Bioinformatics portion

Lab "X" – CRISPR Design

Software required:

- Text editor (Notepad/Notepad++, Sublime Text Editor, Atom, etc.…)
- Browser for web access (Google Chrome, Mozilla Firefox, Microsoft Edge, etc.…)

Recommended you do Computer Protocol #1 at home prior to coming into the lab though not necessary!

Objective:

- ✓ To understand bioinformatic tools that available online for CRISPR design
- ✓ Be able to create, design and customize gRNA and its appropriate primers

Introduction:

In this dry lab portion of the CRISPR protocol, we will be looking at online tools that help you, the user, select and design genes and its corresponding guide RNA (gRNA) sites, protospacer adjacent motif (PAM) sites and primers. You will need access to Benchling (account required, may require you to sign-up using a UofT/university associated email) for this protocol.

Both bioinformatics (creation of biological software's/applications and tools) and computational biology (the study of biology using computational techniques) are an ever-growing field with every new discovery in the field of science; whether it be computer science/programming, math, physics, biology, etc. With each discovery and innovative technique, new software and applications must be created with its own unique or modified algorithms to solve different problems that can be found accompanying the respective issues. One example of this comes with CRISPR.

When CRISPR became a keen interest of many labs, tools and software's had to be created to accompany the mass need for a tool to help determine and design the optimal CRISPR sequences. This included finding and creating optimal gRNA sites with corresponding and appropriate PAM sites based on certain parameters defined by the scientist that matched their specific experiment. Here, you will be using Benchling's CRISPR design tool[1] to find and create your gRNA sites based on specific PAM sites to learn how to target and select specific and appropriate gRNA.

Supplementary Information #1: Many different tools!

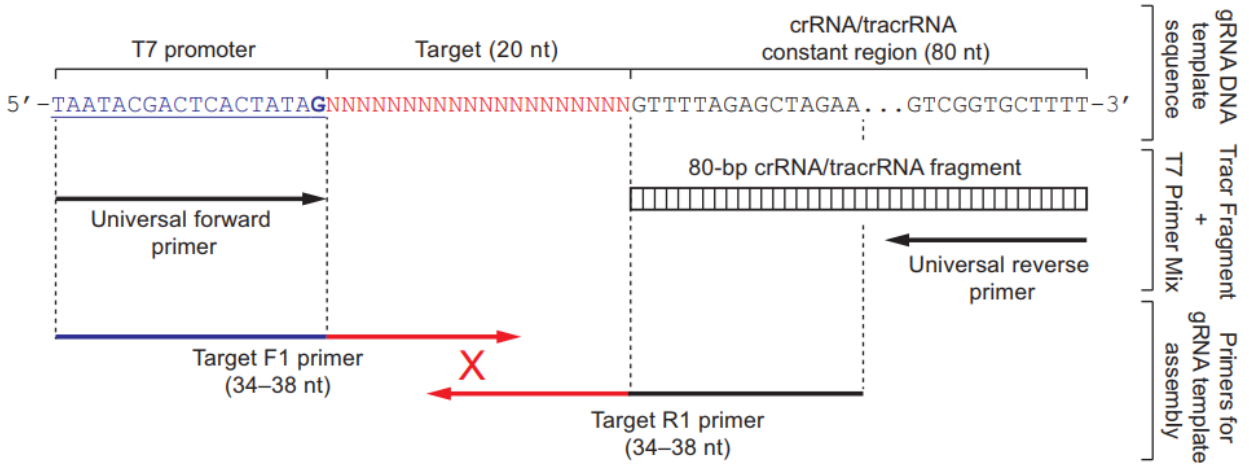
There are many different online web-based applications and local-client based software you can utilize to create your desired gRNA and PAM sites as well as your primers which all have their own benefits and disadvantages.

We are specifically looking at Benchling, a virtual lab notebook web-tool that began as a iGEM project by MIT in 2012[2]. Not only does Benchling provide the user with a clean, user-friendly interface, it provides the user, you, with many different tools such as plasmid design and cutting, genome visualization, primer design, band/digest prediction, sequence alignment, CRISPR design and many other tools with multiple in-depth customizable options specific to your options that utilize many different algorithms that have become commonly accepted in the field of bioinformatics.

Another tool that exists is the ThermoFisher GeneArt CRISPR Search and Design Tool[3]. This tool allows you to insert a gene or a gene name and search through it for an all possible CRISPR sites that work with their products. This tool also provides you with a table of recommended sequences and how many binding sites that specific CRISPR sequence has. The disadvantage to this web-tool is that it does not provide the user, you, with the options to customize the settings in depth to specify what regions, without you manually checking, you want the gRNA to bind to. This tool is optimal if you are planning on buying products directly from ThermoFisher as it tells the user if the users' sequence is compatible with their products and provides the user with Genomic Cleavage Detection (GCD) primers to be utilized in GCD assays (by GCD Kits, a technique used to rapidly determine if locus-specific double-stranded breaks occur in their genome edits[4]) to purchase with their CRISPR product.

There are many more options that exist out there, like MIT's own and crucial MIT CRISPR Design tool[5] or Galaxy's CRISPR Recognition Tool[6], that you can utilize to create your own gRNA with PAM sites and its appropriate primers. Once you understand what conditions your design preparations you require for your CRISPR project, you will understand which tool would be optimal for your design process.

You will also be learning to design your own target forward (F1) and reverse (R1) primers similar to how they were designed in your wet lab protocol; utilizing ThermoFisher's GeneArt Precision gRNA Synthesis' forward and reverse oligonucleotides for PCR assembly guide (as seen in the figure below). These primers allow for amplification of your desired target DNA sequence. The overlapping of the two primers you see within the target DNA sequence occurs because it allows the sequences to not only anneal later but act both as a primer and a template in PCR to generate a full-length DNA template.[11] These two primers are called nested PCRs primers because they are "nested" within the sequence. In general, nested PCRs role is to provide intrinsic product carryover to ensure all the target gene is covered as well as provide maximum concentration of PCR products[17].



Though, to create an optimal gRNA, you must be able to understand what makes an optimal gRNA binding sites. When segmented sequences align, there is always the possibility that it can be bind to multiple position along the genome. If the sequence binds on your specified location, this is called on-target binding while if it binds to an undesired location, this is called off-target location. To determine if your gRNA is optimal for what you desire, you would want as little as possible off-target binding sites as possible and the most efficient on-target binding sites. This can be calculated manually by doing a sequence alignment and seeing where the binds to on the genome to determine off-target binding and how well it can encompass the on-target sequence but thankfully, there are algorithms that exist that does these calculations for you.

The algorithms we will be using, which are the default setting for Benchling, are the CFD on-target scoring algorithm[7] and the MIT Scoring Method for off-target scoring[8]. For both variables, you would want an output as close to 100 as possible as that would represent the best possible scoring and sequence for your CRISPR experiment.

When looking at the on-target sites, you would want to choose a site that is within 20% of your max value (therefore, if your highest guide sequence has a on-target score of 100, you would want to choose something that is at least 80 or higher) though, it would not be optimal to choose something below 40[8].

When looking at off-target sites using the CFD algorithm, the scoring does depend on what PAM sequence you are looking for. Let's say you are looking for a common PAM site like NGG, you would want a off-target site that is relatively high: near or above 80. While, if you are looking at a less common PAM site, a lower off-target score can be considered acceptable relative to the other off-target scores that have been calculated for your specified gene[9]. Do beware, the lower the off-target score, the more likely the gRNA will bind to an undesired location.

Supplementary Information #2: Algorithms involved

Do not be scared off by the word "algorithm". Though it is commonly thrown around to make something sound fancy or complicated, all it means it a set of rules for an operation or calculation to follow to determine an output.

The MIT Scoring Method for off-target scoring is rather simple. Each mismatch would have a weighted value to it based on its position along its sequence ($W[e]$) and be accounted towards the score of a single hit (a single hit is defined by the whole sequence that contains mismatches, not a single mismatch) for each mismatch in that single hit (n_{mm}^2). This can be shown using the equation (don't worry too much about the equation, you will *not* be tested on it)[8, 14]:

$$s_{hit} = \prod_{e \in M}(1 - W[e]) * \frac{1}{\left(\frac{(19 - \overline{d})}{19} * 4 + 1\right)} * \frac{1}{n_{mm}^2}$$

From there, the scores would be aggregated using an assigned score where the more mismatches you have, the smaller the number would be. Therefor, the less the number of single hits you have (less mismatches), the score would be closer to a theoretical 100 therefor, that would be the best option for a guide sequence. This can be visualized using the equation[8, 14]:

$$S_{guide} = \frac{100}{100 + \sum_{i=1}^{n_{mm}} S_{hit}(h_i)}$$

/*** I cannot find the breakdown of the CFD equations to determine its output, going to need more time to understand "CFD Rule Sets" and "CFD Poolsets" to determine how they calculated the on-target scores ***/

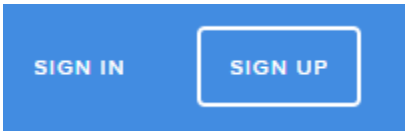
The software that you will be required to have installed onto your computer will be a text editor and a web browser. Your operating system does not matter but you may not be able to run certain applications on mobile devices.

Computer Protocol #1: Signing up and understanding NCBI and Benchling

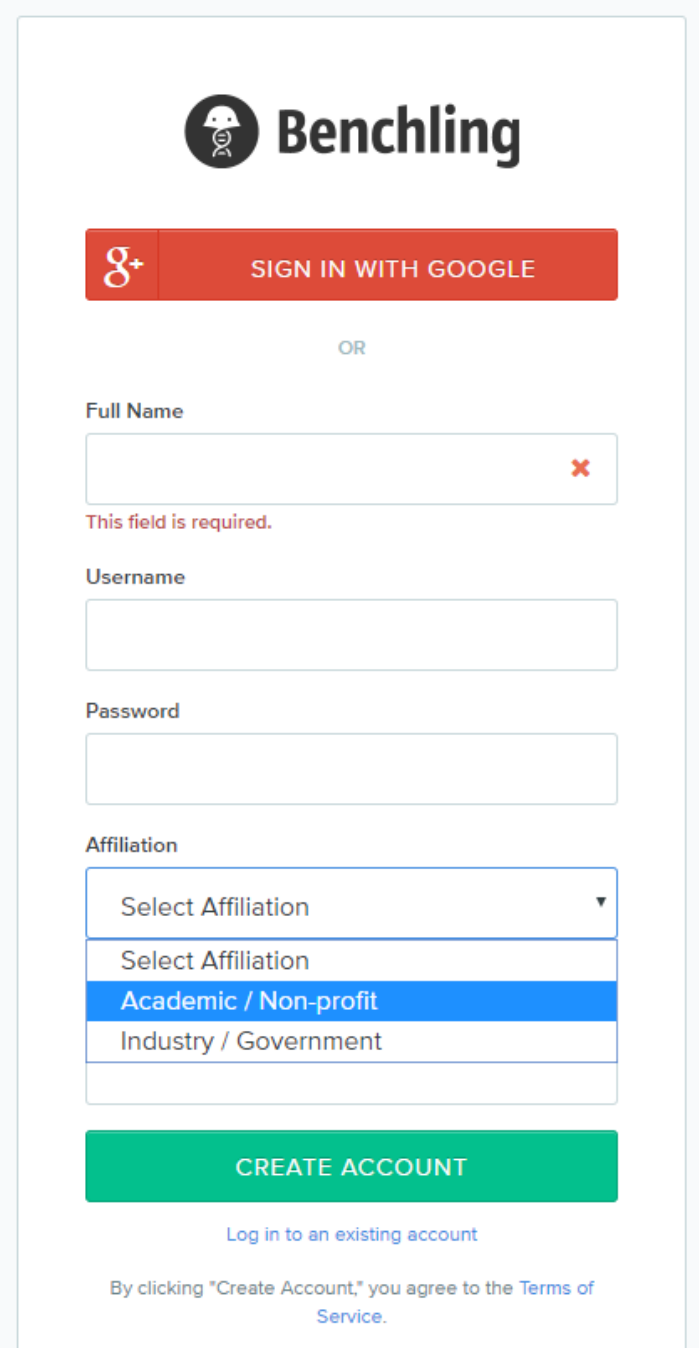
The web-tool that we will be using today is called Benchling. Benchling is a virtual lab notebook that is used by the academia and the industry to maintain their lab notes while providing easy-to-use and user-friendly bioinformatic tools[1]. In Computer Protocol #1, you will be brought through the basics of and how-to sign-up for Benchling.

1. Go to the following link: https://benchling.com/
2. On the top right of the page, you should see an option to "SIGN UP", click that button

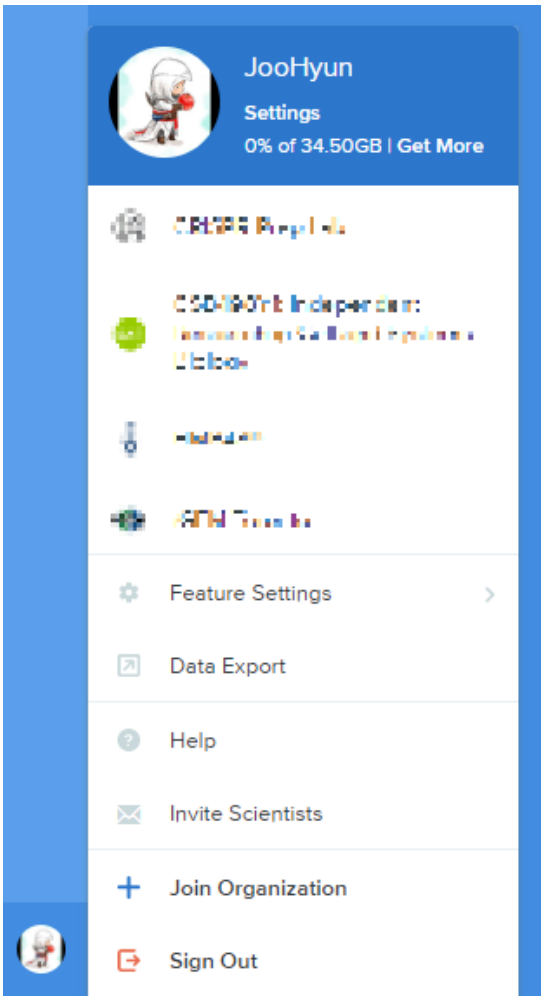
*****IT IS REQUIRED FOR YOU TO SIGN-UP TO USE BENCHLING*****



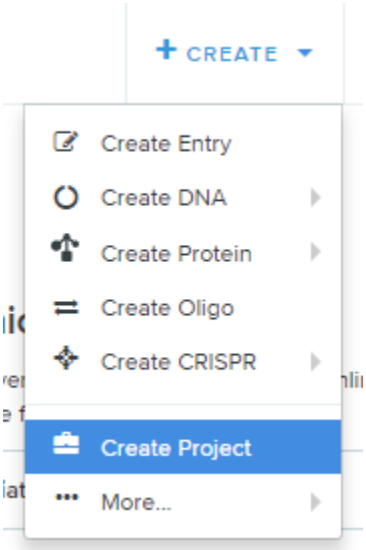
3. On the sign-up page, change your affiliation “Academic/Non-profit” and follow the rest of the on-screen instructions to complete your sign-up
    - If you are running into any issues during sign-up, please contact your TA



4. Once you have completed your sign-up onto Benchling, follow on-screen prompts to finish setting up your personal Benchling account
5. You are going to want to create a new project. Click on your profile icon on the bottom left of Benchling and go to “Settings” (not “Feature Settings” but “Settings”)



6. On the top right, click on the “Create” dropdown and choose the “Create Project” option



- It may be important to note that every time you work with Benchling, you do *NOT* need to create a new project. Later into this tutorial you will learn how to create folders within your project. If your work is still related to the same topic/field as your

project, it helps with your organization to create folders and sub-folders opposed to new projects each time.
- *Though*: This is dependent on you, the user, on your preferred method of organization as there is no formal way of keep your files organized relating to this type of work

7. From here, just follow the on-screen instructions on how to create your project. We recommend you call your project something recognizable like "HMB CRISPR Teaching Lab" or with your course code. Once you have completed this, please move onto Computer Protocol #2 where you will be designing your gRNA and primers.

Computer Protocol #2: Creating a gRNA and primers

The sequence we are working on is called enhanced blue fluorescence protein (eBFP). First, you are going to create your own file that will contain the following sequence:

TCCACCCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATT
GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACC
GGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCCACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACTTCAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCG
CCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGG
TACCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAA
ACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTAC
AATAAAGCAATAGCATCACAATTTCACAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGG

1. In any text editor like Notepad/Notepad++, Sublime Text or Atom, write the following:
   ```
   > eBFP Sequence
   ```
2. Then, create two new lines (press Enter twice) and copy & paste the above sequence into the text editor, it should look as the following:
   ```
   > eBFP Sequence

   TCCACCCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCG
   CCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGAT
   CCGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
   TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
   GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCCACGGCGTGCAGTGC
   TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
   CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
   CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAAC
   GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCG
   TGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAG
   CACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGG
   ATCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTA
   CCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTT
   AAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCT
   TATAATGGTTACAATAAAGCAATAGCATCACAATTTCACAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGG
   ```

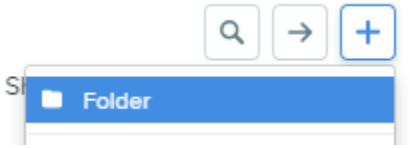
3. Choose "Save As" (typically found under the "File" dropdown option), and save your file as either a text file (.txt) or as a fasta file (.fasta). If you are using Notepad++, Sublime or Atom, please save your file as a fasta file. If you are using Notepad or another text editor, please save it as a text file.

Supplementary Information #3: FASTA

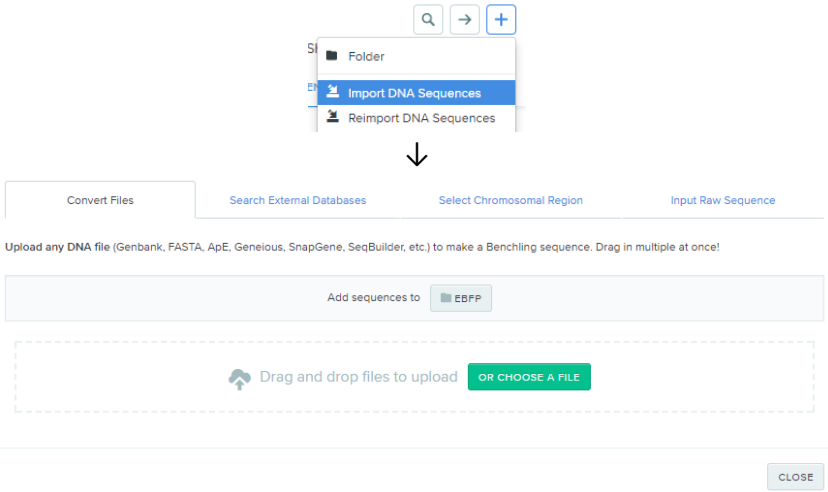
FASTA files are a common type of file format that is used in bioinformatics and computational biology. It is a sequence record file that consists of a single-line description (indicated by the ">") followed by its sequence.

These file formats are extremely important and crucial as they are easy to use and easy to parse through with clear indications for many software and applications to read and understand. Many web-apps and software's accept these types of file formats, so it is important that you understand how they are structured, how you can modify them and how you can create them for your own sequences/data.

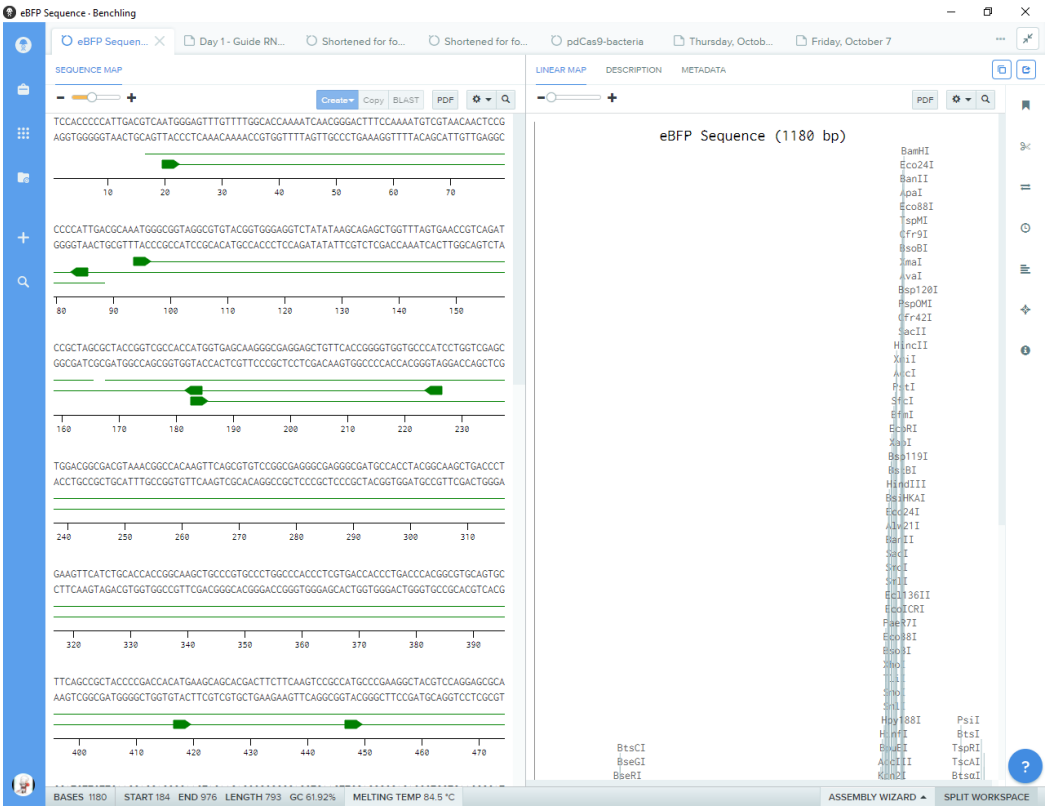
4. In Benchling, under the appropriate Project, click on the "Inventory" tab and create a folder (you do not have to create a folder, but it may help keep you organized)
    I) To create a folder, you will click on the plus symbol while the "Projects" viewer is opened



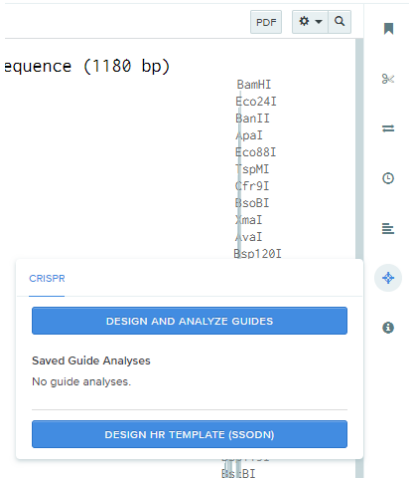
5. While in the new folder, click on the plus in the "Projects" viewer and "Import DNA Sequence". When this modal opens, while in the "Convert Files" tab, either drag and drop your file or browse/choose your eBFP (.fasta or .txt) file to upload



6. When the file is finished uploading, open the sequence. It should appear in the inventory's folder that you uploaded it to. The sequence screen should appear similar to the following:

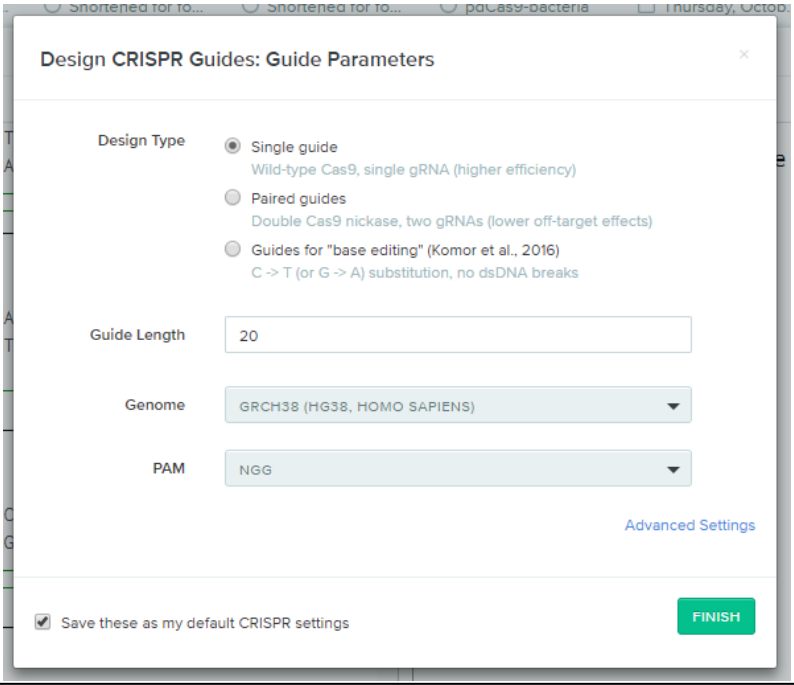


7. On the right side of Benchling while your sequence is open, you should see the following button with the glyph icon "✥". This button opens a CRISPR design tool which would allow you to create and design guide sequences. Open this and click on "Design and analyze guides"



8. For now, keep the default settings on the "Design CRISPR Guides: Guide Parameters" modal and click on the green "Finish" button. If you wish to learn more about the "Advance Settings", see the supplementary information box #1 and #4.

   ***** Default settings may not be the same for all people, please check if your default settings are the same as below. Specifically, check that the genome is "GRCH38 (HG38, HOMO SAPIENS)"*****

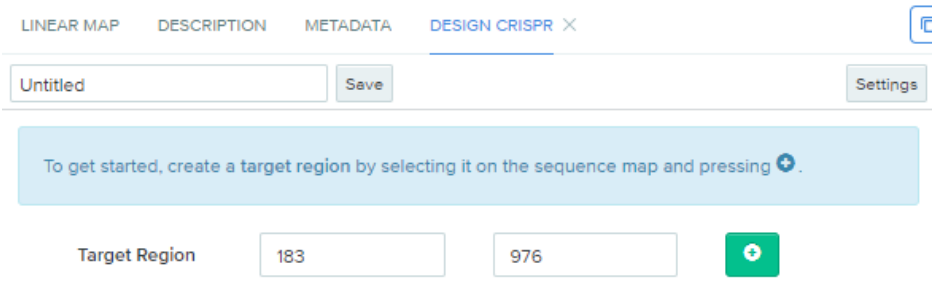


Supplementary Information #4: Advanced Settings

Though in your protocol today you will only be using the default settings designing your gRNA, in the future, you might want to utilize the "Advanced Settings" that Benchling has available for you to utilize. These include:

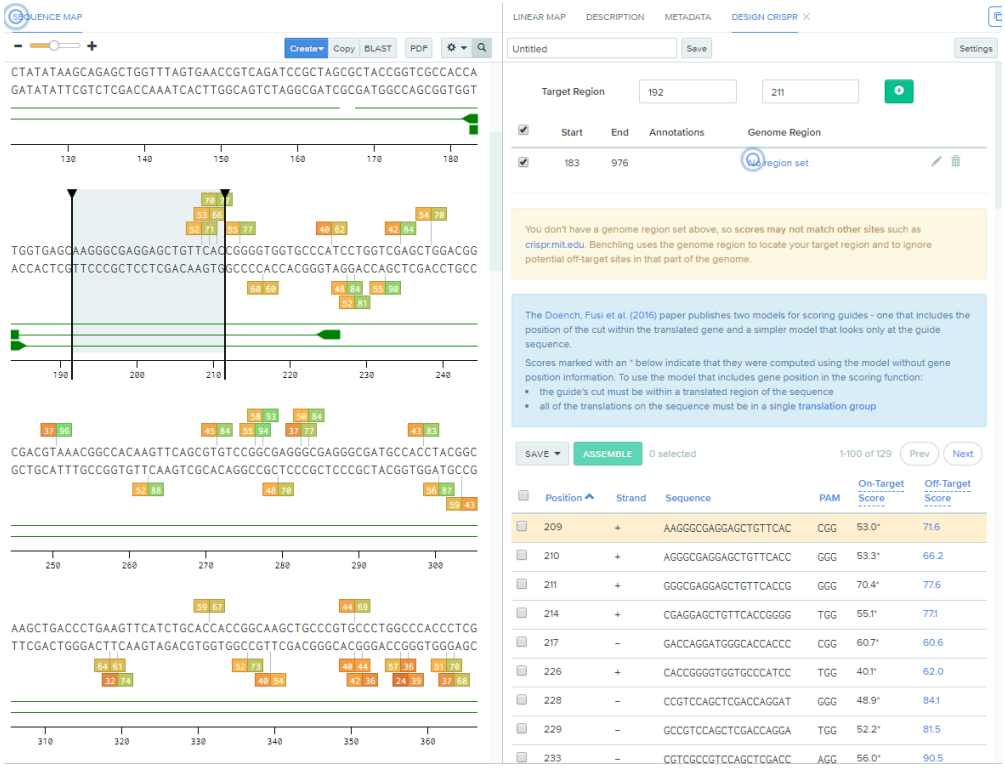
1. Masked Regions on or off (default off)
    - A masked region is a region of a sequence/gene that has skipping repeats or low complexity regions
2. Change off-target scoring algorithm
    - Change from MIT Scoring Method to CFD off-target scoring method
3. Turn of optimized scoring for on-target scores (by default this is on)
    - If you are utilizing SpCas9, a recent paper has determined how to optimize CFD off-target scoring to allow for maximum activity, you can turn this option off if you would want[15]
4. Display certain properties on the sequence
    - In this setting, there is a threshold setting this prevents any on-target below 10 and off-target below 25 from being shown, you can change these values here
5. Guide composition
    - Nucleotide proportion minimum and maximum for gRNA settings can be determined here (default is 0% min, 80% max)

9. Our target site that we wish to have our CRISPR gRNA modify is at sequence position 382, changing His66 to Tyr66. Therefore, we would want our gRNA to contain that that nucleotide position. Therefore, we are creating a target region along the protein region of nucleotide position 183 to 976. Put this as your "Target Region" and click the green button

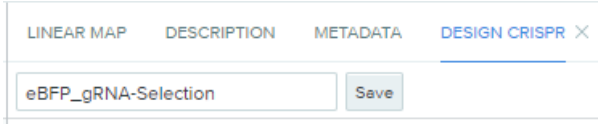


Question: Do you think the target region has to be the region that contains the sequence that encodes the protein product of the sequence? Would a shorter target region work (let's say 300-500)? How about a larger target region (let's say 1-1000)? Do you think one be more optimal over another and why?

10. Once the target region has finished loading (do let it load for a short bit, more sequences will appear over time), you should see a list of potential guide sequences become available. Look through them and click on different ones (do *not* click on "Save" or "Assemble" yet), you would notice different regions of the sequence map will be selected, these correspond to the guide sequence.

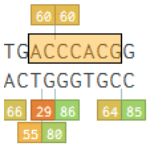


* It is recommended that you save your "Design CRISPR" tab in-case you need to work on this at a later point in time or your browser/computer/network crashes. This can be done at the top of the "Design CRISPR" tab:



11. As mentioned before, the modification we wish to create a gRNA for is at position 382 (a cytosine), changing Tyr66 to His66. Therefore, you would want to look for a sequence that contains this nucleotide position. We are going to choose a sequence based on strand direction, on- and off-target score as well as, most importantly, position.
    * Normally you would have to know exactly what you are looking for before you start designing a gRNA and its primers but due to this being a tutorial, we are giving you this information

    I) First, we would look at position, we would want a sequence that contains our target position. Click through the options and see which contain position 382
        * While nothing is highlighted in the sequence map tab, you can click on the 🔍 in the top right of this tab and find "ACCCACG" to have the sequence you want to be looking for highlighted. The cytosine in the middle of the "ACCCACG" is the exact position you are looking for, not this whole sequence



    II) If you are looking to use CRISPR to cut and modify/change a single (or few) nucleotides, the desired nucleotide should be within 3-8 nucleotides of the PAM site which is where the cut-site is located

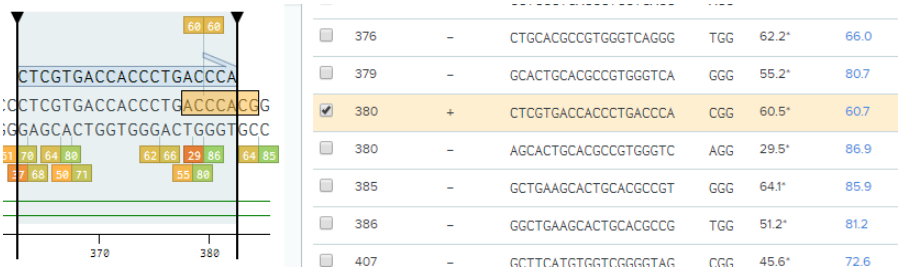
    III) Next you would look at on-target and off-target scores. The higher the score, the better. Generally, you should look for an on-target score above 40 as anything lower would be advised against using. Optimally, you would want a score of 80 or higher for off-target.
        * You can click on the off-target score to see other potential off-target binding sites in the genome for a selected species from Step 8

    Question: Why is it important to have an high off-target score?

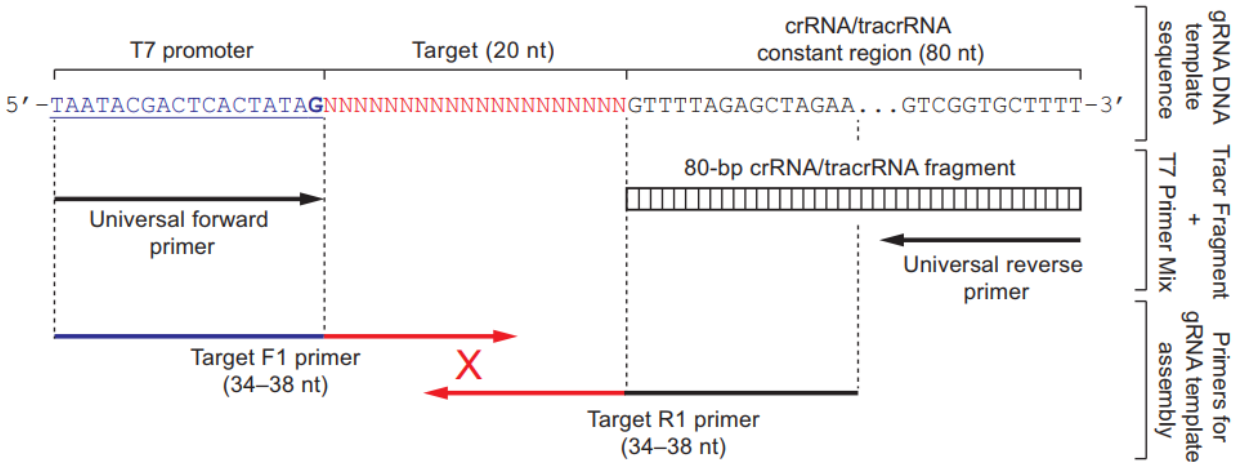
    IV) Once you have found potential sequences, you would look at the strand direction (the top sequence indicates the sense-strand (+) while bottom is the antisense-strand (-)). We are looking for a sequence that matches the same side your target gene is on. Typically, on Benchling, this is the sense-strand (+), the top strand.

    Question: Let's say your target region is on the sense-strand and *you are trying to change/modify a nucleotide, would using the antisense strand still work? Why?*
        * In the instance that no sense-strand is available, you can use an antisense strand. Refer to Alternative Example Protocol #1 for an example of how to do this with an antisense strand.

12. From the sub-steps given in Step 11, you should only be left with one option available; a guide sequence that reads 5'-CTCGTGACCACCCTGACCCA-3' (PAM: 5'-CGG-3') on the position strand starting at position 380 with an off-target score of 60.7. Select the checkbox



13. Once you have found your desired CRISPR's target DNA sequence, you will begin designing your target forward (F1) and reverse (R1) primers, the nested PCR primers. Following is a reiteration of the format of how the primers will be designed. Try and understand it before moving on:



Supplementary Information #4: T7 Promoters

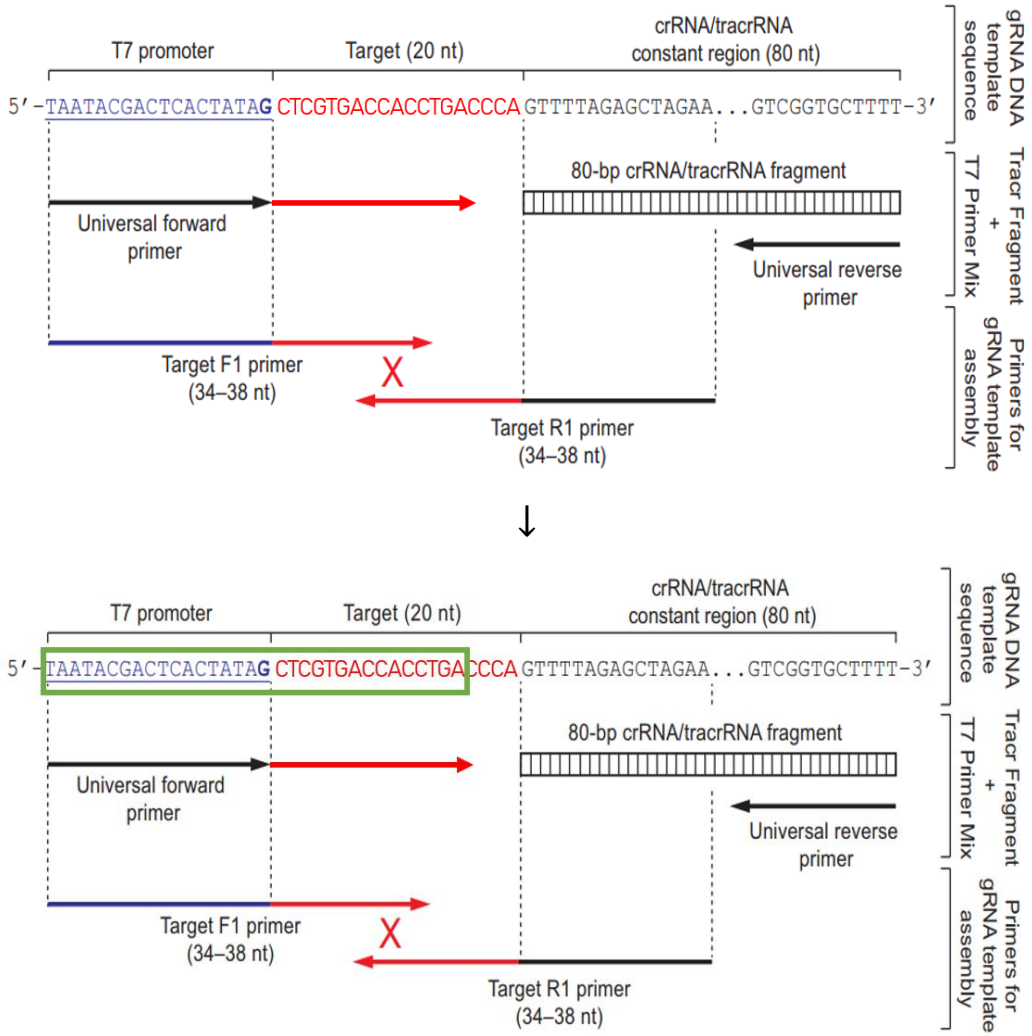
T7 promoters are commonly used in synthetic biology and in biological techniques as it a promoter that produces a high level of gene expression when T7 RNA Polymerase (T7-RNAP), a highly specific polymerase with a very low error rate, is present in its environment.

When used in a synthetic/artificial sequence, such as our CRISPR's gRNA DNA template sequence, it can enable a high level of expression of our target gene with its tracrRNA attached to it.

14. First, we are going to design the target forward primer (or the Target F1 primer). But it is important to first state some rules that come with designing Target F1 primers. First, the Target F1 primers could be between 34-38 nucleotides long where the first 18 are fixed from the T7 promoter[11]. Second, the Target F1 primers target sequence (everything after the first 18 nucleotides) *SHOULD* (does not have to be) be equal to or less than 20 bases long as it has been found to lead to improved specificity and lower off-target cutting rates[12]. Lastly, if the target sequence contains a guanine at the 5' end, it is your choice to keep it in your target sequence in the primer as it will not put your sequence at a disadvantage/inefficient without it[11].
    I) The Target F1 primer always begins with 5'-TAATACGACTCACTATAG-3' (T7 promoter)

II) Then you will be isolating the first 16-20 nucleotides of your designated eBFP target sequence. For simplicity, we are just going to use the first 16 nucleotides of our designated eBFP target sequence (5'-CTCGTGACCACCCTGACCCA-3'): 5'-CTCGTGACCACCCTGA-3'

III) Put those two sequences from I) and II) together to have your Target F1 Primer: 5'-TAATACGACTCACTATAGCTCGTGACCACCCTGA-3'



15. Next, we will be designing the target reverse primer (or the Target R1 primer). The following are some rules that come with designing Target R1 primers. First, the Target R1 primers could be between 34-38 nucleotides long where the first 15 are fixed as from the reverse compliment of the tracrRNA constant region[11]. Second, the Target R1 primers target sequence (everything after the first 15 nucleotides) *SHOULD* (does not have to be) be equal to or less than 20 bases long as it has been found to lead to improved specificity and lower off-target cutting rates[12]. There are two ways you can approach this:

    I) Reversing complementing everything first then stitching the sequence together:

        i. The Target R1 primer always begins with 5'-TTCTAGCTCTAAAAC-3'

ii. Then you will be isolating the first 19-20 nucleotides of the reverse compliment of your designated eBFP target sequence. You can write down the reverse compliment sequence yourself or use an online tool like Bioinformatics.org's Reverse Complement converter[13] to do it for you. Therefore, we will be using the whole reverse compliment sequence: 5'-TGGGTCAGGGTGGTCACGAG-3'

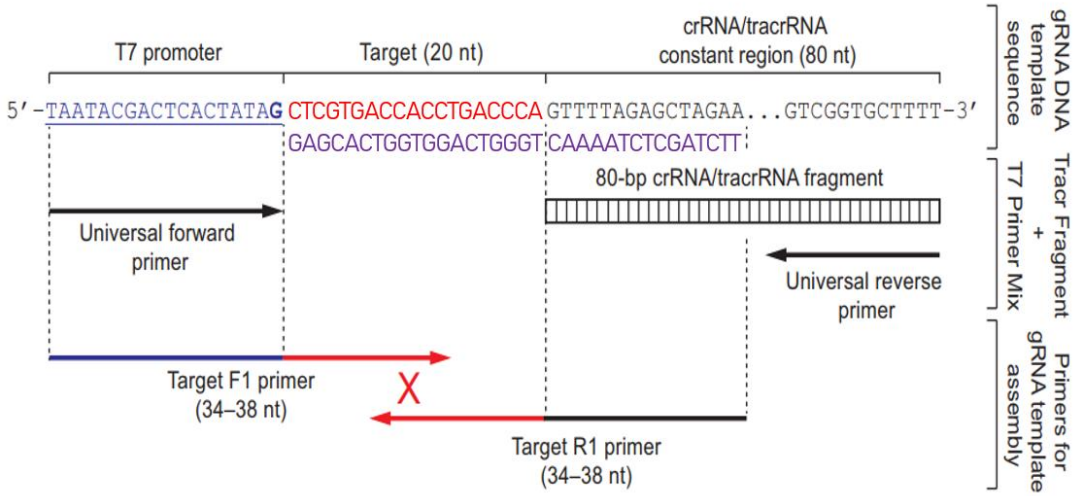
iii. Put those two sequences from I.i) and I.ii) together to have your Target R1 Primer: 5'-TTCTAGCTCTAAAACTGGGTCAGGGTGGTCACGAG-3'

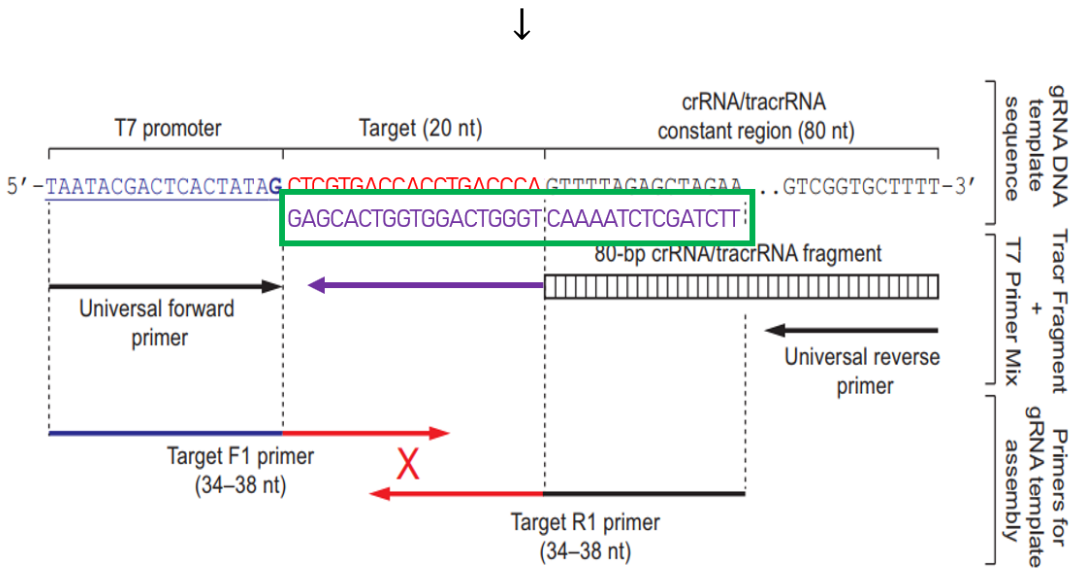
II) Putting the sequence together first then reverse complementing it:

i. The Target R1 primer always ends with the 5'-GTTTTAGAGCTAGAA-3'

ii. Then you will be isolating the first 19-20 nucleotides of your designated eBFP target sequence (5'-CTCGTGACCACCCTGACCCA-3') and adding the sequence from II.i) onto the end of that sequence: 5'-CTCGTGACCACCCTGACCCAGTTTTAGAGCTAGAA -3'

iii. Next, you will be generating the reverse complement sequence from what you generated in II.ii) which you can do either by writing down the reverse compliment sequence yourself or use an online tool like Bioinformatics.org's Reverse Complement converter[13]. Therefore, you would get something like this for your Target R1 Primer: 5'-TTCTAGCTCTAAAACTGGGTCAGGGTGGTCACGAG-3'





Alternative Example Protocol #1: A more complicated example of step 11 with ACTN3

The following alterative protocol consists of almost all the same steps with a few differences including what to do if Benchling provides you with multiple gRNA sequences for your desired site. You are not required to do this protocol but if you are struggling to understand or would like to get more practice with Benchling and creating gRNA, this alternative protocol can help you.

The following protocol is designed for the following gene with an alternative step 11: actinin alpha 3 (ACTN3). This gene is an alpha-actin binding protein that is involved in crosslinking actin in thin filaments, functioning as a structural component of sarcomere Z line in skeletal muscle. A mutation in ACTN3 leads to a deficiency which is found to cause elite athletic performance in sprinters and endurance runners[18]. Here, we are trying to create to a gRNA to fix for the ACTN3 mutation and determine if this gene is reversible by changing the thymine back to a cytosine at position 567.

GGAAAAGGGCTATGAGGACTGGCTGCTCTCGGAGATCCGGCGCCTGCAGCGACTCCAGCACCTGGCTGAGAAGTTCCGGCAGAAGG
CCTCCCTGCACGAAGCCTGGACCCGGGGAAAGGAGGAGATGCTGAGCCAGCGCGACTACGATTCGGCTTTGCTACAGGAGGTGCGG
GCGTTGCTGCGGCGCCACGAGGCCTTTGAGAGCGACCTGGCGGCGCACCAGGACCGCGTGGAGCACATTGCCGCGCTGGCCCAGGA
GCTCAATGAGCTGGACTACCACGAGGCAGCCTCAGTGAATAGCCGCTGCCAGGCCATCTGCGATCAGTGGGACAACCTGGGCACCC
TGACCCAGAAGAGGCGGGATGCGCTAGAGCGGATGGAGAAGCTCCTGGAGACCATTGACCGGCTGCAACTGGAGTTTGCCCGGCGG
GCCGCGCCCTTCAACAACTGGCTGGATGGTGCCGTGGAGGACCTGCAGGACGTGTGGCTGGTACACTCTGTGGAGGAGACCCAGAG
CCTGCTGACAGCGCACGATCAGTTCAAGGCAACACTGCCCGAGGCTGACTGAGAGCGAGGTGCCATCATGGGCATCCAGGGTGAGA
TCCAGAAGATCTGCCAGACGTATGGGCTGCGGCCCTGCTCCACCAATCCCTACATCACCCTCAGCCCGCAGGACATCAACACCAAG
TGGGATATGGTCCGAAAGCTGGTGCCCAGCTGTGACCAGACACTGCAGGAGGAGCTGGCACGGCAGCAGGTAAACGAGAGGCTCCG
GCGACAGTTTGCGGCCCAGGCCAATGCCATTGGACCCTGGATCCAGGCGAAGGTGGAGGAAGTGGGGCGGCTGGCAGCAGGGCTAG
CTGGCTCTCTGGAGGAGCAGATGGCTGGGCTACGGCAGCAGGAGCAGAAC

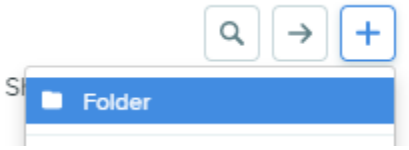
1. In any text editor like Notepad/Notepad++, Sublime Text or Atom, write the following:
   `> ACTN3 Sequence`
2. Then, create two new lines (press Enter twice) and copy & paste the above sequence into the text editor, it should look as the following:
   `> ACTN3 Sequence`

   GGAAAAGGGCTATGAGGACTGGCTGCTCTCGGAGATCCGGCGCCTGCAGCGACTCCAGCACCTGGCTGAGAAGTTCCGG
   CAGAAGGCCTCCCTGCACGAAGCCTGGACCCGGGGAAAGGAGGAGATGCTGAGCCAGCGCGACTACGATTCGGCTTTGC

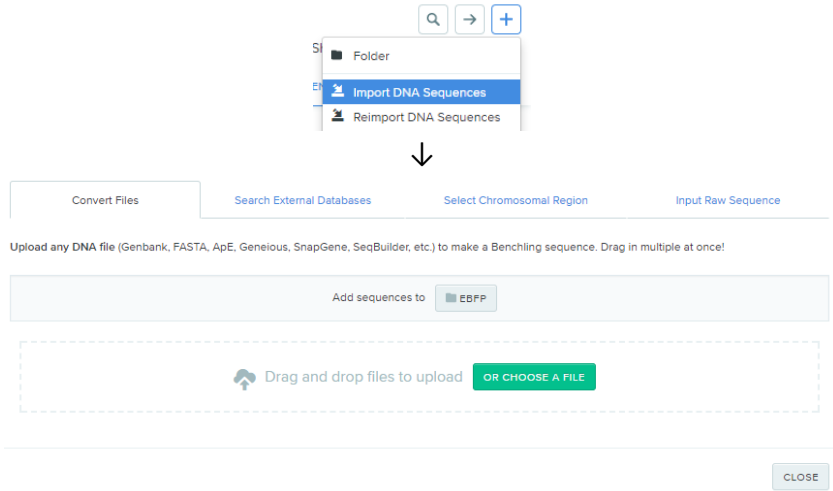
TACAGGAGGTGCGGGCGTTGCTGCGGCGCCACGAGGCCTTTGAGAGCGACCTGGCGGCGCACCAGGACCGCGTGGAGCA
CATTGCCGCGCTGGCCCAGGAGCTCAATGAGCTGGACTACCACGAGGCAGCCTCAGTGAATAGCCGCTGCCAGGCCATC
TGCGATCAGTGGGACAACCTGGGCACCCTGACCCAGAAGAGGCGGGATGCGCTAGAGCGGATGGAGAAGCTCCTGGAGA
CCATTGACCGGCTGCAACTGGAGTTTGCCCGGCGGGCCGCGCCCTTCAACAACTGGCTGGATGGTGCCGTGGAGGACCT
GCAGGACGTGTGGCTGGTACACTCTGTGGAGGAGACCCAGAGCCTGCTGACAGCGCACGATCAGTTCAAGGCAACACTG
CCCGAGGCTGACTGAGAGCGAGGTGCCATCATGGGCATCCAGGGTGAGATCCAGAAGATCTGCCAGACGTATGGGCTGC
GGCCCTGCTCCACCAATCCCTACATCACCCTCAGCCCGCAGGACATCAACACCAAGTGGGATATGGTCCGAAAGCTGGT
GCCCAGCTGTGACCAGACACTGCAGGAGGAGCTGGCACGGCAGCAGGTAAACGAGAGGCTCCGGCGACAGTTTGCGGCC
CAGGCCAATGCCATTGGACCCTGGATCCAGGCGAAGGTGGAGGAAGTGGGGCGGCTGGCAGCAGGGCTAGCTGGCTCTC
TGGAGGAGCAGATGGCTGGGCTACGGCAGCAGGAGCAGAAC

3. Choose “Save As” (typically found under the “File” dropdown option), and save your file as either a text file (.txt) or as a fasta file (.fasta). If you are using Notepad++, Sublime or Atom, please save your file as a fasta file. If you are using Notepad or another text editor, please save it as a text file.

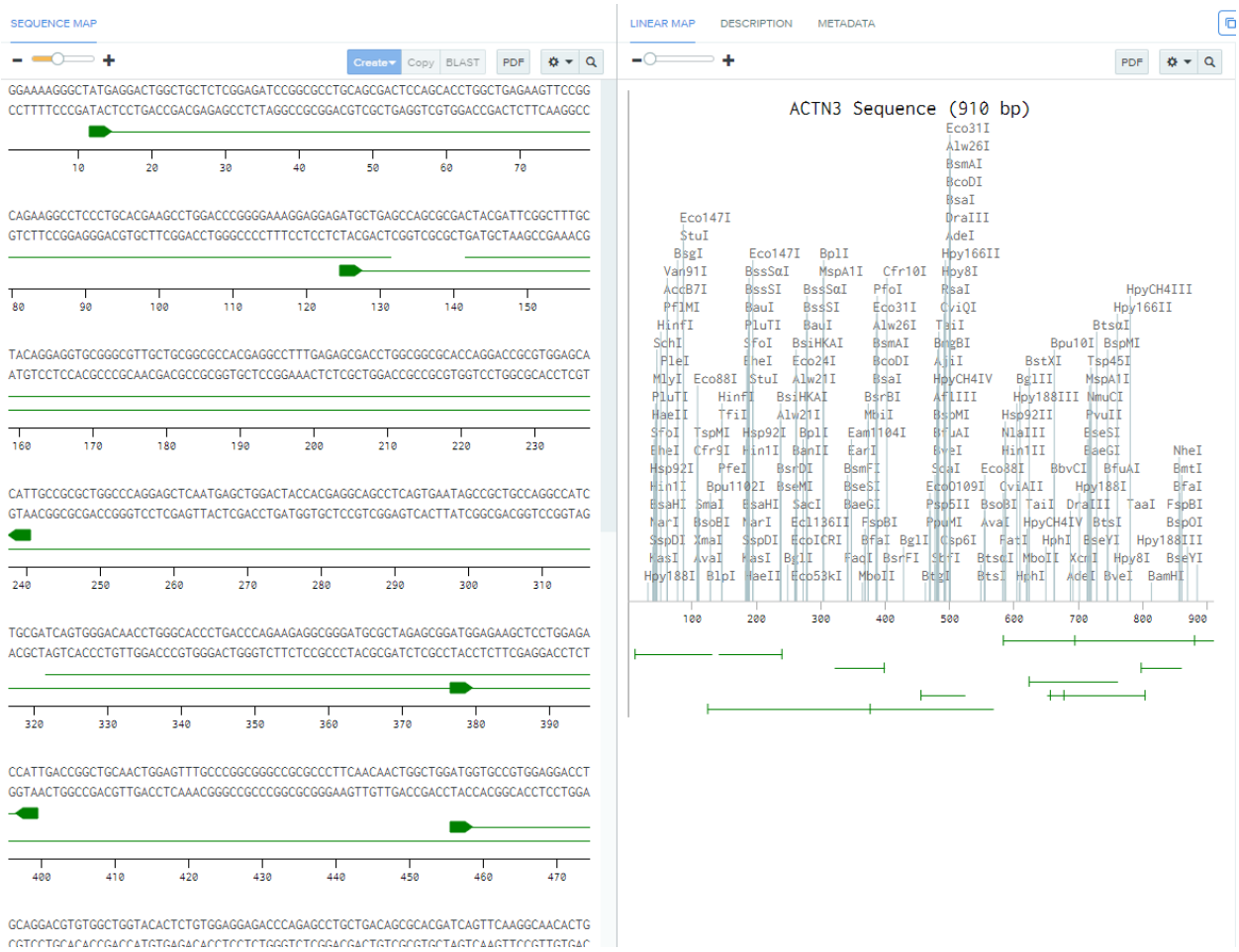
4. In Benchling, under the appropriate Project, click on the “Inventory” tab and create a folder (you do not have to create a folder, but it may help keep you organized)
    I) To create a folder, you will click on the plus symbol while the “Projects” viewer is opened



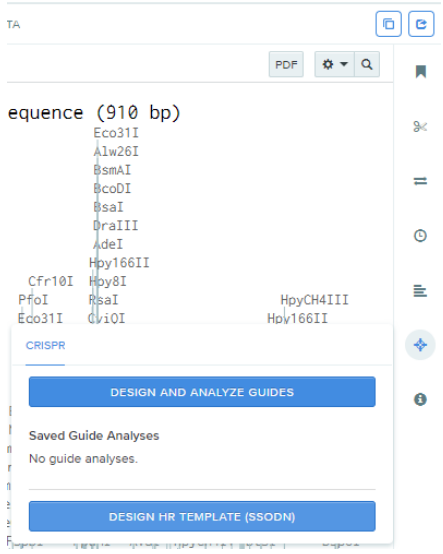
5. While in the new folder, click on the plus in the “Projects” viewer and “Import DNA Sequence”. When this modal opens, while in the “Convert Files” tab, either drag and drop your file or browse/choose your eBFP (.fasta or .txt) file to upload



6. When the file is finished uploading, open the sequence. It should appear in the inventory’s folder that you uploaded it to. The sequence screen should appear similar to the following:

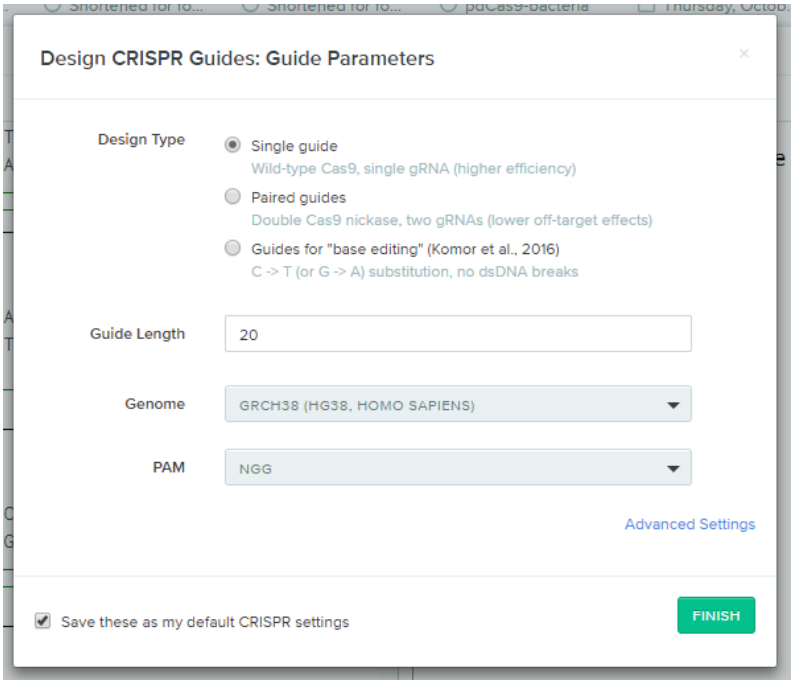


7. On the right side of Benchling while your sequence is open, you should see the following button with the glyph icon "✥". This button opens a CRISPR design tool which would allow you to create and design guide sequences. Open this and click on "Design and analyze guides"

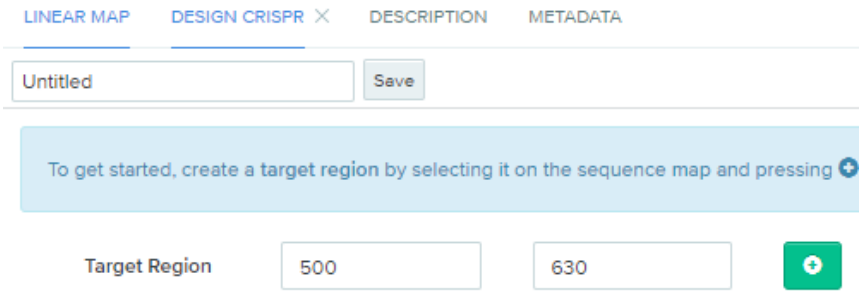


8. For now, keep the default settings on the "Design CRISPR Guides: Guide Parameters" modal and click on the green "Finish" button. If you wish to learn more about the "Advance Settings", see the supplementary information box #1 and #4.

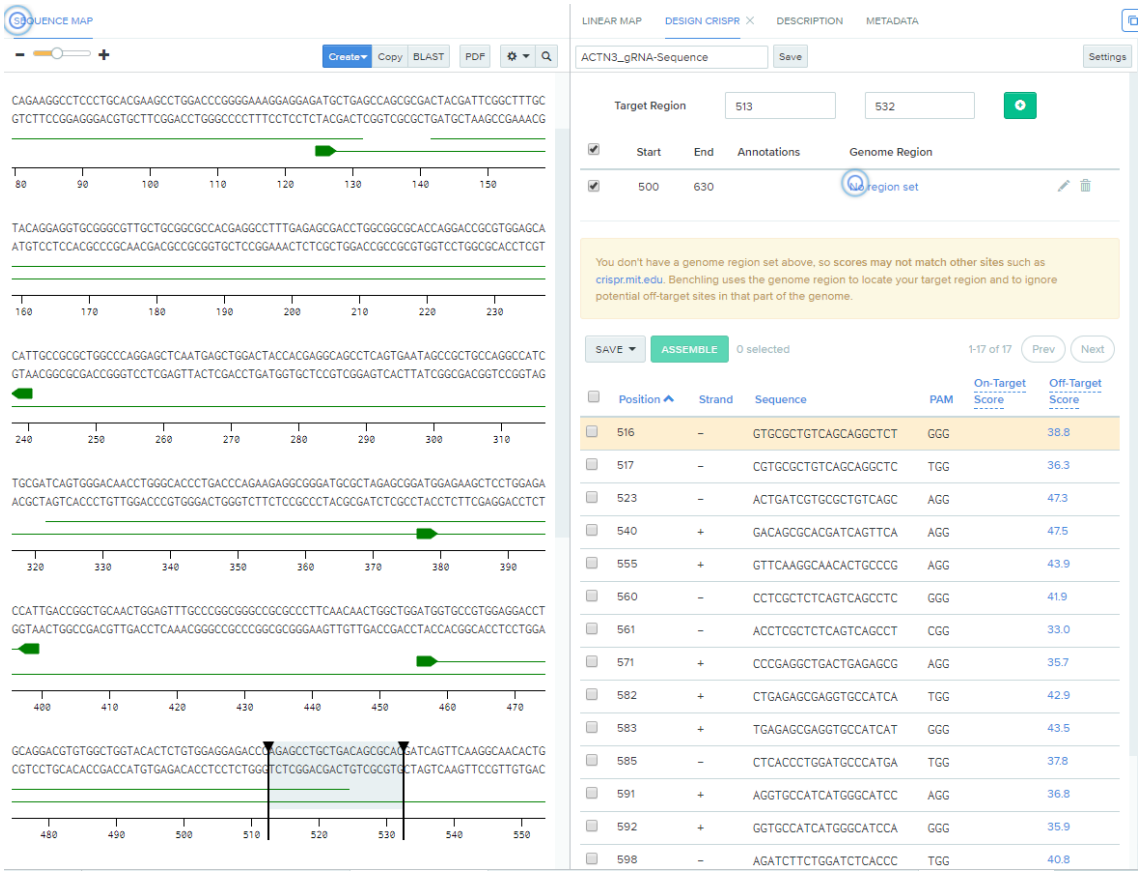
   ***** Default settings may not be the same for all people, please check if your default settings are the same as below. Specifically, check that the genome is "GRCH38 (HG38, HOMO SAPIENS)"*****



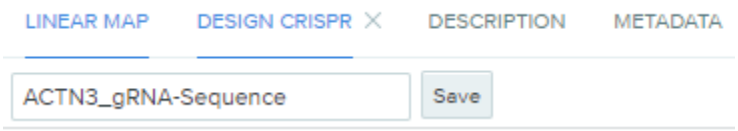
9. Our target site that we wish to have our CRISPR gRNA modify is at sequence position 567, changing Ter577 to Arg577. Therefore, we would want our gRNA to contain that that nucleotide position (from an T back to a C). Therefore, we are creating a target region along the protein region of nucleotide position 500 to 630. Put this as your "Target Region" and click the green button



10. Once the target region has finished loading (do let it load for a short bit, more sequences will appear over time), you should see a list of potential guide sequences become available. Look through them and click on different ones (do *not* click on "Save" or "Assemble" yet), you would notice different regions of the sequence map will be selected, these correspond to the guide sequence.



* It is <u>recommended</u> that you save your "Design CRISPR" tab in-case you need to work on this at a later point in time or your browser/computer/network crashes. This can be done at the top of the "Design CRISPR" tab:



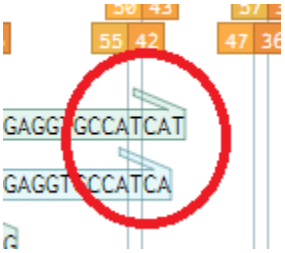
11. As mentioned before, the modification we wish to create a gRNA for is at position 567 (an thymine), changing Ter577 back to a Arg577. Therefore, you would want to look for a sequence that contains this nucleotide position. We are going to choose a sequence based on strand direction, on- and off-target score as well as, most importantly, position.
    * Normally you would have to know exactly what you are looking for before you start designing a gRNA and its primers but due to this being a tutorial, we are giving you this information
    I) First, we would look at position, we would want a sequence that contains our target position. Click through the options and see which contain position 567
        * While nothing is highlighted in the sequence map tab, you can click on the 🔍 in the top right of this tab and find "TGAGAGCGAG" to have the sequence you want to be looking for highlighted. The thymine in the <u>front</u> of the "TGAGAGCGAG" is the exact position you are looking for, not this whole sequence

* In this instance, we are given 5 options to choose our gRNA sequence from:

* Depending on the type of experiment you are running, you may choose more than one gRNA to use but in the case of our protocol, we are trying to find the most optimal gRNA to modify our gene, the following steps help you find the optimal gRNA

II) If you are looking to use CRISPR to cut and modify/change a single (or few) nucleotides, the desired nucleotide should be within 3-8 nucleotides of the PAM site which is where the cut-site is located

* To visually count this, you could count 3-8 nucleotides into the gRNA sequence from the start of the sequence (this is indicated by the arrow up on the gRNA sequence on Benchling)



* By cutting down the possible gRNA's by which contain our desire cut-site, it leaves us with two options remaining:

III) Next you would look at on-target and off-target scores. The higher the score, the better. Generally, you should look for an on-target score above 40 as anything lower would be advised against using. Optimally, you would want a score of 80 or higher for off-target.

* You can click on the off-target score to see other potential off-target binding sites in the genome for a selected species from Step 8
* Here are the on-target and off-target scores for the two-remaining gRNA's:

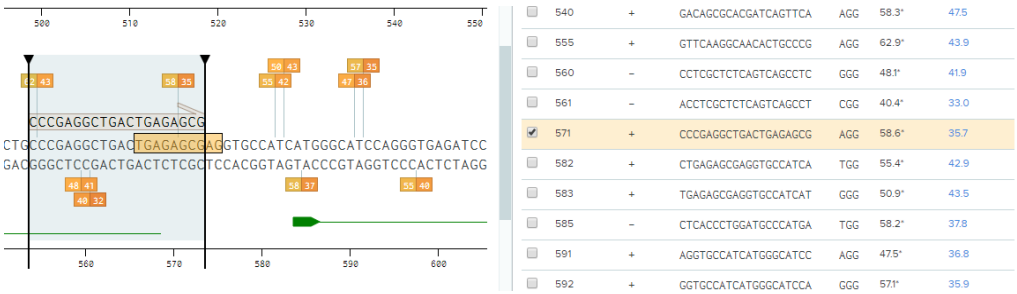
☑	561	–	ACCTCGCTCTCAGTCAGCCT	CGG	40.4*	33.0
☑	571	+	CCCGAGGCTGACTGAGAGCG	AGG	58.6*	35.7

* Though both have an on-target score above 40 so both are usable, using the range that we are using, the maximum on-target score is 62.9, therefore knowing that that our on-target score should be within 20% of the highest on-target score, us on-target score should be no lower than 50.3, the eliminates the antisense-strand (-) option leaving us with only one option remaining:

IV) Once you have found potential sequences, you would look at the strand direction (the top sequence indicates the sense-strand (+) while bottom is the antisense-strand (-)). We are looking for a sequence that matches the same side your target gene is on. Typically, on Benchling, this is the sense-strand (+), the top strand

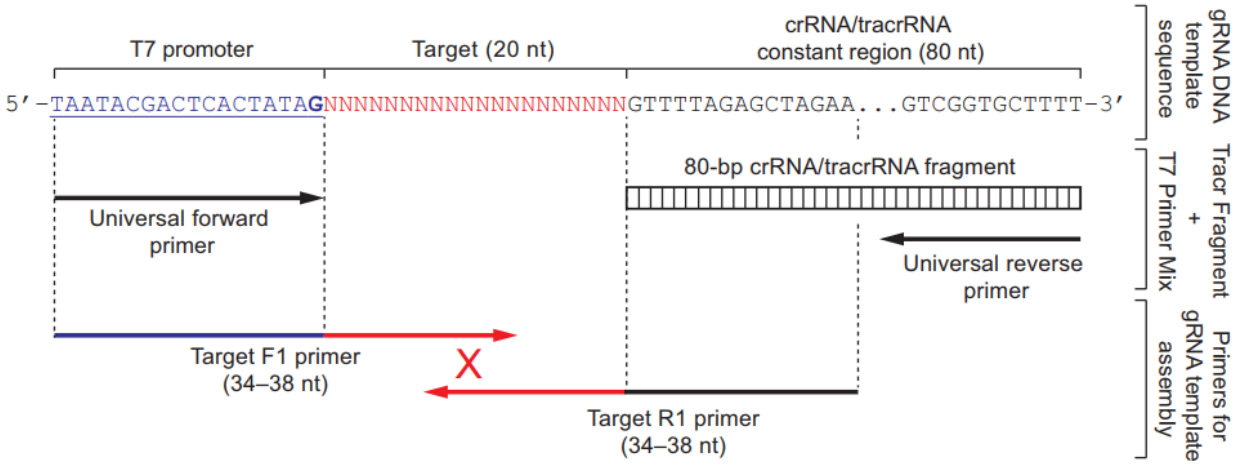
* Antisense-strands (-) still work and will still give you the same output but if you have the option between an antisense-strand (-) and a sense-strand (+), the sense-strand (+) is the more optimal option

12. From the sub-steps given in Step 11, you should only be left with one option available; a guide sequence that reads 5'- CCCGAGGCTGACTGAGAGCG-3' (PAM: 5'-AGG-3') on the position strand starting at position 571 with an off-target score of 35.7. Select the checkbox



540	+	GACAGCGCACGATCAGTTCA	AGG	58.3°	47.5
555	+	GTTCAAGGCAACACTGCCCG	AGG	62.9°	43.9
560	–	CCTCGCTCTCAGTCAGCCTC	GGG	48.1°	41.9
561	–	ACCTCGCTCTCAGTCAGCCT	CGG	40.4°	33.0
571	+	CCCGAGGCTGACTGAGAGCG	AGG	58.6°	35.7
582	+	CTGAGAGCGAGGTGCCATCA	TGG	55.4°	42.9
583	+	TGAGAGCGAGGTGCCATCAT	GGG	50.9°	43.5
585	–	CTCACCCTGGATGCCCATGA	TGG	58.2°	37.8
591	+	AGGTGCCATCATGGGCATCC	AGG	47.5°	36.8
592	+	GGTGCCATCATGGGCATCCA	GGG	57.1°	35.9

13. Once you have found your desired CRISPR's target DNA sequence, you will begin designing your target forward (F1) and reverse (R1) primers, the nested PCR primers. Following is a reiteration of the format of how the primers will be designed. Try and understand it before moving on:



14. First, we are going to design the target forward primer (or the Target F1 primer). But it is important to first state some rules that come with designing Target F1 primers. First, the Target F1 primers could be between 34-38 nucleotides long where the first 18 are fixed from the T7 promoter[11]. Second, the Target F1 primers target sequence (everything after the first 18 nucleotides) *SHOULD* (does not have to be) be equal to or less than 20 bases long as it has been found to lead to improved specificity and lower off-target cutting rates[12]. Lastly, if the target sequence contains a guanine at the 5' end, it is your choice to keep it in your target sequence in the primer as it will not put your sequence at a disadvantage/inefficient without it[11].
    I) The Target F1 primer always begins with 5'-TAATACGACTCACTATAG-3' (T7 promoter)
    II) Then you will be isolating the first 16-20 nucleotides of your designated eBFP target sequence. For simplicity, we are just going to use the first 16 nucleotides of our designated eBFP target sequence (5'-CCCGAGGCTGACTGAGAGCG-3'): 5'-CCCGAGGCTGACTGAG-3'
    III) Put those two sequences from I) and II) together to have your Target F1 Primer: 5'-TAATACGACTCACTATAGCCCGAGGCTGACTGAG -3'

15. Next, we will be designing the target reverse primer (or the Target R1 primer). The following are some rules that come with designing Target R1 primers. First, the Target R1 primers could be between 34-38 nucleotides long where the first 15 are fixed as from the reverse compliment of the tracrRNA constant region[11]. Second, the Target R1 primers target sequence (everything after

the first 15 nucleotides) *SHOULD* (does not have to be) be equal to or less than 20 bases long as it has been found to lead to improved specificity and lower off-target cutting rates[12].

I) The Target R1 primer always ends with the 5'-GTTTTAGAGCTAGAA-3'

II) Then you will be isolating the first 19-20 nucleotides of your designated eBFP target sequence (5'-CCCGAGGCTGACTGAGAGCG-3') and adding the sequence from I) onto the end of that sequence: 5'-CCCGAGGCTGACTGAGAGCGGTTTTAGAGCTAGAA -3'

III) Next, you will be generating the reverse complement sequence from what you generated in II.ii) which you can do either by writing down the reverse compliment sequence yourself or use an online tool like Bioinformatics.org's Reverse Complement converter[13]. Therefore, you would get something like this for your Target R1 Primer: 5'-TTCTAGCTCTAAAACCGCTCTCAGTCAGCCTCGGG-3'

Conclusion:

You have successfully found and created a gRNA for eBFP with its respective target forward (F1) and reverse (R1) primers using online tools which can be used in the ThermoFisher GeneArt Precision gRNA Synthesis!

With this knowledge, you should be able to apply this protocol to any simple CRISPR design protocol that you wish to create, design and modify your CRISPR designs without the need of purchasing tools or kits, or hiring a third party to create you're your gRNA. Using other CRISPR kits from other companies, modifications will be needed to be made but the general concept stays the same.

Assignment:

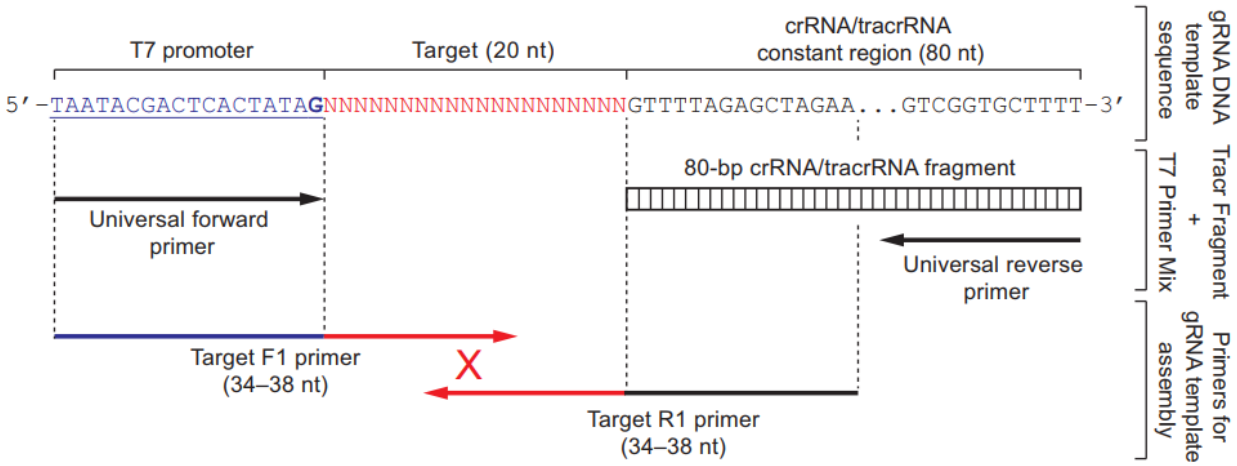
Using what you learned from the dry lab protocol, below is a list of genes that are have mutations in their sequence. Your objective is to find these mutations, design a gRNA and its corresponding F1/R1 primers so *YOU* can change it back into it's wildtype gene.

Instructions for gRNA creation:

- Using Benchling, you will create a gRNA using the settings given in the protocol above (if you changed any, please note that down)
- Of the gRNA you chose, also write down its PAM sequence, on-target and off-target score and its strand direction

Instructions for forward/reverse (F1/R1) primers:

- You will ***NOT*** be using the primers given from Benchling but rather create your own. Using the illustration below, you will create forward (F1) and reverse (R1) primers based on the way they are designed in the illustration



Genes:

*****PLEASE NOTE: THE SEQUENCES PROVIDED ARE CONDENSED VERSION OF THE GENE'S ACTUAL SEQUENCES WITH MUTATIONS/CHANGES ALREADY MADE. EACH GENE NAME WILL BE HYPERLINKED TO THEIR GENE INFORMATION PAGE BUT THE GENE MAY NOT EXACTLY MATCH WHAT YOU SEE. BLAST RESULTS WILL STILL SHOW THE SAME GENE AS DESCRIBED THOUGH. PLEASE USE THESE GENES HERE OPPOSED TO WHAT YOU SEE ONLINE*****

- HBB – Sickle cell anemia
  - Sickle cell anemia is caused by a SNP in [human] hemoglobin beta-chain at nucleotide position 20 from A → T causing a Glu7Val change (research does suggest there are more than one possible SNP such as a nucleotide position that 19G→A causing Glu7Lys[16]) which can cause the red blood cells to change into a sickle-like shape.
  - Target site:
    - Nucleotide position 73's thymine back to an alanine (highlighted green)
  - Sequence:

        - ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTC CTGAGGTGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGG CCCTGGGCAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGT CCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCT TTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACT GTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGG CCCATCACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTG TGGCTAATGCCCTGGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGG TTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGAT TCTGCCTAATAAAAAACATTTATTTTCATTGC

- CCR5 – C-C motif chemokine receptor 5
    - The CCR gene encodes for the beta-chemokine receptor family which is expressed in T cells and macrophages, playing an important role as a co-receptor for macrophage-tropic viruses (like HIV) to enter the host cells. A defective version (changing position Ser185 into Ile185) of this gene leads to an increase HIV infection resistance.
    - Target site:
        - Nucleotide position 210's guanine into a thymine (highlighted green)
    - Sequence:
        - GAATCTTCTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTGTCGTCCATGCTGTGTTTGCTT TAAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTG CGTCTCTCCCAGGAATCATCTTTACCAGATCTCAAAAAGAAGGTCTTCATTACACCTGCAGCTCTC ATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGC TGGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAATCCTAAAAACTCTGCTTCGGTGTC GAAATGAGAAGAAGAGGCACAGGGCTGTGAGGCTTATCTTCACCATCATGATTGTTTATTTTCTCT TCTGGGCTCCCTACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAATAATT GCAGTAGCTCTAACAGGTTGGACCAAGC

- ANKK1 – Ankyrin repeat and kinase domain
    - The gene encodes for the for a protein kinase superfamily involved in a signaling transduction pathway involved in dopamine receptors. A SNP leading to a missense variant of Glu713 to Ly713 has been associated with a reduced number of dopamine binding in the brain[19]
    - Target site:
        - Nucleotide position 201's alanine back to a guanine (highlighted green)
    - Sequence:
        - TGCAGTGTGGGGCTGACCCCAATGCTGCAGAGCAGTCAGGCTGGACACCCCTCCACCTGGCGGTCC AGAGGAGCACCTTCCTGAGTGTCATCAACCTCCTAGAACATCACGCAAATGTCCACGCCCGCAACA AGGTGGGCTGGACACCCGCCCACCTGGCCGCCCTCAAGGGCAACACAGCCATCCTCAAAGTGCTGG TCAAGGCAGGCGCCCAGCTGGACGTCCAGGATGGAGTGAGCTGCACACCCCTGCAACTGGCCCTCC GCAGCCGAAAGCAGGGCATCATGTCCTTCCTAGAGGGCAAGGAGCCGTCAGTGGCCACTCTGGGTG GTTCTAAGCCAGGAGCCGAGATGGAAATTTAGACAACTTGGCCAGCCGTGGTGGCTCACGTCTGTA ATCCCAGCACTTTGGGAGGCTGAG

- APOE – Apolipoprotein E
    - This gene encodes for a major apoprotein of chylomicron and is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. Mutations in this gene generally result in increased plasma cholesterol and triglycerides and one specific SNP, Arg176 to Cys176, has been found be associated with Alzheimer's disease[20]
    - Target site:
        - Nucleotide position 222's thymine back to a cystine (highlighted green)
    - Sequence:
        - CGGTGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCGGCGCAGGCCCGGCTGGGCG CGGACATGGAGGACGTGTGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGCC AGAGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCC GCGATGCCGATGACCTGCAGAAGTGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGC

GCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCGCCA
CTGTGGGCTCCCTGGCCGGCCAGCCGCTACAGGAGCGGGCCCAGGCCTGGGGCGAGCGGCTGCGCG
CGCGGATGGAGGAGATGGGCAGCCGGACCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGCGG
AGGTGCGCGCCAAGCTGGAGGAGCAGGCCCAGCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCC
GCCTCAAGAGCTGGTTCGAGCCCCTGGTGGAA

References:

1. Benchling.com
2. http://2012e.igem.org/Team:MIT_E
3. https://www.thermofisher.com/ca/en/home/life-science/genome-editing/geneart-crispr/geneart-crispr-search-and-design-tool.html
4. https://www.thermofisher.com/order/catalog/product/A24372
5. http://crispr.mit.edu/
6. https://github.com/bgruening/galaxytools/tree/master/tools/crt
7. Doench, Fusi et al 2016
8. Hsu et al 2013
9. https://www-ncbi-nlm-nih-gov.myaccess.library.utoronto.ca/pmc/articles/PMC4744125/
10. https://www.ncbi.nlm.nih.gov/
11. https://www.thermofisher.com/order/catalog/product/A29377
12. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3988262/
13. https://www.bioinformatics.org/sms/rev_comp.html
14. http://crispr.mit.edu/about
15. http://www.nature.com/nbt/journal/v34/n2/full/nbt.3437.html
16. https://www.ncbi.nlm.nih.gov/books/NBK1377/#sickle.Molecular_Genetics
17. https://www.ncbi.nlm.nih.gov/pubmed/7920237
18. https://www.ncbi.nlm.nih.gov/pubmed/18637739
19. https://www.ncbi.nlm.nih.gov/pubmed/9672901?dopt=Abstract
20. https://www.ncbi.nlm.nih.gov/pubmed/16603077?dopt=Abstract