Boston University

OpenBU **http://open.bu.edu**

Boston University Theses & Dissertations Boston University Theses & Dissertations

2023

Developing novel biosensing elements for molecular diagnostics

https://hdl.handle.net/2144/48023

Downloaded from OpenBU. Boston University's institutional repository.

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

DEVELOPING NOVEL BIOSENSING ELEMENTS FOR MOLECULAR DIAGNOSTICS

by

KAIYUE WU

B.S., Wuhan University, 2017

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2023

© 2023 by
KAIYUE WU
All rights reserved

Approved by

First Reader ____
Alexander A. Green, Ph.D.
Assistant Professor of Biomedical Engineering

Second Reader ____
Wilson W. Wong, Ph.D.
Associate Professor of Biomedical Engineering

DEDICATION

I want to dedicate this work to my grandparents Lianjia Ding and Suying Chen, who provided early inspirations that led me to life sciences. I want to also dedicate this work to the rest of my family members and friends who have supported me along this journey.

ACKNOWLEDGMENTS

I am incredibly thankful to the people who have helped and supported me during my Ph.D. journey. First of all, I would like to express my sincerest gratitude to my Ph.D. advisor, Dr. Alexander Green, for all the guidance and support he provided to my research and my career. I still remember the talk he gave during the orientation seminar. That really inspired me and has driven me to pursue a Ph.D. in synthetic biology. As a mentor and role model, he is always resourceful and inspirational and provides enough freedom for me to chase my research goals. During my Ph.D., I have the opportunities to work on many interesting and fulfilling projects that really broaden my skillsets and prepare me for the next career. It has been a huge pleasure to work with Alex.

In addition to that, I want to thank my committee members for their trust, help and support in my Ph.D. studies. Dr. Po-Lin Chiu and Dr. Yuval Mazor have provided a lot of useful feedback during the first three years of my Ph.D. at Arizona State University. Dr. Wilson Wong, Dr. John Ngo and Dr. Keith Pardee have served as my committee members since my transfer to BU and provided academic support toward my Ph.D. completion.

I would also like to acknowledge everyone whom I worked with in the Green lab. Dr. Duo Ma, Dr. Fanhong, Dr. Annie Tang, Dr. Yuzhou, Dr. Soma Chaudhary and Dr. Kristie Swingle have provided a wealth of guidance when I just started to work on synthetic biology with no prior experience. Special thanks to my labmate, roommate and good friend, Zhaoqing Yan for his kind support in and out of the laboratory throughout the 5 years. Also, I want to thank Yudan Li and Jinan Qin for their contributions on the research projects. It was a great pleasure to work with all the members of the Green lab.

Lastly, thank you to my dear family members for always being supportive. Their support and love have always motivated me to overcome challenges and difficulties.

DEVELOPING NOVEL BIOSENSING ELEMENTS FOR MOLECULAR DIAGNOSTICS

KAIYUE WU

Boston University Graduate School of Arts and Sciences, 2023

Major Professor: Alexander A. Green, Assistant Professor of Biomedical Engineering

ABSTRACT

Diagnostics are critical tools to assist in the identification of pathogens, the assessment of medical conditions, and helping to inform therapeutic decisions. Nevertheless, commonly used molecular diagnostics often require sophisticated instruments and skilled technicians, and therefore can only be done in centralized, well-equipped laboratories, which leads to long turnaround times, increased costs, and limited accessibility. These limitations have motivated the development of rapid, low-cost, decentralized diagnostics that are more widely accessible, affordable, and suitable for point-of-care applications.

Synthetic biology, by creating rationally designed biological components that can sense disease markers, provides innovative and promising diagnostic solutions to achieve highly sensitive and specific detection for targets of interest, while at the same time being time- and cost-efficient, field-deployable, and shelf-stable. This dissertation focuses on the development of novel biosensing elements and their diagnostic applications. First, I introduce the methods for the computational design of riboregulators using automated algorithms. Followed by that, I describe the development, optimization, and applications of toehold-switch-based platforms for the detection of coccidioides, noroviruses, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Next, I introduce the development of an ultra-specific riboregulator system termed single-nucleotide specific programmable riboregulators (SNIPRs) and their use for detecting different variants of concern of SARS-CoV-2. It is shown that riboregulators can be ideal solutions for various pathogen diagnostics with comparable accuracy and reduced cost. Lastly, I describe the use of peptide reporters derived from split protein systems to detect gene mutations. By incorporating peptide reporters into amplification primers, detection can be achieved by a quick isothermal amplification step and cell-free gene expression. Together, this research brings advancements in diagnostics based on riboregulators and cell-free systems that will increase the accessibility of these essential healthcare tools.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGMENTS v

ABSTRACT vii

TABLE OF CONTENTS ix

LIST OF FIGURES xiii

LIST OF ABBREVIATIONS xvi

CHAPTER 1 INTRODUCTION 1

1.1 Chapter overview 1

1.2 Synthetic biology on RNA-based biological regulation 3

1.3 Synthetic biology on cell-free gene expression 5

1.4 Affordable nucleic-acid diagnostics 7

1.4.1 Paper-based cell-free diagnostics 8

1.4.2 CRISPR-based diagnostics 8

1.5 Figures 10

CHAPTER 2 COMPUTATIONAL DESIGN OF RNA TOEHOLD-MEDIATED TRANSLATIONAL ACTIVATORS 15

2.1 Abstract 15

2.2 Introduction 16

2.3 Materials 22

2.4 Methods 23

2.4.1 Installation of software packages 23

2.4.2 Computational design of toehold switches 24

2.4.3 Computational design of SNIPRs 27

2.4.4 Notes 30

2.5 Figures 33

CHAPTER 3 APPLICATIONS OF TOEHOLD-SWITCH-BASED CELL-FREE DIAGNOSTICS FOR PATHOGEN DETECTION 38

3.1 Introduction 38

3.2 Results 41

3.2.1 Rapid, low-cost detection of valley fever using toehold switches 41

3.2.2 Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment 42

3.2.4 Interfacing Gel Switch Resonators with Cell-Free Toehold Switches for SARS-CoV-2 detection 44

3.2.5 Detection of SARS-CoV-2 with LAMP amplicons 47

3.3 Discussion 48

3.4 Materials and Methods 50

3.5 Figures 58

CHAPTER 4 DEVELOPMENT AND APPLICATIONS OF SINGLE-NUCLEOTIDE-SPECIFIC RIBOREGULATORS FOR SARS-COV-2 VARIANT DETECTION . 74

4.1 Abstract 74

4.2 Introduction 74
4.3 Results 78
4.3.1 Developing SARS-CoV-2 variant diagnostics with SNIPRs 78
4.3.2 Integrating isothermal amplification for highly sensitive and specific variant detection 80
4.3.3 Clinical assessment of SNIPRs in the detection of variants 81
4.3.4 Rapid development of SNIPR platform for BA.1 and BA.2 distinction 82
4.4 Discussion 84
4.5 Materials and Methods 86
4.6 Acknowledgments 89
4.7 Figures 90
CHAPTER 5 RAPID AND LOW-COST DETECTION OF MUTATIONS USING TRANSLATION BASED SEQUENCE VERIFICATION ASSAYS 97
5.1 Abstract 97
5.2 Introduction 98
5.3 Results 101
5.3.1 Implementing a Mutation-Specific Translation-Based Assay 101
5.3.2 Rapid, isothermal detection of mutations 106
5.3.3 Implementing signal self-calibration, long-range sequence verification, and room temperature operation 109
5.3.4 Mutation detection in colorimetric, paper-based reactions 111

5.3.5 Rapid, specific, and sensitive detection of mutations in clinical applications....... 112

5.4 Discussion 113

5.5 Methods.. 116

5.6 Acknowledgements 121

5.7 Figures.. 123

CHAPTER 6 CONCLUSIONS AND FUTURE DIRECTIONS 134

BIBLIOGRAPHY 139

CURRICULUM VITAE 154

LIST OF FIGURES

Figure 1.5.1 An example of a riboswitch functioning as thermosensors. 10

Figure 1.5.2 An example of a conventional riboregulator. 11

Figure 1.5.3 Schematic of toehold switches. 12

Figure 1.5.4 Schematic of loop-initiated RNA activators... 13

Figure 1.5.5 Schematic of an activated Cas13a protein cleaving 6-FAM labeled RNA probes with fluorescence or later flow output formats. 14

Figure 2.5.1. Schematic of the toehold switch mechanism.. 33

Figure 2.5.2 Structural information for Series A and Series B toehold switches. 34

Figure 2.5.3. Schematic of the SNIPR mechanism... 35

Figure 2.5.4. Schematic of the computational design of toehold switches against user-specified target sequences. 36

Figure 2.5.5 Schematic of the computational design of SNIPRs.................................... 37

Figure 3.5.1 Schematic of valley fever diagnostics using toehold switches with visual readout... 58

Figure 3.5.2 Screening of toehold switches against valley fever using cell-free systems with colorimetric output. 59

Figure 3.5.3 Time course measurement of the top performing toehold switch E9........... 60

Figure 3.5.4 Time course measurement of OD575 in cell-free reactions from different concentrations. 61

Figure 3.5.5 Color change photographs of valley fever toehold switch E9 captured at 0 hour and at 2 hours. 62

Figure 3.5.6. Overview of the norovirus detection assay using paper-based cell-free transcription-translation reactions. 63

Figure 3.5.7 Schematic of full-length lacZ and α-complementation, and the time-course measurement of OD 575 using lacZα and full lacZ as the reporter, respectively..... 64

Figure 3.5.8. Implementation of a synbody-based capture and concentration method for norovirus detection. 65

Figure 3.5.9. A gel-switch-resonator (GSR) test flow diagram. 67

Figure 3.5.10 Toehold switch sensor screening and validation in cell-free systems. 68

Figure 3.5.11. Demonstrating mail-in SARS-CoV-2 detection with toehold switch and gel-switch resonator system. 69

Figure 3.5.12 Schematic of toehold-switch cell-free detection with LAMP reactions..... 70

Figure 3.5.13 ON/OFF ratio of 142 toehold switches targeting ssDNA of LAMP products. 71

Figure 3.5.14. Sensitivity of D7 switch targeting the Spike gene. 72

Figure 3.5.15. Sensitivity of G12 switch targeting the human 18s rRNA gene. 73

Figure 4.7.1 Schematic of the SNIPR-based SARS-CoV-2 variant detection assay. 90

Figure 4.7.2 Design and screening of SNIPR sensors for variant identification. 91

Figure 4.7.3 The integration of isothermal amplification for sensitive and specific detection. 93

Figure 4.7.4 Validation of SNIPR assays on patient samples.. 94

Figure 4.7.5 Development of dual-fluorescent SNIPR assays for distinguishing Omicron BA.1 and BA.2 subvariants. 95

Figure 5.7.1 Schematic representation of detection of mutations using TRANSLATR. 123

Figure 5.7.2 Split fluorescent proteins are ideal reporters for detection of mutations.... 124

Figure 5.7.3 Implementing RPA to detect different mutations by TRANSLATR 126

Figure 5.7.4 Signal self-calibration by dual split fluorescent peptide reporter.............. 128

Figure 5.7.5 Using LacZ-alpha peptide reporter to enable paper-based and colorimetric output. 130

Figure 5.7.6 Streamlined workflow for the detection of clinical samples. 131

Figure 5.7.7 Detection of a nonsense mutation in the RB1 gene.................................. 132

Figure 5.7.8 Fluorescence response for BRCA2 6174delT in cell-free reactions and E. coli... 133

LIST OF ABBREVIATIONS

CFPS	Cell-free Protein Synthesis
CPRG	Chlorophenol red-β-D-galactopyranoside
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
LAMP	Loop-mediated Isothermal Amplification
NUPACK	Nucleic Acid Package
NASBA	Nucleic Acid Sequence Based Application
RBS	Ribosome Binding Site
PCR	Polymerase Chain Reaction
RPA	Recombinase Polymerase Amplification
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2

CHAPTER 1 INTRODUCTION

1.1 Chapter overview

Cell-free synthetic biology has enabled gene expression in a test tube. When combined with biosensors, they can be used for diagnostics. In this dissertation, I describe the development of biosensing elements for cell-free diagnostics, including the development, optimization and applications of novel and existing technologies to expand nucleic acid-based diagnostics through the use of synthetic biology toolkits.

In chapter 1, I will introduce the current progress of RNA-based biological regulation and cell-free gene expression in synthetic biology. I will then outline the current affordable diagnostics enabled by synthetic biology, including paper-based cell-free diagnostics and CRISPR-based diagnostics.

In chapter 2, I will go through the computational design methods for two riboregulators that respond to nucleic acid molecules and regulate gene expression: toehold switches and single-nucleotide specific programmable riboregulators. Based on a nucleic acid designed suite termed NUPACK, the compiled design algorithms allow fast generation and analysis of riboregulator sequences that respond to specific nucleic acid targets.

In chapter 3, I will introduce the development and application of toehold switches for the use of pathogen diagnostics, including norovirus, valley fever, SARS-CoV-2. Through the incorporation of paper-based cell-free systems, toehold switches can be used as sensors for pathogen nucleic acids. The sensors, upon binding their cognate targets, lead to conformation change that will result in activation of the downstream genes that produce measurable signals. Cell-free systems provide a uniquely ideal bed for toehold-switch mediated reactions, enabling in vitro applications towards point-of-care diagnostics.

In chapter 4, I will focus on the development of single-nucleotide specific programmable riboregulators (SNIPRs) and applying it to detection of SARS-CoV-2 variants. SNIPRs are a novel class of riboregulators with the ability to sense and discriminate single nucleotide mutations for their targets. By sophisticated thermodynamic energy control, the SNIPRs operate near chemical equilibrium: a fully matched target leads to a negative reaction free energy that shift the reaction to the ON state which will turn on gene expression. A target with one-nucleotide mismatch will cause an energy penalty that keeps SNIPRs to OFF state with no translation. As an example, we showed that SNIPRs can be applied to the detection of SARS-CoV-2 variants sharing highly similar sequences with few mutations. The high accuracy returned from clinical sample validation suggests SNIPRs can potentially be used as a diagnostic for quick screening of SARS-CoV-2 variants.

In chapter 5, I will introduce a peptide-based reporter system termed TRANslation of Sequence-LAbelled TRanscripts (TRANSLATR) for detection of gene mutations. Obtained from self-assembling split protein systems, peptides are so short that they can be genetically encoded into amplification primers. These modified primers amplify the target sequence and attach the peptide primers downstream of the targets. Mutation status of the target is revealed by cell-free gene expression which rapidly converts genotype to phenotype.

In chapter 6, I will conclude the research done in this research and point out future directions in terms of improving these technologies and expanding the expand the use of them. Also, I will point out future efforts needed to bring cell-free diagnostics from laboratories to clinics or, even more, to the point of care or the point of need.

1.2 Synthetic biology on RNA-based biological regulation

Synthetic biology aims to create, control and program cellular behaviors through the application of engineering principles. The term was first described by Stéphane Leduc in his publication[1], but the field did not actually launch until early 2000s[2]. In two decades, synthetic biology has undergone dramatic growth and has developed to a multidisciplinary field that encompasses methodologies ranging from genetic engineering, system biology, molecular biology and more[3]. To cover every perspective of synthetic biology is extremely hard, so here we aimed to provide some contexts that are relevant to this dissertation by giving a brief introduction to the biological regulation aspect of synthetic biology, specifically RNA-based regulation.

The central dogma of molecular biology implies that the primary function of RNA is to convert the information stored in DNA to proteins. But apparently, the functions of RNA are far more diverse and complex than that. For example, some RNAs are known to be catalytic[4]; others are found to have regulatory roles in cells[4]. In general, RNAs that have functions other than encoding proteins are named non-coding RNAs[5].

Two good examples of non-coding RNAs with regulatory functions are riboswitches and riboregulators. Riboswitches are RNA sensors located on the 5' untranslated region of mRNA that detects environmental (e.g., temperature) and metabolic (e.g., amino acids) signals and respond by structural changes which affect the protein production encoded by the mRNA[6] (**Figure 1.5.1**). They are composed of two canonical domains, a sensor domain that interacts with the signals or ligands and an effector domain that directly acts on gene expression. Riboregulators, on the other hand, share the same domain composition, except that the sensor domain only binds the cognate RNA or DNA stands (**Figure 1.5.2**).

Harnessing these RNA-based regulatory elements, scientists have attempted to create synthetic non-coding RNA regulatory elements for diverse applications. Besides the fact that RNA has already known to regulate biological processes, choosing RNA (instead of DNA or proteins) as the molecule for engineering also benefits from the following aspects: (1) Unlike DNA which is usually double-stranded and requires strict base-pairing, RNA can form diverse structures; (2) Unlike proteins, the composition of RNA is as simple as A, G, U, C; (3) RNA secondary structures are predictable by computer software through the calculation of free energy. (4) Recently, the cost of DNA synthesis has gone down dramatically, and RNA can be simply attained from DNA by transcription.

To date, a variety of *de novo* designed synthetic RNA-based regulatory elements have been developed[7–10]. These regulatory elements can be designed to operate in either transcriptional or translational levels. Here I provide a few examples of those with brief descriptions.

Toehold switches[9] are translational activators for controlling gene expression. They are RNA molecules containing a single-stranded toehold region in their 5’ end, followed by a stem-loop structure and a downstream gene (**Figure 1.5.3**). The ribosome binding site and the start codon sequence in the stem-loop structure block their accessibility from translation machinery and therefore preventing translation of downstream gene. An RNA complementary to the toehold regions and part of the stem served as the trigger of the toehold switch, which opens the stem-loop and exposing the RBS and the start codon for translation.

Loop-initiated RNA activators (LIRAs)[10] are also translation-level gene expression regulators and have the same basic working mechanism as the toehold switches by blocking RBS and start codon accessibility. The biggest difference worth noting is that the toehold region used to initiate trigger and switch binding is now within the loop region, with the RBS and the start codon being placed in the stem (**Figure 1.5.4**). This design allows a stronger translation inhibitory effect.

These gene expression regulators have been shown to function in prokaryotic cells[9–11] and more recently, in eukaryotic cells[12]. They can also be used to perform multi-input logic operations in vivo[13]. Enabled by cell-free systems, these systems can function outside living cells and produce signals that could be used as RNA sensors for diagnostics.

1.3 Synthetic biology on cell-free gene expression

Cell-free expression (CFE), also known as cell-free protein synthesis (CFPS) and cell-free transcription and translation (TX-TL) is a gene expression method using biological machinery in a cell-free system outside living cells. In the year of 1961, the renown experiments of Nirenberg and Matthaei that revealed genetic codes used a bacterial cell extract for cell-free gene expression[14]. Since then, different types of CFE systems from varying organisms have been invented, with E. coli extract being the most popular one to use due to its low cost, high yield, and robust product yield. If categorized by the preparation methods, CFE systems can be divided into two versions: (1) cell extract-based system, in which cells were properly lysed and the transcription and translation components inside the cells are used for expression; (2) purified-enzyme-based system, in which the proteins involved in the expression process are pre-purified and mixed together with other ingredients. Compared to live cell expression, cell-free protein expression systems have several key advantages as summarized below[15]:

1. Easy to perform: gene expression can be done by simple addition and mixing of components
2. Short reaction time: the reactions are typically completed within 1-24 h
3. High and scalable protein yield: maximal yield up to 50mg/ 10mL reaction
4. Ideal for hard-to-express proteins: toxic, membrane, inclusion and insoluble proteins that are considered problematic to express in cells can be expressed in cell-free systems
5. Open system: ingredients can be customized for specific applications, such as the addition of protease inhibitor and non-canonical amino acids.

In recent years, there is a technical renaissance in CFE systems that has contributed to refined CFE systems with high productivity and adaptivity and various applications beyond protein synthesis and into the fields such as education[16,17], vaccines[18], diagnostics[19], metabolic manipulation[20,21], etc. In 2014, Pardee and colleagues developed an in vitro paper-based platform for gene expression using commercially available cell-free systems. The cell-free systems can be freeze-dried onto the papers and be reactivated through rehydration, which provides an inexpensive, sterile and abiotic method for protein synthesis and allows engineered gene regulation circuits to be deployed out of cells. The cell-free systems demonstrate successful production of antimicrobial peptide antibodies, vaccines, showing the potential of cell-free systems for real-world applications.

1.4 Affordable nucleic-acid diagnostics

Entering the second decade, the goal of synthetic biology has been to apply scientific discoveries to real-world applications. We've seen the applications of synthetic biology research in food production, biosensing, therapeutics and diagnostics, etc[22]. This dissertation focuses mainly on the development of rapid, low-cost nucleic acid diagnostic tools using synthetic biology.

Nucleic acid diagnostics are considered sensitive and accurate testing tools compared to others like antigen-based testing. However, with qPCR being the most frequently used nucleic-acid testing methods, it is still hard to bring such test out of the lab due to the requirements of thermocycling equipment and trained personnel. Several nucleic-acid diagnostics were developed to address these limitations and aimed to bring nucleic-acid tests to the point-of-care. In this section, I will provide some backgrounds on current diagnostics enabled by synthetic biology, specifically on two diagnostics that have entered clinical phases[23].

1.4.1 Paper-based cell-free diagnostics

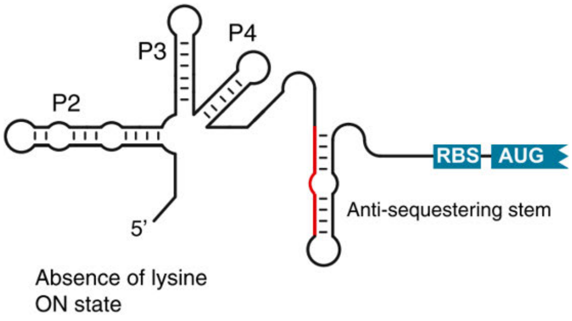
Synthetic riboregulators are ideal RNA sensors that can be designed to target almost any RNA and conditionally regulate gene expression. When combined with cell-free systems, riboregulators can be used to detect and report the presence of nucleic acids from pathogens. Moreover, cell-free reactions can be freeze-dried on a piece of paper for extended storage and transportation term and can be reactivated through rehydration. The properties facilitated the applications of riboregulators and cell-free systems in diagnostics.

In 2016, in response to Zika outbreak, toehold-switch-based cell-free diagnostics were developed[19]. The method used computer designed toehold switches and paper-based cell-free reactions to detect RNA of Zika virus with colorimetric output. With the use of an isothermal amplification method termed NASBA, the method reaches femtomolar detection sensitivity. Since then, different types of riboregulators were used to create affordable diagnostics for pathogens including norovirus[24], HIV[10], chikungunya viruses[25]. Riboregulator-based cell-free diagnostics provided a new avenue towards next generation diagnostics that could provide people with wider access to testing.

1.4.2 CRISPR-based diagnostics

CRIPSR(Clustered Regularly Interspaced Short Palindromic Repeats)-Cas technologies have revolutionized many areas of biological research[26]. Among these CRIPSR-Cas systems, Cas12 and Cas13 have been harnessed for sensitive nucleic acid diagnostics owing to their promiscuous trans- cleavage properties (**Figure 1.5.5**). More specifically, Cas13a or C2c2, upon binding complementary crRNA, will cleave its target RNA in cis, which then triggers the trans cleavage activity to promiscuously degrade RNA. Whereas Cas12a or Cpf1, upon binding commentary crRNA, will cleave its double-stranded DNA target is cis, which then triggers the trans cleavage activity to promiscuously cleave single-stranded DNA. Due to these special collateral cleavage activities, CRSPR Cas13 and Cas12 were used to create nucleic acid diagnostic platforms. In 2017, Gootenberg and colleagues[27] demonstrated that Cas13a can couple with isothermal amplification methods for highly sensitive and specific nucleic acid diagnostics, for which they named it SHERLOCK (specific high sensitivity enzymatic reporter unlocking). Soon after that, DETECTR (DNA endonuclease-targeted CRISPR trans reporter) that uses Cas12a collateral cleavage activity[28] came out. SHERLOCK and DETECTR became the foundations of CRISPR-based diagnostics. To date, there are countless diagnostic research using CRISPR-Cas systems, they are all based on and built upon the basic collateral cleavage activity. The representative ones were reviewed recently[29]. CRISPR-based diagnostics are rapidly growing and expanding, which have the promise of becoming options for future point-of-care diagnostics.

1.5 Figures


P2
P3
P4
RBS
AUG
Anti-sequestering stem
5'
Absence of lysine
ON state
P2
P3
P4
P5
Lysine
Sequestering stem
RBS
AUG
P1
5'
Presence of lysine
OFF state

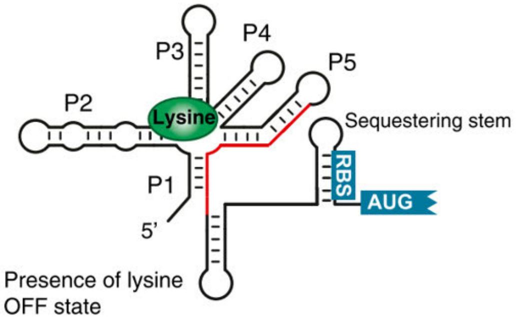


Figure 1.5.1 An example of a riboswitch functioning as thermosensors.

Without the presence of lysine, the riboswitch forms an anti-sequestering hairpin, exposing the RBS and the start codon for normal gene expression. When lysine binds to the riboswitch, the conformation of the hairpin changes, resulting in the formation of a sequestering hairpin, blocking the ribosome binding site. The domain that binds to lysine is usually known as aptamer. Adapted from ref 6 with permission.

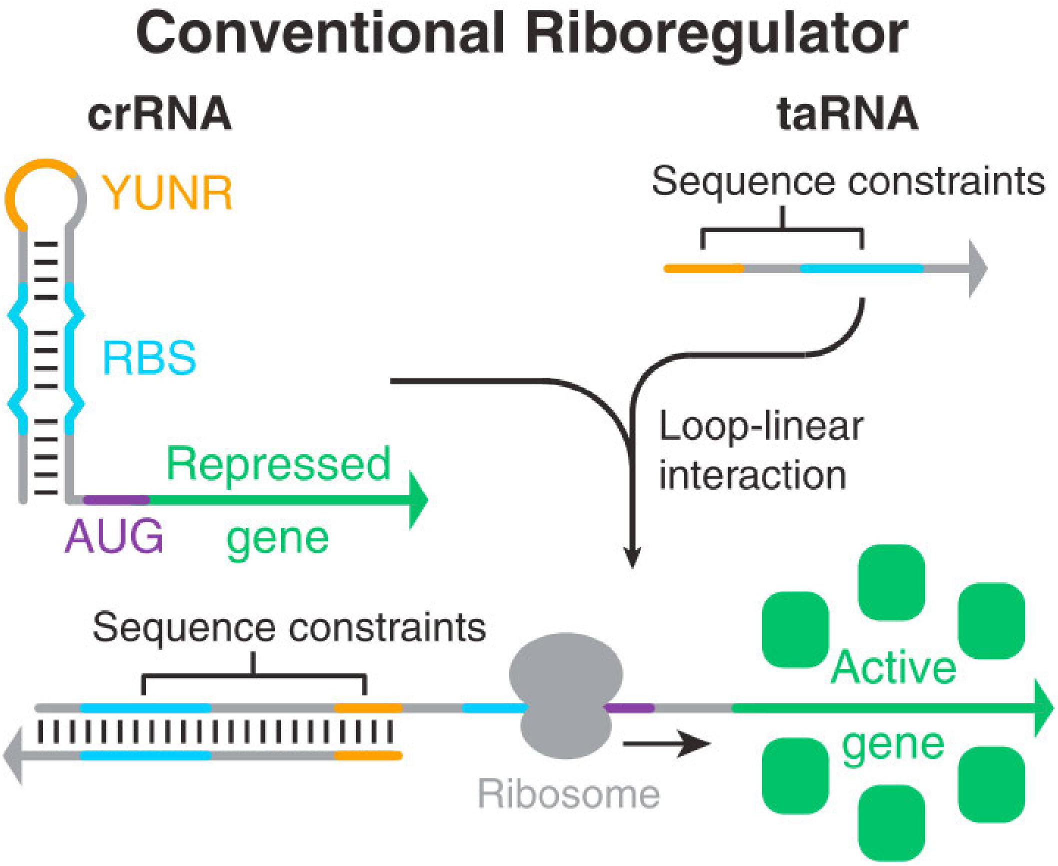

Conventional Riboregulator
crRNA
taRNA
YUNR
Sequence constraints
RBS
Loop-linear
interaction
Repressed
AUG
gene
Sequence constraints
Active
gene
Ribosome

Figure 1.5.2 An example of a conventional riboregulator.

In a conventional riboregulator, the ribosome binding site (RBS) sequence in blue is situated within the stem of the hairpin, including two bulges and the start codon (AUG) in purple is located after the stem. When taRNA presents, it binds to part of the loop and the stem including fixed sequence YUNR and RBS, which loosen the hairpin and permits translation. Adapted from ref. 9 with permission.

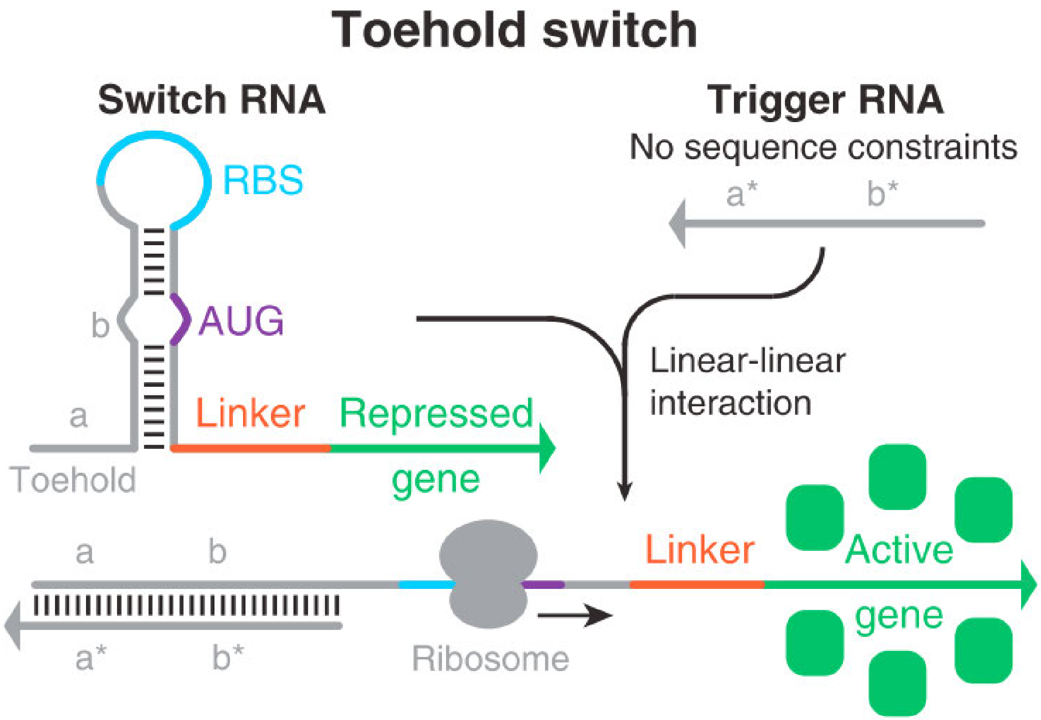

Toehold switch
Switch RNA
RBS
b
AUG
a
Linker
Repressed
gene
Toehold
Trigger RNA
No sequence constraints
a*
b*
Linear-linear
interaction
a
b
a*
b*
Ribosome
Linker
Active
gene

Figure 1.5.3 Schematic of toehold switches.

A typical toehold switch contains a single-stranded toehold region, a hairpin structure, a linker sequence and the downstream reporter gene. The ribosome binding site (RBS) in blue of a toehold switch is situated in the loop region of the hairpin, and the start codon is located within the stem. When a cognate trigger RNA is present, a toehold switch will bind to the trigger and relieve the hairpin, allowing translation to occur. Adapted from ref. 9 with permission.

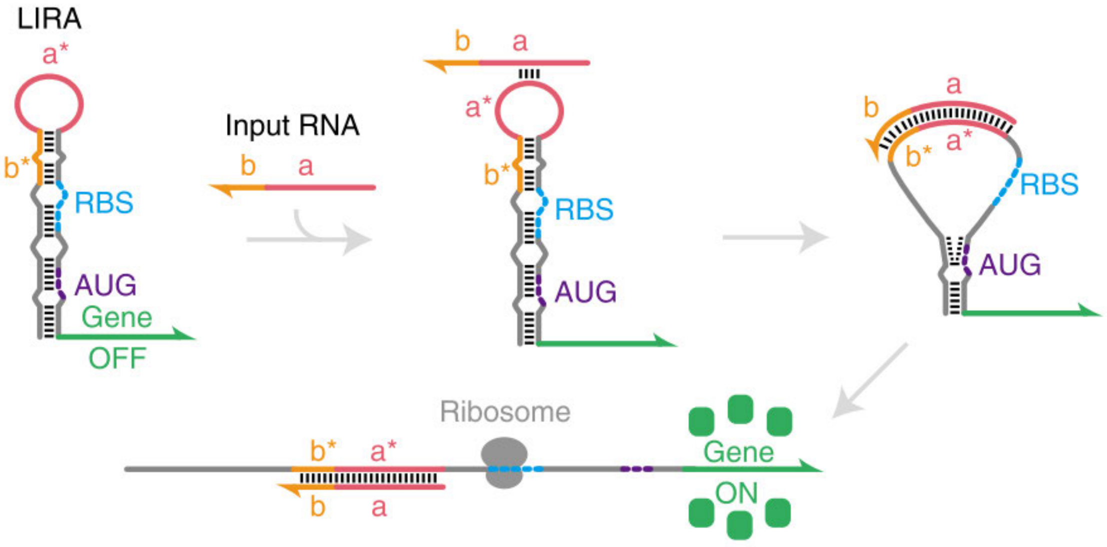

LIRA
a*
b*
RBS
AUG
Gene
OFF
Input RNA
b
a
b
a
a*
b*
RBS
AUG
b
a
b*
a*
RBS
AUG
Ribosome
b*
a*
b
a
Gene
ON

Figure 1.5.4 Schematic of loop-initiated RNA activators.

A typical loop-initiated RNA activator (LIRA) contains a hairpin structure, including key functional motifs like RBS, AUG. The binding of input RNA a and b to the LIRA at a* and b* initiates strand displacement to destabilize the stem and allows RBS to be accessed. Adapted from ref. 10 with permission.

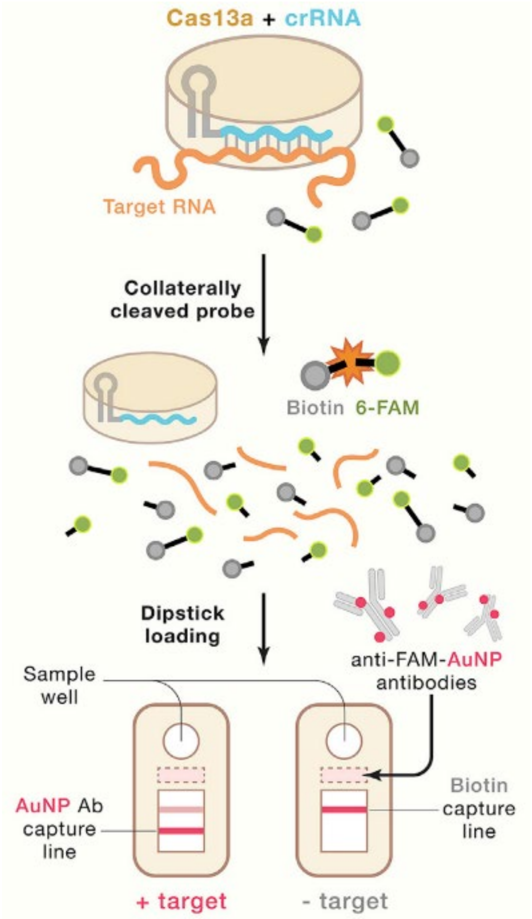

Cas13a + crRNA
Target RNA
Collaterally cleaved probe
Biotin 6-FAM
Dipstick loading
anti-FAM-AuNP antibodies
Sample well
AuNP Ab capture line
Biotin capture line
+ target
- target

Figure 1.5.5 Schematic of an activated Cas13a protein cleaving 6-FAM labeled RNA probes with fluorescence or later flow output formats.

A dsDNA template with its cognate crRNA can activate CRISPR's cis-activity to cleave the dsDNA plus its trans-activity to cleave ssDNA promiscuously. When ssDNA is tagged by Biotin and 6-FAM, it can interface with lateral flow strip to produce visible output, confirming the presence of the dsDNA target. Adapted from ref.23 with permission.

CHAPTER 2 COMPUTATIONAL DESIGN OF RNA TOEHOLD-MEDIATED TRANSLATIONAL ACTIVATORS

In this chapter, I introduce the computational design criteria and procedure for two de novo design translation activators: toehold switches and SNIPRs. This chapter is adapted from the chapter published as Wu, K., Yan, Z., & Green, A. A. (2022). Computational Design of RNA Toehold-Mediated Translation Activators. *Riboregulator Design and Analysis*, 33.

2.1 Abstract

Translation activators are an important class of riboregulators that respond to nucleic acid signals by activating gene expression. Toehold switches and Single-NucleotIde-specific Programmable Riboregulators (SNIPRs) are two types of translation activators that can detect nearly any nucleic acid sequence using interactions initiated by single-stranded domains known as toeholds. Toehold switches operate with high dynamic range, orthogonality, and programmability, making them capable of detecting a variety of pathogens in paper-based cell-free diagnostic assays. SNIPRs are designed to enable the accurate detection of single-nucleotide mutations, making them valuable tools for mutation and drug-resistance assays. Here we describe the computational design process for generating toehold switches and SNIPRs active against different pathogens and mutations of interest. Such riboregulators can be deployed in paper-based diagnostic assays to enable rapid and low-cost disease detection.

2.2 Introduction

Riboregulators are RNA molecules that respond to nucleic acid signals through Watson-Crick base pairing to control gene expression. This capability has allowed riboregulators to be used for constructing genetic circuits *in vivo* and *in vitro* that alter gene expression in response to endogenous transcripts or synthetic exogenous RNAs [30,31]. While earlier work on riboregulators focused on their applications in living cells [32–36], in recent years, there has been a growing interest in applying riboregulators in cell-free transcription/translation systems [37]. Without the constraints of the cellular chassis, cell-free systems provide unfettered access to the translational machinery and do not raise biosafety concerns. Moreover, these systems can remain active for extended periods at room temperature by freeze-drying them onto paper substrates [37], raising the exciting possibility of bringing synthetic biology tools outside the lab in a convenient portable format.

Within cell-free systems, riboregulators usually serve as sensors, activating gene expression once a target nucleic acid marking the presence of a pathogen or mutation is detected. Paper-based cell-free systems powered by riboregulators have been used for detecting the genetic signature of a variety of pathogens, including the Ebola virus [37], Zika virus [19], norovirus [24], and SARS-CoV-2 [38], along with gut microbiota [39] and mutations associated with cancer and drug resistance [40]. Coupling these paper-based molecular assays with isothermal amplification reactions results in diagnostic tests that cost a few dollars each, do not require expensive equipment, and provide test results that can be read out by eye. All these features make them promising technologies for use in global health and deployment in low-resource settings.

Designing riboregulators that respond effectively to pathogens and mutations can be a challenging process. Natural nucleic acid sequences often adopt thermodynamically stable secondary structures that frustrate sensor binding. Furthermore, the structure and translational efficiency of the riboregulator itself changes with each target sequence, typically requiring computational screening of hundreds of potential riboregulators for each pathogen. In this chapter, we describe the main strategies that our laboratory employs for designing riboregulators capable of detecting specific pathogens and mutations. We focus in particular on two types of *de-novo*-designed toehold-mediated translation activators: toehold switches [9], which are optimized for general pathogen detection, and Single-NucleotIde-specific Programmable Riboregulators (SNIPRs) [41], which are optimized for mutation detection. Both riboregulators employ a toehold domain, a free single-stranded sticky end, to initiate binding with their cognate target RNA, enabling them to operate with favorable thermodynamics and reaction kinetics. The following sections will cover the design of these translation activators, from the selection of targets and specification of design parameters to computer-assisted generation, analysis, and selection of candidate designs.

1. Toehold switches

Toehold switch riboregulators provide high dynamic range, orthogonality and programmability [9]. They were designed from first principles based on knowledge of thermodynamic principles and the influence of secondary structures on mRNA expression. These riboregulators contain two components: a trigger RNA and a switch RNA. The switch RNA contains a hairpin-based core sensing unit (**Figure 2.5.1**), which includes the critical toehold domain, and the sequence of the gene to be regulated. In the absence of the trigger, the hairpin conceals the ribosomal binding site (RBS) and the start codon within its loop and stem regions, respectively. The branch migration process that follows trigger binding changes the conformation of the switch and thus exposes the RBS and the start codon to allow gene expression to occur.

Despite sharing a similar secondary structure and design principle, several different versions of toehold switches have been developed with slightly different parameters based on different optimization and selection strategies [9,19,24]. We will discuss the basic design principles and methods for two standard versions of the toehold switch - Series A[9] and Series B[19] – and how these design principles interface with the design algorithms. These principles are generalizable to other toehold switches and riboregulator designs.

The first step of the design process involves the selection of targets and the specification of design parameters. The target of interest can be a fragment of a specific mRNA, rRNA, or viral RNA, preferably displaying weaker secondary structures. Even though the actual trigger binding region for toehold switches is around 30-40 nts, we recommend selecting a longer target (200-300 nt) as the input to allow the design algorithms to generate a large enough number of designs and to maximize the likelihood of getting a riboregulator with good performance.

According to the version of the toehold switch design selected, the specification of structure and sequence parameters varies and are listed as follows:

(1) Series A toehold switches

As shown in **Figure 2.5.2a** starting from the 5' end of the hairpin, 5' single-stranded toehold region is 30 nts, followed by a stem-loop structure, within which, the lower and the upper stems are 9 and 6 bp, respectively and are separated by a 3-nt bulge. The loop region is 11 nt. The trigger binding region is 36 nt, including the full toehold region and the first 6 nt in the lower stem. The upper part of the hairpin (from position 37 to position 71) has the conserved sequence:

GUUAUAGUUAUGAACAGAGGAGACAUAACAUGAAC, which includes the RBS and the start codon sequence. A universal linker sequence is attached to the 3' end of the hairpin: AACCUGGCGGCAGCGCAAAAG.

(2) Series B toehold switches

In Series B toehold switches (**Figure 2.5.2b**), the 5' single-stranded toehold region is 25 nt. The lower and the upper stems are 11 and 5 nt, respectively and are separated by a 3-nt bulge. The loop region is 12 nt. The trigger binding region is also 36 nt, including the full toehold region and the 11-nt lower stem. The conserved sequence in the upper part of the hairpin 37 (from position 37 to position 64):

GGACUUUAGAACAGAGGAGAUAAAGAUG. A universal linker sequence that follows the hairpin is the same as the Series A toehold switches:

AACCUGGCGGCAGCGCAAAAG.

We use the NUPACK software package [42] to assess all of the potential toehold switches for each target RNA computationally. The algorithm moves in 1-nt steps along the target RNA defining the switch RNA based on the sequence of its binding site in the target. For each candidate riboregulator, NUPACK is used to compute a variety of different parameters for the design, including the defect level of the switch hairpin, and the degree of single-strandedness of the toehold, trigger, and activated switch. This process is completed in stage 1 of the design algorithm.

In stage 2, the designs for each target are analyzed to assign them scores based on a set of selection criteria. First, the variable stem region in the toehold switches can generate in-frame stop codons between the start codon and the reporter gene. Therefore, such designs are first discarded. Next, the design parameters calculated during stage 1 are combined via different weighting coefficients into a single numerical score to rank each design. Taking Series A toehold switches as an example, the following defect levels are used to score the designs: (1) d_{sensor}: the defect for the toehold switch hairpin from 5' toehold to the 3' end of output gene, usually the first 29 nt; (2) d_{toehold}: the defect level of the toehold domain of the toehold switch hairpin compared to a fully single-stranded target structure; (3) $d_{target\_site\_unpaired}$: predicts the availability of the binding site within the target, by calculating the probability that the bases in the binding site will be unpaired in the context of the full target transcript. Each of the defect values is first calculated and then normalized by dividing it by the maximal value in the set generated for a given target sequence.

The scoring equation with weighting coefficients is summarized as below, with lower scores corresponding to better expected performance:

$$S = 3d_{sensor} + 4d_{toehold} + 5d_{target\_site\_unpaired} \tag{1}$$

This scoring process allows us to estimate the performance of the toehold switches. However, empirical testing is still necessary to validate the performance of the toehold switches. The Series B toehold switches employ a four-parameter scoring function that was established based on experimental testing of devices for norovirus detection [24]. The Series B scoring function provides a predicted ON/OFF ratio for each candidate design. The final stage of the algorithm selects top designs based on scores, makes final adjustments to the sequence to allow them to be ordered as single-stranded DNA oligos and compiles them into a spreadsheet.

2. SNIPRs

SNIPRs are toehold-mediated translation activators designed to operate with high-specificity and programmability [40]. Each SNIPR is composed of four functional elements (**Figure 2.5.3**): a docking site, an energy-balancing region, a translation-initiation region, and the output gene coding sequence. Translation from the SNIPR is repressed by positioning the RBS within a hairpin loop and the start codon (AUG) within the loop or a hairpin stem. The energy-balancing region is programmed to allow the SNIPR to operate near chemical equilibrium through a pair of forward and reverse toeholds domains and a double-stranded branch-migration region. This construction enables SNIPRs to be exquisitely sensitive to small changes in target sequence that alter the reaction free energy and hence induce a shift in equilibrium.

In the presence of the trigger RNA, hybridization with a complementary single-stranded docking site drives efficient recruitment of the trigger to the SNIPR, placing it in close proximity with the energy-balancing region. The ensuing reaction between the forward toehold domain and trigger RNA leads to a branch migration reaction causes the SNIPR stem to unwind and reach its ON state. Importantly, this OFF-state to ON-state transition with the cognate trigger is designed to have a slightly negative free energy difference of -1 to -2 kcal/mol. On the other hand, the binding of a trigger RNA with any single-nucleotide mutation that causes mismatch within the double-stranded branch migration region and generates a +3 kcal/mol energy penalty. This penalty leads to a positive free difference between the OFF state and ON state, thus preventing the mutated trigger from activating the SNIPR. In addition, in the case of a false activation by a mutant trigger RNA, the reverse toehold domain that is left undisturbed after a displacement reaction can initiate a reverse branch migration reaction to restore the OFF-state secondary structure.

2.3 Materials

All computational design work for the toehold switches and SNIPRs was done using computers running MacOS or Linux. The *in silico* design algorithms are run using MATLAB (installed with the Bioinformatics Toolbox) and they make use of functions in the NUPACK software package version 3.2. MATLAB releases R2013a or later have been used with the software. No other materials are required.

The toehold switch design software can be found at https://github.com/AlexGreenLab/TSGEN.

The SNIPR design software can be found at https://github.com/Albert09111/SNIPR.

2.4 Methods

2.4.1 Installation of software packages

1. Install the NUPACK version 3.2 software package:

First, go to the http://www.nupack.org website and click "Downloads" to go to the software download page.

Second, register to download NUPACK.

Lastly, follow the NUPACK installation instructions to allow the software to run on your computer.

2. Set up the environment variables in MATLAB to enable it to call NUPACK functions (*see* **Note 1**):

First, open MATLAB.

Second, open or create a startup.m file in your MATLAB default user work folder. This folder is defined in the MATLAB variable userpath and startup.m automatically run each time you open MATLAB.

Third, in the startup.m file, set the NUPACKINSTALL variable to the location of your NUPACK installation. To do this, use the following line of code with the location of the package, where [user_name] is your username on the computer:

```
NUPACKINSTALL = '/Users/[user_name]/…/nupack3.2.2';
```

Lastly, add the lines of code below to set NUPACKINSTALL as an environment variable and add the folder with NUPACK binaries to the PATH:

```
setenv('NUPACKINSTALL',NUPACKINSTALL);
setenv('PATH',[getenv('PATH'),sprintf(':%s/build/bin',NUPACKINSTALL)]);
```

The necessary environment variables will now be set whenever you start MATLAB.

2.4.2 Computational design of toehold switches

The general design process is shown in **Figure 2.5.4**. In brief, it can be divided into two sections: (1) selection of targets; (2) specification of design parameters and NUPACK-assisted computational design of toehold switches, which contains three stages: design generation, scoring, and sorting (*see* **Note 2**).

One of the advantages of toehold switches is their ability to target nearly any RNA sequence. The following process is often used to choose the potential toehold switch target regions that likely lead to good switch performance.

1. Determine the organism/species (e.g., genetic materials of a virus, mRNA) of interest and identify the related DNA/RNA target sequence. (*see* **Note 3**)
2. The targets can often be obtained from previous literature if available. Otherwise, they can be obtained by downloading sequences from different sequence databases (e.g., GenBank).
3. Identify the regions that remain conserved from the downloaded sequences. Then compare the conserved sequences with the ones in closely related species to ensure that the sequences for the target pathogen are unique (e.g., for SARS-CoV-2, you would compare it to the other human coronaviruses). This information can be obtained by aligning the sequences using the Basic Local Alignment Search Tool (BLAST) (*see* **Note 4**).

4. For each conserved, unique sequence, generate a target input sequence that is approximately 200 to 300 nt in length, which is ideal for use with toehold switches (*see* **Note 5**).

5. The resulting target sequences can then be entered into the design_input_file.csv file in the toehold switch design software input folder.

6. Open the spreadsheet named design_input_file.csv and specify the desired name of the target input under the Column 1 as indicated in the spreadsheet.

8. Specify the target input RNA sequences in Column 2 and 3.

a. If you have an existing set of amplification primers for your target input RNA, you can use Column 2 to specify the full expected amplicon sequence produced during amplification. Use Column 3 to specify the subsequence that does not include the sites where the amplicon binds to the amplification primers. This step ensures that the algorithm will not generate toehold switches that bind directly to one of the primers and the resulting assay interrogates three regions of the target rather than just two.

b. If you do not have amplification primers for your target sequence, you can use the same sequence in Columns 2 and 3. Alternatively, you can exclude a given number of bases on the 5' and 3' end that correspond to the typical primer length of the amplification method that you plan to use.

9. Input the desired temperature in Column 4 for running the toehold switch reactions in °C (e.g., 37).

10. Input the name of the output gene in Column 5.

11. Input the sequence of the output gene in RNA format in Column 6 (*see* **Note 6**).

12. Enter more targets as needed in the rows below. Leave the rows below blank if there are no more targets.

13. Save the spreadsheet file.

14. Open MATLAB and navigate to the folder that contains the toehold switch design software.

15. Run the function toehold_switch_design_run(num_designs,input_file,options) using the parameter settings of your choice:

(a) num_designs (default value = 6): Use this parameter to specify the number of designs that you want generated for each target sent to the software.

(b) input_file (default value = ‘design_input_file.csv'): Use this parameter to specify the location of the input file that provides the target sequences. The function will search for this file in an ‘input’ subfolder in the current working folder unless otherwise specified.

(c) SeriesA (default value = 1): Defines if the function will generate Series A designs. Set to 0 to not produce these designs.

(d) SeriesB (default value = 1): Defines if the function will generate Series B designs. Set to 0 to not produce these designs.

(e) Parallel (default value = 1): Enables processing with multiple cores on local computer. Set to 0 if you do not want parallelization enabled or do not have the Parallel Processing Toolbox.

(f) Antisense (default value = 0): Set to 1 if you wish to create toehold switch sensors for the antisense sequences of those provided in the input file.

For example, if you want to generate eight toehold switch designs for only Series A for targets specified in the ‘my_sequences.csv’ file in the input folder. You would specify:

```
toehold_switch_design_run(8,'my_sequences.csv','SeriesB',0);
```

For the same set of targets, adding the antisense orientation and Series B designs, the function call would be:

```
toehold_switch_design_run(8,'my_sequences.csv','SeriesB',1,'Antisense',1);
```

16. After the toehold switch design process is complete. Information on the resulting designs will be saved in the form of csv-format spreadsheets within the final_designs subfolder in your current MATLAB working folder. The files SeriesA_sensor_DNA_sequences.csv and SeriesB_sensor_DNA_sequences.csv will contain the DNA templates necessary to order for toehold switch construction. The files SeriesA_final_design_info.csv and SeriesB_final_design_info.csv will contain information on the RNA sequences for each design (*see* **Note 7**).

17. The toehold switch hairpins and targets are usually ≤200 nt so they can be ordered as individual DNA oligonucleotides. The hairpins are later assembled with the reporter gene to create fully functional toehold switches. Necessary primers should be ordered along with the hairpins for the subsequent PCR assembly process. (*see* **Note 8**)

2.4.3 Computational design of SNIPRs

The design process of SNIPRs (**Figure 2.5.5**) is also highly automated and the software is available online for researchers at https://github.com/Albert09111/SNIPR. Follow the installation instructions provided with the SNIPR software for installation. Since the SNIPR software also uses MATLAB to call NUPACK binaries, you can also follow the instructions in Section 3.1 to set up the necessary packages.

1. The designs of SNIPRs require two inputs: wild-type and mutant sequences. SNIPRs are capable of distinguishing mutations with only a single-nucleotide difference between wild-type and mutant sequence.
2. Similarly, the easiest way to find the targets is to look for them in existing literature. Otherwise, they can be selected from sequence databases (e.g., GenBank). The ideal sequences should be highly conserved, except for the mutation of interest.
3. To allow the sequence to be used as the input, tailor the sequence so the mutation position in the target is at least 15 bases away from the 5' end of the RNA. Sequences should be specified in the 5' to 3' direction.
4. Open target_input.csv located in the input folder within the SNIPR-master folder.
5. Specified the name of the output SNIPR sequence in Column 1.
6. Specified the wild-type and the mutated sequence in Column 2 and Column 3, respectively (*see* **Note 9**).
7. Input the first 29 nt of the output gene sequence (e.g., GFP, lacZ) in Column 4 (*see* **Note 10**).
8. Save and close the target_input.csv file.
9. Open the main code SNIPR.m in MATLAB.
10. The code contains several parameters with default values, but they can be customized. Before running the code, review the following parameters and edit as needed.

(a) library_num: this parameter determines the number of the design libraries the code will generate (*see* **Note 11**).

(b): IS_PARALLEL: set to 1 to enable parallel computing by harnessing multiple computer cores or 0 to disable parallel computing.

(c): select_num: this parameter determines the final number of top designs generated and selected from the libraries (*see* **Note 12**).

(d): SNP_TARGET: this parameter determines whether the wild-type sequence or the mutant sequence will activate the reporter gene expression. Set to 1 to allow mutant targets to activate gene expression; set to 0 to allow wild-type targets to turn on gene expression.

11.Save the parameters.

12. Run the SNIPR.m code to start generating SNIPRs design for the specified targets. The SNIPR code generates the designs automatically in four stages: (1) load sequence and generate design script, (2) run design script to generate seq library, (3) extract the design library, and (4) analysis and screening for the best design. The progress of the algorithm can be visualized in real time on the MATLAB command window.

13. After the full running process is completed, find the SNIPR RNA sequence for each target along with other parameters in the output folder.

14. To order SNIPRs as DNA oligos, convert the SNIPR RNA to DNA sequence and append the T7 promoter sequence to the 5’ end (*see* **Note 13**). Necessary primer sequences should also be ordered to assemble the SNIPR hairpin with the corresponding gene.

2.4.4 Notes

1. It is also possible to link MATLAB to the NUPACK functions by modifying the bash configuration file from the command-line terminal program of your choice. This procedure is described in the SNIPR software installation guide, which can be found at https://github.com/Albert09111/SNIPR. If you do choose this method, MATLAB must be started from the command line to set the necessary environment variables defined in the configuration file.

2. The toehold switch design code is available at https://github.com/AlexGreenLab/TSGEN.

3. Toehold switches have mainly been used for detection of RNA, but it is possible to detect DNA sequences. Designs may need to be adjusted for DNA detection based on the weaker binding for DNA/RNA hybridization compared to RNA/RNA hybridization.

4. We recommend selecting at least two regions from the pathogen DNA/RNA to increase the likelihood of getting a functional sensor. In addition, sense and antisense orientations can generally be detected by the toehold switch depending on whether the forward or reverse primer contains a T7 promoter.

5. To ensure the algorithms generate a sufficient number of designs in the pool, we recommend using at least targets that are at least 70 nts. While a longer target is generally better, it is not necessary for the initial screening process and increases the difficulty of synthesizing the target. We recommend using targets that are 200-300 nts as inputs to generate toehold switch design since targets of this length also provide many options for suitable amplification primers.

6. Only up to the first 29 nts of the output gene are considered in the algorithm in order to reduce computation times.

7. The toehold switch design algorithm also outputs to different subfolders spreadsheets containing the individual parameter values (e.g. switch defect level, toehold defect level, etc.) of all of the designs generated for each target sequence. You can use these spreadsheets to implement your own scoring functions for selecting higher-performance designs.

8. We recommend ordering and testing at least six top toehold switch designs generated from the design software to increase the likelihood of getting a toehold switch with very good performance.

9. The wild-type sequence and mutant sequence do not have to be in the same length. The algorithm will identify the mutation and proceed with designing as long as the mutation is position at least 15 nts away from the 5' end of the target sequence.

10. Output gene sequence will be considered in the SNIPR design analysis.

11. We recommend setting the library numbers to at least 10 to allow sufficient numbers of designs to be generated and screened.

12: We recommend setting this number to at least 5. In experimental screening, at least 5 top-ranked designs should be tested to increase the likelihood of getting a functional design.

13. The 5' GGG sequence is universally appended to all SNIPR RNA sequences generated, which represents the first three nucleotides transcribed using T7 RNA polymerases. The T7 promoter sequence we use is

GCGCTAATACGACTCACTATAGGG. Alternatively, other RNA polymerases can be used. We recommend ordering and testing at least six top SNIPR designs generated from the design software to increase the likelihood of getting a SNIPR with very good performance.

2.5 Figures

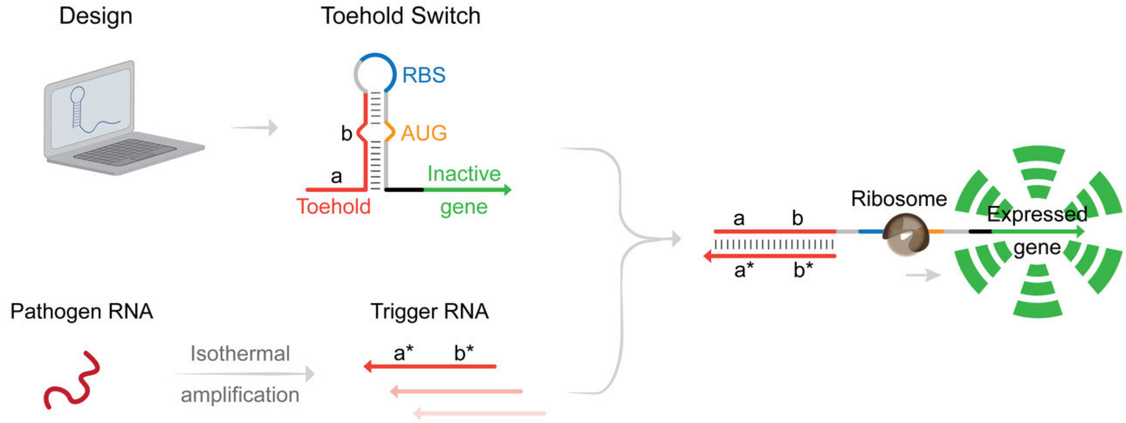
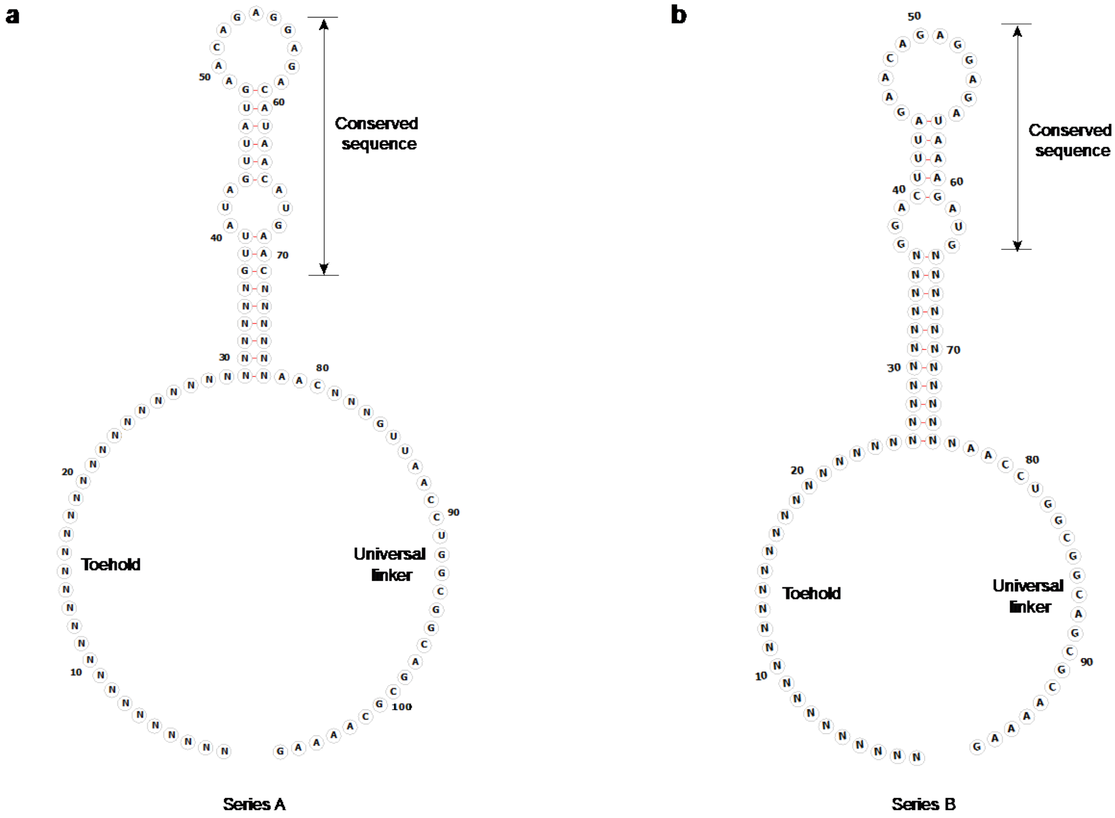

Design
Toehold Switch
RBS
b
AUG
a
Inactive
Toehold
gene
Ribosome
a
b
Expressed
gene
a*
b*
Pathogen RNA
Trigger RNA
Isothermal
amplification
a*
b*

Figure 2.5.1. Schematic of the toehold switch mechanism.

In the presence of trigger RNA, the toehold domain initiates a branch migration and releases the RBS and start codon from a hairpin secondary structure to activate translation.


a
Conserved sequence
Toehold
Universal linker
10
20
30
40
50
60
70
80
90
100
b
Conserved sequence
Toehold
Universal linker
10
20
30
40
50
60
70
80
90

Series A

Series B

Figure 2.5.2 Structural information for Series A and Series B toehold switches.

(a) Structural schematic of the Series A toehold switches. (b) Structural schematic of the Series B toehold switches. (N: undetermined nucleotide, A: adenine, G: guanine, C: cytosine, U: uracil). The diagrams were generated using Forna [43].

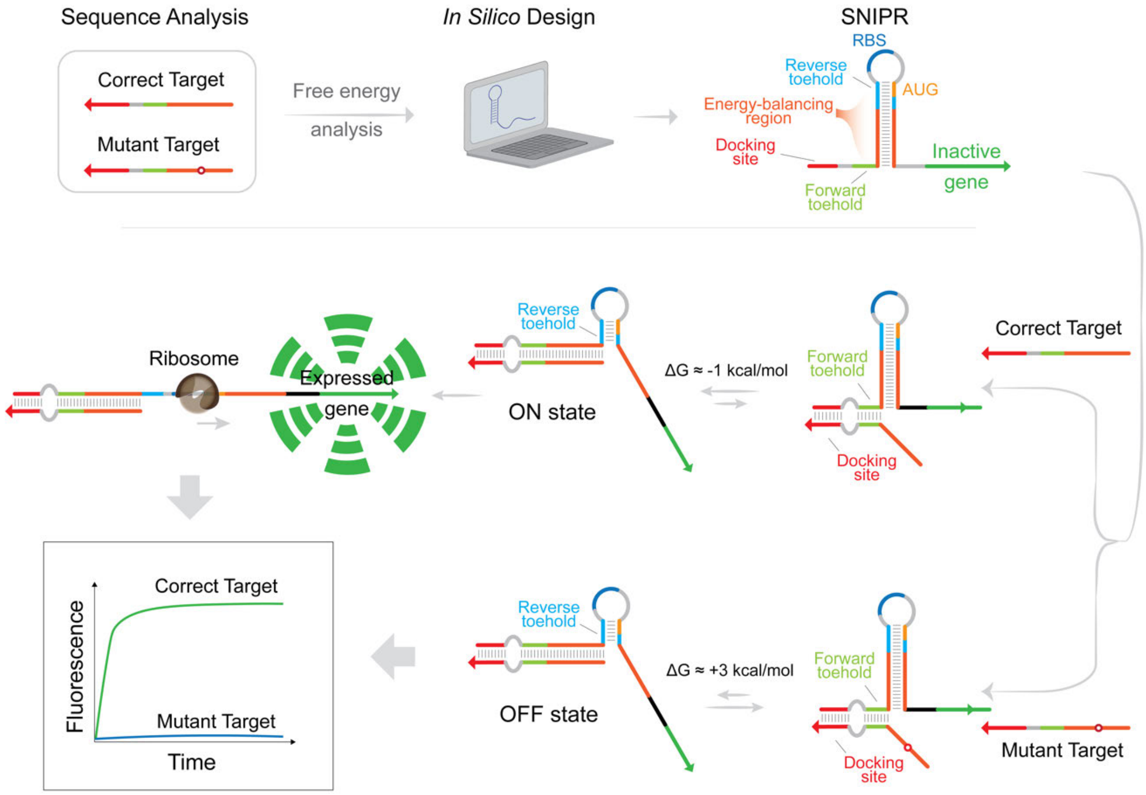

Sequence Analysis
Correct Target
Mutant Target
Free energy analysis
In Silico Design
SNIPR
RBS
Reverse toehold
AUG
Energy-balancing region
Docking site
Forward toehold
Inactive gene
Correct Target
Ribosome
Expressed gene
Reverse toehold
ON state
ΔG ≈ -1 kcal/mol
Forward toehold
Docking site
Correct Target
Fluorescence
Time
Mutant Target
Reverse toehold
OFF state
ΔG ≈ +3 kcal/mol
Forward toehold
Docking site
Mutant Target

Figure 2.5.3. Schematic of the SNIPR mechanism.

The binding of trigger RNA with correct sequence yields an OFF-state to ON-state transition with a slightly negative free energy, and hence releases the hairpin structure. A mutation in the trigger RNA results in an energy penalty leading to a positive reaction free energy, driving the SNIPR toward the OFF state.

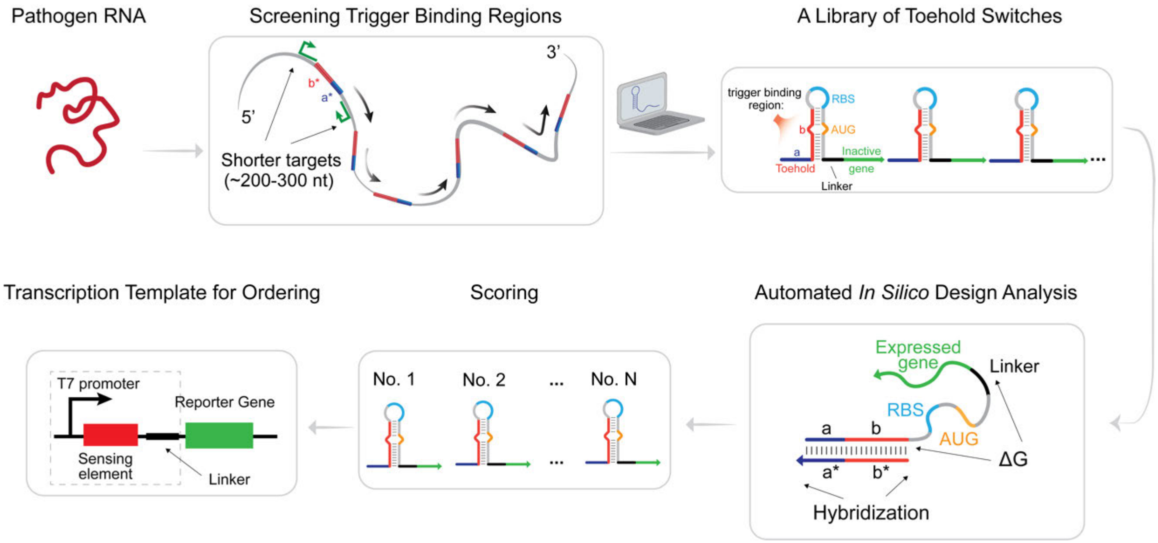

Pathogen RNA
Screening Trigger Binding Regions
5'
3'
b*
a*
Shorter targets
(~200-300 nt)
A Library of Toehold Switches
trigger binding region:
RBS
b
AUG
a
Inactive gene
Toehold
Linker
Transcription Template for Ordering
T7 promoter
Reporter Gene
Sensing element
Linker
Scoring
No. 1
No. 2
...
No. N
...
Automated *In Silico* Design Analysis
Expressed gene
Linker
RBS
a
b
AUG
a*
b*
ΔG
Hybridization

Figure 2.5.4. Schematic of the computational design of toehold switches against user-specified target sequences.

A variety of switches binding to different portions of the target (~200-300 nt) are designed and analyzed. The toehold switches are scored *in silico* and top predicted devices are output in DNA form for ordering.

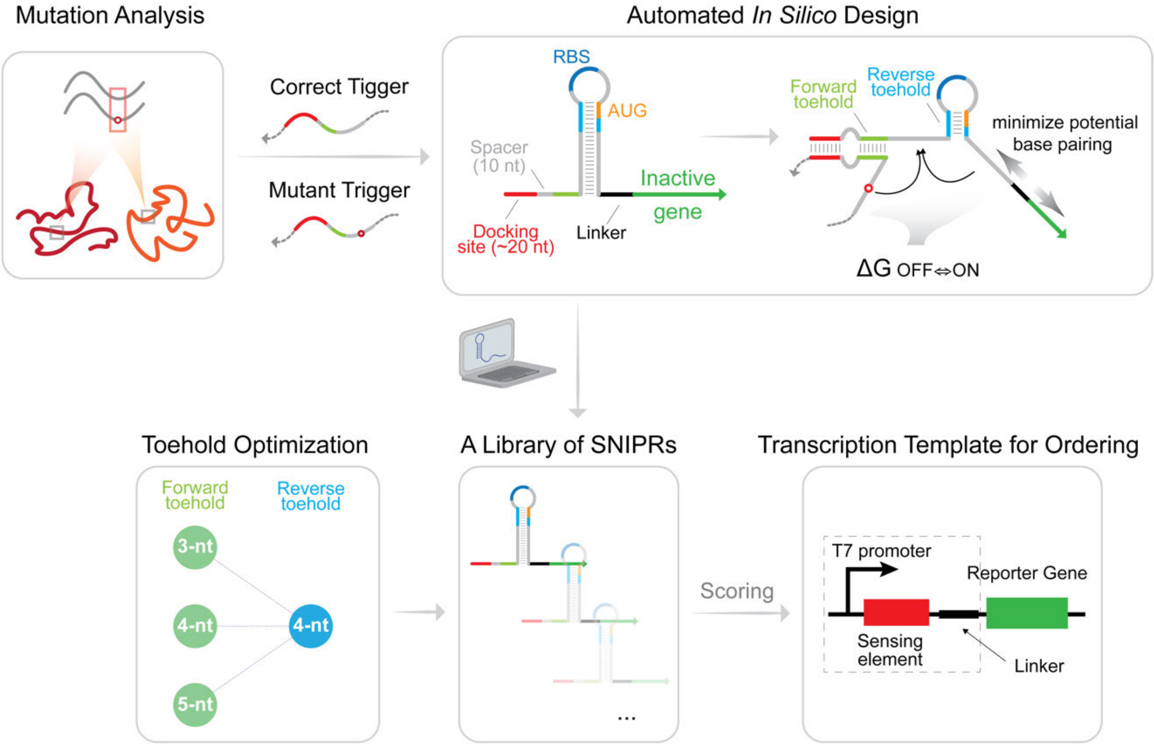

Mutation Analysis
Correct Tigger
Mutant Trigger
Automated *In Silico* Design
RBS
AUG
Spacer (10 nt)
Inactive gene
Docking site (~20 nt)
Linker
Forward toehold
Reverse toehold
minimize potential base pairing
ΔG OFF⇔ON
Toehold Optimization
Forward toehold
Reverse toehold
3-nt
4-nt
5-nt
4-nt
A Library of SNIPRs
...
Scoring
Transcription Template for Ordering
T7 promoter
Reporter Gene
Sensing element
Linker

Figure 2.5.5 Schematic of the computational design of SNIPRs.

The position of the mutation to be detected is identified and positioned to coincide with the middle portion of the SNIPR stem. Three different sets of designs containing different pairs of forward (3-, 4-, 5-nt) and reverse toeholds (4-nt) are generated *in silico*. The sequence of SNIPRs is analyzed and scored based on the free energy difference between the ON state and OFF state, the ensemble defect level, and the secondary structure of the active riboregulator.

CHAPTER 3 APPLICATIONS OF TOEHOLD-SWITCH-BASED CELL-FREE DIAGNOSTICS FOR PATHOGEN DETECTION

Section 3.2.4 is adapted from the article published in ACS sensors as Carr, A.R., Dopp, J.L., Wu, K., et al. Toward Mail-in-Sensors for SARS-CoV-2 Detection: Interfacing Gel Switch Resonators with Cell-Free Toehold Switches. ACS sensors, 7(3), pp.806-815. 2022.

3.1 Introduction

Toehold switches are RNAs containing a stem-loop structure which conceals the ribosome binding site and the start codon, therefore preventing downstream translation[9]. A toehold domain in front of the stem-loop structure allows dynamic interaction between toehold switches and target RNA. The matching RNA will then initiate a strand displacement reaction to open the hairpin and allow translation machinery to access. The ability to sense and respond to RNA targets has allowed toehold switches to be used as biosensors for gene regulation in E. coli cells[9]. With the use of cell-free systems[37,44], which enable gene expression to occur in a test tube, toehold switches have successfully been harnessed to create in vitro diagnostic platforms: including the detection of Ebola virus[37] and Zika virus[19,25]. Toehold-switch-based cell-free diagnostics have shown their potential for next-generation point-of-care diagnostics. In this chapter, I further described the continued development, optimization and applications of the toehold-switch-based diagnostic platforms to combat different pathogens of interest, including coccidioides, norovirus and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Below I will provide a brief introduction to each pathogen and its related disease, as well as the need for detecting them.

Coccidioides, more specifically, Coccidioides posadasii and Coccidioides immitis are the causative agents for valley fever, also known as coccidioidomycosis[45]. The fungi typically grow in dirt and soil and can make a person sick when he or she breathes in the dust that contains the fungi. It is becoming increasingly common in the southwestern United States, most notably in California and Arizona. Even though many cases show mild, flu-like symptoms and can resolve on their own, some will cause severe or even fatal chronic infection or widespread disseminated infection and will require treatment[45,46]. The disease is usually treatable; however, due to its highly similar symptoms to a common cold, diagnosis based on symptoms is challenging.

Norovirus is the etiological agent accounting for most of the acute gastroenteritis[47], which can lead to symptoms like fever, nausea, vomiting and diarrhea. It is highly contagious and is estimated to cause 219,000 deaths within nearly 700 million infections[48]. The virus also concerns the food industry as it can lead to foodborne illnesses[49].

Severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2, the virus that cause the global COVID-19 pandemic, is a strain of coronavirus that cause the respiratory illness. Due to its ability for human-to-human transmission, the disease rapidly spread across the world and resulted in nearly 600 million cases with 6 million deaths[50].

These diseases are either highly contagious and are hard to distinguish based on symptoms, thus making the development of accurate, rapid and scalable diagnostic methods very important. Nucleic acid tests are among the most accurate for pathogen diagnostics as they checked for the genetic materials[51]. The most used and validated qPCR tests are mainly limited to centralized laboratories as they need expensive instruments and trained technicians. The need for transporting samples from the point of acquisition to the testing laboratories also increases the turnaround time. Other methods, such as serology tests, antigen tests or *ex vivo* culture, could have one or several of the following drawbacks: (1) limited sensitivity and specificity; (2) require high biosafety level containment; (3) long turnover time; (4) requiring special and expensive instrument.

In 2016, a streamlined protocol was introduced to detect Zika viruses using the toehold switch platform, providing a promising strategy for low-cost rapid pathogen diagnostics[19]. It demonstrates a way in which synthetic biology can be used to address real-world challenges. Nevertheless, future improvements are needed in some aspects of this technologie: (1) further enhancing the speed of reactions. As described in the paper, RNA amplification can take 2 hr to complete and the toehold reaction will be another 1 hr. (2) increasing the sensitivity of the reactions. With the incorporation of NASBA reactions, the toehold switch detects viral RNA down to 1 fM. However, to ensure the tests can be applied towards more pathogens, a higher sensitivity is desired. Additionally, interfacing toehold switch reactions with other output systems would also be interesting as they expand the application of toehold sensors.

In this chapter, I expanded the use of toehold-switch-based diagnostics to the detection of Coccidioides, norovirus and SARS-CoV-2 with the goals to further improve the limitations mentioned above. The results showed successful detection of all three methods with sensitivity down to attomolar and had a decreased reaction time.

3.2 Results

3.2.1 Rapid, low-cost detection of valley fever using toehold switches

To design a nucleic-acid-based test for Coccidioides, the organism causing valley fever, we first looked at the sequence and the predicted secondary structure from its full genome. We selected two regions whose sequences are unique to the pathogen with predicted low secondary structures. Next, we used the toehold switch design algorithm to generate toehold switch candidates targeting the two regions. A schematic of the workflow is depicted in **Figure 3.5.1**. The toehold switch sequences were synthesized and assembled by polymerase chain reaction (PCR). We screened 31 toehold switch candidates empirically by putting the candidate sensors into cell-free systems with or without the cognate triggers. β-galactosidase was used as the reporter gene for the switches with the color-changing substrate termed Chlorophenol red-β-D-galactopyranoside (CPRG) that changes the color from yellow to dark red when catalyzed by the enzyme. The color difference for the switch is quantified and ranked by the difference in OD 575 nm measurement as shown in **Figure 3.5.2**. The cell-free experiment identifies switch E9 as the top-performing candidate (**Figure 3.5.3**).

Keeping in mind the goal of increasing the speed of the assay, we selected RPA as the amplification method: RPA uses recombinase to initiate strand break and primer-template interaction; strand-displacing DNA polymerase allows rapid generation of DNA amplicons as short as 10 minutes[52]. Compared to the usual 2-hr reaction time of NASBA, RPA can significantly shorten the time to results. As toehold switches usually interact with RNA, we attached a T7 promoter sequence to the forward primer to enable transcription for the amplicons. Through the incorporation of RPA, we showed the limit of detection of toehold switch assay is between 50 fM – 50 aM (**Figure 3.5.4**). The colorimetric response of the reactions was recorded in **Figure 3.5.5.**

3.2.2 Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment

Norovirus is highly contagious and can easily cause outbreaks within communities. Deployable diagnostics are conducive to combating these viruses. We reasoned that toehold-switch-based diagnostics that examine the RNA of norovirus would be an accurate and convenient approach for norovirus detection.

A similar computerized design and experimental screening process have been adopted to obtain toehold switches targeting norovirus RNA. We also systematically explored NASBA and RT-RPA as options for isothermal amplification (**Figure 3.5.6**). Compared to the previous work, we made two improvements to the procedure mentioned I will mention below.

First, we used lacZ-α as the reporter in place of full-length β-galactosidase (or lacZ). lacZ-α is a small fragment of lacZ which can self-assemble with its complement lacZ- ω to reconstitute a functional full-length lacZ[53]. As lacZ-α is one-tenth the size of the full-length lacZ, we hypothesized the speed and yield of transcription and translation of the small lacZ-α would increase in this resource-limited cell-free expression environment. lacZ-ω protein will be pre-purified and supplied as an additive. To test this hypothesis, we performed a comparison test using toehold switches ligated with a full-length lacZ reporter and a lacZ-α reporter, respectively (**Figure 3.5.7**). It can be seen that when lacZ-α is used as the reporter for toehold switches, the activation speed is significantly improved.

Second, we employed a synbody-based viral enrichment method to improve the overall sensitivity of the detection assay (**Figure 3.5.8A**). Synbody is biotin-labeled synthetic peptide affinity ligands with similar affinity and specificity to antibodies[54,55]. Better than antibodies, synbodies have broad cross-affinity for multiple norovirus genotypes. As a method to capture and concentrate norovirus for better detection sensitivity, we used biotin-labelled synbody ASU1052 which has been developed for binding to multiple norovirus strains to concentrate the norovirus from stool samples. After the concentration, we performed amplification and cell-free reactions. The results (**Figure 3.5.8B-C**) show that samples that undergone synbody enrichment show significantly improved sensitivity.

3.2.4 Interfacing Gel Switch Resonators with Cell-Free Toehold Switches for SARS-CoV-2 detection

Diagnostics have been a central method to combat the COVID-19 pandemic, yet decentralized, low-cost and point-of-care diagnostics are still needed. With the current limitations of nucleic-acid-based tests, we sought to create a test that could achieve the following: (1) sampling of the virus at home without PPE; (2) mailing in a sealed envelope and (3) reading at a centralized location through the sealed envelope without making contact to samples. Toehold switches function as a highly specific and programmable nucleic-acid sensor, allowing us to quickly develop a sensing platform targeting SARS-CoV-2 RNA. The gene being regulated by toehold switches can be simply changed to meet different needs. A schematic of toehold switches interfacing with gel switch resonators is depicted in **Figure 3.5.9.**

To further test the hypothesis, toehold switches were designed with specificity to the SARS-CoV-2 N-gene. Linear DNA constructs harboring the toehold sequences were initially screened in PURExpress cell-free expression system using LacZ as a reporter gene, which catalyzes the cleavage of a yellow substrate termed chlorophenol red-β-D-galactopyranoside and produces a purple product (**Figure 3.5.10a**). This initial screening identified which switches provided the strongest response while maintaining a low background signal. Of the 11 switches tested, the N-gene-S1 switch is able to quickly activate gene expression upon trigger activation and provided an ON/OFF ratio above 40; therefore, the S1 toehold switch was selected for further testing (**Figure 3.5.10b**). Since the resonators work by detecting the degradation of gelatin, the next step was to identify a protease that could both efficiently degrade the gelatin substrate and be expressed using CFE. We used a protease called subtilisin, which degrades gelatin and causes a resonator frequency swiftly measurable by the designated reader.

Although the platform described here focuses on the detection of SARS-CoV-2, the approach can be adapted for the detection of other pathogens in general. The following steps can be taken to adapt the toehold switch-GSR platform to target other viruses. First, identify the RNA/DNA sequences specific to the targets of interest from published literature or sequence databases. Second, use these sequences to generate target-specific toehold switches using NUPACK-based algorithms. Third, assemble the toehold switches with fluorogenic or chromogenic reporters for rapid empirical validation using cell-free systems. Based on the performance, functional toehold switches can then be selected and assembled with the appropriate hydrolase reporter (e.g., subtilisin) sequences to interface with gel switch resonators. The toehold sensor development stage only takes approximately 5 days and therefore allows a timely response to new pathogens once their sequences are known. The new toehold sequences would be formulated and dried on the cards, and they would be ready for shipment. The extent of toehold switch activation is correlated with the amount of trigger RNA present in the reaction, and this must be amplified from low viral copy counts to provide enough sensitivity.

Amplification of trigger RNA was done using RT-RPA as it is an isothermal amplification method that can theoretically be performed at room temperature, or for better results, at body temperature[56]. It is important to ensure that the RPA reaction produces enough cDNA at a fast enough rate to support downstream reverse transcription of trigger RNA. Multiple combinations of forward and reverse primers for toehold switches encoding N-gene with S1 trigger sequence were designed in silico and screened for their efficiency. RT-RPA was conducted for 10 min at 30 °C before being heat-inactivated and run on a gel. The primer combinations resulting in the brightest bands were determined to be the most efficient because of their ability to produce amplified cDNA quickly (data not shown). These primers were used in all subsequent experiments. This unit operation would need to be simplified for a practical mail-in platform as, in the current state, it requires a combination of several reagents in a tube off the card and subsequent incubation at a set temperature. For optimal workflow as a simple mail-in platform, the virus lysis and RNA amplification should be incorporated into the paper platform, where only the liquid sample is applied and incubated at room temperature.18 The switches containing SBT(n) were expressed in PURExpress for initial proof of concept studies. These studies used toehold switches encoded as linear expression templates and expressed using the PURExpress kit.

The GSR card was tested through the mailing system. The sensor was initially placed in an envelope in Ames, IA and scanned before being shipped overnight to collaborators in Toronto, ON. There a cell-free reaction was placed on top of the GSR card in an incubator according to previously mentioned protocols for gel degradation experiments. After the test was complete the sensors were sealed in an envelope and shipped to be scanned in Ames, IA inside of sealed envelope, **Figure 3.5.11a**. The results demonstrated the ability to use the GSR card as a mail-in testing platform utilizing existing mailing infrastructure.

Viral samples from different coronaviruses were tested to ascertain the toehold switches selectivity towards SARS-CoV-2. Two samples were from separate SARS-CoV-2 isolates: 2019-nCoV/USA-WA1/2020 from cell lysate and Hong Kong/VM20001061/2020 from cell culture fluid. Samples were purified and subject to RT-RPA using primers for the N C7 switch followed by in vitro transcription. These transcripts were added to in-house CFE reactions as triggers. After incubation, the CFE reactions were transferred to the GSR cards. The GSR cards with toehold switch were able to detect SARS-CoV-2 virus strains (HK and WA) over SARS-CoV-1 and seasonal coronaviruses (**Figure 3.5.11b**).

3.2.5 Detection of SARS-CoV-2 with LAMP amplicons

Of all isothermal amplification reactions, loop-mediated isothermal amplification (LAMP)[57] has been used by a lot of point-of-care SARS-CoV-2 nucleic-acid diagnostics[58–60] due to its high specificity and reliability. Operating at 60-65 °C, LAMP uses 4-6 primers recognizing 6-8 distinct regions for a highly specific binding and amplification of target sequences. The specially designed primers form “loop structure”, allowing exponential extension to form a long DNA product containing repetitive target sequences that are suitable for a variety of downstream applications.

To integrate LAMP assay for toehold-switch diagnostics, one challenge is that the toehold switches are previously designed to target linear ssRNA. Although, in theory, it can work with ssDNA, it has not been demonstrated before. Plus, the ssDNA region in the dump-bell shape LAMP reaction amplificons is within the loop region, which might be difficult for toehold switches to bind.

To design toehold switches that would work for the ssDNA loop region of LAMP products, we extended the toehold domain used to initiate strand-displacement reactions. This change should encourage the binding between ssDNA target and the toehold switch RNA. To increase the likelihood of getting a functional switch, we designed a total of 142 switches targeting SARS-CoV-2 Spike, Nucleocapsid, Open Reading Frame 1b (ORF1b), as well as human 18s rRNA and β-actin mRNA as sample control. A schematic workflow is described in **Figure 3.5.13**. The screening (**Figure 3.5.14**) has yielded switches that can reach an ON/OFF ratio above 10. Within them, we selected a top-performing SARS-CoV-2 Spike toehold switch (D7) and a human 18s rRNA toehold switch (G8) for further characterization.

Next, we validated these switches against LAMP amplicons to see how they react with long, structured DNA containing the targets. Using the different concentrations of input RNA, we determined the sensitivity of the D7 switch to be between 25 aM-2.5 aM. The G8 switch can detect down to 40 aM 18s rRNA control (**Figure 3.5.15**).

3.3 Discussion

In this chapter, I discussed the application of toehold-switch-based diagnostics in the detection of I describe the detection of fungi Coccidioides, norovirus, SARS-CoV-2, and pathogens that are impacting and concerning society. In the detection of fungi Coccidioides, we used RT-RPA as an alternative amplification method to increase the reaction speed and sensitivity of the assay. For norovirus detection, we applied synbody-based viral enrichment to concentrate the virus, which increases the overall sensitivity of the assay. In addition, we use lacZα in place of full-length lacZ, which results in an increase in protein expression. At last, we demonstrate the detection of SARS-CoV-2 using toehold switches as the sensors and gel switch resonators as the reporter. The combination allows contact-free testing, which has the potential to increase testing privacy and save consumables.

Future advancements are required to address the following challenges. First, in our development of valley fever diagnostics, we have yet to be successful in applying our valley fever diagnostics in clinical samples. We have tried using deactivated serum samples from infected patients as targets but were unable to identify the pathogen after isothermal amplification and cell-free reactions. This is likely due to the fact that the fungi Coccidioides reside mainly in respiratory tracts and seldom enter the blood, resulting in low concentrations of targeting materials. Also, the serum contains lots of other molecules that could impact the reaction. We could potentially sample sputum and use it to detect pathogens.

In the norovirus detection assay, although we have demonstrated sufficient sensitivity for detection from clinical samples, the sample-to-result time is from 3 to 6 hours, which is substantially longer than many other diagnostic methods. The extended assay time is a combined result of the current synbody-based enrichment technique, isothermal amplification and cell-free reactions. Further optimization of the enrichment methods, designing toehold switches that can turn on quicker and stronger and implementing more efficient isothermal amplification techniques could be some of the solutions to the time issue.

For SARS-CoV-2 detection with gel switch resonators, the detection method has the potential to be used in a mail-in format to enable a more convenient and safe collection of epidemiological data. However, to enable that, further efforts to adapt viral sampling, RNA extraction, and isothermal amplification for decentralized applications are required. The implementation of lysate-based extracts will further lower the cost which is essential for point-of-care diagnostics.

For LAMP-activated toehold switches, they can enable high signal activation for the detection of SARS-CoV-2; however, they lack the ability to distinguish among the variants which share highly similar sequences. This issue is addressed by the technique developed in Chapter 4.

3.4 Materials and Methods

Computational design of toehold switches

An updated version of the selection algorithm described previously[9] was used to identify toehold switches. The algorithm facilitated selection of six promising designs from a set of over 100 candidate toehold switches generated from each target RNA. Candidate devices were designed to bind to a 36-nt continuous region of the target RNA. Putative toehold switches were generated at 1-nt increments along the target RNA and multiple ensemble defect levels were computed for each sensor based on its deviation from the ideal secondary structure of the toehold switch. Ensemble defects were calculated for the toehold switch 5’ end through to the 3’ end of the hairpin ($d_{\text{min\_sensor}}$), the toehold domain of the toehold switch (d_{toehold}), the binding site of the toehold switch within the target RNA ($d_{\text{binding\_site}}$), and the toehold switch region starting with the base immediately 3' of the target RNA binding site and extending 31 nts beyond the last base on the 3' end of the hairpin ($d_{\text{active\_sensor}}$). The parameter $d_{\text{active\_sensor}}$ was intended to provide a measure of any secondary structures in the activated toehold switch that could interfere with translation after binding to the target RNA.

In addition to ensemble defects, the equilibrium fraction f of target/toehold switch complexes in a system with equimolar concentrations of target and toehold switch RNAs was calculated as a measure of the affinity of the two RNAs. In practice, this parameter was almost always equal to 1. Designs that produced in-frame stop codons in the output gene were eliminated from further consideration. Each of the parameters was then normalized such that their maximum value across the set of putative designs for a given target RNA was equal to 1. These normalized parameters, designated by an overscore, were then inserted into a scoring function s:

$$s = 5\bar{d}_{toehold} + 4\bar{d}_{active\_sensor} + 2\bar{d}_{min\_sensor} + 2\bar{d}_{binding\_site} + (1 - f)$$

Toehold switches displaying the lowest values of s and screened to have $f > 0.9$ were selected for experimental testing.

Toehold switch plasmid construction

Plasmids and DNA templates for transcription were constructed using conventional molecular biology techniques. Synthetic DNA (Integrated DNA Technologies) encoding the toehold switch sensors was amplified by PCR and inserted into plasmids using Gibson assembly with 30-bp overlap regions as described previously. The sequences of the plasmids were confirmed using Sanger sequencing.

NASBA

Reaction buffer (Life Sciences, NECB-24; 33.5%), nucleotide mix (Life Sciences NECN-24; 16.5%), RNase inhibitor (Roche, 03335402001; 0.5%), 12.5 mM of each DNA primer (2%), nuclease free water (2.5%), and RNA amplicon (20%) were assembled at 4°C and incubated at 65°C for 2 min, followed by a 10-min incubation at 41°C. Enzyme Mix (Life Sciences NEC-1-24; 25%) was then added to the reaction (for a final volume of 5 µL), and the mixture was incubated at 41°C for 2 hr. The amplified product was then diluted 1:6 in water and applied to paper disks containing the cell-free system and DNA for the toehold switch.

RT-RPA

RT-RPA experiments used the commercial TwistAmp Basic RT kit (TwistDx). Reactions were prepared by combining 10 µM forward primer (4.8%), 10 µM reverse primer (4.8%), rehydration buffer, RNase Inhibitor (Roche, 03335402001; 4.4%), and RNA amplicon (22%) at room temperature and transferring the mixture to the freeze-dried reaction pellet. After mixing, 2.5 µL of 280 mM magnesium acetate (5%) was added to start the reaction and it was incubated at 41˚C for 5-7 minutes. The reaction tube was then inverted vigorously 8-10 times, spun down briefly, and returned to incubation at 41°C for 2 hr. The amplified product was then diluted 1:6 in water and applied to paper disks containing the cell-free system and DNA for the toehold switch.

LAMP

15 µL LAMP reactions were used for all reactions unless otherwise noticed and were assembled with the following components: 1.6 µM FIP and BIP, 0.2 µM F3 and B3, 0.4 µM LF and LB, 15 µL WarmStart LAMP 2X Master Mix (E1700, New England Biolabs), 12 µL target RNA or water. Reactions were incubated at 61 °C for 30 minutes.

Synbody enrichment

Synbody-based viral enrichment A 30-µL volume of MyOne Streptavidin C1 streptavidin-coated magnetic beads (Life Technologies, U.S.A.), corresponding to 2.1×10^8 to 3.6×10^8 total beads, was added to Protein LowBind tubes (Eppendorf, U.S.A.). The bead storage solution was removed and the beads were washed three times with 1 mL of PBST (0.05% Tween 20 in 1x phosphate-buffered saline). The beads were then blocked with 3% BSA in PBST overnight at 4°C. The following day, the beads were suspended in fresh 3% BSA in PBST and blocked for an additional 2 hours. The beads were then washed three times with PBST and suspended in 30 µL of 1x PBS (phosphate-buffered saline) to yield a final suspension of blocked magnetic beads.

A dilution series of virus particles ranging from $1{:}10^3$ to $1{:}10^7$ was prepared by first taking a 1-µL aliquot of a norovirus GII.4 Sydney positive stool sample and diluting it into 1 mL of PBS. The resulting $1{:}10^3$ sample was serially diluted by factors of ten into PBS to generate the rest of the dilution series. Biotin-labeled synbody ASU1052 was then added to a concentration of 1 µM into each diluted sample and incubated with shaking for 1 hour at room temperature. The solutions were then added to the blocked streptavidin-coated magnetic beads and shaken for an additional 15 minutes at room temperature. The beads were washed three times with PBST and one time with PBS and then suspended with 50 µL water. The beads were incubated for 2 min at 95°C to release the viral RNA for analysis. 50 µL of each stool dilution was also incubated for 2 minutes at 95°C and used for comparison.

Cell-free reactions

All experiments were performed using PURExpress in vitro protein synthesis kits (New England Biolabs, E6800L). A typical 5 µL liquid-phase cell-free reaction contains 2 µL of PURExpress Solution A (40%, v/v), 1.5 µL PURExpress Solution B (30%, v/v), 0.1µL of RNase inhibitor (40 U/µl, 03335402001, Roche), 0.8 µL of DNA amplicons. The rest of the volume was filled with ultrapure water. In situations where a lacZα reporter was used, pre-purified lacZω protein was supplied directly into the cell-free system at 1µM. After thoroughly mixing, 4µL of the cell-free reactions were then transferred to respective wells in a 384-well microplate. When using lacZα or full-length lacZ as reporters, chlorophenol red-β-D-galactopyranoside (10884308001, Roche) is used as the substrate at 0.6mg/mL, and OD575 was used for signal measurements.

For paper-based cell-free experiments, 2-mm paper disks (Whatman, 1442-042) were prepared with a disposable biopsy punch and were placed into the wells of a 384-well microplate. The cell-free reactions were scaled down to a total volume of 2 µL based on the above recipe and were applied to the paper disks respectively. Photographs were taken with a smartphone camera from the bottom of the plate. Pictures were taken using smartphones.

Gel switch mail-in tests

Before being tested with a sample, GSR chips were scanned in envelopes in custom-built scanner to get their baseline resonant frequency. The scanner consisted of a vector network analyzer (Copper Mountain TR1300), linear translation stage (THORLABS KMTS50E/M), and two-coil antenna/holder. The vector network analyzer was setup to scan between 1 and 250 MHz using 5000 points. The linear stage translated the chip 2.5 mm in 0.5 mm steps which was used to average out the noise of the signal due to misalignment between the reader and chip since the chip would be removed and replaced in the reader to be tested with a sample. The scanner saved the frequency sweep of the S11 scattering parameter at all five steps of the linear translation stage motion. Those five scans were then averaged and the average was fitted to a quadratic model which was used to determine the resonant frequency of the chip.

The chips were tested with either samples of bacterial protease (Carolina Biologic) or subtilisin produced in cell-free reactions (either from BL21 DE3 Star extract or using NEB PurExpress kit). For bacterial protease samples the enzyme cocktail was dilute with PBS buffer in a 1:10 mixture and cell-free reactions were always done at 10 μL reaction volumes diluted down with 20 μL of PBS for a working volume of 30 μL on chip. In each port on the GSR chip 30 μL of sample was placed using a pipette. The chip was then placed in a petri dish with a wetted Kim wipe to prevent sample loss due to evaporation. The samples were then incubated at 30 °C in for 2 hours. The chips were then taken out of their petri dishes, excess fluid was dabbed away from ports using Kim wipe and sealed in an envelope. To simulate overnight mailing time the sealed chips were scanned after 24 hours of incubation to analyze potential shifts in resonant frequency.

Coronavirus samples for GSR testing

All viral samples were obtained from BEI Resources/ATCC unless otherwise noted. Two heat inactivated SARS-CoV-2 samples were used for detection: heat inactivated culture fluid containing Hong Kong/VM20001061/2020 (ZeptoMextrix) and cell lysate containing 2019-nCoV/USA-WA1/2020. Other viral samples used to screen for selectivity included irradiated SARS-CoV-1 in PBS, genomic RNA from coronavirus OC43, genomic RNA from coronavirus NL63, and genomic RNA from coronavirus 229E.

GSR mail-in sensor fabrication

Materials for the GSR chip include printer paper, temperature tolerant PET (Melinex ST505) double stick tape (ATack clear), gelatin (Great Value), parafilm (Millipore Sigma), conductive ink (DP 5028), and solid ink. Resonant sensor were screen printed on printer paper using protocol20. The resonant sensor was adhered to a temperature tolerant PET using douple stick tape. Wax printed channels were printed using solid ink in a wax printer (Xerox ColorQube) and adhered on top of the resonator with a laser cut piece of double stick tape. The laser cut pattern consisted of cross cuts in the same pattern and aligned to be under the four channels of the wax printed paper. Gelatin 10 wt% was mixed with double deionized water and mixed in microwave until dissolved and cast in petri dish. Cured gelatin was then cut using a spatula and placed in the edges of the four wax printed channels on top of the assembled GSR chip. The gelatin was desiccated down under a vacuum for one hour. Finally, parafilm with ports laser cut out for the desiccated gelatin was adhered to the top of the GSR chip to prevent samples from spilling out over the desiccated gelatin during application.

3.5 Figures

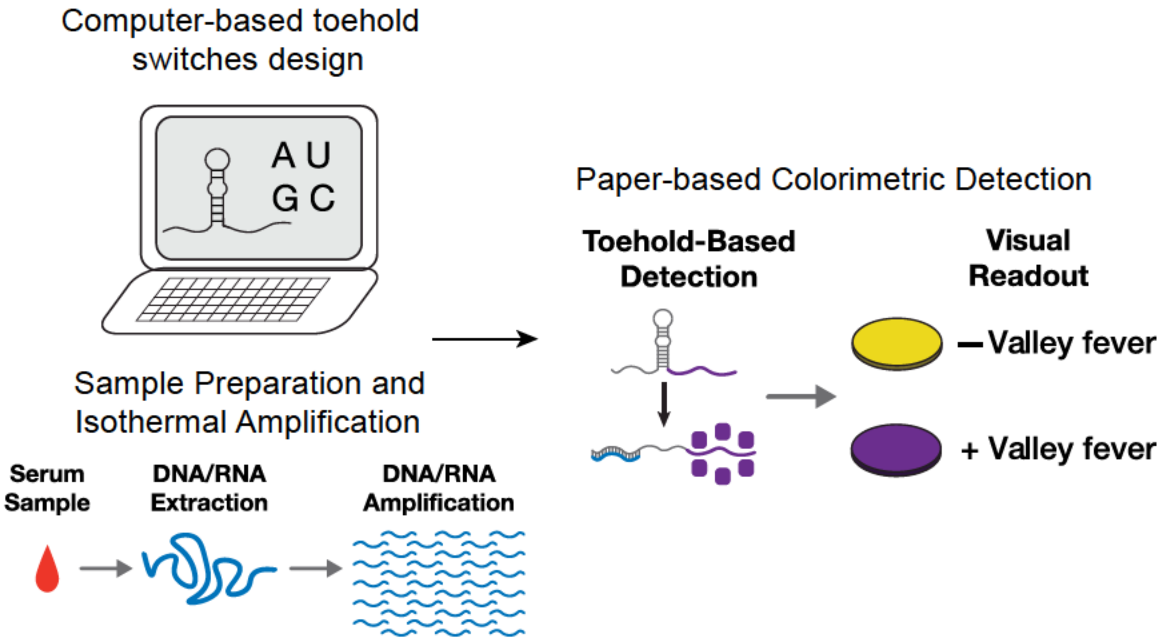

Computer-based toehold switches design
A U
G C
Sample Preparation and Isothermal Amplification
Serum Sample
DNA/RNA Extraction
DNA/RNA Amplification
Paper-based Colorimetric Detection
Toehold-Based Detection
Visual Readout
– Valley fever
+ Valley fever

Figure 3.5.1 Schematic of valley fever diagnostics using toehold switches with visual readout.

Valley fever diagnostics enabled by toehold switches consist of three steps: (1) computer-based toehold switches design; (2) sample preparation and isothermal amplification; (3) paper-based colorimetric detection by cell-free systems. With lacZ as the output gene, the cell-free reactions will turn purple when target pathogen DNA are detected. Without target pathogen DNA, the reactions will remain yellow.

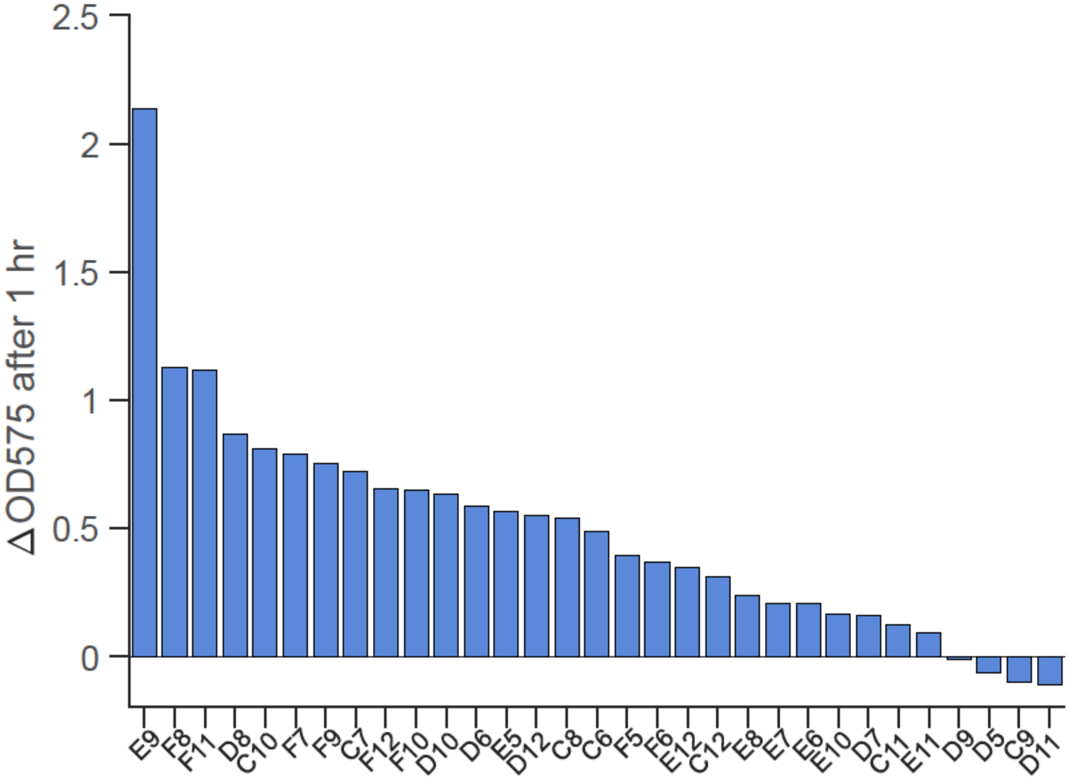


Figure 3.5.2 Screening of toehold switches against valley fever using cell-free systems with colorimetric output.

A total of 31 toehold switches were screened against valley fever pathogens. Bars represent the OD575 difference between 1 h and 0 h.

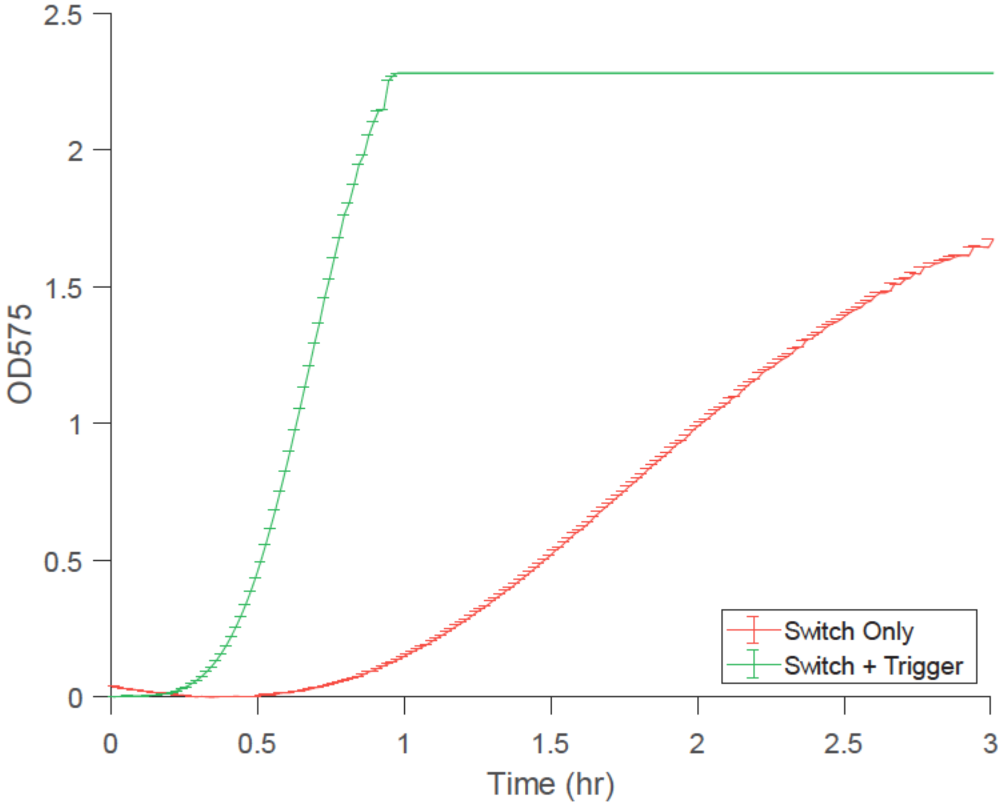


Figure 3.5.3 Time course measurement of the top performing toehold switch E9.

The green curve represents E9 switch with cognate trigger RNA. The red curve represents trigger E9 switch alone without triggers.

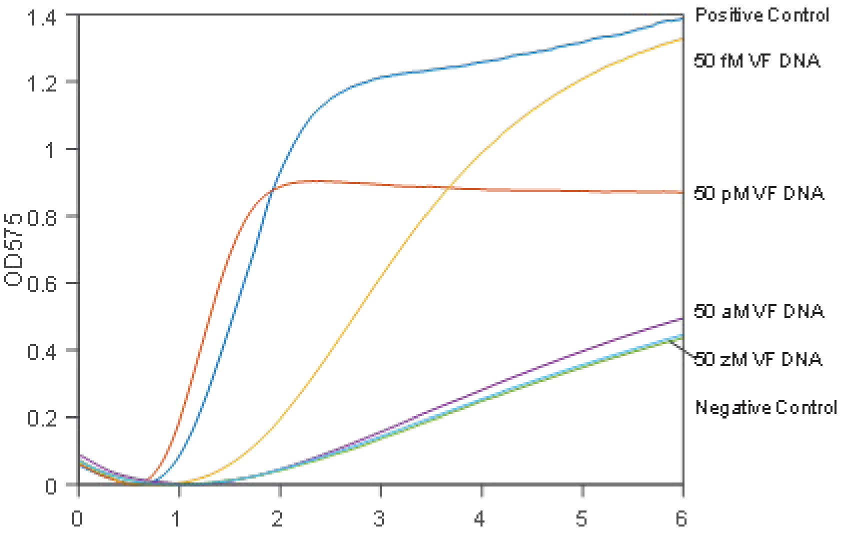


Figure 3.5.4 Time course measurement of OD575 in cell-free reactions from different concentrations.

Concentrations represent the input valley fever (VF) DNA template amount in RPA reactions. Curves represent mean value of triplicate measurements.

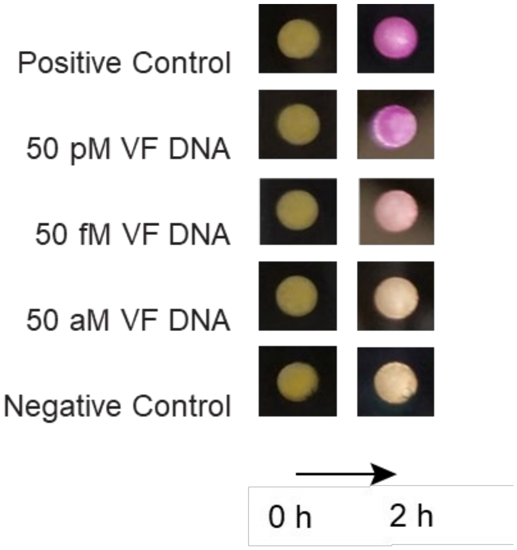

Positive Control
50 pM VF DNA
50 fM VF DNA
50 aM VF DNA
Negative Control
0 h
2 h

Figure 3.5.5 Color change photographs of valley fever toehold switch E9 captured at 0 hour and at 2 hours.

For input concentration at 50 pM and 50 fM, the color change from yellow to read can be visualized within 2 hours. Pictures were taken with smartphones.

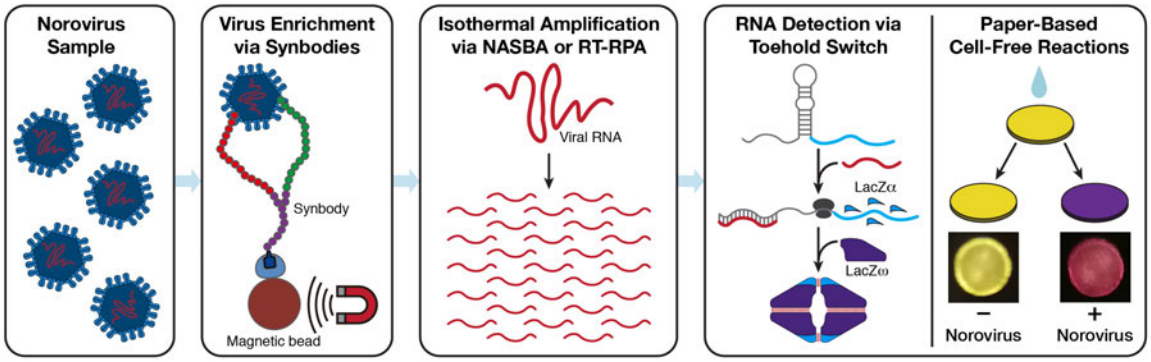

Norovirus Sample
Virus Enrichment via Synbodies
Synbody
Magnetic bead
Isothermal Amplification via NASBA or RT-RPA
Viral RNA
RNA Detection via Toehold Switch
LacZα
LacZω
Paper-Based Cell-Free Reactions
−
Norovirus
+
Norovirus

Figure 3.5.6. Overview of the norovirus detection assay using paper-based cell-free transcription-translation reactions.

A norovirus sample is first enriched using synbodies and viral RNA amplified isothermally using nucleic acid sequence-based amplification (NASBA) or reverse transcriptase recombinase polymerase amplification (RT-RPA). The amplified nucleic acids are added to paper-based cell-free reactions where norovirus RNAs are detected by sequence-specific toehold switches. The toehold switches generate the lacZα peptide, which produces a purple-colored product after complementation with lacZω. Samples positive for norovirus can be identified by their purple color following the assay.

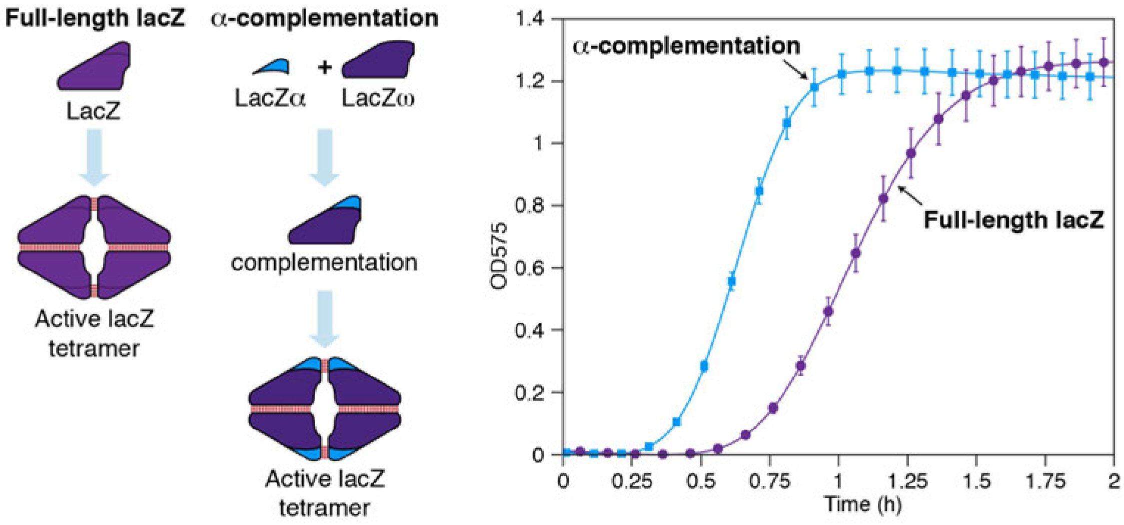


Figure 3.5.7 Schematic of full-length lacZ and α-complementation, and the time-course measurement of OD 575 using lacZα and full lacZ as the reporter, respectively.

The translated lacZ forms a homo-tetramer to execute catalytic function. In α-complementation, lacZ is split into lacZα and lacZω, both of which are non-functional on their own. When present together, they self-assemble into lacZ fragments, which can form homotetramers to regain catalytic activity. Through the use of short lacZα as a reporter for toehold switches, the speed of signal generation in the cell-free reactions is significantly improved. n=3 technical replicates; curves represent mean ± SD.

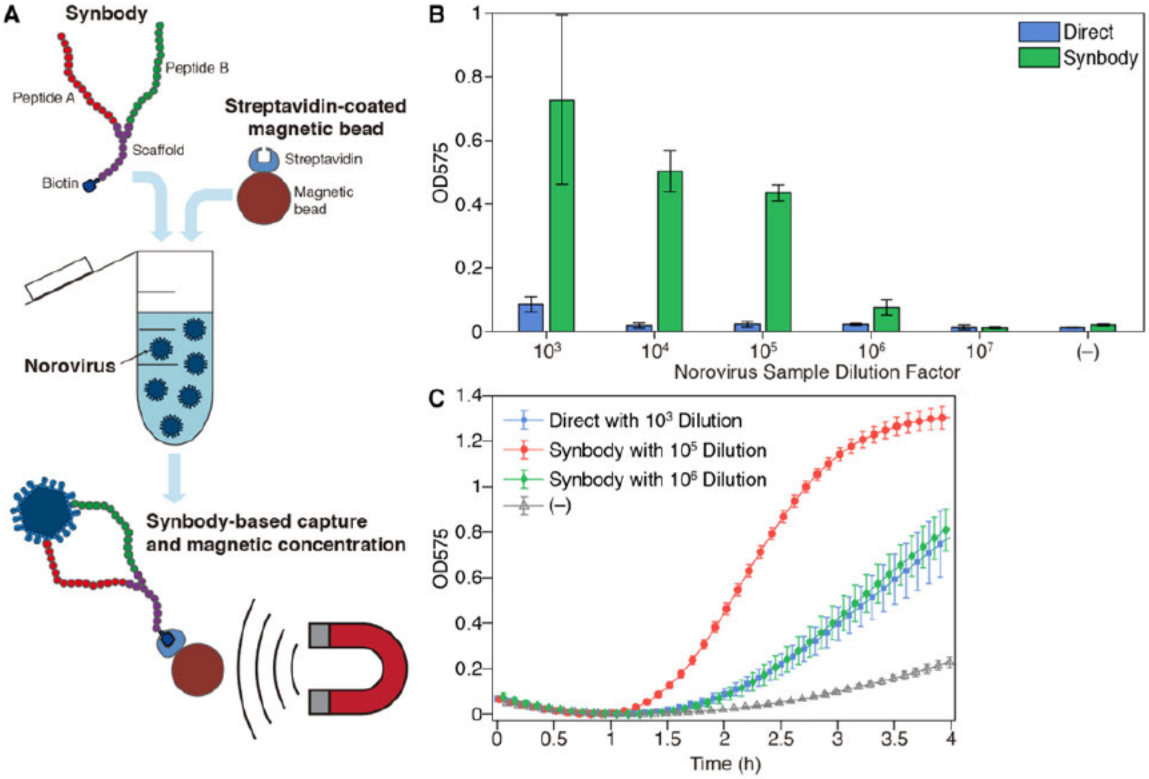


Figure 3.5.8. Implementation of a synbody-based capture and concentration method for norovirus detection.

A, Illustration of the synbody enrichment technique. Biotin-labelled synbodies engineered to recognize diverse norovirus genotypes are used to bind to virus particles in a dilute solution and are in turn captured by streptavidin-coated magnetic beads. Magnetic capture enables concentration of the captured norovirus particles. B, Measurement of OD575 after two-hour cell-free reactions with toehold switch S2. Samples subject to synbody-based concentration and samples used directly without concentration were amplified by NASBA. The negative control (–) is a water-only sample. C, Time-course measurements of OD575 for synbody-concentrated samples compared to samples used directly. OD575 for a sample used directly after 1000-fold dilution is comparable to a concentrated sample initially diluted by 10^6-fold.

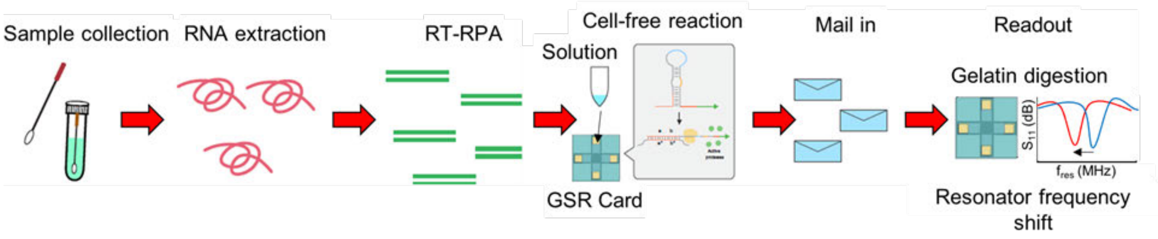

Sample collection
RNA extraction
RT-RPA
Cell-free reaction
Solution
GSR Card
Mail in
Readout
Gelatin digestion
S_{11} (dB)
f_{res} (MHz)
Resonator frequency shift

Figure 3.5.9. A gel-switch-resonator (GSR) test flow diagram.

The workflow starts with sample collection, extraction of viral RNA material, isothermal amplification via RT-RPA, cell-free reaction and sample application on GSR, mail-in test, and lastly readout monitoring frequency shifts.

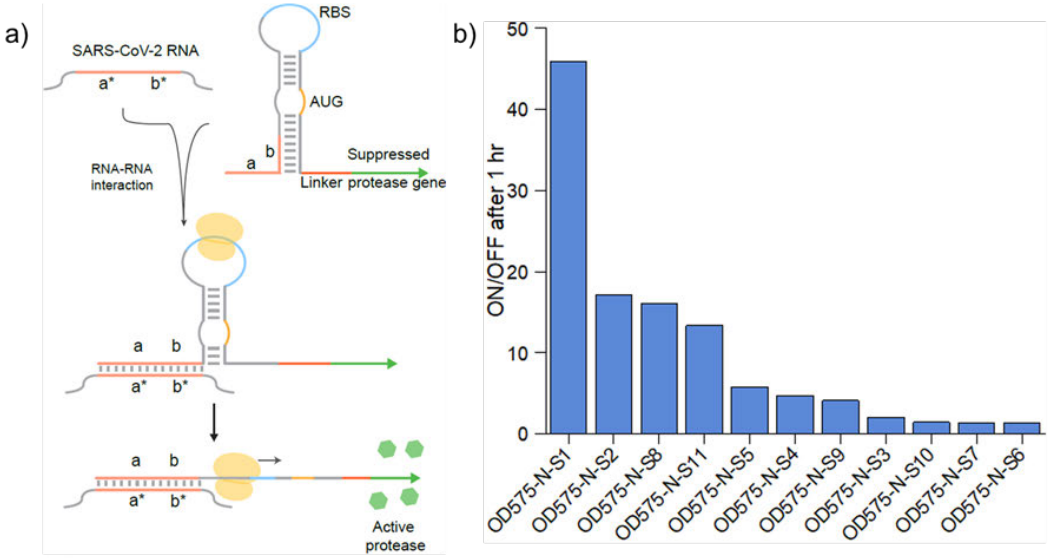


Figure 3.5.10 Toehold switch sensor screening and validation in cell-free systems.

a) Schematic diagram of toehold switch mechanism. Cognate SARS-CoV-2 RNA sequence binds to the toehold switch and exposes RBS and the start codon for translation initiation.

b) Toehold switch initial screening of designed toehold switch and trigger sequences. Candidate toehold switches were ranked by fluorescence ON/OFF. Bars represent fluorescence ratio between ON and OFF.

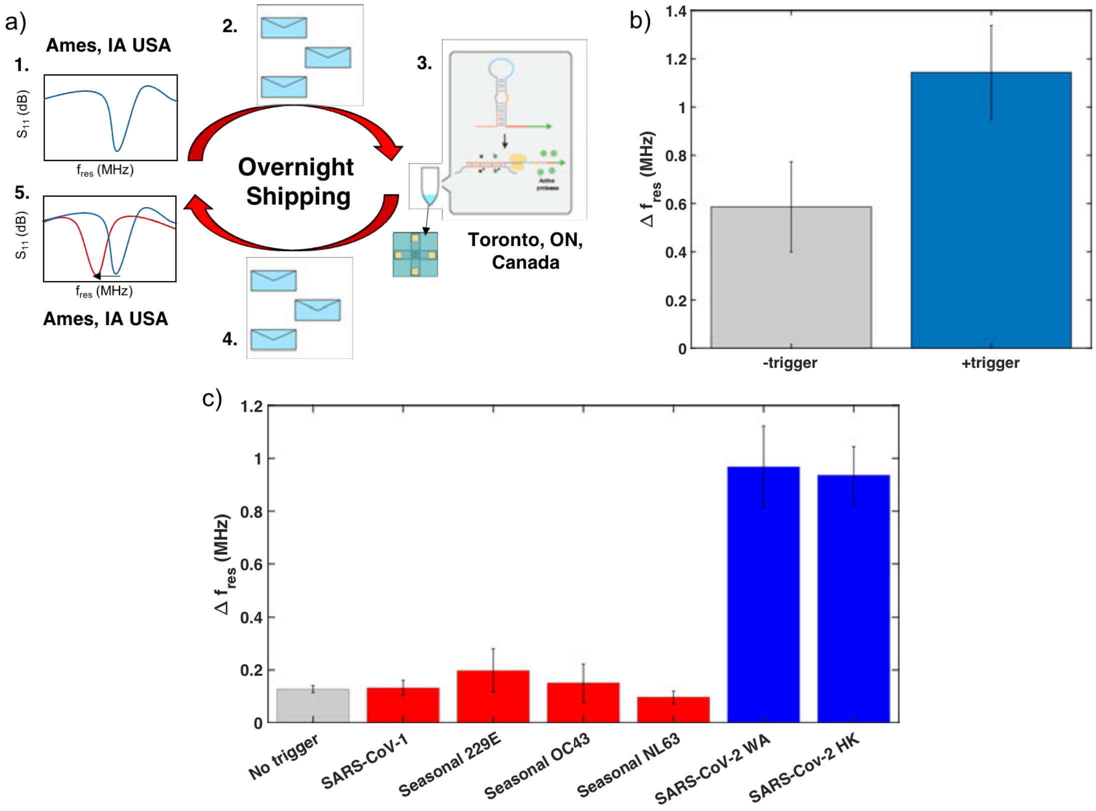


Figure 3.5.11. Demonstrating mail-in SARS-CoV-2 detection with toehold switch and gel-switch resonator system.

a) Mail-in test using overnight shipping between Ames, IA and Toronto, ON. b) Frequency shift results from mail-in test. c) Selectivity of GSR card sensor to different coronavirus SARS (SARS-CoV-1), seasonal (229E, OC43, NL63), and SARS-CoV-2 (WA, HK).

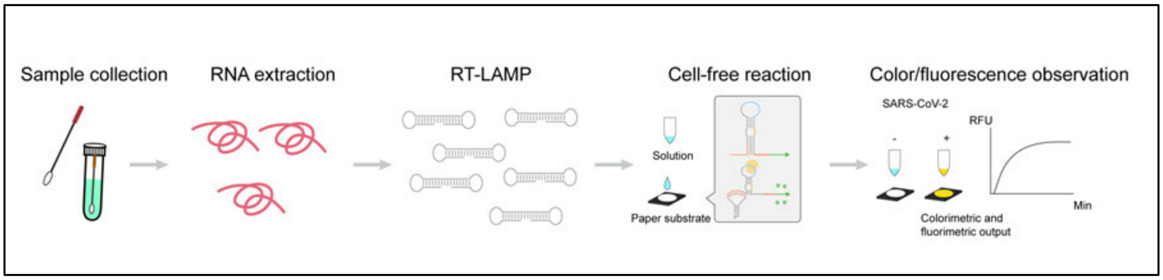

Sample collection
RNA extraction
RT-LAMP
Cell-free reaction
Color/fluorescence observation
Solution
Paper substrate
SARS-CoV-2
-
+
RFU
Min
Colorimetric and fluorimetric output

Figure 3.5.12 Schematic of toehold-switch cell-free detection with LAMP reactions.

The detection workflow consists of sample collection, RNA extraction, RT-LAMP reaction, cell-free reaction and readout by fluorometric or colorimetric.

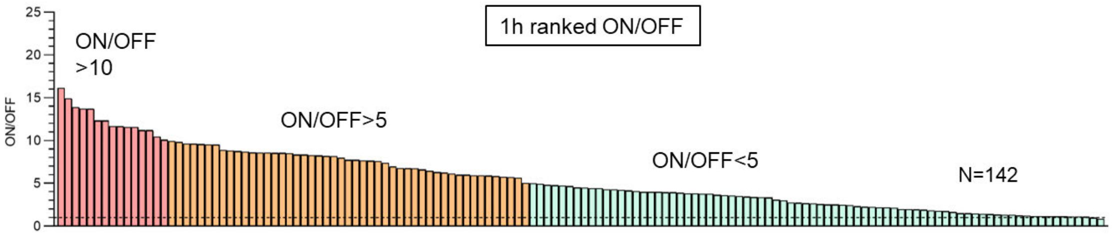


Figure 3.5.13 ON/OFF ratio of 142 toehold switches targeting ssDNA of LAMP products.

A total of 142 toehold switches were screened and plotted based on fluorescence ON/OFF. Bars represent the fluorescence ON/OFF in a singular experiment at 1 hour. Red: ON/OFF> 10; Orange: 10>ON/OFF>5; Cyan: ON/OFF<5.

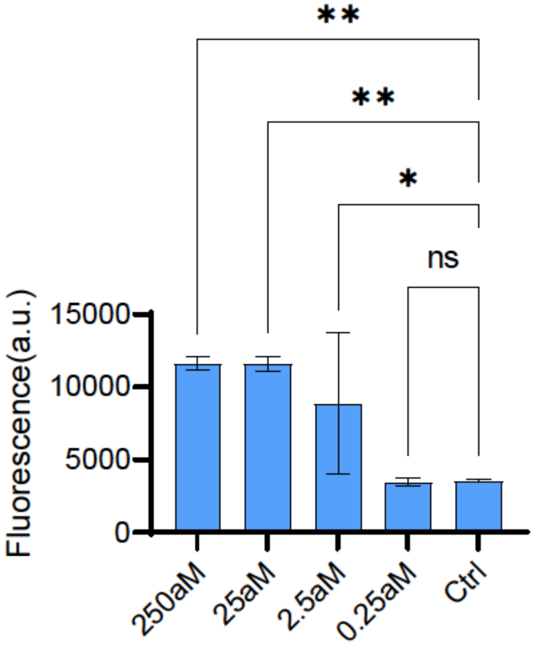


Figure 3.5.14. Sensitivity of D7 switch targeting the Spike gene.

Different concentrations of contrived SARS-CoV-2 viruses were used as templates for RT-RPA reactions. Fluorescence was measured at 2 hr for each reaction. Bars represent mean ± SD. Two-sample t test was performed between the control group and the experimental group. n=3.

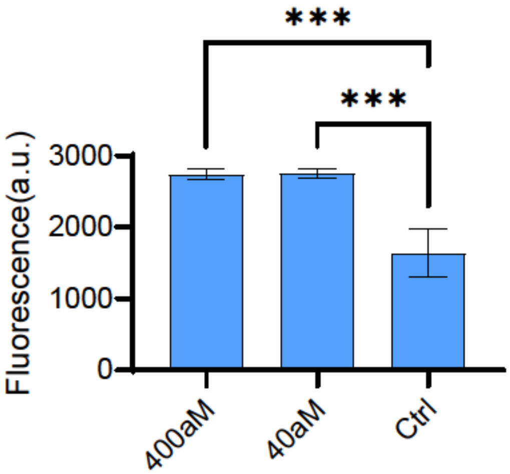


Figure 3.5.15. Sensitivity of G12 switch targeting the human 18s rRNA gene.

Different concentrations of human 18s rRNA were used as templates for RT-RPA reactions. Fluorescence was measured at 2 hr for each reaction. Bars represent mean ± SD. Two-sample t-tests were performed between the control groups within 1 hour. n=3.

CHAPTER 4 DEVELOPMENT AND APPLICATIONS OF SINGLE-NUCLEOTIDE-SPECIFIC RIBOREGULATORS FOR SARS-COV-2 VARIANT DETECTION

4.1 Abstract

COVID-19 has lasted for two and a half years and the emergence of variants and subvariants has put an uncertainty to the disease. Rapid nucleic acid diagnostics are central in monitoring and controlling the spread, yet well-established, field-deployable detection methods that can distinguish variants and subvariants are still limited. Here, we use ultraspecific riboregulators and cell-free protein synthesis to develop a deployable and affordable variant-specific SARS-CoV-2 diagnostic. The assay precisely detects Omicron, Delta, Gamma, Beta and Alpha variants and provide attomolar sensitivity with reverse transcription-recombinase polymerase amplification (RT-RPA). Assay validation against patient samples has shown 96.88 % sensitivity and 100% specificity (compared to sequencing) against the Omicron variant and 97.06% sensitivity and 100% specificity against the Delta variant. Lastly, we demonstrate a two-channel fluorescent assay to distinguish Omicron BA.1 and BA.2 subvariants. Riboregulator-based cell-free tests can be developed for new targets in less than a week and can potentially be deployed to low-resource areas for quick and affordable variant testing.

4.2 Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is an RNA virus causing the COVID-19 pandemic over the world starting in December 2019 and has led to over 6 million death as of August 2022[50]. The quick evolution of SARS-CoV-2 has led to multiple lineages with altered characteristics, some of which have been recognized as variants of concern (VOC) that have significant changes in transmissibility, virulence, ability of immune evasion, etc[61,62]. The first described VOC, alpha variant (B.1.1.7 or 501Y.V1), has shown to be 40-80% more infectious[15,63] than the original virus. The Beta (B.1.351 or 501.V2) variant followed has been associated with increases in hospitalizations and deaths in South Africa during the country's second wave[64] and decreases in vaccine efficacy[65–67]. The Gamma variant (P.1 or 501Y.V3), first documented in Brazil, has been estimated to be 1.7-2.4 times more transmissible than other local strains[68]. More recently, the Delta variant (B.1.617.2 or 478K.V1) and the Omicron variants have become the dominating strain at respective time points and have led to infection surges globally. Both variants can decrease vaccine effectiveness[69–71]. Omicron variants have by now further evolved into several sublineages, including BA.1, BA.2, BA.4, etc[72,73] and new ones are continuing to emerge[72]. These variants have or are likely to contribute to the repeated surges in COVID-19 cases. The continuous emergence of new variants also greatly challenges the development of effective vaccines and therapeutics, and the recovery of economy, and puts uncertainty into the fight against the pandemic.

Throughout the outbreak, diagnostics have proven to be an irreplaceable tool to fight the highly contractible disease and guard public health. To date, the quantitative polymerase chain reaction test (qPCR) remains the gold standard test for general COVID-19 diagnosis in clinical practice due to its accuracy and sensitivity. Antigen-based tests, on the other hand, have become another convenient alternative as they are fast, simple to use and can get results from home, but they have limited sensitivity and specificity. Different SARS-CoV-2 variants vary in infectivity, virulence, antigenicity that could impact epidemiological and risk assessment of the current pandemic, decisions on effective treatment and public health policy, making the detection of variants a new demand on diagnostic development. In addition, while sequencing[74,75] and qPCR[76,77] can both be used for variant detection, the requirements of trained personnel and expensive equipment limit their use to centralized laboratories. As a result, nucleic acid tests that are quick, low-cost and portable while carrying variant identification ability are highly preferred and needed for increasing the throughput and access of diagnostics, especially in resource-limited fields.

Recently, several fast and low-cost nucleic-acid tests with the ability to distinguish SARS-CoV-2 variants were developed, aiming to increase the accessibility and portability while retaining the accuracy and specificity of nucleic-acid tests. Among them, some diagnostics were developed using enzymes that can bridge nucleic-acid signal inputs and measurable signal outputs. For example, using the promiscuous cleavage ability of Cas13 and Cas12 enzymes to cleave ssRNA or ssDNA upon binding cognate nucleic acid targets, CRISPR/Cas proteins have been harnessed for nucleic-acid detection of different pathogens including SARS-CoV-2[58–60,78,79], and some of them have shown abilities to detect SARS-CoV-2 variants[78,80–84]. Also, SplintR ligases that ligate ssDNA splinted by complementary RNA strands were coupled with transcription and the fluorescent aptamer system for SARS-CoV-2 RNA variant detection[85]. In addition, a urease-based method[86] also provides a solution for variant detection with the use of a strand-displacing probe that releases silver ion, a urease inhibitor, upon recognizing the matching target. However, CRISPR-based diagnostics require a conserved protospacer adjacent motif closed to the target nucleic acid sequence, which could limit the choice of targets in some cases; also, few studies have systematically explored variant detection using CRISPR enzymes in clinical samples; additionally, multiplexing can be a challenge for CRISPR-based tests. Others are more of proof-of-concept studies with promises but require extra time for optimization and validation. So far, none of these rapid tests have demonstrated the ability to distinguish Omicron subvariants.

Aside from these nucleic-acid responsive enzymes, engineered riboregulator systems which function based on RNA-RNA interactions are also powerful tools for nucleic acid sensing[9–11,40]. These riboregulator systems activate/repress gene expression on the binding of RNA targets. Coupling freeze-dryable, paper-based cell-free systems[87], riboregulators demonstrated the ability to detect various pathogens, such as Ebola virus[37], Zika virus[19], human immunodeficiency viruses[10], norovirus[24], as well as gut bacteria[39] and gene mutations[40]. The diagnostic scheme is field-validated with Zika and Chikungunya virus[25]. Riboregulator-based cell-free diagnostics for general COVID-19 diagnosis are available[10,38,88]; however, the current strategies lack the ability to precisely identify variants and have not gone through systematic patient sample validation.

Here, we demonstrate the detection of SARS-CoV-2 variants through riboregulator-based diagnostics with systematic clinical sample validation. The single-nucleotide specific programmable riboregulators (SNIPRs) work in identifying Alpha, Beta, Gamma, Delta and Omicron variants. With the incorporation of RT-RPA, the assay can detect viral RNA at attomolar sensitivity and is specific to the target variant. When validated with clinical samples, the assay is 96.88% sensitive and 100% specific against Omicron variants and 97.06% sensitive and 100% specific against Delta variants. In light of the growing concern of Omicron subvariants, we developed a two-channel fluorescent assay for distinguishing Omicron BA.1 and BA.2 strains, demonstrating the multiplexing ability of SNIPR sensors. The streamlined and matured workflow of riboregulator-based cell-free diagnostics and automated design algorithms enable diagnostic development in less than a week with good target adaptability without sequence constraints. The quick-to-design, easy-to-use deployable and multiplexing features may increase the volume of variant screening and testing, especially for resource-limited areas with few centralized laboratories.

4.3 Results

4.3.1 Developing SARS-CoV-2 variant diagnostics with SNIPRs

Programmable riboregulators have been used for pathogen diagnostics, including Ebola virus[37], Zika[19] virus and norovirus[24]. The field validation of riboregulator-based diagnostic assays for Zika and Chikungunya virus provided 98.5% accuracies[25], which are equivalent to those of the real-time quantitative PCR, suggesting the potential use of the assay to provide quick and affordable diagnostic in the field of need. In response to the ongoing SARS-CoV-2 variant waves, we have devised a sensor development pipeline for accurate SARS-CoV-2 variant detection using SNIPRs. The sensor development process can be completed within a week when giving the sequence of the targeted variant

(**Figure 4.7.1a**). The diagnostic can be used to quickly obtain variant information with a few simple steps that could be completed without the need for expensive lab instruments (**Figure 4.7.1b**)

SNIPRs are one class of synthetic riboregulators that regulate gene expression based on the binding of trigger RNA and can discriminate the single-nucleotide difference between two RNAs (**Figure 4.7.2a**). In its original OFF state, SNIPRs form a hairpin that conceals the ribosome binding site and the start codon, preventing the downstream gene translation. The targeting RNA will trigger the SNIPRs to an ON state, where the hairpin will unwind, and the RBS and the start codon will be available for translation. The system operates near chemical equilibrium: the binding of a fully matched target RNA will lead to a slightly negative reaction free energy, promoting the transition from an OFF to an ON state; the unmatched trigger with at least a 1-nt difference will generate an energy penalty that results in a shift of reaction free energy from negative to positive and discourage the transition. In SARS-CoV-2 variant detection, we reasoned the distinct or characteristic mutations the variants carry could be used as targets for SNIPRs to distinguish the variant from one another.

The initial sensor selection process constitutes computational selection and empirical screening (**Figure 4.7.2b**). We started by examining the genomic sequence and identifying the characteristic mutations in the Spike gene for Omicron, Delta, Gamma, Beta and Alpha variants, respectively. Sequences containing the mutations were input to the SNIPR design software[40] to generate candidate sensors. The top candidate sensor sequences were synthesized and assembled with a reporter gene. Next, the performance of the SNIPR sensor sequences was screened empirically with cell-free systems with the presence of target variant RNA (ON state) or non-target variant RNA (OFF state). By ranking the fluorescence ON/OFF generated by the sensors over 2 hours, we were able to obtain a lead candidate for every variant (**Figure 4.7.2c-d**).

4.3.2 Integrating isothermal amplification for highly sensitive and specific variant detection

Nucleic acid amplification methods allow the target RNA to amplify, which increases the sensitivity of detection. Besides PCR, isothermal amplification methods[57,89,90] can allow quick and efficient amplification to occur at a constant temperature and are therefore more ideal for portable and point-of-care settings. Recombinase polymerase amplification[56] employs a recombinase to allow binding of the primers and DNA templates plus a strand-displacing DNA polymerase for amplification (**Figure 4.7.3a**). Compared to other isothermal amplification methods, RT-RPA reactions provide maximal amplification power at only 37-42 °C and can operate at room temperature[91], having lower temperature requirements than other methods like LAMP, HDA and NASBA. For this reason, we chose to explore the use of RT-RPA to amplify target RPA to increase target sensitivity for cell-free detections.

To test if RT-RPA is suited for amplification of SARS-CoV-2 RNA, we obeyed the workflow outlined in **Figure 4.7.3b**. First, we designed a set of RPA primers for Omicron, Delta, Gamma and Beta variants. As RPA generates double-stranded DNA products, a T7 promoter sequence was added to the forward primers to enable transcription into RNA that interacts with the SNIPRs. For each variant, we evaluated the performance of different primer pairs by measuring their amplification efficiency against the same amount of RNA templates with the same amplification time. The amplification efficiency is indicated by the fluorescence signals generated by the SNIPR cell-free reactions. We selected the primer pairs that generated the highest amount of fluorescence in a given cell-free reaction time. To further characterize the SNIPR assay integrated with an amplification step, we went on to test the sensitivity and specificity. When coupled with RT-RPA, we show that SNIPR assays for Omicron, Delta, Beta, Gamma variant detection all reached attomolar sensitivity (**Figure 4.7.3c-f**). Omicron variants contained many mutations, including K417N, which led to the activation of K417N sensors when Omicron variant RNA was present. Other SNIPR sensors are specific to their target variant and show clear discrimination of fluorescence signals between the target variant and other variants (**Figure 4.7.3g**). In addition, we showed that SNIPR detection platform is compatible with other isothermal amplification methods and can be adapted based on needs.

4.3.3 Clinical assessment of SNIPRs in the detection of variants

To date, sequencing remains the central method for discovering new variants, and a small number of qPCR tests are mutation-specific. While accurate, both methods suffered from high instrument and labor requirements. We sought to further validate and determine the performance of our SNIPR platform with clinical SARS-CoV-2 variant samples for potential applications in low-cost, deployable variant testing. We devised a workflow for clinical sample detection using the SNIPR platform on two most recent variants, Delta and Omicron (**Figure 4.7.4a**). The extracted samples were amplified using RT-RPA, followed by cell-free reaction using the validated sensors (**Figure 4.7.4b-c**). We also performed qRT-PCR in parallel to assess sample integrity and viral copy numbers. Samples with positive qRT-PCR results were regarded as valid samples and their results from SNIPR assays were further compared to the results from the previous sequencing. The SNIPR assay is 96.88% sensitive and 100% specific against Omicron variants and 97.06% sensitive and 100% specific against Delta variants (**Figure 4.7.4d**).

4.3.4 Rapid development of SNIPR platform for BA.1 and BA.2 distinction

Since its first appearance, the Omicron variant has evolved into multiple sublineages, including but not limited to BA.1, BA.2, BA.3, BA.4 and BA.5[72,73]. Diagnostics capable of distinguishing subvariants that have a highly similar sequence with altered infectivity and antigenicity are still lacking. To enable subvariant detection capacity, we sought to create a two-output fluorescent assay to distinguish BA.1 and BA.2. The previously developed Omicron (N969K) sensor can be used to identify Omicron variant, but this mutation is present in both BA.1 and BA.2, causing both targets to activate the sensor (**Figure 4.7.5a**). Therefore, we intended to use this sensor for general Omicron variant detection. To further distinguish BA.1 and BA.2, we selected another mutation G496S in the Spike gene that is present only in the BA.1 variant and followed the sensor development workflow to identify the lead sensor and RPA primers.

The lead G496S sensor showed a clear fluorescence difference between BA.1 and BA.2 (**Figure 4.7.5b**) with a sensitivity down to 0.4 aM.

To create an additional fluorescent channel for a multiplex SNIPR reaction, we looked into swapping the fluorescent reporter of the N969K sensor from GFP to mCherry with different strategies: (1) we preserved the first 21nt DNA sequence of the GFP gene of GFP and add the mCherry gene sequence directly after; (2) we deleted the GFP prefix sequence in the design and replaced with the full-length mCherry gene; (3) we deleted the GFP prefix sequence and replaced with a 21-nt linear linker followed by the mCherry DNA sequence. Of the three strategies, the first and the third design work is to provide red fluorescence signals with the presence of Omicron variants (**Figure 4.7.5c**). With two sensors targeting different mutations and generating different fluorescence in hand, we moved on to implementing a cell-free reaction with two channels for the detection of the BA.1 and BA.2 strain. The red fluorescence channel detects the N969K mutation in the Spike gene and functions to distinguish the Omicron variant from the others. The green fluorescence channel detects the G496S mutation in the Spike gene and functions to discriminate between the BA.1 and BA.2 strain. Together, for BA.1 subvariants, both the red fluorescence channel and green fluorescence channel will be activated; for BA.2 subvariants, only the red fluorescence channel will be activated; for non-omicron variants, neither channel will be activated (**Figure 4.7.5d**). We mixed two sensors in one cell-free reactions with BA.1, BA.2 and wild--type target added, respectively. The results showed that both green and red fluorescent channels were activated for BA.1 strain but only red fluorescence is activated for BA.2 strain (**Figure 4.7.5e**). Fluorescence remains minimal for wild-type virus, suggesting this one--pot two-channel cell-free reaction can interrogate more than one mutation at a time and enable subvariant detection.

4.4 Discussion

The emergence of the SARS-CoV-2 variant has altered the properties of the virus, causing the vaccines to be ineffective and contributing to more infections worldwide. Expandable diagnostics that can provide variant identity will help reduce the infection numbers and direct a more appropriate treatment. In this research, we developed a cell-free based nucleic-acid diagnostic for quick identification of SARS-CoV-2 variants. The SNIPR assay combines a 30-min RT-RPA followed by cell-free reactions and enables sensitive (attomolar) and specific detection of a wide range of SARS-CoV-2 variants of concern by targeting respective characteristic mutations, including Alpha, Beta, Gamma, Delta, and Omicron. We systematically validated the assay with patient samples and demonstrated 96.88% sensitivity and 100% specificity for the Omicron variant and 97.06% sensitivity and 100% specificity for the Delta variant. We expect the robustness and programmability of the assay could offer an alternative to qPCR and other variant detection methods to allow quick and effective screening of the emerging and new variants of concern.

We also devised a dual-channel fluorescence assay to achieve Omicron subvariant detection, demonstrating the potential of SNIPRs to create a multiplex diagnostic assay. This was conveniently achieved by swapping the downstream reporter gene of existing sensors without having to redesign the sensors for that specific gene. We demonstrated two fluorescent colors using GFP and mCherry, but conceivably, other fluorescent proteins (e.g., BFP, YFP) can also be integrated to allow multiple fluorescence outputs in a single cell-free reaction.

Through eight years of development and validation, riboregulator-based cell-free diagnostics have gradually matured and have grown to be promising strategies for quick, low-cost and portable diagnostics, as demonstrated in the detection of a diverse set of target pathogens; in this research, we contribute our efforts to stop the unprecedented pandemic by providing a diagnostic solution to SARS-CoV-2 by using SNIPR-based diagnostic. Like other riboregulator-based cell-free diagnostics, the SNIPR assay also benefits from its wide adaptability: first, the target for SNIPRs does not have a sequence constraint, which allows SNIPRs to be designed against diverse pathogens and mutations; second, the reactions can be freeze-dried and be hosted on paper discs[37,87] for transportation and storage needs; third, the ability to change the output scheme without redesigning the sensors may allow SNIPR sensors to interface with other compatible reporter systems, such as portable visible color[19], electrochemical circuits[92], glucometer[38] and resonators[88]. The publicly available design software and the streamlined procedure should also provide wider aaccess to the science community.

However, we do note that the clinical samples used in the validation were kit-extracted. Several research[60,85,93] has pointed out the possibility of using thermal or chemical lysis for sample preparation without the extraction step. These quick lysis methods can be tested with the SNIPR assays to further simplify the testing procedure. We also note that a nucleic acid amplification step is required to achieve maximal sensitivity. Although amplifications are often required for nucleic-acid tests, there are some novel strategies exploring nucleic-acid-amplification-free technologies[79,85] and achieving reasonable sensitivity. Future advances are needed to enable point-of-care diagnostics with cell-free systems.

4.5 Materials and Methods

In silico SNIPR design and DNA synthesis

SNIPR sensor DNA was designed using computer algorithms available online (https://github.com/Albert09111/SNIPR) and methods describedpreviously[41,94]. Briefly, the characteristic mutations for each variant were identified and selected (See Supplemental Information S1 for sequence and mutations used in the design process). The input target sequence was extracted as 200-bp up and downstream of the mutation, with the mutation placed in the center of the sequence. SNIPR sensors were designed to activate gene expression upon binding to the RNA carrying the mutations. At least 6 sensor sequences were selected for each mutation were selected. SNIPR sensors and primers were synthesized as DNA oligos from IDT (Integrated DNA Technologies, lnc).

SNIPR sensor assembly

SNIPR hairpins and reporters were assembled into full-length sensors using PCR. First, the DNA hairpins were amplified with PCR using universal PCR primers. Second, reporters (e.g., GFP or mCherry) containing overlapping sequences and the amplicons from the first step were added at a ratio of roughly 1:1 as templates and assembled by PCR into full-length sensors. Gel electrophoresis was used to validate the assembled PCR products. (See Supplemental Information S2 for primer sequence and the PCR protocol).

RT-RPA

TwistAmp Basic (TABAS03KIT, TwistDx) was used for all RT-RPA reactions. Synthetic SARS-CoV-2 RNA controls for the wild-type (Control 2, MN908947.3), Alpha (Control 15, EPI_ISL_601443), Beta (Control 16, EPI_ISL_678597), Gamma(Control 17, EPI_ISL_792683) Delta (Control 23, EPI_ISL_1544014), Omicron BA.1(Control 48, EPI_ISL_6841980), Omicron BA.2(Control 50, EPI_ISL_7190366) purchased from Twist Biosciences were used for RT-RPA primer screening, sensitivity and specificity testing. The RT-RPA reaction protocol was adapted from the manufacturer's protocol and described in detail in Supplemental S3. In brief, 29.5 µL rehydration buffer was added to resuspend lyophilized enzyme pellets, followed by the addition of 2.4 µL forward and reverse primers (10 µM), and 0.5 µL M-MuLV reverse transcriptase (M0253S, New England Biolabs). After the enzymes are follow dissolved, fully mix the reactions and distribute 10.45 µL to a new PCR strip. Then add 3.8 µL RNA template to the RPA reaction. At last, add 0.75 µL MgOAc to make a 15 µL RPA reaction and mix completely. The reactions were incubated at 42 °C for 30 minutes unless otherwise stated.

Cell-free reactions

All the cell-free reactions were performed using PURExpress In Vitro Protein Synthesis Kits (E6800L, New England Biolabs). To assemble a 5 µL cell-free reaction, 2 µL PURExpress Solution A (40% v/v) and 1.5 µL Solution B (30% v/v) were added based on the manufacturer's protocol. The RNase inhibitor was added to a concentration of 1U/µL and SNIPR DNA was added at a concentration from 5-30 ng/µL. For sensor screening, 0.5µL of unpurified PCR products were straightly added to the cell-free reactions. As for target RNA, 2 µM pre-transcribed synthetic RNA was added to the reactions. When coupling RT-RPA, 0.9 µL RPA amplicons were added to the cell-free reactions. Molecular biology grade water was used to bring the total reaction volume to 5 µL. After thoroughly mixing, 4 µL cell-free reactions were transferred to a 384-well black/clear bottom microplate (3542, Corning) and fluorescence was measured at 485 nm for excitation wavelength and 525 nm for emission wavelength for at least 2 hours.

Plasmid construction

Plasmid construction was performed using Gibson Assembly as previously described[9]. pUCIDT was used as the vector for SNIPR sensor plasmids and pET-15b was used as the vector for synthetic target plasmids. The universal primers for backbone and insert PCR primers were designed to contain 30-bp Gibson overlaps. Gibson assembly reactions were performed at 50 ° C for 30 minutes and transformed into E. coli DH5α competent cells. Sanger sequencing was used to identify transformants with correct sequences. The sensors and trigger plasmids are available in Addgene.

Experimental Design

In the screening of SNIPR sensors and integration of RT-RPA, synthetic SARS-CoV-2 RNA controls were used as the targets and experiments were replicated technically at least three times. The means and standard deviation of replicates were used for graphing. In the clinical sample tests, at least two technical replicates were performed for each sample. The means and standard deviation of replicates were used for graphing.

Graphical and Statistical Analysis

Graphical and statistical analysis were done using Prism 9.4 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com. Unless noted otherwise, significance was determined using two-sample t-tests.

4.6 Acknowledgments

The authors would like to thank Zachary Ticktin for assistance in patient sample testing.

4.7 Figures

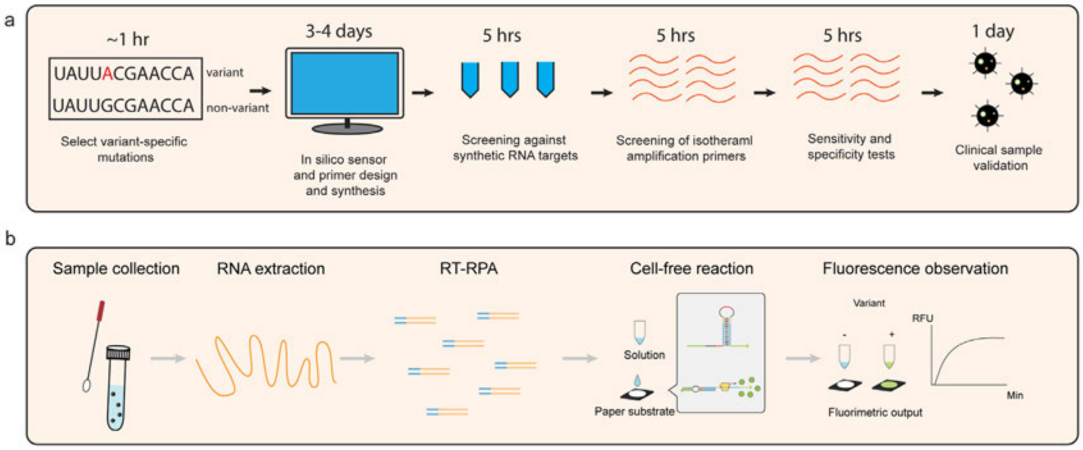

a
~1 hr
UAUUACGAACCA
variant
UAUUGCGAACCA
non-variant
Select variant-specific mutations
3-4 days
In silico sensor and primer design and synthesis
5 hrs
Screening against synthetic RNA targets
5 hrs
Screening of isotheraml amplification primers
5 hrs
Sensitivity and specificity tests
1 day
Clinical sample validation
b
Sample collection
RNA extraction
RT-RPA
Cell-free reaction
Solution
Paper substrate
Fluorescence observation
Variant
-
+
RFU
Min
Fluorimetric output

Figure 4.7.1 Schematic of the SNIPR-based SARS-CoV-2 variant detection assay.

a. SNIPR sensor development workflow for SARS-CoV-2 variant detection. b. Proposed diagnostic workflow for actual application of SNIPRs for SARS-CoV-2 variant detection.

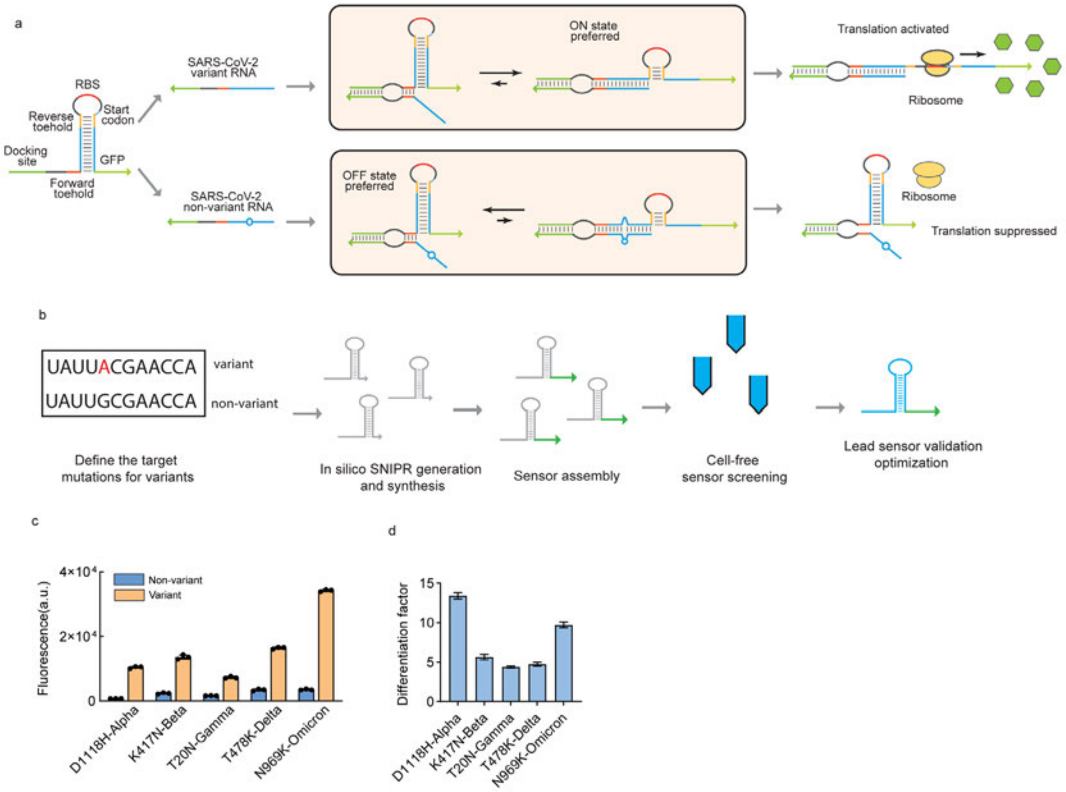

a
RBS
Reverse toehold
Start codon
Docking site
GFP
Forward toehold
SARS-CoV-2 variant RNA
SARS-CoV-2 non-variant RNA
ON state preferred
OFF state preferred
Translation activated
Ribosome
Ribosome
Translation suppressed
b
UAUUACGAACCA
UAUUGCGAACCA
variant
non-variant
Define the target mutations for variants
In silico SNIPR generation and synthesis
Sensor assembly
Cell-free sensor screening
Lead sensor validation optimization
c
Fluorescence(a.u.)
4×10⁴
2×10⁴
0
Non-variant
Variant
D1118H-Alpha
K417N-Beta
T20N-Gamma
T478K-Delta
N969K-Omicron
d
Differentiation factor
15
10
5
0
D1118H-Alpha
K417N-Beta
T20N-Gamma
T478K-Delta
N969K-Omicron

Figure 4.7.2 Design and screening of SNIPR sensors for variant identification.

a. Working principles of SNIPRs. SNIPRs contain a docking site (green) to facilitate its binding to target RNA and a hairpin that structurally conceals the ribosome binding site and the start codon to prevent downstream gene (GFP) translation. When target RNA binds, the energy balancing region containing forward (orange) and reverse (yellow) toehold regions enables target discrimination down to single-nucleotide difference. The perfectly matched target will preferably turn ON downstream GFP expression by binding and dissociating the SNIPR hairpins driven by negative free energy. Targets with mutations are unable to trigger the OFF to ON state conversion as a result of energy penalty. b. The procedure for design and screening SNIPRs for SARS-CoV-2 variants. c-

d. Fluorescence response (c) and the fluorescence ON/OFF (d) of SNIPRs targeting Omicron (N969K), Delta (T478K), Gamma (T20N), Beta (K417N), Alpha (D1118H). (n = 3 technical replicates; bars represent arithmetic mean ± SD in Fig. c and arithmetic mean ± SEM in Fig. d).

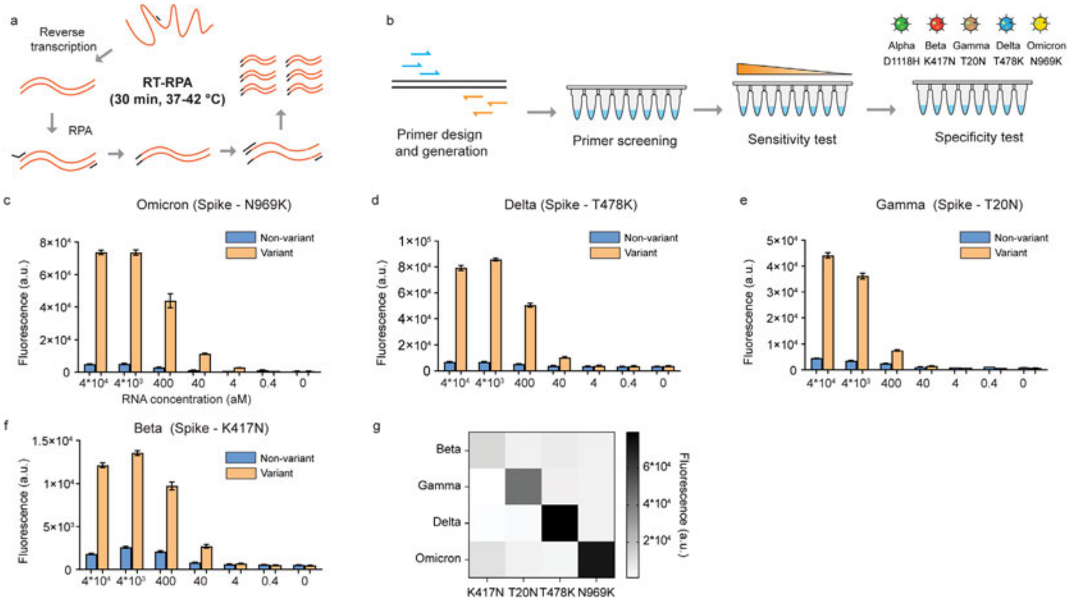
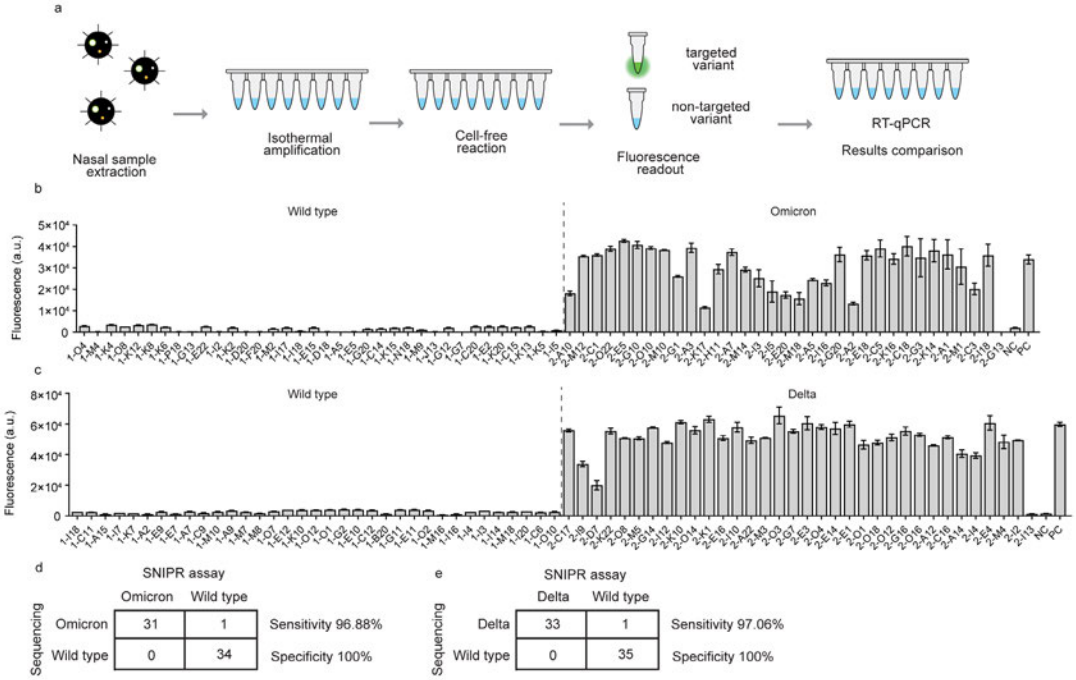


Figure 4.7.3 The integration of isothermal amplification for sensitive and specific detection.

a. The working principle of RT-RPA. Reverse transcription converts RNA to cDNA and recombinase enable primer-template binding. Strand-displacing DNA polymerase performed exponential amplification. b. The procedure of integrating and characterizing RT-RPA with SNIPR assays. c-f. Sensitivity of SNIPR assays for detection of the Omicron (c), Delta (d), Gamma (T20N), Beta (K417N) variant. (n = 3 technical replicates; bars represent arithmetic mean ± SD). g. Heat map depicting the fluorescence responses of sensors to their corresponding and non-corresponding variant targets. Each variant was amplified by RT-RPA and the amplicons were used for cell-free reactions with different SNIPR sensors.



	Omicron	Wild type	
Omicron	31	1	Sensitivity 96.88%
Wild type	0	34	Specificity 100%

	Delta	Wild type	
Delta	33	1	Sensitivity 97.06%
Wild type	0	35	Specificity 100%

Figure 4.7.4 Validation of SNIPR assays on patient samples.

a. Procedure for validating SNIPR assays with clinical samples. b. Fluorescence response for Omicron variants and wild type in patient samples using Omicron (N969K) sensor (n = 2 technical replicates; bars represent arithmetic mean ± SD). c. Fluorescence response for Delta variants and wild type in patient samples using Delta (T478K) (n = 2 technical replicates; bars represent arithmetic mean ± SD). d. Sensitivity and specificity table.

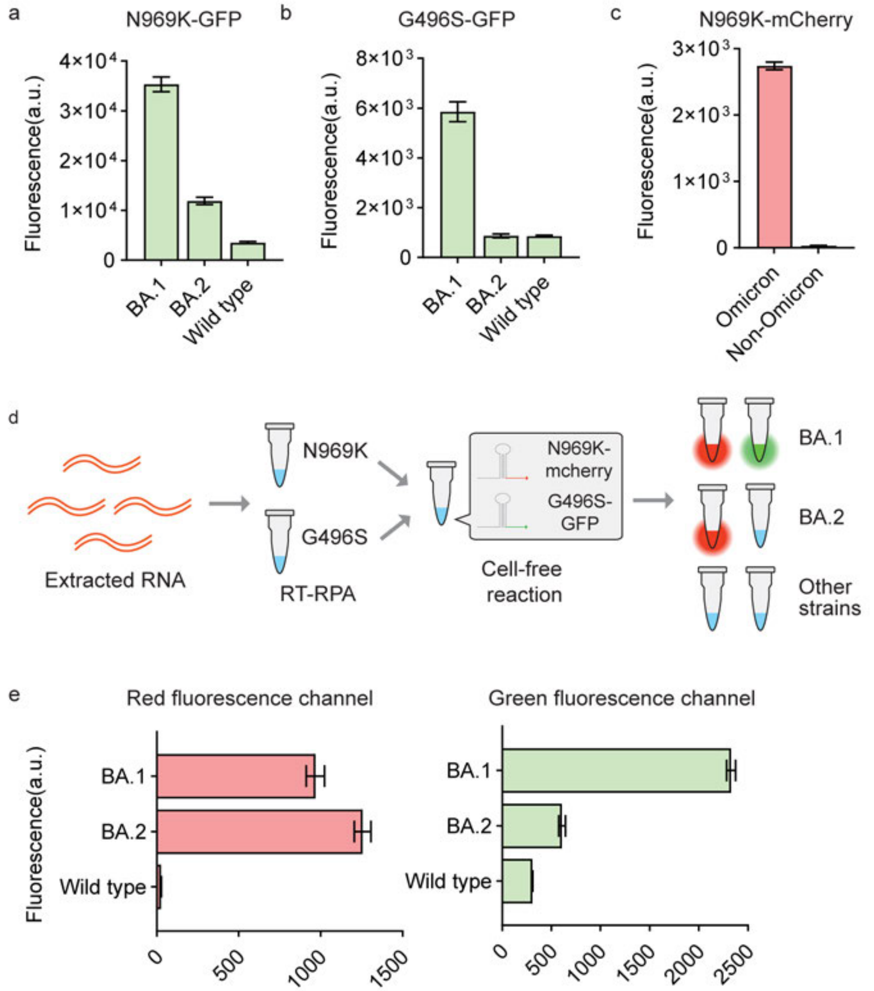


Figure 4.7.5 Development of dual-fluorescent SNIPR assays for distinguishing Omicron BA.1 and BA.2 subvariants.

a-b. Green fluorescence response of the sensors targeting N969K(a) and G496S(b) mutations against SARS-CoV-2 BA.1, BA.2 and wild type strain. (n = 3 technical replicates; bars represent arithmetic mean ± SD). c. Red fluorescence response of N969K sensor attached to a mCherry reporter. (n = 3 technical replicates; bars represent arithmetic mean ± SD). d. The workflow of one-pot two-channel detection for SARS-

CoV-2 Omicron BA.1 and BA.2 strain. e. Fluorescence responses for two-channel reactions.

CHAPTER 5 RAPID AND LOW-COST DETECTION OF MUTATIONS USING TRANSLATION BASED SEQUENCE VERIFICATION ASSAYS

5.1 Abstract

Nucleic acid detection provides accurate information to identify pathogens, direct treatment and manage diseases. In particular, subtle genetic changes often serve as signatures for diseases such as cancer. However, rapid, low-cost, instrument-free diagnostics with the ability to precisely distinguish genetic variations are still limited. Here, we report a rapid, low-cost, translation-based diagnostic platform for detecting mutations termed TRANslation of Sequence-LAbelled TRanscripts (TRANSLATR). We coupled short translation and reporting elements into primers to allow isothermal amplification and tagging simultaneously and employ cell-free translation machinery for fast fluorometric or colorimetric output. TRANSLATR enables facile detection of a wide range of frameshift, nonsense and missense mutations including those causing cancer and cystic fibrosis. The use of two split fluorescent protein reporters enables self-calibration of signals for stable output and allows visual discrimination of mutations by fluorescence difference. Moreover, a 38-residue LacZα peptide can be implemented along with a paper-based platform to provide colorimetric output. Lastly, TRANSLATR can operate at room temperature and is used to detect clinically relevant mutations in blood samples. TRANSLATR provides a novel strategy that has the potential to provide a more affordable, accessible and expandable diagnostic to the public.

5.2 Introduction

Translation is a key step in gene expression that allows the conversion of genotype to phenotype. During translation, ribosomes decipher genetic codes and produce corresponding amino acid sequences. When genetic mutations occur, expression patternsd can change and protein can lose functionality, causing diseases. Frameshift and nonsense mutations are two common types of mutations that cause the open reading frame to shift or terminate prematurely. These mutations are extremely likely to cause large-scale translation error and are therefore highly detrimental. In fact, within the two well-studied tumor suppressor gene BRCA1 and BRCA2 , frameshift mutations take up more than half of all the mutations, followed by nonsense mutations[95], both of which can deactivate the proteins and lead to cancer. Additionally, missense mutations and non-frameshift insertions or deletions in key residues are also contributors to diseases. Genetic mutations are causes to various diseases which are often inheritable. For example, in Ashkenazi Jews, the three hereditary founder frameshift mutations BRCA 185 delAG, BRCA1 5382 insC and BRCA 6174delT have a high population frequency accounting for breast-ovarian cancer[96]; cystic fibrosis is another example of inherited genetic diseases causing severe damages to lungs and other organs. Being able to precisely capture those mutations is critical for disease diagnosis, early intervention and treatment guidance. Nevertheless, basic diagnostics are still very unapproachable for many in the world. An assessment of ten countries across three continents completed by World Health Organization concluded only 1% of the health centers and clinics were considered to have full-service readiness for basic diagnostic tests[97]. Developing rapid, low-cost, expandable mutation diagnostic assay will further facilitate modern medicine and promote patient health globally.

Recent advances in cell-free synthetic biology have paved ways to rapid, low-cost and field-deployable diagnostics[23]. Using cell-free protein synthesis systems, transcription and translation can be conveniently carried out in test tubes[98]; the ability to lyophilize the reaction on paper substrates and reactivate with simple rehydration further increase their portability, stability and applicability[37,87]. Based on this, low-cost riboregulator-based nucleic-acid diagnostics have been developed for the detection of various pathogens, including Ebola virus[9], norovirus[24], Zika virus[19] and SARS-CoV-2[38]. Highly specific riboregulators capable of distinguishing single nucleotide polymorphism (SNP) can be used to detect genetic mutations[41]. However, the speed of the riboregulator-based cell-free diagnostic platform is, to some extent, hindered by strand-displacement reaction and the length of the transcripts.

Translation-free enzymatic cascade circuit systems have provided another avenue for cell-free diagnostics. Several CRISPR-associated proteins have been exploited for such purpose[28,59,99,100]. Specific High Sensitivity Enzymatic Reporter UnLOCKING (SHERLOCK) and DNA Endonuclease Targeted CRISPR Trans Reporter (DETECTR) use promiscuous cleavage activity of ssRNA and ssDNA for Cas13a and Cas12a, respectively, upon recognizing specific nucleic acid targets, to provide output signals. Such detection systems use mismatches between the crRNA and the target sequence to permit mutation detections and generate signals faster.

These emerging diagnostics have provided new opportunities for developing point-of-care testing for clinical applications. While these rapid, affordable diagnostics have great promise for implementation to low-infrastructure areas where resources are limited, there are still some challenges. For example, these assays operate at above room temperature (e.g., 37 °C- 65°C), necessitating the use of a heating device. Additionally, since most of the current technologies identify mutations based on nucleic acid base pairing and mismatches[41,99], the gene length it can examine at a time is limited.

Translation is an ideal way to assess and verify DNA sequences. About two decades ago, researchers have harnessed translation machinery to revealed genetic mutations by using PCR to amplify the gene, using in vitro translation to express the gene and gels to examining the proteins for mutation results[101]. However, instrument requirement and laborious procedures have limited its speed, portability and applications. Recent developments in isothermal amplification methods and freeze-dryable cell-free systems lower the expense, complexity of the procedures and requirement of instruments, which allow us to use translation for diagnostics more easily.

Here, we harnessed the power of cell-free translation for quick sequence-verification to create a rapid and low-cost nucleic acid diagnostic assay that specifically detects genetic mutations named (TRANSLATR). TRANSLATR uses *de novo* designed primers to append peptide reporters and gene expression elements to the target DNA/RNA sequence during amplification and exploits cell-free translation machinery to examine the amplicons and inform mutation status. We first used a short fluorescent tag GFP 11 as proof of concept for TRANSLATR and then demonstrated the detection of frameshift, nonsense and missense mutations using isothermal amplification methods like RPA, including those contributed to carcinogenesis, cystic fibrosis and HIV drug resistance. Use of a red fluorescent tag allows for the self-calibration of signals to improve the accuracy of results and convenient signal interpretation. Moreover, we showed that the use of a 38-amino-acid lacZα enables colorimetric reading and that TRANSLATR is compatible with a freeze-dried paper-based platform. Lastly, we showed the ability to operate TRANSLAR at room temperature and tested it against clinically relevant mutations in human blood samples. Our results indicate a rapid, low-cost translation-based sequence-verification platform against a variety of gene mutations. This method has the potential to provide a simple, deployable and affordable point-of-care diagnostic to the public.

5.3 Results

5.3.1 Implementing a Mutation-Specific Translation-Based Assay

With recent advances in the ability to make cell-free transcription-translation systems low cost, stable, and active in portable formats[37,87], we sought to develop an assay that harnessed the sequence-verification capabilities of translation itself to return a clear optical test result. Previous paper-based cell-free nucleic acid testing systems have relied on RNA-based strand-displacement reactions for sequence checking[9,41], which can increase the time to result and require complex nucleic acid design procedures to achieve robust point mutation detection. In principle, a purely translation-based detection method could avoid delays caused by RNA-RNA interactions while exploiting the inherent sequence-processing capacity of the translating ribosome.

Implementing the TRANSLATR assay required two essential elements. First, we made use of split protein reporter systems that enable the translation of short peptides (≤59 residues) to generate optical signals upon binding to a much larger protein complement. Several such split fluorescent proteins that provide strong signals have been reported recently for protein tagging in optical microscopies, such as GFP11[102] and sfCherry$_2$11[103,104] which complement GFP1-10 and sfCherry$_2$1-10 or sfCherry$_{3C}$(1-10), respectively. Moreover, truncated versions[105,106] of the widely used lacZ-α peptide[107] have also been found to complement the larger lacZ-ω fragment to generate colorimetric and fluorescent signals by cleaving appropriate reporter substrates. Second, since amplification is almost always required for the sensitive detection of DNA or RNA from clinical samples[19,24,27,100,108], we made use of extended amplification primers with lengths of up to 200 nucleotides to generate the templates for translation. Primers of this length can be readily prepared by commercial DNA synthesis and are sufficiently long to encode the necessary transcription and translation signals in the resulting amplicons.

The general composition of the amplification primers used for TRANSLATR assay is shown in **Figure 5.7.1**. A promoter site for T7 RNA polymerase is incorporated into the 5’ overhang, followed by a prokaryotic ribosome binding site (RBS) and a start codon sequence (**Figure 5.7.1a**). These overhang sequences thus provide signals to start transcription and translation. The reverse primer overhang consists of the reverse complement of the reporter peptide and an in-frame stop codon (**Figure 5.7.1b**). After transcription, these sequences generate the reporter signal and terminate translation. Based on the specific amplification strategy used, the 3’ forward and reverse primers can be designed following primer design criteria for PCR or recombinase polymerase amplification (RPA) to bind and specifically amplify the sequence of interest. Primers are checked for secondary structures to ensure effective binding to the template. Following amplification from a sample, transcription and translation of the resulting amplified product will only generate an output signal if the ribosome can translate through to the end of the amplicon, which contains the peptide reporter sequence (**Figure 5.7.1c**). Importantly, since the sequence between the start codon and peptide is generated from the template being amplified, the template must provide a functional open reading frame to ensure the reporter is produced.

Depending on the design of the primers, TRANSLATR enables the detection of mutations through two modalities (**Figure 5.7.1d**). The first modality can be broadly applied to identify frameshift and nonsense mutations, with the former causing shifting of the correct reading frame and the latter causing premature termination of translation. Under this modality, primers are designed to amplify the sequence containing an open reading frame of interest (ORFI). For instance, we can define that the signal turns ON when the wild-type sequence is present, ensuring that the downstream reporter will be in-frame after amplification of the wild-type. Therefore, when any frameshift or nonsense mutation is present in the amplicon, the correct reading frame will shift or terminate prematurely, disrupting the expression of downstream reporter peptide and leading to an OFF signal. In principle, this approach can also be designed to turn ON in the presence of a particular mutation: for frameshift mutations, we can design primers to have the frameshifted ORFI to be in-frame with the peptide reporter; for nonsense mutations, we can locate a stop codon present in the wild type but not the nonsense mutant, which can occur when the template is amplified in the antisense orientation or from a shifted reading frame.

For the second modality, primers are designed with more matches to the sequence of interest with the 3’ end of the forward primer positioned such that it coincides with the point mutation. Therefore, primers will selectively bind and extend along the target sequence but not the mutant, permitting efficient amplification and readout from samples bearing the sequence of interest. The second modality can be used to detect missense mutations and non-frameshift insertions or deletions, in addition to frameshift and nonsense mutations. Conceivably, the signals can be designed to turn ON for mutants by designing primers with exact matches to the mutant instead. Together, these two modalities allow TRANSLATR to cover most genetic mutations causing diseases.

We first investigated split fluorescent reporter protein systems developed for protein tagging and live-cell imaging[102–104,109] for use in TRANSLATR assays. These systems divide fluorescent reporters into individual fragments that are non-fluorescent. When both fragments are present, they spontaneously self-assemble to produce a detectable fluorescent signal. One fragment in the split pair is often relatively short (~16 residues) to facilitate tag addition and reduce the interference with the fusion protein. For example, by isolating the eleventh β−strand, green fluorescent protein (GFP) can be split into two fragments: GFP11 and GFP1-10, with GFP11 being a short, 16-amino-acid peptide[102,109] (**Figure 5.7.2a**). Importantly, we hypothesized that the short length of GFP11 and other split reporter proteins could enable us to incorporate the peptide sequence into amplification primers to provide a translation-based readout for nucleic acid testing.

To test whether the split peptides can function to discriminate mutations when present downstream of the sequence of interest, we first chose the split GFP system and began to investigate the ability of GFP11 peptides to discriminate human gene mutations under prokaryotic expression systems.

We employed the TRANSLATR primer design scheme (**Figure 5.7.2b**) to test the functionality of GFP11-embedded primers to discern BRCA mutations using PCR. Primers with the GFP11 tag should amplify and append the peptide sequence to both wild-type and mutant templates; however, only the wild type will translate GFP11 and generate a strong green fluorescence signal (**Figure 5.7.2c**). The primers specifically amplified both the wild-type and the mutant genes using PCR (**Figure 5.7.2d**). We observed the rapid and much stronger generation of green fluorescence for the wild-type amplicons versus the 6174delT mutants (**Figure 5.7.2e**). Notably, the fluorescence output from the reactions is sufficiently strong to be detected using a smartphone camera using blue-light illumination and an optical filter to exclude the excitation wavelengths. These results suggest GFP11 can be genetically encoded to primers and used as a reporter peptide for TRANSLATR to identify mutations.

By replacing the reporter peptide with two tandem sfCherry$_2$11 in the reverse primers and using its self-assembling complement sfCherry$_{3C}$(1-10) protein[104], we designed primers and showed the detection of mutations in BRCA genes with red fluorescence output (**Figure 5.7.2f**). This indicates that the strategy is generally applicable to other split fluorescent protein systems to enable orthogonal detection of multiple mutations.

5.3.2 Rapid, isothermal detection of mutations

While PCR provides high sensitivity, it is limited in terms of amplification time and the need for expensive thermal cycling equipment. The ability to use isothermal amplification, which allows for rapid nucleic acid amplification at constant temperatures, is crucial for rapid, low-cost diagnostics. RPA[89], which uses recombinases to enable binding between primers and DNA targets, is an attractive isothermal amplification method for DNA or RNA (**Figure 5.7.3a**) that permits rapid and sensitive amplification for diagnostics near human body temperature (37-42 °C) [24,41,99,110] and has even been shown to work at 25 °C[91].

To investigate if RPA can be integrated with the TRANSLATR approach, we followed RPA primer design criteria[89] and generated primers with 30- to 35-nt binding sites along with the required transcription and translation elements. The primers were incubated with the template in RPA reactions for 5 minutes. Then, we applied this approach to detect various pathogenic frameshift mutations across BRCA1 and BRCA2 genes and found that all groups tested show clear differences in signals between the wild-type and mutant samples (**Figure 5.7.3b**), indicating it is generally applicable to target different mutations. Additionally, by designing primers that bind to different regions, we can redefine the “correct” open reading frame and have the mutant sequence turn on the gene expression. A frameshift variant c.597dup in the SLC19A3 gene can cause thiamine metabolism dysfunction syndrome 2 (THMD2), a disease that is otherwise hard to diagnose without DNA sequence information[111]. Through rational primer design, we successfully generated strong fluorescence from the SLC19A3 c.597dup mutation while the wild-type signal remained low (**Figure 5.7.3c**).

Next, we tested TRANSLATR against nonsense mutations. Nonsense mutations in the amplicon stop translation and prevent formation of the reporter peptides, turning off fluorescence. We used this approach to target a pathogenic nonsense mutation c.4111C>T in the BRCA2 gene and successfully identified it (**Figure 5.7.3d**). Additionally, nonsense mutations are an important class of mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene that specifically lead to unstable or shortened RNA and therefore cause little to no production of the CFTR protein. Around 22% of those with cystic fibrosis will carry at least one mutation of this class. However, this type of mutation can be treated with readthrough compounds. Therefore, being able to detect mutations like this will the guide proper selection of drugs and promote precision medicine. Upon the presence of the nonsense mutations W1282X and G542X in the CFTR gene, fluorescence was significantly decreased (**Figure 5.7.3e**), a result of early translation termination. The nonsense mutation in another tumor suppressor gene RB1 also yielded a similar result (**Figure 5.7.7**).

Lastly, we tested TRANSLATR against missense mutations and non-frameshift insertions or deletions. Unlike frameshift and nonsense mutations, missense mutations and non-frameshift insertions or deletions only lead to the change of a codon but do not impact downstream translation. As a result, we employed the second detection modality of TRANSLATR. Existing literature has shown that 3' mismatches in the primers can severely impair the efficiency of RPA, resulting in little or no amplification[112]. We hypothesized that this amplification efficiency bias can be used to distinguish missense mutations. To test this approach, we positioned the mutation at the 3' end of the forward primers and designed primers to have more matches to wild-type sequences than the mutant sequences at the 3' extremity of the forward primers with positioned at the mutation location (**Figure 5.7.3g**). As a result, the amplification bias should lead to fluorescence signal differences upon amplicon translation. During our empirical testing, we found that one mismatch is usually not sufficient to generate substantial amplicon differences, so we manually introduce one or two additional mutations to the primers. Using the strategy, we successfully detected H63D in the HFE gene (**Figure 5.7.3h**), a common mutation associated with hemochromatosis[113]. In addition, ΔF508 is the most common mutation in the CFTR gene that is present in one or both alleles in approximately 90% of cystic fibrosis patients[114]. This mutation deletes a single codon but does not alter the downstream open reading frame. We designed the primers to have an exact match at the 3' end with the wild-type sequence but not the mutant, generating sufficient amplification efficiency difference to allow the detection of this mutation (**Figure 5.7.3i**). Together, these results indicate that TRANSLATR assay can be used to detect various types of mutations, including frameshift, nonsense, missense mutations and non-frameshift insertions or deletions with single-nucleotide specificity through rational design of primers and the sequence-reading properties of translation.

5.3.3 Implementing signal self-calibration, long-range sequence verification, and room temperature operation

Detection signals are sometimes influenced by factors such as sample quality, and nucleic acid amplification yield, which can potentially affect test accuracy. To counter these potential issues, we sought an additional TRANSLATR fluorescence signal to help quantify the amount of amplicon generated and serve as a normalization signal for the GFP11 reporter. We positioned a short fluorescent tag, sfCherry$_2$11, between the start codon and priming site in the forward primer (**Figure 5.7.4a**). By placing this red-fluorescent tag upstream of the ORFI, this strategy should provide a fluorescent signal that corresponds to the overall level of translation and is independent of any downstream translational termination, enabling system self-calibration (**Figure 5.7.4b**). Using the forward primers with sfCherry$_2$11 tags, we observed strong red fluorescence for both the wild type and the mutant in cell-free reactions. We also found that red fluorescence signals were similar between the mutant and wild-type samples, suggesting that total translation events and amplification yields are similar (**Figure 5.7.4c**). After normalizing the green fluorescence signals to the red fluorescence, we observed a much stronger calibrated signal from the wild-type sample compared to the mutant samples, as expected (**Figure 5.7.4d-e**). Using this two-color self-calibration approach, it is possible to achieve more accurate and quantitative measurements that can be used to inform WT/MT ratio in the samples. This capability can be very useful when a heterozygous mutant is present and also provides a fail-safe measure to prevent errors such as amplification failure from affecting the results.

In principle, TRANSLATR assays should enable mutation detection over the full length of an open reading frame. To examine the length of sequence that TRANSLATR can assess at a time, we introduced a one-nucleotide insertion in the lacZ gene and used RPA to amplify gene fragments of different lengths. The amplicons were added to cell-free reactions and fluorescence was measured (**Figure 5.7.4f**). We found that TRANSLATR can amplify and scan up to 300 bp sequence for frameshift and nonsense mutations (**Figure 5.7.4g**).

Moreover, we exploited the ability of both RPA and cell-free reactions to operate at room temperature. We performed all the RPA and the cell-free reactions under room temperature (~22 °C) and 25 °C and tested if the reactions could occur. We extended the RPA reaction time to 30 minutes and measured the fluorescence of the cell-free reactions afterwards. Although a longer time is required, TRANSLATR can function to discriminate mutations at room temperature (**Figure 5.7.4h**).

Taken together, this dual fluorescent readout detection system not only provides a way to normalize the input DNA amount but also offers a more quantitative method in identifying frameshift and nonsense mutations. In addition, TRANSLATR can scan up to 300 bp sequence for frameshift and nonsense mutations and is capable of functioning at room temperature.

5.3.4 Mutation detection in colorimetric, paper-based reactions

Although the fluorescent split proteins provide strong signals in TRANSLATR assays, colorimetric reporter systems are attractive readout options since they can be viewed directly by the eye without requiring any additional light sources or optical filters. Pardee et al.[19] have demonstrated the use of β-galactosidase to produce yellow-to-red colorimetric output in paper-based cell-free systems which is an easy-to-read signal for diagnostics, while Ma et al.[24] took advantage of lacZ α-complementation for faster cell-free reaction signals (**Figure 5.8.5a**). To implement a colorimetric TRANSLATR assay, we employed a minimal length lacZ-α peptide of 38 residues with proven activity[115] and tested its activity in paper-based cell-free reactions supplemented with lacZ-ω. Upon production, the lacZ-α peptides will reassemble with lacZ-ω protein supplied in the cell-free system to form functional β-galactosidase, which catalyzes a colorimetric reaction, cleavage of the initially yellow substrate chlorophenol red-β-D-galactopyranoside to dark red (**Figure 5.8.5b**).

We designed the primers to match the 3029_3030del mutant in the BRCA1 gene and used the minimal lacZ-α as the downstream reporter peptide. Consistent with the design, the OD575 value of the mutant increased rapidly compared with the wild type, indicating production of lacZ-α as expected for the mutant and a color change from yellow to dark red can be visualized by eye in 15-20 minutes (**Figure 5.7.5c**).

5.3.5 Rapid, specific, and sensitive detection of mutations in clinical applications

To facilitate use with clinical samples, we implemented the general workflow shown in **Figure 5.7.6a** to design and develop a TRANSLATR assay for patient serum samples. First, primers targeting the specific DNA/RNA sequence of interest are designed *in silico*. The clinical samples are then subjected to lysis and extraction. The extracted DNA/RNA can then be added to the amplification mix. For improved sensitivity, we adopted a two-step RPA strategy where short RPA primers are used for first-round amplification and then the peptide reporter primers were used to further amplify the target sequence. Amplification products are diluted and added directly to the cell-free reaction. Fluorescence or colorimetric signals can then be observed, measured, and recorded. We determined the limit of detection of the workflow to be 7.5 aM (**Figure 5.7.6b**).

Lastly, we showed the detection of BRCA mutations in clinical samples. These mutations occur in the germline and can lead to an increased risk of breast and ovarian cancers. Using the self-calibrating TRANSLATR system, we successfully captured various mutations in the patients, shown by fluorescence differences. In **Figure 5.7.6 c**, **d and e**, fluorescence was designed to turn on when wild type is present. The clinical mutant samples are higher than the synthetic samples tested and the ratio of the fluorescence signal of clinical mutant to wild type is around 0.4-0.6, suggesting the mutation is heterozygous. In fact, the germline BRCA mutations in humans are usually heterozygous, meaning only one allele is mutated. Homozygous germline mutants on

BRCA1/2 have lethal effects[116]. Signals can be swapped to have the mutants turn on (Figure 6f and g) fluorescence expression based on clinical needs.

5.4 Discussion

We have developed a rapid and low-cost translation-based sequence verification method that employs isothermal reactions. TRANSLATR can rapidly detect frameshift, nonsense, and missense mutations. Compared with conventional assays, the assay has several desirable features. Inheriting the advantages of cell-free-based assays, it is inexpensive and temperature stable, costing only ~\$3 per test and using chemical components that are readily lyophilized. The cost can be further lower by using homemade PURE cell-free systems[117] (0.09 USD/μL). The small sizes of the peptide reporters allow distinguishable signals to generate much faster (15-30 mins) compared to other cell-free assays using full-length reporters (typically at least 1 hour)[24,41]. It provides sensitive and semi-quantitative measurements with visible readouts that can be observed directly by the eye or with simple equipment. Notably, the full procedure can be operated under room temperature, further facilitating use in low-resource settings. TRANSLATR can examine gene fragments up to 300 bp at a time for frameshift and nonsense mutations.

The primer design process is simple and straightforward. The highly flexible primer design criteria and various mutation detection schemes (e.g., the use of primer binding affinity, translational frameshifting events, generation of stop codons and omission of the start codon) have allowed us to detect frameshift, nonsense and missense mutations against a variety of targets. Importantly, the open reading frame can be defined and modified in primer designs and does not necessarily need to be the natural open reading frame, which allowed us to switch signals between the wild type and the mutant and provided more possibilities in the design.

TRANSLATR can detect frameshift and nonsense mutations in genes around 300 bp – 400 bp at a time, longer than most other mutation detection methods based on base-pairing and nucleotide mismatches[100,118]. We expect that this length limit can be extended by using RPA formulations tailored for improved production of long amplicons rather than short ones. Also, long peptide sequences generated by extended assay open reading frames could interfere with assembly of the split proteins.

Our experiments also showed a novel application of self-associating split fluorescence protein systems, which were previously primarily used in *in vivo* studies. The incorporation of GFP11[102,119] and sfCherry$_2$11[103,104] in primers enables them to be readily inserted into any coding sequences to serve as reporters. These tags can report mutations in the coding sequence by cell-free reactions and therefore are useful for cell-free nucleic acid diagnostics. The fluorescence output can further be coupled with a low-cost LED illuminator[120] to allow facile reading and meet point-of-need demand. The split sfCherry3C(1-10)[104] we used is a further improved version of sfCherry (1-10)[102] and sfCherry2(1-10)[103] for enhanced complementation efficiency. Consistent with the observation in vivo, we saw lower complementation efficiency in cell-free reactions. Screening and finding split fluorescent proteins with greater complementation efficiency in cell-free systems will further shorten the time for readout. Moreover, the discovery of new and orthogonal self-associating split fluorescent protein systems[121] with different colors allows multiplexed fluorescent color output and could potentially be harnessed for detecting multiple mutations in parallel.

When testing the platform in a cell-free system, we observed some leaky fluorescence in the mutant samples, which suggests some expression of the reporter. To further investigate the cause, we insert the fragments of sequence containing 6174delT mutations and wild-type in the BRCA2 gene with downstream GFP11 sequence into plasmids and co-transformed them with GFP1-10 plasmids into Escherichia coli, respectively. We observed strong green fluorescence for cells having wild-type sequences (**Figure 5.7.8a**), a result of normal expression of GFP11. In contrast, cells with 6174delT mutants only provided signals similar to cell autofluorescence, a result of the single nucleotide deletion. The leaky fluorescence is significant lower in cell environment. Worrying that potential wild-type template contaminations to the PCR/RPA products or artifacts generated by the amplification reactions may contributed to the leaky expression in GFP11, we put the purified plasmids used in in vivo fluorescent study to the cell-free expression and measure the fluorescence. We can still observe the leaky expression (**Figure 5.7.8b**). The ratio of the mutant fluorescence to the wild-type remains similar (**Figure 5.7.8c**), suggesting that leakage was not from contaminations or artifacts. In fact, the leaky expression in cell-free system is not uncommon. Hong et al.[41] have also observed lower differentiation factors in cell-free reactions as compared to E. coli cells. At this point, we hypothesized that the resources in cell-free reactions limit the overall expression of the wild type and that the translational regulatory system is not as stringent as that in intact cells, which leads to a very small amount of reporter production. These could be improved by supplying additional translation factors that promote translation fidelity and termination efficiency.

Currently, TRANSLATR aims to provide accessible genetic testing for germline mutations. Further improvement of sensitivity could allow us to identify somatic mutations taking up only a portion of the genes.

Overall, TRANSLATR presents an easy-to-design, easy-to-perform, easy-to-observe method for point-of-care diagnostics that can potentially provide wide-accessibility to genetic testing. Its properties allow it to potentially be a reliable point-of-care testing method. It can be very useful in the rapid screening of frequent mutations in a disease, or some known mutations present in other family members. Moreover, it can be used to quickly identify diseases that are otherwise hard to diagnose based on symptoms, decreasing patient suffering and improving outcomes. Additionally, novel therapeutics such as gene editing technologies are rapidly emerging to treat genetic diseases by correcting specific gene mutations. In this case, pairing a rapid mutation diagnostic test can inform treatment selection and promote precision medicine.

5.5 Methods

Polymerase chain reaction

PCR primers were designed using NCBI Primer-BLAST with primer melting temperature set at 57 °C. Reporter elements were added and primer secondary structures were checked. Phusion High-Fidelity PCR Master Mix (NEB, M0531S) was used for all reactions. Forward and reverse primers were added at 500 nM in a 50-µL reaction before being placed in a thermocycler. We followed the manufacturer's protocol with an annealing temperature at 57°C and an extension time of 30 s/kb.

Recombinase polymerase reaction

RPA primers were designed based on recommended the manufacturer's conditions with NCBI-Primer BLAST. For RPA primers containing peptide reporter elements, primer secondary structures were predicted and checked to ensure the internal interactions of primers will not prevent binding to the target sequence. For the best performance of RPA, primer screening is recommended. Primers were ordered as single-stranded DNA oligos from IDT (Integrated DNA technologies, lnc).

RPA kits were sourced from TwistDx (TABAS03KIT) and M-MuLV reverse transcriptase was purchased from New England Biolabs (M0253L). RPA and RT-RPA reactions were run following the manufacturer's protocol. With the exception that the rehydration buffer was added to the lyophilized enzyme pellets to allow the use of smaller reaction volumes (e.g., 10µL, 20µL) prior to the addition of other components. For synthetic templates, a final concentration of 2 around 1nM was used unless otherwise noted. For extracted serum samples, 2µL was used in a 15µL reaction. The reactions were run at 15µl, 37 °C for 20 minutes unless otherwise noted.

For two-step RPA, the first reaction was run using standard RPA primers at 10µL, 37 °C for 10 minutes. 1 µL of the resulting products was straightly added to the second RPA reaction (10µL) containing peptide reporter primers and was incubated for another 20 minutes at 37 °C.

Cell-free reactions

All experiments were performed using PURExpress in vitro protein synthesis kits (New England Biolabs, E6800L). A typical 5 µL liquid-phase cell-free reaction contains 2 µL of PURExpress Solution A (40%, v/v), 1.5 µL PURExpress Solution B (30%, v/v), 0.1µL of RNase inhibitor (40 U/µl, 03335402001, Roche), 0.6µL of purified GFP1-10 proteins (65µM) (for dual channel fluorescent experiments, 0.4µL of purified GFP1-10 proteins and 0.6 µL of sfCherry3C(1-10) proteins (65µM) were added) and 0.4 µL of DNA amplicons. The rest of the volume was filled with ultrapure water. RT-RPA, RPA and PCR products were diluted 5-fold before adding to cell-free reactions. After thoroughly mixing, 4µL of the cell-free reactions were then transferred to respective wells in a 384-well microplate. When using split fluorescent proteins as reporters, green fluorescent signals were measured at Ex485nm/Em525nm with 20 nm bandwidth for both for 4 hours at 37 °C using a plate reader (BioTek Neo2). Red fluorescence signals, where applicable, were measured at 570 nm excitation wavelength and 611 nm emission wavelength with 20 nm bandwidth. When using lacZ-α peptides as reporters, chlorophenol red-β-D-galactopyranoside (10884308001, Roche) is used as the substrate at 0.6mg/mL, and OD575 was used for signal measurements.

For paper-based cell-free experiments, 2-mm paper disks (Whatman, 1442-042) were prepared with a disposable biopsy punch and were placed into the wells of a 384-well microplate. The cell-free reactions were scaled down to a total volume of 2 µL based on the above recipe and were applied to the paper disks respectively. Photographs were taken with a smartphone camera from the bottom of the plate.

To visualize fluorescence, we used a blue light transilluminator (G6600, Life Technologies). Fluorescence and colorimetric photos were captured by a smartphone.

Plasmid construction

DNA oligonucleotides used as primers and insert templates are purchased from IDT (Integrated DNA technologies, lnc). pET15b, pCOLADuet-1 and pCDF were used as backbones. Insert and backbone sequences were amplified using primers with 30-bp overlapping Gibson domains. Gibson assembly was then used for ligation previously[9]. Plasmids were chemically transformed into DH5α competent cells and were selected on an agar plate with appropriate antibiotic resistance. Sequences of successful transformants were confirmed with Sanger sequencing and mini-prep before use.

In vivo fluorescent measurement

The insert sequence containing GFP11 was purchased as ssDNA oligos (Integrated DNA Technologies). Plasmids were constructed via Gibson assembly[122] using methods described previously[9]. Successful constructs were identified by Sanger sequencing and were co-transformed into E. coli BL21(DE3) cells with GFP1-10 plasmids for overnight incubation. Single colonies were selected to grow in LB media with appropriate antibiotics to log phase before the addition of 0.1mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After 3-hour induction, cells were diluted in phosphate-buffered saline and were transferred to a 384-well plate. In vivo fluorescent measurement was completed using flow cytometry. The green fluorescence of ~40,000 cells was captured individually to obtain the geometric mean.

Protein Expression and Purification

The pG110 plasmid encoding GFP 1-10 was a gift from Dr. Jeremy Mills (ASU). The plasmids were transformed to BL21(DE3) competent cells and a selected single colony was inoculated overnight in 6 mL LB media with appropriate antibiotics. To start expression, 4 mL of the overnight culture was added to 400 mL LB media with appropriate antibiotics and incubated at 37 °C, 250 rpm. After OD600 reached 0.6-0.8, add 1mM IPTG to the culture to induce expression of GFP 1-10 at 30 °C, 250 rpm for 6 hours. The resulting culture was centrifuged to obtain cell pellets and was either purified directly or stored at - 80 °C. The amino acid sequence for sfCherry3C (1-10) from Ref. 37 was codon-optimized for Escherichia coli protein expression and synthesized by Genscript. lacZω plasmids driven by a T7 promoter were from our previous publication[24] (Addgene, 118815). The expression procedure for those proteins remains the same except lacZω was purified using a lacZ-deficient E.coli BL21(DE3) strain[123]. All plasmids contain a 6×his tag upstream or downstream of the target proteins for purifications. For his-tagged protein purification, the cell pellets were resuspended in Buffer A and sonicated to released proteins. All proteins were purified with FPLC using Ni-NTA columns. Purified proteins were quantified using a Qubit 4 fluorometer. and were aliquoted and stored in -80 °C.

Clinical serum sample preparation and extraction

Serum samples for BRCA1/BRCA2 variant detection were obtained from Banner MD Anderson Cancer Center with informed consent from donors. The study was approved by the Banner Health Institutional Review Board. The genomic DNA was extracted from serum samples using a QIAamp DNA Blood mini kit following the manufacturer protocol.

GFP signal normalization and calibration

GFP signal was completed using the following formula:

(1) Assuming three replicates were performed, set the wild-type cherry signal to be 1, thus normalized wild-type GFP signals remain the same, denoted as:

$$cal_{wt_{gfp(1)}} = raw\_wt_{\_gfp(1)}/1$$

$$cal\_wt_{gfp(2)} = raw\_wt_{\_gfp(2)}/1$$

$$cal_{wt_{gfp(3)}} = raw\_wt_{\_gfp(3)}/1$$

(2) Calculate relative ratio of wild type cherry signals to mutant cherry signals:

$$wt_{\overline{cherry}} = mean(wt_{cherry(1)} + wt_{cherry(2)} + wt_{cherry(3)})$$

$$cal\_mut_{gfp(1)} = mut_{gfp(1)}/(mut_{cherry(1)}/wt_{\overline{cherry}})$$

$$cal\_mut_{gfp(2)} = mut_{gfp(2)}/(mut_{cherry(2)}/wt_{\overline{cherry}})$$

$$cal\_mut_{gfp(3)} = mut_{gfp(3)}/(mut_{cherry(3)}/wt_{\overline{cherry}})$$

Data analysis

Unpaired t-tests were performed to determine the significance level using Prism version 9.2.0 for Windows, GraphPad Software, San Diego, California, USA, www.graphpad.com. Data were plotted using MATLAB or Prism.

5.6 Acknowledgements

The pG110 plasmid for GFP1-10 expression is a gift from Dr. Jeremy Mills (ASU). We thank Justin Vigar for purifying the omega protein used in the colorimetric reactions. This work was supported by funds to A.A.G. from an NIH Director’s New Innovator Award (1DP2GM126892), the Gates Foundation (OPP1160667), an Arizona Biomedical Research Commission New Investigator Award (ADHS16-162400), an Alfred P. Sloan Fellowship (FG-2017-9108), a DARPA Young Faculty Award (D17AP00026), Gordon and Betty Moore Foundation funds (#6984), NIH funds (1R21AI136571), and Arizona State University.

5.7 Figures

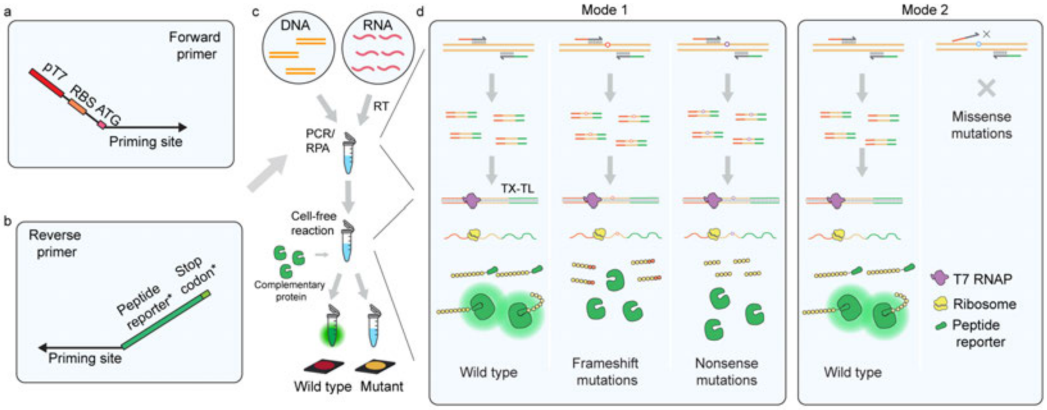

a
Forward primer
pT7
RBS ATG
Priming site
b
Reverse primer
Peptide reporter*
Stop codon*
Priming site
c
DNA
RNA
RT
PCR/ RPA
Cell-free reaction
Complementary protein
Wild type
Mutant
d
Mode 1
TX-TL
Wild type
Frameshift mutations
Nonsense mutations
Mode 2
Missense mutations
T7 RNAP
Ribosome
Peptide reporter
Wild type

Figure 5.7.1 Schematic representation of detection of mutations using TRANSLATR.

(a) Design scheme of a forward TRANSLATR primer. A T7 promoter, a prokaryotic ribosome binding site (RBS) and a start codon (ATG) sequence were sequentially encoded in the 5’ primer overhang.

(b) Design scheme of a reverse TRANSLATR primer. A peptide reporter and a stop codon sequence were reverse complementarily encoded in the 5’ primer overhang.

(c) Workflow for detection of mutation by TRANSLATR. RT: reverse transcription

(d) Detection principles of the TRANSLATR system. Two detection modes were developed to allow detection of different mutation categories.

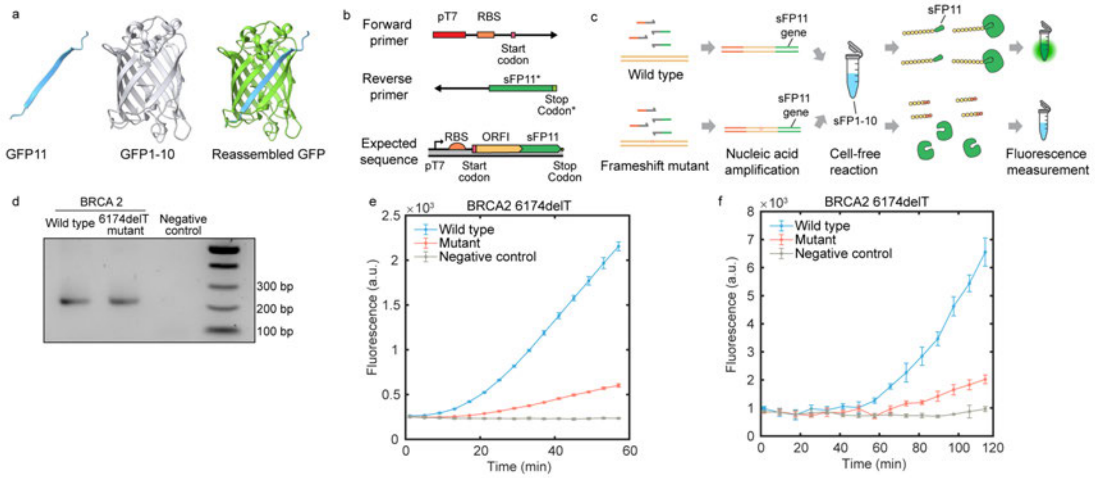


Figure 5.7.2 Split fluorescent proteins are ideal reporters for detection of mutations.

(a) The structure of the 11th β-strand of GFP (GFP11), GFP1-10 and the re-assembled fluorescent GFP. Crystal structures were adapted from a superfolder green fluorescent protein (PDB:2B3P) using UCSF Chimera.

(b) Primer design scheme and expected sequence from amplification. GFP11 or sfCherry$_2$11 was used as the peptide reporter. Other necessary elements were included to allow ORFOI to transcribe and translate. ORFOI: open reading frame of interest. sFP: split fluorescent protein.

(c) Detection of frameshift mutations based on fluorescence output. Successful generation of fluorescence is dependent on the correct generation of peptide reporters.

(d) DNA amplicons from the PCR reaction using GFP-embedded TRANSLATR primers targeting BRCA2 6174delT mutant.

(e-f) The fluorescence response of TRANSLATR targeting BRCA2 6174delT using either GFP11(e) or tandem sfCherry$_2$11(f) as the reporter (n = 3 technical replicates; bars represent arithmetic mean ± SD).

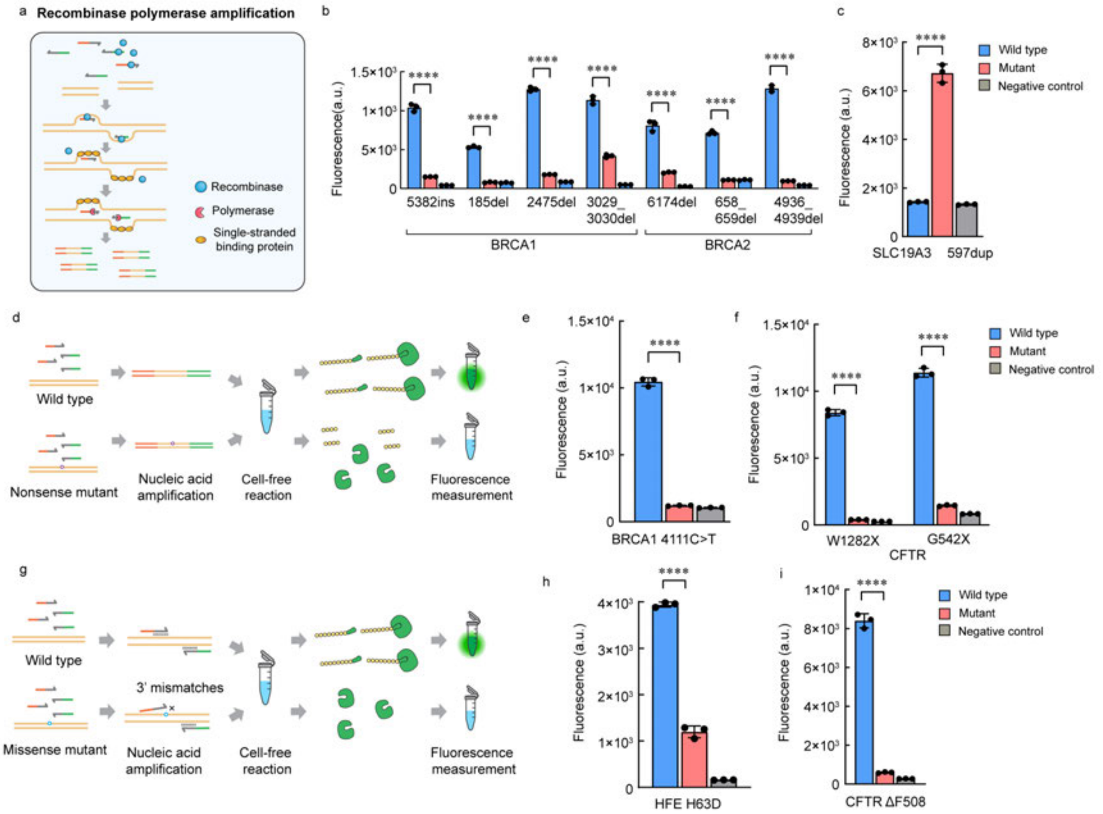


Figure 5.7.3 Implementing RPA to detect different mutations by TRANSLATR

(a) Schematic of recombinase polymerase amplification with TRANSLATR primers.

(b) Fluorescent detection of 7 pathogenic frameshift mutations in BRCA1 or BRCA2 gene amplified from RPA using GFP11 peptide reporters after 60 minutes (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

(c) Fluorescent detection of 597dup in the SLC19A3 gene using TRANSLATR primers designed to turn on fluorescence when mutant is present.

(d) Detection of nonsense mutations by TRANSLATR. Nonsense mutations cause premature translation termination and prevent peptide reporters from generation.

(e-f) Fluorescent detection of nonsense 4111C>T in the BRCA1gene (e) and W1282X and G542X (f) in the CFTR gene using TRANSLATR. Fluorescence was plotted at 60 minutes (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

(g) Detection of missense mutations by TRANSLATR. The 3' end of the primers are designed to have more matches to either the wild type or the mutant sequences, resulting in difference in amplification efficiency and fluorescence difference.

(h-i) Fluorescent detection of a missense mutation H63D in the HFE gene and non-frameshift deletion delF508 in the CFTR gene. Fluorescence was plotted at 60 minutes. (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

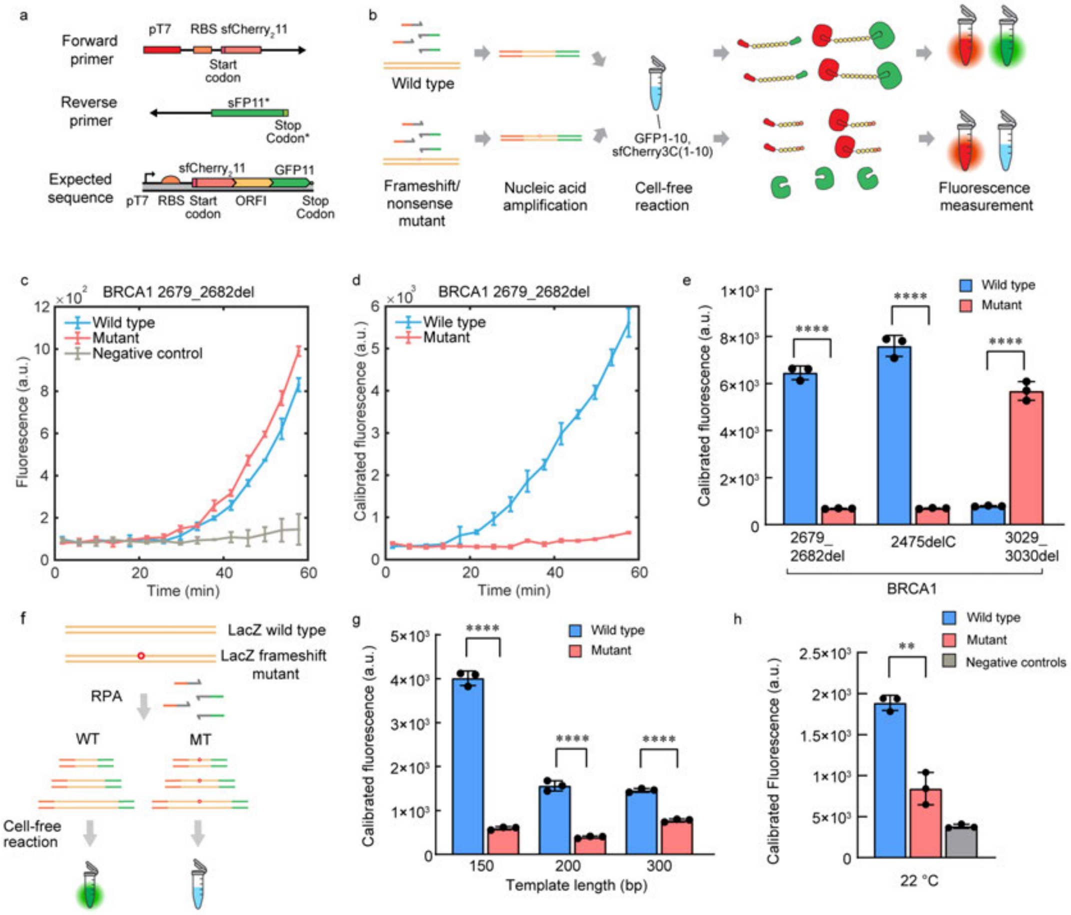


Figure 5.7.4 Signal self-calibration by dual split fluorescent peptide reporter.

(a) Primer design scheme for a self-calibrating TRANSLATR using dual fluorescent peptide reporter systems. A 16-amino-acid sfCherry$_2$11 peptide was inserted in the forward primer immediately after the start codon.

(b) Schematic of self-calibrating detection of mutations using dual-fluorescent TRANSLATR. sfCherry$_2$11 peptides will be translated as long as ORFOI is amplified, GFP11 peptides will be selectively translated.

(c) Red fluorescence responses for the detection of BRCA1 2679_2682del. Fluorescence is produced in both wild type and mutant group.

(d) Green fluorescence responses after calibration (n = 3 technical replicates; bars represent arithmetic mean ± SD).

(e) Fluorescent detection of 2679_2682del, 2475delC, 3029_3030del in the BRCA1 gene. Fluorescence was plotted at 60 minutes. TRANSLATR was designed to turn on fluorescence when wild type is present for 2679_2682del, 2475delC and to be on when mutant is present for 3029_3030del (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

(f) Schematic of investigating mutation sensing range using lacZ and its frameshift mutant. TRANSLATR primers were designed to amplify different length of lacZ gene fragments with the mutation positioned in the middle.

(g) Fluorescent detection of mutations in different template lengths. Fluorescence was plotted at 60 minutes (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

(h) Fluorescent detection of mutation at room temperature 22 °C. Fluorescence was plotted at 2 hours (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

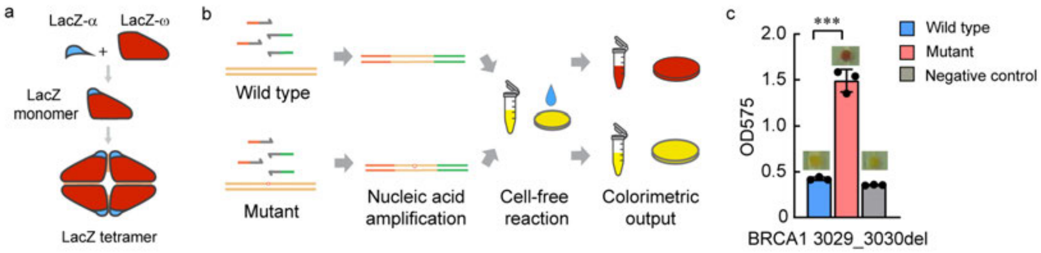

a
LacZ-α
LacZ-ω
LacZ monomer
LacZ tetramer
b
Wild type
Mutant
Nucleic acid amplification
Cell-free reaction
Colorimetric output
c
***
2.0
1.5
1.0
0.5
0
OD575
Wild type
Mutant
Negative control
BRCA1 3029_3030del

Figure 5.7.5 Using LacZ-alpha peptide reporter to enable paper-based and colorimetric output.

(a) Schematic of α-complementation by a 38-amin-acid lacZ-α peptide and lacZ-Ω protein that constitute lacZ proteins.

(b) Schematic of detection of mutations using lacZα-peptide with solution- or paper-based cell-free reactions.

(c) Colorimetric detection of 3029_3030del in the BRCA1. OD575 was plotted after 20 minutes (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD). Photographs were captured at the same time.

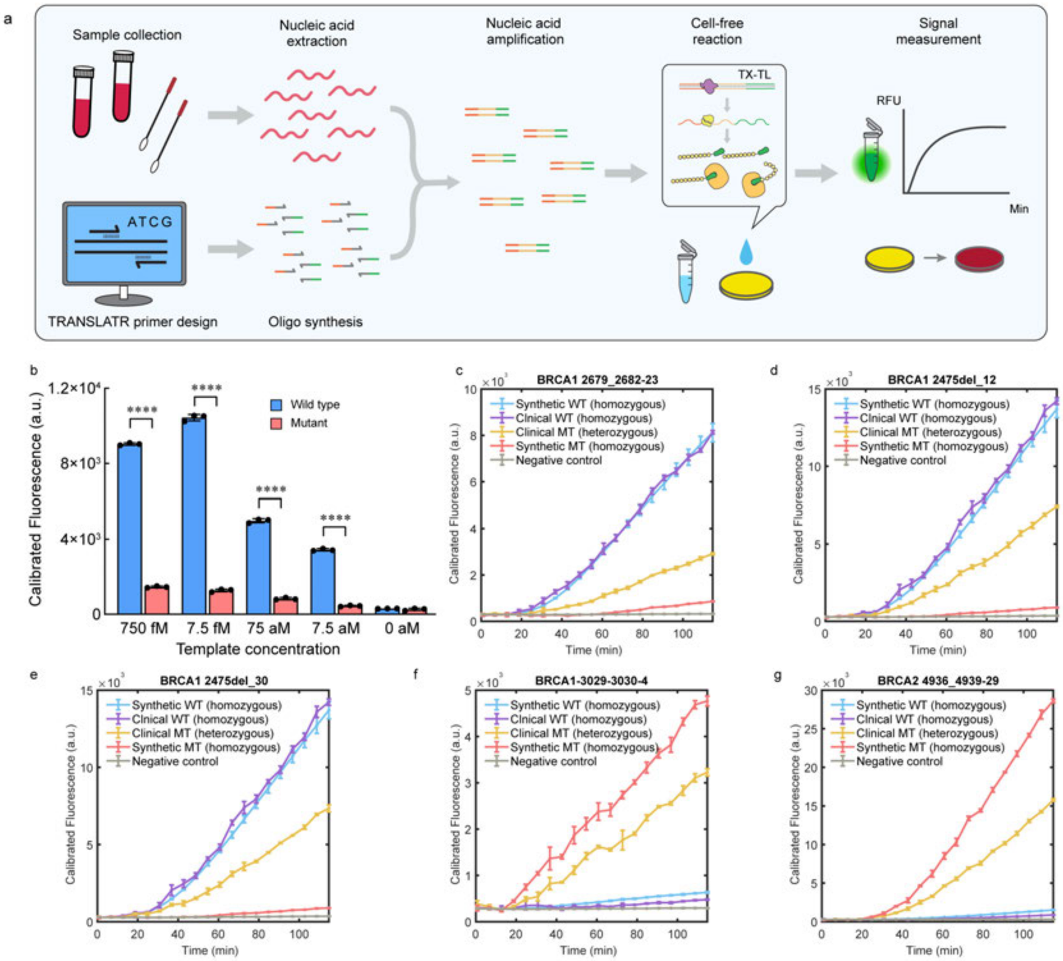


Figure 5.7.6 Streamlined workflow for the detection of clinical samples.

(a) Schematic of TRANSLATR detection workflow.

(b) Sensitivity of the TRANSALTR platform using dual fluorescent peptide reporters. (unpaired two-sample t-tests were used to determine the statistical significance; n = 3 technical replicates; bars represent arithmetic mean ± SD).

(c-g) Fluorescent detection of mutations in the patient blood samples. Synthetic wild type and synthetic mutant templates were run as the controls (n = 3 technical replicates; bars represent arithmetic mean ± SD).

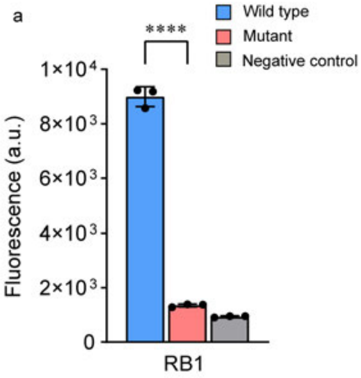


Figure 5.7.7 Detection of a nonsense mutation in the RB1 gene

(n = 3 technical replicates; bars represent arithmetic mean ± SD). Photographs were captured at the same time.

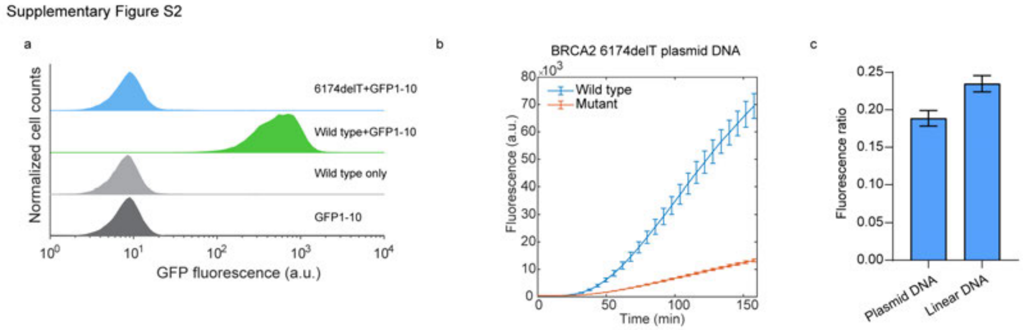


Figure 5.7.8 Fluorescence response for BRCA2 6174delT in cell-free reactions and E. coli.

(n = 3 technical replicates; bars represent arithmetic mean ± SD). Photographs were captured at the same time).

(a) Flow cytometry green fluorescence histograms for cells transformed with BRCA2 wild type and BRCA2 6174delT mutant, tagged by GFP11 downstream. GFP11 sequence is designed to be in frame with wild type ORF. Controls were transformed with either wild-type plasmids or GFP1-10 plasmids as comparisons. At least 3 biological replicates were performed with similar results.

(b) Fluorescence response in cell-free reactions using constructed wild-type and mutant plasmids for detection of 6174delT in the BRCA2 gene.

(c) The fluorescence ratio of mutant and wild type for using plasmid DNA and linear DNA in cell-free reactions. Fluorescence was plotted at 1 hour (n = 3 technical replicates; bars represent arithmetic mean ± SD).

CHAPTER 6 CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, I have described the development and enhancement of cell-free diagnostics with the combination of different approaches. The recent advancement of cell-free systems allows them to be used for in vitro diagnostics in portable settings.

In chapter 2, I described the computational methods for designing toehold switch riboregulators and single-nucleotide specific programmable riboregulators. The computerized methods enable user-friendly design process and quick generation of target sensors. The protocol described here provide details that walk users through the design process.

When the cell-free systems are combined with toehold switches, I describe the detection of fungi Coccidioides, norovirus, SARS-CoV-2, pathogens that are impacting and concerning the society. In the detection of fungi Coccidioides, we used RT-RPA as an alternative amplification method to increase the reaction speed and sensitivity of the assay. For norovirus detection, we applied synbody-based viral enrichment to concentrate the virus, which increases the overall sensitivity of the assay. In addition, we use lacZα in place of full-length lacZ, which results in an increase in protein expression. At last, we demonstrate the detection of SARS-CoV-2 using toehold switches as the sensors and gel switch resonators as the reporter. The combination allows contact-free testing which has the potential to increase testing privacy and save consumables.

With the goal to further increase sequence specificity, we developed another type of de novo-designed riboregulator systems called single-nucleotide specific programmable riboregulators (SNIPRs) to address insufficient specificity issue of current riboregulators. SNIPRs can reliably resolve single-nucleotide differences in sequence in living prokaryotic cells and in vitro in cell-free systems. To demonstrate the potential applicability in real-world challenges, we worked on the development of SARS-CoV-2 variant diagnostics based on SNIPRs. Results have shown SNIPRs are capable of discerning five variants of concern with high sensitivity when integrating an isothermal amplification method. In the subsequent tests with clinical samples, SNIPRs accurately identified Omicron and Delta variants, achieving accuracy above 96.5%, providing a reliable and accurate assay for alternative portable and low-cost assays.

In the last chapter, we devised a new method for detecting mutations using split proteins and cell-free systems, which we termed TRANSLAR (TRANslation of Sequence-LAbelled Transcripts). By incorporating the short peptide reporters derived from split protein systems into amplification primers, we used isothermal methods to amplify and tag the targets and read their mutation status by simply transcribing and translating them. Using TRANSLATR, we successfully captured frameshift mutations, nonsense mutations and missense mutations. TRANSLATR is compatible with fluorescent and colorimetric output, and results can be provided in less than 15 minutes. TRANSLAR features another promising cell-free based strategy for quick mutation screening.

Together, this research brings advancements in diagnostics based on riboregulators and cell-free systems that will increase the accessibility of these important healthcare tools. However, future efforts are needed to push these diagnostics to actual clinical or point-of-care applications. Below I listed some directions which we could work on further.

First, the steps of the cell-free assay could be reduced. Currently, the workflow of cell-free diagnostics includes sampling, nucleic acid extraction, isothermal amplification, and cell-free reactions. For a laboratory-based test, a four-step assay can be easily accommodated. However, for a point-of-care test, for maximal convenience of users, steps could be further reduced. For example, one-pot or amplification-free CRISPR assays[58,124] have been described. These assays either manage to enable isothermal amplification and downstream reactions in one reaction, or they use multiple sensor probes to amplify the signal to obviate the need for nucleic acid amplification steps. Compared to antigen-based lateral flow tests, in which sampling and reading are straightforward and simple, point-of-care nucleic-acid tests still require more efforts to be there, and reducing steps required for sample-to-result is one critical aim.

Second, quick and reliable sample lysis and extraction method are needed. For laboratory-based qPCR tests, nucleic-acid extraction using columns or beads can be reliably performed by high-throughput robotic systems. Whereas in point-of-care settings, the common and reliable extraction methods cannot be integrated, and such methods are lacking. Scientists have tried thermal lysis[85,86] and chemical lysis[85] as quick extraction methods to satisfy point-of-care purposes, but how to maintain comparable sample quality and quantity that could affect assay sensitivity and reliability remain a challenge. Besides trying to implement and optimize thermal lysis and chemical lysis for use in cell-free diagnostic, we should also seek for new strategies that allow convenient yet reliable viral lysis and extraction methods to further push nucleic-acid tests to point-of-care settings.

Third, the speed of the cell-free assays could be increased. In this dissertation, we tried using small reporter system to enhance speed of the assay; we also tried different isothermal amplification methods in an attempt to reduce the total assay time. However, for riboregulator-based diagnostics, the current reaction time is around 2 hours. For TRANSLATR assays, reaction time can be shortened to 20-30 minutes. To compete with 15-minute antigen-based test, we should continue to improve the speed of the cell-free assays.

Fourth, the cost of the assay could be further reduced. Currently, the cell-free diagnostics are among $3 per tests. The low cost of the assay is largely owing to the small volume of the cell-free reactions. However, the current cell-free reactions used commercialized recombinant-protein-based PURE system which is among the most expensive version of prokaryotic cell-free system. To further reduce the cost, we could use home-made PURE system[117] or lysate-based cell-free system which could bring down the cost by 100-fold or more.

Fifth, from the computation design aspect, more efforts can be done to refine the design process, which can provide a more accurate predication of the switch performance and reduce experimental work. This can be achieved by applying machine learning algorithms to study a large-set of switch performance. To be more specific, we can design a oligo pool with thousands of toehold switch sequences, perform a high-throughput screening using FACS and classify the performance of those switches based on output signals. Then we can use them to generate a data set for machine learning algorithms. The algorithms will hopefully provide some critical parameters impacting switch performance.

Last but not least, in parallel to the biological part of diagnostics, advances in materials, electronics, optics can also benefit synthetic-biology based diagnostics and the integration of these technologies can function together to enable a more user-friendly, field-deployable, automated, simple, reliable and diverse diagnostic systems. For example, Collins and coworkers worked on wearable materials for biomolecule detection[125]. Hsu and coworkers integrated microfluidic system to CRISPR-based diagnostics that allows the assay to self-operate. These all shown how advances from different disciplines could be combined to yield greater benefits and improvements.

BIBLIOGRAPHY

1. Théorie Physico-chimique de la Vie et Générations Spontanées. *Nature 1911 86:2169* **86**, 410–410 (1911).
2. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nature Reviews Microbiology 2014 12:5* **12**, 381–390 (2014).
3. Milestones in synthetic (micro)biology. *Nature Reviews Microbiology 2014 12:5* **12**, 309–309 (2014).
4. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature reviews. Genetics* **8**, 776 (2007).
5. Jabri, E. Small, but in control. *Nature Reviews Molecular Cell Biology 2005 6:5* **6**, 361–361 (2005).
6. Caron, M. P. *et al.* Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America* **109**, (2012).
7. Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology 2004 22:7* **22**, 841–847 (2004).
8. Chappell, J., Takahashi, M. K. & Lucks, J. B. Creating small transcription activating RNAs. *Nature Chemical Biology* **11**, 214–220 (2015).
9. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: De-novo-designed regulators of gene expression. *Cell* **159**, 925–939 (2014).
10. Ma, D. *et al.* Multi-arm RNA junctions encoding molecular logic unconstrained by

input sequence for versatile cell-free diagnostics. *Nature Biomedical Engineering 2022 6:3* **6**, 298–309 (2022).

11. Kim, J. *et al.* De novo-designed translation-repressing riboregulators for multi-input cellular logic. *Nature Chemical Biology* **15**, 1173–1182 (2019).
12. Zhao, E. M. *et al.* RNA-responsive elements for eukaryotic translational control. *Nature Biotechnology 2021* 1–7 (2021) doi:10.1038/s41587-021-01068-2.
13. Green, A. A. *et al.* Complex cellular logic computation using ribocomputing devices. *Nature* **548**, 117–121 (2017).
14. NIRENBERG, M. W. & MATTHAEI, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 1588–1602 (1961).
15. Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature 2021 593:7858* **593**, 266–269 (2021).
16. Stark, J. C. *et al.* BioBits^{TM} Bright: A fluorescent synthetic biology education kit. *Science Advances* **4**, 33 (2018).
17. Huang, A. *et al.* Biobits^{TM} explorer: A modular synthetic biology education kit. *Science Advances* **4**, (2018).
18. Stark, J. C. *et al. Stark et al On-demand biomanufacturing of protective conjugate vaccines*. *Sci. Adv* vol. 7 http://advances.sciencemag.org/ (2021).
19. Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable

Biomolecular Components. *Cell* **165**, 1255–1266 (2016).

20. Zubi, Y. S. *et al.* Metal-responsive regulation of enzyme catalysis using genetically encoded chemical switches. *Nature Communications* **13**, 1864 (2022).
21. Karim, A. S. & Jewett, M. C. Cell-Free Synthetic Biology for Pathway Prototyping. *Methods in enzymology* **608**, 31–57 (2018).
22. Meng, F. & Ellis, T. The second decade of synthetic biology: 2010–2020. *Nature Communications 2020 11:1* **11**, 1–4 (2020).
23. Tan, X., Letendre, J. H., Collins, J. J. & Wong, W. W. Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics. *Cell* **184**, 881–898 (2021).
24. Ma, D., Shen, L., Wu, K., Diehnelt, C. W. & Green, A. A. Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. *Synthetic Biology* **3**, (2018).
25. Karlikow, M. *et al.* Field validation of the performance of paper-based tests for the detection of the Zika and chikungunya viruses in serum samples. *Nature Biomedical Engineering* **6**, 246–256 (2022).
26. Ishino, Y., Krupovic, M. & Forterre, P. History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *Journal of Bacteriology* **200**, (2018).
27. Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438–442 (2017).

28. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).

29. Kaminski, M. M., Abudayyeh, O. O., Gootenberg, J. S., Zhang, F. & Collins, J. J. CRISPR-based diagnostics. *Nature Biomedical Engineering* **5**, 643–656 (2021).

30. Chappell, J., Watters, K. E., Takahashi, M. K. & Lucks, J. B. A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Current Opinion in Chemical Biology* **28**, 47–56 (2015).

31. Green, A. A. Synthetic bionanotechnology: synthetic biology finds a toehold in nanotechnology. *Emerging Topics in Life Sciences* **3**, 507–516 (2019).

32. Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology* **22**, 841–847 (2004).

33. Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D. & Arkin, A. P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences* **108**, 8617–8622 (2011).

34. Callura, J. M., Cantor, C. R. & Collins, J. J. Genetic switchboard for synthetic biology applications. *Proceedings of the National Academy of Sciences* **109**, 5850–5855 (2012).

35. Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R. & Collins, J. J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences* **107**, 15898–15903 (2010).

36. Chappell, J., Takahashi, M. K. & Lucks, J. B. Creating small transcription

activating RNAs. *Nature Chemical Biology* **11**, 214–220 (2015).

37. Pardee, K. *et al.* Paper-based synthetic gene networks. *Cell* **159**, 940–954 (2014).

38. Amalfitano, E. *et al.* A glucose meter interface for point-of-care gene circuit-based diagnostics. *Nature Communications* **12**, 1–10 (2021).

39. Takahashi, M. K. *et al.* A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. *Nature Communications* **9**, 1–12 (2018).

40. Hong, F. *et al.* Precise and Programmable Detection of Mutations Using Ultraspecific Riboregulators. *Cell* **180**, 1018-1032.e16 (2020).

41. Hong, F. *et al.* Precise and Programmable Detection of Mutations Using Ultraspecific Riboregulators. *Cell* **180**, 1018-1032.e16 (2020).

42. Zadeh, J. N., Wolfe, B. R. & Pierce, N. A. Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry* **32**, 439–452 (2011).

43. Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).

44. Kazuta, Y., Matsuura, T., Ichihashi, N. & Yomo, T. Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. **118**, 554–557 (2014).

45. Nguyen, C. *et al.* Recent advances in our understanding of the environmental,

epidemiological, immunological, and clinical dimensions of coccidioidomycosis. *Clinical microbiology reviews* **26**, 505–525 (2013).

46. Tsang, C. A. *et al.* Enhanced Surveillance of Coccidioidomycosis, Arizona, USA, 2007–2008. *Emerging Infectious Diseases* **16**, 1738 (2010).
47. Vinjé, J. Advances in laboratory methods for detection and typing of norovirus. *Journal of Clinical Microbiology* vol. 53 373–381 (2015).
48. Bartsch, S. M., Lopman, B. A., Ozawa, S., Hall, A. J. & Lee, B. Y. Global economic burden of norovirus gastroenteritis. *PLoS ONE* **11**, (2016).
49. Scallan, E. *et al.* Foodborne Illness Acquired in the United States—Major Pathogens. *Emerging Infectious Diseases* **17**, 7 (2011).
50. Johns Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering(CSSE) at Johns Hopkins University. https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6.
51. Lisboa Bastos, M. *et al.* Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *BMJ* **370**, 2516 (2020).
52. Miao, F. *et al.* Rapid and Sensitive Recombinase Polymerase Amplification Combined With Lateral Flow Strip for Detecting African Swine Fever Virus. *Frontiers in Microbiology* **10**, 1004 (2019).
53. Juers, D. H., Matthews, B. W. & Huber, R. E. LacZ β-galactosidase: Structure and function of an enzyme of historical and molecular biological importance. *Protein*

Science **21**, 1792–1807 (2012).

54. Rabinowitz, J. A., Lainson, J. C., Johnston, S. A. & Diehnelt, C. W. Non-natural amino acid peptide microarrays to discover Ebola virus glycoprotein ligands. *Chemical Communications* **54**, 1417–1420 (2018).

55. Gupta, N. *et al.* Whole-Virus Screening to Develop Synbodies for the Influenza Virus. *Bioconjugate chemistry* **27**, 2505–2512 (2016).

56. Piepenburg, O., Williams, C. H., Stemple, D. L. & Armes, N. A. DNA detection using recombination proteins. *PLoS Biology* **4**, 1115–1121 (2006).

57. Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic acids research* **28**, 63 (2000).

58. Joung, J. *et al.* Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. *New England Journal of Medicine* **383**, 1492–1494 (2020).

59. Broughton, J. P. *et al.* CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology* (2020) doi:10.1038/s41587-020-0513-4.

60. Chandrasekaran, S. S. *et al.* Rapid detection of SARS-CoV-2 RNA in saliva via Cas13. *Nature Biomedical Engineering 2022* 1–13 (2022) doi:10.1038/s41551-022-00917-y.

61. Gupta, R. K. Will SARS-CoV-2 variants of concern affect the promise of vaccines? *Nature Reviews Immunology 2021 21:6* **21**, 340–341 (2021).

62. Fontanet, A. *et al.* SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet* **397**, 952–954 (2021).

63. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, (2021).

64. Abu-Raddad, L. J. *et al.* Severity, Criticality, and Fatality of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Beta Variant. *Clinical Infectious Diseases* (2021) doi:10.1093/CID/CIAB909.

65. Madhi, S. A. *et al.* Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *New England Journal of Medicine* **384**, 1885–1898 (2021).

66. Cohen, J. South Africa suspends use of AstraZeneca's COVID-19 vaccine after it fails to clearly stop virus variant. *Science* (2021) doi:10.1126/SCIENCE.ABG9559.

67. Wadman, M. Novavax vaccine delivers 89% efficacy against COVID-19 in U.K.—but is less potent in South Africa. *Science* (2021) doi:10.1126/SCIENCE.ABG8101.

68. Faria, N. R. *et al.* Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science (New York, N.y.)* **372**, 815 (2021).

69. Pouwels, K. B. *et al.* Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nature Medicine 2021 27:12* **27**, 2127–2135 (2021).

70. Lopez Bernal, J. *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine* **385**, 585–594 (2021).

71. Andeweg, S. P. *et al.* Protection of COVID-19 vaccination and previous infection

against Omicron BA.1, BA.2 and Delta SARS-CoV-2 infections. *Nature Communications 2022 13:1* **13**, 1–9 (2022).

72. Vitiello, A., Ferrara, F., Auti, A. M., Di Domenico, M. & Boccellino, M. Advances in the Omicron variant development. *Journal of Internal Medicine* **292**, 81 (2022).

73. Mohapatra, R. K. *et al.* The recently emerged BA.4 and BA.5 lineages of Omicron and their global health concerns amid the ongoing wave of COVID-19 pandemic – Correspondence. *International Journal of Surgery (London, England)* **103**, 106698 (2022).

74. Daniels, R. S. *et al.* A Sanger sequencing protocol for SARS-CoV-2 S-gene. *Influenza and other respiratory viruses* **15**, 707–710 (2021).

75. Methods for the detection and characterisation of SARS-CoV-2 variants - second update. https://www.ecdc.europa.eu/en/publications-data/methods-detection-and-characterisation-sars-cov-2-variants-second-update.

76. Erster, O. *et al.* Rapid and High-Throughput Reverse Transcriptase Quantitative PCR (RT-qPCR) Assay for Identification and Differentiation between SARS-CoV-2 Variants B.1.1.7 and B.1.351. *Microbiology Spectrum* **9**, (2021).

77. Korukluoglu, G. *et al.* 40 minutes RT-qPCR Assay for Screening Spike N501Y and HV69-70del Mutations. *bioRxiv* 2021.01.26.428302 (2021) doi:10.1101/2021.01.26.428302.

78. De Puig, H. *et al.* Minimally instrumented SHERLOCK (miSHERLOCK) for

CRISPR-based point-of-care diagnosis of SARS-CoV-2 and emerging variants. *Science Advances* **7**, 23–26 (2021).

79. Fozouni, P. *et al.* Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell* **184**, 323-333.e9 (2021).

80. Casati, B. *et al.* Rapid, adaptable and sensitive Cas13-based COVID-19 diagnostics using ADESSO. *Nature Communications 2022 13:1* **13**, 1–11 (2022).

81. Arizti-Sanz, J. *et al.* Simplified Cas13-based assays for the fast identification of SARS-CoV-2 and its variants. *Nature Biomedical Engineering 2022* 1–12 (2022) doi:10.1038/s41551-022-00889-z.

82. Chakraborty, D., Agrawal, A. & Maiti, S. Rapid identification and tracking of SARS-CoV-2 variants of concern. *The Lancet* **397**, 1346–1347 (2021).

83. Liang, Y. *et al.* Detection of Major SARS-CoV-2 Variants of Concern in Clinical Samples via CRISPR-Cas12a-Mediated Mutation-Specific Assay. *ACS Synthetic Biology* (2021) doi:10.1021/ACSSYNBIO.1C00643/SUPPL_FILE/SB1C00643_SI_001.PDF.

84. Welch, N. L. *et al.* Multiplexed CRISPR-based microfluidic platform for clinical testing of respiratory viruses and identification of SARS-CoV-2 variants. *Nature Medicine 2022 28:5* **28**, 1083–1094 (2022).

85. Woo, C. H., Jang, S., Shin, G., Jung, G. Y. & Lee, J. W. Sensitive fluorescence detection of SARS-CoV-2 RNA in clinical samples via one-pot isothermal ligation and transcription. *Nature Biomedical Engineering* **4**, 1168–1179 (2020).

86. Zhang, T. *et al.* A paper-based assay for the colorimetric detection of SARS-CoV-2 variants at single-nucleotide resolution. *Nature Biomedical Engineering 2022* 1–11 (2022) doi:10.1038/s41551-022-00907-0.

87. Pardee, K. *et al.* Portable, On-Demand Biomolecular Manufacturing. *Cell* **167**, 248-259.e12 (2016).

88. Carr, A. R. *et al.* Toward Mail-in-Sensors for SARS-CoV-2 Detection: Interfacing Gel Switch Resonators with Cell-Free Toehold Switches. *ACS Sensors* **7**, 806–815 (2022).

89. Piepenburg, O., Williams, C. H., Stemple, D. L. & Armes, N. A. DNA Detection Using Recombination Proteins. *PLoS Biology* **4**, e204 (2006).

90. Compton, J. Nucleic acid sequence-based amplification. *Nature* vol. 350 91–92 (1991).

91. Wu, Y., Liu, S.-X., Wang, F. & Zeng, M.-S. Room Temperature Detection of Plasma Epstein-Barr Virus DNA with CRISPR-Cas13. *Clinical chemistry* **000**, 2018–2019 (2019).

92. Mousavi, S. *et al.* A multiplexed, electrochemical interface for gene-circuit-based sensors. doi:10.1038/s41557-019-0366-y.

93. Ganguli, A. *et al.* Rapid isothermal amplification and portable detection system for SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 22727–22735 (2020).

94. Wu, K., Yan, Z. & Green, A. A. Computational Design of RNA Toehold-Mediated

Translation Activators. *Methods in Molecular Biology* **2518**, 33–47 (2022).

95. Zanzottera, C. *et al.* Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations . *Human Mutation* **39**, 593–620 (2018).

96. Walsh, T. *et al.* Genetic Predisposition to Breast Cancer Due to Mutations Other Than BRCA1 and BRCA2 Founder Alleles Among Ashkenazi Jewish Women. *JAMA Oncology* **3**, 1647–1653 (2017).

97. Leslie, H. H., Spiegelman, D., Kruk, M. E. & Leslie, H. H. Republic of Tanzania. 738–748 (2017).

98. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature biotechnology* **19**, 751–755 (2001).

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

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

101. Roelfsema, J. H., Peters, D. J. M. & Breuning, M. H. Detection of translation terminating mutations in the PKD1 gene. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* **11 Suppl 6**, 5–9 (1996).

102. Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nature Communications* **7**, 1–9 (2016).

103. Feng, S. *et al.* Improved split fluorescent proteins for endogenous protein labeling.

Nature Communications **8**, (2017).

104. Feng, S. *et al.* Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses. *Communications Biology* **2**, 1–12 (2019).

105. Welply$, J. K., Fowler, A. V & Zabing, I. *P-Galactosidase a-Complementation OVERLAPPING SEQUENCES*. THE JOURNAL OF BIOLOGICAL CHEMISTRY* vol. 256 (1981).

106. Nishiyama, K., Ichihashi, N., Kazuta, Y. & Yomo, T. Development of a reporter peptide that catalytically produces a fluorescent signal through α-complementation. *Protein Science : A Publication of the Protein Society* **24**, 599 (2015).

107. Langley, K. E., Villarejo, M. R., Fowler, A. V, Zamenhof, P. J. & Zabin, I. *Molecular Basis of O-Galactosidase a-Complementation (protein sequencing/protein conformation/deletion mutant)*. vol. 72 (1975).

108. Chen, J. S. *et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity*. http://science.sciencemag.org/.

109. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. (2005) doi:10.1038/nbt1044.

110. Kim, Y. *et al.* Single-strand RPA for rapid and sensitive detection of SARS-CoV-2 RNA. *medRxiv* 2020.08.17.20177006 (2020) doi:10.1101/2020.08.17.20177006.

111. Owen, M. J. *et al.* Rapid Sequencing-Based Diagnosis of Thiamine Metabolism

Dysfunction Syndrome. *New England Journal of Medicine* **384**, 2159–2161 (2021).

112. Daher, R. K., Stewart, G., Boissinot, M., Boudreau, D. K. & Bergeron, M. G. Influence of sequence mismatches on the specificity of recombinase polymerase amplification technology. *Molecular and Cellular Probes* **29**, 116–121 (2015).

113. Castiella, A. *et al.* Impact of H63D mutations, magnetic resonance and metabolic syndrome among outpatient referrals for elevated serum ferritin in the Basque Country. *Annals of Hepatology* **14**, 333–339 (2015).

114. Riordan, J. R. CFTR Function and Prospects for Therapy. *https://doi.org/10.1146/annurev.biochem.75.103004.142532* **77**, 701–726 (2008).

115. Welply, J. K., Fowler, A. V & Zabing, I. P-Galactosidase a-Complementation. **256**, 6804–6810 (1981).

116. Sokolenko, A. P. *et al.* Large family with both parents affected by distinct BRCA1 mutations: Implications for genetic testing. *Hereditary Cancer in Clinical Practice* **7**, 2 (2009).

117. Lavickova, B. & Maerkl, S. J. A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-Free System. *ACS Synthetic Biology* **8**, 455–462 (2019).

118. Khodakov, D., Li, J., Zhang, J. X. & Zhang, D. Y. Highly multiplexed rapid DNA detection with single-nucleotide specificity via convective PCR in a portable device. *Nature Biomedical Engineering* **5**, 702–712 (2021).

119. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection

with engineered self-assembling fragments of green fluorescent protein. (2005) doi:10.1038/nbt1044.

120. Jung, J. K. *et al.* Cell-free biosensors for rapid detection of water contaminants. *Nature Biotechnology* **38**, (2020).

121. Tamura, R., Jiang, F., Xie, J. & Kamiyama, D. Multiplexed labeling of cellular proteins with split fluorescent protein tags. doi:10.1038/s42003-021-01780-4.

122. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).

123. Didovyk, A., Tonooka, T. & Tsimring, L. Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression. (2017) doi:10.1021/acssynbio.7b00253.

124. Fozouni, P. *et al.* Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell* **184**, 323-333.e9 (2021).

125. Nguyen, P. Q. *et al.* Wearable materials with embedded synthetic biology sensors for biomolecule detection. *Nature Biotechnology 2021 39:11* **39**, 1366–1374 (2021).

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]