Development of Cell Lines for use in COVID-19 Drug Development and Research

London McGill, MS

Staff Scientist II, Gene Editing Institute, ChristianaCare

Background

In late 2019 the SARS-CoV-2 virus sparked the global COVID-19 pandemic, with the first United States cases being seen in early 2020 eventually leading to over 7 million cases throughout the country, and accounting for over 600,000 COVID-19 related deaths.[1,2] To address this novel virus, new testing technologies were developed encompassing the unique RNA and DNA targeting capabilities of the CRISPR/CAS system to detect viral particles. This same CRISPR/CAS technology, however, could also be used to create cell lines modelling COVID-19 which would provide pharmaceutical manufacturers a tool to perform drug screening assays in addition to providing researchers a tool for in vivo research on the mechanism of the SARS-CoV-2 virus that causes COVID-19.

In mid-2020 a surge in COVID-19 cases lead New Castle County, Delaware to establish the COVID CARES grant. The grant aimed to provide funding to reduce the impact of SARS-CoV-2 (COVID-19) within the community, help to protect individuals that were deemed vulnerable to the virus, and provide jobs to those impacted by the virus. Using this grant, six target genes believed to be host factors in the COVID-19 virus were identified, and six knock out cell lines that could be pivotal in helping those researching COVID-19 were created. It was believed that cell lines without these genes would be beneficial tools in COVID-19 research. A preliminary study using CRISPR-based loss-of-function screenings within the genome found numerous host genes implemented in the ability of COVID-19 to infect the host.[3] The genes identified play roles in viral binding, spike cleavage and membrane fusion, endosome recycling, and Golgi and endoplasmic reticulum trafficking. Another group released a separate study in which a CRISPR genome-wide screening of infection and cell death was done for SARS-CoV-2, SARS-CoV-1, and MERS-CoV using Vero-E6 cells,[4] this study brought to light genes implicated as viral receptors, facilitation in viral entry, and genes participating in antiviral activity. These two studies, in addition to others, lead us to move forward with our six genes (Table 1). Vero-E6 (kidney epithelial cells from the African Green Monkey) were used as the host cell line, due to their popularity and common use as cell-based infection models for SARS virus research.[4] The Vero cell lines also have an increased presence of ACE2 receptors,[7] which was identified as one of the targets for this work.

Table 1. Knock Out Cell Line Targets[3,5,6]

Gene Symbol	Gene Name	Presumed Mechanism in SARS-CoV-2 (COVID-19)
ACE2	Angiotensin I Converting Enzyme 2	Receptor for binding of SARS-CoV-2 via endocytosis
ARIDIA	AT-Rich Interaction Domain IA	Helicase and ATPase nuclear receptor binding

CTSL	Cathepsin L	Lysosomal cysteine proteinase spike cleavage and membrane fusion
JMJD6	Jumonji Domain-Containing Protein 6	Protein hydroxylases or histone demethylase
KDM6A	Lysine Demethylase 6A	Helicase and ATPase chromatin remodeling
SMARCC1	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin Subfamily C Member 1	Helicase and ATPase chromatin remodeling

CRISPR/Cas System

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/ Cas9 (CRISPR-associated) system is derived from a naturally occurring defense system found in bacteria.[8] Bacterial protection from bacteriophage and other viruses is mediated by small RNAs that are present within this system. These RNAs detect and act on foreign DNA from bacteriophage and pathogens that invade the bacteria.[8] This adaptive immunity used by bacteria has been well studied and the naturally occurring system has been repurposed into a model that can be used at the bench by scientists to edit the eukaryotic genome. Although the CRISPR/Cas technology is among the newest of the genome engineering tools, it has proven pivotal to the successful modification of mammalian genomes on a large scale and with a much shorter time frame than seen with previous genome editing technologies such as Zinc Finger Nuclease and TALENS.

As the technology has become more popular, new Cas proteins have been identified, each with their own unique capabilities. Of these proteins, SpCas9, has become the primary protein used for eukaryotic cell genome engineering work.[9] This highly studied protein cleaves double-stranded DNA (dsDNA) bluntly, which then allows for an insertion or deletion to be made at the double-strand break or 'cut site' due to nonhomologous end-joining (NHEJ). To achieve this, specific RNAs known as gRNAs are designed to carry the SpCas9 to the cleavage site using the same premise originally seen in bacteria. The gRNA contains of a 20-nucleotide sequence that binds to the DNA of interest. PAM restrictive sequences traditionally consist of NGG, where N can be any nucleotide, and repeat with high frequency throughout the genome. The 20 nucleotides that follow act to create a unique binding site and help to prevent the CRISPR/Cas from working elsewhere in the genome.

gRNA Design

To develop the most efficient knockout for these targets, a dual gRNA method was used. When using a dual gRNA construct to inhibit or 'knockout' a gene, the proximity of the two gRNAs to one another and the reading frame disruption created by the gRNA's must both be considered. A reading frame consists of a set of three nucleotides (or nucleotide triplicates) that create a codon and code for an amino acid, whereas an open reading frame (ORF) is considered the segment of triplicates from the start codon (methionine) to the stop codon (nonsense codon: UGA, UAA, or UAG). Any disruption to the genomic sequence can cause a nucleotide change in a triplicate, and then induce an early stop codon downstream which would shorten the ORF.

For these projects the gRNAs were designed so that the SpCas9 molecules from each complex would not interfere with one another's activity when binding to and cleaving the DNA, thus nsuring that both Cas molecules can access the DNA simultaneously and ensuring that they will both actively cleave the DNA without impeding one another's function. When the gRNAs are located too closely to one another, the gRNAs can work against one another and lead to a lower knockout efficiency, because only one molecule can access the DNA at a time and therefore one gRNA could potentially correct the activity of the other gRNA. These gRNAs were also designed to account for a single base insertion that is commonly seen during the double stranded break repair mechanism after CRISPR/Cas targeting. This single nucleotide insertion can inadvertently shift the disrupted reading frame back into place, leading to the complete ORF being transcribed and the protein being made. To ensure that a frame shifting deletion was made in the DNA, codons were targeted based on the 'two thirds rule', where segments equaling n number of triplicates plus two thirds of a triplicate were targeted (Figure 1). This allowed for the deletion of two nucleotides within the targeted reading frame and would still leave a one nucleotide discrepancy in the reading frame if an additional base pair was added during the double strand break repair mechanism.
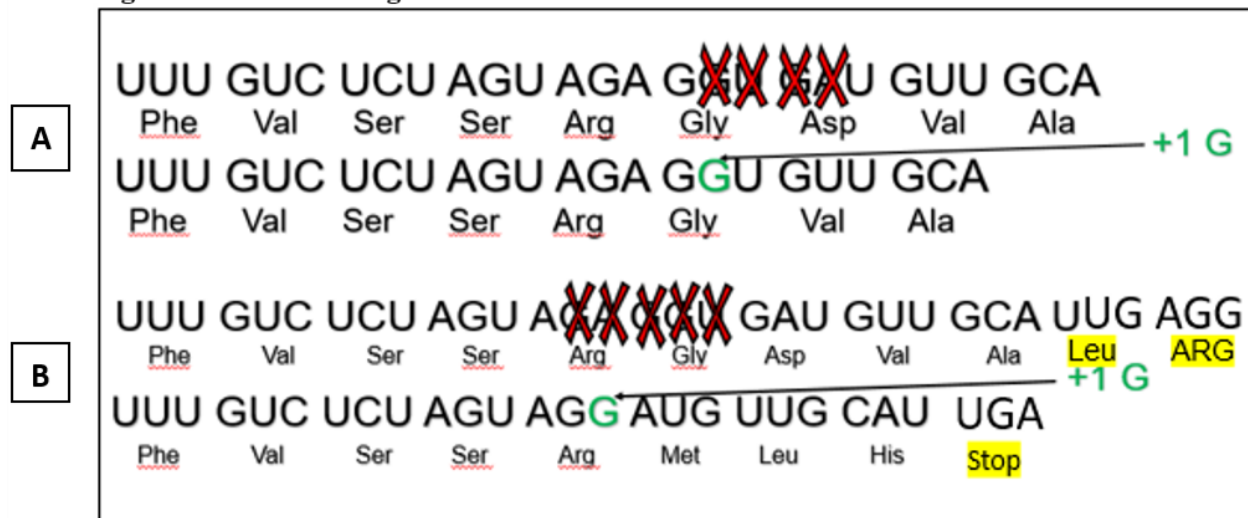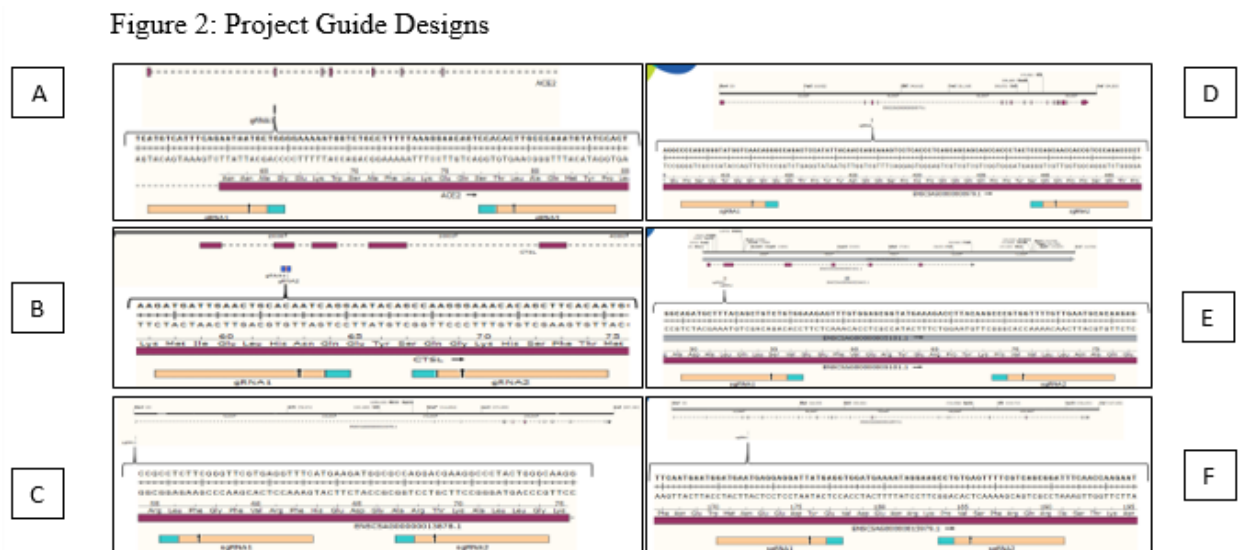
Figure 1. Deletion Design

Figure 1 gives a visual representation of the process used to design a frameshifting deletion. The DNA sequence and the amino acids that are coded for can be seen. A) The one base pair insertion is not accounted for, so when this indel pattern occurs the DNA sequence is knocked back into frame and the protein is still made. B) The one base pair insertion is accounted for and even when the indel pattern occurs the DNA sequence remains shifted out of frame and an early stop codon is induced.

of a gRNA, and the location within the gene that the gRNA will target. gRNAs were selected that had no off-target activity and targeted within the exon of interest (Figure 2). An off-target location is any region within the genome where a gRNA can act on the DNA outside of the anticipated target, typically a gRNA that has fewer than three mismatches in the sequence will be avoided to help ensure that off-target activity does not occur. While aligning the gRNA, the algorithm will show the exact codons that the gRNA will target within the gene, as well as the exon that the codons are within. Finally, the activity level of the gRNA is a measurement of how well the gRNA will act on and cleave the DNA at the target cut site. The algorithm that Synthego's program uses can predict activity levels, but the gRNA activity is truly tested by doing a single gRNA transfection where only the gRNA of interest is used, and the amount of wild type or untouched DNA is measured within the cell. A gRNA that has a higher activity level is traditionally predicted to work better than a low activity gRNA within the cell, but we have case studies that show otherwise.

Figure 2. Project Guide Designs

Figure 2 shows the guide designs for each project. The PAM sequences are identified by the short red or teal portions, and the guide sequences are identified as the yellow or royal blue portions. A) ACE2 guide sequences with PAM site oriented towards one another. B) CTSL guide sequences with PAM site oriented towards one another. C) KDM6A guide sequences with PAM site of guide 2 oriented towards guide sequence of guide 1. D) ARID1A guide sequences with PAM site oriented towards one another. E) JMJD6 with PAM sites oriented towards one another. F) SMARCC1 with PAM site oriented towards one another.

To produce our cell lines, the CRISPR/Cas must first be introduced into the cells. In order to do so the CRISPR/Cas is first assembled, using the gRNA designed with the method above and the purified SpCas9 protein. They are complexed into a ribonucleic protein (RNP) and delivered into the cell nuclease via electroporation, a method which uses electrical current to create temporary pores in the cell membrane. These pores allow the RNP to pass through the cell membrane and into the cell nucleus, where it is able to perform the cleavage. From here, the populations of transfected cells are left to recover. Once the cells appear healthy, a sample of the population is taken and analyzed for 'efficiency,' or the amount of the population that is showing change to the DNA. In this case, the efficiency was based off the population showing a frameshift, which

was determined by Sanger sequencing at the genomic level. From here this information is used to perform statistical analysis.

In the case of the cell lines being produced here, we used the methodology described above, and were able to achieve a high knockout efficiency across all six projects: with the knockout efficiency in the transfected populations averaging 80.95%, and efficiencies ranging from 63% to 100% knockout of the gene across the different gene targets. Using this information, a statistical estimate for clonal sequencing was made using the known allele number for the parental cell line in addition to the transfection efficiency:

Knock Out Efficiency^n = Chance of K0 Clone

(where n represents the number of known allele copies in the cell line)

From here an estimate was made of how many clones would need to be screened to find one that contained a homozygous knock out of the desired gene, keeping in mind that some cell lines including the Vero-E6 line have a varying allele number (n) meaning the chance must be viewed as a statistical range (Table 2). At this point, the bulk populations of transfected cells were sorted into 96-well plates with a seeding density of one cell per well. The cells once expanded are considered cloned as they are all genetically identical. During the clonal expansion process, some cells are harvested for their DNA, which undergoes Sanger Sequencing to determine if the knockout is present within that clonal cell line. Upon completing Sanger sequencing, an inhouse program called DECODR: Deconvolution of Complex DNA Repair is used to align the clonal sequence to the original parental cell line sequence. This program allows us to see when indel patterns are present and we can then determine if the desired indel pattern has been achieved. When a clone is identified as having a complete or 'homozygous' knockout, it was further expanded and banked. At the point of banking the cells are frozen at (-80) degrees Celsius using a controlled freeze method for short term storage and moved into liquid nitrogen where they will remain long term until thawed to be used for research and drug assays.

Table 2. Evaluation of Clonal Chance

Gene	Efficiency	Allele Variant	Chances of Finding a Clone
ACE2	.633	2	40.1%
ACE2	.633	3	25.4%
ACE2	.633	4	16.1%

Table 2 uses ACE2 as an example for the evaluation of the chance of finding a clone. The allele number and the chance of finding a clone are based on the number of alleles present in the cell line. This is used in determining how many clones to sequence, as well as how many 96 well plates to sort cells into at the beginning of the project. This process was done for all six projects.

At the completion of developing all six cell lines, the number of clones sequenced was compared to the number of clones that were found, this data was then compared to the statistical estimate made in the beginning of the project (Table 3). Discrepancy in actual efficiency seen and estimated efficiency can occur when only a small amount of the clonal population has been sequenced or 'screened' like in this case, where only a certain number of clones were needed prior to the sequencing process being terminated. Discrepancies can occur due to the growth characteristics of the cells, variance in copy number among cells, the selection process among

plates where the healthier looking clonal colonies are expanded out of the 96-well plates first, initial cell death at time of single cell plating, and several other external factors.

Table 3. NCC Vero-E6 CARES Act Grant Results

Gene Symbol	Total Clones Screened	Homozygous Clones Identified	Knock Out Efficiency		
			Estimated from Bulk	*Predicted* Clonal	Actual Clonal
ACE2	28	6	63%	16.1% - 40.1%	21.4%
ARID1A	12	2	68%	21.3% - 46.2%	16.6%
CTSL	11	3	68.2%	21.6% - 46.5%	27.2%
JMJD6	15	10	100%	100%	66.7%
KDM6A	23	16	95.5%	83.2% - 91.2%	69.5%
SMARCC1	5	4	91%	68.5% - 82%	80%

Table 3 shows the sequencing results and clonality for each of the six projects. Highlighted efficiencies represent those in which fell the estimated efficiency.

Clones are sequenced as they expand, meaning when clones containing the desired mutation grow at slower rates than others containing the mutation, the slower growing clone may never make it to the point of sequencing, thus lowering the actual efficiency unless the entire population of clones is screened. During screening we also find many times that clones will contain allele copies that have the deletion but will also contain one or more allele copies that are non-frameshifting. When these heterozygous clones occur, the gene is still often able to be transcribed and a functional protein is made therefore a knockout is not seen.

Conclusion

Genome editing technologies have quickly developed over the years from small reactions that took months, to larger scale reactions that can be done in hours and yield results in just a few days. The CRISPR/Cas technology, the newest and most efficient of these technologies, came into play around 2013[8] quickly improving the capabilities of gene editing technologies but much is still being learned about it. Harnessing this technology allows for many potential disease cures and research models that will continue to improve the global healthcare system. As part of the fight against genetic disease and viral pathogens, we have proven that it is possible to efficiently knockout the genes that are believed to be responsible for the SARS-CoV-2 virus to infect the host allowing the global pandemic to persist. It is our hope that these cell lines that have been created will be used to help find pharmaceutical therapeutics and uncover the viral mechanisms that will allow the medical field to put an end to the global pandemic that has been responsible for so many deaths.

Ms. McGill can be contacted at London.mcgill@christianacare.org

References

1. Awadasseid, A., Wu, Y., Tanaka, Y., & Zhang, W. (2021, January 1). Current advances in the development of SARS-CoV-2 vaccines. *International Journal of Biological Sciences*, *17*(1), 8–19. PubMed https://doi.org/10.7150/ijbs.52569

2. Centers for Disease Control and Prevention. (2020, April 3). *Provisional Death Counts for Coronavirus Disease (COVID-19)*. https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm

3. Daniloski, Z., Jordan, T. X., Wessels, H. H., Hoagland, D. A., Kasela, S., Legut, M., . . . Sanjana, N. E. (2021, January 7). Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell*, *184*(1), 92–105.e16. PubMed https://doi.org/10.1016/j.cell.2020.10.030

4. Wei, J., Alfajaro, M. M., DeWeirdt, P. C., Hanna, R. E., Lu-Culligan, W. J., Cai, W. L., . . . Wilen, C. B. (2021, January 7). Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cell*, *184*(1), 76–91.e13. PubMed https://doi.org/10.1016/j.cell.2020.10.028

5. Kai, H., & Kai, M. (2020, July). Interactions of coronaviruses with ACE2, angiotensin II, and RAS inhibitors-lessons from available evidence and insights into COVID-19. *Hypertens Res*, *43*(7), 648–654. PubMed https://doi.org/10.1038/s41440-020-0455-8

6. National Center for Biotechnology Information. (n.d.). https://www.ncbi.nlm.nih.gov/

7. Ren, X., Glende, J., Al-Falah, M., de Vries, V., Schwegmann-Wessels, C., Qu, X., . . . Herrler, G. (2006, June). Analysis of ACE2 in polarized epithelial cells: Surface expression and function as receptor for severe acute respiratory syndrome-associated coronavirus. *The Journal of General Virology*, *87*(6), 1691–1695. PubMed https://doi.org/10.1099/vir.0.81749-0

8. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012, August 17). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. PubMed https://doi.org/10.1126/science.1225829

9. Hsu, P. D., Lander, E. S., & Zhang, F. (2014, June 5). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262–1278. PubMed https://doi.org/10.1016/j.cell.2014.05.010