ValidNESs: Validated NES-containing proteins, functional NES sites and NES predictions

Contents

1. About ValidNESs
2. Data access
   2.1 Retrieve data from table of NES sites
      2.1.1 Pre-classified categories of NESs
      2.1.2 Supported regular expression metacharacters
   2.2 Retrieve data from table of NES-containing proteins
3. NES prediction
4. Questions, Comments, Problems

1. About ValidNESs

ValidNESs is an integration of up-to-date data and the state-of-the-art predictor of NES. It was developed in easy-to-use and searchable mode, allowing users to retrieve particular data of interest. In addition, ValidNESs provides an online prediction tool implemented by NESsential, the state-of-the-art predictor of the classical NES at both protein and site level.
2. Data access

2.1 Retrieve data from table of NES sites

ValidNESs provides two search functions for retrieving particular NESs of interest. The “search-by-category” function is an intuitive drop-down list allowing users to select and display by pre-classified category, classical or non-classical NESs. The “search-by-pattern” function, on the other hand, is a function with regular expression support allowing users to retrieve all NES sites containing specific pattern.

Step 1 - Click on the “NES-containing region” button on the menu bar

<table>
<thead>
<tr>
<th>ValidNESs:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Validated NES-containing proteins, functional NES sites and NES predictions</strong></td>
</tr>
<tr>
<td><strong>Home</strong></td>
</tr>
</tbody>
</table>

**Display by category**

Choose type of NESs from the drop-down list

- Classical NESs if the experimental validated region overlaps with a short NES motif, otherwise Non-classical NESs.

Type of NESs: A drop-down list to select and display NES sites by different type of NESs

**Search by pattern**

Try L[A-Z][2-9][A-Z][A-Z] to retrieve NES sites containing this pattern: L2[3][A-Z].

A search textbox to input an inquiry sequence pattern

Step 2 - Choose search function (ValidNESs display all entries by default)

Pattern: L[A-Z][2-9][A-Z][A-Z]

Total 72 NES sites

Download data in text file format

Search result in text file format is available for download after clicking the hyperlink “text file”

**Experimentally validated NES sites**

Search result in an online simplified table

*Each combination that impacts NES activity either alone or together are marked in red

<table>
<thead>
<tr>
<th>ACC</th>
<th>SEQUENCE AND POSITION OF NES-CONTAINING REGION</th>
<th>REFERENCE DDS (PDB)</th>
<th>REFERENCE DDS (PDB)</th>
<th>VALIDATED BY LEPTOSTOMON TREATMENT</th>
<th>REFERENCE DDS (PDB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S002</td>
<td>350-1400[A-Z][A-Z][1-80]-202</td>
<td>104627</td>
<td>1D12</td>
<td>No</td>
<td>10875108 11307949 11847229</td>
</tr>
<tr>
<td>S004</td>
<td>137-1801SIL, SIL, T1-0-146</td>
<td>014010</td>
<td>-</td>
<td>No</td>
<td>11205759</td>
</tr>
<tr>
<td>S005</td>
<td>1097IL[SIL][A-Z][A-Z][1-80]-451</td>
<td>009974-18</td>
<td>-</td>
<td>Yes</td>
<td>10330933</td>
</tr>
<tr>
<td>S011</td>
<td>51-415IL[A-Z][A-Z][1-80]-33</td>
<td>105806</td>
<td>-</td>
<td>Yes</td>
<td>75628891 18798609</td>
</tr>
</tbody>
</table>

Cross-links to external databases and the original papers

Cross-links to more information at the site and protein level
2.1.1 Pre-classified categories of NESs

Some previous work has defined a consensus sequence for NESs as [LIVFM]-x-(2,3)-[LIVFM]-x(2,3)-[LIVFM]-x-[LIVFM] where x is any amino acid. However, we found that 43% of NESs in ValidNESs deviate from this consensus sequence. We therefore defined a short consensus pattern [LIVFM]-x(2,3)-[LIVFM]-x-[LIVFM], hereafter denoted as the “NES motif”, containing the region bounded by the second and fourth hydrophobic positions of the former consensus, a region which has been shown to affect NES activity strongly. In ValidNESs, we use this generalized consensus pattern to divide experimentally determined NES sites into two categories: classical if the experimental validated region contains or overlaps with a consensus match, otherwise non-classical. This definition of classical NES is justified by the dramatically improved sensitivity (from 57% to 86%).

2.1.1 Supported regular expression metacharacters

<table>
<thead>
<tr>
<th>Metacharacter</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ and ]</td>
<td>Match anything inside the square brackets for ONE character position once and only once, for example, [LI] means match the target to “L” and if that does not match then match the target to “I”; [A-Z] means match to any character in the range A to Z.</td>
</tr>
<tr>
<td>{ n }</td>
<td>Matches the preceding character n times exactly</td>
</tr>
<tr>
<td>{ n, m }</td>
<td>Matches the preceding character at least n times but not more than m times.</td>
</tr>
</tbody>
</table>

e.g. Search by the pattern [LIVFM][A-Z]{2}[LIVFM][A-Z][LIVFM] and retrieve

38-LALKLAGLDI-47  
137-GIDLSGLTLQ-146  
835-ESLEEEELDVLVDDEGG-851

e.g. Search by the pattern [LIVFM][A-Z]{3}[LIVFM][A-Z][LIVFM] and retrieve

339-EMFRELNEALELKD-352  
188-SISLSFDESALCVP-202  
835-ESLEEEELDVLVDDEGG-851

[Note] the underlined region denotes the matching region
2.2 Retrieve data from table of NES-containing proteins

ValidNESs provides “search-by-keyword” functions for retrieving particular NES-containing proteins of interest. Again, the “search-by-category” function is an intuitive drop-down list allowing users to select and display by UniProtKB keywords.

**Step 1** - Click on the “NES-containing protein” button on the menu bar

![ValidNESs: Validated NES-containing proteins, functional NES sites and NES predictions](image)

**Step 2** - Choose search function (ValidNESs display all entries by default)

![A drop-down list to select and display NES-containing proteins by different UniProtKB keywords](image)

**Step 3** - Retrieve search result

![The search result in text file format and protein sequence with site information is available for download after clicking the hyperlink “text file” and “FASTA” respectively](image)

![Proteins containing experimentally validated NESs](image)

---

**Cross-links to more information at the site and protein level**
3. NES prediction

ValidNESs provides online prediction of NES based on NESsential, the state-of-the-art predictor of NES. After successful submission and processing, users can find the prediction results, at both protein and the site level, and an easy explanation about how to interpret them.

**Step 1** - Click on the “NES prediction” button on the menu bar

**Step 2** – Input the UniProt ID or the sequence of your protein

**Step 3** - Retrieve search result

![ValidNESs interface](image)

*NEESential score at the protein level indicates how likely the given protein contains a NES and ranges from 0 to 1.
*NEESential score of each putative site indicates the probability to be a functional NES and ranges from 0 to 1.
4. Questions, Comments, Problems

E-mail: szuchin.fu@gmail.com