

Component Specification for crispr_array_generator

Willow Chernoske

Software Components

Functions

checkgRNA(grna: str, grnaexcelsheet: str {excel sheet title})

- Checks if a gRNA has any errors (too long, too short, non nucleotide bases, TTC cut site within guide region) and removes preceding TTC cut site if present.
- Inputs: gRNA sequence as string or an excel sheet listing gRNAs.
- Output: a print statement describing any issues. String: inputted gRNA if no error or editing gRNA if fixable error.

checkArray(oligos: excel sheet)

- Checks if array oligos have any errors (too long, too short, non nucleotide bases, TTC cut site within guide region, non complementary sequences).
- Inputs: an excel sheet listing gRNAs.
- Output: an excel sheet and print statement describing any issues.

getReverseComplement(dna: str)

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

getArrayFromgRNA(grnaexcelsheet: str {excel sheet title}, grna1: str, grna2: str, grna3: str, grna4: str, grna5: str, grna6: str, grna7: str, grna8: str, grna9: str)

- Checks up to 9 input gRNAs for errors and generates ready-to-order forward and reverse oligos from gRNAs.
- Inputs: up to 9 gRNA sequences as strings or an excel sheet listing gRNAs.
- Output: print statement describing any issues with the input gRNAs and a print statement detailing all forward and reverse oligos. An excel sheet detailing any issues if present and with all oligos listed.

getArrayFromDNA(dna: str, number: int)

- Finds all potential gRNAs in a DNA input, chooses the most optimal targets, and generates corresponding array oligos for the number of gRNAs the user wants.
- Inputs: string - dna of target sequence and an integer - number of gRNAs the user wants in their array (up to 9).

- Output: print statement describing any issues with the input DNA and a print statement detailing all forward and reverse oligos if no issues. An excel sheet detailing any issues if present and with all oligos listed.

Interactions to Accomplish Use Cases

Use Case: Generate ready-to-order CRISPR array oligos or sequence from gRNAs submitted as an excel file. The software will first load the excel workbook and retrieve the gRNAs listed within the workbook. Then, it will call the checkgRNA function to check for non-base characters, sequence errors, or length errors. Then getReverseCompliment 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 assembled and print statements will note if any issues were found and will output each forward and reverse oligo. An updated excel workbook will also be outputted to detail any issues if present and listing all ready-to-order oligos. This is helpful because many companies accept excel workbooks when ordering plates of oligos.

Preliminary Plan

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

Plan

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