Instructions for windows:

Here we have presented a computational tool in MATLAB and python for the design of polyvalent guide RNAs which is optimized for activity at multiple viral sites, while also avoiding interactions with the host genome or transcriptome.

Software needed to be installed for running the code:

1. Install Python for windows
2. Go to the Anaconda Distribution Page at <https://www.anaconda.com/products/individual>. Click Download and select the latest Python version. Please ensure that you check the box that says "add to PATH" when installing on PC.
3. Libraries that needed to be installed in Python to run the code –

2.1) pandas - pandas library is available as a part of the latest anaconda package. It can also be installed using conda and pip, the two main tools that install python packages.

* Using conda – Type, conda install pandas, in the windows command prompt (cmd).
* Using pip – If you are using pip, Pandas can also be installed using pip. pip is a package management system used to install and manage software packages/libraries written in Python. These files are stored in a large “on-line repository” termed as Python Package Index (PyPI). To install using pip type, !pip install pandas in the Jupyter Notebook App or type pip install pandas on windows command prompt (cmd).

1. NCBI Standalone Blast
2. Download and install BLAST 2.8.1+ installers for your machine which is available from NCBI at <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>.
3. Set up the reference human genome database for the blast search using the command, makeblastdb -in GRCh38\_latest\_rna.fna/fasta\_file -dbtype nucl -parse\_seqids -out Human\_NCBI\_rnadb/database\_file. Add the created database to the system path.
4. Make the blastn query from the code (python) to find blast hits to the human transcriptomes/human DNA using subprocess call in windows.
5. Use the blastdbcmd command using subprocess call in windows from code(python) to retrieve the sequence parts from the created reference human transcriptome/ human DNA database.
6. Vienna RNA Package
7. Install the RNAfold installer compatible for your PC available through the website. Add the path of the package installation directory to your PATH variable manually.
8. Use the subprocess call command from the script (python) to make queries in the RNAfold software.

How the pgRNA design code works:

Input

1. To run the code, you need to provide an initial input in the form of a csv file which contains three columns mainly, potential gRNAs, their respective scores, and positions relative to a target sequence.

The model input for SARS-CoV-2 DNA genome sequence is available for reference as [CovidCasRxguides.csv](ttt):

1. To run the Cutting Frequency Determination (CFD) calculations, you need to input two csv files for Cas13d design, [Mismatch\_Scores.csv](tt) and [Adjacent\_Mismatch.csv](ttt) respectively.

Output will have two main elements:

1. Initially it provides a csv containing homologous target pairs which are in the top quartile. The pairs are sorted according to the respective guide efficiency score. The model output is shown as a csv file, [df\_pairs\_sorted\_top\_quartile.csv](ttt).
2. The final output is a csv file which gives a list of potential gRNAs with following characteristics:
3. High relative activity at two homologous targets in a given viral sequence.
4. Low predicted relative activity at potential human “off-targets”.
5. Biophysical characteristics such as high GC content, direct repeats, secondary structure energy etc. that suggest high CRISPR activity for potential antiviral application.
6. High relative activity across clinical strain variants

The model output can be referred at [df\_crRNA\_withhits.csv](ttt).

Running the code:

The code can be run in following two ways:

1. Run in the Jupyter Notebook - The Jupyter Notebook App can be launched by typing in a terminal (cmd on Windows). Change the Jupyter Notebook startup folder by using the command cd /the\_folder\_name in the command prompt (Windows). Once Jupyter Notebook is launched, it will automatically open a browser window and will show the following page. Click on new and then on “Python 3”. Run each cell in the jupyter notebook to get all the outputs.
2. Running through the windows terminal - navigate to the folder containing the python/Cas13\_pgRNA.py by typing in the terminal, cd path/to/ Cas13\_pgRNA.py.

Once you have navigated to the folder, run the python file in the windows terminal (cmd) by typing, python Cas13\_pgRNA.py

Common errors:

1. *ModuleNotFoundError: No module named 'pandas' when import pandas* – Make sure you have pandas installed by typing pandas –-h in the windows terminal(cmd).
2. *ImportError: No module named RNA* – Check your RNAfold installation by typing RNAfold –-h in windows terminal(cmd).
3. *BLAST Database error: No alias or index file found for nucleotide database* –Make sure you have given the correct path to the database while making the blastn query from your python script.
4. *df\_crRNA1["Free Energy"] = energy\_list, length mismatch error,* make sure you do not have a file named energyoutput.txt in the folder containg the python file/jupyter notebook file.