Applications and Improvements of CRISPR/Cas9 and Cas12a Gene Editing Technologies

by

Amanda Krysler

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Pharmacology

University of Alberta

Abstract

Background: The CRISPR/Cas9 system has emerged as a revolutionary genetic engineering technology capable of editing various cell types, creating disease models, and more recently, changing human DNA[1]. It consists of three pieces: the Cas9 endonuclease, the gRNA, and the DNA target. In short, the Cas9 protein is 'guided' by the gRNA to cut a specified DNA target. This system has risen above other genetic engineering tools due to the feasibility of programming the 20-nt sequence of the gRNA to target a complementary DNA sequence. This programmability provides researchers with an easy and efficient way to knockout and study proteins of interest, as demonstrated in **Chapter 3** of this thesis. Despite the promise of this new system, there are some major roadblocks preventing easy adaptation to the clinical setting. One of these challenges is the natural variability of our genome, as it limits the researcher's ability to target all versions of a highly polymorphic DNA sequence[2]. In **Chapter 4**, we propose the use of universal bases, ie: bases which can pair with any of the naturally found DNA bases, to target SNPs[3]. This would allow for one gRNA to target all four versions of a DNA sequence instead of having to design and deliver four separate gRNAs. This would further expand applications of the CRISPR system, as outlined in **Chapter 5**.

Methods: In **Chapter 3**, gRNAs were designed to target the early exons of the gene of interest, for each collaborating lab. After the transfection of the desired cell line, the knockout cell populations were screened and sorted with FACs. Each single clone population was then tested for efficient disruption of the genomic DNA and subsequent

lack of the desired protein. In **Chapter 4**, gRNAs with universal and degenerate bases were designed to target highly polymorphic genes: *HLA* and *ABO*. DNA sequences containing naturally occurring SNPs were synthesized for each gene. *In vitro* cleavage assays were performed to determine the activity of Cas9 on each DNA sequence using different modified gRNAs. A specificity profile for all the modified gRNAs was then created using an *in vitro* high through-put assay. The best gRNA from these two assays was further validated by testing in cells. Finally, the best combination and type of modification was applied to an alternative CRISPR system, Cas12a, and used in a detection assay targeting a polymorphic section of the *HIV-1* genome.

Results: In **Chapter 3**, 3 different protein knockout cell lines (CRMP2A, FAM120B, and B4GALNT1) were created and fully validated. One final cell line has a predicted 50% knockdown of SF3B4 protein based on the genomic DNA results, however, further protein validation is yet to be performed. In **Chapter 4**, we show that the addition of universal bases resulted in increased activity at SNP targets which would not have been cut by the current wildtype gRNA. The addition of these bases resulted in selective degeneracy at the DNA positions complementary to the positions of the incorporated modifications in the gRNA. The best gRNA from these experiments contained 3 ribose inosines. Application of these ribose inosines to the Cas12a gRNAs also resulted in successful cleavage of only the relevant SNP targets. This was reflected in the DETECTR[4] assay where a gRNA targeting a polymorphic region of the *HIV-1* genome was used.

Conclusions: As shown in **Chapter 3**, CRISPR/Cas9 knockout cell lines were successfully created for each desired protein and can now be used in functional studies. These will be used to investigate disease pathology, drug pathways, mitochondrial gene interactions, and lung cancer characteristics. In **Chapter 4**, we demonstrate that the incorporation of universal bases results in increased cleavage activity of their respective SNP target(s) without losing overall targeting specificity. This was validated using both *in vitro* and cellular assays and also applied to a detection assay developed for clinical use. Overall, the addition of universal and degenerate bases addresses the problem of targeting the innate diversity of the human and viral genome and introduces additional applications of the CRISPR/Cas9 and Cas12a system for both bench and clinical research. These future applications are described in **Chapter 5**.

Preface

(Mandatory due to collaborative work)

Chapter 3 of this thesis was conducted as collaborations with research labs at the University of Alberta. *Part I: CRMP2A Knockouts* was led by Dr. Evangelos Michelakis supervising Aristeidis Boukouris. *Part II: FAM120B Knockouts* was led by Khushwant Singh Bhullar and co-supervised by Dr. Basil Hubbard. *Part III: B4GALNT1 Knockouts* was led by Dr. Simonetta Sipione supervising Noam Steinberg and Vaibhavi Kadam. *Part IV: SF3B4 Knockouts* was led by Dr. Andrew MacMillan supervising Ayat Omar. For all collaborations: the design, creation, and DNA validation of the CRISPR knock-out cells was done by myself. Protein validation was done by the student(s) involved in the collaboration. Further functional experiments have/are being performed in the respective labs who led each collaboration.

Some of the research conducted for **Chapter 4** was through a collaboration led by Dr. Juan Jovel at The Applied Genomics Core, University of Alberta. The high throughput heat maps referred to in Chapter 4 were designed by myself, with the assistance of Christopher Cromwell, and Dr. Juan Jovel. The data analysis and creation of the code for these figures were conducted by Dr. Juan Jovel. The original template for the heat maps and protocol for the assay was created by Christopher Cromwell[5].

Chapter 4 of this thesis will be submitted as Amanda R. Krysler, Christopher R. Cromwell, Tommy Tu, Juan Jovel, Basil P. Hubbard. “Guide RNAs containing universal

bases enable Cas9/Cas12a recognition of polymorphic sequences". Dr. Basil Hubbard and I conceived the study and designed experiments to examine how the inclusion of universal bases into crRNAs affects Cas9/Cas12a activity and specificity. I performed all experiments studying the activity, kinetics, and thermodynamics of universal-base modified crRNAs *in vitro* and in cells, and was assisted by Christopher Cromwell and Tommy Tu. Dr. Juan Jovel assisted with high-throughput sequencing and subsequent data analysis, as described above. All authors contributed to the writing of the manuscript.

Acknowledgements

I would like to foremost thank my supervisor, Dr. Basil Hubbard, for his mentorship and invaluable guidance throughout my undergraduate and graduate studies. I was incredibly lucky to receive a position in your lab and have grown exponentially as a researcher throughout this degree. I am also grateful for the assistance provided by my committee members, Dr. Harley Kurata and Dr. Toshifumi Yokota. As well, I would like to thank my student mentor in the lab, Christopher Cromwell, for his invaluable encouragement and support when learning new protocols and for always answering my never-ending questions. Finally, I would like to thank my family and friends for always supporting me in my studies. Thank you to my partner for always listening and pushing me to strive for my best self throughout this degree. I couldn't have finished this incredible journey, as the woman I am today, without all of the support I received from the wonderful people in my life.

Table of Contents

List of Tables

List of Figures

Glossary of Terms

Term	Definition
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas9/12a/13	CRISPR associated protein 9/12a/13
gRNA	Guide RNA
nt(s)	Nucleotide(s)
bp(s)	Base pair(s)
SNP	Single nucleotide polymorphism
FACs	Fluorescence-activated cell sorting
DETECTR	DNA Endonuclease-Targeted CRISPR Trans Reporter
TALE	Transcription activator-like effector
PAM	Protospacer motif
crRNA	CRISPR RNA
tracrRNA	Trans-activating CRISPR RNA
sgRNA	Single guide RNA
GWAS	Genome wide association study
dCas9	Deactivated CRISPR associated protein 9
RNP	Ribonucleoprotein complex

NHEJ	Nonhomologous end-joining pathway
HDR	Homology directed repair
indels	Insertions and/or deletions
nCas9	Nickase CRISPR associated protein 9
KO	Knockout
KD	Knockdown
KI	Knock-in
ssDNA	Single-stranded DNA
dsDNA	Double-stranded DNA
SHERLOCK	Specific high-sensitivity enzymatic reporter unlocking
CORDS	Cas12a-based On-site and Rapid Detection System
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
CAR-T cell	Chimeric Antigen Receptor Engineered T cell
PI	Protein immunoprecipitation
tRNA	Transfer ribonucleic acid
siRNA	Small interfering RNA
BNA	Bridged nucleic acid
HIV	Human immunodeficiency virus
A549 cells	Adenocarcinomic human alveolar basal epithelial cells
HEK 293T cells	Human embryonic kidney 293T cells that expresses a mutant version of the SV40 large T antigen

N2A cells	Mouse neuroblastoma Neuro-2a cells
MC3T3 cells	Osteoblast precursor cell line from Mus musculus calvaria
ICLs	Interstrand DNA cross-links
snRNPs	Small nuclear ribonucleoproteins
NS	Nager's syndrome
5-NI	5-nitroindole
3-NP	3-nitropyrrole
rI	Ribose inosine
dI	Deoxyribose inosine
mI	2'O methyl inosine
dN	5-nitroindole
dK	Degenerate base K
dP	Degenerate base P

CHAPTER 1: Introduction

Overview of the CRISPR system

Discovery & Development

Discovery

The discovery of restriction enzymes in the 1970s opened up the molecular biology field to the possibility of being able to cut and edit DNA in a test tube[6]. This along with the discovery that these cut sites can be used to disrupt genes, or incorporate external DNA, allowed researchers to begin editing and characterizing genes[7]. This field advanced further with the study of mega nucleases[8], double-stranded endonucleases which can recognize stretches of 14-40 nts of DNA, zinc fingers[9], zinc ion-dependent small protein motifs that each recognize 3 nts of DNA, and TALE proteins[10], small protein motifs which can recognize a single nt of DNA. While mega nucleases already can make double-stranded breaks, both zinc fingers and TALEs were conjugated to the cleavage domain of the restriction enzyme Fok I, to produce this double-stranded cut in the DNA[11]. The challenge presented by these genetic engineering tools is the required customization of each protein for every new DNA target, which is costly and requires complex cloning, preventing wide adoption of these systems to the scientific community. Therefore, the more recent discovery of the CRISPR system revolutionized the genetic engineering field, as it is simple and adaptable to targeting most areas of the genome, without significant redesign.

The CRISPR (clustered regularly interspaced short palindromic repeat) system was initially discovered in bacteria and archaea as CRISPR repeat and spacer sequences[12,13]. Dr. Francisco Mojic was the first researcher to report that sequences of bacteriophage DNA matched these CRISPR spacers[14]. Based on this finding, he hypothesized these CRISPR sequences played a role in the adaptive immunity of bacteria and archaea. During the same year, it was observed that each spacer contained the same sequence on the end, termed the PAM, which is now known as a critical piece of the CRISPR targeting mechanism[15]. As well, in *Streptococcus thermophilus* bacteria, researchers found novel cas genes that encoded a large protein with predicted nuclease activity, now known as Cas9. Furthermore, Dr. Mojic's original hypothesis was confirmed experimentally by scientists who were investigating how bacteria involved in yogurt creation would respond to phage attacks, a common problem in this industry[16]. They elucidated that CRISPR was an adaptive immune system, using pieces of invading bacteriophage DNA, and incorporating them into a 'memory bank' CRISPR array. Using this array, the bacteria can utilize a CRISPR associated protein (Cas9) to fight off any future attacks. These findings sparked a renewed interest in the system and researchers rapidly filled in the knowledge gaps surrounding the biological mechanism. Firstly, John van der Oost and colleagues found that the spacer sequences were actually transcribed into a targeting piece of RNA termed crRNA[17]. Secondly, Marraffini and Sontheimer demonstrated the target molecule was in fact DNA instead of the previously believed RNA interference mechanism[18]. Thirdly, Moineau and colleagues determined that Cas9 makes a blunt double-stranded break 3 base pairs upstream of the PAM sequence[19]. Finally, the Emmanuelle Charpentier group

discovered the second piece of RNA, termed tracrRNA, which base pairs with the crRNA, and is also necessary for guiding the Cas9 protein[20]. Once the initial biochemical mechanism of the CRISPR/Cas9 system was elucidated, researchers began to determine the functionality of this system in other species and found it to be conserved when cloned into bacteria that did not initially contain it[21]. Based on these findings, the scientific community began harnessing this new system as a tool for genomic engineering.

Development of the CRISPR/Cas9 system for genome engineering

GWAS studies have exposed that many complex and chronic illnesses, such as T1 & T2 diabetes, inflammatory bowel disease, cancers, and obesity, can be connected to monogenetic changes in our DNA[22,23]. In addition, these studies revealed many novel rare disease-causing mutations in the human genome[24]. These discoveries solidified the understanding that changes in our DNA can result in differing disease phenotypes, as well as, report on the involvement of certain heritable genes/biological mechanisms underlying these conditions. With the increasing availability of DNA sequencing technology, researchers and healthcare professionals are now able to better inform treatment decisions of these diseases, based on an individual's genomic information[25]. This knowledge has sparked a growing movement toward individualized medicine. Furthermore, to take advantage of this wealth of genomic information, researchers have been pursuing ways to manipulate the DNA sequences, to improve symptoms, cure disease, and create models to better understand gene regulation and biochemical mechanisms[26]. This pursuit of DNA editing has resulted in the exponential expansion of

the field of genomic engineering and the creation of multiple genomic reference databases[27].

One of the most promising players in the field of genetic engineering is a bacterial defence system that targets phage DNA; the CRISPR-Cas9 system[28,29]. In 2012, two research groups found that by changing only 20 nts of the crRNA, scientists can essentially program Cas9 to cut a chosen stretch of DNA[30,31]. Using this knowledge, Dr. Zhang and his team became the first group to demonstrate genetic engineering of eukaryotic cells was possible using the CRISPR/Cas9 system[32]. They also revealed that it can be programmed to target multiple loci in one CRISPR array and it could drive homology-directed repair by providing the cells with an exogenous DNA template. Based on these results, scientists have now adapted the CRISPR/Cas9 system to successfully target a wide range of organisms including mice, plants, zebrafish, as well as the human genome[33]. This system has been refined and harnessed to allow for better targeting of disease-causing mutations, as well as knocking out/in proteins of interest, for easy cell and mouse disease models[34]. As well, this system allows for easy modification of plant genomes which have led to the creation of nutrient-rich and weather-resistant versions of common plants[35]. In conclusion, due to the ease of programmability and the simplicity of the system compared to its predecessors, CRISPR/Cas9 has become the foremost leader for effective engineering of genomes.

Biochemical Mechanism

The CRISPR/Cas9 system is composed of three main components: 1. Cas9, 2. gRNA, and 3. DNA target. As shown in **Figure 1**, the Cas9 protein is a DNA endonuclease

which results in a blunt double-stranded cut on the targeted DNA sequence. The gRNA is made up of both mature crRNA and tracrRNA[31]. The mature crRNA is normally transcribed from spacer sequences in the bacterial genome and is composed of both the 20 nts DNA targeting sequence and a small repeat section which base pairs with the tracrRNA[20]. The tracrRNA is normally responsible for the maturation of the crRNA and Cas9 recruitment in bacteria. This non-coding dual-RNA guide system is necessary for the Cas9 to target and cleave the intended DNA sequence. However, for a simplified design, these two RNA pieces can be combined into a simpler chimeric piece of RNA called a sgRNA[31]. Cas9 contains two nuclease domains (RuvC and HNH)[36]. RuvC cuts the non-target DNA strand and HNH cuts the target strand. Both of these domains require magnesium ions to catalyze the cleavage reaction. These regions can also be mutated to form “deactivated” Cas9 (dCas9) by single amino acid changes in the HNH (H840A) and the RuvC domain (D10A). This mutated form retains binding capacity while destroying the protein’s endonuclease ability. As well, one of these amino acid mutations, in either domain, creates a single DNA strand cutting nickase Cas9 (nCas9) which is useful for easier insertion of an external DNA template.

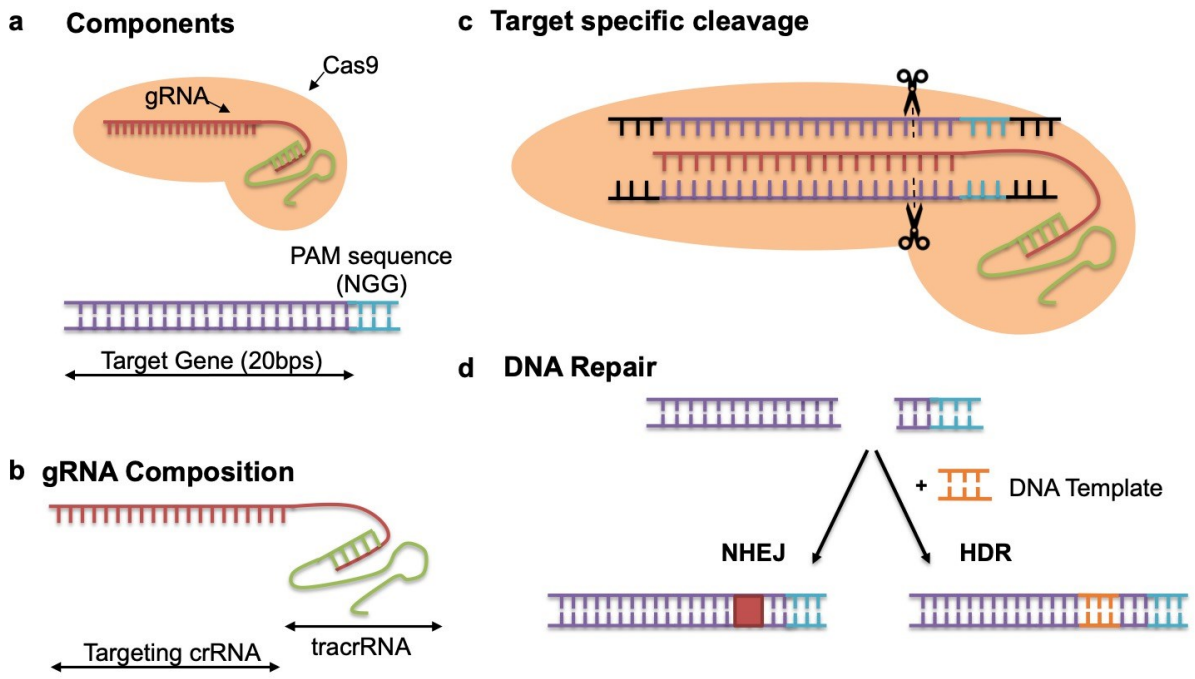


Figure 1. Overview of CRISPR/Cas9 composition and mechanism. a Components of the CRISPR/Cas9 system: consists of the Cas9 endonuclease, gRNA, and complementary DNA target containing an "NGG" PAM directly 3′ of the target region. **b** gRNA composition: crRNA (red) contains a programmable 17-20 nts and a scaffold tracrRNA which connects the crRNA and the Cas9 endonuclease. **c** After locating a complementary DNA target containing a 3′ NGG PAM, the Cas9 endonuclease creates a blunt dsDNA cut, 3 base pairs upstream of the PAM. **d** Cleavage of the DNA results in the activation of cellular repair mechanisms. Firstly, without exogenous DNA material, cells mainly repair these cuts with the error-prone non-homologous end-joining mechanism (NHEJ), resulting in the formation of random insertions or deletions (indels). Otherwise, given a DNA template, the cells can use high fidelity homology-directed repair (HDR) to insert this template into the broken DNA strand.

To initiate DNA cleavage, the Cas9 in combination with the gRNA, form a DNA surveillance ribonucleoprotein complex (RNP)[36,37]. gRNA loading onto Cas9 drives a conformational change from an inactive form to a DNA recognition confirmation. The protein makes extensive interactions with the backbone of the gRNA and places the first 10 nts of the crRNA (seed region) into A-form to prepare for initial DNA interrogation. As well, the PAM-interacting sites of Cas9: R1333 and R1335 are positioned to make contact with the target DNA. One of the hallmark specifications of the CRISPR system is the presence of the 3-4 nts protospacer adjacent motif (PAM) which must be present either 3′ or 5′ of the target region of the DNA, depending on the Cas system used. In the original bacterial immune system, this is used to identify “self” and “non-self” DNA[38]. This PAM is the first piece identified and initiates local DNA unwinding. After finding the PAM, the first 8-10 base pairs of the target DNA strand are probed for matches to the seed region of the crRNA[36,37]. Time spent interrogating a sequence depends on the complementary between the crRNA and DNA and the subsequent formation of an A-form heteroduplex[36,39]. Therefore, this initial region is less amenable to mismatches. If sufficient homology occurs in the seed region, this triggers further invasion of the gRNA to form Watson-Crick base pairs with the target DNA strand[36,37]. This eventually results in an R-loop made up of a pseudo-A-form heteroduplex and a loose non-target DNA strand. This DNA:RNA hybrid duplex is recognized by Cas9 based on its geometry. If the proper formation of the complementary heteroduplex is found, and binding of the non-target DNA strand to the Cas9 occurs, the Cas9-gRNA-DNA complex is stabilized in a closed active conformation. This change moves the HNH domain into a position to cut the target strand of DNA 3 base pairs upstream of the PAM, which in turn brings the

non-target strand into contact with the RuvC domain, leading to concerted double-stranded DNA (dsDNA) cleavage. The RNP remains on the DNA strand until it is recycled by the cell. This can also be achieved *in vitro* by purifying the reaction.

As shown in **Figure 1**, after blunt dsDNA cleavage occurs, the cell attempts to repair the DNA through two mechanisms. The most common repair mechanism utilized is the NHEJ[36]. This is an error-prone pathway that usually results in the formation of random insertions or deletions (indels) near the cut site. Researchers can take advantage of this pathway to create DNA disruptions into a gene of interest. With sufficient disruptions, the cell is unable to transcribe the proper DNA sequence, which ultimately leads to either insufficient or faulty protein translation. However, if the cells are provided with an exogenous DNA template, then the high fidelity homology-directed repair pathway can be initiated (**Figure 1**)[36]. This allows researchers the ability to precisely edit the genome to fix or delete specific pieces of DNA. Both of these pathways can be easily manipulated to create knockout (KO) or knock-ins (KI) to further study the desired protein.

Other CRISPR systems

Cas12a

Variations beyond the CRISPR/Cas9 system are emerging as researchers dive deeper into the world of CRISPR. They are split into two categories: Class I (Type I & III), which utilize multiple-protein complexes, and Class 2 (Type II, IV, V, VI), which are the preferred class for genome engineering, as they only require a single Cas protein[40]. The most common form utilized in research, and the focus of this thesis, is Type II:

Streptococcus pyogenes Cas9. Another notable Class 2 system, which is also utilized in this thesis, is known as Type V: Cas12a/Cpf1. Two versions of this Cas12a/Cpf1 family were discovered to have efficient editing of human cells[41]. These came from *Acidaminococcus* and *Lachnospiraceae* and are named AsCpf1 and lbCpf1, respectively.

Cas12a differs from the original Cas9 mechanism in three major ways: gRNA composition, targeting and cleavage mechanisms, and DNAse activity of Cas12a. Firstly, the smaller Cas12a (~135KDa vs ~160KDa Cas9) only uses a crRNA (42-44 nts) which is matured in a tracrRNA independent mechanism (**Figure 2**)[41]. The targeting spacer region, following a 19 nts repeat, is 23-25 nts long. This is in contrast to the Cas9 system which requires 20 nts of targeting region followed by ~22 nts of repeat sequence[20]. As well, Cas12a requires a 5′ TTTN PAM instead of a 3′ NGG PAM[41]. Genome off-target analysis also shows this system is more stringent for the number of mismatches (1-2) tolerated versus the Cas9 system (5-6)[42]. Secondly, the RuvC-like domain used for cleavage is similar to Cas9, however, Cas12a lacks the HNH domain[41]. As such, single amino acid mutations (D917A or E1006A) completely abolish its cleavage activity. This RuvC-like domain is therefore responsible for the cleavage of both DNA strands and is found to result in a staggered cut, with a 4-8 nts 5′ overhang on the non-target strand (**Figure 2**). This creation of sticky ends is important as it improves the efficiency of homology-directed repair, by allowing for easier integration of the DNA template[42]. Finally, after cleavage of the intended DNA target, Cas12a has the unique ability to non-specifically cleave single-stranded DNA (ssDNA)[4]. The variant with the strongest ssDNase capability, lbCpf1, can also non-specifically cleave double-

stranded (dsDNA) and single-stranded RNA over long periods of time. This non-specific DNase activity has been taken advantage of as a method of detection, by introducing an ssDNA fluorophore/quencher system, to determine whether the intended DNA target has been successfully cleaved. In conclusion, Cas12a provides researchers with more opportunities to fine-tune the genome editing capabilities and potential applications of the CRISPR system.

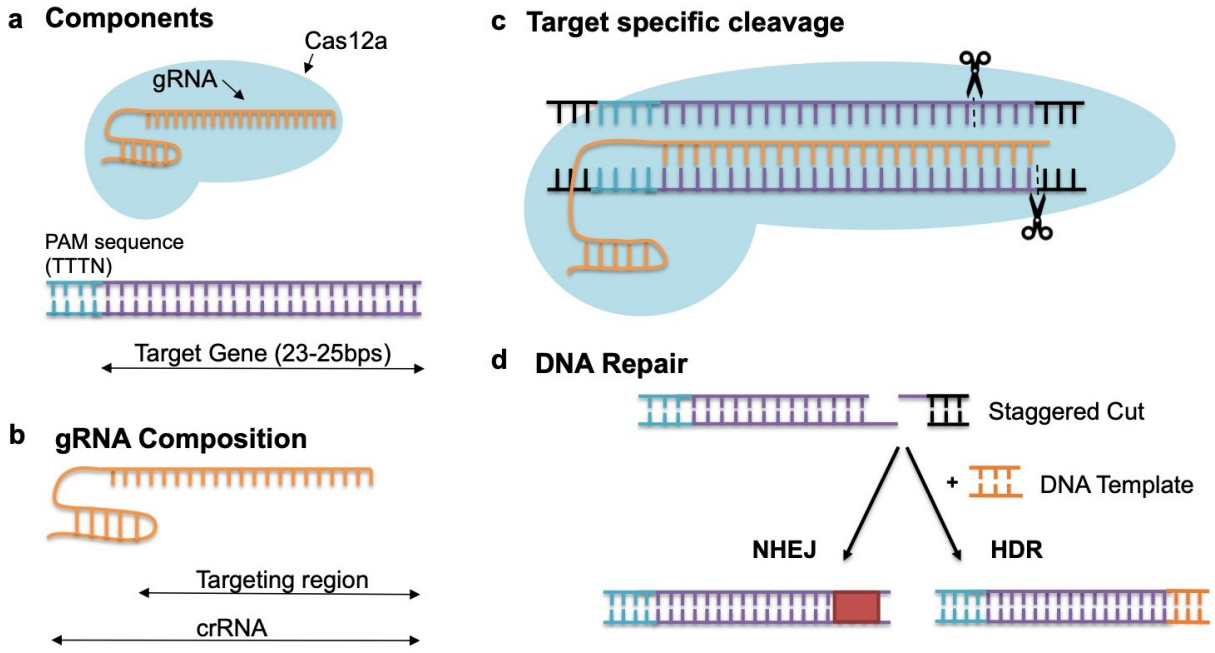


Figure 2. Overview of CRISPR/Cas12a composition and mechanism. a Components of the CRISPR/Cas12a system: consists of the Cas12a endonuclease, gRNA, and complementary DNA target containing a “TTTN” PAM directly 5′ of the target region. **b** gRNA composition: crRNA (red) contains programmable 23-25 nts and 19 nts of a scaffold to interact with the Cas12a endonuclease. **c** After locating a complementary DNA target containing a 5′ TTTN PAM, the Cas12a endonuclease creates a staggered dsDNA cut. **d** Cleavage of the DNA results in the activation of cell repair mechanisms. Without exogenous DNA material, cells mainly repair these cuts

with the error-prone NHEJ, resulting in the formation of random indels. Given a DNA template, the cells can use HDR to insert the template into the broken DNA strand.

Cas13 and others

There are now numerous Cas proteins and homologs which have different specifications than the original Cas9[43]. Notably, Cas13 has been found to target RNA instead of DNA, allowing for the targeting of viral and messenger RNA[44,45]. This protein also demonstrates a non-specific RNase activity, similar to Cas12a, after cleavage of the target RNA strand. Proteins such as CasX[46] and Cas12a[41] have different PAM specifications (TTCN & TTTN) which allow for a wider range of targetable areas of the genome. Both of these Cas proteins also result in a staggered cut allowing for easier insertion of a DNA template[41,46]. However, CasX does not exhibit the ssDNA activity that is seen in Cas12a and is much smaller than either of the other DNA targeting Cas proteins at <1000 amino acids (<100 KDa), allowing for easier delivery of the protein complex[46]. With the discovery of these new Cas proteins, the CRISPR field has incorporated new applications for genome engineering and beyond, involving animals, plants, and viral research[47].

Applications to Research

Genome Engineering

The broadest application for this genetic engineering technology is the creation of disease models. Using CRISPR/Cas9 to KO or KI proteins is efficient and easy to do, compared to previous methods[48]. This method is generally used for the rapid creation of

cells with loss of function mutations or insertion of reporter genes. This allows researchers the opportunity to study how these proteins or mutations impact a disease or biochemical pathways. In **Chapter 3** of this thesis, I portray how to create and validate KO cell lines for 4 different proteins. This is achieved by delivering a ribonucleic complex (RNP), or plasmids consisting of Cas9 and gRNA components, into the cell line of study[48]. The gRNA is designed to efficiently target the region of interest (intron or exon) of the desired gene. After transfection with this RNP complex, cells are screened to determine single clonal populations which have the greatest gene disruption effect, as a result of random indel formation. After verification of DNA indels, the protein levels are examined to determine whether these mutations prevent the cell from creating functional protein. These cells can then be used as a KO model to study different drugs, biochemical mechanisms, and protein function[49]. On the other hand, knock-ins are usually synthesized to report on protein production or specifically change the DNA of a cell. To accomplish this, researchers must provide a partially homologous DNA template for insertion, to initiate the HDR pathway[50]. To improve upon insertion efficiency at the cut site, nCas9[51] or other staggered cut Cas enzymes are used (ex: Cas12a)[42]. After verification of correct insertions, these cells can be used as a method for protein level detection, using a fluorophore marker such as GFP, or to study the effect of specific DNA changes on disease[50]. The CRISPR/Cas9 system can also be applied to plants and animals to create protein KO or KI *in vivo* models[52,53].

Besides the creation of indels, CRISPR/Cas9 has been modified to regulate gene expression and make single base changes in genomic DNA. This is achieved by coupling dCas9, which retains binding ability without cleavage activity, to transcription

effectors, epigenetic modifiers, or fluorescent markers[28]. As well, using just dCas9 and a gRNA complex, researchers can regulate gene activity through a process called CRISPR interference or activation[54]. This involves designing a gRNA to target a transcription factor or regulator binding site and recruiting dCas9 to subsequently block the ability for genetic regulators to bind and modify transcription of the gene(s). In addition, nCas9 can be used in conjunction with base editors to allow for single base changes such as C to T or A to G[55]. These methods allow for more direct DNA editing, without the need for a double-stranded break, preventing potential off-target cleavage made by the wildtype Cas9. Using these Cas9 variants, researchers can now fix mutations, modify epigenetics of a DNA strand, and influence genetic regulation.

Clinical Applications

Due to further development of the CRISPR system and modifications of the Cas9 protein, the capabilities of this genetic engineering tool have become more diverse and widespread. They now include the ability to modify crops, detect disease, and create in vivo gene regulation and modification[56]. As well, different versions of the CRISPR system provide unique platforms for cutting and pasting human genomic DNA in order to fix disease-causing mutations or engineer immune cells to improve targeting of certain cancer cells[28]. In fact, scientists have already started applying this technology in clinical trials[56,57]. These include CRISPR/Cas9 engineering of immune cells such as T-cells[58,59] and natural killer cells[60] to improve upon previous cancer immunotherapy treatments. As well, CAR-T (Chimeric Antigen Receptor Engineered T cell) therapies have become more efficient and easier to create using the CRISPR/Cas9 system[61].

Furthermore, CRISPR is playing a role in potentially treating hereditary diseases such as sickle cell anemia and thalassemia (ClinicalTrials.gov: **NCT03655678**, **NCT03745287**, and **NCT03728322**)[57]. Finally, the development of Cas12a and Cas13 allows for new DNA and RNA sequence targeting capability and hold great potential for efficient diagnostic testing of infectious disease[62]. Overall, CRISPR applications are showing great promise for agriculture, alternative treatment, and detection of disease.

Detection Screens

Recently, applications of the CRISPR system have expanded past the original Cas9 protein. Cas12 and Cas13 have been in the spotlight due to their smaller size and ability to non-specifically degrade ssDNA/ssRNA[63]. This non-specific degradation occurs after targeting of the original oligonucleotide sequence and has been harnessed as a reporter system for effective cleavage. One of these systems, SHERLOCK, utilizes Cas13 and its ability to target small RNA sequences which allows for a wider range of therapeutic targets, such as viral RNA[64]. This has been used to detect attomolar amounts of the Zika virus and infectious bacteria[65]. DETECTR utilizes the DNA targeting Cas12a, in combination with an ssDNA fluorophore/quencher reporter, to inform effective targeting of infectious DNA sequences[4]. In fact, a version of this system, CORDS, has already been used to effectively detect the African Swine Flu[66]. Notably, all these systems use recombinant polymerase amplification (RPA) to simplify the reporter assay[67]. The use of this polymerase system allows for wider clinical use as it does not require a temperature above 37˚C, reducing the need for expensive thermocyclers, and only requires 10 minutes to complete. In fact, field-deployable versions of the SHERLOCK system have

already been developed and tested[68]. Therefore, these diagnostic tools are cheap, sensitive, and allow for the rapid detection of communicable diseases. As well, in light of current events, Dr. Zhang and his group have opened access to their SHERLOCK protocol to allow for potential rapid detection of COVID-19[69]. The DETECTR system was also recently published as a way to rapidly and accurately detect SARS-Cov-2 using Cas12[70]. Finally, a CRISPR/Cas13 prophylactic strategy called PAC-MAN (Prophylactic Antiviral CRISPR in huMAN cells) has been shown to degrade both influenza A and SARS-Cov-2 viral RNA[71]. Overall, through the discovery of Cas12a and Cas13, the CRISPR system can now be used as an efficient means of detection or degradation of viral and bacterial genomes, which as we have seen in recent circumstances, can be invaluable in the midst of an outbreak.

Challenges

Delivery

The CRISPR/Cas9 system is not without its limitations[40]. One of the challenges faced when applying this to clinical use, is the ability to efficiently deliver both Cas9 and gRNA, without detrimental side effects. Current delivery systems are split into three categories; physical (ex: microinjection & electroporation), viral (ex: adeno-associated viruses & lentivirus), and non-viral (ex: liposomes, gold nanoparticles, & cell-penetrating peptides)[72]. The most common approach for cellular work is physical delivery due to its high efficiency, however, *in vivo* work relies on viral vectors as physical delivery damages the cellular membranes[72,73]. More recently, the focus has been on non-viral methods as they eliminate any immunogenicity or toxicity concerns that arise from using

viral particles[72]. This being said, more research needs to be done to increase the efficiency of these methods, both viral and non-viral, as they are highly impacted by the size and nature of their cargo. This is the reason researchers are dedicated to finding smaller Cas proteins and reducing the amount of gRNA necessary to cut multiple DNA targets. Therefore, despite these advances in delivery, there is still a problem being faced by having to target multiple regions of the genome at one time. To target multiple regions or versions of a genetic sequence, researchers must design and deliver multiple gRNAs, without creating any additional off-targets. Even with advances in the creation of multiplex gRNA arrays, researchers are still faced with many challenges[74]. These include the technical challenge of gRNA array design, an increased presence of unwanted off-targets, a limited number of Cas9 molecules, and an overall decrease in individual gRNA efficiency. These challenges would be greatly improved if researchers only needed to deliver a single gRNA. Overall, there is still work to be done to create an efficient delivery system for human applications.

Specificity

One of the main roadblocks for the clinical use of the CRISPR/Cas9 system is the presence of off-targets[1]. Due to the origin of the system as a bacterial defence mechanism against highly mutable viral DNA, it has been found to allow <5 mismatches in the target region, while still resulting in DNA cleavage. These off-targets create a large problem in human genomes as unexpected cleavage of DNA outside of the programmed sequence could result in unknown and detrimental mutations[75,76]. As this system is now being adapted to mammalian genomes, researchers have made efforts

to improve cleavage specificity. These efforts include variations of the Cas9 protein[77], changes to the length and composition of the gRNA[78], and recently, coupling Cas9 to artificial inhibitory domains[79]. Mutations of the Cas9 protein to eSpCas9 (enhanced specificity)[80] and HFCas9 (high fidelity)[81] have resulted in less off-target effects, however, they also exhibit lower activity compared to the original Cas9 in human cells. More recent variants include evoCas9[82], HypaCas9 (hyper-accurate)[83], and Sniper-Cas9[77]. These variants retain both increased specificity and desired activity in human cells. Other approaches to increased specificity involve editing the gRNA[78]. In previous work, we have shown that the addition of bridged nucleic acids (BNAs) can result in more specific targeting[5]. Due to the bulky structure of BNAs, the initial RNA:DNA heteroduplex is forced into the desired A-form, and during the interrogation of off-target sequences, spends less time in both the open/unzipped conformation as well as the closed/fully zipped cleavage competent conformation[5,39]. This shortened dwell time results in an inability to form a cleavage competent conformation on off-target sequences, reducing overall off-target activity. As well, researchers have found that shortening of the crRNA to 17 nucleotides can result in less off-target cleavage[84]. This is due to the shortened non-seed region which is known to be more tolerant of mismatches in the DNA. Finally, using weakened anti-CRISPR proteins, researchers have improved gRNAs which previously created many off-targets, by kinetically insulating the open and closed conformations[79]. Similar to what is seen with BNAs, the slower kinetics of the complex makes it harder for mismatches to result in a cleavage competent conformational change[5,79]. Overall, by using a combination of modified gRNA and Cas9 protein, researchers can reduce unwanted off-target cleavage.

Genetic Variation

Finally, the increase in specificity has outlined a new problem when targeting regions with high mutation rates or areas of the genome which contain SNPs[85,86]. The original system allows for targeting of <5 mismatches, which can result in harmful and unknown off-targets. As well, the prevalence of these off-targets is hard to predict as the number of tolerated mismatches are highly variable between crRNAs. However, the downside of having a highly specific system is the inability to target highly polymorphic sequences. Currently, in order to target a region that may have 1-4 SNPs, the creation and delivery of multiple gRNAs is required. As discussed above, having to design and deliver multiple gRNAs presents many challenges, including reducing the overall efficacy of the system[74,87]. Furthermore, variants in the PAM and the targeting region of the DNA have been shown to destroy both targeting and cleavage capabilities of CRISPR/Cas9[87,88,89]. In **Chapter 4** of this thesis, I will be focusing on how to combat the challenge presented by the innate variation of the human genome[86], as well as the evolving polymorphic bacterial[90] and viral[91] genomes. A typical human genome diverges from the reference genome at 4-5 million sites[92]. The majority of these variations are SNPs. These are single base changes in the DNA and are present in a minority of the population (<50%). Importantly, these changes play a large role in identifying DNA samples, determination of certain phenotypes of genes, and disease variation[93]. Although these are only found in a minority, scientists need to account for these changes, as they can result in being unable to effectively treat or target up to 50% of the population. As well, since these variants can play a role in drug resistance, disease susceptibility, and drug pharmacokinetics, SNPs have become very relevant to individualized medicine[93,94].

Therefore, to create an effective system to combat disease, you have to take care to consider these variants, as they are likely found in patients which could benefit the most from alternative treatments. Overall, there is a necessity to provide a solution to targeting highly polymorphic regions of the genomes without losing the desired specificity of the system.

Scope of the Thesis

Outline of Chapters

We utilized the CRISPR/Cas9 system to create KO cell lines of 3 different proteins: CRMP2A, FAM120B, and B4GALNT1 (**Chapter 3**). As well, a 50% knockdown (KD) of SF3B4 was produced, as full KO of this protein results in lethality. These cell lines will be used by our collaborators to investigate the relationship between genotype and phenotype of different conditions. In **Chapter 4**, we attempt to extend the capabilities of the CRISPR system by incorporating universal bases into the targeting region of the gRNA. The feasibility of using these bases is tested in both the CRISPR/Cas9 and Cas12a system. We demonstrate that universal bases can be tolerated in both Cas9 and Cas12a gRNAs, and can be used to effectively target and detect polymorphic DNA sequences, without losing overall specificity. Potential applications and future experiments using these bases are outlined in **Chapter 5**.

Introduction to the Problem

In **Chapter 4**, we attempt to simplify and advance the applications of the CRISPR/Cas9 system by addressing the problem of genetic diversity in mammalian, viral, and bacterial genomes[62,85,87,89]. Notably, in 2017, the effects of polymorphisms in a human DNA target was investigated in the CRISPR/Cas9 system[2]. The presence of certain SNPs eradicated the activity of multiple gRNAs at that DNA target. This study suggests that innate diversity in the human genome presents a real problem for the effectiveness of the CRISPR/Cas9 system. This is especially true in areas of the genome which have greater diversity, such as genes involved in the immune[95] and blood systems[96]. This is important to note, as these systems are the basis of disorders that could be treated by the use of CRISPR/Cas9, including auto-immune disorders, primary immunodeficiency, and hematological diseases such as sickle cell anemia and beta-thalassemia[49]. Therefore, finding a solution to this innate diversity would improve the potential benefits of this promising genetic engineering tool for human genomes.

Furthermore, another promising application for CRISPR/Cas9 is targeting viral genomes, in order to eradicate or further treat viral-sourced diseases. In fact, viruses such as HIV contain common mutations that create drug resistance to current treatments[97]. This opens up a need for new therapies and CRISPR/Cas9 has taken on the challenge. Unfortunately, viruses such as *HIV-1/2* are highly polymorphic, leading to a problem in consistently targeting their genomes[98]. In fact, CRISPR has already been used to target *HIV-1* and has run into problems with viral escape[99,100]. This is due to the presence of mutations forming around the original CRISPR/Cas9 target site, which facilitates viral escape, and prevents complete eradication of the disease. As well,

studies have further investigated the limitations of targeting a sequence in the *HIV-1* genome using CRISPR/Cas9[101]. They concluded single gRNAs were overall ineffective and there was only a marked decrease in viral activity after the use of 2< gRNAs. In conclusion, developing a way to target a sequence of DNA which contains polymorphisms or mutations, without leading to off-target cleavage, would increase the targeting capacity of the CRISPR/Cas9 system against viral genomes and provide necessary treatments for viral-sourced disease.

Universal Bases

Our proposed solution to the problem of genetic diversity is to replace relevant bases in the gRNA with universal bases. Universal and degenerate bases are bases that can base-pair with all the natural DNA bases or ≥ 2 DNA bases, respectively. Currently, the most 'universal' base is 5'-nitroindole as it does not participate in hydrogen bonding and only interacts through base stacking interactions[102,103]. This means it does not discriminate between pairing with the natural DNA bases (A, C, T, G). Another universal base, which can form hydrogen bonds with all 4 DNA bases, is called inosine[104,105]. It is naturally occurring in tRNAs and is used to create wobble base pairs[106,107]. It is also the by-product of A→I RNA editing[108]. These bases are made up of hypoxanthine and either a ribose, deoxyribose, or 2'O-methyl sugar. While inosine does have the ability to form hydrogen bonds with each natural DNA base, they do have a base pair stability preference of I-C > I-A > I-T ≈ I-G[105,109]. Finally, degenerate bases can indiscriminately pair with a family of bases (purines or pyrimidines)[110]. Degenerate base K is a purine mimic and can, therefore, base pair with both C and T. Whereas degenerate base P is a

pyrimidine mimic and can pair with A and G[111]. Universal bases are traditionally used in PCR primers to allow researchers more flexibility in amplifying regions of unknown or polymorphic DNA[110,112,113,114]. Deoxyribose inosine has also been used in siRNA and has been shown to improve the stability of the overall RNA sequence[115]. As well, degenerate bases were originally developed in order to have a better way to mimic an "N" base as they bring much-needed stability to the DNA duplex[110]. We hypothesize that incorporation of one or more of these bases into crRNA positions, which base-pair with SNP locations in a DNA target, will allow for improved targeting of these highly polymorphic regions.

Research Questions

The following questions will be addressed in Chapter 3 of this thesis:

1. Can CRISPR/Cas9 be used to create a KO of CRMP2A in A549 cells?
2. Can CRISPR/Cas9 be used to create a KO of FAM120B in HEK293T cells?
3. Does a KO of FAM120B impact the interaction of a small synthetic peptide with mitochondrial gene *TFAM*?
4. Can CRISPR/Cas9 be used to create a KO of B4GALNT1 in N2a cells?
5. Can CRISPR/Cas9 be used to create a 50% KD of SF3B4 in MC3T3 cells?

The following questions will be addressed in Chapter 4 of this thesis:

1. Does the addition of universal bases into the crRNA negatively impact overall activity?

2. Does the incorporation of universal bases allow us to target all the polymorphic versions of a CRISPR/Cas9 DNA target?
3. Does the incorporation of the modifications negatively impact the overall specificity profile of the gRNA?
4. Are these modified gRNAs tolerable in cells?
5. Can we incorporate these bases into multiple CRISPR systems with similar results?
6. Can we use these modified gRNAs to detect and report on polymorphic viral targets?

Overall Significance of Research

Chapter 3

By providing KOs of specific proteins, our collaborators will be able to determine if there are any significant correlations between the protein of interest and the phenotype of the disease. This delivers valuable insight into the role these proteins may have in facilitating phenotypes or cellular mechanisms that have not been extensively studied before. For example, Dr. Michelakis' lab will be using CRMP2A KOs to determine if this protein changes the characteristics of lung cancer, which has not been previously investigated. Graduate student Khushwant Singh Bhullar has already determined there is an interaction between his small peptide of study and FAM120B, using an IP (protein immunoprecipitation) assay pull-down, and will be using these FAM120B KOs to further elucidate this relationship. Dr. Sipiones' lab will be using their B4GALNT1 KOs to investigate the unknown relationship between complex gangliosides and extracellular

vesicle secretion. Finally, Dr. MacMillans' student Ayat Omar will be using a 50% KD of SF3B4 to determine the impact of this protein on the phenotypes seen in Nager's syndrome. Overall, these models of loss of functional protein provide each lab with a tool to investigate novel molecular changes and disease pathophysiology.

Chapter 4

Currently, there is no good way to target SNPs in a DNA target without using ≥ 2 gRNAs. Although modifications have been incorporated into the gRNA to improve specificity[78], they have never been used to target polymorphic DNA. As well, universal and degenerate bases have never been incorporated into a crRNA sequence before. Therefore, we are unable to tell how these bases would interact with the RNA:DNA heteroduplex which forms between the crRNA and the DNA target. We can infer from previous applications of universal bases in PCR primers that these modifications may not negatively affect the gRNA base-pairing ability, however, this has yet to be investigated in a system of protein and RNA such as the CRISPR system. In addition, characterizing the cleavage specificity profile of these modified gRNAs will provide more information regarding the impact of incorporating these modifications into the CRISPR system. Finally, being able to target polymorphic bacterial and viral genomes opens up new possibilities for previously developed detection assays. Overall, the investigation of universal bases in crRNAs will provide new insight into the capabilities of modified gRNAs in the CRISPR system.

CHAPTER 2: Materials and Methods

Chapter 3

Creation of KO cells

gRNA were designed to target the protein of interest and were assembled using equimolar amounts of Alt-R CRISPR-Cas9 tracrRNA, labelled with an ATTO 550 (IDT), and Alt-R CRISPR-Cas9 crRNA (IDT) and then annealed as previously described[5]. The sequences of crRNAs designed are listed in **Table 1**. RNPs were created by mixing equimolar quantities of Cas9 (NEB) and the previously assembled gRNA. This was transfected into the chosen cell type using Lipofectamine CRISPRMAX (ThermoFisher), according to the manufacturer's instructions. After 24 hours of incubation at 37˚C, 5% CO_2, cells were sorted based on the ATTO 550 fluorescent marker on the tracrRNA using a BD FACS Aria III instrument (Flow Cytometry Core, University of Alberta). Positive cells were single-sorted onto a 96-well plate and grown for 2-4 weeks. Once confluence of cells reached ~50%, they were moved to a 48-well, and then a 24-well plate. Once the cells on the 24-well plate reached ~80% confluence, they were split 50/50, with half being transferred to a 6-well plate and half spun down for DNA extraction. DNA was extracted from the cells using QuickExtract™ DNA Extraction Solution (Lucigen), following the manufacturer's instructions. Briefly, 500 µL of cells were spun down at 300 x g for 5 mins. The pellet was washed with PBS pH 7.4 (Gibco) and then resuspended in 100 µL of QuickExtract™ (Lucigen). This solution was placed in a thermocycler and the following protocol was run: 65˚C for 10 mins, 68˚C for 10

mins, and 98˚C for 3 mins. The resulting DNA was used directly as a template for PCR reactions amplifying the Cas9 target site. These PCR reactions were purified with QIAquick PCR Purification Kit (Qiagen) and analyzed by Sanger sequencing to confirm the formation of gene-disrupting indels following Cas9 cleavage. PCR primers used are listed in **Table 2**. Clone populations with demonstrated indel formation were further confirmed for protein knock-out by Western or dot blot.

T7 Endonuclease I assay

A bulk sample of all FACs sorted positive cells were kept to determine the efficiency of the gRNA, as previously described[5]. Briefly, 48 hours after FACs sorting, gDNA was extracted from the cell population, as described above. PCR reactions surrounding the Cas9 target site were performed and purified with QIAquick PCR Purification Kit (Qiagen). 200 ng of the resulting PCR product was denatured and reannealed using the following thermocycler program: 95˚C for 5mins, ramp -2˚C/sec to 85˚C for 1 sec, ramp -0.1˚C/sec to 25˚C for 1 sec. 1 µL of T7 Endonuclease I (NEB) was added to the reaction and incubated at 37˚C for 15 mins. The resulting DNA product was analyzed on a 1.5% TAE agarose gel. Indel percentages were calculated as indel (%) = 100 × (1-(1-fraction cut)0.5).

Table 1. Sequences of crRNAs used in the creation of KO cell lines.

Name	Sequence (5′→3′)
CRMP2A-crRNA	rArCrArUrGrCrCrArCrArGrArArUrUrUrCrUrGrCrGrUrUrUrUrArG rArGrCrUrArUrGrCrU

FAM120B-crRNA	rArArGrArGrCrCrArGrArArArUrArCrArGrGrUrUrGrUrUrUrUrArGrArGrCrUrArU
B4GALNT1-crRNA	rCrArGrGrArUrGrCrGrGrCrUrArGrArCrCrGrCrCrGrUrUrUrUrArGrArGrCrUrArU
SF3B4-crRNA	rUrUrArGrArUrGrCrCrArCrGrGrUrGrUrArCrGrUrGrGrUrUrUrUrArGrArGrCrUrArUrArGrGrCrU

Table 2. PCR primer sequences used to validate the KO cell lines.

Name	Sequence (5′→3′)
CRMP2A-F	GACTTAGGGACTGGCAGACG
CRMP2A-R	CCCACCTCTCAAGCTCAAGG
FAM120B-F	CGTACGTATGCCAGCCCTTT
FAM120B-R	TTAGTCAAGGCCAGAGCAGC
B4GALNT-F	AGAGAGGCGGAAGAAAGGA
B4GALNT-R	GTGCGAACTAGGGCCAATTA
mSF3B4-F	GTACCGGACCAGCAAGAATG
mSF3B4-R	TAGCCTGGGGAATCAATGAA

Chapter 4

Design of gRNAs

gRNAs were rationally designed based on clinical polymorphism data for the *HLA*[116], *ABO*[116], and *HIV-1*[117] gene sets. Two of the four SNPs chosen for the *ABO* target site are found in the most common *ABO* alleles and are subsequently linked to changes in blood type[118]. The polymorphisms seen in the *HIV-1* gene set are linked to the formation

of drug-resistant mutations in the protease domain[119]. Cas9 crRNAs were designed based on the presence of a 3′-NGG PAM directly adjacent to a 20bp target site for the *HLA-B* and *ABO* genes. Cas12a crRNAs were designed based on a 5′-TTTN present directly adjacent to the 23 bp target site for the *HIV-1* gene. Sequences for these crRNAs can be found in **Appendix A**.

Cloning of DNA targets

Forward and reverse ssDNA target inserts were designed for Cas9 target sites including SNPs. The oligos used to make the DNA targets are listed in **Appendix B**. The forward and reverse ssDNA sequences were annealed by heating to 95˚C for 5 mins, then cooled to 25˚C over 1 hour. To create the targets, pUC19 (Invitrogen) plasmid and annealed dsDNA target inserts (IDT) were double-digested with *HindIII* and *XbaI* (NEB). These were ligated and then transformed into DH5α *E. coli*. Confirmation of proper insertion was performed by Sanger sequencing.

In vitro cleavage assays (Cas9)

In vitro cleavage assays were performed as previously described[5]. Briefly, plasmid templates containing DNA targets were amplified with pUC19F/R primers listed in **Appendix B**. gRNAs were created by mixing equimolar amounts of tracrRNA (IDT) and crRNA (Genelink) in Nuclease Free Duplex Buffer (IDT), then heated to 95˚C for 5 min and cooled to 25˚C over 1 hour. Sequences for crRNAs and tracrRNA are listed in **Appendix A**. Each reaction consisted of the amplified 5 nM DNA target with 40 nM Cas9 protein and 80 nM gRNA. Initially, Cas9 and gRNA were incubated in 1x NEB 3.1

buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl2, 100 µg/ml BSA, pH 7.9) at 25˚C for 10 mins. DNA template was then added and the reaction was incubated at 37˚C for 3 hours. Reactions were stopped by purifying the DNA with MinElute PCR Purification Kit (Qiagen). Cleavage products were run on a 1% agarose gel and imaged with an Amersham Imager 600 (GE Healthcare). Densitometry was performed using Image J.

In vitro cleavage assays (Cas9 kinetics)

These reactions were performed as described above. Reactions were incubated at 37˚C for 0, 5, 30, 120, and 180 mins.

In vitro cleavage assays (Cas12a)

These reactions were performed as described above with slight modifications. Sequences for crRNAs are listed in **Appendix A**. Each reaction consisted of amplified 10 nM DNA target with 100 nM Cas12a protein and 125 nM gRNA. Reactions were incubated at 37˚C for 30 mins.

Expression and purification of *S. pyogenes* Cas9 and humanized *Lachnospiraceae bacterium* Cpf1/Cas12a

Recombinant Cas9 was purified as previously described[120]. Briefly, *E. coli* Rosetta (DE3) cells were transformed with a plasmid encoding the *S. pyogenes Cas9* gene fused to an N-terminal 6xHis-tag, MBP, and TEV site (Addgene #39312 and #39318, respectively). 25 mL of LB broth containing 25 µg mL^{-1} of kanamycin was inoculated

and grown overnight (~16 hours) at 37˚C. These cells were diluted 1:100 in the same growth media and grown at 37˚C until OD_{600} of 0.8 before moving to 18˚C for 30 mins. Protein production was induced by the addition of 0.5 mM final concentration of isopropyl-ß-D-1-thiogalactopyranoside (IPTG). After induction for 16 hours, the cells were harvested by centrifugation for 15 min at 2700 × g and resuspended in 15 mL/L culture lysis buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 5 mM imidazole, pH 8.0) supplemented with lysozyme and 0.1 M PMSF. This was incubated on ice for 30 mins before being further lysed with sonication (30 secs pulse-on and 60 secs pulse-off for 7.5 mins at 60% amplitude) and centrifuged at 30 000 x g for 1 hour to obtain cleared lysate. This lysate was applied to a 1 mL HisTrap FF Crude column (GE Healthcare) attached to an AKTA Start System (GE Healthcare), washed (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 10 mM imidazole, pH 8.0), and eluted with a single concentration of imidazole (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 250 mM imidazole, pH 8.0). Fractions containing Cas9 were pooled, TEV protease was added, and this was dialyzed into ion-exchange buffer overnight (20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% (v/v) glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA). After dialysis, the sample was centrifuged to remove cleaved MBP. The supernatant was then loaded onto a 1 mL HiTrap FF column (GE Healthcare), washed (20 mM HEPES-KOH, pH 7.5, 100 mM NaCl), and eluted with a 0-50% gradient of NaCl (20 mM HEPES-KOH, pH 7.5, 1 M NaCl). Fractions containing purified Cas9 were concentrated using a 50kDa centrifugal filter (Pall). During concentration, the buffer was exchanged for storage buffer (20 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 1 mM DTT). Concentrated protein was aliquoted and stored at -80˚C.

Humanized Cpf1 was purified as previously described[121]. Briefly, *E. coli* Rosetta (DE3) pLyseS (EMD Millipore) cells were transformed with a plasmid encoding humanized *Lachnospiraceae bacterium* Cpf1 fused to an N-terminal 6xHis-tag, MBP, TEV site, and C-terminal NLS and HA tag (Addgene # 90096). 25 mL of Terrific broth containing 100 μg mL^{-1} of carbenicillin was inoculated and grown overnight (~16 hours) at 37˚C. These cells were diluted 1:100 in the same growth media and grown at 37˚C until OD_{600} of 0.2. This was moved to 21˚C and grown until an OD_{600} of 0.6 before induction with 0.5 mM final concentration IPTG to induce MBP-Cpf1 expression for 14-18 hours. After induction, the cells were harvested by centrifugation for 15 min at 2700 × g and resuspended in 50 mL/L culture of lysis buffer (50 mM HEPES pH 7, 2 M NaCl, 5 mM MgCl2, 20 mM imidazole, pH 8.0), supplemented with lysozyme and 0.1 M PMSF. Cells were lysed and purified as described above. lbCpf1 was stored in Cpf1 storage buffer (50 mM Tris-HCl pH7.5, 2 mM DTT, 5% glycerol, 500 mM NaCl).

Library for high-throughput specificity profiling

Pre-selection libraries were generated as previously described[3]. Briefly, 10 pmol of each partially randomized oligo (IDT, sequences are listed in **Appendix B**) was circularized with CircLigase II ssDNA Ligase Kit (Epicenter). 5 pmol of the circularized ssDNA was used as a template for the Illustra TempliPhi Amplification Kit (GE Healthcare) according to the manufacturer's protocol. The resulting amplified libraries were quantified with a Qubit 2.0 Fluorometer (Invitrogen).

In vitro high-throughput specificity profiling

A specificity profile of the modified crRNAs was created as previously described[3]. Briefly, 200 nM of the pre-selection library was incubated with 1000 nM gRNA and 1000nM Cas9 in NEB Buffer 3.1 for 1 hr at 37˚C to create the post-selection library. In addition, 200nM of the library was incubated with 2U of BspMI using the same reaction conditions as above, to create the final pre-selection library. Both digestion reactions were purified with QiaQuick PCR Purification Kit (Qiagen) and ligated to 10 pmol of barcoded adaptor S50X-F/R (post-selection) or lib_adapter1 with ABO/HLA_lib_adapter2 (pre-selection) using 1000U of T4 DNA Ligase (NEB) for 16hrs at room temperature. Ligation reactions were purified with MinElute PCR Purification Kit (Qiagen) then amplified using primer PE2_short with barcoded primer HLA/ABO-N70X (post-selection) or primer lib_PCR_F with barcoded ABO/HLA_PCR_R (pre-selection) using Q5 Hot Start High-Fidelity Master Mix (NEB). Products were gel extracted and purified using MinElute Gel Extraction Kit (Qiagen) and quantified with a Qubit 2.0 Fluorometer. Finished libraries were run on a HiSeq 2000 (Novogene), demultiplexed, and analyzed as previously described[6]. The sequences used for this protocol are listed in **Appendix B**.

Determination of crRNA–DNA heteroduplex melting temperature

Equimolar amounts of crRNA and complementary ssDNA (**Appendix A and B**) were combined in Duplex Buffer (30 mM HEPES, pH 7.5, 100 mM Potassium Acetate) (IDT) to a final concentration of 2 μM. 100x SYBR Green I was then added for a final concentration of 10x. The solution was added to a CFX96 Real-Time System (BioRad).

The following program was run: 5 min at 95 °C, then cooled to 25 °C at 0.1 °C s^{-1} to anneal the RNA/DNA heteroduplex. The heteroduplex was then heated to 45˚C and subsequently heated at a rate of 0.1 °C s^{-1} to 95 °C. The SYBR Green I fluorescent signal was used to generate a melt curve from which a Tm for each combination was determined.

RFP/GFP reporter assay

Target sites were cloned into a pRGS backbone (PNA Bio Inc.) containing an RFP reporter and 2 out-of-frame GFP reporters[122]. gRNA was annealed as described above. HeLa-Cas9 cells were cultured in high glucose DMEM media with pyruvate (Gibco) supplemented with 10% FBS/1× pen-strep/1× glutamine (Gibco) and 5 µg/mL Blasticidin S HCl (Gibco) at 37˚C in 5% CO_2. Transfection of the HeLa-Cas9 cells was performed using DharmaFECT Duo (Dharmacon), according to manufacturer instructions for the CRISPR system. Cleavage % was determined using an Attune NxT Flow Cytometer (Invitrogen) based on the %GFP+/%RFP+ cells. Sequences used to create DNA targets are listed in **Appendix B**.

DETECTR assay

DETECTR assays were performed as previously described, with minor modifications[4]. Briefly, target constructs were created with a pUC19 backbone as described above. Recombinase Polymerase Amplification (RPA) reactions were performed using the target plasmid constructs as the template and pUC19 RPA F/R primers. This reaction was incubated at 37˚C for 10 mins. 250 nM LbCas12a, 312.5 nM crRNA, and 250 nM

ssDNA-FQ reporter was incubated at 25˚C for 10mins and added directly to the reaction. They were then incubated at 37˚C in a fluorescent plate reader Spectramax i3 (Molecular Devices) for 2 hrs with measurements taken every 2 mins (λ_{ex}: 535nm; λ_{em}: 595nm). The sequences used are listed in **Appendix B**.

CHAPTER 3: CRISPR/Cas9 Knockout Cell Lines

Introduction

As part of my work in Dr. Hubbard's lab, I was responsible for collaborating with labs at the University of Alberta to create cellular models of different protein KOs or KDs, using the CRISPR/Cas9 system. This is achieved by designing a gRNA which will allow Cas9 to target a region of DNA within the early exons of the desired gene. As described previously, due to the error-prone nature of NHEJ, cleavage of the DNA results in the formation of indels. If these indels create a frameshift of the premature stop codon(s) in the genetic DNA, then the translation of functional protein is effectively destroyed. This leads to decreased or null levels of this protein in the cells and can be used as a KD or KO model to study disease.

Part I: CRMP2A Knockouts

Background

Dihydropyrimidinase-related protein 2 (DPYSL2), also known as collapsin response mediator protein 2 (CRMP2), is the second member of the Dihydropyrimidinase-related protein family, which consists of five different proteins (DPYSL1-5)[123]. DPYSL2 has two main isoforms in human cells: the short isoform DPYSL2B and the longer but less prevalent DPYSL2A. DPYSL2A and B have been extensively studied in neurons, where they have been shown to play a role in neuronal polarization[124], regulate axon growth at baseline[125] and after injury[126]. Due to the critical role it plays in neuronal development, CRMP2A has been implicated in neuronal diseases, such as Alzheimer's, schizophrenia,

and multiple sclerosis[123]. However, much less is known about their role in cancer. A study in 2015 found an unusual splice variant of CRMP2A was phosphorylated in certain tumour cells (lung, breast, and lymphoma) indicating an oncogenic mechanism involving CRMP2A protein[127]. To further investigate this role, CRMP2A KO cells were created in A549 (human lung carcinoma) cells.

Results

DNA Validation

To verify the transfection of the FACs sorted cells, a bulk sample of ATTO550 positive A549 cells were taken as a sample of transfection efficiency. Using a T7E1 mismatch detection assay, the percent of indel formation was calculated based on the number of cleaved mismatches found in the re-annealed genomic DNA (**Figure 3**). The presence of indels indicates there were successful cleavage and repair of the programmed target site. Once each clone population was cultured to a sufficient size, the genomic DNA of each clone was screened with Sanger sequencing, to validate the presence of gene interruption. As shown in **Figure 4**, after successful cleavage of the target region, the error-prone DNA repair pathway NHEJ results in copies of DNA that contain random indels. This can result in the presence of “N”s immediately downstream of the target region if only one DNA strand contains the indel resulting from the improper repair. Using the wildtype trace as a control, each clone trace was run through ICE analysis (Synthego), to provide a percent knockout (%KO) score. The top clones from this analysis were handed off for further protein KD validation.

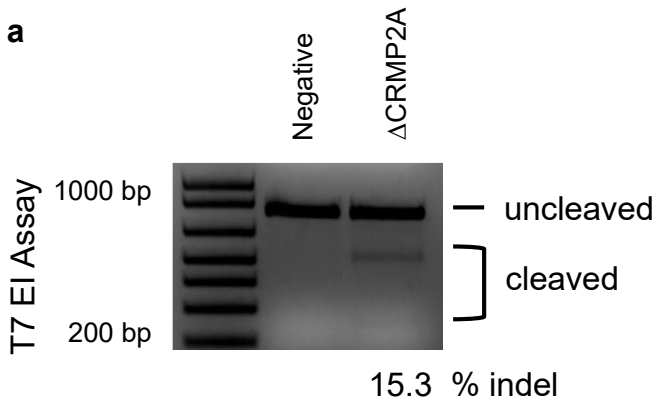


Figure 3. T7EI assay on bulk ΔCRMP2A cell genomic DNA. a DNA extracted from FACs sorted A549 cells was used in a T7EI mismatch detection assay to determine the % indel formation. 200 ng of the purified PCR product was reannealed to create mismatch bubbles that are cleaved by the T7EI endonuclease. Parental A549 cells were used as a negative control. Cleavage product was run on 1.5% TAE gel at 100V for 45 mins. Densitometry was performed with Image J software and %indel was calculated.

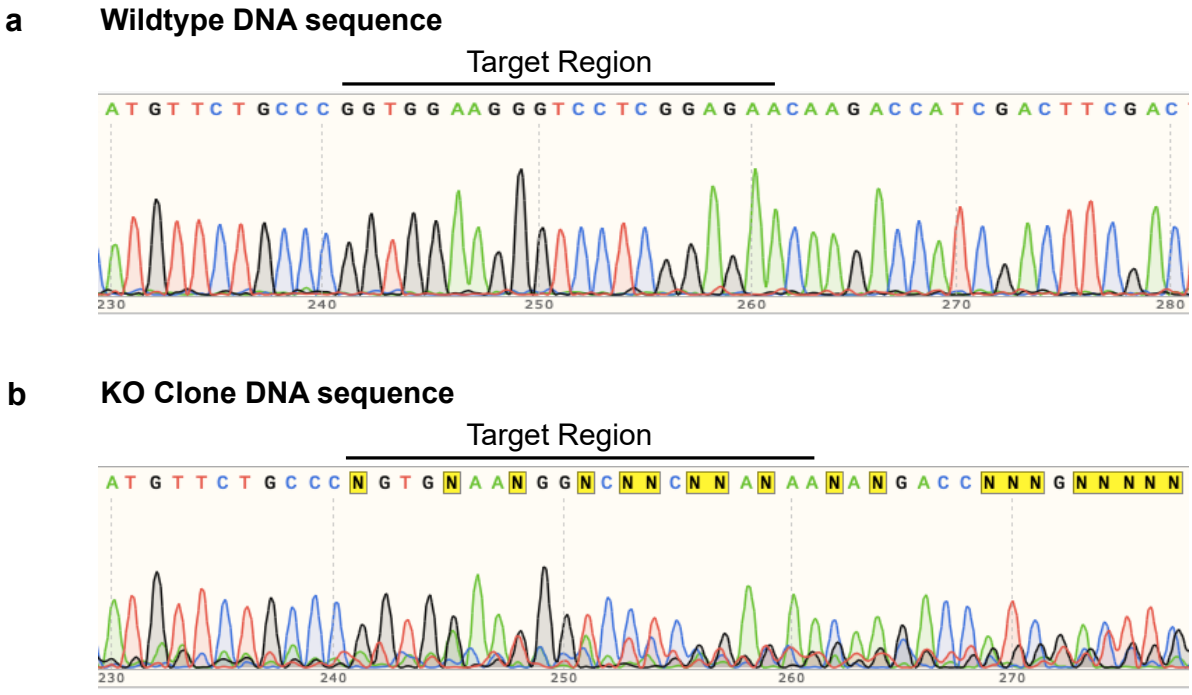


Figure 4. Screening results for KO clone #10. a Representative Sanger sequencing trace for wildtype and **b** FACs selected clone #10 (A549 cells). Region of DNA targeted by crRNA is outlined. “N” indicates the presence of multiple bases. Multiple clones were screened and selected for protein validation based on the %KO score calculated by ICE Analysis (Synthego).

Protein Validation

Protein validation was completed by Aristeidis Boukouris in Dr. Michelakis’s lab. As shown in **Figure 5**, immunoblot blot analysis using a CRMP2A antibody was performed to confirm the complete KO of CRMP2A, compared to parental A549 cells. Clone #10 shows a complete eradication of protein production, suggesting this clone can be used as a model of CRMP2A/DPYSL2 KO (**Figure 5**).

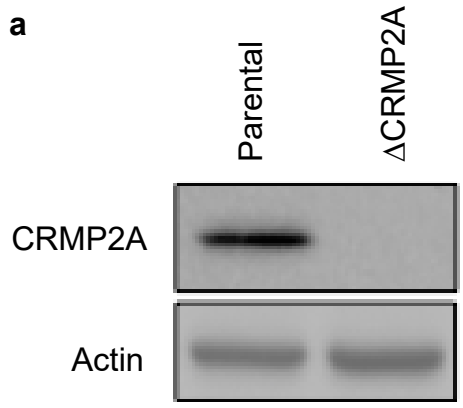


Figure 5. Validation of KO clone #10 protein levels. an Immunoblot using a CRMP2A antibody on A549 lysates. KO clone #10 does not contain detectable levels of CRMP2A compared to parental A549 cells. Actin was used as a loading control.

Future Directions

Further experiments will be completed in Dr. Michelakis's lab using clone #10 as a permanent loss-of-function model in lung cancer cells. This cell line will be used *in vitro* and *in vivo* to study the effect of CRMP2A levels on cell migration, invasion, and metastasis in lung carcinoma cells.

Part II: FAM120B Knockouts

Background

FAM120B, also known as CCPG/SAN1/PGCC1, is localized in the nucleus and is a novel PPARγ coactivator[128]. It is also a vital 5' exonuclease that functions in response to inter-strand DNA cross-links (ICLs). ICLs prevent both replication and transcription and are usually mended by the Fanconi anemia (FA) pathway[129]. FAM120B works with Senataxin, a helicase, in a FA independent pathway, to create necessary resistance to ICLs[130]. However, limited information is available on the detailed function and cellular interactions of this gene. The purpose of this project is to study the effect of a small synthetic peptide on mitochondrial function, possibly via FAM120B. In the preliminary experiments, it was observed that this peptide interacted with FAM120B, as evidenced by a pull-down IP assay. Following the validation of the agarose-bead pull-down assay, we made a FAM120B/PGCC1 KO in HEK293T (human embryonic kidney) cells to further study the relationship between the small synthetic peptide and FAM120B protein.

Results

DNA Validation

Creation and validation of FAM120B KOs was performed as described in Part I, using HEK293T cells. As shown in **Figure 6**, the transfection was successful in creating indels in the target gene. However, due to the low incidence of indels (<15%), many clones were screened to find ones with a complete disruption of DNA. Furthermore, the Sanger trace of clone #2 represents the distinct disruption of the genomic DNA directly downstream of the target region, which was seen in 6 of the clones screened (**Figure 7**). This disruption of the DNA indicates a high chance of protein KO and was further verified by Western blot analysis to measure FAM120B protein levels (**Figure 8**).

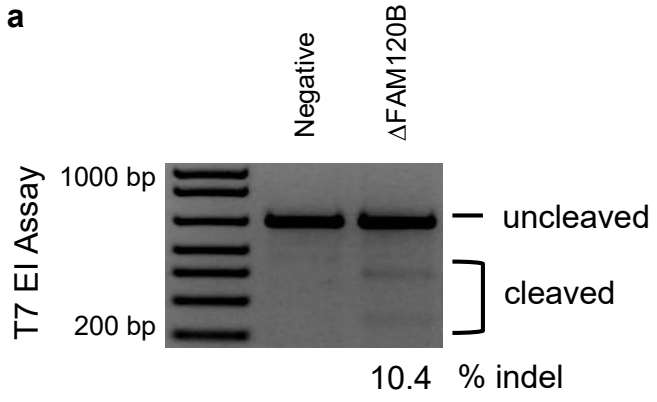
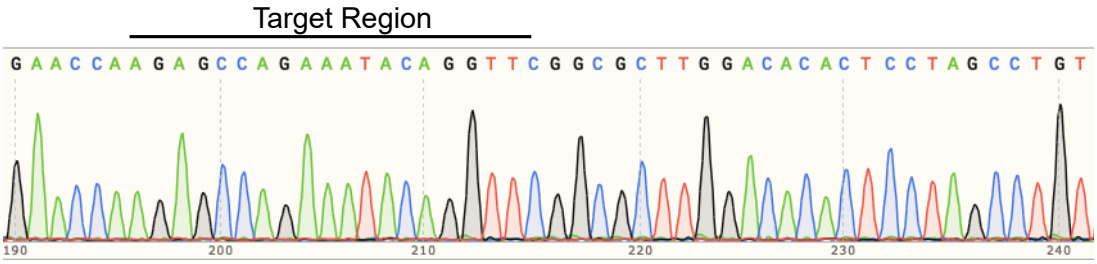


Figure 6. T7EI assay on bulk ΔFAM120B cell genomic DNA. a DNA extracted from FACs sorted HEK293T cells was used in a T7EI mismatch detection assay to determine the % indel formation. 200 ng of purified the PCR product was reannealed to create mismatch bubbles that are cleaved by the T7EI endonuclease. Parental HEK293T cells were used as a negative control. Cleavage product was run on 1.5% TAE gel at 100V for 45 mins. Densitometry was performed by Image J software and %indel was calculated.

a **Wildtype DNA sequence**



b **KO Clone DNA sequence**

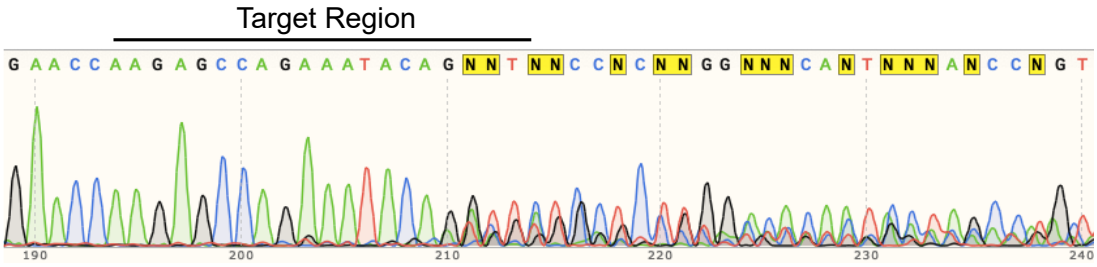


Figure 7. Screening for potential KO clones using HEK293T genomic DNA. a Representative Sanger sequencing trace for wildtype and **b** FACs selected clone #2 (HEK293T cells). Region of DNA targeted by crRNA is outlined. “N” indicates the presence of multiple bases. Multiple clones were screened and selected for protein validation based on the %KO score calculated by ICE Analysis (Synthego).

Protein and Functional Validation

To further validate the potential KOs, Western blot analysis was performed by Khushwant Singh Bhullar, using a FAM120B antibody. Clone #2 and 5 have a complete KO of FAM120B/PGCC1 protein compared to wildtype levels and can be used as a KO cellular model (**Figure 8**). Using clones #2 and 5, functional verification was performed to determine the effect of FAM120B levels on the ability of the small synthetic peptide to activate a mitochondrial gene *TFAM* (**Figure 9**). As shown in **Figure 9**, the peptide

induction of TFAM protein levels in wildtype (WT) HEK293T cells is abrogated by the use of FAM120B KOs.

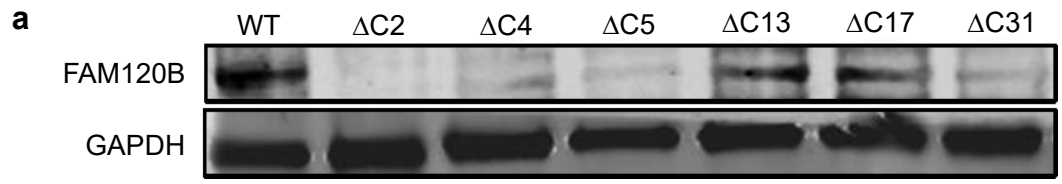


Figure 8. Validation of potential KO clones protein levels. a Western blot using a FAM120B antibody on HEK293T lysates. Clones #2 and 5 (ΔC2 and ΔC5) have significantly depleted levels of FAM120B compared to wildtype (WT) HEK293T cells. GAPDH was used as a loading control.

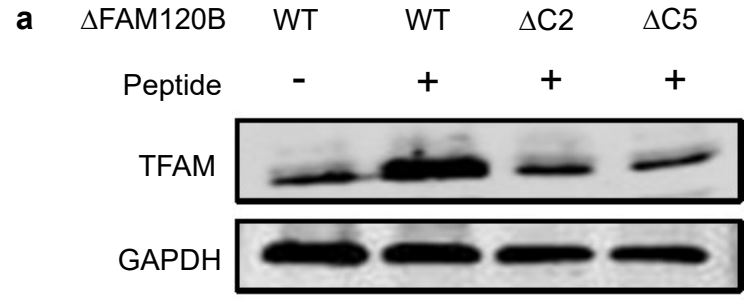


Figure 9. Functional study of FAM120B levels on target gene *TFAM*. an Induction of TFAM levels is seen using 50 μM peptide in wildtype (WT) HEK293Ts. This effect is abolished by the use of FAM120B KO clones #2 and 5 (ΔC2 and ΔC5). GAPDH was used as a loading control.

Future Directions

Further experiments will be performed by graduate student Khushwant Singh Bhullar. He will be using FAM120B KO clones #2 and 5 to further investigate the role of FAM120B in facilitating the effects of a small synthetic peptide, used in **Figure 9**, on

mitochondrial function. To achieve this goal, experiments will explore the effects of this small synthetic peptide via FAM120B on mitochondrial genes and potential drug pathway(s).

Part III: B4GALNT1 Knockouts

Background

B4galnt1 encodes for the enzyme B4GALNT1, also known as GM2/GD2 synthase, found in the ganglioside biosynthetic pathway[131]. It is responsible for synthesizing ganglioside GM2 from ganglioside GM3. This enzyme is therefore essential for the subsequent synthesis of complex gangliosides present in the brain (including GM1, GD1a, GD1b and GT1b). Mutations in *B4galnt1* underlie neurodegenerative diseases in both humans and mice[132,133]. Both gangliosides and extracellular vesicles (EVs) have been indicated in neurodegenerative diseases, such as Alzheimer's pathophysiology, however, research has not been done on the effect they have on one another[131,134]. EVs are membrane-bound nanoparticles shed by most cells[135]. They are involved in cell to cell communication, modulation of the immune system, and elimination of toxic misfolded proteins from cells. Dr. Sipione and her lab are testing whether cellular gangliosides affect the ability of cells to secrete extracellular vesicles (EVs). CRISPR/Cas9 B4GALNT1 KO in N2a (mouse neuroblastoma) cells were created to investigate the relationship between complex gangliosides and EV secretion.

Results

DNA Validation

Creation and validation of B4GALNT1 KOs (coded for by *B4galnt1)* were performed as described in Part I, using N2a cells. Similarly to the FAM120B KOs, the T7EI assay results showed a lower overall indel formation (**Figure 10**). Therefore, multiple clones (20-30) were screened with Sanger sequencing to provide the highest chance of finding clones with a complete KO of protein. As shown in **Figure 11**, clone #16 represents one of the 2 clones which contained a disruption of the genomic DNA and was used in protein validation.

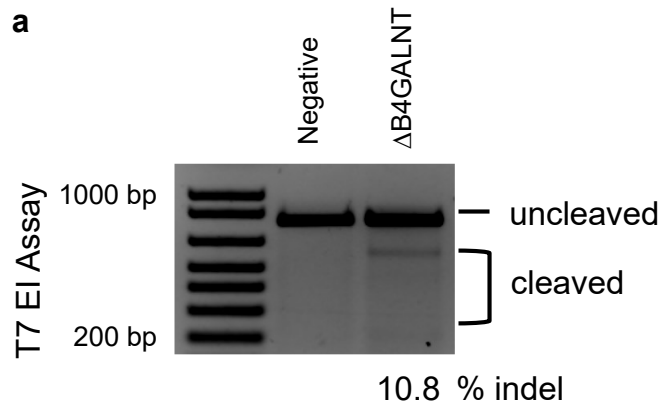
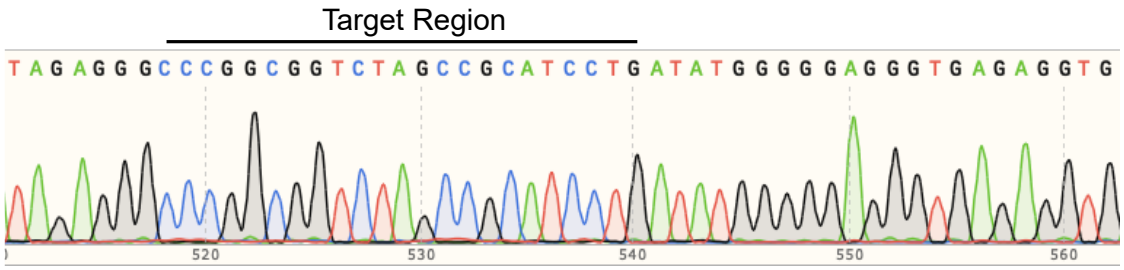


Figure 10. T7EI assay on bulk ΔB4GALNT cell genomic DNA. a DNA extracted from FACs sorted N2a cells were used in a T7EI mismatch detection assay to determine the % indel formation. 200 ng of the purified PCR product was reannealed to create a mismatch bubble that is cleaved by the T7EI endonuclease. Parental N2a cells were used as a negative control. Cleavage product was run on 1.5% TAE gel at 100V for 45 mins. Densitometry was performed with Image J software and %indel was calculated.

a **Wildtype DNA sequence**



b **KO Clone DNA sequence**

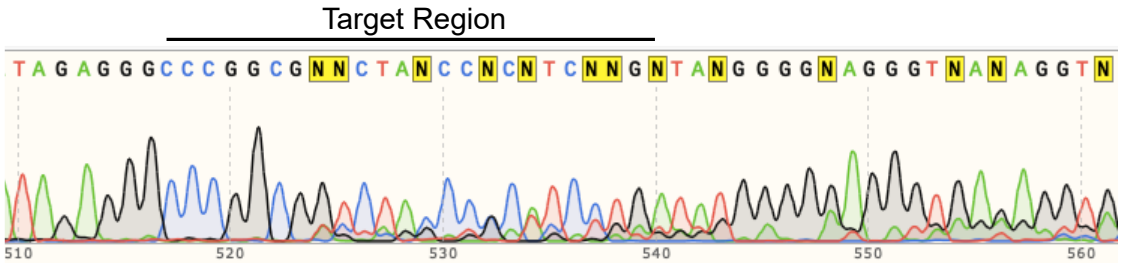


Figure 11. Screening results for KO clone #16. a Representative Sanger sequencing trace for wildtype and **b** FACs selected clone #16 (N2a cells). Region of DNA targeted by crRNA is outlined. “N” indicates the presence of multiple bases. Multiple clones were screened and selected for protein validation based on the %KO score calculated by ICE Analysis (Synthego).

Functional Validation

Figure 12 shows a dot blot analysis using clone #16, created by graduate student Noam Steinberg. Compared to wildtype GM1 levels, the levels of this ganglioside present in the KOs are distinctly lower. This indicates these clone populations contain much less B4GALNT1 and are a functional KD of GM1 production.

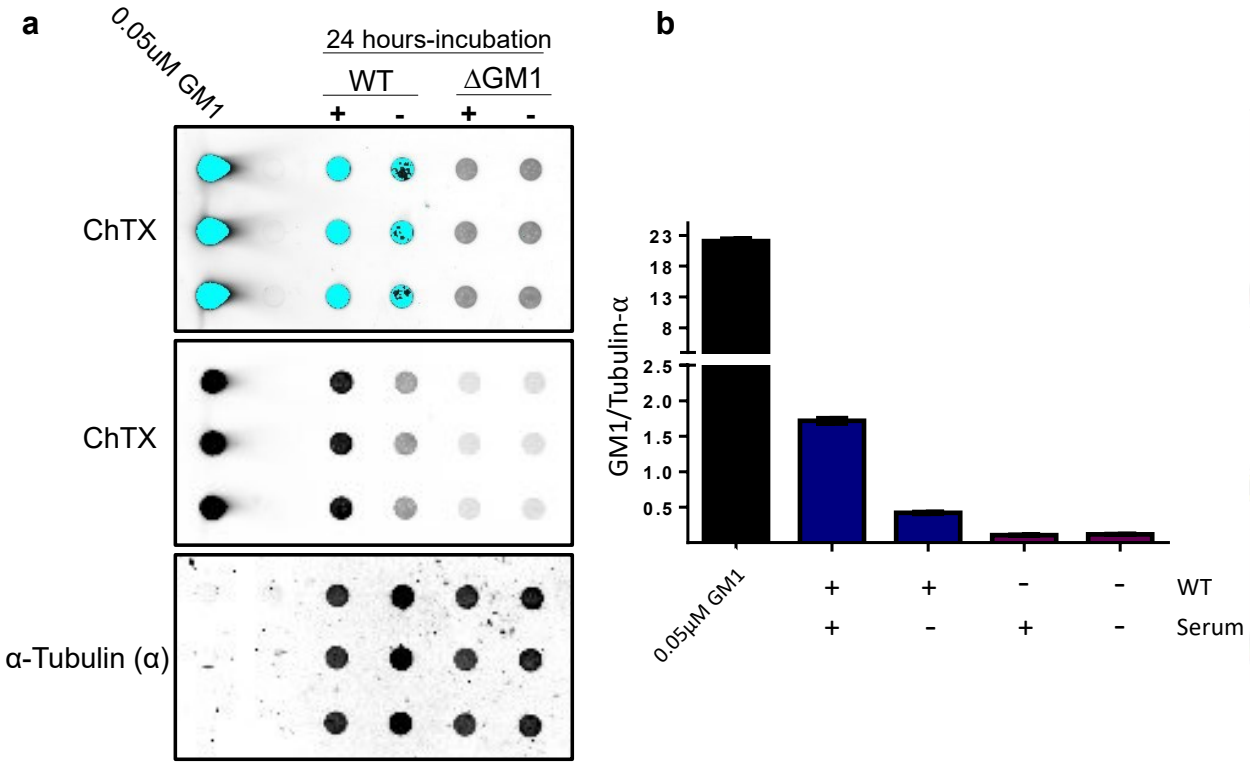


Figure 12. Functional validation of KO clone #16 GM1 levels. a Dot blot analysis of N2a lysates cultured under serum-free conditions for 24 hours. 1 µg protein was dot blotted onto nitrocellulose membrane and incubated with 1:2,500 biotinylated cholera toxin B (ChTX, Invitrogen, C34779) and 1:2,000 Tubulin alpha (#2125 clone 11H10, Cell Signaling). Then the membrane was incubated with 1:5,000 streptavidin 800 (926-32230, Li-Cor) and 1:1,500 Anti-Rabbit IRDye Goat (Licor 926-68071). As a positive control, 0.05 µM of GM1 was loaded to the membrane. The upper panel represents the saturated signal of GM1. Alpha tubulin was used as a loading control. **b** Graph representing the amount of GM1 found under regular and serum-free conditions compared to a positive control (0.05 µM GM1). The amount of GM1 is depleted compared to wild-type (WT) lysates under each condition.

Future Directions

The KO cells generated were used to reinforce studies showing ganglioside synthesis was decreased using an inhibitor of the ganglioside biosynthetic pathway. They also

confirmed that complex gangliosides, which are substantially decreased in B4GALNT1 KO cells, are important for EV secretion. Further experiments will be performed to delve deeper into the connection of complex gangliosides and EVs.

Part IV: SF3B4 Knockouts

Background

Pre-mRNA splicing is an essential process by which introns are excised and exons are ligated to process a pre-mRNA into a mature mRNA, which subsequently encodes the functional proteins[136]. This process is carried out by protein machinery known as the spliceosome. The spliceosome is composed of uracil rich snRNPs (small nuclear ribonucleoproteins): U1, U2, U2/U6 and U5. SF3B4 is an essential component of the heptameric SF3B subcomplex of the U2 snRNP[137]. SF3B4's homolog depletion in yeast is lethal and human SF3B4 knock out cells never survive[138]. Recently, haploinsufficiency of SF3B4 was reported as a cause for Nager's syndrome[138,139]. Nager's syndrome (NS) was first described by Nager and De Reynier in 1948. It is a rare developmental condition (<100 cases in 2012) characterized by Acrofacial dysostoses: malformations of the craniofacial skeleton and the limbs. 18 unique mutations of SF3B4 were identified in NS patients: frameshift, 2 nonsense, one splicing and one missense mutation. Furthermore, a recent study reported the reduction of some neural crest genes expression in SF3B4 depleted Xenopous embryos[140]. Following the embryonic stage, tadpoles manifested hypoplasia of craniofacial cartilage that is usually derived from neural crest cells. A phenotype manifesting similar features to NS. Due to the lethality of SF3B4 KOs, 50% CRISPR/Cas9 KD in MC3T3 (mouse

osteoblast precursor) cells were created, to further study the effect of SF3B4 haploinsufficiency on NS.

Results

DNA Validation

Creation and validation of SF3B4 KDs was performed as described in Part I, using MC3T3 cells. **Figure 13** depicts a successful transfection of the MC3T3 cells. As well, screening of the clones using Sanger sequencing resulted in 5 clones with a disruption of the genomic DNA downstream of the target site (**Figure 14**). Analysis of these Sanger sequencing traces indicated %KO scores of ~50% for each clone using ICE analysis (Synthego). These clones will undergo protein validation in Dr. MacMillan's lab to confirm 50% KD of the SF3B4 protein.

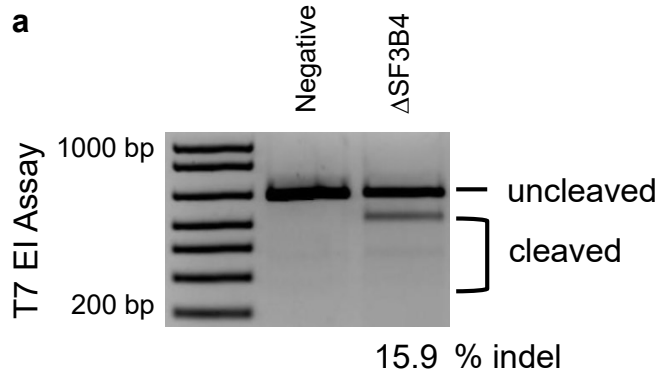


Figure 13. T7 EI assay on bulk ΔSF3B4 cell genomic DNA. a DNA extracted from FACs sorted MC3T3 cells was used in a T7EI mismatch detection assay to determine the amount of indel formation. 200 ng of the purified PCR product was reannealed to create mismatch bubbles that are cleaved by the T7EI endonuclease. Parental MC3T3 cells were used as a negative control. Cleavage product was run on 1.5% TAE gel at

100V for 45 mins. Densitometry was performed by Image J software and %indel was calculated.

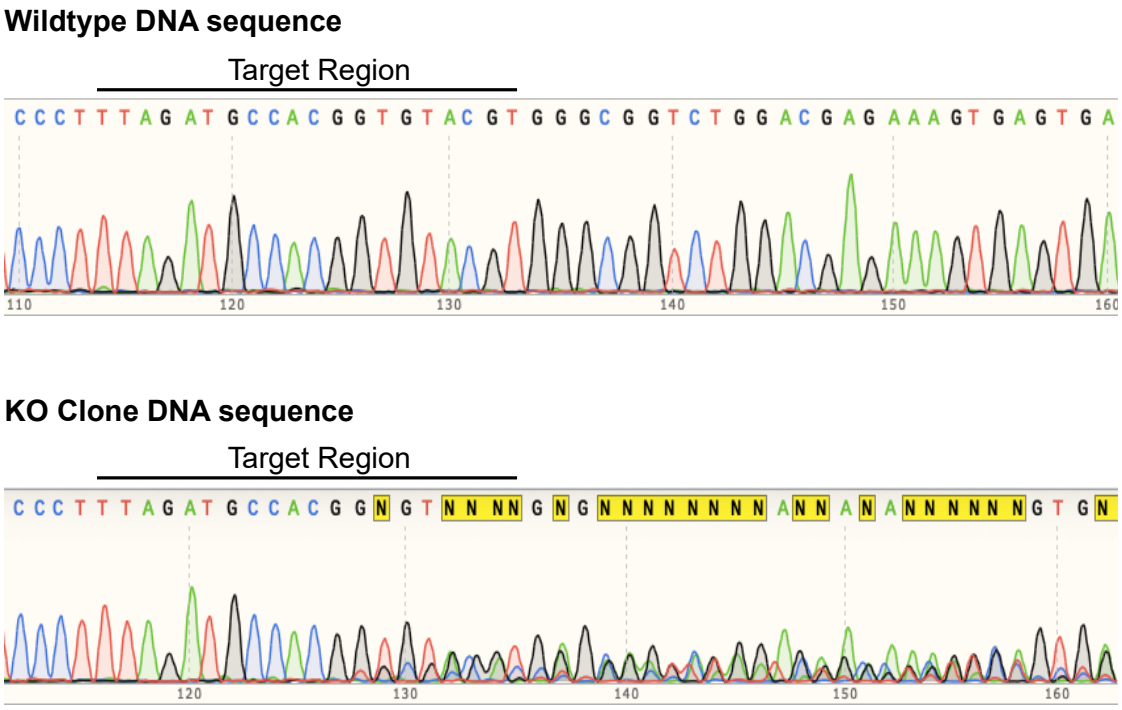


Figure 14. Screening for potential KO clones using MC3T3 genomic DNA. a Representative Sanger sequencing trace for wildtype and **b** FACs selected clone #2 (MC3T3 cells). Region of DNA targeted by crRNA is outlined. “N” indicates the presence of multiple bases. Multiple clones were screened and selected for protein validation based on the %KO score calculated by ICE Analysis (Synthego).

Future Directions

Protein validation will be performed on these cells by Ayat Omar. If ~50% KD is confirmed at the protein level then these cells will be used as a mouse bone cell model of SF3B4 depletion. These will be utilized to further investigate the impact lowered SF3B4 levels have on NS pathophysiology.

CHAPTER 4: Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences.

Background

Clustered Regularly Interspaced Palindromic Repeat (CRISPR) systems play an important role in adaptive immunity in prokaryotes[141] and have been effectively repurposed as gene and RNA-editing tools[13,44]. Over 400 CRISPR-Cas homologs, comprised of enzymes with different nucleic acid binding specificities and cleavage mechanisms have been annotated[43] and grouped into two classes[40]. Class I systems employ multi-subunit nuclease complexes, while Class II systems, more widely used for gene editing, rely on a single effector protein[40]. Class II systems may be further divided into subtypes (e.g. II-A, II-B, V, VI)[40]. Cas9, a type II-A system, directs DNA cleavage using two separately expressed RNA elements: a crRNA that contains a 20-bp RNA sequence complementary to the target DNA sequence, and a transactivating crRNA (tracrRNA) that bridges the Cas9-crRNA interaction[32]. Target recognition by Cas9 involves binding a PAM sequence (5'-NGG-'3 in *S. pyogenes*), followed by hybridization of the 20-bp spacer sequence to the target. Formation of a fully paired duplex induces conformational changes that activate the RuvC and HNH nuclease domains in Cas9, leading to dsDNA cleavage[31,37]. In contrast, the Type V system Cas12a (Cpf1) employs a single RuvC active site to induce staggered cuts within the target and non-target strands[40]. Cas12a recognizes a T-rich PAM (5'TTN'3), uses a 20-24 bp spacer sequence, does not require a tracrRNA, and can process its own pre-cRNA[41]. Unlike

Cas9, Cas12a unleashes indiscriminate single-stranded DNase activity *in vitro* upon nuclease activation[4].

Both Cas9 and Cas12a have been used to edit the genomes of diverse organisms ranging from plants to mammals and both show promise as therapeutics to treat genetic disease[142,143]. In addition, the collateral DNase activity of Cas12a has been exploited to generate a diagnostic platform for the detection of aberrant mutations or pathogen DNA sequences[4]. Briefly, the DNA endonuclease-targeted CRISPR trans reporter (DETECTR) system links activation of Cas12a nuclease activity to *trans* cleavage of a single-stranded DNA substrate containing flanking fluorophore and quencher moieties; when combined with isothermal amplification, this system achieves attomolar DNA detection sensitivity[4]. One of the primary obstacles to translating CRISPR/Cas systems to clinical applications has been concern over off-target DNA cleavage, which could have detrimental health consequences for therapeutics, and yield false-positive results for diagnostics[76]. As a result, much work has been done to improve the specificity of these systems through protein engineering or evolution[80,81], or engineering or chemical modification of guide RNAs[5]. For example, guide RNAs with engineered secondary structures improve Cas12a specificity[144], and incorporation of DNA[145] or BNA[5] into Cas9 gRNAs improves specificity.

While single-nucleotide precision is desirable for many nucleic acid targeting applications[144], there are other instances where recognition of a discrete 20-bp sequence may be limiting. First, CRISPR/Cas9 can be sensitive to naturally occurring SNPs within the PAM-proximal portion of a guide sequence[88,146,147]. Since SNPs occur roughly every 300bp in the human genome[148], a CRISPR/Cas9 therapeutic designed for

one patient may be ineffective for another. Indeed, a recent test of 263 therapeutically-relevant guide RNAs revealed that >16% failed to cleave the on-target site in at least one of 7700 haplotypes tested[2]. This creates a potential for resistance to treatment, especially in highly polymorphic regions, such as the immune and blood systems. Second, the high degree of natural genetic diversity present in pathogens such as *HIV-1* greatly complicates antiviral treatment or diagnostic detection using CRISPR/Cas systems[98,99]. Finally, studies have shown that even successful cleavage of *HIV-1* DNA sequences using CRISPR/Cas9 can result in mutations that accelerate viral escape and render the virus resistant to the original guide RNA[100,101]. These scenarios highlight the need for additional CRISPR/Cas capabilities that allow sequences to be targeted more flexibly.

In nature, partially degenerate recognition of mRNA codons by the tRNA anticodon loop is achieved through the inclusion of inosine (I) nucleotides (containing the hypoxanthine base)[105,107]. Inosine also plays a role in RNA editing[149], and acts as a DNA damage intermediate following adenosine deamination[108]. Characterized as a 'universal base', inosine forms 2-hydrogen bonds with all 4 canonical bases with a slight I-C > I-A > I-T ≈ I-G bias in stability[105]. Inosine has been successfully applied to the design of degenerate PCR primers and diagnostic probes, as well as in DNA sequencing[105]. Synthetic universal bases such as 5-nitroindole (5-NI) and the related 3-nitropyrrole (3-NP) have also been developed[114]. 5-NI lacks the ability to form any hydrogen bonds but adopts as a standard *anti* configuration with the opposing nucleotide and acts to stabilize hydrophobic base stacking[103]. While more destabilizing in certain contexts, 5-NI bases appear to be devoid of any base-pairing bias[103,150]. Other

synthetic bases have been developed to exhibit partial degeneracy, including base K (2-amino-6-methoxyaminopurine) and base P (6H,8H-3,4-dihydro-pyrimido[4,5-c] [4,5-c] [1,2]oxazin-7-one), which shows a preference for C/T and A/G pairing, respectively[110,111].

Based on their present use in nucleic acid amplification and detection technologies[105], we wondered if universal bases could be harnessed to impart Cas systems with the ability to target multiple related sequences with an individual guide RNA. Here, we show that multiple different universal bases can be tolerated within both Cas9 and Cas12a guide RNAs, depending on the context. We demonstrate the applicability of this strategy to targeting multiple naturally occurring SNPs within the *ABO* gene using a single guide RNA both *in vitro* and in cells, and to the detection of multiple strains of *HIV-1* using a single probe in the DETECTR system. Using high-throughput specificity profiling we show that inclusion of universal bases imparts selective degeneracy at the site of incorporation, without otherwise substantially altering gRNA specificity. Our results outline a new strategy for expanding the capabilities of CRISPR/Cas to the recognition of nucleic acid targets with high variability and those for which only incomplete sequence information is available.

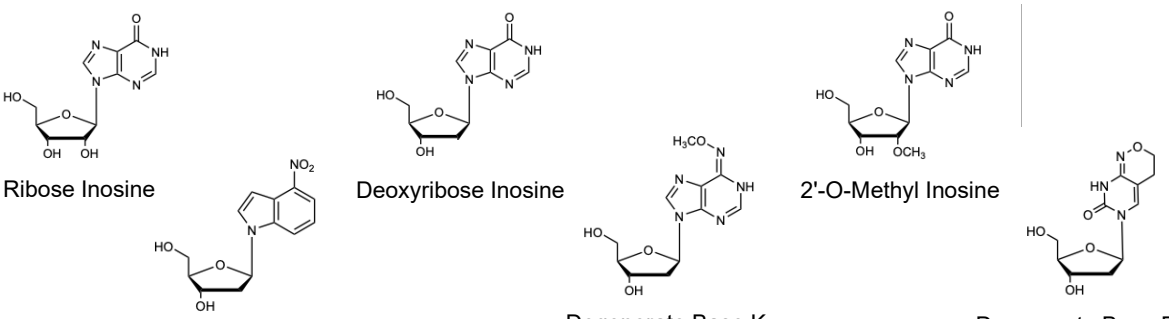
Results

Context-dependent tolerance of inosine bases in Cas9 crRNAs enables degenerate targeting

Past studies have shown that both sugar (e.g. locked/bridged nucleic acids, DNA)[151,5] and backbone (2-OMe phosphonoacetate, phosphorothioate)[152,153] chemical

modifications can be tolerated by Cas9. Moreover, several of these modifications have been demonstrated to improve Cas9 DNA cleavage specificity in a localized and predictable manner by increasing the stringency of base-pairing interactions[5]. Based on these findings, we hypothesized that inclusion of inosine or its deoxy and 2'OMe derivatives, as well as synthetic universal bases (deoxyNI, deoxyK, deoxyP) (**Figure 15A**), would present a viable approach for imparting Cas9 guide RNAs with selective degeneracy. To begin, we selected a highly polymorphic sequence from the *ABO* gene that determines the most clinically important blood group system in mammals[118]. We generated a series of 16 DNA target sequences (ABO-T1-16) containing single or multiple prevalent SNPs within that region (**Figures 15B, 16A**). Next, we tested the ability of Cas9 to cleave these sequences using an unmodified guide RNA (ABO-RNA) corresponding to the canonical, most highly abundant target sequence (ABO-T1). Consistent with previous studies on Cas9 specificity[5,154], we found that in addition to cleaving the perfectly matched sequence (ABO-T1), sequences with single nucleotide mismatches were also cleaved but those with multiple mismatches were not (**Figure 16B**).

a



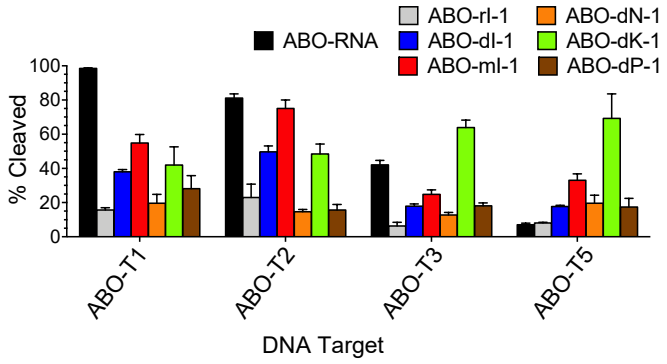
b

Target	DNA Sequence (5′→3′)	Allele Frequency (%)
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	54.21
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	25.50
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	15.98
ABO-T4	CATGGAGTTCCGCGACCAC**ATGG**	0.015
ABO-T5	CATGGAG**A**TCCGCGACCAC**ATGG**	0.0038
ABO-T6	CATGGAG**A**TCCGCGACCA**TATGG**	0.00061

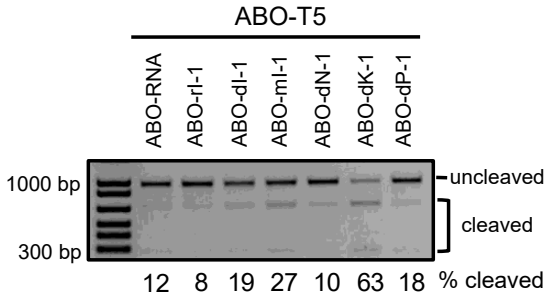
c

[*] = Modification

crRNA: 5′-CAUGGAG[*]UCCGCGACCAC[*]-3′



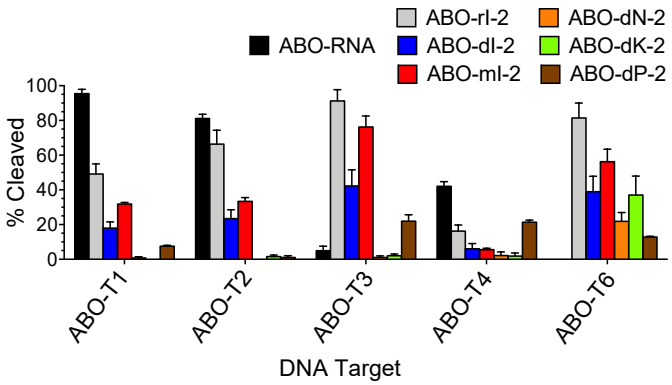
d **ABO-RNA:** 5′-CAUGGAGUUCCGCGACCACG-3′



e

[*] = Modification

crRNA: 5′-CAUGGAG[*]UCCGCGACCA[*][*]-3′



f **ABO-RNA:** 5′-CAUGGAGUUCCGCGACCACG-3′

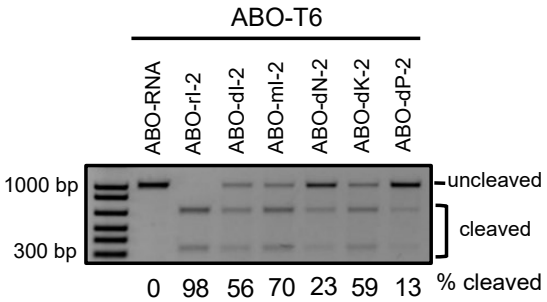


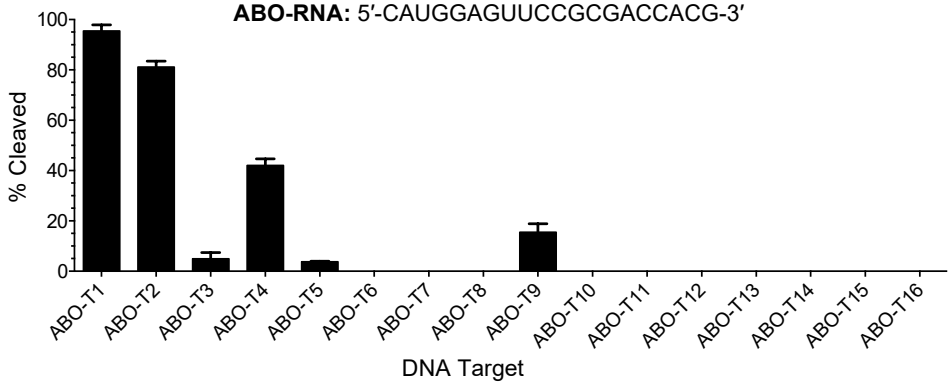
Figure 15. The incorporation of modifications restores the ability to cleave targets containing single nucleotide polymorphisms (SNPs). a Diagrams and names of chemical modifications used as universal bases. **b** List of DNA targets designed for the *ABO* gene based on clinical polymorphism data. SNPs are indicated with red lettering. Allele frequency is based on both current minor allele frequency collected in all populations and calculated probability that multiple SNPs will occur. Graphs representing *in vitro* cleavage of SNP DNA targets using gRNAs containing **c** double and **e** triple modifications. *In vitro* cleavage assays were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM). Individual data points are shown as SEM (n=2). Modification positions in the crRNA sequence are indicated with [*]. rI = Ribose Inosine, dI = Deoxyribose Inosine, mI = 2'-O-methyl Inosine, dN = 5'-Nitroindole, dK = Degenerate Base K, dP = Degenerate Base P. Representative gel of *in vitro* cleavage activity for crRNAs containing **d** 2 modifications against DNA target ABO-T5 or **f** 3 modifications against DNA target ABO-T8. For all *in vitro* cleavage assay gels, the top band is uncleaved DNA substrate, and the bottom two bands are cleaved DNA substrate. Quantification of cleavage percentages was calculated using densitometry software (ImageJ).

To generate gRNAs capable of simultaneously targeting complex natural sequence variants bearing multiple SNPs, we designed a series of crRNAs in which 1-4 inosine bases were incorporated at sites overlapping with known SNPs (**Figure 16C**). Initially, we tested the activity of these modified crRNAs using Cas9 with the naturally found ABO-T1 sequence. We found that inclusion of a single inosine was tolerated in all instances, albeit with reduced activity, while substitutions of 2 or 3 inosines were tolerated in certain positional context (**Figure 16D**). These results demonstrate that the inclusion of inosine bases into crRNAs can be tolerated, depending on number and position, and can enable Cas9 cleavage of multiple sequences using an individual gRNA.

a

Target	DNA Sequence (5′→3′)	# of SNPs	Target	DNA Sequence (5′→3′)	# of SNPs
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	0	**ABO-T9**	CATGGAGTTC**T**GCGACCACG**TGG**	1
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	1	**ABO-T10**	CATGGAG**A**TC**T**GCGACCACG**TGG**	2
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	1	**ABO-T11**	CATGGAGTTC**T**GCGACCA**T**G**TGG**	2
ABO-T4	CATGGAGTTCCGCGACCAC**A**TGG	1	**ABO-T12**	CATGGAGTTC**T**GCGACCAC**A**TGG	2
ABO-T5	CATGGAG**A**TCCGCGACCAC**A**TGG	2	**ABO-T13**	CATGGAG**A**TC**T**GCGACCA**T**G**TGG**	3
ABO-T6	CATGGAG**A**TCCGCGACCA**TA**TGG	3	**ABO-T14**	CATGGAG**A**TC**T**GCGACCAC**A**TGG	3
ABO-T7	CATGGAG**A**TCCGCGACCA**T**G**TGG**	2	**ABO-T15**	CATGGAGTTC**T**GCGACCA**TA**TGG	3
ABO-T8	CATGGAGTTCCGCGACCA**TA**TGG	2	**ABO-T16**	CATGGAG**A**TC**T**GCGACCA**TA**TGG	4

b



c

Name	crRNA Sequence (5′→3′)	Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG	**ABO-rI-8**	CAUGGAG[I]UCCGCGACCA[I]G
ABO-rI-1	CAUGGAG[I]UCCGCGACCAC[I]	**ABO-rI-9**	CAUGGAGUUC[I]GCGACCA[I]G
ABO-rI-2	CAUGGAG[I]UCCGCGACCA[I][I]	**ABO-rI-10**	CAUGGAGUUC[I]GCGACCAC[I]
ABO-rI-3	CAUGGAG[I]UCCGCGACCACG	**ABO-rI-11**	CAUGGAGUUCCGCGACCA[I][I]
ABO-rI-4	CAUGGAGUUC[I]GCGACCACG	**ABO-rI-12**	CAUGGAG[I]UC[I]GCGACCA[I]G
ABO-rI-5	CAUGGAGUUCCGCGACCA[I]G	**ABO-rI-13**	CAUGGAG[I]UC[I]GCGACCAC[I]
ABO-rI-6	CAUGGAGUUCCGCGACCAC[I]	**ABO-rI-14**	CAUGGAGUUC[I]GCGACCA[I][I]
ABO-rI-7	CAUGGAG[I]UC[I]GCGACCACG	**ABO-rI-15**	CAUGGAG[I]UC[I]GCGACCA[I][I]

d

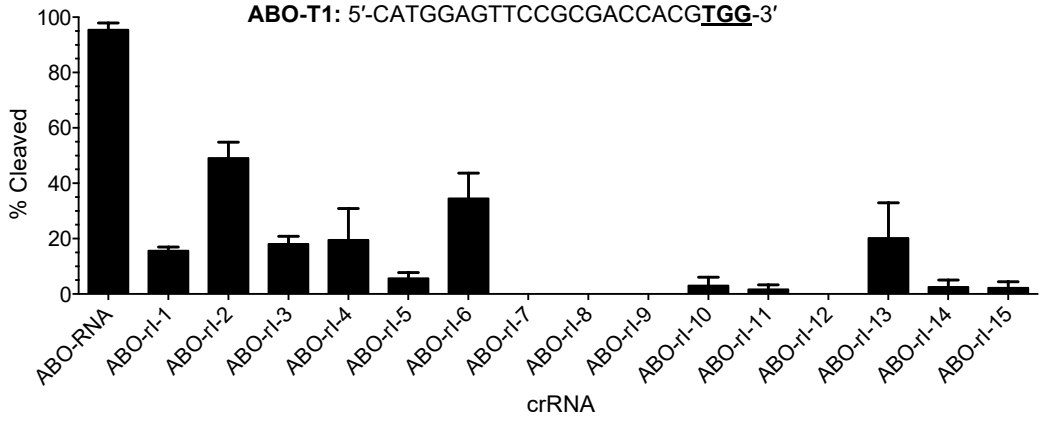


Figure 16. *In vitro* activity of unmodified and modified *ABO* crRNAs. a List of DNA targets designed for the *ABO* gene based on clinical polymorphism data (NCBI). SNPs are indicated with red lettering. The number of SNPs is based on wildtype DNA target: ABO-T1. **b** Graph representing cleavage of SNP DNA targets using unmodified crRNA: ABO-RNA. **c** List of modified crRNAs with ribose inosine modification(s) indicated by red [I]. **d** Graph representing cleavage of ABO-T1 when using ribose inosine modified crRNAs containing different numbers and/or positions of modifications. *In vitro* cleavage assays were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM). Quantification of cleavage percentages was calculated using densitometry software (ImageJ), SEM (n=2).

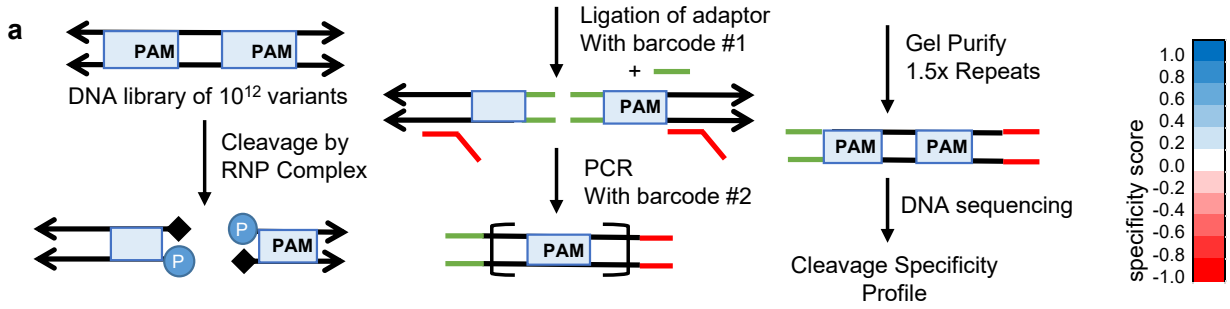
Multiple types of universal bases can be incorporated into Cas9 guide RNAs to allow for multi-sequence targeting

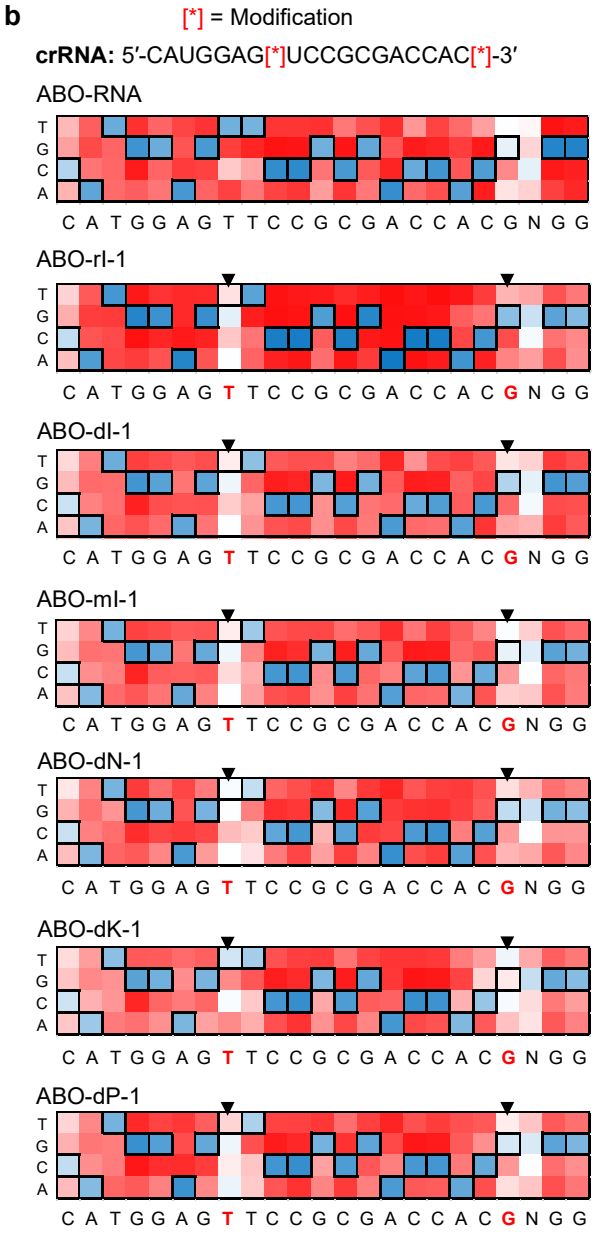
Numerous inosine derivates as well as synthetic bases have been successfully applied to various PCR and nucleic acid detection applications[114]. To test the possibility that some of these engineered nucleotides could be incorporated into Cas9 crRNAs to enable multi-sequence targeting, we selected two *ABO* sequences bearing 2 (ABO-T5) or 3 (ABO-T6) substitutions relative to ABO-T1 (**Figure 15B**) and designed a panel of corresponding crRNAs in which inosine (ABO-rI-1, ABO-rI-2), deoxyinosine (ABO-dI-1, ABO-dI-2), 2'OMe inosine (ABO-mI-1, ABO-mI-2), 5'nitroindole (ABO-dN-1, ABO-dN-2), dK (ABO-dK-1, ABO-dK-2), or dP (ABO-dP-1, ABO-dP-2) bases were included at the sites of mismatch. Using ABO-rI-1, ABO-dI-1, ABO-mI-1, ABO-dN-1, ABO-dK-1, and ABO-dPI-1, we assayed Cas9 cleavage activity on ABO-T1, the corresponding ABO-T5 double SNP variant sequence, and sequences containing each mismatch in isolation (ABO-T2, ABO-T3). As shown in **Figure 15C, D**, Cas9 cleaved ABO-T5 using ABO-mI-1 and ABO-dK-1 with 2 and 5-fold increased activity versus ABO-RNA, respectively.

Importantly, both of these crRNAs also supported Cas9 cleavage of the single variant (ABO-T2, ABO-T3) and canonical sequences (ABO-T1), maintaining activity within a twofold range of ABO-RNA (**Figure 15C**). Similarly, we found that ABO-rI-2, ABO-mI-2, and ABO-dK-2 guided efficient Cas9 cleavage of ABO-T6 (>50% compared to 0% with ABO-RNA) (**Figure 15E, F**). ABO-rI-2 and ABO-mI-2 were also able to direct the cleavage of ABO-RNA and the single variant sequences ABO-T2 and ABO-T3, but not ABO-T4 (**Figure 15D**). These results demonstrate that universal bases with diverse chemistries can be incorporated into crRNAs to allow simultaneous targeting of complex SNP variants *in vitro*.

The inclusion of universal bases into crRNAs alters specificity only at the site of incorporation.

A prerequisite for the practical application of guide RNAs containing universal bases to targeting SNPs is that they must influence overall Cas9 specificity in a localized and predictable manner. That is to say, they should impart selective degeneracy rather than globally impacting the precision of Cas9 DNA cleavage. To examine if this is the case we employed a previously described high-throughput specificity profiling assay[5,76,120] that measures Cas9 cleavage of a library of $>10^{12}$ off-target sequences, containing a 10-fold coverage of all sequences with ≤8 mutations relative to the ABO-T1 sequence (**Figure 17A**).


a
PAM
PAM
DNA library of 10^12 variants
Cleavage by
RNP Complex
P
P
PAM
Ligation of adaptor
With barcode #1
+
PAM
PCR
With barcode #2
PAM
Gel Purify
1.5x Repeats
PAM
PAM
DNA sequencing
Cleavage Specificity
Profile
specificity score
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0


b
[*] = Modification
crRNA: 5′-CAUGGAG[*]UCCGCGACCAC[*]-3′
ABO-RNA
ABO-rI-1
ABO-dI-1
ABO-mI-1
ABO-dN-1
ABO-dK-1
ABO-dP-1
T G C A
C A T G G A G T T C C G C G A C C A C G N G G

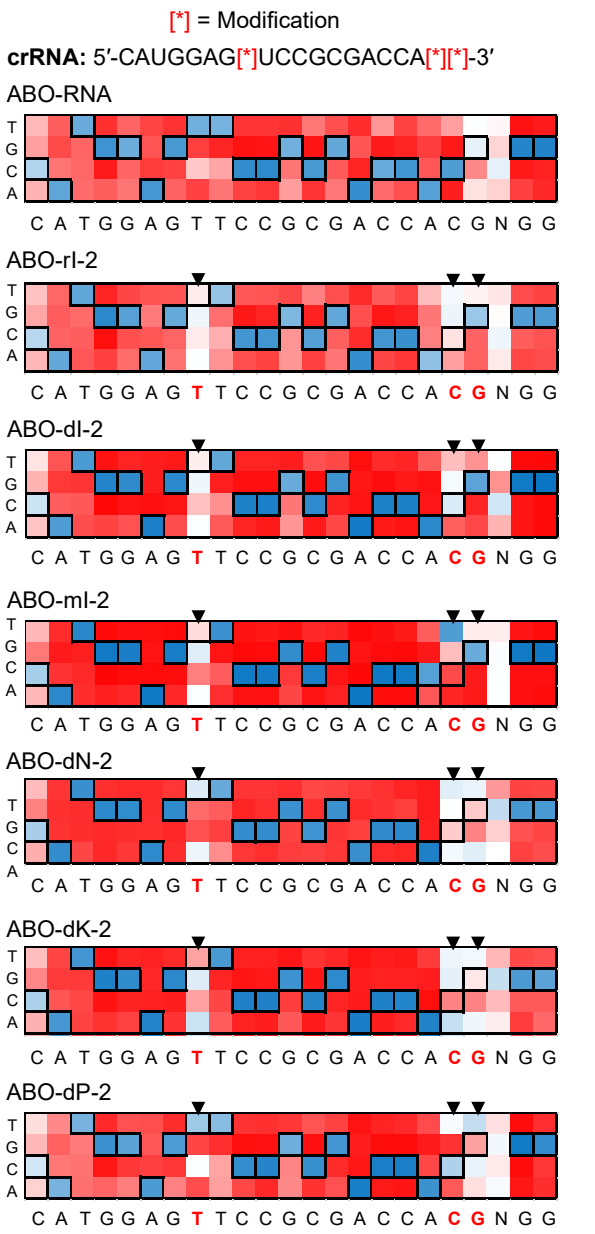

c
[*] = Modification
crRNA: 5′-CAUGGAG[*]UCCGCGACCA[*][*]-3′
ABO-RNA
ABO-rI-2
ABO-dI-2
ABO-mI-2
ABO-dN-2
ABO-dK-2
ABO-dP-2
T G C A
C A T G G A G T T C C G C G A C C A C G N G G

Figure 17. *In vitro* specificity profile for *ABO* crRNAs containing multiple modifications. a Representative diagram depicting workflow for generating *in vitro* cleavage libraries for next-generation sequencing using 1000 nM of gRNA and Cas9. Initial library contain >10^{12} DNA targets (200 nM). Ligated adaptors and PCR primers contain barcodes to demultiplex the resulting sequencing data. Final pre- and post-selection libraries were sequenced on a HiSeq (Illumina). Heat maps representing the specificity profile of each **b** double or **c** triple substituted modified gRNA. These maps are created by counting each DNA base present across each position of the 20 base pair target region and the 3' NGG PAM. Specificity scores of 1.0 (dark blue) correspond to 100% reads containing the indicated base while scores of −1.0 (dark red) correspond to 0% reads containing the indicated base at each position of the target. Scores of 0 (white) denotes no preference for the indicated base. These positions of selective degeneracy are indicated by arrows. Black boxes denote the intended target nucleotide.

We performed the assay on all 13 crRNAs listed in **Figure 15C, D**. To gain insight into the specificity profile of each crRNA, we used the datasets for each to calculate enrichment scores for each base at each position within the ABO-T1 sequence, and generated specificity heatmaps to visualize the results. Strikingly, we found that in nearly all cases the specificity profile for the crRNAs containing universal bases was similar to that of ABO-RNA at all positions except those that overlapped with the locations of the universal bases. Moreover, the substitution of the indicated PAM-distal uracil with inosine (ABO-rI-1 and ABO-rI-2), deoxyinosine (ABO-dI-1 and ABO-dI-2), or 2'O methylinosine (ABO-mI-1 and ABO-mI-2) rendered the crRNA virtually non-specific at this position (**Figure 17B, C**) and was associated with a difference in specificity score of <-0.6 (**Figure 18**). Similar results were observed when the indicated PAM-proximal cytosine base was replaced by a universal base (**Figure 17C**).

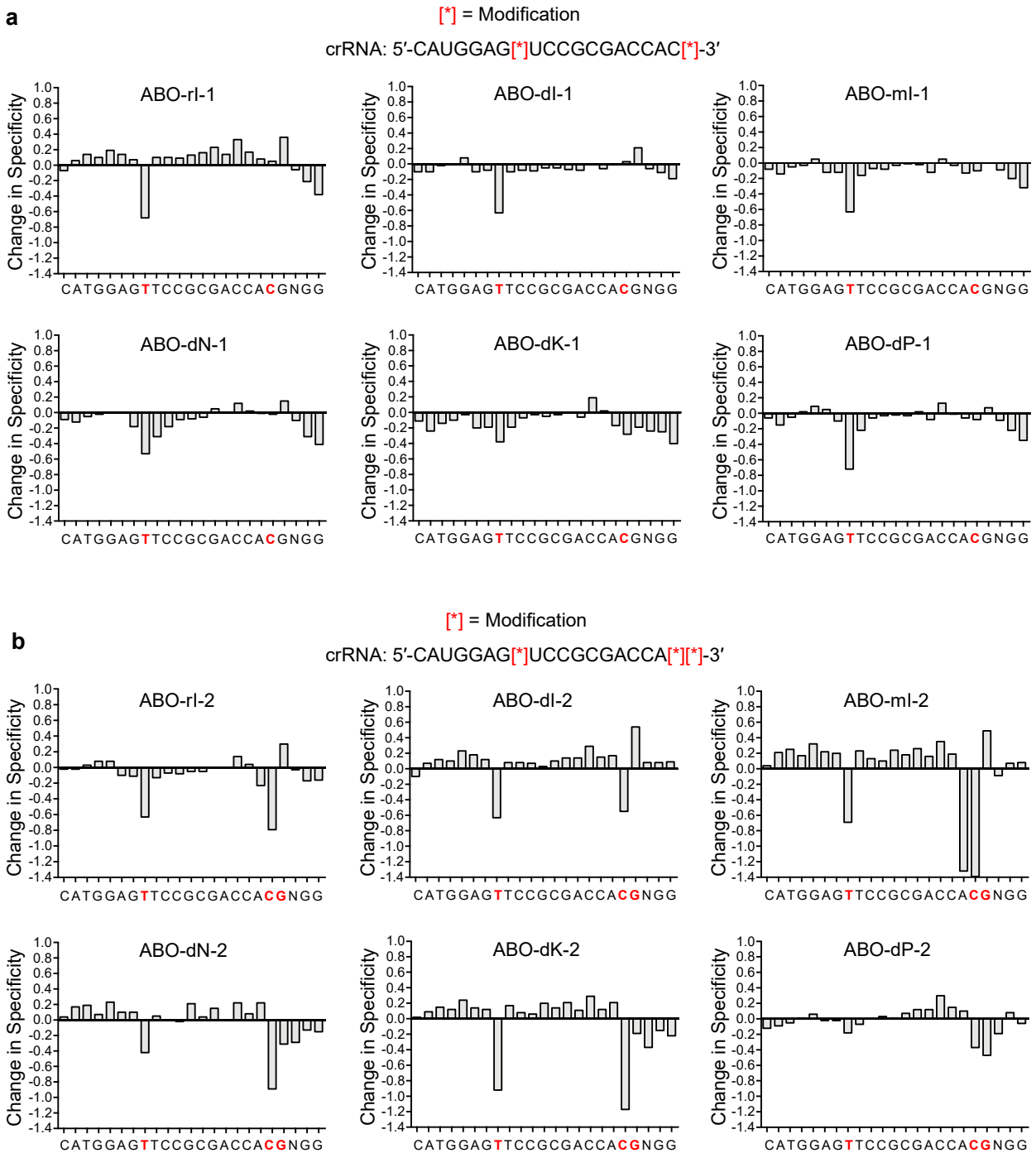


Figure 18. Change in specificity score of *ABO* crRNAs with multiple modifications compared to ABO-RNA. a Bar graphs showing the quantitative difference in specificity score at each position in the DNA target site for modified crRNAs with double or **b** triple modifications. Red [*] denotes the modification position in the crRNA sequence. SNP

locations in the 20 base-pair DNA target are indicated with red lettering. The PAM is shown as “NGG” on the 3′ end of the target. A score of zero indicates no change in specificity. The difference in specificity was calculated as the specificity score(modified)−specificity score (ABO-RNA). The specificity scoring of each nucleotide position is relative to the pre-selection control library data.

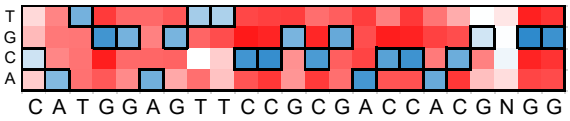
This effect was strongest for ABO-rI-2, ABO-dN-2, and ABO-dK-2, whose specificity scores at that position decreased by 1.2, 0.6, and 1.2, respectively, versus ABO-RNA. Specificity at the indicated PAM-proximal guanine position was less affected by substitutions with universal bases, ostensibly due to an initial lack of specificity at this position in ABO-RNA (**Figures 17B, C and 18**). We also performed high-throughput specificity profiling on the entire collection of ribose inosine-modified crRNAs (ABO-rI-1-15) listed in **Figure 16C** (**Figures 19, 20**), as well as a separate set of 8 cRNAs targeting SNPs of the major histocompatibility complex *HLA* (**Figures 21, 22**). The results of all of these experiments were internally consistent, and reveal that inclusion of universal bases in crRNAs imparts selective degeneracy at the site of incorporation, and otherwise preserves specificity.

a

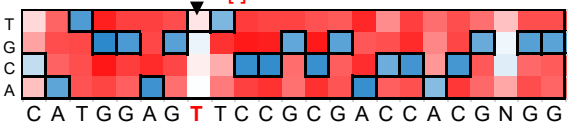
Target	DNA Target Sequence	Allele Frequency (%)
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	54.21
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	25.50
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	15.98
ABO-T4	CATGGAGTTCCGCGACCAC**A**TGG	0.015
ABO-T5	CATGGAG**A**TCCGCGACCAC**A**TGG	0.0038
ABO-T6	CATGGAG**A**TCCGCGACCA**TA**TGG	0.00061
ABO-T7	CATGGAG**A**TCCGCGACCA**T**A**TGG**	4.08
ABO-T8	CATGGAGATCCGCGACCA**TA**TGG	0.0024
ABO-T9	CATGGAGTTC**T**GCGACCACG**TGG**	0.14
ABO-T10	CATGGAG**A**TC**T**GCGACCACG**TGG**	0.037
ABO-T11	CATGGAGTTC**T**GCGACCA**T**G**TGG**	0.022
ABO-T12	CATGGAGTTC**T**GCGACCAC**A**TGG	0.0000021
ABO-T13	CATGGAG**A**TC**T**GCGACCA**T**G**TGG**	0.0059
ABO-T14	CATGGAG**A**TC**T**GCGACCAC**A**TGG	0.0000056
ABO-T15	CATGGAGTTC**T**GCGACCA**TA**TGG	0.0000033
ABO-T16	CATGGAG**A**TC**T**GCGACCA**TA**TGG	0.00000089

b

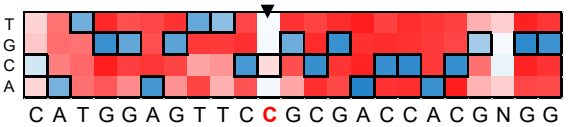
ABO-RNA
crRNA: 5′-CAUGGAGUUCCGCGACCACG-3′



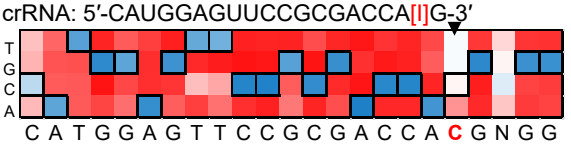
ABO-rI-3
crRNA: 5′-CAUGGAG[I]UCCGCGACCACG-3′



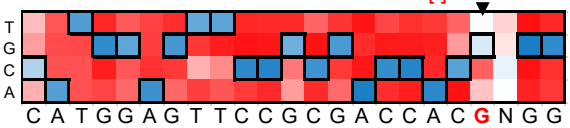
ABO-rI-4
crRNA: 5′-CAUGGAGUUC[*]GCGACCACG-3′



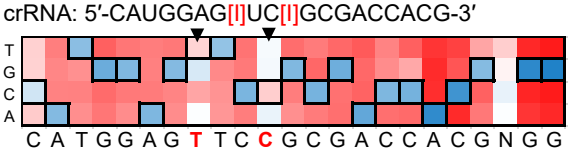
ABO-rI-5
crRNA: 5′-CAUGGAGUUCCGCGACCA[I]G-3′



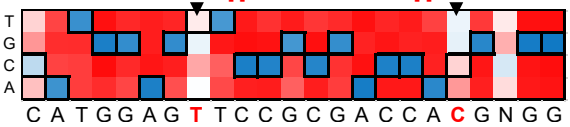
ABO-rI-6
crRNA: 5′-CAUGGAGUUCCGCGACCAC[I]-3′



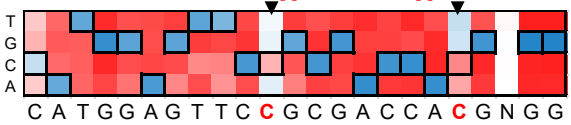
ABO-rI-7
crRNA: 5′-CAUGGAG[I]UC[I]GCGACCACG-3′



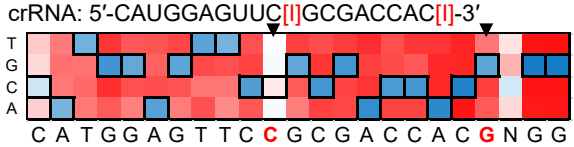
ABO-rI-8
crRNA: 5′-CAUGGAG[I]UCCGCGACCA[I]G-3′



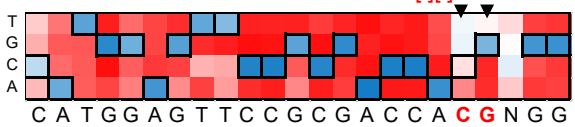
ABO-rI-9
crRNA: 5′-CAUGGAGUUC[I]GCGACCA[I]G-3′



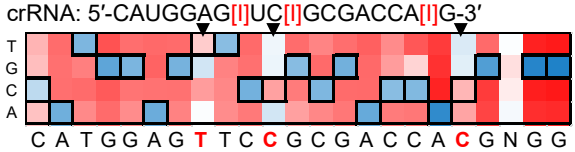
ABO-rI-10
crRNA: 5′-CAUGGAGUUC[I]GCGACCAC[I]-3′



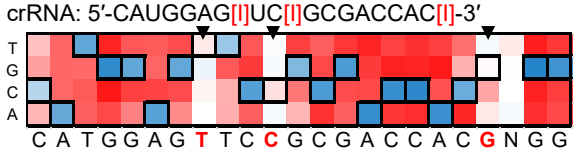
ABO-rI-11
crRNA: 5′-CAUGGAGUUCCGCGACCA[I][I]-3′



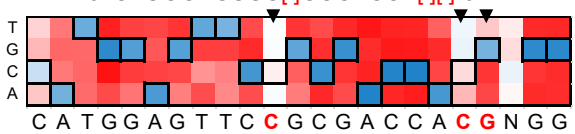
ABO-rI-12
crRNA: 5′-CAUGGAG[I]UC[I]GCGACCA[I]G-3′



ABO-rI-13
crRNA: 5′-CAUGGAG[I]UC[I]GCGACCAC[I]-3′



ABO-rI-14
crRNA: 5′-CAUGGAGUUC[I]GCGACCA[I][I]-3′



ABO-rI-15
crRNA: 5′-CAUGGAG[I]UC[I]GCGACCA[I][I]-3′

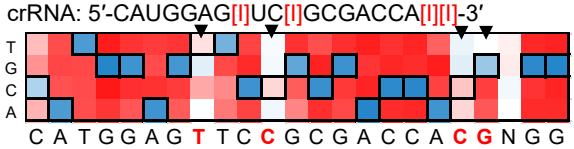


Figure 19. *In vitro* specificity profile for ribose inosine modified gRNAs in the *ABO* gene set. a List of all DNA targets designed for the *ABO* gene. The position of SNPs in the DNA target is indicated with red lettering. Allele frequency is based on current minor allele frequency collected in all populations and probability calculation that multiple SNPs will occur. **b** Heat maps representing the specificity profile of each gRNA. Red [I] indicates a ribose inosine modification. These maps are created by counting each DNA base present across each position of the 20 base pair target region and PAM. Specificity scores of 1.0 (dark blue) correspond to 100% reads containing the indicated base while scores of −1.0 (dark red) correspond to 0% reads containing the indicated base at each position of the target sequence. A score of 0 (white) denotes no preference for the indicated base. These positions of selective degeneracy are indicated by arrows. Black boxes denote the intended target nucleotide.

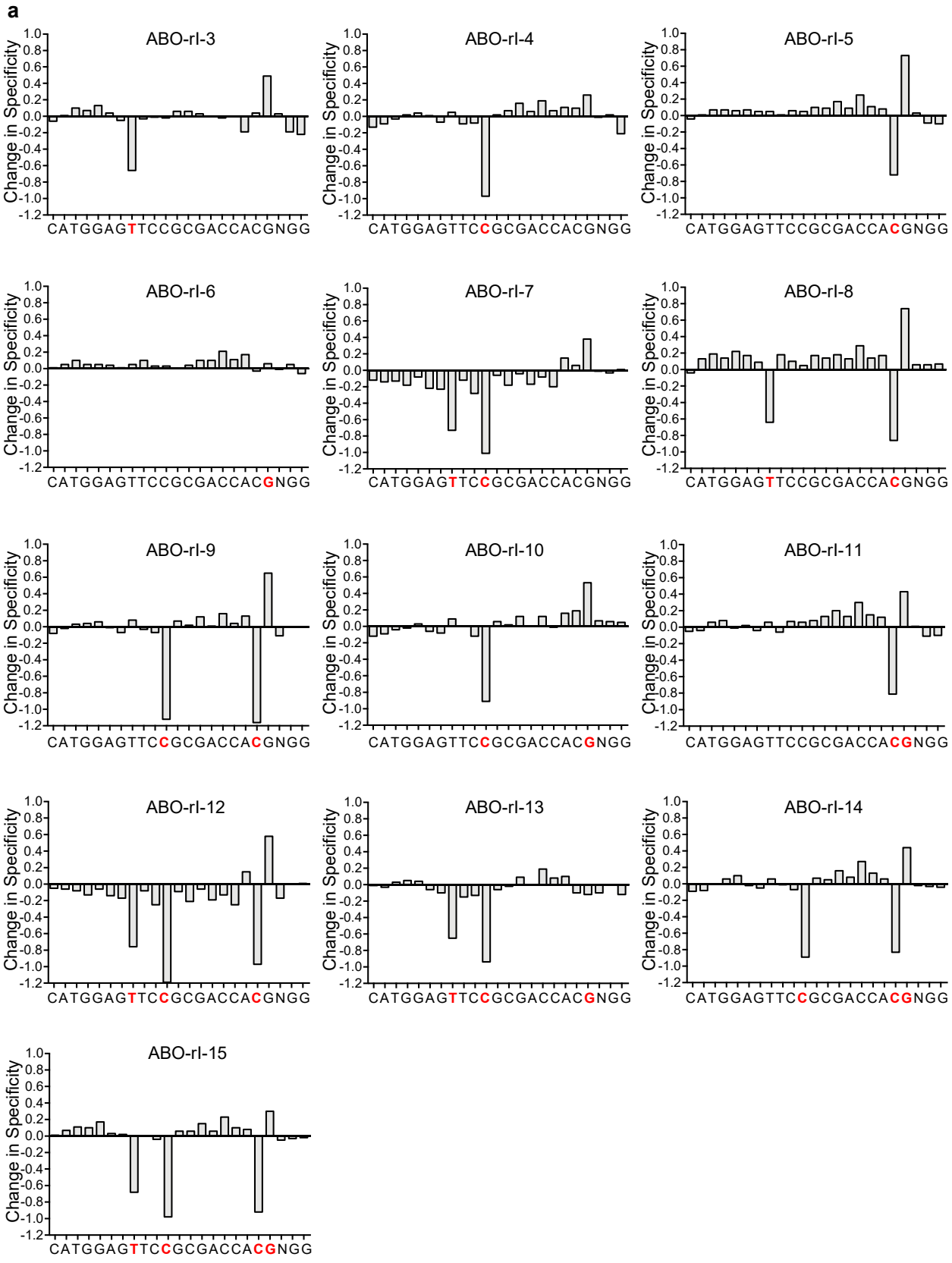

a
ABO-rI-3
ABO-rI-4
ABO-rI-5
ABO-rI-6
ABO-rI-7
ABO-rI-8
ABO-rI-9
ABO-rI-10
ABO-rI-11
ABO-rI-12
ABO-rI-13
ABO-rI-14
ABO-rI-15
Change in Specificity
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG
CATGGAGTTCCGCGACCACGNGG

Figure 20. Change in specificity score of ribose inosine modified *ABO* crRNAs compared to ABO-RNA. a Bar graphs showing the quantitative difference in specificity score at each position in the DNA target site for ribose inosine modified crRNAs. SNP locations in the 20 bp DNA target are indicated with red lettering. The PAM is shown as “NGG” on the 3′ end of the target. A score of zero indicates no change in specificity. The difference in specificity was calculated as the specificity score(modified)−specificity score(ABO-RNA). The specificity scoring of each nucleotide position is relative to the pre-selection control library data.

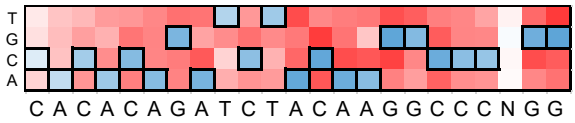
a

Target	DNA Target Sequence	Allele Frequency (%)
HLA-T1	CACACAGATCTACAAGGCCC**AGG**	0.50
HLA-T2	**G**ACACAGATCTACAAGGCCC**AGG**	45.49
HLA-T3	CACACAGATCT**C**CAAGGCCC**AGG**	36.70
HLA-T4	CACACAGATCTACAAGGCC**A**AGG	74.20
HLA-T5	CACACAGATCT**C**CAAGGCC**A**AGG	27.23
HLA-T6	**G**ACACAGATCTACAAGGCC**A**AGG	33.75
HLA-T7	**G**ACACAGATCT**C**CAAGGCCC**AGG**	16.69
HLA-T8	**G**ACACAGATCT**C**CAAGGCC**A**AGG	12.39

b

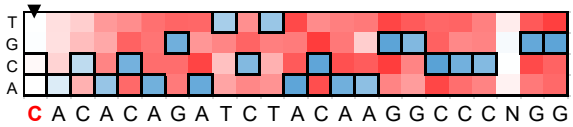
HLA-RNA

crRNA: 5'-CACACAGAUCUACAAGGCCC-3'



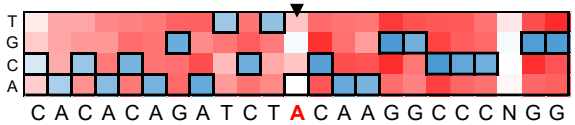
HLA-rI-1

crRNA: 5'-[I]ACACAGAUCUACAAGGCCC-3'



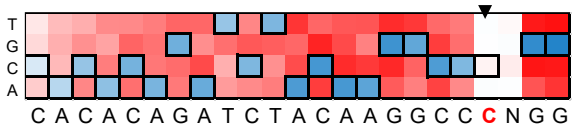
HLA-rI-2

crRNA: 5'-CACACAGAUCU[I]CAAGGCCC-3'



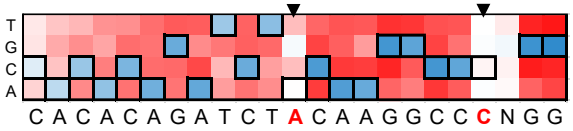
HLA-rI-3

crRNA: 5'-CACACAGAUCUACAAGGCC[I]-3'



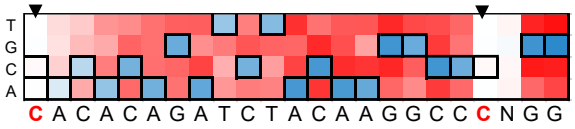
HLA-rI-4

crRNA: 5'-CACACAGAUCU[I]CAAGGCC[I]-3'



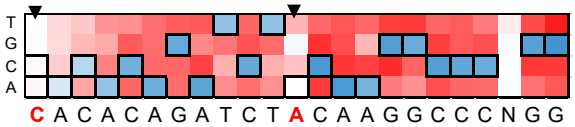
HLA-rI-5

crRNA: 5'-[I]ACACAGAUCUACAAGGCC[I]-3'



HLA-rI-6

crRNA: 5'-[I]ACACAGAUCU[I]CAAGGCCC -3'



HLA-rI-7

crRNA: 5'-[I]ACACAGAUCU[I]CAAGGCC[I]-3'

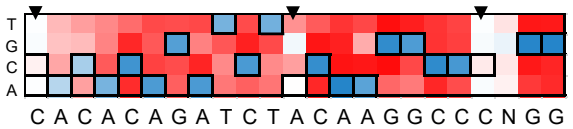


Figure 21. *In vitro* specificity profile for ribose inosine modified *HLA* crRNAs. a List of all DNA targets designed for the *HLA* gene. The positions of SNPs in the DNA target are indicated with red lettering. Allele frequency is based on current minor allele frequency collected in all populations and probability calculation that multiple SNPs will occur. **b** Heat maps representing the specificity profile of each gRNA. Red [I] indicates a ribose inosine modification. These maps are created by counting each DNA base present across each position of the 20 bp target region and PAM. Specificity scores of 1.0 (dark blue) correspond to 100% reads containing the indicated base while scores of −1.0 (dark red) correspond to 0% reads containing the indicated base at each position of the target sequence. A score of 0 (white) denotes no preference for the indicated base. These positions of selective degeneracy are indicated by arrows. Black boxes denote the intended target nucleotide.

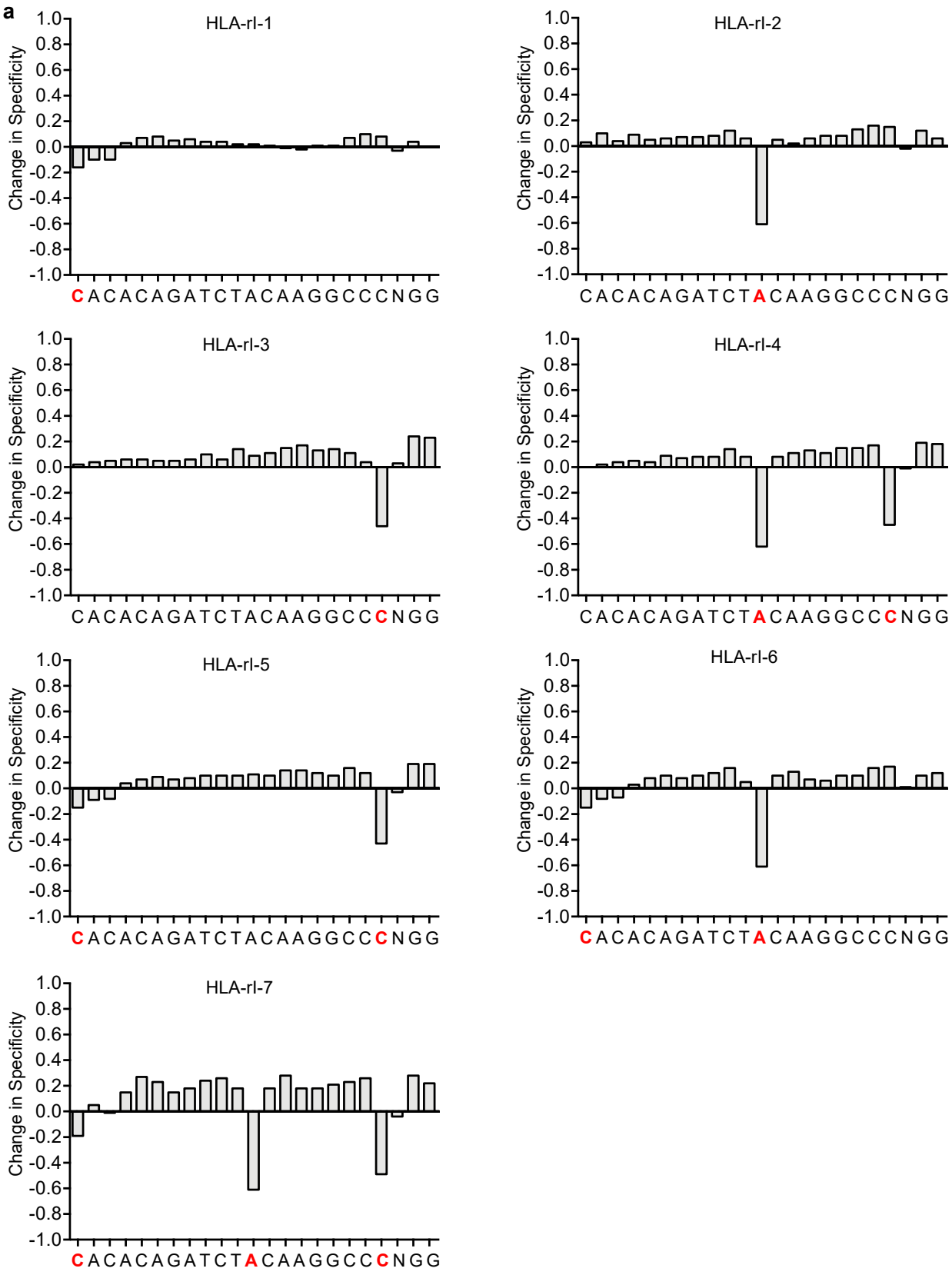

a
HLA-rI-1
HLA-rI-2
HLA-rI-3
HLA-rI-4
HLA-rI-5
HLA-rI-6
HLA-rI-7
Change in Specificity
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
CACACAGATCTACAAGGCCCNGG
CACACAGATCTACAAGGCCCNGG
CACACAGATCTACAAGGCCCNGG
CACACAGATCTACAAGGCCCNGG
CACACAGATCTACAAGGCCCNGG
CACACAGATCTACAAGGCCCNGG
CACACAGATCTACAAGGCCCNGG

Figure 22. Change in specificity score of ribose inosine modified *HLA* crRNAs compared to HLA-RNA. a Bar graphs showing the quantitative difference in specificity score at each position in the DNA target for ribose inosine modified crRNAs. SNP locations in the 20 base-pair DNA target are indicated with red lettering. The PAM is shown as “NGG” on the 3′ end of the target. A score of zero indicates no change in specificity. The difference in specificity was calculated as the specificity score(modified)−specificity score(HLA-RNA). The specificity scoring of each nucleotide position is relative to the pre-selection control library data.

A single crRNA containing multiple ribose inosine bases can induce high-efficiency DNA cleavage of polymorphic gene variants in cells.

Knowing that inclusion of ribose inosine bases into crRNAs could impart selective degeneracy while broadly maintaining cleavage specificity *in vitro*, we sought to translate our findings to cells. First, we selected ABO-rI-2, which bears 3 inosine modifications, and tested its ability to direct Cas9 cleavage of ABO-T1, the corresponding triple SNP variant (ABO-T6), 3 double SNP sequences (ABO-T5,7,8) and 3 single SNP sequences (ABO-T2, T3, T4) (**Figure 23A**). We found ABO-rI-2 directed >50% cleavage of 6/8 sequences tested, the exceptions being ABO-T4 (~20%) and ABO-T5 (<10%) (**Figure 23B**). In contrast, ABO-RNA only supported robust Cas9 cleavage of its matched sequence (ABO-T1) and ABO-T2 (**Figure 23B**).

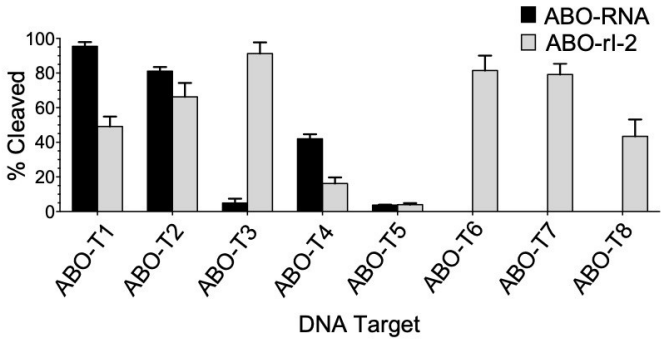
a

Target	DNA Sequence (5′→3′)	Allele Frequency (%)
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	54.21
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	25.50
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	15.98
ABO-T4	CATGGAGTTCCGCGACCAC**A**T**GG**	0.015
ABO-T5	CATGGAG**A**TCCGCGACCAC**A**T**GG**	0.0038
ABO-T6	CATGGAG**A**TCCGCGACCA**TA**T**GG**	0.00061
ABO-T7	CATGGAG**A**TCCGCGACCA**T**AT**GG**	4.08
ABO-T8	CATGGAGATCCGCGACCA**TA**T**GG**	0.0024

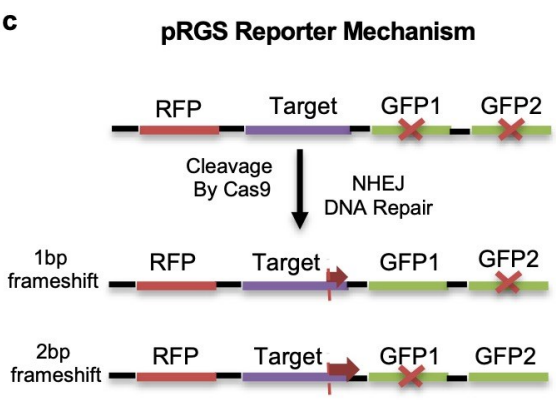
b

ABO-RNA: 5′-CAUGGAGUUCCGCGACCACG-3′

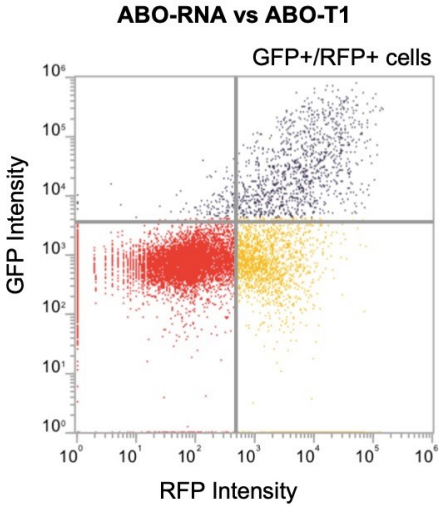
ABO-rI-2: 5′-CAUGGAG[I]UCCGCGACCA[I][I]-3′



c



d



e

Target	ABO-RNA (%GFP+/%RFP+)	ABO-rI-2 (%GFP+/%RFP+)
ABO-T1	36.82 ± 0.75	26.59 ± 0.73
ABO-T2	27.06 ± 0.30	29.40 ± 0.94
ABO-T3	19.27 ± 1.31	39.24 ± 0.33
ABO-T4	25.86 ± 0.43	15.75 ± 0.67
ABO-T5	13.94 ± 1.82	21.43 ± 1.02
ABO-T6	5.21 ± 0.84	35.76 ± 1.09
ABO-T7	6.01 ± 0.56	46.35 ± 0.52
ABO-T8	5.45 ± 0.46	26.65 ± 1.48

Figure 23. Effect of ribose inosine modifications on cleavage activity in vitro and in cells. a List of DNA targets used in the following experiments. SNPs are indicated with red lettering. **b** Graph representing *in vitro* cleavage of DNA targets when using ABO-RNA or ABO-rI-2. Red [I] denotes ribose inosine position in crRNA sequences. *In vitro* cleavage assays were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM). Individual data points are shown as SEM (n=2). **c** Representative diagram of cellular reporter system using HeLa-Cas9 cells. DNA target is cloned into a plasmid containing a constituent RFP reporter and two out-of-frame GFP reporters. Cleavage by Cas9 stimulates error-prone non-homologous end joining (NHEJ) DNA repair, resulting in indel formation. Indel formation shifts one of the GFP reporters into frame resulting in GFP fluorescence. **d** Representative plot for fluorescence-activated cell sorting (FACS) of HeLa-Cas9 cells based on GFP and RFP fluorescence. Dual positive cells appear in the top right quadrant. **e** Table of %GFP+/RFP+ cells for unmodified and modified gRNAs against SNP DNA targets listed above. Data is presented based on the number of events that are %GFP+/all %RFP+ cells (SD, n=3).

Next, we adapted a previously established fluorescence-based reporter system to evaluate the cleavage of these 8 sequences in cells[122]. We cloned all of the target DNAs into a plasmid in which sequences were flanked by an in-frame mRFP gene at the 5’ end and two out of frame eGFP genes at the 3’ end. Past work has shown that double-strand breaks formed in-between these genes can be repaired by non-homologous end-joining (NHEJ), resulting in frameshift mutations that generate a multifluorescent mRFP-eGFP fusion protein (**Figure 23C**)[122]. We co-transfected all 8 constructs with either ABO-RNA or ABO-rI-2 into HeLa cells stably expressing Cas9 and used fluorescence-activated cell sorting (FACS) to quantify the resulting cell populations (**Figure 23D, E**). Using the ABO-RNA, Cas9 cleaved 3/8 sequences with

>25% efficiency (ABO-T1, ABO-T2, ABO-T4). However, 7/8 sequences were cleaved with >25% efficiency, and ABO-T4 RNA was cleaved at 16% efficiency when the ABO-rI-2 guide RNA was used. Cleavage of the ABO-T6, ABO-T7, and ABO-T8 DNA sequences were increased by factors of >7, 8, and 5-fold using ABO-rI-2 compared to ABO-RNA, underscoring the magnitude of the effect. These experiments establish a viable new approach for targeting multiple gene variants in cells using a single guide RNA.

DETECTR probes containing universal bases identify evolved variants of a pathogen

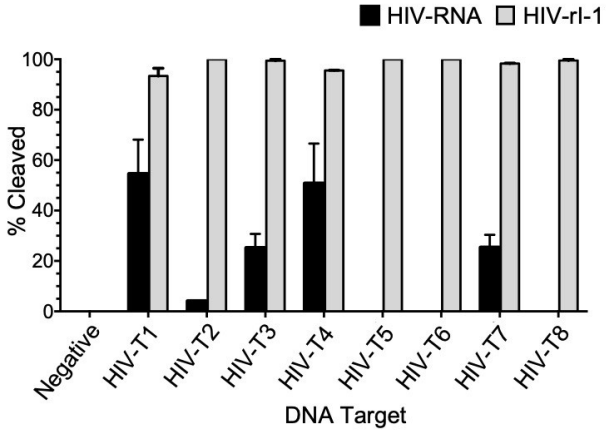
In addition to its application as a gene-editing agent, Cas12a/Cfp1 has also been used for diagnostic purposes as part of the DETECTR system[4]. Point-of-need technologies using this platform to diagnose swine flu[66] as well as COVID-19[70] have now been successfully deployed. However, the detection of evolved pathogen sequences, a problem that has been mitigated through the use of universal bases in PCR-based diagnostic probes, is a challenge that has yet to be addressed for DETECTR. We reasoned that inosine bases could be incorporated into Cas12a guide RNAs to impart them with selectively degenerate targeting capabilities to circumvent this limitation. To test this possibility, we selected a DNA sequence from the *HIV-1* protease gene and identified 7 clinically-relevant sequence variants bearing 1,2, or 3 SNPs encoding protease inhibitor drug resistance mutations[117,119] (**Figure 24A**).

a

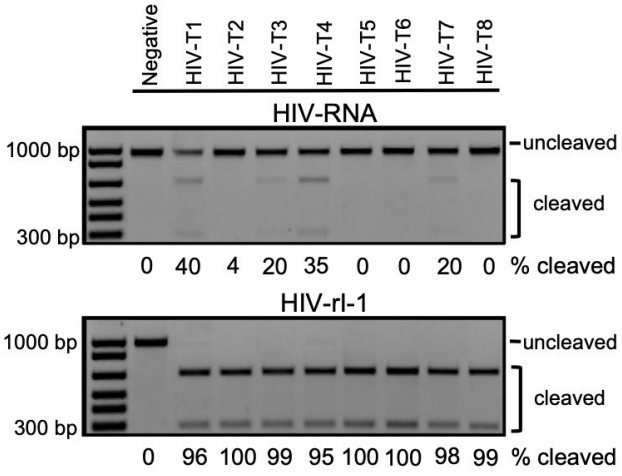
Target	DNA Sequence (5′→3′)	Protease Inhibitor Drug Resistant Mutation(s)
HIV-T1	**TTTG**ATAAAACCTCCAATTCCCCCTAT	N/A
HIV-T2	**TTTG**ATAA**G**ACCTCCAATTCCCCCTAT	F53L
HIV-T3	**TTTG**ATAAAACCTCCAATTCCC**A**CTAT	G48V
HIV-T4	**TTTG**ATAAAACCTCCAATTCCCCCTA**C**	I47V
HIV-T5	**TTTG**ATAA**G**ACCTCCAATTCCC**A**CTAT	F53L, G48V
HIV-T6	**TTTG**ATAA**G**ACCTCCAATTCCCCCTA**C**	F53L, I47V
HIV-T7	**TTTG**ATAAAACCTCCAATTCCC**A**CTA**C**	G48V, I47V,
HIV-T8	**TTTG**ATAA**G**ACCTCCAATTCCC**A**CTA**C**	F53L, G48V, I47V

b

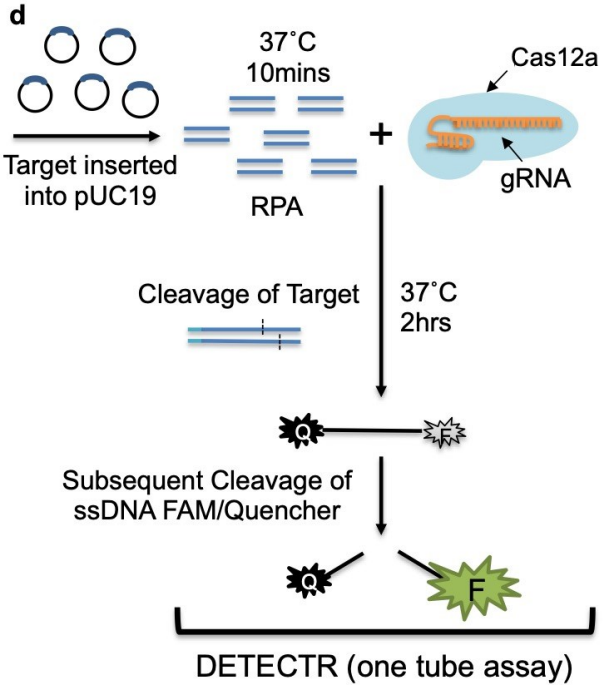
HIV-RNA: 5′-AUAAAACCUCCAAUUCCCCCUAU-3′
HIV-rI-1: 5′-AUAA[I]ACCUCCAAUUCCC[I]CUA[I]-3′



c



d



e

HIV-RNA: 5′-AUAAAACCUCCAAUUCCCCCUAU-3′
HIV-rI-1: 5′-AUAA[I]ACCUCCAAUUCCC[I]CUA[I]-3′

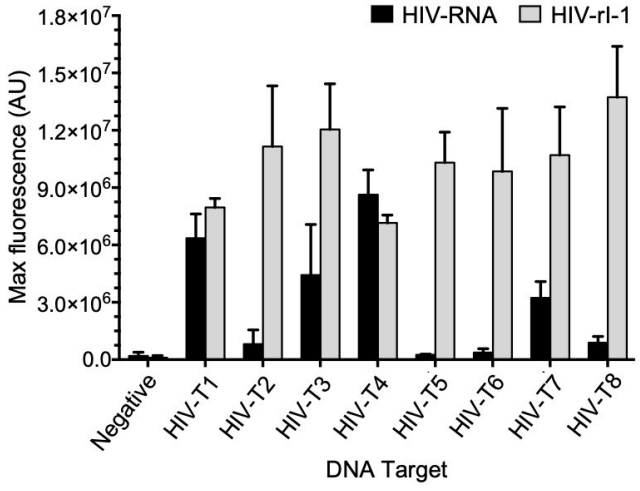


Figure 24. Incorporation of universal bases into the CRISPR/Cas12a system for the detection of *HIV-1* DNA targets. a List of DNA targets designed based on polymorphisms found in a database of *HIV-1* patient DNA samples. SNPs are indicated with red lettering. Resulting drug-resistant mutations are derived from the *HIV-1* mutation database. **b** Graph representing *in vitro* cleavage of DNA targets when using modified and unmodified gRNAs. Red [I] indicates ribose inosine position in crRNA. *In vitro* cleavage assays were performed using fixed concentrations of gRNA (125 nM) and Cas9 (100 nM). Individual data points (SEM, n=2). **c** Representative gel of *in vitro* cleavage activity for modified and modified HIV crRNAs. Quantification of cleavage percentages was calculated using densitometry software (ImageJ). **d** Representative diagram of DETECTR assay. Amplification of plasmid containing the dsDNA target is performed with recombinant polymerase amplification (RPA). Cleavage of dsDNA target results in non-specific degradation of ssDNA reporter by Cas12a. Resulting fluorescence (λ_{ex}: 535nm; λ_{em}: 595nm) is measured every 2 minutes for 2 hours. **e** Graph of fluorescent measurements indicating *in vitro* cleavage of the intended target sequence using unmodified and ribose inosine modified gRNAs. Background subtracted maximal fluorescence values for each crRNA against each relevant DNA target are shown. DNA target was amplified and added to 250 nM Cas12a, 312.5 nM gRNA, and 250 nM ssDNA-FQ reporter. Individual data are shown as SD (n=3). The negative DNA target is designed to have no similarity to targeting regions of the crRNAs used.

Next, we synthesized two crRNAs, HIV-RNA to target the canonical sequence, and HIV-rI-1, which contains 3 ribose inosine substitutions designed to allow for targeting of both the canonical and evolved variant sequences. Interestingly, both Cas12a cleavage activity and reaction kinetics (**Figure 25**) on the canonical sequence (HIV-T1) were equivalent using either of the crRNAs. An *in vitro* cleavage assay of all target sequences using HIV-RNA with Cas12a revealed that HIV-T1, HIV-T3, HIV-T4, HIV-T7 were cleaved at efficiencies of 55%, 30%, 55%, and ~0%, respectively (**Figure 24B, C**).

In stark contrast, all 8 sequences were fully cleaved when HIV-rI-1 was used as the guide RNA (**Figure 24B, C**). We then set out to port these probes into the DETECTR system, which is outlined in **Figure 24D**. To simulate pathogen DNA, we cloned each of our 8 target sequences into pUC19 plasmids and performed RPA as described in the protocol[4]. Next, we set up individual reactions containing each DNA sample paired with either HIV-RNA or HIV-rI-1 probes in the presence of a fluorescent detection substrate. As shown in **Figure 24E**, the HIV-rI-1 probe positively identified all 8 of the *HIV-1* variant sequences, while the HIV-RNA probe only identified 3 sequences and provided false negatives for the other 5 variants. Collectively, these data demonstrate the applicability of crRNAs containing inosine modifications to the detection of evolved or hypervariable gene sequences within pathogens using the DETECTR system.

a

Target	DNA Sequence (5′→3′)	Protease Inhibitor Drug Resistant Mutation(s)
HIV-T1	**TTTG**ATAAAACCTCCAATTCCCCCTAT	N/A
HIV-T4	**TTTG**ATAAAACCTCCAATTCCCCCTA**C**	I47V

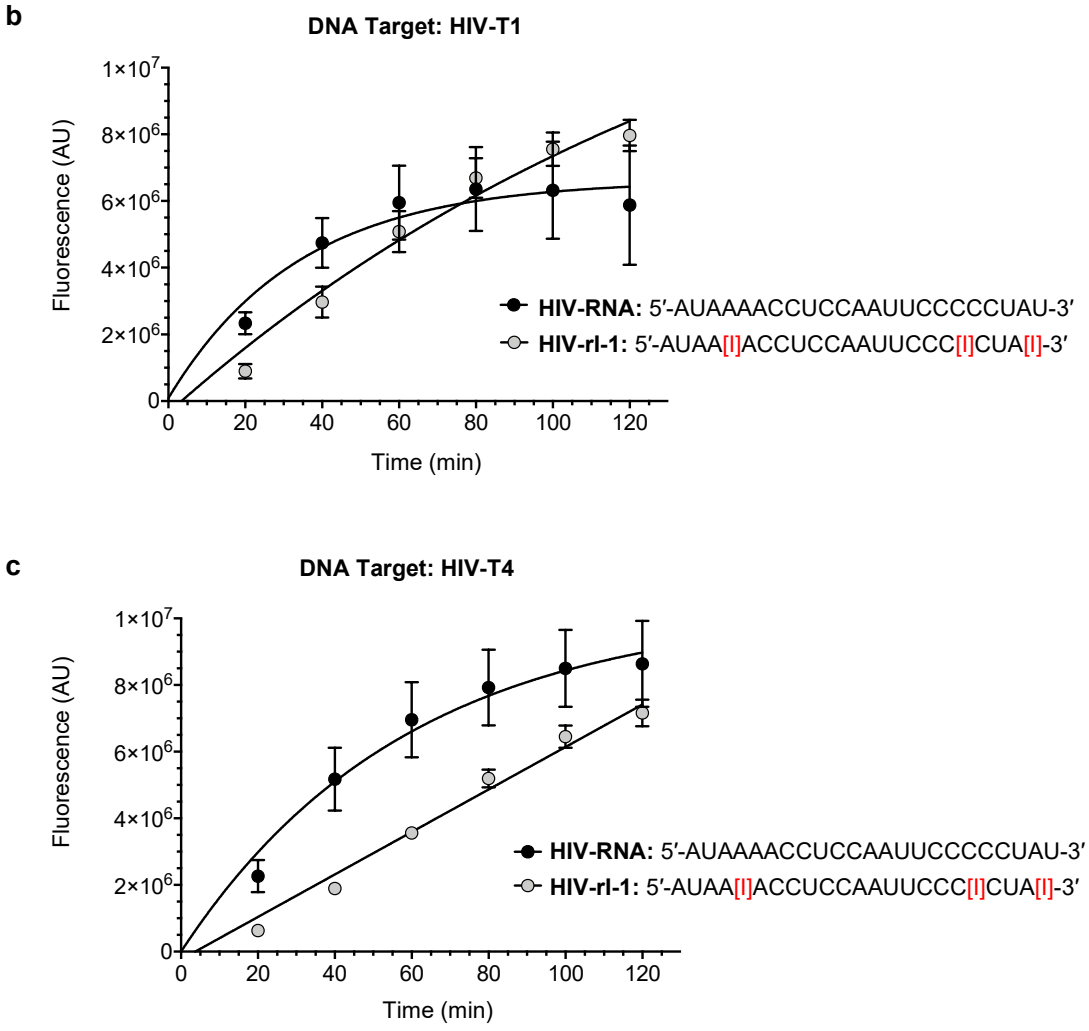


Figure 25. Effect of ribose inosine modified gRNAs on the kinetics of Cas12a using DETECTR assay. a List of DNA targets used in the DETECTR experiments. Red text indicates the presence of a polymorphism resulting in a drug resistance mutation. Graph of fluorescence indicating cleavage of **b** HIV-T1 or **c** HIV-T4 DNA targets using HIV-RNA and HIV-rI-1 gRNAs. Red [I] indicates a ribose inosine modification. Each

DNA target was amplified and added to 250 nM Cas12a, 312.5 nM gRNA, and 250 nM ssDNA-FQ reporter. Background subtracted fluorescence measurements were taken every 2 mins. Individual data points are shown for every 20 mins (SD, n=3).

Discussion

Cas9 tolerates several chemical alterations within the guide segment of its crRNA, including 2'O methylation[152], 2'O-4'C linkages[5], and 2'deoxyribose modifications[151], as well as phosphate backbone modifications such as phosophorothioate[152] and phosphonoacetate[153]. This flexibility is quite remarkable given that the enzyme forms several direct contacts with sugar moieties within this region (e.g. via Thr404) and over 10 interactions with the guide segment phosphate backbone[155]. In contrast, crystal structures have elucidated only a couple of interactions between Cas9 amino acids and bases present within the gRNA targeting region. This is true for Cas12a/Cpf1 also, where only 2 direct interactions between Gln286 and Asn175 and bases within the targeting portion of the crRNA, have been reported[156]. Here, we provide the first evidence that several classes of chemically unrelated bases may be tolerated within crRNAs. Importantly, this effect appears to be dependent on the chemical structure of the base (e.g. I, NI, K, P), as well as the number and positioning of the substitutions.

We show that while ribose inosine and its derivates can be incorporated into guide RNAs to impart selective degeneracy, this capability is context-dependent (**Figures 15, 17**). The causes underlying the reduction in cleavage activity observed for several inosine-modified crRNAs are likely multifarious, involving structural, thermodynamic, and kinetic aspects. First, it is possible that the degenerate base-pairing ability of inosine substitutions within a crRNA could lead to the emergence of

low-abundance RNA secondary structures. Second, with certain positional exceptions (e.g. addition of Inosine to the end of a Watson-Crick helix), changing A-U to I-U and G-U to I-C base pairs in a nucleic acid duplex is thermodynamically destabilizing[157]. Moreover, the magnitude of this energy change is dependent on the nearest-neighbour 5' and 3' bases, following a general decreasing stability trend of G-C > C-G > A-T >T-A[105]. This effect was evident in a comparison of the activity of crRNAs bearing single inosine substitutions at different locations (**Figure 26**). While our preliminary melting analysis did not reveal a major drop in duplex T_m in the case of ABO-rI-1 (2 substitutions) **(Figure 27**), it is possible that this effect could predominate in crRNAs with more substitutions or different sequence contexts.

a

Name	Sequence (5'→3')	Activity on ABO-T1 (±SEM)	[I] Base Pairing	5' Base Pairing	3' Base Pairing
ABO-RNA	CAUGGAGUUCCGCGACCACG(G)	95 ± 2			
ABO-rI-3	CAUGGAGUUCCGCGACCA[I]G	6 ± 2	I·G	A·T	**G·C**
ABO-rI-4	CAUGGAGUUC[I]GCGACCACG	20 ± 11	I·G	**C·G**	**G·C**
ABO-rI-5	CAUGGAG[I]UCCGCGACCACG	18 ± 3	**I·A**	**G·C**	T·A
ABO-rI-6	CAUGGAGUUCCGCGACCAC[I]	34 ± 9	**I·C**	**C·G**	**G·C**
HLA-RNA	CACACAGAUCUACAAGGCCC(G)	68 ± 8			
HLA-rI-1	[I]ACACAGAUCUACAAGGCCC	89 ± 3	I·G	N/A	A·T
HLA-rI-2	CACACAGAUCU[I]CAAGGCCC	75 ± 6	I·T	T·A	**C·G**
HLA-rI-3	CACACAGAUCUACAAGGCC[I]	35 ± 7	I·G	**C·G**	**G·C**

b

Position	Stability Trends
[I] Base Pairing	**I·C** > **I·A** > I·T ≈ I· G > I·I
5' Base Pairing	**G·C** > **C·G** > A·T > T·A
3' Base Pairing	**G·C** > **C·G** > A·T > T·A

Figure 26. Context dependency of single inosine substitutions. a Table of single inosine substituted crRNAs. Inosine modifications are indicated by [I]. Activity is based on *in vitro* cleavage assays using single concentrations of Cas9 (40 nM) and gRNA (80 nM). Data are presented as SEM n=2. **b** Stability trends of inosine base pairings and surrounding base pairs seen in PCR primers. Red text indicates a preferred base pairing.

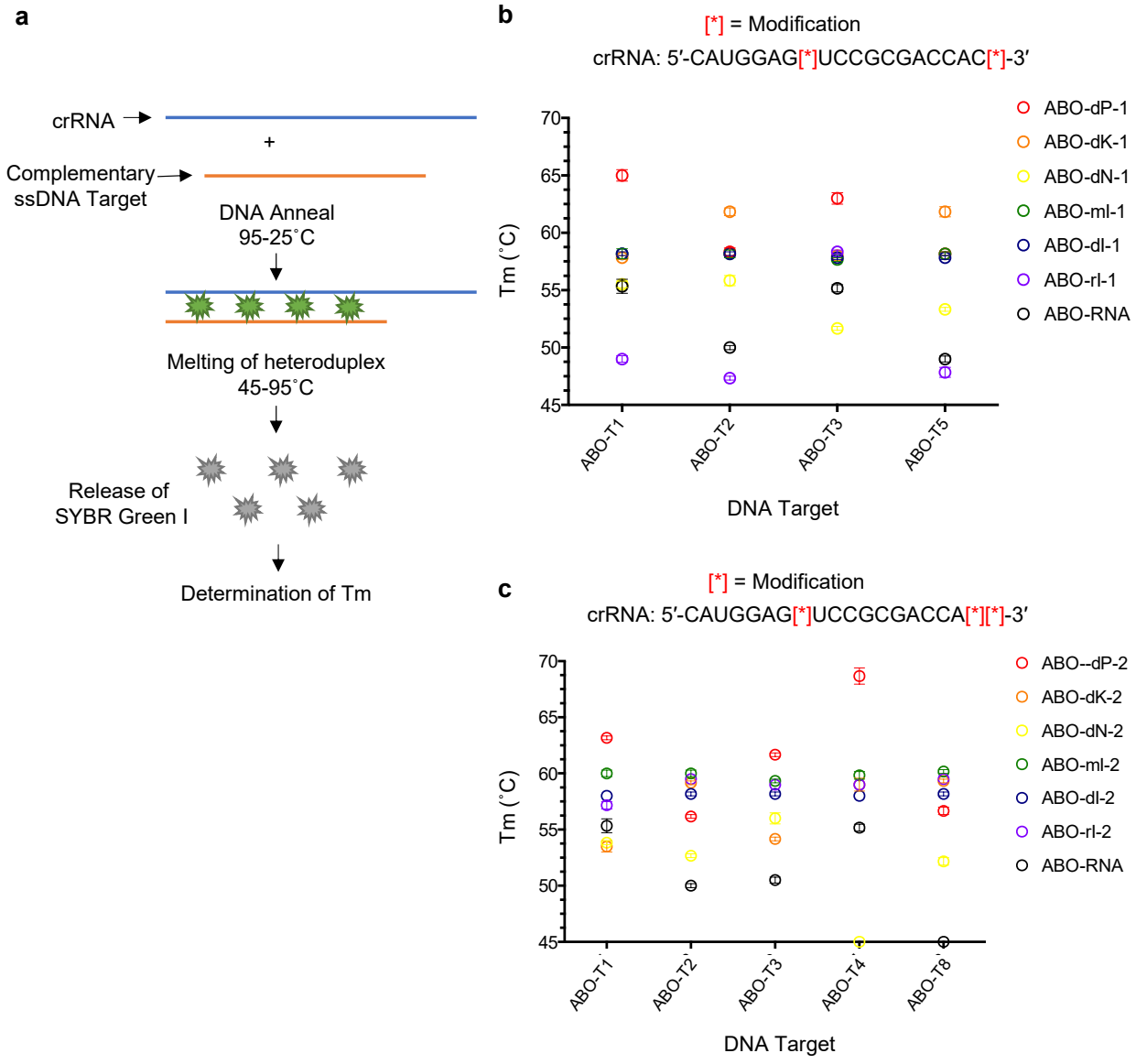


Figure 27. Effect of universal bases on crRNA:DNA heteroduplex melting temperature. a Representative diagram of the experiment used to determine the

melting temperature. Heteroduplexes of ssDNA target (orange) and crRNA (blue) are annealed together. SYBR Green I (green) intercalates into the duplex and results in maximal fluorescence. Subsequent heating of the duplex releases the SYBR Green I (gray). **b** Experimentally determined melting temperatures for double substituted and **c** triple substituted crRNAs against listed DNA targets (**Figure 16A**). [*] represents the position of a universal base. Individual data points are shown (SEM, n=3).

Synthetic universal base analogs were designed to improve upon the properties of inosine[150]. While isolated 5-NI substitutions are more destabilizing than inosine, due to the lack of hydrogen bonding, 5-NI pairing is completely unbiased. In addition, short contiguous stretches of 5-NI are more tolerated than contiguous stretches of inosine[114]. Single deoxyK and deoxyP substitutions were shown to allow for partially degenerate C/T and A/G pairing, respectively, without any significant loss in duplex thermodynamic stability[110,111]. In this work, we demonstrate that the inclusion of any of these bases can be tolerated to a certain extent (**Figure 15**) and used to enable degenerate targeting by Cas9 (**Figures 17, 18**). Similar to inosine substitutions, we did not observe any substantial decrease in duplex T_m with these synthetic bases (**Figure 27**). However, this potential destabilization could have been offset by the additional 2'deoxy modifications[151]. Overall, it is likely that decreases in the basal DNA cleavage activity of Cas9 observed when using crRNAs containing dN, dP, and dK substitutions result from the potential causes listed above for inosine.

The 1000 genomes project identified over 85 million SNPs, 3.6 million short indels, and 60,000 structural variants[92]. Equally incredible is the fact that certain viruses can have mutation rates in the order of 1 in 10^{-3}, roughly 1 million times higher than in humans[158]. These examples highlight the enormous pool of existing genetic diversity, as

well as its continuous expansion. While current CRISPR/Cas technology is not adequately equipped to deal with the challenge of this complexity, this study outlines a viable solution. We show that incorporation of universal bases into crRNAs can enable multi-sequence targeting *in vitro* and in cells using individual crRNAs. This could facilitate the development and approval of single gene editing therapeutics designed to work on all individuals in a population (**Figures 15, 17, 23**). Furthermore, we demonstrate how degenerate crRNA technology can be applied to the design of diagnostic Cas12a DETECTR probes (**Figure 24**). We envision that this could help reduce the false-negative detection rate of the system by imparting the platform with the flexibility to take into account pathogen evolution (either documented or predicted). Finally, future work could aim to assess if contiguous stretches of universal bases can be incorporated into crRNAs to reduce the functional portion of the fixed 20-bp spacer sequence. This would be useful for targeting shorter sequences, in which 20-bp of sequence information may be unavailable, genes with small indels, or even for the design of artificial transcription factors that mimic the shorter consensus sequence of many natural DNA-binding domains[76].

In sum, this work details the first demonstration that incorporation of non-canonical, universal nucleic acid bases can be tolerated in Cas9/Cas12a guide RNAs and modify targeting specificity. We show that the inclusion of distinct types of universal bases into individual guide RNAs can impart Cas enzymes with the ability to cleave the whole series of polymorphic gene variants *in vitro* and in cells. Furthermore, we delineate how this technology can be applied to diagnostics to circumvent false-negative results caused by pathogen evolution. By relaxing the current restrictions of

guide RNA targeting, we anticipate this work will enable many new applications of Cas9, Cas12a/Cpf1, and potentially other CRISPR systems.

CHAPTER 5: General Discussion

Overview of Chapter 3

In **Chapter 3**, I outlined four collaborations completed throughout my degree, to create KO cell lines using CRISPR/Cas9. I demonstrated the methods for validation of single cell clonal populations at both the DNA and protein level. Full KOs were successfully created for the CRMP2A and FAM120B proteins (**Figures 5, 8**). While the N2a cell line created contained significantly lowered levels of B4GALNT1/GM1 (**Figure 12**). Finally, based on Sanger sequencing analysis, MC3T3 cells are expected to contain 50% KD of SF3B4 protein (**Figure 14**). Furthermore, functional analysis was performed on FAM120B KOs which demonstrated their potential role in facilitating the effects of a small synthetic peptide on mitochondrial genes (**Figure 9**). Overall, these cell lines will provide valuable cellular models of reduced or null protein levels for each collaborator's lab. They will hopefully give these researchers insight into the affect CRMP2A levels have on lung cancer characteristics, FAM120B interaction in facilitating mitochondrial gene effects, ganglioside involvement in extracellular vesicle movement, and the role of SF3B4 haploinsufficiency in Nager's Syndrome.

Overview of Chapter 4

Currently, one of the challenges facing the CRISPR system is the inability to efficiently target areas of DNA containing innate polymorphism, without causing off-target effects. In **Chapter 4**, we present a potential solution to this problem in the form of universal bases. Following the outline in **Figure 28**, we started with designing crRNAs, which have specific substitutions of the natural RNA with universal or degenerate bases, and

we determined the tolerable numbers and placements of each potential base *in vitro* (**Figures 15, 16**). This was completed by using an *in vitro* cleavage assay to determine the amount of activity Cas9 had in combination with each of the modified gRNAs. This was done against DNA targets that contained natural SNPs that would not have been cleaved by the wildtype gRNA (**Figure 16**). We found that gRNAs with ribose inosine substitutions were able to target both the naturally found DNA target (ABO-T1) and any targets that contained SNPs relevant to the modified positions in the crRNA (**Figures 15, 16**). Next, a specificity profile was determined for each modified gRNA to ensure that the degeneracy occurring in the cleavage assays was not a loss of global targeting specificity. Data from a library containing $>10^{\wedge 12}$ targets showed the modified gRNAs did not lose the ability to target specifically (**Figures 17-22**). Non-bias towards a DNA base was observed at the positions of the crRNA which contained the modifications. Therefore, both the challenge of targeting polymorphic regions and limiting off-target cleavage is tackled by including these modifications at specific locations. We then investigated if these modified gRNAs were tolerated in cells, by using an RFP/GFP reporter assay (**Figure 23**). Using the best gRNA from the previous *in vitro* assays (ABO-rI-2), it was found that similar results observed *in vitro*, were echoed by the cellular assay (**Figure 23**). Thus, these modified gRNAs are not impacting the ability to efficiently deliver an RNP complex into cells. Finally, we decided to apply our modified gRNAs to the Cas12a system to determine whether they could be widely adapted to more clinical applications (**Figure 24**). We utilized a previously created DETECTR system, which is used to detect short DNA sequences, for screening viral and bacterial related disease[4]. We found our ribose inosine modified gRNA worked better in the

Cas12a system and was able to effectively eliminate false-negative results seen when using HIV-RNA (**Figure 24**). This finding creates an opportunity to screen multiple versions of viral and bacterial DNA without using more than one gRNA. Overall, our findings provide researchers with the ability to effectively and specifically target multiple versions of a polymorphic DNA target without producing harmful off-targets. Furthermore, these modifications can be used both *in vitro* and in cells and are very well tolerated in both CRISPR/Cas9 and Cas12a systems. In conclusion, the use of universal base modifications opens up the potential for targeting polymorphic regions, for both clinical detection and gene editing purposes.

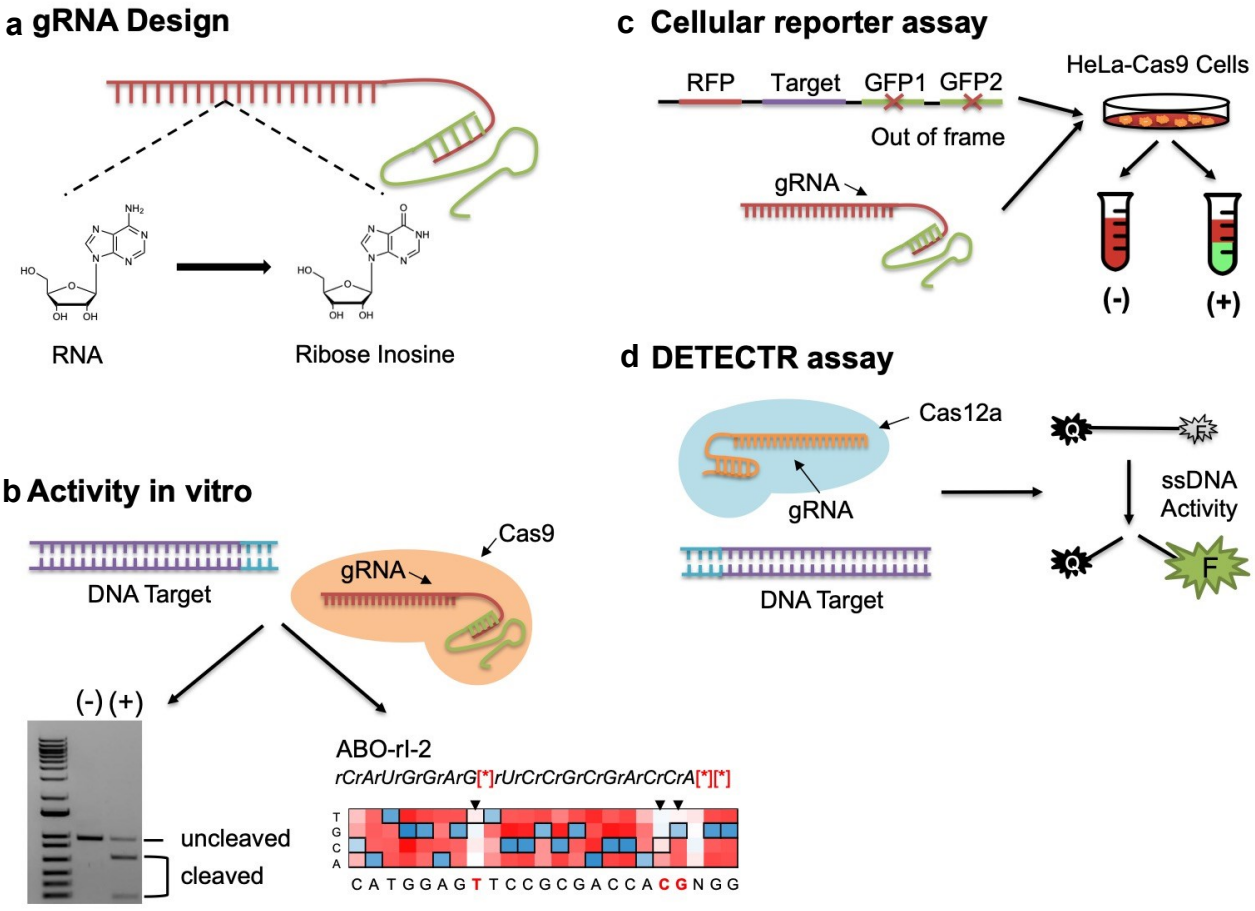


Figure 28. Overview of Chapter 4 workflow. a Diagram depicting the initial design of gRNA. Universal bases were substituted for original RNA bases based on the position of relevant SNPs in the corresponding DNA target. **b** *In vitro* assays used to determine the cleavage activity of Cas9 using each modified gRNA. gRNAs were initially tested *in vitro* (depicted on the left) and then a specificity profile (right) was created to determine if the incorporation of the modifications resulted in an overall decrease of specificity. **c** Workflow of cellular reporter assay. The best gRNA from the *in vitro* assays was verified in cells. **d** Verification of universal bases in a different CRISPR system. Ribose inosines were used in the CRISPR/Cas12a system to screen multiple versions of a highly polymorphic *HIV-1* DNA target with the DETECTR assay.

Limitations

While our results are promising, there are a few limiting factors for testing universal bases in the CRISPR system. First, due to the magnitude of crRNAs required, being able to provide a thorough outline of where the modifications may be better tolerated in the crRNA is not feasible for this research. Research shows that in PCR primers, the stability of inosine substitutions is based on the base pairs surrounding them, and this is also seen in our crRNAs (**Figure 26**)[105]. We also show that the kinetics of Cas9 with these modified crRNAs may play a role in the activity of the complex (**Figure 29**). However, being able to further test all the combinations in a 20 base pair crRNA would provide a lot more detail into the effect various combinations have on activity. Studying more combinations and possibly more genetic targets would give valuable information for the future design of universal base modified crRNAs. Furthermore, combining this comprehensive profile with a CRISPR design tool, such as the recently developed SNP-CRISPR[159], would exponentially expand the ability of researchers to target SNPs.

Currently, this tool takes variant data for a targeting crRNA sequence and provides efficiency scores for sgRNAs targeting both the reference genome and SNPs. Although this idea is promising, SNPs still present a major challenge for targeting and cleavage ability when there are ≥ 2 present in the target region[2] (**Figure 16**). By using universal bases in the positions of variants of interest, researchers could more easily design efficient sgRNAs, without being limited by low on and off-target scores. Finally, although we have shown that universal bases do not result in an overall non-specific targeting profile (**Figures 17-22**), further examination of all potential off-targets *in vivo* would solidify this claim. Off-targets are a major obstacle in the use of CRISPR in human clinical applications[1]. We have effectively used these bases in the DETECTR system to prevent false negatives in a polymorphic genome (**Figure 24**), however, in order to use these bases in cells, a more extensive profile of potential off-targets is needed. This could be accomplished by using a system such as GUIDE-seq to provide an overarching insight on any detrimental effects in cells[160]. Overall, more testing of these bases in different crRNAs and *in vivo* is required to create a more unanimous understanding of how to utilize them in the CRISPR system.

Future Directions

Future Experiments

To better understand why the inclusion of inosine in crRNAs reduces activity in certain instances, future studies will be done using a series of thermodynamic, kinetic, and biophysical experiments using the three crRNAs: ABO-RNA, ABO-rI-1, and ABO-rI-8. First, we will evaluate the effect of inosine substitution on crRNA hybridization to ABO-

T1 in the absence of Cas9. Since I*G and I*A pairs have been shown to decrease thermodynamic duplex stability by 0.84 kcal/mol and 0.52 kcal/mol compared to C*G and A*T pairs, respectively[105], we wonder if a substantial drop in duplex T_m could preclude annealing and explain the decreased activity. As shown in **Figure 27**, preliminary melting experiments show T_m values for hybridization of double substituted crRNAs are increased by the addition of certain modifications. Second, we will alter the crRNA:tracrRNA ratio to examine if guide RNA annealing was impaired by structural changes caused by the inclusion of inosine. Third, we will investigate if Cas9 target binding was affected by performing titrations of guide RNA as well the as whole RNP complex.

Cas9 DNA cleavage may be divided into several stages: 1) guide RNA loading, 2) PAM searching, 3) DNA melting and PAM-proximal hybridization, 4) full duplex formation, and 5) structural rearrangement leading to nuclease activation[161]. If crRNA/tracrRNA hybridization and overall RNP target binding are ruled out as potential causes for the reduced activity of ABO-rI-1 and ABO-rI-8, we will use a previously validated single-molecule fluorescence resonance energy transfer (smFRET) assay to pinpoint which phase of the DNA cleavage reaction was being influenced by incorporation of inosine. Briefly, this assay uses a Cy5-labeled crRNA which is complexed to Cas9 and a Cy3-labeled substrate that is immobilized on a quartz surface[161]. Changes in FRET that occur between the dye pairs during R-loop formation correspond to the transitional dynamics between the partially zipped (“open”; low energy) and the fully zipped (“zipped”; high energy) conformations of the Cas9 complex[5,161]. Next, we will study the kinetics of the open-zipped transition by analyzing

single-molecule time trajectories. Time-courses of ABO-T1 cleavage by Cas9 using double substitutions of universal bases revealed a decrease in V_{max} relative to ABO-RNA, confirming this assertion (**Figure 29**). These preliminary results are in agreement with independent studies on distinct crRNA chemical modifications[5], and with mFRET analysis, would imply that inclusion of inosine into crRNAs may lead to lower cleavage activity due to slower kinetics. Overall, these future experiments will determine the reason why the incorporation of inosines may lead to lower activity and could comment on how to incorporate these modifications in the future.

a

Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG
ABO-[*]-1	CAUGGAG[*]UCCGCGACCAC[*]

b

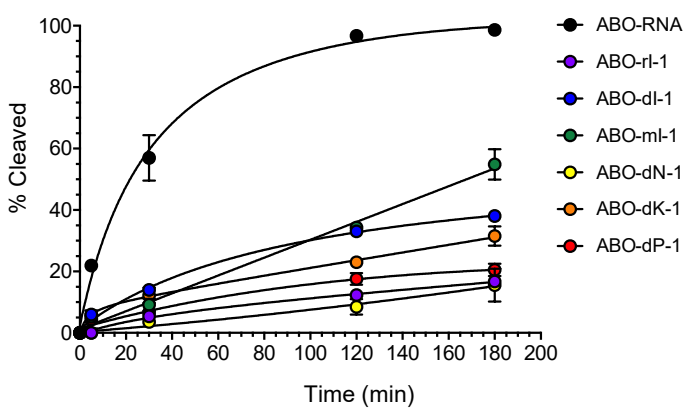


Figure 29. Effect of modified crRNAs on the kinetics of Cas9 *in vitro*. a List of crRNA sequences used. [*] represents the position of the universal base. **b** Graph showing ABO-T1 cleavage over 3 hours. *In vitro* cleavage assays were performed using fixed concentrations gRNA (80 nM) and Cas9 (40 nM) for time points of 0 min, 5 mins, 30 mins, 120 mins, and 180 mins. Individual data points are shown (SEM, n=2).

Applications of Research

We have established that universal bases can be utilized in both the Cas9 and Cas12a CRISPR systems. Future applications of using universal bases are outlined in **Figure 30**. One of which is the screening of highly polymorphic regions including viral and bacterial genomes. Potentially, these could be used in both DNA and RNA targeting screens after testing of universal bases in the CRISPR/Cas13 system. This would allow for even faster and cheaper testing as only one gRNA would be required to target 4 versions of an SNP. Secondly, CRISPR/Cas9 has already been used clinically to target viral genomes such as *HIV-1*[91] and other infectious diseases[62]. Having access to targeting highly mutable regions of the *HIV-1* genome would solve problems associated with reduced single gRNA efficiency against these viruses[99,101]. Finally, polymorphic regions of human genomic DNA presented a problem that is currently being solved by a lacklustre multiple gRNA approach[87]. By incorporating universal bases such as ribose inosines into the gRNA, researchers no longer have to design and deliver multiple gRNAs for the same DNA target region, simplifying this process. In conclusion, many current and future applications of the CRISPR system can be improved upon or further adapted by the use of universal base modified gRNAs.

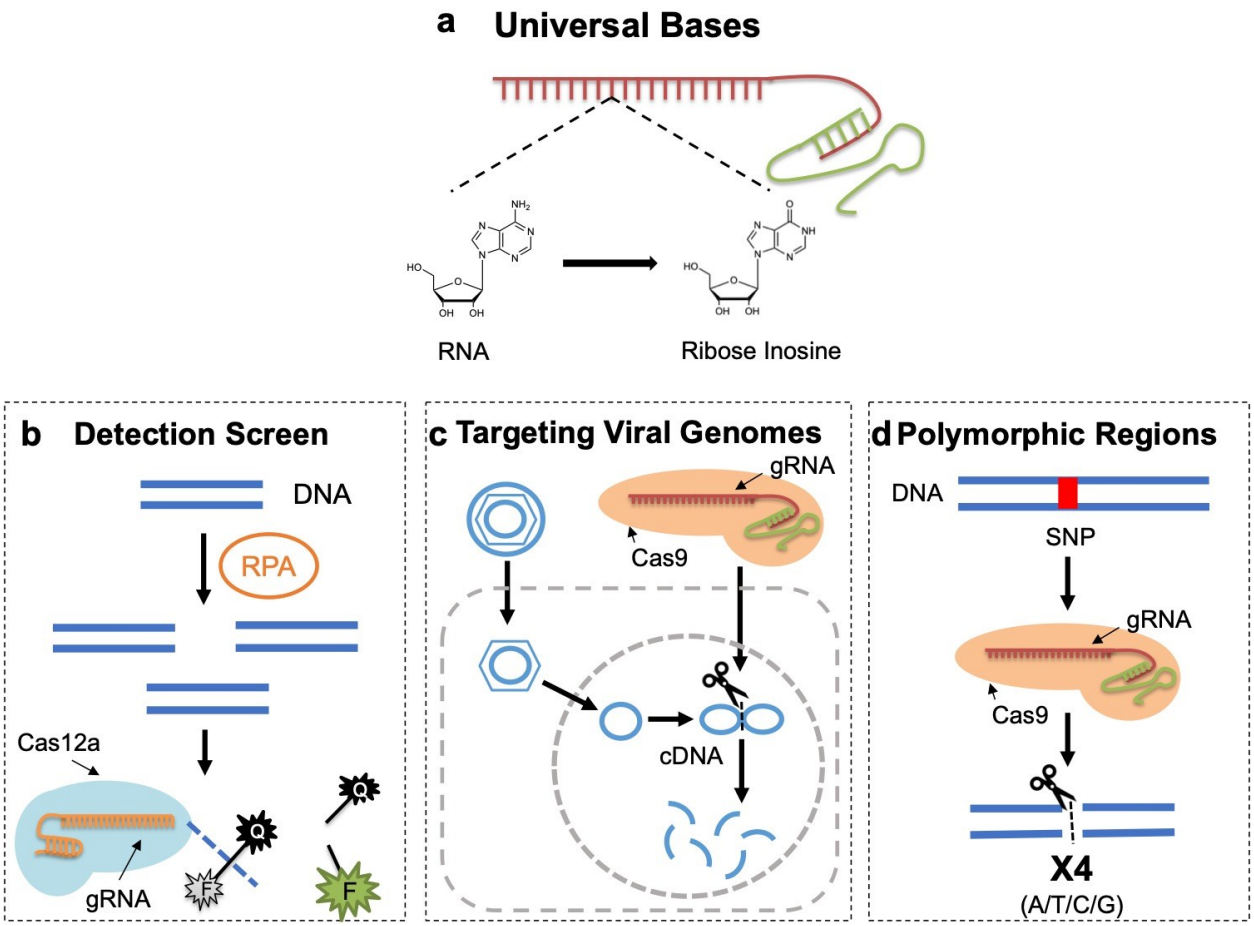


Figure 30. Future applications of universal bases in the CRISPR system. a Diagram depicting the substitution of RNA bases for ribose inosines in CRISPR gRNA. **b** Diagram representing the application of universal bases in current detection screens. **c** Application of universal bases in the CRISPR/Cas9 system to target highly polymorphic viral cDNA. **d** Utilizing universal bases to target and cleave all versions of an SNP in a DNA target.

References

1. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* **23**, R40-46 (2014).
2. Lessard, S. *et al.* Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11257–E11266 (2017).
3. Loakes, D. SURVEY AND SUMMARY. *Nucleic Acids Res.* **29**, 2437–2447 (2001).
4. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
5. Cromwell, C. R. *et al.* Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat. Commun.* **9**, 1448 (2018).
6. Smith, H. O. & Welcox, K. W. A Restriction enzyme from Hemophilus influenzae: I. Purification and general properties. *J. Mol. Biol.* **51**, 379–391 (1970).
7. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* **42**, 3–19 (2014).
8. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
9. Klug, A. & Rhodes, D. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 473–482 (1987).

10. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512 (2009).

11. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160 (1996).

12. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).

13. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

14. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).

15. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiol. Read. Engl.* **151**, 2551–2561 (2005).

16. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).

17. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).

18. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).

19. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).

20. Deltcheva, E. *et al.* CRISPR RNA maturation by trans -encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

21. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**, 9275–9282 (2011).

22. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).

23. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

24. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic Mapping in Human Disease. *Science* **322**, 881–888 (2008).

25. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).

26. Humbert, O., Davis, L. & Maizels, N. Targeted gene therapies: tools, applications, optimization. *Crit. Rev. Biochem. Mol. Biol.* **47**, 264–281 (2012).

27. Prakash, V., Moore, M. & Yáñez-Muñoz, R. J. Current Progress in Therapeutic Gene Editing for Monogenic Diseases. *Mol. Ther.* **24**, 465–474 (2016).

28. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1–13 (2018).

29. Kotagama, O. W., Jayasinghe, C. D. & Abeysinghe, T. Era of Genomic Medicine: A Narrative Review on CRISPR Technology as a Potential Therapeutic Tool for

Human Diseases. *BioMed Research International* vol. 2019 e1369682 https://www.hindawi.com/journals/bmri/2019/1369682/ (2019).

30. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2579-2586 (2012).

31. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

32. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

33. Ma, Y., Zhang, L. & Huang, X. Genome modification by CRISPR/Cas9. *FEBS J.* **281**, 5186–5193 (2014).

34. Miura, H., Quadros, R. M., Gurumurthy, C. B. & Ohtsuka, M. Easi -CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nat. Protoc.* **13**, 195–215 (2018).

35. Zhang, Y., Malzahn, A. A., Sretenovic, S. & Qi, Y. The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants* **5**, 778–794 (2019).

36. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).

37. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).

38. Marraffini, L. A. & Sontheimer, E. J. Self vs. non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).

39. O'Reilly, D. *et al.* Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity. *Nucleic Acids Res.* **47**, 546–558 (2019).

40. Moon, S. B., Kim, D. Y., Ko, J.-H. & Kim, Y.-S. Recent advances in the CRISPR genome editing tool set. *Exp. Mol. Med.* **51**, 1–11 (2019).

41. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).

42. Tóth, E. *et al.* Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biol. Direct* **11**, (2016).

43. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).

44. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284 (2017).

45. Freije, C. A. *et al.* Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Mol. Cell* **76**, 826-837.e11 (2019).

46. Liu, J.-J. *et al.* CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).

47. Murugan, K., Babu, K., Sundaresan, R., Rajan, R. & Sashital, D. G. The revolution continues: Newly discovered systems expand the CRISPR-Cas toolkit. *Mol. Cell* **68**, 15–25 (2017).

48. Giuliano, C. J., Lin, A., Girish, V. & Sheltzer, J. M. Generating Single Cell–Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. *Curr. Protoc. Mol. Biol.* **128**, e100 (2019).

49. Rodríguez-Rodríguez, D. R., Ramírez-Solís, R., Garza-Elizondo, M. A., Garza-Rodríguez, M. D. L. & Barrera-Saldaña, H. A. Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review). *Int. J. Mol. Med.* **43**, 1559–1574 (2019).

50. Koch, B. *et al.* Generation and validation of homozygous fluorescent knock-in cells using CRISPR/Cas9 genome editing. *Nat. Protoc.* **13**, 1465–1487 (2018).

51. Hyodo, T. *et al.* Tandem Paired Nicking Promotes Precise Genome Editing with Scarce Interference by p53. *Cell Rep.* **30**, 1195-1207.e7 (2020).

52. Bortesi, L. & Fischer, R. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **33**, 41–52 (2015).

53. Matsuda, T. & Oinuma, I. Optimized CRISPR/Cas9-mediated in vivo genome engineering applicable to monitoring dynamics of endogenous proteins in the mouse neural tissues. *Sci. Rep.* **9**, 1–13 (2019).

54. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).

55. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

56. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).

57. You, L. *et al.* Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Mol. Ther. - Methods Clin. Dev.* **13**, 359–370 (2019).

58. Salas-Mckee, J. *et al.* CRISPR/Cas9-based genome editing in the era of CAR T cell immunotherapy. *Hum. Vaccines Immunother.* **15**, 1126–1132 (2019).

59. Liu, X. & Zhao, Y. CRISPR/Cas9 genome editing: Fueling the revolution in cancer immunotherapy. *Curr. Res. Transl. Med.* **66**, 39–42 (2018).

60. Afolabi, L. O., Adeshakin, A. O., Sani, M. M., Bi, J. & Wan, X. Genetic reprogramming for NK cell cancer immunotherapy with CRISPR/Cas9. *Immunology* **158**, 63–69 (2019).

61. Gao, Q. *et al.* Therapeutic potential of CRISPR/Cas9 gene editing in engineered T-cell therapy. *Cancer Med.* **8**, 4254–4264 (2019).

62. Strich, J. R. & Chertow, D. S. CRISPR-Cas Biology and Its Application to Infectious Diseases. *J. Clin. Microbiol.* **57**, (2019).

63. Terns, M. P. CRISPR-Based Technologies: Impact of RNA-Targeting Systems. *Mol. Cell* **72**, 404–412 (2018).

64. Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360**, 439–444 (2018).

65. Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438–442 (2017).

66. Bai, J. *et al.* Cas12a-Based On-Site and Rapid Nucleic Acid Detection of African Swine Fever. *Front. Microbiol.* **10**, (2019).

67. Hatoum-Aslan, A. CRISPR Methods for Nucleic Acid Detection Herald the Future of Molecular Diagnostics. *Clin. Chem.* **64**, 1681–1683 (2018).

68. Myhrvold, C. *et al.* Field-deployable viral diagnostics using CRISPR-Cas13. *Science* **360**, 444–448 (2018).

69. Enabling coronavirus detection using CRISPR-Cas13: An open-access SHERLOCK research protocol. *MIT McGovern Institute* https://mcgovern.mit.edu/2020/02/14/enabling-coronavirus-detection-using-crispr-cas13-an-open-access-sherlock-research-protocol/ (2020).

70. Broughton, J. P. *et al.* CRISPR–Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* 1–5 (2020) doi:10.1038/s41587-020-0513-4.

71. Abbott, T. R. *et al. Development of CRISPR as a prophylactic strategy to combat novel coronavirus and influenza*. http://biorxiv.org/lookup/doi/10.1101/2020.03.13.991307 (2020) doi:10.1101/2020.03.13.991307.

72. Lino, C. A., Harper, J. C., Carney, J. P. & Timlin, J. A. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv.* **25**, 1234–1257 (2018).

73. Tong, S., Moyo, B., Lee, C. M., Leong, K. & Bao, G. Engineered materials for in vivo delivery of genome-editing machinery. *Nat. Rev. Mater.* **4**, 726–737 (2019).

74. McCarty, N. S., Graham, A. E., Studená, L. & Ledesma-Amaro, R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1–13 (2020).

75. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. - Nucleic Acids* **4**, e264 (2015).

76. Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **12**, 939–942 (2015).

77. Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 1–10 (2018).

78. Filippova, J., Matveeva, A., Zhuravlev, E. & Stepanov, G. Guide RNA modification as a way to improve CRISPR/Cas9-based genome-editing systems. *Biochimie* **167**, 49–60 (2019).

79. Aschenbrenner, S. *et al.* Coupling Cas9 to artificial inhibitory domains enhances CRISPR-Cas9 target specificity. *Sci. Adv.* **6**, eaay0187 (2020).

80. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).

81. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

82. Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).

83. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

84. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).

85. Scott, D. A. & Zhang, F. Implications of human genetic variation in CRISPR-based therapeutic genome editing. *Nat. Med.* **23**, 1095–1101 (2017).

86. Zheng, T. *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci. Rep.* **7**, 40638 (2017).

87. Canver, M. C., Joung, J. K. & Pinello, L. Impact of Genetic Variation on CRISPR-Cas Targeting. *CRISPR J.* **1**, 159–170 (2018).

88. Canver, M. C. *et al.* Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.* **49**, 625–634 (2017).

89. Yang, L. *et al.* Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat. Commun.* **5**, 1–6 (2014).

90. Pursey, E., Sünderhauf, D., Gaze, W. H., Westra, E. R. & Houte, S. van. CRISPR-Cas antimicrobials: Challenges and future prospects. *PLOS Pathog.* **14**, e1006990 (2018).

91. Panfil, A. R., London, J. A., Green, P. L. & Yoder, K. E. CRISPR/Cas9 Genome Editing to Disable the Latent HIV-1 Provirus. *Front. Microbiol.* **9**, (2018).

92. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

93. Shastry, B. S. SNPs in disease gene mapping, medicinal drug development and evolution. *J. Hum. Genet.* **52**, 871–880 (2007).

94. Nonnemacher, M. R. *et al.* HIV-1 Promoter Single Nucleotide Polymorphisms Are Associated with Clinical Disease Severity. *PLOS ONE* **11**, e0150835 (2016).

95. Jin, P. & Wang, E. Polymorphism in clinical immunology – From HLA typing to immunogenetic profiling. *J. Transl. Med.* **1**, 8 (2003).

96. Daniels, G. The molecular genetics of blood group polymorphism. *Transpl. Immunol.* **14**, 143–153 (2005).

97. Wainberg, M. A. & Brenner, B. G. The Impact of HIV Genetic Polymorphisms and Subtype Differences on the Occurrence of Resistance to Antiretroviral Drugs. *Molecular Biology International* vol. 2012 e256982 https://www.hindawi.com/journals/mbi/2012/256982/ (2012).

98. Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J. & Sanjuán, R. Extremely High Mutation Rate of HIV-1 In Vivo. *PLOS Biol.* **13**, e1002251 (2015).

99. Darcis, G. *et al.* The Impact of HIV-1 Genetic Diversity on CRISPR-Cas9 Antiviral Activity and Viral Escape. *Viruses* **11**, (2019).

100. Wang, Z. *et al.* CRISPR/Cas9-Derived Mutations Both Inhibit HIV-1 Replication and Accelerate Viral Escape. *Cell Rep.* **15**, 481–489 (2016).

101. De Silva Feelixge, H. S., Stone, D., Roychoudhury, P., Aubert, M. & Jerome, K. R. CRISPR/Cas9 and Genome Editing for Viral Disease - Is Resistance Futile? *ACS Infect. Dis.* **4**, 871–880 (2018).

102. Loakes, D. & Brown, D. M. 5-Nitroindole as an universal base analogue. *Nucleic Acids Res.* **22**, 4039–4043 (1994).

103. Gallego, J. & Loakes, D. Solution structure and dynamics of DNA duplexes containing the universal base analogues 5-nitroindole and 5-nitroindole 3-carboxamide. *Nucleic Acids Res.* **35**, 2904–2912 (2007).

104. Oda, Y., Uesugi, S., Ikehara, M., Kawase, Y. & Ohtsuka, E. NMR studies for identification of dI:dG mismatch base-pairing structure in DNA. *Nucleic Acids Res.* **19**, 5263–5267 (1991).

105. Watkins, N. E. & SantaLucia, J. Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.* **33**, 6258–6267 (2005).

106. Rozov, A. *et al.* Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* **7**, (2016).

107. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).

108. Gatsiou, A., Vlachogiannis, N., Lunella, F. F., Sachse, M. & Stellos, K. Adenosine-to-Inosine RNA Editing in Health and Disease. *Antioxid. Redox Signal.* **29**, 846–863 (2018).

109. Martin, F. H., Castro, M. M., Aboul-ela, F. & Tinoco, I. Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.* **13**, 8927–8938 (1985).

110. Lin, P. K. & Brown, D. M. Synthesis of oligodeoxyribonucleotides containing degenerate bases and their use as primers in the polymerase chain reaction. *Nucleic Acids Res.* **20**, 5149–5152 (1992).

111. Lin, P. K. & Brown, D. M. Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogues. *Nucleic Acids Res.* **17**, 10373–10383 (1989).

112. Ben-Dov, E. & Kushmaro, A. Inosine at Different Primer Positions to Study Structure and Diversity of Prokaryotic Populations. *Curr. Issues Mol. Biol.* **17**, 53–56 (2015).

113. Liu, H. & Nichols, R. PCR amplification using deoxyinosine to replace an entire codon and at ambiguous positions. *BioTechniques* **16**, 24–26 (1994).

114. Loakes, D., Brown, D. M., Linde, S. & Hill, F. 3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR. *Nucleic Acids Res.* **23**, 2361–2366 (1995).

115. Kawasaki, A. M. *et al.* Uniformly modified 2'-deoxy-2'-fluoro phosphorothioate oligonucleotides as nuclease-resistant antisense compounds with high affinity and specificity for RNA targets. *J. Med. Chem.* **36**, 831–841 (1993).

116. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

117. Shafer, R. W. Rationale and uses of a public HIV drug-resistance database. *J. Infect. Dis.* **194 Suppl 1**, S51-58 (2006).

118. Yip, S. P. Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles. *Blood* **95**, 1487–1492 (2000).

119. Shafer, R. W., Jung, D. R. & Betts, B. J. Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. *Nat. Med.* **6**, 1290–1292 (2000).

120. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).

121. Zetsche, B. *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).

122. Ramakrishna, S. *et al.* Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat. Commun.* **5**, 1–10 (2014).

123. Ip, J. P. K., Fu, A. K. Y. & Ip, N. Y. CRMP2: Functional Roles in Neural Development and Therapeutic Potential in Neurological Diseases. *The Neuroscientist* **20**, 589–598 (2014).

124. Arimura, N. & Kaibuchi, K. Neuronal polarity: from extracellular signals to intracellular mechanisms. *Nat. Rev. Neurosci.* **8**, 194–205 (2007).

125. Bretin, S. *et al.* Differential expression of CRMP1, CRMP2A, CRMP2B, and CRMP5 in axons or dendrites of distinct neurons in the mouse brain. *J. Comp. Neurol.* **486**, 1–17 (2005).

126. Indraswari, F., Wong, P. T. H., Yap, E., Ng, Y. K. & Dheen, S. T. Upregulation of Dpysl2 and Spna2 gene expression in the rat brain after ischemic stroke. *Neurochem. Int.* **55**, 235–242 (2009).

127. Grant, N. J. *et al.* Phosphorylation of a splice variant of collapsin response mediator protein 2 in the nucleus of tumour cells links cyclin dependent kinase-5 to oncogenesis. *BMC Cancer* **15**, 885 (2015).

128. Li, D., Kang, Q. & Wang, D.-M. Constitutive coactivator of peroxisome proliferator-activated receptor (PPARgamma), a novel coactivator of PPARgamma that promotes adipogenesis. *Mol. Endocrinol. Baltim. Md* **21**, 2320–2333 (2007).

129. Ceccaldi, R., Sarangi, P. & D'Andrea, A. D. The Fanconi anaemia pathway: new players and new functions. *Nat. Rev. Mol. Cell Biol.* **17**, 337–349 (2016).

130. Andrews, A. M., McCartney, H. J., Errington, T. M., D'Andrea, A. D. & Macara, I. G. A senataxin-associated exonuclease SAN1 is required for resistance to DNA interstrand cross-links. *Nat. Commun.* **9**, 2592 (2018).

131. Yamaguchi, T. *et al.* Expression of B4GALNT1, an essential glycosyltransferase for the synthesis of complex gangliosides, suppresses BACE1 degradation and modulates APP processing. *Sci. Rep.* **6**, 1–12 (2016).

132. Ledeen, R. W. & Wu, G. Gangliosides, α-Synuclein, and Parkinson's Disease. *Prog. Mol. Biol. Transl. Sci.* **156**, 435–454 (2018).

133. Bhuiyan, R. H. *et al.* Loss of Enzyme Activity in Mutated B4GALNT1 Gene Products in Patients with Hereditary Spastic Paraplegia Results in Relatively Mild Neurological Disorders: Similarity with Phenotypes of B4galnt1 Knockout Mice. *Neuroscience* **397**, 94–106 (2019).

134. Dinkins, M. B., Wang, G. & Bieberich, E. Sphingolipid-enriched extracellular vesicles and Alzheimer's disease: A decade of research. *J. Alzheimers Dis. JAD* **60**, 757–768 (2017).

135. van Niel, G., D'Angelo, G. & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat. Rev. Mol. Cell Biol.* **19**, 213–228 (2018).

136. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).

137. Xiong, F. & Li, S. SF3b4: A Versatile Player in Eukaryotic Cells. *Front. Cell Dev. Biol.* **8**, 14 (2020).

138. Petit, F. *et al.* Nager syndrome: confirmation of SF3B4 haploinsufficiency as the major cause. *Clin. Genet.* **86**, 246–251 (2014).

139. Bernier, F. P. *et al.* Haploinsufficiency of SF3B4, a Component of the Pre-mRNA Spliceosomal Complex, Causes Nager Syndrome. *Am. J. Hum. Genet.* **90**, 925–933 (2012).

140. Devotta, A., Juraver-Geslin, H., Gonzalez, J. A., Hong, C.-S. & Saint-Jeannet, J.-P. Sf3b4-depleted Xenopus embryos: A model to study the pathogenesis of craniofacial defects in Nager syndrome. *Dev. Biol.* **415**, 371–382 (2016).

141. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).

142. Alok, A. *et al.* The Rise of the CRISPR/Cpf1 System for Efficient Genome Editing in Plants. *Front. Plant Sci.* **11**, 264 (2020).

143. Nakade, S., Yamamoto, T. & Sakuma, T. Cas9, Cpf1 and C2c1/2/3-What's next? *Bioengineered* **8**, 265–273 (2017).

144. Kocak, D. D. *et al.* Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat. Biotechnol.* **37**, 657–666 (2019).

145. Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat. Chem. Biol.* **14**, 311–316 (2018).

146. Christie, K. A. *et al.* Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. *Sci. Rep.* **7**, 16174 (2017).

147. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).

148. Lindpaintner, K. Genetics in drug discovery and development: challenge and promise of individualizing treatment in common complex diseases. *Br. Med. Bull.* **55**, 471–491 (1999).

149. Clutterbuck, D. R., Leroy, A., O'Connell, M. A. & Semple, C. a. M. A bioinformatic screen for novel A-I RNA editing sites reveals recoding editing in BC10. *Bioinforma. Oxf. Engl.* **21**, 2590–2595 (2005).

150. Vallone, P. M. & Benight, A. S. Melting studies of short DNA hairpins containing the universal base 5-nitroindole. *Nucleic Acids Res.* **27**, 3589–3596 (1999).

151. Rueda, F. O. *et al.* Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat. Commun.* **8**, 1610 (2017).

152. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).

153. Ryan, D. E. *et al.* Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.* **46**, 792–803 (2018).

154. Ivanov, I. E. *et al.* Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5853–5860 (2020).

155. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).

156. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962 (2016).

157. Wright, D. J., Force, C. R. & Znosko, B. M. Stability of RNA duplexes containing inosine·cytosine pairs. *Nucleic Acids Res.* **46**, 12099–12108 (2018).

158. Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* **16**, e3000003 (2018).

159. Chen, C.-L. *et al.* SNP-CRISPR: A Web Tool for SNP-Specific Genome Editing. *G3 Genes Genomes Genet.* **10**, 489–494 (2020).

160. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

161. Lim, Y. *et al.* Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* **7**, 13350 (2016).

Appendix A: Unmodified, modified crRNAs, and tracrRNA sequences used.

[rI] = Ribose inosine, [dI] = Deoxyinosine, [mI] = 2'O methyl inosine, [dN] = 5′-nitroindole, [dK] = Degenerate base K, [dP] = Degenerate base P.

Name	Sequence (5′→3′)
HLA-RNA	rCrArCrArCrArGrArUrCrUrArCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-1	[rI]rArCrArCrArGrArUrCrUrArCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-2	rCrArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-3	rCrArCrArCrArGrArUrCrUrArCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-4	rCrArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-5	[rI]rArCrArCrArGrArUrCrUrArCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-6	[rI]rArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-7	[rI]rArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrC[rI] rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-RNA	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-1	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dI-1	rCrArUrGrGrArG[dI]rUrCrCrGrCrGrArCrCrArC[dI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-mI-1	rCrArUrGrGrArG[mI]rUrCrCrGrCrGrArCrCrArC[mI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dN-1	rCrArUrGrGrArG[5NitInd]rUrCrCrGrCrGrArCrCrArC[5NitInd]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dK-1	rCrArUrGrGrArG[dK]rUrCrCrGrCrGrArCrCrArC[dK]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dP-1	rCrArUrGrGrArG[dP]rUrCrCrGrCrGrArCrCrArC[dP]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-2	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dI-2	rCrArUrGrGrArG[dI]rUrCrCrGrCrGrArCrCrA[dI][dI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-mI-2	rCrArUrGrGrArG[mI]rUrCrCrGrCrGrArCrCrA[mI][mI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU

ABO-dN-2	rCrArUrGrGrArG[5NitInd]rUrCrCrGrCrGrArCrCrA[5NitInd][5NitInd]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dK-2	rCrArUrGrGrArG[dK]rUrCrCrGrCrGrArCrCrA[dK][dK]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dP-2	rCrArUrGrGrArG[dP]rUrCrCrGrCrGrArCrCrA[dP][dP]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-3	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-4	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-5	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-6	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-7	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-8	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-9	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-10	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-11	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-12	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-13	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-14	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-15	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HIV-RNA	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArArArArCrCrUrCrCrArArUrUrCrCrCrCrUrArU
HIV-rI-1	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArA[rI]rArCrCrUrCrCrArArUrUrCrCrC[rI]rCrUrA[rI]
HIV-rI-2	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArA[rI]rArCrCrUrCrCrA[rI]rUrUrCrCrCrCrCrUrArU
HIV-rI-3	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArA[rI]rArCrCrUrCrCrArArUrUrCrCrC[rI]rCrUrArU
HIV-rI-4	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArArArArCrCrUrCrCrA[rI]rUrUrCrCrCrCrCrUrA[rI]
HIV-rI-5	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArArArArCrCrUrCrCrArArUrUrCrCrC[rI]rCrUrA[rI]

HIV-rI-6	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArA[rI]rArCrCrUrCrCrA[rI]rUrUrCrCrCrCrCrUrA[rI]
HIV-rI-7	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArArArArArCrCrUrCrCrA[rI]rUrUrCrCrC[rI]rCrUrA[rI]
tracrRNA	rArGrCrArUrArGrCrArArGrUrUrArArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrUrUrU

Appendix B: Oligonucleotide sequences used.

Name	Sequences (5′→3′)
Oligos used to create invitro cleavage assay DNA target constructs.	
ABO-T1-F	GCCGAAGCTTCTCCACGTGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T1-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCACGTGGAGAAGCTTCGGC
ABO-T2-F	GCCGAAGCTTCTCCACGTGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T2-R	GGCCTCTAGAAGCATGGAGATCCGCGACCACGTGGAGAAGCTTCGGC
ABO-T3-F	GCCGAAGCTTCTCCACATGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T3-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCATGTGGAGAAGCTTCGGC
ABO-T4-F	GCCGAAGCTTCTCCATGTGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T4-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCACATGGAGAAGCTTCGGC
ABO-T5-F	GCCGAAGCTTCTCCATGTGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T5-R	GGCCTCTAGAAGCATGGAGATCCGCGACCACATGGAGAAGCTTCGGC
ABO-T6-F	GCCGAAGCTTCTCCATATGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T6-R	GGCCTCTAGAAGCATGGAGATCCGCGACCATATGGAGAAGCTTCGGC
ABO-T7-F	GCCGAAGCTTCTCCACATGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T7-R	GGCCTCTAGAAGCATGGAGATCCGCGACCATGTGGAGAAGCTTCGGC

ABO-T8-F	GCCGAAGCTTCTCCATATGGTCGCGGAACTCCATGCTTCTAGA GGCC
ABO-T8-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCATATGGAGAAGCTT CGGC
ABO-T9-F	GCCGAAGCTTCTCCACGTGGTCGCAGAACTCCATGCTTCTAGA GGCC
ABO-T9-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCACGTGGAGAAGCTT CGGC
ABO-T10-F	GCCGAAGCTTCTCCACGTGGTCGCAGATCTCCATGCTTCTAGA GGCC
ABO-T10-R	GGCCTCTAGAAGCATGGAGATCTGCGACCACGTGGAGAAGCTT CGGC
ABO-T11-F	GCCGAAGCTTCTCCACATGGTCGCAGAACTCCATGCTTCTAGA GGCC
ABO-T11-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCATGTGGAGAAGCTT CGGC
ABO-T12-F	GCCGAAGCTTCTCCATGTGGTCGCAGAACTCCATGCTTCTAGA GGCC
ABO-T12-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCACATGGAGAAGCTT CGGC
ABO-T13-F	GCCGAAGCTTCTCCACATGGTCGCAGATCTCCATGCTTCTAGA GGCC
ABO-T13-R	GGCCTCTAGAAGCATGGAGATCTGCGACCATGTGGAGAAGCTT CGGC
ABO-T14-F	GCCGAAGCTTCTCCATGTGGTCGCAGATCTCCATGCTTCTAGA GGCC
ABO-T14-R	GGCCTCTAGAAGCATGGAGATCTGCGACCACATGGAGAAGCTT CGGC
ABO-T15-F	GCCGAAGCTTCTCCATATGGTCGCAGAACTCCATGCTTCTAGA GGCC
ABO-T15-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCATATGGAGAAGCTT CGGC
ABO-T16-F	GCCGAAGCTTCTCCATATGGTCGCAGATCTCCATGCTTCTAGA GGCC
ABO-T16-R	GGCCTCTAGAAGCATGGAGATCTGCGACCATATGGAGAAGCTT CGGC
HIV-T1-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCCCTAT C TTC TAG AGG CC
HIV-T1-R	GGC CTC TAG AAG ATAGGGGGAATTGGAGGTTTTATCAAA A GAA GCT TCG GC
HIV-T2-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCCCTAT C TTC TAG AGG CC
HIV-T2-R	GGC CTC TAG AAG ATAGGGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC

HIV-T3-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCACTAT C TTC TAG AGG CC
HIV-T3-R	GGC CTC TAG AAG ATAGTGGGAATTGGAGGTTTTATCAAA A GAA GCT TCG GC
HIV-T4-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCCCTAC C TTC TAG AGG CC
HIV-T4-R	GGC CTC TAG AAG GTAGGGGGAATTGGAGGTTTTATCAAA A GAA GCT TCG GC
HIV-T5-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCACTAT C TTC TAG AGG CC
HIV-T5-R	GGC CTC TAG AAG ATAGTGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC
HIV-T6-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCCCTAC C TTC TAG AGG CC
HIV-T6-R	GGC CTC TAG AAG GTAGGGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC
HIV-T7-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCACTAC C TTC TAG AGG CC
HIV-T7-R	GGC CTC TAG AAG GTAGTGGGAATTGGAGGTTTTATCAAA A GAA GCT TCG GC
HIV-T8-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCACTAC C TTC TAG AGG CC
HIV-T8-R	GGC CTC TAG AAG GTAGTGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC
Negative-F	GCC GAA GCT TCT TTTGATTCTTGCTCTGCTCTCTTCGTC C TTC TAG AGG CC
Negative-R	GGC CTC TAG AAG GACGAAGAGAGCAGAGCAAGAATCAAA A GAA GCT TCG GC
pUC19_F	CAGCGAGTCAGTGAGCGA
pUC19_R	GCGACACGGAAATGTTGAATACTCAT
pUC19RPA_F	AGCGGATAACAATTTCACACAGGAAACAGC
pUC19RPA_R	TAACGCCAGGGTTTTCCCAGTCACGACGTT

Name	Sequences(5′→3′)
Oligos used in invitro high through-put assay experiments.	
HLA-library	/5′Phos/*TTGTGT*NNNNC*C*NG*G*G*C*C*T*T*G*T*A*G*A*T *C*T*G*T*G*T*G*T*G*NNNNACCTGCCGAG*TTGTGT*
ABO-library	/5′Phos/*AGAGAA*NNNNC*C*NC*G*T*G*G*T*C*G*C*G*G*A* A*C*T*C*C*A*T*G*NNNNACCTGCCGAG*AGAGAA*
S501-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***TAGATCGC***

S501-R	GCGATCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGG
S502-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***C****TCTCTAT*
S502-R	ATAGAGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGG
S503-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***TATCCTCT***
S503-R	AGAGGATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGG
S504-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***AGAGTAGA***
S504-R	TCTACTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT GTAGATCTCGGTGG
S505-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***GTAAGGAG***
S505-R	CTCCTTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT GTAGATCTCGGTGG
S506-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***ACTGCATA***
S506-R	TATGCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGG
S507-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***AAGGAGTA***
S507-R	TACTCCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT GTAGATCTCGGTGG
S508-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT***CTAAGCCT***
S508-R	AGGCTTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGG
HLA-N701	CAAGCAGAAGACGGCATACGAGAT***TCGCCTTA***ACCTGCC GAG*TTGTGT*
HLA-N702	CAAGCAGAAGACGGCATACGAGAT***CGTACTAG***ACCTGC CGAG*TTGTGT*
HLA-N703	CAAGCAGAAGACGGCATACGAGAT***TTCTGCCT***ACCTGCC GAG*TTGTGT*
ABO-N705	CAAGCAGAAGACGGCATACGAGAT***AGGAGTCC***ACCTGC CGAG*AGAGAA*
ABO-N706	CAAGCAGAAGACGGCATACGAGAT***CATGCCTA***ACCTGC CGAG*AGAGAA*
ABO-N707	CAAGCAGAAGACGGCATACGAGAT***GTAGAGAG***ACCTGC CGAG*AGAGAA*
ABO-N708	CAAGCAGAAGACGGCATACGAGAT***CCTCTCTG***ACCTGCC GAG*AGAGAA*
PE2_short	AAT GAT ACG GCG ACC ACC GA

HLA_sel_PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA G*TT GTG T*
ABO_sel_PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA G*AG AGA A*
Lib_adaptor1	GAC GGC ATA CGA GAT
HLA_lib_adaptor2	*TTG T*AT CTC GTA TGC CGT CTT CTG CTT G
ABO_lib_adaptor2	*AGA G*AT CTC GTA TGC CGT CTT CTG CTT G
lib_PCR_F	CAA GCA GAA GAC GGC ATA CGA GAT
HLA_PCR_R	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA G*TT GTG T*
ABO_PCR_R	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA G*AG AGA A*

NNNNNNN = Library barcode, ***NNNNNNNN*** = target barcode
An asterisks (*) indicates that the preceding nucleotide was incorporated as a hand mix of bases consisting of 79 mol % of the indicated base, and 7 mol % of each of the other three natural bases. “/5Phos/“ denotes a 5′ phosphate group added to the sequence.

Name	Sequences(5′→3′)
Oligos used to create DNA target constructs in cellular reporter assay.	
ABO-T1-REP-F	AATTCC CATGGAGTTCCGCGACCACGTGG AGGAG
ABO-T1-REP-R	GATCCTCCT CCACGTGGTCGCGGAACTCCATG GG
ABO-T2-REP-F	AATTCC CATGGAGATCCGCGACCACGTGG AGGAG
ABO-T2-REP-R	GATCCTCCT CCACGTGGTCGCGGATCTCCATG GG
ABO-T3-REP-F	AATTCC CATGGAGTTCCGCGACCATGTGG AGGAG
ABO-T3-REP-R	GATCCTCCT CCACATGGTCGCGGAACTCCATG GG
ABO-T4-REP-F	AATTCC CATGGAGTTCCGCGACCACATGG AGGAG
ABO-T4-REP-R	GATCCTCCT CCATGTGGTCGCGGAACTCCATG GG
ABO-T5-REP-F	AATTCC CATGGAGATCCGCGACCACATGG AGGAG
ABO-T5-REP-R	GATCCTCCT CCATGTGGTCGCGGATCTCCATG GG
ABO-T6-REP-F	AATTCC CATGGAGATCCGCGACCATATGG AGGAG
ABO-T6-REP-R	GATCCTCCT CCATATGGTCGCGGATCTCCATG GG
ABO-T7-REP-F	AATTCC CATGGAGATCCGCGACCATGTGG AGGAG
ABO-T7-REP-R	GATCCTCCT CCACATGGTCGCGGATCTCCATG GG
ABO-T8-REP-F	AATTCC CATGGAGTTCCGCGACCATATGG AGGAG
ABO-T8-REP-R	GATCCTCCT CCATATGGTCGCGGAACTCCATG GG

Name	Sequences(5′→3′)
Oligos used as ssDNA target in melting temperature assay experiments.	

ABO-T1-Tm	CCCACGTGGTCGCGGAACTCCATGT
ABO-T2-Tm	CCCACGTGGTCGCGGATCTCCATGT
ABO-T3-Tm	CCCACATGGTCGCGGAACTCCATGT
ABO-T4-Tm	CCCATGTGGTCGCGGAACTCCATGT
ABO-T5-Tm	CCCATGTGGTCGCGGATCTCCATGT
ABO-T6-Tm	CCCATATGGTCGCGGATCTCCATGT
ABO-T7-Tm	CCCACATGGTCGCGGATCTCCATGT
ABO-T8-Tm	CCCATATGGTCGCGGAACTCCATGT