In silico prediction of COVID-19 test efficiency with DinoKnot

Tara Newman
University of Victoria
taranewman@uvic.ca

Hiu Fung Kevin Chang
University of Alberta

Hosna Jabbari
University of Victoria
jabbari@uvic.ca

Abstract—The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel coronavirus spreading across the world causing the disease COVID-19. COVID-19 is diagnosed by quantitative reverse-transcription polymer chain reaction (qRT-PCR) testing which utilizes different primer-probe sets depending on the assay used. Using in silico analysis we aimed to determine how the secondary structure of the SARS-CoV-2 RNA genome affects the interaction between the reverse primer during qRT-PCR and how it relates to the experimental primer-probe test efficiencies. We introduce the program DinoKnot (Duplex Interaction of Nucleic acids with pseudoKnots) that follows the hierarchical folding hypothesis to predict the secondary structure of two interacting nucleic acid strands (DNA/RNA). DinoKnot is the first program that utilizes stable stems in both strands as a guide to predict their interaction structure. Using DinoKnot we predicted the interaction of the reverse primers used in four common COVID-19 qRT-PCR tests with the SARS-CoV-2 RNA genome. In addition, we predicted how 12 mutations in the primer/probe binding region may affect the primer/probe ability and subsequent SARS-CoV-2 detection. We identified three mutations that may prevent primer binding, reducing the ability for SARS-CoV-2 detection. Furthermore, we investigated the effect of mutations in two variants of concern (UK and South Africa) on the efficacy of the existing primer-probe sets. Despite mutations, we did not detect deviation in primer binding when compared to the reference target strand. We believe our contributions can aid in the design of more sensitive SARS-CoV-2 diagnosis tests.

Index Terms—SARS-CoV-2, COVID-19, qRT-PCR, Diagnosis

I. INTRODUCTION

SARS-CoV-2 is a positive sense RNA virus and is most commonly detected using qRT-PCR on samples collected by nasopharyngeal swabs [1]. During qRT-PCR, the reverse primer first binds to the positive sense SARS-CoV-2 RNA genome so that the reverse transcriptase (RT) can use the primer to generate the complementary DNA (cDNA) of the negative sense strand as shown in Fig 1 [2].

The primer-probe set used to detect SARS-CoV-2 depends on which qRT-PCR assay is used. The China Center for Disease Control (China CDC), United States CDC (US CDC), Charité Institute of Virology, Universitätsmedizin Berlin (Charité), and Hong Kong University (HKU) have developed the most common assays to detect SARS-CoV-2 that all target different areas of the SARS-CoV-2 genome [3]. In a recent publication, Vogels *et al.* compared the analytical efficiencies and sensitivities of the primer-probe sets used in these assays [3].

By studying how the reverse primer interacts with the RNA genome, we can predict if the generation of the initial cDNA template from the RNA strand proceeds as expected during the qRT-PCR test. We do not investigate the interaction of the forward primer and the probes with the SARS-CoV-2 genome because the important interactions of these oligonucleotides are involved in binding to the cDNA, not the positive sense RNA genome.
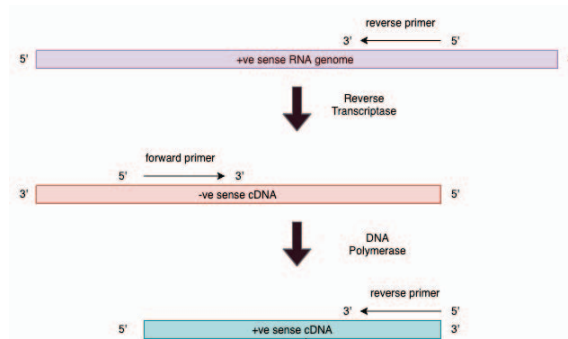

Fig. 1. **Interaction of the primers to the RNA/cDNA strands during qRT-PCR amplification.** The reverse primer first binds to the target complementary sequence on the SARS-CoV-2 positive (+ve) sense RNA genome. The reverse transcriptase then generates the negative (-ve) sense complementary DNA (cDNA) strand. The forward primer then binds to the negative sense cDNA strand and the DNA polymerase generates the positive sense cDNA strand. The reverse primer binds to the complementary target sequence on the positive sense cDNA and the DNA polymerase generates a new negative sense cDNA strand [2]. This process repeats for strand amplification during qRT-PCR.

We aimed to determine whether the RNA structure of the SARS-CoV-2 genome affects the binding of the reverse primers in the qRT-PCR assay and whether this correlated to the analytical efficiencies and sensitivities shown by Vogels *et al.* [3]. To study the structure of such interactions, we developed DinoKnot, a program that given two nucleic acid strands predicts their interaction structure. We used DinoKnot to predict how the RNA structure of SARS-CoV-2 changes when the reverse primer binds to the RNA genome and whether the primer binds to the expected location. We further predicted how mutations in the primer/probe binding region of the SARS-CoV-2 genome provided by Vogels *et al.* could affect the primer/probe sensitivity for SARS-CoV-2 detection [3].

First, we introduce DinoKnot and our experiment set up. We then present the interaction structures predicted by DinoKnot. We discuss the interactions and which mutations may decrease SARS-CoV-2 detection during qRT-PCR testing. Finally, we suggest the design of a more sensitive COVID-19 test and future work that can be done using DinoKnot.

II. MATERIALS AND METHODS

Interaction of the reverse primer and the SARS-CoV-2 genome is an example of a duplex DNA/RNA interaction, in which both strands (i.e. the RNA genome and the cDNA transcript) can be structured. Their structures can change upon interaction with one another to accommodate formation of more stable base pairings. Existing tools that predict structure of interaction in two molecules mostly focus on similar strands (i.e. DNA/DNA or RNA/RNA) [4, 5, 6, 7, 8, 9], and merely focus on the interaction site (i.e. ignoring the intramolecular structures) [10, 11, 12, 13, 14]. Our DinoKnot (Duplex Interaction of Nucleic acids with pseudoKnots) aims to address both of such shortcomings.

In this section we first provide an overview of the problem of secondary structure prediction for two interacting molecules. We then introduce DinoKnot, and explain how it overcomes the shortcomings of the existing methods on prediction of the secondary structure for two different types of nucleic acid strands. Finally, we provide details of our in silico system and experiment setup.

A. Secondary Structure Prediction for Interacting Molecules

An RNA is a single stranded molecule with two distinct ends, namely 5' and 3'. This single stranded molecule folds back onto itself by forming *intramolecular* base pairs. An RNA molecule is represented by a sequence, S, of its four bases, Adenine (A), Cytosine (C), Guanine (G) and Uracil (U) arranged on a line (representing the backbone) from 5' (left) to 3' (right) ends. The length of the RNA molecule is denoted by n and each base of the RNA sequence is referred to by its index i, $1 \leq i \leq n$. Complementary bases bind (form hydrogen bonds) and form base pairs ($A.U$, $C.G$, and $G.U$). Here "." represents a pairing of the two bases. A *secondary structure*, R, is then defined as a set of base pairs $i.j$, $1 \leq i < j \leq n$; $i.j$ and $k.j$ can belong to the same set if and only if $i = k$ i.e. each base may pair at most with one other base. If $i.j$ and $k.l$ are two base pairs of a secondary structure, R, such that $1 \leq i < k < j < l \leq n$, then $i.j$ crosses $k.l$. A *pseudoknotted* secondary structure refers to a structure with crossing base pairs. A *pseudoknot-free* secondary structure refers to a structure without crossing base pairs. The RNA structure forms because it is energetically favourable for bases to form paired helices. Different base pairing patterns in a secondary structure define different loop types. We note that the principles of structure prediction are essentially the same for single stranded DNAs. Computational secondary structure prediction from the base sequence is often done by finding the energetically most stable (minimum free energy) secondary structure, when each loop is assigned

an energy value. These energy values are kept in various tables and are known as energy parameters. Some parameter sets have been derived directly from experiments, and others are extrapolated based on experimentally determined values. Energy parameters are strand type specific, i.e. similar loops in an RNA molecule have different assigned energy than the ones in a DNA molecule. Existing minimum free energy (MFE) structure prediction methods find the minimum free structure for a given sequence from the pool of all possible structures. Methods for MFE pseudoknot-free structure prediction use dynamic programming to find the MFE structure [15, 16]. Since prediction of the MFE pseudoknotted structure is NP-hard [17, 18] and even inapproximable [19], methods for pseudoknotted MFE structure prediction focus on a restricted class of structures [20, 21, 22, 23, 24].

We can represent the interacting secondary structure of two nucleic acid sequences similarly, by concatenating the two strands together and keeping track of the gap between the two strands with a linker. When two strands are of similar type, the energy calculation of the concatenated sequence will be similar to that of a single strand of the same type, except for loops containing the gapped region (as they are not true loops when sequences are not concatenated). When two strands of different types interact (i.e a DNA strand binding to an RNA strand) the situation is more complicated as there is currently no energy parameters available for loops formed between the two strands.

B. DinoKnot

DinoKnot follows the relaxed hierarchical folding hypothesis [25] for prediction of the minimum free energy structure of two interacting nucleic acid strands. Following this hypothesis an RNA molecule first forms simple pseudoknot-free base pairs before forming more complex and possibly pseudoknotted structures [26]. During this process some of the originally formed base pairs may open up to accommodate more stable pairings. Existing methods based on hierarchical folding, namely HFold [23] and Iterative HFold [25], focus on single RNA structure prediction. Our DinoKnot is to the best of our knowledge the first program that follows the relaxed hierarchical folding hypothesis for prediction of pseudoknotted structure of two interacting nucleic acid molecules. Similar to HFold and Iterative HFold, DinoKnot handles a large class of pseudoknotted structures called density-2 structures [23]. This class of structures include a wide range of commonly found pseudoknotted structures including H-type pseudoknots and kissing hairpins with arbitrary nested substructures.

DinoKnot, takes a pair of nucleic acid sequences as input and returns their interaction structure with its corresponding free energy value. Each sequence can be of type RNA or DNA. We note that the minimum free energy structure of two input strands may not involve any interaction if it is energetically more favourable for each sequence to form intramolecular base pairs.

The user can optionally provide a pseudoknot-free input structure (in addition to the input sequence) if such in-

formation is available to guide the prediction. If no input structure is provided by the user, DinoKnot will generate up to 20 pseudoknot-free secondary structures (i.e. energetically favourable stems) by default for each strand. Considering all possible combinations of these structures, DinoKnot creates up to 400 sequence-structure combinations for the two strands. For each sequence-structure combination, DinoKnot (1) finds a pseudoknot-free structure that when combined to the input structure provides the minimum free energy structure given the input structure; (2) explores iteratively adding and removing base pairs to and from the input structure in search of lower energy structures than that found in part (1). To achieve this, DinoKnot follows four methods (all biologically sound and similar to the underlying methods of Iterative HFold [25]). Following these steps, DinoKnot finds multiple structures (sorted by their free energy) for the interacting structures. The output structure (in dot-bracket format) is the minimum free energy structure among this set of structures. Note that DinoKnot's output structure is guided by the originally found energetically favourable stems, and may not include the input structure for the given sequences.

DinoKnot employs the energy parameters of Andronescu et al. HotKnots V2.0 [27] for RNA structures and MultiRNAFold energy parameters [5, 28] for DNA structures. Energy parameters for perfect hybrid stacks (DNA/RNA) were obtained from [29, 30]. Similar to the work of Lorenz *et al.* [31], we estimated the energy parameters for loops formed between an RNA and a DNA molecule to be the average of similar loops formed intramolecularly in an RNA and a DNA molecule. All source code and energy parameters are available at https://github.com/HosnaJabbari/DinoKnot.

C. System and Experiment Setup

The SARS-CoV-2 reference genome NC_045512.2 was obtained from the National Center for Biotechnology Information GenBank database [32]. The sequence of the RNA transcripts used by Vogels *et al.* [3] to determine the primer efficiency were input into the programs Iterative HFold [25] to predict their secondary structure. This output is referred to as the secondary structure of the transcript prior to the interaction with the reverse primer. The primer sequences were obtained from the list of World Health Organization (WHO) protocols to diagnose COVID-19 [33]. The locations where the primers bind on the reference genome NC_045512.2 and the corresponding RNA transcript area are stated in Table I.

qRT-PCR involves the use of forward and reverse primers, a probe, reverse transcriptase (RT) and DNA polymerase. Forward and reverse primers are DNA oligonucleotides that bind to complementary sequences in order to amplify a section of the SARS-CoV-2 genome. The reverse primer is a DNA oligonucleotide that interacts with both the positive sense SARS-CoV-2 RNA genome and the positive sense cDNA transcript as shown in Fig 1. Therefore the reverse primer is involved in both RNA/DNA and DNA/DNA interactions. The forward primer is a DNA oligonucleotide interacts with the negative sense cDNA transcript. The probe is a fluorescent-

TABLE I
Primer binding location on the SARS-CoV-2 reference genome NC_045512.2 [32], THE TRANSCRIPT LOCATION IS THE AREA FROM THE REFERENCE GENOME INPUT INTO DINOKNOT TO PREDICT THE REVERSE PRIMER/SARS-CoV-2 RNA GENOME INTERACTION STRUCTURE. THE TRANSCRIPT LOCATIONS ARE BASED ON LAB PROTOCOLS USED IN [3].

Gene	Transcript location	Primer	Primer location
nsp10	13,122 - 13,825	CCDC-ORF1-F	13,342 - 13,362
		CCDC-ORF1-R	13,442 - 13,460
RdRp	15,094 - 15,976	RdRp-SARSr-F	15,431 − 15,452
		RdRp-SARSr-R	15,491 - 15,517
nsp14	18,447 - 19,294	HKU-ORF1-F	18,778 - 18,797
		HKU-ORF1-R	18,889 - 18,909
Envelope (E)	26,207 - 27,116	E-Sarbeco-F	26,269 - 26,294
		E-Sarbeco-R	26,360 - 26,381
Nucleocapsid (N)	28,068 - 29,430	CCDC-N-F	28,881 − 28,902
		CCDC-N-R	28,958 - 28,979
		HKU-N-F	29,145 - 29,166
		HKU-N-R	29,145 - 29,166
		2019-nCoV_N1-F	28,287 - 28,306
		2019-nCoV_N1-R	28,335 - 28,358
		2019-nCoV_N2-F	29,164 - 29,183
		2019-nCoV_N2-R	29,213 - 29,230
		2019-nCoV_N3-F	28,681 - 28,702
		2019-nCoV_N3-R	28,732 - 28,752

labelled DNA oligonucleotide that may bind to either the positive or negative sense cDNA transcript, depending on how it is designed. Therefore, the forward primer and probe are only involved in DNA/DNA interactions. RT is an enzyme that uses an RNA as a template to generate a complementary DNA strand. DNA polymerase is an enzyme that uses a DNA template to generate a complementary DNA strand.

DinoKnot was used with the RNA transcript region and the reverse primer to determine the location where the reverse primer binds and the secondary interaction structure. Two of the reverse primers, HKU-ORF1-R and RdRp-SARS-R, contain degenerate bases which means there are a mixture of oligonucleotides that contain different bases at the degenerate base position [34]. In these cases, all possible degenerate base combinations were predicted with DinoKnot. The dot-bracket output was visualized using VARNA [35] in arc format in which RNA backbone is represented by a horizontal line and base pairs are presented as arcs that connect the two bases.

D. Mutations

Vogels *et al.* listed mutations in the primer/probe binding area of the SARS-CoV-2 genome that occur at a frequency of greater than 0.1% [3]. To predict if these mutations affect primer/probe binding, and thus the sensitivity of SARS-CoV-2 detection, the mutated sequence of the transcript along with the affected primer/probe was entered into DinoKnot. The RNA/DNA setting was used to predict the interaction between the mutated RNA transcript and the reverse primer. The DNA/DNA setting in DinoKnot was used to predict the interaction structure between the mutated cDNA transcript and the corresponding primer/probe. The dot-bracket output was visualized using VARNA [35] in arc format.

The GISAID next hCoV-19 app was used to look at the frequency of mutations in the SARS-CoV-2 genome [36].

To investigate the effect of variants of concern on the primer-probe sets, a complete genome of SARS-CoV-2 B.1.1.7

lineage (UK variant) was obtained from the National Center for Biotechnology Information GenBank database, (accession ID: MW487270.1) [32]. A genome from the B.1.315 lineage (South African variant) was obtained from the GISAID database (accession ID: EPI_ISL_860693) [36]. BLASTn [32] was used to align the 5 gene areas of the RNA transcripts from the NC_045512.2 reference genome against the MW487270.1 and EPI_ISL_860693 sequences to detect any mutations that occurred in the RNA transcript gene areas. To predict if these changes may affect primer binding, the gene areas of the variants with mutations, along with the affected primer, was entered into DinoKnot.

E. Data Availability

All structure images, primer sequences and output data are available at https://github.com/HosnaJabbari/DinoKnot_data.

III. RESULTS

Here we first present interaction results of the reference SARS-CoV-2 genome with the four primer-probe sets. Then we explore the effect of mutations on efficacy of these primer-probe sets.

A. Reverse primer interactions with SARS-CoV-2 genome

DinoKnot predicted the interaction structures represented in Fig 2 to occur between the reverse primer (in red) with the corresponding region around its target sequence (in green). Fig. 2 represents the interaction site only. As shown in the figure, except HKU-N-R and E-Sarbeco-R, all interaction structures are pseudoknot-free.

As confirmation of stability of primer binding, we further calculated the net free energy of primer binding, $\mathcal{E}$, as follows based on the energy values shown in Table II.

$$\mathcal{E} = \Delta G_{intermolecular} - \sum \Delta G_{intramolecular}$$

The net free energy of primer binding is negative (i.e. binding is stable) in most cases except HKU-N-R when primer is allowed to form intramolecular structure (i.e. at 37°C). In addition, the net free energy of binding is negligible in case of E-Sarbeco-R at 37°C.

Vogels *et al.* experimentally found that all of the primer-probe sets had comparable analytical efficiencies that were all above 90% [3]. All primer-probe sets had comparable analytical sensitivities with a limit of detection of $100,000$ SARS-CoV-2 viral copies/mL except for the RdRp-SARSr set which had the lowest sensitivity [3].

All of the reverse primers were predicted by DinoKnot to interact with their expected target region. The 2019-nCoV_N1-R, 2019-nCoV_ N3-R, CCDC-N-R and CCDC-ORF1-R primers fully paired to their target region. This prediction agrees with the analytical efficiency and sensitivity results found by Vogels *et al.* [3].

Partial reverse primer mismatching: The HKU-ORF1-R and the 2019-nCoV_N2-R primer were predicted to pair to their target with single base mismatches. The HKU-ORF1-R primer contains the degenerate base R which means
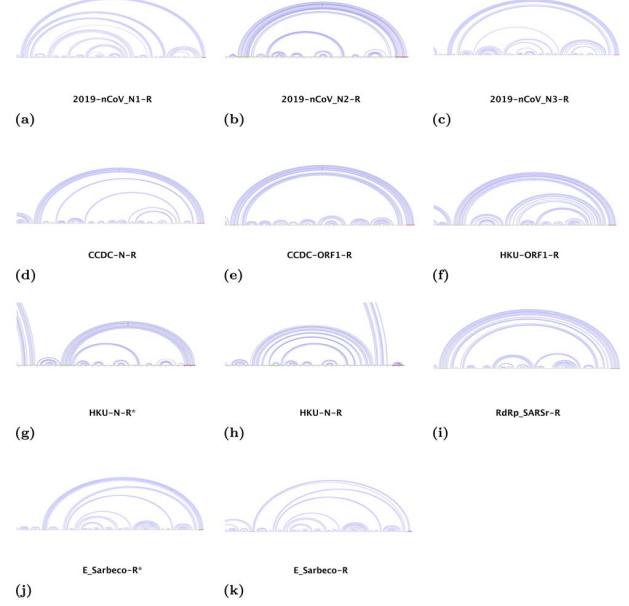

Fig. 2. **Interaction structure predicted by DinoKnot of the SARS-CoV-2 transcript genome area targeted by the reverse primer.** The expected target region of the reverse primer is highlighted in green and the reverse primer sequence is highlighted in red. The nucleic acid sequences of the RNA transcript and reverse primer were input into DinoKnot to produce the output interaction structures. The E-Sarbeco-R* and HKU-N-R* primer structures were input into DinoKnot as unfolded to simulate the primer structure during the 95°C denaturation step of the qRT-PCR assay due to primer mismatch prediction under default conditions of 37°C.

TABLE II
Energies (kcal/mol) of RNA transcript, reverse primer and reverse primer/transcript interaction structures. THE TRANSCRIPT ENERGY AND PRIMER ENERGY IS THE ENERGY OF THE TRANSCRIPT AND PRIMER STRUCTURES BEFORE THE INTERACTION. THE PRIMER BINDING ENERGY IS ENERGY OF THE INTERACTION STRUCTURE. THE TRANSCRIPT ENERGIES WERE PREDICTED BY ITERATIVE HFOLD [25], PRIMER ENERGIES PREDICTED WERE BY SIMFOLD [4], AND PRIMER BINDING ENERGIES WERE PREDICTED BY DINOKNOT.

Primer	Transcript Energy (kcal/mol)	Primer Energy (kcal/mol)	Primer Binding Energy (kcal/mol)
2019-nCoV_N1-R	-267.24	-5.90	-277.67
2019-nCoV_N2-R	-267.24	0	-276.06
2019-nCoV_N3-R	-267.24	-3.75	-272.3
CCDC-N-R	-267.24	-1.30	-274.67
CCDC-ORF1-R	-133.39	0.2	-138.54
HKU-ORF1-R	-138.89	-3.10	-151.76
HKU-N-R	-267.24	0*	-280.82
		-3.30	-267.91
RdRP-SARSr-R	-150.03	-3.70	-162.37
E-Sarbeco-R	-148.74	0*	-167.39
		-1.50	-150.48

*E-Sarbeco and HKU-N-R primer structure forced unfolded for primer to bind to targeted area.

the primer may contain an A or a G at this position. The target base is a U in the SARS-CoV-2 genome. The interaction structure was predicted twice, both with an A and G in the position of degenerate base R. The output produced the expected match, except when the primer was set as a G, the

G.U base pair did not bind but it did not affect the rest of the primer area from binding to the target area. The last base at the 3' end of the 2019-nCoV_N2-R primer did not pair with any nucleotide but the rest of the primer was predicted to bind to its expected target region. Although these primers contain a single predicted mismatch, the analytical efficiencies and sensitivities of these primer probe sets were comparable to the other primer-probes tested experimentally [3].

HKU-N-R and E-Sarbeco-R: The HKU-N-R and E-Sarbeco-R primers did not pair as expected to their target region when DinoKnot was not given an input structure (i.e. when both strands were free to assume possible structures before interacting with one another). The HKU-N-R primer was predicted to bind to itself, rather than its target region. When the HKU-N-R primer structure was forced to be unfolded, the first base of the HKU-N-R primer at the 5' end did not bind to its expected nucleotide but the rest of the primer interacted with the target region as expected. Forcing the primer to be unfolded is a prediction of the primer structure after the qRT-PCR test denaturation step at 95°C in the work done by Vogels *et al.* [3] The predicted interaction structure for the E-Sarbeco-R primer when Dinoknot was not given an input structure results in the bases at positions 8-15 in the reverse primer binding as expected to the target region. However, bases 5-7 paired with bases 16-18 and the remaining bases of the 22bp primer remained unpaired. When the primer structure was input into DinoKnot as unfolded, as done for the HKU-N-R primer, the E-Sarbeco-R primer fully paired to its target region. Both the HKU-N-R and E-Sarbeco primer-probe sets were shown experimentally by Vogels *et al.* to have analytical efficiencies and sensitivities that are comparable to the other primer-probe sets that were predicted by DinoKnot to bind completely to their target region; this supports the structures predicted with their primer structure is forced unfolded [3].

RdRp-SARSr-R: The RdRp-SARSr-R primer contains two degenerate bases, R and S. This means that an A or G may be present in the position of the R and a C or G may be present in the position of the S. All combinations were given as input to DinoKnot and the predicted structures are shown in Fig 3.

Fig 3 represents the interaction site only. The RdRp-SARSr-R primer interacted as expected with its target region when the base at the R position was entered into DinoKnot as an A, and the base at the S position was entered as a G. The target nucleotide is a U at both of the positions where the R and S base bind. The structures were also predicted with an A input at the position of the degenerate base S. This change was suggested by Vogels *et al.* to possibly increase the sensitivity of this primer-probe set because it was found experimentally to have the lowest sensitivity of all the primer-probe sets [3]. However, this change resulted in the primer not binding to its target region as predicted by DinoKnot.

B. Mutations in primer/probe binding regions

The mutations listed by Vogels *et al.* are shown in Table III, along with the interaction structure energies [3].
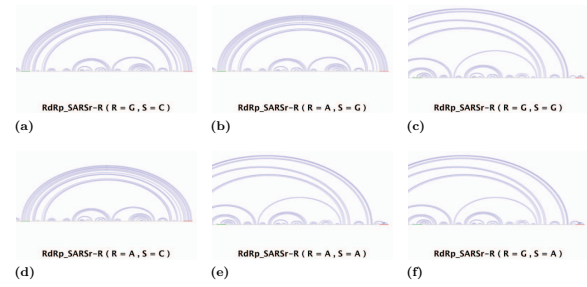

Fig. 3. **Interaction structures of the RdRp-SARS-R primer predicted by DinoKnot with all of the possible base combinations from the degenerate primers.** The RdRp-SARS-R primer contains the degenerate bases R and S, which means an A or G may be present at the R position and a C or G may be present at the S position. All possible combinaions were predicted. The S position was also input as an A to predict if this change may increase primer sensitivity.

TABLE III
Structure energy differences predicted with mutations in the primer binding region of the SARS-CoV-2 genome obtained from Vogels *et al.* [3] . THE PRIMER BINDING ENERGY WAS PREDICTED BY DINOKNOT. THE PROBE/TRANSCRIPT AND FORWARD PRIMER/TRANSCRIPT INTERACTIONS ARE DNA/DNA INTERACTIONS SINCE THESE OLIGONUCLEOTIDES INTERACT WITH THE CDNA STRANDS [2]. MUTATIONS IN THE REVERSE PRIMER BINDING REGION INCLUDE BOTH DNA/DNA AND DNA/RNA INTERACTION SINCE THE REVERSE PRIMER INTERACTS WITH THE SARS-COV-2 GENOME ALONG WITH THE NEGATIVE SENSE CDNA STRAND [2]

Primer-probe	Mutation in ref. genome	Mutation Position	Primer Binding Energy (kcal/mol)	Interaction type
CCDC-N-F	no mutation		-283.27	DNA/DNA
	G → A	28,881	-282.43	DNA/DNA
	G → A	28,882	-281.6	DNA/DNA
	G → C	28,883	-279.62	DNA/DNA
CCDC-ORF1-F	no mutation		-138.05	DNA/DNA
	C → T	13,358	-133.77	DNA/DNA
CCDC-ORF1-P	no mutation		-144.82	DNA/DNA
	T → G	13,402	-147.82	DNA/DNA
E-Sarbeco-R	no mutation		-167.39	RNA/DNA
	no mutation		-167.77	DNA/DNA
	G → T	26,370	-149.57	RNA/DNA
	G → T	26,370	-149.49	DNA/DNA
HKU-N-F			-279.36	DNA/DNA
	T → C	29,148	-275.41	DNA/DNA
2019-nCoV_N1-P	no mutation		-282.37	DNA/DNA
	C → T	28,311	-279.49	DNA /DNA
2019-nCoV_N1-R	no mutation		-277.67	RNA/DNA
	no mutation		-277.85	DNA/DNA
	C → A	28,344	-273.22	RNA/DNA
	C → A	28,344	-273.4	DNA/DNA
2019-nCoV_N3-F	no mutation		-280.31	DNA/DNA
	T → C	28,688	-279.86	DNA/DNA
2019-nCoV_N3-R	no mutation		-272.3	RNA/DNA
	no mutation		-273.04	DNA/DNA
	C → T	28,739	-267.95	RNA/DNA
	C → T	28,739	-268.57	DNA/DNA

The resulting structures (interaction site only) are shown in Fig 4. The effects of mutations on the interaction structures are compared to the primer binding ability of the primer/probes to the reference genome NC_045512.2. All primer/probes paired as expected to the target area of the reference genome with the exception of the 2019-nCoV_N3-F primer which has one unpaired base at the 3' end.

Mutations that cause no change to the interaction

2019_nCoV-N1-P 28,311C>T (a)
2019_nCoV-N1-R 28,344C>A (DNA) (b)
2019_nCoV-N1-R 28,344C>A (RNA) (c)
2019_nCoV-N3-F 28,688T>C (d)
2019_nCoV-N3-R 28,739C>T (DNA) (e)
2019_nCoV-N3-R 28,739C>T (RNA) (f)
CCDC_N-F 28,881G>A (g)
CCDC_N-F 28,882G>A (h)
CCDC_N-F 28,883G>C (i)
CCDC-ORF1-F 13,358C>T (j)
CCDC_ORF1-P 13,402T>G (k)
E_Sarbeco_R 26,370G>T (DNA) (l)
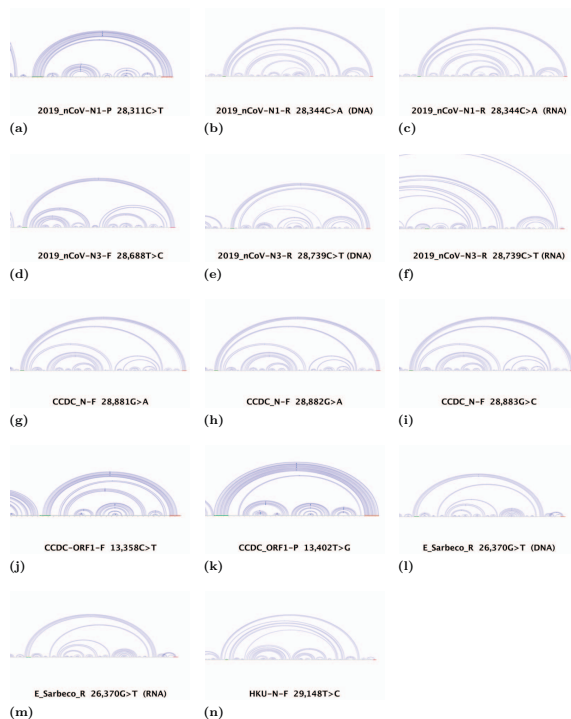E_Sarbeco_R 26,370G>T (RNA) (m)
HKU-N-F 29,148T>C (n)

Fig. 4. **Interaction structures of primer/probes with the transcript regions containing mutations in the primer/probe binding region predicted by DinoKnot.** The expected target region is highlighted in green and the primer/probe sequence is highlighted in red. The mutated base is highlighted in pink.

structure: The mutations in the primer binding region of the CCDC-N-F at base pair positions $28,881$ and $28,882$ had no affect on the primer binding ability.

The mutation in the primer binding region of 2019-nCoV_N3-F mutation at base pair position $28,688$ did not change the structure. The primer is still able to bind to the mutated base, but the final base at the 3' end of the primer remains unpaired. However, this unpaired base is the same in the structure of the 2019-nCoV_N3-F primer interacting with the reference genome NC_045512.2.

Mutations that cause partial mismatching: The following mutations resulted in a single mismatch between the primer/probe and the mutated base: CCDC-ORF1-P at base position $13,402$, 2019-nCoV_N1-R at base position $28,344$ and CCDC-N-F at base position $28,883$. The other bases in the probe/primer were still able to pair with their target nucleotides.

The mutation in the primer binding region of CCDC-ORF1-F at base position $13,358$ affects the ability of the 5 bases at the 3' end of the primer to bind. The remaining 16 out of 21 bases of the forward primer are still able to bind to their target nucleotides.

The mutation in the probe binding region of 2019-nCoV_N1-P at base position $28,311$ is predicted to cause a mismatch of the first three bases at the 5' end of the probe. The remaining 21 nucleotides of the probe still bind as expected.

Mutations that cause complete mismatching: The mutation in the primer binding region of E-Sarbeco-R at base pair position $26,370$ is predicted to result in a complete mismatch between the reverse primer and the SARS-CoV-2 genome. The primer does not bind to its target. There is some partial binding 600 bp downstream of the target region and partial base pairing of the primer to itself. The mutation also causes the same complete mismatch between the primer and the positive sense cDNA strand.

The mutation in the primer binding region of HKU-N-F at base position $29,148$ is predicted to result in complete mismatch between the forward primer and the negative sense cDNA strand.

The mutation in the primer binding region of 2019-nCoV_N3-R mutation at base position $28,739$ is predicted to result in a complete mismatch between the reverse primer and the SARS-CoV-2 RNA genome. The primer does not bind to the target area. However, the reverse primer is able to bind to the target area on the positive sense cDNA, except with one base pair mismatch at the mutated base.

B.1.1.7 and B.1.315 variants of concern: Base pair differences from the reference genome were detected in the RdRp and N gene areas of the B.1.1.7 variant. There was one base difference in the RdRp transcript area. The equivalent area on the MW487270.1 virus for the RdRp region is 15034–15916bp. There were 9 base differences in the N gene region which is 27999–29361 bp in the MW487270.1 virus. The E, nsp14 and nsp10 transcript area sequences aligned with 100% identity to the reference genome.

The N, nsp14 and E gene areas differed in the B.1.315 variant from the reference genome. In the N gene region there were 3bp differences between the reference genome and EPI_ISL_860693 gene area of 28013-29375 bp. In the nsp14 gene region there was 1 base difference in the EPI_ISL_860693 gene area of 18401– 19240 bp. In the E gene region there were 2 base differences in the EPI_ISL_860693 gene area of $26152 - 26961$ bp. The nsp10 and RdRp transcript areas aligned with 100% identity to the reference genome.

None of the mutations in the RdRp and N gene areas of MW487270.1 (B.1.1.7 variant) and the N, nsp14 and E gene areas of EPI_ISL_860693 (B.1.315 variant) were predicted to affect primer binding. The primer binding results predicted by DinoKnot were the same as shown for the reference genome.

IV. DISCUSSION

We used DinoKnot to predict the secondary structure of the interaction between RNA transcripts from the SARS-CoV-2 genome and the reverse primers from nine primer-probe sets in order to see if the RNA structure of SARS-CoV-2 affected reverse primer binding. In addition, we predicted how mutations listed by Vogels *et al.* in the primer/probe binding regions may affect the binding ability of the primer/probes, and thus, SARS-CoV-2 detection. We found that all reverse primers were able to interact with their target regions and predicted partial mismatching to occur between some reverse primers.

475

We also predicted three mutations to prevent the ability of the primer/probe to bind to their target region.

Reverse primer interaction with the SARS-CoV-2 RNA genome: The following primers were predicted to have partial mismatching in the primer binding region: 2019-nCoV_N2-R, HKU-N-R, E-Sarbeco-R, and RdRp-SARSr-R. The 2019-nCoV_N2, E-Sarbeco and HKU-N primer-probe sets had analytical efficiencies and sensitivities that are similar to the other primer-probe sets [3]. The RdRp-SARSr primer-probe set had an analytical efficiency comparable to the other primer-probe sets but a reduced sensitivity [3].

The single base mismatch in the 2019-nCoV_N2-R primer is unlikely to affect the ability of the reverse transcriptase to generate the cDNA strand. The reverse transcriptase generates the cDNA strand in the 5' to 3' direction so a mismatch at the 3' end is unlikely to affect strand synthesis. As well, this primer-probe sets had high analytical efficiency and sensitivity shown by Vogels *et al.* which supports that this mismatch is unlikely to have an effect on cDNA strand synthesis [3]

The mismatching in the E-Sarbeco-R and HKU-N-R primers occur when the primer is assumed to have structure prior to interaction with the target. However, when the primer was forced in DinoKnot as unfolded, the program predicts the E-Sarbeco-R primer to fully bind to its target sequence and the HKU-N -R primer to bind with a single mismatch at the 5' end of the primer. Since the qRT-PCR assay used by Vogels *et al.* has a denaturation step of 95°C and an annealing step of 55°C, it can be assumed that the primer is fully unfolded at those temperatures because the E-Sarbeco-R and HKU-N-R had primer efficiencies comparable to the other primers as determined by Vogels *et al.* [3].

The RdRp-SARSr-R primer contains two degenerate bases, R and S. Vogels *et al.* determined that 990 out of 992 of their SARS-CoV-2 clinical samples contain a T at the position in the reference genome NC_045512.2 where the S base binds to its target [3]. This primer-probe set had the lowest sensitivity out of all of the primers [3]. During the qRT-PCR test by Vogels *et al.* this primer-probe set had the highest Ct values, 6-10 Ct values higher than the other sets, and is unable to detect low viral amounts [3]. DinoKnot was run with all possible combinations of the position of the degenerate base R input as an A or G and S input as a C or G shown in Fig 3.

The combination that produced the expected match was when an A was input at the R position and a G was input at the S postion. Vogels *et al.* proposed that changing the S to an A in the reverse primer could increase the sensitivity of the primer-probe set [3]. To test this, the primer was given as input to DinoKnot with an A input at the S position. All combinations where S was input as an A resulted in the primer not binding to the target area. The result was the primer binding to itself and partial binding to the 5' end of the transcript, over 400 bases downstream from the target area. Therefore, this result indicates that changing the S to an A is unlikely to increase the sensitivity of the primer even though the target base pair is a U in the RNA genome and a T on the cDNA transcript. Based on these results, we hypothesize that the low sensitivity issue of the RdRp-SARSr primer-probe set may be due to the predicted primer mismatch when a G is present in the R and S position of the primer. Although the concentration of each base combination at the degenerate R and S positions is not stated in the protocol, if the assumption is made that the four combinations are present in equal amounts, this may explain the low sensitivity issue of the RdRp primer probe-set [33, 3]. The base combination where a G is present at both the R and S position is predicted to mismatch to the target region. This would lower the RdRp-SARSr-R concentration that is capable of binding to its target region. If the effective concentration of the RdRp-SARSr-R primer is lower than the other primer-probe sets, this may explain why this set had low sensitivity issues due to higher Ct values shown experimentally by Vogels *et al.* [3]. Therefore, we suggest that changing the base at the R position to an A and the base at the S position to a G may increase the RdRp-SARSr primer-probe set sensitivity since this base combination was predicted to perfectly bind to its target region.

Mutations in the expected primer/probe binding region: Vogels *et al.* looked at 992 clinical samples and identified mutations in the expected primer binding regions that could decrease the primer sensitivity for SARS-CoV-2 detection [3]. We used DinoKnot to predict how those mutations affect the ability of the primer/probe to bind to its target sequence. Mutations in the expected binding region of the forward primer and probe were DNA/DNA interactions and mutations in the reverse primer expected binding region show both the RNA/DNA interaction and the DNA/DNA interaction.

Based on the interaction structures predicted by DinoKnot, the mutations in the primer binding regions of the E-Sarbeco-R, N-HKU-F, and 2019-nCoV_N3-R primers are most likely to have the greatest effect on decreasing primer sensitivity for SARS-CoV-2 detection. This is because the interaction structures predicted by DinoKnot show no binding of the primer to the target area. All mutations resulted in an increase in the energy of binding, except for the mutation in the CCDC-ORF1-P binding region, which lowered the energy of binding. The energy increase means that binding is not as favourable as before and in environments where there is competition for binding, binding may not happen. The mutations in E-Sarbeco-R, N-HKU-F, and 2019-nCoV_N3-R binding regions caused the greatest increase in binding energy.

The mutation in the primer binding region of the E-Sarbeco-R primer at base position 26,370 is predicted by DinoKnot to cause a complete mismatch between the reverse primer and the RNA genome, as well as between the reverse primer and positive sense cDNA strand. Therefore, this single base mutation may cause the mutated SARS-CoV-2 strains to not be detected by the E-Sarbeco primer-probe set, especially since the reverse primer is needed for the generation of the cDNA strand. Even if the forward primer and probe can still bind as expected to their target sequence, if the reverse primer cannot bind to the RNA genome, the reverse transcriptase will not be able to generate the negative sense cDNA strand for the forward primer to bind to.

The mutation in the HKU-N-F primer binding region at position $29,148$ was predicted by DinoKnot to result in the forward primer not being able to bind to its target region. Without a functional forward primer, this would result in the inability to generate a negative sense cDNA copy, and therefore no amplification of either the positive sense or negative sense cDNA strand. As shown in Fig 1, the negative sense cDNA is required for the generation of the positive sense cDNA strand and any subsequent amplification. Even if the reverse primer and probe bind as expected to their target region, there would only be the generation of one positive sense cDNA strand and no amplification.

The mutation in the 2019-nCoV_N3-R primer binding region at position $28,739$ is predicted to cause a complete mismatch between the reverse primer and the RNA genome. However, the reverse primer is able to bind to the target area on the positive sense cDNA, except with one base pair mismatch at the mutated base. This mutation may cause this strain not to be detected by this primer-probe set because if the reverse primer does not bind to the target area on the SARS-CoV-2 genome, the reverse transcriptase cannot generate the positive sense cDNA strand. Therefore, the binding of the reverse primer to the negative sense cDNA strand interaction would not occur.

The CCDC-ORF1-F primer had a mismatch at the 3' end of the primer causing the 5 bases at 3' end to remain unpaired. The DNA polymerase generates the complementary strand in the 5'-3' direction. Since the 5' end still binds, the full target area may still be generated. In this primer-probe set, the probe binds to the same area of the cDNA that the forward primer binds so having the full area of the transcript generated is important. The 5 bases at the 3' end did not align to any other part of the reference genome in a NCBI BLAST search so it is unlikely but still possible that the tail could bind to another part of the genome and hinder the ability of the DNA polymerase to generate the cDNA strand, preventing SARS-CoV-2 detection [32]. This mutation should be tested experimentally in order to determine its effect on the primer-probe set's ability for SARS-CoV-2 detection.

B.1.1.7 and B.1.315 variants of concern: Since the variants did not change the predicted primer binding, we believe the primers should be able to have the same efficiency as detection of the reference genome based on our in silico prediction.

Viral Load and SARS-CoV-2 detection: Recent work by Pan *et al.* on 82 infected individuals with SARS-CoV-2 determined the viral load in sputum and swab samples to peak on day 5-6 after symptom onset, with a range of 10^4 to 10^7 copies per mL [37]. This is earlier than SARS-CoV which peaks approximately 10 days after symptom onset [37]. Vogels *et al.* determined the qRT-PCR primer-probe sets to have a limit of detection of 10^5 copies per mL except for the RdRp-SARSr primer-probe set which tested negative in samples containing 10^0-10^2 viral RNA copies/μL concentrations (which is equivalent to $100 - 100,000$ viral RNA copies/mL) [3]. Pan *et al.* determined that early after the onset of symptoms, the viral load is greater than 1×10^6 copies per mL [37]. This

is sufficient for detection by the primer-probe sets, except for RdRp-SARSr [3]. However, respiratory samples from 80 patients from different stages of infection showed a range of 641 copies per mL to 1.34×10^{11} copies per mL [37]. Therefore, the more time that passes after symptom onset may result in the viral load to be below the limit of detection during testing. If patients are tested a greater number of days after symptom onset when the viral load may be below the primer-probe set limit of detection determined by Vogels *et al.*, this could result in a false negative test. However, nasopharyngeal swabs are the most common ways to collect patient samples and the work by Pan *et al.* only looked at one nasal swab which was tested 3 days after symptom onset with a total of 1.69×10^5 copies per mL [1, 37].

In a patient led survey, 27.5% of respondents believed they had COVID-19 but tested negative for the disease [38]. This group of respondents were on average tested later, by day 16, compared to the 21% of respondents who believed they had COVID-19 and also tested positive, who were on average tested by day 10 [38]. As is a limitation of this survey we cannot verify whether the respondents who believed they had COVID-19 were in fact infected with SARS-CoV-2. However, since they were tested later, this may support the hypothesis that the viral load may be below the limit of detection if the test is taken a greater number of days after symptom onset. Therefore false negative tests are likely an issue with having enough viral load in the sample rather than an issue with the primer interaction with the SARS-CoV-2 genome since DinoKnot predicted correct primer binding for most of the primers at the higher temperatures.

We hypothesize that a SARS-CoV-2 aptamer test is beneficial to provide a lower limit of detection. Aptamers are single stranded RNA or DNA nucleotides (10-100nt) that are able to bind to targets such as viruses and proteins [39]. Aptamers' binding specificity is ensured by their secondary and tertiary structure [39]. An RNA aptamer test designed for the SARS-CoV Nucleocapsid protein showed a detection limit of 2 pg/mL [40]. A recent aptamer test for Norovirus, a positive sense RNA virus, has a limit of detection of 200 viral copies/mL [41]. This is lower than the qRT-PCR limit of detection of determined by Vogels *et al.* by a magnitude of 10^3 [3]. The lower limit of detection that is possible with aptamer tests would be beneficial in testing patients later after symptom onset.

We believe that an aptamer test can be designed for SARS-CoV-2 that both improves the detection sensitivity and provides a more rapid test compared to the qRT-PCR tests currently in use. The areas targeted in the primer-probe sets may be good targets for aptamer design, since they are specific to SARS-CoV-2. Zooming in on these regions in the GISAID h-CoV-19 app, it can be shown that these areas generally have few mutations, as compared to the number of mutations that can be seen in the entire genome as shown in Fig 5 [36].

Our algorithm and findings may be beneficial in a starting place for aptamer design. The HKU-ORF1 primer-probe set was the only set to have its reverse primer completely bind
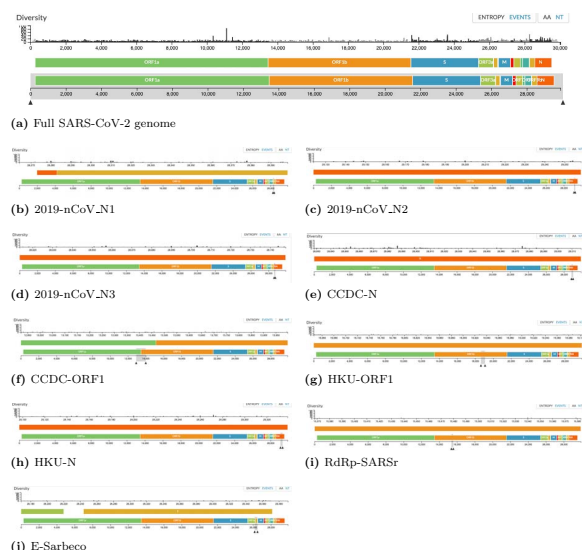
(a) Full SARS-CoV-2 genome

(b) 2019-nCoV_N1

(c) 2019-nCoV_N2

(d) 2019-nCoV_N3

(e) CCDC-N

(f) CCDC-ORF1

(g) HKU-ORF1

(h) HKU-N

(i) RdRp-SARSr

(j) E-Sarbeco

Fig. 5. **Screenshots from GISAID hCoV-19 global genetic epidemiology mutation events.** Mutation events reported to GISAID for nucleic acid positions of the (a) full SARS-CoV-2 genome and (b-j) zoomed in sections of the primer-probe sets as of March 28, 2021.

its target sequence and not contain any mutations in the primer/probe target areas in the 992 clinical samples looked at by Vogels *et al.* [3]. The GISAID h-CoV-19 app also shows the area targeted by this primer-probe set to have a low number of mutation events [36]. The 2019-nCoV_N2 primer-probe set did not have any mutations in the primer/probe binding regions listed by Vogels *et al.* but the reverse primer did contain the one base pair mismatch to the reference genome. The areas targeted by these primer-probe sets may be suitable targets for designing a SARS-CoV-2 aptamer test since these areas are specific for SARS-CoV-2 and contain few mutations.

V. CONCLUSION

We introduced the program DinoKnot that is able to predict the secondary structure of two interacting nucleotide sequences. By in silico prediction we provided evidence to support that during COVID-19 qRT-PCR tests, the reverse primers investigated here are binding to their expected target region on the SARS-CoV-2 genome. Our prediction supports that binding is not prevented by the secondary structure of the SARS-CoV-2 RNA genome. This supports the idea that the correct area of the SARS-CoV-2 genome is amplified during the qRT-PCR tests and the RNA structure does not prevent SARS-CoV-2 detection. Our prediction also provides a hypothesis to explain the low sensitivity issue of the RdRp-SARSr primer-probe set and we suggest changing the degenerate base positions in the RdRp-SARSr-R from an R to an A and an S to a G to potentially increase the sensitivity.

We also predicted how different mutations in the primer binding region affected primer binding and that the SARS-CoV-2 strains with mutations in the E-Sarbeco-R, N-HKU-F, and 2019-nCoV_N3-R primer binding regions may reduce the sensitivity of these COVID-19 tests for SARS-CoV-2 detection. Our in silico results did not detect any problem with primer-probe binding to the two variants of concern.

Finally we hypothesize that an aptamer test may be a better way to test for COVID-19 since it can provide a lower limit of detection. Further research is required to quantify the viral load at each stage of infection and determine at what point the qRT-PCR can no longer detect COVID-19 so that conclusions can be made on the impact of the limit of detection affecting false negative test results.

DinoKnot can be used as a tool to aid in the prevention of false positive results during the design COVID-19 and other nucleic acid based testing. A false positive test would involve the non-specific binding of the primer-probe sets to a source other than the target in the patient sample. Future work can be done using our program DinoKnot to predict the interaction between primer/probes with other pathogens and nucleic acid sequences that may be present in patient samples to determine what the primer/probes may be capable of binding to produce false positive results.

REFERENCES

[1] W. Wang et al. "Detection of SARS-CoV-2 in Different Types of Clinical Specimens." In: *JAMA* 323 (18 2020), pp. 1843–1844.

[2] S.A. Bustin and R. Mueller. "Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis." In: *Clin Sci (Lond)* 109 (4 2020), pp. 365–379.

[3] C.B.F. Vogels et al. "Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets." In: *Nat Microbiol* (2020).

[4] M. Andronescu et al. "RNAsoft: A suite of RNA secondary structure prediction and design software tools." In: *Nucleic Acids Res* 31.13 (2003), pp. 3416–3422.

[5] M. Andronescu, Zhi C., and A. Condon. "Secondary structure prediction of interacting RNA molecules." In: *Journal of molecular biology* 345.5 (2005), pp. 987–1001.

[6] R.M. Dirks et al. "Thermodynamic Analysis of Interacting Nucleic Acid Strands". In: *SIAM Rev.* 49.1 (2007), pp. 65–88.

[7] C. Alkan et al. "RNA-RNA Interaction Prediction and Antisense RNA Target Search". In: *Journal of Computational Biology* 13.2 (2006), pp. 267–282.

[8] Y. Kato, T. Akutsu, and H. Seki. "A grammatical approach to RNA-RNA interaction prediction". In: *Pattern Recognition* 42.4 (2009), pp. 531–538.

[9] H. Chitsaz et al. "A partition function algorithm for interacting nucleic acid strands". In: *Bioinformatics* 25.12 (2009), pp. i365–i373.

[10] R.A. Dimitrov and M. Zuker. "Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids". In: *Biophysical Journal* 87.1 (2004), pp. 215–226.

[11] M. Rehmsmeier et al. "Fast and effective prediction of microRNA/target duplexes." In: *RNA* 10.10 (2004), pp. 1507–1517.

[12] R. Lorenz et al. "ViennaRNA Package 2.0". In: *Algorithms for Molecular Biology* 6.1 (2011), p. 26.

[13] J.S. Reuter and D.H. Mathews. "RNAstructure: software for RNA secondary structure prediction and analysis." In: *BMC bioinformatics* 11.1 (2010), pp. 129+.

[14] M.B. Kery et al. "TargetRNA2: identifying targets of small regulatory RNAs in bacteria". In: *Nucleic acids research* 42.Web Server issue (2014), W124–W129.

[15] R. Nussinov and A.B. Jacobson. "Fast algorithm for predicting the secondary structure of single-stranded RNA." In: *Proceedings of the National Academy of Sciences of the United States of America* 77.11 (1980), pp. 6309–6313.

[16] M. Zuker and P. Stiegler. "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information." In: *Nucleic acids research* 9.1 (1981), pp. 133–148.

[17] T. Akutsu. "Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots," in: *Discrete Applied Mathematics* 104.1-3 (2000), pp. 45–62.

[18] R.B. Lyngsø and C.N.S. Pedersen. "Pseudoknots in RNA secondary structures". In: *Proceedings of the fourth annual international conference on Computational molecular biology*. 2000, pp. 201–209.

[19] S. Sheikh, R. Backofen, and Y. Ponty. "Impact of the Energy Model on the Complexity of RNA Folding with Pseudoknots". In: *Combinatorial Pattern Matching*. Ed. by Juha Kärkkäinen and Jens Stoye. Vol. 7354. 2012, pp. 321–333.

[20] E. Rivas and S.R. Eddy. "A dynamic programming algorithm for RNA structure prediction including pseudoknots." In: *Journal of molecular biology* 285.5 (1999), pp. 2053–2068.

[21] R.M. Dirks and N.A. Pierce. "A partition function algorithm for nucleic acid secondary structure including pseudoknots". In: *J. Comput. Chem.* 24.13 (2003), pp. 1664–1677.

[22] J. Reeder and R. Giegerich. "Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics". In: *BMC Bioinformatics* 5.1 (2004), pp. 104+.

[23] H. Jabbari, A. Condon, and S. Zhao. "Novel and efficient RNA secondary structure prediction using hierarchical folding." In: *J Comput Biol* 15.2 (2008), pp. 139–163.

[24] H. Jabbari et al. "Knotty: efficient and accurate prediction of complex RNA pseudoknot structures". In: *Bioinformatics* 34.22 (2018), pp. 3849–3856.

[25] H. Jabbari and A. Condon. "A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures". In: *BMC Bioinformatics* 15.1 (2014), pp. 147+.

[26] I. Tinoco and C. Bustamante. "How RNA folds". In: *Journal of Molecular Biology* 293.2 (1999), pp. 271–281.

[27] M. Andronescu et al. "Computational approaches for RNA energy parameter estimation". In: *RNA* 16.12 (2010), pp. 2304–2318.

[28] N. Sugimoto et al. "Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes". In: *Nucleic Acids Research* 24.22 (1996), pp. 4501–4505.

[29] F.H. Martin and I.J. Tinoco. "DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription". In: *Nucleic Acids Research* 8.10 (1980), pp. 2295–2300.

[30] N. Sugimoto et al. "Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes". In: *Biochemistry* 34.35 (1995), pp. 11211–11216.

[31] R. Lorenz, I.L. Hofacker, and S.H. Bernhart. "Folding RNA/DNA hybrid duplexes". In: *Bioinformatics* 28.19 (2012), pp. 2530–2531.

[32] NCBI Resource Coordinators. "Database resources of the National Center for Biotechnology Information". In: *Nucleic Acids Res.* 44 (2016), (D1):D7–D19.

[33] WHO. *WHO - Coronavirus disease (COVID-19) technical guidance: Laboratory testing for 2019-nCoV in humans.* Accessed:2020-08-27. 2020 (Retrieved Sept 6, 2020).

[34] J.A. Iserte et al. "Family-Specific Degenerate Primer Design: A Tool to Design Consensus Degenerated Oligonucleotides". In: *Biotechnology Research International* 2013 (2013), p. 9.

[35] K. Darty, A. Denise, and Y. Ponty. "VARNA: Interactive drawing and editing of the RNA secondary structure." In: *Bioinformatics* (15 2009), pp. 1974–1975.

[36] S. Elbe and G. Buckland-Merrett. *Data, disease and diplomacy: GISAID's innovative contribution to global health.* 2017.

[37] Y. Pan et al. "Viral load of SARS-CoV-2 in clinical samples". In: *The Lancet - Correspondence* 20 (4 2020), pp. 411–412.

[38] G. Assaf et al. *An Analysis of the Prolonged COVID-19 Symptoms Survey by Patient-Led Research Team.* Accessed: 2020-8-25. 2020. URL: https : / / patientresearchcovid19.com/research/report-1/.

[39] X. Zou et al. "Application of Aptamers in Virus Detection and Antiviral Therapy". In: *Front. Microbiol.* 10 (2019), p. 1462.

[40] D-G. Ahn et al. "RNA aptamer-based sensitive detection of SARS coronavirus nucleocapsid protein". In: *Analyst* 134 (2009), pp. 1896–1901.

[41] P. Weerathunge et al. "Ultrasensitive Colorimetric Detection of Murine Norovirus Using NanoZyme Aptasensor". In: *Anal. Chem.* 91 (5 2019), pp. 3270–3276.