoidin I-like domains 3 (EDIL3) that reveal the relationship between clear cell renal cell carcinoma (ccRCC) and EMT markers are limited. The aim of our study was to reveal the correlation between tumor and EMT markers (E-cadherin and vimentin) and EDIL3 expression. Additionally, we evaluated target gene expression levels and mutational profiles in kidney cancer tissue and normal tissue.

Materials and Methods: We investigated the mutational profile and mRNA expression of EDIL3 and compared them with that of VIM and CDH1 in 523 patients with ccRCC using validated bioinformatics analysis. Additionally, Polymorphism Phenotyping v2 (PolyPhen-2), Screening for NonAcceptable Polymorphisms (SNAP) were used to predict and confirm the pathogenicity of the mutations detected. Studies were performed in silico using bioinformatics tools.

Results: EDIL3 and VIM expression was statistically significantly higher in the healthy group and exhibited a positive correlation in patients with ccRCC. Patients with elevated VIM and CDH1 expression and low EDIL3 expression had prolonged survival time. In addition, 7 mutations were detected in the evaluated genes, 6 of which had potential pathogenic features.

Conclusion: Our study provides insights for further experimental studies. EDIL3 can be used as a diagnostic or prognostic indicator of cancer development to help cure renal clear cell cancer.

Keywords: Epidermal growth factor-like repeats and discoidin I-like domains 3, epithelial-mesenchymal transition, clear cell renal cell carcinoma, mutation, gene expression

INTRODUCTION

In adults, clear cell renal cell carcinoma (ccRCC) constitutes 3%–4% of cancers and is the most common cancer of the kidney, accounting for approximately 65%–70% of all renal malignancies (1, 2). Cancer progression toward the invasive and metastatic is associated with the reactivation of EMT in renal cancer (3, 4). Acquisition of EMT characteristics involves increased expression of mesenchymal markers, including N-cadherin and vimentin (VIM), and decreased expression of epithelial markers, especially E-cadherin also known cadherin-1 (CDH1). Upregulation of N-cadherin and VIM and downregulation of E-cadherin are hallmarks of EMT, which involves phenotypic alterations, such as loss of cell-cell adhesion and gain of cell migration capabilities (5–7). Additionally, it involves the growth of blood vessels, angiogenesis of the tumor, and migration of abnormal kidney cancer cells.

Developmentally regulated endothelial cell locus 1 (DEL-1), also known as epidermal growth factor-like repeats and discoidin I-like domains 3 (EDIL3), is a glycoprotein associated with the endothelial cell surface and extracellular matrix and is expressed on some tissues during embryonic development (8). In adults, the expression of EDIL3 is limited to normal tissues except certain types of cancer such as carcinoembryonic antigen (CEA) (9). EDIL3 is a novel angiogenic factor, which plays a role in modulating pathological angiogenesis (10). Moreover, EDIL3 has been shown to induce EMT in many types of tumor tissues (10–14). EDIL3 is overexpressed during the transition from kidney intraepithelial neoplasia to invasive carcinoma and causes a significant increase in proliferation, angiogenesis, invasion, and metastasis (15).

We hypothesized that EDIL3 serves as a small molecule regulator of the EMT program in renal cancer. To identify characteristics of ccRCC due to EDIL3-mediated EMT in cancer progression, tissue samples from patients with ccRCC were evaluated for EMT marker gene expression, including CDH1, VIM, and EDIL3, and their mutational profiles were compared with that of healthy tissue samples. Additionally, survival analysis was performed to evaluate the prognostic significance of EDIL3.
MATERIALS and METHODS

Downloading Patient Data

The cBio Cancer Genomics Portal (http://cbióportal.org) hosts more than 40 datasets from The Cancer Genome Atlas Program (TCGA) and other large-scale genomic studies and makes them available for bulk download. The TCGA dataset is a sub-collection of the open-access genomics archive of The cBio Cancer Genomics Portal. For analysis of EDIL3, VIM, and CDH1 mutations, “TCGA, PanCancer Atlas” datasets were selected. Datasets for patients with ccRCC were downloaded on April 2020 from the cBio Cancer Genomics Portal. The “OncoPrint” tab displays a general view of genetic alterations within each sample of studied genes. The characteristic features of these patients were downloaded from the web browser; data are shown in Table 1.

Mutational Analysis of EDIL3, VIM, and CDH1 in Patients with ccRCC

The cBio Cancer Genomics Portal (http://cbióportal.org) includes data from National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI) libraries. Data from approximately 5,000 tumors are available to cancer researchers and in-library users for interactive cancer genomics studies (16).

Possible mutations in EDIL3, VIM, and CDH1 were detected on the web interface for genome sequencing data of 523 patients with ccRCC, which was available in the portal. We analyzed the mutation scale using OncoPrint and Mutation algorithms.

Bioinformatics Analysis for EDIL3, VIM, and CDH1 in Patients with ccRCC

Polymorphism Phenotyping v2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2), Screening for NonAcceptable Polymorphisms (SNAP) (https://www.rostlab.org/services/SNAP/), and Catalog of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic) databases were used to demonstrate the potential pathogenicity of mutations in our target genes, namely EDIL3, VIM, and CDH1. PolyPhen-2, which can be accessed from a web server, predicts the effect of potential mutations on the stability and activity of synthesized proteins. We combined evolutionary and structural information of these potential amino acid substitution mutations. Moreover, it evaluates the functional interpretation of single nucleotide polymorphisms (SNPs), maps coding SNPs to transcripts, extracts annotations of protein sequence and structure information, and constructs conservation profiles. The potential of missense mutations was estimated using a scoring system based on three parameters, including probably damaging, possibly damaging, and benign or unknown, using this program (17). SNAP is a neural network-based online tool. It is a novel sequence aligner and measures synonymous and non-synonymous substitution rates on a set of codon-aligned nucleotide sequences by considering sequence variety. The most important single feature for SNAP software prediction is conservation and constructs conservation profiles. The potential of missense mutations across species using the “Multiple sequence alignment” tool.

Analysis of EDIL3, VIM, and CDH1 mRNA Expression in Patients with ccRCC

GEPIA (http://gepia.cancer-pku.cn/index.html) is a new interactive web server for analyzing RNA sequencing expression data from 9,736 tumors and 8,587 normal samples from the TCGA and GTEx projects using a standard processing pipeline. This advanced web server can be used for various applications, including transcript expression analysis, profiling based on cancer types or pathological stages, dimensionality reduction, analogous gene detection, correlation analysis, and overall survival analysis (19). Levels of EDIL3, VIM, and CDH1 mRNA from 523 patients with ccRCC and 100 normal tissue were represented as box-plot graphs using the GEPIA database. Transcripts per million (TPM) was used to quantify mRNA expression. Gene expression data were first log2 (TPM + 1) transformed for differential analysis and log2FC was defined as median (tumor tissue) – median (normal tissue). Genes with higher [log2FC] values and lower q values than pre-set thresholds were considered differentially expressed genes.

Analyses of correlation between the expression of EDIL3 and other target genes were performed and p-values were obtained through this web server. Finally, survival analyses of target genes, including EDIL3, VIM, and CDH1, based on change in expression were carried out using the software.

Statistical Analysis

Statistical analyses for data evaluation were performed using the GEPIA database. One-way ANOVA was performed for differential analysis, using the disease state (tumor or normal) as a variable for evaluating differential gene expression. Overall survival analysis was carried out using Kaplan–Meier curves. The logrank test was used for comparing the low and high expression groups. Correlation analysis was performed using Pearson’s correlation test using the electronically connected database. A value of p<0.05 was considered statistically significant.

RESULTS

Results of Mutation Analysis for EDIL3, VIM, and CDH1 in Patients with ccRCC

To detect genetic alterations in EDIL3, VIM, and CDH1 in ccRCC samples, genome sequencing data of 523 patients with ccRCC were examined through the cBioPortal interface. A total of 1.9% of patients with ccRCC had 7 mutations, namely 4 missense, 1 nonsense, and 2 frame shifts, in EDIL3, VIM, and CDH1. Detailed information on detected mutations is summarized in Table 2. EDIL3 was the most frequently mutated, with a 1.3% change. The nonsense p.G185* mutation created a termination codon at amino acid 185 of EDIL3, leading to the formation of a truncated protein (Fig. 1). Frameshift mutations p.I289Hfs*27 and p.P127Af5s*41, which were detected in EDIL3 and CDH1, respectively, shifted the reading frame, leading to the formation of truncated proteins. As presented in Figure 2, all detected mutations in EDIL3 are present on protein domains encoded by the gene. The p.N315S missense mutation in CDH1 is located in an extracellular region. However, the p.P127Af5s*41 frame-shift change is on a location that may lead to the early terminated protein formation at such a level, which can impact the synthe-
sis, placement and function of the protein of E-cadherin. The p.L380F missense mutation in VIM is located on the filament domain. All mutations were saved as somatic mutations in the COSMIC database. Domain architecture of proteins and mutations found in ccRCC are shown in Figure 2.

**Results of Bioinformatics Analysis for EDIL3, VIM, and CDH1 in Patients with ccRCC**

PolyPhen-2 analysis revealed that 6 of 7 missense mutations detected, which are presented in detail in Table 1, might pathogenically be Probably Damaging because their score is close to 1. However, 3 of 4 missense mutations, which were detected using SNAP analysis, were determined as being affected because their scores ranged from 0 to 100.

We compared amino acid substitutions caused by missense mutations in different species using the “Multiple sequence alignment” option, which is included in the PolyPhen-2 program. Our findings revealed that p.Q187R, p.T343M, and p.L380F in EDIL3 were present at critically important locations and were evolutionarily conserved across species. Predicted pathogenic characteristics and evaluations of evolutionary conservation, which were carried out using the PolyPhen-2 software, are shown in Figure 3a–d. However, the p.N315S missense change in CDH1 was identified as “benign” and thus, is not at a critical localization in the evolutionary process (Fig. 4).

**Table 1.** Demographic, clinical and genetic data of patients with clear cell renal cell carcinoma (ccRCC)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patient data n=523</th>
<th>Characteristic</th>
<th>Patient data n=523</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td>Stage IV</td>
<td></td>
</tr>
<tr>
<td>Male/Female/NA</td>
<td>326/186/11</td>
<td>83</td>
<td>15.8</td>
</tr>
</tbody>
</table>
| Diagnosis age, years                   | 60 (range, 29–90)  | NA                                     | 11                 | 2.1
| Race category                          |                    | Tumor stage code                       |                    |
| White                                  | 442                | T1                                     | 16                 | 3
| Black or African American              | 55                 | T1a                                    | 127                | 24.2|
| Asian                                  | 8                  | T1b                                    | 103                | 19.6|
| NA                                     | 18                 | T3                                     | 3                  | 0.5
|                                      |                    | T3a                                    | 117                | 22.3|
| Sample type                            |                    | T3b                                    | 51                 | 9.7
| Primary                                | 512                | T3c                                    | 2                  | 0.3
| NA                                     | 11                 | T2                                     | 48                 | 9.1
| Overall survival status                |                    | T2a                                    | 9                  | 1.7
| Living                                 | 353                | T2b                                    | 4                  | 0.7
| Deceased                               | 170                | T4                                     | 7                  | 1.3
| Overall Survival (months)-median       | 6.11               | NA                                     | 36                 | 6.8|
| Metastasis stage code                  |                    |                                        |                    |
| M0                                     | 403                | EDCI3 Mutation                         | 4                  | 1.13|
| MX                                     | 29                 | EDCI3 Amplification                    | 0                  | 0
| M1                                     | 78                 | EDCI3 Deep deletion                    | 0                  | 0
| M1a                                    | –                  | VIM Mutation                           | 1                  | 0.28|
| M1b                                    | –                  | VIM Amplification                      | 0                  | 0
| NA                                     | 13                 | VIM Deep deletion                      | 0                  | 0
| Neoplasm disease stage cancer code     |                    | CDH1 Mutation                          | 1                  | 0.28|
| Stage I                                | 249                | CDH1 Amplification                     | 0                  | 0
| Stage II                               | 54                 | CDH1 Deep deletion                     | 0                  | 0
| Stage III                              | 126                |                                        |                    |

MX: Distant metastasis cannot be assessed; M0: No distant metastasis; M1: Distant metastasis; M1a: Distant metastasis to lung on opposite side of the primary tumor, pleural lymph nodes or malignant or pericardial effusion; M1b: Distant metastasis; NA: Not applicable

**Analysis of EDIL3, VIM, and CDH1 mRNA Expression in Patients with ccRCC**

We compared EDIL3, VIM, and CDH1 mRNA expression profiles of 523 patients with ccRCC with that of 100 healthy samples. Levels of EDIL3 and VIM mRNA were found to be considerably higher in patients with ccRCC when compared to samples from healthy individuals. However, the level of CDH1 mRNA was high in the healthy sample group. Analysis of expression profiles of all three genes revealed statistical significance among study groups (Fig. 4a, p<0.05). Analysis of the
Table 2. Epidermal growth factor-like repeats and discoidin I-like domains 3 (EDIL3), vimentin (VIM), and cadherin-1 (CDH1) mutations in patients with clear cell renal cell carcinoma (ccRCC)

<table>
<thead>
<tr>
<th>No</th>
<th>Gene</th>
<th>Nt alteration</th>
<th>Rs number</th>
<th>Alteration type</th>
<th>Localization</th>
<th>AA position</th>
<th>Previously determined disease</th>
<th>Clinical characteristics of patients with the mutation</th>
<th>Clinical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-1</td>
<td>EDIL3</td>
<td>c.559C&gt;A</td>
<td>COSV56902716</td>
<td>Missense mutation</td>
<td>F5_F8 Type C</td>
<td>p.Q187R</td>
<td>ccRCC M1 Decesead T4</td>
<td>Overall survival months: 5</td>
<td>Probably Damaging (score 1.00) Effect (score 0.99) Pathogenic</td>
</tr>
<tr>
<td>M-2</td>
<td>EDIL3</td>
<td>c.1028C&gt;T</td>
<td>COSV56879041</td>
<td>Missense mutation</td>
<td>F5_F8 Type C</td>
<td>p.T343M</td>
<td>ccRCC M0 Living T3A</td>
<td>Overall survival months: 3</td>
<td>Probably Damaging (score 0.98) Effect (score 0.99) Pathogenic</td>
</tr>
<tr>
<td>M-3</td>
<td>EDIL3</td>
<td>c.553G&gt;T</td>
<td>COSV56908142</td>
<td>Nonsense mutation</td>
<td>F5_F8 Type C</td>
<td>p.G185*</td>
<td>ccRCC M1 Decesead T1B</td>
<td>Overall survival months: 10</td>
<td>Probably Damaging (score 1.00) – Pathogenic (score 0.98)</td>
</tr>
<tr>
<td>M-4</td>
<td>EDIL3</td>
<td>c.865ins</td>
<td>N/A</td>
<td>Frame Shift mutation</td>
<td>F5_F8 Type C</td>
<td>p.I289Hfs*27</td>
<td>ccRCC M0 Living T2</td>
<td>Overall survival months: 49</td>
<td>Probably Damaging (score 1.00) N/A</td>
</tr>
<tr>
<td>M-5</td>
<td>VIM</td>
<td>c.1138C&gt;T</td>
<td>COSV56407096</td>
<td>Missense mutation</td>
<td>Filament</td>
<td>p.L380F</td>
<td>ccRCC M0 Living T2</td>
<td>Overall survival months: 83</td>
<td>Probably Damaging (score 1.00) Effect (score 23) Pathogenic (score 0.94)</td>
</tr>
<tr>
<td>M-6</td>
<td>CDH1</td>
<td>c.377dup (Insertion)</td>
<td>COSV55727472</td>
<td>Frame Shift mutation</td>
<td>PRE</td>
<td>p.P127As*41</td>
<td>ccRCC M1 Decesead T1B</td>
<td>Overall survival months: 43</td>
<td>Probably Damaging (score 1.00) – N/A</td>
</tr>
<tr>
<td>M-7</td>
<td>CDH1</td>
<td>c.944A&gt;G</td>
<td>COSV55727472</td>
<td>Missense mutation</td>
<td>Cadherin</td>
<td>p.N315S</td>
<td>ccRCC M0 Living T3b</td>
<td>Overall survival months: 129</td>
<td>Benign (0.00) Neutral (-65) Neutral (score 0.17)</td>
</tr>
</tbody>
</table>

M: Mutation; COSV: The genomic mutation identifier; PRE: Precursor peptit; ccRCC: Clear cell renal cell carcinoma; M0: No distant metastasis; M1: Distant metastasis; T describes the size of the tumor; N/A: Not available
relationship between expression profiles of VIM and CDH1 and that of EDIL3 independently using Pearson’s correlation test revealed a positive correlation between VIM and EDIL3 expression (Fig. 4b, r=0.31). Results from survival analysis revealed that overall survival times and disease-free survival times of patients with ccRCC with low EDIL3 were longer than those with high EDIL3 (Fig. 5a). Patients with ccRCC and high VIM and CDH1 expression had longer survival, which was statistically significant, than those with low expression levels (Fig. 5b, c). Details of overall survival data of 7 patients with ccRCC are not available from the data server. Therefore, the overall survival graphs were constructed by reducing the sample size from 523 to 516 patients with ccRCC.

**DISCUSSION**

EMT plays a critical role in tumorigenesis (5, 6, 20). EDIL3, a potent stimulator of EMT, has been shown to have important roles in various biological mechanisms (21). Targeting this small mediator may be a novel therapeutic strategy for antitumor treatment in renal cancer. Studies on the prognostic significance of EDIL3 and its role in EMT are limited. A study evaluated the prognostic significance of EDIL3 expression and demonstrated the relationship between EDIL3 and EMT in lung adenocarcinoma (22). Our in silico analysis results support this data. EDIL3-high positive patients had lower survival rates than did EDIL3-low positive or negative patients. This status may indicate the prognostic significance of EDIL3 expression. Furthermore, our findings suggest that the increased expression of EDIL3 in malignant tissue during the transition from epithelial to invasive mesenchymal features is associated with the EMT phenotype. EDIL3 expression correlated with levels of CDH1 and VIM expression. Based on these findings, EDIL3 is an important target for the treatment of ccRCC. A study confirmed that EDIL3 expression can support characteristics of EMT, in terms of decreasing E-cadherin and increasing vimentin expression in patients with lung cancer (22).

Although an association between EDIL3 expression and the EMT program has been observed in several cancer types (11, 21, 22), no study has examined mutational profiles of EDIL3, VIM, and CDH1 and their expression patterns in patients with ccRCC. First, we evaluated mutational profiles of EDIL3, VIM, and CDH1 from genome sequencing data, which are available in TGCA datasets, of 523 patients with ccRCC. The mutation ratio was 1.9% in the
ccRCC patient group, with EDIL3 being the most mutated gene at 1.3%. Mutations were detected in the genes studied, especially in sequences encoding important domains. Results from evolutionary analysis across species revealed that p.Q187R, p.T343M, and p.L380F missense mutation in VIM were evolutionarily conserved. In addition, according to functional potential pathogenic effect analysis, we determined that it may cause the formation of impaired or unstable EDIL3 and VIM. Mutations in the three genes were recorded as somatic mutations in the COSMIC database. The p.G185* nonsense mutation in EDIL3 is a truncating mutation, which leads to the creation of a stop codon. We believe that this mutation is of clinical significance. We detected p.I289Hfs*27 and p.P127Afs*41 frame-shift mutations in EDIL3 and CDH1, respectively. Studies have reported that frame-shift mutations in CDH1 lead to the development of diffuse stomach cancer (23, 24). E-cadherin, which is encoded by CDH1, is a glycoprotein found in the epithelium of all mammals. E-cadherin belongs to the family of the cell adhesion molecules and is the first identified member of this family. The intracellular domain of E-cadherin consists of 151 amino acids and is linked to the intracellular actin skeleton through α-, β-, and γ-catenin. The extracellular domain of E-cadherin comprises 554 amino acids and communicates with molecules of neighboring E-cadherin (23, 24, 25). The p.P127Afs*41 mutation arises when incomplete synthesis of the extracellular portion of the protein occurs. We believe that this causes protein function disruption. Therefore, the bioinformatics-based design of our study is the limitation of our study.

Results from bioinformatics analysis should be considered in the light of some limitations. The limitation concerns the lack of experimental study. Demographic statistics, including gender, race, age, and survival status, of healthy individuals were not available in the database, which is a statistically limiting factor in our bioinformatics study. However, our study has some strengths. Systematic cancer genomic projects, including TCGA, enabled the analysis of specific tumor types. Identification of disease-specific loci has led to the discovery of oncogenes, identification of molecular subtypes of tumors, and discovery of biomarkers based on transcriptomic, proteomics, and epigenomic changes. Studies on mutations in regions can provide valuable information.

In conclusion, our findings suggest that EDIL3 is a novel molecule associated with the EMT phenotype and can be a novel target in ccRCC. A large-scale wet-lab study is warranted to elucidate the roles of EDIL3, VIM, and CAD in ccRCC for discovering drugs for renal cancer.

Acknowledgements: The data used in our study are obtained from public database the TCGA Research Network: https://www.cancer.gov/tcga. We thank the TCGA and GEPIA databases for the availability of the data.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – RIS; Design – RIS, DFAB; Supervision – RIS, DFAB; Data Collection and/or Processing – RIS, DFAB; Analysis and/or Interpretation – RIS, DFAB; Literature Search – RIS, DFAB; Writing – RIS, DFAB; Critical Reviews – RIS.
Figure 4. Validation of mRNA expression levels of (a) epidermal growth factor-like repeats and discoidin I-like domains 3 (EDIL3), vimentin (VIM), and cadherin-1 (CDH1) in clear cell renal cell carcinoma (ccRCC) tissues and normal tissues using GEPIA. The four box plots are based on 523 ccRCC samples (red) and 100 normal samples (black). KIRC: Kidney renal clear cell carcinoma. (b) Pearson’s correlation analyses of VIM and CDH1 mRNA expression with that of EDIL3 (*indicates p<0.05)

Figure 5. Kaplan–Meier survival analysis conducted with high and low mRNA expression of (a) epidermal growth factor-like repeats and discoidin I-like domains 3 (EDIL3), (b) vimentin (VIM), and (c) cadherin-1 (CDH1) regarding their association with overall survival in 523 patients with clear cell renal cell carcinoma (ccRCC). Data are presented as hazard ratio with 95% confidence interval. Logrank p<0.01 was considered statistically significant. (n: Sample size; logrank: Mantel–Cox test; HR (high): Hazard Ratio; p-value (hazard ratio))

TPM: Transcripts per million
REFERENCES


