Discovering the molecular basis of CRISPR-Cas13b for precise silencing of tumour transcripts

By Wenxin Hu
ORCID: 0000-0002-0330-5200

Supervisors: Dr Mohamed Fareh

Prof Joseph Trapani

A/Prof Ilia Voskoboinik

COMPREHENSIVE CANCER DOCTOR OF PHILOSOPHY

December 2022

Peter MacCallum Cancer Centre
and
The Sir Peter MacCallum Department of Oncology, The University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Comprehensive Cancer Doctor of Philosophy

Abstract

Cancers are caused by the accumulation of genomic mutations and gene expression dysregulations. In many patients, standard therapies such as chemotherapy often result in severe toxicity due to off-target effects that takes place in non-tumour tissues. Additionally, certain genetic aberrations confer resistance to these standard therapies and mediate cancer relapse. Recently, advanced sequencing technologies led to a better understanding of tumour genomics, which opened up a new era of precision and personalized medicine. These insights from tumour genomics enabled the identification of several hundreds of aberrant oncogenes called tumour drivers that confer clonal expansions and malignancy, which are considered high-priority targets of precision oncology.

Currently, there are two main precision medicine approaches to target tumour drivers. The first is adoptive cell transfer involving *in vitro* engineering patients' immune cells to recognize tumour antigens and lyse tumour cells, but this approach is restricted to tumour antigens on the cell. The second is the use of small inhibitory molecules and antibodies to target tumour drivers at a protein level. However, the drug development process using conventional pipelines is time-consuming, limiting the number of cancer drugs that are available to target newly identified tumour drivers in a personalized manner. These limitations highlight the need to develop a new design-flexible precision medicine approach to target tumour drivers in a personalized manner.

The recent breakthroughs in the field of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) offer a promising opportunity to precisely target tumour driver genes. Theoretically, the most extensively studied CRISPR-Cas9 can cleave any oncogene and induces its loss of function. Undoubtedly, it has been a powerful tool to study cancer functional genomics in basic research with immense potentials as modular cancer therapeutics. Nevertheless, safety concerns due to off-target effects and irreversible chromosomal instabilities largely constrain the therapeutic applications of classical CRISPR-Cas9. On the other hand, the discovery of a novel RNA-guided RNA targeting CRISPR-Cas13 may address some of the limitations of Cas9 to target tumour drivers at the RNA level.

In the introduction of this thesis (Chapter 1), I discuss the molecular basis of various CRISPR effectors with special focus on RNA-targeting enzymes. I also highlight how recent scientific discoveries in CRISPR filed translate into various tumour precision medicine applications. Finally, I chart progress in the CRISPR-Cas13 field and discuss its advantages and limitations as a programmable tool to target tumour drivers in a personalized manner.

In my Ph.D. research project, we hypothesized that the CRISPR-*Psp*Cas13b nucleoprotein complex may be re-engineered to silence tumour driver transcripts with single-nucleotide accuracy. However, the poor understanding of the molecular principles governing *Psp*Cas13b target recognition

and cleavage limits its utility and development as programmable cancer therapeutics. Thus, this research focuses on revealing the molecular mechanisms of *Psp*Cas13b in order to reprogram this ribonuclease to target tumour transcripts that are currently 'undruggable'.

Excitingly, our Single-Base Tiled crRNA screens (SiBTil), unbiased computational analysis, and comprehensive spacer-target mutagenesis revealed key determinants of *Psp*Cas13b activity. *De novo* design of crRNAs harbouring base-paired or mismatched guanosine bases at key spacer positions greatly enhances the silencing efficacy of otherwise inefficient crRNAs, expanding the targeting spectrum of this enzyme. We also reveal the interface between mismatch tolerance and intolerance, which unlocks an unexpected single-base precision targeting capability of this RNA nuclease. Notably, our *de novo* design principles enable potent and selective silencing of various gene fusion transcripts and their downstream oncogenic networks, without off-targeting of non-translocated variants that share extensive sequence homology. We demonstrate that *Psp*Cas13b targeting the breakpoint of fusion transcripts enables efficient suppression of ancestral and single-nucleotide mutants (e.g., BCR-ABL1 T315I) that often drive clinical cancer relapse. Our transcriptomic and proteomic analyses suggest *Psp*Cas13b is highly specific and has no global off-targeting or collateral activity against endogenous transcripts and proteins. Collectively, my Ph.D. research provides new design principles for *Psp*Cas13b programming to specifically recognise and degrade any '*undruggable'* fusion oncogenic transcript, thus providing a new conceptual framework for personalized oncology.

Declaration

This is to certify that:

i. This thesis comprises only my original work towards the PhD except where indicated.

ii. Due acknowledgment has been made in the text to all other material used.

iii. This thesis is less than 100,000 words in length, exclusive of tables, figures, bibliographies, and appendices.

Wenxin Hu, December 2022

Preface

Publications from part of this thesis.

I would like to acknowledge the contributions of the following people to the completion of my thesis. The research questions in this thesis were formulated together with my primary supervisor Dr Mohamed Fareh. The work in the thesis was completed with the co-supervision of Professor Joseph Trapani and Associate Professor Ilia Voskoboinik. The contributions from others are acknowledged below.

Chapter 2, Chapter 3 and Chapter 4 include an author in-revision manuscript at *Nature Communications* (preprint available at *BioRxiv*).

Wenxin Hu, Amit Kumar, Shijiao Qi, Teresa Sadras, Joshua ML Casan, David Ma, Gurjeet Jagjeet Singh, Lauren M Brown, Michelle Haber, Ilia Voskoboinik, Joseph A Trapani, Paul G Ekert, Mohamed Fareh (2022). Single-base precision design of CRISPR-Cas13b enables the systematic silencing of oncogenic fusions. *bioRxiv* 2022.06.22.497105; doi: https://doi.org/10.1101/2022.06.22.497105

Author	**Proportion**	**Contributions**
Wenxin Hu	60%	Study design; performed the experiments; Data collection, analysis and interpretation; manuscript writing and revision
Amit Kumar	4%	Creating tools and bioinformatics analysis
Shijiao Qi	3%	Bioinformatics analysis
Teresa Sadras	2%	Study design and creating tools: cloned Imatinib-resistant BCR-ABL1 gene fusion.
Joshua ML Casan	1.5%	Manuscript revision
David Ma	1.5%	Creating tools and bioinformatics analysis
Gurjeet Jagjeet Singh	0.5%	Creating tools
Lauren M Brown	0.5%	Creating tools and manuscript revision

Michelle Haber	1%	Funding support
Ilia Voskoboinik	3%	Study design and manuscript revision
Joseph A Trapani	4%	Supervised the study; Study design and manuscript revision
Paul G Ekert	4%	Supervised the study; Study design and manuscript revision
Mohamed Fareh	15%	Conceived the study; supervised the study; Study design; performed the experiments; Data collection, analysis and interpretation; manuscript writing and revision

Chapter 4.4.2: Discoveries made in this chapter are based on preliminary findings of a follow-up project of Chapter 2, Chapter 3, and Chapter 4.3.1.

This work was supported by a Cancer Council Victoria Ventures grant (ID 829606), a Peter MacCallum Cancer centre strategic plan funding in partnership with the Children's Cancer Institute Australia (CCIA), a Peter MacCallum Foundation grant (ID 2119), and a National Health and Medical Research Council (NHMRC) of Australia through a program.

Acknowledgments

A long, tough trip−filled with excitement, uncertainty, discovery, and reflection. I wish to acknowledge all those who helped me complete the journey.

First and foremost, I would like to thank my supervisor Dr Mohamed Fareh, who has formulated the research questions for my PhD, and guided me in driving the project forward. Exchanging thoughts with his always makes me inspired and motivated. His expertise, knowledge, and insight have been invaluable, and his encouragement and constructive feedback have helped me to push through challenging times. Thank you for your patience, dedication, and unwavering support. Second, I want to acknowledge my secondary supervisors, Prof Joseph Trapani, A/Prof Ilia Voskoboinik and A/Prof Paul Ekert who have provided invaluable advice for my project. It is a great honour for me to study under their supervision.

I would also like to thank my mentor A/Prof Louise Cheng and the members of my dissertation committee, Prof Rob Ramsay, A/Prof Paul Beavis, Dr Stephen Luen, Dr Viv Sutton, Dr Dane Newman, Dr Jason Li, and Dr Criselle DSouza, for their invaluable feedback and suggestions, which have improved the quality of my research. Their expertise, knowledge, and insights have been instrumental in shaping my research and expanding my understanding of the subject matter.

I would like to acknowledge the support from the Flow Cytometry Facility and the Molecular Genomics Core at Peter MacCallum Cancer Centre and Mass Spectrometry and Proteomics Facility at the Bio21 Institute, University of Melbourne.

I am deeply grateful to the Petermac Research Education Admin, and especially Dr Erika Cretney, for providing an enriching and stimulating academic environment. Their support, encouragement, and resources have been instrumental in my success.

I am also thankful for my friends and colleagues, Dr Joshua Casan, Dr Carolyn Shembrey Gurjeet JagjeetSingh, Sandra Verschoor, Annette Ciccone, Dr Teresa Sadras and Fatimah Jalud for their unwavering support, encouragement, and laughter during my PhD journey. Their shared experiences and sense of humor have made this journey more enjoyable and fulfilling.

I would like to express my appreciation to the University of Melbourne and Peter MacCallum Cancer Centre for providing financial support for my research. This support allowed me to pursue my research interests and allowed me to focus on my studies.

I am also grateful to my parents for their constant love and support throughout my academic journey. Their unwavering encouragement and understanding have been a source of inspiration and motivation during difficult times of Covid-19 pandemic.

In conclusion, I express my heartfelt gratitude to all those who have contributed to my academic growth and success. I am humbled by your support and encouragement, and I hope to pay it forward in the future.

Abbreviations

ACTB/β-actin	beta-actin
AcrVIA	anti-CRISPR type VI-A
ADAR2-DD	RNA type 2 deaminase domain
ALKBH5	AlkB homolog 5
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukaemia
A.U.	arbitrary units
Cas	CRISPR-associated protein
CIRTS	CRISPR-Cas-inspired RNA targeting system
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CRY2PHR	cryptochrome circadian regulator
CIB1	calcium-and integrin-binding 1
CYB5A	cytochrome b5 form A
DOX	Doxycycline
DNA	deoxyribonucleic acid
DR	direct repeat
DSB	double-strand break
dsDNA	double-strand DNA
eGFP	enhanced green fluorescent protein
EGFR	epidermal growth factor receptor
FC	fold change
FISH	fluorescence *in situ* hybridization
FTO	obesity-associated protein
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
CARMEN	combinatorial arrayed reactions for multiplexed evaluation of nucleic acids
CLIP-seq	cross-linking immunoprecipitation sequencing
Gluc	glucose
cryo-EM	cryogenic electron microscopy
COVID-19	Coronavirus Disease 2019

CRUIS	CRISPR-based RNA-United Interacting System
DMPK	dystrophia myotonic protein kinase
HEDGES	high-level extended duration gene expression system
HEK293T/293T	human embryonic kidney cells with SV40 large T antigen
HEPN	higher eukaryotes and prokaryotes nucleotide-binding
HIV-1	Human immunodeficiency virus 1
HTH	helix-turn-helix domain
HUDSON	heating unextracted diagnostic samples to obliterate nucleases
IGF2BP2	insulin-Like Growth Factor 2 MRNA Binding Protein 2
lncRNA	long non-coding RNA
IAV	influenza A virus
IVT	*in-vitro* transcription
IDL	interdomain linker
KRAB	Krüppel-associated box
LASV	Lassa virus
LiveFISH	live-cell fluorescent in situ hybridization
m^6A	N^6-methyladenosine
mES	mouse embroynic stem cells
MEF	mouse embryonic fibroblasts
MFE	minimal free energy
miRNA	micro-RNA
mRNA	messenger RNA
ncRNA	non-coding RNA
NES	nuclear export signal
NHEJ	error-prone non-homologues end joining
NLS	nuclear localization signal
NT	non-targeting
nt	nucleotide
NTD	N-terminal domain
NUC	nuclease lobe
NS3	non-structural protein 3
PafA	proteasomal accessory factor A

PAM	protospacer adjacent motif
PAC-MAN	prophylactic antiviral CRISPR in human cells
Pcsk9	proprotein convertase subtilisin/kexin type 9
PFS	protospacer flanking sequence
PKR	PKR
Pten	phosphatase and tensin homolog
Ptbp1	polypyrimidine tract-binding protein 1
PWM	position weight matrix
poly-A	polyadenylation
RBP	RNA-binding proteins
REC	recognition lobe
REPAIR	RNA-editing nucleoprotein complex RNA Editing for programmable A to I replacement
RESCUE	RNA Editing for specific C to U exchange
RGC	retinal ganglion cells retinal ganglion cells
RPA	recombinase polymerase amplification
RIP-seq	RNA immunoprecipitation sequencing
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
RuvC	RuvC-like nuclease domain
RT-PCR	reverse transcription polymerase chain reaction
SARS-Cov-2	severe acute respiratory syndrome coronavirus 2
SHERLOCK	specific high sensitivity enzymatic reporter unlocking
shRNA	short hairpin RNA
sgRNA	single-guide RNA
SiBTil	Single-Base Tiled crRNA screens
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
snoRNA	small nucleolar RNAs
ssRNA	single-stranded RNA
TH	transmembrane helix

TRM	targeted RNA methylation
UTR	untranslated region
vlincRNA	very long non-coding RNA
WT	wildtype
WYL	WYL domain
YTHDF	YTH domain-containing family protein

Table of contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 1

Chapter 4

Appendix

Chapter 1

Literature review

Abstract

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and CRISPR-associated (Cas) effector proteins can be used as tools that precisely manipulate eukaryotic gene expression. Recent breakthrough discoveries have identified a novel single-effector Class 2 type VI CRISPR family, termed CRISPR-Cas13, that targets RNA instead of DNA. To date, six VI subtypes (A-D, X, and Y) have been described as programmable RNA-targeting endonucleases, which share limited sequence homology. Cas13 enzymes possess two enzymatically distinct RNase domains and form a nucleoprotein complex with a CRISPR RNA (crRNA) capable of degrading single-stranded RNA (ssRNA) with high efficacy and specificity. More recent studies have further engineered Cas13 as a versatile RNA toolbox for precise RNA silencing, RNA editing, nucleic acid detection, and live-cell RNA imaging. Importantly, compared to DNA-targeting CRISPR effectors, Cas13 offers a promising opportunity to manipulate transcriptomes precisely and efficiently without altering DNA genomic sequences. In fact, the off-targeting and permanent chromatin rearrangements that are often associated with CRISPR tools that target DNA limit their broad translation (e.g., in patient DNA editing), and highlights potential applications of Cas13 tools that transiently target RNA, and which may provide higher specificity and superior safety characteristics. In this perspective, current knowledge of the molecular basis of Cas13 crRNA processing, target RNA recognition, and cleavage is summarised, together with new applications for RNA manipulation. The potential uses of CRIPSR-Cas 13, and its limitations in basic research, biomedicine, and beyond are also discussed.

1.1 Introduction

In 1987, Ishino and his colleagues first described what would later be termed Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and CRISPR-associated (Cas) effector proteins when they accidentally cloned a CRISPR sequence with their target gene isozyme conversion of alkaline phosphatase (IAP) in *E. coli*[1]. They found that the repeated sequences (later termed 'direct repeats') were arranged consecutively[1]. Similar sequence patterns were later discovered by independent research groups in a range of bacterial and archaeal species including *M. tuberculosis*, *Halferax*, and *Haloarula* and researchers recognized the diversity of sequences (later termed 'spacers') that interrupt direct repeats among different strains[2]. In 2001, Mojica and Jansen proposed the common term CRISPR to avoid confusion from numerous synonyms describing various repeats in the literature[3]. In parallel, a group of homologous genes encoding helicase and nuclease domains were observed to accompany the CRISPR sequences and were named CRISPR-associated (Cas) proteins[4]. In 2005, three independent research groups demonstrated that the spacer sequences were derived from the DNA sequence of viruses that had previously infected a given bacterium, which suggested a possible role for the CRISPR-Cas

system as nucleic acid-based adaptive immunity in prokaryotes that would silence invading bacteriophages and plasmids[5-7]. This hypothesis was experimentally validated in 2007, and since then the molecular mechanisms of CRISPR-Cas systems have been unravelled in great detail[8]. In 2012, a groundbreaking paper by Doudna, Charpentier, and their colleagues demonstrated how they reprogrammed the CRISPR enzyme Cas9 using a dual tracrRNA:crRNA system to target and cleave specific DNA sequences. This innovation showcased the potential of Cas9 as a programmable RNA-guided genome editing tool, sparking a transformative impact on biological research, biotechnology, and gene therapy. In recognition of its profound significance, this discovery was awarded the Nobel Prize of Chemistry in 2020[9].

The discovery of CRISPR effectors offers a great opportunity to achieve sequence-specific targeting of any gene of interest at the DNA or RNA level[10,11]. CRISPR guide RNAs (crRNAs) and Cas proteins are found in approximately 50% of bacteria and nearly all archaea and confer upon these unicellular organisms a nucleic acid-based form of adaptive immunity to neutralise invading viruses[8,12]. Overall, these effectors are organized into an array of DNA sequences that contain multiple copies of alternating semi-palindromic repeat sequences and spacer sequences, the latter derived from exogenous viral genetic elements. The process of incorporating new small exogenous genetic fragments into the 5' end of the parental CRISPR array by recruiting Cas1 and Cas2 encoded in the CRISPR-Cas loci is called spacer acquisition[13]. The CRISPR array is transcribed into pre-crRNA molecules that are further processed into individual mature crRNAs containing a repeat and a spacer RNA motif. crRNAs are loaded into Cas effectors and guide the process of target recognition through Watson-Crick RNA-DNA or RNA-RNA base-pairing (depending on the CRISPR-Cas type). The base pairing with the target triggers the activation of Cas protein nuclease domains to cleave target nucleic acids in a sequence-specific manner[11].

So far, over 130 subtypes of CRISPR-Cas have been discovered which are classified into two main classes. Class 1 systems comprise types I, III, and IV and utilize an assembled protein complex with multiple Cas proteins whose complexity makes it challenging to re-engineer as programmable genome editing tools. In contrast, Class 2 systems (types II, V, and VI) use a single Cas effector protein[8,12]. The Class 2 system has attracted intensive investigations due to its design simplicity and facile reprogramming thanks to its single enzyme guided by a single crRNA[14]. Among class 2 CRISPR enzymes, a type II enzyme called Cas9 that targets double-stranded DNA (dsDNA) is the most studied and widely used for targeted genome editing in various organisms. This nucleoprotein contains a Cas9 nuclease protein, and two RNA molecules called CRISPR RNA (crRNA) and tracrRNA[15]. Researchers further engineered the two RNA molecules into a single-guide RNA (sgRNA) which contains a 20-nucleotide (nt) spacer sequence complementary to the target DNA sequence. Target recognition and cleavage by Cas9 require that the targeted sequence is flanked by a 3-nucleotide long sequence called protospacer adjacent motif (PAM)[15]. For example, *Streptococcus pyogenes* Cas9 (*Sp*Cas9) requires a

downstream 'NGG' sequence to cleave the target DNA sequence[16]. PAMs are believed to exclusively exist in the invading viral genetic material, which enables bacteria to distinguish self from nonself target sequences, thereby preventing self-cleavage events[17]. The spacer sequence of sgRNA hybridizes with one strand of the dsDNA to form an R-loop sequence. The basepairing between the spacer and target DNA initiates conformational changes that activate two Cas9 nuclease domains –a McrA-like HNH (HNH) and a RuvC-like (RuvC) nuclease domains – which create a double-strand break (DSB) at the targeted site. Following this DNA DSB, a host error-prone non-homologues end joining (NHEJ) DNA repair machinery is recruited to repair the DNA damage at this specific genomic location. While repairing the damage, NHEJ often introduces indels, creating frameshifts and/or stop codons that lead to gene inactivation[9,10]. In 2013, *Sp*Cas9 was first reprogrammed for precise gene knockout in mammalian cells by Zhang's Lab and Church's Lab, which opened a new era of CRISPR-based genome editing, revolutionizing classical molecular biology in various biological fields[18]. These two back-to-back studies developed *Sp*Cas9-mediated genome editing by incorporating two nuclear localization signals (NLSs) for efficient nuclear localization, optimizing a full-length fusion sgRNA with 3' end hairpin, and targeting multiple genes simultaneously by CRISPR arrays with cognate spacers[19,20]. Beyond genome editing based on basic Cas9 nuclease activity, other recent breakthroughs in CRISPR-Cas-derived genome editing tools include the development of base editors, prime editors, and transposases/recombinases that have greatly expanded the scope of targeted genomic DNA modifications that can be achieved (for details refer to this excellent review[21]). However, DNA editing still faces obstacles such as targeting site restrictions due to the requirement of PAM sequences, and safety and ethical concerns about permanent genomic alterations in somatic and germline cells during clinical applications[21]. Therefore, there remains a need for more highly specific and efficient RNA-targeting CRISPR systems.

In 2015, computational data mining of microbial metagenomic databases for previously uncharacterized Class 2 CRISPR-Cas loci led to the identification of a new single-effector CRISPR-Cas, which further enlarged the Class 2 family[22]. This novel Cas protein was first named C2c2 (later known as Cas13)[23,24]. In contrast to Cas9 which has two DNase domains, Cas13 possesses two 'higher eukaryote or prokaryote nucleotide-binding domains' (HEPN) that act as RNase domains and exclusively cleave single-stranded RNA (ssRNA)[23]. Following the discovery of the first Cas13 orthologue called Cas13a, further computational analyses were conducted to identify other RNA-targeting CRISPR-Cas13 orthologues. By 2021, five further Cas13 effectors (Cas13b, Cas13c, Cas13d, Cas13x, and Cas13y) were discovered and designated as Class 2 type VI (A-D, X, and Y)[25-28]. Recent studies have further investigated the structure of some Cas13 orthologues, the molecular structure of the Cas13-crRNA nucleoprotein complex, and engineered these programmable RNA nucleases for various applications[29-32].

1.2 Diversity of CRISPR-Cas13 systems

To date, six subtypes of CRISPR type VI (Cas13a, Cas13b, Cas13c, Cas13d, Cas13x, and Cas13y) have been discovered and characterized, described briefly below[25-28] (**Figure 1.1**).

1.2.1 Cas13a

CRISPR-Cas13a was first-discovered and the most studied type VI sub-type[33]. An early study suggested Cas13a effectors have two distinct clades defined by their amino acid homology, alpha and beta. However, some Cas13a effectors fall outside these two clades and have ambiguous sequence features, indicating a need for clearer classification for type VI-A subtype[34]. A later study found that the crRNA DRs exist in two clusters (clusters 1 and 2) for type VI-A. The classification by crRNA DRs overlaps but distinguishes from the clade classification based on Cas13a effectors. The cluster 1 crRNAs-directed Cas13a effectors preferentially cleave at uridine during *trans*-cleavage (also known as collateral activity, discussed below in **Chapter 1.3.7**) while the cluster 2 CRISPR-Cas13a effectors show an adenosine preference. Interestingly, clusters 1 and 2 have markedly different degrees of promiscuity regarding the loading of a non-cognate Cas13a crRNAs. crRNAs of cluster 1 are broadly exchangeable between different Cas13a orthologues[35]. For example, *Lwa*Ca13a and *Lbu*Cas13a both can be directed to cleave target RNAs by the non-cognate crRNA of *Ppr*Cas13a[35]. In comparison, cluster 2 members have much narrower flexibility and exhibit very limited activity when reprogrammed with a non-cognate crRNA[35].

1.2.2 Cas13b

Further computational phylogenetics studies revealed a group of type VI effectors with distinct features from Cas13a, named Cas13b. The identity of accessory proteins located in type VI-B CRISPR loci indicates two distinct branches in the phylogenetic tree of Cas13b[26]. Cas13b effectors can be further classified into two categories: VI-B1 with accessory protein Csx27 in CRISPR loci (e.g., *Bergeyella zoohelcum* Cas13b or *Bz*Cas13b) and VI-B2 with Csx28 loci (e.g., *Prevotella buccae* Cas13b or *Pbu*Cas13b) (**Figure 1.1**). Csx27 represses, while Csx28 enhances the cleavage activity of Cas13b[26]. Repression of VI-B1 Cas13b activity by Csx27 might be an important regulatory mechanism of phage interference. On the other hand, Csx28 might enhance the collateral activity of VI-B2 Cas13b to inactivate viral RNAs or to promote bacterial dormancy. Csx27 and Csx28 might directly interact with Cas13b effectors based on their predicted structure[26]. However, the molecular basis that determines the mode of action of these accessory proteins remains largely unknown and highlights the need for further structural and biochemical investigations[26,33]. In August 2021, an additional group of ultracompact Cas13b subtypes called Cas13bt were discovered, which are characterized by their small sizes (775-804 amino acids (aa)) and the absence of accessory proteins[36] (**Figure 1.1**). Cas13bt has a similar size to previously characterized compact Cas13 effectors including Cas13d (average size 928 aa), Cas13x (775-805 aa), and Cas13y (790-803 aa)[27,28,36]. These genes encoding ultra-compact proteins are

compatible with adeno-associated virus (AAV) packaging for *in vivo* delivery compared to their orthologues Cas13a (~ 1250 aa) and Cas13b (~1150 aa)[24,26,37].

1.2.3 Cas13d

This was the first discovered compact type VI effector, with a polypeptide chain shorter than 1000 aa[27,37]. Similar to type VI-B2, most of the VI-D effectors have a small accessory protein termed WYL domain-containing protein adjacent to the CRISPR loci in the host genome, believed to enhance target RNA cleavage (**Figure 1.1**). A structural study revealed that the three-domain architecture (NTD, WTL, and CTD domains) of *Ruminococcus* sp. WYL1 (*Rsp*WYL1) increases the binding affinity of Cas13d to ssRNAs. The revealed structural feature of the WYL domain indicated its ability to bind target ssRNA or crRNA. *Rsp*WYL1 may facilitate the cleavage activity of *Rsp*Cas13d by enhancing the interaction between Cas13d and the target RNAs. Alternatively, *Rsp*WYL1 may improve the accessibility of target RNAs to Cas13d by increasing the surrounding ssRNA concentration[38].

1.2.4 Cas13x/y

In May 2021, computational analysis of genomic data from uncharacterized environmental microbes identified the latest subtypes, X and -Y. Seven candidates were found and classified into VI-X (Cas13x.1 and Cas13x.2) and VI-Y (Cas13y.1 to Cas13y.5)[28] (**Figure 1.1**). To date, structural and mechanistic studies on Cas13x and y are very limited. Notably, phylogenetic analysis of Cas13x, y and Cas13bt shows extensive protein sequence homology (more than 60%) and crRNA similarity, particularly between Cas13x.1, and Cas13bt.3, where the two proteins exhibit 99.48% protein sequence identity, with only 2-aa difference. The very high sequence similarity raises the possibility that the minimal differences reflect polymorphisms encoding the one functional protein, which may require reconsideration of their nomenclature by the CRISPR community.

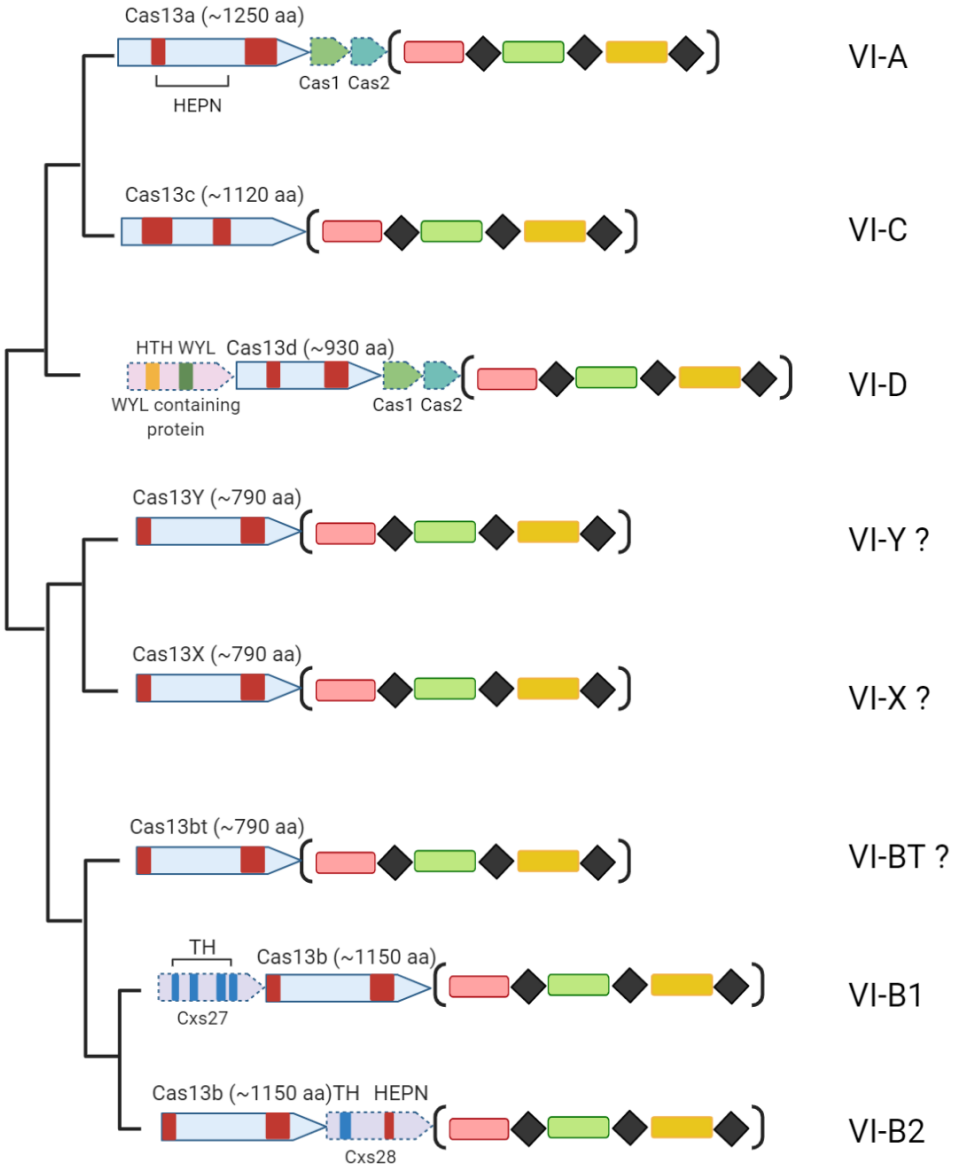


Figure 1. 1. Diversity and classification of type VI CRISPR-Cas subtypes.

The phylogenetic tree illustrating the evolutionary relationships between known Type VI genomic loci (branch lengths and evolutionary distances are not indicated, and genes size is not shown in scale). The classification of VI-X, Y and BT may require further investigation due to the extensive sequence identity (indicated by question marks). The type VI-C CRISPR loci remain largely uncharacterized. Cas13 gene cartoons (light blue) are labelled with gene names and the approximate protein size is indicated. The approximate locations of the two HEPN domains (in red) within each Cas effector gene are indicated. Dash outlined genes (*Cas1*, *Cas2 Csx27* and *Csx28*) suggest that these genes are absent in some of the indicated CRISPR subtype loci. Within each CRISPR array (in brackets), black diamonds represent the conserved direct repeat sequences, while red, green and yellow rectangles represent distinctive spacer sequences. Abbreviations:

HEPN, Higher Eukaryotes and Prokaryotes Nucleotide-binding domain; HTH, helix-turn-helix domain; WYL, WYL domain; TH, putative transmembrane helices. Adapted from[25,28,33].

1.3 RNA-guided RNA targeting by CRISPR-Cas13

Recent studies have offered a basic understanding of CRISPR-Cas13 structures and functions. Briefly, the CRISPR array of the Cas13 locus is the DNA sequence that contains multiple copies of alternating semi-palindromic repeat sequences and spacer sequences that are derived from exogenous genetic materials[39]. The CRISPR array is first transcribed into pre-CRISPR RNA (crRNA) molecules that are further processed into individual mature crRNAs containing a single direct repeat (DR) and a spacer RNA motif. crRNAs are loaded into Cas13 effectors and guide the process of target recognition through Watson-Crick RNA-RNA base pairing[23]. Base pairing with the target triggers the activation of Cas13 HEPN nuclease domains to *cis*-cleave target RNA in a sequence-specific manner[40]. In addition, some Cas13 orthologues have been shown to possess a *trans*-cleavage (known as collateral) activity which can degrade neighbouring ssRNAs *in vitro* and in bacteria[25,33] (**Figure 1.2**).

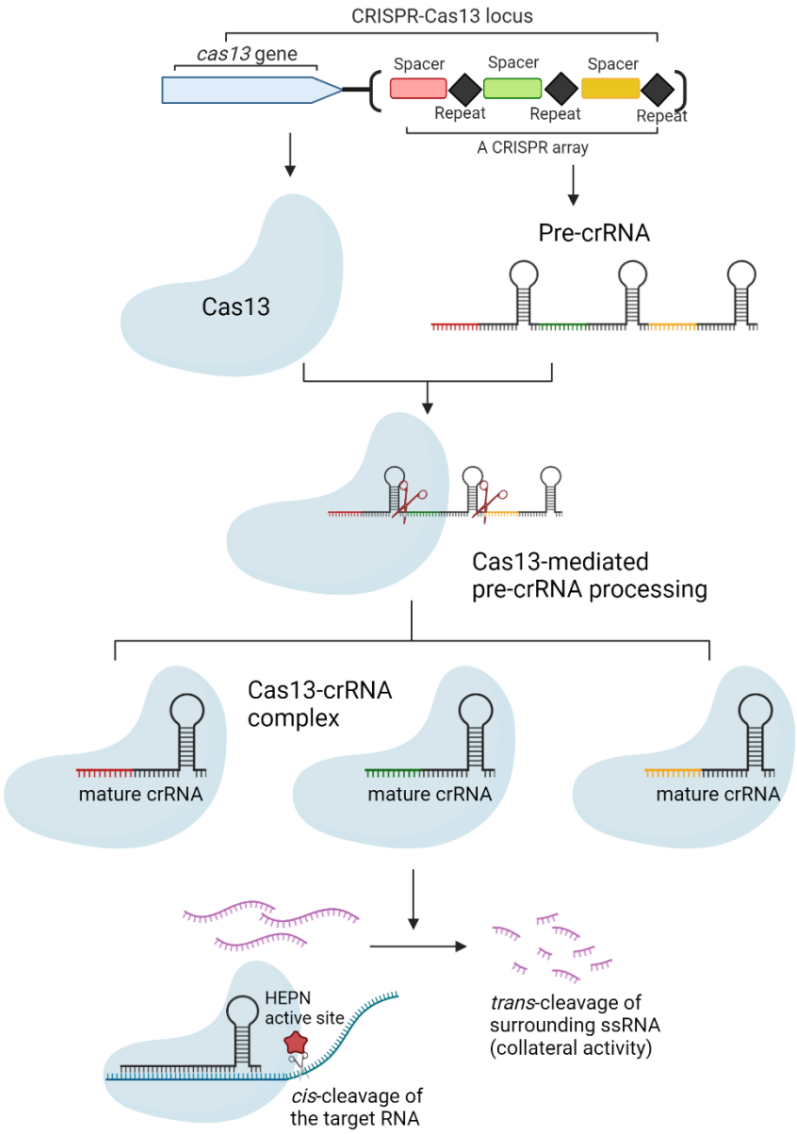


Figure 1. 2. An overview of RNA-dependent RNA targeting by CRISPR-Cas13.

When a phage infects a bacterium, short sequences from the phage's genetic material are integrated into the CRISPR array, where they become new spacers. Within the CRISPR array, two successive spacers are separated by a conserved direct repeat (DR) sequence. Several spacer and DR sequences are co-transcribed into a single pre-crRNA which is further cleaved into multiple mature crRNAs each contains a single DR and a unique spacer sequence. The Cas13 protein recognises the DR to initiate the loading process of crRNA, which forms a functional CRISPR-Cas13 nucleoprotein complex. When a new infection by the same phage occurs, the spacer sequence directs the Cas13 protein to recognize viral RNAs through spacer-target RNA-RNA base pairing,

which mediates Cas13 conformational change, resulting in the activation of HEPN nuclease domains and the cleavage of the target viral ssRNA (cis cleavage). Some Cas13 orthologues are shown to possess an additional trans cleavage mechanism which degrade the surrounding ssRNA in a nonspecific manner (from both viral and host origin). This trans cleavage is believed to mediate host dormancy and limit further viral replication.

1.3.1 Structural arrangement of Cas13 effectors

Unlike other CRISPR sub-types, many type VI loci possess a CRISPR array and a single Cas13 effector but lack *cas1* and *cas2* genes which typically play important roles in spacer acquisition process[24,26,27] (**Figure 1.1**). For example, the VI-B CRISPR arrays contain spacers distinct from closely related bacterial strains, suggesting that spacer acquisition does occur in VI-B loci. This process might be spontaneous or by recruiting Cas1 and Cas2 encoded in other CRISPR-Cas loci from the same genome[26]. All Cas13 proteins are characterized by two HEPN domains (amino acid motif *RXXXXH*) and form a nucleoprotein complex with a single crRNA[24,26,27]. Cas13 proteins have two functionally independent RNase domains. The first RNase is responsible for pre-crRNA cleavage to generate mature crRNAs. This RNA processing activity is not observed in CRISPR-Cas9[41]. For instance, *Sp*Cas9 pre-crRNA cleavage is mediated by a tracrRNA molecule and an independent housekeeping RNase III. The tracrRNA binds to the complementary repeat sequences in the pre-crRNA and forms an RNA duplex, which is recognized and cleaved by RNase III[42]. The second RNase activity of the two Cas13 HEPN domains allows cleavage of the target RNA upon spacer-target base-pairing[23,25] (**Figure 1.2**). Structural data have revealed that Cas13 proteins adopt a bi-lobed structure containing one crRNA recognition lobe (REC) and one nuclease lobe (NUC), located at various positions of the linear proteins[23,43-45] (**Figure 1.3**). Interestingly, although CRISPR-Cas13 subtypes have extensive functional similarities, Cas13 proteins exhibit low sequence identity (11%-16%) and have different structural conformations and features[25].

In Cas13a proteins, the REC lobe contains the N-terminal domain (NTD) and Helical-1 domain while the NUC lobe comprises the HEPN-1 domain, HEPN-2 domain, Helical-2 domain, and Linker domain (also known as Helical-3 domain). The Helical-2 domain separates the HEPN-1 domain into two subdomains (HEPN-1 I and II). The HEPN-1 and HEPN-2 domains connected by the Linker domain formed the outer surface groove where the two RXXXXH nuclease motifs localize[46-48]. The ternary structure of Cas13d is less resolved due to its compact size but still maintains the bi-lobed feature like other type VI effectors. The domain organization of REC and NUC lobes in Cas13d is similar to Cas13a except for a largely shortened Helical-1 domain. The HEPN-1 domain function as a bridge between the two lobes[45,49]. In contrast, the Cas13b effectors domain arrangement is significantly distinct

from other type VI effectors. The HEPN-1 and 2 domains locate at the extreme N and C protein termini (**Figure 1.3**), and the two RXXXXH motifs are relatively close to each other after protein folding[44,50,51]. Apart from the two HEPN domains, the Cas13b subtypes described so far (*Bz*Cas13b and *Pbu*Cas13b) have different names of other domains due to different characterisation approaches by independent studies, Helical-1, Helical-2, RRI-1, RRP-2, and LR in *Bz*Cas13b, and interdomain linker (IDL), Lid, Helical-1, Helical-2 in *Pbu*Cas13b[44,51] (**Figure 1.3**). Both *Bz*Cas13b and *Pbu*Cas13b fold into a pyramidal structure with a large positively charged channel located at the bottom[44,51].

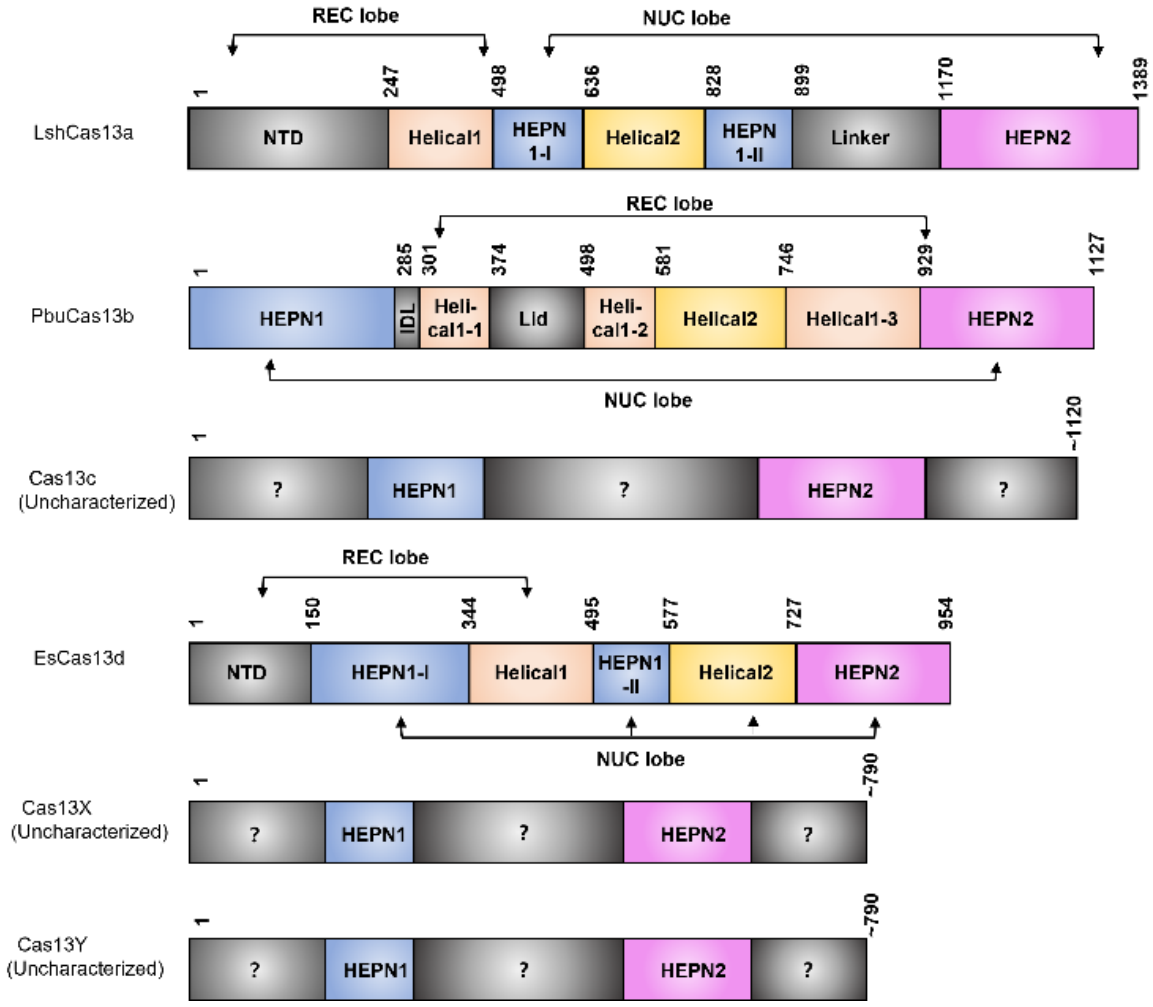


Figure 1. 3. Linear structural domains of type VI CRISPR-Cas effectors and their interactions.

*Lsh*Cas13a, *Pbu*Cas13b and *Es*Cas13d are used as examples. The architectures of the type VI-C, X and Y effectors are incomplete as they remain uncharacterized. The illustrations are approximate to scale, and REC and NUC lobes are annotated. conserved domains are highlighted with the same colour code unless otherwise stated. Abbreviations: NTD, N-terminal domain; HEPN, Higher

Eukaryotes and Prokaryotes Nucleotide-binding domain; IDL, interdomain linker; REC, crRNA recognition lobe; nuclease lobe

1.3.2 Cas13 crRNAs

The crRNAs of type VI family members are simpler than those in type II. Instead of a crRNA:tracrRNA duplex, type VI crRNA is a single RNA molecule containing a DR sequence and a spacer sequence[22,50,52]. The DR sequence forms a highly ordered stem-loop structure that facilitates the loading of crRNA to Cas13 by interactions between conserved RNA bases and specific amino acid residues within Cas13 proteins. Notably, the DR of the Cas13a and Cas13d crRNA forms a stem-loop containing a 2nt-bulge at the stem base and with a 5' and 3' flanking sequence up to 5nt[46-48]. Additionally, the DR of Cas13b and Cas13x crRNA is at the 3' end, while Cas13a, c and d have their DRs at the opposite 5' end[28,44]. The DR stems of Cas13b, Cas13x, and Cas13y are much longer than other subtypes and have multiple unpaired regions, resulting in an additional internal loop[28,44,45,48]. (**Figure 1.4**). The DR lengths of Cas13a crRNAs are heterogeneous, varying from 28 to 36nt. In contrast, the DR lengths of VI-B (36nt) and -D subtypes (30nt) are conserved within each subtype. On the other hand, the length of the spacer which confers the recognition and binding of Cas13 to the target RNA[24,26,43,53], varies between different subtypes but are all longer than the ~20-nt spacer in type II family members, providing better targeting specificity than Cas9 or the classical eukaryotic RNAi [10,25,35,39].

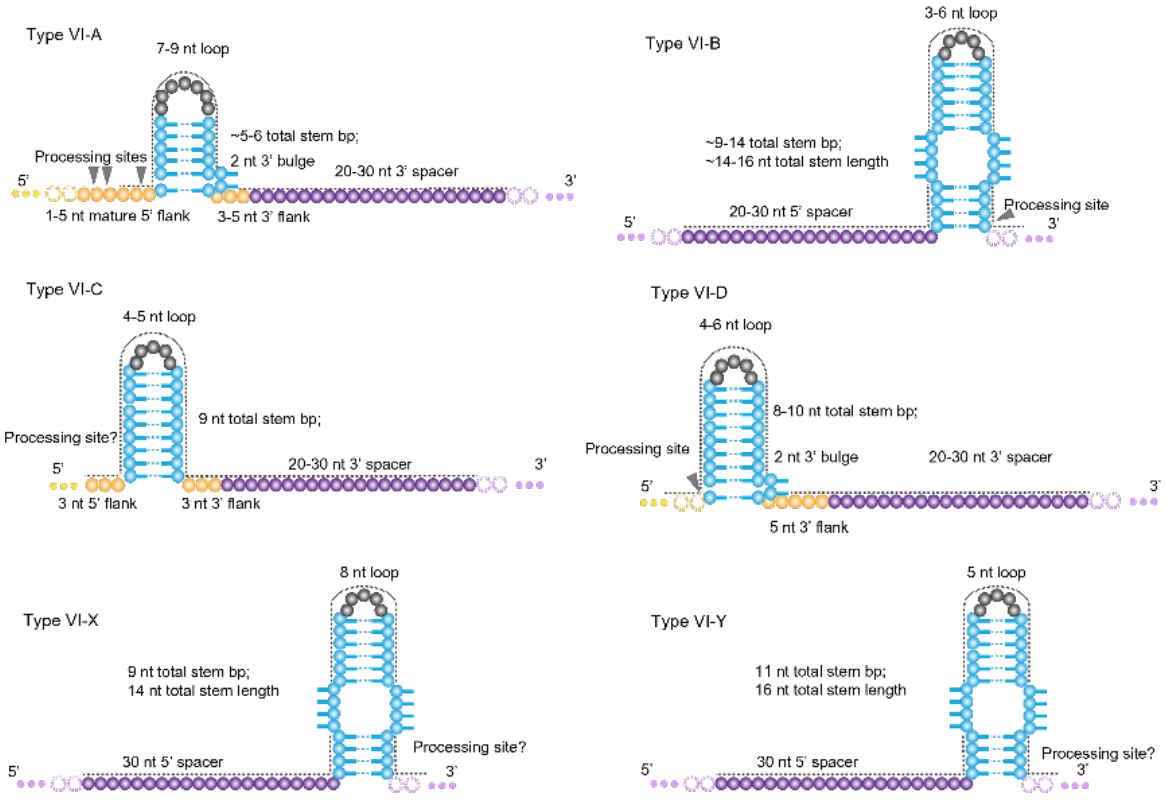


Figure 1. 4. Structural characteristics of Type VI crRNAs.

Major similarities and differences between crRNAs from each subtype are illustrated in the schematics. Spacer length is variable among orthologues and can further vary when expressed in *E. coli* due to trimming by host nucleases with the ranges shown in the schematics. Stems in the type VI-A and D systems contain unpaired bases and bulges. Accurate crRNA processing sites are indicated by arrows. The processing sites of types VI-C, X and Y remain unknown. (Adapted from[25]).

1.3.3 Cas13 and crRNA interaction

Several cryogenic electron microscopy (cryo-EM) studies have solved the structures of *Leptotrichia buccalis Cas13a* (*Lbu*Cas13a), *Lachnospiraceae bacterium* Cas13a (*Lba*Cas13a), *Leptotrichia shahii* Cas13a (*Lsh*Cas13a), *Bz*Cas13b, *Pbu*Cas13b, *Eubacterium siraeum* Cas13d (*Es*Cas13d) and *Ruminococcus sp*. Cas13d (*Ur*Cas13d), which gives insights into how Cas13 effectors interact with their crRNAs. In the presence of crRNA, Cas13 effectors form a "surveillance complex", poised for searching and identifying complementary target RNAs[27,44,47-49,51,54].

Generally, in sub-type VI-A, the binding of the cognate crRNA is mainly stabilized by non-covalent contacts between evolutionarily conserved bases in the DR and a strong negatively charged cleft formed by an NTD and the Helical-I domain in REC lobe[23,47,48] (for review, see Ref[25]). Cas13a

undergoes significant conformational changes to form a more compact structure to stabilize the binding of crRNA, close the crRNA-binding cleft and facilitate target RNA binding (**Figure 1.5a**). Compared to the apo (crRNA free) structure, the binary structure of the *Lsh*Cas13a-crRNA complex shows that the binding of crRNA to the protein results in substantial rearrangement of the Helical-2 domain and the Linker domain. The two domains come towards each other from the proximal to the distal end of the NTD domain and adopt a tweezers-like configuration while the HEPN-2 domain covers the crRNA-binding cleft [46].

In a more compact Cas13d effector such as *Es*Cas13d, the DR recognition is mainly mediated by the NTD in the REC lobe due to missing a large portion of the Helical-I domain with additional contacts with HEPN-1 and HEPN-2 domains in the NUC lobe. The spacer region immediately downstream of the DR locates within a channel between Helical-1 in the REC lobe and Helical-2 in the NUC lobe[27,45]. Similar to *Es*Cas13d, the *Ur*Cas13d crRNA is located within a positively charged channel formed by REC and NUC lobes[49]. Unlike Cas13a or Cas13b crRNAs, where their DRs are buried within the protein, the distal stem and the loop of Cas13d DRs (*Es*Cas13d: the first 2-nt base pairs and a 5-nt loop; *Ur*Cas13d: the first 4-nt base pairs and a 4-nt loop) protrude out into the environmental solvent from the protein (**Figure 1.5c**). Interestingly, hydrated Mg2+ ions aid in stabilizing DR confirmation of type VI-D crRNA and are essential for target RNA cleavage[45,49].

Recently, the high-resolution structures of Cas13b effectors revealed that the crRNA binding is coordinated by Helical-I, Helical-II, and lid domains[44]. Both *Bz*Cas13b and *Pbu*Cas13b adopt a triangular shape with a highly positive channel that accommodates the central L-shaped crRNA where the hairpin loop formed by DR is perpendicular to the spacer region[44,51]. The DR and 3' end of the spacer is located within the protein and make extensive intermolecular interactions with all domains except HEPN-2. The crRNA DR is mainly stabilized by the RRI-2 domain (Lid and Helicl-1 III in *Pbu*Cas13b) and locates within the channel formed by Helical-2, RRI-1, RR1-2 domains, and the linker region[44,51](**Figure 1.5b**). The 3' segment of the spacer is surrounded by Helical-1, HEPN-1, RRI-1, and RRI-2 domains on one side, and by the Helical-2 domain on the opposite side[51]. Unlike the flexible exchangeability between non-cognate crRNA DR and Cas13 effectors in the previously mentioned type VI-A systems, the interactions between Cas13b and crRNA DR are more ortholog-specific. However, those interactions mainly function in stabilizing the nucleic acid phosphate backbone, suggesting that the crRNA DR of sub-type VI-B is recognized in a structure-specific, rather than a sequence-specific manner[44,51].

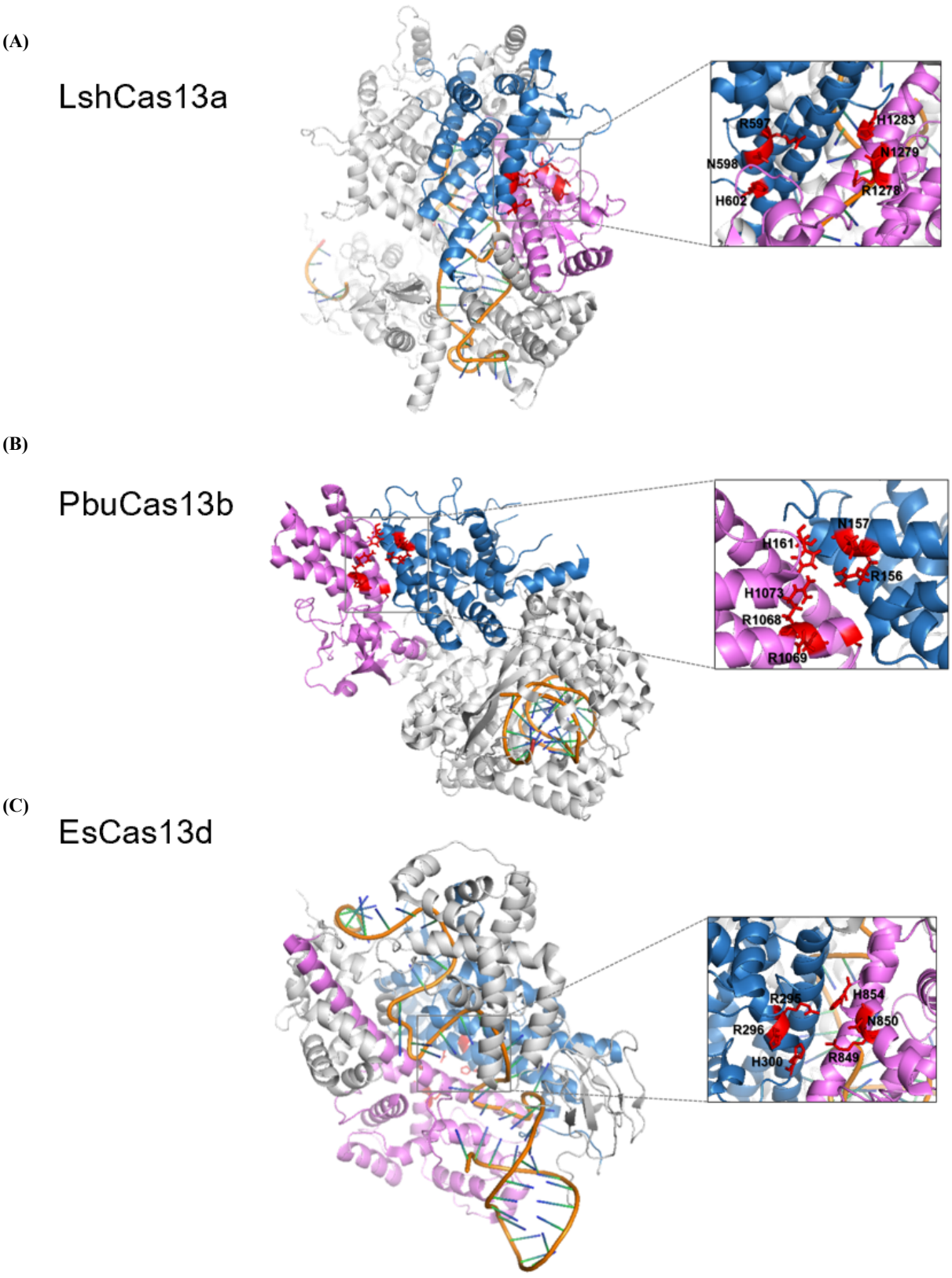


Figure 1. 5. High-resolution binary structure of Cas13 proteins interacting with their cognate crRNAs.

(A) LshCas13a (PDB ID:5WTJ)[46] (B) PbuCas13b (PDB ID:6DTD)[44], and (C) EsCas13d (PDB ID: 6E9E)[45] as examples. Insets show active site residues oriented with catalytic residues aligned (red). (Figure generated by Pymol)

1.3.4 Target RNA recognition and cleavage by Cas13-crRNA complex

The recognition of the target RNA by the Cas13-crRNA complex through RNA-RNA basepairing, the activation of Cas13 HEPN RNases, and the cleavage of the target RNA all involve serial conformation changes in the target RNA-crRNA duplex and the Cas13 effector. In the case of Cas13a, the 5' and 3' ends of the spacer sequence are buried within the protein, while the structurally disordered central portion appears to be exposed to solvent, suggesting its flexible movement within this nucleoprotein complex[47,48]. It is proposed that the central segment initiates the spacer-target RNA hybridization process, which drives a conformational change in Cas13a to further expose the 5' and 3' ends for full basepairing with the target. In line with that, the solvent-exposed region is more sensitive to mismatches due to its presumed role in duplex nucleation and propagation of target RNA-crRNA basepairing, prompting some studies to refer to this region as the seed region in comparison to the 2-8 nt region within the Ago2-siRNA nucleoprotein complex that mediates target recognition in eukaryotic RNAi. For example, the *Lbu*Cas13a has a 9-15nt seed region within its spacer, while *Lsh*Cas13a also has a central seed region although its nucleotide position is not well defined[46,47]. The formation of crRNA-target RNA duplex mediates major displacement of both crRNA and Cas13a domains, which triggers HEPN activation and target cleavage[46-48,55] (**Figure 1.6a**). The catalytic R-XXXXH motifs in two HEPN domains are located at distal positions from each other within the *Lbu*Cas13a binary complex. During the transition to the ternary complex through spacer-target basepairing, the structural rearrangements bring HEPN-1 and HEPN-2 domains into close proximity to form a concave RNase active site on the outer surface of the protein[47]. In contrast to Cas9 and Cas12 effectors whose nuclease active sites are located close to the target DNA-crRNA duplex, the active sites of Cas13 proteins are far away from the central channel that wraps the RNA-RNA duplex[47]. In the ternary structure of the *Lbu*Cas13a-crRNA-target RNA complex, the 1-24 base pairs of the duplex fits in the positively charged channel in the NUC lobe and are surrounded by the Helical-2, Helical-3, and HEPN-1 domains with physical contact with the protein through 7-15nt, 18-24nt of the spacer and 11-21nt of the target RNA. The 5' end of the target RNA extending from the crRNA-target RNA duplex is captured by a HEPN-1 β-hairpin in the catalytic active site[47].

A similar solvent-exposed motif has been observed in the type VI-D systems to describe target recognition and cleavage processes in Cas13d (**Figure 1.6c**). Different from the Cas13a binary configuration, a large proportion of crRNA spacer is solvent-accessible in the *Es*Cas13d-crRNA complex, suggesting that the base pairing with the target RNA may initiate at multiple locations along

the spacer sequence. An initial mutagenesis study of target binding and cleavage by *Es*Cas13d did not find a clear seed region on its spacer. *Es*Cas13d likely undergoes partial conformational changes once 18-nt complementarity is reached and attains maximum activity upon the formation of a 22-nt long crRNA-target RNA-RNA duplex[45]. A later study of the *Ur*Cas13d binary complex suggested that both central 4-8nt and 3' end 14-22 nt regions on the spacer are exposed to the environmental solvent. In line with that, these two solvent-exposed spacer regions (5-8nt and 13-22nt) exhibit higher mismatch intolerance characteristics which is suggestive of potential seed[49]. A recent library screen investigated *Ruminococcus flavefaciens Rfx*Cas13d (also known as CasRx) targeting rules in mammalian cells and indicated that the seed may reside within a central region (15-21nt) of the spacer[56]. To accommodate the target RNA in the positively charged channel and transit to a cleavage-competent conformation, Cas13d domains undertake numerous conformational movements by forming new interactions with the crRNA and target RNA. In the ternary *Es*Cas13d complex, the most dramatic change occurs within the Helical-1 domain, which shifts outward to accept target RNA binding within an expanded cleft. Further stabilization is achieved by subtler rearrangements of the HEPN-1, Helical-2, and HEPN-2 domains, particularly regions proximally to the 3' end of the spacer. Comparing the positions of HEPN domains between binary and ternary complexes revealed that the catalytic residues of HEPN-2 move proximal to the corresponding catalytic residues on HEPN-1 to generate a bipartite active site facing externally[45]. In addition, *Ur*Cas13d crystal structures indicate that Cas13d requires a relatively small structure rearrangement to achieve the cleavage-competent conformation, which is consistent with its compact size compared to *Lsh*Cas13a and *Bz*Cas13b[49].

In contrast to Cas13a and d effectors, the model of crRNA-mediated target recognition and cleavage is controversial for Cas13b[44,51]. (**Figure 1.6b**). Slaymaker *et al*. suggested that a conformation change in Cas13b happens first to expose the 3' end of the crRNA spacer, which allows the nucleation of target RNA-spacer basepairing. The initial binding mediates a structural rearrangement that opens the HEPN-1 and Helical-2 domains to allow the target RNA to access the positively charged central channel. The rest of the spacer sequence gradually pairs with the target RNA to achieve full conformation changes and activation of the *Pbu*Cas13b[44]. However, this model was challenged since the it didn't describe how the initial confirmation change happens to expose the buried 3' end of spacer[33]. Moreover, researchers questioned whether the channel between the HEPN-1 and Helical-2 is flexible enough to accommodate the target RNA given the movement of the HEPN-1 domain is restricted by the IDL[33]. Furthermore, this model cannot apply to *Bz*Cas13b-mediated target recognition since the channel between the HEPN-1 and Helical-2 domains does not exist in the binary structure of the *Bz*Cas13b-crRNA complex[51]. There is structural evidence that suggests the existence of a solvent-exposed motif within the spacer sequence of Cas13b systems too, which could mediate the initiation of the target recognition process. The 9-15nt of *Bz*Cas13b spacer is solvent-exposed while the 12-17nt of *Pbu*Cas13b spacer is sensitive to single nucleotide mismatches, which indicates that the type IV-B

systems may have central seed regions as observed in sub-type VI-A[44,51]. Collectively, the target RNA may hybridize with the central part of the spacer and progressively extend towards the 5' and 3' ends[33]. To date, the ternary architecture of Cas13b has not been solved yet to provide structural insights into the HEPN nuclease site activation and RNA cleavage after target RNA binding. Similarly, little is known about crRNA binding and target recognition for types VI-C, X, and Y due to the lack of high-resolution structures[25].

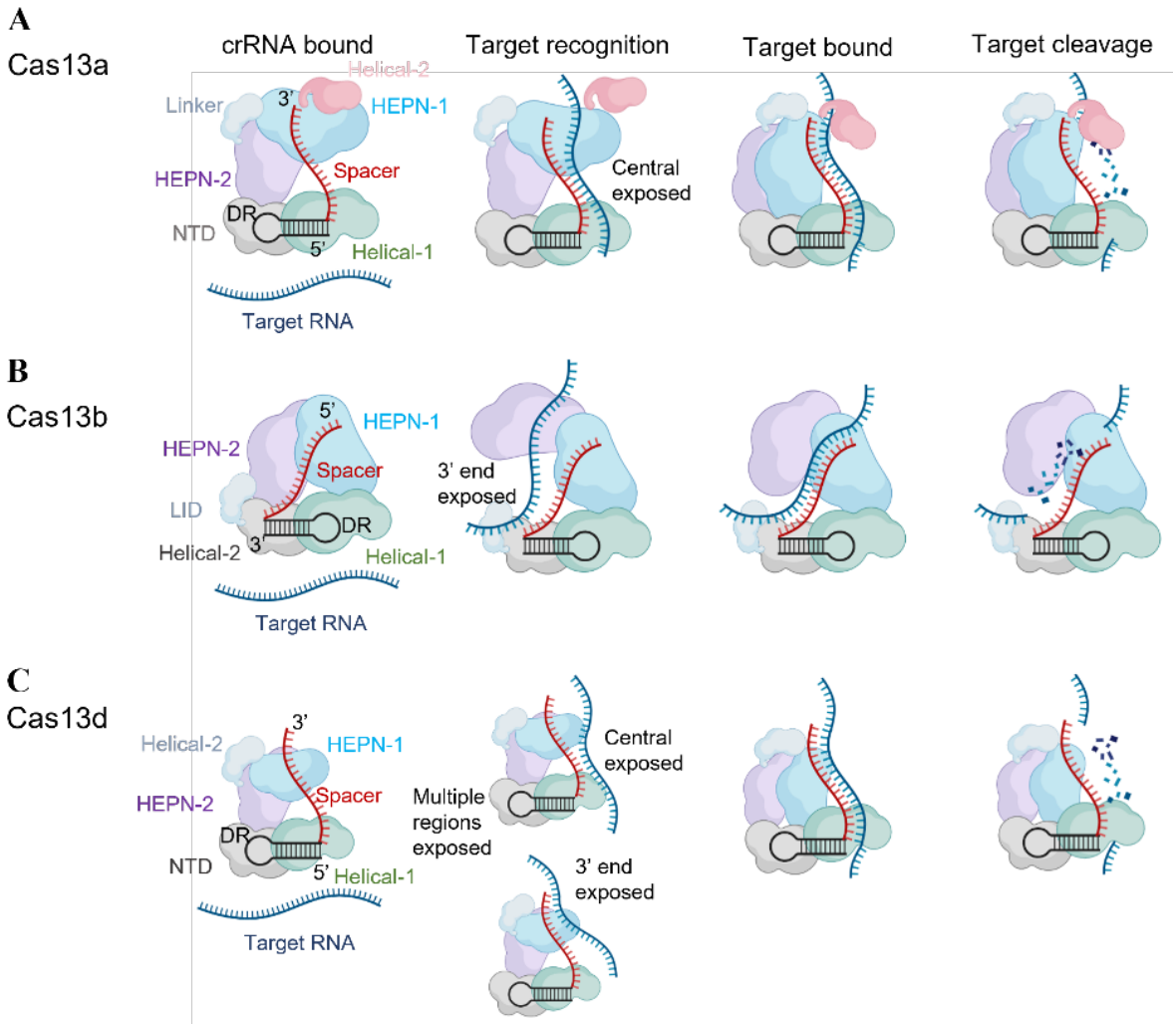


Figure 1. 6. Schematics showing structural mechanisms of RNA targeting by Cas13a, Cas13b and Cas13d effectors.

(A) In Cas13a, the binding of crRNA is mainly mediated by the NTD domain and Helical-1 domain. The solvent-exposed central segment of the spacer initiates the crRNA-target RNA hybridization. Cas13a effectors undergo further conformational changes for full-length base pairing, which triggers HEPN activation and target cleavage. (B) A cartoon depicting the predicted dynamics of Cas13b, although it remains controversial due to contradictive studies. Helical-1, Helical-2, and Lid

domains recognize the crRNA. HEPN1 opens to accept the target RNA into the central channel to hybridize with the guide RNA spacer. HEPN1 closes to allow for Cas13b-mediated cleavage to occur. (C) The crRNA loading and target cleavage in type VI-D systems (Cas13d) are similar to type VI-A. However, the target recognition may initiate at multiple regions of the spacer. Domains and RNAs are coloured according to annotations.

1.3.5 Protospacer flanking sites

Many CRISPR effectors require a short sequence called protospacer flanking site (PFS) or protospacer adjacent motif (PAM) for the activation of the nuclease domain and cleavage. PFS or PAM sequences are present in the genetic material of the invaders next to the spacer binding sites, which enables CRISPR effectors in the host to distinguish self from nonself sequences, thereby preventing cleavage of host nucleic acids[17]. CRISPR-Cas9 requires 2-5nt long PAM sequences adjacent to the spacer binding site on the target[57]. For example, *Sp*Cas9 requires a downstream 'NGG' sequence in the non-complementary strand of sgRNA to cleave the target DNA sequence[16].

In CRISPR DNA interference, a sequence complementary to the spacer is stored in the host genome (antisense strand) so that the absence of a PAM/PFS inhibits CRISPR effectors from targeting the host genome. In CRISPR RNA interference, however, PAM/PFS sequence may not be required as the target transcript that marches the spacer sequence is not expressed by the host. Thus, it remains unclear if RNA-targeting CRISPR-Cas13 enzymes require a PFS sequence for efficient RNA silencing.

Several studies have used library screens in bacteria and found increased silencing efficiency in the presence of PFS for Cas13a and Cas13b[24,26,29,36,39]. For instance, *Lsh*Cas13a depends on a 3'H (not G) base immediately flanking the spacer to activate its nuclease activity[24]. *Pbu*Cas13b has a 5' PFS preference for non-C nucleotide and a weak AA preference at the 3′ of the target sequence[26,29]. However, such PFS constraints did not significantly affect RNA targeting by Cas13a or Cas13b in the mammalian cells [25,29,36,39]. Notably, high throughput screens in bacteria found no evidence of PFS in Cas13d, while the PFS in Cas13c, Cas13x, and Cas13y remain to be investigated[25,28,45]. A small-scale screen conducted in mammalian cells also indicated the absence of PFS for efficient RNA silencing with Cas13x.1[28]. Most available studies suggest either permissive or the absence of PFS sequences required for CRISPR-Cas13 activation. While additional investigations are required to strengthen this conclusion, the lack of a fitness advantage via the acquisition of PFS throughout the evolution of these RNA targeting CRISPR appears to support this conclusion.

1.3.6 Targeting rules

Currently, the target recognition process employed by Cas13 remains poorly understood, limiting the rational design of efficient crRNAs as would be required to target aberrant tumour RNAs in mammalian

cells. One significant factor may be the accessibility of highly structured regions on target RNAs as Cas13 lacks a helicase domain to unwind double-stranded RNA motifs to facilitate target recognition and spacer-target basepairing[58]. A screen targeting essential genes in *E.coli* indicated that spacers of Cas13b with a secondary structure near the protospacer had reduced targeting efficiency[26]. Similarly, crRNAs targeting regions with low secondary structure showed better knockdown efficiency[39]. A pooled library screen was performed where *Rfx*Cas13d loaded with tiled crRNAs targeting GFP and three cell surface markers (CD46, CD55, and CD71) in mammalian cells to study the targeting rules of *Rfx*Cas13d for optimal guide design (https://cas13design.nygenome.org/). They indicated that the target RNA context and crRNA features may constrain target knockdown efficacy[59]. In a more recent study, Wei *et al*. performed a larger positive selection screen to investigate *Rfx*Cas13d targeting rules using over 127,000 single-nucleotide resolution crRNAs targeting 55 essential genes. Deep machine learning that incorporates various parameters including secondary structure, target site position, and target isoforms enabled the authors to develop a model to predict highly effective crRNAs with >90% accuracy and a crRNA design tool (http://RNAtargeting.org)[60]. As both studies listed above claimed the development of highly accurate crRNA design tools for *Rfx*Cas13d, it would be informative if the scientific community independently compare the prediction accuracy of these two algorithms in back-to-back experiments. In addition to the crRNA features (e.g., crRNA nucleotide composition, crRNA folding), the target RNA sequences and RNA-RNA hybridization, other factors such as RNA-binding proteins which could act as roadblocks and impair target accessibility should also be taken into consideration during the prediction of potent crRNAs by machine learning algorithms.

Although *Rfx*Cas13d targeting rules have been investigated, the molecular targeting rules of other type VI sub-types remain largely unknown. High-resolution structures of Cas13 binary and ternary complexes give insights into general conformational changes of Cas13 during interaction with crRNA and target RNA. However, mechanisms underlying the binding of the target RNA, conformational rearrangement of HEPN nuclease sites and target RNA cleavage remain largely unknown. Apart from uncovering crRNA design rules by pooled library screens, identifying and characterizing the critical motifs of crRNA spacer for target RNA binding and HEPN nuclease site activation are particularly crucial for designing efficient crRNAs. The lack of these knowledge limits the potential for future Cas13 applications.

1.3.7 Collateral (trans-cleavage) activity

In addition to on-target cis-cleavage, some Cas13 orthologues have been shown to possess a 'collateral activity' that indiscriminately degrades spatially proximal RNAs. This collateral cleavage of adjacent RNAs is also found in CRISPR-Cas12a but not in Cas9[23,24,26,27,61]. It is believed that, after phage infection, the on-target activity of Cas13 directly silences the phage's RNA through spacer-target basepairing, whereas the collateral activity triggers a global degradation of host transcripts to prevent further spread of the phage. This mechanism leads to cell dormancy or suicide, protecting the bacterial

community at the expense of an altruistic individual bacterium[62]. Mechanistically, the collateral activity is triggered by spacer-target basepairing and subsequent conformational changes in Cas13 structure that exposes an activated RNA nuclease domain to the surface of this protein, which it interacts with and cleaves various host RNAs that are in the vicinity. In that sense, the collateral activity of Cas13 is still highly target-dependent and becomes unleashed only in infected bacteria.

This collateral activity of various RNA-targeting CRISPR effectors has been well-documented in bacteria and in *in-vitro* biochemical assays. In fact, various researchers took advantage of this collateral activity to develop ultra-sensitive *in-vitro* diagnostic devices (e.g., SHERLOCK, described in detail in **chapter 1.5.1.2**) in which target recognition activates the Cas13 nuclease domain that cleaves an RNA reporter in the reaction mix. Cleavage of the RNA reporter leads to the physical separation of a fluorophore and a quencher, resulting in fluorescence activation[63].

However, this collateral activity remains controversial in mammalian cells as various studies reported conflicting findings. Initial investigations of Cas13 that used transcriptomic analysis to investigate its on- and off-target potential reported extremely high on-target activity and no collateral degradation of cellular RNAs, suggesting high specificity[26,29,31,35,37,39]. However, a study published in 2019 first reported collateral activity in eukaryotic cells when targeting EGPF by *Lwa*Cas13a in human glioma cells[64]. Interestingly, the authors suggested that the collateral activity might be cell line dependent since they did not observe this activity in HEK 293T cells[64]. More recently, collateral activities induced by *Lwa*Cas13a, *Psp*Cas13b, *Rfx*Cas13d, and Casx.1 have been reported in various mammalian cells including HepG2, AT2, B16F10, GL261 cells, and mouse ES cells by assessing cell viability, RNA integrity, or transcriptome sequencing[28,65-68].

There are several hypotheses to explain these apparently contradictory observations. Further studies are required to test these hypotheses:

1) Cas13's collateral activity largely may depend on cell type and target RNA abundance, as the amplitude of collateral activity was positively correlated with the expression level of the target[28,65,66,68]. For instance, collateral activities of *Lwa*Cas13a, *Rfx*Cas13d, and Cas13x.1 were observed when targeting overexpressed exogenous transcripts but were undetectable when targeting endogenous RNA in HEK293T cells[28].

2) Conflicting results may arise due to the lack of a proper internal control in RNA sequencing. A non-targeting crRNA commonly used as a negative control for assessing the transcriptomic specificity of CRISPR effector may only reflect off-target activity but not collateral activity if a targeting crRNA corresponding to an mRNA actively expressed in that cell is not included as a control[66]. A recent preprint publication used a spike-in control strategy in RNA sequencing to normalize the RNA extracted from HEK293T cells expressing *Rfx*Cas13d to spike-in mouse cell RNA. They observed that the entire human transcriptome was downregulated on average by 46% compared to spiked control

RNA encoding commonly used internal controls such as *GAPDH* and *ACTB*[66]. Although this approach is interesting, it doesn't directly capture endogenous RNA degradation in live cells when Cas13 nuclease domain is activated.

3) Assessing collateral activity at different time points may give varying results. For example, collateral cleavage may occur soon after the initial expression of Cas13 but then be inactivated by endogenous regulatory pathways induced in response to widespread RNA degradation[33]. In an analogous fashion, Cas13 mRNA may itself be cleaved by collateral activity, which would act as a negative feedback loop regulating its nuclease activity[68].

4) Collateral cleavage may occur only when Cas13 effectors and target/environment RNAs are co-localized spatially and temporally, greatly limiting the impact on most endogenous mRNA transcripts.

1.4 A comparison of CRISPR-Cas13 and other RNA interference tools

1.4.1 RNAi shares some commonalities but has many differences from CRISPR Cas13

Two decades before the discovery of Cas13, small non-coding RNAs termed microRNAs (miRNAs) were identified to regulate RNA expression in nearly all eukaryotes by a process called RNA interference (RNAi)[69]. Similar to the RNA-guided targeting mechanism used by Cas13 in bacteria[70], in eukaryotic cells, RNAi processed to a miRNA is loaded into the Argonaut protein as part of an RNA-induced silencing complex (RISC) in which the miRNA guides target recognition and cleavage through miRNA-RNA basepairing[69]. Shortly after the discovery of eukaryotic RNAi, synthetic small interference RNA (siRNA) and small hairpin RNA (shRNA) mimicking natural miRNAs were developed and introduced into cells to reprogram the endogenous RISC complex to silence any transcript of interest[71] (**Figure 1.7**). Since then, RNAi technology has been established as a strong tool to knock down a single RNA or a multitude of transcripts using pooled shRNA screens[72].

Recent studies that have adapted CRISPR-Cas13 to mammalian cells have demonstrated potent gene silencing at the transcript level. Both Cas13 and eukaryotic RNAi can achieve high knockdown efficiency of a target transcript[58], however, CRISPR-Cas13 has numerous practical advantages over eukaryotic RNAi. One of the limitations of eukaryotic RNAi therapeutics is poor specificity dictated by the very short seed region (7-8nt) of siRNA/shRNA, which typically leads to the recognition and off-target silencing of hundreds of endogenous genes. This is because an 8-nt recognition sequence (A, G, C, or U at eight consecutive positions) can result in no more than 4e8 (approximately 65,000) different 8-mer sequences, whereas the human transcriptome consists of several billion bases. After a multi-billion dollar investment in eukaryotic RNAi, the lack of specificity of RNAi precipitated the premature termination of several RNAi programs by Big Pharma[73,74]. In contrast to eukaryotic RNAi, Cas13 has

a 22-30 nt spacer sequence that mediates target recognition with low mismatch tolerance and greatly reduces the probability of off-targeting[24,44,58]. *Lsh*Cas13a and *Psp*Cas13b have a spacer central seed region that is even sensitive to single, double, or consecutive mismatches[29]. Recently, a study found that *Rfx*Cas13d-mediated knockdown was diminished significantly by a single-nucleotide mismatch between nucleotide 15 and 21 of the spacer[59]. Consistent with these mutagenesis studies, *Psp*Cas13b and *Lwa*Cas13a RNA silencing in mammalian cells also demonstrated negligible off-target effects trascriptome-wide[29,39]. Notably, a recent study compared the off-target potential of shRNA and RfxCas13 and found no off-target events associated with *Rfx*Cas13d knockdown of *ANXA4, whereas shRNA knockdown exhibited* over 900 off-target transcripts[37]. In addition to its higher specificity, Cas13 enzymes can be tagged with a nuclear localization signal (NLS) or a nuclear export signal (NES) sequence to cleave targets localized in either the nucleus (e.g. pre-mRNAs and long non-coding RNAs) or cytoplasm (e.g. mRNAs)[29,31,39]. Last but not least, si/shRNA can hijack the cellular RISC machinery and may impair gene expression regulation through competition with natural miRNA processes[75]. Therefore, Cas13 may be a more suitable approach to target pathogenic RNA.

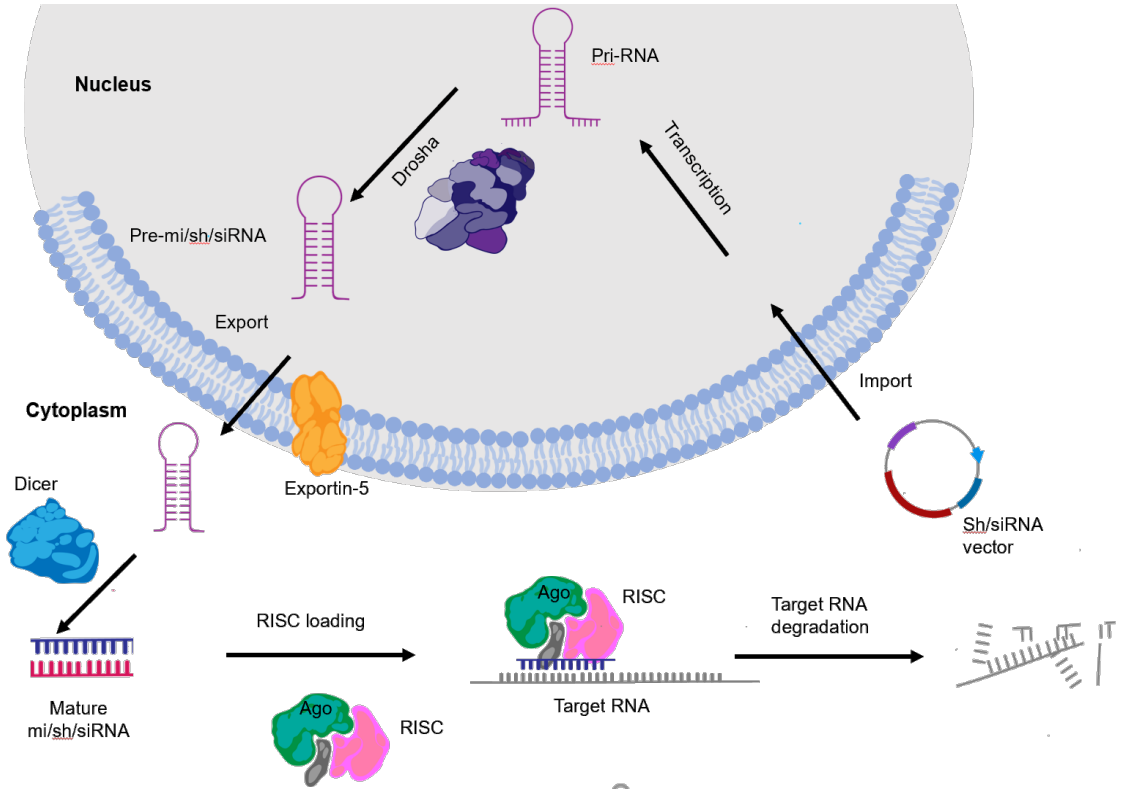


Figure 1. 7. RISC-mediated mRNA degradation.

A pri-RNA (endogenous or sh/siRNA expression vector transcribed) is processed into the pre-mi/sh/siRNA by the RNase III enzyme Drosha. Exportin-5 mediates the export of pre-mi/shRNA to the cytoplasm where Dicer cleaves the terminal loop to generate a mature miRNA/siRNA/shRNA. This duplex RNA is loaded into Argonaut protein within the multi-subunit RISC complex. The loaded

miRNA/shRNA/siRNA directs RISC to target RNA with sequence complementarity, resulting in target RNA translational repression and RNA degradation.

1.4.2 Cas7-11: a new RNA-targeting CRISPR effector

Among the known CRISPR families, only Class 1 type III, and Class 2 type VI enable targeted RNA cleavage. Type III members possess multiple effector subunits that coordinate crRNA binding, maturation, and cleavage of DNA or RNA targets[76,77]. *In vitro*, RNA interference is executed by multiple Cas7 (Csm3) proteins of the ribonucleoprotein complex at 6-nt intervals, whereas the DNA cleavage is mediated by Cas10[78].

Very recently, Özcan *et al.* discovered subtype III-E, which is characterized by a single effector protein Cas7-11 that emerged from the fusion of a putative Cas11 (Csm2) domain and four Cas7 proteins[65]. Like many other CRISPR families, the type III-E loci contain a *cas1* gene[79] and a caspase-like protease Csx29 that acts as a negative regulator of RNA interference activity, similar to the role of Csx27 protein in type VI-B1. The study observed that deletion of the *Csx29* gene in the *Desulfonema ishimotonii* Cas7-11 (*Di*Cas7-11) locus improved the target cleavage activity by *Di*Cas7-11 in both bacteria and *in vitro*[26,65]. Very recently, three back-to-back studies demonstrated that the target RNA binding to Cas7-11 activates the protease activity of Csx29, resulting in the cleavage of Csx30, a protein expressed from the same locus[80-82]. Cleaved Csx30 can interact with the RNA polymerase to remodel gene expression in response to phage infection. *Di*Cas7-11 lacks a Cas10 component and therefore does not cleave DNA. Like Cas13, *Di*Cas7-11 alone can process pre-CRISPR RNA into 32-38nt mature crRNA and only targets ssRNAs due to the absence of RNA helicase domains[65]. Furthermore, *Di*Cas7-11 appears to lack collateral activity *in vitro* or in bacteria[65]. This study claimed that *Di*Cas7-11 has robust RNA interference and editing in mammalian cells and outperformed the specificity of other RNA-targeting tools (shRNA and Cas13), although these conclusions have not been independently tested by other research groups[65].

1.5 Cas13 applications in biological research and biotechnology

CRISPR-Cas13 is the first naturally occurring CRISPR discovered that exclusively targets ssRNA with high efficacy and specificity, offering many opportunities for targeted RNA manipulation in research, biotechnology, and biomedicine[25,37]. Recent studies have harnessed this novel RNA toolbox in a wide range of RNA applications including RNA silencing, fast and robust nucleic acid detection, RNA base-editing, and live-cell RNA imaging[29,39,53,58,83-85].

1.5.1 Applications of active Cas13 effectors

Cellular RNA is fundamental in various biological processes. Protein coding RNA or messenger RNA (mRNA) conveys genetic information for protein synthesis[86]. On the other hand, non-coding RNA (ncRNA) such as transfer RNA (tRNA), ribosomal RNA (rRNA), and long non-coding RNA (lncRNA) play essential roles in regulating gene expression and catalysing biological reactions[86]. The use of conventional RNAi as a versatile tool for mRNA and ncRNA manipulation is hampered by its lack of specificity[72,73]. CRISPR-Cas13 however possesses high specificity and can provide a flexible and robust module for investigating cellular RNAs in mammalian cells[87,88].

1.5.1.1 Transcriptome targeting in mammalian cells

The most straightforward application of CRISPR-Cas13 is RNA knockdown (**Figure 1.8a**). In 2017, Abudayyeh *et al.* evaluated the RNA targeting efficiency of 15 Cas13a orthologues in eukaryotic cells and found *Lwa*Cas13a to be the most effective. They demonstrated that *Lwa*Cas13a can be reprogrammed to cleave endogenous transcripts *KRAS*, *CXCR4*, and *PPIB* with various efficiencies (40%-84%), which was comparable to position-matched shRNAs. However, gene silencing with *Lwa*Cas13a required a fusion with monomeric super folder green fluorescent protein (msfGFP) that was thought to enhance protein folding and stability in human cells[39]. Later, a larger screen of 21 Cas13a orthologues, 15 Cas13b orthologues, and 7 Cas13c orthologues concluded that *Psp*Cas13b outperformed the silencing efficiency of *Lwa*Cas13a in that it achieved 90% knockdown of luciferase reporter by *Psp*Cas13b versus 40.1% by *Lwa*Cas13a-msfGFP; furthermore, *Psp*Cas13b did not require an msfGFP to retain function in mammalian cells[29]. Subsequently, researchers discovered a compact orthologue called *Rfx*Cas13d which showed a robust knockdown of 14 endogenous transcripts including *B4GALNT1* and nuclear lncRNAs *HOTTIP* and *MALAT1* in human cells. The initial study claimed that *Rfx*Cas13d had consistently favourable targeting activity (96%) in HEK293 FT cells across 3 transcripts relative to shRNAs (65%), dCas9-mediated RNA interference (53%), *Lwa*Cas13a-msfGFP (80%) and *Psp*Cas13b (66%)[37]. More recent studies suggested that the variability in silencing potency of Cas13a, b, and d effectors was likely due to the different cell models, crRNA design, and target RNAs employed[28,65,89]. In fact, as these orthologues have little sequence homology and their crRNA design principles remain largely unknown, it is not valid to compare their potency based on one or a few silencing assays. We favour the concept that the evolutionary arms race between bacterial immunity and bacteriophages shaped the potency of these RNA targeting tools, selecting orthologues that all can achieve high silencing should we discover the design principles and best cell model to demonstrate their maximum efficiency. Even the newly identified Cas13bt1, Cas13bt3, Cas13x.1, and Cas13y.1 demonstrated comparable knockdown efficiency to previously characterized Cas13 effectors in human cells[28,36].

Recent advances in CRISPR pooled library screens have revolutionized the scale of forward genetic screening, enabling the interrogation of thousands of genes in a single experiment. CRISPR-

Cas13 is a new player in the arena of forward genetic screens that expands the screen to both coding and non-coding RNAs and brings a new dimension of reversibility. Just 5 years after the first paper describing Cas13, Xu *et al.* published the first pooled library screen in early 2020 where they applied *Lwa*Cas13a to target a subset of very long non-coding RNAs (vlincRNAs) and protein-coding mRNAs expressed in K562 cells in response to anti-cancer drugs. This dropout screen enabled the identification of up to 60% of the tested mRNAs and 64% vlincRNAs as important for cellular survival in response to anti-cancer drug treatments[87]. Shortly after, another *Rfx*Cas13d screen in HEK293 FT and HeLa cells revealed a group of essential circular RNAs (circRNAs) that control cell growth in a cell-type-dependent manner. The design of a crRNA library targeting the sequence spanning the specific back-splicing junction sites in 762 circRNAs, they were able to reprogram *Rfx*Cas13d to perform large-scale, loss-of-function screens of circRNA in human cells. Importantly, the study found a common oncogenic circRNA in different cell types, circFAM120A, which promotes cell proliferation by preventing the FAM120A mRNA from binding the translation inhibitor IGF2BP2[89]. Other studies further exploited the power of library screens to investigate the targeting rules of *Rfx*Cas13d[59,60]. Although these initial reports showed the great potential of Cas13 forward genetic screens, the delivery of a single copy of crRNA per cell can limit the penetrance of silencing as target recognition and cleavage is highly dependent on the abundance of intracellular crRNAs. This copy number limitation is irrelevant for Cas9 screens as only two DNA alleles require editing, whereas efficient RNA silencing with Cas13 typically requires the cleavage of several thousand RNA copies present in a cell.

1.5.1.2 Nucleic acid detection

Upon phage RNA recognition, the promiscuous collateral activity of Cas13 leads to the degradation of nearby host RNAs to plunge the infected bacterium into a dormant state, thereby limiting viral replication. The Zhang Lab harnessed this collateral activity of *Lwa*Cas13a to develop a fast and cost-effective diagnostic tool referred to as Specific High Sensitivity Enzymatic Reporter UnLOCKING (SHERLOCK) that detects a specific sequence of nucleic acids with zeptomolar to attomolar sensitivity[53,90]. In a SHERLOCK reaction, an RNA sample is first amplified and converted to cDNA to increase sensitivity, using recombinase polymerase amplification (RPA) followed by T7 transcription. Then, an ssRNA reporter is added which contains a fluorophore and a quencher at proximity. Recognition of the target sequence by Cas13 activates its two nucleases which cleave all nearby ssRNA. The cleavage of the ssRNA linker within the reporter liberates the fluorophore from the quencher, leading to photon emissions and a fluorescence signal, which when detected indicates the presence of the target sequence[53] (**Figure 1.8b**). Later, SHERLOCK was upgraded to SHERLOCKv2 to simultaneously detect several ssRNA and dsDNA targets in a single reaction with increased sensitivity thanks to a combination of Cas effectors (*Lwa*Cas13a, *Psm*Cas13b, *Cca*Cas13b, and *As*Cas12a and Csm6) and four di-nucleotide fluorescent reporters[90]. Furthermore, a simplified sample preparation protocol called Heating Unextracted Diagnostic Samples to Obliterate Nucleases (HUDSON) was

developed that couples with SHERLOCK for viral detection from patients' bodily fluids within 2 hours using a lateral flow strip[91]. Recently, this protocol was optimized as a diagnostic platform named SHINE (SHERLOCK and HUDSON integration to navigate epidemics) which further shorten the reaction time to 50 minutes while maintaining specificity and sensitivity. SHINE's results are visualised by fluorescence and interpreted by a companion smartphone application, making this tool convenient and user-friendly[92].

SHERLOCK is a versatile tool to detect DNA and RNA in basic research, clinical diagnosis, and for genotyping and detection of mutations that confer drug resistance or drive genetic diseases[53,90,93-97]. Compared to the conventional, lab-based *reverse transcription* polymerase chain reaction (RT-PCR) which is the gold standard diagnostic tool for viral and bacterial infections, SHERLOCK is faster, more accurate, and requires minimal equipment[63]. So far, SHERLOCK has been used to detect Zika virus, Dengue virus, BK polyomavirus, cytomegalovirus, Ebola virus, and Lassa virus (LASV) in patient liquid biopsy[53,91,98,99]. Strikingly, SHERLOCK was adopted in a clinical study to detect Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) in the Coronavirus Disease 2019 (COVID-19) and showed 100% accuracy compared to RT-PCR results, which highlights the potential of SHERLOCK as a portable and multiplexable method for on-site, large-scale diagnosis of emerging diseases[100].

Apart from SHERLOCK, numerous Cas13/crRNA-mediated nucleic acid detection methods have been developed. For example, *Lbu*Cas13a has been used to detect and differentiate miRNA with single nucleotide polymorphisms (SNPs), promising a great potential for early diagnosis of miRNA-related disease[101]. Importantly, a SHERLOCK-based multiplexable, large-scale diagnostic platform termed Combinatorial Arrayed Reactions for Multiplexed Evaluation of Nucleic acids (CARMEN) has been developed to simultaneously test for 169 human viruses, and to identify mutations that drive drug-resistance in HIV. In the CARMEN platform, Cas13 effectors, crRNA, and a single-stranded RNA reporter are assembled in nanolitre droplets. The detection mixes self-organize in a microwell array to pair with droplets of PCR or RPA amplified samples, testing each sample against more than 4,500 crRNA in replicates[102].

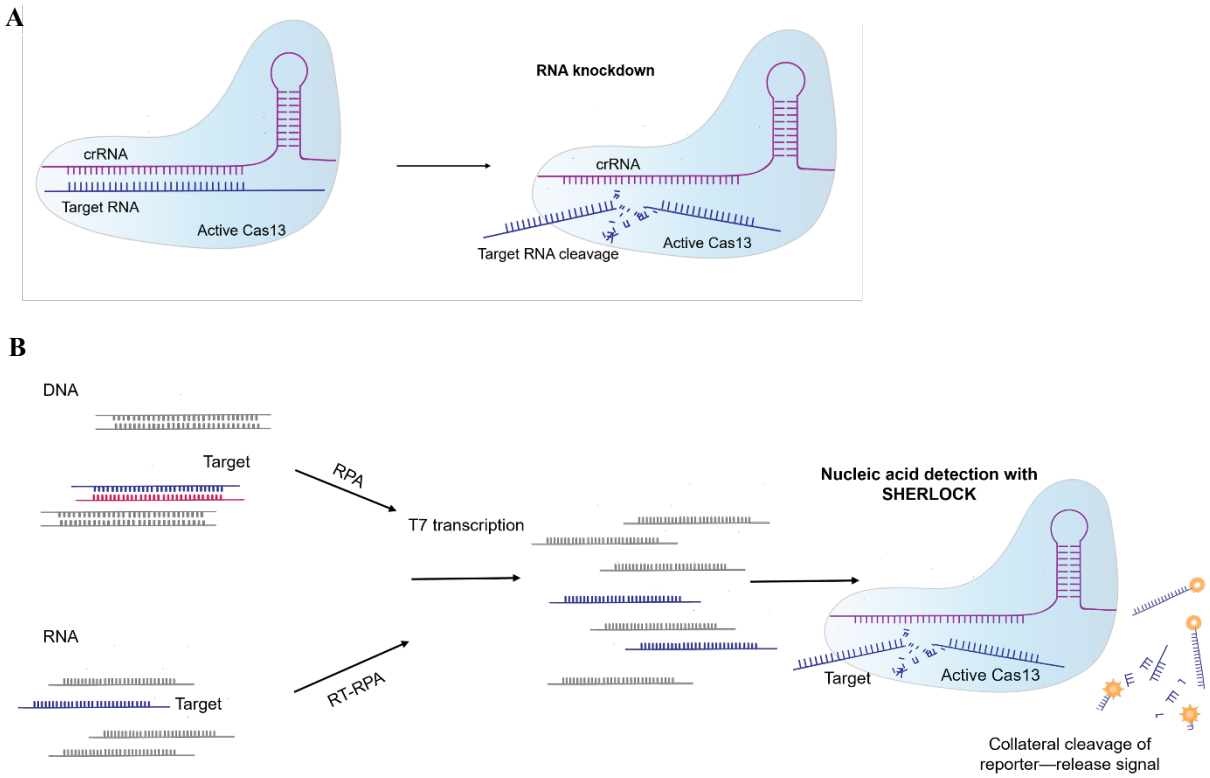


Figure 1. 8. Applications of Cas13 effectors.

(A) Type VI CRISPR-Cas systems can be used to knock down cellular RNA of interest in numerous ways to study RNA functions. (B) A schematic demonstrating nucleic acid detection using Specific High Sensitivity Enzymatic Reporter UnLOCKING (SHERLOCK). Briefly, nucleic acid samples are first amplified and converted to cDNA by recombinase polymerase amplification (RPA) or reverse-transcription (RT)-RPA, after which reverted to RNA by T7 transcription. Samples are then incubated with target-specific crRNA, Cas13 effectors, and RNA reporters containing a fluorophore-quencher. The presence of the target of interest triggers Cas13-mediated collateral cleavage of RNA reporters, resulting in fluoresce emission. The intensity of the fluorescence is proportional to target RNA abundance in the reaction, providing a quantitative measurement of target RNA expression.

1.5.2 Applications of catalytically inactive Cas13 (dCas13) effectors

Catalytically inactive Cas13 (or dead Cas13, dCas13) is engineered through the introduction of inactivating mutations into its nuclease domains. However, dCas13 enzymes retain their specific RNA-binding activity and can be programmed to bind to any transcript of interest[70]. By exploiting this programmable RNA-binding capability, further RNA manipulation applications have been developed to investigate various RNA processes including post-transcriptional modifications, RNA regulation,

RNA dynamics, and RNA stability in live cells. Below is provided an overview of dCas13 applications for precise RNA manipulations (refer to this review[85] for further details).

1.5.2.1 RNA editing: REPAIR and RESCUE

Cox and colleagues first demonstrated that d*Psp*Cas13b could be fused to an adenosine deaminase acting on RNA type 2 deaminase domain (ADAR2-DD) containing an E488Q hyper-activating mutation to edit endogenous transcripts[29]. ADAR2-DD deaminates adenosine (A) to inosine (I) which is read by ribosomes during protein translation as guanosine (G). They termed this RNA-editing nucleoprotein complex RNA Editing for Programmable A to I Replacement (REPAIR) (**Figure 1.9a**). By using a long spacer (50nt) with an A-C mismatch at the edited site, REPAIR was able to achieve a 50% editing rate in reporter RNAs and 23%-35% rates in transcripts that encode a pathogenic single-nucleotide mutation (e.g., 878G to A mutation in X-linked nephrogenic diabetes insipidus and a 1517 G to A mutation in Fanconi anaemia) in HEK293 FT cells. However, one concern of REPAIR was the significant off-target effects of altering other adenosine bases proximal to the targeted site. To address this limitation in REPAIRv2, a screen of ADAR2-DD variants was conducted to identify new ADAR2-DD mutants with higher specificity. An additional T375G point mutation was introduced into ADAR2-DD (E488Q) which significantly reduced the number of off-target events from 183,855 to 20 transcriptome-wide[29]. Later, a second RNA editing platform called RNA Editing for Specific C to U Exchange (RESCUE) was developed by fusing *Riemerella anatipestifer* (d*Ran*Cas13b) to a modified ADAR2. The authors used a directed protein evolution mutagenesis approach to develop a C to U editor capable of cytidine deamination, which further expands the possibilities of RNA base editing. The editing specificity of RESCUE was further improved by rational mutagenesis of ADAR2-DD at residues interacting with the RNA target, including a critical mutation S375A (referred to as RESCUE-S)[75]. Similarly, REPAIR/RESCUE.t1 and REPAIR/ RESCUE. t3 editors have been generated with Cas13bt.1 and Cas13bt.3 fused to ADAR2-DD variants, respectively[36]. Subsequentially, more Compact RNA base editors (865 aa) adenosine deaminase mxABE and cytosine deaminase mxCBE (truncated dCas13X.1 fused with ADAR2 variants) could achieve comparable or higher base conversions efficiency and specificity compared to REPAIRv2 (1,388 aa), d*Rfx*Cas13d-ADAR2-DD (E488Q/T375G) (1,375 aa) and RESCUE-S (1,495 aa) in HEK293T cells. The compact size of these RNA editors may offer higher editing efficiency *in vivo* as its size is compatible with the adeno-associated virus as a delivery modality for CRISPR-based therapeutics[28].

1.5.2.2 Live cell RNA tracking and imaging

Another valuable application of dCas13 is RNA tracking and imaging to study the dynamics of RNA throughout its life cycle in live cells (**Figure 1.9b**). RNA fluorescence *in situ* hybridization (FISH) is the gold standard method for tracking the spatial localization of RNAs. However, this technique requires the fixation and permeabilization of the cells, which is incompatible with dynamic RNA imaging in live cells[103]. RNA visualization and tracking in live cells can be achieved with the MS2-MCP system using

modified chimeric transcripts where the RNA of interest is fused to an MS2 binding site sequence forming stem-loop structures which allow the binding of bacteriophage-derived MS2 coat proteins (MCP) (for review, refer to[104]). In 2017, Abudayyeh *et al.* tagged d*Lwa*Cas13a orthologue with msfGFP fluorescent protein to track RNA localization by fluorescence microscopy. To reduce the background signal from free dCas13, the authors added negative feedback (NF) comprising a zinc finger binding site and KRAB domain in the dCas13 construct (d*Lwa*Cas13a-NF-msfGFP) to improve dCas13 translocation from the nucleus to the cytoplasm. When reprogrammed with a crRNA to bind to *ACTB* mRNA, d*Lwa*Cas13a-NF-msfGFP translocated from the nucleus to the stress granule likely due to the binding to *ACTB* mRNA target that resides in the stress granules[39]. More recently, Wang and colleagues developed CRISPR live-cell fluorescent *in situ* hybridization (LiveFISH) where d*Rfx*Cas13d was utilized to visualize RNA transcription of an artificial Doxycycline (DOX)-inducible gene cassette containing an MS2-repeat sequence in real-time[84]. Shortly after, Chen Lab demonstrated that d*Psp*Cas13b and d*Pgu*Cas13b could be programmed to label the endogenous mRNAs *NEAT1*, *MUC4*, and *GCN.* By combing different dCas13 orthologues or dCas13 and MS2 bacteriophage coat proteins, they were able to image RNAs in cells with dual colours. Moreover, the combination of dCas13 and dCas9 enabled simultaneous visualization of genomic DNA and nascent RNA transcripts arising at the same locus in real time[30,105].

1.5.2.3 Epitranscriptomic modification and regulation

The DNA binding programmability was leveraged by different research groups to engineer numerous tools for targeted chromatin epigenetic modification and gene regulation including CRISPRi and CRISPRoff for gene silencing, CRISPRa, and CRISPRon for gene activation. In these tools, dCas9 is fused to transcriptional regulators such as Krüppel-associated box (KRAB) and VP64, where dCas9 enables sequence-specific binding to the chromatin, binging the transcriptional regulator close to the targeted gene for enhanced epigenetic chromatin modifications.

On the other hand, dCas13 opens new opportunities for precise posttranscriptional modification of individual RNA to better understand the molecular basis of epitranscriptomic regulation and its role in various cellular processes[106]. N^6-methyladenosine (m^6A) is a common, dynamic, and selective epigenetic mark on mRNA orchestrated by various enzymes called m^6A reader, writer, and eraser. m^6A modification plays an important role in mRNA processing events such as alternative splicing, folding, translation, and degradation of transcripts[32]. Rauch *et al.* targeted methylated sites in RNA transcript by programmable m^6A reader proteins consisting of dCas13b and N-terminal domains of YTH domain-containing family protein 1 and 2 (YTHDF1 and YTHDF2). They demonstrated that dCas13b-YTHDF1 enhances the translation of the target transcripts while dCas13b-YTHDF2 promotes its degradation[107]. Li *et al*. developed a chimeric protein, named dm^{6}ACRISPR by fusing d*Psp*Cas13b to m^6A demethylase AlkB homolog 5 (ALKBH5), specifically demethylating m^6A of targeted mRNA such as cytochrome b5 form A (CYB5A) and β-catenin-encoding CTNNB1 to increase their stability or translation[106]. In

addition, a photoactivatable RNA m^6A editing platform, termed PAMEC, was engineered to spatiotemporally control m^6A modification in response to blue light. PAMEC is a two-component system: **(1)** an RNA binding complex consisting of the d*Pgu*Cas13b fused a the light-sensitive protein called calcium-and integrin-binding 1 (CIB1), and **(2)** cryptochrome circadian regulator (CRY2PHR) which, when stimulated with blue light, can heterodimerize with CIB1 fused to an m^6A effector such as m^6A demethylase fat-mass and obesity-associated protein (FTO) to achieve m^6A erasure or the METTL3-METTL14 m^6A methyltransferase complex for m^6A writing. The targeted RNA methylation (TRM) systems engineered by Wilson *et al.* utilize the truncated catalytic domain of methyl transferase METTL3 and METTL3-METTL14 methyltransferase complex linked to the nucleus- and cytoplasm localized d*Psp*Cas13b, respectively (**Figure 1.9c**). These two editors mediated selective m^6A RNA methylation on transcripts such as *ACTB*, *GAPDH*, *FOXM1*, and *SOX2*, which alters transcript abundance and induces alternative splicing[108].

1.5.2.4 RNA-protein mapping

RNA-binding proteins (RBPs) are major players that regulate the structure, localization, and function of various RNAs such as lncRNA, small nucleolar RNAs (snoRNAs), and mRNA through direct binding. Protein-centric methods such as cross-linking immunoprecipitation sequencing (CLIP-seq) and RNA immunoprecipitation sequencing (RIP-seq) have identified hundreds of new RNAs that bind to proteins of interest[109]. dCas13 provides a complementary platform to facilitate the identification of novel RBPs in an RNA-centric manner through the pulldown of the dCas13-target complex together with its RBP partners. CRISPR-based RNA-United Interacting System (CRUIS) uses the d*Lwa*Cas13a-crRNA complex to bind target RNA. Simultaneously, the bacterial ligase proteasomal accessory factor A (PafA) fused to d*Lwa*Cas13a catalyses the binding of the small prokaryotic ubiquitin-like protein PupE to RBPs at close proximity, which is then identified by mass spectrometry[110]. Han *et al.* developed a three-component fusion protein generated by fusing d*Rfx*Cas13d to the double-stranded RNA-binding domain of human protein kinase R (PKR) to enhance the binding of the target RNA, whereas a modified plant peroxidase APEX2 can mediate the biotinylation of neighbouring RBPs[111](**Figure 1.9d**).

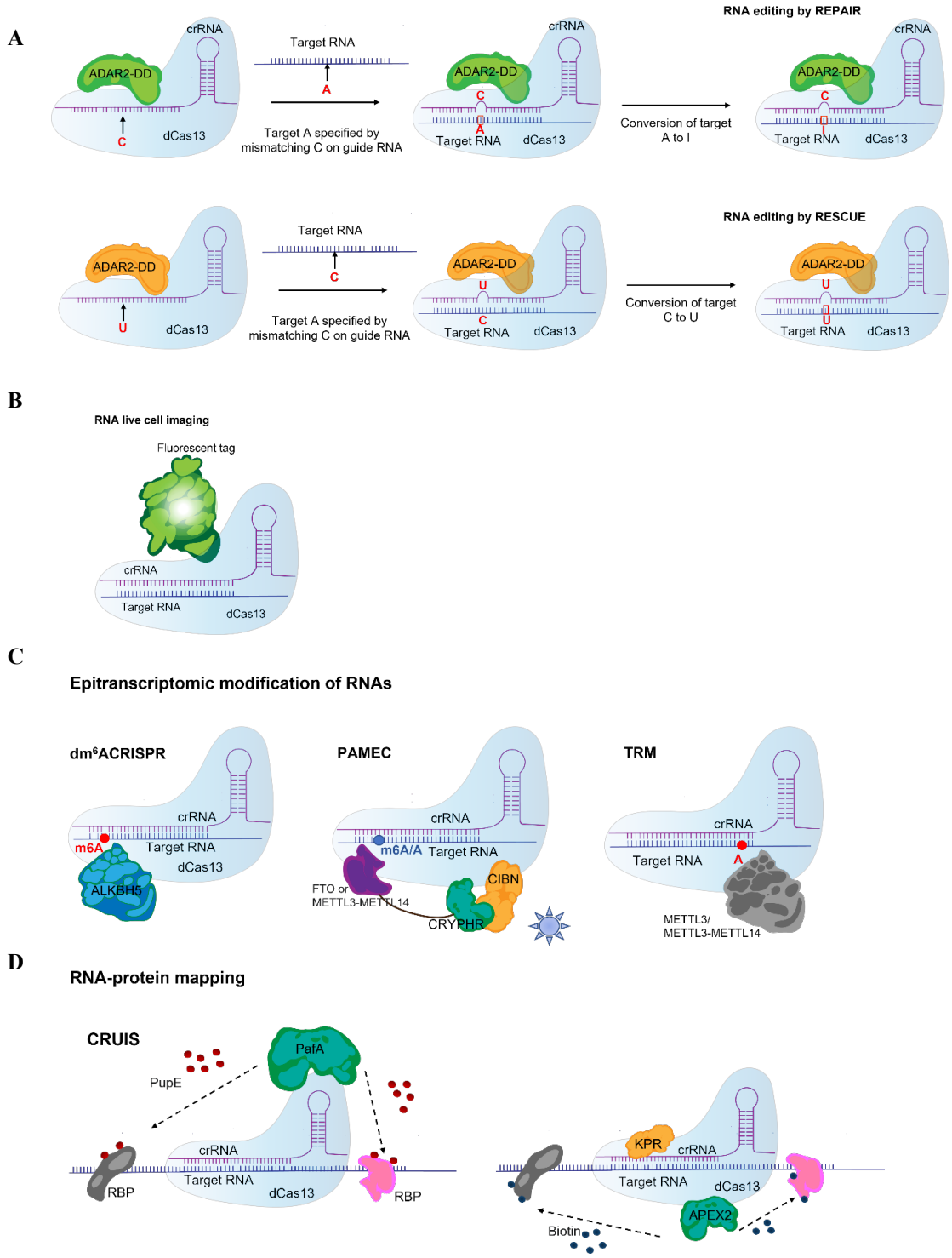


Figure 1. 9. Applications of catalytically inactive Cas13 (dCas13).

(A) dCas13 proteins can be used for RNA base editing through its fusion with various deaminase domains of human ADAR (ADAR-DD). In REPAIR systems, a target-specific crRNA directs the dCas13b–ADAR-DD fusion protein to a specific region of the targeted transcript. A mismatched cytidine (C) in the crRNA opposite the target adenosine (A) results in target adenosine deamination to inosine (A-to-I editing). In RESCUE systems, a mismatched uridine (U) in the crRNA opposite the

target cytosine (C) results in -C to-U editing. (B) dCas13b fused to fluorescent proteins loaded with a target-specific crRNA can be used to track RNA dynamics in live cells with high spatiotemporal resolution. (C) dCas13 fusion with other RNA editing enzymes (e.g., m^6A demethylase AlkB homolog 5 (ALKBH5), m^6A demethylase fat-mass and obesity-associated protein (FTO) and METTL3-METTL14 m^6A methyltransferase) enables site-specific epitranscriptomic RNA modifications. (D) dCas13 fusion with effectors that can ubiquitinate (e.g., proteasomal accessory factor A (PafA)) or biotinylate (e.g., the modified plant peroxidase APEX2) proteins at the proximity of dCas13-crRNA-target ternary complex enabling the mapping of proteins that interacts with a specific RNA molecule.

1.6 Therapeutic potentials of CRISPR-Cas13

Target recognition and cleavage by Cas13 is based on the complementary base-pairing between crRNAs and target sequences, which offers design flexibility to target various pathogenic RNAs. With the recent advances in sequencing technologies that are currently incorporated into standard-of-care protocols, the genetic aberrations that drive human pathologies can be rapidly identified in each patient. With the sequence information in hand and the design flexibility of Cas13, it became feasible in principle to develop personalized Cas13-based drugs designed specifically to counter the genetic aberrations that drive various human diseases[58]. Importantly, such RNA-targeting precision medicine programs that leverage a patient's genetic information may enable the rapid development of personalized drugs in an actionable time frame, outperforming classical drug development pipelines that are slow and expensive[33,39].

1.6.1 Precise and personalized therapeutics

In the last decade, a remarkable breakthrough was achieved through the development of various CRISPR-Cas technologies for precise manipulations of the genetic code, which has changed our perspectives on future diagnosis and therapeutics for mutation-driven diseases[112-115]. For example, CRISPR-Cas9 has shown encouraging results in a recent clinical trial in treating patients suffering from sickle cell disease that is caused by a single mutation in the beta-globin gene resulting in misfolded haemoglobin and distorted red blood cells[116]. Hematopoietic stem cells were isolated from the patient, *ex vivo* edited with CRISPR-Cas9 to produce high levels of foetal haemoglobin in red blood cells and re-infused into the patient's circulation. Preliminary data of this CRISPR gene therapy have shown that 46% of the haemoglobin in the patient are edited foetal haemoglobin, which was expressed in 94% of the red blood cells[117]. While *ex vivo* editing of patient's cells appears to be successful, targeting disease-causing genes by *in vivo* delivery of CRISPR-Cas9 is challenging for several reasons[118,119]. The irreversible DNA alterations by Cas9 have given rise to major safety concerns[120]. The two major limitations of CRISPR-Cas9 are off-target effects and chromosome instability due to Cas9-induced

double-stranded DNA breaks[121]. Although Cas9 was expected to have tightly regulated nuclease activity by the 20-nt sgRNA, several studies have shown that 3 to 5 mismatches can be tolerated between sgRNA and the target in the PAM-distal regions[41,122-124], which gives rise to far higher more off-target events than previously anticipated. In addition, Cas9-mediated on-target and off-target DSB can cause unpredictable chromosomal rearrangements and megabase-scale chromosomal truncations[125]. Given the permanent gene disruption by Cas9, these side-effects can lead to loss of genome integrity with risks outweighing the benefits [118,119]. The ethical issue of intentional or unintentional editing of the inherited human genome with DNA editors further obstructs the likely wider use of these technologies[126,127]. Given the reasons discussed above, there is a consensus among the scientific community to refrain from using classical Cas9 in patients until the development of other DNA-editing tools that have higher specificity and safety profiles.

The recent breakthrough of CRISPR-Cas13 may offer a reversible, safer, and more specific approach for *in vivo* targeting of disease-causing genes at the RNA level without altering the genomic DNA[22,40,128]. Human diseases can be caused by various genetic aberrations, including point mutations of a single base, gene duplication, insertions, deletions of DNA segments, and chromosomal rearrangements, resulting in disease-specific transcripts that would be ideal targets for Cas13[129,130]. Principally, the crRNAs of Cas13 can be engineered to target any aberrant pathogenic RNA[29,58,131]. For example, a proof-of-concept study demonstrated that *Psp*Cas13b could be reengineered to cleave RNAs that were transcribed from an aberrant expansion of trinucleotide repeat sequences[132]. In their study, the team chose to target CUG repeat RNAs resulting from a noncoding CTG tandem repeat in the 3' UTR of the human dystrophia myotonic protein kinase gene (*DMPK*). These extensive RNAs misfold and aggregate to form toxic nuclear foci, causing myotonic dystrophy type 1[132]. They demonstrated that *Psp*Cas13b containing an N-terminal NLS induced a significant reduction of RNA nuclear foci labelled by m-Cherry in COS7 cells[132].

CRISPR-Cas13 may represent an exciting platform for personalized precision medicine for cancer treatment by targeting tumour-specific mutations[58,133]. A pilot study found that reprogrammed *Lwa*Cas13a was able to knock down 94% of KRAS mRNA harbouring a G12D point mutation in AsPC-1 pancreatic cancer cells. However, this crRNA also showed a 60% off-target knockdown of the wildtype *KRAS* transcripts in HEK293T cells. Importantly, a single mismatch in this crRNA at position 14 could exhibit 70% *KRAS*-G12D transcript knockdown while showing no off-target effect on wildtype *KRAS* transcripts, which suggested that crRNAs could be designed to degrade tumour transcripts harbouring single-nucleotide mutations while sparing the wild-type transcripts that share extensive sequence homology[133].

The proprotein convertase subtilisin/kexin type (PCSK9) binds and induces degradation of the low-density lipoprotein receptor. The gain-of-function mutation of PCSK9 has been recently

discovered as the third factor associated with hypercholesterolemia[134]. A study demonstrated that *Rfx*Cas13d could knock down transcripts of metabolism regulatory genes including *Pcsk9*, phosphatase and tensin homolog (*Pten)*, lncRNA *LncLstr in mouse hepatocytes.* Furthermore, the AAV-mediated delivery of *RfxCas13d* and *Pcsk9*-targeting crRNAs into mouse liver successfully decreased the level of PCSK9 in the serum, resulting in a significant reduction of cholesterol levels in the serum of mice as a potential treatment of this metabolic diseases[100]. Similarly, *Rfx*Cas13d has been used to silence the polypyrimidine tract-binding protein 1 (*Ptbp1*) transcript, resulting in the conversion of Müller glia into retinal ganglion cells (RGCs) *in vivo*. Further *Ptbp1* knockdown through AAV-mediated delivery alleviated motor defects in a Parkinson's disease mouse model[135].

Apart from active Cas13 effectors, re-engineered dCas13 such as REPAIR and RESCUE can potentially be used for therapeutic applications to correct SNP-driven diseases[29]. Remarkably, two tumour-driven transcripts encoding epidermal growth factor receptor (EGFR) and MYC positively regulate cell proliferation by m^6A posttranscriptional modification, whereas dm^6ACRISPR (d*Psp*Cas13b-ALKBH5 fusion protein)-mediated demethylation of EGFR and MYC transcripts suppresses cancer cell proliferation[106].

1.6.2 Cas13 Anti-viral therapeutics

Conventional antiviral drugs are made of small inhibitory molecules or recombinant antibodies that target viral proteins[136]. The current drug development process is slow and cannot meet the urgent need for efficient and specific antivirals when facing pandemics or viral outbreaks such as SARS-CoV-2, Zika, and Ebola, which warrants the development of a multiplexable and fast-adapting antiviral platform[137,138]. Cas13 naturally functions as an adaptive immune system in bacteria to silence viral RNA molecules, indicating its potential as an anti-viral tool in human cells[31]. Importantly, Cas13 can process its pre-crRNA molecules that are transcribed from the CRISPR array and encodes various spacer sequences. Leveraging this crRNA maturation mechanism opens the possibility of multiplexing to target multiple regions on viral RNAs simultaneously to increase efficacy and neutralise the notorious mutation-driven viral escape[27,45].

Several studies have explored the RNA-targeting properties of Cas13 to treat viral infections. Kou's Lab has employed *Lwa*Cas13a to target the non-structural protein 3 (NS3) region of the Dengue viral genomic RNA, which efficiently inhibited viral replication[135]. Abbott *et al*. developed a *Rfx*Cas13d-mediated strategy, named Prophylactic Antiviral CRISPR in Human cells (PAC-MAN) to target SARS-CoV-2 and influenza A virus (IAV). They defined a group of conserved regions between the SARS-CoV-2 genomes from 47 patients and synthesized 40 crRNAs in these regions. They demonstrated that PAC-MAN could inhibit coronaviral genomic fragment expression and reduce the replication of IAV in A549 human lung epithelial cells[138]. In this study, they used a bioinformatic approach to design a minimum number of 6 crRNAs that target approximately 90% of known

coronavirus genomes (2776). By combing multiple crRNAs targeting conserved regions, it is feasible to cover most of the strains[138]. The targeting of highly conserved sequences can also prevent mutation-driven viral escape, which may offer an advantage over conventional vaccines, recombinant antibodies, and small inhibitory molecules. However, they did not assess the activity of Cas13 against infectious SARS-CoV-2 in live cells[138]. In another study, Blanchard *et al*. delivered polymer-formulated *Lbu*Cas13a mRNA and crRNA targeting influenza virus polymerase, SARS-CoV-2 nucleocapsid to the respiratory tract using a nebulizer. *Lbu*Cas13a degraded influenza RNA in lung tissue efficiently in mice and reduced SARS-CoV-2 replication and symptoms in hamsters[137]. More recently, our lab demonstrated that reprogrammed *Psp*Cas13b effectors and crRNAs targeting the spike, RNA-dependent RNA polymerase subunits, and nucleocapsid transcripts suppressed viral replication in mammalian cells infected with ancestral SARS-CoV-2 (Wuhan strain) and its emerging variant of concern B.1.1.7. Interestingly, the comprehensive mutagenesis of crRNA-target RNA interaction revealed that single-nucleotide mismatches do not impair the potency of single crRNA, therefore this mismatch tolerance capability of *Psp*Cas13b can circumvent mutation-driven viral escape commonly seen in SARS-CoV-2 and other viruses. We demonstrated that a single crRNA can simultaneously suppress ancestral and mutated SARS-CoV-2 strains in infected mammalian cells, including the Spike D614G mutant, even if the mutation occurs within the targeted region[139].

1.7 Conclusion

Compared to the previous technologies such as CRISPR-Cas9 and RNAi, CRISPR-Cas13 is a multiplexable and robust toolkit for RNA-targeting with considerably reduced safety concerns regarding off-target effects, which provides tremendous opportunities for therapeutic RNA manipulations in a precise manner. Since the discovery of CRISPR-Cas13, there have been a great interest from the scientific community in these new tools, which revealed parts of the molecular basis of CRISPR-Cas13 (**Table 1.1**). However, given the recent discovery of new CRISPR orthologues with distinct sequences, structures, and action mechanisms, exciting new questions have emerged around their reprogramming rules, efficiency, specificity, and collateral activity in mammalian cells, which require deeper investigations. Therefore, in the next decade CRISPR researchers are anticipated to elucidate the underlying basis of target recognition process of various Cas13 orthologues and use this knowledge for the rational design of efficient reprogramming of Cas13. Ultimately, future research will unravel design principles of Cas13 tools with higher efficiency, specificity, and no collateral activity that may transform tomorrow's precision medicine and personalised oncology.

Table 1. 1. Molecular characterized CRISPR-Cas13 effectors

Effector	Species	Size (aa)	Spacer (nt)	Direct repeat (nt)	Seed regions	PFS	Collateral activity in mammalian cells	Crystal structure (PDB ID)	Applications	Refs
*Lwa*Cas13a	*Leptotrichia wadei*	1182	28 (in vitro); 29-30 native	36	13-24	Weak 3′ H (A/C/U) preference	Low collateral activities in HEK293T cells with target transiently overexpressed; undetectable collateral activity for endogenous RNA knockdown in HEK293 T cells; Detectable collateral activity in U87 glioblastoma cells when targeting stably expressed EGFP	NA	RNA knockdown in mammalian cells; SHERLOCK (V2); Live cell RNA tracking; RNA-protein mapping	28,39,53, 64,90,100
*Lbu*Cas13a	*Leptotrichia buccalis*	1159	28	31	9-15	3'C (-1)' is unavailable to base pair with crRNA	Unclear	5XWY (binary complex) /5XWP (ternary complex)	Anti-viral activities in hamster	23,47,55, 137
*Lsh*Cas13a	*Leptotrichia shahii*	1389	28	28	Central (further mutagenesis study required)	3' H (A/C/U) preference	Unclear	5WTJ (Apo protein) /5WTK (Binary complex)	NA	24,35,46
*Lba*Cas13a	*Lachnospiraceae bacterium*	1437	24	28	Unclear	Unclear	Unclear	5W1I/5W1H /5WLH (Binary complex)	NA	48

*Lse*Cas13a	*Listeria seeligeri*	1120	15-18	29	Unclear	Unclear	Unclear	Unclear	NA	[22,39]
Hhe Cas13a	*Herbinix hemicellulosilytica*	1285	Unclear	32	Unclear	Unclear	Unclear	Unclear	NA	[35]
*Bz*Cas13b	*Bergeyella zoohelcum*	1224	22	36	Unclear	5′-recognition of D (G/U/A); 3′-motif requirement of NAN or NNA	Unclear	6AAY (binary complex)	NA	[26,51]
*Pbu*Cas13b	*Prevotella buccae*	1127	26	36	12-17	5' D (G/U/A) preference; slight 3' AA preference	Unclear	6DTD (binary complex)	NA	[26,29,44]
*Psp*Cas13b	*Prevotella sp. P5–125*	1133	30	36	12-26	Not required	The collateral activity was detected in mouse ES cells, and U87 cells but not HEK293 FT cells when targeting transiently overexpressed luciferase reporter constructs	Unclear	REPAIR; REPAIRv2; Live cell RNA imaging; RNA knockdown in mammalian cells; Antiviral activity in mammalian cells; RNA epitranscriptomic modification	[29,65,75,108,139]
*Rfx*Cas13d (CasRx)	*Ruminococcus flavefaci*	972	22	30	15-21	Not required	Detectable collateral activities in HEK293T	Unclear	Live cell RNA imaging (LiveFISH);	[28,30,37,59,65,66,]

	ens XPD3002						cells with transient target overexpression; undetectable collateral activity for endogenous RNA knockdown in HEK293T cells Relatively higher collateral activity in HepG2 cells and U87 cells with transient target overexpression		RNA knockdown in mammalian cells; dRfxCas13d-ADAR2-DD (E488Q/T375G) RNA editor; RNA-protein mapping; Antiviral activity in mammalian cells	84,88,100,138
Cas13bt.1	*NA*	804	30	36	Unclear	Weak 5′ D (A/G/U) preference in *E.coli*; not required mammalian cells	Unclear	Unclear	RESCUE.t1; REPAIR.t1; RNA knockdown in mammalian cells	36
Cas13bt.3	*NA*	775	30	36	Unclear	Weak 5′ D (A/G/U) preference in *E.coli*; not required in mammalian cells	Unclear	Unclear	RESCUE.t3; REPAIR.t3; RNA knockdown in mammalian cells	36
*Es*Cas13d	*Eubacterium siraeum*	954	23 (Most common); 20-30	30	Investigated, but lack of a clear seed	Not required	Unclear	NDB:6E9E/6E9F Apo protein, binary, and	NA	27,45,58

			(native)					ternary complex		
*Rsp*Cas13d (*Ur*Cas13d)	*Ruminococcus sp.*	922	20 (in vitro); 31 (native)	30	5-8;13-22	Not required	Unclear	6IV9 binary complex	NA	[27,49,58]
Cas13x.1	Uncultivated microbes	775	30	36	Unclear	Not required in mammalian cells	Low collateral activities in mammalian (HEK293T) cells with transient target overexpression; undetectable collateral activity for endogenous RNA knockdown	Unclear	RNA knockdown in the mammalian cells; Antiviral activity in mammalian cells; compact RNA editors (truncated dCas13X.1 fused with ADAR2-DD variants)	[28]
Cas13y.1	Uncultivated microbes	790	30	36	Unclear	Unclear	Unclear	Unclear	RNA knockdown in mammalian cells	[28]

1.8 Aims and hypotheses

Recent advances in next-generation sequencing enable rapid identification of oncogenic transcripts in individual cancer patients within actionable time frames[140-142]. However, many such driver mutations cannot be targeted due to the lack of specific inhibitory molecules[143,144]. For example, numerous fusion genes generated by chromosomal translocations demonstrate cogent oncogenic activity but also remain largely '*undruggable*' through conventional small molecule drug-based therapeutic approaches[145]. Personalized targeting of these fusion structural variants at the protein level has proven to be challenging[146], and in the rare cases where small molecule inhibitors are devised, treatment often rapidly results in drug resistance and disease relapse[147]. As an alternative, my thesis aimed to test whether the unique chimeric sequence generated at the breakpoint/junction of two genes at the transcript level might represent a tractable and impactful target for sequence-specific silencing with

programmable RNA nucleases.

In terms of feasibility, the programmable RNA-guided targeting offered by CRIPSR-Cas13 exclusively degrades ssRNAs with high efficacy and specificity[22,24], while recent studies have deployed Cas13 in a variety of targeted RNA manipulations including, nucleic-acid detection[23,53] and precise RNA base editing[39,108], and viral suppression[31,139]. The efficiency and reversibility of RNA targeting with Cas13 potentially offer a way to specifically edit oncogenes without risking permanent alteration of the genome in somatic and germline cells, an inherent limitation of DNA-editing CRISPR enzymes[148,149]. In particular, Cas13 is highly attractive for targeting aberrant transcripts arising from gene fusions that drive various human genetic diseases including cancer, as the breakpoint region represents a region of extensive divergence from the two wildtype genes from which it was formed. Among all the characterised Cas13 orthologues, the *Psp*Cas13b appears promising as it possesses low collateral activity, high silencing efficiency, and specificity compared to other orthologues such as *Rfx*Cas13d, underscoring its potential for targeted gene silencing in human cells[68].

A hypothesis central to this body of work is that the CRISPR-Cas13 nucleoprotein complex may be re-engineered to silence oncogenic tumour drivers with single-nucleotide accuracy. However, the poor understanding of the molecular principles governing *Psp*Cas13b target recognition and cleavage has to this point limited its utility and development as programmable cancer therapeutics. To address this need, my PhD research focuses on revealing the molecular mechanisms underpinning the activity of *Psp*Cas13b in order to reprogram this enzyme to target tumour transcripts that are currently 'undruggable'. Innovative approaches used in my study include the development of tiled crRNA library screens and crRNA mutagenesis studies to achieve the following two major aims:

Aim 1: Understand the molecular basis of CRISPR Cas13 nuclease activity, including crRNA design rules to ensure the efficient design of crRNAs.

Aim 2: Equipped with this knowledge, reprogram the *Psp*Cas13b-crRNA complex to silence tumour fusion transcripts and reverse the phenotypic feature of cancer cells *in vitro*.

Aim 3: Investigate the specificity of CRISPR-*Psp*Cas13b nuclease activity in mammalian cells

Achievement of these aims would be a cornerstone toward the development of a novel RNA-based therapeutic platform to silence tumour transcripts in a personalized manner.

Chapter 2

Rules and principles that govern RNA recognition and silencing by CRISPR-*Psp*Cas13b

Abstract

CRISPR-Cas13 is a newly discovered member of the CRISPR-Cas family of bacterial nucleases. The target recognition process employed by *Psp*Cas13b remains poorly understood, limiting our ability to rationally design efficient crRNAs that can target and degrade oncogenic RNAs in mammalian cells. To address this deficiency, a comprehensive study was undertaken to understand the molecular mechanism of target recognition used by the *Psp*Cas13b-crRNA complex, with the overall aim of understanding the molecular 'rules' that govern crRNA loading, target recognition, and target cleavage. The mcherry mRNA was first targeted as a model system, as it provides an accurate, rapid, and easily quantifiable readout of silencing efficiency. Using 61 single-nucleotide tiled crRNAs targeting a defined region of mcherry mRNA, crRNAs that showed a range of silencing efficiency ranging from very low (<2%) to very high (>98%) were identified. Similarly, variable silencing efficiency was also observed when targeting other transcripts such as the Spike (SARS-CoV-2) and BCR-ABL1 transcripts with single-nucleotide tiled crRNAs. To explore the molecular basis for this variable efficiency, a bioinformatics analysis was performed for >200 crRNAs to establish the crRNA sequence parameters that governed *Psp*Cas13b-crRNA complex-mediated targeting and mRNA degradation. A comparison of the potent and non-potent crRNAs revealed that a PFS is not required for efficient *Psp*Cas13b silencing. However, the presence of a 5' 'GG' was consistently associated with far greater mRNA silencing, whereas a 'poly-C' motif at the 5' and near the centre of the spacer resulted in poor *Psp*Cas13b crRNA silencing. Surprisingly, introducing a 5' GG into otherwise ineffective crRNAs restored their silencing efficiency in many instances, even when complementarity to the target sequence was reduced at this location. Spacer nucleotide preferences observed in the above screens were validated by using the findings to design crRNAs targeting eGFP and tagBFP, and *in-vitro* transcribed crRNA in HEK 293T cells. Incorporating these features resulted in the design of predictably highly potent crRNAs for the different mRNA transcripts. To further the impact of these discoveries, an open-access and user-friendly algorithm (https://cas13b.github.io/) was developed to assist the wider scientific community with the design of predictably potent crRNAs for use with *Psp*Cas13b.

2.1 Introduction

The type VI CRISPR-Cas13 effectors are programmable RNA-guided targeting enzymes that exclusively degrade single-stranded RNAs (ssRNAs) with high efficacy and specificity[22,40]. Recent studies have deployed Cas13 systems in a variety of targeted RNA manipulations including, nucleic-acid detection[23,53,90], precise RNA base editing[75,108], and viral suppression[31,137,139]. The efficiency and reversibility of RNA targeting with Cas13 represent a promising modality to specifically edit oncogenes without risking permanent alteration of the genome in somatic and germline cells, an inherent limitation of DNA-editing CRISPR enzymes[31,125,149]. Therefore, Cas13 is highly attractive for targeting aberrant fusion transcripts that drive various human genetic diseases including cancer.

Currently, the specifics of target recognition employed by Cas13 are poorly understood, limiting the rational design of efficient crRNAs targeting tumour-associated aberrant RNAs. One significant factor limiting efficacy may be the accessibility of highly structured regions on target RNAs as Cas13 lacks a helicase domain to unwind double-stranded RNA motifs and allow for crRNA binding[58]. A screen targeting essential genes in *E.coli* indicated that spacers of *Psp*Cas13b with a secondary structure near the protospacer had reduced targeting efficiency[26]. Similarly, crRNAs targeting regions with low secondary structure showed better knockdown in a pool of crRNAs tiled along the target transcript sequence[39]. One study carried out four pooled screens targeting GFP and three cell surface markers CD46, CD55, and CD71 by *Rfx*Cas13d with tiled crRNAs to systematically investigate the targeting rules for optimal guide design (https://cas13design.nygenome.org/). The study determined that the target RNA context may constrain target knockdown efficacy[59]. More recently, Wei *et al*. performed a larger positive selection screen to investigate *Rfx*Cas13d targeting rules by tiling 55 previously identified essential genes with over 127,000 crRNAs at single-nucleotide resolution. By deep machine learning of the large-scale dataset with the incorporation of parameters including secondary structure, free energy, target site position, and target isoform percent, the team developed a model to predict highly effective crRNAs with >90% accuracy and released a crRNA design tool (http://RNAtargeting.org)[150].

However, the targeting rules for other type VI effectors including *Psp*Cas13b remain unaddressed. A recent study demonstrated that *Rfx*Cas13d exhibited collateral activity in mammalian cells, as knocking down gene expression at the mRNA level using *Rfx*Cas13d in adult brain neurons caused the death of mice[151]. Importantly, a previous study suggested that the *Psp*Cas13b orthologue appears to possess higher silencing efficiency and specificity, and lower collateral activity compared to *Rfx*Cas13d, underscoring its potential suitability for targeted gene silencing in human cells[152]. Consequently, the experiments in this Chapter aim to correct the poor understanding of the molecular principles governing *Psp*Cas13b target recognition and cleavage, with the ultimate aim of developing this tool for highly selective silencing targeting of oncogenic tumour transcripts.

2.2 Materials and Methods

2.2.1 Design and cloning of crRNAs for *Psp*Cas13b

The design and cloning of *Psp*Cas13b crRNAs were as previously described[29]. Briefly, individual guide RNAs were cloned into the pC0043-*Psp*Cas13b crRNA backbone (addgene#103854, a gift from Feng Zhang lab, later referred to as crRNA backbone) which contains a *Psp*Cas13b crRNA direct repeat (DR) sequence and two BbsI restriction sites for the cloning of spacer sequences. A total of 20 µg crRNA backbone was digested to completion with BbsI (NEB, R3539) following the manufacturer's instructions for 2 hours at 37C°. Backbone linearization was checked with 1% agarose gel. The digested

DNA was purified with NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, 740609.50), aliquoted, and stored at -20C° until required.

For crRNA cloning, forward and reverse single-stranded DNA oligonucleotides containing CACC and CAAC overhangs respectively, were ordered from Sigma or IDT (Integrated DNA Technologies). 1.5 µL of each of the forward and reverse DNA oligonucleotides (each at 100 µM) were annealed in 47 µL annealing buffer (5 µl NEB buffer 3.1 and 42 µL H_2O) by 5 min incubation at 95 °C and slowly cool down in the heating block overnight. 1 µL of the annealed oligonucleotides were ligated with 0.04 ng digested *Psp*Cas13b crRNA backbone in 10 µL of T4 ligation buffer (3 h, RT) (Promega, M1801). All *Psp*Cas13b crRNA spacer sequences used in this study are listed in **Appendix Table 1**. All crRNAs and *Psp*Cas13b clones generated in this study were verified by Sanger sequencing (AGRF, AUSTRALIA). The primers used for PCR and Sanger sequencing are listed in **Appendix Table 3**.

2.2.2 Plasmid amplification and purification

Plasmid amplification and purification were as described previously[139]. Briefly, TOP10 or Stbl3 bacteria were used for transformation. Chemically-competent bacteria (30uL) were transformed by heat shock at 42 °C for 45 s with 5–10 µL ligated plasmid DNA, followed by 2 min on ice. The transformed bacteria were incubated in 500 µL LB medium containing 75 µg/mL ampicillin (Sigma-Aldrich, A9393) for 1 h at 37 °C in a shaking incubator (200 rpm). The bacteria were pelleted by centrifugation at 6,000 rpm for 1 min at room temperature (RT), re-suspended in 100 µL of LB, plated onto a pre-warmed 10 cm LB agar plate containing 75 µg/mL ampicillin, and incubated at 37 °C overnight. The next day, single colonies were picked and transferred into bacterial starter cultures and incubated for ~6 h for mini-prep (Macherey-Nagel, NucleoSpin Plasmid Mini kit for plasmid DNA, 740588.50) or maxi-prep (Macherey-Nagel**,** NucleoBond Xtra Maxi Plus, 740416.50) DNA purification according to the standard manufacturer's protocol.

2.2.3 Cell culture

HEK 293 T (ATCC CRL-3216) cell line was cultured in DMEM high glucose media (Thermo Fisher, 11965092) containing 10% heat-inactivated fetal bovine serum (FBS, Thermo Fisher, 10100147), 100mg/ml Penicillin/-Streptomycin (Thermo Fisher, 151401220), and 2mM GlutaMAX (Thermo Fisher, A1286001). HEK 293 T cells were maintained at confluency between 20 and 80% in 37 °C incubators with 10% CO_2. Cells were routinely tested and were negative for Mycoplasma.

2.2.4 crRNA *in-vitro* transcription

The DNA template for crRNA in-vitro transcription was manufactured by IDT as complementary single-stranded DNA oligonucleotides (100 µM) encoding the TAATACGACTCACTATA T7 promoter, followed by the spacer and direct repeat sequences. A total of 1.5 µL of the forward and reverse DNA oligonucleotides (100 µM) were added to 47 µL annealing buffer (5 µL NEB buffer 3.1 and 42 µL H_2O). The two strands were first denatured for 5 min at 95 °C followed by a gradual cool

down in the heating block overnight for annealing. 1.5 µL of the annealed DNA was used as a template for in-vitro transcription using the mMESSAGE mMACHINE T7 Transcription Kit (Thermo Fisher, AM1344) according to the manufacturer's instructions. We performed a 6-hour incubation to maximize the yield due to small DNA templates, followed by DNase treatment. In-vitro transcribed crRNAs were purified with Monarch RNA Clean-up kits (NEB, T2040L) to remove enzymes and unincorporated nucleotides according to the manufacturer's instructions. The purified crRNAs were quantified with NanoDrop 2000/2000c Spectrophotometers (Thermo Fisher, ND-2000), and crRNA length/integrity was evaluated using 1% agarose gel electrophoresis.

2.2.5 RNA silencing assays using plasmid transfections

Gene transfections were performed using an optimized Lipofectamine 3000 transfection protocol (Thermo Fisher, L3000015). For RNA silencing in HEK 293T, cells were plated at approximately 30,000 cells/100 µL/96-well in tissue culture-treated flat-bottom 96-well plates (Corning) 18 h prior to transfection. For each well, a total of 100 ng DNA plasmids (22 ng of *Psp*Cas13b-NES-3xFLAG-T2A-BFP (addgene #173029) or pC0046-EF1a-*Psp*Cas13b-NES-HIV (addgene #103862), 22 ng crRNA plasmid, and 56 ng of the target gene plasmid) were mixed with 0.2 µL P3000 reagent in Opti-MEM Serum-free Medium (Thermo Fisher, 31985070) to a total of 5 µL (mix1). Separately, 4.7 µL of Opti-MEM was mixed with 0.3 µL Lipofectamine 3000 (mix2). Mix1 and mix2 were added together and incubated for 20 min at room temperature, then 10 µL of transfection mixture was added to each well. **Appendix Table 4** summarizes the transfection conditions used in 96, 24 and 12-well plates. After transfection, cells were incubated at 37C°, 10% CO_2, and the transfection efficiency was monitored 24-72 hours post-transfection by fluorescence microscopy.

2.2.6 RNA silencing assays using *in-vitro* transcribed (IVT) crRNAs

Similar transfection conditions were used except that crRNA plasmid was substituted with 200 ng of IVT crRNA of interest. The choice of 200 ng of IVT crRNA was based on prior optimization assays where 200 ng IVT crRNA achieved silencing.

2.2.7 Fluorescence microscopy

For RNA silencing experiments, fluorescence intensity was monitored using an EVOS M5000 FL Cell Imaging System (Thermo Fisher). Images were taken 48 h post-transfection, and the fluorescence intensity of each image was quantified using a lab-written macro in ImageJ software. Briefly, all images obtained from a single experiment are simultaneously processed using a batch mode macro. First, images were converted to 8-bit, threshold adjusted, then converted to black and white using the Convert to Mask function, and fluorescence intensity per pixel was measured using the Analyze Particles function. Each single mean fluorescence intensity was obtained from four different fields of view for each crRNA and subsequently normalized to the non-targeting (NT) control crRNA. A 50% or greater reduction in fluorescence intensity was considered biologically relevant.

2.2.8 RNA extraction, cDNA synthesis, and RT-PCR

Total RNA was isolated from $5x10^5$ to $1x10^6$ cells using the NucleoSpin RNA Plus (MACHEREY-NAGEL, 740984.50) or *Quick*-RNA Miniprep Kit (*Zymo* Research, R1055) following the manufacturer's instructions. For crRNA RT-PCR, total RNA was isolated by standard Trizol-chloroform extraction according to manufacturer's instruction (Thermo Fisher, 15596026), followed by Dnase treatment with RQ1 RNase-Free DNase according to manufacturer's instruction (Promega, M6101). Comparison of DNase treated and untreated RNA samples using RT-QPCR confirmed that DNAse treatment removed contaminant DNA plasmids (data not shown).

1µg of total RNA was converted to cDNA using the high-capacity cDNA reverse transcription kit (*Thermo Fisher,* 4368814) following the manufacturer's instructions. Quantitative RT-PCR reaction was performed in duplicate in a StepOne Real-Time PCR system (*Thermo Fisher*) using PowerUp™ SYBR™ Green Master Mix (*Thermo Fisher*, A25742). The total reaction mixture contains 0.2µl cDNA, 0.6µM forward primer, and 0.6 µM reverse primer. Primers for RT-PCR are detailed in **Appendix Table 3**.

2.2.9 Prediction of RNA secondary structure, RNA MFE, and RNA-RNA hybridization energy

RNAfold was used to predict the MFE of crRNA spacer, crRNA (DR and spacer), and the 70nt target region in the target RNA (20nt up/downstream from the 30nt-spacer-binding region). RNAfold was also used to predict the secondary structure of crRNAs and the target regions in the target RNAs. RNAplex[153] and intaRNA[154] were used to predict the hybridization energy and interaction energy between crRNA spacer and target RNA, respectively.

2.2.10 *Psp*Cas13b crRNA design tool

The design of *in silico* prediction code was based on the experimental data presented below. Briefly, the algorithm generates single-nucleotide tiled spacer sequences for any input DNA or RNA sequence using HTML5, CSS, and other JavaScript libraries. The program then removes all spacer sequences that possess more than three consecutive T bases (>3T) that are predicted to act as a transcription termination signal and yield premature crRNAs. The algorithm scores the remaining spacer sequences based on their nucleotide composition and position. Spacers with a G nucleotide at the first or second positions receive a maximum score of +20 each. In contrast, a C nucleotide at spacer position 1, 2, 3, or 4 receives a penalty score of -20 each. Additionally, C bases at positions 11, 12, 15, 16, and 17 receive a -5 score each. All other nucleotides or spacer positions that did not show any distribution bias in the potent and ineffective crRNA receive a score of 0. The algorithm then calculates the cumulative score for each spacer and ranks them accordingly. As a result, the top spacers with high scores are enriched with G bases at 1st and 2nd positions, and depleted from C bases at positions 1, 2, 3, 4, 11, 12, 15, 16, and 17, and are predicted to yield potent silencing. Conversely, the lowest-scoring spacers at the bottom of the

list are enriched with C bases at positions 1, 2, 3, 4, 11, 12, 15, 16, and 17 and are predicted to yield ineffective silencing. The prediction accuracy of the algorithm is supported by *in silico* analysis and functional validation data in **Figures 2.4-Figure 2.10**.

To predict crRNAs that would be potent against a full-length transcript, the full target sequence was used for prediction using the design tool. For crRNAs detecting the breakpoint of fusion transcripts, a 50-nucleotide sequence spanning the breakpoint (25-nucleotide matching each partner gene) was used for prediction to avoid potential off-targeting of untranslocated variants. This *Psp*Cas13b crRNA design tool is open-source and available to the wider scientific community at https://cas13b.github.io/.

The prediction power of this *in-silico* tool is dependent on the number of available targeted sites per transcript. For instance, a 1000-nucleotide long transcript provides ~970 binding sites (with full spacer-target basepairing) and crRNA design possibilities, from which the algorithm can choose spacer sequences that highly satisfy the criteria of potent crRNA. However, targeting the breakpoint of a gene fusion restricts the design to only ~50 crRNAs that can exclusively base pair with the breakpoint sequence. This restriction in crRNA design to exclusively base pair with the breakpoint is critical to avoid off-targeting of untranslocated partners (e.g., BCR or ABL1). Therefore, from a probability perspective, the likelihood of finding highly potent crRNAs targeting the breakpoint of gene fusion mRNAs is lower than when targeting other long transcripts without a restricted binding site. Therefore, we anticipate that the targeting prediction is less accurate when designing crRNAs targeting the breakpoint of gene fusion transcripts. Nonetheless, this *in-silico* tool can maximize the probability of finding the most potent crRNAs among a subset of possible crRNAs that can base pair with the breakpoint of a fusion gene. In addition, the user can substitute the 1st and 2nd nucleotides of the spacer with G bases if higher potency is required.

2.2.11 Data analysis

Data analyses and visualizations (graphs) were performed in GraphPad Prism software version 9 unless stated otherwise. Specific statistical tests and numbers of independent biological replicates are mentioned in respective figure legends. The silencing efficiency of various crRNAs was analyzed using one-way ANOVA followed by Dunnett's multiple comparison tests where we compare every mean to a control mean as indicated in the Figures (95% confidence interval). The *P* values (P) are indicated in the Figures. $P < 0.05$ is considered statistically significant. Pearson correlation coefficient was used to analyze the correlation between the crRNA silencing efficiency and potential parameters including crRNA MFE, target MFE, crRNA spacer MFE, crRNA-target RNA hybridization/interaction energy, crRNA spacer GC content, and A/U/G/C content. The R package 'ggseqlogo' was used to assess nucleotide preference in crRNA spacer and PFS sequences[155]. Delta probability graphs of spacer nucleotides were generated with Matplotlib[156]. Delta probability is the change of frequency of the base

at each spacer nucleotide position in the potent and ineffective groups compared to the unfiltered crRNAs.

2.2.12 Code Availability

The bioinformatics codes for the design of predicted potent crRNA are available at https://github.com/david-ma/cas13. crRNA design web server described here is available at: https://cas13b.github.io/.

2.3 Results

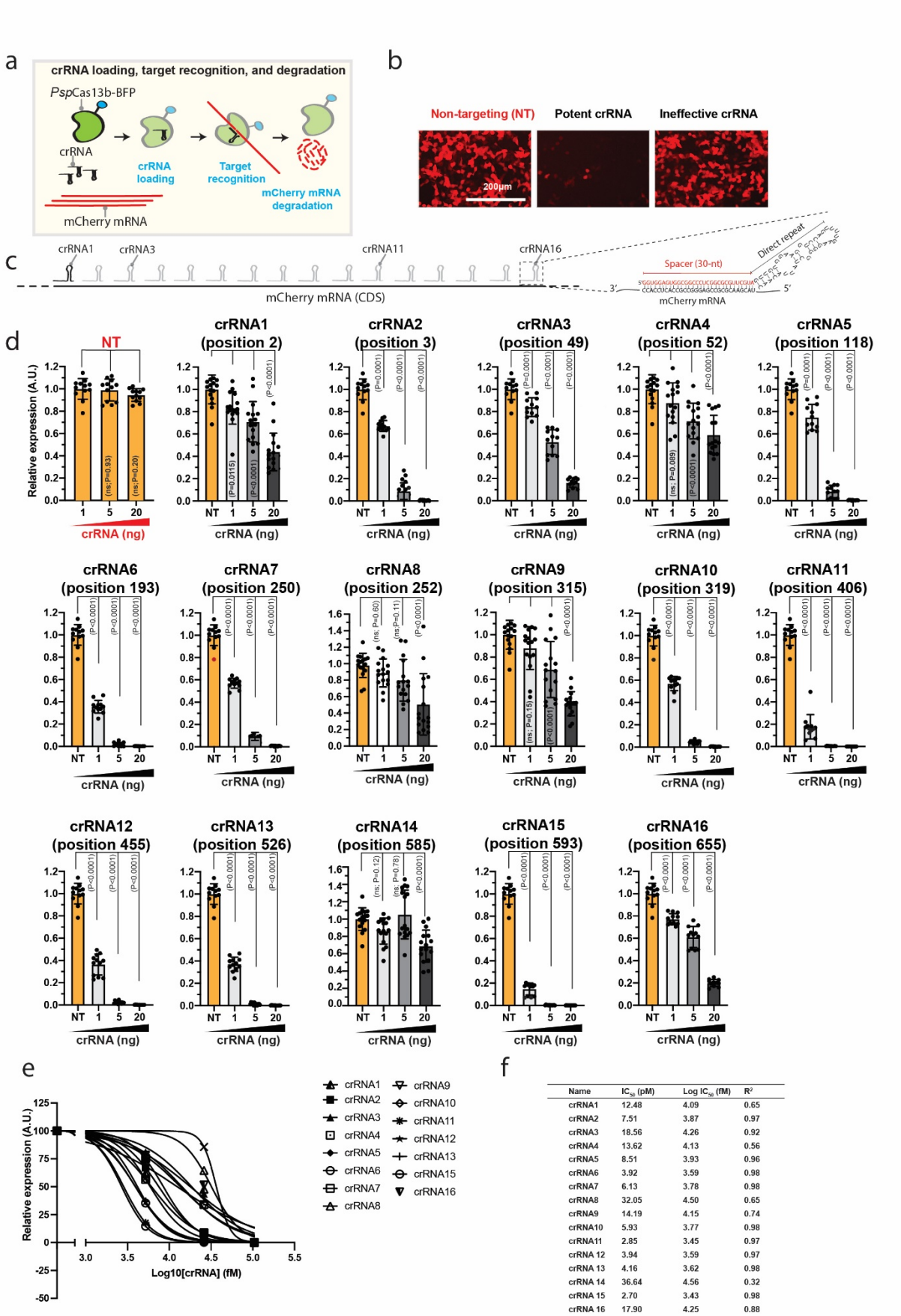
2.3.1 crRNAs silencing efficiency is highly variable.

To elucidate *Psp*Cas13b crRNA design principles, a quantitative fluorescence-based silencing assay was developed, in which we targeted the transcript of the mCherry reporter gene. To achieve this, HEK 293T cells were transfected with three plasmids encoding mCherry, *Psp*Cas13b-BFP, and either non-targeting (NT) or mCherry-targeting crRNA (**Figure 2.1a**). Fluorescence microscopy analysis of cells transfected with mCherry targeting crRNAs showed pronounced silencing, but no appreciable silencing in cells expressing NT crRNAs (**Figure 2.1b**).

Next, it was determined whether parameters such as the efficiency of crRNA transcription, crRNA loading, spacer nucleotide composition, target accessibility, and the presence of a potential protospacer-flanking sequence (PFS) may influence the efficiency of *Psp*Cas13b and lead to variability in the silencing profiles of various crRNAs. 16 crRNAs with spacer sequences that fully base pair with the coding sequence of the mCherry mRNA at various positions were designed (**Figure 2.1c**). To accurately determine the silencing efficacy of each crRNA in this cohort, crRNA dose-dependent silencing assays were carried out in which cells were transfected with 0, 1, 5, or 20 ng of each of 16 mCherry-targeting crRNAs, and silencing efficiency was measured. crRNA potency was determined by calculation of the IC50 value, the dose that achieved 50% degradation of the target RNA, as determined in this case by relative fluorescence intensity. mCherry-targeting crRNAs demonstrated dose-dependent silencing, but there were marked differences in the silencing efficacy of the various crRNAs, even when designed to target immediately adjacent RNA locations (**Figure 2.1d-2.1f**). For example, crRNA6, crRNA11, crRNA12, crRNA13, and crRNA15 were extremely potent and degraded the majority of mCherry mRNA at a very low dose of 1 ng plasmid (IC50 5.2 pM). Conversely, crRNA1, crRNA4, crRNA8, crRNA9, and crRNA14 were inefficient and failed to completely degrade mCherry mRNA even at higher doses of 5 and 20 ng (IC50 26 and 104 pM, respectively) (**Figure 2.1d**). Surprisingly, although crRNA14 and crRNA15 target neighbouring regions, separated by just 8 nucleotides, their silencing efficiencies differed markedly. Whereas 5ng of crRNA15 silenced >99% of mCherry expression ($P<0.0001$), the same amount of crRNA14 produced no significant silencing

(P=0.78) (**Figure 2.1e-2.1f**). As these two crRNAs target spatially adjacent sequences, this finding suggested there are key determinants of *Psp*Cas13b efficacy beyond target accessibility. Identifying

such determinants was considered to be crucial for efficient reprogramming.



Name	IC_{50} (pM)	Log IC_{50} (fM)	R^2
crRNA1	12.48	4.09	0.65
crRNA2	7.51	3.87	0.97
crRNA3	18.56	4.26	0.92
crRNA4	13.62	4.13	0.56
crRNA5	8.51	3.93	0.96
crRNA6	3.92	3.59	0.98
crRNA7	6.13	3.78	0.98
crRNA8	32.05	4.50	0.65
crRNA9	14.19	4.15	0.74
crRNA10	5.93	3.77	0.98
crRNA11	2.85	3.45	0.97
crRNA 12	3.94	3.59	0.97
crRNA 13	4.16	3.62	0.98
crRNA 14	36.64	4.56	0.32
crRNA 15	2.70	3.43	0.98
crRNA 16	17.90	4.25	0.88

Figure 2. 1 Sixteen crRNAs targeting mCherry transcript revealed diverse silencing efficiencies by *Psp*Cas13b.

(a) Schematic of *Psp*Cas13b silencing assay used to track the recognition and degradation of mCherry RNA. (b) Representative fluorescence microscopy images show the silencing of mCherry transcripts with a potent or ineffective targeting crRNA versus a non-targeting (NT) control crRNA in HEK 293T cells. (c) Schematic of the 16 *Psp*Cas13b crRNAs targeting mCherry RNA. The detailed spacer and direct repeat of a representative crRNA is indicated in the right schematic. (d) crRNA dose-dependent silencing of mCherry transcripts with either NT (orange) or targeting crRNAs (grey); Data points in the graph are normalized mean fluorescence from 4 representative fields of view imaged in $N = 3$ or 4. The data are represented in arbitrary units (A.U.). Errors are SD and p-values of one-way Anova test are indicated (95% confidence interval). N is the number of independent biological experiments. (e) Dose-response fitting of mCherry silencing data with non-targeting crRNA (NT) and 16 targeting crRNAs in d. (f) The IC_{50} values for 16 crRNAs targeting mCherry transcripts obtained from fitting in e.

2.3.2 Single-base tiled crRNAs reveal hidden parameters regulating mRNA silencing

To further understand the spectrum of crRNA potency, the silencing activity of *Psp*Cas13b was investigated across a defined targeted region of mRNA, as it was reasoned that silencing efficiency is likely intrinsic to the spatial characteristics of the crRNA sequence and binding sites. The study focused on crRNA12 (binding position #455) and crRNA16 (#655) which exhibited high and moderate silencing, respectively (**Figure 2.1c-2.1f**). 3-nucleotide resolution tiled crRNAs were designed to span a 30-nucleotide target region surrounding crRNA12 and crRNA16 binding positions (**Figure 2.2a**). In this tiled design, the binding sites of each adjacent crRNAs were spaced by just 3 nucleotides, thus their silencing profiles should reveal the relationship between efficacy, spacer-target sequence, and target accessibility. Considerable heterogeneity in the potency of these tiled crRNAs was again noted despite their proximity, with some adjacent crRNAs possessing contrasting silencing efficacy. These data indicated that physical barriers such as RNA binding proteins or structured RNA motifs were unlikely to explain the fluctuation in silencing between spatially adjacent crRNAs (**Figure 2.2b&2.2c**).

To further enhance understanding, the spatial resolution of this approach was extended by designing 61 tiled crRNAs with single-base incremental targeting nucleotides 424 to 485 of the mCherry coding sequence (**Figure 2.2d**). Consistent with previous results, markedly diverse silencing was noted for neighbouring crRNAs. For instance, crRNA13 achieved > 95% silencing, but shifting the

targeted region by only 1 nucleotide (crRNA14) reduced efficiency to ~30%. Similarly, crRNA51 yielded ~99% silencing while the adjacent crRNA52 had minimal activity (**Figure 2.2d**). These data strengthened the hypothesis that silencing efficacy is unlikely to be solely dependent on target accessibility and that other factors including specific nucleotide positions within the spacer or target, and a possible PFS, may all influence key steps of target silencing such as crRNA transcription, loading, and target recognition.

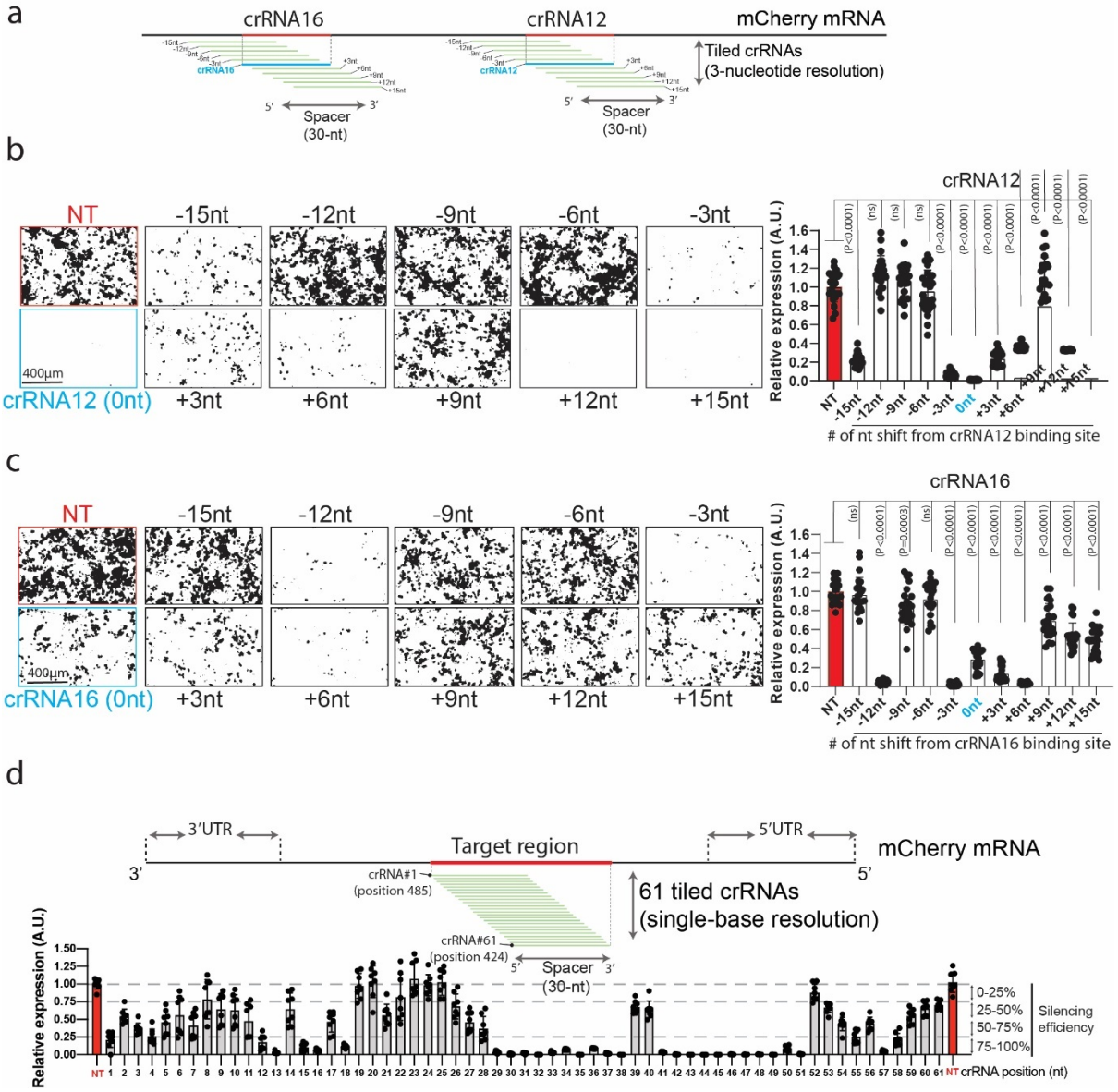


Figure 2. 2. Single-base tiled crRNAs revealed key design principles.

(a) The schematic illustrates mCherry RNA regions covered by 3-nucleotide increments tiled crRNAs. (b-c, left) Representative fluorescence microscopy images show the silencing of mCherry transcripts with tiled crRNAs targeting the region surrounding crRNA12 (b) and crRNA 16 (c). Images are processed for quantification using ImageJ. Similar results were obtained in N=4. (b-c, right).

Quantification of silencing efficiency with tiled crRNAs targeting the mCherry regions surrounding crRNA12 (left panel) and crRNA16 (right panel). Data points in the graph are normalized mean fluorescence from 4 representative fields of view per experiment imaged in $N = 4$. The data are represented in arbitrary units (A.U.). Errors are SD and p-values of one-way Anova test are indicated (95% confidence interval). (d) The schematic shows the sequence of mCherry RNA covered by 61 single-nucleotide resolved tiled crRNAs around the targeted region of crRNA12. The graph quantifies the silencing efficiency obtained with 61 tiled crRNAs in HEK 293T cells. Data points in the graph are normalized mean fluorescence from 4 representative fields of view imaged in $N = 2$. The data are represented in arbitrary units (A.U.). Errors are SD with a 95% confidence interval. N is the number of independent biological experiments.

2.3.3 *In silico* analysis of 201 crRNAs revealed key design principles

To further explore the factors that dictate *Psp*Cas13b silencing efficiency, the dataset was expanded by analysing the silencing profiles of 201 further crRNAs targeting various transcripts[139]. First, several characteristics that may influence *Psp*Cas13b silencing efficiency in unfiltered crRNAs population were examined, including the predicted folding of crRNA, target folding, spacer-target hybridization energy (**Figure 2.3a-2.3e**), and spacer nucleotide content (A, U, C, G, and CG) (**Figure 2.3f-2.3j**). The Pearson correlation was used to identify any relationship between predicted folding and *Psp*Cas13b silencing efficiency. The data revealed a moderate positive correlation between the minimum free energy (MFE) of the crRNA and *Psp*Cas13b silencing efficiency (r=0.19; p=0.0050) (**Figure 2.3b**). Similarly, target accessibility (folding of a 70-nt RNA sequence surrounding the targeted region) exhibited a modest correlation with the silencing efficiency of crRNA (r=0.16; p=0.0247) (**Figure 2.3c**). Together, these data suggest that the folding of the crRNA and the targeted sequence into complex secondary structures can only moderately limit *Psp*Cas13b silencing efficiency, possibly perturbing crRNA loading or target accessibility. Next, a similar approach was used to analyse the effect of differential ribonucleotide abundance within the spacer on crRNA activity. A (r=0.14; p=0.0427) and G (r=0.17; p=0.0137) nucleotide contents were positively correlated with *Psp*Cas13b silencing, whereas spacers with enriched C content exhibited a strong negative correlation with crRNA potency (r=-0.30; p<0.0001) (**Figure 2.3f-2.3j**). These data indicate that spacer nucleotide content is likely a key determinant of *Psp*Cas13b silencing and suggest spacer enrichment with C bases is likely to impede their potency.

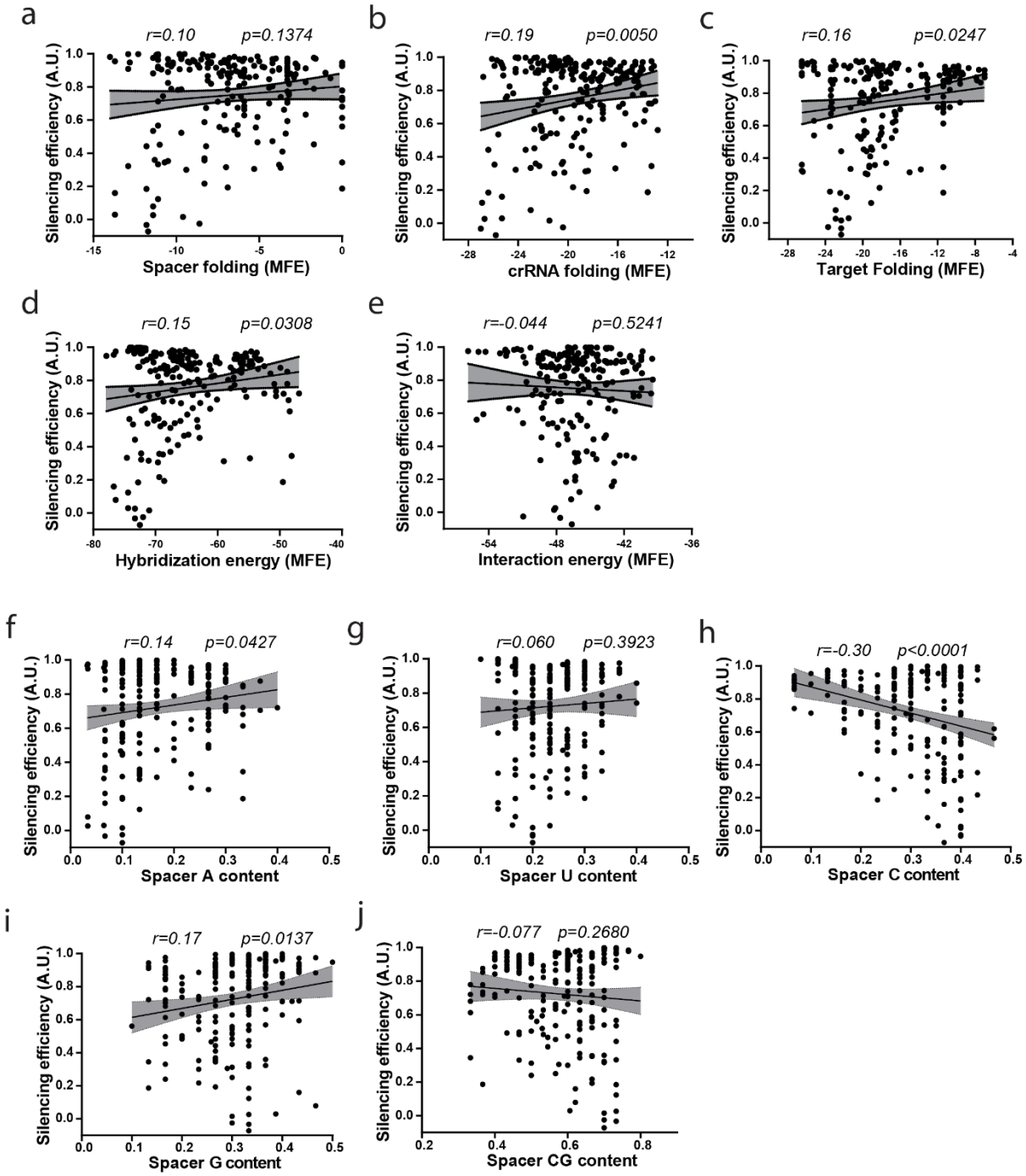


Figure 2. 3. *in silico* analysis of 201 crRNAs by predicted folding of crRNA, target folding, spacer-target hybridization energy.

Pearson correlation analysis between (a) crRNAs silencing efficiency and spacer folding, (b) entire crRNA folding (spacer and direct repeat), (c) target sequence folding, (d) spacer-target hybridization, (e) and spacer-target interaction. Data points in the graph are values of the silencing efficiency of individual crRNAs and their predicted folding (MFE) or hybridization/interaction energy. Pearson correlation analysis between spacer silencing efficiency and the content of spacer in (f) A

bases, (g) U bases, (h) C bases, (i) G bases, and (j) CG bases. Data points in the graph show the silencing efficiency and base content of individual spacer sequences. *r* (correlation coefficient) and *p*-value (95% confidence interval) are indicated in each graph.

Next, the 201 crRNAs were ranked in order of silencing efficiency. crRNAs that achieved >90% silencing efficiency were designated 'potent', and those with < 50% efficiency were considered 'ineffective'. crRNAs with intermediate silencing (50 to 90%) were excluded from the analysis. Molecular features capable of differentiating potent and ineffective crRNAs were examined (**Figure 2.4a**). Many CRISPR variants possess an upstream or downstream PFS that restricts targeting activity and prevents degradation of their own nucleic acids[157]. To investigate the potential presence of a PFS that could constrain *Psp*Cas13b silencing, weight matrix plots were generated that analyse nucleotide composition up to four bases upstream or downstream of the targeted sequence in the highly potent and ineffective cohorts of crRNAs. There was no appreciable bias in nucleotide composition at these target flanking sites, indicating that *Psp*Cas13b activity is not subject to PFS regulation in mammalian cells (**Figure 2.4b**).

Last, it was determined whether the nucleotide composition of the spacer could influence *Psp*Cas13b silencing efficiency. Overall, G nucleotides were enriched in the spacer sequences of potent crRNAs compared to the unfiltered total crRNAs (**Figure 2.5a-2.5c**). Conversely, C nucleotides were enriched in the ineffective crRNAs (**Figure 2.5a, 2.5d, & 2.5e**). This data reveals that potent crRNA favours G over C bases, however the positions of the enriched/depleted G or C nucleotides in effective and ineffective crRNAs remained unknown. To study this parameter further, an unbiased analysis of the nucleotide composition at all 30 positions of the spacer was carried out for potent and ineffective crRNAs. Weight matrix plots and Delta probability analysis were used to compare spacer nucleotide composition at all positions in filtered and unfiltered samples (**Figure 2.4c-2.4f**), and marked differences were identified between the two crRNA cohorts. G bases at the 5'end, particularly a GG sequence at the first and second positions were strongly associated with highly potent crRNAs. Conversely, G nucleotides were depleted, and C bases enriched at the 5'end of ineffective spacers. Ineffective crRNAs also had a significant enrichment of C bases at positions 11, 12, 15, 16, and 17 (**Figure 2.4c-2.4f**). These data thus revealed key nucleotide positions that may determine the potency of crRNAs, and which could serve as predictive parameters of crRNA potency.

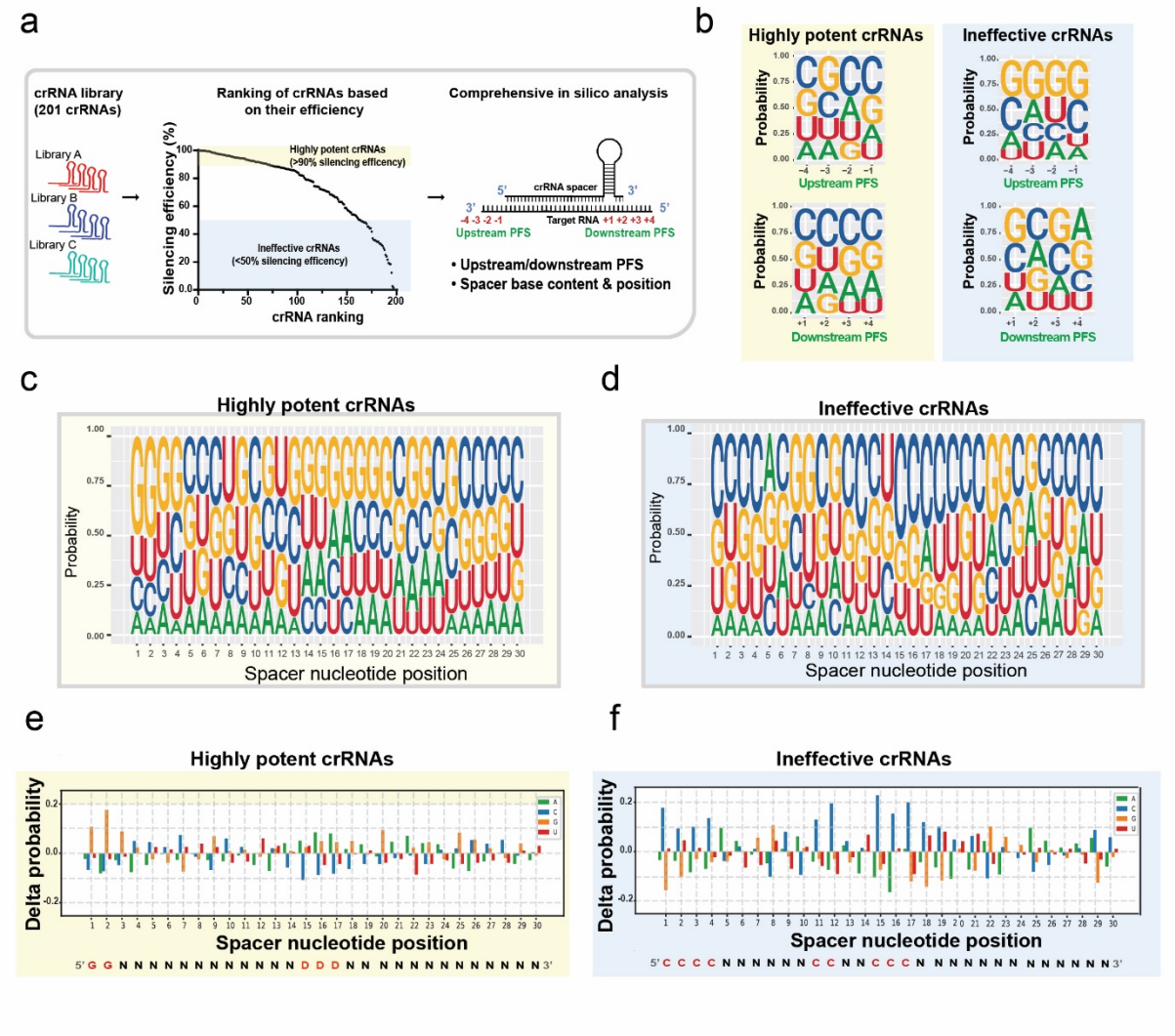


Figure 2. 4. *in silico* analysis of crRNA silencing profiles revealed key design principles.

(a) Schematic of the bioinformatics pipeline used to investigate various parameters that impact *Psp*Cas13b silencing. PFS positions (4 nucleotides surrounding the 5' and 3' end of the targeted region that basepair with the spacer) are indicated. 201 crRNAs are ranked based on their silencing efficiency. The highly potent crRNAs that achieved >90% silencing efficiency (yellow box) and the ineffective crRNAs that achieved <50% silencing efficiency (blue box) are analysed for PFS and spacer nucleotide positions. (b) Position Weight Matrices (PWMs) depicting the positional nucleotide probabilities of upstream or downstream PFS in either the highly potent (left panel) or ineffective crRNAs (right panel). (c-d) Position Weight Matrices (PWMs) depicting the positional nucleotide probabilities of either the highly potent (c) or the ineffective (d) crRNA spacer sequences. (e-f) Delta nucleotide probabilities of the highly potent (e) and ineffective (f) crRNA spacer sequences that compare filtered spacer nucleotide positions to the baseline nucleotide distribution. A, C, G, and U bases in the histograms are represented with green, blue, orange, and red colours, respectively.

Spacer nucleotide positions-based formulas to predict highly potent and ineffective spacer sequences. The red colour in the spacer sequence highlights the position of key residues.

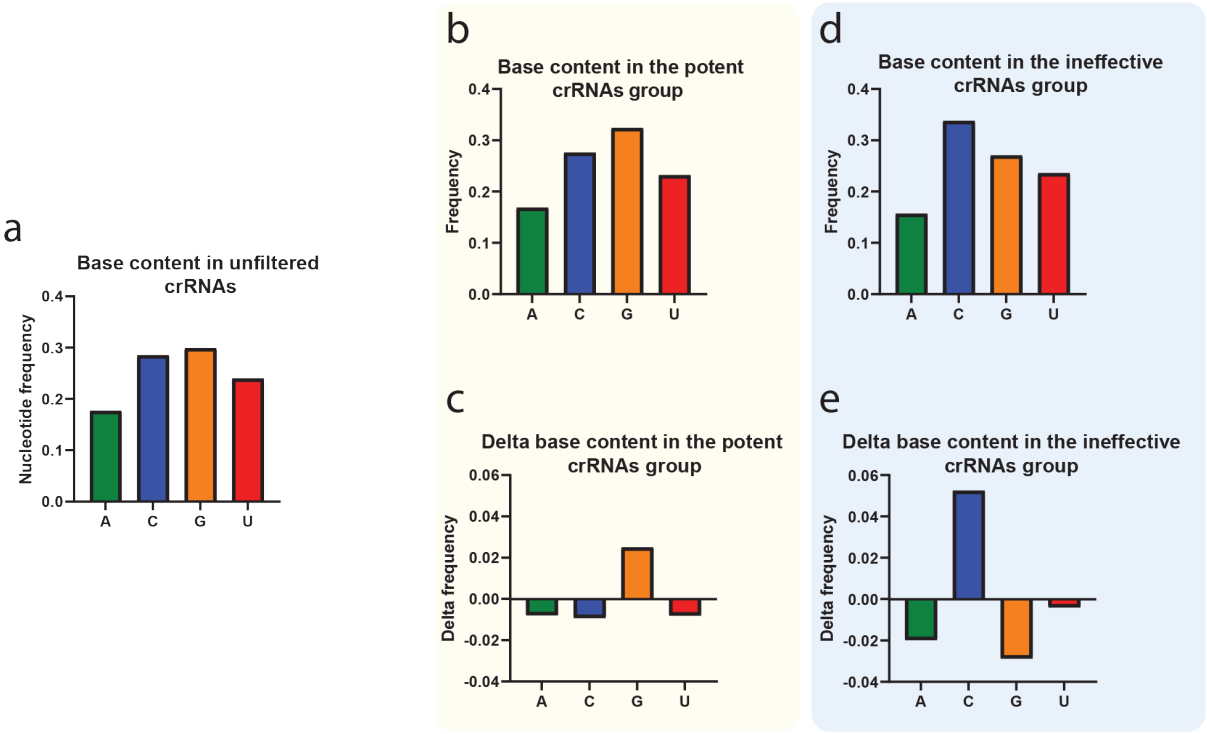


Figure 2. 5. The frequency of A, C, G, and U bases in crRNA spacer sequences.

(a) The frequency of A, C, G, and U bases in unfiltered crRNA spacer sequences. The frequency (b) and delta frequency (c) of each base in the spacer sequences of filtered potent crRNAs. The frequency (d) and delta frequency (e) of each base in the spacer sequences of filtered ineffective crRNAs.

2.3.4 Functional validation of crRNA prediction and design

The above *in silico* analysis enabled us to generate a consensus sequence that might predict potent and ineffective crRNAs. Potent crRNAs should include GG sequence at the first and second position of the spacer and lack C bases in positions 11, 12, 15, 16, and 17 (***GG***NNNNNNNN***DD***NN***DDD***NNNNNNNNNNNN; D is a G, U, or A nucleotide; N is any nucleotide). Also, crRNAs containing C in spacer positions 1, 2, 3, 4, 11, 12, 15, 16, and 17 were predicted to have poor silencing efficiency (***CCCC***NNNNNN***CC***NN***CCC***NNNNNNNNNNNN). The predictive accuracy of this algorithm was then tested through a prospective unbiased design of crRNAs targeting EGFP and TagBFP, two mRNA targets not investigated previously. Notably, out of 21

predicted potent crRNAs, 20 achieved high silencing efficiency for either EGFP or TagBFP mRNA. Conversely, the majority of crRNAs predicted to be inefficient failed to efficiently silence EGFP or TagBFP transcripts (**Figure 2.6a-2.6f**). By formulating our prediction from a pre-existing dataset, and validating its accuracy on hitherto untested transcripts, the spacer nucleotide prediction was found to be both accurate and generalisable, demonstrating its utility in crRNA design for any transcript of interest.

The efficiency of this newly designed tool was then compared to the benchmark crRNA design tool currently available for *Rfx*Cas13d[59] (**Figure 2.6g**). The 10 crRNAs predicted to be most potent for *Rfx*Cas13d targeting mCherry on average achieved silencing of 80.7% (**Figure 2.6h**). By comparison, the new methodology described above for *Psp*Cas13b on average resulted in ~87.8% silencing efficiency (data pooled for EGFP and TagBFP, **Figure 2.6c, 2.6f**) thus outperforming the *Rfx*Cas13d design, and further validating the accuracy of the new prediction tool (**Figure 2.6c & 2.6f**).

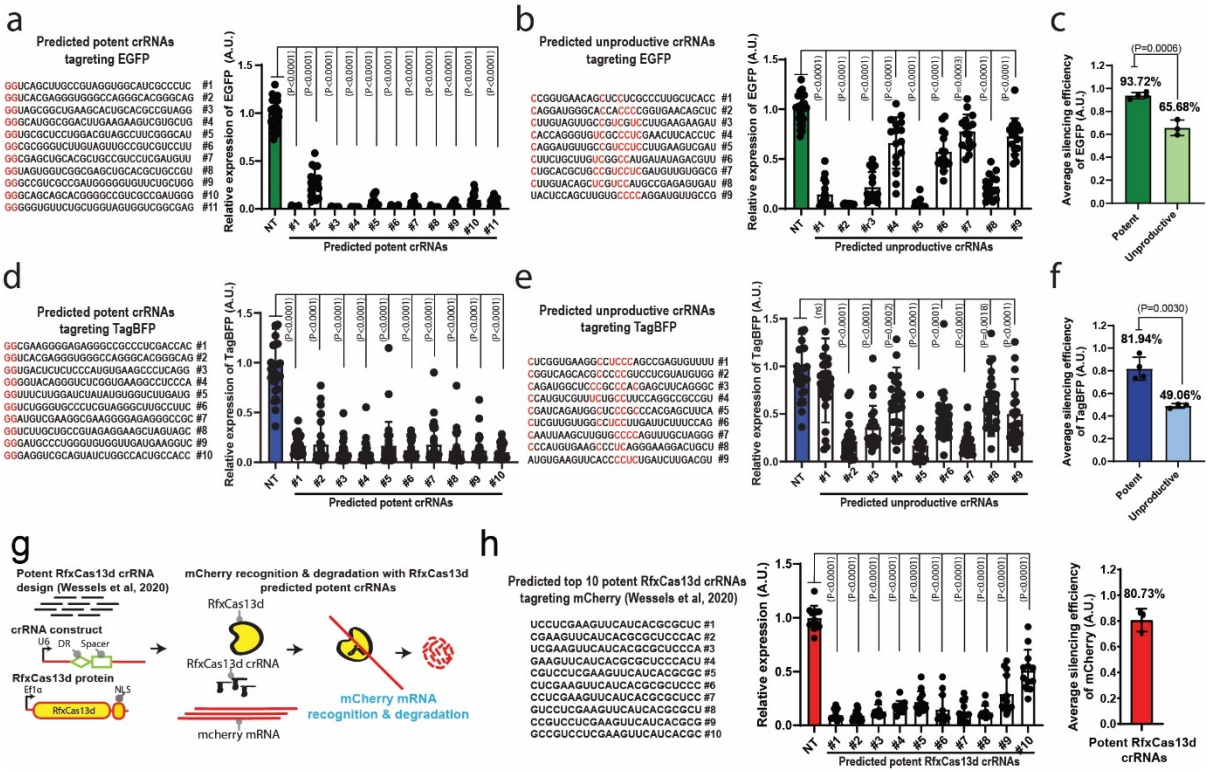


Figure 2. 6. Functional validation of *Psp*Cas13b crRNA prediction and design.

(a-c) Validation of spacer nucleotide preferences through the targeting EGFP in HEK 293T cells. (a) Prospective design and validation of predicted potent crRNAs harbouring a 'GG' motif at 5' end of spacers targeting EGFP transcript. (b) Prospective design and validation of predicted ineffective crRNAs lacking 5' GG motif and harbouring 'C' bases at the central region of spacers targeting EGFP transcript. Data points in the graph are mean fluorescence from 4 representative field of views per

condition imaged; *N*=3 or 4. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of one-way Anova test are indicated (95% confidence interval). (c) The average silencing efficiency of predicted potent and ineffective crRNAs targeting EGFP transcripts. *N*= 3 or 4; Data are normalized means and errors are SE; Results are analysed by unpaired two-tailed Student's t-test (95% confidence interval). (d-f) Validation of spacer nucleotide preferences revealed in the screen by targeting TagBFP transcript in HEK 293T cells. (d) Prospective design and validation of predicted potent crRNAs harbouring a 'GG' motif at 5' end of spacers targeting TagBFP transcript. (e) Prospective design and validation of predicted ineffective crRNAs lacking 5' GG motif and harbouring 'C' bases at the central region of spacers targeting TagBFP transcript. Data points in the graph are mean fluorescence from 4 representative fields of view per condition imaged; *N*=3 or 4. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of one-way Anova test are indicated (95% confidence interval). (f) The average silencing efficiency of predicted potent and ineffective crRNAs targeting TagBFP transcripts. Data points in the graph represent independent biological replicates. *N*= 3 or 4; Data are normalized means and errors are SE; Results are analysed by unpaired two-tailed Student's t-test (95% confidence interval). (g) Schematic of *Rfx*Cas13d silencing assay to target mCherry transcript in HEK 293T cells using potent crRNAs predicted by the *Rfx*Cas13d guide prediction platform published by Wessels et al[36]. (h) The spacer sequences of the top 10 potent *Rfx*Cas13d crRNAs targeting mCherry transcripts predicted with Wessels et al tool[36] (left). Data points in the graph (middle) are mean fluorescence from 4 representative field of views per condition imaged; *N*=3. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of one-way Anova test are indicated (95% confidence interval). The average silencing efficiency of predicted potent *Rfx*Cas13d crRNAs targeting mCherry transcripts (Wessels et al tool[36]) is shown at the right-side graph. Data points in the graph represent independent biological replicates. *N*= 3; Data are normalized means and errors are SE (95% confidence interval).

To further investigate whether specifying a G-rich motif at the 5'end of a crRNAs would increase crRNA potency whereas C-rich motifs would decrease it. 11 crRNAs encoding GG sequence at positions 1 and 2 were altered to CC by spacer mutagenesis. This caused a substantial compromise in the silencing efficiency of most of the crRNAs (**Figure 2.7a**). For instance, 5 ng of WT crRNA7 achieved ~95% silencing, whereas replacing GG with CC led to a complete loss of silencing (**Figure 2.7a**). When 3 or any 2 G bases at the 5'end of the spacer were altered to C, silencing was reduced

by >99% and ~70% respectively, however introducing only a single C at position 1, 2, or 3 has no significant effect on crRNA potency (**Figure 2.7b-2.7c**).

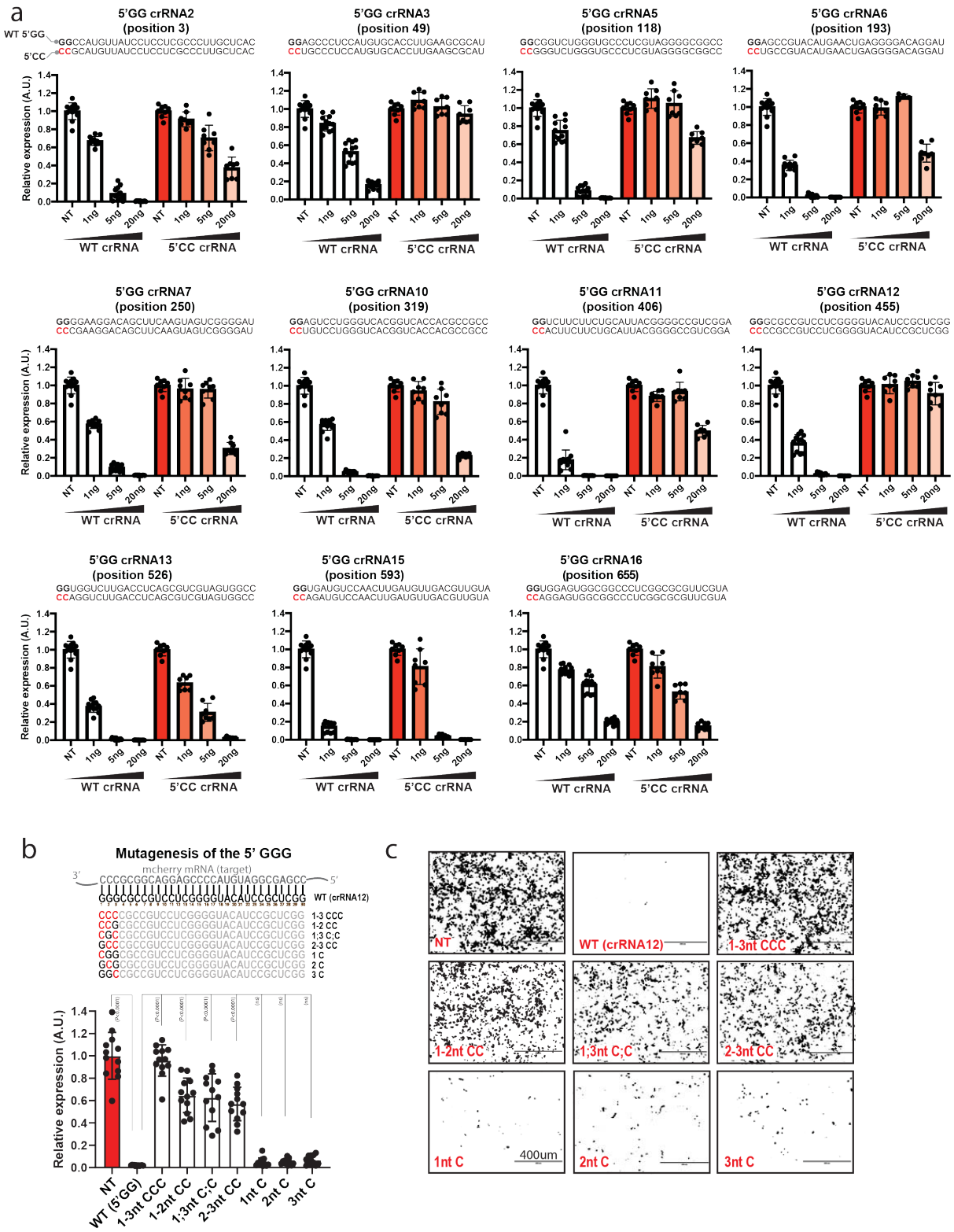


Figure 2. 7. The dose-dependent silencing assays show that the substitution of G to C bases at the 5' end of the spacer greatly compromised their silencing efficiency.

(a) Dose-dependent silencing of mCherry transcript with non-targeting crRNA (NT), 11 unmodified crRNA (WT) that possess a GG sequence at their 5'end (white bars), or the same 11 crRNAs were the 5' GG sequence of the spacer is mutated to 5' CC through 1-3 nucleotides mutagenesis. Data points in the graph are normalized mean fluorescence from 4 different field of views imaged in *N*=2. The data are represented in arbitrary units (A.U.). Errors are SD with 95% confidence interval. (b) Mutagenesis analysis of spacer 1-3 nucleotides (5' end) examining the impact of C to G substitutions on crRNA silencing efficiency. The nucleotides in red highlight mismatch positions in the spacer sequence. Data points in the graph are mean fluorescence from 4 representative field of views per condition imaged; Errors are SD and p-values of one-way Anova test are indicated (95% confidence interval). *N* = 3. The data are represented in arbitrary units (A.U.). *N* is the number of independent biological replicates. (c) Representative fluorescence microscopy images show the silencing efficiency of the mCherry transcripts with NT, WT, and mutant crRNAs in HEK 293T cells. NT is a non-targeting control crRNA. Scale bar = 400μm. Similar results were obtained in 3 independent experiments in HEK 293T cells.

Next, ineffective crRNAs lacking a GG sequence at their 5'end were modified either by inserting an additional G at position 1, altering the 1st nucleotide to a G, or substituting the 1st and 2nd nucleotides to GG (**Figure 2.8a-2.8g**). Surprisingly, it was demonstrated that any G sequences at the 5'end of the spacer greatly increased crRNA potency, even if this introduced a spacer-target mismatch (**Figure 2.8a-2.8g**). A Previous study[158] suggested that the U6 promotor has a preference for G or A immediately adjacent to the transcription starting site. Therefore, the systematic design of crRNA with two G nucleotides at their 5' end may increase crRNA potency through enhanced transcription. To determine whether the improved silencing by crRNAs harbouring a G-rich 5'end might be secondary to changes in crRNA abundance, the expression levels of the original and mutated crRNAs were estimated using quantitative real-time PCR (RT-PCR). Although the difference was not statistically significant, an increase in crRNA abundance was observed when a 5' G-rich motif was present (**Figure 2.8h-2.8l**).

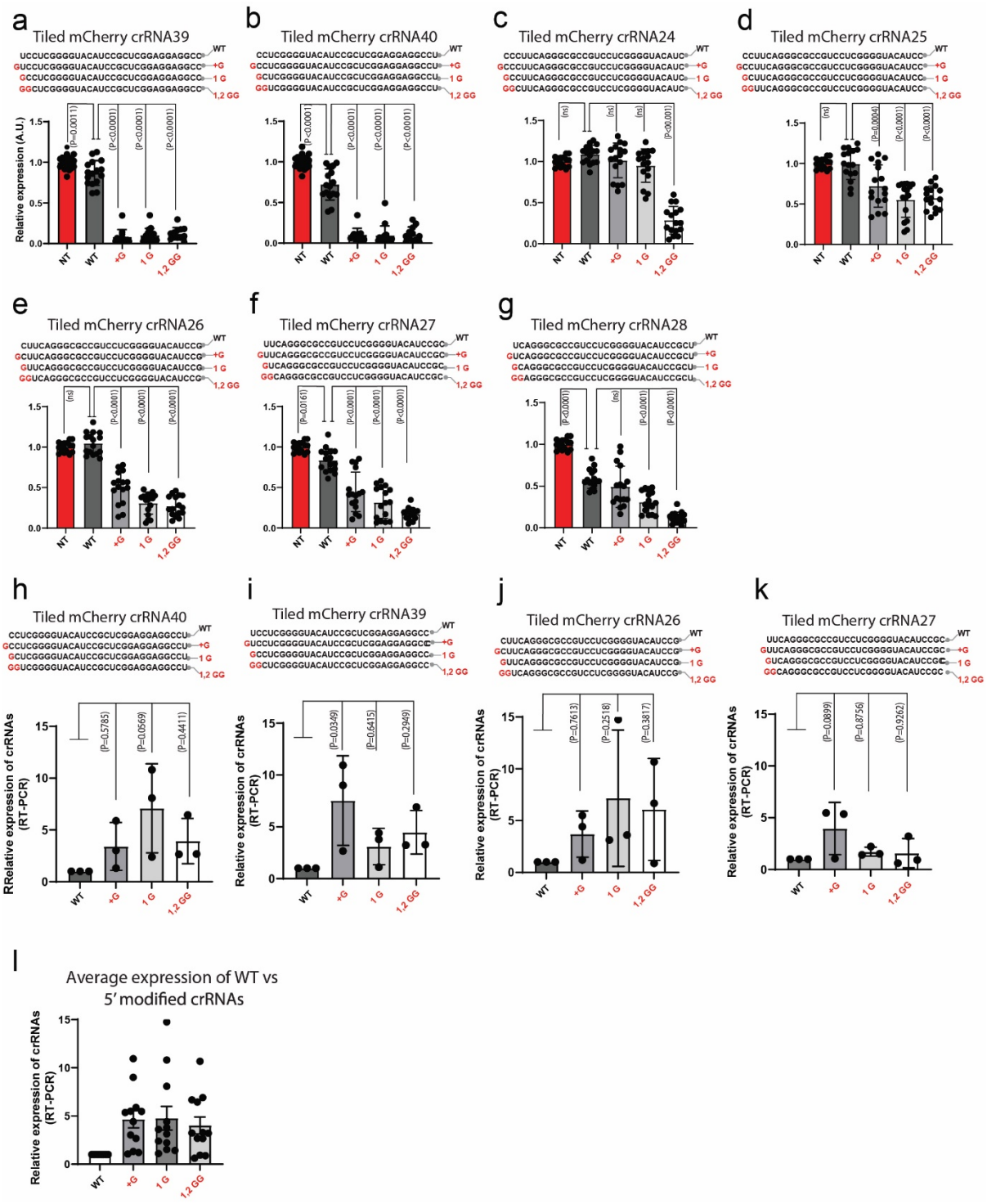


Figure 2. 8. 5' GG motifs on crRNA spacers improved PspCas13b efficiency partially through enhancing crRNA transcription.

(a-g) Incorporation of a G-rich motif at the 5' end of ineffective spacer sequences targeting mCherry through G-nucleotide insertion or substitution greatly enhanced their silencing efficiency. Data points in the graph are mean fluorescence from 4 representative field of views per condition imaged; *N*=3 or 4. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of

one-way Anova test are indicated (95% confidence interval). (h-k) RT-PCR analysis to assess the expression of WT crRNAs, crRNAs with an extra G at their 5' end (31-mer spacer), crRNAs with the first spacer nucleotide substituted to a G (30-mer spacer), and crRNAs with the first and second spacer nucleotides substituted to GG (30-mer spacer). crRNAs expression was measured 48h post-transfection in HEK 293T cells, $N = 3$. Data are normalized means and errors are SEM; Results analysed with one-way Anova test with p-value indicated (95% confidence interval). (l) Averaged expression (from a-d) of unmodified or modified crRNAs harbouring G-rich 5'end.

To further understand whether a 5' GG can enhance crRNA potency beyond transcription upregulation, four pairs of crRNAs were *in vitro* transcribed, in each instance either possessing or not possessing a 5' GG (**Figure 2.9a-2.9b**). Equal amounts of purified *in-vitro* transcribed crRNAs were co-transfected into HEK 293T cells together with *Psp*Cas13b plasmid to compare their silencing efficiency. 24h post-transfection, all four crRNAs with a 5'GG sequence achieved significantly higher silencing efficiency than their unmodified counterparts (**Figure 2.9c**). At 48h, the silencing efficiency of all the *in-vitro* transcribed crRNAs had diminished (probably due to their gradual degradation by cellular nucleases), but two of four crRNAs with a 5'GG remained more potent than the unmodified crRNAs (**Figure 2.9d**). Together, these data suggested that incorporating a 5'GG enhances both crRNA transcription and *Psp*Cas13b potency, the latter possibly through increased crRNA loading, target recognition, and/or target cleavage.

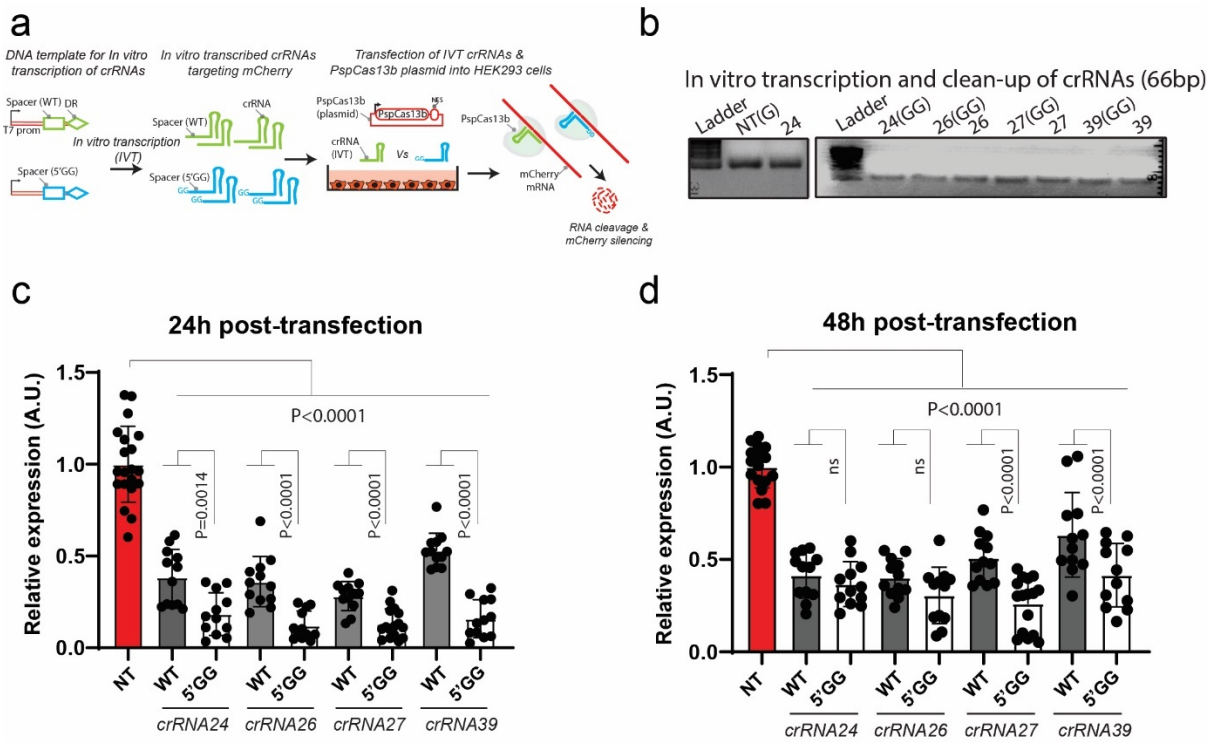


Figure 2. 9. 5' GG motifs on spacers of in-vitro transcribed crRNAs improved *Psp*Cas13b efficiency.

(a) Schematic of *Psp*Cas13b silencing assay performed in HEK 293T cells using *in-vitro* transcribed (IVT) crRNAs with or without a 5' GG motif. (b) Quality control analysis of *in-vitro* transcribed crRNAs using 1% agarose gel electrophoresis. (c-d) Assessment of the silencing efficiency of *in-vitro* transcribed crRNA pairs that either fully basepair with the target (WT) or harbour a mismatched 5' GG motif in their spacer sequence. The silencing efficiency was monitored at 24h (c) and 48h (d) timepoint post transfection of IVT crRNAs. Data points in the graph are mean fluorescence from 4 representative field of views per condition imaged; $N=3$. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of unpaired two-tailed Student's t-test are indicated (95% confidence interval). N is the number of independent biological replicates.

In addition to mCherry, C to G substitutions at key spacer positions (1, 2, 11, 15, 16, 17) were also shown to improve the silencing efficiency of crRNAs targeting other transcripts (**Figure 2.10a-2.10l**). Indeed, when crRNA design choices were limited, the *de novo* incorporation of mismatched G bases at these positions substantially increased their potency despite introducing nucleotide mismatches with the target. The silencing assays in **Figure 2.10a-2.10l** show that the substitution of nucleotides at spacer positions 1 and 2 with GG enhanced the potency of six different crRNAs targeting gene-fusion transcripts. Likewise, the replacement of 'C' at positions 11, 12, 15, 16, 17, or 18 with G enhanced the

potency of six crRNAs targeting the breakpoint of oncogenic gene fusion transcripts BCR-ABL-1 and SNX2-ABL1. These data confirmed the importance of the 5' G-rich motif and the avoidance of C at the central region of the crRNA to maximise silencing efficiency, supporting the findings in **Figure 2.4 & Figure 2.6**. The incorporation of up to 3 mismatched G bases per spacer (5'end and/or central region) was found to enhance the potency of crRNAs, but beyond 3-nucleotide substitutions, the benefit became compromised (**Figure 2.10c, 2.10e**), likely due to the destabilization of spacer-target interaction, consistent with mutagenesis data in **Figure 4.1 (chapter 4)** and **Figure 3.2a, 3.2c (chapter 3)**.

To facilitate the use of this optimized and validated spacer nucleotide-based formula for potent crRNA design, our laboratory created a user-friendly webpage (https://cas13b.github.io/) and made it freely available to the research community. This *in-silico* tool requires only the targeted sequence as input to create single-base tiled spacer sequences and rank them based on their predicted potency (see **Chapter 2.2.10**).

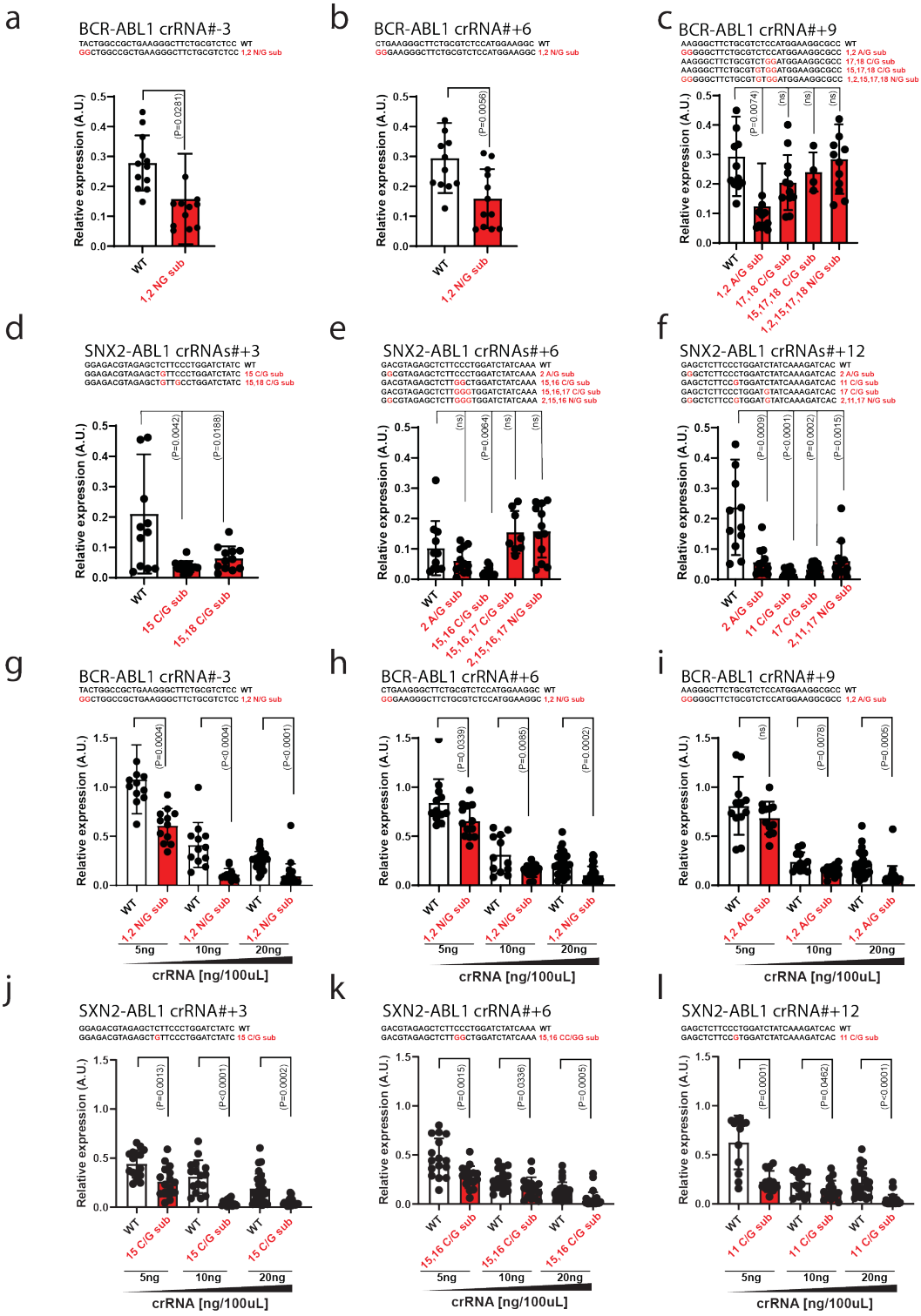


Figure 2. 10. Incorporation of target-mismatched 'G' bases at the 5'end of spacer sequence greatly enhances *Psp*Cas13b crRNA efficiency.

(a-f) Comparison of silencing efficiencies of crRNAs targeting the breakpoint of gene fusion transcripts with or without incorporation of mismatched G-bases at the 5'end and/or central regions

of the spacer. Data points in the graphs are mean fluorescence from 4 representative field of views per condition imaged; *N*=3 or 4. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of unpaired two-tailed Student's t-test are indicated (95% confidence interval). (g-l) crRNA dose-dependent silencing of gene fusion transcripts with or without incorporation of mismatched G-bases at the 5'end and/or central regions of the spacer. Data points in the graphs are mean fluorescence from 4 representative field of views per condition imaged; *N*=3 or 4. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of unpaired two-tailed Student's t-test are indicated (95% confidence interval). Data points in the graphs are mean fluorescence from 4 representative field of views per condition imaged; *N*=3 or 4. *N* is the number of independent biological replicates.

2.4 Discussion

Our study elucidates key molecular principles of *Psp*Cas13b RNA silencing that enabled *de novo* design of crRNAs. Overall, Cas13 nucleoproteins can offer effective and specific silencing of targeted transcripts without the risk of permanently altering genomic DNA[125,148,149]. The single-nucleotide resolution tiled crRNA screens and comprehensive *in silico* analysis performed above have demonstrated that, unlike other CRISPR orthologs[159], *Psp*Cas13b silencing is not restricted by a defined PFS (PAM-like) sequence, underlining its greater design flexibility (**Figure 2.4**). Among the strongest predictors of Cas13 activity, the presence of a 'G-rich' motif at the 5'end of the spacer was identified as a key determinant of crRNA potency (**Figure 2.4, Figure 2.6, Figure 2.8-2.9**). Notably, *de novo* designed crRNAs harbouring target-matched or target-mismatched 'GG' sequences at the 1st and 2nd nucleotide positions of the spacer greatly enhanced the silencing potency of otherwise poorly effective crRNAs (**Figure 2.8**). The ability of this target mismatched 'GG' motif to rescue the potency of certain ineffective crRNAs expands the range of effective crRNAs for a given target, which is particularly important for narrowly defined target sequences, especially when targeting the breakpoint region of fusion transcripts, RNA isoforms, or single-nucleotide variants. It was also revealed that 'C' nucleotides at the 5'end or central spacer positions can reduce the efficacy of crRNAs (**Figure 2.4, Figure 2.6-2.7**). The substitution of these 'C' bases with 'G' can enhance crRNA potency despite the introduction of mismatches that do not exceed 3 (**Figure 2.10**).

The thermodynamic stability of the spacer-target duplex is unlikely to explain the superior silencing obtained with spacers containing a 5' G-rich motif, given a G-C or C-G basepairing contributes the same number of hydrogen bonds for the stability of spacer-target duplex. Yet, our data show that 5' G-rich spacers achieved higher potency than those that are C-rich. In fact, when

substituting ribonucleotides at the 5'end of an ineffective spacer with mismatched G bases systematically improved silencing despite introducing mismatches that were inherently unfavourable for duplex RNA stability (**Figure 2.8**). This improved silencing with mismatched G bases at the 5'end of the spacer suggests that the primary function of this motif may not be to contribute to spacer-target basepairing. It was also shown that the intracellular abundance of 5'GG crRNAs was increased, possibly due to their enhanced transcription (**Figure 2.8**). This is consistent with a previous report showing that Pol III promoters can yield higher transcription rate when the transcribed small RNA has a 5' A or G[158]. Direct transfection of *in-vitro* transcribed crRNAs showed that this G-rich motif can indeed further improve crRNA potency through other mechanisms that may include enhanced crRNA stability, loading (affinity with *Psp*Cas13b), or by conferring greater nuclease activity beyond target recognition (**Figure 2.9**).

Integrating these discoveries into otherwise generic design processes enabled us to a consistent generation of highly potent crRNAs with an average silencing efficiency of ~87% (**Figure 2.6**), outperforming the benchmark prediction algorithm available for *Rfx*Cas13d orthologue[56]. To maximise the benefit of these findings for the entire research community and to simplify the process of crRNA optimisation for *Psp*Cas13b, an open-access and user-friendly webpage has been established (https://cas13b.github.io/), based on the aforementioned design criteria.

Chapter 3

Silencing tumour fusion transcripts with Reprogrammed CRISPR-*Psp*Cas13b

Abstract

Precision oncology programs can rapidly identify oncogenic gene fusions in individual patients[140-142]. However, despite their established oncogenicity, the vast majority of gene fusions remain '*undruggable*' due to the lack of specific inhibitory small molecules[143,144]. The unique chimeric sequences encoded on fusion gene transcripts are not found in normal cells, which potentially makes them an ideal target for Cas13 nucleases. In principle, designing crRNAs whose sequence is readily determined by their complementarity with an RNA target sequence unique to a patient's cancer should enable a far more rapid and flexible approach to developing personalised therapeutics than the current painstaking and expensive alternative of small molecule drug design and development. The *de novo* crRNA design principles described in **Chapter 2** are used here to demonstrate the potent and selective silencing of various gene fusion transcripts and consequently, their downstream oncogenic networks. This was achieved without off-targeting of non-translocated (wild type) gene variants that share extensive sequence homology with the oncogenic fusion product. It is demonstrated that the rational design of crRNAs using the previously described targeting rules can specifically and potently target the breakpoint of various fusion transcripts including BCR-ABL1 p210 that drive chronic myeloid leukemia and suppress their downstream oncogenic signalling networks without any impact on untranslocated wild type BCR and ABL1 mRNAs. Furthermore, a single crRNA targeting the BCR-ABL1 mRNA breakpoint suppressed the signalling activity of both parental and the *Imatinib*-resistant T315I mutant (which often drives treatment resistance and clinical relapse) with similar potency. Together, the results demonstrate the utility, potency, and high specificity of *Psp*Cas13b in suppressing various fusion oncogenic drivers including drug-resistant mutants in a personalized manner.

3.1 Introduction

Cancers are caused by the accumulation of genomic mutations and gene expression dysregulations[160]. In many patients, standard therapies such as chemotherapy can often result in extensive off-target effects and select for further genetic aberrations that confer resistance to these therapies[161,162]. Advanced sequencing technology and an understanding of tumour genomics have brought in a new era of precision and personalized medicine[163]. Aberrant genes called tumour drivers that confer clonal expansion and malignant spread are particularly considered appropriate targets for cancer treatments[164]. Tumours develop by a process termed clonal evolution, in which genetic mutations initially result in a range of deleterious, neutral, or advantageous phenotypes. Darwinian natural selection then acts on the diverse population of tumour cells to confer growth or survival advantages on certain clones within the relevant tissue-specific microenvironment[160]. Such mutations that give fitness advantages to tumour cells are called 'tumour driver' mutations. Various clones compete for resources in the microenvironment and the ones that accumulate advantageous tumour driver mutations ultimately form a tumour mass and may metastasize to other organs[164].

Given the above, great efforts are made to identify tumour drivers and many have been revealed with the recent revolution of genome-wide screens. As a result, visualizing the Philadelphia translocation in chronic myeloid leukaemia by low-resolution cytogenetic techniques has been replaced by reading single-base substitutions in genes such as BRAF through whole genome sequencing[165]. Importantly, next-generation sequencing techniques usher in a new era of high-throughput, cost-effective genome sequencing, which is increasingly accelerating the identification of new tumour-driver mutations despite the frequently extensive inter/intra-tumour heterogeneity[163]. So far, over 500 tumour driver mutations have been identified in the tumour genomes of multiple patients and can be classified into four categories: point mutation of a single base, insertions and deletions of DNA segments, chromosomal rearrangements, and abnormalities of gene copy number[129]. These mutations typically cause hyperactivity or overexpression of oncoproteins or loss of function in tumour suppressors. Overall, the definition of so many tumour drivers should pave the way for drug development, and a move away from a one-size-fits-all strategy to personalized, precise and targeted therapy[163]. Recently, small molecule drugs or biologicals that specifically target various tumour driver mutations have been developed, improving survival rates dramatically, especially in the first few months or years of therapy. For example, small molecule drugs *vemurafenib* and *dabrafenib* target the very common BRAF V600E point mutation present in up to 40% of advanced melanomas[166], *imatinib* targets the BCR-ABL translocation product in chronic myeloid leukaemia[167] and monoclonal antibody *trastuzumab* targets human epidermal growth factor receptor 2 (HER2) in breast cancer[168].

Despite some considerable advances in structure-based drug design, the translation of cancer genomic discoveries to the clinic remains challenging. Currently, most drugs including small molecules and monoclonal antibodies target tumour drivers at the protein level. The process of conventional small-molecule drug development is time-consuming and often requires solving a target protein's structure and multiplexed high throughput screening of libraries of small molecules[169]. For example, it took nearly 10 years to develop *vemurafenib* targeting BRAF V600E after it was identified as a tumour driver mutation[166]. In fact, over 500 consensus tumour drivers have been identified, but only approximately 30 drugs targeting a small percentage of them are available on the market[129]. The added revelation of intra and inter-tumour heterogeneity across various tumours, different patients, and different stages of tumour development[170,171] highlights even more, the need to tailor specific treatment for individual clinical settings. Unfortunately, only a small proportion of patients carrying particular tumour driver mutations can currently receive a targeted therapy and it is impossible to design a protein-targeting drug for each patient in a practical timeframe[172]. Therefore, there is an urgent need for a new drug strategy that provides high specificity, and ease of design and that can be multiplexed for personalized use.

In a relatively short time, impressive breakthroughs in CRISPR-Cas biology and therapeutic development have offered a promising opportunity to precisely target tumour-driver genes[10,11]. The most extensively studied CRISPR-Cas9 ortholog can in principle disrupt any oncogene or defective

tumour suppressor gene and induce its loss of function at the DNA level, while also providing a powerful tool to study cancer functional genomics in basic research[9]. Nevertheless, the irreversibility of the approach and safety concerns related to off-target effects and chromosomal instability largely constrain its therapeutic applications[10]. A major hypothesis tested in the current study is that highly specific, RNA-guided targeting of oncogenic RNA by a suitable CRISPR-Cas13 ortholog may address the limitations of Cas9 to target the tumour genome[22,25,40,128].

Cas13 effectors are programmable RNA-guided targeting enzymes that exclusively degrade ssRNAs with high efficacy and specificity[22,40]. The reversibility of RNA targeting with Cas13 represents a promising property of the approach, that would permit oncogene suppression without risking permanent alteration of the genome in somatic and germline cells, an inherent limitation of DNA-editing CRISPR enzymes[125,148,149]. The work presented below tests the possibility that Cas13 can effectively target aberrant mRNA transcripts that drive various human genetic diseases including cancer. This includes mRNA transcribed from gene fusions caused by chromosome rearrangements and which ultimately contribute to tumourigenesis by producing oncogenic fusion proteins[173]. In theory, Cas13 can degrade any type of tumour transcripts including overexpressed, point-mutated, and fusion transcripts. We believe that aberrant transcripts that are tumour-specific (e.g., gene fusions) are ideal targets as their expression is restricted to tumour cells but not normal cells. However, overexpressed transcripts are not exclusively expressed in tumour cells. Using Cas13 to target overexpressed transcripts will degrade these transcripts in both tumour and normal cells, which may cause some side effects. However, if only the tumour cells are addicted to the overexpression of this transcript, it is possible to find a therapeutic window where its knockdown only impacts tumour cells but not normal cells.

In **Chapter 2**, unbiased *in silico* analysis was used to reveal key determinants of *Psp*Cas13b potency and specificity, based on 'rules' and principles for effective crRNA design. On this basis, it is proposed that it should be possible to rationally design crRNAs that can degrade oncogenic fusion transcripts with near single-nucleotide precision while sparing their wildtype counterpart.

3.2 Materials and Methods

The design and cloning of all *Psp*Cas13b crRNAs used in the studies below are shown **(Appendix Table 1**). Methods for plasmid amplification and purification, quantification of RNA silencing, fluorescence microscopy, RNA extraction, cDNA synthesis, and RT-PCR are described in Chapters **2.2.1 to 2.2.10**.

3.2.1 Cloning of BCR-ABL1 T315I, BCR-ABL1, ABL1, and BCR fragments

Sequences for BCR-ABL1, ABL1, and BCR were derived from the full-length BCR-ABL1 P190 DNA sequence[174]. DNA was synthesised commercially by IDT, digested to completion with EcoRI/BamHI,

gel-purified, and ligated using T4 DNA ligase into the corresponding restriction sites of MSCV-IRES-mCherry, MSCV-IRES-eGFP, and MSCV-IRES-tagBFP plasmid DNA, in frame with 3xHA tag sequence (Promega, R6011/Promega, R6021). Ligated DNA was used to transform chemically competent bacteria (TOP10 or Stbl3) and positive clones were screened by PCR and verified by Sanger sequencing (AGRF, AUSTRALIA). The sequences of these DNA constructs and the primers used for PCR and Sanger sequencing are shown in **Appendix Tables 2** and **3**, respectively. To generate the BCR-ABL1 T315I mutant construct, site-directed mutagenesis was performed by using the Phusion Site-Directed Mutagenesis Kit (Thermo Fisher F541) with the following primers following 5' phosphorylation: (T315I_Forward: 5'CCCCGTTCTATATCATCATCGAGTTCATGACCTAC3' and T315I_Reverse: 5'GCTCCCGGGTGCAGACCCCAAGGAG 3').

3.2.2 Cell culture

HEK 293T cells (ATCC CRL-3216) were cultured in DMEM high glucose media (Thermo Fisher, 11965092) containing 10% heat-inactivated FBS (Thermo Fisher, 10100147), 100 mg/mL Penicillin/-Streptomycin (Thermo Fisher, 151401220), and 2 mM GlutaMAX (Thermo Fisher, A1286001) at 20-80% confluency at 37 °C and 10% CO_2. K562 (ATCC CCL-243) cells were cultured in RPMI 1640 media (Thermo Fisher, 11875083) containing 10% heat-inactivated fetal bovine serum and 100mg/mL Penicillin/-Streptomycin at 20-80% confluence at 37 °C and 5% CO_2. Cells were routinely tested for Mycoplasma and remained negative throughout.

3.2.3 Cell Nucleofection

All RNA silencing experiments in K562 cells were performed using optimized SF cell line 4D-Nucleofector X Kit S protocol (Lonza, V4XC-2032). For each reaction, $1x10^6$ cells resuspended in 20 µL SF Cell Line 4D-Nucleofector™ X Solution containing 3.5 µg crRNA plasmid were nucleofected using program FF-120, 4D-Nulceofector X unit (Lonza, AAF-1003X). Cells were placed in the incubator overnight for recovery and to enable crRNA expression. The next day, the cells were nucleofected in the same way with 4D-Nucleofector X Solution containing 3.5 µg Cas13-BFP plasmid DNA. Cells were incubated at 37 °C and 5% CO_2 for 48 hours, FACS sorted, and RNA silencing efficiency was monitored by RT-PCR, western blot, or FACS analysis.

3.2.4 Western Blotting

Cells were washed three times with ice-cold PBS and lysed on ice in RIPA lysis buffer [50 mM Tris (Sigma-Aldrich, T1530), pH 8.0, 150 mM NaCl, 1% NP-40 (Sigma-Aldrich, I18896), 0.1% SDS, 0.5% sodium deoxycholate (Sigma-Aldrich, D6750)] containing protease inhibitor cocktail (Roche, 04693159001) and phosphatase inhibitor cocktail (Roche, 4906845001). Samples were incubated for 30min at 4 °C with rotation (25 rpm) and centrifuged at 16,000 g for 10 min, 4 °C. Supernatant was transferred to a new tube. Protein concentrations were quantified using the Pierce BCA Protein Assay

Kit (Thermo Fisher, 23225) according to the manufacturer's instructions. A total of 10 µg of protein diluted in 1x Bolt LDS sample buffer (Thermo Fisher, B007) and 1x Bolt sample reducing agent (Thermo Fisher, B009) were denatured at 95 °C for 5 min. Samples were resolved by Bolt Bis-Tris Plus 4–12% gels (Thermo Fisher, NW04120BOX) in 1x MES SDS running buffer (Thermo Fisher, B0002) and transferred to 0.45 µM PVDF membranes (Thermo Fisher, 88518) by a Trans-Blot Semi-Dry electrophoretic transfer cell (Bio-Rad) at 20 Volt for 30 min. Alternatively, samples were resolved by 4-15% Criterion TGX Precast Midi Protein gels (Bio-Rad, 5671084) in 1x Tris/glycine/SDS running buffer (Bio-Rad, 1610732) and transferred to 0.20 µM nitrocellulose membranes (Bio-Rad, 1704159) by a Trans-Blot Turbo Transfer System (Bio-Rad) with a HIGH MW protocol. Membranes were incubated in blocking buffer 5% (w/v) BSA (Sigma-Aldrich, A3059) in TBST with 0.15% Tween 20 (Sigma-Aldrich, P1379) for 1h at RT and probed overnight with primary antibodies at 4 °C. Membranes were washed three times in TBST with 0.15% Tween20, followed by incubation with fluorophore-conjugated or HRP-conjugated secondary antibodies for 1h at RT. Membranes were washed in TBST (0.15% Tween20) three times and fluorescence or chemiluminescence was detected using the Odyssey CLx Imager 9140 (Li-cor), iBright CL1500 Imaging System (Thermo Fisher), or ChemiDoc Imaging System (Bio-Rad). The antibodies used for western blots **are listed (Appendix Table 5)**.

3.2.5 Cell flow cytometry

For intracellular staining, ~1x10^6 cells were resuspended in Fixation Buffer (BioLegend, 420801) and permeabilized with True-Phos Perm Buffer (BioLegend, 425401) according to the manufacturer's instructions. Samples were then incubated in 50 µL PBS/2% (v/v) FBS containing primary antibody for 1 hour at RT, followed by washing twice with 200µl PBS/2% (v/v) FBS. Next, samples were incubated with 50 µL PBS/2% (v/v) FBS containing an Alexa fluor 647 goat anti-rabbit IgG (H+L) secondary antibody (CAT# A21245) for 30 minutes at RT, followed by washing twice with 200µl PBS/2% (v/v) FBS. Finally, cells were re-suspended in 200 µL of PBS containing 2% FBS for flow cytometry analysis. All samples were analyzed by FACSymphony Cell Analyzer A5 or A3 (BD Biosciences). For cell sorting, cells were resuspended in PBS/2% (v/v) FBS containing 1 ng/mL propidium iodide (PI) in 5 mL FACS tubes, sorted using a BD FACSAria Fusion 5 or Fusion 3 instrument, and re-collected in fresh media for RT-PCR, western blotting, or proliferation assays. All flow cytometry profiles were analyzed using FlowJo V10 software (Tree Star Inc). The antibodies used for cell flow cytometry are listed **(Appendix Table 5)**.

3.2.6 Resazurin-based proliferation assay

Resazurin reagent was used to quantify K562 cell proliferation. Resazurin reagent contains 75 mg of Resazurin (Sigma, R7017), 12.5 mg of methylene blue (Sigma, MB-1), 164.5 mg of potassium hexacyanoferrate (III) (Sigma, P8131), and 211 mg potassium hexacyanoferrate (II) trihydrate (Sigma, P9387) in 500 mL PBS. On Day 0, 10,000 K562 cells were seeded in triplicate in 200 µL fresh media

per well in a 96-well plate. 40 µL of resazurin reagent was added to each well and incubated at 37 °C and 5% CO_2 for 4 hours, and absorbance was measured at 0, 24, 48, 72, and 96 hours using Cytation3 Plate Reader (BioTek, wavelengths 550 nm and 590 nm). *Imatinib* (1 µM final concentration) was used as a positive control to inhibit BCR-ABL1 and K562 proliferation and DMSO (0.01%) was the vehicle control.

3.2.7 Data analysis

Data analyses and displays were performed in GraphPad Prism software version 9 unless stated otherwise. Specific statistical tests and numbers of independent biological replicates are mentioned in the associated figure legends. The silencing efficiency of various crRNAs was assessed statistically using one-way ANOVA followed by Dunnett's multiple comparison tests where every mean was compared to a control mean as indicated in the Figures (95% confidence interval). The *P* values (P) are indicated in the Figures. $P < 0.05$ was considered significant.

3.3 Results

3.3.1 Efficient silencing of oncogenic fusion drivers

The breakpoint at the interface between the two genes contributing to the BCR-ABL1 gene fusion offers a unique targetable sequence at the RNA level, and enabled the design of crRNAs targeting the breakpoint characterised previously in 3 independent fusions: BCR-ABL1 p190, SFPQ-ABL1 and SNX2-ABL1, all of which are established drivers of acute lymphoblastic leukaemia (ALL)[175-177]. The DNA sequences encompassing the breakpoint were each cloned into an IRES-eGFP reporter plasmid that also includes in series an IRES-eGFP sequence. The IRES sequence ensures that eGFP protein is expressed independently of upstream coding sequences present on the same mRNA. When the chimeric RNA is cleaved by *Psp*Cas13b at the breakpoint, the resultant RNA fragments lose key features such as their 5'cap, 5'/3'UTRs, and polyA tail which impairs their intracellular stability and/or compromises their translation on the ribosome. As a result, both the cleaved fusion gene sequence and the downstream reporter gene (in this case, eGFP) undergo degradation and/or translational repression (**Figure 3.1**). Therefore, efficient cleavage of the fusion transcript by *Psp*Cas13b is also accompanied by the loss of eGFP fluorescence.

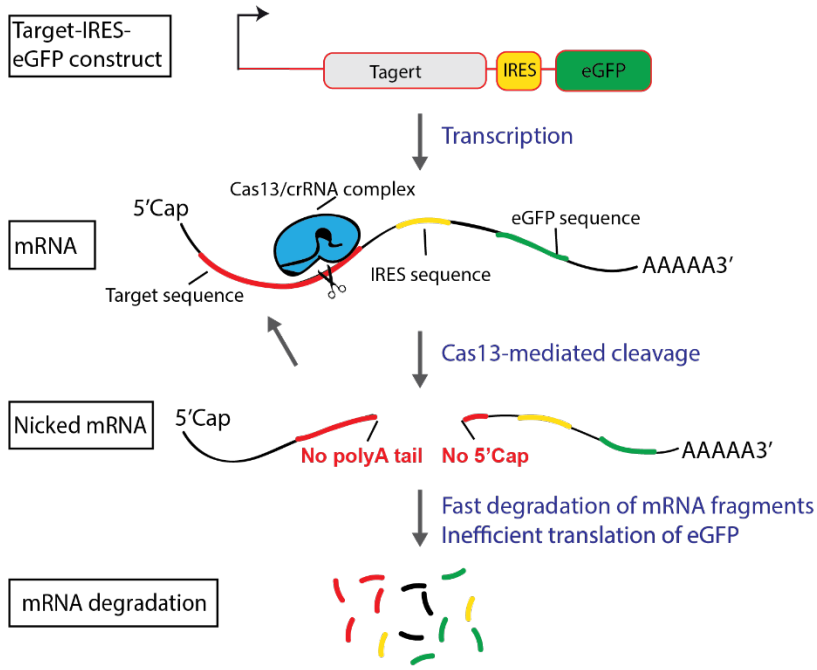


Figure 3. 1. Schematic of fluorescent reporter assay used to track degradation of various fusion transcripts.

The fusion genes were each cloned into an IRES-eGFP reporter construct, which is subsequently transcribed into a single chimeric RNA encoding a fusion protein, IRES sequence, and eGFP. *Psp*Cas13b cleaves the chimeric RNA at the fusion gene breakpoint. The RNA fragments lose key RNA features like their 5'cap, 5'/3'UTRs, and polyA tail, and therefore undergo RNA degradation and/or inefficient protein translation.

---

9 crRNAs tiled across the breakpoint were designed for each ABL1 fusion gene, in which the binding sites of consecutive crRNAs were separated by 3 nucleotides (3-nucleotide tiled crRNAs). Using fluorescence microscopy, highly efficient silencing of all 3 fusion genes was achieved, although the silencing efficiency varied with the crRNAs (**Figure 3.2a-3.2c**). Analysis of mRNA levels by RT-PCR confirmed that high silencing efficiency was achieved with many crRNAs (**Figure 3.2d-3.2f**). Comparing fluorescence and RT-PCR data confirmed efficient silencing, although moderate differences in the magnitude of silencing occurred for some, possibly due to a non-linear correlation between mRNA abundance and eGFP protein levels. Western blotting further confirmed high silencing of BCR-ABL1 at the protein level, with -12, -9, and +12 crRNAs showing the best silencing (**Figure 3.2g**). Suppressed STAT5 and ERK phosphorylation, a hallmark of BCR-ABL1-dependent oncogenic signalling (**Figure 3.2h**), confirmed that potent crRNA efficiently suppressed BCR-ABL1 and its downstream signalling network (**Figure 3.2i**). *Imatinib*, a small inhibitory molecule that blocks the

tyrosine kinase domain of ABL1 (**Figure 3.2h**), inhibited BCR-ABL1 mediated phosphorylation of STAT5 and ERK without altering the expression of BCR-ABL1 protein, whereas *Psp*Cas13b crRNAs efficiently silenced BCR-ABL1 protein expression as well as phosphorylation of STAT5 and ERK (**Figure 3.2i**). Interestingly, the most potent crRNA+12 showed even greater suppression of STAT5 phosphorylation than the dose of *Imatinib* used in this experiment, consistent with its highly efficient depletion of BCR-ABL1 protein through mRNA silencing (**Figure 3.2i**).

Next, single-nucleotide tiled crRNAs targeting the BCR-ABL1 breakpoint were tested. Forty-one single-base tiled crRNAs across the breakpoint (**Figure 3.2j**). Once again, the silencing efficiency varied considerably, even for neighbouring crRNAs. For instance, despite 96.6% sequence homology and only a single nucleotide position shift, crRNA+14 achieved >90% silencing while crRNA+15 exhibited none, confirmed with both fluorescence microscopy and Western blotting (**Figure 3.2j & 3.2k**). Consistent with these findings, crRNA+14 also potently inhibited STAT5 phosphorylation (**Figure 3.2k**). The contrasting silencing activity obtained with single-base resolved crRNAs within the same targeted region indicated that featured other than the RNA sequence itself can profoundly influence *Psp*Cas13b activity.

Next, the design principles described above **in Chapter 2.3.4** were applied to the *de novo* design of crRNAs that would efficiently target an otherwise undruggable oncogenic fusion transcript, RUNX1-RUNX1T1 which recruits chromatin corepressors to block cell differentiation in acute myeloid leukaemia (AML)[178], or NPM-ALK which encodes a constitutively active tyrosine kinase that promotes anaplastic large cell lymphoma[179]. Five crRNAs predicted to be efficient and to target, each mRNA breakpoint were tested. Both fluorescence microscopy and western blotting confirmed efficient silencing of both RUNX1-RUNX1T1 and NPM-ALK using this predictive approach (**Figure 3.2l-3.2q**). These results further showed that the design principles developed for *Psp*Cas13b and crRNA could be useful in silencing a wide variety of gene fusion transcripts.

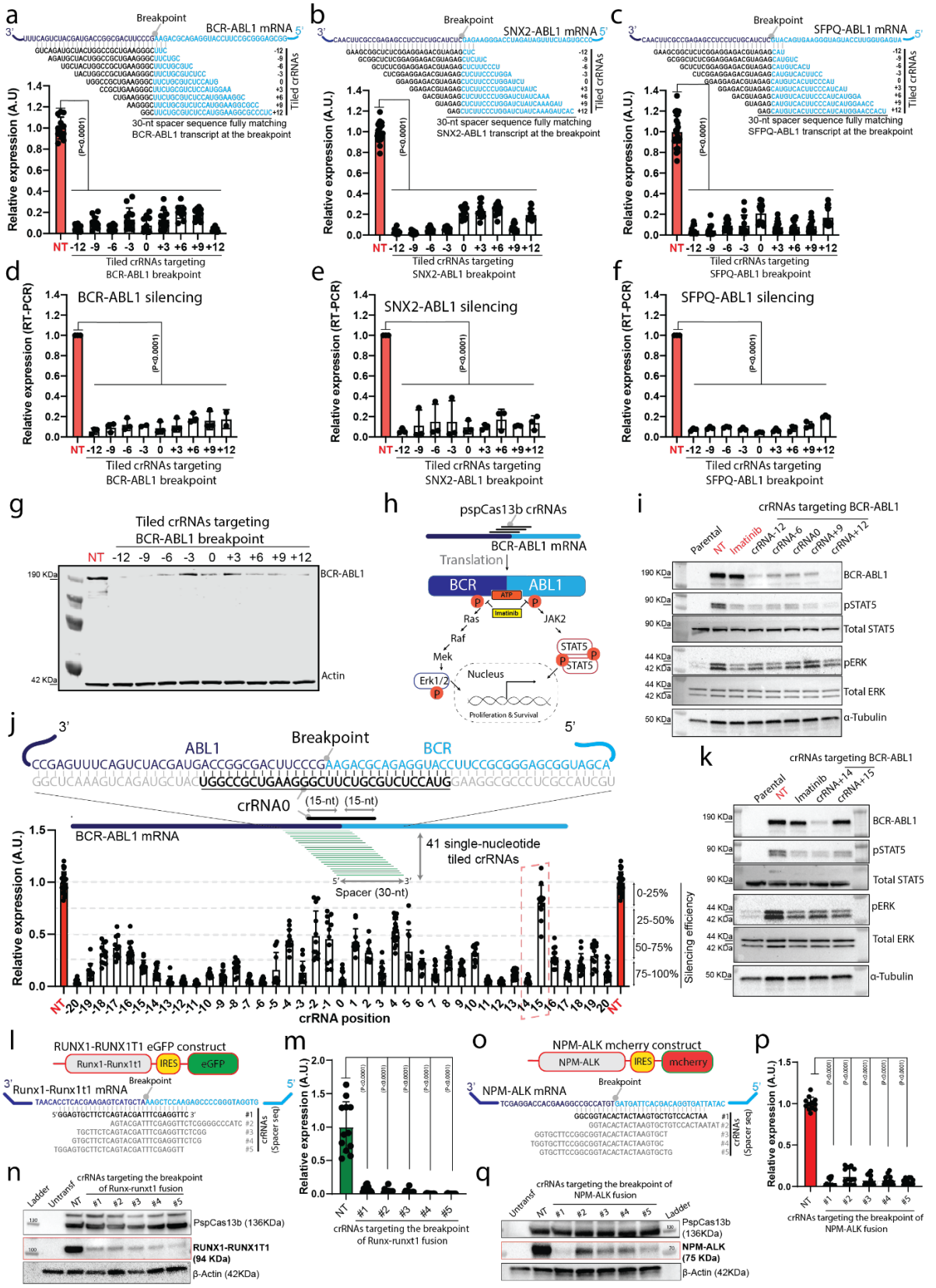


Figure 3. 2. Reprogrammed *Psp*Cas13b suppresses fusion gene transcripts with high efficiency.

(a-c) EGFP reporter assay to assess the silencing efficiency of tiled *Psp*Cas13b crRNAs with 3-nucleotide resolution targeting the breakpoint region of fusion transcripts BCR-ABL1 **(a)**, SNX2-ABL1 **(b)** and SFPQ-ABL1 **(c)**. Data points in the graphs are mean fluorescence from 4 representative fields of view per condition imaged; $N = 3$. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of one-way Anova test are indicated (95% confidence interval). **(d-f)** RT-PCR assays measuring silencing efficiency of tiled *Psp*Cas13b crRNAs (3-nucleotide resolution) targeting the breakpoint regions of fusion transcripts BCR-ABL1 **(d)**, SNX2-ABL1 **(e)** and SFPQ-ABL1 **(f)**; $N = 3$; Data are normalized means and errors are SEM; Results are analysed by one-way Anova test with *p*-values indicated (95% confidence interval). **(g)** Representative Western blot analysis to examine the expression level of BCR-ABL1 protein in HEK 293T cells expressing tiled crRNAs with 3-nucleotide increment targeting the breakpoint region of BCR-ABL1 transcripts 24 h post-transfection; $N = 3$. **(h)** Schematic of BCR-ABL1 dependent phosphorylation of ERK and Stat proteins, and inhibition of BCR-ABL1 oncogenic activity with Imatinib, a potent competitor of ATP binding to ABL kinase. **(i)** Representative Western blot analysis to examine the suppression of BCR-ABL1 expression and the subsequent inhibition of STAT5 and ERK phosphorylation in HEK 293T cells expressing BCR-ABL1, *Psp*Cas13b and either NT or crRNA targeting the BCR-ABL1 24 h post-transfection. HEK 293T cells expressing BCR-ABL1 and *Psp*Cas13b treated with 1μM imatinib for 4 hours were used as a positive control. Parental cells are HEK 293T cells transfected with *Psp*Cas13b, NT and a random control plasmid. This condition shows the baseline expression of pSTAT5 and pERK in a BCR-ABL1 independent manner; $N = 2$. **(j)** 41 single-nucleotide tiled crRNAs targeting the mRNA region surrounding the breakpoint of BCR-ABL1 to reveal the relationship between RNA sequence, accessibility, and *Psp*Cas13b silencing efficiency. The schematic shows the sequence of BCR-ABL1 RNA covered by 41 tiled crRNAs and RNA-RNA duplex formed by spacer-target interaction. The red dashed box highlights two adjacent crRNAs (14 & 15) with markedly contrasted silencing efficiency. Data points in the graph are normalized mean fluorescence from 4 representative fields of view imaged in $N = 3$. The data are represented in arbitrary units (A.U.). Errors are SD with a 95% confidence interval. **(k)** Representative western blot analysis to examine the silencing efficiency of single-base resolved crRNAs 14 & 15 that target BCR-ABL1 mRNA. crRNA potency is examined through the silencing of the BCR-ABL1 protein and phosphorylation of STAT5 and ERK proteins. Cells expressing BCR-ABL1, *Psp*Cas13b and either NT or crRNA targeting the BCR-ABL1 were harvested for WB analysis 24 h post-transfection. 1μM imatinib treatment for 4

hours was used as a positive control to inhibit BCR-ABL1 kinase activity. Parental cells are HEK 293T cells transfected with *Psp*Cas13b, NT and a control plasmid to examine the baseline expression of pSTAT5 and pERK in a BCR-ABL1 independent manner; *N* = 3. (l; o) Schematic of predicted *Psp*Cas13b crRNAs targeting the breakpoint sequence of RUNX1-RUNX1T1 (i) and NPM-ALK (o) mRNAs, respectively. (m; p) Quantification of the silencing efficiency of predicted potent crRNAs targeting RUNX1-RUNX1T1 (m) and NPM-ALK (p) fusion transcripts in HEK 293T cells. Data points in the graph are normalized mean fluorescence from 4 representative fields of view per experiment imaged in *N* = 3. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of one-way Anova test are indicated (95% confidence interval). (n; q) Representative western blot analysis to examine the expression level of RUNX1-RUNX1T1 (n) and NPM-ALK (q) in HEK 293T cells; *N*=3. *N* is the number of independent biological experiments.

3.3.2 *Psp*Cas13b accurately discriminates between a fusion RNA and its untranslocated partners

To determine whether *psp*Cas13b is sensitive to mismatches with oncogenic fusion RNAs, 3, 4, 5, 6, 7, 10, or 14 non-consecutive mismatches were introduced between the spacer of BCR-ABL1 crRNA (crBCR-ABL1) and the target breakpoint sequence (**Figure 3.3a**). The data revealed that 3 nucleotide mismatches were well tolerated, whereas incorporating 4 or more non-consecutive mismatches drastically impaired crRNA silencing efficiency (**Figure 3.3a**). 3 consecutive nucleotide mismatches at various positions did not affect the silencing of BCR-ABL1. 6 consecutive nucleotide mismatches at the 3'end (25-30) or the central region (12-17) led to a notable loss of silencing, while 9 consecutive nucleotide mismatches dramatically curtailed silencing irrespective of position (**Figure 3.3b**). Western blotting that assessed BCR-ABL1 protein expression confirmed these data and confirmed that 3-nucleotide mismatches were well tolerated, while 4-nucleotide mismatches or more led to substantially or completely preserved protein levels (**Figure 3.3c**).

To extend this study on targeting specificity, the impact of the crRNAs targeting BCR-ABL1 was assessed on wildtype (untranslocated) BCR and ABL1 mRNA expressed in the same cells. Constructs encoding partial mRNA sequences for BCR-ABL1, ABL1 alone, and BCR alone, were cloned in frame with mCherry, eGFP, or TagBFP fluorescence reporters, respectively (**Figure 3.3d-3.3g**). Three crRNAs targeting each of the BCR-ABL1 breakpoint, BCR (crBCR), or ABL1 (crABL1) were tested to assess targeting specificity (**Figure 3.3d**), as would be indicated by reduced fluorescence signals generated from mCherry, eGFP, or TagBFP. As anticipated, all 3 crRNAs silenced the *bona fide* BCR-ABL1 transcript as it possessed fully matched 30 nt binding sites for all three crRNAs (**Figure**

3.3e). However, ABL1 and BCR transcripts were silenced only by their cognate crABL1 and crBCR crRNAs (**Figure 3.3f-3.3g**); crBCR-ABL1, which targeted the breakpoint sequence had no effect on either BCR or ABL1 wildtype transcripts despite the 15-nucleotide sequence basepairing that remained (**Figure 3.3f-3.3g**). Western blotting again confirmed these findings at the protein level (**Figure 3.3e-3.3g**).

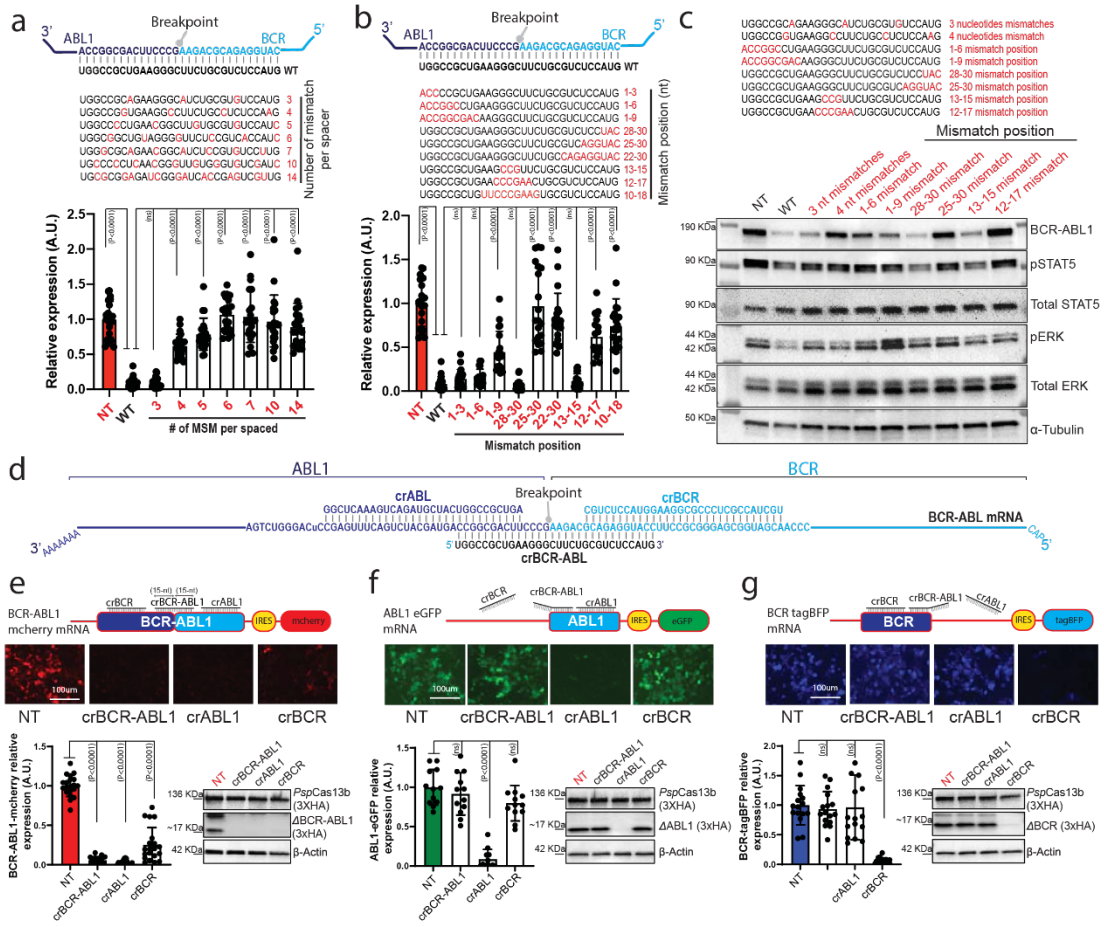


Figure 3. 3. The targeting of the breakpoint can efficiently discriminate between translocated tumour RNAs and wild type variants despite extensive sequence homology.

(a–b) Comprehensive analysis of spacer-target interaction examining specificity and mismatch tolerance of *Psp*Cas13b crRNAs targeting the breakpoint region of BCR-ABL1 transcript. The nucleotides in red highlight various mismatch positions in the spacer sequence. Data points in the graph are mean fluorescence from 4 representative field of views per condition imaged; $N = 4$. The data are represented in arbitrary units (A.U.). Errors are SD and p-values of one-way Anova test are indicated (95% confidence interval). (c) Western blot analysis examining the expression level of BCR-ABL1 protein and phosphorylation status of STAT5 and ERK in HEK 293T cells expressing crRNAs

with various mismatches 24 h post-transfection; *N* = 3. (d) The Schematic illustrates the binding sites of crRNAs designed to exclusively target BCR-ABL1 breakpoint (crBCR-ABL1), untranslocated BCR (crBCR), and untranslocated ABL1 (crABL1). (e-g) Three colour fluorescence-based reporter assays to assess the on-target specificity of crRNA targeting the breakpoint region of BCR-ABL1 and potential off-targeting of untranslocated BCR and ABL1 transcripts in HEK 293T cells 48 h post-transfection. The schematics show BCR-ABL1-mCherry mRNA (e), ABL1-eGFP mRNA, (f) and BCR-TagBFP mRNA (g) and their interaction with crBCR, crBCR-ABL1 and crABL1crRNAs through full, partial, or no spacer-target basepairing. Representative fluorescence microscopy images show the silencing efficiency of crBCR, crBCR-ABL1 and crABL1 crRNAs targeting BCR-ABL1-mCherry (e), ABL1-eGFP (f) and BCR-TagBFP transcripts (g) in HEK 293T cells. Scale bar = 100 µm. Similar results were obtained in 3 independent experiments. The histograms quantify the silencing of BCR-ABL1-mCherry (e), ABL1-eGFP (f) and BCR-TagBFP (g). Data points are normalized mean fluorescence from 4 representative fields of view per condition imaged. The data are represented in arbitrary units (A.U.). Errors are SD and *p*-values of one-way Anova test are indicated (95% confidence interval). *N* = 3. The right panels show representative western blot analyses to examine the expression levels of ΔBCR-ABL, ΔABL1, and ΔBCR proteins from data in e-g; *N* = 3.

3.3.3 High potency persists against drug-resistant point-mutated fusion transcripts

Acquired drug resistance to all approved ABL1 kinase inhibitors arises through secondary mutations and remains a major challenge in the treatment of BCR-ABL1-driven leukemias[180]. For instance, the BCR-ABL1 kinase domain mutation Thr315Ile (T315I) confers resistance to *Imatinib* and drives tumour relapse[181]. It was hypothesised that, unlike *Imatinib*, targeting the breakpoint of BCR-ABL1 transcript with potent crRNAs should remain effective against both ancestral and T315I BCR-ABL1 variants as the point mutation maps away from the targeted sequences. We tested the potency of either *Imatinib* or three *Psp*cas13b crRNAs targeting the breakpoint of BCR-ABL1. As anticipated, *Imatinib* efficiently inhibited the oncogenic signalling of ancestral BCR-ABL1 but failed to suppress downstream signalling by T315I BCR-ABL1 (**Figure 3.4**). Notably, all three *Psp*Cas13b crRNAs inhibited the expression of both the ancestral and T315I BCR-ABL1 proteins and their downstream oncogenic signalling as exemplified by pSTAT5 and pERK inhibition. Consistent with previous findings, crRNA-12, and crRNA+12 achieved the highest inhibition (**Figure 3.4**). These results also suggested that the persisting STAT5 and ERK phosphorylation was due to residual T315I protein remaining under these targeting conditions, which possesses greater activity than the ancestral BCR-ABL1. Using a higher dose of the potent crRNA may overcome this residual T315I signalling, as the

degree of *Psp*Cas13b silencing is highly dependent on crRNA abundance (**Figure 2.1**). Together, these data demonstrated that targeting the breakpoint of the BCR-ABL1 transcript can potentially avoid or overcome drug resistance that frequently triggers clinical relapse.

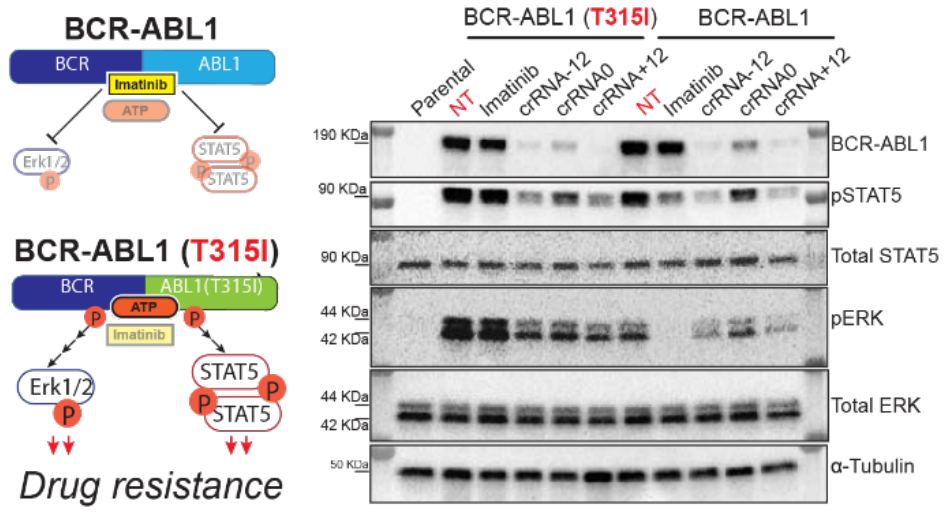


Figure 3. 4 Representative western blot analysis to examine the suppression of Imatinib-resistant T315I BCR-ABL1 with PspCas13b.

The schematic (left) illustrates the *Imatinib*-sensitivity or *Imatinib*-resistance of WT and T315I variants, respectively. BCR-ABL1 expression and STAT5/ERK phosphorylation status were examined in HEK 293T cells expressing WT or T315I BCR-ABL1 variants, *Psp*Cas13b, and either NT or crRNAs targeting the BCR-ABL1 breakpoint 24 h post-transfection. HEK 293T cells expressing BCR-ABL1 variants and *Psp*Cas13b were treated with 1μM imatinib for 4 hours as a positive control. Parental cells are HEK 293T cells transfected with *Psp*Cas13b, NT and a control plasmid, which shows the baseline expression of pSTAT5 and pERK in a BCR-ABL1 independent manner; $N = 3$. N is the number of independent biological replicates.

3.3.4 Potent silencing of fusion transcripts in patient-derived cancer cells

The above findings were generated from artificial over-expression models established in HEK-293T cells, but it remained to be established whether similar effects would occur in a patient-derived hemopoietic cancer cell line driven by the same gene translocation. K562 cells were originally derived from a 53-year-old patient with chronic myeloid leukaemia in blast crisis, and express the p210 fusion variant of BCR-ABL1 in which nucleotide 3315 of the BCR coding sequence and nucleotide 80 of the ABL1 coding sequence are juxtaposed in frame. First, the full-length BCR-ABL1 p210 sequence was cloned into the IRES-eGFP reporter construct, and the silencing efficiency of a predicted potent crRNA

(crRNA0) targeting the breakpoint region was assessed. crRNA0 recognises 15 nucleotides of BCR and 15 contiguous nucleotides of ABL1 at the breakpoint (**Figure 3.5a**). Fluorescence microscopy confirmed highly potent (>95%) silencing in HEK 293T cells (**Figure 3.5a**). crRNA0 was then tested for its ability to silence p210 mRNA in K562 cells. As K562 cells are refractory to conventional transfection and lentiviral transduction of *Psp*Cas13b is ineffective[35], sequential nucleofection was used as a new approach to deliver *Psp*Cas13b and crRNA plasmids, followed by cell sorting of the *Psp*Cas13b-expressing sub-population, using BFP as a surrogate marker (**Figure 3.5b**). Western blotting demonstrated efficient silencing of BCR-ABL1 p210, and resulting inhibition of STAT5 and ERK phosphorylation in the K562 cells, but the expression of wildtype ABL1 protein was unaffected (**Figure 3.5c**). Consistent with the western blots, RT-PCR confirmed efficient silencing of BCR-ABL1 mRNA, but no silencing of untranslocated BCR and ABL1 mRNA (**Figure 3.5d**). Flow cytometry also confirmed efficient suppression of STAT5 and ERK phosphorylation with *Psp*Cas13b, at a level similar to that achieved with *Imatinib* (**Figure 3.5e-3.5f**, **Figure 3.6)**. As K562 cells are highly dependent on BCR-ABL1 p210 for proliferation and survival, silencing with *Psp*Cas13b largely impaired K562 cell proliferation over 96h, as was also seen with *Imatinib* treatment (**Figure 3.5g-3.5i**). Together, these data confirmed that *Psp*Cas13b can potently and specifically silence fusion drivers in a patient-derived cancer cell line and in so doing, shut down oncogenic signalling to an extent that impairs cell proliferation but has no effect on untranslocated wildtype mRNA transcripts.

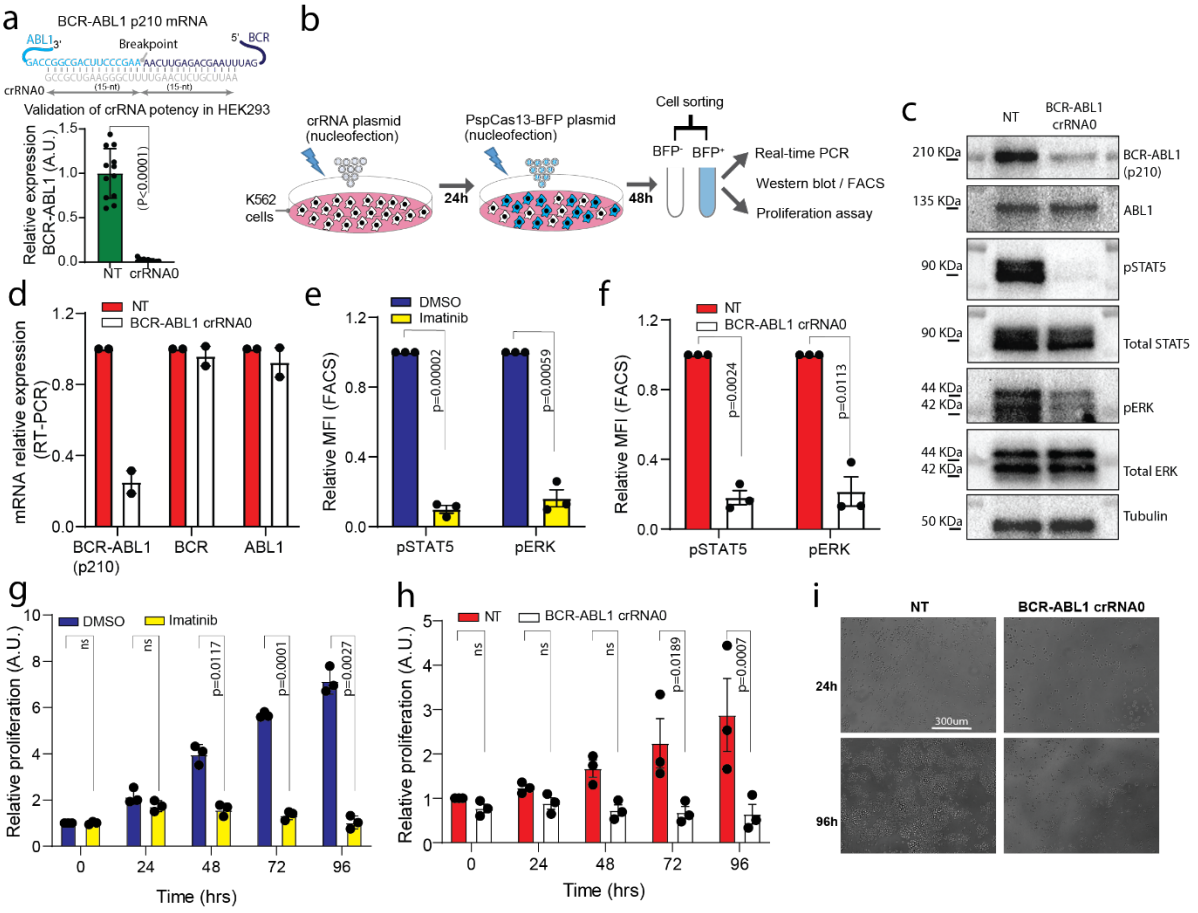


Figure 3. 5. *Psp*Cas13b-mediated silencing of BCR-ABL1 in the patient-derived cell line K562 suppresses downstream oncogenic signalling and cell proliferation.

(a) The schematic illustrates the binding site of crRNA0 to the breakpoint of BCR-ABL1 p210 mRNA. The histogram quantifies the silencing of BCR-ABL1 p210 mRNA in HEK 293T cells with predicted potent crRNA0. (b) Schematic of experimental pipelines to assess *Psp*Cas13b-mediated silencing of BCR-ABL1 p210 in K562 cells using this potent crRNA0. (c) Representative western blot analysis to examine the expression levels of BCR-ABL1 p210, untranslocated ABL1, and STAT5/ERK phosphorylation in K562 cells expressing either non-targeting crRNA (NT) or crRNA0 targeting BCR-ABl1 p210 mRNA; *N*=2. (d) RT-PCR assays to measure the expression levels of BCR-ABL1 p210, untranslocated BCR, and untranslocated ABL1 mRNAs in K562 cells expressing non-targeting (NT) or crRNA0 targeting BCR-ABL1 p210 mRNA; *N*=2. Data are normalized means and errors are SEM. (e-f) Intracellular cell flow cytometry measuring STAT5 and ERK phosphorylation levels in K562 cells following *Imatinib* treatment (e) or BCR-ABL1 p210 silencing with *Psp*Cas13b/crRNA0 (f); *N*=3. Data are normalized mean fluorescence intensity (MFI) and errors are SEM; Results are analyses by

unpaired two-tailed Student's t-test (95% confidence interval). Alarm blue assays to assess K562 cell proliferation following *imatinib* treatment (g) or BCR-ABL1 p210 mRNA silencing with *Psp*Cas13b/crRNA0 (h) throughout a 96h time window; *N*=3. Data are normalized means and errors are SEM; Results are analysed by two-way ANOVA test (95% confidence interval). (i) Representative microscopy images show K562 cell density at 24 and 96h time points post *Psp*Cas13b/crRNA delivery in cells expressing either non-targeting (NT) or crRNA0 targeting the BCR-ABL1 p210 mRNA. Scale bar = 300 μm. Similar results were obtained in *N*=3. *N* is the number of independent biological experiments.

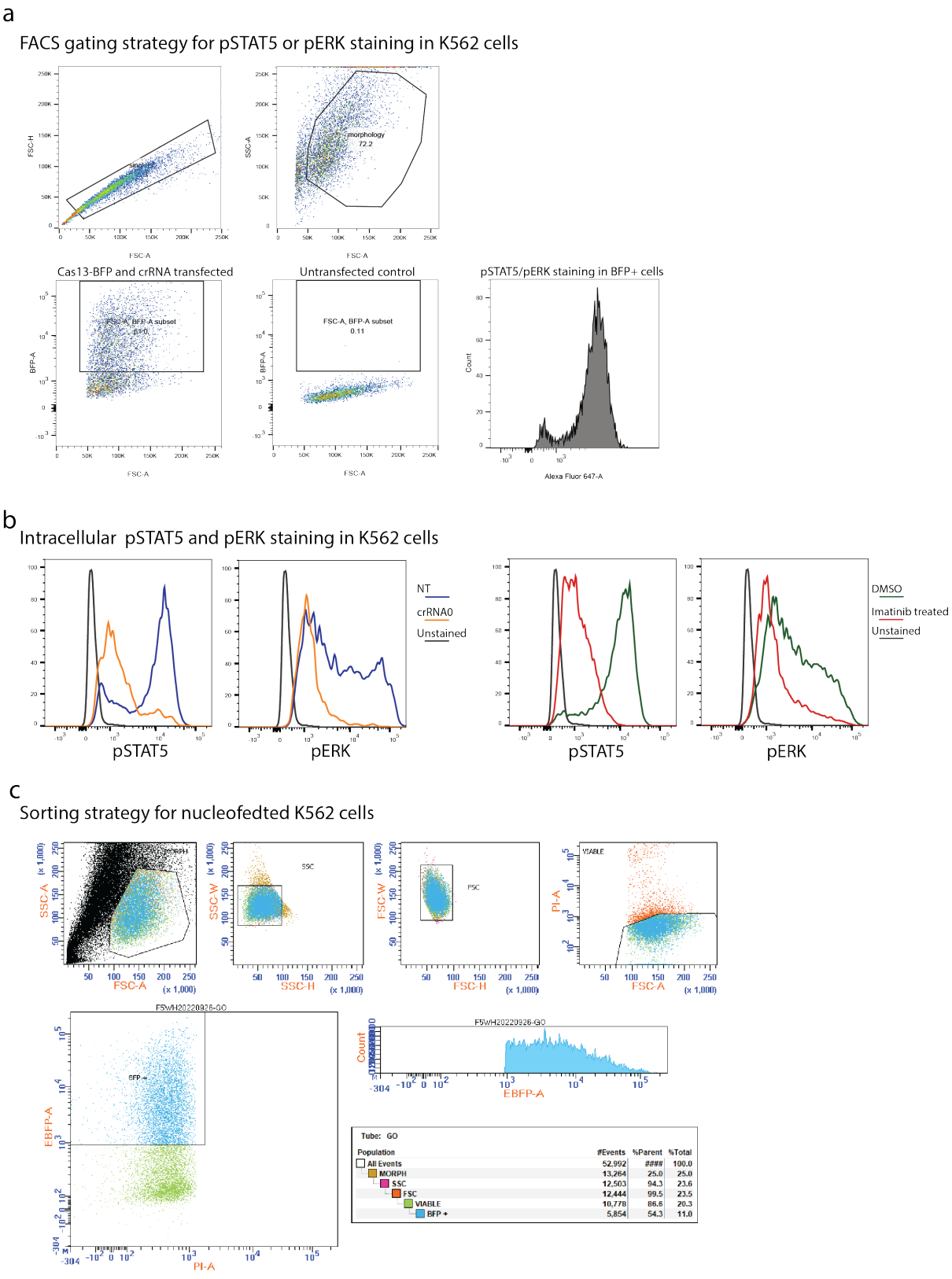


Figure 3. 6. (related to Figure 3.5). FACS gating strategy.

(a) FACS gating strategy for intracellular pSTAT5 and pERK staining in K562 cells post nucleofection with *Psp*Cas13b and crRNA. The singlets were gated to exclude aggregations, followed by gating

on K562 cell morphology, and gating on BFP positive population. In the BFP positive population, expression levels of pSTAT5 and pERK was probed. The right panel displays a representative pSTAT5 FACS histogram obtained with this gating strategy. (b) Representative histograms of pSTAT5 and pERK staining in K562 cells expressing *Psp*Cas13/crRNA (left panel) or treated with *Imatinib* (right panel). (c) FACS sorting strategy for *Psp*Cas13b (BFP) positive population in nucleofected K562 cells. K562 cells were stained with Propidium Iodide (PI) for cell viability. K562 cells were gated by morphology followed by gating on PI-negative population to exclude dead cells from the sorted cell population. Then BFP positive population that express *Psp*Cas13b nuclease for sorting was gated.

3.4 Discussion

A highly desirable attribute of *Psp*Cas13b-mediated target silencing would be its potential to silence treatment-resistant or otherwise '*undruggable*' oncogenic gene fusions[140,146,182]. Over 3,297 unique fusion transcripts across multiple cancer types have been described[32], most of which remain '*undruggable*' but could in principle be targeted by personalized *Psp*Cas13b drugs. Here, it was shown that *Psp*Cas13b can efficiently recognise and silence five different fusion transcripts including BCR-ABL1, a well-established driver of leukaemia, with *in vitro* potency comparable to that of *Imatinib*[167,176,183-185] (**Figure 3.2-Figure 3.5**). BCR-ABL1 transcript silencing led to subsequent depletion of the fusion protein and thereby inhibited the phosphorylation and activation of downstream STAT5 and ERK signalling that is a hallmark of BCR-ABL1 driven cancers, and in these experiments, were at least as effective as the tyrosine kinase inhibitor *Imatinib* (**Figure 3.2-Figure 3.5**). The use of *imatinib* in treating BCR-ABL1-positive leukaemia can be associated with severe toxicity due to the non-specific inhibition of ABL1 and ABL1-like kinases in normal tissues[186]. Conversely, custom-designed crRNAs targeting the breakpoint of BCR-ABL1 p190 in HEK 293T cells and BCR-ABL1 p210 in K562 cells exhibited high potency and exquisite specificity (**Figure 3.2-Figure 3.5**). *Psp*Cas13b silenced the fusion drivers, downregulated oncogenic signalling pathways, and impaired cell proliferation without any detectable off-targeting of untranslocated BCR or ABL1 transcripts in K562 cells. Additional western blots, transcriptomic, proteomic, and phosphoproteomic analysis could be performed to further investigate the global impact of BCR-ABL1 inhibition with Cas13 in K562 cells.

In addition to drug toxicity, acquired resistance through secondary mutations remains a major challenge in the management of cancers. For example, the BCR-ABL1 kinase domain "gatekeeper" mutation T315I confers resistance to first-line ABL1 inhibitors, including *Imatinib,* and often drives

tumour relapse in patients[181]. Unlike *Imatinib*, potent *Psp*Cas13b crRNAs targeting the breakpoint of BCR-ABL1 transcript were highly effective against both the parental and T315I BCR-ABL1 variants and suppressed their downstream oncogenic networks (**Figure 3.4-3.5**). These data demonstrate the great potential of this *Psp*Cas13b tool to silence '*undruggable*' fusion transcripts that drive various tumours and overcome their mutation-driven drug resistance (**Figure 3.7**).

However, Cas13 therapeutic is in its infancy and further optimization of dosing and delivery modality is required to enhance its potency *in vitro* and *in vivo*. After these optimizations, Cas13 therapeutics may outperform the efficacy and specificity of *imatinib*, especially in terms of reducing side effects and maintaining potency against BCR-ABL1 variants that drive resistance. Further experiments will be performed to optimize the dosing and delivery both *in vivo* and *in vitro* as well as benchmarking Cas13 alone against *imatinib* and a combination of Cas13 and *imatinib*.

Excitingly, the breakpoint sequence of a given fusion driver can be easily obtained from patients' tumour RNA sequencing[182], which would enable the manufacture of personalized *Psp*Cas13b drugs in just a few days (**Figure 3.7**). Assuming the problem of effective delivery can be overcome *in vivo*, such a drug development pipeline would be suitable for both cancer discovery as well as potential clinical translation in an actionable time frame and for a relatively modest cost.

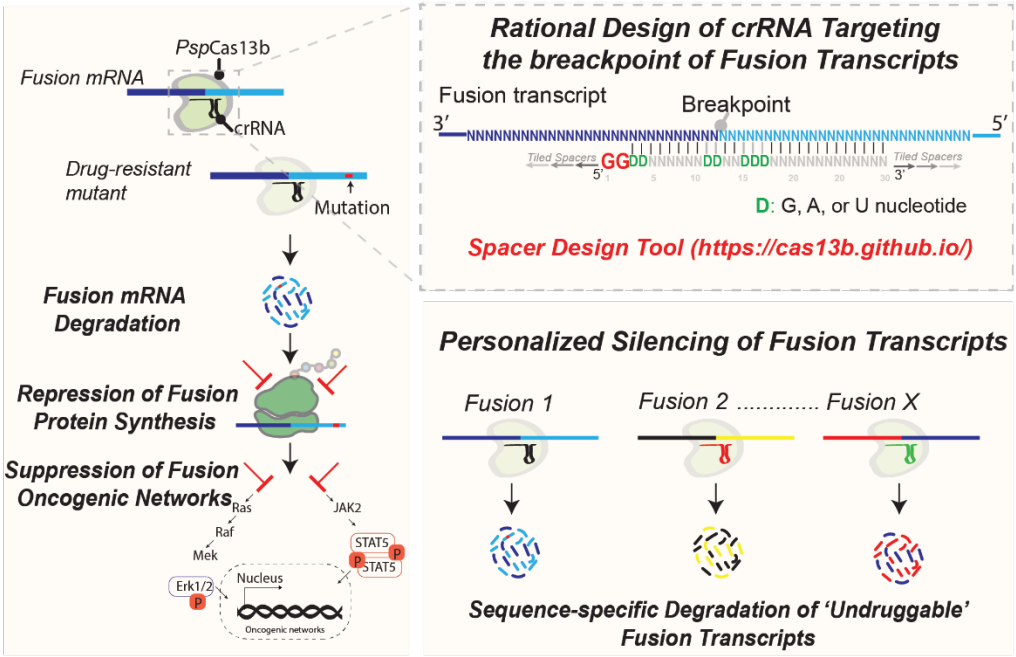


Figure 3. 7. Design principles for potent, specific, and personalized silencing of various gene fusion mRNAs with programmable *Psp*Cas13b nuclease.

The left panel shows the rational design of crRNAs targeting the breakpoint of gene fusion transcripts through spacer-target basepairing, which enables efficient and specific silencing of these oncogenic drivers, as well as drug-resistant fusion mutants. *Psp*Cas13b-mediated RNA degradation suppresses fusion protein translation and their downstream oncogenic networks. This targeting of the breakpoint strategy avoids the off targeting of untranslocated gene partners that are expressed in normal tissues. The top right panel summarizes spacer nucleotide positions that are important for the rational design of potent crRNA, which are integrated into our *in-silico* tool (https://cas13b.github.io/). The substitution of the first two nucleotides of the spacer with a GG sequence can enhance the potency of otherwise inefficient crRNAs despite target mismatch. The bottom right panel shows the versatility and therapeutic potential of *Psp*Cas13b to silence various '*undruggable*' gene fusion transcripts and their oncogenic networks in a personalised manner through precise targeting of fusion's breakpoint sequence that is exclusively expressed in tumour cells.

Chapter 4

Parameters that dictate the specificity of CRISPR-*Psp*Cas13b nuclease activity in mammalian cells

Abstract

*Psp*Cas13b is a promising programmable RNA nuclease that may enable efficient and specific transcriptome editing for biotechnology and therapeutic applications. However, a thorough understanding of *Psp*Cas13b's potential for off-target and collateral nuclease activities is essential if transcriptome editing applications are ever to be considered safe for human clinical trials. In **Chapter 3**, *Psp*Cas13b was programmed to successfully target the breakpoint of oncogenic fusion transcripts and associated mutations that result in resistance to small drug therapy. This was achieved without off-targeting the parental wild-type mRNAs that contribute to the gene fusion, as the incomplete basepairing of the spacer with these wildtype mRNA fails to activate PspCas13b nuclease domain. However, the precise level of mismatch tolerance of *Psp*Cas13b to maintain the catalytic activity and trigger target mRNA degradation was not determined. Understanding the level of mismatch tolerance for *Psp*Cas13b should uncover further new design principles that might ultimately enable single-base precision RNA targeting and building a probability-based predictive model for transcriptome-wide off-target effects. Related to these issues, recent reports have shown several Cas13 orthologues to also possess non-specific 'collateral cleavage' activity that can trigger indiscriminate degradation of RNA molecules at close proximity to the HEPN nuclease domains. Here, comprehensive mutagenesis, RNA sequencing, and mass spectrometry approaches were used to investigate the specificity of *Psp*Cas13, to help define the interface between mismatch tolerance/intolerance, and reveal its on-target versus collateral activity in human cells with great details. It was found that *Psp*Cas13b can tolerate up to 3-nt mismatches within its 30-nt long crRNA spacer without loss of activity against the cognate target sequence. Importantly, this was achieved without incurring any off-target or collateral nuclease activities against the proteome of human cells. Together, the study reveals that extraordinarily precise targeting can be achieved using *Psp*Cas13 without adverse cellular impacts and supports its deployment for accurate manipulation of the eukaryotic transcriptome.

4.1 Introduction

The specificity of various Cas13 orthologues can be hindered by two intrinsic properties of these enzymes, which are (**i**) spacer mismatch tolerance and (**ii**) collateral nuclease activity that could lead to two distinct RNA off-targeting scenarios.

4.1.1 Spacer mismatch tolerance

The specificity of RNA-guided nucleases including Cas13 is determined by the length of their spacer sequence and its promiscuity when sequence mismatches occur. From the first principle, nucleases with long spacer sequences should provide higher specificity than counterparts with shorter spacer sequences. In fact, argonaut nucleoprotein that orchestrates eukaryotic RNAi has only moderate specificity due to a 7-8 nt surface-exposed RNA motif called the 'seed region' in its miRNA/siRNA, which coordinates target search and target recognition processes[73,74]. This seed sequence is too short to offer high

specificity as there are theoretically close to a million potential RNA targets within the mammalian transcriptome that could base-pair exactly with it[74]. On the other hand, Cas13 nucleases are known to possess longer spacer sequences that range from 22 to 30 nt, which offer many orders of higher specificity than the classic eukaryotic RNAi. Nonetheless, Cas13 enzymes can tolerate several mismatches with the target, which may compromise its absolute specificity. Indeed, the trade-off between spacer length and mismatch tolerance determines the relative specificity of these RNA-guided RNA targeting tools. To evaluate Cas13 specificity and the likelihood of off-targeting other endogenous RNA molecules that share partial sequence homology with a given target, a comprehensive mutagenesis study was carried out where defined numbers of consecutive or non-consecutive mismatches were introduced at different positions of the spacer-target RNA-RNA duplex.

Previous studies have shown that the orthologues *Lwa*Cas13a and *Rfx*Cas13d are highly specific and tolerate only 2-3 mismatches while maintaining efficient on-target cleavage[39,59]. However, these studies were performed using pooled library screens using a low multiplicity of infection (MOI) to ensure delivery of a single crRNA per cell, and research labs including our own have demonstrated that Cas13-mediated RNA interference is highly dependent on the intracellular concentration of crRNAs[139,187] (**Figure 2.1**). In fact, potent silencing with Cas13 nucleases requires the delivery of several copies of the crRNA template (as plasmid or lentiviruses) as several thousand copies of the target mRNA need to be identified and degraded in a typical cell. Cas13 pooled library screens are limited by the delivery of a single copy of crRNA per cell, which may yield poor silencing and inconclusive results[187]. Indeed, pooled library screens are more suitable for studying DNA editing enzymes (e.g., Cas9) as only two DNA alleles require editing in most instances[188].

4.1.2 Indiscriminate 'collateral' ribonuclease activity

In addition to the issue of mismatch tolerance discussed above, some Cas13 orthologues have been shown to possess collateral activity, in which spacer-target basepairing can trigger indiscriminate degradation of host RNAs that are at the vicinity of their activated HEPN domains. It has been proposed that the collateral activity of Cas13 may have evolved to impose cell dormancy in a bacterium infected with bacteriophages[62]. This may prevent further replication and dissemination of the bacteriophages, which would ultimately protect the bacterial community at the cost of an altruistic individual bacterium[62]. Therefore, there is an important distinction between the collateral cleavage and the off-target effects caused by mismatch discussed discussed above. In principle, a Cas13 orthologue may have very specific on-target activity due to low or no mismatch tolerance, and yet exhibit strong target-dependent collateral cleavage that leads to off-target degradation of multiple cellular RNAs in the vicinity. For example, *Lwa*Cas13 has been shown to possess strong collateral activity in bacteria and *in vitro,* which was exploited to develop highly sensitive diagnostic tools known as SHERLOCK to detect ssRNAs[53,63]. In this assay, the target recognition process activates the HEPN nuclease domains of *Lwa*Cas13a, which induces the collateral degradation of a reporter ssRNA in the test tube[53]. It is

remarkable that this broad collateral activity still occurs only when the spacer of a crRNA recognises its complementary target sequence. In this scenario, the collateral activity could not occur in any cell in which the gene that encodes the target mRNA is transcriptionally silent.

Cas13 can have versatile applications for precise RNA manipulations in mammalian cells. For instance, some RNA silencing applications such as pooled library screens would require Cas13 orthologues that are free of any collateral activity to unambiguously elucidate the biological role of each target RNA. In contrast, Cas13 orthologues with pronounced collateral activity in mammalian cells may still prove valuable for other purposes such as inducing programmable suicide of virus-infected or tumour cells that express aberrant RNAs that are not present in other cells. While collateral activity is well studied in biochemical *in-vitro* assays and in bacteria[24,90], its extent in mammalian cells remains controversial[28,189]. Unlike bacteria and *in vitro* assays, eukaryotic cells have a higher cellular organization and compartmentalized architecture where RNA concentration and RNA-protein interactions are tightly regulated, which may affect Cas13 collateral activity. Also, the various Cas13 orthologues studied arise from a plethora of bacterial species, have low sequence homology, distinct molecular architectures, and mechanisms of action, and so, may exhibit varying collateral activity in mammalian cells. It is therefore perhaps not surprising that the recent studies that have investigated the collateral activity of various Cas13 enzymes in mammalian cells have produced some contradictory findings. Adding to the complexity, multiple Cas13 orthologues have been studied in different cell types, targeting mRNAs with varying abundance, and utilising different measures of effectiveness[67,151,152,190]. Most of these studies measured fluorescence signals arising from overexpressed exogenous reporter genes as a proxy of collateral activity, so the true level of collateral activity on endogenous mammalian genes and the proteins they encode remains elusive. Altogether, the current poor understanding of *Psp*Cas13b on-target specificity, mismatch tolerance, and collateral activity limits its use as a precise RNA editing tool. To address this matter, the experiments presented in this chapter aim to reveal the specificity of *Psp*Cas13b, its off-targeting potential, and the extent of its collateral activity in mammalian cells.

4.2 Materials and Methods

crRNAs for *Psp*Cas13b **are listed (Appendix Table 1**). Methods for plasmid amplification and purification, cell culture, RNA silencing, fluorescence microscopy, RNA extraction, cDNA synthesis, RT-PCR and western blotting please are described in **Chapters 2.2.1 to 2.2.10**.

4.2.1 Assessing RNA integrity

RNA extracted from cells was analysed on an Agilent 2200 Tapestation (Agilent, G2964AA) with RNA ScreenTape (Agilent, 5067- 5576) and an RNA ScreenTape Ladder (Agilent, 5067- 5578) according to manufacturer's instruction.

4.2.2 RNA sequencing

Total RNA was isolated as described in Chapter 2.2.8, and measured by Qubit assays to determine dsDNA contamination. RNA was processed by the Molecular Genomics Facility (PeterMac) using the QuantSeq 3′ mRNA-Seq Library Prep Kit (Lexogen) and sequenced on an Illumina NextSeq 500. The raw data files were used and referenced against the modified Human reference transcriptome (GRCh37 ver.19) with eGFP, BCR-ABL1 p190 RNA sequences added using Galaxy Australia HISAT2 functions. Genes without more than 1 CPM in at least 3 samples were considered to be insignificant and filtered out. Read counts were normalized for effective library size, and differentially expressed genes (DEGs) were analyzed using Galaxy. Differential expression analysis between targeting and non-targeting samples was performed using the voom-limma (v3.42.2) pipeline[191]. Genes with log2 fold change (FC) >1 were considered upregulated while genes with log2 FC <-1 were considered downregulated. Adjusted p-value (with 0.05 false discovery rate) < 0.05 was considered statistically significant.

4.2.3 Mass spectrometry

Protein lysates (50 µg in RIPR buffer) underwent chemical reduction for 45 min at 37°C by adding tris(2-carboxyethyl) phosphine to 10mM final concentration, then alkylated for 45 minutes at 37°C in the dark using iodoacetamide (55mM final concentration). Samples were then acidified by the addition of phosphoric acid (2.5% final concentration). S-Trap binding buffer (90% MeOH, 100 mM final TEAB, pH 7.1) was added to the acidified lysate, and samples were passed through Suspension trapping (S-trap) columns. Bound elements in the samples were washed three times with the S-Trap binding buffer and digested overnight at 37°C by adding trypsin (in 50 mM TEAB) at a 1:10 ratio. The resultant peptides were eluted with 40 µL of 50 mM TEAB, followed by 40 µL 0.2% aqueous formic acid, and finally with 35 µL of 50% acetonitrile containing 0.2% formic acid. Lastly, samples were centrifuged, lyophilized, and reconstituted in 2% acetonitrile, and 0.1% trifluoroacetic acid.

Samples were analyzed by LC-MS/MS using Orbitrap Exploris 480 (Thermo Scientific) fitted with nanoflow reversed-phase-HPLC (Ultimate 3000 RSLC, Dionex). The nano-LC system was equipped with an Acclaim Pepmap nano-trap column (Dionex – C18, 100 Å, 75 µm × 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex – C18, 100 Å, 75 µm × 50 cm). Typically for each LC-MS/MS run, 1 µL of the peptide mix was loaded onto the enrichment (trap) column at an isocratic flow of 5 µL/min of 2% Acetonitrile containing 0.05% trifluoroacetic acid for 6 min before the analytical column is switched in-line. The buffers used for the LC were 0.1% v/v formic acid in water (solvent A) and 100% Acetonitrile/0.1% formic acid v/v (Solvent B). The gradient used was 3% B to 23% B for 89 min, 23% B to 40% B in 10 min, 40% B to 80% B in 5 min, and maintained at 80% B for the final 5 min before equilibration for 10 min at 2% B prior to the next analysis. All spectra were acquired in positive mode with full scan MS spectra scanning from m/z 300-1600 at 120,000 resolution with AGC target of $3x10^6$ with a maximum accumulation time of 25 ms. The peptide ions with charge

state ≥2-6 were isolated with an isolation window of 1.2 m/z and fragmented with a normalized collision energy of 30 at 15,000 resolution.

The raw data files were searched against the modified Human reference proteomes (UP000005640 with eGFP, BCR-ABL1 p190, and *Psp*Cas13b amino acid sequences added manually) using MaxQuant-Andromeda (version 1.6.17.0). Search parameters were set as follows: variable modifications: methionine oxidation; fixed modifications: cysteine carbamidomethylation; trypsin digestion with a maximum of 2 missed cleavages. Differential expression analysis between targeting and non-targeting samples was performed using Perseus (Version 2.0.7.0)[192]. Proteins with log2 FC >1 were considered upregulated while those with log2 FC <-1 were considered downregulated. Adjusted p-value (with 0.01 false discovery rate) < 0.05 was considered statistically significant. Genes that were significantly up-or down-regulated between targeting and NT crRNAs were visualised in both a volcano plot and a linear regression plot.

4.3 Results

4.3.1 Comprehensive mutagenesis of spacer-target interaction

To investigate crRNA spacer promiscuity and its consequences for *Psp*Cas13b targeting, a comprehensive series of mismatches were introduced at various positions of the spacer-target RNA-RNA duplex. This approach would test the hypothesis that *Psp*Cas13b can tolerate a small number of sequence mismatches without affecting its catalytic activity, whereas a greater number of mismatches would be disruptive, either through spacer-target destabilisation or by constraining the conformational changes necessary for HEPN nuclease activation (**Figure 4.1**). To test this proposition, a quantitative fluorescence-based silencing assay was established, in which mCherry reporter mRNA was targeted. HEK 293T cells were co-transfected with plasmids encoding mCherry, *Psp*Cas13b-BFP, and either non-targeting (NT) or mCherry-targeting crRNA. Firstly, 3, 6, 9, 12, 15, 18, 21, 24, 27, or 30-nt successive mismatches were introduced extending from each end of the spacer (**Figure 4.2a & b**). 3-nt mismatches at the 3'end did not affect silencing efficiency but mismatches greater than 3-nt completely abrogated crRNA silencing (**Figure 4.2a**). In contrast, all 5'end mismatches tested caused complete loss of silencing, including 3-nt mismatches (**Figure 4.2b**). Based on earlier findings (**Figure 2.4 & Figure 2.7**), silencing loss due to a 3-nt mismatch at the 5'end may be attributable to the substitution of a GGG motif by CCC rather than a numerical spacer-target mismatch, reaffirming the importance of a G-rich 5'end of potent crRNAs.

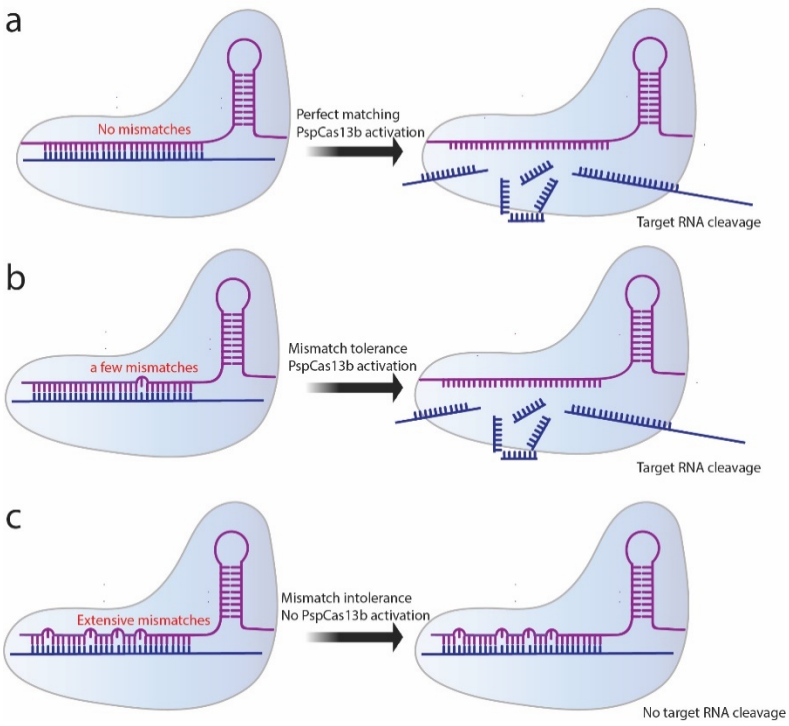


Figure 4. 1. Schematic representation of the predicted fates of target RNAs that fully match the *Psp*Cas13b crRNA spacer sequence (a), have few (b) or extensive (c) mismatches with the spacer.

---

crRNA constructs harbouring 6-nt, 5-nt, 4-nt, and 3-nt mismatches at different spacer positions were next tested (**Figure 4.2c-f**). Overall, 6-nt mismatches compromised the efficiency of *Psp*Cas13b regardless of their position (**Figure 4.2c**). 5-nt mismatches at positions 6-10, 11-15, and 26-30 exhibited a partial loss of silencing, while mismatches at positions 1-5, 16-20, and 21-25 led to a near-complete or complete loss of silencing (**Figure 4.2d**). 4-nt mismatches at positions 9-12, 13-16, and 17-20 retained partial silencing, whereas mismatches at positions 1-4, 5-8, 21-24, and 25-28 yielded a complete loss of silencing (**Figure 4.2e**). By contrast, 3-nt mismatches at various spacer positions were well tolerated, causing no or minor loss of silencing, except at positions 1-3, as explained above (**Figure 4.2f**).

Next, the impact of non-consecutive mismatches was investigated by introducing 2, 3, 4, 5, 6, 7, 10, or 15 non-consecutive mismatches spread throughout the spacer (**Figure 4.2g**). In general, 2, 3, and 4 non-consecutive mismatches had minimal effect on silencing but introducing more than 4 non-consecutive mismatches led to substantial or complete loss of silencing. Likewise, introducing multiple 2 or 3 nucleotide mismatches completely abolished silencing (**Figure 4.2g**). These data suggest that

introducing >4-nt non-consecutive mismatches critically destabilises spacer-target interactions and compromise *Psp*Cas13b activity. From a probabilistic point of view, the data also suggest that endogenous targets with partial sequence homology are very unlikely to be impacted by off-target silencing due to the required minimum of ~25 consecutive or non-consecutive nucleotide basepairing.

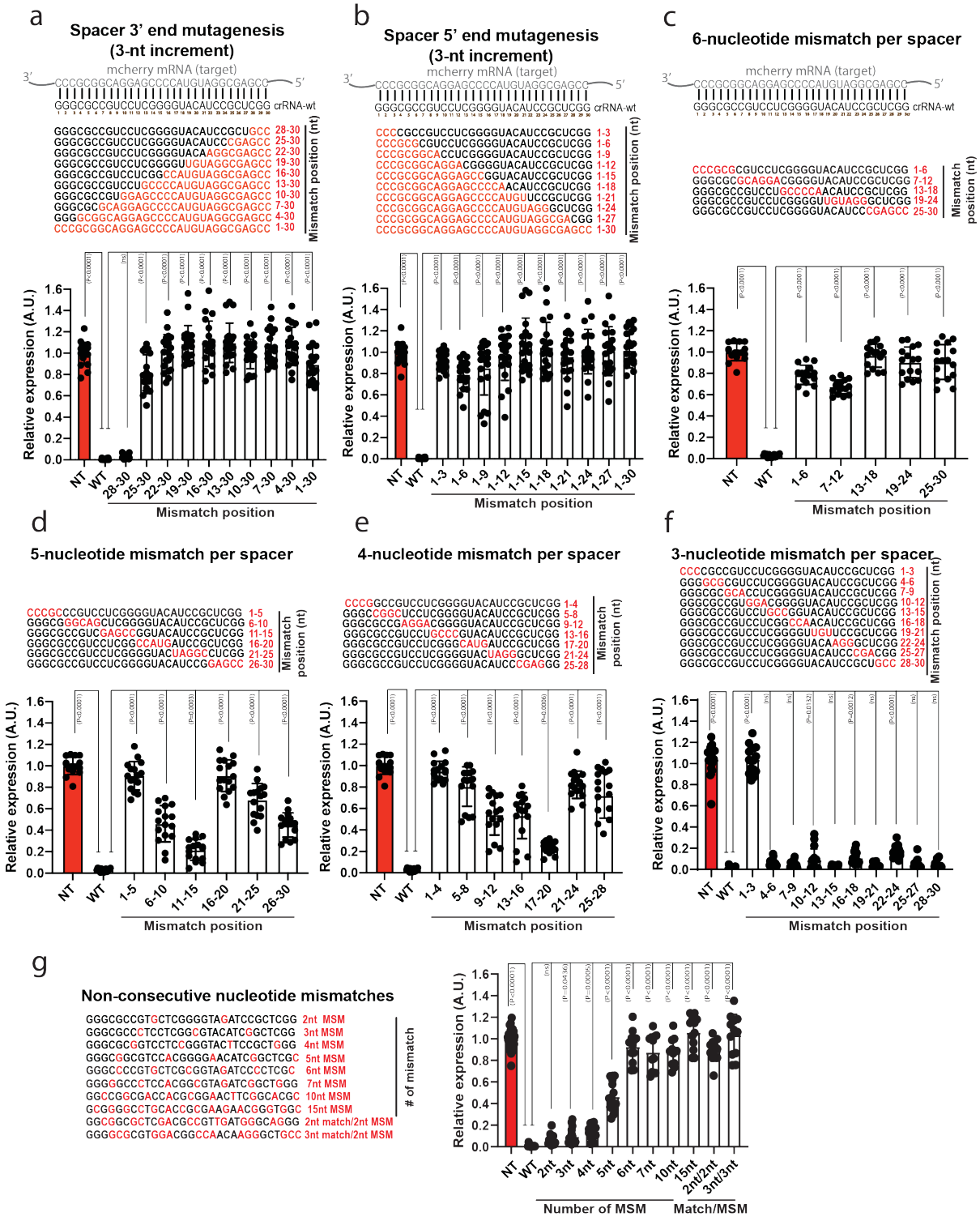


Figure 4. 2. Limited *Psp*Cas13b spacer-target mismatch can be tolerated to maintain specificity and on target nuclease activity.

Comprehensive analysis of spacer-target interaction examining specificity and mismatch tolerance of *Psp*Cas13b at various positions of crRNA spacer sequence. Perturbation of spacer-target interaction through spacer mutagenesis to introduce 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30-

nucleotide consecutive mismatch at the 3' (a) and 5' (b) end of the spacer. Perturbation of spacer-target interaction through spacer mutagenesis to introduce 6 (c), 5 (d), 4 (e) 4 or 3 (f) consecutive mismatches at various spacer positions. (g) Perturbation of spacer-target interaction by introducing various numbers of non-consecutive nucleotide mismatches at different positions. The nucleotides in red highlight mismatch positions in the spacer sequence. Data points in the graph are mean fluorescence from 4 representative fields of view per condition imaged; $N = 3$ or 4 independent experiments. The data are represented in arbitrary units (A.U.). Errors are SD and p-values of one-way Anova test are indicated (95% confidence interval).

4.3.2 Analysis of global cellular RNA integrity when a specific mRNA is targeted

Target-activated *Rfx*Cas13d has been shown to possess collateral activity that led to the cleavage of the endogenous ribosomal RNA 28S (rRNA 28S) in HEK 293T cells[151]. To determine whether *Psp*Cas13b or *Rfx*Cas13d have similar collateral activity against rRNA when they are activated by spacer-target basepairing, HEK 293T cells were transfected with plasmids encoding BCR-ABL1 p190, a targeting and non-targeting control crRNA directed against it (crBCR-ABL1) and either *Psp*Cas13b or *Rfx*Cas13d. 48 hours post-transfection, total cellular RNA was extracted for TapeStation analysis. It was shown that upon BCR-ABL1 target recognition and nuclease activation, *Rfx*Cas13d exhibited collateral activity manifested as partial nicking of 28S rRNA. However, *Psp*Cas13b showed no 28S RNA cleavage activity (**Figure 4.3**). This finding was in agreement with recent work that also observed differential collateral activities for the two Cas13 orthologues against ectopically overexpressed reporter genes[152]. This result also prompted further studies to determine whether cognate (on-target) *Psp*Cas13b nuclease activation can affect endogenous RNA more globally.

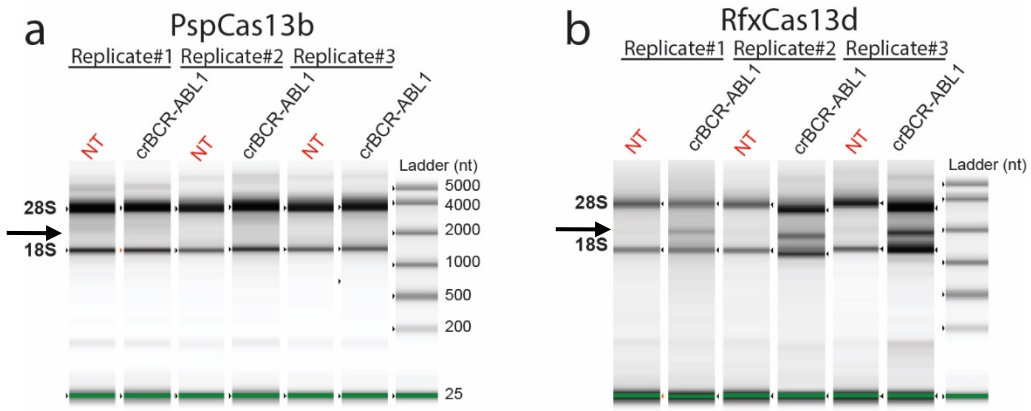


Figure 4. 3. *Psp*Cas13b demonstrates no 28s rRNA cleavage while *Rfx*Cas13d cleaves 28s rRNA upon target recognition and nuclease activation.

Total RNA was extracted 48 hours post-transfection from HEK 293T cells expressing either *Psp*Cas13b (a) or *Rfx*Cas13d (b) with a BCR-ABL1 p190 targeting crRNA and a non-targeting (NT) crRNA and quantified by Agilent 2200 TapeStation. The cleavage product of 28s rRNA is indicated by an arrow. *N*=3 (*N* represents the number of independent experiments performed for each orthologue).

4.3.3 Transcriptomics analysis suggests that *Psp*Cas13b activation has a minimal impact on endogenous mRNA expression

To better determine the impact of *Psp*Cas13b nuclease activation on endogenous cellular RNA, 3' poly-A RNA capture sequencing was performed on HEK 293T cells that had been transfected as above (**Section 4.3.2**) (**Figure 4.4a**). The resulting volcano and linear regression plots revealed a limited number of downregulated RNAs following targeted-activated *Psp*Cas13b following crRNA0 basepairing with the BCR-ABL1 transcript (**Figure 4.4b**). Only 25 of ~12700 transcripts (0.19%) were significantly reduced under these conditions, while the levels of 50 were increased **(Table 4.1)**. The upregulated 50 transcripts were very unlikely related to the nuclease activity of *Psp*Cas13b. Gene ontology analysis suggested the down- and upregulated genes did not cluster (less than 4 genes were grouped) to a particular cellular component, biological process, or molecular function.

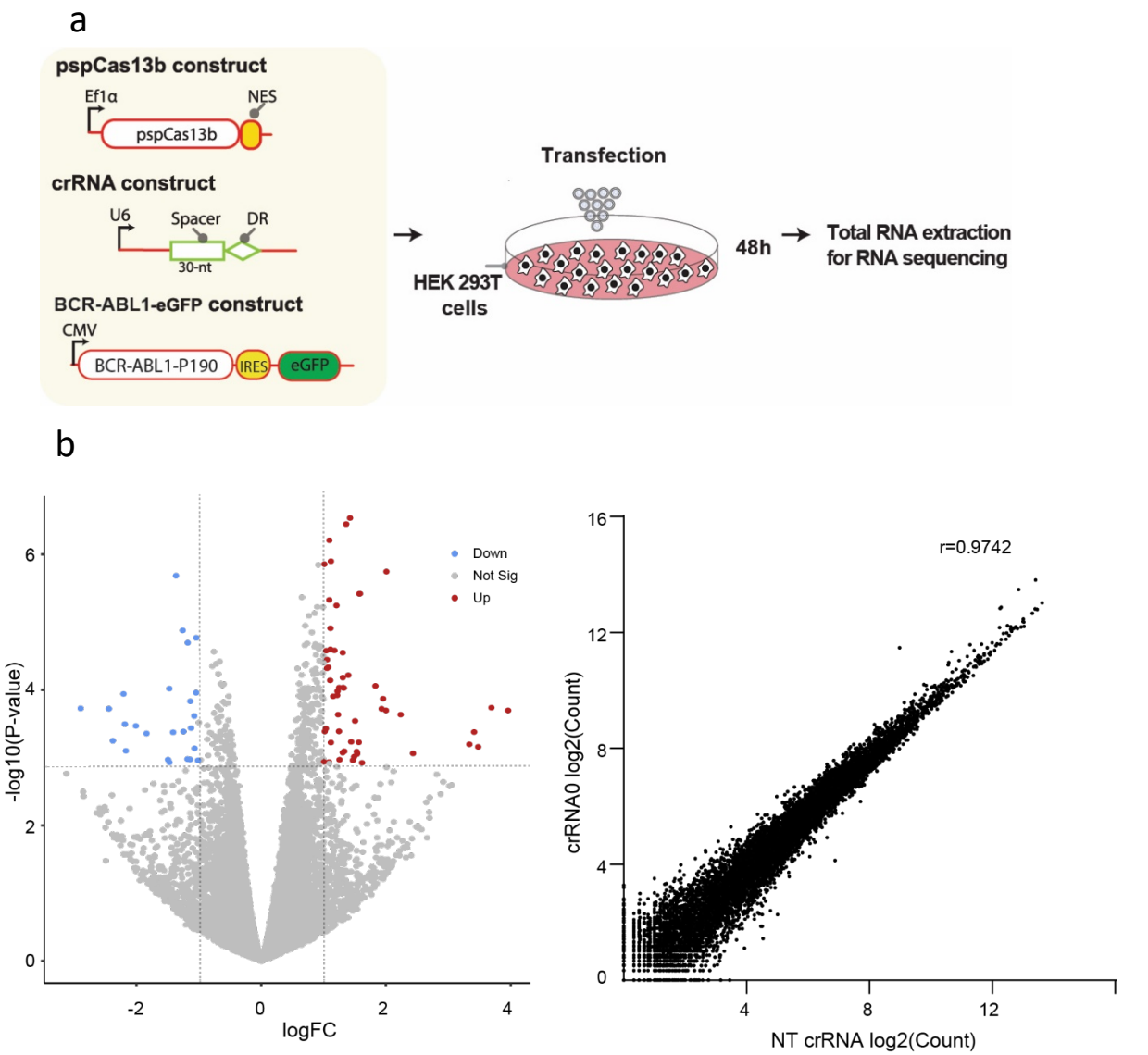


Figure 4. 4. Transcriptomic analysis suggested that *Psp*Cas13b activation have minimal impacts on endogenous transcripts.

(a) Schematic of RNA sequencing assays to investigate the off-target effects and collateral activities of *Psp*Cas13b in HEK 293T cells transfected with plasmids coding *Psp*Cas13b, BCR-ABL1 p190-IRES-eGFP and either a crRNA0 (targeting the breakpoint of BCR-ABL1 p190 transcript) or NT crRNA. RNA was extracted 48 hours post-transfection. A volcano plot (b, left) and linear regression plot (b, right) of 3' polyA RNA-sequencing analysis of HEK 293T cells expressing *Psp*Cas13b with a BCR-ABL1 p190 targeting crRNA (crRNA) versus a non-targeting (NT) crRNA. Each data point represents a gene, and genes with log2FC > 1 (upregulation, indicated in red) or log2FC < -1 (downregulation, indicated blue), and p-value < 0.05 are considered as significant differentially expressed transcripts. *N*=3; (*N* represents the number of biological replicates).

Table 4. 1. Specific mRNAs shown to be up- or down-regulated upon RNA sequencing, ranked in order of increasing disparity with control

	Gene symbol	Gene name	Log2FC	FC
No changes 12678 genes				
Downregulation 25 genes	IRF9	interferon regulatory factor 9	-2.89	0.1349035
	IFIT1	interferon induced protein with tetratricopeptide repeats 1	-2.44	0.1842837
	NR4A1	nuclear receptor subfamily 4 group A member 1	-2.37	0.1934456
	SOCS3	suppressor of cytokine signaling 3	-2.20	0.2176376
	KANK3	KN motif and ankyrin repeat domains 3	-2.18	0.2206757
	IFIT3	interferon induced protein with tetratricopeptide repeats 3	-2.16	0.2237563
	ACTA2	actin alpha 2, smooth muscle	-2.00	0.25
	STARD5	StAR related lipid transfer domain containing 5	-1.83	0.2812646
	KLHL17	kelch like family member 17	-1.49	0.3560125
	DDB2	damage specific DNA binding protein 2	-1.47	0.3609823
	CLEC16A	C-type lectin domain containing 16A	-1.46	0.3634931
	RBPMS	RNA binding protein, mRNA processing factor	-1.41	0.3763117
	SNHG15	small nucleolar RNA host gene 15	-1.36	0.3895823
	SNHG17	small nucleolar RNA host gene 17	-1.25	0.4204482
	UCKL1	uridine-cytidine kinase 1 like 1	-1.24	0.4233727
	CROCCP2	CROCC pseudogene 2	-1.18	0.4413515
	CHAC1	ChaC glutathione specific gamma-glutamylcyclotransferase 1	-1.17	0.4444213
	ZNF32	zinc finger protein 32	-1.13	0.4569157
	SMPD4BP	sphingomyelin phosphodiesterase 4B, pseudogene	-1.13	0.4569157
	GIHCG	GIHCG inhibitor of miR-200b/200a/429 expression	-1.12	0.4600938
	PISD	phosphatidylserine decarboxylase	-1.07	0.476319
	PLEKHH3	pleckstrin homology, MyTH4 and FERM domain containing H3	-1.06	0.4796321
	CCDC92	coiled-coil domain containing 92	-1.04	0.4863275
	PCBP4	poly(rC) binding protein 4	-1.04	0.4863275
	HTRA1	HtrA serine peptidase 1	-1.01	0.4965462
	SKIL	SKI like proto-oncogene	1.01	2.0139111
	LRRC58	leucine rich repeat containing 58	1.01	2.0139111
	MAP3K1	mitogen-activated protein kinase 1	1.01	2.0139111
	ATP13A3	ATPase 13A3	1.03	2.0420243

Upregulation 50 genes	HERC1	HECT and RLD domain containing E3 ubiquitin protein ligase family member 1	1.04	2.0562277
	SHPRH	SNF2 histone linker PHD RING helicase	1.05	2.0705298
	ZNF714	zinc finger protein 714	1.05	2.0705298
	DPP8	dipeptidyl peptidase 8	1.07	2.0994334
	MYBL1	MYB proto-oncogene like 1	1.08	2.1140361
	CHD9	chromodomain helicase DNA binding protein 9	1.09	2.1287404
	SGK3	serum/glucocorticoid regulated kinase family member 3	1.10	2.1435469
	LIN52	lin-52 DREAM MuvB core complex component	1.10	2.1435469
	SNTB2	syntrophin beta 2	1.11	2.1584565
	DGCR8	DGCR8 microprocessor complex subunit	1.11	2.1584565
	TET2	tet methylcytosine dioxygenase 2	1.11	2.1584565
	OTUD3	OTU deubiquitinase 3	1.15	2.2191389
	CEP135	centrosomal protein 135	1.17	2.250117
	ZNF430	zinc finger protein 430	1.20	2.2973967
	POLQ	DNA polymerase theta	1.21	2.3133764
	CLCN5	chloride voltage-gated channel 5	1.22	2.3294672
	FKBP14	FKBP prolyl isomerase 14	1.23	2.3456699
	MEGF9	multiple EGF like domains 9	1.24	2.3619853
	NCR3LG1	natural killer cell cytotoxicity receptor 3 ligand 1	1.24	2.3619853
	BACH2	BTB domain and CNC homolog 2	1.25	2.3784142
	HBEGF	heparin binding EGF like growth factor	1.30	2.4622888
	FAM76B	family with sequence similarity 76 member B	1.30	2.4622888
	FOXO3B	forkhead box O3B	1.30	2.4622888
	GFOD1	glucose-fructose oxidoreductase domain containing 1	1.31	2.4794154
	LONRF2	LON peptidase N-terminal domain and ring finger 2	1.32	2.4966611
	CCNT1	cyclin T1	1.33	2.5140267
	TMEM245	transmembrane protein 245	1.36	2.5668518
	ZC3H7A	zinc finger CCCH-type containing 7A	1.42	2.6758551
	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing	1.44	2.7132087
	ZBED6	zinc finger BED-type containing 6	1.46	2.7510836
	GABPB1-AS1	GABPB1 antisense RNA 1	1.50	2.8284271
	USP31	ubiquitin specific peptidase 31	1.53	2.8878584

	C3orf62	chromosome 3 open reading frame 62	1.53	2.8878584
	LRRIQ1	leucine rich repeats and IQ motif containing 1	1.53	2.8878584
	MIR17HG	miR-17-92a-1 cluster host gene	1.57	2.9690471
	KLHL20	kelch like family member 20	1.58	2.9896985
	RNY4	RNA, Ro60-associated Y4	1.83	3.5553707
	GOLGA2P5	GOLGA2 pseudogene 5	1.92	3.7842306
	DNAJC27	DnaJ heat shock protein family (Hsp40) member C27	1.95	3.8637453
	HYMAI	hydatidiform mole associated and imprinted	2.00	4
	NBPF10	NBPF member 10	2.00	4
	LOC401261	uncharacterized LOC401261	2.23	4.6913398
	SNORD46	small nucleolar RNA, C/D box 46	2.43	5.3889343
	H3C2	H3 clustered histone 2	3.33	10.056107
	FOXP1-IT1	FOXP1 intronic transcript 1	3.41	10.629487
	GCNA	germ cell nuclear acidic peptidase	3.60	12.125733

Surprisingly, the RNA sequencing data also suggested that the expression levels of eGFP and BCR-ABL1 p190 mRNA were not reduced by *Psp*Cas13b-mediated cleavage (**Table 4.1**), which was contradictory to the results obtained from RT-PCR (**Figure 3.2d**), fluorescence reporter assays (**Figure 3.2a**) and western blotting (**Figure 3.2g**) in experiments carried out under very similar conditions (**Figure 4.4a**). To examine this issue further, RT-PCR was carried out on the same samples used for 3' capture RNA sequencing. The RT-PCR carried out using primers flaking the fusion breakpoint showed that BCR-ABL1 p190 expression was indeed significantly reduced (**Figure 4.5a**), consistent with previous results (**Figure 3.2d**). The eGFP and BCR-ABL1 p190 proteins are encoded in a single mRNA species (BCR-ABL1 p190-IRES-eGFP) and cleavage of the mRNA at the BCR-ABL1 p190 breakpoint site was anticipated to result in quick degradation of the two resultant RNA fragments, as *Psp*Cas13b exposes unprotected 5' and 3' ends that are then susceptible to cellular exonucleases. Surprisingly, RT-PCR indicated that eGFP mRNA abundance was actually increased, in contrast to the BCR-ABL1 p190 product (**Figure 4.5b and 4.5c**).

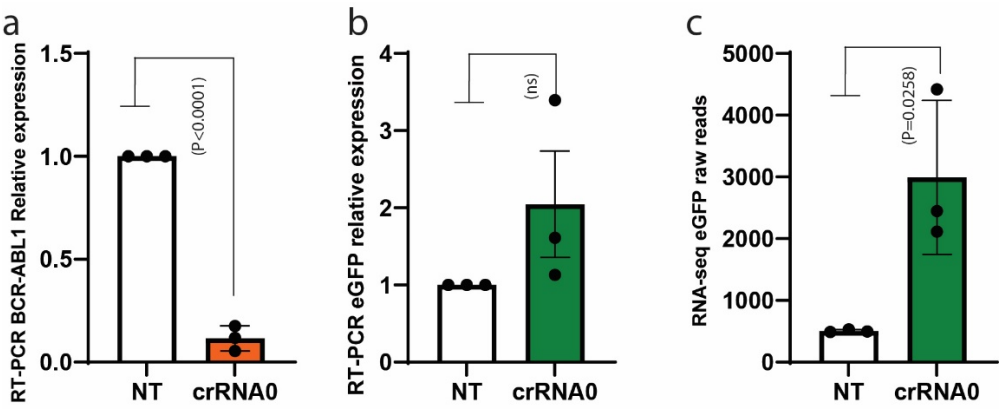


Figure 4. 5. RT-PCR assays to measure the expression levels of BCR-ABL1 p190-IRES-eGFP using primers flanking the breakpoint region (a) or the mid-region of eGFP mRNA (b) in HEK 293T cells expressing non-targeting (NT) or crRNA0 targeting BCR-ABL1 p190 mRNA. (c) Raw read numbers for eGFP mRNA in the same experiment. *N*=3. Data are normalized means and errors are SE. Results are analyses by unpaired two-tailed Student's t-test (95% confidence interval). *N* = the number of independent experiments.

Based on this interesting observation, it was hypothesized that when BCR-ABL1 p190-IRES-eGFP mRNA is cleaved at the BCR-ABL1 p190 breakpoint, the two resultant RNA fragments may become stabilized, extending their half-life and leading to the higher expression detected by RT-PCR and RNA sequencing (**Figure 4.6**). But this hypothesis challenges the current dogma, as RNA molecules lacking a poly-A tail and 5'cap typically undergo rapid degradation by cellular nucleases[193,194]. Despite the increase in eGFP mRNA after *Psp*Cas13b cleavage of BCR-ABL1, a drastic loss of eGFP fluorescence was consistently observed, as a likely result of inefficient translation of eGFP mRNA which lacks the 5' cap and 5'UTR, features critical for ribosomal recruitment and protein translation[195]. On the other hand, the apparently contradictory results for the BCR-ABL1 p190 transcript obtained by RT-PCR (reduced expression) and RNA sequencing (unchanged expression) might simply reflect where the primers achieving these results were sited. The primers used in RT-PCR flanked the breakpoint region of BCR-ABL1 190. In principle, they should bind to the two (now separate) cleaved fragments and so would fail to amplify a cDNA product. By contrast, random primers were used to generate short-read 3' RNA sequences, which may have still been generated from the remaining RNA fragments, thus underestimating the extent of RNA cleavage by *Psp*Cas13b (**Figure 4.6**).

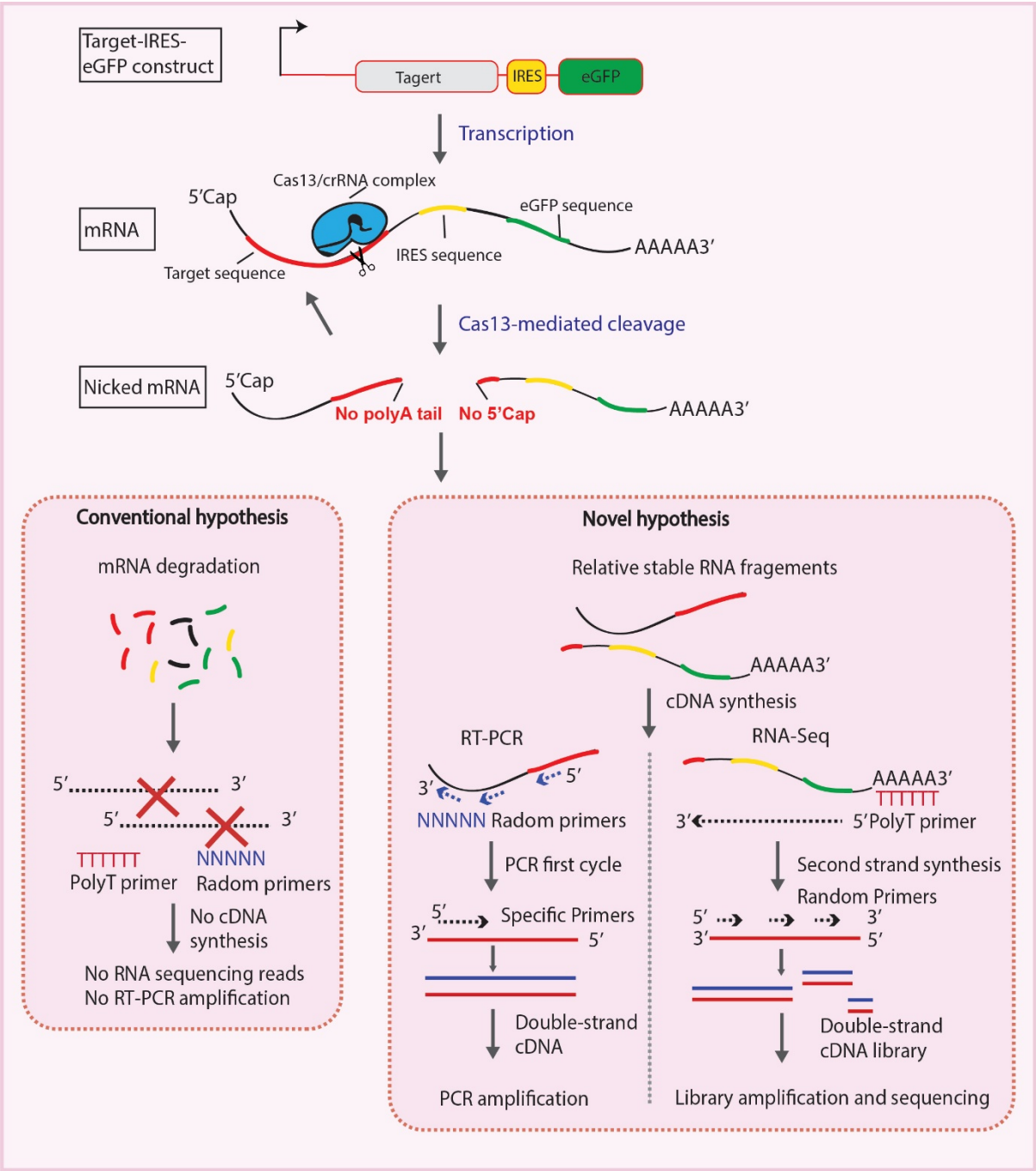


Figure 4. 6. Schematic of two proposed models to explain the fate of RNA fragments after *Psp*Cas13b cleavage and its impact on RT-PCR and 3' poly-A RNA sequencing assays.

The conventional view is that once the target RNA is cleaved by *Psp*Cas13b, the two fragments will be rapidly degraded by cellular exonucleases due to their unprotected 3' or 5' ends, and would not be captured by the RT-PCR or RNA sequencing. Rather, a novel hypothesis that may better explain the data is that the cleaved RNA fragments are stabilized by unknown factors, leading to efficient cDNA synthesis and capture by both RT-PCR and RNA sequencing.

To test our hypothesis, RT-PCR was carried out with 6 sets of primers designed to detect different regions of the full-length BCR-ABL1 p190 transcript (**Figure 4.7a**). RT-PCR confirmed that only the primers flanking and immediately adjacent to the breakpoint were able to substantiate RNA cleavage, however, the other five primer sets did not detect significant alterations of BCR-ABL1 p190 mRNA level (**Figure 4.7b**). Similar results were obtained with SFPQ-ABL1-IRES-eGFP mRNA, showing no apparent reduction in SFPQ-ABL1 and eGFP mRNA when the PCR primers did not flank the cleavage site near the breakpoint (**Figure 4.7d and e**). Together, this data suggests that *Psp*Cas13b cleavage generates two RNA fragments that remain stable in the intracellular environment. However, the major reduction in eGFP fluorescence indicated that the translation of these RNA fragments is repressed (**Figure 3.2a & 3.2g**). We propose that cleavage by *Psp*Cas13b nicks the RNA target at the precise location encoded by the crRNA spacer, however, the resultant RNA fragments can persist and can be detected by RNA sequencing and RT-PCR, but cannot be translated into proteins due to the loss of crucial translation initiation features such as 5' UTR, 5' Cap and 3' Poly-A tail. The findings confirm that *Psp*Cas13b mediates precise nicking of the RNA target at the spacer binding site, which compromises ribosomal RNA translation. Further transcriptomic analysis including long-read sequencing and depletion of short ribosomal RNA sequences may reveal the precise cleavage site with single-base resolution.

The transcriptomic data also confirmed the differential expression of very few genes following *Psp*Cas13b nuclease activation, implying only a very limited impact on off-target or collateral activities on endogenous mRNAs. However, the non-linear relationship between RNA and protein expression means the impact of the off-targeting or collateral activity of *Psp*Cas13b on protein homeostasis cannot be confidently inferred from the transcriptomic data. Mass spectrometry proteomics would be required to measure any such impact, as investigated below.

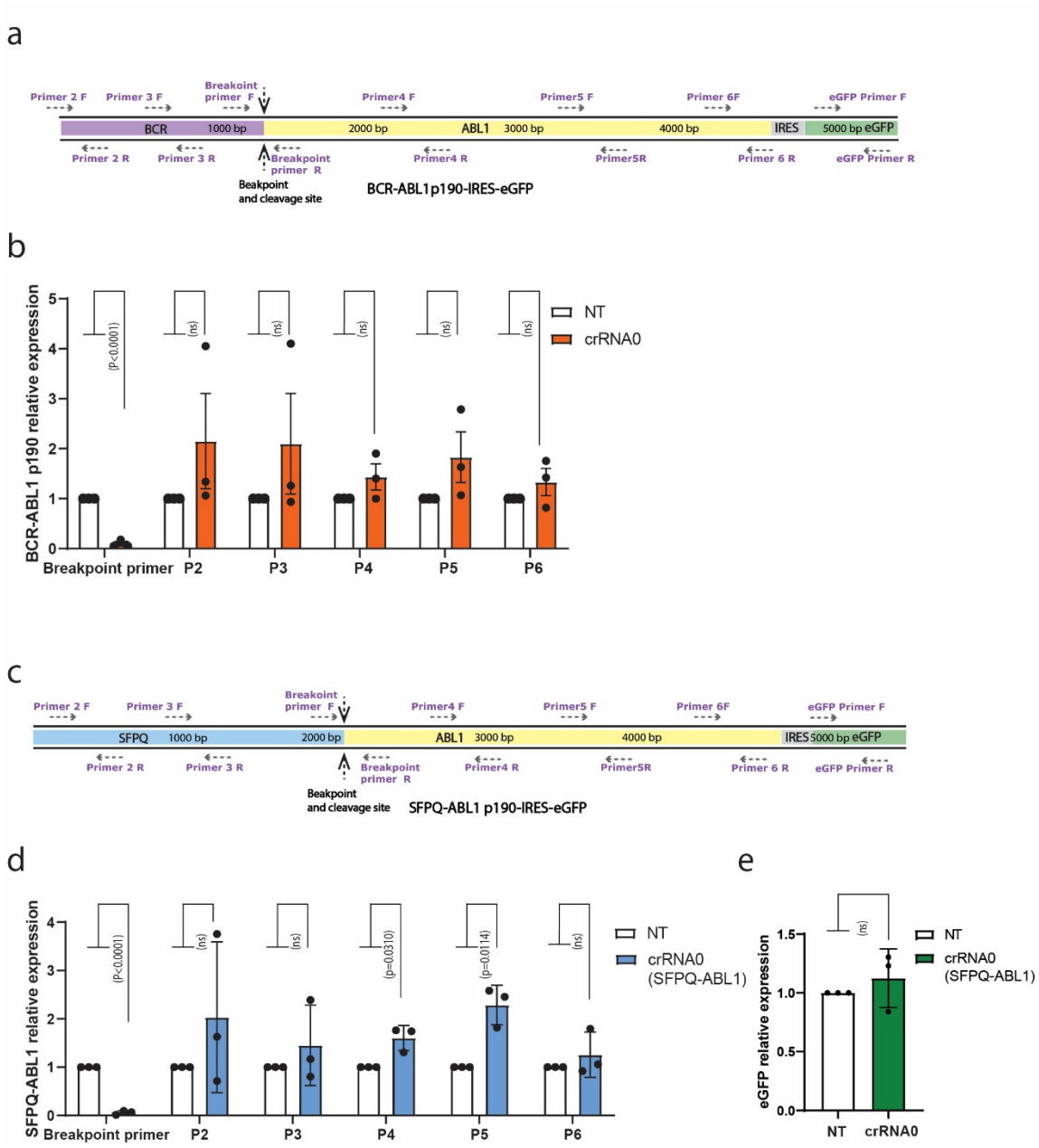


Figure 4. 7. Schematic showing the design of RT-PCR primers binding different regions on BCR-ABL1 p190 **(a)** and SFPQ-ABL1 transcripts **(c)** to determine the fate of cleavage RNA fragments. RT-PCR assays to measure the expression levels of BCR-ABL1 p190-IRES-eGFP **(b)** and SFPQ-ABL1-IRES-eGFP **(d)** using different primers, and eGFP **(e)** using primers binding to the middle region of eGFP sequence in HEK 293T cells expressing a non-targeting (NT) or crRNAs targeting the breakpoint region of fusion transcript (crRNA0); *N*=3. Data are normalized means and errors are SEM. Results are analysed by unpaired two-tailed Student's t-test (95% confidence interval). *N* represents the number of biological replicates).

4.3.4 Proteomic analysis reveals the high specificity and lack of collateral activity of targeted *Psp*Cas13b

To gain a global view on the impact of the collateral activity on protein expression levels, mass spectrometry was used for proteomics analysis of HEK 293T cells when *Psp*Cas13b was co-expressed with non-targeted (NT) crRNA or crRNA0, which targets the BCR-ABL1 mRNA breakpoint. There was no evidence of off-target protein degradation by crRNA0 compared to the non-targeting crRNA **(Figure 4.8)**. Consistent with previous fluorescence and western blot data (**Figure 3.2a & 3.2g**), the only significantly repressed proteins were the target BCR-ABL1 p190 (81% reduction) and eGFP (89% reduction) that are encoded on the same mRNA expression construct and encoded mRNA (**Figure 4.8**). Interestingly, among 3837 endogenous proteins that were detected in this mass spectrometry analysis, none showed significantly reduced abundance. Although not quite meeting the two-fold-change threshold considered biologically relevant, expression of exogenous *Psp*Cas13b was reduced by 48% in the targeting condition compared to the non-targeting samples (**Figure 4.8**). Western blotting of the same samples used for mass spectrometry revealed ~85% and ~18% reduction of BCR-ABL1 and *Psp*Cas13b proteins expression, respectively (**Figure 4.9**). The minor differences in the efficiency of silencing measured with mass spectrometry and western blot analyses were considered likely due to the differential sensitivity of the two technologies, and that western blotting cannot be strictly considered a quantitative technique. The reduced expression of *Psp*Cas13b itself indicates that some of the *Psp*Cas13 mRNA was cleaved by its own collateral activity, which appears to be restricted to overexpressed exogenous, but not endogenous transcripts.

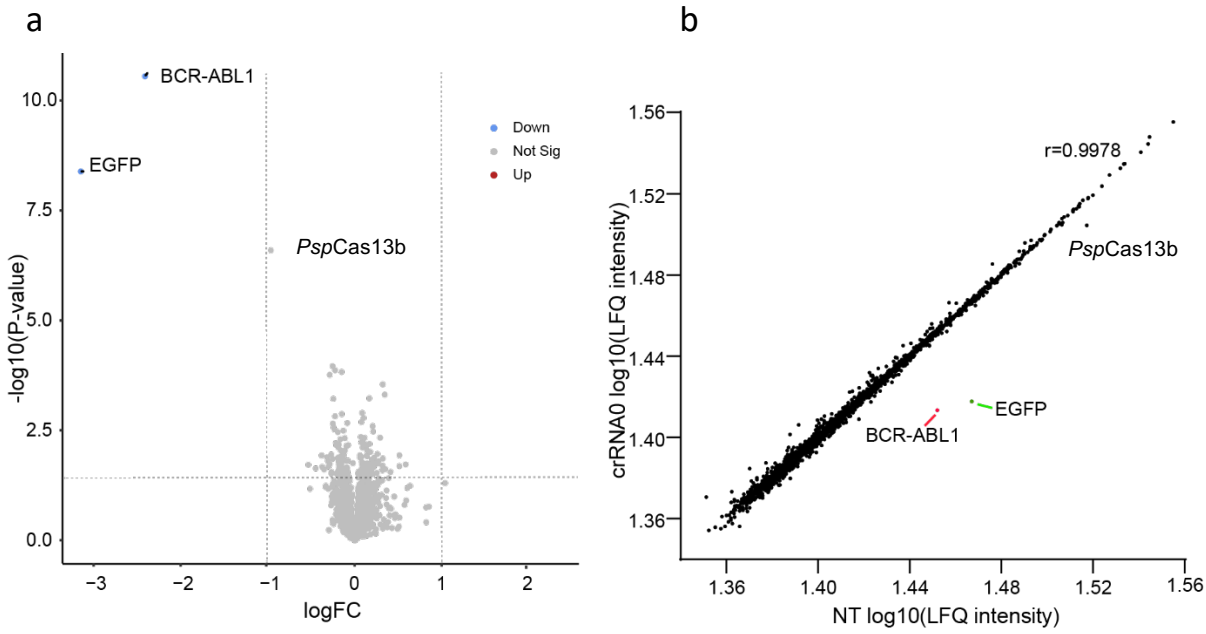


Figure 4. 8. Proteomic analysis revealed that *Psp*Cas13b nuclease domain activation does not affect the expression of endogenous proteins.

(a) Volcano plot of mass spectrometry analysis of HEK 293T cells expressing *Psp*Cas13b with BCR-ABL1 p190 targeting crRNA (crRNA0) versus a non-targeting (NT) crRNA. Each data point represents a protein, and proteins with log2FC > 1 (upregulation, indicated in red) or log2FC < -1 (downregulation, indicated blue), and p-value < 0.05 are considered as significant differentially expressed proteins. BCR-ABL1, eGFP and *Psp*Cas13b are labelled in the plot. (b) Linear regression plot of mass spectrometry analysis of HEK 293T cells expressing *Psp*Cas13b with BCR-ABL1 p190 targeting crRNA (crRNA0) versus a non-targeting (NT) crRNA. BCR-ABL1 and eGFP are highlighted in red and green, respectively. *Psp*Cas13b are labelled in the plot. N=5 (N represents the number of biological replicates).

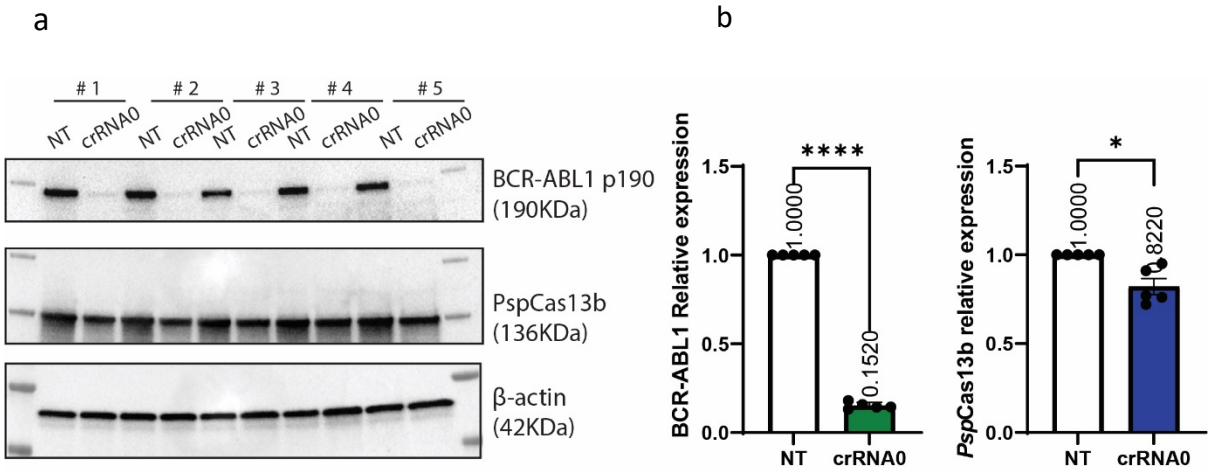


Figure 4. 9. *Psp*Cas13b demonstrated strong silencing against BCR-ABL1 p190 transcript with minor collateral cleavage against *Psp*Cas13b transcript but not against endogenous transcripts.

(a) Western blot analysis examining the expression level of *Psp*Cas13b and BCR-ABL1 p190 proteins in HEK 293T cells expressing either BCR-ABL1 p190 targeting crRNA (crRNA0) or a non-targeting (NT) crRNA. (b) Quantification of BCR-ABL1 and *Psp*Cas13b protein levels. Data points are the band intensity ratio of *Psp*Cas13b:β-actin of one biological replicate from crRNA0 condition normalised to NT crRNA condition; N=5 (N represents the number of biological replicates performed). All data are analysed 48 hours post-transfection, error bars represent SEM and p-values are calculated by unpaired two-tailed Student's t-test.

The lack of any change in the proteomic analysis indicates that global endogenous mRNA degradation is not induced by the collateral activity of *Psp*Cas13b, providing strong evidence that endogenous RNAs of HEK 293T cells are likely to be resilient to *Psp*Cas13b collateral activity. The minor collateral activity observed in mass spectrometry and western blotting experiments was directed only against *Psp*CAs13b itself, suggesting that non-specific nuclease activity may affect overexpressed exogenous RNAs, but not endogenous cellular RNAs.

4.4 Discussion

The specificity of spacer-target interactions is core to the capability of *Psp*Cas13b to selectively silence target RNAs in the RNA-crowded cellular environment[196], especially when one also considers the sequence homology that frequently occurs between various RNAs (**Figure 4.2**). Although some *Psp*Cas13b spacer sequences can tolerate ~3 to 5 consecutive or non-consecutive nucleotide

mismatches with the target, the 30-nucleotide spacer length enables this protein to retain remarkably high specificity (**Figure 4.2. Table 4.2**).

Table 4. 2. Predicted probabilities of *Psp*Cas13b off-targeting a random transcript in Human transcriptome.

The off-target transcript will require ~25-nt basepairing with the spacer sequence. The number of the transcript meeting this requirement in the Human coding or transcribed (coding + non-coding transcriptome) genome is close to 0.

Minimum basepairing required for PspCas13b activation (mismatch tolerance threshold)	**Probability of off-targeting other RNA in the human protein coding transcriptome (size ~$6x10^7$ -nt)**	**Probability of off-targeting other RNA in the human transcriptome (coding & non-coding) (size ~$2.3x10^9$- nt)**
26-nt basepairing required If four mismatches tolerance)	P=($6x10^7$)/(4^{27}) =$1.332x10^{-8}$ (Chance of occurrence: 0.000001332%)	P=($2.3x10^9$)/(4^{26})=$5.107x10^{-7}$ (Chance of occurrence: 0.00005107%)
25-nt basepairing required (Five mismatches tolerance)	P=($6x10^7$)/(4^{25}) =$5.329\ x10^{-8}$ (Chance of occurrence: 0.000005329%)	P=($2.3x10^9$)/(4^{25})=$2.042x10^{-6}$ (Chance of occurrence: 0.0002042%)
24-nt basepairing required (Six mismatches tolerance)	P=($6x10^7$)/(4^{24}) =$2.131x10^{-7}$ (Chance of occurrence: 0.00002131%)	P=($2.3x10^9$)/(4^{24})=$8.171x10^{-6}$ (Chance of occurrence: 0.0008171%)

In many RNA-guided nucleases such as Cas9, Cas12, and Ago2, mismatch intolerance is determined by surface-exposed mismatch-sensitive 'pre-seed' or 'seed' regions within their guide RNA. These micro-features facilitate efficient target search and recognition processes through the nucleation of RNA-DNA or RNA-RNA basepairing[39,40]. The mutagenesis findings above did not reveal any unique 'pre-seed' or 'seed'-like region within the spacer of *Psp*Cas13b that is highly sensitive to single, double, or triple nucleotide mismatches, other than the 5'GG motif discussed above (**Figure 4.2**). Instead, silencing activity appears to be governed by the number of consecutive or non-consecutive mismatches at various locations of the spacer. The mismatch intolerance occurs when ~3-5 unpaired nucleotides are introduced between the spacer and the target, which creates a sharp decline in the silencing activity, possibly due to duplex destabilisation (**Figure 4.2**). Future investigations with high spatio-temporal resolution may further reveal the dynamics of *Psp*Cas13b target recognition, and how these nucleotide mismatches impair its enzymatic activity. Overall, the comprehensive mutagenesis study performed here revealed the interface between mismatch tolerance and intolerance, and provided a blueprint for

crRNA design to silence various pathogenic transcripts with near single-base resolution, further expanding the spectrum of mammalian targets that might safely and effectively be silenced[182].

The collateral activity is well studied in biochemical *in-vitro* assays[12] and bacteria[10]. However, [-6]the extent to which it occurs in mammalian cells remains controversial[48,49]. We showed that upon BCR-ABL1 target recognition and nuclease activation, *Rfx*Cas13d orthologue exhibited some collateral activity that partially nicked the 28S rRNA, whereas *Psp*Cas13b did not (**Figure 4.3**). This suggested that no or very limited collateral activity of *Psp*Cas13b occurs in mammalian cells and that *Psp*Cas13b may offer better specificity than *Rfx*Cas13d. This observation supported recent work that examined the collateral activity of these two Cas13 orthologues against ectopically overexpressed reporter genes[50]. It is worth noting that as previous studies investigated collateral activity in mammalian cells using overexpressed exogenous genes that may lead to high target abundance and distinct cellular localization, little is known about whether the collateral activity also applies to endogenous RNA.

While *Psp*Cas13b nuclease activation and the associated collateral activity led to a moderate degradation of highly expressed exogenous RNAs such as *Psp*Cas13b itself, our proteomic analysis unambiguously suggested that *Psp*Cas13b collateral activity did not lead to a global downregulation of endogenous proteins in mammalian cells, which makes this orthologue a suitable tool for targeted RNA degradation and protein translation repression (**Figure 4.8**). A previous study has reported such observation of *Rfx*Cas13d where they observed the collateral activity of *Rfx*Cas13d was activated to cleave its own mRNA when *Rfx*Cas13d-crRNA complex bound and cleaved the target mRNA[151]. It is possible that the collateral activity had been activated as indicated by the minor reduction in *Psp*Cas13b protein expression in targeting crRNA condition, but it was too weak to be detected against the endogenous genes due to their lower expression or localisation. The likelihood of a specific transcript being collaterally cleaved by *Psp*Cas13 may be determined by its proximity (transcript abundance and its localization) to *psp*Cas13. Any potential collateral activities against various exogenous genes require further systematic investigation.

While the overall results are very reassuring with respect to any potential adverse effect from off targeting or collateral activity of activated *Psp*Cas13b, the complete significance of our transcriptomic analysis is limited by an incomplete understanding of the fate of RNA fragments of the cognate target mRNA following *Psp*Cas13b cleavage, which remains incompletely understood. In the conventional view, the cleavage of an mRNA molecule by endonucleases should seriously destabilize it due to the exposure of its 5'and 3'ends[193,194]. We observed no reduction of BCR-ABL1 and eGFP mRNAs in RNA sequencing data, other than when the RT-PCR primers flanked the putative *Psp*Cas13b cleavage site (**Table 4.1, and Figure 4.7**). However, the fluorescence reporter and western blot analyses both demonstrated potent and significant depletion of BCR-ABL1 and eGFP proteins (**Figure 3.2a and Figure 3.2g**). The targeted design of the RT-PCR primers used in the above studies generated key data

to explain the apparent discrepancy in data generated from RNA-seq, RT-PCR, and western blotting/fluorescence assays. One way of rationally explaining all of the findings is that post *Psp*Cas13 on-target cleavage, the two RNA fragments remain stable and resist total degradation until the time at which the samples were harvested (**Figure 4.6 and Figure 4.7**). However, protein translation was impaired as the cleavage removes RNA features that are critical for protein translation including 5'cap, 5/3' UTRs, and Poly-A tail. This observation challenges the concept that RNA cleavage with endonucleases invariably leads to mRNA destabilization and rapid degradation by cellular exonucleases[39,194]. Therefore, it is clear that 3'poly-A RNA sequencing using random primers or RT-PCR carried out using primers that do not bind very close to the cleavage site are unsuitable quantification of *Psp*Cas13b knockdown or assessing off-targeting. Consistent with our finding, a recent study also reported that primers bound to different regions on the target sequence detected different expression levels of the same target RNA by RT-PCR and the primers flanking the target region demonstrated the most significant reduction of the target RNA[197]. Therefore, to directly assess *Psp*Cas13b on-target and collateral nuclease activity and pinpoint the precise on target cleavage point in mammalian cells, transcriptome-wide, alternative deep sequencing technologies need to be considered including long read sequencing. The mRNAs shown to be up-or down-regulated in the RNA-seq were not shown to be deregulated in the proteomic analysis. Only the target protein BCR-ABL1 and EGFP were shown downregulated in the proteomic analysis but not the endogenous proteins. We believe the genes shown to be deregulated at the RNA level but not at the protein level could be due to two reasons:

(i) Non-linear relationship between RNA and protein expression.
(ii) Artifact due to RNA sequencing.

In a follow up validation study, we will use western blot analysis to further investigate protein expression profile of the deregulated genes shown in RNA-seq.

Overall, the comprehensive mutagenesis, transcriptomic and proteomic studies presented here revealed the interface between mismatch tolerance and intolerance and demonstrated that *Psp*Cas13 RNA silencing is highly specific with no discernible impact of its collateral activity on endogenous gene expression. Therefore, our investigation of *Psp*Cas13b specificity was reassuring when considering using the technology for therapeutic purposes, and highlights its usefulness as a precise and specific RNA silencing tool (**Figure 4.10**). Future proteomic analyses in other eukaryotic cell lines and primary cells are required for a deeper understanding of the on-target and collateral activity of various other Cas13 orthologues, whose properties in these regards may differ substantially from *Psp*Cas13b.

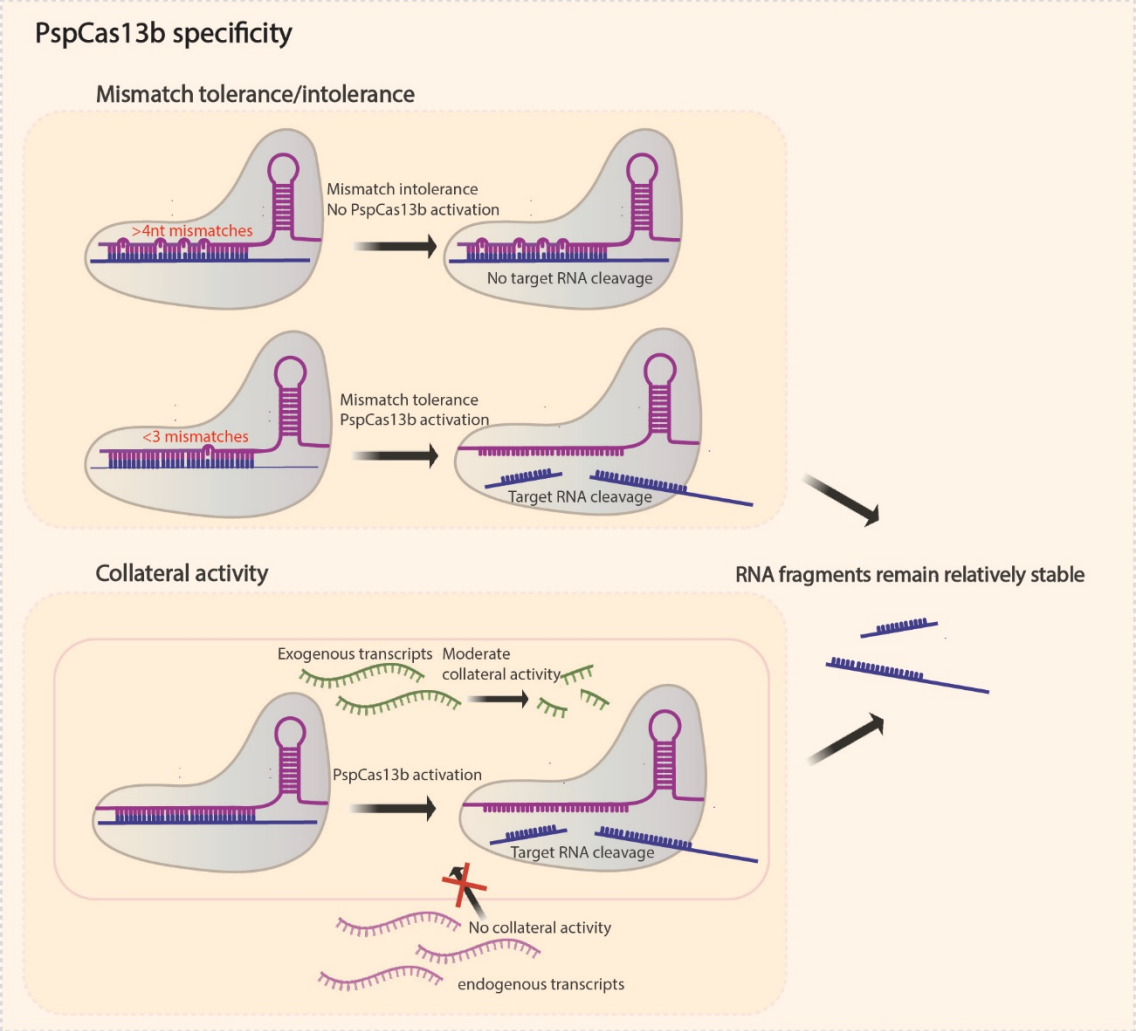


Figure 4. 10. Proposed schematic of the *Psp*Cas13b specificity in mammalian cells.

The on-target silencing activity by *Psp*Cas13b is highly specific and is sensitive to the number of consecutive or non-consecutive mismatches at various locations of the spacer. The mismatch intolerance occurs when ~3-5 unpaired nucleotides are introduced between the spacer and the target, which creates a sharp decline in the silencing activity. On the other hand, *Psp*Cas13b nuclease activation and the associated collateral activity led to a moderate degradation of highly expressed exogenous RNAs such as *Psp*Cas13b itself. However, our proteomic analysis suggested that *Psp*Cas13b collateral activity did not lead to a global endogenous protein downregulation in mammalian cells. Interestingly, our result suggested that *Psp*Cas13b-mediated cleavage (either on-target or collateral) generate relatively stable RNA fragments, which is contradictory to conventional concept that RNA cleavage with endonucleases leads to RNA destabilization and rapid degradation by cellular exonucleases.

Chapter 5

Discussion and conclusions

5.1 Overview

Over the last two decades, major advances in cancer genomics have contributed to the evolution of cancer treatments from systemic, non-targeted therapies into precise and personalized treatments[198]. Aberrations in a group of genes called tumour drivers including the gain of function proto-oncogenes and loss of function tumour suppressor genes result in the expansion of malignant cell clones and are thus considered appropriate targets for precise and personalized cancer treatments[199].

Currently, most drug development projects choose oncogenic proteins as targets. This process requires structural information, the development of informative biochemical or functional assays and methodical high throughput screening of libraries comprising collections of synthetic chemicals or, less frequently, natural products. Though laborious and expensive, the approach enabled the development of several very effective cancer drugs[200]. However, the development of protein-targeting drugs is slow, expensive, and technically challenging[201]. Protein-targeting drugs may block sites on proteins to prevent their conformational changes to active status, interact directly or allosterically with an enzyme active site, or block a domain responsible for ligand binding or formation of a heterologous signalling complex. This process is structure-based and therefore needs a highly iterative process of establishing structure/activity relationships of the drug to fit the desired protein sub-domain or pocket[202,203]. With fast-evolving sequencing techniques, unique mutations in one or a few individuals arise continually, and new tumour driver mechanisms constantly emerge[204]. However, it is impossible to develop a drug for each mutant protein. Extensive research has been conducted to target tumour drivers at a different stage in their genesis, for example at the DNA and mRNA stages[205]. Since DNA and mRNAs are nucleic acid sequences, the design of a targeting nucleic acid probe would be based on sequence specificity rather than protein structure, which would in principle be easier and time-saving. Most importantly, this strategy would allow us to target mutant DNA and RNA based on any individual patient's genetic profile.

Since its discovery in 2015, there has been huge enthusiasm from the scientific community regarding the potential of CRISPR-Cas13 systems in basic research, biotechnology, and medical sciences[24], which led to exploration of how/whether various orthologues could be used to manipulate RNA in a sequence-specific manner[43]. Among various Cas13 orthologues, *Psp*Cas13b is a multiplexable and robust toolkit for RNA-targeting with higher specificity and very limited off-target effects, which offers great opportunities for RNA manipulation including precise targeting of oncogenic transcripts[26,29]. However, little is known regarding the molecular basis that ultimately governs *Psp*Cas13b's ribonuclease activity. One critical issue that remains poorly understood is what molecular principles are required for the design of highly potent and specific crRNAs for efficient and precise

gene silencing. If these 'rules' were understood, then the rational design of efficient crRNAs for various applications including targeting oncogenic transcript would become far more feasible.

In **Chapter 2**, the targeting rules of *Psp*Cas13b were comprehensively investigated and the key design principles revealed were integrated into an online crRNA design algorithm (https://cas13b.github.io/). By ranking the most potent predicted *Psp*Cas13b crRNAs, this web tool will assist the wider community in designing their experiments. In **Chapter 3**, the design principles described in **Chapter 2** were used to enable the potent and selective silencing of various gene fusion transcripts and their downstream oncogenic networks, without off-targeting of untranslocated parental transcripts that by necessity share extensive sequence homology with parts of the target. *Psp*Cas13b nucleoproteins targeting the breakpoint of fusion transcripts enabled the efficient suppression of both 'conventional' (e.g., BCR-ABL1) gene fusion transcripts, as well as secondary mutants (e.g., BCR-ABL1- T315I) that often drive clinical relapse after an initial response to a targeted cancer drug. **In Chapter 4**, *Psp*Cas13b specificity was investigated through (i) comprehensive and systematic introduction of mismatches in spacer-target basepairing, and (ii) assessment of the impact of nuclease activation on global cellular mRNA and proteins expression using transcriptomic and proteomic analyses. This study provided new design principles for *Psp*Cas13b programming to specifically recognise and degrade any '*undruggable'* fusion oncogenic transcript with high specificity, thus providing a new conceptual framework for personalized targeting of oncogenic gene fusions at the RNA level.

5.2 Targeting tumour transcripts with the single-base precision design of *Psp*Cas13b

Next-generation sequencing technologies have identified over 500 common tumour-driver mutations and revealed significant inter/intra-tumour heterogeneities[129,170,171]. The translation of discoveries to the clinic remains challenging due to the complexity of conventional structure-based drug design and development[202]. For example, in a recent study, only 13.60% of patients with advanced or metastatic cancers were eligible for and only 7.4% respond to targeted therapy for their own specific tumour driver mutation/s[206]. Furthermore, designing a protein-targeting drug specific for each tumour driver identified is impossible with current drug design and validation technologies[172]. Therefore, there is enormous unmet need for new drugs that are highly specific and offer flexible design to target single or multiple oncogenic drivers in a personalized manner.

5.2.1 Targeting oncogenic genes at the chromatin level

Cancer gene therapy is defined as the genetic modification of patients' somatic cells to treat cancer[207]. Early gene therapy in the 1980s involved the delivery of therapeutic DNA by using viruses (e.g., adeno-associated virus: AAV) or through direct injection of plasmid DNA encoding wildtype genes. However,

these approaches were unsuccessful due to host immune responses, gene silencing, and insertional mutagenesis[208]. More recently, numerous studies have explored targeting tumour driver genes at the DNA level by designing a guide sequence that base-pairs with the target sequence (e.g., CRISPR) or a series of DNA binding domains (e.g., TALE binding domain and zinc finger binding domain) to direct a nuclease effector (e.g., Cas, TAL effector and zinc finger nuclease) to specifically to the target DNA sequence[209,210]. The design of these nuclease effectors is based on sequence, needs no structural information and should therefore in principle be a highly flexible and efficient way to adapt a patient's genetic information for their own personalized treatment.

Nuclease-based DNA editing including zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have shown promise as potential therapeutics[209,210]. For example, in an ongoing phase I clinical trial, researchers intravenously delivered AAVs carrying a programmed ZFN to insert a wild type copy of the Iduronate 2-Sulfatase gene into hepatocytes of patients with Mucopolysaccharidosis II[211]. However, both ZFN and TALEN technologies achieve only moderate editing efficiency and, as they do not rely on guide RNA for target recognition, still require complicated protein engineering pipelines[212].

The discovery of CRISPR-Cas9 offered an opportunity to overcome many of these issues and achieve sequence-specific targeting with a relatively simple design and high efficacy[10,11]. DNA targeting CRISPR-Cas9 consists of a Cas9 protein and an engineered sgRNA which contains a 20-nucleotide spacer sequence complementary to the target DNA flanked by a PAM sequence[15]. The sgRNA hybridizes with one DNA strand and two Cas9 nuclease domains introduce a DSB on the target DNA[10]. The application of Cas9-based editing provides new approaches to manipulate genes of interest and opens new possibilities for gene therapies for mutation-driven diseases. For example, CRISPR-Cas9 has shown encouraging results in a recent trial treating patients with sickle cell disease that is caused by a single mutation in the beta-globin gene resulting in malformed haemoglobin and distorted red blood cells. Hematopoietic stem cells were isolated from the patient, *ex vivo* engineered with CRISPR-Cas9 to produce high levels of fetal haemoglobin in red blood cells and re-infused into the patient. Preliminary data have shown that this therapy produced 46% of the haemoglobin in the patient as edited fetal haemoglobin, which was expressed in 94% of the red blood cells[117]. While *ex vivo* editing of patient's cells appears to be attractive and feasible, *in vivo* delivery of CRISPR-Cas9 to target tumour driver mutations in patients remains more challenging. The irreversible DNA alterations caused by Cas9 have raised major safety concerns related to its off-target effects and associated chromosome instability[121]. Although Cas9 has a 20-nt long spacer sequence, several studies have shown that 3 to 5 mismatches between sgRNA and the target in the PAM-distal regions can be tolerated[41,122-124], which lowers its targeting specificity and greatly raises the likelihood of off-target effects. Depending on the site affected, these errors may have relatively minor impacts, but may cause unexpected chromosomal rearrangements and even megabase-scale chromosomal truncations[125]. Given the permanence of gene

disruption by Cas9, these side effects can lead to loss of genome integrity and are a serious limitation to Cas9 gene therapy. Importantly, studies have shown that Cas9-based editing in cell lines could lead to the emergence of cells that have deficient p53-mediated DNA damage responses[213,214]. More recently, a patient died in an FDA-approved clinical trial (CRD-TMH-001) of Cas9-based treatment of Duchenne muscular dystrophy, although the cause of death is currently under investigation[215].

5.2.2 Target tumour drivers at the RNA level

Considering the limitations of Cas9, RNA-dependent RNA targeting CRISPR effectors offer a great alternative, as they potentially offer reversible and sequence-specific targeting of tumour drivers at the RNA level. According to this concept, ingeniously engineered *Sp*Cas9 was first demonstrated in 2014 to cleave ssRNA by providing an oligonucleotide containing a PAM sequence[216]. Later, modified *Staphylococcus aureus* Cas9 (*Sau*Cas9) was shown to cleave ssRNA without the facilitation by PAM, further simplifying the application[217-219]. However, programmed RNA targeting using Cas9 effectors still retains the intrinsic target recognition properties of Cas9, so off-targeting of DNA remains likely[217].

Almost three decades ago, miRNA was uncovered as a master regulator of RNA expression in nearly all eukaryotes by a process called RNA interference (RNAi). RNAi emerged as the golden standard technology to knock down a specific RNA in eukaryotic cells. Since its discovery, academic research and biopharma massively invested in research programs to develop eukaryotic RNAi as personalized medicine[72]. However, as discussed in **Chapter 1**, RNAi has strong off-target effects, which can alter the expression of hundreds of genes transcriptomically[220]. Secondly, exogenously introduced RNAi is shown to be capable of hijacking the native RNAi mechanism, disrupting normal gene expression through competition with host miRNA. These side effects have been reported to cause liver failure and the death of human α-1 antitrypsin-transgenic mice[221,222].

CRISPR-Cas13 offers an unparalleled opportunity to achieve personalized targeting of pathogenic RNAs. The Cas13-crRNA nucleoprotein complex can be reprogrammed to potentially cleave any mRNAs with high specificity and efficiency. The very highly specificity of RNA interference holds the promise of targeting tumour transcripts that are typically considered "undruggable", including aberrant fusion transcripts that drive various human genetic diseases. The spacer of Cas13 mediates sequence-specific target recognition, as nuclease activation and target degradation require near-complete base-pairing with the target sequence. In fact, Cas13a orthologues were recently used to silence fusion gene transcripts in cell lines including FGFR3-TACC3 (glioblastoma), EML4-ALK (lung cancer) and SS18-SSX2 (synovial sarcoma)[223-225]. However, the design principles for systematic, potent, and specific silencing of gene fusions without the suppression of untranslocated fusion partners remained unknown until the current study. In this study, SIBTIL screening was devised which, through unbiased *in silico* analysis and comprehensive spacer-target mutagenesis, revealed key determinants of *Psp*Cas13b potency and specificity. *De novo* design of

crRNAs incorporating these newly discovered RNA features was shown to greatly enhance the silencing efficiency of otherwise poorly effective crRNAs, expanding the targeting spectrum of conventional *Psp*Cas13b. Importantly, it was shown that rational design of crRNAs could specifically and potently target the breakpoint of various fusion transcripts including BCR-ABL1 P210, BCR-ABL1 P190, SNX2-ABL1, SFPQ-ABL1, RUNX1-RUNX1T1 and NPM-ALK (**Chapter 3**).

*Psp*Cas13b also provides an approach to overcome drug-resistant relapse caused by secondary mutations. In our study, it was shown that the crRNA targeting the BCR-ABL1 breakpoint sequence was equally potent for the BCR-ABL1 T315I variant which confers resistance to first-line ABL1 inhibitors, including *Imatinib,* and often drives clinical relapse[181]. In this case, crRNA potency was not compromised since the T315I point mutation occurs in the kinase domain which is located away from the breakpoint sequence that *Psp*Cas13b recognises. Therefore, *Psp*Cas13b crRNA can be designed to target an oncogene at one location but enable the targeting of both the ancestral transcript and derivatives that carry secondary mutations. Additionally, if the oncogenic transcript becomes mutated at the original crRNA binding region, *Psp*Cas13b may still be effective as the mutagenesis study described above demonstrated that a crRNAs may tolerate ~3 to 5 mismatches. Moreover, targeting oncogenic RNA with *Psp*Cas13b offers an opportunity to multiplex crRNAs to simultaneously target multiple drivers. In fact, the mismatch tolerance mechanism described above, combined with multiplexing was effective at reprogramming *Psp*Cas13b to suppress SARS-CoV-2 replication and circumvent its mutational escape[139]. Here, crRNA multiplexing enabled simultaneous targeting of several RNA locations of the SARS-Cov-2 viral genome, which increased the probability of high silencing while making viral mutational escape far less likely[139].

The potential use of *Psp*Cas13b to silence other forms of oncogenic transcript including point-mutated or single-nucleotide variants (SNVs) was also demonstrated. Specific targeting of SNVs is challenging since the SNV and wild-type transcripts differ by only one nucleotide and CRISPR enzymes are intrinsically tolerant to single-nucleotide mismatch. A comprehensive mutagenesis study revealed that it is possible to define the interface between mismatch tolerance and intolerance for a given SNV. 3-5 unpaired nucleotides can typically be introduced between the spacer and the target with little effect on potency, however a very sharp decline in silencing activity occurs when even a single further difference is introduced, possibly due to RNA-RNA duplex destabilisation. The observation that one additional mismatch can shift the nuclease from active to inactive state provides a window in which to design crRNAs that can specifically cleave point-mutant transcript but not their wildtype counterparts. The hypothesis remains to be tested with more 'real-world' examples, but this study proposes that rationally designing a crRNA with a few (n) mismatches against a given point-mutant transcript would permit its cleavage due to mismatch tolerance, whereas having n+1 mismatches with the wildtype sequence would spare it from destruction (**Figure 5.1**). A single additional mismatch is unlikely to change the thermodynamic stability of spacer-target duplex in the cell. However, the bulky mismatched

nucleotides could impair the conformational changes that are required for nuclease domain activation. But this process of target binding is protein assisted, which introduces an additional variable that is likely to affect the thermodynamic stability of spacer-target. Thus, excluding *Psp*Cas13b contribution to the thermodynamic stability may not be the correct approach. While this hypothesis is plausible, it requires experimental validation by decoupling binding and cleavage events. In our future work we will use biochemical assays (i.e., EMSA vs denaturating PAGE, and pulldown using catalytically active vs inactive *Psp*Cas13b) to determine *Psp*Cas13 binding and cleavage activity while introducing various mismatches between the spacer and the target. With the current data, it is difficult to conclude whether it is purely due to the destabilization of target-crRNA duplex or steric clashes of RNA with Cas13 protein inducing an allosteric shift in protein conformation. Future investigations are required to validate the design of crRNAs capable of achieving single-base resolution silencing against oncogenic SNVs such as the currently 'undruggable' KRAS mutants.

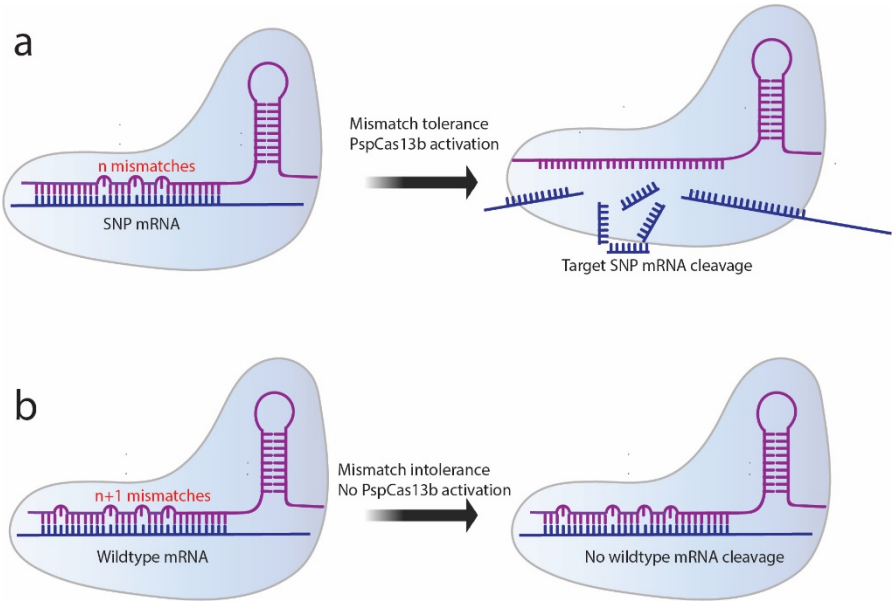


Figure 5. 1. Schematics illustrating specific targeting of tumour mRNAs with single-base precision design of *Psp*Cas13b.

(a) The design of crRNA with a few (n) mismatches with the target SNV mRNA will trigger *Psp*Cas13b nuclease activation and mRNA cleavage due to its mismatch tolerance. (b) The same crRNA is designed to achieve n+1 mismatches with the wildtype mRNA, which will not trigger *Psp*Cas13b nuclease activity due to mismatch intolerance as n+1 is beyond the tolerance threshold.

5.3 Challenges and future perspectives

The structural and functional characterization of CRISPR-Cas13 tools in this study has provided the foundation for programming them for a wide range of RNA manipulations. However, additional challenges need to be overcome for their potential use as therapeutics. For example, the molecular mechanisms of Cas13-mediated targeting are also not yet fully understood, with unsolved questions regarding the rules for efficient reprogramming, off-target effects, and the status of the controversial collateral activity in cells and *in vivo*. Also targeted *in vivo* delivery of CRISPR-Cas13-based therapeutics and their potential immunogenicity in mammals may require further investigation and optimization.

5.3.1 Delivery of CRISPR-Cas13

Although Cas13 has great potential to target disease-causing transcripts, successfully targeted *in vivo* delivery is a critical challenge for clinical implementation[43,58,70]. Various approaches have been explored to deliver CIRIPSR-based therapeutics *in vivo*, including viral-based delivery and non-viral delivery. Recently, a pilot study delivered a lentiviral vector encoding *Lwa*Cas13a and its crRNA to mice bearing glioma and inhibited the formation of intracranial tumours in mice[64]. Another group directly injected CRISPR-Cas13 RNPs targeting KRAS and CRISPRMAX reagents into mouse pancreatic tumours and significantly reduced tumour size[133]. Cas9 and crRNA plasmids were delivered by hydrodynamic delivery of DNA solution intravenously to a mouse liver cancer model and achieved editing of the target gene[226]. More recently, a novel liposome-based gene delivery platform called high-level extended duration gene expression system (HEDGES) was developed, which allowed long-term (3 weeks to 1.5 years) expression of therapeutic levels of proteins encoded by delivered transgenes[227].

However, more practicable and safer delivery approaches are required for human applications. For example, AAVs are small replication-defective, non-enveloped ssDNA viruses with low immunogenicity. Additionally, AAVs encoding therapeutic vectors can infect both dividing and quiescent cells but are unable to integrate into the genome due to removal of *rep* and *cap* viral genes[228]. As discussed in **Chapter 1**, the relatively small-sized Cas13d can be packaged into AAV backbone for *in vivo* delivery[27]. However, for other larger Cas13 orthologues and Cas effectors such as *Psp*Cas13b, other solutions are required due to the limited capacity of AAV vector[220]. Recently, enormous efforts have been made to develop non-viral approaches for delivering nucleic acid-based therapies including Cas9 and RNAi gene therapies and mRNA vaccines, which can be translated into Cas13 settings.

A well-documented non-viral approach is lipid-based nanoparticles (LNPs) that encapsulate CRISPR/Cas components or RNAi by electrostatic interaction between the negatively charged nucleic acid phosphate backbone and the positively charged lipids. LNP can be artificial or naturally occurring such as extracellular vesicles and exosomes that play important roles in transferring biomolecules[229,230].

Cellular transport by nanoparticles is typically mediated by endocytosis and micropinocytosis and they have been shown to efficiently deliver DNA, RNA, proteins, carbohydrates, and lipids[226,231].

Among the various molecules that can be delivered by LNPs, mRNA has emerged as a promising therapeutic cargo[232]. mRNA has several advantages over proteins and DNA for gene transfer and expression. For example, mRNA is efficiently translated to produce sufficient amount of a protein of interest, has no possibility of genomic integration and its function is not dependent on cell proliferation[232,233]. A recent study demonstrated that the combined delivery of lipid nanoparticles containing Cas9 mRNA and AAVs encoding sgRNA can correct the pathogenic Fah-splicing mutation in a mouse model of tyrosinemia[27]. Another study demonstrated that by chemically altering sgRNA synthesis and delivering it with Cas9 as mRNA or protein instead of a DNA plasmid by nanoparticles, the genome editing rate in both human primary T cells and hematopoietic stem and progenitor cells could be reduced significantly[234]. Excitingly, the first siRNA drug *Patisiran* delivered by LNP was approved by the FDA in 2018 to treat the neuronal disorder, hereditary transthyretin-mediated amyloidosis, which paved the way for their use to deliver other genetic payloads to treat various human genetic diseases and may prove useful for Cas13 mRNA delivery[235]. Notably, LNP–mRNA vaccines mRNA-1273 and BNT162b against COVID-19 are now on the market, which marks a milestone for mRNA therapeutics[236,237]. These methods can be adapted to deliver CRISPR systems transiently, which is safer than viral-based delivery and may be subject to fewer regulatory hurdles, as the LNP cannot replicate.

Depending on *Psp*Cas13b's pharmacokinetics and efficiency of targeting cancer cells, it should be possible to optimise the dosing of *Psp*Cas13b mRNA and its crRNA in a manner similar to conventional drugs. This transient silencing of an oncogenic driver would be comparable to the administration of small inhibitory molecules that targets oncogenic proteins (e.g., *Imatinib*) which of course also need to be appropriately dosed. Moreover, Cui *et al.* developed LNP encapsulating CRISPR–*Rfx*Cas13d to target lung protease cathepsin L (*Ctsl*) mRNA to block SARS-CoV-2 infection in mice with high specificity and safety[238]. More recently, next-generation LNPs have been developed to encapsulate and deliver nucleic acid payloads to various tissues and cells in a targeted manner, including RNA-encoding CRISPR effectors. Inclusion of a cationic lipid (DOTAP) in LNP selectively delivers mRNA to lung[239].

Here I conducted a proof-of-concept study to characterize the feasibility of targeting gene fusions at the molecular level. Our current research addresses the challenges of efficient *in vivo* delivery through the use of lipid nanoparticles and Cas13 mRNA therapeutics.

5.3.2 Potential *in vivo* cytotoxicity related to immunogenicity, collateral activity, and off-target effects

5.3.2.1 Collateral activity and off-target effects

One of the main current limitations of CRISPR-Cas13-based therapeutics is the concern on their potential collateral and off-target activities[240]. The literature so far shows conflicting results regarding Cas13 collateral and off-target activities; clearly, the specificity of Cas13-mediated cleavage requires further study. Kushawah *et al*. first assessed the efficiency of *Rfx*Cas13d-mediated downregulation of transcripts of interest in embryos of animal models including zebrafish, medaka, killifish, and mice. They found an average of 76% knockdown while observing no signs of toxicity, non-specific developmental defects, off-target effects, or collateral activity[68]. Very recently, Morelli *et al*. demonstrated that intrastriatal AAV-delivery of *Rfx*Cas13d and its crRNA selectively reduces mutant *HTT* mRNA and protein levels in the *corpus striatum*, in a mouse model of Huntington's disease. The phenotypic improvements including improved motor coordination and attenuated striatal atrophy lasted for eight months without adverse effects and, importantly, this study reported very few off-target transcriptomic alterations[241]. Conversely, a recent study on *BioRxiv* (awaiting peer review), has reported that knocking down gene expression using *Rfx*Cas13d in adult neurons caused death in the Sleepy (*Sik3*$^{Slp/+}$) mouse strain. This was not the result of loss-of-function of the target gene, but rather due to collateral activity that induced cleavage of 28S rRNA[151]. Nevertheless, two Cas13 effectors demonstrating minimal collateral activity (Cas13d-N2V8 and Cas13X-M17V6) in transgenic mice and AAV-mediated somatic cell targeting were recently engineered via mutagenesis and screening of hundreds of Cas13 variants, which provides a safer option for Cas13-based therapeutics[67]. There is no doubt that over time, this relatively young field will mature, and new tools will emerge with higher on-target efficiency and no or very limited off-targeting or collateral activity in both basic research and therapeutic contexts.

Whilst the collateral activity of some Cas13 orthologues (e.g., *Lwa*Cas13a and *Rfx*Cas13d) is well established in *in vitro* assays where it is used as a diagnostic tool, its relevance for mammalian cells has remained unclear. Indeed, apart from Cas13d, the collateral activity of other Cas13 orthologues in mammalian cells remained poorly investigated and controversial[28]. Certain studies have reported that *Psp*Cas13b appears to have no or very limited 'collateral activity' compared to *Rfx*Cas13d[152]. The latter study aligns with the data presented in the current study, showing that upon target recognition and nuclease activation, *Rfx*Cas13d exhibited some levels of collateral activity that partially cleaved the 28S rRNA, while *Psp*Cas13b tested under the same experimental conditions did not. Moreover, the majority of reports to date on mammalian cells examined the effect of the collateral activity on over-expressed exogenous RNAs, but little has been known about its impact on endogenous RNAs. There has been a lack of a comprehensive transcriptomic and proteomic analysis to investigate the impact of the collateral activity of various Cas13 orthologues on global endogenous proteins in both cell lines and primary cells.

Therefore, future work should focus on investigating collateral activities of various Cas13 effectors in a cell-type-dependent manner.

5.3.2.2 Immunogenicity

Apart from toxicity induced by off-target effects and collateral activity, immunogenicity is another potential issue of bacterial CRISPR effectors. Pre-existing antibodies and T cells directed against Cas9 proteins have been reported in previous human studies[242-244], but this hasn't prevented further development of Cas9 and its use today in clinical trials[245]. Cas13 as a bacterial protein may exhibit some levels of immunogenicity in humans that need to be considered as a risk factor[240]. But there are several approaches to mitigate the immunogenicity, which include developing immunotolerant Cas proteins by mutating immune-dominant epitopes[244] or using immune-orthogonal Cas orthologues[246]. Ferdosi *et al*. identified two immunodominant *Sp*Cas9 epitopes in pre-existing human CD8+T cells and further modified *Sp*Cas9 by targeted mutation of immunodominant epitopes while preserving its function and specificity[244]. Alternatively, Moreno *et al*. computationally predicted Cas9 orthologues that might not activate immunity and demonstrated efficient gene editing by three candidate Cas9 orthologues in mice previously immunized against AAV vector and Cas9 cargo[246]. Previous studies aimed at lowering immunogenicity of Cas9 can be applied to Cas13 to uncover or generate Cas13 orthologues that are not immunogenic. Recent milestones in artificial intelligence-enabled *de novo* design and functional validation of new unnatural proteins, may enable the creation of programmable Cas13-like effectors from the human proteome[247]. In addition, it is possible to safely combine Cas13-based therapeutics with immunosuppressant drugs.

5.3.2.3 CRISPR inhibitors

The discovery of phage-generated anti-CRISPR type VI-A (AcrVIA) proteins provides the potential option to tune targeting duration and efficiency for optimal RNA manipulation and reduce the unwanted effect of constitutively active Cas13a effectors. Remarkably, AcrVIAs also significantly reduced the efficiency of d*Lwa*Cas13a-mediated RNA A-to-I editing[248]. Similar phage-derived Cas13 inhibitors may also exist for other type VI systems during evolutionary phage-bacteria interactions[248]. Therefore, it will be worthwhile to identify and characterise Cas13 antagonist proteins for other Cas13 orthologues to downregulate the strongest manifestations of Cas13-mediated targeting, collateral, or immunogenic activities.

In conclusion, *Psp*Cas13b design flexibility, specificity, and the apparent lack of collateral activity described in this study provide a new conceptual framework for precise RNA silencing or editing, which could transform the landscape of personalized medicine through targeting previously '*undruggable'* or drug-resistant oncogenic RNAs including fusion transcripts. However, further investigations are needed to better understand *Psp*Cas13b activity in a variety of additional human cell

lines, primary cells, and animal models, which in turn will require the development of efficient targeted delivery strategies *in vivo*.

References

1 Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology* **169**, 5429-5433, doi:10.1128/jb.169.12.5429-5433.1987 (1987).

2 Groenen, P. M. A., Bunschoten, A. E., Soolingen, D. v. & Errtbden, J. D. A. v. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Molecular Microbiology* **10**, 1057-1065, doi:10.1111/j.1365-2958.1993.tb00976.x (1993).

3 Mojica, F. J. M. & Rodriguez-Valera, F. The discovery of CRISPR in archaea and bacteria. *The FEBS Journal* **283**, 3162-3169, doi:10.1111/febs.13766 (2016).

4 Jansen, R., Embden, J. D. A. v., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43**, 1565-1575, doi:10.1046/j.1365-2958.2002.02839.x (2002).

5 Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. **151**, 653-663, doi:https://doi.org/10.1099/mic.0.27437-0 (2005).

6 Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. **151**, 2551-2561, doi:https://doi.org/10.1099/mic.0.28048-0 (2005).

7 Mojica, F. J. M., Díez-Villaseñor, C. s., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution* **60**, 174-182, doi:10.1007/s00239-004-0046-3 (2005).

8 Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. **315**, 1709-1712, doi:10.1126/science.1138140 %J Science (2007).

9 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

10 Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Human Molecular Genetics* **23**, R40-R46, doi:10.1093/hmg/ddu125 %J Human Molecular Genetics (2014).

11 Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239-1259, doi:10.1016/j.cell.2017.11.032 (2018).

12 Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. **322**, 1843-1845, doi:10.1126/science.1165771 %J Science (2008).

13 McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nature Reviews Microbiology* **17**, 7-12, doi:10.1038/s41579-018-0071-7 (2019).

14 Lino, C. A., Harper, J. C., Carney, J. P. & Timlin, J. A. Delivering CRISPR: a review of the challenges and approaches. *Drug delivery* **25**, 1234-1257, doi:10.1080/10717544.2018.1474964 (2018).

15 Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2579-E2586, doi:10.1073/pnas.1208507109 (2012).

16 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).

17 Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nature methods* **10**, 957-963, doi:10.1038/nmeth.2649 (2013).

18 Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods* **10**, 1116-1121, doi:10.1038/nmeth.2681 (2013).
19 Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. **339**, 819-823, doi:doi:10.1126/science.1231143 (2013).
20 Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
21 Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology* **38**, 824-844, doi:10.1038/s41587-020-0561-9 (2020).
22 Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385-397, doi:10.1016/j.molcel.2015.10.008 (2015).
23 East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270-273, doi:10.1038/nature19802 (2016).
24 Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016).
25 O'Connell, M. R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. *J Mol Biol* **431**, 66-87, doi:10.1016/j.jmb.2018.06.029 (2019).
26 Smargon, A. A. *et al.* Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell* **65**, 618-630.e617, doi:10.1016/j.molcel.2016.12.023 (2017).
27 Yan, W. X. *et al.* Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Molecular cell* **70**, 327-339.e325, doi:10.1016/j.molcel.2018.02.028 (2018).
28 Xu, C. *et al.* Programmable RNA editing with compact CRISPR–Cas13 systems from uncultivated microbes. *Nature Methods* **18**, 499-506, doi:10.1038/s41592-021-01124-4 (2021).
29 Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science (New York, N.Y.)* **358**, 1019-1027, doi:10.1126/science.aaq0180 (2017).
30 Yang, L. Z. *et al.* Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Mol Cell* **76**, 981-997.e987, doi:10.1016/j.molcel.2019.10.024 (2019).
31 Freije, C. A. *et al.* Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Molecular Cell* **76**, 826-837.e811, doi:https://doi.org/10.1016/j.molcel.2019.09.013 (2019).
32 Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nature reviews. Molecular cell biology* **20**, 608-624, doi:10.1038/s41580-019-0168-5 (2019).
33 Perčulija, V., Lin, J., Zhang, B. & Ouyang, S. Functional Features and Current Applications of the RNA-Targeting Type VI CRISPR-Cas Systems. *Advanced Science* **8**, 2004685, doi:https://doi.org/10.1002/advs.202004685 (2021).
34 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).
35 East-Seletsky, A., O'Connell, M. R., Burstein, D., Knott, G. J. & Doudna, J. A. RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. *Molecular cell* **66**, 373-383.e373, doi:10.1016/j.molcel.2017.04.008 (2017).
36 Kannan, S. *et al.* Compact RNA editors with small Cas13 proteins. *Nature Biotechnology*, doi:10.1038/s41587-021-01030-2 (2021).
37 Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676.e614, doi:10.1016/j.cell.2018.02.033 (2018).
38 Zhang, H., Dong, C., Li, L., Wasney, G. A. & Min, J. Structural insights into the modulatory role of the accessory protein WYL1 in the Type VI-D CRISPR-Cas system. *Nucleic Acids Research* **47**, 5420-5428, doi:10.1093/nar/gkz269 (2019).
39 Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284, doi:10.1038/nature24049 (2017).

40 Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. **353**, aaf5573, doi:10.1126/science.aaf5573 %J Science (2016).
41 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
42 Karvelis, T. *et al.* crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA Biol* **10**, 841-851, doi:10.4161/rna.24203 (2013).
43 Murugan, K., Babu, K., Sundaresan, R., Rajan, R. & Sashital, D. G. The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. *Molecular cell* **68**, 15-25, doi:10.1016/j.molcel.2017.09.007 (2017).
44 Slaymaker, I. M. *et al.* High-Resolution Structure of Cas13b and Biochemical Characterization of RNA Targeting and Cleavage. *Cell reports* **26**, 3741-3751.e3745, doi:https://doi.org/10.1016/j.celrep.2019.02.094 (2019).
45 Zhang, C. *et al.* Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. *Cell* **175**, 212-223.e217, doi:10.1016/j.cell.2018.09.001 (2018).
46 Liu, L. *et al.* Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168**, 121-134.e112, doi:10.1016/j.cell.2016.12.031 (2017).
47 Liu, L. *et al.* The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170**, 714-726.e710, doi:10.1016/j.cell.2017.06.050 (2017).
48 Knott, G. J. *et al.* Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nature structural & molecular biology* **24**, 825-833, doi:10.1038/nsmb.3466 (2017).
49 Zhang, B. *et al.* Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. *Nature Communications* **10**, 2544, doi:10.1038/s41467-019-10507-3 (2019).
50 Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017).
51 Zhang, B. *et al.* Structural insights into Cas13b-guided CRISPR RNA maturation and recognition. *Cell Res* **28**, 1198-1201, doi:10.1038/s41422-018-0109-4 (2018).
52 O'Connell, M. R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. *Journal of Molecular Biology* **431**, 66-87, doi:https://doi.org/10.1016/j.jmb.2018.06.029 (2019).
53 Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442, doi:10.1126/science.aam9321 (2017).
54 Wang, B. *et al.* Structural basis for self-cleavage prevention by tag:anti-tag pairing complementarity in type VI Cas13 CRISPR systems. *Molecular Cell* **81**, 1100-1115.e1105, doi:10.1016/j.molcel.2020.12.033 (2021).
55 Tambe, A., East-Seletsky, A., Knott, G. J., Doudna, J. A. & O'Connell, M. R. RNA Binding and HEPN-Nuclease Activation Are Decoupled in CRISPR-Cas13a. *Cell reports* **24**, 1025-1036, doi:10.1016/j.celrep.2018.06.105 (2018).
56 Wessels, H. H. *et al.* Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat Biotechnol* **38**, 722-727, doi:10.1038/s41587-020-0456-9 (2020).
57 Shah, S. A., Erdmann, S., Mojica, F. J. M. & Garrett, R. A. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol* **10**, 891-899, doi:10.4161/rna.23764 (2013).
58 Granados-Riveron, J. T. & Aquino-Jarquin, G. CRISPR–Cas13 Precision Transcriptome Engineering in Cancer. **78**, 4107-4113, doi:10.1158/0008-5472.CAN-18-0785 %J Cancer Research (2018).
59 Wessels, H.-H. *et al.* Massively parallel Cas13 screens reveal principles for guide RNA design. *Nature Biotechnology*, doi:10.1038/s41587-020-0456-9 (2020).
60 Wei, J. *et al.* Deep learning and CRISPR-Cas13d ortholog discovery for optimized RNA targeting. *bioRxiv*, 2021.2009.2014.460134, doi:10.1101/2021.09.14.460134 (2022).
61 Li, S.-Y. *et al.* CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res* **28**, 491-493, doi:10.1038/s41422-018-0022-x (2018).

62 Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570**, 241-245, doi:10.1038/s41586-019-1257-5 (2019).
63 Gronowski, A. M. Who or What is SHERLOCK? *EJIFCC* **29**, 201-204 (2018).
64 Wang, Q. *et al.* The CRISPR-Cas13a Gene-Editing System Induces Collateral Cleavage of RNA in Glioma Cells. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)* **6**, 1901299-1901299, doi:10.1002/advs.201901299 (2019).
65 Özcan, A. *et al.* Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature* **597**, 720-725, doi:10.1038/s41586-021-03886-5 (2021).
66 Shi, P. *et al.* RNA-guided cell targeting with CRISPR/RfxCas13d collateral activity in human cells. *bioRxiv*, 2021.2011.2030.470032, doi:10.1101/2021.11.30.470032 (2021).
67 Tong, H. *et al.* High-fidelity Cas13 variants for targeted RNA degradation with minimal collateral effect. *bioRxiv*, 2021.2012.2018.473271, doi:10.1101/2021.12.18.473271 (2021).
68 Kushawah, G. *et al.* CRISPR-Cas13d Induces Efficient mRNA Knockdown in Animal Embryos. *Developmental Cell* **54**, 805-817.e807, doi:10.1016/j.devcel.2020.07.013 (2020).
69 Felekkis, K., Touvana, E., Stefanou, C. & Deltas, C. microRNAs: a newly described class of encoded molecules that play a role in health and disease. *Hippokratia* **14**, 236-240 (2010).
70 Terns, M. P. CRISPR-Based Technologies: Impact of RNA-Targeting Systems. *Molecular cell* **72**, 404-412, doi:10.1016/j.molcel.2018.09.018 (2018).
71 Wilson, R. C. & Doudna, J. A. Molecular mechanisms of RNA interference. *Annual review of biophysics* **42**, 217-239, doi:10.1146/annurev-biophys-083012-130404 (2013).
72 Agrawal, N. *et al.* RNA interference: biology, mechanism, and applications. *Microbiology and molecular biology reviews : MMBR* **67**, 657-685, doi:10.1128/mmbr.67.4.657-685.2003 (2003).
73 Qiu, S., Adema, C. M. & Lane, T. A computational study of off-target effects of RNA interference. *Nucleic Acids Research* **33**, 1834-1847, doi:10.1093/nar/gki324 (2005).
74 Jackson, A. L. & Linsley, P. S. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews Drug Discovery* **9**, 57-67, doi:10.1038/nrd3010 (2010).
75 Abudayyeh, O. O. *et al.* A cytosine deaminase for programmable single-base RNA editing. **365**, 382-386, doi:10.1126/science.aax7063 %J Science (2019).
76 Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945-956, doi:10.1016/j.cell.2009.07.040 (2009).
77 Staals, Raymond H. J. *et al.* RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Molecular Cell* **56**, 518-530, doi:10.1016/j.molcel.2014.10.005 (2014).
78 Samai, P. *et al.* Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**, 1164-1174, doi:10.1016/j.cell.2015.04.027 (2015).
79 Silas, S. *et al.* Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. *Science* **351**, aad4234, doi:10.1126/science.aad4234 (2016).
80 Wang, X. *et al.* Target RNA-guided protease activity in type III-E CRISPR–Cas system. *Nucleic Acids Research* **50**, 12913-12923, doi:10.1093/nar/gkac1151 (2022).
81 Strecker, J. *et al.* RNA-activated protein cleavage with a CRISPR-associated endopeptidase. *Science* **378**, 874-881, doi:10.1126/science.add7450 (2022).
82 Kato, K. *et al.* RNA-triggered protein cleavage and cell growth arrest by the type III-E CRISPR nuclease-protease. *Science* **378**, 882-889, doi:10.1126/science.add7347 (2022).
83 Wolter, F. & Puchta, H. The CRISPR/Cas revolution reaches the RNA world: Cas13, a new Swiss Army knife for plant biologists. *Plant J* **94**, 767-775, doi:10.1111/tpj.13899 (2018).
84 Wang, H. *et al.* CRISPR-mediated live imaging of genome editing and transcription. **365**, 1301-1305, doi:10.1126/science.aax7852 %J Science (2019).

85 Tang, T., Han, Y., Wang, Y., Huang, H. & Qian, P. Programmable System of Cas13-Mediated RNA Modification and Its Biological and Biomedical Applications. **9**, doi:10.3389/fcell.2021.677587 (2021).
86 Li, J. & Liu, C. Coding or Noncoding, the Converging Concepts of RNAs. **10**, doi:10.3389/fgene.2019.00496 (2019).
87 Xu, D. *et al.* A CRISPR/Cas13-based approach demonstrates biological relevance of vlinc class of long non-coding RNAs in anticancer drug response. *Scientific Reports* **10**, 1794, doi:10.1038/s41598-020-58104-5 (2020).
88 Li, S. *et al.* Screening for functional circular RNAs using the CRISPR-Cas13 system. 2020.2003.2023.002865, doi:10.1101/2020.03.23.002865 %J bioRxiv (2020).
89 Li, S. *et al.* Screening for functional circular RNAs using the CRISPR–Cas13 system. *Nature Methods* **18**, 51-59, doi:10.1038/s41592-020-01011-4 (2021).
90 Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360**, 439-444, doi:10.1126/science.aaq0179 (2018).
91 Myhrvold, C. *et al.* Field-deployable viral diagnostics using CRISPR-Cas13. *Science* **360**, 444-448, doi:10.1126/science.aas8836 (2018).
92 Arizti-Sanz, J. *et al.* Integrated sample inactivation, amplification, and Cas13-based detection of SARS-CoV-2. *bioRxiv*, 2020.2005.2028.119131, doi:10.1101/2020.05.28.119131 (2020).
93 Liu, Y. *et al.* CRISPR-Cas13a Nanomachine Based Simple Technology for Avian Influenza A (H7N9) Virus On-Site Detection. *Journal of Biomedical Nanotechnology* **15**, 790-798, doi:10.1166/jbn.2019.2742 (2019).
94 Wu, Y., Liu, S.-X., Wang, F. & Zeng, M.-S. Room Temperature Detection of Plasma Epstein–Barr Virus DNA with CRISPR–Cas13. *Clinical Chemistry* **65**, 591-592, doi:10.1373/clinchem.2018.299347 (2019).
95 Freije, C. A. *et al.* Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Mol Cell* **76**, 826-837.e811, doi:10.1016/j.molcel.2019.09.013 (2019).
96 Qin, P. *et al.* Rapid and Fully Microfluidic Ebola Virus Detection with CRISPR-Cas13a. *ACS sensors* **4**, 1048-1054, doi:10.1021/acssensors.9b00239 (2019).
97 Kaminski, M. M. *et al.* A CRISPR-based assay for the detection of opportunistic infections post-transplantation and for the monitoring of transplant rejection. *Nature biomedical engineering* **4**, 601-609, doi:10.1038/s41551-020-0546-5 (2020).
98 Shen, M. *et al.* Recent advances and perspectives of nucleic acid detection for coronavirus. *Journal of Pharmaceutical Analysis*, doi:https://doi.org/10.1016/j.jpha.2020.02.010 (2020).
99 Barnes, K. G. *et al.* Deployable CRISPR-Cas13a diagnostic tools to detect and report Ebola and Lassa virus cases in real-time. *Nature Communications* **11**, 4131, doi:10.1038/s41467-020-17994-9 (2020).
100 Patchsung, M. *et al.* Clinical validation of a Cas13-based assay for the detection of SARS-CoV-2 RNA. *Nature biomedical engineering* **4**, 1140-1149, doi:10.1038/s41551-020-00603-x (2020).
101 Shan, Y., Zhou, X., Huang, R. & Xing, D. High-Fidelity and Rapid Quantification of miRNA Combining crRNA Programmability and CRISPR/Cas13a trans-Cleavage Activity. *Analytical Chemistry* **91**, 5278-5285, doi:10.1021/acs.analchem.9b00073 (2019).
102 Ackerman, C. M. *et al.* Massively multiplexed nucleic acid detection with Cas13. *Nature* **582**, 277-282, doi:10.1038/s41586-020-2279-8 (2020).
103 Cui, C., Shu, W. & Li, P. Fluorescence In situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications. *Frontiers in cell and developmental biology* **4**, 89-89, doi:10.3389/fcell.2016.00089 (2016).
104 George, L., Indig, F. E., Abdelmohsen, K. & Gorospe, M. Intracellular RNA-tracking methods. *Open biology* **8**, doi:10.1098/rsob.180104 (2018).

105 Wang, Y., Yang, L.-Z. & Chen, L.-L. Protocol for Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 System. *STAR Protoc* **1**, 100037-100037, doi:10.1016/j.xpro.2020.100037 (2020).
106 Li, J. *et al.* Targeted mRNA demethylation using an engineered dCas13b-ALKBH5 fusion protein. *Nucleic Acids Res* **48**, 5684-5694, doi:10.1093/nar/gkaa269 (2020).
107 Rauch, S. & Dickinson, B. C. in *Methods in Enzymology* Vol. 621 (ed Arun K. Shukla) 1-16 (Academic Press, 2019).
108 Wilson, C., Chen, P. J., Miao, Z. & Liu, D. R. Programmable m6A modification of cellular RNAs with a Cas13-directed methyltransferase. *Nature Biotechnology* **38**, 1431-1440, doi:10.1038/s41587-020-0572-6 (2020).
109 Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA-protein interactions. *Nat Methods* **16**, 225-234, doi:10.1038/s41592-019-0330-1 (2019).
110 Zhang, Z. *et al.* Capturing RNA-protein interaction via CRUIS. *Nucleic Acids Res* **48**, e52, doi:10.1093/nar/gkaa143 (2020).
111 Han, S. *et al.* RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 22068-22079, doi:10.1073/pnas.2006617117 (2020).
112 Khadempar, S. *et al.* CRISPR-Cas9 in genome editing: Its function and medical applications. *J Cell Physiol* **234**, 5751-5761, doi:10.1002/jcp.27476 (2019).
113 Yang, G. & Huang, X. Methods and applications of CRISPR/Cas system for genome editing in stem cells. *Cell regeneration (London, England)* **8**, 33-41, doi:10.1016/j.cr.2019.08.001 (2019).
114 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).
115 Li, H. L., Nakano, T. & Hotta, A. Genetic correction using engineered nucleases for gene therapy applications. *Dev Growth Differ* **56**, 63-77, doi:10.1111/dgd.12107 (2014).
116 Demirci, S., Leonard, A., Haro-Mora, J. J., Uchida, N. & Tisdale, J. F. CRISPR/Cas9 for Sickle Cell Disease: Applications, Future Possibilities, and Challenges. *Adv Exp Med Biol* **1144**, 37-52, doi:10.1007/5584_2018_331 (2019).
117 Karponi, G. & Zogas, N. Gene Therapy For Beta-Thalassemia: Updated Perspectives. *The application of clinical genetics* **12**, 167-180, doi:10.2147/TACG.S178546 (2019).
118 Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21**, 121-131, doi:10.1038/nm.3793 (2015).
119 Cribbs, A. P. & Perera, S. M. W. Science and Bioethics of CRISPR-Cas9 Gene Editing: An Analysis Towards Separating Facts and Fiction. *The Yale journal of biology and medicine* **90**, 625-634 (2017).
120 Li, X.-J., Tu, Z., Yang, W. & Li, S. CRISPR: Established Editor of Human Embryos? *Cell stem cell* **21**, 295-296, doi:10.1016/j.stem.2017.08.007 (2017).
121 Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids* **4**, e264, doi:https://doi.org/10.1038/mtna.2015.37 (2015).
122 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology* **31**, 833-838, doi:10.1038/nbt.2675 (2013).
123 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
124 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
125 Cullot, G. *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature Communications* **10**, 1136, doi:10.1038/s41467-019-09006-2 (2019).

126 Ayanoğlu, F. B., Elçin, A. E. & Elçin, Y. M. Bioethical issues in genome editing by CRISPR-Cas9 technology. *Turkish journal of biology = Turk biyoloji dergisi* **44**, 110-120, doi:10.3906/biy-1912-52 (2020).
127 Caplan, A. L., Parent, B., Shen, M. & Plunkett, C. No time to waste—the ethical challenges created by CRISPR. *EMBO reports* **16**, 1421-1426, doi:https://doi.org/10.15252/embr.201541337 (2015).
128 Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology* **13**, 722-736, doi:10.1038/nrmicro3569 (2015).
129 Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nature reviews. Drug discovery* **16**, 19-34, doi:10.1038/nrd.2016.230 (2017).
130 Podlaha, O., Riester, M., De, S. & Michor, F. Evolution of the cancer genome. *Trends in genetics : TIG* **28**, 155-163, doi:10.1016/j.tig.2012.01.003 (2012).
131 Huang, C.-H., Lee, K.-C. & Doudna, J. A. Applications of CRISPR-Cas Enzymes in Cancer Therapeutics and Detection. *Trends in cancer* **4**, 499-512, doi:10.1016/j.trecan.2018.05.006 (2018).
132 Poosala, P., Lindley, S. R., Anderson, K. M. & Anderson, D. M. Targeting Toxic Nuclear RNA Foci with CRISPR-Cas13 to Treat Myotonic Dystrophy. 716514, doi:10.1101/716514 %J bioRxiv (2019).
133 Zhao, X. *et al.* A CRISPR-Cas13a system for efficient and specific therapeutic targeting of mutant KRAS for pancreatic cancer treatment. *Cancer Lett* **431**, 171-181, doi:10.1016/j.canlet.2018.05.042 (2018).
134 Tibolla, G., Norata, G. D., Artali, R., Meneghetti, F. & Catapano, A. L. Proprotein convertase subtilisin/kexin type 9 (PCSK9): from structure-function relation to therapeutic inhibition. *Nutrition, metabolism, and cardiovascular diseases : NMCD* **21**, 835-843, doi:10.1016/j.numecd.2011.06.002 (2011).
135 Li, H. *et al.* CRISPR-Cas13a Cleavage of Dengue Virus NS3 Gene Efficiently Inhibits Viral Replication. *Molecular Therapy - Nucleic Acids* **19**, 1460-1469, doi:10.1016/j.omtn.2020.01.028 (2020).
136 Kamath, A. V. Translational pharmacokinetics and pharmacodynamics of monoclonal antibodies. *Drug Discovery Today: Technologies* **21-22**, 75-83, doi:https://doi.org/10.1016/j.ddtec.2016.09.004 (2016).
137 Blanchard, E. L. *et al.* Treatment of influenza and SARS-CoV-2 infections via mRNA-encoded Cas13a in rodents. *Nature Biotechnology* **39**, 717-726, doi:10.1038/s41587-021-00822-w (2021).
138 Abbott, T. R. *et al.* Development of CRISPR as a prophylactic strategy to combat novel coronavirus and influenza. 2020.2003.2013.991307, doi:10.1101/2020.03.13.991307 %J bioRxiv (2020).
139 Fareh, M. *et al.* Reprogrammed CRISPR-Cas13b suppresses SARS-CoV-2 replication and circumvents its mutational escape through mismatch tolerance. *Nature Communications* **12**, 4270, doi:10.1038/s41467-021-24577-9 (2021).
140 Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371-376, doi:10.1038/nature25795 (2018).
141 Wong, M. *et al.* Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nature medicine* **26**, 1742-1753, doi:10.1038/s41591-020-1072-4 (2020).
142 Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311-317, doi:10.1038/nature22973 (2017).
143 Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J. & Der, C. J. Drugging the undruggable RAS: Mission possible? *Nature reviews. Drug discovery* **13**, 828-851, doi:10.1038/nrd4389 (2014).
144 Bushweller, J. H. Targeting transcription factors in cancer — from undruggable to reality. *Nature Reviews Cancer* **19**, 611-624, doi:10.1038/s41568-019-0196-7 (2019).

145 Tomlins, S. A. *et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia (New York, N.Y.)* **10**, 177-188, doi:10.1593/neo.07822 (2008).
146 Lazo, J. S. & Sharlow, E. R. Drugging Undruggable Molecular Cancer Targets. *Annual review of pharmacology and toxicology* **56**, 23-40, doi:10.1146/annurev-pharmtox-010715-103440 (2016).
147 Rosenzweig, S. A. Acquired Resistance to Drugs Targeting Tyrosine Kinases. *Advances in cancer research* **138**, 71-98, doi:10.1016/bs.acr.2018.02.003 (2018).
148 CRISPR Causes Unexpected Genomic Damage. *Cancer discovery* **8**, Of2, doi:10.1158/2159-8290.Cd-nb2018-106 (2018).
149 Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765-771, doi:10.1038/nbt.4192 (2018).
150 Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503-2519.e2517, doi:10.1016/j.cell.2021.03.025 (2021).
151 Li, Y. *et al.* Collateral cleavage of 28s rRNA by RfxCas13d causes death of mice. *bioRxiv*, 2022.2001.2017.476700, doi:10.1101/2022.01.17.476700 (2022).
152 Ai, Y., Liang, D. & Wilusz, J. E. CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. *Nucleic Acids Research* **50**, e65-e65, doi:10.1093/nar/gkac159 (2022).
153 Tafer, H. & Hofacker, I. L. RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics (Oxford, England)* **24**, 2657-2663, doi:10.1093/bioinformatics/btn193 (2008).
154 Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res* **45**, W435-w439, doi:10.1093/nar/gkx279 (2017).
155 Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645-3647, doi:10.1093/bioinformatics/btx469 %J Bioinformatics (2017).
156 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95, doi:10.1109/MCSE.2007.55 (2007).
157 Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics* **46**, 505-529, doi:10.1146/annurev-biophys-062215-010822 (2017).
158 Ma, H. *et al.* Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Molecular therapy. Nucleic acids* **3**, e161, doi:10.1038/mtna.2014.12 (2014).
159 Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications* **12**, 555, doi:10.1038/s41467-020-20633-y (2021).
160 Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306-313, doi:10.1038/nature10762 (2012).
161 Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* **108**, 479-485, doi:10.1038/bjc.2012.581 (2013).
162 Blagosklonny, M. V. Why therapeutic response may not prolong the life of a cancer patient: selection for oncogenic resistance. *Cell Cycle* **4**, 1693-1698, doi:10.4161/cc.4.12.2259 (2005).
163 Jacoby, M. A., Duncavage, E. J. & Walter, M. J. Implications of Tumor Clonal Heterogeneity in the Era of Next-Generation Sequencing. *Trends Cancer* **1**, 231-241, doi:10.1016/j.trecan.2015.10.006 (2015).
164 Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).
165 Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724, doi:10.1038/nature07943 (2009).
166 Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *The New England journal of medicine* **364**, 2507-2516, doi:10.1056/NEJMoa1103782 (2011).
167 Gora-Tybor, J. & Robak, T. Targeted drugs in chronic myeloid leukemia. *Curr Med Chem* **15**, 3036-3051, doi:10.2174/092986708786848578 (2008).

168 Damodaran, S. & Olson, E. M. Targeting the human epidermal growth factor receptor 2 pathway in breast cancer. *Hospital practice (1995)* **40**, 7-15, doi:10.3810/hp.2012.10.997 (2012).
169 Hoelder, S., Clarke, P. A. & Workman, P. Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Molecular oncology* **6**, 155-176, doi:10.1016/j.molonc.2012.02.004 (2012).
170 De Palma, M. & Hanahan, D. The biology of personalized cancer medicine: facing individual complexities underlying hallmark capabilities. *Mol Oncol* **6**, 111-127, doi:10.1016/j.molonc.2012.01.011 (2012).
171 McDermott, U., Downing, J. R. & Stratton, M. R. Genomics and the continuum of cancer care. *The New England journal of medicine* **364**, 340-350, doi:10.1056/NEJMra0907178 (2011).
172 Nussinov, R., Jang, H., Tsai, C.-J. & Cheng, F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLoS computational biology* **15**, e1006658-e1006658, doi:10.1371/journal.pcbi.1006658 (2019).
173 Parker, B. C. & Zhang, W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chinese journal of cancer* **32**, 594-603, doi:10.5732/cjc.013.10178 (2013).
174 Brown, L. M. *et al.* SFPQ-ABL1 and BCR-ABL1 use different signaling networks to drive B-cell acute lymphoblastic leukemia. *Blood Adv* **6**, 2373-2387, doi:10.1182/bloodadvances.2021006076 (2022).
175 Ernst, T. *et al.* Identification of FOXP1 and SNX2 as novel ABL1 fusion partners in acute lymphoblastic leukaemia. *British journal of haematology* **153**, 43-46, doi:10.1111/j.1365-2141.2010.08457.x (2011).
176 Duhoux, F. P. *et al.* The t(1;9)(p34;q34) fusing ABL1 with SFPQ, a pre-mRNA processing gene, is recurrent in acute lymphoblastic leukemias. *Leukemia research* **35**, e114-117, doi:10.1016/j.leukres.2011.02.011 (2011).
177 De Braekeleer, E. *et al.* A new partner gene fused to ABL1 in a t(1;9)(q24;q34)-associated B-cell acute lymphoblastic leukemia. *Leukemia* **21**, 2220-2221, doi:10.1038/sj.leu.2404773 (2007).
178 Al-Harbi, S., Aljurf, M., Mohty, M., Almohareb, F. & Ahmed, S. O. A. An update on the molecular pathogenesis and potential therapeutic targeting of AML with t(8;21)(q22;q22.1);RUNX1-RUNX1T1. *Blood Advances* **4**, 229-238, doi:10.1182/bloodadvances.2019000168 (2020).
179 Pearson, J. D., Lee, J. K., Bacani, J. T., Lai, R. & Ingham, R. J. NPM-ALK: The Prototypic Member of a Family of Oncogenic Fusion Tyrosine Kinases. *Journal of signal transduction* **2012**, 123253, doi:10.1155/2012/123253 (2012).
180 Druker, B. J. *et al.* Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *The New England journal of medicine* **355**, 2408-2417, doi:10.1056/NEJMoa062867 (2006).
181 Nicolini, F. E. *et al.* Clinical outcome of 27 imatinib mesylate-resistant chronic myelogenous leukemia patients harboring a T315I BCR-ABL mutation. *Haematologica* **92**, 1238-1241, doi:10.3324/haematol.11369 (2007).
182 Calabrese, C. *et al.* Genomic basis for RNA alterations in cancer. *Nature* **578**, 129-136, doi:10.1038/s41586-020-1970-0 (2020).
183 Papadopoulos, P., Ridge, S. A., Boucher, C. A., Stocking, C. & Wiedemann, L. M. The novel activation of ABL by fusion to an ets-related gene, TEL. *Cancer Res* **55**, 34-38 (1995).
184 Parker, B. C. & Zhang, W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chinese journal of cancer* **32**, 594-603, doi:10.5732/cjc.013.10178 (2013).

185 Druker, B. J. Imatinib as a paradigm of targeted therapies. *Advances in cancer research* **91**, 1-30, doi:10.1016/s0065-230x(04)91001-9 (2004).
186 Hensley, M. L. & Ford, J. M. Imatinib treatment: specific issues related to safety, fertility, and pregnancy. *Seminars in hematology* **40**, 21-25, doi:10.1053/shem.2003.50038 (2003).
187 Li, Z. *et al.* Intrinsic RNA targeting constrains the utility of CRISPR-Cas13 systems. *bioRxiv*, 2022.2005.2014.491940, doi:10.1101/2022.05.14.491940 (2022).
188 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).
189 Kelley, C. P., Haerle, M. C. & Wang, E. T. Negative autoregulation mitigates collateral RNase activity of repeat-targeting CRISPR-Cas13d in mammalian cells. *Cell reports* **40**, 111226, doi:10.1016/j.celrep.2022.111226 (2022).
190 Bot, J. F., Zhao, Z., Kammeron, D., Shang, P. & Geijsen, N. Collateral RNA cleavage by CRISPR-Cas13 allows selective cell elimination. *bioRxiv*, 2023.2001.2019.524716, doi:10.1101/2023.01.19.524716 (2023).
191 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).
192 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).
193 Lima, S. A. *et al.* Short poly(A) tails are a conserved feature of highly expressed genes. *Nature structural & molecular biology* **24**, 1057-1063, doi:10.1038/nsmb.3499 (2017).
194 Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* **136**, 763-776, doi:10.1016/j.cell.2009.01.019 (2009).
195 Leppek, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature reviews. Molecular cell biology* **19**, 158-174, doi:10.1038/nrm.2017.103 (2018).
196 Fareh, M. *et al.* TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments. *Nature Communications* **7**, 13694, doi:10.1038/ncomms13694 (2016).
197 Burris, B. J. D., Molina Vargas, A. M., Park, B. J. & O'Connell, M. R. Optimization of specific RNA knockdown in mammalian cells with CRISPR-Cas13. *Methods* **206**, 58-68, doi:https://doi.org/10.1016/j.ymeth.2022.08.007 (2022).
198 Verma, M. Personalized medicine and cancer. *Journal of personalized medicine* **2**, 1-14, doi:10.3390/jpm2010001 (2012).
199 Jin, J. *et al.* Identification of Genetic Mutations in Cancer: Challenge and Opportunity in the New Era of Targeted Therapy. *Frontiers in oncology* **9**, 263, doi:10.3389/fonc.2019.00263 (2019).
200 Jeon, J. *et al.* A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Medicine* **6**, 57, doi:10.1186/s13073-014-0057-7 (2014).
201 Schilsky, R. L., Allen, J., Benner, J., Sigal, E. & McClellan, M. Commentary: tackling the challenges of developing targeted therapies for cancer. *The oncologist* **15**, 484-487, doi:10.1634/theoncologist.2010-0079 (2010).
202 Zhong, L. *et al.* Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. *Signal Transduction and Targeted Therapy* **6**, 201, doi:10.1038/s41392-021-00572-w (2021).
203 Nussinov, R., Jang, H., Tsai, C.-J. & Cheng, F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLOS Computational Biology* **15**, e1006658, doi:10.1371/journal.pcbi.1006658 (2019).

204 Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27-38, doi:10.1016/j.cell.2013.09.006 (2013).
205 Katti, A., Diaz, B. J., Caragine, C. M., Sanjana, N. E. & Dow, L. E. CRISPR in cancer biology and therapy. *Nature Reviews Cancer* **22**, 259-279, doi:10.1038/s41568-022-00441-w (2022).
206 Haslam, A., Kim, M. S. & Prasad, V. Updated estimates of eligibility for and response to genome-targeted oncology drugs among US cancer patients, 2006-2020. *Annals of Oncology* **32**, 926-932, doi:10.1016/j.annonc.2021.04.003 (2021).
207 Amer, M. H. Gene therapy for cancer: present status and future perspective. *Molecular and cellular therapies* **2**, 27-27, doi:10.1186/2052-8426-2-27 (2014).
208 Mali, S. Delivery systems for gene therapy. *Indian journal of human genetics* **19**, 3-8, doi:10.4103/0971-6866.112870 (2013).
209 Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P. & Zakian, S. M. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta naturae* **6**, 19-40 (2014).
210 Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188**, 773-782, doi:10.1534/genetics.111.131433 (2011).
211 Sawamoto, K., Chen, H.-H., Alméciga-Díaz, C. J., Mason, R. W. & Tomatsu, S. Gene therapy for Mucopolysaccharidoses. *Molecular genetics and metabolism* **123**, 59-68, doi:10.1016/j.ymgme.2017.12.434 (2018).
212 Lee, J., Chung, J.-H., Kim, H. M., Kim, D.-W. & Kim, H. Designed nucleases for targeted genome editing. *Plant Biotechnology Journal* **14**, 448-462, doi:10.1111/pbi.12465 (2016).
213 Enache, O. M. *et al.* Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nature Genetics* **52**, 662-668, doi:10.1038/s41588-020-0623-4 (2020).
214 Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature medicine* **24**, 927-930, doi:10.1038/s41591-018-0049-z (2018).
215 Philippidis, A. Brother of Cure Rare Disease CEO Dies in Trial of Duchenne Muscular Dystrophy Therapy. *Human Gene Therapy* **33**, 1224-1227, doi:10.1089/hum.2022.29228.bfs (2022).
216 Ghodsizadeh, O. Cleaving RNA with Cas9. *Nature Methods* **11**, 1090-1090, doi:10.1038/nmeth.3164 (2014).
217 Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A. & Doudna, J. A. RNA-dependent RNA targeting by CRISPR-Cas9. *eLife* **7**, e32724, doi:10.7554/eLife.32724 (2018).
218 Mout, R. *et al.* Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing. *ACS Nano* **11**, 2452-2458, doi:10.1021/acsnano.6b07600 (2017).
219 Nelles, D. A., Fang, M. Y., Aigner, S. & Yeo, G. W. Applications of Cas9 as an RNA-programmed RNA-binding protein. *Bioessays* **37**, 732-739, doi:10.1002/bies.201500001 (2015).
220 Sudbery, I., Enright, A. J., Fraser, A. G. & Dunham, I. Systematic analysis of off-target effects in an RNAi screen reveals microRNAs affecting sensitivity to TRAIL-induced apoptosis. *BMC Genomics* **11**, 175, doi:10.1186/1471-2164-11-175 (2010).
221 Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537-541, doi:10.1038/nature04791 (2006).
222 Aagaard, L. & Rossi, J. J. RNAi therapeutics: principles, prospects and challenges. *Adv Drug Deliv Rev* **59**, 75-86, doi:10.1016/j.addr.2007.03.005 (2007).
223 Wu, Y. *et al.* Precise editing of FGFR3-TACC3 fusion genes with CRISPR-Cas13a in glioblastoma. *Molecular therapy : the journal of the American Society of Gene Therapy* **29**, 3305-3318, doi:10.1016/j.ymthe.2021.07.002 (2021).
224 Saifullah, Sakari, M., Suzuki, T., Yano, S. & Tsukahara, T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. *International journal of molecular sciences* **21**, doi:10.3390/ijms21238904 (2020).

225 Bandaru, S. *et al.* Structure-based design of gRNA for Cas13. *Scientific Reports* **10**, 11610, doi:10.1038/s41598-020-68459-4 (2020).
226 Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat Biotechnol* **34**, 328-333, doi:10.1038/nbt.3471 (2016).
227 Handumrongkul, C. *et al.* Durable multitransgene expression in vivo using systemic, nonviral DNA delivery. **5**, eaax0217, doi:10.1126/sciadv.aax0217 %J Science Advances (2019).
228 Deyle, D. R. & Russell, D. W. Adeno-associated virus vector integration. *Current opinion in molecular therapeutics* **11**, 442-447 (2009).
229 Eoh, J. & Gu, L. Biomaterials as vectors for the delivery of CRISPR–Cas9. *Biomaterials Science* **7**, 1240-1261, doi:10.1039/C8BM01310A (2019).
230 Stahl, P. D. & Raposo, G. Extracellular Vesicles: Exosomes and Microvesicles, Integrators of Homeostasis. **34**, 169-177, doi:10.1152/physiol.00045.2018 (2019).
231 Khorsandi, S. E. *et al.* Minimally invasive and selective hydrodynamic gene therapy of liver segments in the pig and human. *Cancer Gene Ther* **15**, 225-230, doi:10.1038/sj.cgt.7701119 (2008).
232 Hou, X., Zaks, T., Langer, R. & Dong, Y. Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials* **6**, 1078-1094, doi:10.1038/s41578-021-00358-0 (2021).
233 Wadhwa, A., Aljabbari, A., Lokras, A., Foged, C. & Thakur, A. Opportunities and Challenges in the Delivery of mRNA-based Vaccines. *Pharmaceutics* **12**, doi:10.3390/pharmaceutics12020102 (2020).
234 Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology* **33**, 985-989, doi:10.1038/nbt.3290 (2015).
235 Kristen, A. V. *et al.* Patisiran, an RNAi therapeutic for the treatment of hereditary transthyretin-mediated amyloidosis. **9**, 5-23, doi:10.2217/nmt-2018-0033 (2019).
236 Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine* **384**, 403-416, doi:10.1056/NEJMoa2035389 (2020).
237 Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* **383**, 2603-2615, doi:10.1056/NEJMoa2034577 (2020).
238 Cui, Z. *et al.* Cas13d knockdown of lung protease Ctsl prevents and treats SARS-CoV-2 infection. *Nature Chemical Biology* **18**, 1056-1064, doi:10.1038/s41589-022-01094-4 (2022).
239 Cheng, Q. *et al.* Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nature nanotechnology* **15**, 313-320, doi:10.1038/s41565-020-0669-6 (2020).
240 Smargon, A. A., Shi, Y. J. & Yeo, G. W. RNA-targeting CRISPR systems from metagenomic discovery to transcriptomic engineering. *Nat Cell Biol* **22**, 143-150, doi:10.1038/s41556-019-0454-7 (2020).
241 Morelli, K. H. *et al.* An RNA-targeting CRISPR–Cas13d system alleviates disease-related phenotypes in Huntington's disease models. *Nature Neuroscience*, doi:10.1038/s41593-022-01207-1 (2022).
242 Wagner, D. L. *et al.* High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nature medicine* **25**, 242-248, doi:10.1038/s41591-018-0204-6 (2019).
243 Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature medicine* **25**, 249-254, doi:10.1038/s41591-018-0326-x (2019).
244 Ferdosi, S. R. *et al.* Multifunctional CRISPR-Cas9 with engineered immunosilenced human T cell epitopes. *Nat Commun* **10**, 1842, doi:10.1038/s41467-019-09693-x (2019).
245 You, L. *et al.* Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Molecular therapy. Methods & clinical development* **13**, 359-370, doi:10.1016/j.omtm.2019.02.008 (2019).

246 Moreno, A. M. *et al.* Immune-orthogonal orthologues of AAV capsids and of Cas9 circumvent the immune response to the administration of gene therapy. *Nature biomedical engineering* **3**, 806-816, doi:10.1038/s41551-019-0431-2 (2019).
247 Bhisetti, G. & Fang, C. in *Artificial Intelligence in Drug Design* (ed Alexander Heifetz) 409-419 (Springer US, 2022).
248 Lin, P. *et al.* CRISPR-Cas13 Inhibitors Block RNA Editing in Bacteria and Mammalian Cells. *Molecular Cell* **78**, 850-861.e855, doi:https://doi.org/10.1016/j.molcel.2020.03.033 (2020).

Appendix

Appendix Table 1. crRNA spacer sequences used in this study

Name	Spacer sequence (forward strand)	crRNA Target/Description
NT crRNA	TAGATTGCTGTTCTACCAAGTAATCCATCA	Non targeting crRNA
PspmCherry1 (matching position 2)	CCATGTTATCCTCCTCGCCCTTGCTCACCA	crRNA targeting mCherry
PspmCherry2 (matching position 3)	GGCCATGTTATCCTCCTCGCCCTTGCTCAC	crRNA targeting mCherry
PspmCherry3 (matching position 49)	GGAGCCCTCCATGTGCACCTTGAAGCGCAT	crRNA targeting mCherry
PspmCherry4 (matching position 52)	CACGGAGCCCTCCATGTGCACCTTGAAGCG	crRNA targeting mCherry
PspmCherry5 (matching position 118)	GGCGGTCTGGGTGCCCTCGTAGGGGCGGCC	crRNA targeting mCherry
PspmCherry6 (matching position 193)	GGAGCCGTACATGAACTGAGGGGACAGGAT	crRNA targeting mCherry
PspmCherry7 (matching position 250)	GGGGAAGGACAGCTTCAAGTAGTCGGGGAT	crRNA targeting mCherry
PspmCherry8 (matching position 252)	TCGGGGAAGGACAGCTTCAAGTAGTCGGGG	crRNA targeting mCherry
PspmCherry9 (matching position 319)	GGAGTCCTGGGTCACGGTCACCACGCCGCC	crRNA targeting mCherry
PspmCherry10 (matching position 315)	TCCTGGGTCACGGTCACCACGCCGCCGTCC	crRNA targeting mCherry
PspmCherry11 (matching position 406)	GGTCTTCTTCTGCATTACGGGGCCGTCGGA	crRNA targeting mCherry
PspmCherry12 (matching position 455)	GGGCGCCGTCCTCGGGGTACATCCGCTCGG	crRNA targeting mCherry
PspmCherry13 (matching position 526)	GGTGGTCTTGACCTCAGCGTCGTAGTGGCC	crRNA targeting mCherry
PspmCherry14 (matching position 593)	TCCAACTTGATGTTGACGTTGTAGGCGCCG	crRNA targeting mCherry
PspmCherry15 (matching position 585)	GGTGATGTCCAACTTGATGTTGACGTTGTA	crRNA targeting mCherry
PspmCherry16 (matching position 655)	GGTGGAGTGGCGGCCCTCGGCGCGTTCGTA	crRNA targeting mCherry
Pspmcherry12(-15nt)	TGATCTCGCCCTTCAGGGCGCCGTCCTCGG	crRNA targeting mCherry
Pspmcherry12(-12nt)	TCTCGCCCTTCAGGGCGCCGTCCTCGGGGT	Tiled crRNA targeting mCherry
Pspmcherry12(-9nt)	CGCCCTTCAGGGCGCCGTCCTCGGGGTACA	Tiled crRNA targeting mCherry

Pspmcherry12(-6nt)	CCTTCAGGGCGCCGTCCTCGGGGTACATCC	Tiled crRNA targeting mCherry
Pspmcherry12(-3nt)	TCAGGGCGCCGTCCTCGGGGTACATCCGCT	Tiled crRNA targeting mCherry
Pspmcherry12(+3nt)	CGCCGTCCTCGGGGTACATCCGCTCGGAGG	Tiled crRNA targeting mCherry
Pspmcherr128(+6nt)	CGTCCTCGGGGTACATCCGCTCGGAGGAGG	Tiled crRNA targeting mCherry
Pspmcherry12(+9nt)	CCTCGGGGTACATCCGCTCGGAGGAGGCCT	Tiled crRNA targeting mCherry
Pspmcherry12(+12nt)	CGGGGTACATCCGCTCGGAGGAGGCCTCCC	Tiled crRNA targeting mCherry
Pspmcherry12(+15nt)	GGTACATCCGCTCGGAGGAGGCCTCCCAGC	Tiled crRNA targeting mCherry
Pspmcherry16(-15nt)	CTCGTCCATGCCGCCGGTGGAGTGGCGGCC	Tiled crRNA targeting mCherry
Pspmcherry16(-12nt)	GTCCATGCCGCCGGTGGAGTGGCGGCCCTC	Tiled crRNA targeting mCherry
Pspmcherry16(-9nt)	CATGCCGCCGGTGGAGTGGCGGCCCTCGGC	Tiled crRNA targeting mCherry
Pspmcherry16(-6nt))	GCCGCCGGTGGAGTGGCGGCCCTCGGCGCG	Tiled crRNA targeting mCherry
Pspmcherry16(-3nt)	GCCGGTGGAGTGGCGGCCCTCGGCGCGTTC	Tiled crRNA targeting mCherry
Pspmcherry16(+3nt)	GGAGTGGCGGCCCTCGGCGCGTTCGTACTG	Tiled crRNA targeting mCherry
Pspmcherry16(+6nt)	GTGGCGGCCCTCGGCGCGTTCGTACTGTTC	Tiled crRNA targeting mCherry
Pspmcherry16(+9nt)_	GCGGCCCTCGGCGCGTTCGTACTGTTCCAC	Tiled crRNA targeting mCherry
Pspmcherry16(+12nt)	GCCCTCGGCGCGTTCGTACTGTTCCACGAT	Tiled crRNA targeting mCherry
Pspmcherry16(+15nt)	CTCGGCGCGTTCGTACTGTTCCACGATGGT	Tiled crRNA targeting mCherry
PspmCherry tiled1 (matching position 485)	GCTTCAGCCTCTGCTTGATCTCGCCCTTCA	Tiled crRNA targeting mCherry
PspmCherry tiled2 (matching position 484)	CTTCAGCCTCTGCTTGATCTCGCCCTTCAG	Tiled crRNA targeting mCherry

PspmCherry tiled3 (matching position 483)	TTCAGCCTCTGCTTGATCTCGCCCTTCAGG	Tiled crRNA targeting mCherry
PspmCherry tiled4 (matching position 482)	TCAGCCTCTGCTTGATCTCGCCCTTCAGGG	Tiled crRNA targeting mCherry
PspmCherry tiled5 (matching position 481)	CAGCCTCTGCTTGATCTCGCCCTTCAGGGC	Tiled crRNA targeting mCherry
PspmCherry tiled6 (matching position 480)	AGCCTCTGCTTGATCTCGCCCTTCAGGGCG	Tiled crRNA targeting mCherry
PspmCherry tiled7 (matching position 479)	GCCTCTGCTTGATCTCGCCCTTCAGGGCGC	Tiled crRNA targeting mCherry
PspmCherry tiled8 (matching position 478)	CCTCTGCTTGATCTCGCCCTTCAGGGCGCC	Tiled crRNA targeting mCherry
PspmCherry tiled9 (matching position 477)	CTCTGCTTGATCTCGCCCTTCAGGGCGCCG	Tiled crRNA targeting mCherry
PspmCherry tiled10 (matching position 476)	TCTGCTTGATCTCGCCCTTCAGGGCGCCGT	Tiled crRNA targeting mCherry
PspmCherry tiled11 (matching position 475)	CTGCTTGATCTCGCCCTTCAGGGCGCCGTC	Tiled crRNA targeting mCherry
PspmCherry tiled12 (matching position 474)	TGCTTGATCTCGCCCTTCAGGGCGCCGTCC	Tiled crRNA targeting mCherry
PspmCherry tiled13 (matching position 473)	GCTTGATCTCGCCCTTCAGGGCGCCGTCCT	Tiled crRNA targeting mCherry
PspmCherry tiled14 (matching position 472)	CTTGATCTCGCCCTTCAGGGCGCCGTCCTC	Tiled crRNA targeting mCherry
PspmCherry titling15 (matching position 471)	TTGATCTCGCCCTTCAGGGCGCCGTCCTCG	Tiled crRNA targeting mCherry
PspmCherry tiled16 (matching position 470)	TGATCTCGCCCTTCAGGGCGCCGTCCTCGG	Tiled crRNA targeting mCherry
PspmCherry tiled17 (matching position 469)	GATCTCGCCCTTCAGGGCGCCGTCCTCGGG	Tiled crRNA targeting mCherry
PspmCherry tiled18 (matching position 468)	ATCTCGCCCTTCAGGGCGCCGTCCTCGGGG	Tiled crRNA targeting mCherry
PspmCherry tiled19 (matching position 467)	TCTCGCCCTTCAGGGCGCCGTCCTCGGGGT	Tiled crRNA targeting mCherry
PspmCherry tiled20 (matching position 466)	CTCGCCCTTCAGGGCGCCGTCCTCGGGGTA	Tiled crRNA targeting mCherry
PspmCherry tiled21 (matching position 465)	TCGCCCTTCAGGGCGCCGTCCTCGGGGTAC	Tiled crRNA targeting mCherry
PspmCherry tiled22 (matching position 464)	CGCCCTTCAGGGCGCCGTCCTCGGGGTACA	Tiled crRNA targeting mCherry
PspmCherry tiled23 (matching position 463)	GCCCTTCAGGGCGCCGTCCTCGGGGTACAT	Tiled crRNA targeting mCherry
PspmCherry tiled24 (matching position 462)	CCCTTCAGGGCGCCGTCCTCGGGGTACATC	Tiled crRNA targeting mCherry
PspmCherry tiled25 (matching position 461)	CCTTCAGGGCGCCGTCCTCGGGGTACATCC	Tiled crRNA targeting mCherry
PspmCherry tiled26 (matching position 460)	CTTCAGGGCGCCGTCCTCGGGGTACATCCG	Tiled crRNA targeting mCherry

PspmCherry tiled27 (matching position 459)	TTCAGGGCGCCGTCCTCGGGGTACATCCGC	Tiled crRNA targeting mCherry
PspmCherry tiled28 (matching position 458)	TCAGGGCGCCGTCCTCGGGGTACATCCGCT	Tiled crRNA targeting mCherry
PspmCherry tiled29 (matching position 457)	CAGGGCGCCGTCCTCGGGGTACATCCGCTC	Tiled crRNA targeting mCherry
PspmCherry tiled30 (matching position 456)	AGGGCGCCGTCCTCGGGGTACATCCGCTCG	Tiled crRNA targeting mCherry
PspmCherry tiled31 (matching position 455)	GGGCGCCGTCCTCGGGGTACATCCGCTCGG	Tiled crRNA targeting mCherry
PspmCherry tiled32 (matching position 454)	GGCGCCGTCCTCGGGGTACATCCGCTCGGA	Tiled crRNA targeting mCherry
PspmCherry tiled33 (matching position 453)	GCGCCGTCCTCGGGGTACATCCGCTCGGAG	Tiled crRNA targeting mCherry
PspmCherry tiled34 (matching position 452)	CGCCGTCCTCGGGGTACATCCGCTCGGAGG	Tiled crRNA targeting mCherry
PspmCherry tiled35 (matching position 451)	GCCGTCCTCGGGGTACATCCGCTCGGAGGA	Tiled crRNA targeting mCherry
PspmCherry tiled36 (matching position 450)	CCGTCCTCGGGGTACATCCGCTCGGAGGAG	Tiled crRNA targeting mCherry
PspmCherry tiled37 (matching position 449)	CGTCCTCGGGGTACATCCGCTCGGAGGAGG	Tiled crRNA targeting mCherry
PspmCherry tiled38 (matching position 448)	GTCCTCGGGGTACATCCGCTCGGAGGAGGC	Tiled crRNA targeting mCherry
PspmCherry tiled39 (matching position 447)	TCCTCGGGGTACATCCGCTCGGAGGAGGCC	Tiled crRNA targeting mCherry
PspmCherry tiled40 (matching position 446)	CCTCGGGGTACATCCGCTCGGAGGAGGCCT	Tiled crRNA targeting mCherry
PspmCherry tiled41 (matching position 445)	CTCGGGGTACATCCGCTCGGAGGAGGCCTC	Tiled crRNA targeting mCherry
PspmCherry tiled42 (matching position 444)	TCGGGGTACATCCGCTCGGAGGAGGCCTCC	Tiled crRNA targeting mCherry
PspmCherry tiled43 (matching position 443)	CGGGGTACATCCGCTCGGAGGAGGCCTCCC	Tiled crRNA targeting mCherry
PspmCherry tiled44 (matching position 442)	GGGGTACATCCGCTCGGAGGAGGCCTCCCA	Tiled crRNA targeting mCherry
PspmCherry tiled 45 (matching position 441)	GGGTACATCCGCTCGGAGGAGGCCTCCCAG	Tiled crRNA targeting mCherry
PspmCherry tiled46 (matching position 440)	GGTACATCCGCTCGGAGGAGGCCTCCCAGC	Tiled crRNA targeting mCherry
PspmCherry tiled47 (matching position 439)	GTACATCCGCTCGGAGGAGGCCTCCCAGCC	Tiled crRNA targeting mCherry
PspmCherry tiled48 (matching position 438)	TACATCCGCTCGGAGGAGGCCTCCCAGCCC	Tiled crRNA targeting mCherry
PspmCherry tiled49 (matching position 437)	ACATCCGCTCGGAGGAGGCCTCCCAGCCCA	Tiled crRNA targeting mCherry
PspmCherry tiled50 (matching position 436)	CATCCGCTCGGAGGAGGCCTCCCAGCCCAT	Tiled crRNA targeting mCherry

PspmCherry tiled51 (matching position 435)	ATCCGCTCGGAGGAGGCCTCCCAGCCCATC	Tiled crRNA targeting mCherry
PspmCherry tiled52 (matching position 434)	TCCGCTCGGAGGAGGCCTCCCAGCCCATCG	Tiled crRNA targeting mCherry
PspmCherry tiled53 (matching position 432)	CCGCTCGGAGGAGGCCTCCCAGCCCATCGT	Tiled crRNA targeting mCherry
PspmCherry tiled54 (matching position 431)	CGCTCGGAGGAGGCCTCCCAGCCCATCGTC	Tiled crRNA targeting mCherry
PspmCherry tiled55 (matching position 430)	GCTCGGAGGAGGCCTCCCAGCCCATCGTCT	Tiled crRNA targeting mCherry
PspmCherry tiled56 (matching position 429)	CTCGGAGGAGGCCTCCCAGCCCATCGTCTT	Tiled crRNA targeting mCherry
PspmCherry tiled57 (matching position 428)	TCGGAGGAGGCCTCCCAGCCCATCGTCTTC	Tiled crRNA targeting mCherry
PspmCherry tiled58 (matching position 427)	CGGAGGAGGCCTCCCAGCCCATCGTCTTCT	Tiled crRNA targeting mCherry
PspmCherry tiled59 (matching position 426)	GGAGGAGGCCTCCCAGCCCATCGTCTTCTT	Tiled crRNA targeting mCherry
PspmCherry tiled60 (matching position 425)	GAGGAGGCCTCCCAGCCCATCGTCTTCTTC	Tiled crRNA targeting mCherry
PspmCherry tiled61 (matching position 424)	AGGAGGCCTCCCAGCCCATCGTCTTCTTCT	Tiled crRNA targeting mCherry
BCR-ABL1P190 crRNA-20	GGCUCAAAGUCAGAUGCUACUGGCCGCUGA	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-19	GCUCAAAGUCAGAUGCUACUGGCCGCUGAA	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-18	CUCAAAGUCAGAUGCUACUGGCCGCUGAAG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-17	UCAAAGUCAGAUGCUACUGGCCGCUGAAGG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-16	CAAAGUCAGAUGCUACUGGCCGCUGAAGGG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-15	AAAGUCAGAUGCUACUGGCCGCUGAAGGGC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-14	AAGUCAGAUGCUACUGGCCGCUGAAGGGCU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-13	AGUCAGAUGCUACUGGCCGCUGAAGGGCUU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-12	GUCAGAUGCUACUGGCCGCUGAAGGGCUUC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-11	UCAGAUGCUACUGGCCGCUGAAGGGCUUCU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-10	CAGAUGCUACUGGCCGCUGAAGGGCUUCUG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-9	AGAUGCUACUGGCCGCUGAAGGGCUUCUGC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-8	GAUGCUACUGGCCGCUGAAGGGCUUCUGCG	Tiled crRNA targeting BCR-ABL1

BCR-ABL1P190 crRNA-7	AUGCUACUGGCCGCUGAAGGGCUUCUGCGU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-6	UGCUACUGGCCGCUGAAGGGCUUCUGCGUC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-5	GCUACUGGCCGCUGAAGGGCUUCUGCGUCU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-4	CUACUGGCCGCUGAAGGGCUUCUGCGUCUC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-3	UACUGGCCGCUGAAGGGCUUCUGCGUCUCC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-2	ACUGGCCGCUGAAGGGCUUCUGCGUCUCCA	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA-1	CUGGCCGCUGAAGGGCUUCUGCGUCUCCAU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA0	UGGCCGCUGAAGGGCUUCUGCGUCUCCAUG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA1	GGCCGCUGAAGGGCUUCUGCGUCUCCAUGG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA2	GCCGCUGAAGGGCUUCUGCGUCUCCAUGGA	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA3	CCGCUGAAGGGCUUCUGCGUCUCCAUGGAA	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA4	CGCUGAAGGGCUUCUGCGUCUCCAUGGAAG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA5	GCUGAAGGGCUUCUGCGUCUCCAUGGAAGG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA6	CUGAAGGGCUUCUGCGUCUCCAUGGAAGGC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA7	UGAAGGGCUUCUGCGUCUCCAUGGAAGGCG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA8	GAAGGGCUUCUGCGUCUCCAUGGAAGGCGC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA9	AAGGGCUUCUGCGUCUCCAUGGAAGGCGCC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA10	AGGGCUUCUGCGUCUCCAUGGAAGGCGCCC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA11	GGGCUUCUGCGUCUCCAUGGAAGGCGCCCU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA12	GGCUUCUGCGUCUCCAUGGAAGGCGCCCUC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA13	GCUUCUGCGUCUCCAUGGAAGGCGCCCUCG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA14	CUUCUGCGUCUCCAUGGAAGGCGCCCUCGC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA15	UUCUGCGUCUCCAUGGAAGGCGCCCUCGCC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA16	UCUGCGUCUCCAUGGAAGGCGCCCUCGCCA	Tiled crRNA targeting BCR-ABL1

BCR-ABL1P190 crRNA17	CUGCGUCUCCAUGGAAGGCGCCCUCGCCAU	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA18	UGCGUCUCCAUGGAAGGCGCCCUCGCCAUC	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA19	GCGUCUCCAUGGAAGGCGCCCUCGCCAUCG	Tiled crRNA targeting BCR-ABL1
BCR-ABL1P190 crRNA20	CGUCUCCAUGGAAGGCGCCCUCGCCAUCGU	Tiled crRNA targeting BCR-ABL1
SNX2-ABL1-12	GAAGCGGCUCUCGGAGGAGACGUAGAGCUC	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1-9	GCGGCUCUCGGAGGAGACGUAGAGCUCUUC	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1-6	GCUCUCGGAGGAGACGUAGAGCUCUUCCCU	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1-3	CUCGGAGGAGACGUAGAGCUCUUCCCUGGA	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1+0	GGAGGAGACGUAGAGCUCUUCCCUGGAUCU	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1+3	GGAGACGUAGAGCUCUUCCCUGGAUCUAUC	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1+6	GACGUAGAGCUCUUCCCUGGAUCUAUCAAA	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1+9	GUAGAGCUCUUCCCUGGAUCUAUCAAAGAU	Tiled crRNA targeting SNX2-ABL1
SNX2-ABL1+12	GAGCUCUUCCCUGGAUCUAUCAAAGAUCAC	Tiled crRNA targeting SNX2-ABL1
SFPQ-ABL1-12	GAAGCGGCUCUCGGAGGAGACGUAGAGCAU	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1-9	GCGGCUCUCGGAGGAGACGUAGAGCAUGUC	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1-6	GCUCUCGGAGGAGACGUAGAGCAUGUCACU	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1-3	CUCGGAGGAGACGUAGAGCAUGUCACUUCC	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1+0	GGAGGAGACGUAGAGCAUGUCACUUCCCAU	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1+3	GGAGACGUAGAGCAUGUCACUUCCCAUCAU	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1+6	GACGUAGAGCAUGUCACUUCCCAUCAUGGA	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1+9	GUAGAGCAUGUCACUUCCCAUCAUGGAACC	Tiled crRNA targeting SFPQ-ABL1
SFPQ-ABL1+12	GAGCAUGUCACUUCCCAUCAUGGAACCACU	Tiled crRNA targeting SFPQ-ABL1
Spike crRNA -TIL1	GTCGCGCACCAGGTTGATTGGGGTGTGCTT	Tiled crRNA targeting Spike
Spike crRNA-TIL2	TCGCGCACCAGGTTGATTGGGGTGTGCTTG	Tiled crRNA targeting Spike

Spike crRNA-TIL3	CGCGCACCAGGTTGATTGGGGTGTGCTTGG	Tiled crRNA targeting Spike
Spike crRNA-TIL4	GCGCACCAGGTTGATTGGGGTGTGCTTGGA	Tiled crRNA targeting Spike
Spike crRNA-TILF5	CGCACCAGGTTGATTGGGGTGTGCTTGGAG	Tiled crRNA targeting Spike
Spike crRNA-TIL6	GCACCAGGTTGATTGGGGTGTGCTTGGAGT	Tiled crRNA targeting Spike
Spike crRNA-TIL7	CACCAGGTTGATTGGGGTGTGCTTGGAGTA	Tiled crRNA targeting Spike
Spike crRNA-TIL8	ACCAGGTTGATTGGGGTGTGCTTGGAGTAG	Tiled crRNA targeting Spike
Spike crRNA-TILF9	CCAGGTTGATTGGGGTGTGCTTGGAGTAGA	Tiled crRNA targeting Spike
Spike crRNA-TIL10	CAGGTTGATTGGGGTGTGCTTGGAGTAGAT	Tiled crRNA targeting Spike
Spike crRNA-TIL11	AGGTTGATTGGGGTGTGCTTGGAGTAGATC	Tiled crRNA targeting Spike
Spike crRNA-TIL12	GGTTGATTGGGGTGTGCTTGGAGTAGATCT	Tiled crRNA targeting Spike
Spike crRNA-TIL13	GTTGATTGGGGTGTGCTTGGAGTAGATCTT	Tiled crRNA targeting Spike
Spike crRNA-TIL14	TTGATTGGGGTGTGCTTGGAGTAGATCTTA	Tiled crRNA targeting Spike
Spike crRNA-TIL15	TGATTGGGGTGTGCTTGGAGTAGATCTTAA	Tiled crRNA targeting Spike
Spike crRNA-TIL16	GATTGGGGTGTGCTTGGAGTAGATCTTAAA	Tiled crRNA targeting Spike
Spike crRNA-TIL17	ATTGGGGTGTGCTTGGAGTAGATCTTAAAG	Tiled crRNA targeting Spike
Spike crRNA-TIL18	TTGGGGTGTGCTTGGAGTAGATCTTAAAGT	Tiled crRNA targeting Spike
Spike crRNA-TIL19	TGGGGTGTGCTTGGAGTAGATCTTAAAGTA	Tiled crRNA targeting Spike
Spike crRNA-TIL20	GGGGTGTGCTTGGAGTAGATCTTAAAGTAG	Tiled crRNA targeting Spike
Spike crRNA-TIL21	GGGTGTGCTTGGAGTAGATCTTAAAGTAGC	Tiled crRNA targeting Spike
Spike crRNA-TIL22	GGTGTGCTTGGAGTAGATCTTAAAGTAGCC	Tiled crRNA targeting Spike
Spike crRNA-TIL23	GTGTGCTTGGAGTAGATCTTAAAGTAGCCA	Tiled crRNA targeting Spike
Spike crRNA-TIL24	TGTGCTTGGAGTAGATCTTAAAGTAGCCAT	Tiled crRNA targeting Spike
Spike crRNA-TIL25	GTGCTTGGAGTAGATCTTAAAGTAGCCATC	Tiled crRNA targeting Spike
Spike crRNA-TIL26	TGCTTGGAGTAGATCTTAAAGTAGCCATCG	Tiled crRNA targeting Spike

Spike crRNA-TIL27	GCTTGGAGTAGATCTTAAAGTAGCCATCGA	Tiled crRNA targeting Spike
Spike crRNA-TIL28	CTTGGAGTAGATCTTAAAGTAGCCATCGAT	Tiled crRNA targeting Spike
Spike crRNA-TIL29	TTGGAGTAGATCTTAAAGTAGCCATCGATG	Tiled crRNA targeting Spike
Spike crRNA-TIL30	TGGAGTAGATCTTAAAGTAGCCATCGATGT	Tiled crRNA targeting Spike
Spike crRNA-TIL31	GGAGTAGATCTTAAAGTAGCCATCGATGTT	Tiled crRNA targeting Spike
Spike crRNA-TIL32	GAGTAGATCTTAAAGTAGCCATCGATGTTC	Tiled crRNA targeting Spike
Spike crRNA-TIL33	AGTAGATCTTAAAGTAGCCATCGATGTTCT	Tiled crRNA targeting Spike
Spike crRNA-TIL34	GTAGATCTTAAAGTAGCCATCGATGTTCTT	Tiled crRNA targeting Spike
Spike crRNA-TIL35	TAGATCTTAAAGTAGCCATCGATGTTCTTA	Tiled crRNA targeting Spike
Spike crRNA-TIL36	AGATCTTAAAGTAGCCATCGATGTTCTTAA	Tiled crRNA targeting Spike
Spike crRNA-TIL37	GATCTTAAAGTAGCCATCGATGTTCTTAAA	Tiled crRNA targeting Spike
Spike crRNA-TIL38	ATCTTAAAGTAGCCATCGATGTTCTTAAAC	Tiled crRNA targeting Spike
Spike crRNA-TIL39	TCTTAAAGTAGCCATCGATGTTCTTAAACA	Tiled crRNA targeting Spike
Spike crRNA-TIL40	CTTAAAGTAGCCATCGATGTTCTTAAACAC	Tiled crRNA targeting Spike
Spike crRNA-TIL41	TTAAAGTAGCCATCGATGTTCTTAAACACG	Tiled crRNA targeting Spike
Spike crRNA-TIL42	TAAAGTAGCCATCGATGTTCTTAAACACGA	Tiled crRNA targeting Spike
Spike crRNA-TIL43	AAAGTAGCCATCGATGTTCTTAAACACGAA	Tiled crRNA targeting Spike
Spike crRNA-TIL44	AAGTAGCCATCGATGTTCTTAAACACGAAC	Tiled crRNA targeting Spike
Spike crRNA-TIL45	AGTAGCCATCGATGTTCTTAAACACGAACT	Tiled crRNA targeting Spike
Spike crRNA-TIL46	GTAGCCATCGATGTTCTTAAACACGAACTC	Tiled crRNA targeting Spike
Spike crRNA-TIL47	TAGCCATCGATGTTCTTAAACACGAACTCC	Tiled crRNA targeting Spike
Spike crRNA-TIL48	AGCCATCGATGTTCTTAAACACGAACTCCC	Tiled crRNA targeting Spike
Spike crRNA-TIL49	GCCATCGATGTTCTTAAACACGAACTCCCG	Tiled crRNA targeting Spike
Spike crRNA-TIL50	CCATCGATGTTCTTAAACACGAACTCCCGC	Tiled crRNA targeting Spike

Spike crRNA-TIL51	CATCGATGTTCTTAAACACGAACTCCCGCA	Tiled crRNA targeting Spike
Spike crRNA-TIL52	ATCGATGTTCTTAAACACGAACTCCCGCAG	Tiled crRNA targeting Spike
Spike crRNA-TIL53	TCGATGTTCTTAAACACGAACTCCCGCAGG	Tiled crRNA targeting Spike
Spike crRNA-TIL54	CGATGTTCTTAAACACGAACTCCCGCAGGT	Tiled crRNA targeting Spike
Spike crRNA-TIL55	GATGTTCTTAAACACGAACTCCCGCAGGTT	Tiled crRNA targeting Spike
Spike crRNA-TIL56	ATGTTCTTAAACACGAACTCCCGCAGGTTC	Tiled crRNA targeting Spike
Spike crRNA-TIL57	TGTTCTTAAACACGAACTCCCGCAGGTTCT	Tiled crRNA targeting Spike
Spike crRNA-TIL58	GTTCTTAAACACGAACTCCCGCAGGTTCTT	Tiled crRNA targeting Spike
Spike crRNA-TIL59	TTCTTAAACACGAACTCCCGCAGGTTCTTG	Tiled crRNA targeting Spike
Spike crRNA-TIL60	TCTTAAACACGAACTCCCGCAGGTTCTTGA	Tiled crRNA targeting Spike
Spike crRNA-TIL61	CTTAAACACGAACTCCCGCAGGTTCTTGAA	Tiled crRNA targeting Spike
PspmCherry12 3ntMSM 1-3nt-mismatches	**CCC**CGCCGUCCUCGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 4-6nt-mismatches	GGG**GCG**CGUCCUCGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 7-9nt-mismatches	GGGCGC**GCA**CCUCGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 10-12nt-mismatches	GGGCGCCGU**GGA**CGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 13-15nt-mismatches	GGGCGCCGUCCU**GCC**GGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 16-18nt-mismatches	GGGCGCCGUCCUCGG**CCA**ACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 19-21nt-mismatches	GGGCGCCGUCCUCGGGGU**UGU**UCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 22-24nt-mismatches	GGGCGCCGUCCUCGGGGUACA**AGG**GCUCGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 25-27nt-mismatches	GGGCGCCGUCCUCGGGGUACAUCC**CGA**CGG	mCherry Spacer mutagenesis
PspmCherry12 3ntMSM 28-30nt-mismatches	GGGCGCCGUCCUCGGGGUACAUCCGCU**GCC**	mCherry Spacer mutagenesis
PspmCherry12 4ntMSM_1-4nt-mismatches	**CCCG**GCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 4ntMSM_5-8nt-mismatches	GGGC**CGGC**TCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 4ntMSM_9-12nt-mismatches	GGGCGCCG**AGGA**CGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis

PspmCherry12 4ntMSM_13-16nt-mismatches	GGGCGCCGTCCT**GCCC**GTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 4ntMSM_17-20nt-mismatches	GGGCGCCGTCCTCGGG**CATG**ATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 4ntMSM_21-24nt-mismatches	GGGCGCCGTCCTCGGGGTAC**TAGG**GCTCGG	mCherry Spacer mutagenesis
PspmCherry12 4ntMSM_25-28nt-mismatches	GGGCGCCGTCCTCGGGGTACATCC**CGAG**GG	mCherry Spacer mutagenesis
PspmCherry12 5ntMSM_1-5nt-mismatches	**CCCGC**CCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 5ntMSM_6-10nt-mismatches	GGGCG**GGCAG**CTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 5ntMSM_11-15nt-mismatches	GGGCGCCGTC**GAGCC**GGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 5ntMSM_16-20nt-mismatches	GGGCGCCGTCCTCGG**CCATG**ATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 5ntMSM_21-25nt-mismatches	GGGCGCCGTCCTCGGGGTAC**TAGGC**CTCGG	mCherry Spacer mutagenesis
PspmCherry12 5nt MSM_26-30nt-mismatches	GGGCGCCGTCCTCGGGGTACATCCG**GAGCC**	mCherry Spacer mutagenesis
PspmCherry12 6ntMSM_1-6nt-mismatches	**CCCGCG**CGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 6ntMSM_7-12nt-mismatches	GGGCGC**GCAGGA**CGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 6ntMSM_13-18nt-mismatches	GGGCGCCGTCCT**GCCCCA**ACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12 6ntMSM_19-24nt-mismatches	GGGCGCCGTCCTCGGGGT**TGTAGG**GCTCGG	mCherry Spacer mutagenesis
PspmCherry12 6ntMSM_25-30nt-mismatches	GGGCGCCGTCCTCGGGGTACATCC**CGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 28-30nt-mismatches	GGGCGCCGUCCUCGGGGUACAUCCGCU**GCC**	mCherry Spacer mutagenesis
PspmCherry12 25-30nt-mismatches	GGGCGCCGUCCUCGGGGUACAUCC**CGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 22-30nt-mismatches	GGGCGCCGUCCUCGGGGUACA**AGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 19-30nt-mismatches	GGGCGCCGUCCUCGGGGU**UGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 16-30nt-mismatches	GGGCGCCGUCCUCGG**CCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 13-30nt-mismatches	GGGCGCCGUCCU**GCCCCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 10-30nt-mismatches	GGGCGCCGU**GGAGCCCCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 7-30nt-mismatches	GGGCGC**GCAGGAGCCCCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 4-30nt-mismatches	GGG**GCGGCAGGAGCCCCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis

PspmCherry12 1-30nt-mismatches	**CCCGCGGCAGGAGCCCCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry12 1-3nt-mismatches	**CCC**CGCCGUCCUCGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-6nt-mismatches	**CCCGCG**CGUCCUCGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-9nt–mismatches	**CCCGCGGCA**CCUCGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-12nt-mismatches	**CCCGCGGCAGGA**CGGGGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-15nt-mismatches	**CCCGCGGCAGGAGCC**GGUACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-18nt-mismatches	**CCCGCGGCAGGAGCCCCA**ACAUCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-21nt-mismatches	**CCCGCGGCAGGAGCCCCAUGU**UCCGCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-24nt-mismatches	**CCCGCGGCAGGAGCCCCAUGUAGG**GCUCGG	mCherry Spacer mutagenesis
PspmCherry12 1-27nt-mismatches	**CCCGCGGCAGGAGCCCCAUGUAGGCGA**CGG	mCherry Spacer mutagenesis
PspmCherry12 1-30nt-mismatches	**CCCGCGGCAGGAGCCCCAUGUAGGCGAGCC**	mCherry Spacer mutagenesis
PspmCherry1_1-3nt-mismatches	**CCG**CATGTTATCCTCCTCGCCCTTGCTCAC	mCherry Spacer mutagenesis
PspmCherry3_1-3nt-mismatches	**CCT**GCCCTCCATGTGCACCTTGAAGCGCAT	mCherry Spacer mutagenesis
PspmCherry4_1-3nt-mismatches	**CCG**GGTCTGGGTGCCCTCGTAGGGGCGGCC	mCherry Spacer mutagenesis
PspmCherry6_1-3nt-mismatches	**CCT**GCCGTACATGAACTGAGGGGACAGGAT	mCherry Spacer mutagenesis
PspmCherry7_1-3nt-mismatches	**CCC**GAAGGACAGCTTCAAGTAGTCGGGGAT	mCherry Spacer mutagenesis
PspmCherry9_1-3nt-mismatches	**CCT**GTCCTGGGTCACGGTCACCACGCCGCC	mCherry Spacer mutagenesis
PspmCherry11_1-3nt-mismatches	**CCA**CTTCTTCTGCATTACGGGGCCGTCGGA	mCherry Spacer mutagenesis
PspmCherry12_1-3nt-mismatches	**CCC**CGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry13_1-3nt-mismatches	**CCA**GGTCTTGACCTCAGCGTCGTAGTGGCC	mCherry Spacer mutagenesis
PspmCherry14_1-3nt-mismatches	**CCA**GATGTCCAACTTGATGTTGACGTTGTA	mCherry Spacer mutagenesis
PspmCherry16_1-3nt-mismatches	**CCA**GGAGTGGCGGCCCTCGGCGCGTTCGTA	mCherry Spacer mutagenesis
PspmCherry12_1-2nt-mismatches	**CC**GCGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12_2-3nt-mismatches	G**CC**CGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis

PspmCherry12_1st&3rdnt-mismatches	**C**G**C**CGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12_1st nt-mismatch	**C**GGCGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12_2nd nt-mismatch	G**G**GCGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
PspmCherry12_3rd nt-mismatch	GG**C**CGCCGTCCTCGGGGTACATCCGCTCGG	mCherry Spacer mutagenesis
pspmcherry12 non-consecutive 5 MSM	GGGCGGCGTCCACGGGGAACATCGGCTCGC	mCherry Spacer mutagenesis
pspmcherry12 non-consecutive 4 MSM	GGGCGCGGTCCTCCGGGTACTTCCGCTGGG	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 3 MSM	GGGCGCCCTCCTCGGCGTACATCGGCTCGG	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 2 MSM	GGGCGCCGTGCTCGGGGTAGATCCGCTCGG	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 15 MSM	GCGGGGCCTGCACCGCGAAGAACGGGTGGC	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 10 MSM	GGCCGGCGACCACGCGGAACTTCGGCACGC	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 7 MSM	GGGGGCCCTCCACGGCGTAGATCGGCTGGG	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 6 MSM	GGGCCCCGTGCTCGCGGTAGATCCCCTCGC	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 2 MSM	GGCGGCGCTCGACGCCGTTGATGGGCAGGG	mCherry Spacer mutagenesis
Pspmcherry12 non-consecutive 3 MSM	GGGGCGCGTGGACGGCCAACAAGGGCTGCC	mCherry Spacer mutagenesis
PspmCherry tiled 24 1G	GCCTTCAGGGCGCCGTCCTCGGGGTACATC	Validation of crRNA prediction
PspmCherry tiled 25 1G	GCTTCAGGGCGCCGTCCTCGGGGTACATCC	Validation of crRNA prediction
PspmCherry tiled 26 1G	GTTCAGGGCGCCGTCCTCGGGGTACATCCG	Validation of crRNA prediction
PspCherry tiled 27 1G	GTCAGGGCGCCGTCCTCGGGGTACATCCGC	Validation of crRNA prediction
PspmCherry tiled 28 1G	GCAGGGCGCCGTCCTCGGGGTACATCCGCT	Validation of crRNA prediction
PspmCherry tiled 24 2G	GGCTTCAGGGCGCCGTCCTCGGGGTACATC	Validation of crRNA prediction
PspmCherry tiled 25 1G	GGTTCAGGGCGCCGTCCTCGGGGTACATCC	Validation of crRNA prediction
PspCherry tiled 26 2G	GGTCAGGGCGCCGTCCTCGGGGTACATCCG	Validation of crRNA prediction
PspmCherry tiled 27 2G	GGCAGGGCGCCGTCCTCGGGGTACATCCGC	Validation of crRNA prediction
PspmCherry tiled 28 2G	GGAGGGCGCCGTCCTCGGGGTACATCCGCT	Validation of crRNA prediction

PspCherry tiled 39 extra G	GTCCTCGGGGTACATCCGCTCGGAGGAGGCC	Validation of crRNA prediction
PspmCherry tiled 40 extra G	GCCTCGGGGTACATCCGCTCGGAGGAGGCCT	Validation of crRNA prediction
PspmCherry tiled 39 1G	GCCTCGGGGTACATCCGCTCGGAGGAGGCC	Validation of crRNA prediction
PspmCherry tiled 40 1G	GCTCGGGGTACATCCGCTCGGAGGAGGCCT	Validation of crRNA prediction
PspmCherry tiled 39 2G	GGCTCGGGGTACATCCGCTCGGAGGAGGCC	Validation of crRNA prediction
PspmCherry tiled 40 2G	GGTCGGGGTACATCCGCTCGGAGGAGGCCT	Validation of crRNA prediction
PspmCherry tiled 24 extra G	GCCCTTCAGGGCGCCGTCCTCGGGGTACATC	Validation of crRNA prediction
PspCherry tiled 25 extra G	GCCTTCAGGGCGCCGTCCTCGGGGTACATCC	Validation of crRNA prediction
PspCherry tiled 26 extra G	GCTTCAGGGCGCCGTCCTCGGGGTACATCCG	Validation of crRNA prediction
PspCherry tiled 27 extra G	GTTCAGGGCGCCGTCCTCGGGGTACATCCGC	Validation of crRNA prediction
PspCherry tiled 28 extra G	GTCAGGGCGCCGTCCTCGGGGTACATCCGCT	Validation of crRNA prediction
BCR-ABL1P190 crRNA -3nt G1-2	GGCTGGCCGCTGAAGGGCTTCTGCGTCTCC	Validation of crRNA prediction
BCR-ABL1P190 crRNA +6nt G1-2	GGGAAGGGCTTCTGCGTCTCCATGGAAGGC	Validation of crRNA prediction
BCR-ABL1P190 crRNA +9nt G1-2	GGGGGCTTCTGCGTCTCCATGGAAGGCGCC	Validation of crRNA prediction
BCR-ABL1P190 crRNA +9nt G15-17-18	AAGGGCTTCTGCGTGTGGATGGAAGGCGCC	Validation of crRNA prediction
BCR-ABL1P190 crRNA +9nt G17-18	AAGGGCTTCTGCGTCTGGATGGAAGGCGCC	Validation of crRNA prediction
BCR-ABL1P190 crRNA +9nt G1-2-15-17-18	GGGGGCTTCTGCGTGTGGATGGAAGGCGCC	Validation of crRNA prediction
SNX2-ABL1 crRNA +6nt G2	GGCGTAGAGCTCTTCCCTGGATCTATCAAA	Validation of crRNA prediction
SNX2-ABL1 crRNA +6nt G15-16	GACGTAGAGCTCTTGGCTGGATCTATCAAA	Validation of crRNA prediction
SNX2-ABL1 crRNA +6nt G15-16-17	GACGTAGAGCTCTTGGGTGGATCTATCAAA	Validation of crRNA prediction
SNX2-ABL1 crRNA +6nt G2-15-16-17	GGCGTAGAGCTCTTGGGTGGATCTATCAAA	Validation of crRNA prediction
SNX2-ABL1 crRNA +12nt G2	GGGCTCTTCCCTGGATCTATCAAAGATCAC	Validation of crRNA prediction
SNX2-ABL1 crRNA +12nt G11	GAGCTCTTCCGTGGATCTATCAAAGATCAC	Validation of crRNA prediction
SNX2-ABL1 crRNA +12nt G17	GAGCTCTTCCCTGGATGTATCAAAGATCAC	Validation of crRNA prediction

SNX2-ABL1 crRNA +12nt G2-11-17	GGGCTCTTCCGTGGATGTATCAAAGATCAC	Validation of crRNA prediction
SNX2-ABL1 crRNA +3nt G15-18	GGAGACGTAGAGCTGTTGCCTGGATCTATC	Validation of crRNA prediction
SNX2-ABL1 crRNA +3nt G15	GGAGACGTAGAGCTGTTCCCTGGATCTATC	Validation of crRNA prediction
RUNX1-RUNX1T1 1	GGAGTGCTTCTCAGTACGATTTCGAGGTTC	Validation of crRNA prediction
RUNX1-RUNX1T1 2	AGTACGATTTCGAGGTTCTCGGGGCCCATC	Validation of crRNA prediction
RUNX1-RUNX1T1 3	TGCTTCTCAGTACGATTTCGAGGTTCTCGG	Validation of crRNA prediction
RUNX1-RUNX1T1 4	GTGCTTCTCAGTACGATTTCGAGGTTCTCG	Validation of crRNA prediction
RUNX1-RUNX1T1 5	TGGAGTGCTTCTCAGTACGATTTCGAGGTT	Validation of crRNA prediction
NPM-ALK 1	GGCGGTACACTACTAAGTGCTGTCCACTAA	Validation of crRNA prediction
NPM-ALK 2	GGTACACTACTAAGTGCTGTCCACTAATAT	Validation of crRNA prediction
NPM-ALK 3	GGTGCTTCCGGCGGTACACTACTAAGTGCT	Validation of crRNA prediction
NPM-ALK 4	TGGTGCTTCCGGCGGTACACTACTAAGTGC	Validation of crRNA prediction
NPM-ALK 5	GTGCTTCCGGCGGTACACTACTAAGTGCTG	Validation of crRNA prediction
BCR-ABL1P190 0 1-3nt MSM	ACCCCGCTGAAGGGCTTCTGCGTCTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 1-6nt MSM	ACCGGCCTGAAGGGCTTCTGCGTCTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 1-9nt MSM	ACCGGCGACAAGGGCTTCTGCGTCTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 28-30nt MSM	TGGCCGCTGAAGGGCTTCTGCGTCTCCTAC	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 25-30nt MSM	TGGCCGCTGAAGGGCTTCTGCGTCAGGTAC	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 22-30nt MSM	TGGCCGCTGAAGGGCTTCTGCCAGAGGTAC	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 13-15nt MSM	TGGCCGCTGAAGCCGTTCTGCGTCTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 12-17nt MSM	TGGCCGCTGAACCCGAACTGCGTCTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 11-18nt MSM	TGGCCGCTGTTCCCGAAGTGCGTCTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 3MSM	TGGCCGCAGAAGGGCATCTGCGTGTCCATG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 4MSM	TGGCCGGTGAAGGCCTTCTGCCTCTCCAAG	BCR-ABL1 spacer mutagenesis

BCR-ABL1P190 0 5 MSM	TGGCCCCTGAACGGCTTGTGCGTGTCCATC	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 6 MSM	TGGCGGCTGTAGGGGTTCTCCGTCACCATC	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 7 MSM	TGGGCGCAGAACGGCATCTCCGTGTCCTTG	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 10 MSM	TGCCCCCTCAACGGGTTGTGGGTGTCGATC	BCR-ABL1 spacer mutagenesis
BCR-ABL1P190 0 14nt MSM	TGCGCGGAGATCGGGATCACCGAGTCGTTG	BCR-ABL1 spacer mutagenesis
BCR-ABL1 P210 crRNA0	GCCGCTGAAGGGCTTTTGAACTCTGCTTAA	Target BCR-ABL1 P210 breakpoint
Ineffective EGFP crRNA1	**C**CGGTGAACAG**C**TC**C**T**CG**CCCTTGCTCACC	Validation of crRNA prediction
Ineffective EGFP crRNA2	**C**AGGATGGGCA**C**CA**CCC**CGGTGAACAGCTC	Validation of crRNA prediction
Ineffective EGFP crRNA3	**C**TTGTAGTTGC**C**GT**CGTC**CTTGAAGAAGAT	Validation of crRNA prediction
Ineffective EGFP crRNA4	**C**ACCAGGGTGT**C**GC**CCTC**GAACTTCACCTC	Validation of crRNA prediction
Ineffective EGFP crRNA5	**C**AGGATGTTGC**C**GT**CCTC**CTTGAAGTCGAT	Validation of crRNA prediction
Ineffective EGFP crRNA6	**C**TTCTGCTTGT**C**GG**CCA**TGATATAGACGTT	Validation of crRNA prediction
Ineffective EGFP crRNA7	**C**TGCACGCTGC**C**GT**CCTC**GATGTTGTGGCG	Validation of crRNA prediction
Ineffective EGFP crRNA8	**C**TTGTACAGCT**C**GT**CCA**TGCCGAGAGTGAT	Validation of crRNA prediction
Ineffective EGFP crRNA9	**T**ACTCCAGCTT**G**TG**CCCC**AGGATGTTGCCG	Validation of crRNA prediction
Potent EGFP crRNA1	**GG**TCAGCTTGCCGTAGGTGGCATCGCCCTC	Validation of crRNA prediction
Potent EGFP crRNA2	**GG**TCACGAGGGTGGG**CC**AGGGCACGGGCAG	Validation of crRNA prediction
Potent EGFP crRNA3	**GG**TAGCGGCTGAAG**CA**CTGCACGCCGTAGG	Validation of crRNA prediction
Potent EGFP crRNA4	**GG**GCATGGCGGACTTGAAGAAGTCGTGCTG	Validation of crRNA prediction
Potent EGFP crRNA5	**GG**TGCGCTCCTGGA**C**GTAGCCTTCGGGCAT	Validation of crRNA prediction
Potent EGFP crRNA6	**GG**CGCGGGTCTTGTAGTTGCCGTCGTCCTT	Validation of crRNA prediction
Potent EGFP crRNA7	**GG**CGAGCTGCACGCTG**C**CGTCCTCGATGTT	Validation of crRNA prediction
Potent EGFP crRNA8	**GG**TAGTGGTCGGCGAG**C**TGCACGCTGCCGT	Validation of crRNA prediction
Potent EGFP crRNA9	**GG**GCCGTCGCCGATGGGGGTGTTCTGCTGG	Validation of crRNA prediction

Potent EGFP crRNA10	**GG**GCAGCAGCACGGGG**C**CGTCGCCGATGGG	Validation of crRNA prediction
Potent EGFP crRNA11	**GG**GGGTGTTCTGCTGGTAGTGGTCGGCGAG	Validation of crRNA prediction
Ineffective Tagbfp crRNA1	CTCGGTGAAGGCCTCCCAGCCGAGTGTTTT	Validation of crRNA prediction
Ineffective Tagbfp crRNA2	CGGTCAGCACGCCCCCGTCCTCGTATGTGG	Validation of crRNA prediction
Ineffective Tagbfp crRNA3	CAGATGGCTCCCGCCCACGAGCTTCAGGGC	Validation of crRNA prediction
Ineffective Tagbfp crRNA4	CCATGTCGTTTCTGCCTTCCAGGCCGCCGT	Validation of crRNA prediction
Ineffective Tagbfp crRNA5	CGATCAGATGGCTCCCGCCCACGAGCTTCA	Validation of crRNA prediction
Ineffective Tagbfp crRNA6	CTCGTTGTTGGCCTCCTTGATTCTTTCCAG	Validation of crRNA prediction
Ineffective Tagbfp crRNA7	CAATTAAGCTTGTGCCCCAGTTTGCTAGGG	Validation of crRNA prediction
Ineffective Tagbfp crRNA8	CCCATGTGAAGCCCTCAGGGAAGGACTGCT	Validation of crRNA prediction
Ineffective Tagbfp crRNA9	ATGTGAAGTTCACCCCTCTGATCTTGACGT	Validation of crRNA prediction
Potent Tagbfp crRNA1	GGCGAAGGGGAGAGGGCCGCCCTCGACCAC	Validation of crRNA prediction
Potent Tagbfp crRNA2	GGTCACGAGGGTGGGCCAGGGCACGGGCAG	Validation of crRNA prediction
Potent Tagbfp crRNA3	GGTGACTCTCTCCCATGTGAAGCCCTCAGG	Validation of crRNA prediction
Potent Tagbfp crRNA4	GGGGTACAGGGTCTCGGTGAAGGCCTCCCA	Validation of crRNA prediction
Potent Tagbfp crRNA5	GGTTTCTTGGATCTATATGTGGTCTTGATG	Validation of crRNA prediction
Potent Tagbfp crRNA6	GGTCTGGGTGCCCTCGTAGGGCTTGCCTTC	Validation of crRNA prediction
Potent Tagbfp crRNA7	GGATGTCGAAGGCGAAGGGGAGAGGGCCGC	Validation of crRNA prediction
Potent Tagbfp crRNA8	GGTCTTGCTGCCGTAGAGGAAGCTAGTAGC	Validation of crRNA prediction
Potent Tagbfp crRNA9	GGGATGCCCTGGGTGTGGTTGATGAAGGTC	Validation of crRNA prediction
Potent Tagbfp crRNA10	GGGAGGTCGCAGTATCTGGCCACTGCCACC	Validation of crRNA prediction
RfxCas13d mCherry efficient 1	TCCTCGAAGTTCATCACGCGCTC	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 2	CGAAGTTCATCACGCGCTCCCAC	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 3	TCGAAGTTCATCACGCGCTCCCA	Predicted efficient RfxCas13d crRNA

RfxCas13d mCherry efficient 4	GAAGTTCATCACGCGCTCCCACT	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 5	CGTCCTCGAAGTTCATCACGCGC	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 6	CTCGAAGTTCATCACGCGCTCCC	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 7	CCTCGAAGTTCATCACGCGCTCC	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 8	GTCCTCGAAGTTCATCACGCGCT	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 9	CCGTCCTCGAAGTTCATCACGCG	Predicted efficient RfxCas13d crRNA
RfxCas13d mCherry efficient 10	GCCGTCCTCGAAGTTCATCACGC	Predicted efficient RfxCas13d crRNA
RfxCas13d crBCR-ABL1	GCCGCTGAAGGGCTTCTGCGTCT	Efficient RfxCas13d crRNA targeting BCR-ABL1 p190

Appendix Table 2. Target sequences used in this study

mcherry
atggtgagcatggacgaggaggataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagat cgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgt cccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgt gatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcc cctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaag cagaggctgaagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtc aacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgag ctgtacaagtaa
EGFP
atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagg gcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctac ggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttca aggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggac ggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagat ccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccacta cctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatgg acgagctgtacaagtaa
TagBFP sequence
Atgagcgagctgattaaggagaacatgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagc cctacgagggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccccttcgccttcgacatcctggctactagcttcctctacggcagcaaga ccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggagagagtcaccacatacgaggacgggggcgtgc tgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtcaagatcagaggggtgaacttcacatccaacggccctgtgatgcagaag aaaacactcggctgggaggccttcaccgagaccctgtaccccgctgacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagc catctgatcgcaaacatcaagaccacatatagatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatca aggaggccaacaacgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaattga
RUNX1-RUNX1T1 sequence (full length cloned into MSCV-IRES-EGFP)
atggcatacccatacgacgtgcctgactacgcctcccgtatccccgtagatgccagcacgagccgccgcttcacgccgccttccaccgcgctgagcccaggc aagatgagcgaggcgttgccgctgggcgccccggacgccggcgctgccctggccggcaagctgaggagcggcgaccgcagcatggtggaggtgctgg ccgaccacccgggcgagctggtgcgcaccgacagccccaacttcctctgctccgtgctgcctacgcactggcgctgcaacaagaccctgcccatcgctttca aggtggtggccctaggggatgttccagatggcactctggtcactgtgatggctggcaatgatgaaaactactcggctgagctgagaaatgctaccgcagccat gaagaaccaggttgcaagatttaatgacctcaggtttgtcggtcgaagtggaagagggaaaagcttcactctgaccatcactgtcttcacaaacccaccgcaag tcgccacctaccacagagccatcaaaatcacagtggatgggccccgagaacctcgaaatcgtactgagaagcactccacaatgccagactcacctgtggatg tgaagacgcaatctaggctgactcctccaacaatgccacctcccccaactactcaaggagctccaagaaccagttcatttacaccgacaacgttaactaatggc acgagccattctcctacagccttgaatggcgccccctcaccacccaatggcttcagcaatgggccttcctcttcttcctcctcctctctggctaatcaacagctgcc

cccagcctgtggtgccaggcaactcagcaagctgaaaaggttccttactaccctgcagcagtttggcaatgacatttcacccgagataggagaaagagttcgc
accctcgttctgggactagtgaactccactttgacaattgaagaatttcattccaaactgcaagaagctactaacttcccactgagacctttttgtcatcccatttttga
aggccaacttgcccctgctgcagcgtgagctcctccactgcgcaagactggccaaacagaaccctgcccagtacctcgcccagcatgaacagctgcttctgg
atgccagcaccacctcacctgttgactcctcagagctgcttctcgatgtgaacgaaaacgggaagaggcgaactccagacagaaccaaagaaaatggctttg
acagagagcctttgcactcagaacatccaagcaagcgaccatgcactattagcccaggccagcggtacagtccaaataacggcttatcctaccagcccaatg
gcctgcctcaccctaccccacctccacctcagcattaccgtttggatgatatggccattgcccaccactacagggactcctatcgacaccccagccacaggga
cctcagggacagaaacagacctatggggttgcatggcacacgtcaagaagaaatgattgatcacagactaacagacagagaatgggcagaagagtggaaa
catcttgaccatctgttaaactgcataatggacatggtagaaaaaacaaggcgatctctcaccgtactaaggcggtgtcaagaagcagaccgggaagaattga
attactggatccggcggtacagtgacgccgaggacttaaaaaaaggtggcggcagtagcagccactctaggcagcagagtcccgtcaacccagaccc
agttgcactagacgcgcatcgggaattccttcacaggcctgcgtctggatacgtgccagaggagatctggaagaaagctgaggaggccgtcaatgaggtga
agcgccaggcgatgacggagctgcagaaggccgtgtctgaggcggagcggaaagcccacgacatgatcacaacagagagggccaagatggagcgcac
ggtcgccgaggccaaacggcaggcggcggaggacgcactggcagttatcaatcagcaggaggattcaagcgagagttgctggaattgtggccgtaaagc
gagtgaaacctgcagtggctgtaacacagcccgatactgtggctcattttgccagcacaaagactgggagaagcaccatcacatctgtggacagaccctgca
ggcccagcagcagggagacacacctgcagtcagctcctctgtcacgcccaacagcggggctgggagcccgatggacacaccaccagcagccactccga
ggtcaaccaccccgggaacccccttccaccatagagacaacccctcgctag

NPM-ALK sequence (full length cloned into MSCV-IRES-mcherry)

atggaagattcgatggacatggacatgagccccctgaggccccagaactatcttttcggttgtgaactaaaggccgacaaagattatcactttaaggtggataat
gatgaaaatgagcaccagttatctttaagaacggtcagtttaggggctggtgcaaaggatgagttgcacattgttgaagcagaggcaatgaattacgaaggcag
tccaattaaagtaacactggcaactttgaaaatgtctgtacagccaacggtttcccttgggggctttgaaataacaccaccagtggtcttaaggttgaagtgtggtt
cagggccagtgcatattagtggacagcacttagtagtgtaccgccggaagcaccaggagctgcaagccatgcagatggagctgcagagccctgagtacaag
ctgagcaagctccgcacctcgaccatcatgaccgactacaaccccaactactgctttgctggcaagacctcctccatcagtgacctgaaggaggtgccgcgg
aaaaacatcaccctcattcggggtctgggccatggcgcctttggggaggtgtatgaaggccaggtgtccggaatgcccaacgacccaagccccctgcaagt
ggctgtgaagacgctgcctgaagtgtgctctgaacaggacgaactggatttcctcatggaagccctgatcatcagcaaattcaaccaccagaacattgttcgct
gcattggggtgagcctgcaatccctgccccggttcatcctgctggagctcatggcgggggggagacctcaagtccttcctccgagagacccgccctcgcccga
gccagccctcctccctggccatgctggaccttctgcacgtggctcgggacattgcctgtggctgtcagtatttggaggaaaaccacttcatccaccgagacattg
ctgccagaaactgcctcttgacctgtccaggccctggaagagtggccaagattggagacttcgggatggcccgagacatctacagggcgagctactatagaa
agggaggctgtgccatgctgccagttaagtggatgccccagaggccttcatggaaggaatattcacttctaaaacagacacatggtcctttggagtgctgctat
gggaaatcttttctcttggatatatgccataccccagcaaaagcaaccaggaagttctggagtttgtcaccagtggaggccggatggacccacccaagaactgc
cctgggcctgtataccggataatgactcagtgctggcaacatcagcctgaagacaggcccaactttgccatcattttggagaggattgaatactgcacccagga
cccggatgtaatcaacaccgctttgccgatagaatatggtccacttgtggaagaggaagagaaagtgcctgtgaggcccaaggaccctgaggggttcctcct
ctcctggtctctcaacaggcaaaacgggaggaggagcgcagcccagctgccccaccacctctgcctaccacctcctctggcaaggctgcaaagaaacccac
agctgcagaggtctctgttcgagtccctagagggccggccgtggaaggggacacgtgaatatggcattctctcagtccaaccctccttcggagttgcacaag
gtccacggatccagaaacaagcccaccagcttgtggaacccaacgtacggctcctggtttacagagaaacccaccaaaaagaataatcctatagcaaagaag
gagccacacgacaggggtaacctggggctggagggaagctgtactgtcccacctaacgttgcaactgggagacttccgggggcctcactgctcctagagcc
ctcttcgctgactgccaatatgaaggaggtacctctgttcaggctacgtcacttcccttgtgggaatgtcaattacggctaccagcaacagggcttgcccttagaa
gccgctactgcccctggagctggtcattacgaggataccattctgaaaagcaagaatagcatgaaccagcctgggccctga

SNX2-ABL1 sequence (full length cloned into MSCV-IRES-EGFP)

cgtacgcgtgctcacgtgacgggtccgcgaggcccagctcgcgcagtcgttcgggtgagcgaagatggcggccgagagggaacctcctccgctgggggga
cgggaagcccaccgactttgaggatctggaggacggagaggacctgttcaccagcactgtctccaccctagagtcaagtccatcatctccagaaccagctag
tcttcctgcagaagatattagtgcaaactccaatggcccaaaacccacagaagttgtattagatgatgacagagaagatctttttgcagaagccacagaagaagt
ttctttggacagccctgaaagggaacctatcctatcctcggaaccttctcctgcagtcacacctgtcactcctactacactcattgctcctagaattgaatcaaaga
gtatgtctgctcccgtgatctttgatagatccagggaagagctctacgtctcctccgagagccgcttcaacaccctggccgagttggttcatcatcattcaacggt
ggccgacgggctcatcaccacgctccattatccagccccaaagcgcaacaagcccactgtctatggtgtgtcccccaactacgacaagtgggagatggaac
gcacggacatcaccatgaagcacaagctgggcgggggccagtacggggaggtgtacgagggcgtgtggaagaaatacagcctgacggtggccgtgaag
accttgaaggaggacaccatggaggtggaagagttcttgaaagaagctgcagtcatgaaagagatcaaacaccctaacctggtgcagctccttggggtctgc
acccgggagcccccgttctatatcatcactgagttcatgacctacgggaacctcctggactacctgagggagtgcaaccggcaggaggtgaacgccgtggtg
ctgctgtacatggccactcagatctcgtcagccatggagtacctggagaagaaaaacttcatccacagagatcttgctgcccgaaactgcctggtaggggaga
accacttggtgaaggtagctgattttggcctgagcaggttgatgacaggggacacctacacagcccatgctggagccaagttccccatcaaatggactgcacc
cgagagcctggcctacaacaagttctccatcaagtccgacgtctgggcatttggagtattgctttgggaaattgctacctatggcatgtcccttacccgggaatt
gacctgtcccaggtgtatgagctgctagagaaggactaccgcatggagcgcccagaaggctgcccagagaaggtctatgaactcatgcgagcatgttggca
gtggaatccctctgaccggccctcctttgctgaaatccaccaagcctttgaaacaatgttccaggaatccagtatctcagacgaagtggaaaaggagctgggga
aacaaggcgtccgtggggctgtgagtaccttgctgcaggccccagagctgcccaccaagacgaggacctccaggagagctgcagagcacagagacacca
ctgacgtgcctgagatgcctcactccaagggccagggagagagcgatcctctggaccatgagcctgccgtgtctccattgctccctcgaaaagagcgaggtc
ccccggagggcggcctgaatgaagatgagcgccttctccccaaagacaaaaagaccaacttgttcagcgccttgatcaagaagaagaagaagacagcccc
aaccctcccaaacgcagcagctccttccgggagatggacggccagccggagcgcagaggggccggcgaggaagagggccgagacatcagcaacgg
ggcactggctttcacccccttggacacagctgacccagccaagtccccaaagcccagcaatggggctggggtccccaatggagccctccgggagtccggg
ggctcaggcttccggtctccccacctgtggaagaagtccagcacgctgaccagcagccgcctagccaccggcgaggaggagggcggtggcagctccagc
aagcgcttcctgcgctcttgctccgcctcctgcgttccccatggggccaaggacacggagtggaggtcagtcacgctgcctcgggacttgcagtccacgggga
agacagtttgactcgtccacatttggagggcacaaaagtgagaagccggctctgcctcggaagagggcaggggagaacaggtctgaccaggtgacccgag
gcacagtaacgcctccccccaggctggtgaaaaagaatgaggaagctgctgatgaggtcttcaaagacatcatggagtccagcccgggctccagcccgccc

aacctgactccaaaaccctccggcggcaggtcaccgtggcccctgcctcgggcctccccacaaggaagaagctggaaagggcagtgccttagggaccc
ctgctgcagctgagccagtgacccccaccagcaaagcaggctcaggtgcaccaggggcaccagcaagggccccgccgaggagtccagagtgaggag
gcacaagcactcctctgagtcgccagggagggacaaggggaaattgtccaggctcaaacctgccccgccgcccccaccagcagcctctgcagggaaggc
tggaggaaagccctcgcagagcccgagccaggaggcggccggggaggcagtcctgggcgcaaagacaaaagccacgagtctggttgatgctgtgaaca
gtgacgctgccaagcccagccagccgggagagggcctcaaaaagcccgtgctcccggccactccaaagccacagtccgccaagccgtcggggacccccc
atcagcccagcccccgttccctccacgttgccatcagcatcctcggccctggcaggggaccagccgtcttccaccgccttcatccctctcatatcaacccgagt
gtctcttcggaaaacccgccagcctccagagcggatcgccagcggcgccatcaccaagggcgtggtcctggacagcaccgaggcgctgtgcctcgccatc
tctaggaactccgagcagatggccagccacagcgcagtgctggaggccggcaaaaacctctacacgttctgcgtgagctatgtggattccatccagcaaatg
aggaacaagtttgccttccgagaggccatcaacaaactggagaataatctccgggagcttcagatctgcccggcgacagcaggcagtggtccagcggccac
tcaggacttcagcaagctcctcagttcggtgaaggaaatcagtgacatagtgcagaggtagctcgagctact

P190 BCR-ABL1 sequence (full length cloned into MSCV-IRES-EGFP)

atggtggacccggtgggcttcgcggaggcgtggaaggcgcagttcccggactcagagcccccgcgcatggagctgcgctcagtgggcgacatcgagcag
gagctggagcgctgcaaggcctccattcggcgcctggagcaggaggtgaaccaggagcgcttccgcatgatctacctgcagacgttgctggccaaggaaa
agaagagctatgaccggcagcgatggggcttccggcgcgcggcgcaggcccccgacggcgcctccgagccccgagcgtccgcgtcgcgcccgcagcc
agcgcccgccgacggagccgacccgccgcccgccgaggagcccgaggcccggcccgacggcgagggttctccgggtaaggccaggcccgggaccg
cccgcaggcccggggcagccgcgtcgggggaacgggacgaccggggaccccccgccagcgtggcggcgctcaggtccaacttcgagcggatccgca
agggccatggccagcccggggcggacgccgagaagcccttctacgtgaacgtcgagtttcaccacgagcgcggcctggtgaaggtcaacgacaaagagg
tgtcggaccgcatcagctccctgggcagccaggccatgcagatggagcgcaaaaagtcccagcacggcgcgggctcgagcgtgggggatgcatccaggc
ccccttaccggggacgctcctcggagagcagctgcggcgtcgacggcgactacgaggacgccgagttgaaccccgcttcctgaaggacaacctgatcga
cgccaatggcggtagcaggcccccttggccgcccctggagtaccagccctaccagagcatctacgtcgggggcatgatggaaggggagggcaagggccc
gctcctgcgcagccagagcacctctgagcaggagaagcgccttacctggccccgcaggtcctactccccccggagttttgaggattgcggaggcggctata
ccccggactgcagctccaatgagaacctcacctccagcgaggaggacttctcctctggccagtccagccgcgtgtccccaagccccaccacctaccgcatgt
tccgggacaaaagccgctctccctcgcagaactcgcaacagtccttcgacagcagcagtccccccacgccgcagtgccataagcggcaccggcactgccc
ggttgtcgtgtccgaggccaccatcgtgggcgtccgcaagaccgggcagatctggcccaacgatggcgagggcgccttccatggagacgcagaagccctt
cagcggccagtagcatctgactttgagcctcagggtctgagtgaagccgctcgttggaactccaaggaaaaccttctcgctggacccagtgaaaatgacccca
acctttttcgttgcactgtatgattttgtggccagtggagataacactctaagcataactaaaggtgaaaagctccgggtcttaggctataatcacaatggggaatg
gtgtgaagcccaaaccaaaaatggccaaggctgggtcccaagcaactacatcacgccagtcaacagtctggagaaacactcctggtaccatgggcctgtgtc
ccgcaatgccgctgagtatctgctgagcagcgggatcaatggcagcttcttggtgcgtgagagtgagagcagtcctggccagaggtccatctcgctgagatac
gaagggagggtgtaccattacaggatcaacactgcttctgatggcaagctctacgtctcctccgagagccgcttcaacaccctggccgagttggttcatcatcat
tcaacggtggccgacgggctcatcaccacgctccattatccagccccaaagcgcaacaagcccactgtctatggtgtgtcccccaactacgacaagtgggag
atggaacgcacggacatcaccatgaagcacaagctgggcgggggccagtacggggaggtgtacgagggcgtgtggaagaaatacagcctgacggtggc
cgtgaagaccttgaaggaggacaccatggaggtggaagagttcttgaaagaagctgcagtcatgaaagagatcaaacaccctaacctggtgcagctccttgg
ggtctgcacccgggagcccccgttctatatcatcactgagttcatgacctacgggaacctcctggactacctgagggagtgcaaccggcaggaggtgaacgc
cgtggtgctgctgtacatggccactcagatctcgtcagccatggagtacctggagaagaaaaacttcatccacagagatcttgctgcccgaaactgcctggtag
gggagaaccacttggtgaaggtagctgattttggcctgagcaggttgatgacaggggacacctacacagcccatgctggagccaagttccccatcaaatgga
ctgcacccgagagcctggcctacaacaagttctccatcaagtccgacgtctgggcatttggagtattgctttgggaaattgctacctatggcatgtccccttaccc
gggaattgacctgtcccaggtgtatgagctgctagagaaggactaccgcatggagcgcccagaaggctgcccagagaaggtctatgaactcatgcgagcat
gttggcagtggaatccctctgaccggccctcctttgctgaaatccaccaagcctttgaaacaatgttccaggaatccagtatctcagacgaagtggaaaaggag
ctggggaaacaaggcgtccgtggggctgtgagtaccttgctgcaggccccagagctgcccaccaagacgaggacctccaggagagctgcagagcacaga
gacaccactgacgtgcctgagatgcctcactccaagggccagggagagagcgatcctctggaccatgagcctgccgtgtctccattgctccctcgaaaagag
cgaggtcccccggagggcggcctgaatgaagatgagcgccttctccccaaagacaaaaagaccaacttgttcagcgccttgatcaagaagaagaagaaga
cagccccaaccccctcccaaacgcagcagctccttccggggagatggacggccagccggagcgcagaggggccggcgaggaagagggccgagacatcag
caacggggcactggctttcacccccttggacacagctgacccagccaagtccccaaagcccagcaatggggctggggtccccaatggagccctccgggag
tccgggggctcaggcttccggtctccccacctgtggaagaagtccagcacgctgaccagcagccgcctagccaccggcgaggaggagggcggtggcagc
tccagcaagcgcttcctgcgctcttgctccgcctcctgcgttccccatggggccaaggacacggagtggaggtcagtcacgctgcctcgggacttgcagtcca
cgggaagacagtttgactcgtccacatttggagggcacaaaagtgagaagccggctctgcctcggaagagggcaggggagaacaggtctgaccaggtgac
ccgaggcacagtaacgcctcccccccaggctggtgaaaaagaatgaggaagctgctgatgaggtcttcaaagacatcatggagtccagcccgggctccagcc
cgcccaacctgactccaaaaccctccggcggcaggtcaccgtggcccctgcctcgggcctcccccacaaggaagaagctggaaagggcagtgccttagg
gacccctgctgcagctgagccagtgacccccaccagcaaagcaggctcaggtgcaccaggggcaccagcaagggccccgccgaggagtccagagtga
ggaggcacaagcactcctctgagtcgccagggagggacaaggggaaattgtccaggctcaaacctgccccgccgcccccaccagcagcctctgcaggga
aggctggaggaaagccctcgcagagcccgagccaggaggcggccggggaggcagtcctgggcgcaaagacaaaagccacgagtctggttgatgctgtg
aacagtgacgctgccaagcccagccagccgggagagggcctcaaaaagcccgtgctcccggccactccaaagccacagtccgccaagccgtcggggac
ccccatcagcccagcccccgttccctccacgttgccatcagcatcctcggccctggcaggggaccagccgtcttccaccgccttcatccctctcatatcaaccc
gagtgtctcttcggaaaacccgccagcctccagagcggatcgccagcggcgccatcaccaagggcgtggtcctggacagcaccgaggcgctgtgcctcgc
catctctaggaactccgagcagatggccagccacagcgcagtgctggaggccggcaaaaacctctactcgttctgcgtgagctatgtggattccatccagcaa
atgaggaacaagtttgccttccgagaggccatcaacaaactggagaataatctccgggagcttcagatctgcccggcgacagcaggcagtggtccagcggc
cactcaggacttcagcaagctcctcagttcggtgaaggaaatcagtgacatagtgcagaggtag

P210 BCR-ABL1 sequence (full length cloned into MSCV-IRES-EGFP)

atggtggacccggtgggcttcgcggaggcgtggaaggcgcagttcccggactcagagcccccgcgcatggagctgcgctcagtgggcgacatcgagcag
gagctggagcgctgcaaggcctccattcggcgcctggagcaggaggtgaaccaggagcgcttccgcatgatctacctgcagacgttgctggccaaggaaa

agaagagctatgaccggcagcgatggggcttccggcgcgcggcgcaggcccccgacggcgcctccgagccccgagcgtccgcgtcgcgcccgcagcc
agcgcccgccgacggagccgacccgccgcccgccgaggagcccgaggcccggcccgacggcgagggttctccgggtaaggccaggcccgggaccg
cccgcaggcccggggcagccgcgtcgggggaacgggacgaccggggaccccccgccagcgtggcggcgctcaggtccaacttcgagcggatccgca
agggccatggccagcccggggcggacgccgagaagcccttctacgtgaacgtcgagtttcaccacgagcgcggcctggtgaaggtcaacgacaaagagg
tgtcggaccgcatcagctccctgggcagccaggccatgcagatggagcgcaaaaagtcccagcacggcgcgggctcgagcgtgggggatgcatccaggc
ccccttaccggggacgctcctcggagagcagctgcggcgtcgacggcgactacgaggacgccgagttgaaccccgcttcctgaaggacaacctgatcga
cgccaatggcggtagcaggcccccttggccgcccctggagtaccagccctaccagagcatctacgtcgggggcatgatggaaggggagggcaagggccc
gctcctgcgcagccagagcacctctgagcaggagaagcgccttacctggccccgcaggtcctactccccccggagttttgaggattgcggaggcggctata
ccccggactgcagctccaatgagaacctcacctccagcgaggaggacttctcctctggccagtccagccgcgtgtccccaagccccaccacctaccgcatgt
tccgggacaaaagccgctctccctcgcagaactcgcaacagtccttcgacagcagcagtccccccacgccgcagtgccataagcggcaccggcactgccc
ggttgtcgtgtccgaggccaccatcgtgggcgtccgcaagaccgggcagatctggcccaacgatggcgagggcgccttccatggagacgcagatggctcg
ttcggaacaccacctggatacggctgcgctgcagaccgggcagaggagcagcgccggcaccaagatgggctgccctacattgatgactcgccctcctcatc
gccccacctcagcagcaagggcaggggcagccgggatgcgctggtctcgggagccctggagtccactaaagcgagtgagctggacttggaaaagggctt
ggagatgagaaaatgggtcctgtcgggaatcctggctagcgaggagacttacctgagccacctggaggcactgctgctgcccatgaagcctttgaaagccgc
tgccaccacctctcagccggtgctgacgagtcagcagatcgagaccatcttcttcaaagtgcctgagctctacgagatccacaaggagttctatgatgggctctt
cccccgcgtgcagcagtggagccaccagcagcgggtgggcgacctcttccagaagctggccagccagctgggtgtgtaccgggccttcgtggacaactac
ggagttgccatggaaatggctgagaagtgctgtcaggccaatgctcagtttgcagaaatctccgagaacctgagagccagaagcaacaaagatgccaaggat
ccaacgaccaagaactctctggaaactctgctctacaagcctgtggaccgtgtgacgaggagcacgctggtcctccatgacttgctgaagcacactcctgcca
gccaccctgaccacccccttgctgcaggacgccctccgcatctcacagaacttcctgtccagcatcaatgaggagatcacaccccgacggcagtccatgacgg
tgaagaagggagagcaccggcagctgctgaaggacagcttcatggtggagctggtggaggggcccgcaagctgcgccacgtcttcctgttcaccgacct
gcttctctgcaccaagctcaagaagcagagcggaggcaaaacgcagcagtatgactgcaaatggtacattccgctcacggatctcagcttccagatggtggat
gaactggaggcagtgcccaacatccccctggtgcccgatgaggagctggacgctttgaagatcaagatctcccagatcaagagtgacatccagagagagaa
gagggcgaacaagggcagcaaggctacggagaggctgaagaagaagctgtcggagcaggagtcactgctgctgcttatgtctcccagcatggccttcagg
gtgcacagccgcaacggcaagagttacacgttcctgatctcctctgactatgagcgtgcagagtggagggagaacatccgggagcagcagaagaagtgtttc
agaagcttctccctgacatccgtggagctgcagatgctgaccaactcgtgtgtgaaactccagactgtccacagcattccgctgaccatcaataaggaagatga
tgagtctccggggctctatgggtttctgaatgtcatcgtccactcagccactggatttaagcagagttcaaaagcccttcagcggccagtagcatctgactttgag
cctcagggtctgagtgaagccgctcgttggaactccaaggaaaaccttctcgctggacccagtgaaaatgaccccaacctttttcgttgcactgtatgattttgtgg
ccagtggagataacactctaagcataactaaaggtgaaaagctccgggtcttaggctataatcacaatggggaatggtgtgaagcccaaaccaaaaatggcca
aggctgggtcccaagcaactacatcacgccagtcaacagtctggagaaacactcctggtaccatgggcctgtgtcccgcaatgccgctgagtatctgctgagc
agcgggatcaatggcagcttcttggtgcgtgagagtgagagcagtcctggccagaggtccatctcgctgagatacgaagggagggtgtaccattacaggatc
aacactgcttctgatggcaagctctacgtctcctccgagagccgcttcaacaccctggccgagttggttcatcatcattcaacggtggccgacgggctcatcacc
acgctccattatccagccccaaagcgcaacaagcccactgtctatggtgtgtccccaactacgacaagtgggagatggaacgcacggacatcaccatgaag
cacaagctgggcgggggccagtacggggaggtgtacgagggcgtgtggaagaaatacagcctgacggtggccgtgaagaccttgaaggaggacaccat
ggaggtggaagagttcttgaaagaagctgcagtcatgaaagagatcaaacaccctaacctggtgcagctccttggggtctgcacccgggagcccccgttctat
atcatcactgagttcatgacctacgggaacctcctggactacctgagggagtgcaaccggcaggaggtgaacgccgtggtgctgctgtacatggccactcag
atctcgtcagccatggagtacctggagaagaaaaacttcatccacagagatcttgctgcccgaaactgcctggtaggggagaaccacttggtgaaggtagctg
attttggcctgagcaggttgatgacaggggacacctacacagcccatgctggagccaagttccccatcaaatggactgcacccgagagcctggcctacaaca
agttctccatcaagtccgacgtctgggcatttggagtattgctttgggaaattgctacctatggcatgtcccctttacccgggaattgacctgtcccaggtgtatgag
ctgctagagaaggactaccgcatggagcgcccagaaggctgcccagagaaggtctatgaactcatgcgagcatgttggcagtggaatccctctgaccggcc
ctcctttgctgaaatccaccaagcctttgaaacaatgttccaggaatccagtatctcagacgaagtggaaaaggagctggggaaacaaggcgtccgtggggct
gtgagtaccttgctgcaggccccagagctgcccaccaagacgaggacctccaggagagctgcagagcacagagacaccactgacgtgcctgagatgcctc
actccaagggccagggagagagcgatcctctggaccatgagcctgccgtgtctccattgctccctcgaaaagagcgaggtcccccggagggcggcctgaat
gaagatgagcgccttctccccaaagacaaaaagaccaacttgttcagcgccttgatcaagaagaagaagaagacagccccaaccctcccaaacgcagcag
ctccttccgggagatggacggccagccggagcgcagaggggccggcgaggaagagggccgagacatcagcaacggggcactggctttcacccccttgg
acacagctgacccagccaagtccccaaagcccagcaatggggctggggtccccaatggagccctccgggagtccgggggctcaggcttccggtctcccca
cctgtggaagaagtccagcacgctgaccagcagccgcctagccaccggcgaggaggagggcggtggcagctccagcaagcgcttcctgcgctcttgctcc
gcctcctgcgttccccatggggccaaggacacggagtggaggtcagtcacgctgcctcgggacttgcagtccacgggaagacagtttgactcgtccacattt
ggagggcacaaaagtgagaagccggctctgcctcggaagagggcaggggagaacaggtctgaccaggtgacccgaggcacagtaacgcctccccccag
gctggtgaaaaagaatgaggaagctgctgatgaggtcttcaaagacatcatggagtccagcccgggctccagcccgcccaacctgactccaaaaccccctcc
ggcggcaggtcaccgtggcccctgcctcgggcctcccccacaaggaagaagctggaaagggcagtgccttagggaccccctgctgcagctgagccagtga
cccccaccagcaaagcaggctcaggtgcaccaggggcaccagcaagggccccgccgaggagtccagagtgaggaggcacaagcactcctctgagtcg
ccagggagggacaaggggaaattgtccaggctcaaacctgccccgccgcccccaccagcagcctctgcagggaaggctggaggaaagccctcgcagag
cccgagccaggaggcggccggggaggcagtcctgggcgcaaagacaaaagccacgagtctggttgatgctgtgaacagtgacgctgccaagcccagcc
agccgggagagggcctcaaaaagcccgtgctcccggccactccaaagccacagtccgccaagccgtcggggacccccatcagcccagccccgttccct
ccacgttgccatcagcatcctcggccctggcaggggaccagccgtcttccaccgccttcatccctctcatatcaacccgagtgtctcttcggaaaacccgccag
cctccagagcggatcgccagcggcgccatcaccaagggcgtggtcctggacagcaccgaggcgctgtgcctcgccatctctaggaactccgagcagatgg
ccagccacagcgcagtgctggaggccggcaaaaacctctactcgttctgcgtgagctatgtggattccatccagcaaatgaggaacaagtttgccttccgaga
ggccatcaacaaactggagaataatctccgggagcttcagatctgcccggcgacagcaggcagtggtccagcggccactcaggacttcagcaagctcctca
gttcggtgaaggaaatcagtgacatagtgcagaggtag

SFPQ-ABL1 (full length cloned into MSCV-IRES-EGFP)

acgcgtagcgtcttcttcgcttttgcctcgaccgccccttgaccacagacatgtctcgggatcggttccggagtcgtggcggtggcggtggtggcttccacagg
cgtggaggaggcggcggccgcggcggcctccacgacttccgttctccgccgcccggcatgggcctcaatcagaatcgcggccccatgggtcctggcccg
ggccagagcggccctaagcctccgatcccgccaccgcctccacaccaacagcagcaacagccaccaccgcagcagccaccgccgcagcagccgccac
cgcatcagccgccgccgcatccacagccgcatcagcagcagcagccgccgccaccgccgcaggactcttccaagcccgtcgttgctcagggacccggcc
ccgctcccggagtaggcagcgcaccaccagcctccagctcggccccgcccgccactccaccaacctcgggggccccgccagggtccgggccaggcccg
actccgaccccgccgcctgcagtcacctcggcccctcccggggcgccgccacccaccccgccaagcagcggggtccctaccacacctcctcaggccgga
ggcccgccgcctccgcccgcggcagtcccgggcccgggtccagggcctaagcagggcccaggtccgggtggtcccaaaggcggcaaaatgcctggcg
ggccgaagccaggtggcggcccgggcctaagtacgcctggcggccaccccaagccgccgcatcgaggcggcggggagccccgcgggggccgccag
caccacccgccctaccaccagcagcatcaccaggggcccccgcccggcgggcccggcggccgcagcgaggagaagatctcggactcggaggggtttaa
agccaatttgtctctcttgaggaggcctggagagaaaacttacacacagcgatgtcggttgtttgttgggaatctacctgctgatatcacggaggatgaattcaaa
agactatttgctaaatatggagaaccaggagaagttttttatcaacaaaggcaaaggattcggatttattaagcttgaatctagagctttggctgaaattgccaaagc
cgaactggatgatacacccatgagaggtagacagcttcgagttcgctttgccacacatgctgctgccctttctgttcgtaatctttcaccttatgtttccaatgaactg
ttggaagaagcctttagccaatttggtcctattgaaagggctgttgtaatagtggatgatcgtggaagatctacagggaaaggcattgttgaatttgcttctaagcc
agcagcaagaaaggcatttgaacgatgcagtgaaggtgttttcttactgacgacaactcctcgtccagtcattgtggaaccacttgaacaactagatgatgaaga
tggtcttcctgaaaaacttgcccagaagaatccaatgtatcaaaaggagagagaaacccctcctcgttttgcccagcatggcacgtttgagtacgaatattctcag
cgatggaagtctttggatgaaatggaaaaacagcaaagggaacaagttgaaaaaaacatgaaagatgcaaaagacaaattggaaagtgaaatggaagatgc
ctatcatgaacatcaggcaaatcttttgcgccaagatctgatgagacgacaggaagaattaagacgcatggaagaacttcacaatcaagaaatgcagaaacgt
aaagaaatgcaattgaggcaagaggaggaacgacgtagaagagaggaagagatgatgattcgtcaacgtgagatggaagaacaaatgaggcgccaaaga
gaggaaagttacagccgaatgggctacatggatccacgggaaagagacatgcgaatgggtggcggaggagcaatgaacatgggagatccctatggttcag
gaggccagaaatttccacctctaggaggtggtggtggcataggttatgaagctaatcctggcgttccaccagcaaccatgagtggttccatgatgggaagtgac
atgctctacgtctcctccgagagccgcttcaacaccctggccgagttggttcatcatcattcaacggtggccgacgggctcatcaccacgctccattatccagcc
ccaaagcgcaacaagcccactgtctatggtgtgtcccccaactacgacaagtgggagatggaacgcacggacatcaccatgaagcacaagctgggcgggg
gccagtacggggaggtgtacgagggcgtgtggaagaaatacagcctgacggtggccgtgaagaccttgaaggaggacaccatggaggtggaagagttctt
gaaagaagctgcagtcatgaaagagatcaaacaccctaacctagtgcagctccttggggtctgcacccgggagcccccgttctatatcatcactgagttcatga
cctacgggaacctcctggactacctgagggagtgcaaccggcaggaggtgaacgccgtggtgctgctgtacatggccactcagatctcgtcagccatggagt
acctagagaagaaaaacttcatccacagagatcttgctgcccgaaactgcctggtaggggagaaccacttggtgaaggtagctgattttggcctgagcaggtt
gatgacaggggacacctacacagcccatgctggagccaagttccccatcaaatggactgcacccgagagcctggcctacaacaagttctccatcaagtccga
cgtctgggcatttggagtattgctttgggaaattgctacctatggcatgtcccctttacccgggaattgaccgttcccaggtgtatgagctgctagagaaggactac
cgcatgaagcgcccagaaggctgcccagagaaggtctatgaactcatgcgagcatgttggcagtggaatccctctgaccggccctcctttgctgaaatccacc
aagcctttgaaacaatgttccaggaatccagtatctcagacgaagtggaaaaggagctggggaaacaaggcgtccgtggggctgtgactaccttgctgcagg
ccccagagctgcccaccaagacgaggacctccaggagagctgcagagcacagagacaccactgacgtgcctgagatgcctcactccaagggccaggga
gagagcgatcctctggaccatgagcctgccgtgtctccattgctccctcgaaaagagcgaggtcccccggagggcggcctgaatgaagatgagcgccttctc
cccaaagacaaaaagaccaacttgttcagcgccttgatcaagaagaagaagaagacagccccaaccctcccaaacgcagcagctccttccgggagatgg
acggccagccggagcgcagaggggccggcgaggaagagggccgagacatcagcaacggggcactggctttcacccccttggacacagctgacccagc
caagtccccaaagcccagcaatggggctggggtccccaatggagccctccgggagtccgggggctcaggcttccggtctccccacctgtggaagaagtcc
agcacgctgaccagcagccgcctagccaccggcgaggaggagggcggtggcagctccagcaagcgcttcctgcgctcttgctccgtctcctgcgttcccca
tggggccaaggacacggagtggaggtcagtcacgctgcctcgggacttgcagtccacgggaagacagtttgactcgtccacatttggagggcacaaaagtg
agaagccggctctgcctcggaagagggcaggggagaacaggtctgaccaggtgacccgaggcacagtaacgcctccccccaggctggtgaaaaagaat
gaggaagctgctgatgaggtcttcaaagacatcatggagtccagcccgggctccagcccgcccaacctgactccaaaaccccctccggcggcaggtcaccgt
ggcccctgcctcgggcctccccacaaggaagaagcctggaaaggcagtgccttagggacccctgctgcagctgagccagtgaccccccaccagcaaagc
aggctcaggtgcaccaaggggcaccagcaagggccccgccgaggagtccagagtgaggaggcacaagcactcctctgagtcgccagggagggacaag
gggaaattgtccaagctcaaacctgccccgccgcccccaccagcagcctctgcagggaaggctggaggaaagccctcgcagaggcccggccaggaggc
tgccggggaggcagtcttgggcgcaaagacaaaagccacgagtctggttgatgctgtgaacagtgacgctgccaagcccagccagccggcagagggcct
caaaaagcccgtgctcccggccactccaaagccacaccccgccaagccgtcggggacccccatcagcccagccccgttcccctttccacgttgccatcag
catcctcggccttggcaggggaccagccgtcttccactgccttcatccctctcatatcaacccgagtgtctcttcggaaaacccgccagcctccagagcgggcc
agcggcgccatcaccaagggcgtggtcttggacagcaccgaggcgctgtgcctcgccatctctgggaactccgagcagatggccagccacagcgcagtgc
tggaggccggcaaaaacctctacacgttctgcgtgagctatgtggattccatccagcaaatgaggaacaagtttgccttccgagaggccatcaacaaactgga
gaataatctccgggagcttcagatctgcccggcgtcagcaggcagtggtccggcggccactcaggacttcagcaagctcctcagttcggtgaaggaaatcag
tgacatagtgcagaggtagctcgag

ABL1 (partial)-3XHA tag cloned into MSCV-IRES-EGFP

atggggcagcagcctggaaaagtacttggggaccaaagaaggccaagcttgcctgccctgcatttatcaaaggagcagggaagaaggaatcatcgaggca
tgggggtccacactgcaatgtttttgtggaacatgaagcccttcagcggccagtagcatctgactttgagcctcagggtctgagtgaagccgctcgttggaactc
caaggaaaaccttctcgctggacccagtgaaaatgaccccaacctttcgttgcactgtatgattttgtggccagtggagataacactctaagcataactaaaggt
gaaaagctccgggtcttaggctataatcacaatggggaatggtgtgaagcccaaaccaaaaatggccaaggtagttcctacccatacgatgttccagattacgc
ttatccctacgacgtgcctgattatgcatacccatatgatgtccccgactatgcctaa

BCR (partial)-3XHA tag cloned into MSCV-IRES-tagBFP

atgttccgggacaaaagccgctctccctcgcagaactcgcaacagtccttcgacagcagcagtccccccacgccgcagtgccataagcggcaccggcactg
cccggttgtcgtgtccgaggccaccatcgtgggcgtccgcaagaccgggcagatctggcccaacgatggcgagggcgccttccatggagacgcagatggc
tcgttcggaacaccacctggatacggctgcgctgcagaccgggcagaggagcagcgccggcaccaagatgggctgccctacattgatgactcgccctcctc

atcgccccacctcagcagcaagggcaggggcagccgggatgcgctggtctcgggagccctggagtccactaaagcgagtggtagttcctacccatacgat gttccagattacgcttatccctacgacgtgcctgattatgcatacccatatgatgtccccgactatgcctaa
BCR-ABL1 (partial)-3XHA tag cloned into MSCV-IRES-m-cherry
Atgttccgggacaaaagccgctctccctcgcagaactcgcaacagtccttcgacagcagcagtccccccacgccgcagtgccataagcggcaccggcact gcccggttgtcgtgtccgaggccaccatcgtgggcgtccgcaagaccgggcagatctggcccaacgatggcgagggcgccttccatggagacgcagaag cccttcagcggccagtagcatctgactttgagcctcagggtctgagtgaagccgctcgttggaactccaaggaaaaccttctcgctggacccagtgaaaatga ccccaacctttttcgttgcactgtatgattttgtggccagtggagataacactctaagcataactaaaggtgaaaagctcggtagttcctacccatacgatgttccag attacgcttatccctacgacgtgcctgattatgcatacccatatgatgtccccgactatgcctaa

Appendix Table 3. Primer sequences used in this study

Primer name	Sequence	Description
IRES forward	TGGCTCTCCTCAAGCGTATT	Sanger sequence primer for BCR-ABL1/BCR/ABL1 fragment cloning
M13 reverse	CAGGAAACAGCTATGAC	Sanger sequence primer for crRNA cloning
BCR PCR forward	CAGATCTGGCCCAACGAT	PCR for BCR-ABL1 P210 cDNA
ABL1 PCR reverse	CAGATACTCAGCGGCATTGCGG	PCR for BCR-ABL1 P210 cDNA
BCR-ABL1 P210 RT-PCR forward	GGGCTCTATGGGTTTCTGAATG	RT-PCR for BCR-ABL1 P210 mRNA level
BCR-ABL1 P210 RT-PCR reverse	GAGGCTCAAAGTCAGATGCTACTG	RT-PCR for BCR-ABL1 P210 mRNA level
BCR RT-PCR forward	AGTAAAGCTCTCGGTCAAGTTC	RT-PCR for wild-type BCR mRNA level
BCR RT-PCR reverse	TTGGACCTCTCTCTCTTGGT	RT-PCR for wild-type BCR mRNA level
ABL1 RT-PCR forward	ATGTTGGAGATCTGCCTGAAG	RT-PCR for wild-type ABL1 mRNA level
ABL1 RT-PCR reverse	AAGGGCTTCTTCCAGATAACAG	RT-PCR for wild-type ABL1 mRNA level
BCR-ABL1 p190 RT-PCR forward	CAGATCTGGCCCAACGAT	RT-PCR for BCR-ABL1 P190 mRNA level
BCR-ABL1 p190 RT-PCR reverse	CCTGAGGCTCAAAGTCAGAT	RT-PCR for BCR-ABL1 P190 mRNA level
SFPQ-ABL1 RT-PCR forward	TGGTTCCATGATGGGAAGTG	RT-PCR for SFPQ-ABL1 mRNA level
SFPQ-ABL1 RT-PCR reverse	ATAATGGAGCGTGGTGATGAG	RT-PCR for SFPQ-ABL1 mRNA level
SNX2-ABL1 RT-PCR forward	CTGCTCCCGTGATCTTTGATAG	RT-PCR for SNX2-ABL1 mRNA level

SNX2-ABL1 RT-PCR reverse	GATAATGGAGCGTGGTGATGAG	RT-PCR for SNX2-ABL1 mRNA level
crRNA DR RT-PCR reverse	GTTGTAATAGCCCCTCAAAAC	RT-PCR for crRNA level
crRNA mCherry tiled 39/40 RT-PCR forward	GTACATCCGCTCGGAGGA	RT-PCR for crRNA level
crRNA mCherry RT-PCR tiled 26/27 forward	CCGTCCTCGGGGTACATC	RT-PCR for crRNA level
HSP90A1B RT-PCR forward	AGAAATTGCCCAACTCATGTCC	RT-PCR for HSP90A1B mRNA level (housekeeping)
HSP90A1B RT-PCR reverse	ATCAACTCCCGAAGGAAAATCTC	RT-PCR for HSP90A1B mRNA level (housekeeping)
GAPDH RT-PCR forward	GGAGCGAGATCCCTCCAAAAT	RT-PCR for GAPDH mRNA level (housekeeping)
GAPDH RT-PCR reverse	GGCTGTTGTCATACTTCTCATGG	RT-PCR for GAPDH mRNA level (housekeeping)

Appendix Table 4. Transfection conditions of HEK 293T cells

Component (per well)	**96-well**	**24-well**	**12-well**
# of seeded HEK 293T cells	30,000	150,000	300,000
Plasmids DNA amount	100ng	500ng	1000 ng
P3000 reagent	0.2µL	1µL	2 µl
Lipofectamine 3000 reagent	0.3µL	1.5µL	3 µl
Opti-MEM	Up to 10 µL	Up to 50 µL	Up to 100 µL

Appendix Table 5. Resources Table

Reagent or resource	**Source**	**Identifier**
General chemicals		
Sodium deoxycholate	Sigma-Aldrich	D6750
Ethanol	Sigma-Aldrich	443611
Methanol	Sigma-Aldrich	34860

Glycine	Sigma-Aldrich	G7126
Isopropanol	Sigma-Aldrich	563935
Agarose	Sigma-Aldrich	A9539
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich	L3771
Bovine serum albumin (BSA)	Sigma-Aldrich	A3059
NP-40	Sigma-Aldrich	I8896
Tris	Sigma-Aldrich	T1530
Tween-20	Sigma-Aldrich	P1379
Triton X-100	Sigma-Aldrich	T8787
Resazurin	Sigma-Aldrich	R7017
Methylene blue	Sigma-Aldrich	MB-1
Potassium hexacyanoferrate (III)	Sigma-Aldrich	P8131
Potassium hexacyanoferrate (II) trihydrate	Sigma-Aldrich	P9387
Western blot reagents		
Precision Plus Protein Dual Colour Standards	Bio-Rad	**1610374**
PageRuler Plus Prestained Protein Ladder	Thermo Fisher	26619
Pierce BCA Protein Assay Kit	Thermo Fisher	23225
X4 Blot LDS sample buffer	Thermo Fisher	B0007
X10 Bolt sample reducing agent	Thermo Fisher	B0009
Bolt Bis-Tris Plus 4-12% gel	Thermo Fisher	NW04120BOX
4-15% Criterion TGX Precast Midi Protein gels	Bio-Rad	5671084
20x MES SDS running buffer	Thermo Fisher	B0002
10x Tris/glycine/SDS running buffer	Bio-Rad	1610732
ECL Select Western Blotting Detection Reagent	Cytiva Life Sciences	12393969
Complete mini protease inhibitor cocktail tablet	Roche	4906845001
protease inhibitor cocktail	Roche	04693159001
0.45 μM Polyvinylidene fluoride (PVDF) membranes	Thermo Fisher	88518
FACS reagents		
Fixation Buffer	BioLegend	*420801*
True-Phos Perm Buffer	BioLegend	*425401*
Constructs		
pC0046-EF1a-PspCas13b-NES-HIV	Addgene A gift from Feng Zhang Lab	103862

PspCas13b-NES-HIV-T2A-BFP	Addgene	173029
pC0043-PspCas13b crRNA backbone	Addgene A gift from Feng Zhang Lab	103854
pMSCV-IRES-m-Cherry	Addgene A gift from Dario Vignali Lab	52114
MSCV-IRES-EGFP	A gift from Ilia Voskoboinik Lab	N/A
MSCV-IRES-tagBFP	A gift from Ilia Voskoboinik Lab	N/A
MSCV-BCR-ABL1-p190-3xHA-IRES-mCherry	This paper	N/A
MSCV-ABL1-3xHA-IRES-eGFP	This paper	N/A
MSCV-BCR-3xHA-IRES-tagBFP	This paper	N/A
MSCV-BCR-ABL1-p190-IRES-EGFP	A gift from Paul Ekert Lab	N/A
MSCV-SNX2-ABL1-IRES-EGFP	A gift from Paul Ekert Lab	N/A
MSCV-SFPQ-ABL1-IRES-EGFP	A gift from Paul Ekert Lab	N/A
MSCV-BCR-ABL1-p210-IRES-EGFP	A gift from Paul Ekert Lab	N/A
MigR1-AE(HA-RUNX1-RUNX1T1-IRES-EGFP)	Addgene A gift from Dong-Er Zhang Lab	12431
MSCV-NPM-ALK-IRES-mCherry	This paper	N/A
Cell line and bacteria strain		
HEK293 T cell	American Type Culture Collection (ATCC)	CRL-3216
K562 cell	American Type Culture Collection (ATCC)	CCL-243
Top10 chemically competent *E. coli*	Thermo Fisher	C404010
Stbl3 chemically competent *E. coli*	Thermo Fisher	C737303
Tissue/bacteria culture reagents and materials		
Dulbecco's Modified Eagle's Medium (DMEM) Gibco	Thermo Fisher	11965092
RPMI 1640 media	Thermo Fisher	11875083
GlutaMAX	Thermo Fisher	A1286001
Opti-MEM I Reduced Serum Medium	Thermo Fisher	11058021
TrypLE Express Enzyme (1X)	Thermo Fisher	12604013
Heat-inactivated fetal Bovine Serum (FBS)	Thermo Fisher	10100147
LB broth	Thermo Fisher	10855001
Ampicillin	Sigma-Aldrich	A9393

Penicillin/-Streptomycin	Thermo Fisher	151401220
Cell culture plasticwares	Greiner Bio-One, BD Biosciences, and Corning	NA
Western blot and FACS antibodies		
β actin monoclonal antibody (source: mouse western blot application: 1:2000)	Sigma-Aldrich	A2228
α-Tubulin Antibody (source: rabbit, application: 1:2000)	Sigma-Aldrich	SAB4500087
Phospho-Stat5 (Tyr694) (D47E7) (source: rabbit, application 1:2000 for western blot; 1:400 for FACS)	Cell signalling technology	4322
Stat5 (D2O6Y) (source: rabbit, western blot application 1 :2000)	Cell signalling technology	94205
P44/42 MAPK (Erk1/2) (source: rabbit, western blot application 1:1000)	Cell signalling technology	9102
p44/42 MAPK (Erk1/2) P Thr202/Tyr204 (source: rabbit, application 1:1000 for western blot; 1:400 for FACS)	Cell signalling technology	9101
(Horseradish peroxidase) HRP conjugated goat anti-mouse IgG secondary Antibody (western blot application: 1:10,000)	Abcam	ab97023
HRP conjugated goat anti-rabbit IgG secondary Antibody (application: 1:2000)	Abcam	ab205718
IRDye 680RD Goat anti-Rabbit IgG Secondary Antibody (0.1mg) (western blot application: 1:10,000)	Li-cor	92668071
IRDye 800CW Donkey anti-Mouse IgG Secondary Antibody (0.1mg) (western blot application: 1:10,000)	Li-cor	92632212
Alexa fluor 647 Goat anti-Rabbit IgG Secondary antibody (FACS application: 1:2000)	Thermo Fisher	A21245
Cloning enzymes		
BbsI restriction *enzyme*	New England Biolabs	R3539
NheI-HF restriction enzyme	New England Biolabs	R3131S
EcoRI restriction *enzyme*	Promega	R6011
BamHI restriction *enzyme*	Promega	R6021
T4 DNA ligase	Promega	M1801
DNA oligos for crRNAs cloning or gBLOCK for construct cloning		
crRNA sequences (Appendix Table 1)	Sigma or IDT	N/A
gBlock sequences (*Appendix Table 2)*	Sigma	N/A
Primer sequences (*Appendix Table 3)*	IDT	N/A

Critical commercial assays or kits		
NucleoSpin Gel and PCR Clean-up Kit	Macherey-Nagel	740609.50
NucleoSpin Plasmid Mini kit for plasmid DNA	Macherey-Nagel	740588.50
NucleoBond Xtra Maxi Plus	Macherey-Nagel	**740416.50**
Lipofectamine 3000 transfection kit	Thermo Fisher	L3000015
NucleoSpin RNA Plus	Macherey-Nagel	740984.50
Quick-RNA Miniprep Kit	*Zymo* Research	R1055
high-capacity cDNA reverse transcription kit	*Thermo Fisher*	4368814
PowerUp™ SYBR™ Green Master Mix	*Thermo Fisher*	A25742
Trans-Blot Turbo Midi 0.2 µm Nitrocellulose Transfer Packs	Bio-Rad	1704159
Phusion Site-Directed Mutagenesis Kit	Thermo Fisher	F541
mMESSAGE mMACHINE T7 Transcription Kit	Thermo Fisher	AM1344
Monarch RNA Cleanup kits	New England Biolabs	T2040L
Q5 High-Fidelity DNA kit	New England Biolabs	E0555S
4D-Nucleofector X Kit S	Lonza	V4XC-2032
Software and Algorithms		
Image J	National Institutes of Health	https://imagej.net
Prism software version 9 (GraphPad)	GraphPad Software Inc	https://www.graphpad.com/scientific-software/prism/
Image Lab	Bio-Rad	https://www.bio-rad.com/en-au/product/image-lab-software?ID=KRE6P5E8Z
Image studio	Li-cor	https://www.licor.com/bio/image-studio-lite/
FlowJo V10 software	Tree Star Inc	https://www.flowjo.com/solutions/flowjo/downloads
RNAfold	Vienna RNAfold	http://rna.tbi.univie.ac.at/
RNAplex	RNAplex	http://www.tbi.univie.ac.at/~htafer/
intaRNA	intaRNA	http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp;jsessionid=57B

		A904DBEE77BAA83BBAB1F20F55BAE
R package ‘ggseqlogo’	CRAN-The Comprehensive R archive network	https://cran.r-project.org/web/packages/ggseqlogo/
Matplotlib	Matplotlib	https://matplotlib.org/
Equipment		
EVOS M5000 FL Cell Imaging System	Thermo Fisher	AMF5000
Trans-Blot Semi-Dry electrophoretic transfer cell	Bio-Rad	1703940
Trans-Blot Turbo Transfer System	Bio-Rad	1704150
Odyssey CLx Imager	Li-cor	9140
iBright CL1500 Imaging System	Thermo Fisher	A44114
ChemiDoc Imaging System	Bio-Rad	1708265
StepOne Real-Time PCR system	*Thermo Fisher*	4376357
NanoDrop 2000/2000c Spectrophotometers	*Thermo Fisher*	ND-2000
4D-Nulceofector X unit	Lonza	AAF-1003X
FACSymphony A5 Cell Analyzer	BD Biosciences	NA
FACSymphony A3 Cell Analyzer	BD Biosciences	NA
FACSAria Fusion (5 lasers, 18 colours)	BD Biosciences	NA
FACSAria Fusion (3 lasers, 10 colours)	BD Biosciences	NA
Cytation3 Plate Reader	BioTek	CYT3V