How to use Green Listed (greenlisted.cmm.ki.se/) to design a custom CRISPR screen

Please also look at the provided explanatory videos and demo example to get a more hands-on idea about how to use Green Listed.

1. Choose the included GeCKOv2 library or upload your own reference library.

   GeCKOv2:
   The GeCKOv2 (Sanjana et al., Nature Methods, 2014) are full genomic reference libraries targeting human or mouse genes. Both human and mouse versions are divided into an A and a B library where each library contains 3 unique gRNAs/gene (the A+B thus contains 6 unique gRNAs/gene).

   More information about the GeCKOv2 libraries can be found under the GeCKOv2 info tab.

   Own reference library:
   The reference library has to contain one column with the Gene name and another column with the gRNA/spacer sequence. Other columns could be included, such as a ranking score if applicable. The reference library has to be uploaded as a Tab Delimited Text File (.txt). We often prepare the reference libraries in excel and then save them as a Tab Delimited Text File. Keep in mind that excel can introduce (easily fixed) artefacts in a library by changing the name of genes, if excel e.g. interprets a name as a date (this happens for e.g. the Sept gene family).

   As the Reference library has been successfully uploaded, Green Listed specifies that "The file has been uploaded and is ready to be processed". Sometimes it can take a little while if the internet connection is not optimal, or the server is occupied. Still you need to await the confirmation before you press run.

   As a reference library has been uploaded, define which column that contains gene ID, gRNA and, if applicable ranking. Note that the ranking option is only relevant if the reference library contains a ranking score (the GeCKOv2 does not).

   If you press the Demo link you can download a short demo references library and see how it is formatted, as well as to test using it by following the instructions in the Demo tab. You can also download the formatted GeCKOv2 libraries used in the software under the GeCKOv2 info tab.

   Several excellent reference libraries are presented in publications referred to in the reference list below.

2. Upload a set of genes that the custom screen should target.

   The list of genes need to be entered with one gene per row, i.e. if you manually input them you need to press enter after each gene name.

Note that the “Allow for partial matches” alternative gives the possibility to retrieve a larger set of genes by only searching a common part of the name. This can be useful to retrieve all genes starting with for example Map or Cyp and thereby avoid to have to write each of the members of the respective gene families.

We often design libraries targeting:

- Genes differentially expressed comparing cells stimulated or not with a ligand of interest.
- Genes up or down regulated as cells are differentiating.
- Genes related to particular pathways defined by for example GO-terms (e.g. GO:0016301 / Kinase Activity corresponding to 874 human genes or 818 mouse genes).
- Known drug targets.
- A list of candidate genes from a primary screen.

Examples of web pages we have used to identify interesting sets of genes:

- http://amigo1.geneontology.org/cgi-bin/amigo/go.cgi (to identify GO-terms. We find Amigo 1.8 simpler than Amigo 2.0 for this particular purpose. Genes can be retrieved from GO-terms using e.g. www.ensembl.org/biomart/)
- http://www.pathwaycommons.org/pcviz/ can be used to find genes linked to a gene of interest.
- http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways can be used to extract lists of genes related to BioCarta pathways, note that it is not recently updated.
- http://cgap.nci.nih.gov/Pathways/Kegg_Standard_Pathways can be used to extract lists of genes related to Kegg pathways, note that it is not recently updated.
- http://www.gtexportal.org/home/eqtls/tissue?tissueName=Whole_Blood can be used to identify the top expressed genes in different organs.
- http://dgidb.genome.wustl.edu/druggable_gene_categories can be used to identify druggable genes.
- http://rstats.immgen.org/PopulationComparison/ can be used to identify differentially expressed genes comparing different sorted mouse cell populations.

Also note that Green Listed will remove duplicates entered in the input box. Thus, it does not matter if you enter a list with multiple entries of the same gene. This means that you easily can include several lists of genes from different sources, and any overlapping genes will only be included one time.

3. Ranking (optional)

If a ranking alternative is included in the reference library the ranking option makes it possible to select a subset of gRNAs for each gene, e.g. the 3 gRNAs with the highest score or all gRNAs with a score above a chosen threshold.

Note that the basis for the ranking score comes from the original publication and can therefore be arranged differently. Usually 100 or 1 is the best score and 0 is the lowest. Note also that different investigators have used different algorithms to define what a good gRNA is, and thus which score it should have. It is therefore hard to compare the scores of different reference libraries with each other.

Ranking order defines which number Green Listed should interpret as the highest. The default setting is “Descending” resulting in that number 100 (or 1 for that matter) is considered higher/better than number 0. Changing ranking order to “Ascending” (thus 0 would be considered higher than 1) could be useful e.g. if the ranking is based on a p-value. You might in such case be interested in retrieving e.g. the 5 gRNAs with the lowest p-value.

With the *Interval* setting you can define that gRNAs in a certain interval should be selected, e.g. everything between 50 and 100. We use this feature to select all gRNAs with a score above a chosen threshold.

With the *Limit to top* setting you can define that the software should select a defined number of gRNAs with the highest score. This becomes very useful as some reference libraries can contain >1000 gRNAs for some genes (e.g. the excellent libraries presented in Chari et al., Nature Methods, 2015).

4. Define the output file name (optional)

   The output files will be found in the folder where your computer downloads files.

5. Sequence Modifications

   The *Adaptor sequence* feature lets you define the 5’ and 3’ overhangs you want for your gRNAs. These adaptors should be designed to match your cloning strategy.

   The *Trim* feature allows you to remove a number of bases from the 5´or 3´ end of the gRNA. This is useful if the reference library contains unwanted sequences; some libraries e.g. contain the PAM sequence, which needs to be removed, or if you want to change the first base to a G which has been suggested for better expression.

   The *Identify sequences* feature allows the identification of particular sequences of interest in the gRNA/spacer sequence. This is particularly useful if you are using a restriction enzyme based cloning strategy as gRNA/spacer sequences containing the restriction site will be cleaved in the cloning process and thus lost from the final library. If you want to identify if e.g. a BsaI site (ggtctc) is found in the selected gRNAs, input BsaI:ggtctc. If ggtctc is found in the forward or reverse sequence of the gRNA “BsaI” will be specified in the output list next to the gRNA. Note that you can exchange “BsaI” to whatever you want. If you input hallo:ggtctc in the Identify sequences box “hallo” will be found in the output file if ggtctc was identified.

6. Press run!

   As all necessary parameters have been inserted, press run and wait for Green Listed to complete the job. When it is done press “Download Result” and open the downloaded folder. Do not press the “New job” button before downloading the result because the designed library will then be lost.

7. Output

   Several output files will be found in the downloaded folder. The files and folder will contain the output name you defined (or a default name). Note that the files are in .txt format

which can be hard to read depending on how you open them. A suggestion is to open the .txt file with a text editor (such as Notepad), copy the text as it is and then paste it in excel to get a clearer division in columns.

The *main output file* (with the name that was specified by the user) will contain the suggested custom library as well as information about the different sequences, including suggested short names, complementary sequences, and GC%. The "gRNA target Sequence with adaptors" column contains the suggested single stranded gRNA sequences that could be ordered (preferentially as an oligo pool) to generate the library.

The *"Compact"* output list contains a condensed list defining the genes and the number of gRNA for each gene that was specified by the user.

The *"CutShortList"* is only relevant if ranking is used and a specific number of gRNAs was to be retrieved. The list presents which genes Green Listed was not able to find as many gRNAs as were expected. For instance, if 5 gRNAs/gene were expected, but only 2 gRNAs for gene A were identified this will be identified in the CutShortList. This information could then be used to e.g. rerun the genes with less stringent parameters or with a different reference library.

The *"NotFoundList"* contains a list of genes that were not found in the reference library. This most often can be attributed to that different names were used to define a gene, or there were spelling mistakes. It can also be related to that e.g. a human list of genes were run against a mouse reference list or similar, as not all genes (and gene names) overlap between different species. This information can be used to rerun the non-identified genes with a different reference library. If the gene id column or gRNA sequence column were not correctly identified, the NotFoundList will contain all the genes that you entered.

The *"UserInputParams"* list contains information about the user specified parameters of the library.

8. Troubleshoot

- If Green Listed does not function, make sure your uploaded reference list is a .txt file (tab delimited). If the reference library is in Excel format, just save it as a tab delimited text file.
- If Green Listed crashes as you upload a tab delimited text file reference library, make sure that the Excel file you made it from is not open at the same time.
- If you get an error message using Green Listed (e.g. due to some parameters were missing when you pressed run) it could be good to refresh the webpage and add the parameters again.
- If Green Listed does not identify any suggested gRNAs for your gene list, make sure that the Gene ID and gRNA sequence columns are defined correctly in the Reference List box.
- When you upload a reference library, note that Green Listed needs to have specified "The file has been uploaded and ready to be processed" before you press run. Sometimes this can take a little while if the reference file is very big, the internet connection is not optimal, or the server is occupied.

9. Other comments

Although we use Green Listed to design custom CRISPR screens, it can also be used for other data mining activities trying to retrieve information linked to a searchable term. This could for example include designing shRNA/SiRNA libraries.

Green Listed is only to be used for projects adhering to consensus ethical guidelines.

10. A selection of great CRISPR screen papers that provide reference libraries

- **Chari R** et al. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Chari et al. Nature Methods 2015. link
- **Doench JG** et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nature Biotechnology 2014. link
- **Doench JG** et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology 2016. link
- **Koike-Yusa H** et al. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature Biotechnology 2014. link
- **Konermann S** et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature, 2015 link (gain-of-function screen)
- **Mali P** et al. RNA-Guided Human Genome Engineering via Cas9. Science 2013. link
- **Sanjana NE** et al. Improved vectors and genome-wide libraries for CRISPR screening. Nature Methods 2014. link
- **Shalem O** et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014. link
- **Wang T** et al. Genetic screens in human cells using the CRISPR-Cas9 system. Science 2014. link
- **Wang T** et al. Identification and characterization of essential genes in the human genome. Science 2015. link