🌐 ⬚ A Computational Method for detecting and evaluating Tankyrase-Binding Motifs

Christopher M. Clements[1], Samantha X Shellman[2], Melody H Shellman[3], Yiqun G. Shellman[1,4]

[1]Department of Dermatology, University of Colorado Anshutz Medical Campus, School of Medicine, Aurora, CO 80045;
[2]Department of Computer Science, University of Colorado, Boulder, CO 80309;
[3]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30309;
[4]Charles C. Gates Regenerative Medicine and Stem Cell Biology Institute, University of Colorado Anschutz Medical Campus, School of Medicine, Aurora, CO 80045

◯ Yiqun Shellman

Protocol status: Working
We use this protocol and it's working
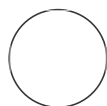
ABSTRACT

Tankyrases are multifunctional proteins within the poly(ADP-ribose) polymerase family. Known tankyrase binders primarily interact with the scaffolding portion of tankyrases, which comprises five ankyrin repeat cluster (ARC) domains. These domains recognize a specific sequence known as the Tankyrase binding motif (TBM), typically following an octapeptide format characterized by an arginine at position 1 and a glycine at position 6. However, extended and nonconventional TBMs have also been reported. At present, there is no system in place that can easily find and score TBMs. This protocol describes how to use a web-based, public-accessible, computational method (https://shellmanlab.github.io/) we developed to locate and rank all types of potential TBMs. As interest in tankyrases continues to grow across various biological fields, our tool empowers researchers to quickly assess how tankyrases may impact their protein of interest.

BEFORE START INSTRUCTIONS

A computer with access to the Internet will be required. Internet access to both https://www.uniprot.org/ and https://shellmanlab.github.io/ will be necessary.

Finding and Scoring Potential Canonical and Extended TBMs

1   Identify the Uniprot code for your protein of interest in the Uniprot database (https://www.uniprot.org/), which is the identifier indicated by the arrow in Figure 1.

Figure 1. Finding the Uniprot code for your protein of interest.

2   Enter Uniprot code(s), one per line, to the text box of the application at https://shellmanlab.github.io/ (Figure 2). For screening multiple proteins simultaneously, a list of Uniprot codes in a column of an Excel spreadsheet can also be copied and pasted into the text box.

Tankyrase Binding Motif (TBM) Search Project

Enter UniProt code(s) below, one code per line:

```
Q9NTX7
P78314
```

[Find and Score Motifs!] 🛈   [Clear]

Input one sequence of 8 amino acids to be scored:

[Score any 8-amino-acid sequence] 🛈   [Clear]

Note: Scores are a predication and do not replace experimental data.

Figure 2. Enter the Uniprot codes, one per line.

3   "Find and Score Motifs!" function will search the sequence of all listed proteins for potential canonical and extended TBMs and provide a score of binding strength for each TBM.

Tankyrase Binding Motif (TBM) Search Project

Enter UniProt code(s) below, one code per line:

```
Q9NTX7
P78314
```

Find and Score Motifs! 🛈   Clear

Input one sequence of 8 amino acids to be scored:

Score any 8-amino-acid sequence 🛈   Clear

Note: Scores are a predication and do not replace experimental data.

Figure 3. The function of "Find and Score Motifs!".

4   The "Find and Score Motifs!" function will output the predicted canonical and then extended TBMs, with one TBM in one row. Each row will include the protein name, Uniprot code, amino acid sequence of the TBM, the starting position of the TBM in the protein, and the score of the binding strength.

Tankyrase Binding Motif (TBM) Search Project

Enter UniProt code(s) below, one code per line:

```
Q9NTX7
P78314
```

[Find and Score Motifs!] 🛈   [Clear]

Input one sequence of 8 amino acids to be scored:

[Score any 8-amino-acid sequence] 🛈   [Clear]

Note: Scores are a predication and do not replace experimental data.

Canonical TBMs: R-x-x-x-x-G-(No Proline)-x:
Protein,Uniprot Code,Motif,Position,Score
RN146_HUMAN,Q9NTX7,RRNEHGRR,156,0.28303108790000003
RN146_HUMAN,Q9NTX7,RSDRSGTD,325,0.366943005
RN146_HUMAN,Q9NTX7,RSVAGGGT,333,0.4537564766
3BP2_HUMAN,P78314,RSPPDGQS,415,0.8031606216999999

Extended TBMs with 5, 6, or 7 amino acids between R and G:
Protein,Uniprot Code,Motif,Position,Score
RN146_HUMAN,Q9NTX7,RESSADGAD,194,0.5436639119
RN146_HUMAN,Q9NTX7,RPLTSVDGQL,220,0.3378512397
RN146_HUMAN,Q9NTX7,RSHRGEGEE,260,0.5874655647
RN146_HUMAN,Q9NTX7,RSVAGGGTV,333,0.3502754821
RN146_HUMAN,Q9NTX7,RSRRPDGQC,346,0.5328374655
3BP2_HUMAN,P78314,RAHSFTSKGPG,222,-0.12898071609999998
3BP2_HUMAN,P78314,REAAMPGLF,369,0.013415978000000009
3BP2_HUMAN,P78314,RQPSQADTGGD,432,0.060330578600000004
3BP2_HUMAN,P78314,RGEPQDGLY,477,0.3258126722
3BP2_HUMAN,P78314,RNSSTKSGKV,488,0.018787878900000006
3BP2_HUMAN,P78314,RHPYGYTGPR,552,-0.13493112929999998

Figure 4. An example of output of "Find and Score Motifs!".

4.1   Convert output into an Excel file with delimited by commas (Figure 5-7). Copy and paste the output into an Excel file. Go to the "Data" tab (A) and select "Text to Columns" (B). Then make sure "delimited" is selected in the pop-up window and click "Next" (C).
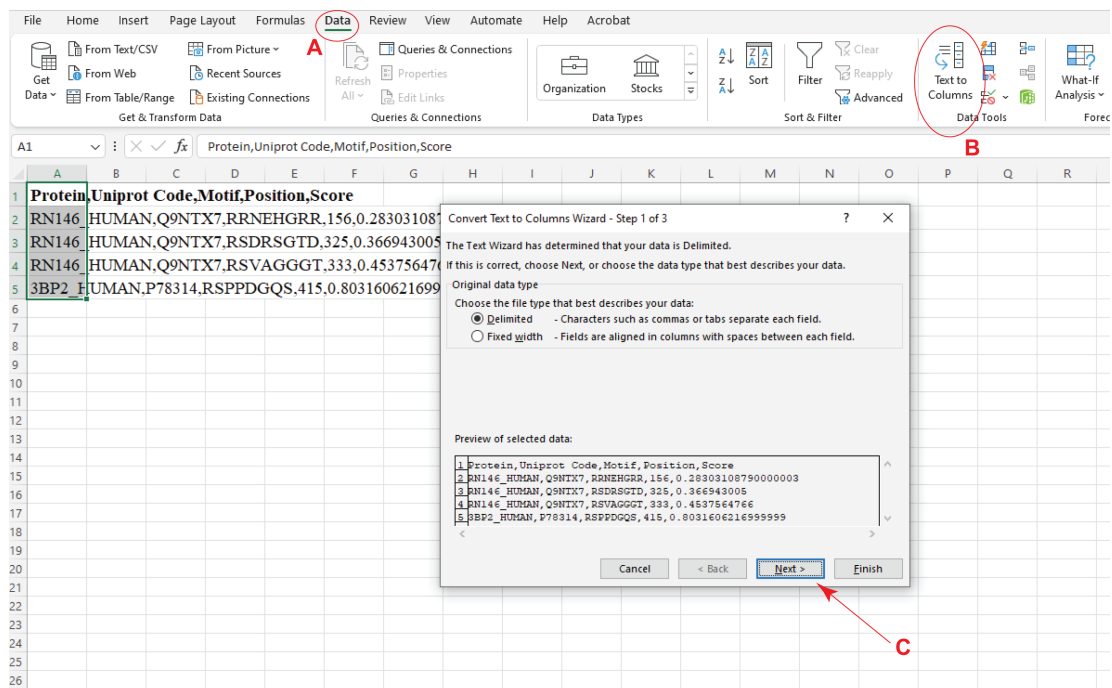
Figure 5. Paste the output into a spreadsheet of an Excel file.

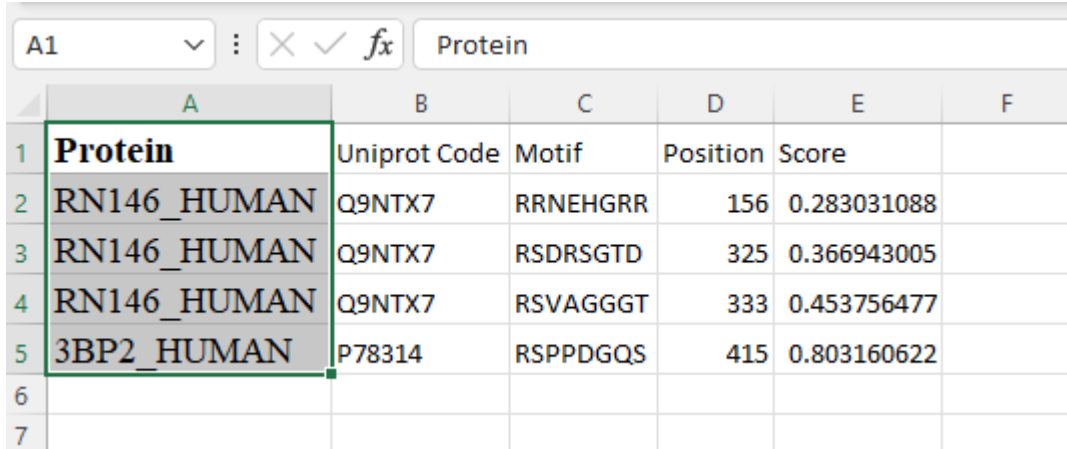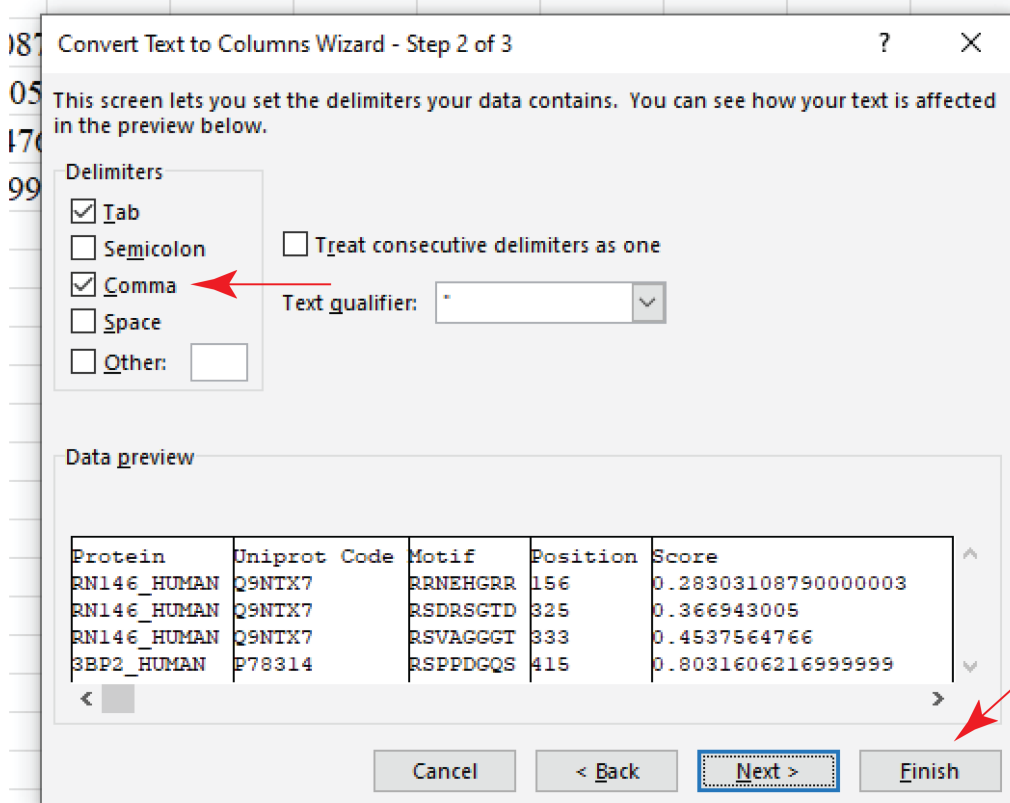**4.2**   Make sure that "Comma" is selected as a delimiter and click "Finish" (Figure 6).

Figure 6. Convert text to column in Excel.

4.3 Data will now be presented as columns for analysis (Figure 7).

Figure 7. An example of output converted into an Excel file.

Manually Score any potential TBM

5 Paste or manually input any 8-residue protein sequence. Select "Score any 8-amino-acid sequence" to output a calculated score of the sequence no matter the sequence composition.

Tankyrase Binding Motif (TBM) Search Project

Enter UniProt code(s) below, one code per line:

Find and Score Motifs! ❶   Clear

Input one sequence of 8 amino acids to be scored:

LSGRTRSS

Score any 8-amino-acid sequence ❶   Clear

Note: Scores are a predication and do not replace experimental data.

Figure 8. The function of "Score any 8-amino-acid sequence."

6   This will provide a raw score for your peptide with no constraints on selection.

Tankyrase Binding Motif (TBM) Search Project

Enter UniProt code(s) below, one code per line:

[Find and Score Motifs!] ⓘ   [Clear]

Input one sequence of 8 amino acids to be scored:

LSGRTRSS

[Score any 8-amino-acid sequence] ⓘ   [Clear]

Note: Scores are a predication and do not replace experimental data.

LSGRTRSS -> score of -0.4671243525

Figure 9. An example of output for "Score any 8-amino-acid sequence" function.