

Component Specification for crispr_array_generator

Willow Chernoske

Software Components

Class

Array: A set of python functions necessary to check or design CRISPR Cas12 guide (g)RNAs or arrays

Functions

extract_excel_data(excel_file: str {excel sheet title without .xlsx})

- Extracts all data from an excel file and isolates the DNA components into a separate array.
- Input: The name of an excel file as a string.
- Output: All DNA components listed in the input excel file compiled into an array.

get_reverse_complement(dna: str)

- Generates the reverse complement of an input DNA sequence
- Inputs: DNA sequence as a string
- Output: Reverse complement of the input DNA sequence as a string

check_grna(excel_file: str {excel sheet title without .xlsx})

- Checks if a gRNA has any errors (too long, too short, TTC cut site within guide region) and removes preceding TTC cut site if present.
- Inputs: The name of an excel file as a string.
- Output: An excel file named “grnacheck.xlsx” listing all processed gRNAs and any errors.

get_array(excel_file: str {excel sheet title without .xlsx})

- Generates ready-to-order forward and reverse oligonucleotides and the full array sequence of an array for up to 9 gRNAs listed in an excel file.
- Inputs: The name of an excel file as a string
- Output: An excel file named “grnacheck.xlsx” listing all processed gRNAs and any errors on the first sheet and the designed array including individual forward and reverse oligonucleotides on the second sheet.

Interactions to Accomplish Use Cases

Use Case: Generate ready-to-order CRISPR array oligos or sequence from gRNAs submitted as an excel file. For this use case the Array.get_array function will be used. This function will first call the extract_excel_data function to load the excel workbook and retrieve the gRNAs listed within the workbook. Then, it will call the check_grna function to check for sequence errors or

length errors. Individual processed gRNAs and any errors will be listed on the first sheet of the output excel file “grnacheck.xlsx.” Then get_reverse_complement will be called to produce the reverse complement sequence of each gRNA. The number of input gRNAs will be counted to determine how to assemble the array oligos. Finally, array oligos will be determined and compiled into the second sheet of the output excel file “grnacheck.xlsx”. This output format is helpful because many companies accept excel workbooks when ordering plates of oligos.

Preliminary Plan (from 11/06/23)

Current status: I have working function prototypes for the checkgRNA, getReverseCompliment, and getArrayFromgRNA functions with string inputs.

Plan

1. Implement excel interface to check_grna function.
2. Import a test function to github repository and verify that I can install with pip
3. Implement excel interface to get_array function.
4. Build get_array_from_dna function if time
5. Import all functions to github repository and again, verify pip install works
6. Final touches and optimizations as necessary/time allows