Development of the CasCADE CRISPR Detection guide RNA Design Pipeline

Price, Colin; Davis, Phil; Shteyman, Alan; Clark, Brian; Russell, Joe; Bradford, Elaine; Lucas, Julie; Winegar, Richard

MRIGlobal, Gaithersburg MD

Background

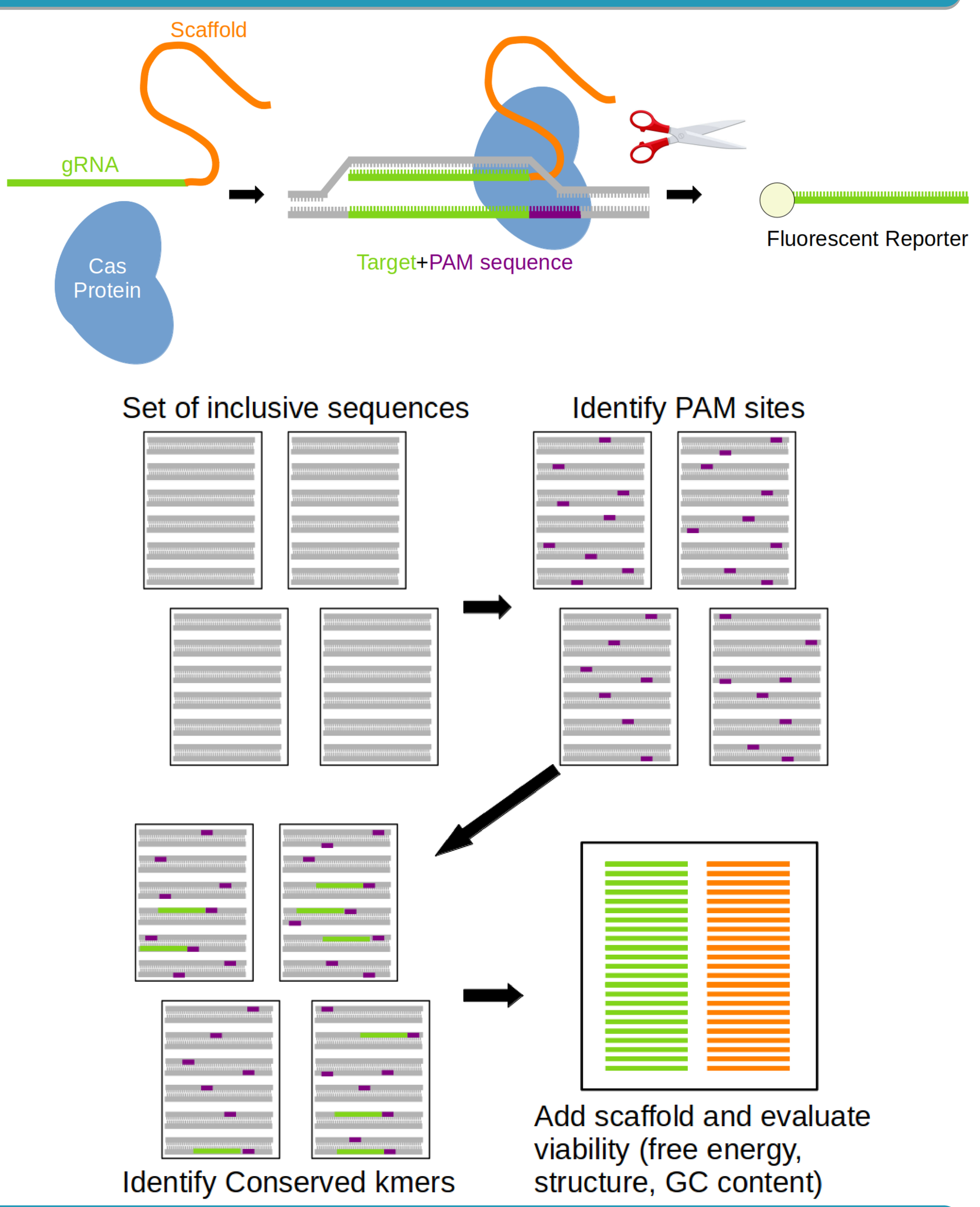
The use of CRISPR as a molecular detection system is emerging as a viable alternative to established methods such as PCR. The adaptation of CRISPR proteins, particularly Cas12 and Cas13, for diagnostic applications leverages their ability to bind and cleave target DNA or RNA sequences, signaling the presence of specific genetic material[1][2]. CRISPR detection assays may be a more desirable technology in some applications relative to PCR, such as biosurveillance[3]. CRISPR detection has shown significant promise in biosurveillance for identifying pathogens with high accuracy, which is crucial for infectious disease response[3]. CRISPR detection has higher specificity due to lack of non-specific amplification, higher sensitivity due to requiring less genetic material for detection, and quicker time-to-completion without requirement of time consuming amplification cycling[4]. However, much like PCR primers, specific guide RNA (gRNA) sequences must be developed per genomic target which can be a lengthy process without computational assistance.

The Cas-CRISPR Automated Design and Evaluation (CasCADE) bioinformatics pipeline is a first of its kind gRNA designer that allows for the selection and filtering of key criteria to include inclusion and exclusion groups, Cas protein type, scaffold sequence, Protospacer Adjacent Motif (PAM), and k-size. CasCADE provides a quick time to answer by finding regions of conservation of sequences the PAM sequence. Conserved sequences are evaluated for properties such as structural stability, free energy, and GC content. CasCADE provides output statistics for the inclusivity to in-group and exclusivity to out-group, which can be generated by a user with data in-hand or can be retrieved with CasCADE functions for TaxID inputs. CasCADE's flexible and time-to-answer in a matter of hours for most tasks make it a premier tool in gRNA assay design.

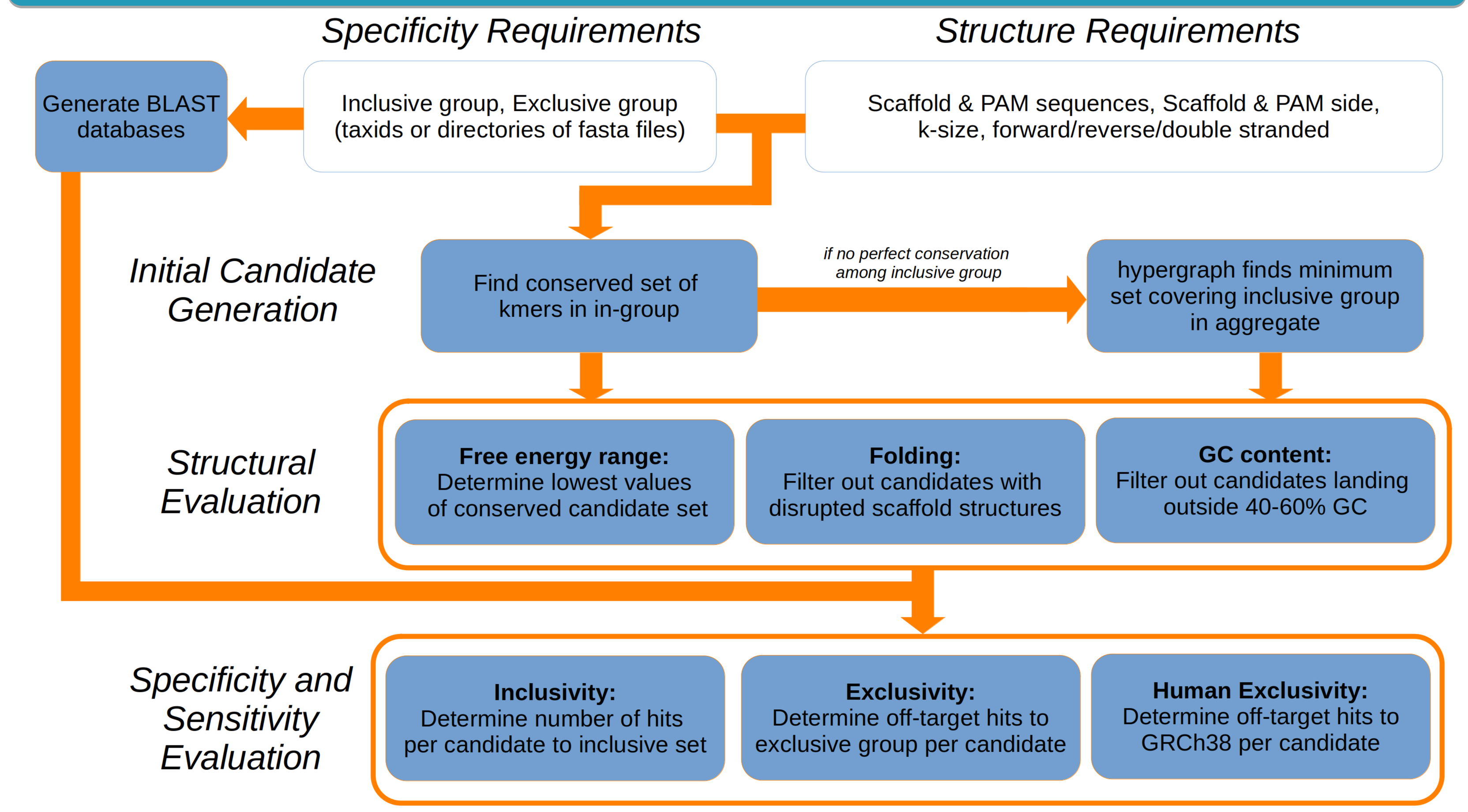
Methods – CRISPR for Detection

CRISPR-based molecular detection systems utilize a guide RNA combined with a scaffold sequence that attach the Cas protein to the target DNA or RNA[1][2]. The presence of a Protospacer Adjacent Motif (PAM) sequence in the target DNA allows the Cas enzyme to recognize and bind to the specific site[2]. Once bound, the Cas protein can trigger a reporter mechanism, such as fluorescence, to signal detection visually or quantitatively[1][2].

Candidate designs are identified by finding the set of kmers that occur in the correct position relative to a PAM sequence match per inclusive group record and are then iteratively inner joined until only fully conserved records remain. The scaffold sequence is then concatenated to each candidate and ViennaRNA is used to evaluate free energy and structural stability properties.



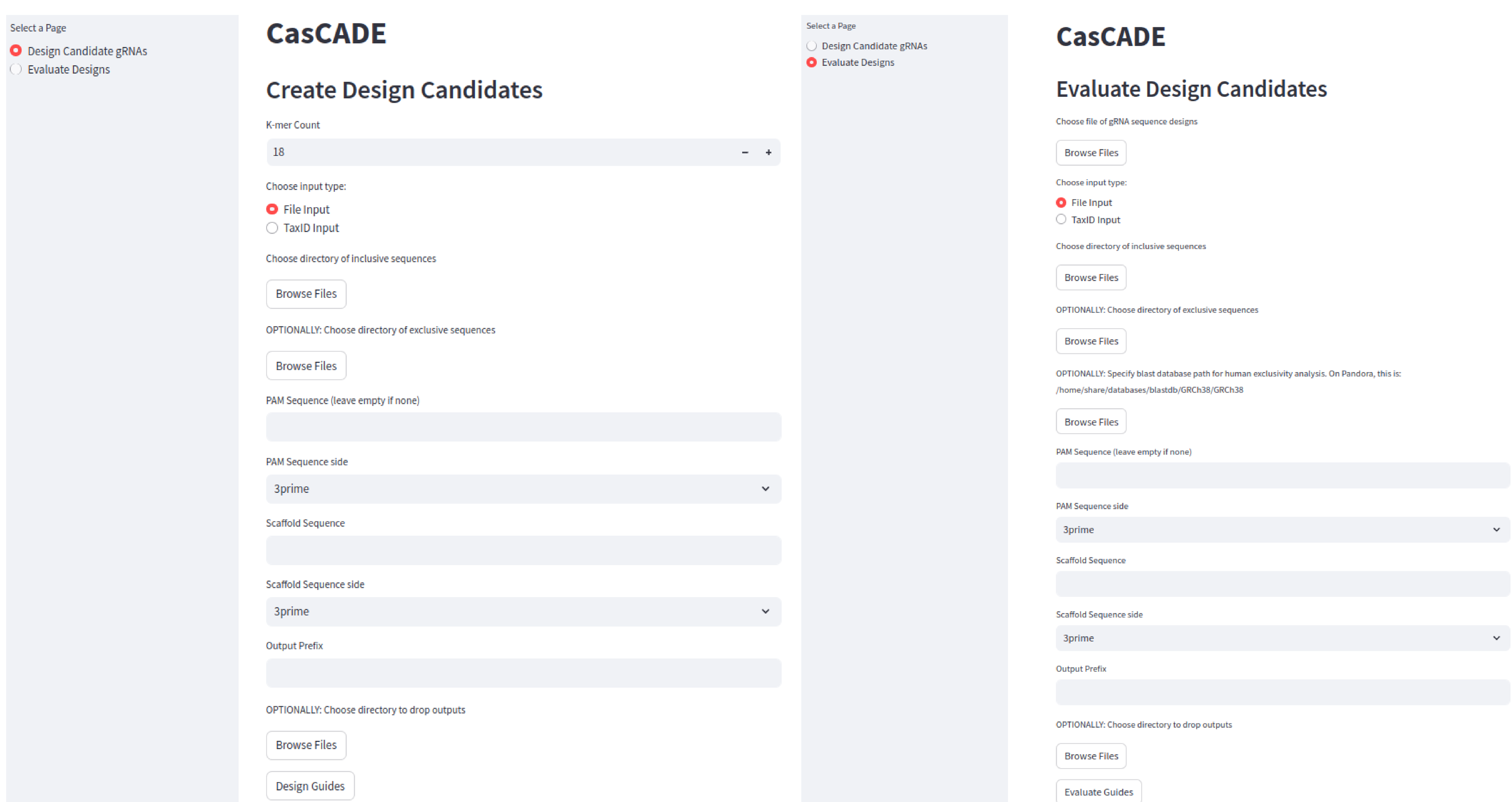
Methods – CasCADE Workflow



The CasCADE pipeline first identifies candidates that are conserved to the in group using set operations on kmer counts. Candidate structural characteristics are then calculated using and filtered. Finally, BLAST is used to evaluate specificity and sensitivity of the remaining candidates.

Methods – CasCADE GUI

In addition to the CLI version of CasCADE for computationally experienced users, CasCADE offers a Streamlit implemented GUI that streamlines the CasCADE pipeline into two simple point-and-click forms. The first form accepts design criteria including the k-mer size, directory or taxid of the inclusive group, an exclusive group, PAM sequence, PAM sequence side, scaffold sequence, and scaffold sequence side in order to end-to-end yield designs with the filtered down best structural, GC, and free energy characteristics. Users can then submit those designs or designs made elsewhere to the second form to sensitivity and specificity. This can again be done using taxid or file browsing input, and can also be screened against a GRCh38 database to confirm human exclusivity.



Results

	gRNA_ID	gRNA_sequence	GC_content	scaffold_structure	scaffold_free_energy	scaffold_seq	gRNA_structure	gRNA_free_energy	scaffold_structure_preserved
2	bmal_genomes_cascade.10092	AAUUUCUACUAAGUGUAGAUGCAACGGCUUUCAUCUCUGG	55	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
3	bmal_genomes_cascade.11812	AAUUUCUACUAAGUGUAGAUGAGAUACGUCGGAGAUAAUG	45	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
4	bmal_genomes_cascade.18727	AAUUUCUACUAAGUGUAGAUAGCGCCUUGUUCGACAGUAC	55	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
5	bmal_genomes_cascade.1873	AAUUUCUACUAAGUGUAGAUGCGCCCUUUUUCUUCGACGA	55	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
6	Borrelia_mayonii_designs.14549	UAAUUUCUACUAAGUGUAGAUUGAACUGUAUGCUCACUACA	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
7	Borrelia_mayonii_designs.16760	UAAUUUCUACUAAGUGUAGAUUGCGAAUGACCAUGUUUAUG	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
8	Borrelia_mayonii_designs.5245	UAAUUUCUACUAAGUGUAGAUGCUUUGGUGAUUGAUGUUUC	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	....((((((((.(((....))).))))))))..........................	-4.7	TRUE
9	Borrelia_mayonii_designs.52889	UAAUUUCUACUAAGUGUAGAUGCUCGUUUCCCCAACUAUUU	45	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
10	Borrelia_mayonii_designs.55106	UAAUUUCUACUAAGUGUAGAUUUGCCUUACCUUCAACCAAC	45	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
11	Borrelia_mayonii_designs.7050	UAAUUUCUACUAAGUGUAGAUAAUCUGGCUCCAAACGAUUU	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
12	Borrelia_miyamotoi_designs.16293	UAAUUUCUACUAAGUGUAGAUCACCAACUGCUUCUAAUCGU	45	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
13	Borrelia_miyamotoi_designs.16944	UAAUUUCUACUAAGUGUAGAUUCUAAAAUUCCCCAUCCAAC	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
14	Borrelia_miyamotoi_designs.17030	UAAUUUCUACUAAGUGUAGAUACAACAACACCAGAAUCCUU	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
15	Borrelia_miyamotoi_designs.17626	UAAUUUCUACUAAGUGUAGAUCGUAUCGUCCAUAGUAGUAA	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
16	Borrelia_miyamotoi_designs.8195	UAAUUUCUACUAAGUGUAGAUCCAAGUCAAUAACAAAUCCC	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
17	Borrelia_miyamotoi_designs.9024	UAAUUUCUACUAAGUGUAGAUAUCCUGAAUCGACGUAAUAG	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
18	bthur_designs.1465	AAUUUCUACUAAGUGUAGAUGAGAACGAAUACCACCGCCA	55	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
19	bthur_designs.1541	AAUUUCUACUAAGUGUAGAUCGAAUCAUAGUAUCCCUCCU	45	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
20	bthur_designs.1646	AAUUUCUACUAAGUGUAGAUCAACCCCCGAAUCUUUCACA	50	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
21	bthur_designs.2188	AAUUUCUACUAAGUGUAGAUAAGUCAUGUCCAUAUCCCCU	45	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
22	bthur_designs.2249	AAUUUCUACUAAGUGUAGAUCCCAAAAUGUUUCGGUUUGU	40	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
23	bthur_designs.3022	AAUUUCUACUAAGUGUAGAUUCAUAACGUGUCAUGGUCAU	40	....(((((.....))))).	-4.7	AAUUUCUACUAAGUGUAGAU	....(((((.....)))))......................	-4.7	TRUE
24	c_bot_chromosome.TTTV.20mer.136	UAAUUUCUACUAAGUGUAGAUUCAUCUUCUGGCAUUGUAAG	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
25	c_bot_chromosome.TTTV.20mer.136	UAAUUUCUACUAAGUGUAGAUGGGUGUCUUCUUUUAAGUUC	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
26	c_bot_chromosome.TTTV.20mer.136	UAAUUUCUACUAAGUGUAGAUCCUUCACCUGUUUUCAUUUC	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
27	c_bot_chromosome.TTTV.20mer.831	UAAUUUCUACUAAGUGUAGAUCUAUAUGUCCUUUUCUUGGC	40	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE
28	c_bot_chromosome.TTTV.20mer.831	UAAUUUCUACUAAGUGUAGAUCAAUCCGAAGACCUUCAUCA	45	.....(((((.....))))).	-4.7	UAAUUUCUACUAAGUGUAGAU	.....(((((.....)))))......................	-4.7	TRUE

The final output of CasCADE is a table report of designs with structural analysis (above) and sensitivty/specificity analysis (right). Each sequence is assigned an ID and has the complete sequence listed along with dot-bracket notation representations of scaffold and gRNA structure. Scaffold and gRNA free energy are also listed, as well as a binary column flagging TRUE or FALSE that the scaffold structure is stable. Percent a GC content is also reported.

label	Inclusivity	Num Hits	Num Possible	total_inclusi	# Off Targets	Exclusivity Offtargets	Hits Human
13373	1	68	68	68	1	Burkholderia pseudomallei	FALSE
13373	1	68	68	68	1	Burkholderia pseudomallei	FALSE
13373	1	68	68	68	1	Burkholderia pseudomallei	FALSE
13373	1	68	68	68	1	Burkholderia pseudomallei	FALSE
1674146	1	4	4	4	0	No off-targets	FALSE
1674146	1	4	4	4	0	No off-targets	FALSE
1674146	1	4	4	4	0	No off-targets	FALSE
1674146	1	4	4	4	0	No off-targets	FALSE
1674146	1	4	4	4	0	No off-targets	FALSE
1674146	1	4	4	4	0	No off-targets	FALSE
47466	1	92	92	92	0	No off-targets	FALSE
47466	1	92	92	92	0	No off-targets	FALSE
47466	1	92	92	92	0	No off-targets	FALSE
47466	1	92	92	92	0	No off-targets	FALSE
47466	1	92	92	92	0	No off-targets	FALSE
47466	1	92	92	92	0	No off-targets	FALSE
1428	1	220	220	220	20	Bacillus cereus, Bacillus mycoides, B	FALSE
1428	1	220	220	220	21	Bacillus pseudomycoides, Bacillus ce	FALSE
1428	1	220	220	220	20	Bacillus cereus, Bacillus sp. ABP14, B	FALSE
1428	1	220	220	220	19	Bacillus cereus, Bacillus sp. AR4-2, B	FALSE
1428	1	220	220	220	20	Bacillus pseudomycoides, Bacillus sp	FALSE
1428	1	220	220	220	21	Bacillus sp. ABP14, Bacillus sp. FDA	FALSE
1491	0.035714286	5	140	5	59	No off-targets	FALSE
1491	0.035714286	5	140	5	0	No off-targets	FALSE
1491	0.035714286	5	140	5	270	No off-targets	FALSE
1491	0.964285714	135	140	135	295	No off-targets	FALSE
1491	0.964285714	135	140	1131	1515	No off-targets	FALSE

For the sensitivity/specificity report, the taxonomic or user specified label appears in the label column. Inclusivity is reported as a percentage of hits to total number of records. For generalized exclusivity, total count of off target hits is reported as well as a comprehensive list of the off target species IDs in the taxonomic label case, or the user input custom label. A second file reports the same information, but with accessions as the label. A stand alone column for the human exclusivity analysis reports either TRUE or FALSE.

Discussion and Future Work

CasCADE to date has been used to generate hundreds of candidate gRNAs to detect a broad range of targets including viral, fungal, bacterial, plant, and animal. The lab team at MRIGlobal is working on validating the performance *in vitro* as well as working towards validating a composite gRNA score number, which ranges 1-10 factoring in all metrics from the CasCADE report. As understanding of relevant criteria for *in vitro* comes to light, CasCADE's modular pipeline is well suited to add additional *in silico* design criteria or tweak existing design criteria. When a sufficient number of evaluated gRNA designs are labeled as successful or unsuccessful, the bioinformatics team will begin development of a machine learning model to predict gRNA success *in vitro* based on the reported metrics.

In the cases where perfectly conserved gRNA designs do not exist, the hypergraph method is currently the alternative method to find a minimum set of designs that in aggregate are wholly inclusive to the inclusive group. Other techniques to find inclusive sets of guides include a decision tree-based method that maximizes coverage for a specific total number of designs and method for allowance for degenerate IUPAC characters in design for scenarios like SNP detection.

MRIGlobal would like to acknowledge funding from the DARPA Biological Technologies Office as part of the Detect It with Gene Editing Technologies (DIGET) program funded under the Naval Information Warfare Center contract N66001-21-1-4048 which is awarded to MRIGlobal. The authors thank Craig Willis, Pamela Winegar, Landon Adebiyi, and Sarah Pope for programmatic support.

1. Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., ... & Zhang, F. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. Science, 356(6336), 438-442. doi:10.1126/science.aam9321.
2. Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science, 360(6387), 436-439. doi:10.1126/science.aar6245.
3. Myhrvold, C., Freije, C. A., Gootenberg, J. S., Abudayyeh, O. O., Metsky, H. C., Durbin, A. F., ... & Sabeti, P. C. (2018). Field-deployable viral diagnostics using CRISPR-Cas13. Science, 360(6387), 444-448. doi:10.1126/science.aas8836.
4. Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., ... & Chiu, C. Y. (2020). CRISPR–Cas12-based detection of SARS-CoV-2. Nature Biotechnology, 38, 870-874. doi:10.1038/s41587-020-0513-4.

Contact Information

Colin Price
T. 240-361-4073
E. cprice@mriglobal.org

MRIGlobal
65 W Watkins Mill Rd
Gaithersburg, MD 20878