Computational Prediction of Tumor-Specific Antigens as Potential Vaccine Candidates against Germ-line Mutations in Endometrial Cancer

Iqra Iftikhar, Aqsa Khalid*, Zainab Bibi, Azhar Mehmood, Muhammad Rizwan, Sajid Khan, Anum Munir

Department of Bioinformatics, Govt. Postgraduate College, Mandian Abbottabad, Pakistan

Received August 19, 2019; Revised September 16, 2019; Accepted September 23, 2019

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract

Endometrial cancer is the fourth most common cancer in women. It arises from the endometrium and accompanied by the abnormal growth of the cells. Signs and symptoms include pelvic pain and abnormal vaginal bleeding. It has two categories. Type 1 tumors are estrogen-dependent and have mutations in PTEN, PIK3CA while Type 2 tumors are more sensitive and have mutations in TP53. Overactivation of the signaling pathway (PI3K) results in anti-apoptosis. Here, this study aims to identify Tumor-Specific Antigen for germline mutations in endometrial cancer which can be used as a potential vaccine candidate. The germline mutations data are obtained from cancer gene census of the cosmic database. Genes mutating with crucial role in endometrial cancer are considered. Peptides libraries are generated using peptide design library. Human leukocyte antigen alleles are identified for the peptide library through NetMHC. Binding affinities of alleles with peptide are determined. Linear regression is performed to generate graphs. PTEN, TP53, PIK3CA, KRAS, and CTNNB1 proved to have critical role. About 575 overlapping peptide libraries are generated and each peptide has a length of 18-20 amino acids. Approximately 58 HLA s are identified, having strong interactions with HLAs. Regression analysis shows that the no. of mutations are directly associated with a binding affinity of peptides. From this, we suggest that the identified TSA can be used as personalized peptide vaccines that directly target the mutated genes in endometrial cancer. This research work can be used in the laboratories for further validation.

Keywords

HLAS, Peptide Vaccine, Endometrial Cancer, Germline Mutations

1. Introduction

Endometrial cancer (EC) is the sixth most commonly diagnosed cancer among women worldwide, arises from the endometrium and accompanied by the abnormal growth of the cell [1, 2]. The uterus is the hollow, pear-shaped pelvic organ in women where fetal development occurs. Endometrial cancer begins in the layer of cells that form the lining (endometrium) of the uterus, sometimes also called uterine cancer, results from the abnormal growth of cells with the ability to invade or spread to other parts of the body. Endometrial cancer occurs most commonly after menopause [3]. Most EC is low grade (G1 or G2) tumor that is diagnosed at early stages [1, 4].

It is fourth most common gynecological malignancy in Europe and in the United State. More than 280000 women are diagnosed every year and 74000 women died from it annually, second-most gynecological cancer in Taiwan [5]. According to the cancer statistics, 2005 and extent of 40880 new cases were diagnosed whereas the mortality rates are 7310 [1]. In 2008 estimated cases of endometrial cancer and the mortality chances are 49560 and 8190 respectively. Most recent data on the incidence rate and death rates lies in the range of 60050 and 10470 independently according to the report of cancer journal [6]. In 2013 approximately 49560 new cases of endometrial cancer and 8190 related deaths are predicted [7].

Endometrial cancer is broadly divided into two groups. Type I called endometrioid carcinomas and type II called uterine carcinoma about 80-90% of EC are endometrioid carcinoma and 2-10% are serious carcinomas [5, 8]. Type I are associated with loss of estrogen/progesterone receptor absence of unopposed estrogen reduced E cadherin expression and mutation in P53, type II carcinoma exhibit extrauterine spread at the time of surgery[7, 9]. Although,
type II is in minority and it is much more aggressive than Type I and has a poor outcome [5].

There are different stages of endometrial cancer. Stage 0: the abnormal cells are found only on the surface of the inner lining of the uterus, stage I: the tumor has grown through the inner lining of the uterus to the endometrium, it may have perfused the myometrium, stage II: the tumor has perfused the cervix, stage III: the tumor has grown through the uterus to reach nearby tissues, such as the vagina, and stage IV: the tumor has invaded the bladder or intestine [10, 11].

Signs and the symptoms include abnormal vaginal bleeding, pain or difficulty when emptying the bladder, discharge of a watery, blood-streaked flow that gradually contains more blood and pain in the pelvic area and during the intercourse [3].

For the detection of uterine cancers cytological and histological methods were considered the best for screening and the earlier diagnosis of the uterine cancers. Squamous cell carcinoma-associated antigen (SCC) is relatively tumor-specific and widely used for monitoring patients with squamous cell carcinoma. There is no particular tumor marker for uterine corpus carcinoma, several tumor markers incorporations including cancer antigen 125 (CA125) may be of greater diagnostic value in cases of uterine corpus carcinoma [12]. There is no specific tumor marker for endometrial cancer, and the investigation of HLA [13] class 1 in endometrial cancer is very limited. This scientific study reports the identification of tumor-specific antigen for the germline mutations in endometrial cancer (uterine corpus carcinoma) that will be of great prognostic value and are considered the potential targets for the anticancer treatment.

2. Materials and Methods

To identify tumor Specific Antigen for germline mutations in endometrial cancer, a step by step methodology was applied shown in Fig.1.
2.1. Screening and Selection of Genes

Genes containing mutations that have been implicated in endometrial cancer from cancer gene census of the COSMIC database were cataloged. COSMIC database curates cancer data on somatic mutation. There were 23986 genes having mutations in endometrial cancer, top five genes with high mutations were selected and identified as: PTEN, PIK3CA, KRAS, CTNNB1, and TP53 [14].

2.2. Peptide Library Generation

Peptide library was generated using the Peptide Design Library. As Peptide Design Library generate overlapping library by breaking the original protein into many equal length overlapping fragments. The protein sequences of those genes were downloaded from UniProt. The peptide length and amino acid overlap were chosen as 18 and 16 respectively. The peptides having hydrophobic composition greater than equal to 50% were selected [15].

2.3. Identification of Human Leukocyte Antigen

Human leukocyte antigen is the complex genes that encode major histocompatibility complex (MHC) proteins. In humans, these proteins are responsible for the control of the immune system. Down-regulation of human leukocyte antigen class I has a vital role in several types of cancer [10]. Human leukocyte antigen alleles were identified using MHCpred, it covers a wide range of human MHC allele peptide specificity model [16]. For HLAs prediction, the sequence of a protein was entered, MHC allele and other parameters were selected, HLAs having IC50 less than 200 were identified for MHC alleles (HLA-A*0201 and HLA-A*0101). Binding affinities of alleles with peptides were determined [13].

2.4. Statistical Analysis

Descriptive analysis has been performed to evaluate expression dependency of the variables i.e. amino acids, confidence level and logIC50 on one another. Pearson correlation was used for evaluating the relationship between an amino acid and their logIC50 based on R-value. R-value should be closed to 1 or -1 [17]. We have performed regression linear modeling to determine the association between amino acids sequence and their log IC50 in our research work. Amino acids were taken as the dependent variable and log IC50 as the independent variable. Linear Modeling was performed to determine the lead peptide as a vaccine target on the basis of a maximum confidence level of prediction [18].

2.5. Protein Peptide Docking

After the recognition of lead peptide through linear modeling, the protein (HLA) peptide docking was done using GalaxyPepDock. GalaxyPepDock is freely accessible [19]. For this purpose, the HLA was downloaded from Protein databank in PDB format. Protein Data Bank is an archive for macromolecular structure [20]. The endometrial protein 4MZR having highest mutations was obtained, the protein preparation was done in Discovery studio by removing non-standard residues, hetero-atoms, water molecules and docking was done to get the structural protein-peptide complexes [21].

3. Results

In this research work, the list of genes involved in endometrial cancer was retrieved. There were 23986 genes having mutations in endometrial cancer. Top five genes with high mutations were selected and identified as germline mutations are: PTEN, PIK3CA, KRAS, CTNNB1, and TP53. Mutated genes are listed in Table 1.

<table>
<thead>
<tr>
<th>Gene Names</th>
<th>Description</th>
<th>Function</th>
<th>Chromosomal Location</th>
<th>Mutated Sample</th>
<th>Samples Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTEN</td>
<td>Phosphatase And Tensin Homolog</td>
<td>Tumor suppressor protein mutated in large no of cancers.</td>
<td>10q 23.31</td>
<td>1260</td>
<td>3311</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha</td>
<td>Provides instructions for making p110 alpha.</td>
<td>3q 26.32</td>
<td>949</td>
<td>3980</td>
</tr>
<tr>
<td>KRAS</td>
<td>KRAS Proto-Oncogene, GTPase</td>
<td>Mutation in KRAS frequently observed i.e. a single substitution results in mutation</td>
<td>12 p 12.1</td>
<td>599</td>
<td>4022</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>Catenin Beta 1</td>
<td>It causes the cell to stop dividing.</td>
<td>3p 22.1</td>
<td>438</td>
<td>2416</td>
</tr>
<tr>
<td>TP53</td>
<td>Tumor Protein P53</td>
<td>It induces cell cycle arrest and apoptosis.</td>
<td>17p 13.1</td>
<td>382</td>
<td>1998</td>
</tr>
</tbody>
</table>

Table 1. Genes with Critical Role in Endometrial Cancer
Overlapping peptides are essential to decide which part of the protein contains essential amino acid and they are characterized by the parameters i.e. peptide length and amino acid overlap. Therefore, the peptide length was taken as 18-20 amino acids whereas amino acid overlap was taken 16. Total 1676 peptide libraries were generated for the above mentioned mutated genes and 575 peptides based on parameter i.e. Hydrophobicity composition greater than or equal to 50 were screened out.

3.1. Prediction of Peptide Binding Affinity with MHCPred 2.0

After submission of the overlapping peptides to MHCPred, binding affinities were checked for the MHC alleles (HLA-A-*0101 and HLA-A-*0201). 898 HLA s for MHC alleles were identified. For HLA-A-*0101 the count was 170 and for HLA-A-*0201 the count was 728. Ability to predict binding affinities will aid in determining the most reputed vaccines.

For the strong association between the variables, the confidence of prediction should be equal to 1 i.e. maximum value. For HLA-A-*0201 the confidence level of prediction was .695 nearest to the standard value, similarly for HLA-A-*0101 the confidence level of prediction was .963 again closest to the maximum level of prediction.

### Table 2. Pearson 2-tailed test for HLA-A-*0201 and HLA-A-*0101

<table>
<thead>
<tr>
<th>Amino acid groups</th>
<th>Statistical Test</th>
<th>LogIC50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson correlation</td>
<td>Sig (2-tailed)</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.022</td>
<td>0.777</td>
</tr>
</tbody>
</table>

For analyzing the differences among the variances Anova Test was performed. In regression analysis, the degree of freedom for HLA-A-*0201 and for HLA-A-*0101 was 1. Whereas Mean square for HLA-A-*0201 was 525682.514 and for HLA-A-*0101 was 166.551.

### Table 3. Analysis of Variance for HLA-A-*0201 and HLA-A-*0101

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>525682.514</td>
<td>1</td>
<td>525682.514</td>
<td>12.336</td>
<td>.000b</td>
</tr>
<tr>
<td>Residual</td>
<td>30937521.50</td>
<td>726</td>
<td>42613.666</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31463204.02</td>
<td>727</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>166.551</td>
<td>1</td>
<td>166.551</td>
<td>.081</td>
<td>.777b</td>
</tr>
<tr>
<td>Residual</td>
<td>346882.061</td>
<td>128</td>
<td>2064.774</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>347048.612</td>
<td>169</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2. Mutation Distribution for HLA-A-*0201 and HLA-A-*0101

On the basis of linear modeling, it was concluded that out of among 728 HLA’S, one HLA (LLLSVLLSV) was having the maximum confidence level of prediction for HLA-A-*0201. The graph in Fig.3 shows there are two HLA’S (GSDDINVT, TTDCLQILA) having the maximum confidence level for HLA-A-*0101.

![Figure 2. Linear Modeling for HLA-A-*0201](image-url)
3.3. Identified Tumor-Specific Antigen

The identified TSAs are mentioned in table 4:

<table>
<thead>
<tr>
<th>TSA</th>
<th>Maximum Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSDDINVVT</td>
<td>2.62</td>
</tr>
<tr>
<td>TTDCLQILA</td>
<td>2.62</td>
</tr>
<tr>
<td>LLLSVLLSV</td>
<td>1</td>
</tr>
</tbody>
</table>

The three HLAs have been identified as tumor-specific antigens based on maximum confidence. These results indicated that untreated endometrial cancer cells acquired a wide range of germline mutations in PIK3CA, KRAS, PTEN, TP53, and CTNNB1. The identified TSAs can be a prognostic marker for the identification of endometrial cancer and can be used as drugs target.

3.4. Docking Analysis

A computational peptide-protein docking was used to analyze structural complexes of 4MZR with peptides to understand the structural basis. The docking analysis results are shown in Fig.4.

3.5. Interacting Residues

Interacting residues in each docked complex are shown in Fig.5. These residues illustrate there is maximum interaction of the peptides with the receptor proteins.
4. Discussion

In this scientific study, the mutated peptides in 23985 genes were examined and checked for the binding affinity of those mutated amino acids to MHC Alleles (HLA-A-0101 and for HLA-A-0201) computationally. Mutated data was taken from cancer gene census of the cosmic database [22]. 23985 genes were having germ-line mutations in endometrial cancer and top 5 genes were given consideration on account of having the highest score.

Mutated peptides are of great significant value and they act as a potential immunological target for the cancer therapy [23]. Therefore, those mutated peptides were selected for further analysis [25]. After the submission of those peptides to MHCPRED 2.0, MHC alleles (HLA-A-0201 and HLA-A-0101) were selected to check binding affinity. The screening was performed based on the parameter i.e. IC50 value. Only those HLAs were screened out with the inhibitory concentration (IC50) less than 200.

For comparison and for the identification of Tumor-specific antigen (TSA) among the total count of 898 HLAs statistical analysis was performed. To determine the strength of association between variables correlation analysis (r) between dependent parameter i.e. a Confidence level of residues and the independent parameter i.e. amino acid was performed which revealed that r value was positive and was showing the strong interaction/association. Based on linear modeling, we could suggest there were only 3 HLAs from the list of 898 having the maximum confidence level. Hence they could be referred to as best TSA against endometrial cancer.

The TSAs were also validated through docking and the result shows that the peptide ligands bind best with the active pocket of the protein. A computational peptide ligand-protein docking was used to analyze and validate the structural complex of receptor protein (target) with peptide-ligands. Our results show that receptor-ligand (peptide) shows strong interaction and the identified TSAs bind best with the active pocket of 4MZR protein. Estimated accuracy for predicted protein-peptide complex of GSDDINVVT, LLLSVLLSV and TTDCLQILA are
0.305, 0.357 and 0.294 respectively.

Figure 5 shows the potential binding sites of TSAs with target protein residues. Maximum of amino-acid residues bind with the identified TSAs, which show strong association [24].

5. Conclusions

In this study, an Insilco approach is formulated for the identification of tumor-specific antigen. This research work provides the best framework for the analysis of those mutated peptides and their binding with HLAs which could be a great significance towards developing personalized peptide vaccines that could directly mark the mutated targets in cancer therapy. This could have a superior effect in mutating target cell population [25]. These identified TSAs (GSDDINVVT, TTDCLQILA, and LLLSVLLSV) are of great prognostic value. In future, results of this study require experimental validation for implementing these personalized vaccines as it was an Insilco work.

6. Abbreviations

CTNNB1: Catenin beta-1
COSMIC: Catalogue of Somatic Mutations in Cancer
EC: Endometrial Cancer
HLA: Human Leukocyte Antigens
IC: Inhibitory Concentration
MHC: Major Histocompatibility Complex
PI3KCA: Phosphatidylinositol 3-kinase
PTEN: Phosphatase And Tensin Homolog

Acknowledgement

The author’s thanks to the Department of Bioinformatics Government Post Graduate College Mandian Abbottabad for conducting this research work. The authors also thank Professor Ghulam Rasool and Professor Amjad Ali Shah for their Cooperation.

REFERENCES


