



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



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

Home NES-containing regions NES-containing proteins NES prediction

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

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 [More details in documentation](#)

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]{2,3}L[A-Z]L` to retrieve NES sites containing this pattern: `Lx{2,3}LxL`
x can be any amino acid while the spacing between 1st and 2nd leucine can be 2 or 3
[More details in documentation](#)

`L[A-Z]{2,3}L[A-Z]L` [A search textbox to input an inquiry sequence pattern](#)

submit

Step 3 - Retrieve search result

Data retrieved [Search result](#)

Pattern: `L[A-Z]{2,3}L[A-Z]L`

Total 73 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](#)

*point-mutations that impair NES activity either alone or together are marked in red

ACC	SEQUENCE AND POSITION OF NES-CONTAINING REGION	REFERENCE DB (UniProt)	REFERENCE DB (PDB)	VALIDATED BY LEPTOMYCIN TREATMENT	REFERENCE (PMID)
S002	339-EMFRELNEALEKD-352	P04637	1A1E	Yes	10075936 11397945 11847229
S004	137-GIDL SGLTLQ-146	Q9W639	-	Yes	11265759
S005	835-ESLEEELDVLVDDEGG-851	O60716-18	-	Yes	10393933
S011	75-LPPIERLTL-83	P05865	-	Yes	7543368 10739668

[Cross-links to more information at the site and protein level](#) [Cross-links to external databases and the original papers](#)

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

Metacharacter	Meaning
[and]	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.
{ n }	Matches the preceding character n times exactly
{ n, m }	Matches the preceding character at least n times but not more than m times.

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

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

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

339-EMFRELNEALELKD-352
188-SISLSFDESLALCVI-202
835-ESLEEELDVLVLDDEGG-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



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



Step 3 - Retrieve search result



Proteins containing experimentally validated NESs ← Search result in an online simplified table

ACC	PROTEIN NAME (GENE NAME)	ORGANISM	REFERENCE DB (UniProt)	SITES
P002	Cellular tumor antigen p53 (TP53)	Homo sapiens (Human)	P04637	339-352
P007	Tyrosine-protein kinase ABL1 (Ab11)	Mus musculus (Mouse)	P00520	1078-1096
P031	Tumor protein p73 (TP73)	Homo sapiens (Human)	015350	384-382

↓
Cross-links to external database

```

*** SITE INFORMATION ***
NES1
38-LALQLASLDI-47
38-####HT 8-47

*** PROTEIN INFORMATION ***
Name
cAMP-dependent protein kinase inhibitor alpha
Organism
Homo sapiens (Human).
Subcellular localization

Reference databases
UniProt:P01925; InAct:P01925;

Sequence
MIDNETTYAD FIASQRTGRR NAIHDILVSS ASDNSNE[24] [25][26]KT EGEEDAGRSS TEQSGEADGE AAKSES
    
```

← 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



ValidNESs:

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



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

Input the UniProt ID

e.g. IPKA_HUMAN

IPKA_HUMAN

submit

Or protein sequence in one-letter code

e.g. MTDVETTYADFIASGRTGRRNAIHDILVSSASGNSNELALKLAGLDINKTEGEEDAQRSSTEQSGEAQGEAAKSES

MTDVETTYADFIASGRTGRRNAIHDILVSSASGNSNELALKLAGLDINKTEGEEDAQRSSTEQSGEAQGEAAKSES

Text input area for protein sequence

submit

Step 3 - Retrieve search result

Submit successfully!!!

The prediction will be done in several ten minutes, please go [Result](#) and bookmark the page

↓ [Link to the result page](#)

Name:
IPKA_HUMAN

Prediction at the protein level:
0.66

Prediction at the site level:

Putative site (Probability)
42-LAGLDI-47 (0.93)
23-IHDILV-28 (0.06)

An easy explanation about how to interpret the prediction result

↓

*NESsential score at the protein level indicates how likely the given protein contains a NES and ranges from 0 to 1
*NESsential 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