UNIVERSITY OF CALIFORNIA
SANTA CRUZ

METHOD DEVELOPMENT TO AID THERAPEUTIC DISCOVERY: A BIOLAYER INTERFEROMETRY SEROLOGY TEST AND CHARACTERIZATION OF AN IMMUNE CHECKPOINT INHIBITOR

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
BIOMOLECULAR ENGINEERING AND BIOINFORMATICS

By

Nicholas Lorig-Roach
December 2023

The Dissertation of Nicholas Lorig-Roach
is approved:

Rebecca DuBois, Ph.D., Chair

Manuel Ares, Ph.D.

Seth Rubin, Ph.D.

Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

METHOD DEVELOPMENT TO AID THERAPEUTIC DISCOVERY:
A BIOLAYER INTERFEROMETRY SEROLOGY TEST AND
CHARACTERIZATION OF AN IMMUNE CHECKPOINT INHIBITOR

Nicholas Lorig-Roach

This research focuses on three primary topics: (1) the development of biolayer interferometry serology tests to detect viral antigen reactivity with microliter scale human serum samples; (2) characterization of targets and epitopes for antibody-based cancer therapeutics with an emphasis on immune checkpoint regulation of natural killer cells via KIR receptors; and (3) preliminary investigation of the human astrovirus 1 RNA-dependent-RNA-polymerase.

BLI-ISA: The biolayer interferometry based antibody test (BLI-ISA) represents a novel, high throughput platform with numerous advantages over conventional enzyme-linked immunosorbent assays (ELISA) in terms of sample requirements, labor, and overall assay time while maintaining high accuracy. Compared to the COV2T antibody test for SARS-CoV-2, BLI-ISA had 94% positive percent agreement and >98% negative percent agreement in detection of serum IgG reactivity to viral antigen. The BLI-ISA antibody test is readily adaptable to new antigens, and was quickly deployed to assess reactivity to a panel of eight human astrovirus serotypes with good agreement to studies using more laborious methods.

Antibody therapeutics for cancer: a comparison of bioinformatic methods for the identification of tumor surface markers using RNAseq reveals statistical methods like DESeq outperform some dimensionality reduction techniques like PCA. Then, the epitope of an immune checkpoint inhibitor antibody is characterized by X-ray crystallography in the context of KIR receptor variants, highlighting the importance of population-wide polymorphism analysis.

HAstV RdRp: expression methods for this under-studied polymerase were developed to aid future structural work and activity assays.

ACKNOWLEDGMENT

I would like to give thanks to my friends and family who supported me during my graduate career.

The work presented in this dissertation was funded by NIH training grant 5T32HG008345-05, a Santa Cruz Cancer Grant, the UCSC Baskin Engineering dissertation year fellowship, and NIH R01 AI144090. Additional funding sources for individual projects and instrumentation can be found in the acknowledgment statement of each chapter section (where applicable).

The text of this dissertation includes reprints of the following previously published material: "Rapid and Sensitive Detection of SARS-CoV-2 Antibodies by Biolayer Interferometry" (Dzimianski, J. V.; Lorig-Roach, N.; O'Rourke, S. M.; Alexander, D. L.; Kimmey, J. M.; DuBois, R. M. Scientific Reports 2020, 10 (1), 21738) and "Human Astrovirus 1–8 Seroprevalence Evaluation in a United States Adult Population" (Meyer, L.; Delgado-Cunningham, K.; Lorig-Roach, N.; Ford, J.; DuBois, R. M. Viruses 2021, 13 (6), 979). These articles are licensed via the Creative Commons BY 4.0 licence which allows their reproduction here with attribution. Preceding the reprints of these articles in their respective chapter sections are more detailed acknowledgments with author contributions listed.

Chapter One

Biolayer Interferometry as a New method of Antibody testing

This chapter discusses the development of an assay that uses biolayer interferometry to perform antibody tests. We show that our method can be used with tiny blood samples to detect exposure to SARS-CoV-2 and the eight canonical strains of human astrovirus (HAstV 1-8). Two peer reviewed publications will be reproduced in this chapter along with an application we submitted to the FDA to apply for Emergency Use Authorization to use our BLI assay for SARS-CoV-2 antibody testing.

1.0.1 Introduction to Biolayer Interferometry

Measuring intermolecular interactions with BLI

The biolayer interferometry (BLI) technique was developed as an alternative to techniques like ELISA (enzyme-linked immunosorbent assay) and SPR (surface plasmon resonance) for detecting intermolecular interactions. BLI can be used to

infer such interactions by exploiting the principle of thin film interference. Thin film interference occurs when light reflects from a surface coated with a thin, partially transmissive material (nanometer to several micrometer thick), resulting in two offset reflection events in which the light reflected from each 'layer' may either constructively or destructively interfere with each other. The extent of constructive or destructive interference is dependent upon the ratio of the wavelength of light to the thickness of the thin film.

In the BLI instrument, white light (i.e. light in the 100-700 nm range) is shone down an optical fiber to a sensor tip with a thin film at its surface (Fig. 1.1. The thin film at the tip is composed of a <50 nm layer of high refractive index material like Ta_2O_5 called the 'reference layer' followed by 500-1000 nm of SiO_2 (glass) that can be modified at its surface to bind analyte ('analyte layer') [36]. Light that reflects off the reference layer and light reflecting from the analyte layer then travels back through the sensor and optical fiber to a broad spectrum UV-Vis detector like a CCD detector which can measure the amplitudes of light waves in the UV-Vis spectrum. Because the constructive or destructive interference that occurs between light reflected from each layer of the thin film is dependent on that film's thickness, binding events on the surface of the sensor tip modulates which wavelengths of light are most enhanced by that interference.

The Sensor Tip Surface

The surface of the BLI sensor tip is a critical part of experiment that allows tethering of materials that enable association of analytes of interest onto the tip surface. The primary approaches to functionalization of the glass surface of the sensor include adsorption of a reactive or useful polymer (like cross linked lipid-

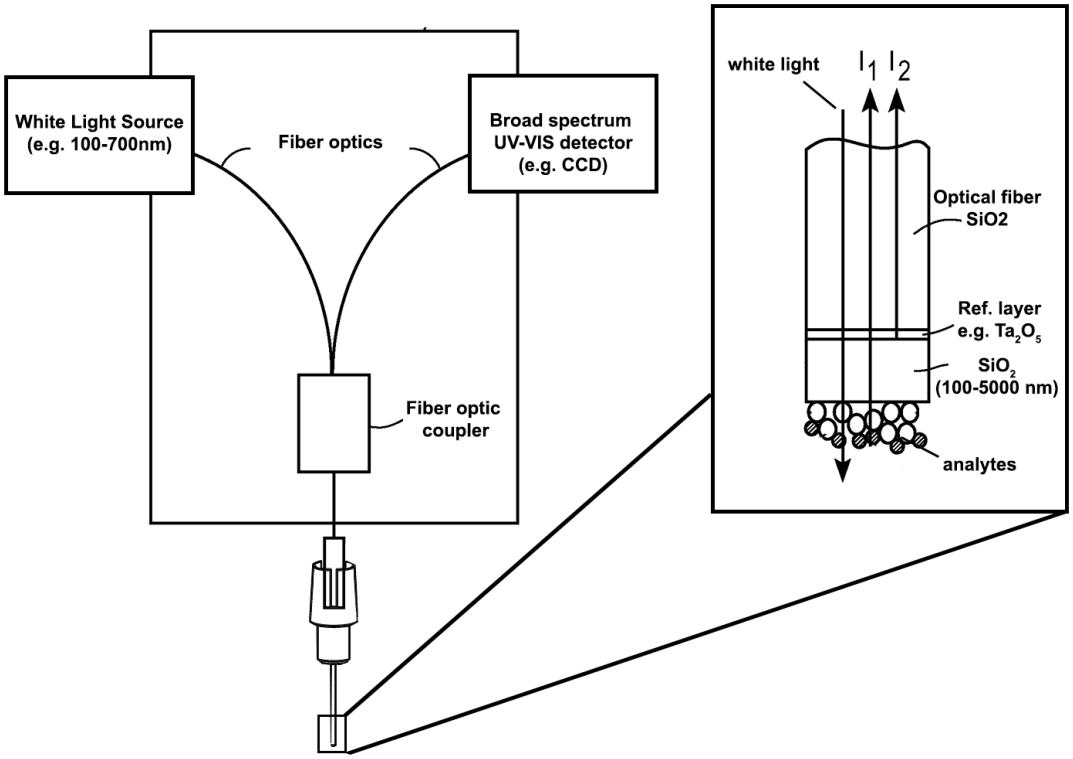


Figure 1.1 Overview of the BLI instrument. White light is transmitted via optical fibers to a detachable sensor and the light reflected back from the sensor tip is measured by a CCD detector. The tip is terminated by a layer of a high refractive index material (relative to glass) like tantalum pentoxide followed by a functionalized glass layer that can form bonds with organic molecules through a variety of chemistries. Light measured by the CCD is the sum of the light reflected from the analyte layer (I_1) and the reference layer (I_2). Created in part from figures in US patent US7319525B2 [36].

biotin matrix) or etching of the SiO2 surface to reveal hydroxyl groups that can be converted to sulfhydryl or amine groups compatible with e.g. maleimide based click chemistry [36].

Commercially available biosensors are based on some variation of these chemistries, but often are pre-conjugated with materials capable of binding common analytes or protein tags. It is important to be aware of the methods of sensor tip functionalization to control for non-specific binding or other unintended binding interactions during experiments. For example we have found that commercial his-tag binding sensor tips appear to have been manufactured using some form of the biotin-streptavidin interaction – samples containing biotin-tagged material also interact with the His-binding sensor!

Extracting Kinetic Parameters from BLI Measurements

The basis for deriving kinetics parameters from BLI data is the binding equilibrium equation (Eq. 1.1) where A is the analyte in solution, B is the ligand immobilized on the sensor tip, AB is the bound complex, and k_a and k_d are the association and dissociation rate constants.

$$\mathrm{A} + \mathrm{B} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{AB} \tag{1.1}$$

Many of the assumptions made for SPR data analysis by O'Shannessy et al [28] hold for the BLI experiment: the concentration of soluble analyte [A] can be assumed to be constant due to the large excess relative to B on the sensor tip, and the BLI signal, R, (i.e. thickening of the thin film at the sensor tip surface and corresponding phase shift of reflected light) should be proportional to the rate of AB formation. Additionally the max observable signal, R_{max}, should be propor-

tional to the initial amount of ligand, B_0, loaded onto the tip surface.

Just as derived by O'Shannessy et al for SPR, the rate of bound complex formation at the tip surface is given by Eq. 1.2, and we can arrive at Eq 1.3 by substituting $[B]$ for $[B]_0 - [AB]$ (concentration of unbound ligand is equal to its initial concentration minus the concentration of complex).

$$\frac{d[AB]}{dt} = k_a[A][B] - k_d[AB] \tag{1.2}$$

$$\frac{d[AB]}{dt} = k_a[A]([B]_0 - [AB]) - k_d[AB] \tag{1.3}$$

Because the concentration of analyte, [A], changes negligibly over the course of the BLI experiment, it can be substituted for a constant, C. The proportionality of [AB] to the phase-shift response, R, and $[B]_0$ to R_{max} allows conversion of Eq. 1.3 to

$$\frac{dR}{dt} = k_aC(R_{max} - R) - k_dR \tag{1.4}$$

$$\frac{dR}{dt} = k_aCR_{max} - (k_aC + k_d)R \tag{1.5}$$

Integration of equation 1.5 gives an expression for the observed BLI signal at time t:

$$R_t = \frac{Ck_aR_{max}[1 - e^{-((Ck_a+k_d)t)}]}{Ck_a + k_d} \tag{1.6}$$

Eq. 1.6 assumes the initial response measured by the BLI instrument is 0, and would therefore require baseline corrected data to be applied. An additional correction parameter could also be added to 1.6 to account for non-zero signal in data being fit. Using a regression analysis approach starting with known parameters and estimates for R_{max}, k_a, and k_d, Eq. 1.6 can be used to find a set of parameters that best fits the experimental data for the association phase of the BLI experiment. Especially in cases where k_d is small, it is useful to observe complex dissociation rate directly, which is done by moving the sensor tip from a solution

containing analyte to one without any. Here, where [A] (i.e. C) is zero, Eq. 1.4 becomes

$$\frac{dR}{dt} = -k_d R \tag{1.7}$$

which, when integrated is

$$R_t = De^{-k_d t} \tag{1.8}$$

where D is effectively the initial signal at the start of the dissociation phase. As before, linear regression can be performed to identify a k_d that best fits the observed data.

We can use equations 1.6 and 1.8 to simulate expected association and dissociation curves for different combinations of k_a and k_d (Fig. 1.2). Note that vastly different curve shapes (i.e. k_a and k_d values) can result in an equivalent K_D, highlighting one of the advantages of BLI over equilibrium based K_D determination methods. These curves can be plotted simply using Python with the code snippet below:

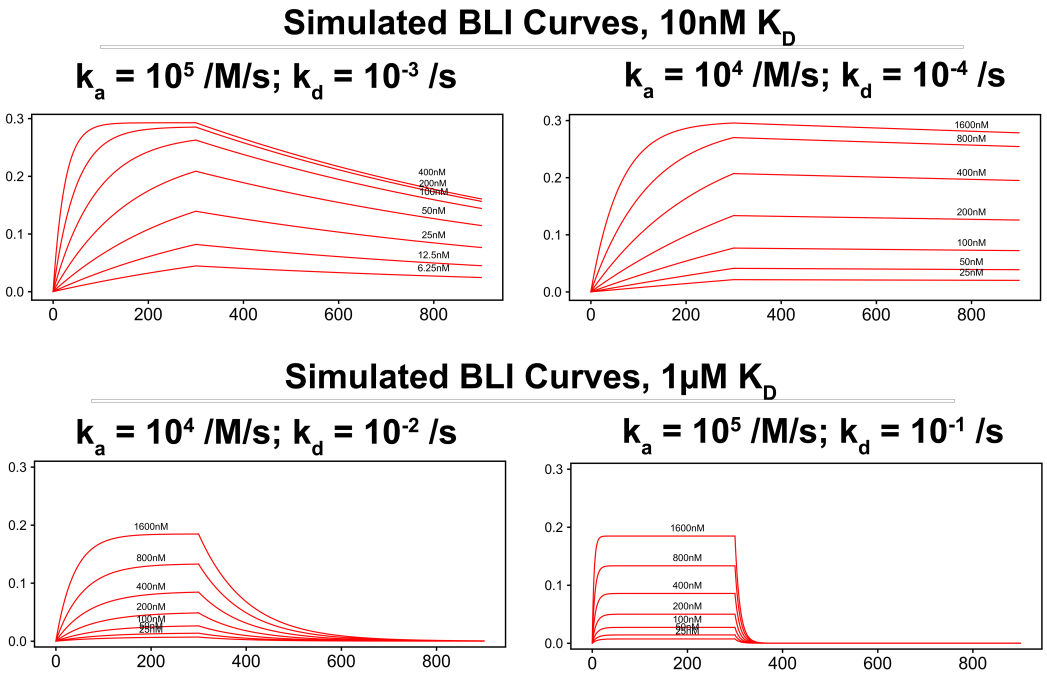


Figure 1.2 Simulated BLI curves. Using equations 1.6 and 1.8, simulated 1:1 binding interactions are plotted for K_D values of 10nM and 1μM with two different pairs of k_a and k_d values each. In each plot Rmax (the theoretical max signal for a given ligand loading onto a sensor tip) was set to 0.3 and concentrations of analyte were varied between 6 and 1600 nM. For clarity, the highest or lowest concentration curves were omitted in some plots.

```
import numpy as np
import matplotlib.pyplot as plt
import math

## Eq 1.6
def association_function(t,Rmax,ka,kd,C):
    return (C*ka*Rmax*(1-math.e**(-((C*ka+kd)*t))))/(C*ka+kd)
## Eq 1.8
def dissociation_function(t,kd,D):
    return D*math.e**(-kd*t)

## use an 'average' BLI association signal of 0.3 nm for Rmax
Rmax = 0.3
## choose ka, kd values
ka, kd = 10.0**5, 10.0**-1
## generate x values for 300 sec association, 600 sec dissociation
x_assoc = np.linspace(0, 300, 600)
x_diss = np.linspace(300, 900, 1200)

concentrations =  [1600, 800, 400, 200, 100, 50, 25, 12.5, 6.25]
## loop over concentration list, plotting a curve for each
for C in concentrations:
    C0 = C*10.0**-9 ## convert to nM
    ##calculate y values for association and plot them
    y_assoc = association_function(x_assoc,Rmax,ka,kd,C0)
    plt.plot(x_assoc, y_assoc, linewidth = 0.6, color = 'red',)
    ## set 'dissociation amplitude' to max signal from assoc.
    D = np.max(y_assoc)
    ##calculate y values for dissociation and plot them
    y_diss = dissociation_function(x_diss-300,kd,D)
    plt.plot(x_diss, y_diss, linewidth = 0.6, color = 'red',)
plt.show()
```

Listing 1.1 Simulating BLI Curves with Python

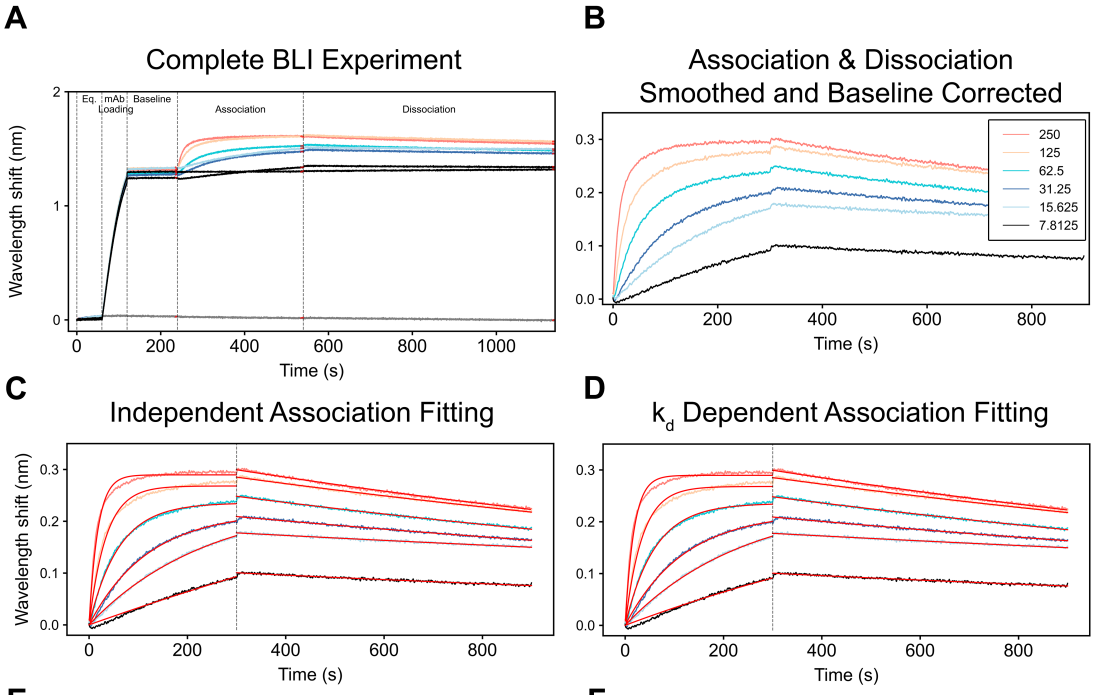
With the model equations for association and dissociation in a BLI experiment established, measured data can be fit by non-linear regression methods (e.g. the least-squares regression algorithm) to identify unknown kinetics parameters. Figure 1.3 shows data from a kinetics experiment in which an antibody is loaded onto a sensor tip and association and dissociation to an analyte is detected. In Fig. 1.3A, the entirety of the BLI experiment is shown. Briefly, a baseline value is measured followed by loading of the ligand antibody onto the tip surface, which can be tracked by the wavelength shift corresponding to the thickening thin film at the tip surface. This loading step defines $[B]_0$ and thus Rmax. Next, a second baseline value is measured to confirm or establish stability of the ligand on the tip. Then the ligand-loaded tip is moved to a solution containing a known

concentration of analyte so that association can be observed. Finally the tip is placed in a solution without analyte to measure dissociation.

To simplify model fitting, the data from the association and dissociation phase for each concentration of analyte are trimmed from the other experiment steps, then aligned such that the start of each association step is zero (Fig. 1.3B). Often baseline correction is also performed where data from no-ligand and/or no-antigen controls are subtracted from experimental data to correct for drift or non-specific binding interactions. After preparing the data, fitting can be performed using equations 1.6 and 1.8 for the association and dissociation steps, respectively (Fig. 1.3C, E). Initial parameters for Rmax, k_a, and k_d can be inferred from the data and literature (a good guess for Rmax is the max signal observed for a given dilution series, assuming saturation is reached; k_a is generally between 10^2 and 10^7 $M^{-1}s^{-1}$; k_d is between 0 and 1 s^{-1}), though the least-squares algorithm typically performs well even without accurate a priori parameter guesses.

The association phase is dependent upon both k_a and k_d, which allows for seemingly accurate fitting where one or both parameters is actually incorrect. The dissociation phase is only dependent on k_d and model fitting across an analyte dilution series will generally give a more consistent k_d measurement (compare k_d1 and k_d2 values in 1.3E). Thus more accurate k_a (and therefore K_D) parameters can be identified by restricting the search space for association fitting based on the k_d discovered during dissociation fitting (Fig. 1.3D, F). Despite nearly imperceptible difference in the appearance of the fit line between Fig. 1.3C and 1.3D, there is a dramatic improvement in k_a and K_D consistency (Fig. 1.3F).

The concept of restraining parameters across phases and experiments to find



E

Independent Association and Dissociation Fitting – Kinetic Parameters

[A] (nM)	Rmax	K_a (M^{-1} s^{-1})	K_d1 (s^{-1}) (Assoc.)	K_d2 (s^{-1}) (Dissoc.)	K_D1 (nM) k_d1/k_a	K_D2 (nM) k_d2/k_a
250	0.32	2.01E+05	5.32E-03	4.90E-04	26.5	2.5
125	0.32	1.92E+05	4.91E-03	4.50E-04	25.6	2.3
62.5	0.39	1.61E+05	6.55E-03	4.90E-04	40.6	3.1
31.25	0.42	1.39E+05	4.02E-03	4.10E-04	28.9	3.0
15.625	0.48	1.34E+05	2.07E-03	2.80E-04	15.5	2.1
7.8125	0.59	7.31E+04	0.00E-23	4.70E-04	0	6.5

F

Dissociation Dependent Fitting – Kinetic Parameters

[A] (nM)	Rmax	K_a (M^{-1} s^{-1})	K_d (s^{-1})	K_D1 (nM) K_a/k_d1
250	0.29	2.20E+05	4.90E-04	2.2
125	0.27	2.28E+05	4.50E-04	2.0
62.5	0.24	2.58E+05	4.90E-04	1.9
31.25	0.23	2.54E+05	4.10E-04	1.6
15.625	0.26	2.48E+05	2.80E-04	1.1
7.8125	0.59	7.71E+04	4.70E-04	6.1

Figure 1.3 Example of BLI data fitting and Kinetic Parameter Extraction. (A) Full BLI experiment in which mAb is loaded onto AHC sensor tips and dipped into a dilution series of analyte. Includes no-ligand and no-analyte controls for baseline correction. (B) Association and dissociation steps from A, with baseline subtraction, data smoothing, and alignment to Association step applied. (C) Application of least-squares regression to the association and dissociation steps, using equations 1.6 and 1.8 respectively. (D) similar to panel C, except the k_d value for the association function was set based on the parameter identified from the corresponding dissociation step. (E) and (F) summarize the kinetics parameters identified by each fitting method, note the significant improvement in k_a and k_D consistency when the dissociation k_d is used to restrict the search space of the association fitting.

the 'true' k_a and k_d can be taken further by identifying a singular value for each that minimizes fit error overall across a series of experiments rather than for an individual experiment or phase. This is referred to as 'global' fitting within the Octet BLI's Data Analysis software. It is unclear to the author whether this global fitting would actually improve k_a and k_d accuracy over averaging dissociation-restricted individual fits. In both strategies, it is important to identify and potentially exclude experiments where model fit is poor or divergent (like the 7.8125 nM [A] experiment in Fig. 1.3 where, putatively, the linear nature of the curve results in an incorrect Rmax assumption during fitting).

Other approaches to kinetic parameter evaluation - steady state

When association or dissociation rates are particularly fast, it can be difficult to fit binding model equations to the observed data accurately. Although the individual k_a and k_d values cannot be inferred without such fitting, it is still possible to find the equilibrium constant K_D. At equilibrium, or steady state conditions, the rate of complex formation and dissociation are equal and the change in measured response, R, is zero (no net change in thickness at the sensor tip). This can be described by setting equation 1.5 to zero:

$$\frac{dR}{dt} = k_a C R_{max} - (k_a C + k_d)R = 0 \tag{1.9}$$

Letting R at equilibrium be defined as R_{eq}, we can rearrange to

$$(k_a C + k_d)R_{eq} = k_a C R_{max} \tag{1.10}$$

Divide each side by k_a, where $k_d/k_a = K_D$

$$(C + K_D)R_{eq} = C R_{max} \tag{1.11}$$

$$R_{eq} = \frac{C R_{max}}{(C + K_D)} \tag{1.12}$$

By running a BLI experiment where the association phase is long enough to reach equilibrium for each concentration in a dilution series (similar to Fig. 1.2, bottom right), the above equation can be used to fit the experimental data, solving for K_D. Calculating EC/IC50 values follows a similar experimental approach to a steady state experiment: a concentration series of analyte is allowed to associate with ligand until equilibrium, where the highest concentrations of analyte in the series reach R_{max} (i.e. increasing [A] further does not increase measured signal).

Beyond Kinetics - BLI for quantification and trace analyte detection

Small assay volumes and lack of fluidics makes biolayer interferometry an excellent platform for molecule detection in a variety of media. For precise quantification, a dilution series of known concentration can be used to establish the relationship between analyte concentration and signal response from the biosensor. We've used this to titer secreted protein during eukaryotic cell expression of numerous constructs. BLI measurements can be performed directly in expression media, cell lysates, even serum.

When analyte concentrations are particularly low or the size of analytes are small, the observed BLI signal may be exceedingly small due to minimal change in the effective thickness at the biosensor tip. In these instances, using a secondary detection reagent, analogous to those from the classic blot techniques and ELISA, can allow unambiguous detection of analyte with high sensitivity. Secondary antibodies, with or without conjugation to another size-increasing construct, improve signal by adding analyte-specific material at the sensor tip surface and causing a more significant phase shift to occur. In the sections to come, we use this idea to detect trace amounts of virus-binding antibodies in human sera via an IgG-specific secondary detection reagent.

1.1 Rapid and sensitive detection of SARS-CoV-2 antibodies by biolayer interferometry

1.1.1 Preface

We started the project described here shortly after the onset of the COVID-19 pandemic and the acquisition of a biolayer interferometry (BLI) instrument obtained via a shared instrumentation grant led by Seth Rubin. BLI is incredibly useful for measuring macromolecular interactions and kinetics, and is marketed as an alternative to techniques like ELISA and SPR. As soon as the first few BLI experiments were completed, excitement about the possibilities of the technique grew. From simple tasks like protein expression detection and titering to kinetics and epitope footprinting, the ease and speed of the technique compared to the alternatives we had access to opened up new doors. At the beginning of the pandemic, our early BLI experiments investigating antibody:virus-antigen interactions quickly cued ideas to try detecting SARS-CoV-2 or antibodies that bound the virus, eventually leading to the BLI-ISA protocol we applied to serology testing in the sections below. At the time, access to campus and the lab was heavily restricted so only John and Sara were able to do wet lab experiments. I was able to contribute by developing a script that automated quality control and scoring of the antibody detection assay as well as assisting with data visualization and interpretation from many of the experiments. Along with the contributions of numerous collaborators, we were able to develop and release our protocol quite quickly as first a pre-print, then a research article in Scientific Reports.

This Section contains a reproduction of a research paper published in Scientific

Reports available online at https://doi.org/10.1038/s41598-020-78895-x.

Citation

Rapid and Sensitive Detection of SARS-CoV-2 Antibodies by Biolayer Interferometry.

Dzimianski, J. V.; Lorig-Roach, N.; O'Rourke, S. M.; Alexander, D. L.; Kimmey, J. M.; DuBois, R. M.

Sci Rep 2020, 10 (1), 21738.

Author contributions

J.V.D., N.L.R. and R.M.D. conceived the study. S.M.O. expressed RBD proteins in CHO cells. J.V.D. and R.M.D. purified RBD-His and RBD-biotin proteins and acquired data. N.L.R. wrote the BLI-plotter program. J.V.D., N.L.R. and R.M.D. analyzed the data. D.L.A. and J.M.K. contributed unique reagents and consulted on experi- mental design. J.V.D. and N.L.R. prepared figures. J.V.D. and R.M.D. wrote the manuscript. All authors edited the manuscript.

1.1.2 Abstract

Serological testing to evaluate antigen-specific antibodies in plasma is generally performed by rapid lateral flow test strips that lack quantitative results or by high complexity immunoassays that are time- and labor-intensive but provide quantitative results. Here, we describe a novel application of biolayer interferometry for the rapid detection of antigen-specific antibody levels in plasma samples, and demonstrate its utility for quantification of SARS-CoV-2 antibodies. Our biolayer

interferometry immunosorbent assay (BLI-ISA) utilizes single-use biosensors in an automated "dip-and-read" format, providing real-time optical measurements of antigen loading, plasma antibody binding, and antibody isotype detection. Complete quantitative results are obtained in less than 20 minutes. BLI-ISA meets or exceeds the performance of high complexity methods such as Enzyme-Linked Immunosorbent Assay (ELISA) and Chemiluminescent Immunoassay. Importantly, our method can be immediately implemented on existing BLI platforms for urgent COVID-19 studies, such as serosurveillance and the evaluation of vaccine candidates. In a broader sense, BLI-ISA can be developed as a novel diagnostic platform to evaluate antibodies and other biomolecules in clinical specimens.

1.1.3 Introduction

In December 2019, a novel coronavirus emerged in Wuhan, China, causing severe respiratory disease with initial reported fatality rates of 2–3%[1]. In the ensuing months the virus became established internationally through travel and community transmission, leading to the declaration of a pandemic by the WHO on March 11, 2020[2]. Officially named severe acute respiratory coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses due to its phylogenetic relatedness to SARS and SARS-like coronaviruses[3], the virus causes coronavirus disease 2019 (COVID-19). As of November 10, 2020, over 51 million cases and over 1.2 million deaths have been reported due to COVID-19, and the disease continues to be a source of economic and societal strain. Efficient and accurate testing is critical to understand the full breadth of impact and to developing coun- termeasures to limit future infections.

Detection of SARS-CoV-2 infection relies predominantly on two approaches: nucleic acid testing, which detects viral RNA, and serological testing, which detects antibodies elicited against SARS-CoV2 antigens. Nucleic acid testing methods were quickly developed after the release of the virus genome[4–6] and serve as the primary, definitive diagnostic tool for active cases of COVID-19. However, due to limitations in nucleic acid testing availability and the occurrence of mild or asymptomatic infections, many cases of COVID-19 are not diagnosed. Thus, serological testing, which detects antibodies elicited by SARS-CoV-2 antigens, have become key to assessing the true extent of SARS-CoV-2 spread within the population[7]. Serological studies have shown that antibodies develop over several weeks following infection with SARS-CoV-2, and that antibody levels can vary significantly between individuals[8–11]. Accurate serological testing is crucial to develop countermeasures against SARS-CoV-2 infection, including the identification and evaluation of donors for convalescent plasma therapy and the development of a SARS-CoV-2 vaccine.

Serological testing methods for SARS-CoV-2 predominantly use the virus nucleocapsid protein, the spike glycoprotein, or fragments thereof such as the spike receptor binding domain (RBD), to probe for antibodies. Methods that utilize the spike, and the RBD in particular, have been shown to correlate with SARS-CoV-2 neutralization assays [7,8,10,12–16]. Serological tests are also distinguished on whether they detect total antibodies, IgG, IgM, or both IgG and IgM. Current tests that have been developed include the Lateral Flow Immunoassay (LFIA), Enzyme-Linked Immunosorbent Assay (ELISA), Immunofluorescent Assay (IFA), and Chemiluminescent Immunoassay (CLIA). LFIA tests present the most rapid turnaround and can be performed with minimal training, with test strip bands

visualized in 15–20 min, making it useful as a point-of-care test[17]. The output, however, is generally a positive/negative binary outcome, and the tests have had a mixed performance in terms of sensitivity [18,19]. ELISA, IFA, and CLIA tests are high complexity laboratory tests that are generally robust in terms of the sensitivity and specificity[19–21]. In addition, these methods provide a semi-quantitative measure of antibody responses that can distinguish between strong and weak responses. However, ELISA, IFA, and CLIA are time-intensive processes requiring 1–5 h, with significant incubation times and washing steps that are performed manually or require automated fluidic platforms. Due to these drawbacks, the development of alternate serological testing methods that are simple, rapid, and quantitative would be advantageous for many applications.

Here, we describe a novel method for measuring antigen-specific antibodies in blood plasma utilizing biolayer interferometry (BLI). BLI is a fiber optics-based biophysical technique designed to measure the affinity between biological molecules. White light is shone down a fiber optic biosensor and the interference between light reflected off two layers—a reference layer and a biological layer—is measured[22]. Binding of molecules to the biosensor surface results in a real-time signal due to the shift in the wavelength of the reflected light. While historically used to precisely measure the kinetics of binding between purified biological molecules, BLI has also been adapted to quantify a target biological molecule in more complex fluids, such as proteins in cell growth media[23,24] and biomolecules in clinical specimens[25–30]. We developed a novel application for this technology, termed biolayer interferometry immunosorbent assay (BLI-ISA), for the rapid and semi-quantitative measurement of SARS-CoV-2 antibodies in plasma. BLI-ISA advantages include a simple "dip-and-read" format that is free of fluidics, provides

real-time measurements of both total antibody levels and specific antibody isotypes in the same assay, and can detect weakly seropositive samples. Importantly, BLI-ISA can be completed in less than 20 min and provides semi-quantitative information on antibody levels. This method can be immediately implemented for urgent SARS-CoV-2 pandemic studies, including evaluation of antibody responses to natural infection and to candidate vaccines, as well as serosurveillance studies. In a broader sense, BLI-ISA can be adapted and multiplexed, allowing for measurement of antibodies against multiple antigens, detection of multiple antibody isotypes, and quantitation of other clinically relevant biological molecules.

1.1.4 Results

Validation of test samples by ELISA.

For the development of serological testing, we obtained 37 plasma or serum samples to serve as our test set for experimental design. This included 10 commercially available convalescent plasma samples from donors who had recovered from COVID-19, each of which had previously tested positive by a CoV2T CLIA test, with CoV2T scores ranging from 8 to 440 (Table 1)1.1. These ten samples served as our presumed seropositive (SP) group. 27 plasma and serum samples collected prior to the COVID-19 pandemic formed the presumed seronegative (SN) group.

To confirm the status of the samples and establish the expected outcomes for future study, we adapted a published ELISA protocol to measure IgG antibody reactivity towards recombinant SARS-CoV-2 spike RBD (Fig. 1, Supplementary Fig. S1)12. All samples were evaluated at 1:50 dilution in biological duplicates with RBD-coated plates as well as control uncoated plates. The majority of pre-pandemic (presumed seronegative) samples did not react to RBD-coated plates,

Table 1. Comparison of antibody reactivity assays with SARS-CoV-2 RBD antigen.

Sample	COV2T (Total Ab)[1]	ELISA (AUC)	BLI-ISA (Total Ab)[2]	BLI-ISA (IgG)[3]	BLI-ISA (IgG)
PBS	N/A	9.1	0.05		-
SP1	222	6,531.00	3.01		+
SP2	153	3,509.80	1.17		+
SP3	8.02	262	0.14		+
SP4	42.8	2,875.10	1.13		+
SP5	121	6,276.40	2.47		+
SP6	103	1,842.80	0.71		+
SP7	72.2	10,621.80	3.68		+
SP8	53.9	835.9	0.38		+
SP9	282	4,961.70	2.47		+
SP10	440	4,346.80	2.5		+

[1]COV2T (Total Ab) test is produced by Ortho Diagnostics
[2]BLI-ISA (Total Ab) data was collected with 1:8 diluted plasma
[3]BLI-ISA (IgG) data was collected with 1:16 diluted plasma

Table 1.1 Comparison of antibody reactivity assays with SARS-CoV-2 RBD antigen. CoV2T (Total Ab) was determined previously with an Anti-SARS-CoV-2 Total test (Ortho Clinical Diagnostics). 2BLI-ISA (Total Ab) data was collected with 1:8 diluted plasma. 3BLI-ISA (IgG) data was collected with 1:16 diluted plasma.

resulting in a signal that was indistinguishable from the background (mean SN OD490 = 0.227). As expected, the presumed positive samples showed a robust signal in ELISA that was substantially higher than the presumed negatives and background in the absence of antigen (Fig. 1a). Intriguingly, however, there were two notable exceptions to these trends. The presumed positive sample SP3 failed to exceed its own background and was indistinguishable from some of the presumed negatives. In contrast, SN22 showed a signal above background that, while weaker than most of the presumed positives, appeared to be more robust than SP3.

We subsequently performed dilution series ELISA and area-under-the-curve (AUC) calculations for all SP samples and a subset of SN samples (SN7, SN12, SN15, and SN22) chosen to represent the diversity of signals observed in the

single-dilution ELISA (Fig. 1.4b,c, Table 1). The dilution series ELISA clarified the differences between highly reactive samples, revealing SP7 as having the most robust anti-CoV-2 spike RBD IgG levels, while SP8 showed moderate levels. In contrast, SP3 overlapped with SN22 and was only slightly more reactive than SN15.

Design and optimization of antibody detection by BLI-ISA.

Having established the range of expected reactivity of the samples for SARS-CoV-2 spike RBD by ELISA, we then developed the BLI-ISA method. We developed this method with the goal of a simple, rapid ($<$ 20 min), and semi-quantitative method to measure antigen-specific antibodies in plasma (Fig. 1.5). The simple "dip-and-read" format allows for antigen and antibody samples to be plated into a 96- or 384-well plate (Supplementary Fig. S2). This plate is then loaded along with single-use biosensors into the BLI instrument, which dips the biosensors into designated plate wells to perform each step. Most notably, the BLI-ISA technology enables real-time measurements throughout the entire experiment (Supplementary Fig. S3). The signal from the antigen loading step serves as a quality control measure and ensures even antigen loading onto each biosensor surface. In addition, two antibody-binding steps can be evaluated by our method; (1) a Total Antibody Binding signal is measured when the antigen-coated biosensors are dipped into plasma samples and (2) a Detection Antibody Binding signal is measured when the biosensors are subsequently dipped into an anti-human IgG secondary antibody reagent.

Several features required optimization including antigen loading, assay buffers,

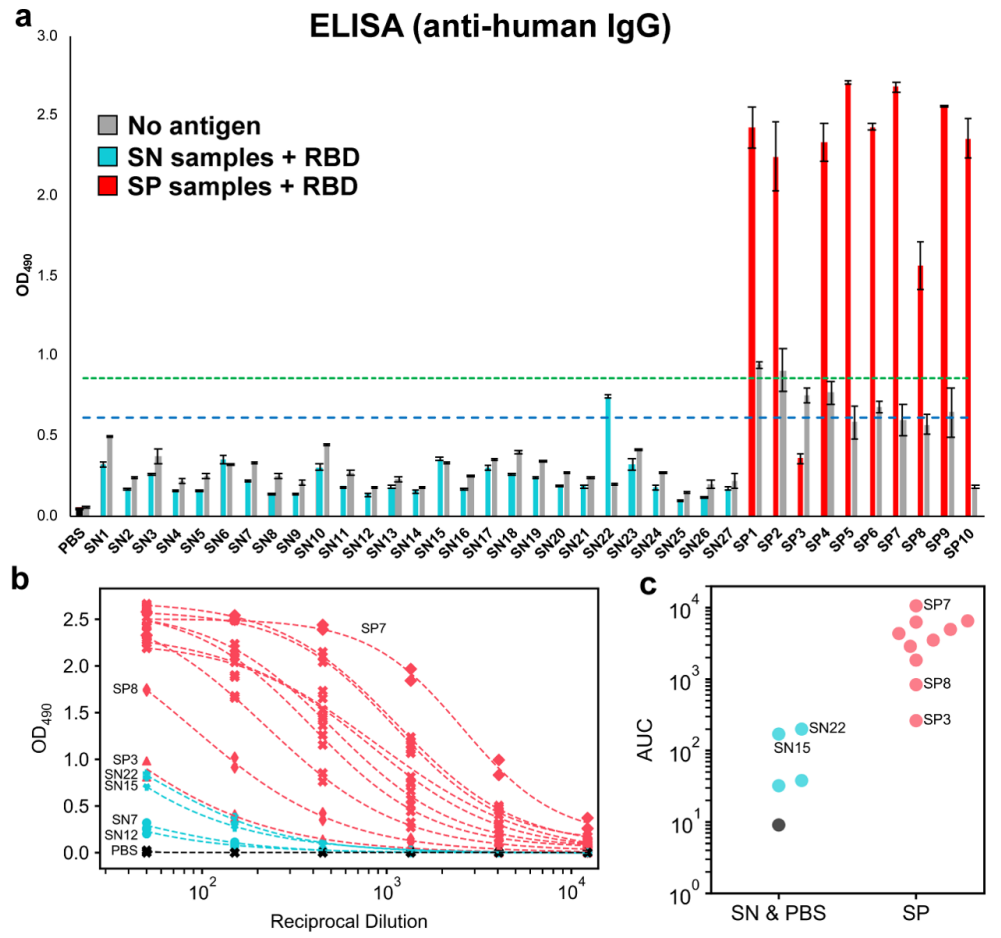


Figure 1.4 ELISA evaluation of SARS-CoV-2 spike RBD reactivity of pre-pandemic and convalescent plasma. (a) Single-dilution ELISA to evaluate the presence of RBD-reactive human IgG in pre-pandemic seronegative (SN, cyan) and convalescent seropositive (SP, red) samples compared to no-antigen controls (grey). The assays were performed with plasma at a 1:50 dilution. Samples were evaluated in biological duplicates and error bars represent one standard deviation from the mean. Blue and green dashed lines represent the mean of seronegative samples plus 3 and 5 standard deviations, respectively. (b) Dilution series ELISA was performed to quantitate RBD-reactive human IgG in plasma. Samples were evaluated in biological duplicates. Dashed curves represent fit lines from a four-parameter logistic regression applied over each series. (c) Data from (b) plotted as area-under-the-curve (AUC).

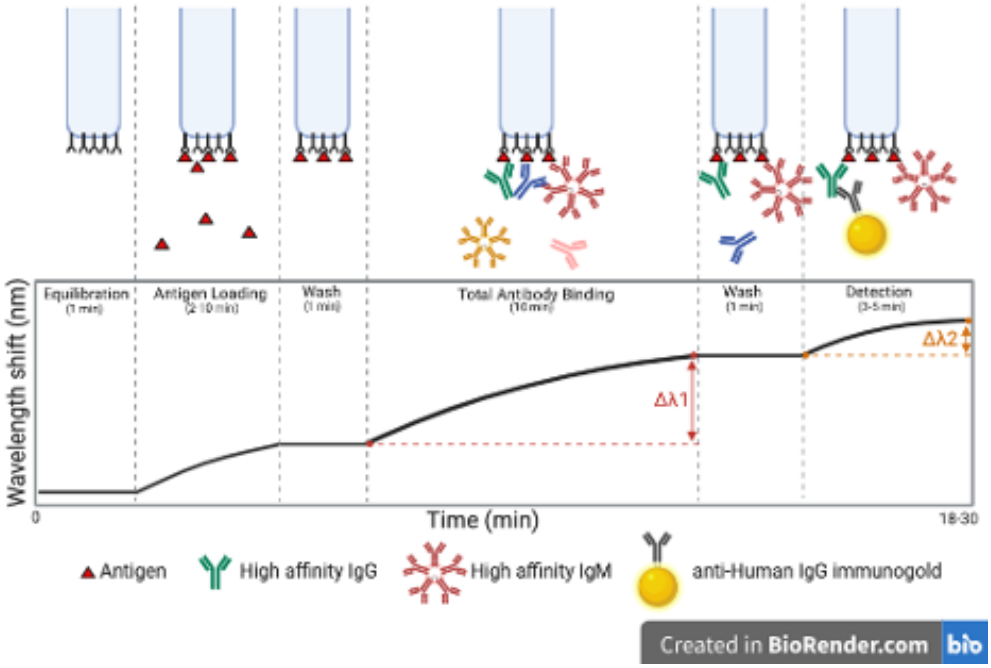


Figure 1.5 Overview of the BLI-ISA experiment. To begin, a tray of fiber optic biosensors and a 96- or 384-well plate of samples are placed into the Octet BLI instrument (Supplementary Fig. S2), and the assay program is run. Throughout the experiment, real-time measurements are recorded as the change in the wavelength of reflected light returning from the biosensor surface. First, biosensors are equilibrated by dipping into wells containing BLI assay buffer. In the antigen loading step, biosensors are dipped into wells containing tagged antigen (e.g. streptavidin SA biosensors dipped into biotinylated antigen). After a wash, antigen-loaded biosensors are placed into diluted plasma, and a Total Antibody Binding signal is measured. After another wash, the antigen– antibody-coated biosensors are dipped into wells containing isotype-specific binding reagents (e.g. colloidal gold-conjugated anti-human IgG), and a Detection signal is measured. Created in BioRender.com.

plasma dilution factors, secondary antibody detection reagents, and times at each assay step. First, we evaluated the loading stability of three recombinant SARS-CoV-2 spike antigens onto BLI biosensor tips: the His-tagged RBD (RBD-His) used for ELISA12, an Avi-tagged RBD with a single biotin modification on the Avi-Tag (RBD-biotin), and the His- tagged prefusion-stabilized spike trimer (pre-

fusion Spike-His)[31,32] (Supplementary Fig. S1). All three constructs exhibited sufficient loading onto their respective anti-penta-His (HIS1K) or (anti-biotin) streptavidin (SA) bio- sensors. However, loading of RBD-His demonstrated considerable downward baseline drift compared to RBD- biotin with SA biosensors (Supplementary Fig. S4). Interestingly, this downward drift was not observed by the prefusion Spike-His, likely due to the stronger anchoring by the three His-tags in this trimeric form. To enable comparison of our BLI-ISA results with the CoV2T CLIA and ELISA, which use the RBD antigen, we further optimized our method using RBD-biotin loaded onto SA biosensors. During buffer optimization, we found that addition of 20–25% ChonBlock, a blocking agent, to the plasma samples reduced background signal from SN samples without affecting SP sample signals (Supplementary Fig. S5). Finally, we developed a novel method to detect human IgG antibody bound to the antigen-coated biosensor. Whereas anti-human antibodies alone or conjugated to enzymes did not yield a signal (data not shown), we discovered that colloidal gold-conjugated anti-human antibody reagents were large enough to give a significant signal in the Detection step (Fig. 1.5, Supplementary Fig. S3).

To confirm the specificity of BLI-ISA, we tested antigen binding with commercially available rabbit antibodies that were raised against the spike proteins of either HCoV-HKU1 or SARS-CoV-2 (Supplementary Fig. S6). Consistent with previous studies showing no cross-reactivity of antibodies against HCoV-HKU1 toward the SARS-CoV-2 spike RBD[8,12], polyclonal antibodies against HCoV-HKU1 spike showed a negligible Total AntibodyBinding signal, whereas both polyclonal and monoclonal rabbit SARS-CoV-2-targeting antibodies resulted in a significantly higher Total Antibody Binding signal. As an additional check

on the specificity, we evaluated the signals by the anti-human IgG secondary antibody in the Detection step. As a positive control, we used the human monoclonal antibody CR3022, which targets the SARS-CoV-1 RBD and has known cross-reactivity with the SARS-CoV-2 RBD33–35. Notably, despite being present at a lower concentration compared to the rabbit antibody samples, CR3022 exhibited a robust signal at the Detection step with the anti-human IgG secondary antibody detection reagent. In contrast, only background signal was observed with the rabbit antibody samples at the Detection step, indicating the high specificity of the anti-human IgG reagent. Moreover, the signal amplification at the Detection step at a low concentration of CR3022 suggests that use of the colloidal gold-conjugated anti-human IgG secondary antibody may enable the detection of RBD-specific antibodies even at low titers, potentially allowing the characterization of weakly seropositive samples.

To ascertain the approximate dynamic range for the detection of RBD-specific antibodies, we tested a concentration series of CR3022 ranging from 0.037 to 27 μg/mL (Supplementary Fig. S7). In the Total Antibody Binding step, antibody concentrations from 0.037 to 3 μg/mL CR3022 demonstrated a linear dose–response relationship, while the Detection step was linear with 0.037–1 μg/mL of CR3022.

Finally, we identified a 1:8 plasma dilution as an optimal screening dilution for BLI-ISA to minimize assay time and sample volume, while maximizing the dynamic range during the Total Antibody Binding step (Fig. 1.6). Our assay utilizes a 10-min Total Antibody Binding step with plasma, however if sample volume is limited, a higher dilution factor could be used, and this could be com-

pensated for with a longer Total Antibody Binding step to achieve the same signal. Similarly, we identified a 1:10 secondary antibody dilution as an optimal dilution to minimize assay time and maximize the signal during the Detection step (data not shown). Importantly, the real-time measurements of the BLI-ISA allow this assay flexibility which, to our knowledge, is not practical with any other currently available serological assays.

Measurement of plasma antibody levels by BLI-ISA. With the method optimized, we then performed BLI-ISA on our test set of 37 plasma samples (Fig. 1.6). All samples were evaluated at 1:8 dilution in biological duplicates with RBD-biotin-coated biosensors as well as control "no antigen" biosensors. In contrast to other serological tests, which detect either total antibodies or isotype-specific antibodies, BLI-ISA gives signals in real-time as plasma antibodies bind, thus allowing for measurements at this Total Antibody Binding step, as well as measurements at the Detection step after addition of anti-human IgG secondary antibody binding. As a result, we report the results of these two steps separately as the overall change in signal at each step (Fig. 1.5).

Data from the Total Antibody Binding step revealed that the strongest SP samples (8 of the 10 samples) could be easily distinguished from the SN samples (Fig. 1.6a). Significantly, the trends obtained with this single dilution in BLI-ISA generally align with the AUC values obtained by dilution-series ELISA (Table 1.1), revealing single- dilution BLI-ISA as a rapid method to identify and differentiate antibody levels of strongly SP samples. Using a cutoff of the mean of seronegative samples plus three standard deviations (mean SN signal = 0.140; standard deviation = 0.088) unambiguously identifies 8 of the 10 SP samples as

positive. SP3 and SP8 do not have sufficient RBD-specific antibody levels to test positive in this step. Two seronegatives (SN15 and SN17) appear slightly positive, likely due to background signal intrinsic to these specific samples given that a similar signal is observed in the absence of antigen. Additionally, using a more conservative cutoff of the SN mean plus five standard deviations removes this ambiguity without impacting the remaining SP samples.

To specifically measure anti-RBD IgG levels, the sensors are then dipped in wells containing colloidal gold conjugated anti-human IgG. Data from this Detection step distinguishes between all 27 SN samples and 10 SP samples (Fig. 3b). Here, all SP samples are positive, even with the more conservative cutoff of the SN mean plus five standard deviations. Notably, SP8 is clearly positive, consistent with ELISA results (Fig. 1a), while SP3 is weakly positive. All the SN samples are below both cutoffs, and we no longer observe nonspecific background positivity ("no antigen" wells do not approach detection threshold).

Interestingly, the signals from the strong SP samples in the Detection step do not strongly correlate with their signals in the Total Antibody Binding step. This suggests a potential saturation effect, where the large number of antibodies bound during the Total Antibody Binding step saturates the biosensor surface and curtails the signal during the Detection step with anti-human IgG. To confirm this, we performed BLI-ISA with a dilution series ranging from 1:4 to 1:32 using SP7 and SP8, which respectively gave the strongest and second weakest signals in the Total Antibody Binding step, but relatively similar signals in the detection step (Fig. 1.6c). As expected, the Total Antibody Binding step showed a predictable correlation between strength of signal and sample dilution for both

SP7 and SP8. In contrast, while this trend is also present for SP8 in the Detection step, there is actually a lower signal for SP7 at the highest concentration of plasma, followed by a rebound before it begins to follow a normal pattern. Thus, there appears to be a maximum combined signal from the antibodies in both steps.

To overcome the saturation effect at the Detection step, we evaluated all ten SP samples at a 1:16 dilution in biological duplicates and found that the signals in the Detection step now align with the respective signals in the Total Antibody Binding step as well as the AUC values obtained by dilution series ELISA (Table 1.1), supporting BLI-ISA as a semi-quantitative technology.

Given the observed boost in signal in SP8 at a 1:4 dilution, we investigated whether a follow-up assay with a higher proportion of plasma could boost the signal in the case of very weak seropositives, such as SP3. We per- formed a dilution series ranging from 1:2 to 1:16 with SP3 and the seronegatives SN20, SN21, SN26, and SN27, which represented the diversity of signals in the single dilution BLI-ISA (Fig. 1.6d). Data from the Total Antibody Binding step were ambiguous, with SP3 slightly separating from SN21 at the 1:2 dilution, although the signal observed in all of the SN samples increased. Data from the Detection step, however, revealed substantial signal improvement of SP3 while maintaining low signal for the SN samples, including SN21, which had produced the most background signal in the single dilution screen.

Overall, our data suggests that BLI-ISA can detect and rank seropositive samples at a single plasma dilution. In the case of strongly seropositive samples at a 1:8 dilution, the Total Antibody Binding step alone appears suf- ficient to clas-

sify them as seropositive and evaluate antibody levels. While the Detection step provides qualitative confirmation of IgG antibodies at a 1:8 dilution, this step can evaluate antigen-specific IgG levels at a 1:16 dilu- tion. Importantly, the Detection step is able to identify all ten SP samples as positive, including moderate and weak seropositive samples, even with a conservative cutoff of the SN mean plus five standard deviations. Similar to ELISA assays, dilution-series BLI-ISA can be used to re-assess ambiguous samples.

Comparison of RBD versus prefusion spike in BLI-ISA.

To confirm that our observations for the RBD would be applicable more generally, we performed the BLI-ISA using a trimeric, prefusion-stabilized spike (prefusion spike) with all the seropositive samples (Fig. 1.7a). Due to sample limitations, we limited our negative controls to the two seronegative samples of which we had the greatest quantity. In general, the Total Antibody Binding step was consistent with what was observed with the RBD. The range in signal between samples was less extreme, which could be due to different sizes of the two antigens, different loading capabilities of the different biosensor surfaces (SA versus HIS1K biosensors), and/or the greater number of epitopes present on the prefusion Spike compared to the RBD. In addition, the Detection step with the anti-human IgG antibody generally reflected the magnitude in the first step, likely as a result of the lower overall signal. Intriguingly, the Detection step resulted in strong downward, negative curves for the SP samples, while remaining flat for the SN samples (Supplementary Fig. S8). Similar inverted signals have been observed in other applications of BLI and are generally associated with binding of large macromolecules or complexes36. This is hypothesized to occur due to a large change

in the optical thickness of the biological layer that results in an unusually large shift in the wavelength. Thus, it is likely that the increased size of the prefusion spike compared to RBD creates a threshold that is crossed during the Detection step. To confirm that the negative curves were dependent on the antigen and not an artifact of the HIS1K biosensors, we performed the assays again using biotinylated, prefusion spike with SA biosensors. As anticipated, similar inverted signals were observed with these biosensors (Supplementary Fig. S8), consistent with a model of inverted signal that is driven predominantly by the thickness of the antigen–antibody biological layer.

Detecting different antibody isotypes.

Although IgG is the most prevalent antibody isotype in circulation, making it ideal for serosurveillance, in some cases it may be desirable to identify different isotypes in a sam- ple or target a specific physiological context where another antibody isotype is most prevalent. IgA, for example, is most predominant in mucosal membranes, but is also present at a low percentage in blood. To assess whether we could specifically detect IgA, we evaluated the signal in the Detection step using a colloidal gold-conjugated anti-IgA antibody (Fig. 1.7b). Not surprisingly, the signals were substantially lower than those observed for the anti-IgG antibody and were undetectable in some cases. However, some samples showed a clear signal, espe- cially several of the strong SP samples. Intriguingly, while the SP samples generally showed a positive signal, SP7 showed a negative curve. This may indicate that the positioning of IgA binding in this sample results in a thicker biological layer upon secondary antibody binding. Alternatively, it may also be consistent with a model of steric competition between weakly bound IgG

and the IgA-specific antibody, or due to differences in antibody epitope usage in this individual sample. Excitingly, these results suggest that different isotypes can be detected from a single sample by BLI-ISA, including those that are not the predominant isotype.

1.1.5 Discussion

Here, we describe BLI-ISA, a novel serological testing method to measure antigen-specific antibodies in plasma utilizing biolayer interferometry. The assay is simple to perform; samples are pipetted into a 96- or 384-well plate, the plate is loaded into the BLI instrument along with single-use biosensors, and the assay program is run, direct- ing the instrument to dip biosensors in different wells for each step (Fig. 2, Supplementary Fig. S2). Moreover, the assay is rapid, with real-time data output and full results in $<$ 20 min following sample preparation. The assay is semi-quantitative and measures relative levels of both total antibodies and specific antibody isotypes in the same assay, and single-dilution BLI-ISA signals from diverse seropositive samples align with AUC values obtained by dilution-series ELISA. We acknowledge the relatively small numbers of pre-pandemic samples used for this proof-of-principal study are not sufficient to precisely define the lower limit of detection, however evaluation of more pre-pandemic samples will ultimately make this limit clear.

Beyond the advantage of short assay times, BLI-ISA has technical advantages over ELISA, IFA, and CLIA in two key ways; (1) BLI-ISA does not require washing of wells or beads, eliminating time-consuming manual washing or use of fluidic

instrumentation, and (2) BLI-ISA does not utilize enzyme-based signal amplification (i.e. HRP), which can vary due to differences in temperature, pH, and manufacturing lots of enzyme-conjugated reagent. As a result, BLI-ISA overcomes the lab-to-lab variability that can occur with other methods and provides a solution to standardize other serological testing methods as well as to perform longitudinal studies of biological samples. We do acknowledge a caveat of our BLI-ISA method is the need for BLI instrumentation, however this instrument is becoming more commonplace at research institutions, and in many cases the acquisition of such an instrument would be beneficial due to reduced labor and faster data acquisition time. Regarding costs, we estimate that evaluation of a sample by ELISA dilution series costs $1 in supplies, whereas evaluation of a sample by BLI-ISA costs $7 in supplies (mainly due to the cost of the biosensor). However, when adding in the cost of a technician's time, we estimate that the total cost to evaluate a sample by ELISA or by BLI-ISA is similar, at $< \$10$ per sample.

In a broader sense, BLI-ISA can be adapted, multiplexed, and performed in a high-throughput fashion. Straightforward adaptions can allow for measurement of antibodies against different antigens, as we demonstrated for RBD and prefusion Spike, and/or detection of different antibody isotypes, as we demonstrated for IgG and IgA antibodies. There is also potential to measure antibodies in other biological specimens (e.g. breastmilk, saliva) and also measure other clinically relevant, non-antibody biological molecules in human and animal specimens. Finally, BLI-ISA has the potential to be performed in a high-throughput fashion. While many institutions have BLI instruments that measure 8 or 16 biosensors at a time, the Octet HTX instrument can measure 96 biosensors at a time. In addition, we found that RBD-biotin could be pre-loaded onto SA biosensors with no loss in signal

over at least three hours (data not shown), suggesting that biosensors could be pre-loaded to eliminate this step from the assay method and save time. Thus, with antigen pre-loading and the use of an Octet HTX instrument, approximately 3000 samples could be analyzed in an 8-h day.

Here, we used BLI-ISA to detect antibodies to SARS-CoV-2 using the spike RBD antigen. We chose this antigen because it is highly selective for antibodies to SARS CoV-2 and because antibodies to the RBD have been shown to correlate with virus neutralization[7,8,10,12–16]. We also found that use of prefusion Spike antigen shows similar trends in seroreactivity. Our BLI-ISA method can be immediately implemented for urgent SARS-CoV-2 serological testing needs. First, our method could be used for serosurveillance studies to evaluate seroconversion in communities. Second, and importantly, BLI-ISA can be used to evaluate antibody responses to natural infection and vaccine candidates to define correlates of immunity to SARS-CoV-2 infection. Finally, we believe that BLI-ISA can be developed as a novel diagnostic platform to evaluate antibodies and other biomolecules in clinical specimens, for example to evaluate plasma antibody levels to inform patients on vaccinations, or to quickly identify and prioritize donors for convalescent plasma therapy donation[37,38].

1.1.6 Methods

Reagents and Supplies

Phosphate buffered saline (PBS) tablets (Sigma P4417), Tween-20 (Fisher BP337), dry milk powder (RPI 50488786), ELISA plates (Corning 3590), Goat anti-Human

IgG Fc HRP (Thermo Fisher A18817), OPD tablets (Pierce PI34006), bovine serum albumin (BSA) (Fisher BP1600), ChonBlock (Chondrex 9068), Biotinylated SARS-CoV-2 protein RBD His AviTag (Acro Biosystems SPD-C82E9), 4 nm Colloidal Gold- AffiPure Goat Anti-Human IgG Fcg fragment specific (Jackson ImmunoResearch 109-185-098) rehydrated in 1 mL deionized water per the manufacturer's instructions, 4 nm Colloidal Gold-AffiPure Goat Anti-Human Serum IgA alpha chain specific (Jackson ImmunoResearch 109-185-011) rehydrated in 1 mL deionized water per the manufacturer's instructions, Human coronavirus spike glycoprotein Antibody, Rabbit PAb, Antigen Affinity Purified (Sino Biological 40021-T60), SARS-CoV-2 (2019-nCoV) spike Antibody, Rabbit PAb, Antigen Affinity Purified (Sino Biological 40589-T62), SARS-CoV-2 (2019-nCoV) spike Antibody, Rabbit MAb (40150- R007), Anti-SARS-CoV S Therapeutic Antibody (CR3022) (Creative Biolabs MRO-1214LC), Octet Anti-Penta- His (HIS1K) sensor tips (Sartorius ForteBio 18-5120), Octet Streptavidin (SA) sensor tips (Sartorius ForteBio 18-5019), tilted bottom (TW384) microplates (Sartorius ForteBio 18-5080), electroporation cuvettes (Max- Cyte SOC4), suspension adapted CHO-S cells (Thermo Fisher R80007). CD-CHO medium (Thermo Fisher 10743029), CD OptiCHO medium (Thermo Fisher 12681011), HisTrap FF (GE Healthcare 17-5286-01), Strep- Trap HP (GE Healthcare 28-9075-47), Superdex 200 Increase GL (GE Healthcare 28-9909-44).

Recombinant SARS-CoV-2 spike proteins

The expression plasmid for the SARS CoV-2 spike RBD- His was obtained from BEI Resources. This pCAGGS plasmid encodes the signal peptide (residues 1–14) and RBD (residues 319–541) of the SARS CoV-2 spike (GenBank: MN908947.3),

fused to a C-terminal 6XHis-tag [12]. To generate the expression plasmid for RBD-biotin, the cDNA encoding the SARS-CoV-2 spike signal peptide, RBD and 6XHis-tag in pCAGGS was sub-cloned by Gibson Assembly into a derivative of pcDNA3.1 [39] in frame with a Strep-tag and AviTag at the C-terminus. The pαH expression plasmid encoding the prefusion-stabilized SARS CoV-2 spike trimer was described previously [31]. Recombinant prefusion-stabilized spike trimer (prefusion Spike-His) produced in ExpiCHO cells was a generous gift from the Almo lab (Albert Einstein College of Medi- cine) [32]. Recombinant prefusion-stabilized spike trimer containing an AviTag biotinylation acceptor sequence (prefusion Spike-biotin) was obtained from BEI Resources and was biotinylated in vitro as described below. Recombinant RBD proteins were produced in suspension adapted CHO-S cells. CHO-S cells were maintained in CD-CHO medium supplemented with 8 mM GlutaMAX, 0.1 mM Hypoxanthine, and 0.016 mM thymidine (HT) in shake flasks using a Khuner shaker incubator at 37 °C, 8% CO2, and 85% humidity. For protein production, CHO-S cells were transfected with purified endotoxin-free DNA using flow electroporation technology (MaxCyte). Transfected cells were grown at 32 °C in CD OptiCHO medium supplemented with 2 mM GlutaMAX, HT supplement, 0.1% pluronic, and 1 mM sodium butyrate, supplementing daily with MaxCyte CHO A Feed (comprised of 0.5% Yeastolate, 2.5% CHO-CD Efficient Feed A, 2 g/L Glucose, and 0.25 mM GlutaMAX). On day 8 post-transfection, cells were centrifuged at 4000×g for 15 min, and the media was 0.22-μ m filtered. For purification of RBD-His, media was diluted with Buffer A (300 mM NaCl, 50 mM NaH2PO4, 20 mM imidazole [pH 7.4]) and loaded onto a HisTrap column. The column was washed with Buffer A, and RBD-His was eluted with a gradient to Buffer B (300 mM NaCl, 50 mM NaH2PO4, 225 mM imidazole [pH 7.4]). RBD-His was further purified by size-exclusion chromatography on a

Superdex 200 column in PBS and the fractions containing pure monomeric RBD-His were pooled and concentrated to 1.03 mg/mL. For purification of RBD-biotin, media was supplemented with 20 mM Tris pH 8.0 and 150 mM NaCl (TBS), 1 mM EDTA, and 27 mg/L BioLock and loaded onto a StrepTrap column. The column was washed with TBS containing 1 mM EDTA, and RBD protein was eluted with a gradient of TBS, 1 mM EDTA, and 2.5 mM desthiobiotin. Additional RBD protein in the media was obtained by dialysis and purification with a HisTrap column as described above. The purest elution fractions from each purification were pooled and dialyzed overnight into PBS. Biotinylation of the AviTag was achieved following published procedures[40]. Briefly, 46 μ M RBD was incubated overnight at room temperature with 3 μ M recombinant GST-BirA biotin ligase in PBS containing 5 mM MgCl2, 25 mM ATP, and 625 μ M D-biotin. RBD-biotin was further purified by size-exclusion chromatography on a Superdex 200 column in PBS and the fractions containing pure monomeric RBD-biotin were pooled and concentrated to 315 μ g/mL. All purified recombinant proteins were aliquoted, flash frozen in liquid nitrogen, and stored at -80 °C until use.

Human samples.

Human plasma and serum samples were obtained from several sources. The pre-pandemic seronegative (SN) panel comprised of de-identified samples selected based on the date of collection, before the emergence of SARS-CoV-2. First, human serum samples (n = 25) collected in 2017 were from study participants enrolled in an Institutional Review Board-approved study for development of Lyme disease and other diagnostic tests. To this end 550 samples were collected from individuals on the East Coast and in the Upper Midwest of the United States

where Lyme disease is endemic41. Samples were obtained from the Lyme Disease Biobank as part of a research collaboration with Ontera Inc. (Santa Cruz, CA, USA). Second, human plasma samples (n = 2) collected in 2016 were from study participants enrolled in an Institutional Review Board-approved study at UCSC. All participants agreed to sample banking and future research use. The convalescent seropositive (SP) panel comprised 10 de-identified plasma samples from nine individuals, purchased from AllCells (Alameda, CA, USA). To be eligible for plasma donation, prospective donors must have either had COVID-19 symptoms resolved at least 28 days after the diagnosis or suspicion of COVID-19 (i.e. no fever, cough, difficulty breathing, etc.) or been 14 days symptom free with a follow up negative nasopharyngeal/PCR test. Each SP sample had been tested by an Anti-SARS-CoV-2 Total test (Ortho Clinical Diagnostics) and was given a CoV2T score (ranging from 8 to 440) (Table 1.1)). All serum and plasma samples were heated at 56 °C for 1 h before use.

ELISA

The ELISA protocol was adapted from a previously published protocol [12]. ELISA plates (96-well) were incubated overnight at 4 °C with 50 μ L per well of 2 μ g/mL RBD-His in PBS. After removal of RBD-His, plates were blocked for 1 h at room temperature with 300 μ L per well of 3% non-fat milk in PBS with 0.1% Tween 20 (PBST). After removal of blocking buffer, 100 μ L plasma/serum samples diluted 1:50 in 3% milk in PBST were added to wells and incubated for 2 h at room temperature with gentle shaking. For dilution series ELISAs, plasma/serum samples were first diluted 1:50 in 3% milk in PBST and then diluted 1:4 in series in 3% milk in PBST. The human monoclonal antibody CR3022 antibody, which is reac-

tive to the RBD of both SARS-CoV-1 and SARS-CoV-2, was used as a positive control[33–35]. The plates were washed three times with PBST. After washing, 100 μ L goat anti-human IgG Fc horseradish peroxidase (HRP) conjugated secondary antibody diluted 1:3000 in 1% milk in PBST was added to each well and incubated for 1 h at room temperature with gentle shak- ing. Plates were again washed three times with PBST. The plates were washed three times with PBST, followed by addition of 100 μ L o-phenylenediamine dihydrochloride (OPD) solution to each well. The substrate was left on the plates for exactly 10 min and then the reaction was stopped by adding 50 μ L per well of 3 M hydrochloric acid. The optical density at 490 nm (OD490) was measured using a Molecular Devices Spectramax plate reader. The background value was set at an OD490 of 0.051 based on the average PBS measurement and subtracted from all data prior to curve fitting and AUC calculations. The AUC values were calculated by fitting a four-parameter logistic regression model to the OD490 data of each sample using the curve fit algorithm from the SciPy Python Library[42] followed by integration between the upper and lower bounds of the data.

BioLayer interferometry immunosorbent assay (BLI-ISA).

BLI-ISA studies were performed on an Octet RED384 instrument at 24 °C with shaking at 1000 rpm. BLI assay buffer consists of 2% BSA in PBST, which was 0.22-μm filtered. Before use, anti-penta-His (HIS1K) or (anti-biotin) streptavidin (SA) biosensors (were loaded into the columns of a biosensor holding plate and pre-hydrated in BLI assay buffer for 10 min. Tilted bottom 384-well microplates were loaded with 45 μL per well. The assay plate was prepared as follows: column 1 (BLI assay buffer), column 2 (2–12 μg/mL RBD or prefusion Spike in BLI assay

buffer), column 3 (25% ChonBlock in BLI assay buffer), column 4 (plasma/serum samples diluted 1:8 in 25% ChonBlock in BLI assay buffer), column 5 (BLI assay buffer), and column 6 (4 nm Colloidal Gold-AffiPure Goat Anti-Human IgG or IgA secondary antibody diluted 1:10 in BLI assay buffer). For dilution series studies, samples were diluted into stock solutions of ChonBlock in BLI assay buffer to yield a 25% ChonBlock solution after plasma dilution. RBD-biotin purchased from Acro Biosystems was used at 2 μ g/mL for 1:8 single-dilution studies. RBD-biotin produced in-house at UCSC was not fully biotinylated and required use at 10 μg/mL to achieve the same loading signal. RBD-biotin produced at UCSC was used for dilution-series studies and anti-human IgA studies. RBD-His and Prefusion Spike-His were used at 10 μg/mL.

The BLI-ISA method was set as follows. Baseline1 (60 s) in column 1 (Equilibration), Loading (120 s or 600 s) in column 2 (Antigen Loading: RBD or prefusion Spike, respectively), Baseline2 (60 s) in column 3 (Wash), Association1 (600 s) in column 4 (Total Antibody Binding), Baseline3 (60 s) in column 5 (Wash), and Association2 (180 s) in column 6 (Detection: anti-human IgG or IgA). Loading of RBD-biotin over 120 s onto SA sensor tips resulted in a wavelength shift signal of 2.2 nm (Supplementary Fig. S4). Loading of RBD-His, prefusion Spike- His, or prefusion Spike-biotin over 600 s resulted in a wavelength shift signal of 1.0 nm (Supplementary Fig. S4). We note that loading density on sensor tips had little effect on antibody binding signals in Association 1 or 2.

A Python program was written to automate analysis of BLI-ISA data. Raw data .csv files were exported from the Octet Data Analysis software and read by our program. Our script determined the Total Antibody Binding (Association 1)

value by subtracting the average wavelength shift of seconds 2–4 of this step from the average shift of the last 5 s of this step. Similarly, the Detection (anti-human IgG or IgA binding) (Association 2) value was determined by subtracting the average of the data 1–2 s after the start of this step from the average of the data from the last 5 s of this step. Complete raw data traces were also plotted and inspected to ensure proper antigen loading onto sensor tips where applicable. This Python program named bli_plotter is available for adaption to other BLI-ISA studies and has been deposited on GitHub (https://github.com/nlorigroach/bli_plotter).

1.1.7 References

1. Wu, F.etal. A new corona virus associated with human respiratory disease in China.Nature 579,265–269. https://doi.org/10.1038/s41586-020-2008-3 (2020).

2. Organization, W.H. WHO Director-General's opening remarks at the media briefing on COVID-19(2020).

3. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5, 536–544. ://doi.org/10.1038/s41564-020-0695-z (2020).

4. Chu, D.K.W. et al. Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. Clin.Chem. 66, 549–555. https://doi.org/10.1093/clinchem/hvaa029 (2020).

5. Waggoner, J.J. et al. Triplex real-time RT-PCR for severe acute respiratory syndrome coronavirus 2. Emerg. Infect. Dis. 26,1633–1635. ://doi.org/10.3201/eid2607.201285 (2020).

6. Corman, V. M. et al. novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. ://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045 (2019).

7. Geurts van Kessel, C.H. et al. An evaluation of COVID-19 serological assays informs future diagnostics and exposure assessment. Nat. Commun. 11, 3436. https://doi.org/10.1038/s41467-020-17317-y (2020).

8. Okba, N.M.A. et al. Severe acute respiratory syndrome coronavirus 2-specific antibody responses in coronavirus disease patients. Emerg. Infect. Dis. 26, 1478–1488. https://doi.org/10.3201/eid2607.200841 (2020).

9. Long, Q.X. et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat.Med. 26, 845–848. https://doi.org/10.1038/s41591-020-0897-1 (2020).

10. Wu, F. et al. Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. medRxiv. https://doi.org/10.1101/2020.03.30.20047365 (2020).

11. Zhao, J. et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin. Infect. Dis. https://doi.org/10.1093/cid/ciaa344 (2020).

12. Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. Nat. Med. https://doi.org/10.1038/s41591-020-0913-5 (2020).

13. Poh, C. M. et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nat. Commun. 11, 2806. https://doi.org/10.1038/s41467-020-16638-2 (2020).

14. Salazar, E. et al. Relationship between anti-spike protein antibody titers and SARS-CoV-2 in vitro virus neutralization in convalescent plasma. BioRxiv. https://doi.org/10.1101/2020.06.08.138990 (2020).

15. Zost, S.J. et al. Potently neutralizing and protective human antibodies against SARS-CoV-2. Nature https://doi.org/10.1038/s41586-020-2548-6 (2020).

16. Zost, S.J. et al. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. Nat. Med. https://doi.org/10.1038/s41591-020-0998-x (2020).

17. Wu, J. L. et al. Four point-of-care lateral flow immunoassays for diagnosis of COVID-19 and for assessing dynamics of antibody responses to SARS-CoV-2. J. Infect. https://doi.org/10.1016/j.jinf.2020.06.023 (2020).

18. FDA. Certain COVID-19 Serology/Antibody Tests Should Not Be Used—Letter to Clinical Laboratory Staff and Health Care Providers (2020).

19. Whitman, J. D. et al. Test performance evaluation of SARS-CoV-2 serological assays. medRxiv. https://doi.org/10.1101/2020.04.25.20074856 (2020).

20. FDA. EUA Authorized Serology Test Performance(2020).

21. Nicol, T. et al. Assessment of SARS-CoV-2 serological tests for the diagnosis of COVID-19 through the evaluation of three immunoassays: two automated immunoassays (Euroimmun and Abbott) and one rapid lateral flow immunoassay (NG Biotech). J. Clin. Virol. 129, 104511. https://doi.org/10.1016/j.jcv.2020.104511 (2020).

22. Abdiche, Y., Malashock, D., Pinkerton, A. & Pons, J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. Anal. Biochem. 377, 209–217. https://doi.org/10.1016/j.ab.2008.03.035 (2008).

23. Coats, M.A.-O. et al. mRNA transfection into CHO-cells reveals production bottlenecks. Biotechnol. J. 15, 1900198 (2020).

24. Carbonetti, S. et al. A method for the isolation and characterization of functional murine monoclonal antibodies by single B cell cloning. J. Immunol. Methods 448, 66–73. https://doi.org/10.1016/j.jim.2017.05.010 (2017).

25. Mechaly, A., Cohen, H., Cohen,O. & Mazor, O. A biolayer interferometry-

based assay for rapid and highly sensitive detection of biowarfare agents. Anal. Biochem. 506, 22–27. https://doi.org/10.1016/j.ab.2016.04.018 (2016).

26. Carney, P.J., Lipatov, A.S., Monto, A.S., Donis, R.O. & Stevens, J. Flexible label-free quantitative assay for antibodies to influenza virus hemagglutinins. Clin. Vaccine Immunol. 17, 1407–1416. https://doi.org/10.1128/CVI.00509-09 (2010).

27. Auer, S. et al. Rapid and sensitive detection of norovirus antibodies in human serum with a biolayer interferometry biosensor. Sens. Actuators B Chem. 221, 507–514. https://doi.org/10.1016/j.snb.2015.06.088 (2015).

28. Gandolfini, I. et al. Rapid biolayer interferometry measurements of urinary CXCL9 to detect cellular infiltrates non invasively after kidney transplantation. Kidney Int. Rep. 2, 1186–1193. https://doi.org/10.1016/j.ekir.2017.06.010 (2017).

29. Gao, S., Zheng, X. & Wu, J. A biolayer interferometry based enzyme linked aptamer sorbent assay for real-time and highly sensitive detection of PDGF-BB. Biosens. Bioelectron. 102, 57–62. https://doi.org/10.1016/j.bios.2017.11.017 (2018).

30. Ziu, I., Laryea, E. T., Alashkar, F., Wu, C. G. & Martic, S. A dip-and-read optical aptasensor for detection of tau protein. Anal. Bioanal. Chem. 412, 1193–1201. https://doi.org/10.1007/s00216-019-02350-8 (2020).

31. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263. https://doi.org/10.1126/science.abb2507 (2020).

32. Herrera, N. G. et al. Characterization of the SARS-CoV-2 S protein: biophysical, biochemical, structural, and antigenic analysis. bioRxiv. https://doi.org/10.1101/2020.06.14.150607 (2020).

33. ter-Meulen, J. et al. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. PLoS Med. 3, e237. https://doi.org/10.1371/journal.pmed.0030237 (2006).

34. Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV.Science 368, 630–633. https://doi.org/10.1126/science.abb7269 (2020).

35. Tian, X. et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg. Microbes Infect. 9, 382–385. https://doi.org/10.1080/22221751.2020.1729069 (2020).

36. ForteBio (2019) (Accessed4November2020); https://www.fortebio.com/sites/default/files/en/assets/app-note/octet-potency-assay-development-qualification-and-validation-strategies.pdf.

37. Cheng, Y. et al. Use of convalescent plasma therapy in SARS patients in Hong Kong.Eur. J. Clin. Microbiol. Infect. Dis. 24,44–46. https://doi.org/10.1007/s10096-004-1271-9 (2005).

38. Leider, J.P., Brunker, P.A. & Ness, P.M. Convalescent transfusion for pandemic influenza: preparing bloodbanks for a new plasma product?. Transfusion 50, 1384–1398. https://doi.org/10.1111/j.1537-2995.2010.02590.x (2010).

39. O'Rourke, S. M. et al. Robotic selection for the rapid development of stable CHO cell lines for HIV vaccine production. PLoS ONE 13, e0197656. https://doi.org/10.1371/journal.pone.0197656(2018).

40. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. Methods Mol. Biol. 1266, 171–184. https://doi.org/10.1007/978-1-4939-2272-7_12 (2015).

41. Horn, E. J. et al. The Lyme Disease Biobank: characterization of 550 patient and control samples from the east coast and upper midwest of the United States.

J. Clin. Microbiol. https://doi.org/10.1128/JCM.00032-20 (2020).

42. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2 (2020).

1.1.8 Acknowledgements

The authors are grateful for the support from the following individuals. Natalia Herrera, Nicholas Morano, Scott Garforth, and Steven Alamo (Albert Einstein College of Medicine) provided recombinant prefusion Spike protein. The Lyme Disease Biobank donated pre-pandemic human serum samples via a research collaboration with Ontera Inc. Phillip Berman shared a derivative of pcDNA3.1 plasmid and allowed use of MaxCyte electroporation instrumentation and Kuhner shaker incubators. The following reagent was produced under HHSN272201400008C and obtained through BEI Resources, NIAID, NIH: Vector pCAGGS Containing the SARS-Related Coronavirus 2, Wuhan-Hu-1 Spike Glycoprotein Receptor Binding Domain (RBD), NR-52309. The following reagent was obtained through BEI Resources, NIAID, NIH: Spike Glycoprotein (Stabilized) from SARS- Related Coronavirus 2, Wuhan-Hu-1 with C-Terminal Histidine and Avi Tags, Recombinant from HEK293F Cells, NR-53524. We thank UCSC colleagues for advice, especially Susan Carpenter, Camilla Forsberg, Ed Green, David Haussler, Sofie Salama, Beth Shapiro, and Josh Stuart. Funding for this research came from the University of California Office of the President, the UC Santa Cruz Office of Research, the Michael-David Family Foundation, and donors to the UC Santa Cruz Respiratory Virus Fund. Funding for the purchase of the Octet

RED384 instrument was supported by the NIH S10 shared instrumentation grant 1S10OD027012-01.

1.1.9 Additional information

Open Access

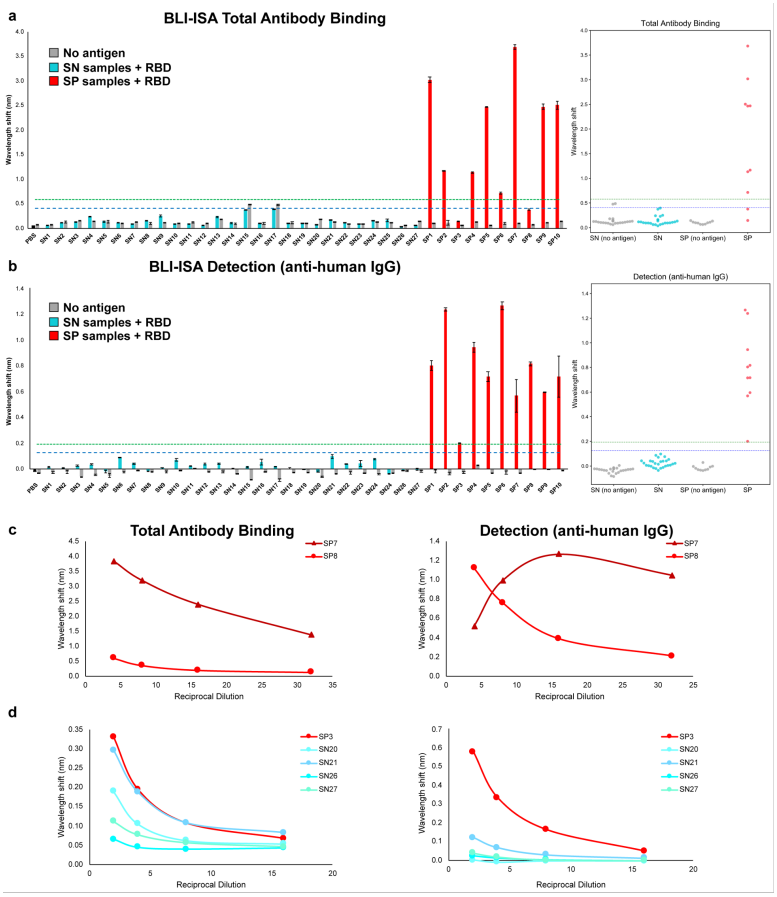


Figure 1.6 BLI-ISA evaluation of SARS-CoV-2 spike RBD reactivity of pre-pandemic and convalescent plasma. (a,b) Single-dilution BLI-ISA to evaluate the presence of RBD-reactive human antibodies in the pre- pandemic seronegative (SN, cyan) and convalescent seropositive (SP, red) samples compared to no-antigen controls (grey). The assays were performed with plasma at a 1:8 dilution. Bars and dots represent the mean of biological duplicates, and error bars represent one standard deviation from the mean. Blue and green dashed lines represent the mean of seronegative samples plus 3 and 5 standard deviations, respectively. (a) The Total Antibody Binding signal is measured when RBD-biotin-loaded SA biosensors are dipped into plasma samples. (b) The Detection signal is measured when RBD-biotin-loaded SA biosensors that had been dipped into plasma are subsequently dipped into colloidal gold-conjugated anti-human IgG. (c) Dilution series BLI-ISA from representative strong (SP7) and moderate (SP8) seropositive samples. (d) Dilution series BLI-ISA from the weakest seropositive sample (SP3) compared to seronegative plasma samples.

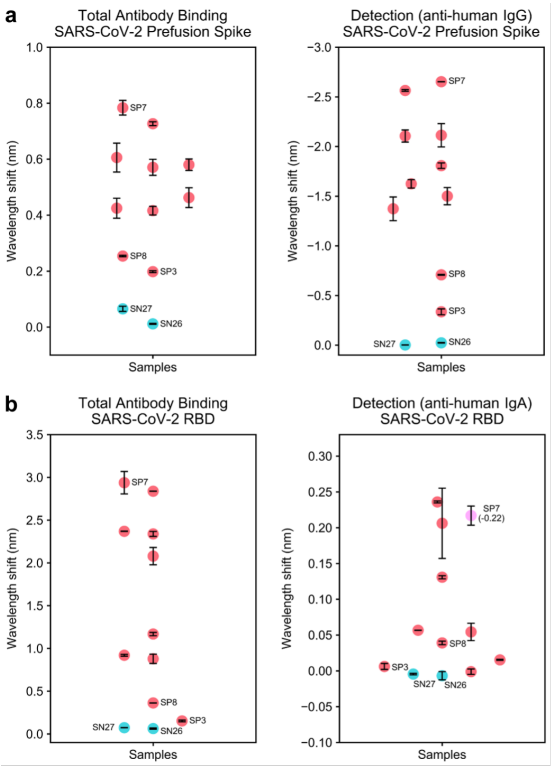


Figure 1.7 BLI-ISA evaluation of plasma antibodies to SARS-CoV-2 prefusion Spike and plasma IgA to SARS- CoV-2 spike RBD. (a) Single-dilution BLI-ISA to evaluate the presence of prefusion Spike-reactive human antibodies in the pre-pandemic seronegative (SN, cyan) and convalescent seropositive (SP, red) samples. The Total Antibody Binding signal (left) is measured when prefusion Spike-His-loaded HIS1K biosensors are dipped into plasma samples. The Detection signal (right) is measured when prefusion Spike-His-loaded HIS1K biosensors that had been dipped into plasma are subsequently dipped into colloidal gold-conjugated anti-human IgG. (b) Single-dilution BLI-ISA to evaluate the presence of RBD-reactive human antibodies in the samples. The Total Antibody Binding signal (left) is measured when RBD-biotin-loaded SA biosensors are dipped into plasma samples. The Detection signal (right) is measured when RBD-biotin-loaded SA biosensors that had been dipped into plasma are subsequently dipped into colloidal gold-conjugated anti-human IgA. The SP7 dot is colored pink to indicate that this sample had a negative signal (value in parentheses) in the Detection step. All assays were performed with plasma at a 1:8 dilution. Dots represent the mean of biological duplicates, and error bars represent one standard deviation from the mean.

1.2 COV2-BLI-ISA EUA application

1.2.1 Preface

Along with our initial publication describing the BLI-ISA experiment, we sought a patent and FDA approval to use the method at UCSC as an antibody test under an Emergency Use Authorization. Shortly after the onset of the COVID-19 pandemic, a public health emergency was declared by the US Department of Health and Human Services which then authorized methods to rapidly approve potential treatments, vaccines, and diagnostic tools. We felt our BLI-ISA technique was a valuable diagnostic tool that fell under that call to action. The goal of the application was to enable larger scale serology studies where we could share the results with our study participants, as well as allow clinical applications of BLI-ISA (e.g. assessment of prior virus exposure or whether vaccination led to adequate antibody production).

Part of our application to the FDA will be reproduced below, along with a discussion of the response we eventually received. It took about a year and a half to receive a response from the FDA, and by then we had moved on from SARS-CoV-2 work. To the credit of the employee who did respond, their response and criticisms were largely valid and quite helpful.

The original application is headed by descriptions of the applicants and brief formal introductions to aspects of the test. Afterwards is a description of the test protocol and data establishing the scoring schema and accuracy of the serology test.

1.2.2 The Application

Purpose for Submission

Emergency Use Authorization (EUA) request for use of the COV2-BLI-ISA in CLIA Laboratory ID# 05D0687329, for the in vitro diagnostic use in the qualitative detection of IgG antibodies to SARS-CoV-2 in EDTA plasma using the Octet BLI System.

Measurand

The COV2-BLI-ISA test is a qualitative serology test that detects IgG antibodies in plasma reactive to the receptor binding domain (RBD) of the SARS-CoV-2 spike protein (Wuhan-Hu-1 isolate).

Proprietary and Established Names

Proprietary Name: COV2-BLI-ISA

Established Name: COV2-BLI-ISA

Proposed Intended Use

The COV2-BLI-ISA is a biolayer interferometry assay intended for qualitative detection of IgG antibodies to SARS-CoV-2 in human plasma with EDTA. The COV2-BLI-ISA is intended for use as an aid in identifying individuals with an adaptive immune response to SARS-CoV-2, indicating recent or prior infection. The COV2-BLI-ISA should not be used to diagnose or exclude acute SARS-CoV-2 infection. At this time, it is unknown for how long antibodies persist following infection and if the presence of antibodies confers protective immunity.

Testing is limited to laboratories certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), 42 U.S.C 263a, that meet the requirements to perform moderate or high complexity tests such as the laboratory at University of California Santa Cruz, Student Health Services, 1156 High Street, Santa Cruz, CA 95064 (CLIA Laboratory ID# 05D0687329).

Results are for the detection of SARS CoV-2 antibodies. IgG antibodies to SARS-CoV-2 are generally detectable in blood several days after initial infection, although the duration of time antibodies are present post-infection is not well characterized. Individuals may have detectable virus present for several weeks following seroconversion.

Laboratories within the United States and its territories are required to report all results to the appropriate public health authorities.

The sensitivity of COV2-BLI-ISA early after infection is unknown. Negative results do not preclude acute SARS-CoV-2 infection. If acute infection is suspected, direct testing for SARS-CoV-2 is necessary.

False positive results for COV2-BLI-ISA may occur due to cross-reactivity from pre-existing antibodies or other possible causes.

1.2.3 Device Description

Biolayer interferometry (BLI) is a light-based technique that allows measurement of intermolecular interactions and kinetics. BLI uses fiber optic tips with surfaces that can be functionalized with a variety of organic reagents and biomolecules including proteins and protein-tag binding molecules. After loading an analyte of interest onto the fiber optic tip, binding kinetics can be inferred by dipping the loaded analyte into solution containing a putative ligand and measuring changes in the reflected light from the tip surface.

In the COV2-BLI-ISA assay (figure 1.8), streptavidin-functionalized fiber optic tips (also called biosensors) are loaded with biotinylated receptor binding domain (RBD) from the spike protein of SARS-CoV-2. The RBD biosensors are next dipped in plasma samples where IgG, other antibody subtypes, and other factors may associate with the SARS-CoV-2 RBD. Finally, the biosensors, now loaded with RBD in complex with any RBD-binding agents from the plasma, are dipped in a reagent containing anti-Human-IgG antibody conjugated to gold nanoparticles. Binding measured in this step determines the outcome of the assay.

The COV2-BLI-ISA uses the following:

- SARS-CoV-2 spike protein (Wuhan-Hu-1 isolate, GenBank: MN908947.3) RBD (residues 319–541) with C-terminal 10-Histidine tag and biotinylated AviTag
- 4 nm Colloidal Gold-AffiPure Goat Anti-Human IgG Fc gamma fragment specific (Jackson ImmunoResearch 109-185-098)

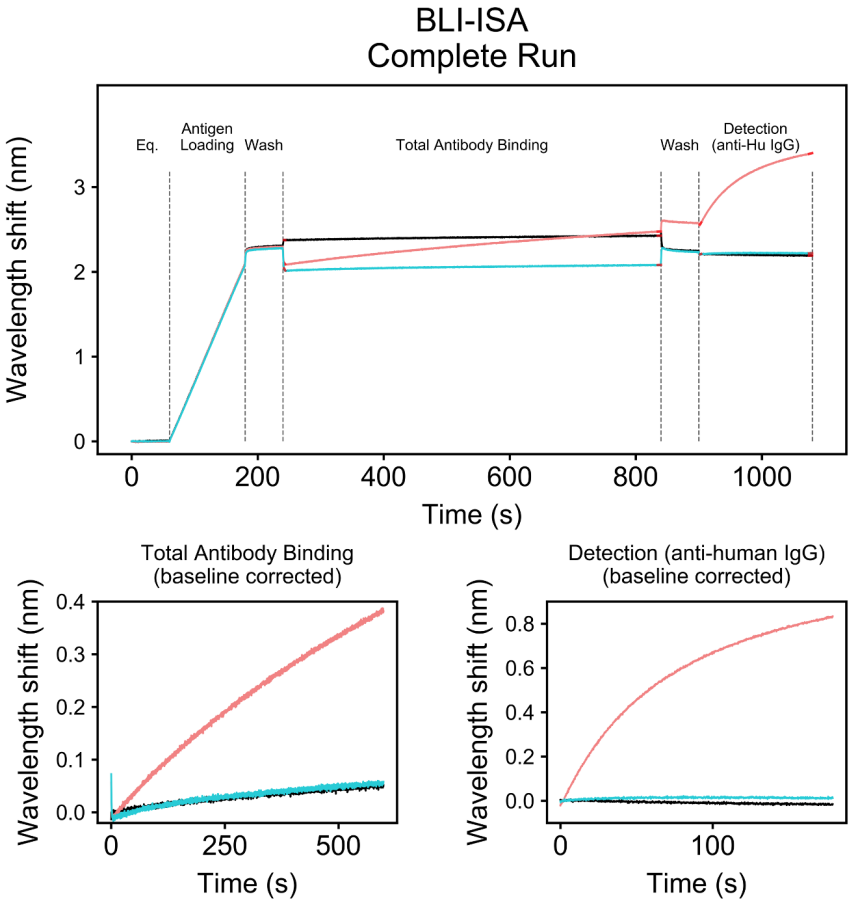


Figure 1.8 Overview of the COV2-BLI-ISA assay. The assay is performed in 6 steps, equilibration of sensor tips to buffer, loading of SARS-CoV-2 RBD antigen, establishment of baseline, total antibody binding (association of antibodies in diluted serum to antigen), a second baseline, and finally secondary detection of antibody subtype (here by an anti-human-IgG antibody conjugated to a gold nanoparticle).

1.2.4 Description of Test Steps

Specimen collection and preparation

1. Collect >200 uL blood by venipuncture. Add to sample tube containing EDTA (to 5 mM). Store on ice or at 4 °C until next step.

2. Centrifuge blood at 2500 x g for 15 minutes. Carefully pipette upper plasma

layer into new specimen tube or well plate suitable for the volume. A minimum of 20 uL to allow replicate assays to be performed for the 384-well format and a minimum of 60 uL is needed for replicate runs in the 96-well format.

3. Heat inactivate plasma at 56 °C for 1 hour.

4. Proceed to assay or store at 4 °C.

Assay Reagent Preparation

BLI-ISA buffer

Materials:

- Phosphate buffered saline (PBS) tablets (Sigma P4417)
- Bovine serum albumin (BSA) (Fisher BP1600)
- Tween-20 (Fisher BP337)

Prepare PBS as directed on tablet package using MilliQ or equivalent purity water. Dissolve 2 grams of BSA in PBS for total combined volume of 100 mL. Add 100 uL Tween-20. Scale as appropriate, 0.22-micron filter after mixing.

25-CB buffer

Materials:

- ChonBlock (Chondrex 9068)
- BLI-ISA buffer (above)

Mix BLI-ISA buffer and ChonBlock in 3:1 ratio (e.g. 75 mL BLI-ISA and 25 mL ChonBlock).

SARS-CoV-2 RBD antigen

Materials:

- 500ug/mL Biotinylated SARS-CoV-2 spike RBD His AviTag (produced as above or acquired from ACROBiosystems SPD-C82E9)
- BLI-ISA buffer (above)

Dilute stock biotinylated SARS-CoV-2 spike RBD to 5 ug/mL. For new RBD lots, a dilution series should be performed in the 2-16 ug/mL range to determine the loading concentration that results in adequate loading onto the biosensors.

IgG detection reagent

Materials:

- 4 nm Colloidal Gold- AffiPure Goat Anti-Human IgG Fc gamma fragment specific (Jackson ImmunoResearch 109-185-098) rehydrated in 1 mL deionized water per the manufacturer's instructions
- BLI-ISA buffer (above)

Dilute Anti-Human IgG reagent 1:9 in BLI-ISA buffer. Eg 40 uL anti-Human IgG in 360 uL BLI-ISA buffer.

Diluted Plasma samples

Materials:

- $>$ 6 uL plasma specimen
- 25-CB buffer

Dilute plasma 1:7 in 25-CB buffer (e.g. 6 uL plasma in 42 uL 25-CB).

Preparing the Assay plate

Reagents and samples are plated in a tilted bottom (TW384) microplate (Sartorius ForteBio 18-5080). Due to the arrangement of the biosensor read head on the Octed RED384 instrument, samples are most easily run in sets of 8-16 with up to 64 total specimens per plate. Each sample requires 6 wells of reagents (antigen well, sample well, detection reagent well, and 3 blocking wells), corresponding to the steps of the COV2-BLI-ISA assay, and must be arranged in a grid-like spacing such that the Octet RED384 biosensor head can access each well in sequence (see figure 1.9 for one example)

For each sample to be tested, fill its 6 columns of reagent wells with 40 uL of the appropriate reagent (Figure 1.9). Column 1 (BLI-ISA buffer), column 2 (SARS-CoV-2 RBD antigen), column 3 (25-CB buffer), column 4 (diluted plasma samples), column 5 (BLI-ISA buffer), and column 6 (IgG detection reagent). Again, because the Octet RED384 BLI instrument read head has 8 channels in 96-well plate spacing, samples are tracked most easily by filling the assay plate in a offset grid pattern. This filling process is facilitated by storing and/or diluting plasma

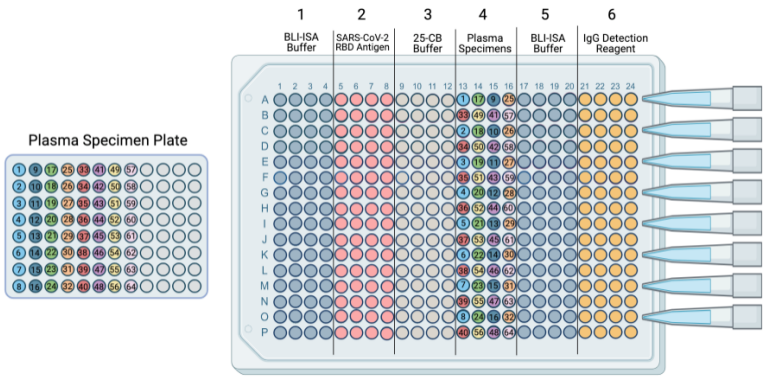


Figure 1.9 COV2-BLI-ISA assay plate set up. There are 6 steps to the assay, which are separated into columns here. During the assay, biosensors in the instrument move across the plate in sets of 8 or 16 spaced in an every-other-well grid. Well color marks the assay wells for a set of 8 sensor channels (which record the data for each specimen).

specimens in 96- or 384-well plates and transferring by multi-channel pipette or liquid handling robot.

Performing the assay

Load the Octet RED384 BLI biosensor tray with an appropriate amount of Octet Streptavidin (SA) sensor tips (Sartorius ForteBio 18-5019), up to 64 per 384 well assay plate, and place inside the instrument over a 96-well microplate (Greiner Bio-One 655209) containing 200 uL BLI-ISA buffer in each well.

Place Assay Plate in the Octet instrument.

Open the Octet Data Acquisition software and load the COV2-BLI-ISA.emf assay method file, optionally entering identifiers for the plasma samples (identifiers can also be assigned by well position after the assay). If setting up the COV2-BLI-ISA method manually (e.g. for less than 64 samples), the method can be programed as follows. Baseline1 (60 s) in column 1 (BLI-ISA buffer), Loading

(120 s) in column 2 (SARS-CoV-2 RBD antigen), Baseline2 (60 s) in column 3 (25-CB buffer), Association1 (600 s) in column 4 (plasma specimen), Baseline3 (60 s) in column 5 (BLI-ISA buffer), and Association2 (180 s) in column 6 (IgG detection reagent).

After the assay is complete, export the raw data file from the Octet HT Analysis software as a .csv file. Next, automated analysis of BLI-ISA data is performed with the bli_plotter.py program. From the command line, navigate to the directory where the raw data file and bli_plotter.py program is stored, then type:

> Python3 bli_plotter.py RAW-DATA-FILE.csv

A number of quality control plots will be generated in the assay directory along with an assay results table. Inspect the QC plots, especially those of controls, and ensure they meet expected assay performance. Next assay results can be read from the table, where detection scores above 0.31 nm correspond to samples with IgG reactive to the SARS-CoV-2 spike RBD.

1.2.5 Methods Appendix

Antigen production

Expressing SARS-CoV-2 RBD antigen:
The expression plasmid for SARS-CoV-2 RBD is a derivative of pcDNA 3.1 containing the SARS-CoV-2 spike RBD (residues 319–541 of Wuhan-Hu-1 isolate, GenBank: MN908947.3) fused to a C-terminal 10 histidine tag and Avi tag with GS linkers (RBD-GS-10H-GS-Avi tag, see full sequence below).

High quality, sterile, expression plasmid DNA is produced using the QIAGEN Plasmid Maxi Kit (QIAGEN Cat. No. 12163) or equivalent plasmid purification platform.

HEK 293F Freestyle (ThermoFisher R79007) cells are electroporated in the presence of the RBD expression plasmid using a MaxCyte electroporator with OC-400 electroporation cuvettes (MaxCyte SOC4) following the manufacturer's instructions, enabling transient expression of the tagged SARS-CoV-2 Spike RBD. After transfection, cells were grown with Freestyle 293 Expression media (Gibco 12338-018) in shake flasks using a Khuner shaker incubator at 37 °C, 8% CO2, and 85% humidity.

Four days after electroporation (or when cell viability drops below 60%), media and cells are centrifuged at 6000 g for 10 minutes and the supernatant decanted and 0.22 micron filtered. If purification is not scheduled immediately, add protease inhibitor and store at 4°C.

Prepare His-tag Wash buffer (20 mM MOPS, 150 mM NaCl, 20 mM imidazole, pH 7.5) and His-tag Elution buffer (20 mM MOPS, 150 mM NaCl, 200 mM imidazole, pH 7.5). Add an appropriate volume of Elution buffer to the filtered 293F supernatant to bring total imidazole concentration to 10 mM, then verify pH is 7.5. To a HisTrap FF column (Cytvia 17531901) equilibrated with His-tag Wash buffer, load the supernatant, wash with 10 column volumes His-tag Wash buffer, elute with 8 column volumes His-tag Elution buffer. Combine and concentrate eluent with high 280nm absorbance.

After running an SDS-PAGE gel to ensure relative purity, the combined and concentrated fractions are dialyzed against 20 mM MOPS, 150 mM NaCl pH 7.5 then biotinylated with GST-BirA. The initial biotinylation reaction consists of 50 uM SARS-CoV-2 spike RBD, 3 uM GST-BirA biotin ligase in PBS containing 5 mM MgCl2, 25 mM ATP, and 625 uM D-biotin and is incubated at room temperature overnight with gentle mixing. After diluting 1:1 with His-tag Wash buffer (above), biotinylated RBD is purified as above with a 1mL HisTrap FF column. Eluted protein is dialyzed to reduce imidazole concentration.

Before freezing, biotinylated RBD loading to streptavidin biosensors is assessed. SARS-CoV-2 spike RBD biotin was diluted into BLI-ISA buffer in the range of 2-10 ug/mL and extent of loading onto Octet Streptavidin (SA) biosensors is measured. To assess loading, SA tips are first equilibrated to BLI-ISA buffer for 10 minutes, then a baseline is recorded for 60 seconds, followed by loading measurement over 120 seconds. The target loading measurement is Δ 2 nm +/- 0.2 nm. Biotinylated spike RBD should be frozen in aliquots that allow 1:50 – 1:100 dilution in BLI-ISA buffer to reach a concentration that leads to satisfactory antigen loading.

SARS-CoV-2 Spike RBD antigen FASTA record

```
> SARS-CoV-2_Spike_RBD_10H-avi
MFMPSSFSYSSWATCWLLCCLIILAKATMFVFLVLLPLVSSQRVQPTESIVRFPNITNLCPFGEVFN
ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQI
APGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGS
```

TPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFGSH

HHHHHHHHHGSGLNDIFEAQKIEWHE*

1.2.6 Control Material(s) to be Used with COV2-BLI-ISA

Controls that are required but not provided with the test kit include an external positive control for the IgG antibody class claimed is needed to both ensure antigen is correctly folded and that the IgG detection reagent is responding in the expected range and should be used with each batch of antigen and/or anti-IgG reagent prepared from thawed stocks. Two dilutions of the CR3022 antibody (Abcam ab273073) near the expected minimum (250 ng/mL) and maximum (1000 ng/mL) detectable concentrations should be run to ensure correct assay performance. Additionally, no-plasma control experiments should also be run to ensure the BLI-ISA and 25-CB buffers are not causing spurious non-specific binding that could result in incorrect assay interpretation.

1.2.7 Interpretation of Results

All test controls should be examined prior to interpretation of patient results. If the controls are not valid, the patient results cannot be interpreted.

The COV2-BLI-ISA cutoff value for a positive result is a BLI-ISA score greater than 32 (Δ0.32 nm). Test results between 17 and 32 are indeterminate and should be examined by a secondary assay or another sample should be tested at a later date.

It is not currently known which COV2-BLI-ISA score correlates with effective SARS-CoV-2 neutralization.

Table 1.2. Interpretation of Results of a the COV2-BLI-ISA assay

BLI-ISA score (Δnm * 100)	Result	Test Result Interpretation
x < 17	Negative	IgG Antibodies for SARS-Cov-2 are not detected.
17 < x < 32	Indeterminate/ Equivocal	Antibodies determination is indeterminate/equivocal with this sample. Test another sample two weeks later
x> 32	Positive	IgG Antibodies for SARS-Cov-2 are detected.

Table 1.2 Interpretation of Results of a the COV2-BLI-ISA assay

Components required but not included with the test

Equipment:

Octet RED384 Biolayer Interferometry instrument

Software:

Octet Data Acquisition Software V11.x or greater

Materials and reagents:

Octet Streptavidin (SA) sensor tips (Sartorius ForteBio 18-5019)

Tilted bottom (TW384) microplates (Sartorius ForteBio 18-5080)

4 nm Colloidal Gold- AffiPure Goat Anti-Human IgG Fcg fragment specific (Jackson ImmunoResearch 109-185-098)

Phosphate buffered saline (PBS) tablets (Sigma P4417)

Bovine serum albumin (BSA) (Fisher BP1600)

Tween-20 (Fisher BP337)

ChonBlock (Chondrex 9068)

The COV2-BLI-ISA has been validated using only the components referenced above. The COV2-BLI-ISA was developed using the SARS-CoV-2 spike RBD (residues 319–541 of the Wuhan-Hu-1 isolate, GenBank: MN908947.3) manufactured in house in Freestyle HEK293 cells and an anti-human IgG-gold monoclonal goat antibody (Jackson ImmunoResearch 109-185-098) that target the Fc-gamma fragment of human IgG.

1.2.8 Manufacturing and Testing Capabilities

Approximate assay time, assuming all reagents are prepped in advance of specimen receipt or during specimen processing:

- 20 minutes – spinning venipuncture draw and transferring plasma to new tube/well plate
- 65 minutes – heat inactivation of plasma
- 20 minutes – prepping assay plate
- 20 minutes – instrument time/assay data collection (20 minutes per 16 samples with Octet RED384, typical to run four 16 sample batches in one plate 80 minutes)
- 20 minutes – data processing and validation

Total time: 2 hr 25 minutes including plasma heat inactivation time. This is a conservative estimate assuming pipetting is done by hand rather than by robot. Assay prep times could be cut significantly by automation and throughput could be increased further by using the 96-channel Octet instrument instead of the 16-channel model.

1.2.9 Performance Evaluation

Plasma Sample Sources and History

Known seronegative plasma was acquired from several sources: the Lyme Disease Biobank (21 samples), Discovery Life Sciences (73 samples), and SeraCare (1 sample). The Lyme Disease Biobank plasma samples were collected as part of an IRB approved study for the development of Lyme disease and diagnostic tests. The 21 samples used here were all collected from people living in Long Island, NY, USA. The Discovery Life Sciences (DLS) samples were all collected prior to December 2019, with the majority (73) collected at their Normal Donor clinic in Huntsville, AL, USA. The DLS samples are split 36/37 male/female and cover ages 19-78. Because all of the pre-pandemic samples used in the study are from the United States where HBV vaccination is prevalent and exposure to influenza, alpha- and beta-coronaviruses, and RSV is common, these samples' results are believed to be adequate to address cross reactivity concerns.

Incidence of HIV and HCV is rarer and not vaccinated against currently. To address this, we acquired 5 plasma samples from patients that were exposed to or currently infected with each disease (determined by antibody and/or PCR test).

Plasma reactive to SARS-CoV-2 was obtained from AllCells (Alameda, CA, USA) and SeraCare (Milford, MA). To be eligible for plasma donation, prospective donors must have either had COVID-19 symptoms resolved at least 28 days after the diagnosis or suspicion of COVID-19 (i.e. no fever, cough, difficulty breathing, etc.) or been 14 days symptom free with a follow up negative nasopharyngeal/PCR test. Each AllCells sample (30 in total) were determined SARS-CoV-2 reactive by an Anti-SARS-CoV-2 Total test (Ortho Clinical Diagnostics) and was given a CoV2T score (ranging from 8 to 440).

A complete list of all samples used along with metadata (age, gender, etc, where available) is provided in tabular form at the end of this document (see tables 6-10).

Analytical Sensitivity and Specificity

a. Reactivity/Inclusivity:

Plasma from people infected with variants of the SARS-CoV-2 virus have not been intentionally evaluated at this time in the COV2-BLI-ISA test. Because the COV2-BLI-ISA test uses the spike RBD to evaluate reactivity, variants with mutations in the RBD or potentially other parts of the spike protein could affect test accuracy if these mutations cause significant conformational changes or otherwise alter important epitope features.

As convalescent plasma from patients that were PCR positive for prevalent

variants of concern like B.1.1.7 becomes available, the performance of the COV2-BLI-ISA test should be evaluated.

b. Cross-Reactivity

The assay was evaluated for potential cross reactivity in specimens with other virus-reactive antibodies. As outlined above, HIV and HCV reactive antibodies were less likely to be present in our U.S. sourced pre-pandemic plasma, so known-reactive plasma was assessed for these viruses. No false positives were observed with the potential cross-reactants listed in the table below.

Clinical Category	Source/ Sample type	Number tested	Number positive with COV2-BLI-ISA
HIV	Discovery Life Sciences, plasma tested positive with Abbot Prism assay and HIV immunoblot	5	0
HCV	Discovery Life Sciences, plasma tested positive with Roche cobas or Abbott Architect l2000	5	0
Total		10	0

Table 1.3 Cross Reactivity Testing

Class Specificity:

The COV2-BLI-ISA antibody test specifically detects IgG antibodies using a goat anti-human Fc antibody conjugated to colloidal gold nanoparticles (Jackson ImmunoResearch 109-185-098). The manufacturer reports that the anti IgG Fc antibodies were purified by immunoaffinity chromatography using IgG Fc as ligand. They also report that the antibody reacts with the Fc portion of IgG heavy chain,

but not Fab by ELISA assay. Reportedly, the antibody does not bind to human IgM, IgA, or other non-immunoglobulin serum proteins.

Clinical Agreement Study:

The COV2-BLI-ISA test was evaluated using a panel of 30 frozen plasma samples reactive to SARS-CoV-2 and 105 frozen plasma samples collected prior to December 2019 (including 5 HIV and 5 HCV reactive plasma samples). The COV2-BLI-ISA test measures the presence of anti SARS-CoV-2 spike RBD IgG antibodies via a change in reflected light (Δnm) in a biolayer interferometry assay after a biosensor-RBD-plasma complex is exposed to an anti-Human-IgG-Fc antibody gold conjugate. To establish a suitable Δnm value to differentiate seronegative plasma from seropositive plasma, a statistical model was made using the assay results from a subset of the pre-pandemic samples. The model samples were selected at random from the 95 pre-December 2019 plasma not known to have HIV or HCV by assigning each sample an index number and generating 22 random numbers between 1 and 95 without replacement using the Python numpy function random.choice(). The seronegative/equivocal cutoff was set as the upper end of the 95% confidence interval of the model group's mean plus five standard deviations. Preliminary studies during the development of the COV2-BLI-ISA led to a belief that there may be a small subset of individuals with background / cross-reactive binding to SARS-CoV-2 RBD antigen complex despite lack of SARS-CoV-2 exposure. In light of this, a more stringent cut-off (model mean plus ten standard deviations) was set for the equivocal/seropositive line to minimize false positive rate.

The results of the clinical agreement study are summarized in the following tables. All SARS-CoV-2 reactive samples were collected more than 14 days after symptom end, so the accuracy of the COV2-BLI-ISA test is not known for the acute or early convalescent period of SARS-CoV-2 infection.

	Comparator Method (Antibody Positive) >14 days post symptom onset	Collected Pre-Dec 2019 (Antibody Negative)			
COV2-BLI-ISA	**IgG +**	**Negative**	**HIV+**	**HCV+**	**Total**
IgG +	29	0	0	0	29
Equivocal	1	1	1	0	3
IgG -	1	74	4	5	84
Total	31	75	5	5	118

Table 1.4 Summary of Clinical Agreement Study Results

Measure	Estimate	Confidence Interval[a]
IgG Sensitivity (PPA)	**93.9% (31/33)**	**(86%; 100%)**
IgG Specificity (NPA)	**98.7% (74/75)**	**(96.1%; 100%)**
NPA including HIV, HCV	**98% (83/85)**	**(94%; 100%)**
Cross-reactivity with HIV+	**0% (0/5)**	
Cross-reactivity with HCV+	**0% (0/5)**	

a. $CI = p \pm Z\sqrt{\frac{pq}{n}}$ $where\ p = \frac{x\ (number\ of\ successes)}{n\ (number\ of\ samples)}$, $q = 1 - p, and\ Z = 1.96\ for\ 95\%\ confidence\ interval$

Table 1.5 Clinical Agreement Study Statistics

Clinical Agreement and Cross Reactivity Line Data

In this section all data from the clinical agreement and cross reactivity study are shown. For this study all samples were run singly. The data are summarized graphically in figure 1.10 while line data for each individual sample is listed in table 1.6.

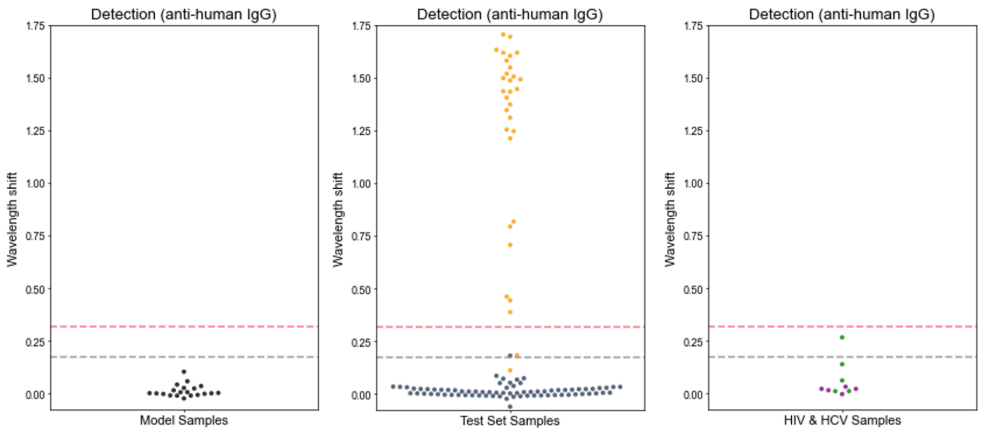


Figure 1.10 Graphical summary of clinical agreement and cross reactivity assay data for the COV2-BLI-ISA test. The raw Dnm score of each of the model samples (left), test set samples (center), and cross-reactivity set samples (right) are plotted as points, with the grey dotted line indicating the equivocal cut-off and the pink line indicating the SARS-CoV-2 reactive cut-off score. In the Test Set plot, dark blue point are samples collected pre December 2019 and gold points were determined SARS-CoV-2 reactive by the COV2T assay. In the HIV and HCV plot, purple corresponds to HCV samples and green HIV.

Table 1.6: COV2-BLI-ISA Clinical Agreement Study Line Data

	Sample ID	BLI-ISA Score	Interpretation	Group (model or test)	COV2T Score
1	DLS-04	0.2	N/A	Model	
2	DLS-05	-0.2	N/A	Model	
3	DLS-11	10.3	N/A	Model	
4	DLS-13	4.2	N/A	Model	
5	DLS-23	-0.9	N/A	Model	
6	DLS-25	0.6	N/A	Model	
7	DLS-28	-0.7	N/A	Model	
8	DLS-31	5.8	N/A	Model	
9	DLS-36	-1	N/A	Model	
10	DLS-38	0	N/A	Model	
11	DLS-49	1.5	N/A	Model	
12	DLS-55	2.6	N/A	Model	
13	DLS-60	3.5	N/A	Model	
14	DLS-65	-0.2	N/A	Model	
15	DLS-68	0.6	N/A	Model	
16	DLS-73	-1	N/A68	Model	

Continued on next page

Table 1.6: COV2-BLI-ISA Clinical Agreement Study Line Data (Continued)

	Sample ID	BLI-ISA Score	Interpretation	Group (model or test)	COV2T Score
17	LDBB-629	-2.4	N/A	Model	
18	LDBB-658	0.1	N/A	Model	
19	LDBB-664	0.3	N/A	Model	
20	LDBB-674	2.4	N/A	Model	
21	DLS-01	7.4	Negative	Test (negative)	
22	DLS-02	2	Negative	Test (negative)	
23	DLS-03	0.2	Negative	Test (negative)	
24	DLS-06	-0.9	Negative	Test (negative)	
25	DLS-07	2.3	Negative	Test (negative)	
26	DLS-08	1.2	Negative	Test (negative)	
27	DLS-09	-0.8	Negative	Test (negative)	
28	DLS-10	7.2	Negative	Test (negative)	
29	DLS-12	0.1	Negative	Test (negative)	
30	DLS-14	0.8	Negative	Test (negative)	
31	DLS-15	-0.6	Negative	Test (negative)	
32	DLS-16	-0.4	Negative	Test (negative)	
33	DLS-17	5.1	Negative	Test (negative)	
34	DLS-18	-0.5	Negative	Test (negative)	
35	DLS-19	2.9	Negative	Test (negative)	
36	DLS-20	8.6	Negative	Test (negative)	
37	DLS-21	0.5	Negative	Test (negative)	
38	DLS-22	3.4	Negative	Test (negative)	
39	DLS-24	3.3	Negative	Test (negative)	
40	DLS-26	2.2	Negative	Test (negative)	
41	DLS-27	6.8	Negative	Test (negative)	
42	DLS-29	2.2	Negative	Test (negative)	
43	DLS-30	1.7	Negative	Test (negative)	
44	DLS-32	5.1	Negative	Test (negative)	
45	DLS-33	1.9	Negative	Test (negative)	
46	DLS-34	-0.4	Negative	Test (negative)	
47	DLS-35	0.4	Negative	Test (negative)	
48	DLS-37	5.3	Negative	Test (negative)	
49	DLS-39	1.8	Negative	Test (negative)	

Continued on next page

Table 1.6: COV2-BLI-ISA Clinical Agreement Study Line Data (Continued)

	Sample ID	BLI-ISA Score	Interpretation	Group (model or test)	COV2T Score
50	DLS-40	1.1	Negative	Test (negative)	
51	DLS-41	3.7	Negative	Test (negative)	
52	DLS-42	1.6	Negative	Test (negative)	
53	DLS-43	-0.5	Negative	Test (negative)	
54	DLS-44	-1.2	Negative	Test (negative)	
55	DLS-45	-0.1	Negative	Test (negative)	
56	DLS-46	0.1	Negative	Test (negative)	
57	DLS-47	-0.8	Negative	Test (negative)	
58	DLS-48	2.3	Negative	Test (negative)	
59	DLS-50	2.7	Negative	Test (negative)	
60	DLS-51	0.6	Negative	Test (negative)	
61	DLS-52	0.5	Negative	Test (negative)	
62	DLS-53	0.3	Negative	Test (negative)	
63	DLS-54	-0.8	Negative	Test (negative)	
64	DLS-56	-1	Negative	Test (negative)	
65	DLS-57	1.1	Negative	Test (negative)	
66	DLS-58	3.1	Negative	Test (negative)	
67	DLS-59	-0.5	Negative	Test (negative)	
68	DLS-61	-0.2	Negative	Test (negative)	
69	DLS-62	3.3	Negative	Test (negative)	
70	DLS-63	-0.7	Negative	Test (negative)	
71	DLS-64	-1.1	Negative	Test (negative)	
72	DLS-66	18.1	Equivocal	Test (negative)	
73	DLS-67	2.5	Negative	Test (negative)	
74	DLS-69	-0.2	Negative	Test (negative)	
75	DLS-70	2.3	Negative	Test (negative)	
76	DLS-71	0.8	Negative	Test (negative)	
77	DLS-72	0.5	Negative	Test (negative)	
78	LDBB-515	1.8	Negative	Test (negative)	
79	LDBB-523	0.2	Negative	Test (negative)	
80	LDBB-526	2.9	Negative	Test (negative)	
81	LDBB-538	1.8	Negative	Test (negative)	
82	LDBB-540	0.2	Negative	Test (negative)	

Continued on next page

Table 1.6: COV2-BLI-ISA Clinical Agreement Study Line Data (Continued)

	Sample ID	BLI-ISA Score	Interpretation	Group (model or test)	COV2T Score
83	LDBB-561	-2.3	Negative	Test (negative)	
84	LDBB-587	2	Negative	Test (negative)	
85	LDBB-610	-0.7	Negative	Test (negative)	
86	LDBB-611	-0.6	Negative	Test (negative)	
87	LDBB-640	-0.8	Negative	Test (negative)	
88	LDBB-657	3.3	Negative	Test (negative)	
89	LDBB-659	0.9	Negative	Test (negative)	
90	LDBB-662	-1.4	Negative	Test (negative)	
91	LDBB-663	-6.1	Negative	Test (negative)	
92	LDBB-668	1.2	Negative	Test (negative)	
93	LDBB-677	0.9	Negative	Test (negative)	
94	LDBB-688	-5	Negative	Test (negative)	
95	SeraCare -	-0.1	Negative	Test (negative)	
96	AC01	151.8	Positive	Test (positive)	222
97	AC02	137.2	Positive	Test (positive)	153
98	AC03	18.3	Equivocal	Test (positive)	8
99	AC04	121.1	Positive	Test (positive)	42.8
100	AC05	158	Positive	Test (positive)	121
101	AC06	149.1	Positive	Test (positive)	103
102	AC07	169.3	Positive	Test (positive)	72.2
103	AC08	79.4	Positive	Test (positive)	53.9
104	AC09	144.5	Positive	Test (positive)	282
105	AC11	11.2	Negative	Test (positive)	5.5
106	AC12	38.7	Positive	Test (positive)	12.9
107	AC14	143.3	Positive	Test (positive)	247
108	AC15	131	Positive	Test (positive)	11.9
109	AC16	161.8	Positive	Test (positive)	19.8
110	AC17	150.4	Positive	Test (positive)	17.8
111	AC18	125.3	Positive	Test (positive)	9.8
112	AC19	170.4	Positive	Test (positive)	18.4
113	AC20	148.6	Positive	Test (positive)	14.4
114	AC21	140.4	Positive	Test (positive)	17
115	AC22	154.7	Positive	Test (positive)	23

Continued on next page

Table 1.6: COV2-BLI-ISA Clinical Agreement Study Line Data (Continued)

	Sample ID	BLI-ISA Score	Interpretation	Group (model or test)	COV2T Score
116	AC23	160.3	Positive	Test (positive)	16
117	AC24	161.7	Positive	Test (positive)	22.1
118	AC25	70.6	Positive	Test (positive)	9.1
119	AC26	44.4	Positive	Test (positive)	1.2
120	AC27	134.5	Positive	Test (positive)	10.1
121	AC28	46.1	Positive	Test (positive)	6.4
122	AC29	163.1	Positive	Test (positive)	20.2
123	AC30	124.5	Positive	Test (positive)	13.5
124	AC31	143.5	Positive	Test (positive)	26.6
125	AC32	149.7	Positive	Test (positive)	20
126	SeraCare +	81.7	Positive	Test (positive)	N/A
127	DHCV-1	2.2	Negative	Test(HCV)	
128	DHCV-2	3.2	Negative	Test(HCV)	
129	DHCV-3	-0.3	Negative	Test(HCV)	
130	DHCV-4	2.2	Negative	Test(HCV)	
131	DHCV-5	1.5	Negative	Test(HCV)	
132	DHIV-1	26.6	Equivocal	Test(HIV)	
133	DHIV-2	1.1	Negative	Test(HIV)	
134	DHIV-3	1.1	Negative	Test(HIV)	
135	DHIV-4	6.2	Negative	Test(HIV)	
136	DHIV-5	13.9	Negative	Test(HIV)	

[1] COV2T score evaluated using SARS-CoV-2 Total antibody assay by AllCells (Alameda, CA, USA).

1.2.10 Discussion of the Application and its Review

The intent of this EUA application was to enable serology studies by ourselves and collaborators as well as provide a new tool to perform antibody tests at e.g. UC clinics. Unfortunately, the relatively long wait for feedback on the application precluded pursuit of some of the studies we wished to perform. When the FDA did respond to our application, a number of minor to significant issues were

brought to our attention that would need to be addressed to get approval to use the BLI-ISA assay for SARS-CoV-2 antibody detection. Although some of the complaints concerned the safety and security of the Octet BLI instrument itself and could not be solved without the manufacturer's input, most of the issues were solvable technical problems. Given the solid performance of the BLI-ISA assay that we observed, it is likely we could have achieved EUA approval with a bit more experience interacting with the FDA and faster feedback.

At the time of application submission, the author's primary concerns were whether our positive control samples were appropriate and whether the manufacturer specified properties of certain reagents (i.e. specificity of the anti-IgG secondary) would be accepted without those properties being explicitly tested in our experiments. Our seropositive control samples were tested by another antibody test with an approved EUA – at the time there were no SARS-CoV-2 antibody tests with anything other than emergency use authorization. Although the demands for EUA approval are high (>90% correct positives, >95% correct negative classification), some error in these tests is acceptable, so error in our test's accuracy would be compounded by that of the intermediate test. With respect to reagents and the instrument, the author's concerns stemmed from the often overlooked 'for laboratory use only' printed on the datasheets for most of our supplies. Each of these concerns were borne out by the response of the FDA.

Inappropriate clinical agreement study design. The most damning complaint we received in our feedback was regarding our sample selection. Rather than using samples tested by another antibody test, the FDA required each sample be

drawn from a person with who tested positive for SARS-CoV-2 infection by a PCR test. It was also important to supply records of where and how those samples were taken and tested in addition to patient infection timelines (i.e. symptom status, dates of PCR tests vs blood draws, etc). We used a panel of commercial serum samples for our positive controls where, to the author's knowledge, these data were not available. Were the project being conducted again, it would have been great to pair up with the lab on campus performing PCR tests to find positive-testing volunteers willing to donate a finger-prick of blood to our project. This would have allowed collection of all the data requested by the FDA and likely led to a larger sample pool, increasing the power of clinical agreement studies we would have performed.

Inadequate use of a 'Reseach Use Only' device. Another significant series of issues the FDA had with our test related to compliance of the Octet BLI instrument itself. Because the Octet was not already approved for diagnostic use, we needed to establish the reproducibility of the data acquisition software, its electrical safety and potential risks to operators of the instrument, extent of electromagnetic interference with other instruments, and the device's 'cyber-security profile'. Although we could not conduct the testing needed to characterize these properties of the BLI instrument, perhaps we could have worked with ForteBio to encourage diagnostic approval. Approval of the Octet for diagnostic assays could open up additional markets for the instrument via sales to hospitals and testing centers, but without deeper market research it is unclear whether they could justify the outlays required to perform these tests for the FDA.

Along similar lines, our reviewer did not feel the IgG detection reagent used was demonstrated to have the immunoglobulin specificity claimed. Although we used a commercial reagent marketed as IgG specific, because that reagent was not registered by the FDA for that purpose we needed to perform experiments to show that class specificity. This we could have addressed by testing the cross-class-reactivity of the detection reagent with a panel of immunoglobulin classes. A potential challenge here is whether that specificity testing needs to be performed with a BLI assay in a similar configuration to the actual BLI-ISA assay or whether orthogonal tests of reagent specificity would be accepted. Either way, we surely could have performed this testing to the satisfaction of the FDA.

Other areas for improvement. The remaining feedback we received on the BLI-ISA EUA application were items that could have been quickly remedied with a faster feedback cycle or a bit more experience on our part. For example, because the BLI-ISA assay used a biotin-streptavidin interaction to mount SARS-CoV-2 antigen to the sensor tips and biotin is present in varying concentrations in serum, our reviewer requested experimental evidence that biotin would not interfere with our measured signal. Although this is a trivial experiment to perform (i.e. expose antigen loaded tips to a dilution series of biotin), a clearer description of the assay mechanism could have assuaged the reviewers concern's too (any signal from biotin would occur during the 'total antibody binding' step and could not contribute to signal measured during the secondary IgG detection step). Our reviewer also requested data demonstrating the shelf life of the reagents in the assay, in particular the long term stability of the positive control antibody CR3022 when frozen, reminding us "note that the EUA applicant takes regulatory responsibility for the external controls even if the controls are sourced from a third party."

Overall, while the study design was flawed due to use of an improper positive comparator test (another antibody test rather than PCR comparator) and the lack of diagnostic approval for the BLI instrument, it seems likely that the BLI-ISA assay could meet regulatory requirements with additional resource investment. If we assume the COV2T test we used as a comparator method adequately identified true positives, the performance of the BLI-ISA antibody test was excellent with >93% positive percent agreement and >98% negative percent agreement with the controls selected. Clearly, though, for a medical diagnostic test every duck should be in a row and a number of ours were askew.

1.3 Human Astrovirus 1-8 Seroprevalence Evaluation in a United States Adult Population

1.3.1 Preface

After completion of the EUA application to use our BLI-ISA method as an antibody test for SARS-CoV-2 antibody detection, we thought to re-use the samples and assay design to do a serological study looking at the seroprevalence of antibodies to the 8 canonical human astrovirus strains. Lena and Kevin had expressed each of the 8 strain's spike proteins as part of other projects, while I had curated a set of serum samples and developed a relatively high-throughput assay that could be performed with the BLI instrument. We were able to combine the two resources, modifying the SARS-CoV-2 assay to use the astrovirus spikes as the antigen probe, allowing a fairly quick study of astrovirus seroprevalence. We had to be more creative with establishing our assay baseline and outcomes as astrovirus is endemic worldwide and we couldn't simply use 'pre-astrovirus' serum as a control like we could do with SARS-CoV-2. In the end, after some troubleshooting, BLI again proved to be a great tool in the serology arsenal. We published the results of our human astrovirus seroprevalence assays as a paper in the journal Viruses. Below is a reproduction of that research paper, which is available online at https://doi.org/10.3390/v13060979.

Citation

Human Astrovirus 1–8 Seroprevalence Evaluation in a United States Adult Population.

Meyer, L.; Delgado-Cunningham, K.; Lorig-Roach, N.; Ford, J.; DuBois, R. M. Viruses 2021, 13 (6), 979.

Author contributions

Conceptualization, R.M.D.; Methodology, L.M., K.D.-C., N.L.-R. and J.F.; Data Analysis, L.M. and N.L.-R.; Writing: Original Draft Preparation, L.M.; Writing: Review and Editing, L.M. and R.M.D.; Visualization, L.M., N.L.-R., K.D.-C. and R.M.D.; Supervision, R.M.D. All authors have read and agreed to the published version of the manuscript.

1.3.2 Abstract

Human astroviruses are an important cause of viral gastroenteritis globally, yet few studies have investigated the serostatus of adults to establish rates of previous infection. Here, we applied biolayer interferometry immunosorbent assay (BLI-ISA), a recently developed serosurveillance technique, to measure the presence of blood plasma IgG antibodies directed towards the human astrovirus capsid spikes from serotypes 1–8 in a cross-sectional sample of a United States adult population. The seroprevalence rates of IgG antibodies were 73% for human astrovirus serotype 1, 62% for serotype 3, 52% for serotype 4, 29% for serotype 5, 27% for serotype 8, 22% for serotype 2, 8% for serotype 6, and 8% for serotype 7. Notably, seroprevalence rates for capsid spike antigens correlate with neutralizing antibody rates determined previously. This work is the first seroprevalence study evaluating all eight classical human astrovirus serotypes.

1.3.3 Introduction

Astroviruses are a diverse family of small, nonenveloped, positive-sense RNA viruses that infect mammalian and avian species [1]. Astrovirus infection is linked to a variety of disease manifestations, growth defects, and mortality in poultry [2]. In mammals, astrovirus infection mainly causes viral gastroenteritis but can also be asymptomatic [3,4] or cause neurological syndromes and encephalitis in rare cases [5,6,7,8]. Human astroviruses are classified into three clades: classical serotypes 1–8, where serotype 1 is the most prevalent globally [9,10,11], as well as the emerging serotypes MLB and VA [5,8].

Human astroviruses (HAstVs) are a leading worldwide cause of viral gastroenteritis but are among the most poorly characterized enteric viruses [12]. Young children, the elderly, and the immunocompromised are most threatened by astrovirus infection, especially in developing countries [13,14,15,16,17,18,19,20]. Worldwide, human astrovirus infection accounts for approximately 2 to 9% of all acute non-bacterial gastroenteritis in healthy children [21]. The United States sees an estimated 3.9 million cases of astrovirus gastroenteritis each year [22]. However, immunofluorescence studies have demonstrated that about 75% of healthy adults have anti-astrovirus antibodies targeting at least one of the eight classical serotypes [23], and another investigation showed that seroprevalence of neutralizing antibodies increases with age [10]. These findings suggest that cases of astrovirus gastroenteritis may be undercounted and that astrovirus disease may actually be a common childhood infection.

Several lines of evidence highlight the importance of anti-astrovirus antibodies

developed in childhood in preventing reinfection in adulthood. Firstly, astrovirus infection is rare in adults [23]. Although approximately 75% of adults have anti-astrovirus antibodies, clinical investigations in healthy volunteers determined that more severe disease is associated with seronegativity for anti-astrovirus antibodies [24,25]. Finally, immunoglobulin replacement therapy resolved a persistent human astrovirus infection in an immunocompromised patient [26]. These studies indicate that a vaccine and therapeutic antibodies could be developed to prevent and/or treat human astrovirus gastroenteritis.

Previous structural studies have defined the structural domains of the human astrovirus capsid (Figure 1.11A,B) [27,28,29,30]. The capsid core domain, which forms the structural icosahedral shell encapsulating the viral genome, remains highly conserved across all eight classical serotypes, ranging from 83.3–97.0% sequence identity between any two serotypes (Table 1.7). In contrast, the capsid spike domain, which forms dimeric protrusions on the surface of the viral particle, is quite variable, with a spread of 41.4–75.7% sequence identity between any two serotypes (Table 1.8). Prior work using enzyme-linked immunosorbent assays showed that both the core and spike domains are antigenic [28].

Serosurveillance investigations bring the frequency of previous human astrovirus infection in a particular cohort into focus, but only a few have been performed on specific astrovirus subsets (classical serotypes 1–7, MLB1, HMOAstV-C (VA1), or turkey astrovirus 2) or in specific populations (turkey growers and abattoir workers) (Table 1.9) [10,23,31,32,33,34,35,36]. Many of the studies focusing on classical human astrovirus were conducted more than 20 years ago. Importantly, no previous serological survey has evaluated seropositivity to all eight classical

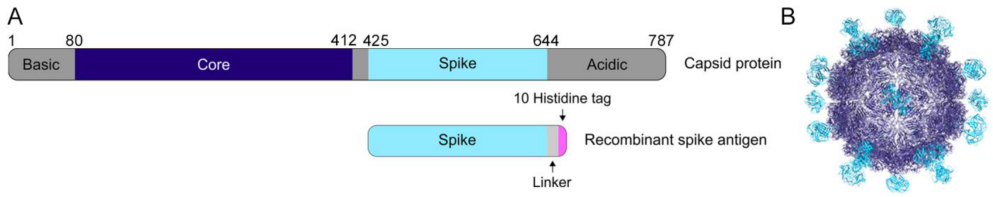


Figure 1.11 The human astrovirus capsid. (A) Schematic of the human astrovirus capsid protein and the recombinant spike antigen used in this study. (B) The mature human astrovirus particle, with the spike in cyan and the core in dark blue (adapted from [28]).

	Core 1	**Core 2**	**Core 3**	**Core 4**	**Core 5**	**Core 6**	**Core 7**	**Core 8**
Core 1	100 %	86. 7 %	91.5 %	86.1 %	90.0 %	90.9 %	87.9 %	86.7 %
Core 2	-	100 %	86.7 %	87.9 %	83.3 %	83.9 %	85.2 %	87.9 %
Core 3	-	-	100 %	87.3 %	90.6 %	91.2 %	92.8 %	88.2 %
Core 4	-	-	-	100 %	84.6 %	86.1 %	84.9 %	97.0 %
Core 5	-	-	-	-	100 %	90.3 %	88.8 %	85.8 %
Core 6	-	-	-	-	-	100 %	87.9 %	87.6 %
Core 7	-	-	-	-	-	-	100 %	85.8 %
Core 8	-	-	-	-	-	-	-	100 %

Table 1.7 Sequence identity matrix across capsid core domains from human astrovirus serotypes 1–8.

	Spike 1	**Spike 2**	**Spike 3**	**Spike 4**	**Spike 5**	**Spike 6**	**Spike 7**	**Spike 8**
Spike 1	100 %	50.2 %	61.2 %	41.4 %	47.0 %	53.1 %	58.9 %	53.0 %
Spike 2	-	100 %	46.7 %	47.9 %	41.8 %	47.0 %	44.9 %	49.8 %
Spike 3	-	-	100 %	46.3 %	48.6 %	54.7 %	75.7 %	55.1 %
Spike 4	-	-	-	100 %	46.0 %	47.0 %	43.5 %	46.1 %
Spike 5	-	-	-	-	100 %	63.34 %	52.8 %	52.1 %
Spike 6	-	-	-	-	-	100 %	56.6 %	57.3 %
Spike 7	-	-	-	-	-	-	100 %	53.3 %
Spike 8	-	-	-	-	-	-	-	100 %

Table 1.8 Sequence identity matrix across capsid spike domains from human astrovirus serotypes 1–8.

human astrovirus serotypes. In addition, these studies used a variety of human astrovirus antigens to determine seroprevalence, including mature human astrovirus particles, the full capsid protein, or capsid protein fragments (Table 1.9).

However, studies that evaluate reactivity to the whole capsid are limited because of the possibility of cross-reactive antibodies that may recognize the conserved core domains of several serotypes. The use of the more variable capsid spike domain as the antigen provides an opportunity to confidently evaluate serotype reactivity across all eight serotypes. Here, we provide an updated report on the seroprevalence of each of the eight classical serotypes in a United States adult population, as determined by biolayer interferometry immunosorbent assay (BLI-ISA) using recombinant human astrovirus 1–8 capsid spikes (Spike 1–Spike 8) as the antigens [37].

1.3.4 Materials and Methods

Human Astrovirus Capsid Sequences Used for Sequence Identity Studies

The accession numbers of sequences used for sequence alignments of the human astrovirus capsid core are as follows for serotypes 1 to 8 (Core- accession number: residues (depository)): Core 1- AAC34717.1: 80-410 (Genbank), Core 2- AZB52195.1: 80-409 (Genbank), Core 3- Q9WFZ0: 82-412 (Uniprot), Core 4- Q3ZN05: 80-410 (Uniprot), Core 5- Q4TWH7: 79-409 (Uniprot), Core 6- AZB52207.1: 80-410 (Genbank), Core 7- Q96818: 81-411 (Uniprot), and Core 8- Q9IFX1: 80-410 (Uniprot). The accession numbers of sequences used for sequence alignments of the human astrovirus capsid spike are as follows for serotypes 1 to 8 (Spike-accession number: residues (depository)): Spike 1-AAC34717.1: 430-644 (Genbank), Spike 2-KY964327.1: 12-228 (Genbank), Spike 3-Q9WFZ0: 432-645 (Uniprot), Spike 4-Q3ZN05: 430-644 (Uniprot), Spike 5-Q4TWH7: 429-641 (Uniprot), Spike 6-AZB52207.1: 430-642 (Genbank), Spike 7-Q96818: 431-644

Study Year Published	Seroprevalence	Size	Antigen
Koopmans et al. [10] 1998	Ages 0-79: Human astrovirus 1: 91 % Human astrovirus 2: 31 % Human astrovirus 3: 69 % Human astrovirus 4: 56 % Human astrovirus 5: 36 % Human astrovirus 6: 16 % Human astrovirus 7: 10 % Human astrovirus 8: ---	242 people ages 0-79	Human astrovirus particles cultured in Caco-2 cells
Mitchell et al. [31] 1999	By age 9: Human astrovirus 1: 94 % Human astrovirus 3: 42 %	393 children	Baculovirus-expressed recombinant capsid proteins
Kriston et al. [32] 1996	By age 5: Human astrovirus 1: 90 % Human astrovirus 6: 10-30 %	273 children and hospital staff	Baculovirus-expressed recombinant capsid proteins
Kobayashi et al. [33] 1999	By age 3: Human astrovirus 1: ~100 % Human astrovirus 2: ~100 %	170 children	Unclear: Article in Japanese
Kurtz et al. [23] 1978	Presence of "astrovirus antibody" By 6 months: 45 % By age 1: 7 % By age 2: 25 % By age 4: 71 % By age 10: 75 % Young adult cohort: 77 %	87 children ages 0-10 70 young adults ages 17-30	Human astrovirus particles cultured in HEK cells
Holtz et al. [34] 2014	By age 17: MLB1: 100 %	395 people ages 0-95	Baculovirus-expressed recombinant capsid protein
Burbelo et al. [35] 2011	HMOAstV-C (VA1): 36 % by age 10 65 % in adults	103 children and 106 adults	Crude protein extracts containing N- and C-terminal capsid fragments
Meliopoulos et al. [36] 2014	By adulthood: Human astrovirus 1: 81 % Turkey astrovirus 2: 26 %	160 turkey growers, turkey meat processing plant workers, and unexposed workers	Baculovirus-expressed recombinant capsid proteins

Table 1.9 Summary of previous human astrovirus serological surveys.

(Uniprot), and Spike 8- Q9IFX1: 431-644 (Uniprot).

Reagents and Supplies

Anti-Penta-His (HIS1K) sensor tips (Sartorius ForteBio, Goettingen, Germany, 18-5120), Benzonase (Merck Millipore, Burlington, MA, USA, 71205), bovine serum albumin (BSA) (Fisher, Hampton, NH, USA, BP1600), CD OptiCHO expression medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA, 12681-029), CHO CD EfficientFeed A (Gibco, Thermo Fisher Scientific, Waltham, MA, USA, A10234-01), ChonBlock (Chondrex, Woodinville, WA, USA, 9068), 4 nm Colloidal Gold-AffiniPure Goat Anti-Human IgG Fcγ fragment-specific (Jackson ImmunoResearch, West Grove, PA, USA, 109-185-170) rehydrated in 1 mL deionized water per the manufacturer's instructions, 4 nm Colloidal Gold-AffiniPure Goat Anti-Rabbit IgG (H + L) (Jackson ImmunoResearch, West Grove, PA, USA, 111-185-144) rehydrated in 1 mL deionized water per the manufacturer's instructions, glucose (Sigma Aldrich, St. Louis, MO, USA, G8769), 1 mL HiTrap Protein A HP column (GE Healthcare, Chicago, IL, USA, 17-0402-03), HT supplement (Gibco, Thermo Fisher Scientific, Waltham, MA, USA, 11067-030), L-glutamine (Gibco, Thermo Fisher Scientific, Waltham, MA, USA, 35050-061), phosphate buffered saline (PBS) tablets (Sigma Aldrich, St. Louis, MO, USA, P4417), Pierce IgG Elution Buffer (Thermo Fisher Scientific, Waltham, MA, USA, 21004), Pierce Normal Rabbit Serum Control (Thermo Fisher Scientific, Waltham, MA, USA, 31884), Pierce Protein A IgG Binding Buffer (Thermo Fisher Scientific, Waltham, MA, USA, 21001), Pluronic F68 (Gibco, Thermo Fisher Scientific, Waltham, MA, USA, 24040-032), protease inhibitor cocktail set V EDTA-free (MilliporeSigma, Burlington, Massachusetts, USA, 539137), sodium butyrate (Sigma-Aldrich B5887), Superdex 75 10/300 and 16/600 (Cytiva, Marlborough, MA, USA), TALON metal affinity resin (GE Healthcare, Chicago, IL, USA, 28-

9574-99), tilted-bottom (TW384) microplates (Sartorius ForteBio, Goettingen, Germany, 18-5080), Tween-20 (Fisher, Hampton, NH, USA, BP337), yeastolate (BD Biosciences, San Jose, CA, USA, 292804).

Expression and Purification of Human Astrovirus Spikes 1–8

Synthetic cDNA encoding Spike 1 (residues 429-645, accession AAC34717.1 (Genbank)), Spike 2 (residues 12-228, accession KY964327.1 (Genbank), a partial HAstV2 capsid sequence), Spike 3 (residues 431-647, accession Q9WFZ0 (Uniprot)), Spike 4 (residues 429-646, accession Q3ZN05 (Uniprot)), Spike 5 (residues 428-644, accession Q4TWH7 (Uniprot)), Spike 6 (residues 429-644, accession AZB52207.1 (Genbank)), Spike 7 (residues 430-646, accession Q96818 (Uniprot)), or Spike 8 (residues 429-647, accession Q9IFX1 (Uniprot)) were cloned separately into pET52b (EMD Millipore) in frame with a C-terminal linker, thrombin cleavage site, and 10-Histidine tag. The plasmids were verified by DNA sequencing. Spike plasmids were transformed into E. coli strain BL21 (DE3). Expression was induced with 1 mM IPTG and carried out at 18 °C for 16 h. Cell pellets were resuspended in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole (Buffer A) containing 2 mM MgCl2, 0.0125 U/μL benzonase, and 1x protease inhibitor cocktail set V EDTA-free. Cells were lysed using ultrasonication, and spikes were batch purified using TALON metal affinity resin. Elutions were carried out using Buffer A with 500 mM imidazole. Spikes were dialyzed into 10 mM Tris-HCl pH 8.0, 150 mM NaCl (TBS) with 1 mM DTT and purified using size exclusion chromatography in TBS with a Superdex 75 16/600 (Spikes 1, 2, 5, 7, and 8) or Superdex 75 10/300 column (Spikes 3, 4, and 6).

Expression and Purification of the SARS-CoV-2 RBD

The SARS-CoV-2 receptor-binding domain (RBD) in frame with a 10-Histidine tag and AviTag at the C-terminus was expressed and purified, as described in [37].

Expression and Purification of Recombinant mAb 3B4

Mouse hybridoma cells producing mAb 3B4 were generated, as reported in [38]. The amino acid sequences of the mAb 3B4 variable regions were identified as described in [39], allowing for recombinant antibody expression. Synthetic cDNA encoding the 3B4 kappa and heavy chain variable regions (Integrated DNA Technologies, Coralville, IA, USA) was cloned by Gibson assembly into the two pCMV-VRC01 antibody backbone vectors for light and for heavy chains, which contain the constant regions of VRC01, a human anti-HIV antibody targeting the gp120 protein [40]. Cloned sequences were in frame with an N-terminal secretion signal sequence. The resulting expression plasmids, pCMV-VRC01_3B4_kappa and pCMV-VRC01_3B4_heavy, contain the variable regions from the original mouse antibody 3B4 and the constant regions from a human IgG1 antibody under the control of the human cytomegalovirus promoter. The plasmids were verified by DNA sequencing. The expression plasmids were used in a 1:1 ratio to electroporate Chinese Hamster Ovary suspension (CHO-S) cells using the MaxCyte system. Recombinant mAb 3B4 was expressed for 8 days by CHO-S cells growing in CD OptiCHO expression medium supplemented with 1 mM sodium butyrate, 8 mM L-glutamine, 1× HT supplement, and 0.1% Pluronic F68 at 32 °C with 125 rpm shaking. Every 24 h, cells were fed with CHO CD EfficientFeed A supplemented with 7 mM L-glutamine, 5.5% glucose, and 23.4 g/L yeastolate. After 8 days, cells were pelleted and medium containing secreted mAb 3B4 was diluted 1:1

with Pierce Protein A IgG Binding Buffer and 0.22-μM filtered. The sample was loaded onto a pre-equilibrated 1 mL HiTrap Protein A HP column. The column was washed with Pierce Protein A IgG Binding Buffer and elution was accomplished with low pH Pierce IgG Elution Buffer. Acidic elutions were neutralized by 2 M Tris pH 8.0 to 10% of the final volume.

Generation of Rabbit Polyclonal Serum to Human Astrovirus 1

Rabbit polyclonal serum to human astrovirus 1 was generated as described in [41].

Human Plasma Samples

Human plasma samples were obtained from Discovery Life Sciences. These 63 de-identified samples were collected from adults ages 19–78 (31 male, 32 female) between 2012–2016 (before the emergence of SARS-CoV-2). Plasma samples were heated at 56 °C for 1 h before use. Appendix table A.1 includes information on patient age, sex, ethnicity, and race as well as date and location of sample collection.

Biolayer Interferometry Immunosorbent Assay

BLI-ISA studies were performed on an Octet RED384 instrument at 22 °C with shaking at 1000 rpm. BLI-ISA assay buffer consists of PBS pH 7.4, 2% BSA, and 0.1% Tween-20, which was 0.22-μm filtered. Before use, Anti-Penta-His (HIS1K) biosensors were placed into the wells of a biosensor-holding plate and pre-hydrated in BLI-ISA buffer for at least 10 min. Tilted-bottom 384-well microplates were loaded with 48 μL per well. The assay plate was prepared as follows: column 1 (BLI-ISA buffer), column 2 (10 nM His-tagged human astrovirus spikes 1–8 or 70 nM His-tagged SARS-CoV-2 RBD in BLI-ISA buffer), column 3 (a 1:4

dilution of ChonBlock in BLI-ISA buffer), column 4 (a 1:8 dilution of plasma samples: 6 μL plasma + 42 μL ChonBlock/BLI-ISA buffer), column 5 (BLI-ISA buffer), and column 6 (4 nm Colloidal Gold-AffiniPure Goat Anti-Human IgG or 4 nm Colloidal Gold-AffiniPure Goat Anti-Rabbit IgG secondary antibody diluted 1:10 in BLI-ISA buffer). The BLI-ISA method was set as follows, with minimal adaptation from [37]. Baseline 1 (60 s) in column 1 (Equilibration), Loading (180 s) in column 2 (Antigen Loading: human astrovirus spikes 1–8 or SARS-CoV-2 RBD), Baseline 2 (60 s) in column 3 (Wash), Association 1 (600 s) in column 4 (Total Antibody Binding), Baseline 3 (60 s) in column 5 (Wash), and Association 2 (180 s) in column 6 (Detection: anti-human IgG-gold or anti-rabbit IgG-gold). Loading of 10 nM spikes or 70 nM SARS-CoV-2 RBD over 180 s onto Anti-Penta-His sensor tips resulted in a wavelength shift loading signal of 0.55 nm. Analysis of BLI-ISA data was performed and automated as previously described in [37].

1.3.5 Results

First, we expressed recombinant Spike 1–Spike 8 antigens in E. coli and purified them by affinity chromatography and size exclusion chromatography. Purity was verified by SDS-PAGE (Figure 1.12A). Comparison of elution volumes for spike antigens to elution volumes for gel filtration standards confirmed that the spikes form dimers in solution (Figure 1.12B,C), consistent with structural studies [27,28,30].

We then used the recombinant spike antigens in a biolayer interferometry immunosorbent assay (BLI-ISA). BLI-ISA is a recently developed method for rapid and semi-quantitative measurement of antibodies in blood plasma directed towards a particular antigen [37]. In a simple dip-and-read format, the binding of

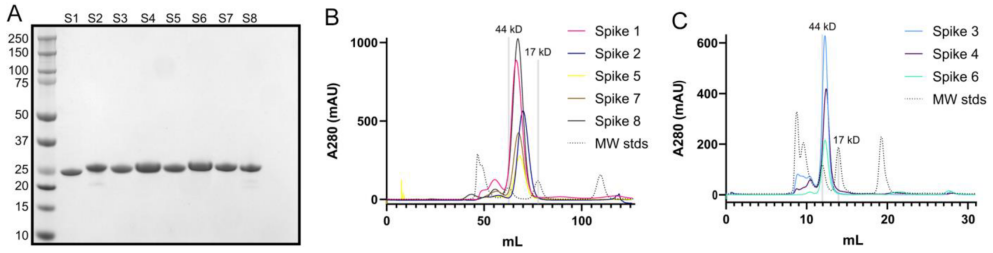


Figure 1.12 Purification of recombinant human astrovirus 1–8 capsid spikes (Spike 1–Spike 8). (A) Reducing SDS-PAGE of purified recombinant Spikes 1–8 (S1–S8). (B) Size exclusion chromatography column (Superdex 75 16/600) traces of Spikes 1, 2, 5, 7, and 8. (C) Size exclusion chromatography column (Superdex 75 10/300) traces of Spikes 3, 4, and 6. Size-exclusion chromatography confirmed that the purified spikes are folded and form dimers in solution.

biomolecules to fiber-optic biosensors is measured in real time through a shift in the wavelength of the reflected light upon binding, with results available in 20 min. After an initial equilibration, the biosensors are loaded with the antigen of interest. Next, the antigen-loaded biosensors are washed and placed into diluted plasma to allow binding of antibodies to the antigen. After another wash, the antigen/antibody-coated biosensors are dipped into wells containing isotype-specific binding reagents, such as colloidal gold-conjugated anti-human IgG, and a wavelength shift detection signal is measured (Figure 3A). BLI-ISA was originally established as a tool for serosurveillance of SARS-CoV-2 [37]. During its development, known SARS-CoV-2 seropositive and seronegative plasma samples were first validated by a standard dilution-series enzyme-linked immunosorbent assay (ELISA) starting at a 1:50 dilution. Importantly, trends observed in subsequent 1:8 single-dilution BLI-ISA measurements of those samples correlated with area-under-the-curve calculations from the dilution-series ELISA, indicating that BLI-ISA can identify the variation in antibody levels in seropositive samples at this plasma dilution.

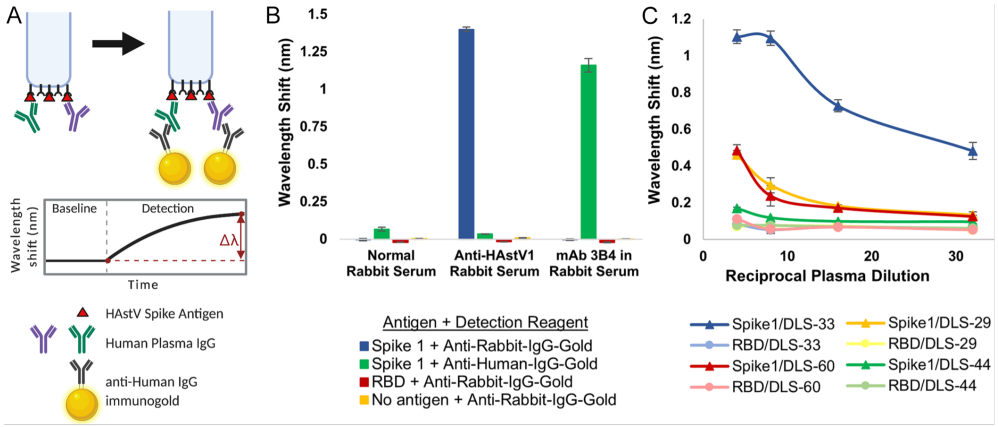


Figure 1.13 Validation of BLI-ISA for the detection of plasma IgG antibodies to human astrovirus capsid spikes. (A) BLI-ISA schematic of the IgG antibody detection step. This image was created with BioRender.com. (B) BLI-ISA detection of control sera. Spike 1, SARS-CoV-2 RBD, or no antigen were loaded onto biosensors and placed into 1:100 normal rabbit serum, 1:100 anti-human astrovirus 1 rabbit serum, or 25 nM mAb 3B4 in 1:100 normal rabbit serum. Bound antibodies were detected with anti-human-IgG-gold or anti-rabbit-IgG-gold. (C) Dilution-series BLI-ISA using Spike 1 (triangles) and RBD (circles) as antigens and representative strong astrovirus 1 seropositive (DLS-33, blue), moderate astrovirus 1 seropositive (DLS-17, cyan, and DLS-27, purple), weak astrovirus 1 seropositive (DLS-29, yellow, and DLS-60, red), and astrovirus 1 seronegative (DLS-44, green) plasma samples.

To confirm whether BLI-ISA can be applied to human astrovirus serosurveillance, we assessed Spike 1 antigen binding with commercially available normal rabbit serum and rabbit serum positive for human astrovirus serotype 1 [41]. As a positive control, monoclonal antibody (mAb) 3B4, a recombinant mouse–human chimeric antibody containing mouse variable regions that target Spike 1, was added to normal rabbit serum (Figure 3B). We also tested the specificity of our detection reagents, colloidal gold-conjugated anti-human IgG and colloidal gold-conjugated anti-rabbit IgG, in a 1:10 dilution. We loaded either Spike 1, the SARS-CoV-2 receptor-binding domain (RBD) negative control antigen, or no

antigen onto the biosensor tips and dipped them into the indicated sera. Background signals were observed in all samples with a 1:100 dilution of normal rabbit serum, indicating a lack of anti-Spike 1 and anti-RBD antibodies in the serum, as well as low non-specific binding of the serum to the empty biosensor tip. When Spike 1 was loaded onto the biosensor and dipped into a 1:100 dilution of anti-human astrovirus 1 rabbit serum, a robust signal emerged when detected with the anti-rabbit-IgG-gold reagent (Figure 3B, blue) but not the anti-human-IgG-gold reagent, indicating the high specificity of the anti-rabbit-IgG-gold reagent. The background signal was observed for RBD-loaded and "No antigen" biosensors. Finally, we added 25 nM mAb 3B4 to a 1:100 dilution of normal rabbit serum into which we dipped Spike 1-loaded biosensors. We noted a strong signal when the sample was detected with the anti-human-IgG-gold reagent (Figure 1.13B, green) but not the anti-rabbit-IgG-gold reagent, showing the precision of the anti-human-IgG-gold reagent. Background signal was exhibited for RBD- and no antigen-loaded biosensors also in this case.

Next, we conducted dilution-series experiments to assess the optimal human plasma dilution to use in BLI-ISA serological assays for human astrovirus. Dzimianski et al. established a 1:8 plasma dilution as optimal after recent SARS-CoV-2 exposure, providing an estimate of the dilution range for serosurveillance studies using other antigens [37]. Sixty-three human plasma samples were collected between 2012–2016 (pre-SARS-CoV-2) from a randomized cross-sectional sample of the population in the United States, adults ages 19–78 (Supplementary Table S1). We performed dilution-series BLI-ISA on a representative subset of these plasma samples using Spike 1 as the antigen (Figure 1.13C). As a control for non-specific binding, the SARS-CoV-2 RBD was loaded onto the biosensors instead of

Spike 1. The selected plasma samples include a strongly astrovirus 1-seropositive sample (DLS-33), two moderately astrovirus 1-seropositive samples (DLS-17 and DLS-27), two weakly astrovirus 1-seropositive samples (DLS-29 and DLS-60), and an astrovirus 1-seronegative sample (DLS-44) to represent the diversity of signals in a single-dilution assay. Each point was performed in duplicate. We loaded either Spike 1 or RBD onto the biosensors and then dipped into 1:4, 1:8, 1:16, or 1:32 dilutions of the indicated plasma. Detection signal at the 1:8 dilution compared to the 1:16 and 1:32 dilutions showed dose-dependent improvement for weak seropositives DLS-29 and DLS-60, as well as moderate seropositives DLS-17 and DLS-27 while maintaining low signal for the DLS-44 seronegative as well as the RBD-loaded background samples. While an even higher signal was observed for these weakly and moderately seropositive samples at the 1:4 dilution, compared to the 1:8 dilution, strong seropositive DLS-33 appears to reach signal saturation at the 1:8 dilution, with little signal improvement at the 1:4 dilution. In addition, seronegative sample DLS-44 and several of the RBD-loaded control samples showed non-specific signals at the 1:4 dilution but remained at background levels with the 1:8, 1:16, and 1:32 dilutions. Thus, we identified the 1:8 plasma dilution as optimal to maximize the dynamic range of the assay.

After validation of the control samples and confirmation of the optimal 1:8 plasma dilution, we used BLI-ISA to determine the seroprevalence of antibodies in the 63 human plasma samples to HAstV spikes from serotypes 1 to 8 (Figure 1.14 and Table 1.10). This investigation, using BLI-ISA to assess the unknown serostatus of individuals, is the first of its kind. Signals for each plasma sample to SARS-CoV-2 RBD were measured to establish each sample's background detection value. To reduce inter-sample variability and prevent false positives from

high-background samples, each sample's spike reactivity signal was subtracted by its own RBD reactivity signal to generate a background-corrected detection value. The variability in plasma reactivity to RBD was used to estimate a seropositivity cut-off of four times the standard deviation of background binding to RBD. Overall, the percentage of samples containing IgG antibodies targeting the spike protein was the highest for human astrovirus serotype 1 (46/63, 73%), followed by serotype 3 (39/63, 62%), serotype 4 (33/63, 52%), serotype 5 (18/63, 29%), serotype 8 (17/63, 27%), serotype 2 (14/63, 22%), serotype 6 (5/63, 8%), and serotype 7 (5/63, 8%), which is in accordance with a comprehensive human astrovirus serosurveillance study conducted by Koopmans et al. in the Netherlands 23 years ago (Table 1.9) [10].

1.3.6 Discussion

Here, we evaluate the seroprevalence of plasma antibodies towards human astrovirus serotypes 1–8 in a cross-sectional sample of United States adults. We adapted BLI-ISA, a recently developed method for the rapid and semi-quantitative measurement of plasma antibodies, to measure IgG antibodies against recombinant human astrovirus 1–8 capsid spikes (Spike 1–Spike 8). We chose Spike 1–Spike 8 as antigens for their ease of production in E. coli and for their variable sequence identity (41.4–75.7%) which should minimize antibody cross-reactivity between spikes from different serotypes. Indeed, most known neutralizing monoclonal antibodies that target the spike are serotype specific [38,41,42], suggesting that total antibody levels against individual spikes should show minimal cross-reactivity. Plasma dilution series confirm that BLI-ISA is semi-quantitative, and a plasma dilution of 1:8 was selected as the ideal dilution to maximize the dynamic

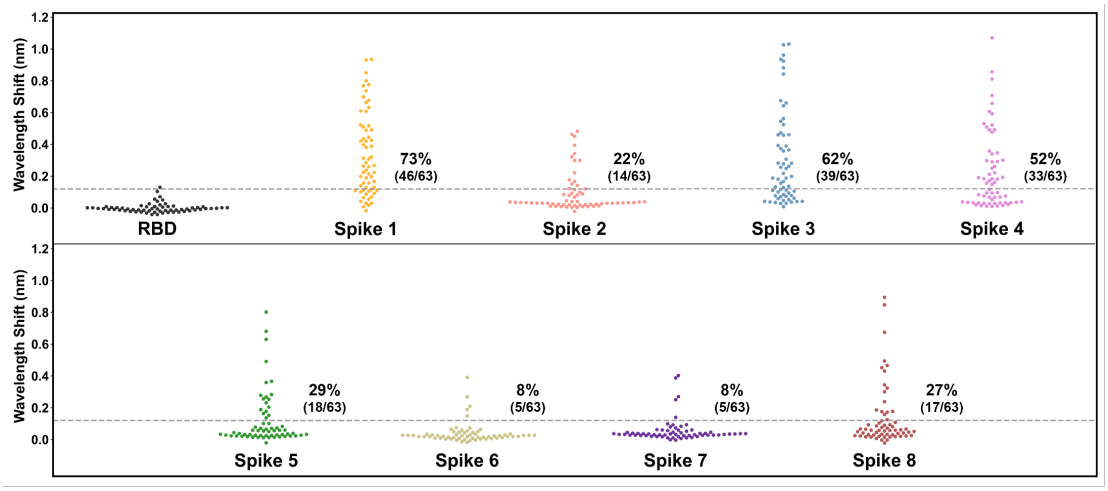


Figure 1.14 Human plasma IgG reactivity to human astrovirus spike proteins. Data are displayed as swarm plots, with each dot representing the signal for one individual for the indicated antigen. For samples with the RBD antigen, each sample's RBD reactivity signal was subtracted by the mean RBD reactivity signal (0.07 nm) to center samples around zero. The dashed line indicates four standard deviations above zero for the RBD samples (0.12 nm), and the printed percentages denote the number of samples above the line. For samples with the spike antigens, each sample's spike reactivity signal was subtracted by its own RBD reactivity signal to generate a background-corrected signal.

range of the BLI-ISA signals and minimize background signal using biosensors coated with a negative control antigen, the SARS-CoV-2 RBD, which should be non-reactive for antibodies in plasma collected before 2019 such as those in this study. Of note, a plasma dilution of 1:8 was also chosen to measure levels of antibodies to the SARS-CoV-2 RBD after recent SARS-CoV-2 exposure during BLI-ISA method development [37]. Time since exposure to human astrovirus is unknown for the samples in our study, yet the optimal plasma dilution remained 1:8. Therefore, BLI-ISA is a sensitive method that can detect antibodies elicited by an immune response which occurred potentially decades before plasma sample collection, while requiring very little plasma (6 μL total per sample per antigen).

		Controls		HASTV Antigens (background corrected)[a]							
	ID	No Antigen	RBD	Spike-1	Spike-2	Spike-3	Spike-4	Spike-5	Spike-6	Spike-7	Spike-8
1	DLS-18	0.13	0.06	0.11	0.04	1.03	0.03	0.02	0.00	0.05	0.03
2	DLS-14	0.10	0.05	0.20	0.03	0.88	0.15	0.03	0.03	0.09	0.07
3	DLS-55	0.07	0.06	0.32	0.15	0.92	0.04	0.04	0.06	0.10	0.18
4	DLS-44	0.05	0.04	0.02	0.03	0.46	0.15	0.02	0.05	0.04	0.17
5	DLS-36	0.14	0.07	0.01	0.09	0.54	0.03	0.01	0.01	0.03	0.06
6	DLS-51	0.08	0.07	0.16	0.02	0.66	0.01	0.02	0.01	0.04	0.02
7	DLS-06	0.11	0.06	0.11	0.04	0.67	0.53	0.06	0.03	0.05	0.09
8	DLS-01	0.14	0.08	0.03	0.11	0.96	0.59	0.03	0.04	0.06	0.32
9	DLS-30	0.13	0.09	0.43	0.02	0.31	0.09	0.26	0.03	0.07	0.07
10	DLS-02	0.20	0.10	0.52	0.03	0.47	0.09	0.27	0.05	0.05	0.10
11	DLS-41	0.10	0.07	0.68	0.03	0.52	0.22	0.28	0.19	0.08	0.08
12	DLS-20	0.09	0.05	0.44	0.03	0.46	0.03	0.03	0.02	0.04	0.03
13	DLS-05	0.37	0.20	0.42	-0.02	0.56	0.07	0.16	0.00	0.01	-0.02
14	DLS-54	0.07	0.07	0.31	0.01	0.37	0.03	0.03	0.03	0.09	0.02
15	DLS-21	0.22	0.17	0.27	0.02	0.46	0.03	-0.02	0.00	0.04	0.12
16	DLS-49	0.09	0.06	0.78	0.02	0.84	0.02	0.06	0.02	0.25	0.06
17	DLS-35	0.09	0.07	0.63	0.14	0.64	0.19	0.01	0.01	0.02	0.01
18	DLS-24	0.12	0.06	0.11	0.03	1.03	0.07	0.19	0.06	0.27	0.89
19	DLS-19	0.08	0.08	0.12	0.00	0.18	0.01	0.14	0.00	0.39	0.47
20	DLS-28	0.06	0.03	0.19	0.03	0.28	0.04	0.02	0.01	0.02	0.67
21	DLS-42	0.07	0.06	0.26	0.03	0.13	0.01	0.02	0.01	0.04	0.43
22	DLS-07	0.09	0.05	0.26	0.01	0.08	0.10	0.68	0.05	0.03	0.05
23	DLS-25	0.10	0.06	0.85	0.46	0.05	0.12	0.63	0.03	0.03	0.09
24	DLS-27	0.11	0.08	0.77	0.01	0.06	0.08	0.02	-0.01	0.01	0.34
25	DLS-34	0.10	0.07	0.70	0.01	0.14	0.03	0.03	0.00	0.02	0.45
26	DLS-23	0.09	0.05	0.93	0.09	0.28	0.07	0.36	0.03	0.04	0.04
27	DLS-33	0.10	0.10	0.93	0.03	0.08	0.01	0.03	-0.02	0.00	0.00
28	DLS-09	0.29	0.12	0.80	0.07	0.10	0.17	0.05	0.07	0.02	0.02
29	DLS-53	0.08	0.06	0.74	0.03	0.04	0.29	0.10	0.06	0.04	0.08
30	DLS-61	0.08	0.07	0.49	0.01	0.03	0.02	0.06	0.01	0.03	0.01
31	DLS-17	0.09	0.06	0.61	0.03	0.19	0.03	0.02	0.01	0.04	0.03
32	DLS-04	0.09	0.04	0.49	0.03	0.25	0.05	0.07	0.02	0.06	0.06
33	DLS-46	0.08	0.06	0.39	0.02	0.17	0.03	0.03	0.02	0.03	0.02
34	DLS-12	0.06	0.03	0.38	0.03	0.11	0.03	0.25	0.03	0.03	0.05
35	DLS-31	0.16	0.14	0.40	0.01	0.07	0.18	0.08	-0.01	0.01	0.00
36	DLS-40	0.09	0.05	0.31	0.02	0.06	0.21	0.07	0.03	0.03	0.03
37	DLS-26	0.12	0.08	0.61	0.04	0.20	0.26	0.06	0.00	0.02	0.17
38	DLS-45	0.10	0.07	0.51	0.12	0.28	0.25	0.07	0.02	0.06	0.30
39	DLS-39	0.13	0.07	0.45	0.17	0.10	0.36	0.02	0.02	0.06	0.06
40	DLS-47	0.09	0.05	0.66	0.01	0.04	0.52	0.03	0.02	0.03	0.02
41	DLS-59	0.08	0.07	0.21	0.02	0.09	0.49	0.03	0.03	0.40	0.02
42	DLS-32	0.11	0.07	0.07	0.00	0.13	0.49	0.01	0.02	0.01	0.07
43	DLS-37	0.11	0.07	0.09	0.03	0.36	0.35	0.05	0.02	0.04	0.24
44	DLS-43	0.10	0.06	0.09	0.03	0.36	0.30	0.03	0.02	0.03	0.04
45	DLS-15	0.08	0.04	0.22	0.08	0.25	0.34	0.23	0.21	0.14	0.16
46	DLS-58	0.08	0.06	0.28	0.22	0.39	0.48	0.04	0.03	0.04	0.12
47	DLS-11	0.08	0.04	0.07	0.45	0.03	0.19	0.18	0.02	0.03	0.06
48	DLS-62	0.13	0.12	0.10	0.32	0.26	0.07	0.03	0.05	0.08	0.08
49	DLS-57	0.07	0.06	0.04	0.04	0.04	0.16	0.04	0.02	0.03	0.19
50	DLS-13	0.09	0.04	0.03	0.12	0.19	0.07	0.15	0.01	0.04	0.05
51	DLS-48	0.09	0.06	0.06	0.09	0.22	0.18	0.07	0.02	0.03	0.02
52	DLS-56	0.07	0.08	0.14	0.34	0.93	0.71	0.80	0.27	0.01	0.11
53	DLS-63	0.07	0.08	0.11	0.48	0.39	0.51	0.49	0.15	0.03	0.04
54	DLS-08	0.08	0.05	0.52	0.30	0.11	1.07	0.28	0.39	0.03	0.85
55	DLS-52	0.22	0.11	-0.02	0.39	0.01	0.81	0.01	0.00	0.00	0.49
56	DLS-50	0.10	0.07	0.42	0.18	0.12	0.66	0.37	0.05	0.04	0.09
57	DLS-03	0.08	0.04	0.17	0.04	0.04	0.85	0.20	0.07	0.03	0.05
58	DLS-22	0.11	0.06	0.16	0.30	0.07	0.61	0.08	0.03	0.06	0.04
59	DLS-29	0.08	0.05	0.20	0.04	0.04	0.18	0.06	0.02	0.03	0.02
60	DLS-10	0.10	0.05	0.22	0.10	0.05	0.30	0.10	0.04	0.03	0.06
61	DLS-16	0.09	0.04	0.13	0.08	0.08	0.29	0.03	0.02	0.03	0.03
62	DLS-38	0.16	0.08	0.24	0.01	0.15	0.03	0.01	0.01	0.02	0.02
63	DLS-60	0.05	0.04	0.13	0.01	0.03	0.01	0.04	0.03	0.02	0.02
64	PBS	0.06	0.05	0.01	0.02	0.02	0.02	0.02	0.03	0.02	0.02

[a]HASTV antigen detection scores were adjusted by subtracting the RBD antigen detection score to control for non-specific binding.

Table 1.10 Heat map of individual human plasma IgG reactivity to human astrovirus spike antigens

This study estimates the seroprevalence rates, in order of prevalence, as 73% for HAstV1, 62% for HAstV3, 52% for HAstV4, 29% for HAstV5, 27% for HAstV8, 22% for HAstV2, 8% for HAstV6, and 8% for HAstV7. These findings are consistent with previous seroprevalence studies, as well as genotypic surveillance studies identifying human astrovirus 1 as the most prevalent serotype/genotype [9,10,11,12]. This result is also consistent with a previous seroprevalence study in identifying human astrovirus serotypes 3, 4, and 5 as having the next highest prevalence after serotype 1 (Table 1.9) [10]. We extend previous work by evaluating seroprevalence for human astrovirus serotype 8 for the first time and determine that it is relatively high at 27%, an unexpected observation given that it was discovered after serotypes 6 and 7, which have relatively low seroprevalence rates of 8% each. We note that seroprevalence rates in this study are likely modest estimates based upon a signal threshold of four times the standard deviation of the samples with a negative control antigen. Indeed, many individual samples had signals for spikes that were two or three times their own background signal yet did not meet the conservative criteria for seropositivity in this study.

Critically, seroprevalence rates for recombinant human astrovirus capsid spike antigens correlate closely with neutralizing antibody rates, determined previously (Table 1.9) [10]. This finding is significant because neutralizing antibodies are often a marker of protection from viral infection. While correlates of protection have yet to be defined for human astrovirus infection, there are some studies suggesting that antibodies are protective [23,24,25,26]. The observation that antibodies to the human astrovirus capsid spike correlate with virus neutralization is not entirely surprising, given previous evidence. Specifically, studies of HAstV-

neutralizing monoclonal antibodies have identified that their epitopes reside on the spike domain or the capsid fragments that contains the spike domain, VP24 and VP26 [38,41,42,43]. Moreover, while both the capsid core and capsid spike domains stimulate antibody production in mice and rabbits, only antibodies to the spike show neutralizing activity [28,38,41]. Thus, plasma IgG levels against human astrovirus capsid spikes may provide a useful estimate of neutralizing antibody levels.

Finally, we present Table 1.10 as a heatmap of antibody levels of each individual for human astrovirus 1–8 capsid spikes. Individuals were clustered using the Ward distance minimization algorithm to identify trends in seropositivity. We find that every individual has reactivity to at least one serotype of spike and, on average, individuals are seropositive for about 3 human astrovirus serotypes. While no sample had reactivity to all 8 spikes, one sample had reactivity to 7 spikes (DLS-15). A few samples were only weakly to moderately seropositive to only one or two spikes (DLS-13, -57, -60). We examined the data for correlations in sample positivity between any two serotypes, indicating potential antibody cross-reactivity, but did not identify any trends. Interestingly, clustering revealed a trend in individuals with high seroreactivity for one spike from either HAstV1, -3, or -4 (which have the highest seroprevalence) in that these individuals rarely have high seroreactivity to another spike from these prevalent serotypes. This result might indicate that a strong antibody response to a first infection provides partial protection to an infection by a different serotype, resulting in a weaker antibody response to the different serotype. However, further research is necessary to explore this possibility.

1.3.7 Funding

This research was funded by NIH grant R01 AI144090. Funding for the purchase of the Octet RED384 instrument was supported by the NIH S10 shared instrumentation grant 1S10OD027012-01. L.M. was supported by an ARCS Foundation Fellowship. K.D.-C. was supported by an NIH Supplement to Promote Diversity. J.F. was supported by a Koret Undergraduate Research Scholarship.

1.3.8 Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the University of California Santa Cruz (protocol 3684, approval date 9 October 2020).

1.3.9 Informed Consent Statement

Informed consent was obtained from all subjects involved in the study, through Discovery Life Sciences.

Additional Acknowledgements

We thank Dorsey Bass for the donation of rabbit polyclonal antibody serum to HAstV1.

1.3.10 References

1. Mendez, E.; Arias, C.F. Astroviruses. In Fields virology, 2 ed.; Howley, P.M., Ed. Lippincott Williams & Wilkins: 2007; pp 982-1000.

2. Koci, M.D.; Schultz-Cherry, S. Avian astroviruses. Avian pathology : journal of the W.V.P.A 2002, 31, 213-227. 3. Woode, G.N.; Pohlenz, J.F.; Gourley,

N.E.; Fagerland, J.A. Astrovirus and breda virus infections of dome cell epithelium of bovine ileum. J Clin Microbiol 1984, 19, 623-630.

4. Cortez, V.; Sharp, B.; Yao, J.; Livingston, B.; Vogel, P.; Schultz-Cherry, S. Characterizing a murine model for astrovirus using viral isolates from persistently infected immunocompromised mice. J Virol 2019, 93.

5. Brown, J.R.; Morfopoulou, S.; Hubb, J.; Emmett, W.A.; Ip, W.; Shah, D.; Brooks, T.; Paine, S.M.; Anderson, G.; Virasami, A., et al. Astrovirus va1/hmo-c: An increasingly recognized neurotropic pathogen in immunocompromised patients. Clin Infect Dis 2015, 60, 881-888.

6. Naccache, S.N.; Peggs, K.S.; Mattes, F.M.; Phadke, R.; Garson, J.A.; Grant, P.; Samayoa, E.; Federman, S.; Miller, S.; Lunn, M.P., et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis 2015, 60, 919-923.

7. Quan, P.L.; Wagner, T.A.; Briese, T.; Torgerson, T.R.; Hornig, M.; Tashmukhamedova, A.; Firth, C.; Palacios, G.; Baisre-De-Leon, A.; Paddock, C.D., et al. Astrovirus encephalitis in boy with x-linked agammaglobulinemia. Emerg Infect Dis 2010, 16, 918-925.

8. Cordey, S.; Vu, D.L.; Schibler, M.; L'Huillier, A.G.; Brito, F.; Docquier, M.; Posfay-Barbe, K.M.; Petty, T.J.; Turin, L.; Zdobnov, E.M., et al. Astrovirus mlb2, a new gastroenteric virus associated with meningitis and disseminated infection. Emerg Infect Dis 2016, 22, 846-853.

9. King, A.M.Q.; Lefkowitz, E.; Adams, M.J.; Carstens, E.B. Virus taxonomy: Classification and nomenclature of viruses. In Ninth report of the International Committee on Taxonomy of Viruses., 9th edition ed.; Elsevier Inc.: Philadelphia, PA, 2011.

10. Koopmans, M.P.; Bijen, M.H.; Monroe, S.S.; Vinje, J. Age-stratified seroprevalence of neutralizing antibodies to astrovirus types 1 to 7 in humans in the netherlands. Clinical and diagnostic laboratory immunology 1998, 5, 33-37.

11. Kurtz, J.B.; Lee, T.W. Human astrovirus serotypes. Lancet 1984, 2, 1405.

12. Cortez, V.; Meliopoulos, V.A.; Karlsson, E.A.; Hargest, V.; Johnson, C.; Schultz-Cherry, S. Astrovirus biology and pathogenesis. Annu Rev Virol 2017, 4, 327-348.

13. Walter, J.E.; Mitchell, D.K. Astrovirus infection in children. Curr Opin Infect Dis 2003, 16, 247-253.

14. Goodgame, R.W. Viral causes of diarrhea. Gastroenterology clinics of North America 2001, 30, 779-795.

15. Gray, J.J.; Wreghitt, T.G.; Cubitt, W.D.; Elliot, P.R. An outbreak of gastroenteritis in a home for the elderly associated with astrovirus type 1 and human calicivirus. J Med Virol 1987, 23, 377-381.

16. Lewis, D.C.; Lightfoot, N.F.; Cubitt, W.D.; Wilson, S.A. Outbreaks of astrovirus type 1 and rotavirus gastroenteritis in a geriatric in-patient population. The Journal of hospital infection 1989, 14, 9-14.

17. Marshall, J.A.; Bruggink, L.D.; Sturge, K.; Subasinghe, N.; Tan, A.; Hogg, G.G. Molecular features of astrovirus associated with a gastroenteritis outbreak in an aged-care centre. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 2007, 26, 67-71.

18. Dennehy, P.H.; Nelson, S.M.; Spangenberger, S.; Noel, J.S.; Monroe, S.S.; Glass, R.I. A prospective case-control study of the role of astrovirus in acute diarrhea among hospitalized young children. J Infect Dis 2001, 184, 10-15.

19. Gallimore, C.I.; Taylor, C.; Gennery, A.R.; Cant, A.J.; Galloway, A.;

Xerry, J.; Adigwe, J.; Gray, J.J. Contamination of the hospital environment with gastroenteric viruses: Comparison of two pediatric wards over a winter season. J Clin Microbiol 2008, 46, 3112-3115.

20. Olortegui, M.P.; Rouhani, S.; Yori, P.P.; Salas, M.S.; Trigoso, D.R.; Mondal, D.; Bodhidatta, L.; Platts-Mills, J.; Samie, A.; Kabir, F., et al. Astrovirus infection and diarrhea in 8 countries. Pediatrics 2018, 141.

21. Bosch, A.; Pinto, R.M.; Guix, S. Human astroviruses. Clinical microbiology reviews 2014, 27, 1048-1074.

22. Mead, P.S.; Slutsker, L.; Dietz, V.; McCaig, L.F.; Bresee, J.S.; Shapiro, C.; Griffin, P.M.; Tauxe, R.V. Food-related illness and death in the united states. Emerg Infect Dis 1999, 5, 607-625.

23. Kurtz, J.; Lee, T. Astrovirus gastroenteritis age distribution of antibody. Med Microbiol Immunol 1978, 166, 227-230.

24. Kurtz, J.B.; Lee, T.W.; Craig, J.W.; Reed, S.E. Astrovirus infection in volunteers. J Med Virol 1979, 3, 221-230.

25. Mitchell, D.K. Astrovirus gastroenteritis. The Pediatric infectious disease journal 2002, 21, 1067-1069.

26. Bjorkholm, M.; Celsing, F.; Runarsson, G.; Waldenstrom, J. Successful intravenous immunoglobulin therapy for severe and persistent astrovirus gastroenteritis after fludarabine treatment in a patient with waldenstrom's macroglobulinemia. International journal of hematology 1995, 62, 117-120.

27. Dong, J.; Dong, L.; Mendez, E.; Tao, Y. Crystal structure of the human astrovirus capsid spike. Proceedings of the National Academy of Sciences of the United States of America 2011, 108, 12681-12686.

28. York, R.L.; Yousefi, P.A.; Bogdanoff, W.; Haile, S.; Tripathi, S.; DuBois, R.M. Structural, mechanistic, and antigenic characterization of the human astro-

virus capsid. J Virol 2015, 90, 2254-2263.

29. Toh, Y.; Harper, J.; Dryden, K.A.; Yeager, M.; Arias, C.F.; Mendez, E.; Tao, Y.J. Crystal structure of the human astrovirus capsid protein. J Virol 2016, 90, 9008-9017.

30. Espinosa, R.; López, T.; Bogdanoff, W.A.; Espinoza, M.A.; López, S.; DuBois, R.M.; Arias, C.F. Isolation of neutralizing monoclonal antibodies to human astrovirus and characterization of virus variants that escape neutralization. J Virol 2018.

31. Mitchell, D.K.; Matson, D.O.; Cubitt, W.D.; Jackson, L.J.; Willcocks, M.M.; Pickering, L.K.; Carter, M.J. Prevalence of antibodies to astrovirus types 1 and 3 in children and adolescents in norfolk, virginia. The Pediatric infectious disease journal 1999, 18, 249-254.

32. Kriston, S.; Willcocks, M.M.; Carter, M.J.; Cubitt, W.D. Seroprevalence of astrovirus types 1 and 6 in london, determined using recombinant virus antigen. Epidemiol Infect 1996, 117, 159-164.

33. Kobayashi, S.; Kobayashi, M.; Araki, K.; Shinozaki, T.; Yanagawa, Y. [antibody prevalence against astrovirus according to age groups]. Kansenshogaku Zasshi 1999, 73, 578-583.

34. Holtz, L.R.; Bauer, I.K.; Jiang, H.; Belshe, R.; Freiden, P.; Schultz-Cherry, S.L.; Wang, D. Seroepidemiology of astrovirus mlb1. Clin Vaccine Immunol 2014, 21, 908-911.

35. Burbelo, P.D.; Ching, K.H.; Esper, F.; Iadarola, M.J.; Delwart, E.; Lipkin, W.I.; Kapoor, A. Serological studies confirm the novel astrovirus hmoastv-c as a highly prevalent human infectious agent. PLoS One 2011, 6, e22576.

36. Meliopoulos, V.A.; Kayali, G.; Burnham, A.; Oshansky, C.M.; Thomas, P.G.; Gray, G.C.; Beck, M.A.; Schultz-Cherry, S. Detection of antibodies against

turkey astrovirus in humans. PLoS One 2014, 9, e96934.

37. Dzimianski, J.V.; Lorig-Roach, N.; O'Rourke, S.M.; Alexander, D.L.; Kimmey, J.M.; DuBois, R.M. Rapid and sensitive detection of sars-cov-2 antibodies by biolayer interferometry. Sci Rep 2020, 10, 21738.

38. Meyer, L.; López, T.; Espinosa, R.; Arias, C.F.; Vollmers, C.; DuBois, R.M. A simplified workflow for monoclonal antibody sequencing. PLoS One 2019, 14, e0218717.

39. Wu, X.; Yang, Z.Y.; Li, Y.; Hogerkorp, C.M.; Schief, W.R.; Seaman, M.S.; Zhou, T.; Schmidt, S.D.; Wu, L.; Xu, L., et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to hiv-1. Science 2010, 329, 856-861.

40. Bass, D.M.; Upadhyayula, U. Characterization of human serotype 1 astrovirus-neutralizing epitopes. J Virol 1997, 71, 8666-8671.

41. Bogdanoff, W.A.; Campos, J.; Perez, E.I.; Yin, L.; Alexander, D.L.; DuBois, R.M. Structure of a human astrovirus capsid-antibody complex and mechanistic insights into virus neutralization. J Virol 2017, 91.

42. Sanchez-Fauquier, A.; Carrascosa, A.L.; Carrascosa, J.L.; Otero, A.; Glass, R.I.; Lopez, J.A.; San Martin, C.; Melero, J.A. Characterization of a human astrovirus serotype 2 structural protein (vp26) that contains an epitope involved in virus neutralization. Virology 1994, 201, 312-320.

43. Bogdanoff, W.A.; Campos, J.; Perez, E.I.; Yin, L.; Alexander, D.L.; DuBois, R.M. Structure of a human astrovirus capsid-antibody complex and mechanistic insights into virus neutralization. J. Virol. 2016, 91, e01859-16.

Chapter Two

Aiding development of antibody-based cancer therapeutics

This chapter highlights my work in the cancer therapeutics space. In the first section, different techniques for identifying tumor-specific cell surface proteins are implemented and evaluated. The goal behind uncovering highly expressed surface proteins on tumors is to both enable diagnostic techniques that may allow early tumor detection, as well as identification of targets for antibody biologics or CAR-T cell therapies. To achieve this goal, a large dataset of RNAseq data from The Cancer Genome Atlas (TCGA) project was interrogated.

In the next section, the epitope of lirilumab, an immune checkpoint inhibitor antibody, is investigated. This antibody is meant to enhance the activity of natural killer (NK) cells by repressing inhibitory signaling of the KIR family of NK cell receptors. Lirilumab is in clinical trials for a variety of cancer indications where it is being tested on its own, as well as in combination with other antibody therapeutics or conventional chemotherapeutic agents. Because NK cells are the immune system's effectors of antibody dependent cellular cytotoxicity, enhancing NK cell activity could pair exceptionally well with other tumor-targeting antibodies.

2.1 Differentiating between tumor biopsies and normal cells in the TCGA dataset using targetable cell-surface gene sets

2.1.1 Preface

This Section contains project work inspired by my coursework in Angela Brooks' RNA bioinformatics class and Chris Vollmer's data visualization course. The goal of the work was to use RNA-seq data from both tumor and normal cell biopsies taken from throughout the body to identify proteins that are uniquely expressed (or at least have uniquely high expression levels) in tumor subtypes. Additionally, I wanted to compare methods for extracting those potential outlier genes – particularly dimensionality reduction techniques like PCA, simpler statistical methods, and (at the time) mainstream differential expression packages (DESeq).

2.1.2 Abstract

The cancer genome atlas project (TCGA) represents one of the largest sources of mRNA-seq data collected from human tumors. The expression profile of CD, receptor tyrosine kinase, and nuclear hormone receptor transcripts within this data set was investigated using a number of methods including DESeq and t-SNE based clustering. These gene families mediate cell-to-cell communication, adhesion, immune recognition, and hormone response among many roles. By identifying unique expression patterns in TCGA tumor cohorts, mechanisms these cancers use to co-opt host resources, evade the immune system, and proliferate may be identified. Highly expressed members of these cell-surface protein families

could also serve as targets for immunotherapies or even small molecule drugs. Here it is shown that TCGA tumor cohorts can be distinguished from TCGA normals using this small, targetable set of 486 genes. Initial insights into the genes within this set that may represent therapeutic targets are discussed.

2.1.3 Introduction

The classical view of cancer genomics is summed up well in Vogelstein's review 'Cancer Genome Landscapes'[41] – the focus has been on identifying highly mutated genes, and in particular, so-called driver genes that are the catalysts for a tissue's transformation into a tumor. Cancer genome sequencing projects like the Cancer Genome Atlas strove to provide the data needed to identify these drivers, and now it is not difficult to identify highly mutated genes among the cancer biopsies sequenced as part of the TCGA project. [44] However, it is still a challenge to correlate even the most probable 'driver' genes to expected survival and a bigger challenge still to choose an appropriate therapy based on a patient's cancer phenotype. Gene transcription data now accessible via mRNA-seq experiments allows another view into the 'life' of a cancer, specifically its transcriptome and by proxy its proteome, which can inform us of the effects of the mutations observed via genome sequencing. The author would argue that a tumor's proteome is a better indicator of what sort of therapy might effectively kill it.

One way to take advantage of transcriptome/proteome data is to find highly expressed cell surface proteins that can be used to identify tumors such that targeted therapies like antibodies can be trained to find and eliminate tumor cells.[33] This strategy has increased in popularity dramatically since the 1990s: up to 2001, approximately 30 human mAbs were clinically evaluated – now the cumulative

number of clinically investigated antibodies is greater than 150 with the majority developed for cancer indications. [26] Despite these advantages, mAbs still face significant hurdles including immunogenicity of the antibodies themselves [34], high production costs, and challenges in formulation and administration. The three basic mechanisms by which antibody therapeutics act is by interruption of a protein function, recruitment of the effector arm of the immune system (e.g. NK cells, granulocytes, complement proteins), or by carrying a 'warhead' to a target. [43] Currently this methodology has been successful primarily in leukemias and lymphomas, where the antibody is targeted to B-cell specific lectins like CD22, but solid tumors have also been successfully targeted like in the HER2-targeting Herceptin for HER2 positive breast cancers. [43]

Here the transcription of 'clusters of differentiation' (CD), receptor tyrosine kinase (RTK), and nuclear hormone receptor (nH) genes are investigated as a proof of concept for methods of identifying tumors with unusual protein expression. Both the CD and RTK gene families are cell surface proteins of which at least one member is already the target of an antibody therapeutic or other therapy in the clinic or in development (e.g. CD20 and HER2;[24] VEGF[24] and PTK7[8]). The CD genes are a family of proteins closely tied to the immune system containing numerous immunoglubulin containing proteins, lectins, selectins, and other proteins involved in cell-cell recognition and communication. [14] RTKs can have dramatic effects on protein expression in response to cytokines, hormones, and growth factors and are known to be important in angiogenesis, development, and oncogenic processes.[21] The nuclear hormone receptor gene set was added to aid in differentiating between androgen dependent tumors like breast cancer. The transcription of these genes was evaluated using TCGA mRNA-seq data processed with the TOIL pipeline[40] that used RSEM [22] to quantify gene expression.

Gene expression data from normal cell biopsies and primary tumors was clustered using t-SNE, which showed generally robust separation between tumors and normal cells, indicating the CD/RTK/nH gene set is suitable to differentiate between some cancers and the normal cells in TCGA cohorts. The genes driving the separation of primary tumors and normal cells were assessed using a number of techniques including analysis with DESeq2. [23]

2.1.4 Methods

The mRNAseq data used in this study is available at the TOIL hub of the UCSC Xena browser (https://toil.xenahubs.net). Specifically, RSEM expected counts data from the TCGA Pan-cancer cohort was used. A metadata table covering most of these samples is available at Synapse (https://www.synapse.org/#!Synapse:syn7248855). To prepare data for analyses, a Python program was written that can remove genes with low variance and subset the bulk RSEM data table by cohort, sample type, and gene set. The output are csv files containing the data matrix subset accompanied by a row-matched metadata subset for easy import into programs like DESeq2. The Python package sklearn was used for the t-SNE and PCA based clustering, which was incorporated into a program that takes the data & metadata output generated previously and plots 2D representations of the computed similarity between samples' gene transcription.

Prior to analysis with DESeq2, RSEM's expected count values were rounded to the nearest integer and all entries were incremented by 1 pseudocount. Here, a single cohort's primary tumor samples were compared to all normals with more than 20 samples in the same cohort. An example DESeq run is shown below, where 'reduced_data.csv' and 'col_data.csv' correspond to the data and metadata

matrices, respectively.

```
> dd <- read.csv("reduced_data.csv", header = TRUE,
                   row.names = 1, check.names = False)
> dd <- round(dd)
> dd <- dd + 1
> col_data <- read.csv("col_data.csv")
> ddsmat <- DESeqDataSetFromMatrix(countData= dd, colData= col_data,
                                        design= ~ sample_type)
> dds <- DESeq(ddsmat, parallel = TRUE)
> results <- results(dds, parallel = TRUE)
> sum(res$padj < 0.01, na.rm = TRUE)
[1] 138
> res_ordered <- results[order(results$padj),]
> res_ordered <- as.data.frame(res_ordered)[1:138,]
> write.csv(res_ordered, file = "deseq_results.csv")
```

After exporting the significant (p-adj < 0.01) differentially expressed genes, they were sorted by greatest-fold increase in gene expression in the primary tumor sample type to show genes that may be targetable or used to identify tumors. In cohorts like BRCA where there are both 'TP' and 'NT' sample types, it is possible to run deseq with the parameter 'design $= \sim$ sample_type $+$ disease' where disease corresponds to cohort.

See Table 2.3 for a full list of TCGA cohorts and the number of samples within each sample type. CD and RTK, and nuclear hormone receptor genes used in this study were ripped from HUGO's gene database, simply using all genes in the respective categories and removing any overlap between categories.[14]

2.1.5 Results

A subset of TCGA tumor cohorts were analyzed using t-SNE based dimensionality reduction to assess whether the CD-RTK-nH gene set would be effective for differentiating between primary tumors and TCGA normals. For this trial analysis, DLBC, PAAD, BRCA, and the lung carcinomas LUAD and LUSC were chosen because these cancers are either high occurrence (lung, breast), high lethality (pancreatic), or act as controls where genes in this gene set are already used clinically to identify or treat disease (DLBC). All four t-SNE analyses (figure 2.1) showed excellent separation from TCGA normals to the author's eye. These same data were also investigated with DESeq2 to potentially identify the genes within the CD-RTK-nH gene set that drove the separation between samples in the t-SNE analyses (Table 2.1 and Table 2.2). Genes with $log2$ fold change higher than 1 are shown for the DLBC cohort and higher than 0.45 for the PAAD, BRCA and LUSC cohorts. In these DESeq analyses, Fc receptor like proteins and TNF-related proteins (like CD70 and TNFRS- family genes) appeared most frequently. Figure 2.2 plots expression of a subset of these 'hits' across all TCGA cohorts, using a visualization provided at firebrowse.org.[3]

Dimensionality reduction techniques like PCA could also be used to identify genes with unique expression patterns. After clustering with PCA, the eigenvectors behind each principle component can be decomposed to pull out the most significant contributory elements to the observed clustering. Here, where the original data are a set of transcript counts for our surface marker gene set, those major contributors to each principle component could be unique targetable genes. In Fig 2.3 the clustering achieved by PCA is shown at the top left. Then, expression across the TCGA dataset was plotted for the genes with the greatest contributions to the

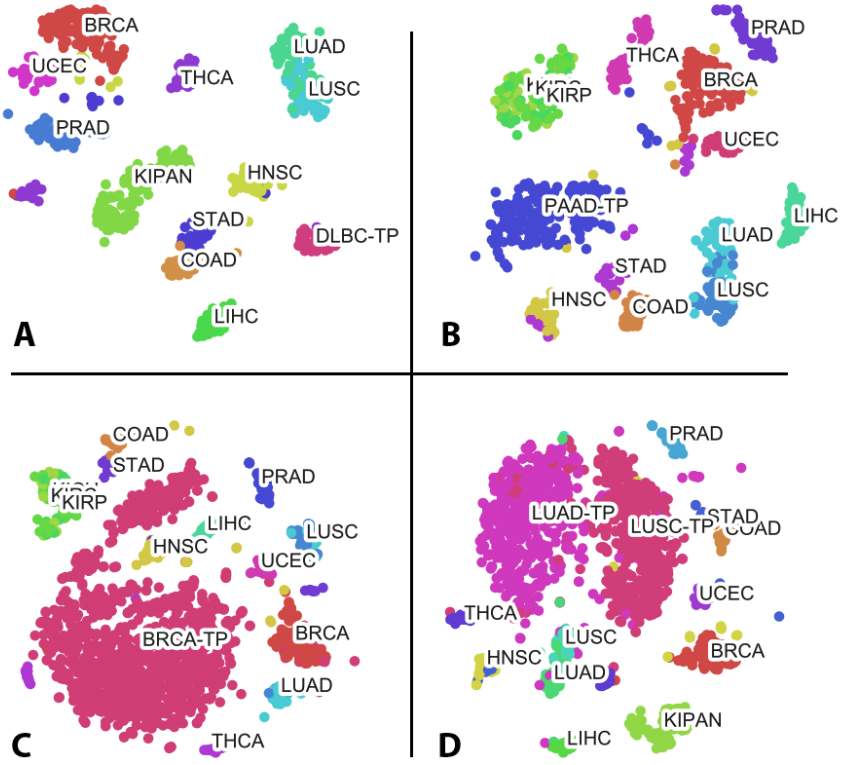


Figure 2.1 Plots showing t-SNE based dimensionality reduction using the CD-RTK-nH gene set with TCGA normals and A: DLBC primary tumors, B: pancreatic adenocarcinoma tumors, C: breast adenocarcinoma tumors, and D: lung adenocarcinoma and squamous cell carcinomatumors . t-SNE was performed using the scikit-learn python package with a euclidean-based distance matrix, perplexity = 30, and default settings for remaining parameters. Plotted using matplotlib and seaborn, labels were applied automatically based on the coordinates of the median value for a given cohort:sample-type.

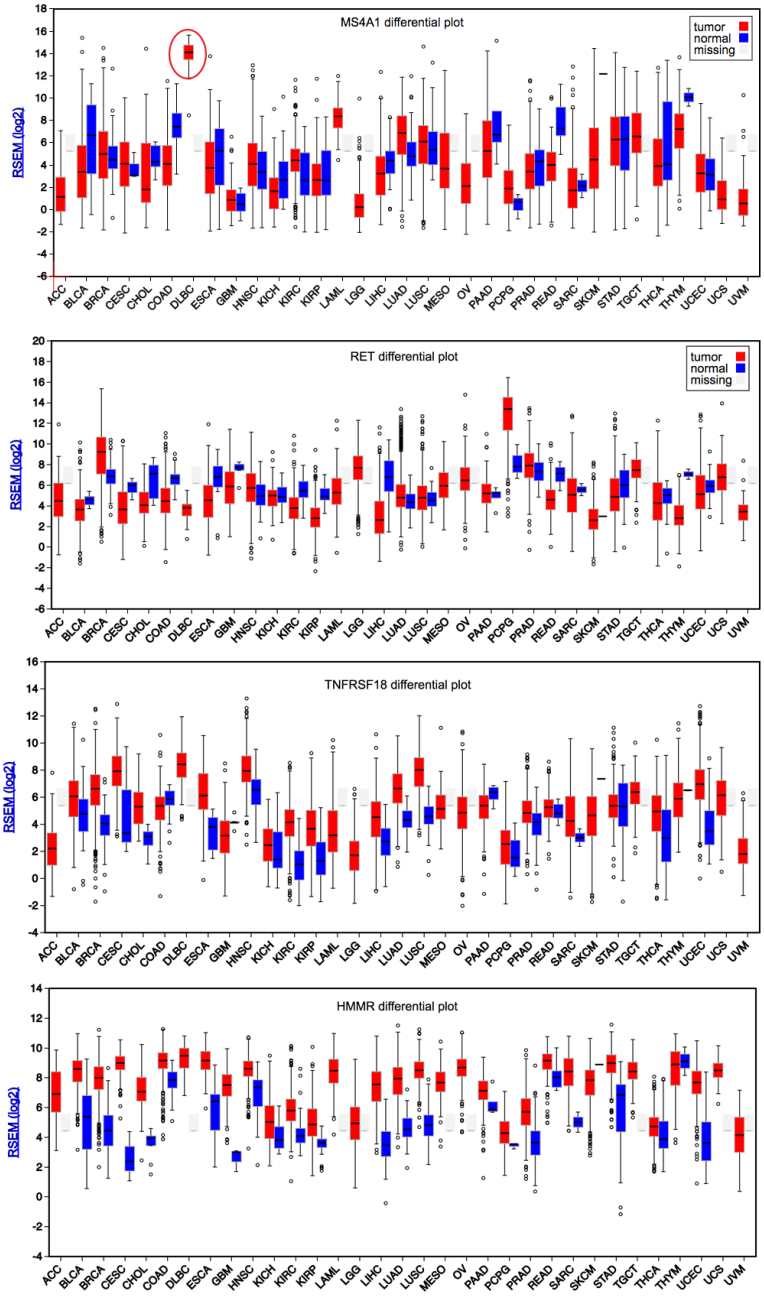


Figure 2.2 Expression *log*2 histograms across all TCGA cohorts (excluding duplicate 'pan' cohorts e.g. KIPAN, COADREAD) for MS4A1 (CD20), RET, TNFSF18, and HMMR. Made at firebrowse.org/viewGene.html.

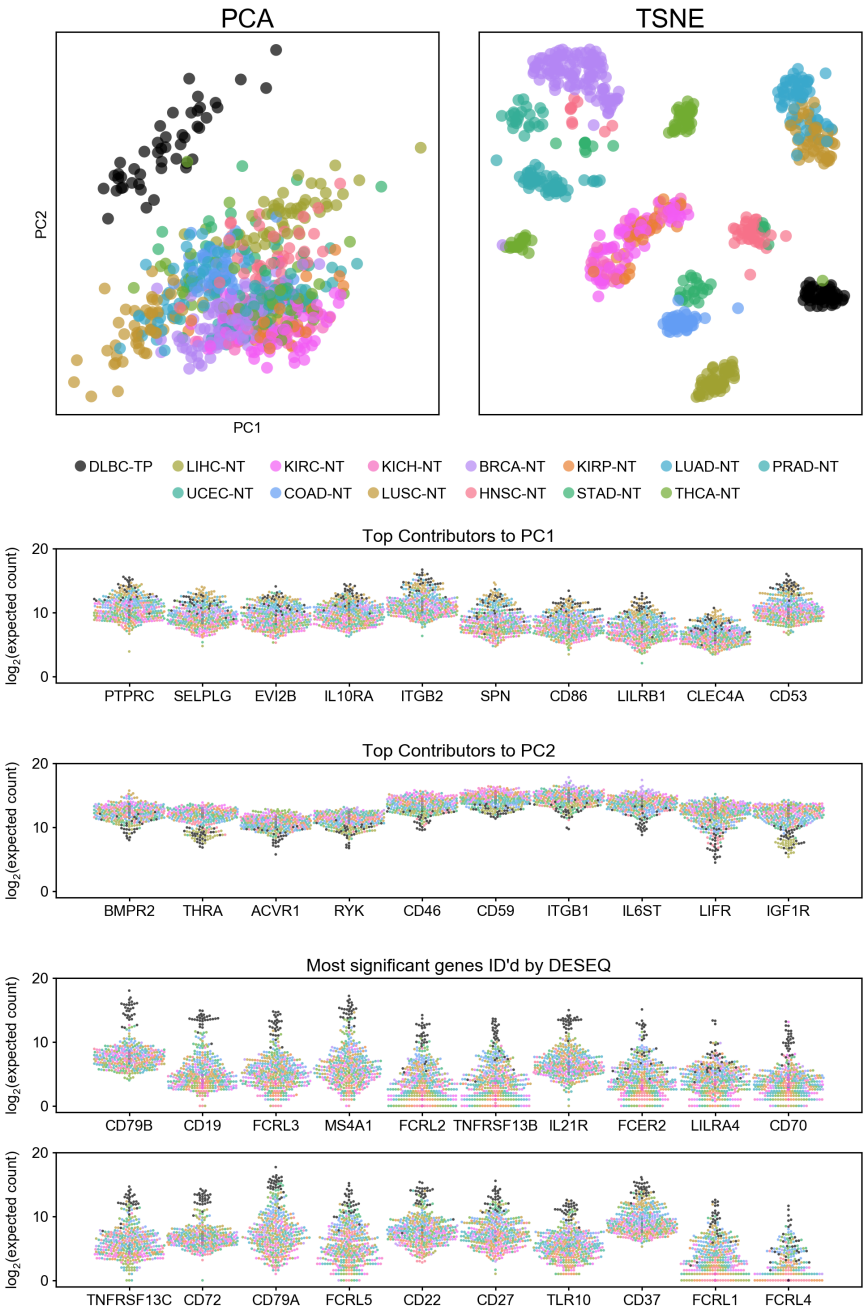


Figure 2.3 Identifying Differential Expression of Tumor Antigens - Comparing PCA and DEseq. RNAseq data from the set of $\tilde{5}00$ CD, RTK, and nH genes was used to generate the PCA, TSNE, and swarm plots. The expression of genes with the largest contributions to the first two principle components and the genes with the most significant p-values identified by DESeq2 are plotted.

Cohort	Gene	log_2 fold change	p-adj	Gene description
DLBC-TP	CD79B	7.82	0.0E+00	CD79b molecule
	CD19	7.63	1.9E-140	CD19 molecule
	FCRL3	7.42	6.0E-176	Fc receptor like 3
	MS4A1	7.25	1.0E-103	membrane spanning 4-domains A1
	FCRL2	7.03	1.0E-122	Fc receptor like 2
	TNFRSF13B	6.78	2.7E-113	TNF receptor superfamily member 13B
	IL21R	6.59	1.6E-261	interleukin 21 receptor
	FCER2	6.45	1.2E-92	Fc fragment of IgE receptor II
	LILRA4	6.38	4.0E-144	leukocyte immunoglobulin like receptor A4
	CD70	6.32	6.0E-132	CD70 molecule
	TNFRSF13C	6.3	3.3E-125	TNF receptor superfamily member 13C
	CD72	6.29	0.0E+00	CD72 molecule
	CD79A	6.11	7.6E-72	CD79a molecule
	FCRL5	5.96	1.5E-91	Fc receptor like 5
	CD22	5.48	2.1E-96	CD22 molecule
	CD27	5.31	9.4E-112	CD27 molecule
	TLR10	5.28	8.4E-120	toll like receptor 10
	CD37	5.25	2.0E-180	CD37 molecule
	FCRL1	5.24	7.4E-67	Fc receptor like 1
PAAD-TP	ADAM8	3.31	0.0E+00	ADAM metallopeptidase domain 8
	SEMA7A	2.91	4.0E-205	semaphorin 7A (John Milton Hagen blood group)
	CEACAM6	2.83	1.2E-26	carcinoembryonic antigen related cell adhesion molecule 6
	MST1R	2.72	2.8E-98	macrophage stimulating 1 receptor
	TNFSF11	2.58	5.0E-51	TNF superfamily member 11
	PLAUR	2.25	1.4E-106	plasminogen activator, urokinase receptor
	MUC1	1.84	6.2E-35	mucin 1, cell surface associated
	CD70	1.79	2.1E-49	CD70 molecule
	ITGB4	1.75	4.7E-43	integrin subunit beta 4
	IL2RA	1.67	1.5E-71	interleukin 2 receptor subunit alpha
	CEACAM5	1.6	3.3E-07	carcinoembryonic antigen related cell adhesion molecule 5
	CCR8	1.55	6.1E-61	C-C motif chemokine receptor 8
	IL2RG	1.52	2.1E-49	interleukin 2 receptor subunit gamma
	TNFSF4	1.36	6.5E-48	TNF superfamily member 4
	LAIR2	1.35	2.1E-29	leukocyte associated immunoglobulin like receptor 2
	HMMR	1.32	4.0E-25	hyaluronan mediated motility receptor
	CD72	1.29	2.6E-52	CD72 molecule
	CD55	1.29	6.6E-35	CD55 molecule (Cromer blood group)
	FUT3	1.25	7.2E-09	fucosyltransferase 3 (Lewis blood group)

Table 2.1 "Highly expressed" CD/RTK/nH genes identified using DESeq2 from the TCGA cohorts DLBC and PAAD primary tumors. DESeq was run comparing the cohorts 'TP' samples to 'NT' samples from all cohorts with at least 20 normal samples. Gene descriptions from HUGO. [14]

first two principle components of the cluster. For comparison, expression of the surface receptor genes identified by DESeq are similarly plotted below. Among the top 10 genes identified by both methodologies, there was no overlap between the two approaches.

2.1.6 Discussion

It is exciting to see how well the primary tumors in Fig. 2.1 were separable from TCGA normal using t-SNE with this small gene set. However, it is unclear

Cohort	Gene	log_2 fold change	p-adj	Gene description
LUSC -TP	NR0B1	6.94	0.0E+00	nuclear receptor subfamily 0 group B member 1
	TNFRSF18	3.55	6.5E-302	TNF receptor superfamily member 18
	NR5A1	3.12	1.7E-145	nuclear receptor subfamily 5 group A member 1
	TFRC	3.1	0.0E+00	transferrin receptor
	HMMR	2.85	7.7E-267	hyaluronan mediated motility receptor
	NR2E1	2.74	6.0E-118	nuclear receptor subfamily 2 group E member 1
	CD70	2.37	4.5E-126	CD70 molecule
	SLC7A5	2.33	1.8E-114	solute carrier family 7 member 5
	TNFRSF9	2.19	2.0E-190	TNF receptor superfamily member 9
	CCR8	2.16	2.0E-228	C-C motif chemokine receptor 8
	EPHA8	2.16	6.9E-82	EPH receptor A8
	ADAM8	2.15	1.3E-265	ADAM metallopeptidase domain 8
	FZD10	2.15	9.4E-68	frizzled class receptor 10
	ITGB4	2.08	1.4E-135	integrin subunit beta 4
	IL2RA	2.04	1.9E-199	interleukin 2 receptor subunit alpha
	FCRL5	2.02	2.5E-81	Fc receptor like 5
	CD109	1.96	2.8E-132	CD109 molecule
	ATP1B3	1.87	2.3E-150	ATPase Na+/K+ transporting subunit beta 3
	CLEC7A	1.77	6.7E-136	C-type lectin domain containing 7A
	TNFRSF13C	1.76	1.7E-78	TNF receptor superfamily member 13C
	ADAM17	1.67	4.4E-217	ADAM metallopeptidase domain 17
BRCA-TP	RET	4.08	0.0E+00	ret proto-oncogene
	EPHA8	3.92	2.6E-272	EPH receptor A8
	ESR1	3.63	8.8E-261	estrogen receptor 1
	BMPR1B	3.4	1.8E-162	bone morphogenetic protein receptor type 1B
	TNFRSF18	2.92	5.3E-258	TNF receptor superfamily member 18
	HMMR	2.69	0.0E+00	hyaluronan mediated motility receptor
	CCR8	2.64	2.1E-301	C-C motif chemokine receptor 8
	FLT3	2.33	1.3E-151	fms related tyrosine kinase 3
	PGR	1.96	1.4E-61	progesterone receptor
	TNFSF4	1.93	0.0E+00	TNF superfamily member 4
	ERBB2	1.85	1.2E-129	erb-b2 receptor tyrosine kinase 2
	ADAM8	1.83	3.0E-295	ADAM metallopeptidase domain 8
	MUC1	1.74	1.1E-96	mucin 1, cell surface associated
	NR2E3	1.74	6.1E-112	nuclear receptor subfamily 2 group E member 3
	IFITM1	1.7	2.8E-181	interferon induced transmembrane protein 1
	ERBB3	1.68	1.4E-195	erb-b2 receptor tyrosine kinase 3
	NR5A1	1.67	3.0E-56	nuclear receptor subfamily 5 group A member 1
	IGF1R	1.67	4.0E-106	insulin like growth factor 1 receptor
	RARA	1.6	3.0E-198	retinoic acid receptor alpha

Table 2.2 "Highly expressed" CD/RTK/nH genes identified using DESeq2 from the TCGA cohorts LUSC and BRCA primary tumors. DESeq was run comparing the cohorts 'TP' samples to 'NT' samples from all cohorts with at least 20 normal samples. Gene descriptions from HUGO. [14]

Cohort	sample type TP	TR	TB	TAP	TM	TAM	NT	Description
BRCA	1093	-	-	-	7	-	112	Breast AC
KIPAN	889	-	-	2	-	-	129	Pan-kidney cohort
COADREAD	623	2	-	-	1	-	51	Colorectal AC
STES	599	-	-	-	1	-	46	Stomach and Esophageal C
UCEC	545	1	-	-	-	-	35	Uterine Corpus Endometrial C
LGG	516	14	-	-	-	-	-	Brain Lower Grade Glioma
HNSC	520	-	-	-	2	-	44	Head and neck SCC
LUAD	515	2	-	-	-	-	59	Lung AC
LUSC	501	-	-	-	-	-	51	Lung SCC
THCA	501	-	-	-	8	-	59	Thyroid C
PRAD	497	-	-	-	1	-	52	Prostate AC
BLCA	408	-	-	-	-	-	19	Bladder Urothelial C
LIHC	371	2	-	-	-	-	5	Liver hepatocellular C
CESC	304	-	-	-	2	-	3	Cervical SCC & endocervical AC
OV	303	4	-	-	-	-	-	Ovarian serous cyst-AC
SARC	259	3	1	-	-	-	2	Sarcoma
PAAD	178	4	-	-	1	-	4	Pancreatic AC
LAML	-	-	173	-	-	-	-	Acute Myeloid Leukemia
TGCT	150	-	-	6	-	-	-	Testicular Germ Cell Tumors
THYM	120	-	-	-	-	-	2	Thymoma
MESO	87	-	-	-	-	-	-	Mesothelioma
ACC	79	-	-	-	-	-	-	Adrenocortical AC
UCS	57	-	-	-	-	-	-	Uterine Carcinosarcoma
DLBC	48	-	-	-	-	-	-	Diffuse Large B-cell Lymphoma
CHOL	36	-	-	-	-	-	9	Cholangiocarcinoma

Table 2.3 Summary of TCGA mRNAseq data used in this study. Counts for each sample type reflect those served by the Firebrowse python API as of March 2017 and do not include filtered samples (e.g replicates or redactions).[3] This 'level 3' Ilummina HiSeq/Ga2 data is preprocessed with the RSEM [22] method and served as log_2 transformed floats. When available, 'pan' cohorts like KIPAN, COADREAD, and STES were used rather than their component cohorts. AC = adenocarcinoma; C = carcinoma; SCC = squamous cell carcinoma.

from that analysis whether the gene features that allow these samples to be distinguished from each other could be viable therapeutic targets. An ideal target would have significantly higher expression among tumors, giving a wide therapeutic window. Based on the significant ($p < 0.01$) genes identified with DESeq, there are nearly equivalent number of up- and down-regulated genes between normals and DLBC (174 vs 171), PAAD (72 vs 79), LUSC (66 vs 72), and BRCA (23 vs 22). Despite many genes showing significantly higher expression, only the DLBC tumor cohort had genes with more than 2-fold greater expression than the aggregate normals.

The DLBC cohort in many ways acted as a positive control for this analysis, where current, clinically approved antibody therapeutic targets like MS4A1 (more commonly CD20) stand out from both normal cells in these cohorts and the other tumor cells (table 2.1, figure 2.2). Among these genes and cohorts, no aggregate primary tumor expression data had a distribution that paralleled hits like MS4A1 in DLBC, where nearly the entirety of the e-$log2$ distribution was outside of the upper percentile of all other cohorts. In light of this, the majority of the genes identified in this work may not be suitable for targeted therapies. However, the expression trends are significant enough to warrant further investigation into the biology backing the observed trends.

Other semi-validating hits include TNFR family members: at least one subtype is highly expressed in each cohort assessed here. TNF is a known mediator of inflammation and angiogenesis and a potent modulator of immune cell activation with a known role in cancer progression. [16] The carcinoembryonic antigen molecules (CEACAM5/6) were also over-expressed in the pancreatic solid tumors assessed with DESeq, another family of molecules known to be associated with cancers.[35] CEACAMs are named due to their expression during fetal develop-

ment, though some expression, particularly in epithelial cells, can be observed into adulthood. However, these proteins are reportedly (and according to the data analyzed here) markers of tumorogenesis and are thought to enable tumor growth and metastasis by stimulating release of tumorogenic cytokines from immune and other cell types. [35] In the BRCA cohort, many known players appear including RET and estrogen and progesterone receptors, in addition to numerous other highly significant leads.

One of the main flaws with the mRNA-seq data used in this work is the lack of normal cell biopsies for many tissue types, particularly CNS, immune system, and germ line (i.e. cervical, ovarian, testicular) tissues. This calls into question many of the genes deemed significant in the DESeq analyses, however it is clear that it is possible to distinguish between the normals available here and DLBC, PAAD, BRCA and lung primary tumors. The TOIL pipeline has also run the entirety of the GETx[37] mRNAseq data with the same RSEM based quantification methodology, which may be a suitable source of additional normals. Their suitability can be assessed by clustering GTEx normals with those available through TCGA – if the normals do not cluster well then additional primary data may be needed for comprehensive expression analyses.

Another important point to recall is that these values are based on transcript abundance which may not accurately reflect the state of the proteome. Additional experiments are required to verify these transcripts are being translated before any could be considered targets. However, the data here suggest that rapid and accurate clustering of tumor and normal cells can be achieved using a modestly sized, biologically relevant gene set. This is useful for analysis of patient biopsies and could perhaps be used to aid in early identification of cancers.

Though the work presented here is preliminary, further development of differ-

ential transcription analysis and gene family based phenotyping appears to be a worthwhile pursuit in uncovering novel cancer cell biology and perhaps even new therapeutic targets.

2.1.7 Acknowledgments

Thanks to Josh Stuart and John Vivian for both curating the data used in this work, and pointing me toward it.

2.2 Structural basis for the activity and specificity of the immune checkpoint inhibitor lirilumab

2.2.1 Preface

The project described in this section started as a collaboration with Trellis Biosciences who were pursuing a follow-on drug to Innate Pharma's immune checkpoint inhibitor antibody, lirilumab. Lirilumab was developed with cancer indications in mind. The antibody binds to a number of receptors in the KIR family, with the highest affinity for the inhibitory receptors KIR2DL1, -2, and -3. The basic premise of lirilumab's action is to heighten natural killer cell cytotoxic activity toward tumors via blockade of those inhibitory receptors. The goal of the overall project was to compare lirilumab's selectivity and epitope to Trellis' candidate follow-on antibodies as well as aid development of immune checkpoint inhibitors targeting the KIR receptor family generally. In the research article reproduced here, lirilumab's epitope was mapped using x-ray crystallography and key amino acids that define selectivity across the KIR family were investigated along with population-wide variants identified through analysis of single nucleotide variants from large-scale genome sequencing efforts.

This Section contains a reproduction of a research paper submitted for publication with Scientific Reports.At the time of this thesis' submission, favorable reviews have been received and manuscript acceptance is anticipated.

Citation

Structural basis for the activity and specificity of the immune checkpoint inhibitor lirilumab

Lorig-Roach, N.; Harpell, N.; DuBois, R. M.

Scientific Reports

Author contributions

N.L.R. and R.M.D. conceived the study. N.L.R. and N.M.H. cloned and expressed antibody constructs. N.L.R. expressed and purified receptors, performed crystallographic screening and acquired biolayer interferometry data. N.L.R. and R.M.D. analyzed the data. N.L.R. prepared figures. N.L.R. and R.M.D. wrote the manuscript. All authors edited the manuscript.

2.2.2 Abstract

The clinical success of immune checkpoint inhibitors has underscored the key role of the immune system in controlling cancer. Current FDA-approved immune checkpoint inhibitors target the regulatory receptor pathways of cytotoxic T-cells to enhance their anticancer responses. Despite an abundance of evidence that natural killer (NK) cells can also mediate potent anticancer activities, there are no FDA-approved inhibitors targeting NK cell specific checkpoint pathways. Lirilumab, the most clinically advanced NK cell checkpoint inhibitor, targets inhibitory killer immunoglobulin-like receptors (KIRs), however it has yet to conclusively demonstrate clinical efficacy. Here we describe the crystal structure of lirilumab in complex with the inhibitory KIR2DL3, revealing the precise epitope

of lirilumab and the molecular mechanisms underlying KIR checkpoint blockade. Notably, the epitope includes several key amino acids that vary across the human population, and binding studies demonstrate the importance of these amino acids for lirilumab binding. These studies reveal how KIR variations in patients could influence the clinical efficacy of lirilumab and reveal general concepts for the development of immune checkpoint inhibitors targeting NK cells.

2.2.3 Introduction

Despite decades of progress in the detection and treatment of cancer, its myriad of forms and variations persist as some of the biggest threats to human health [1]. While small molecule chemotherapeutic agents remain the frontline of non-surgical treatment, a number of monoclonal antibody biologics have become available for specialized indications, including antibodies targeting cancer cells themselves (e.g. rituximab targeting CD20 in leukemia and lymphoma; trastuzumab targeting HER2 for breast cancer) or targeting growth factors that promote tumor growth (e.g. bevacizumab targeting VEGF-A). Monoclonal antibodies' precise and selective targeting of almost any molecular motif, their relative ease of production, and ability to pair with immune effector systems continue to make them strong therapeutic candidates.

In the past decade, a new class of monoclonal antibody biologic termed immune checkpoint inhibitors has emerged. Immune checkpoint inhibitors modulate regulatory receptors on cytotoxic T-cells and natural killer (NK) cells to improve clearance of cancer cells [2–6]. Cytotoxic T-cells and NK cells play a pivotal role in the immune system's response to malignancy, using a variety of strategies to

surveil for the presence of non-self with a mixture of activating and inhibitory receptors, often interacting with MHC molecules [7]. In the presence of strong activating receptor signals or a lack of inhibitory receptor signals, T-cells and NK cells are activated and secrete cytolytic molecules, cytokines, and chemokines that kill nearby cancerous cells and recruit additional immune effector cells [8]. Immune checkpoint inhibitors seek to tip the balance further toward activation by blocking inhibitory receptors such as CTLA-4, PD-1, LAG-3, and KIRs, and preventing tumors from escaping detection via inhibitory signaling [4]. The FDA has approved the use of several antibodies that target PD-1/PD-L1 (nivolumab, pembrolizumab, atezolizumab, and others), CTLA-4 (ipilimumab, tremelimumab), and LAG-3 (relatlimab) for a number of indications, including melanoma, head and neck cancer, lung cancer, renal cancer and liver cancer [9].

Among immune checkpoint inhibitor targets, the killer immunoglobulin-like receptor (KIR) family is a candidate for the enhancement of NK cell mediated tumor clearance [10]. KIRs recognize the basally expressed MHC class I molecules, conferring NK cells the ability to detect self. The KIR family is composed of a number of both inhibitory and activating receptors which differ in the number of extracellular Ig-like domains they contain (KIR2DL1 vs KIR3DL1), the length of their cytoplasmic tail (KIR2DL1 vs KIR2DS1), and sequence variation (KIR2DL1 vs KIR2DL3). It has been known for some time that there is significant population level diversity in both KIR and MHC genes, and different KIR receptors and their variants pair with distinct HLA molecules [11,12]. The binding kinetics between KIRs and HLA are characterized by rapid on- and off-rates with affinities ranging from high nanomolar to micromolar depending on the variant pairing and the peptide presented by the HLA [13,14].

Some of the first evidence that blocking KIR function could improve cancer outcomes came from hematopoietic stem cell transplants in leukemia patients where there was a KIR:HLA repertoire mismatch – patients receiving transplants with an alloreactive NK cell population showed improved disease-free survival rates and reduced rate of relapse [15,16]. Early in vitro experiments demonstrated that KIR-targeted antibody fragments could promote NK-mediated cytotoxicity toward transfected tumor cell lines expressing HLA Cw3 [17]. In mouse models, blocking the mouse KIR receptor equivalent, Ly49, led to NK-cell dependent clearance of RMA lymphoma cells but not syngeneic B6 spleen cells [18]. Data like these led to interest in the development of anti-KIR monoclonal antibody therapeutics to improve tumor clearance by blocking signaling of inhibitory KIR receptors on NK cells (Fig. 1a) [19,20].

Lirilumab (IPH2102, descended from 1-7F9) is a monoclonal antibody therapeutic candidate that was developed to target the two primary inhibitory KIRs (KIR2DL1 and KIR2DL2/3) (Fig. 2.4a). To obtain such an antibody, mice bearing human genomic IgG loci were immunized with BW5417 cells expressing KIR2DL1, followed by 3 boosters with the extracellular region of recombinant KIR2DL3 [21]. The monoclonal antibody selected for development demonstrated high affinity binding to KIR2DL1 and -L3, weaker binding to KIR2DS4, and no binding to the divergent KIR2DL4 or any KIR3D receptor [19,20]. The variable domains of the anti-KIR antibody were cloned into an IgG4 constant region framework to avoid antibody dependent cytotoxicity toward NK cells [19]. In preclinical studies, lirilumab enhanced in vitro lysis of patient-derived acute myeloid leukemia (AML) blasts when administered with IL-2 activated, HLA-matched, NK-cells. Similarly, NOD-SCID mice challenged with both AML cells and NK

cells were rescued by lirilumab where control mice succumbed to leukemia [21]. Phase I clinical trials demonstrated initial safety of lirilumab, prompting a number of phase II clinical trials using lirilumab alone and in combination with other chemotherapeutics [21–24]. While lackluster efficacies of lirilumab in these trials have likely hindered its advancement into phase III clinical trials, a lack of understanding of lirilumab's epitope and specificities for different KIR variants and usages in different patients may be preventing additional phase II clinicals in more stratified patient populations.

Here we determined the crystal structure of the lirilumab variable domains bound to the inhibitory KIR2DL3. The lirilumab epitope overlaps with the binding site for HLA, revealing the mechanism of inhibition by lirilumab. Further investigation of the lirilumab epitope reveals several key amino acids that vary across the human population-level KIR repertoire, and binding studies demonstrate the importance of these amino acids for lirilumab binding. Altogether, our data support the further development of lirilumab and other KIR-targeting immune checkpoint inhibitors.

2.2.4 Results

To define the epitope of lirilumab, we first produced a recombinant antigen binding fragment (Fab) encoding the variable domains of lirilumab fused to the constant domains of a human IgG1 Fab. We termed this antibody construct FabLirilumab. We also produced the recombinant ectodomain of KIR2DL3, which is reported to bind lirilumab with high affinity [19,20]. A complex of FabLirilumab and

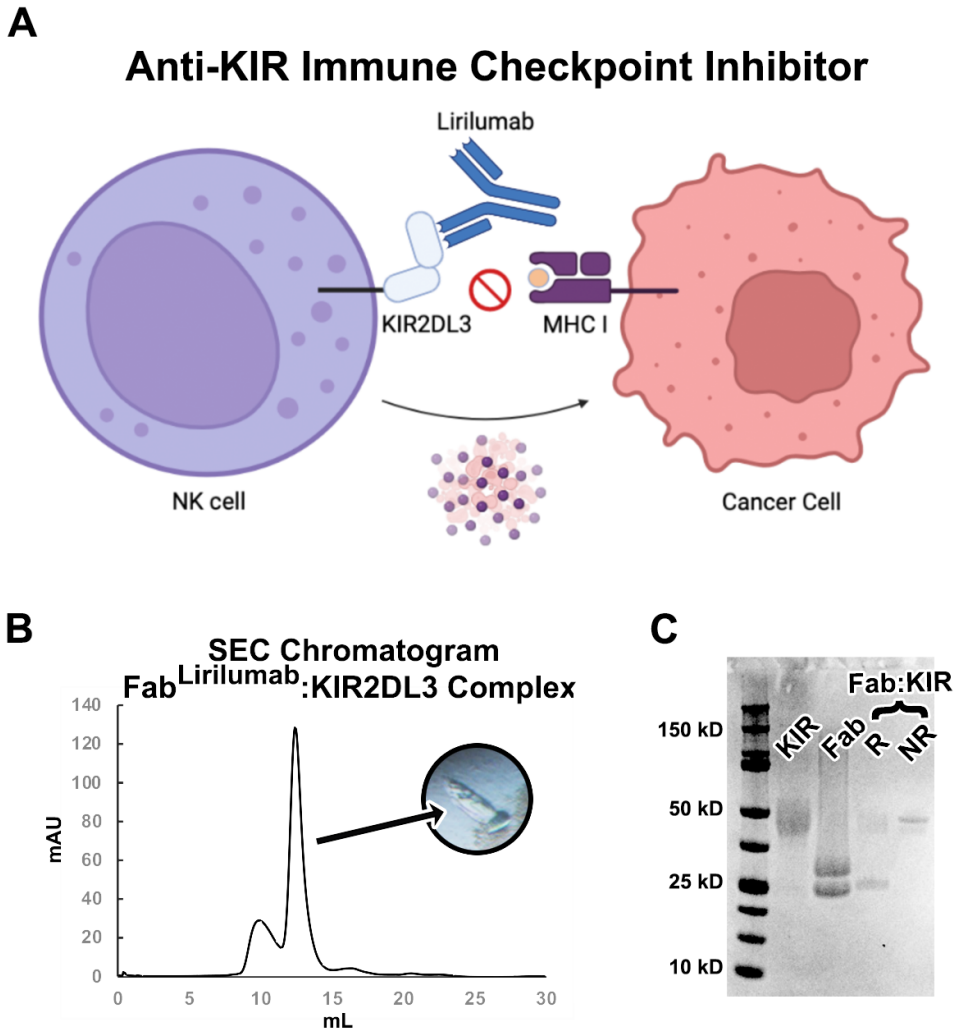


Figure 2.4 The immune checkpoint inhibitor lirilumab binds to NK cell inhibitory KIR receptors. (A) Graphical summary describing the mechanism of the immune checkpoint inhibitor antibody lirilumab targeting inhibitory KIR receptors on NK cells, which promotes NK cell-mediated clearance of cancer cells. Panel made with BioRender. (B) Size exclusion chromatography purification and crystal formation of the FabLirilumab – KIR2DL3 complex. (C) Coomassie-stained SDS-PAGE of KIR2DL3 (KIR), FabLirilumab in a reducing buffer (Fab), and the purified FabLirilumab – KIR2DL3 complex (Fab:KIR) in a reducing (R) or non-reducing (NR) buffer.

KIR2DL3 was purified by size exclusion chromatography and the presence of intact KIR2DL3 and FabLirilumab was confirmed by SDS-PAGE (Fig. 2.4b-c). The FabLirilumab – KIR2DL3 complex was screened in crystallization trials and after 3 weeks crystals were obtained. X-ray diffraction data were collected and molecular replacement was used to determine the crystal structure of the FabLirilumab – KIR2DL3 complex at 2.8 Å resolution (Fig. 2.5, Table 2.4). During molecular replacement, a solution could only be obtained when the KIR2DL3 structure was split into its two separate domains. The final structure is a complex of FabLirilumab with only domain 1 of KIR2DL3 (Fig. 2.5), and analysis of the crystallographic packing reveals that the other domain could not fit. Thus, we suspect that proteolysis of the complex occurred during crystallization trials and the extended time required for crystallization. Nevertheless, the resulting structure from the crystallized complex enabled detailed analyses of the lirilumab epitope.

To understand the mechanism of inhibitory action by lirilumab, we compared the FabLirilumab – KIR2DL3 structure to known HLA - KIR complex structures (Fig. 2.5). While the lirilumab epitope is limited to domain 1 of KIR2DL3 (Fig. 2.5, left), the epitope overlaps with approximately 134 Å2 of the binding site for HLA, which spans both domains 1 and 2 of KIR2DL3 (Fig. 2.5, center). Thus, lirilumab directly blocks the binding site for HLA. We further sought to understand why lirilumab does not bind to KIR3D receptors, despite the significant homology between KIR2DL3 and some KIR3D receptors. Comparison to a structure of HLA bound to a KIR3D receptor reveals that lirilumab's epitope is wholly obscured by the third Ig-like domain in KIR3Ds (Fig. 2.5, right), explaining its inability to bind receptors in that class. We further investigated the molecular

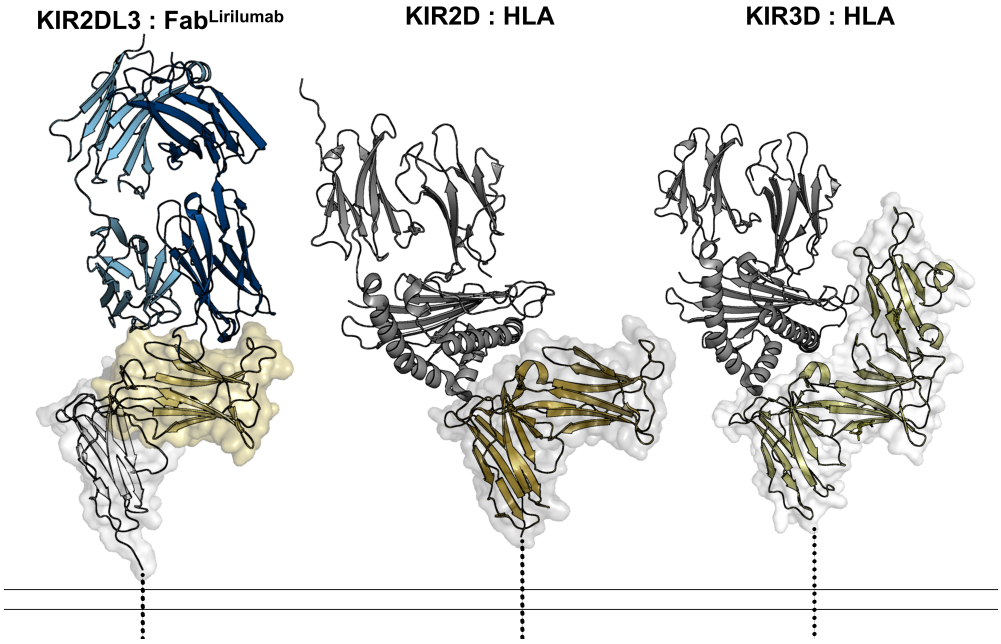


Figure 2.5 The lirilumab epitope overlaps with the HLA binding site but is not exposed on KIR3D class receptors. (Left) KIR2DL3 bound to FabLirulumab, with the FabLirulumab heavy chain colored navy, light chain colored light blue, KIR2DL3 domain 1 colored gold and domain 2 colored white. (Center) KIR2DL1 bound to HLA-Cw4 (PDB 1im9)[38], with HLA colored gray and KIR2DL1 colored bronze. (Right) KIR3DL1 bound to HLA-B*57:03 (PDB 6v3j) [39], with HLA colored gray and KIR3DL1 colored olive. Dashed lines represent the locations of transmembrane domains. There is an approximate 134 Å2 overlap between the lirilumab epitope and the HLA binding site, revealing how lirilumab can block HLA binding. Also, these structures reveal that the lirilumab epitope is occluded by the additional domain of KIR3D receptors, explaining why lirilumab does not bind KIR3D receptors.

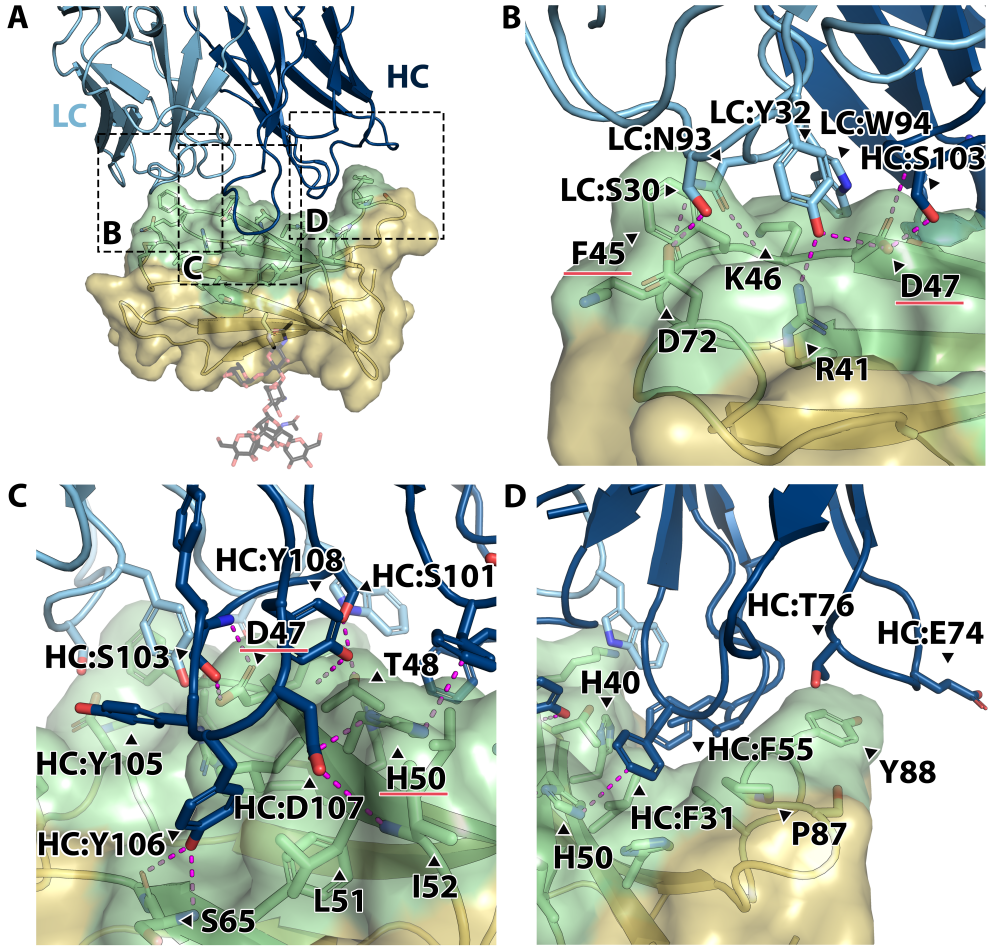


Figure 2.6 Intermolecular interactions at the lirilumab:KIR2DL3 interface. (A) Interface of the FabLirilumab – KIR2DL3 complex, with the FabLirulumab heavy chain colored navy, light chain colored light blue, KIR2DL3 domain 1 colored gold, and KIR2DL3 epitope amino acids colored green. The KIR2DL3 N-glycosylation site on N63 is shown as sticks. Dashed boxes highlight the location of close-up panels B-D. (B) Close-up view of lirilumab light chain interactions with KIR2DL3, with notable hydrogen bonding interactions across the interface. (C) Close-up view of lirilumab heavy chain CDR-H3 loop interactions with KIR2DL3. Note the hydrogen bonding and hydrophobic interactions by the loop residues 103-108 (SYYYDY). (D) Close-up view of lirilumab heavy chain CDR-H1 and -H2 loop interactions with KIR2DL3, which are mostly hydrophobic. In panels B-C, hydrogen bonds are denoted by magenta dashed lines.

Data Collection	**Fab^{lirilumab} : KIR2DL3 complex (PDB ID 8TUI)**
Space group	$P12_11$
a, b, c (Å)	58.38, 84.27, 62.27
α, β, γ (°)	90.00, 115.86, 90.00
Resolution (Å)[a]	50.00-2.75 (2.80 - 2.75)
R_{merge} (%)[a]	0.138 (0.920)
Mean I/σ(I)[a]	14.0 (1.7)
Completeness (%)[a]	98.5 (98.1)
Redundancy[a]	6.5 (5.9)
$CC_{1/2}$	0.94 (0.77)
Refinement	
Resolution (Å)	44.58 - 2.751
Number of reflections	13574
R_{work}/R_{free}[b]	0.2112 / 0.2548
Atoms	4176
Protein	4086
Water	85
Ligand	5
Mean B factor (Å^2)	68.5
Protein	67.96
Water	45.98
Ligand	95.76
R.m.s deviation	
Bond lengths (Å)	0.008
Bond angles (°)	1.13
Ramachandran (%)	
Favored	94.9
Allowed	5.1
Outliers	0

[a] The values in parentheses are for the outermost shell.
[b] Rfree is the Rwork based on 5% of the data excluded from the refinement.

Table 2.4 Crystallography data collection and refinement statistics.

interactions that govern lirilumab's interactions with KIRs (Fig. 2.6). Lirilumab binds to a 954 Å2 conformational epitope on domain 1 of KIR2DL3 that is distal from its single N-linked glycosylation site (Fig. 2.6a). No significant structural changes to KIR2DL3 are observed when in complex with lirilumab compared to its structure alone (RMSD of 0.7 Å2 and TM-score of 0.96 for KIR2DL3 domain 1 structural alignments) [25]. Five CDR loops of lirilumab interact with KIR2DL3, however the heavy chain CDR loops account for the majority of the epitope footprint, covering 659 Å2 of the binding interface. In the light chain, CDR-L1 and -L3 interact with KIR2DL3, making up 295 Å2 of the epitope.

Many of the key interactions at the lirilumab:KIR2DL3 interface form an extensive hydrogen bond network. Lirilumab's light chain S30 and Y32 from CDR-L1 and N93 from CDR-L3 form hydrogen bonds with KIR side chains D72, R41, and D47, as well as K46 backbone (Fig. 2.6b). Lirilumab's heavy chain S101 and S103 from CDR-H3 form hydrogen bonds with KIR2DL3 side chains T48 and D47, respectively (Fig. 2.6c). Also, lirilumab's D107 from CDR-H3 forms hydrogen bonds with KIR2DL3 side chain H50 and I52 backbone (Fig. 2.6c). Finally, two tyrosines, Y106 and Y108, in a surprisingly tyrosine rich heavy chain CDR-H3 along with Y105, form hydrogen bonds with KIR S65 and T48 backbones, respectively (Fig. 2.6c).

In addition to this hydrogen bond network, there are a number of hydrophobic and pi-stacking interactions at the lirilumab:KIR2DL3 interface. First, lirilumab's heavy chain CDR-H1 and -H2 loops nestle within a KIR groove and form pi-stacking interactions between F31 and KIR H50 as well as between F55 and KIR H40 (Fig. 2.6d). Interestingly, lirilumab's heavy chain CDR-H1 and -H2 loops as well as the heavy chain framework region 3 loop containing E74 and T76 also interact with KIR P87 and Y88 (Fig. 2.6d). The heavy chain CDR-H3 Y106 tucks into a hydrophobic pocket formed by KIR L49, L51, and F64 (Fig. 2.6c). Finally, KIR F45 side chain lies between the lirilumab CDR-L1 and -L3 loops while KIR K46 forms a cation-pi interaction with W94 from CDR-L3 (Fig. 2.6b).

To understand the structural basis for lirilumab binding to other inhibitory KIRs, we performed a sequence alignment of KIR2D class receptors and mapped these differences onto the structure (Fig. 2.7). This alignment revealed an intrigu-

ing overlap between the lirilumab:KIR interface and the occurrence of amino acid variation in distinct KIRs (Fig. 2.7c). First, comparison of sequences from different KIR2D class receptors along with literature-reported binding affinities with lirilumab identifies several amino acids that may affect affinity (Fig. 2.7a,c and Table 2.5) [20]. For example, KIR2DL1 has three amino acid differences in the lirilumab epitope (K44 to M44, K46 to N46, and H50 to R50, in KIR2DL3 numbering) yet it has only 2-3-fold lower affinity for lirilumab compared to KIR2DL3, suggesting that those sequence differences do not significantly affect affinity (Fig. 2.7c and Table 2.5). Indeed, the structure reveals that these amino acids are on the periphery of the epitope and could accommodate these side chain differences (Fig. 2.7a). On the other hand, KIR2DS4 also has three amino acid differences in the lirilumab epitope (K46 to N46, D47 to N47, and D72 to V72, in KIR2DL3 numbering), yet it has a 1000-fold lower affinity for lirilumab compared to KIR2DL3. One difference, K46 to N46, is found in KIR2DL1 and is therefore unlikely to significantly affect affinity (Fig. 2.7a). D72 is on the periphery and makes a hydrogen bond with lirilumab's light chain S30 from CDR-L1, and while the change to valine in KIR2DS4 would lose this bond, it is also unlikely to significantly affect affinity. In contrast, D47 is in the center of the epitope, is completely masked by lirilumab, and makes a number of hydrogen-bond interactions with lirilumab (Fig. 2.6b,c). Thus, a difference of D47, a negatively-charged hydrogen-bind-acceptor side chain, to asparagine, an uncharged polar side chain with hydrogen-bond donor and acceptor capabilities, and disruption of hydrogen bond networks, is the most likely reason for the 1000-fold lower affinity of lirilumab for KIR2DS4.

To evaluate population level diversity between KIR variants, we performed a sequence alignment of non-synonymous variants within KIR2D class receptors

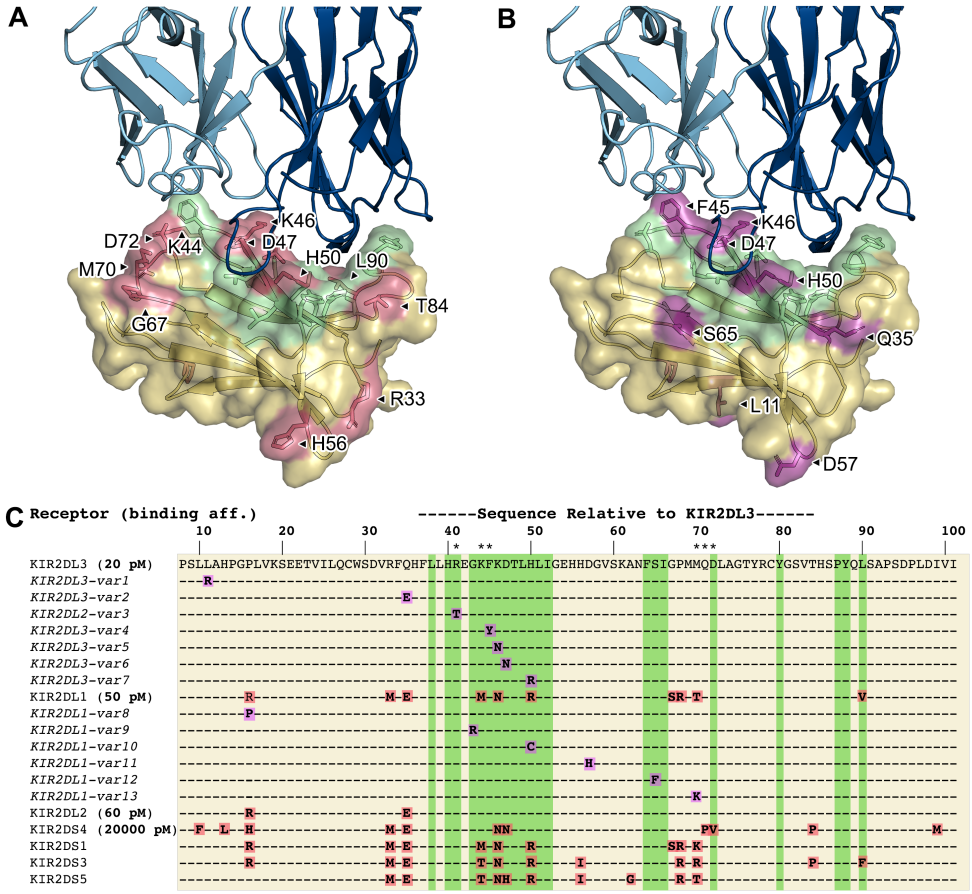


Figure 2.7 Polymorphism between KIR2D class receptors and their variants is often part of the lirilumab epitope. (A,B) Lirilumab-similar binding footprint (green) overlaid with polymorphic sites between KIR2D class receptors (red) and sites of KIR2DL3 and KIR2DL1 variant polymorphisms (magenta). (C) A sequence alignment of KIR2D receptors and variants (from EMBL-EBI IPD-KIR database, release 2.12) that result in amino acid mutations in the KIR domain bound by lirilumab. KIR2DL1 and KIR2DL2 variant mutations are displayed relative to their respective parent receptor sequences, while mutations in all other aligned sequences are relative to KIR2DL3. Sequence numbering corresponds to the KIR2DL3 sequence. Asterisks (*) above the sequence alignment indicate amino acids involved in KIR:HLA binding. Alignment coloring corresponds to the surface colors in panels A and B. Binding affinity constants as previously determined by ELISA (Table 2.5) are shown in bold font next to the KIR2D receptor name. See Table 2.7 for variant accession numbers and listing of other high frequency KIR2D variants.

and mapped these differences onto the structure (Fig. 2.7, for additional variants and their frequencies see supplementary Table S2). This alignment again revealed an intriguing overlap between the lirilumab:KIR interface and the occurrence of amino acid variation in distinct KIRs (Fig. 2.7c) [26]. First, we observed a KIR2DL3 D47N variant, which is the same amino acid change that we hypothesize is the reason for the greatly reduced affinity of KIR2DS4 for lirilumab (described above)(Fig. 2.7b). In addition, there is a KIR2DL3 F45Y variant, which would extend the phenyl ring with a hydroxyl group that we hypothesize could affect binding by clashing or forcing a shift of the nearby S28 on the lirilumab CDR-L1 (Fig. 2.7b). Finally, we observed a KIR2DL3 H50R variant. H50 interacts with residues in lirilumab's CDR-H1 and -H3 loops (Fig. 2.6c, 2.7b), and while change to an arginine is not expected to sterically clash, it may affect these hydrogen bonding and pi-stacking interactions with lirilumab.

To test these observations, we produced several KIR2DL3 proteins with variant mutations in the lirilumab epitope and tested them for binding to lirilumab (IgG1) utilizing biolayer interferometry (BLI) (Fig. 2.8, supplementary Fig. 2.9, Table 2.6). Dissociation constants are compared to reported values (Table 2.5) [19,20]. By BLI, a dissociation constant (KD) of 2.8 nM was determined for the interaction of lirilumab (IgG1) and wild-type KIR2DL3, which is consistent with the published value of 3.5 nM determined by surface plasmon resonance (SPR). In contrast, the affinity determined for the interaction of lirilumab (IgG1) and KIR2DL3 D47N was dramatically reduced, with a KD of 1650 nM and a notably faster off-rate (Table 2.5, Fig. 2.8). Thus, the single functional group change from carboxylic acid to amide dramatically disrupted the lirilumab epitope. In addition, both the KIR2DL3 F45Y and H50R variant mutations modestly weakened antibody binding by 3-4-fold (Table 2.5, Fig. 2.8). The lirilumab affinity for

	K_D (nM)		
	SPR[a]	*ELISA*[b]	*BLI*[c]
KIR2DL1	10.3	0.05	10.0[d]
KIR2DL2		0.06	--
KIR2DL3	3.5	0.02	2.8
KIR2DS4		20	--
KIR2DL3 F45Y			9.9
KIR2DL3 D47N			1650[e]
KIR2DL3 H50R			11.2

[a]Surface plasmon resonance (SPR) values reported in [19] using immobilized antibody 1-7F9 and KIR2D class receptor as the analyte.
[b]ELISA K_D values as reported in [20]. KIR2D class receptors were adhered to the ELISA plate, after which lirilumab was added at different concentration, allowing for enhanced binding by avidity. KIR2DS4 measurement was performed with a KIR2DS4-Fc fusion construct.
[c]Except where otherwise noted, K_Ds were determined by global fitting of a 1:1 langmuir binding model to two-fold dilution series data. Values reported are the average of 3 technical replicates.
[d]Commercially sourced KIR2DL1 (Sino Biological 13145-H08H).
[e]K_D constant determined by steady state model. In samples where a 1:1 binding model fit well, steady state K_D and Langmuir-derived K_D were near-equivalent values.

Table 2.5 Binding affinities between Lirilumab (IgG1) and KIR receptors.

KIR2DL3 H50R is consistent with the affinity for KIR2DL1, which also contains the H50R mutation (Table 2.5, Fig. 2.8). While our data are very similar to published SPR affinities, we note that there are significant differences with published ELISA affinities (Table 2.5). This difference is most likely due to different methodologies, with ELISA experiments performed by coating plates with KIR proteins and binding with lirilumab antibody, resulting in lower KD values due to contributions of antibody avidity. Nevertheless, our BLI data are consistent to the trends in ELISA data, whereby KIR2DL3 has the highest affinity, and the KIR2DL3 D47N variant mutation that is also found in KIR2DS4 results in a dramatically reduced affinity for lirilumab.

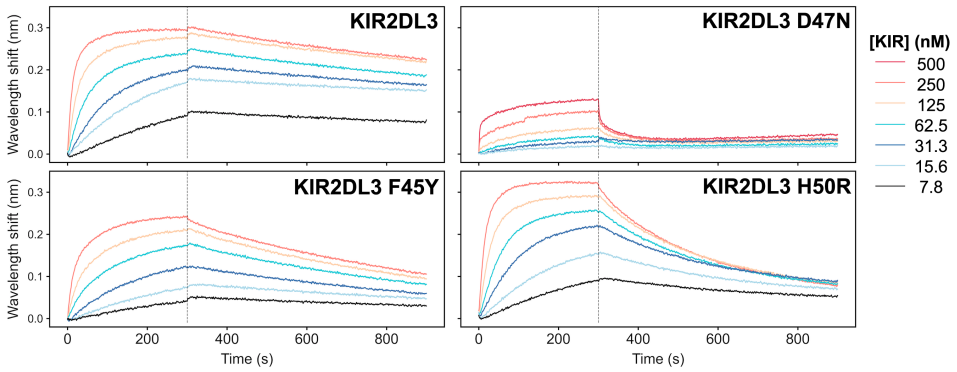


Figure 2.8 The high-affinity interaction between lirilumab and KIR2DL3 is affected by polymorphisms present in KIR2DL3 variants and in KIR2DS4. Biolayer interferometry kinetics plots are shown for lirilumab (IgG1) and three KIR2DL3 mutants in neutral pH conditions. In each assay, IgG1 lirilumab was loaded onto AHC sensor tips and equilibrated to assay buffer prior to analyte binding. Each data series consists of 6 two-fold dilutions starting at an analyte concentration of 250 nM (except KIR2DL3 D47N, where the peak concentration is 500 nM). Representative plots from among three assay replicates are shown.

2.2.5 Discussion

In therapeutic targets with significant population-level genetic diversity, the structural determination of binding epitopes is an important consideration, allowing more precise administration of therapeutics when paired with patient genome profiling, or prioritization during the discovery stage, enabling selection of a more broad-spectrum therapeutic candidate. The KIR family of receptors are relatively variable and an individual's NK cells may include receptors bearing polymorphisms or may only express a subset of receptors from an individual's repertoire [11]. This variability could result in a situation where therapeutic antibodies targeting KIRs may not be equally effective for all individuals. Here, we have defined the epitope for lirilumab, a pan-KIR2D immune checkpoint inhibitor currently in phase II clinical trials, and we have uncovered KIR2D variants that may influence

its clinical efficacy.

We focused our binding studies on variants within the primary inhibitory KIRs, KIR2DL1 and KIR2DL2/3, that fall within the epitope of lirilumab. First, we found that the KIR2DL3 D47N variant has a dramatic >500-fold weaker affinity for lirilumab compared to the wild-type KIR2DL3, suggesting that individuals with this variant may respond differently to lirilumab therapy. While the D47N mutation is rare (6 in 1000 in East Asian population), it is an important reminder of how an individual's genomic variability could influence therapeutic outcomes (see table 2.7 for a list of inhibitory KIR variants and global population frequencies). We also investigated the KIR2DL3 F45Y and H50R variants that are much more common (5 to 15% of the population overall, respectively, with higher frequencies in some sub populations), and we observed 3-4-fold reductions in affinity for lirilumab, which may also be able to influence patient response to lirilumab therapy [27]. Although there are several other inhibitory KIR2D variants within the lirilumab epitope that could be explored, such as KIR2DL3 R41T, K46N, KIR2DL1 G43R, R50C, S65F, and T70K (all in KIR2DL3 numbering)(Fig. 2.7c), our structural observations suggest that these variants will have modest or no effects on lirilumab binding. In addition to polymorphisms directly within the lirilumab epitope, there is evidence that mutations distal to the epitope site can affect KIR:HLA binding, putatively through changes in secondary and tertiary structure, indicating distal variants may warrant additional investigation [28].

Despite demonstrating safety in pre-clinical and phase I clinical trials, phase II trials involving lirilumab often did not demonstrate significant improvement over the standard of care (clinical trial IDs NCT02399917, NCT01687387, NCT01592370)

29. Our studies support that patient genome profiling for inhibitory KIR receptor variation may enable patient stratification for individuals with a matching lirilumab epitope and improve outcomes. Our structure also provides a roadmap to optimize lirilumab affinity and/or specificity and alter clinical outcomes. For example, affinity maturation of the lirilumab light chain CDR-L1 and -L2 loops may increase affinity and epitope overlap with the HLA binding site (Fig. 2). However, we caution that a higher-affinity anti-KIR antibody could be detrimental, as there is evidence of inhibitory receptor down-regulation when KIRs are continuously occupied by an antibody and unable to signal [10]. This prolonged KIR blockade and reduction in KIR expression may result in NK cell anergy [10]. One potential solution is to combine anti-KIR antibody therapy with stimulatory cytokines [10]. Another possible solution that is enabled by our studies is the engineering of lirilumab to have a weaker affinity, thereby increasing NK cell's cytotoxic activity while still allowing intermittent HLA sampling by inhibitory KIRs. Engineered versions of lirilumab, as well other anti-KIR antibodies with differing affinities and KIR specificities, could be used to explore these hypotheses further in pre-clinical studies [20].

The clear importance of NK cell activity in tumor clearance and disease-free survival rates remains a strong motivator to develop therapeutics that can modulate NK cell activity. Our studies reveal general concepts for the development of inhibitory KIR-targeting immune checkpoint inhibitors, most notably epitope evaluation and deep investigation of population scale variation, which could lead to improved clinical outcomes through precision application of next generation antibody therapeutics.

2.2.6 Methods

Study Design

This study was performed to investigate the epitope of the KIR-blocking monoclonal antibody lirilumab to support improved therapeutic administration parameters and/or future development of novel antibodies that may have different or improved therapeutic characteristics. We hypothesized that some feature of the lirilumab epitope may be contributing to its underwhelming performance in stage II clinical trials, either through inadequate blocking of the HLA binding site or overlap with disruptive polymorphic sites present in patients' inhibitory KIR receptor repertoire. In vitro binding experiments involved technical replicates performed in triplicate. Various Fablirilumab – KIR2DL3 complexes were subjected to crystallographic screening under the assumption that molecular replacement methodologies would allow structure determination, as structures of KIR2DL3 and human antibody Fab fragments had been solved previously. The successful structure determination then informed design of KIR receptor mutants and subsequent kinetics assays. Experiments were not blinded and experimenters were aware of experimental conditions.

Expression and purification of recombinant KIR2DL3

The expression plasmid for recombinant KIR2DL3 is a derivative of pcDNA 3.1 containing a CMV promoter, CCR5 signal sequence, and a codon-optimized gene encoding the human KIR2DL3 ectodomain (residues 1-224, RefSeq: NP_056952.2) fused to a C-terminal c-Myc tag and a 6-histidine tag, similar to that expressed by Maenaka et al 30. Mutant KIR plasmids were generated by site-directed mutagenesis. The plasmids were verified by Sanger sequencing. Sterile, endotoxin free

expression plasmid DNA was produced using the Invitrogen Plasmid Maxiprep Kit (Invitrogen A31231).

CHO-S (Thermo Fisher R80007) cells were electroporated in the presence of the KIR2DL3 expression plasmid using a MaxCyte electroporator with OC-400 electroporation cuvettes (MaxCyte SOC4) following the manufacturer's instructions, enabling transient expression of the recombinant KIR2DL3 protein. After transfection, cells were grown with OptiCHO Expression media (Thermo Fisher 12681011, supplemented with 8mM L-glutamine, 0.1 mM Sodium Hypoxanthine, 16 μ M thymidine) in shake flasks using a Khuner shaker incubator at 32 °C, 8% CO2, and 85% humidity. The day after transfection, filter sterilized sodium butyrate was added to the cell culture media to a final concentration of 1 mM. Each day after transfection, cells were fed with 3% volume of an enriched feed media (containing 7 mM L-glutamine, 5.5% glucose, and 23.4g/L yeastolate in CHO CD Efficient Feed). Approximately six days after electroporation (or when cell viability drops below 60%), media and cells were centrifuged at 6000 g for 10 minutes and the supernatant was decanted and 0.22-micron filtered.

An appropriate volume of Elution buffer (20 mM MOPS, 150 mM NaCl, 200 mM imidazole, pH 7.5) was added to the filtered CHO-S supernatant to bring total imidazole concentration to 10 mM. Using an AKTA Pure FPLC instrument, a HisTrap FF column (Cytvia 17531901) was equilibrated with Wash buffer (20 mM MOPS, 150 mM NaCl, 20 mM imidazole, pH 7.5), and the CHO-S supernatant was loaded, washed with 10 column volumes (CV) Histidine Wash buffer, and KIR2DL3 protein was eluted using a 5 CV linear gradient to 100% Elution buffer. The elution was monitored by absorbance at 280nm and the elution fractions

corresponding to the 280nm peak were assessed by SDS-PAGE. The purest elution fractions were combined, concentrated in a 10kD spin concentrator, and dialyzed into 20 mM MOPS, 150 mM NaCl pH 7.5.

Expression and purification of recombinant lirilumab (IgG1) and FabLirilumab

Synthetic cDNA encoding the heavy and light chain variable regions of lirilumab (GSRS UNII identifier S9XDI9W918) was cloned by Gibson Assembly into the pCMV-VRC01 antibody heavy and light chain plasmids, in place of the variable regions of antibody VRC01, a human anti-HIV IgG1 antibody targeting the gp120 protein [31]. For the FabLirilumab heavy chain, only the variable region and the constant heavy 1 region ending with residues DKKVEPKSC was included, followed by an alanine-serine linker, C-terminal thrombin cleavage site, and a Twin Strep-tag. All antibody sequences were in-frame with an N-terminal signal sequence. The plasmids were verified by Sanger sequencing. Sterile, endotoxin free expression plasmid DNA was produced using the QIAGEN Plasmid Maxi Kit (QIAGEN Cat. No. 12163).

HEK 293F Freestyle (ThermoFisher R79007) cells were electroporated in the presence of the lirilumab heavy and light chain expression plasmids using a MaxCyte electroporator with OC-400 electroporation cuvettes (MaxCyte SOC4) following the manufacturer's instructions. After transfection, cells were grown in Freestyle 293 Expression media (Gibco 12338-018) in vented shake flasks using a Khuner shaker incubator at 37 °C, 8% CO2, and 85% humidity. Four days after electroporation, media and cells were centrifuged at 6000 g for 10 minutes and the supernatant was decanted and 0.22-micron filtered.

For purification of lirilumab (IgG1), filtered supernatant was incubated with protein A resin (Pierce 20334), eluted with glycine pH 3.1, and immediately neutralized with 1M Tris pH 9.0. Lirilumab (IgG1) was dialyzed into phosphate buffered saline pH 7.4 and purity was assessed by SDS-PAGE.

For purification of FabLirilumab, filtered 293F supernatant was diluted 1:1 with Strep Wash buffer (20 mM MOPS, 150 mM NaCl, 1 mM EDTA, pH 7.5) and 1.6 mL of Biolock (IBA Life Sciences 2-0205-050) was added per liter of 293F media. The diluted supernatant was loaded onto a StrepTrap column (GE Healthcare 28-9075-47) equilibrated with Strep Wash buffer, the column was washed with 10 CV Strep Wash buffer, and the FabLirilumab was eluted with 8 CV Elution buffer (20 mM MOPS, 150 mM NaCl, 1mM EDTA, 2.5 mM desthiobiotin, pH 7.5). Eluant fractions with high 280nm absorbance were combined, concentrated in a 10kD spin concentrator, and dialyzed into 20 mM MOPS, 150 mM NaCl pH 7.5.

Structure determination of the FabLirilumab – KIR2DL3 complex

Purified, dialyzed FabLirilumab and KIR2DL3 were combined in a 1.2:1 molar ratio and incubated at 4 °C for 1 hour to allow complex formation. The mixture was injected onto a gel filtration column (Superdex 200 10/300 GL, GE Healthcare 28-9909-44) pre-equilibrated in 20 mM MOPS, 150 mM NaCl pH 7.5 and fractions were collected. The contents of fractions were assessed by SDS-PAGE, and fractions containing both FabLirilumab and KIR2DL3 were pooled and concentrated in a 10kD spin concentrator to 7.5 mg/ml.

A sitting drop crystal screen was set up using the concentrated FabLirilumab – KIR2DL3 complex where each drop contained a 1:1 mixture of the complex and the screening condition buffer. After several weeks crystals were observed in a condition containing 200 mM Potassium Sodium Tartrate and 20%(w/v) PEG 3350. The crystal was transferred to a cryoprotectant solution of 0.2 M potassium sodium tartrate, 20% (w/v) PEG 3350, 6% ethylene glycol, 6% DMSO, and 6% glycerol and then flash frozen in liquid nitrogen. Diffraction data were collected at cryogenic temperature at the Advanced Photon Source on beamline 23-ID-D using a wavelength of 1.03 Å.

Diffraction data from a single crystal were processed with HKL2000 [32]. The FabLirilumab – KIR2DL3 complex structure was solved by molecular replacement using a Fab structure extracted from PDB 6XOC [33] and KIR2DL3 domains extracted from PDB 1B6U 30 with the PHASER [34] program. The structure was then partially rebuilt with the auto-build program within the PHENIX suite and completed manually using Coot followed by refinement using PHENIX.refine and Coot [35,36].

Biolayer Interferometry (BLI) Binding Assays

Wild-type and mutant KIR proteins used in biolayer interferometry (BLI) assays were produced similarly to the methods described above, except they were expressed in HEK 293F cells instead of CHO-S cells. BLI assays were performed with an Octet RED384 instrument with temperature set to 25 °C and shaking at 1000 rpm for all assays. The assay buffer used was composed of 20 mM MOPS (pH 7.5 at 22 °C), 150 mM NaCl, 0.1% BSA, and 0.02% Tween-20. Anti-human IgG biosensors (AHC) were hydrated in assay buffer for at least 10 minutes before the

start of each BLI experiment. Assays were performed in either black tilted-bottom 384-well plates or black flat-bottomed 96-well plates which using 40 μ L or 200 μ L of sample, respectively. To determine binding affinities of lirilumab (IgG1) for each KIR receptor protein, BLI experiments were performed as follows: (1) pre-hydrated AHC biosensors were dipped into assay buffer for 120 s to establish a baseline; (2) biosensor were dipped into 7.5 μ g/mL lirilumab (IgG1) in assay buffer for 60 s to load the antibody onto the biosensors; (3) biosensors were dipped into assay buffer for 60 s to confirm stable antibody loading and establish a new baseline; (4) lirilumab (IgG1)-loaded biosensors were dipped in wells containing a concentration series of KIR protein in assay buffer for 300 s to determine the association rate (5) biosensors are dipped in assay buffer for 600 s to determine the dissociation rate. No-KIR-antigen and no-antibody controls were performed so baseline signal drift could be accounted for and to confirm that non-specific binding does not occur between antigens and biosensors. A global association 1:1 model was used to fit at least 4 curves per replicate to determine the on- and off-rates and calculate the dissociation constant (KD). All binding assays were performed in technical triplicates, and the average KD of the three replicates are reported in Table 2.5. Curve fitting was unsuccessful for the KIR2DL3 D47N mutant, preventing KD determination with on- and off-rates. Instead, a steady state equilibrium model was utilized.

Sequence alignment of KIR receptors and variants

Protein sequences used in the KIR receptor alignment were downloaded from RefSeq's non-redundant protein database: KIR2DL1 (NP_055033.2), KIR2DL2 (NP_055034.2), KIR2DL3 (NP_056952.2), KIR2DS1 (NP_055327.1), KIR2DS3 (NP_036445.1),

KIR2DS4 (NP_036446.3), KIR2DS5 (NP_055328.2). Sequences were aligned with MAFFT using the L-INS-I strategy and BLOSUM62 substitution matrix [37]. KIR receptor variants were identified and their occurrence estimated by cross-referencing the IPD-KIR database with NCBI's dbSNP via the variation viewer browser [26]. The SNP accession numbers corresponding to KIR2DL3 polymorphic sites are rs528413442 (D47N), rs202032116 (L11R), rs35719984 (Q35E), R41T (rs76843526), rs78713511 (F45Y), rs145638569 (K46N), rs138897134 (H50R); accession numbers for KIR2DL1 sites are rs35509911 (R16P), rs1481853508 (G43R), rs375476159 (R50C), rs760965171 (D57H), rs765722009 (S65F), rs687485 (T70K). A table listing additional high frequency variants is shown in table 2.7. The 1000Genomes_30x study was the preferred source of allele frequency data in this study [27].

2.2.7 Supplemental Figures and Tables

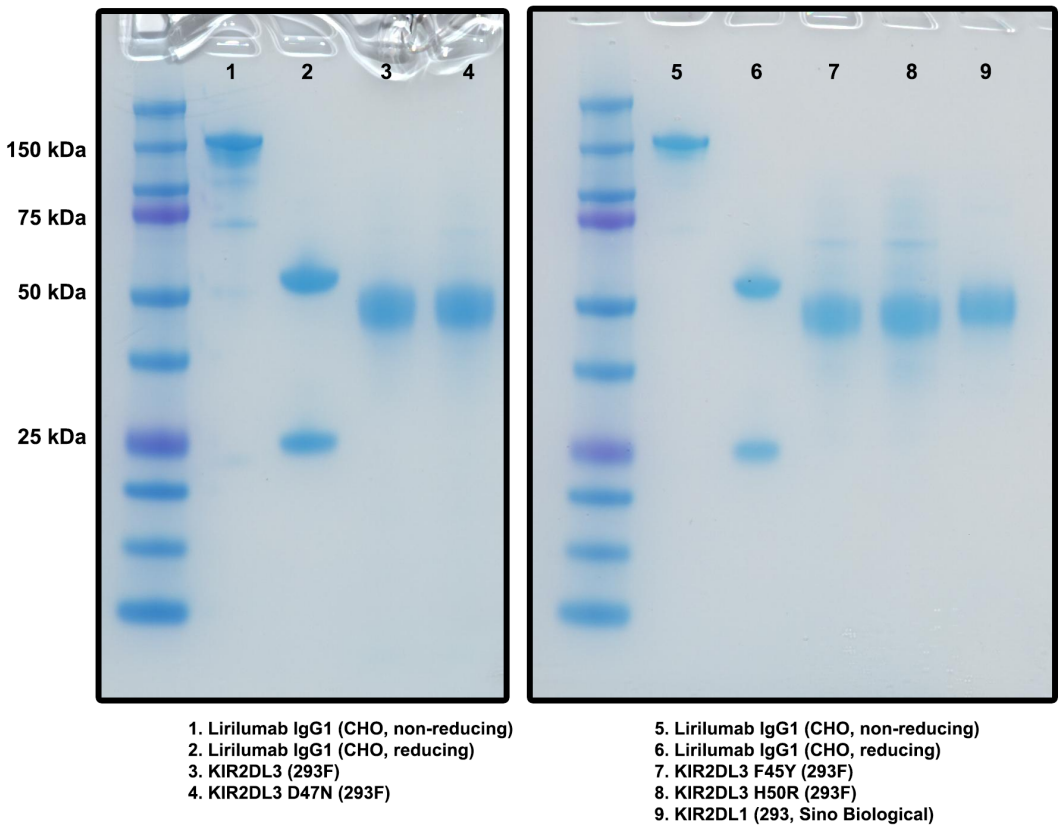


Figure 2.9 SDS-PAGE gels showing lirilumab IgG1, KIR receptors used for BLI assays. Lirilumab IgG1 was purified from CHO-S cell supernatant using protein A resin. KIR receptors were purified from HEK293F supernatant using EDTA-resistant Ni-NTA followed by size exclusion chromatography.

	Model Fit Parameters				
	K_D (nM)[a]	R^2	X^2	k_a (1/Ms)	k_d (1/s)
KIR2DL3	1.64	0.9952	0.5141	2.240E05	3.663E-04
	2.83	0.9954	0.3523	2.197E05	6.219E-04
	3.78	0.9959	0.1989	2.363E05	8.920E-04
KIR2DL1	11.5	0.9912	1.0485	8.319E04	9.537E-04
	9.6	0.9932	0.7627	1.038E05	9.946E-04
	8.8	0.9942	0.6410	1.074E05	9.475E-04
KIR2DL3 F45Y	10.25	0.9932	0.6602	1.218E05	1.249E-03
	10.8	0.9943	0.4065	1.437E05	1.554E-03
	8.5	0.9895	0.3643	1.705E05	1.450E-03
KIR2DL3 D47N[b]	1400	0.9991	0	--	--
	1900	0.9973	0.0001	--	--
KIR2DL3 H50R	7.98	0.9795	2.12	2.442E05	1.948E-03
	10.6	0.9818	1.311	2.308E05	2.437E-03
	15	0.9774	2.3512	2.116E05	3.175E-03

[a]K_D determined using Octet Data Analysis HT software's 1:1 binding model with the 'global' fit setting and using at least 4 curves for each experiment
[b]K_D determined using a steady state binding model due to poor fitting of the 1:1 model

Table 2.6 BLI kinetics model parameters for binding between lirilumab (IgG1) and KIR2DL1, KIR2DL3, and KIR2DL3 variants.

Var#[a]	KIR	Variant accession number	Genome location (GRCh38.p14)	Mutation	Minor Allele Frequency, Global (1000genomes_30x)
	KIR2DL1	rs666590	54775402	A>G = H182R	0.291
	KIR2DL1	rs147072532	54775344	G>A = D163N	0.277
	KIR2DL1	rs111799279	54775317	C>A = P154T	0.276
	KIR2DL1	rs11673144	54775198	T>C = L114P	0.266
	KIR2DL1	rs199644153	54775213	T>A = L119N	0.261
	KIR2DL1	rs597681	54778656	A>G = K216E	0.220
	KIR2DL1	rs687885	54773593	G>T = V90L	0.210
	KIR2DL1	rs574769826	54775225	A>G = N123S	0.199
Var13	KIR2DL1	rs687485	54773534	C>A = T70K	0.194
	KIR2DL1	rs2916003	54775357	G>A = G167D	0.179
	KIR2DL1	rs200746024	54775315	G>T = G153V	0.158
	KIR2DL1	rs538559888	54775318	C>G = P154R	0.155
	KIR2DL1	rs34721508	54783002	C>T = R245C	0.155
Var8	KIR2DL1	rs35509911	54773372	G>C = R16P	0.155
	KIR2DL1	rs62121640	54775329	G>A = G158R	0.151
	KIR2DL1	rs2756923	54783763	A>G = T312A	0.151
	KIR2DL1	rs2304224	54769863	Signal sequence	0.146
	KIR2DL1	rs202205602	54775305	C>T = L150F	0.065
Var10	KIR2DL1	rs375476159	54773473	C>T = R50C	No data
Var9	KIR2DL1	rs1481853508	54773452	G>A = G43R; G>C = G43R	No data
Var11	KIR2DL1	rs760965171	54773494	G>C = D57H	No data
Var12	KIR2DL1	rs765722009	54773519	C>T = S65F	No data
	KIR2DL3	rs3810343	54738571	Signal sequence	0.361
Var2	KIR2DL2/3	rs35719984	54742075	C>G = Q35E	0.340
	KIR2DL3	rs1049267	54752446	G>A = R297H; G>C = R297P	0.236 (A) 0.160 (C)
	KIR2DL2/3	rs200686594	54743929	C>T = R148C	0.194
	KIR2DL2/3	rs150145497	54744086	C>T = T200I	0.149
Var4	KIR2DL2/3	rs78713511	54742106	T>A = F45Y	0.135
	KIR2DL2/3	rs1555897648	54742019	C>G = P16R	0.134
	KIR2DL2/3	rs201504521	54743909	G>C = R141T	0.120
	KIR2DL3	rs35861855	54747356	C>T = P208L	0.115
	KIR2DL2/3	rs200359536	54744000	C>A = H171N	0.096
	KIR2DL2/3	rs200426472	54743878	C>T = R131W	0.096
	KIR2DL3	rs2365227	54752222	T>C = V255A	0.085
	KIR2DL3	rs4020198	54752265	G>C = R279S	0.079
Var7	KIR2DL2/3	rs138897134	54742121	A>G = H50R	0.057
Var1	KIR2DL2/3	rs202032116	54742004	T>G = L11R	0.053
Var3	KIR2DL2/3	rs76843526	54742094	G>C = R41T	0.006
Var6	KIR2DL2/3	rs528413442	54742111	G>A = D47N	0.001
Var5	KIR2DL2/3	rs145638569	54742110	G>C = K46N	0.0002

Table 2.7 High frequency (>5%) inhibitory KIR2D receptor missense variants and other variants of interest.

2.2.8 References and Notes

1. SEER*Explorer: An interactive website for SEER cancer statistics. https://seer.cancer.gov/statistics-network/explorer/ (2023).

2. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. Nat Rev Cancer 12, 278–287 (2012).

3. Zahavi, D. & Weiner, L. Monoclonal Antibodies in Cancer Therapy. Antibodies (Basel) 9, 34 (2020).

4. Marin-Acevedo, J. A., Kimbrough, E. O. & Lou, Y. Next generation of immune checkpoint inhibitors and beyond. Journal of Hematology & Oncology 14, 45 (2021).

5. de Miguel, M. & Calvo, E. Clinical Challenges of Immune Checkpoint Inhibitors. Cancer Cell 38, 326–333 (2020).

6. Galluzzi, L., Humeau, J., Buqué, A., Zitvogel, L. & Kroemer, G. Immunostimulation with chemotherapy in the era of immune checkpoint inhibitors. Nat Rev Clin Oncol 17, 725–741 (2020).

7. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune Checkpoint Blockade: A Common Denominator Approach to Cancer Therapy. Cancer Cell 27, 450–461 (2015).

8. Paul, S. & Lal, G. The Molecular Mechanism of Natural Killer Cells Function and Its Importance in Cancer Immunotherapy. Front. Immunol. 8, 1124 (2017).

9. Baldwin, X. L., Spanheimer, P. M. & Downs-Canner, S. A Review of Immune Checkpoint Blockade for the General Surgeon. Journal of Surgical Research 281, 289–298 (2023).

10. Muntasell, A. et al. Targeting NK-cell checkpoints for cancer immunotherapy. Current Opinion in Immunology 45, 73–81 (2017). 11. Uhrberg, M. et al. Human

diversity in killer cell inhibitory receptor genes. Immunity 7, 753–763 (1997).

12. Trowsdale, J. Genetic and functional relationships between MHC and NK receptor genes. Immunity 15, 363–374 (2001).

13. Valés-Gómez, M., Reyburn, H. T., Mandelboim, M. & Strominger, J. L. Kinetics of Interaction of HLA-C Ligands with Natural Killer Cell Inhibitory Receptors. Immunity 9, 337–344 (1998).

14. Maenaka, K. et al. Killer Cell Immunoglobulin Receptors and T Cell Receptors Bind Peptide-Major Histocompatibility Complex Class I with Distinct Thermodynamic and Kinetic Properties*. Journal of Biological Chemistry 274, 28329–28334 (1999).

15. A, M., F, L. & L, M. Human NK cells: from HLA class I-specific killer Ig-like receptors to the therapy of acute leukemias. Immunological reviews 224, (2008).

16. Foley, B. et al. The biology of NK cells and their receptors affects clinical outcomes after hematopoietic cell transplantation (HCT). Immunological Reviews 258, 45–63 (2014).

17. Moretta, A. et al. P58 molecules as putative receptors for major histocompatibility complex (MHC) class I molecules in human natural killer (NK) cells. Anti-p58 antibodies reconstitute lysis of MHC class I-protected cells in NK clones displaying different specificities. J Exp Med 178, 597–604 (1993).

18. Vahlne, G. et al. In vivo tumor cell rejection induced by NK cell inhibitory receptor blockade: Maintained tolerance to normal cells even in the presence of IL-2. European Journal of Immunology 40, 813–823 (2010).

19. Moretta, A. et al. Antibodies Binding to Receptors Kir2dl1, -2, 3 but Not Kir2ds4 and Their Therapeutic Use. (2006).

20. Ryser, S., Estellés, A., Tenorio, E., Kauvar, L. M. & Gishizky, M. L. High affinity anti-TIM-3 and anti-KIR monoclonal antibodies cloned from healthy hu-

man individuals. PLOS ONE 12, e0181464 (2017).

21. Romagné, F. et al. Preclinical characterization of 1-7F9, a novel human anti–KIR receptor therapeutic antibody that augments natural killer–mediated killing of tumor cells. Blood 114, 2667–2677 (2009).

22. Benson, D. M., Jr et al. IPH2101, a novel anti-inhibitory KIR antibody, and lenalidomide combine to enhance the natural killer cell versus multiple myeloma effect. Blood 118, 6387–6391 (2011).

23. Carlsten, M. et al. Checkpoint Inhibition of KIR2D with the Monoclonal Antibody IPH2101 Induces Contraction and Hyporesponsiveness of NK Cells in Patients with Myeloma. Clin Cancer Res 22, 5211–5222 (2016).

24. Vey, N. et al. A phase 1 study of lirilumab (antibody against killer immunoglobulin-like receptor antibody KIR2D; IPH2102) in patients with solid tumors and hematologic malignancies. Oncotarget 9, 17675–17688 (2018).

25. Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. Nat Methods 19, 1109–1115 (2022).

26. Robinson, J., Halliwell, J. A., McWilliam, H., Lopez, R. & Marsh, S. G. E. IPD—the Immuno Polymorphism Database. Nucleic Acids Res 41, D1234–D1240 (2013).

27. Byrska-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell 185, 3426-3440.e19 (2022).

28. Moesta, A. K. et al. Synergistic Polymorphism at Two Positions Distal to the Ligand-Binding Site Makes KIR2DL2 a Stronger Receptor for HLA-C Than KIR2DL31. The Journal of Immunology 180, 3969–3979 (2008).

29. Wong, J. K. M., Dolcetti, R., Rhee, H., Simpson, F. & Souza-Fonseca-

Guimaraes, F. Weaponizing natural killer cells for solid cancer immunotherapy. Trends in Cancer 9, 111–121 (2023).

30. Maenaka, K., Juji, T., Stuart, D. I. & Jones, E. Y. Crystal structure of the human p58 killer cell inhibitory receptor (KIR2DL3) specific for HLA-Cw3-related MHC class I. Structure 7, 391–398 (1999).

31. Wu, X. et al. Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. Science 329, 856–861 (2010).

32. Otwinowski, Z. & Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. in Methods in Enzymology vol. 276 307–326 (Academic Press, 1997).

33. Seydoux, E. et al. Development of a VRC01-class germline targeting immunogen derived from anti-idiotypic antibodies. Cell Reports 35, (2021).

34. McCoy, A. J. et al. Phaser crystallographic software. J Appl Cryst 40, 658–674 (2007).

35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Cryst D 66, 486–501 (2010).

36. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Cryst D 75, 861–877 (2019).

37. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20, 1160–1166 (2019).

38. Fan, Q. R., Long, E. O. & Wiley, D. C. Crystal structure of the human natural killer cell inhibitory receptor KIR2DL1–HLA-Cw4 complex. Nat Immunol 2, 452–460 (2001).

39. Saunders, P. M. et al. The molecular basis of how buried human leukocyte

antigen polymorphism modulates natural killer cell function. Proceedings of the National Academy of Sciences 117, 11636–11647 (2020).

2.2.9 Acknowledgements

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The authors thank Dr. John Dzimianski for assistance with crystallographic data processing.

2.2.10 Funding

A Santa Cruz Cancer Grant to R.M.D. funded this research. N.L.R. was supported by a NIH T32 training grant (5T32HG008345-05) and a UCSC Baskin Engineering dissertation year fellowship. N.H. was supported by a Doug Drexler Chemistry Scholarship. Funding for the purchase of the Octet RED384 instrument was supported by the NIH S10 shared instrumentation grant 1S10OD027012-01.

Chapter Three

Toward Structural and Biochemical Characterization of the HAstV +ssRNA RNA-dependent-RNA-polymerase

3.1 Preface

Sometime after our BLI-ISA papers were wrapped up and the lirilumab story was still in the crystal screening phase, I was interested in developing additional projects. Being a structure-focused lab, the approach I took was to screen viral genomes, looking for regions containing predicted proteins that did not intersect with any protein data bank (PDB) entries. Next, secondary structure predictions from those regions were used to again query the PDB to assess whether there were related structures that could be learned from or whether the predicted protein contained putatively ordered domains at all. Among the (apparently) low-hanging fruit identified this way was the astrovirus RNA-dependent RNA-polymerase – a

great fit for the lab, which is already funded to do human astrovirus research, and the RdRp is an attractive therapeutic target on top of the prospect of structural characterization.

3.2 Introduction and Taxonomy of +ssRNA viruses

Astrovirus is a classified as a member of Group IV single-stranded positive-sense RNA viruses. This group encompass a diverse group of viruses that include numerous human and animal pathogens 3.1. Within the group are several families, including Picornaviridae, Togaviridae, Coronaviridae, Hepeviridae, Caliciviridae, Flaviviridae, and Astroviridae [2]. These viruses share numerous genomic and proteomic features, but are far from uniform. The often poly-adenylated genomes of these viruses RNA can serve as mRNA directly, allowing for immediate translation of the viral genome upon entry into host cells, often not requiring anything but the RNA molecule (and sometimes VPg) to competently replicate new viral particles [2]. Viruses in the Astroviridae, Picornaviridae, Togaviridae, and Flaviviridae families translate their genome in the form of large polyprotein precursors that are processed into individual functional proteins by a combination of viral and host proteases. The +ssRNA virus group contain both enveloped (Coronaviridae, Flaviviridae, Togaviridae) and non enveloped families (Astroviridae, Caliciviridae, Hepeviridae, and Picornaviridae).

Viruses in this group that cause human disease cover a broad array of presentations, though there are several recurring themes: gastroenteritis (Astrovirus, Torovirus, Norovirus), encephalitis (alphaviruses in the Semliki forest virus complex, Japanese encephalitis virus, many tick-borne viruses), flu-like illness (Rhi-

novirus, Coxsackievirus, SARS variants), and hepatitis (Hepatitis A, C, E, and G). Perhaps the most significant virus in the group is poliovirus, an enterovirus from the Picornaviridae family that can 'escape' its typical replication environment in the gastrointestinal tract and infect the central nervous system, leading to paralytic poliomyelitis. Other members of the group are major contributors to endemic colds and gastroenteritis (e.g. rhinoviruses, and coronaviruses; norovirus, astrovirus) [42].

3.2.1 Structures of Group IV Viral Polymerases

The feature shared by all +ssRNA viruses is some form of RNA-dependent RNA-polymerase (RdRp) that is responsible for genome replication. Although primary sequence identity between polymerases can be extraordinarily low, the polymerase fold is highly conserved across the diverse group of viral families [38, 39]. The astrovirus RdRp is predicted to have a high degree of similarity to that of rabbit hemorrhagic fever disease virus, Sapovirus, and foot-and-mouth disease virus by secondary structure alignment (with HHpred https://toolkit.tuebingen.mpg.de/tools/hhpred). The x-ray crystal structure of rabbit hemorrhagic fever disease virus (member of Calciviridae; an RdRp) solved in 2001 revealed the typical 'right-hand' topology seen in other polymerases like those of poliovirus and hepatitis C virus. [27]

Sapovirus, a member of the caliciviridae family and a causitive agent of gastroentiritis, also has a polymerase that displays the typical right hand structure, though its c-terminal domain is located within the active site cleft. [12] Foot and mouth disease virus is a picornavirus with a 5' UTR and polyadenylated 3' end

Family	Genus	Notable Members
Astroviridae	Avastrovirus	TAstV
	Mamostrovirus	HAstV
Bromoviridae	–	Plant viruses
Caliciviridae	Lagovirus	
	Nebovirus	
	Norovirus	
	Sapovirus	
	Vesivirus	
Closteroviridae	–	Plant viruses
Flaviviridae	Flavivirus	Dengue
	Flavivirus	JEV/WNV
	Flavivirus	YFV
	Flavivirus	Zika
	Hepacivirus	HCV
	Pegivirus	
	Pestivirus	BVDV
Hepeviridae	Orthohepevirus	Hepatitis E Virus
Leviviridae	–	Phages
Luteoviridae	–	Plant viruses
Matonaviridae	Rubivirus	Rubella virus
Narnaviridae	–	Fungal viruses
Nidovirales	Arterivirus	EAV
	Coronaviridae	SARS-CoV
	Coronaviridae	MERS-CoV
Picornaviridae	Apthovirus	FMDV
	Enterovirus	EV71
	Enterovirus	Human poliovirus 1
	Enterovirus	Rhinovirus
		various plant viruses
Potyviridae	Potyvirus	Plant viruses
Togaviridae	Alphavirus	CHIKV
	Alphavirus	SINV
Tombusviridae	–	Plant viruses
Tymovirales	–	Plant viruses
Virgaviridae	–	Plant viruses

Table 3.1 List of the majority of +ssRNA Families. Compiled from the NCBI virus sequence data set in Wolf et al. [45]

that uses a protein primer of 20-24 aa to initiate RNA synthesis. A co crystal of the FMDV polymerase and its VPg showed VPg bound within the RNA-binding cleft, in the palm of the RdRp. [9] While the N-terminal end of the FMDV VPg could be accurately placed, electron density was poor for the C-terminus leading out of the channel.

Using the structural alignment tool US-align [46], crystal structures of +ssRNA polymerases were scored for similarity in a pairwise fashion and clustered using that similarity score as a quasi-distance (Fig. 3.1). The TM-score measure of structural similarity implies a comparable fold with a score of 0.5 and an identical fold with a score of 1.0. Aside from the fungal and dsRNA polymerase outgroups (2R7T, 2J7N), each of the structures compared shared the same fold by the TM-score metric and the picornaviridae, caliciviridae, flaviviridae, and birnaviridae RdRps were nearly identical.

The typical description of polymerase folds in general is that of a partially closed hand, with finger, palm, and thumb domains (Fig. 3.2). The RNA molecule lays upon the palm with the fingers and thumb wrapping around it. Most of the variation that is observed in ++ssRNA viral RdRps occurs in the finger and thumb domains while the catalytic domain within the palm motif is highly conserved. Among viruses that use a VPg, several RdRp-VPg co-crystals have been solved. These RdRp-Vpg structures imply diverse binding modes – some crystals show vpg within the pocket(2F8E [9]) while in others VPg binds the outer surface(4IKA[5], 3CDW[15])).

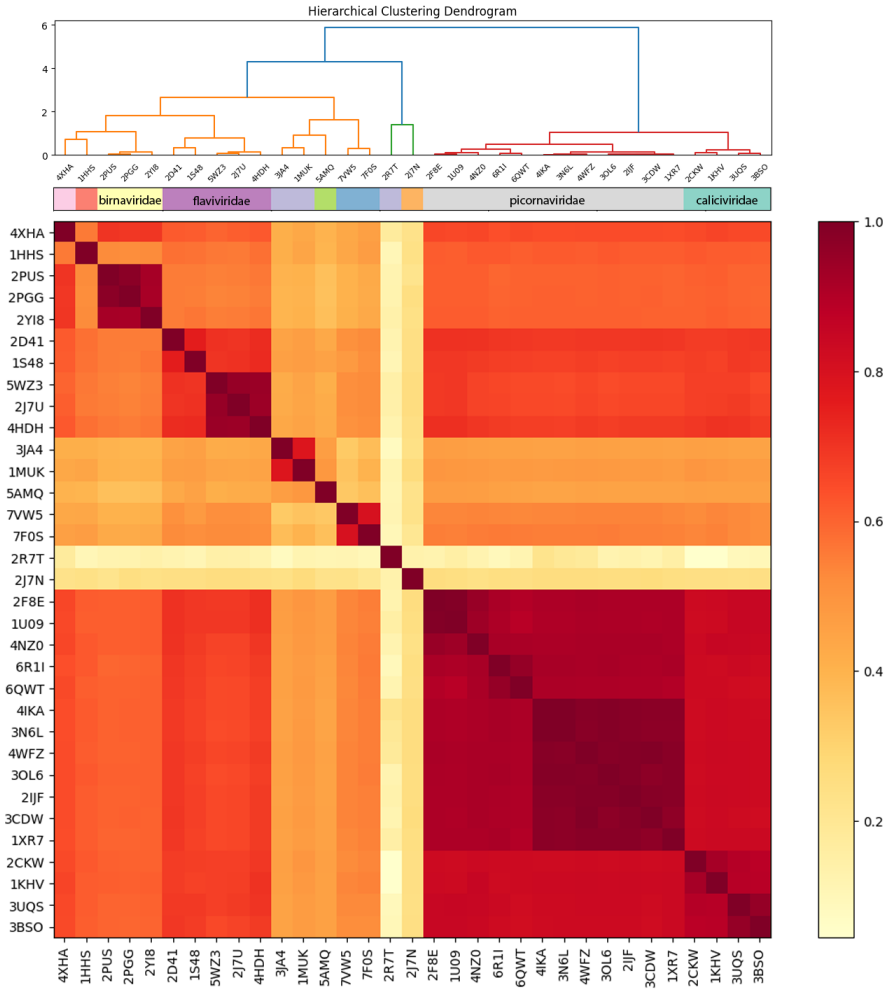


Figure 3.1 Clustering RdRp x-ray structures based on similarity. Pairwise structural similarity scores were computed using USalign [46], the average of each pair's normalized TMscore was clustered using the Ward algorithm and is displayed in the heatmap. Proteins with a TMscore higher than 0.5 generally have the same fold and a score of 1 is the most similar. Neurospora (fungus, 2J7N) and reoviridae (dsRNA, 2R7T, 3JA4) polymerases were included as outgroups. See Table 3.3 for a list of the included Pdb structures.

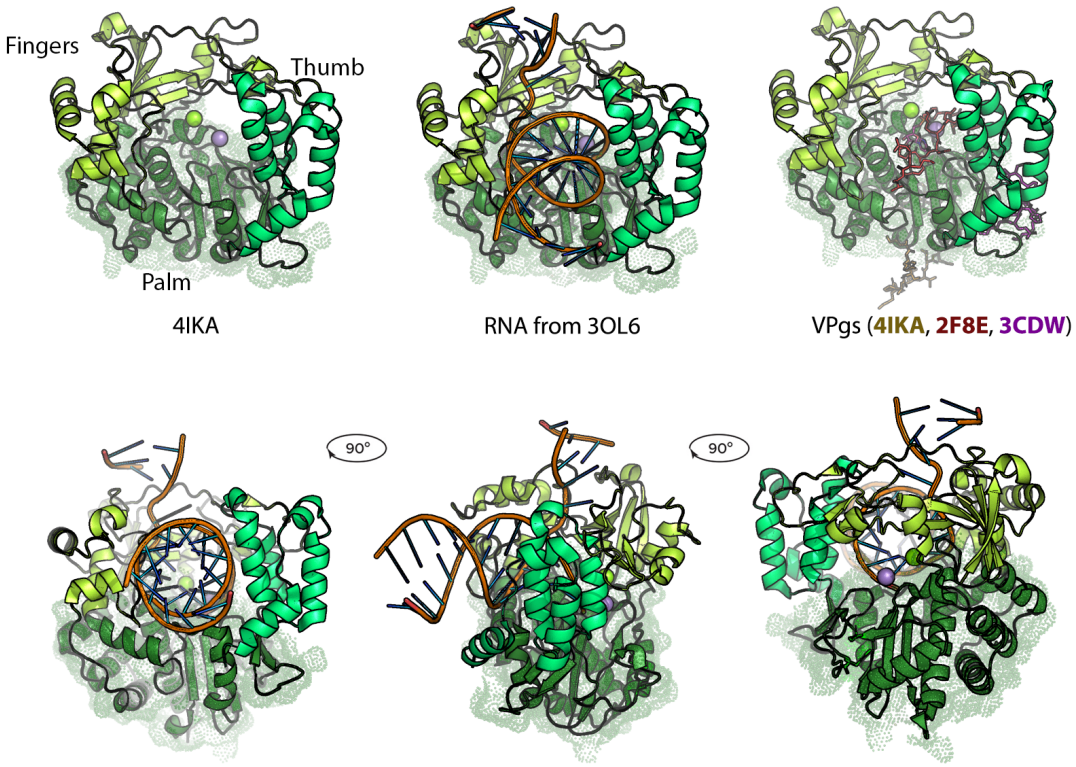


Figure 3.2 Picornaviridae RdRp model based on 4IKA[5], 2F8E[9], and 3OL6[13] structures. The palm, finger, and thumb 'domains' labelled by comparison to the phi6 polymerase (PDB 1HHS) as described in Butcher et al.[4]

VPg

Astrovirus

When the Human Astrovirus 2 genome was sequenced, the authors noted a lack of helicase and methyltransferase genes, suggesting HAstVs may use a VPg priming system like viruses in the Caliciviridae, Picornaviridae, and Potyviridae families.[17] Much later, computational analysis of HAstV genomes suggested presence of a VPg[25], which was then verified experimentally by isolating Astrovirus genetic material followed by RNAse treatment and gel electrophoresis [11]. Astrovirus Vpg appears necessary for priming translation, as proteinase K treatment of the Astroviral genome abrogates infectivity in a CaCo2 cell based in vitro model [11].

Researchers studying norovirus (Caliciviridae) found evidence that its Vpg facilitated 'polymerisation' of norovirus RdRp which is claimed to be important for efficient genome replication. [20] These authors showed this by glutaraldehyde crosslinking followed by Western blot analysis using anti-RdRp and anti-Vpg sera and also claimed nM affinity constants both for RdRp-RdRp and RdRp-Vpg interactions measured by SPR.

Studies of Vpg binding and uridylation in Enterovirus A71 (Picornaviridae) showed a distinct Vpg binding site on the base of the RdRp palm domain. Mutations of amino acids in this region, far from the canonical catalytic center, abolished Vpg uridylation activity and impaired virus replication in vitro.[5] The authors also claim uridylation activity could be rescued by addition of an rdrp lacking polymerase activity, they say this suggests a 'trans' mechanism of uridylation where

one rdrp binds the vpg while another uridylates, but don't fully follow their logic – though there are a number of sources pointing toward cooperative action of a series of RdRps in different viruses, e.g. norovirus above.

Structural information gleaned from xray diffraction data allowed researchers to perform site directed mutagenesis on FMDV polymerase and its VPg to determine a number of amino acid residues important for uridylation activity.[9]

3.3 Preliminary Experiments with HAstV1 RdRp and Vpg

3.3.1 Construct design and cloning

The amino acid sequences chosen for the HAstV1 RdRp and VPg constructs were derived from the gene annotations at UniProt entry Q67726 for the HAstV1 ORF1AB, which encodes the polyprotein containing the non-structural proteins. Of the approximately 7kb genome, ORF1AB makes up $\tilde{4}$.5kb and is thought to encode at least 6 proteins, including the RdRp, VPg, serine protease, and several less characterized proteins (Fig. 3.3). Literature describing the astrovirus RdRp and VPg is sparse, with no information regarding heterologous expression of either protein known to the author at the outset of this work. To assist gene design and improve chances of a stable and soluble RdRp expression, a predicted structure was generated with AlphaFold and aligned to other experimental RdRp structures[18]. Both the N- and C-termini are near the channel where template RNA binds (Fig. 3.4), making it prudent to try placing purification tags at either terminus.

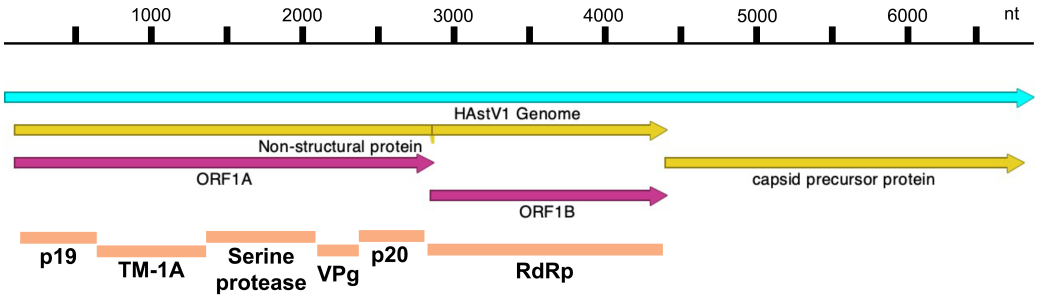


Figure 3.3 The HAstV genome and putative proteome. ORF and gene annotations based on UniProt proteome ID UP000001650.

To start, three core constructs were planned for the RdRp: the full length gene as annotated on UniProt, an N-terminal truncation (aa 10-503), and a C-terminal truncation (aa 1-496). Each of these constructs could be simply cloned from the full length gene into vectors with a variety of tag types and orientations. The 503 amino acid sequence for the RdRp was codon optimized using a machine learning method developed by Fu et al [10]. This method essentially makes a translation model using the target organism's coding genes as the training corpus, which can then be used to translate a desired sequence into 'codon language' for that organism. The output scored better than GeneWiz's (a popular commercial gene vendor with codon optimization service) codon optimizer's output by the codon adaptation index (CAI) metric, which scores a gene's codon usage compared to the usage within a reference set of highly expressed genes [29]. However, CAI isn't strictly correlated to protein expression, so there's undoubtedly room for improved codon optimization if information beyond genomic codon frequency were incorporated. Several modifications had to be made where difficult-to-synthesize sequences of codons were chosen by the ML model.

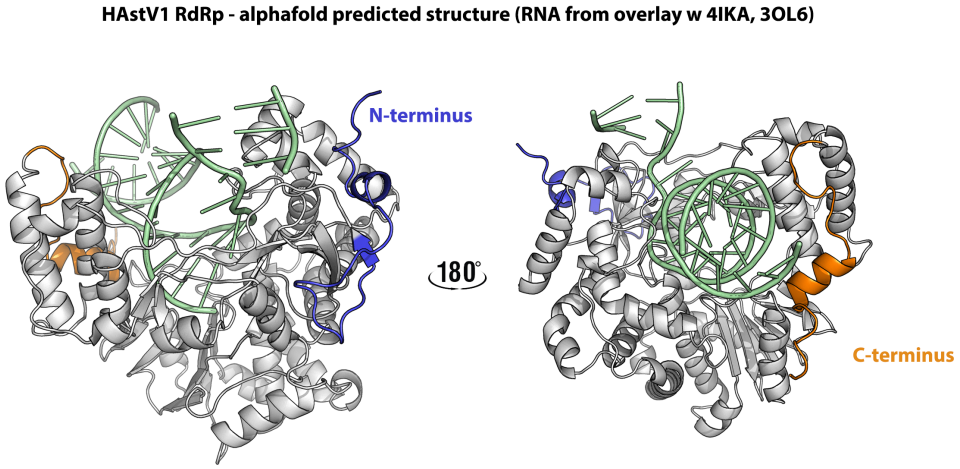


Figure 3.4 The predicted structure of HAstV1 RdRp with N- and C-terminii highlighted.

Gibson cloning strategies were used to generate N- and C-terminally histidine tagged RdRp genes in bacterial expression vectors from a commercially synthesized gene product. Initially, thrombin cleavage sites were included between the RdRp and His tag, though poor protease performance during expression trials led to re-cloning with TEV cleavage sites instead.

3.3.2 HAstV1 RdRp Expression trials & buffer optimization

RdRp constructs cloned into pBR322 based bacterial expression vectors were used for expression trials in BL21 variants. The first trials used C-terminally tagged HAstV RdRp with a thrombin cleavage site. The construct's pI of 6.7 informed the buffer choice and pH (40 mM HEPES, 300 mM NaCl, 5% glycerol, 10 mM imidazole; pH 8). Among the first four small scale expressions, yields varied widely, with one expression apparently yielding >5mg/L while others were unproductive.

Freshly transformed colonies (or glycerol stocks made from fresh transformants) apparently solved the inconsistency in expression, however the next issue was that the purified RdRp was contaminated with DNA or RNA (high 260nm absorbance observed) and was unstable and precipitated during 4 C storage.

To address presumed nucleic acid contamination, the lysis buffer was optimized for improved benzonase nuclease performance (first goal was crystallographic screening, wasn't concerned with residual nuclease). Benzonase requires Mg and performs best in low salt (<150 mM combined monovalent cations), low phosphate buffers [Benzonase Nuclease User Protocol TB261]. The benzonase compatible lysis buffer consisted of 25 mM HEPES, 100 mM NaCl, 2.5% glycerol, 10 mM imidazole, 1 mM MgCl2, and protease inhibitors. Improvement in nucleic acid contamination as measured by A260 absorbance after His-tag affinity purification improved only marginally under these conditions and significant precipitation was still observed after relatively short periods in refrigerate storage. Next, high salt lysis buffer conditions were trialed – salt can disrupt charge-based interactions (like those potentially occurring between negatively charged nucleic acid and the RdRp) and has been observed to improve stability of proteins in solution [e.g. HAstV core protein]. Including 600 mM NaCl in the lysis and wash buffer resulted in much cleaner UV absorbance profiles in affinity purified RdRp fractions. Precipitation rate was slowed dramatically, what previously occurred overnight now took several days to a week.

The high salt buffer led to the first expressions that went into crystal trays. During this time, the RdRp constructs contained thrombin cleavage sites for proteolytic tag removal – regrettably, yield of thrombin based cleavage was poor and

seemed to trigger precipitation events. However, full length His-tagged RdRp as well as thrombin cleaved polymerase was set up in crystal screens while the construct was re-cloned into a vector with TEV cleavage sites for tag removal. Size exclusion chromatography prior to crystallographic screening revealed presence of aggregate RdRp. Around this time, polymerase activity assays were being researched, and a common element of nearly all polymerase buffers was some form of reducing agent. An investigation of the predicted HAstV1 RdRp structure revealed presence of extensive unpaired cysteine residues decorating the RdRp surface – suggesting the precipitation and aggregation observed throughout expression trials may be attributable to RdRp cross-linking through those free cysteines.

After addition of reducing agent to all buffers used during purification (beta-mercaptoethanol during affinity purification stages for resin compatibility, then DTT at the SEC stage), RdRp stability after purification further improved and putative aggregate observed during SEC was nearly eliminated (Fig. 3.5). After these optimizations (in particular high salt during lysis, and inclusion of reducing agent), a protocol that produces clean and relatively stable HAstV1 RdRp had been developed (Fig. 3.6). For RNA assays where DEPC treated buffers may be used, it is advisable to switch the buffering molecule to MOPS as HEPES (and Tris) react with DEPC and may interefere with RNAase inactivation.

HAstV1 VPg purification The full length VPg gene as annotated in UniProt entry Q67726 was expressed via a pET28 vector in E. coli. Little optimization has been performed, the 10H-TEV-VPg construct expresses weakly (1 mg/L in preliminary trials) but can be obtained with relative purity by affinity chromatography (3.7).

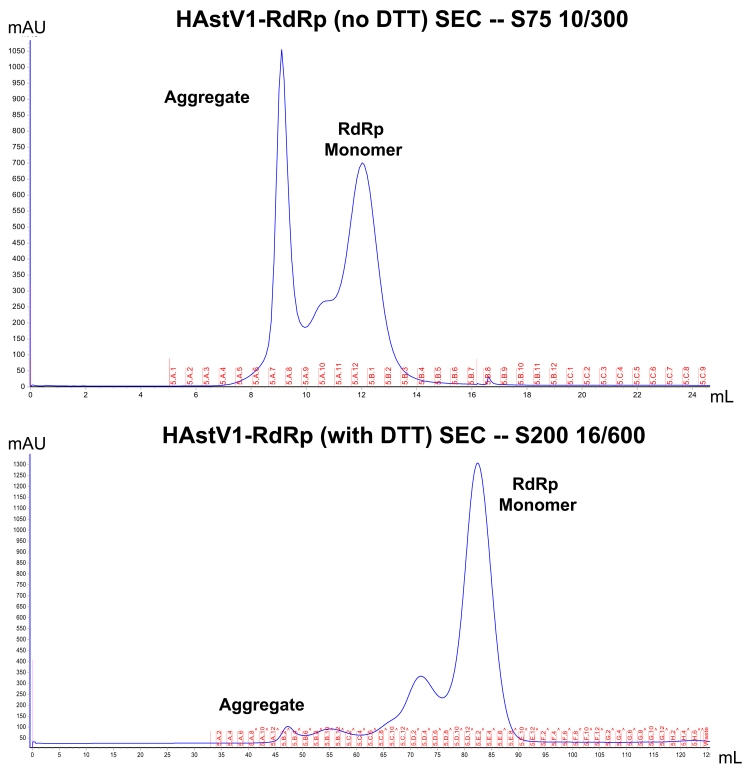


Figure 3.5 Reducing agent mitigates HAstV RdRp aggregation. Size exclusion chromatograms from two HAstV1 RdRp purifications where reducing agents were omitted (top) or included (bottom) in all buffers. Notably, the aggregate peak's area was miniscule when reducing agent is used. Elution times differ due to column particle size and column volume differences between runs.

3.3.3 Crystallographic Screening

As expression conditions were optimized, purified RdRp from those expression trials was used for crystallographic screening. Microvolume, sitting-drop screening plates were prepared using the Midwest Center for Structural Genomics (MCSG) suite of buffer conditions. The RdRp constructs, concentrations, and sample buffer conditions for the screens tested so far are summarized in Table 3.2. Unfortunately the solubility issues observed during purification persisted, and the majority of screened conditions quickly developed signs of aggregate formation

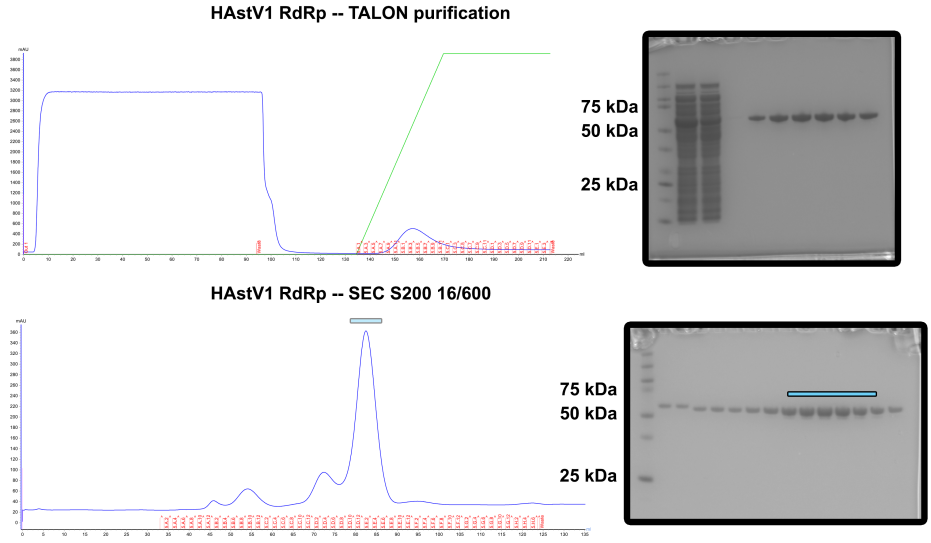


Figure 3.6 Representative affinity chromatography and size exclusion chromatograms for optimized HAstV RdRp purification. After buffer optimization, affinity purification of HIS-tagged RdRp with TALON resin followed by tag proteolysis and size exclusion chromatography results in high purity RdRp monomer.

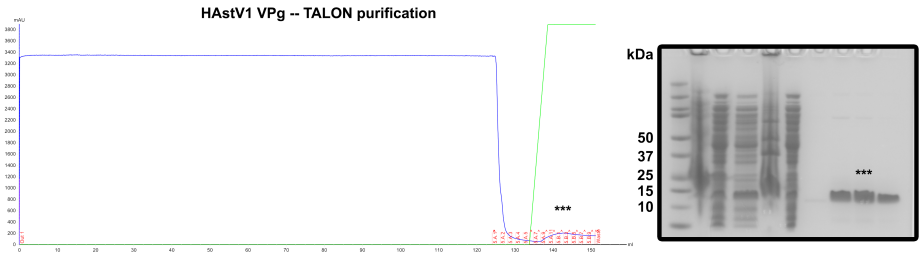


Figure 3.7 HAstV1 VPg affinity chromatography and SDS-PAGE gel

Construct	Concentration(s)	Buffer Notes	Screens
HAstV1-RdRp(1-503)-Thr-10H	5 mg/mL	300 mM NaCl	MCSG1; -2; -3
HAstV1-RdRp(1-503)-/-Thr-10H (cut)	8 mg/mL	600 mM NaCl	MCSG1; -2; -3; -4
10H-TEV-HAstV1-RdRp(1-503)	7.5 mg/mL	600 mM NaCl	MCSG1; -2; -3; -4
10H-TEV-/-HAstV1-RdRp(1-503) (cut)	5, 7.5 mg/mL	1 M NaCl	MCSG1; -2; -3; -4
10H-TEV-/-HAstV1-RdRp(1-503) (cut)	4.7, 9.5, 15 mg/mL	500 mM NaCl, 2 mM DTT	MCSG1; -2; -3; -4
10H-TEV-/-HAstV1-RdRp(1-503) (cut)	3, 5.5, 8 mg/mL	2 mM DTT, 1mM ea MgCl2, MnCl2, UMP	MCSG1; -2; -3; -4
10H-TEV-/-HAstV1-RdRp(1-496) (cut)	2.4, 4.9, 7.9 mg/mL	500 mM NaCl, 2 mM DTT	MCSG1; -2; -3; -4
10H-TEV-/-HAstV1-RdRp(1-496) (cut)	3, 4.6 mg/mL	2 mM DTT, 1mM ea MgCl2, MnCl2, UMP	MCSG1; -2;

Table 3.2 List of RdRp constructs and crystal screening conditions.

and loss of solubility (Fig. 3.8). The inclusion of DTT slowed the rate of 'nebula' formation by several days, perhaps there is a combination of construct and screening condition where crystals form before aggregation begins. Exploration of more persistent reducing agents like TCEP may also be warranted.

So far, only a few RdRp constructs were screened: N- and C-terminally tagged, full length RdRp (1-503); full length RdRp with tags proteolytically removed from either terminus; RdRp (1-496) with and without N-terminus tag. The constructs with N-terminal truncation (10-503, 10-496) do not appear to solubly express. Magnesium and manganese, known polymerase co-factors, were included in two series of screening conditions for both the (1-503) and (1-496) RdRp constructs. Further C-terminal truncation, screening of a catalytically dead mutant (e.g. D360A, D361A) with and without RNA template, and additional or alternate reducing agent in sample buffers are all attractive crystal screening conditions remaining to be explored.

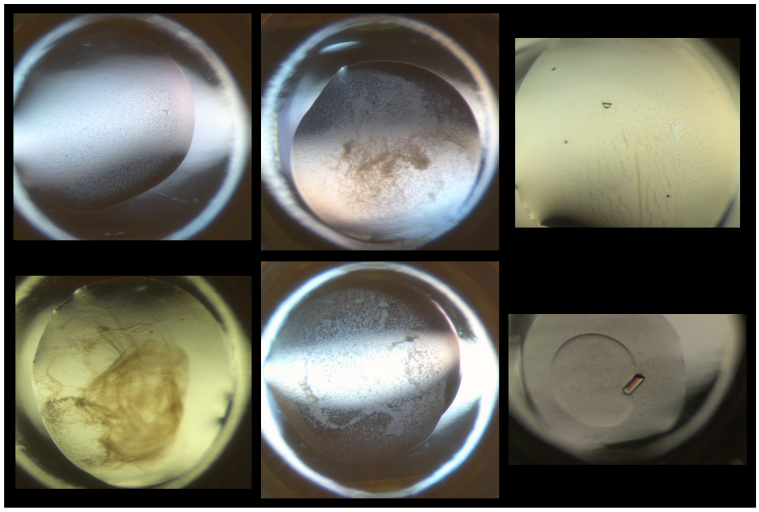


Figure 3.8 Representative images of HAstV1 RdRp crystal screen conditions. The vast majority of conditions assessed in the MCSG buffer suite led to formation of nebulous precipitation across a variety of RdRp concentrations and sample buffers. A limited number of 'hits' were observed and sent to a synchrotron for analysis, however each were either inorganic or non-diffracting crystals.

3.3.4 Activity assay development

RNA template design

Viral RNA-dependent-RNA-polymerases have been shown to be competent to replicate RNA with many configurations of templates and primers. Some RdRps appear to have preferred initiation site sequences while others are capable of de novo initiation with generic sequences. For Human Astrovirus, little is known about the RdRps requirements and template preferences. Our goals are two-fold: to characterize the biochemical activity of the HAstV RdRp and to identify the simplest set of conditions where the RdRp is measurably active so that a high throughput inhibitor screen can be developed.

To that end, it makes sense to explore a variety of template-primer configu-

rations. On the simpler end, homopolymer templates with and without primers should be assessed. Not only are homopolymer templates affordable in bulk, but polyA and polyU templates are particularly relevant for astrovirus whose genome is typically poly-adenylated. It may also be possible to use a self-priming RNA template, where successful transcription actually extends the template molecule rather than making complementary copies, which is attractive for gel based assays. It is also possible that the HAstV RdRp prefers particular initiation sequences, putatively present at either end of the astrovirus genome. Alignments of canonical human astrovirus strains Fig. 3.9 show regions of very high sequence conservation at both the 5' and 3' ends of the genome. These regions and their complements are excellent candidates for RdRp initiation sites and should be explored with and without priming (and polyadenylation should be considered).

Cost of RNA synthesis is prohibitive and error-prone, so in vitro transcription of template candidates is an attractive route to template generation. To this end, T7 RNA polymerase can be used to transcribe target RNAs in large quantity from plasmid or other DNA templates. Here, DNA templates were developed for use with a P266L T7 polymerase mutant reported to have higher yields due to fewer aborted transcripts [6]. The ideal T7 initiation sequences lead to a number of additional 5' bases being prepended to a given target sequence, and T7 termination does not always occur at uniform positions. In some cases these extra bases are acceptable, but if a precise RNA sequence is desired, a common practice is including self-cleaving RNA sequences flanking the target RNA. At the 5' end, a hammerhead ribozyme can be designed and inserted while the 3' end can either be controlled by addition of a GlmS or hepatititis delta ribozyme or by simply cutting template DNA with a restriction enzyme (3.10B). For some RNAs where the

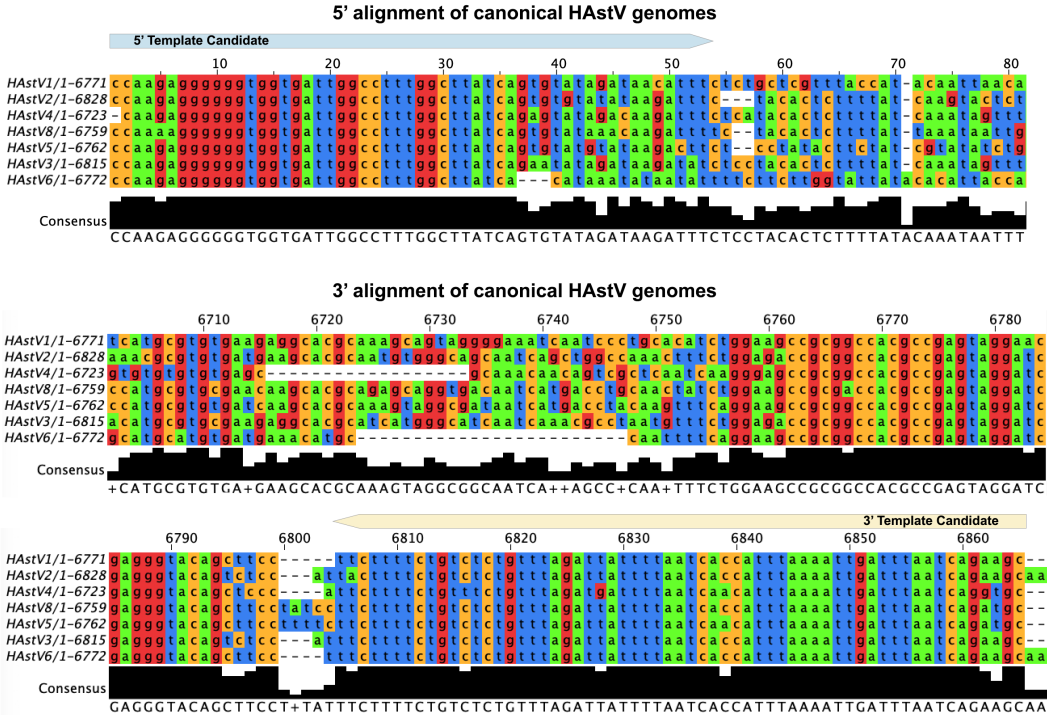


Figure 3.9 Alignment of canonical HAstV genomes reveals high conservation at the 5' and 3' ends. Sequences aligned with MAFFT using the E-INS-i strategy (mafft.cbrc.jp). Accession numbers of sequences used in alignment: MK059949.1 (HAstV1), L13745.1 (HAstV2), AY720891.1 (HAstV4), AF260508.1 (HAstV8), JQ403108.1 (HAstV5), AF141381.1 (HAstV3), HM237363.1 (HAstV6).

standard precipitation and re-solublization or denaturing gel electrophoresis purification protocols may lead to heterogeneous or otherwise non-native RNA folds, affinity purification platforms can be used to purify RNA while retaining native RNA structure. One method employs an RNA motif from the MS2 bacteriophage genome that binds tightly to the MS2 coat protein, which can be exploited as an RNA affinity tag [1]. Target RNAs can be transcribed with an inducible, self-cleaving ribozyme like GlmS followed by the MS2-coat binding motif then purified via incubation with a tagged MS2 coat protein (e.g. histidine or strep tag). The RNA-MS2 protein complex is then itself pulled from solution onto beads complementary to the protein tag. Target RNA is can be eluted by activating the inducible self-cleaving ribozyme, freeing the natively folded RNA.

In vitro transcription of RNA templates HAstV1 RNA fragments were transcribed from NgoMIV cut, maxi-prepped plasmid templates with house-made T7 polymerase P266L. First generation templates did not include 5' ribozyme, and would be expected to bear several additional 5' bases making up the optimal transcription start sequence outlined previously. An important detail "discovered" during T7 transcription trials necessary for high yield transcription was the pH of homemade rNTP stocks (thanks to Sara O'Rourke for her insight here). The low pKa of the rNTPs led to very acidic solutions when prepared from the dry sodium salts, which led to inappropriate pH in the final T7 reaction buffer due to the high concentration of total rNTPs present (16+ mM). Adjustment of rNTP stock pH with NaOH led to a dramatic improvement in T7 transcription activity. Preliminary transcription attempts with the HAstV1 mock genome and 5' primer fragment templates led to nearly 50 μg yields from 50 μL reaction volumes after DNAse treatment and subsequent phenol-chloroform extraction and precipitation.

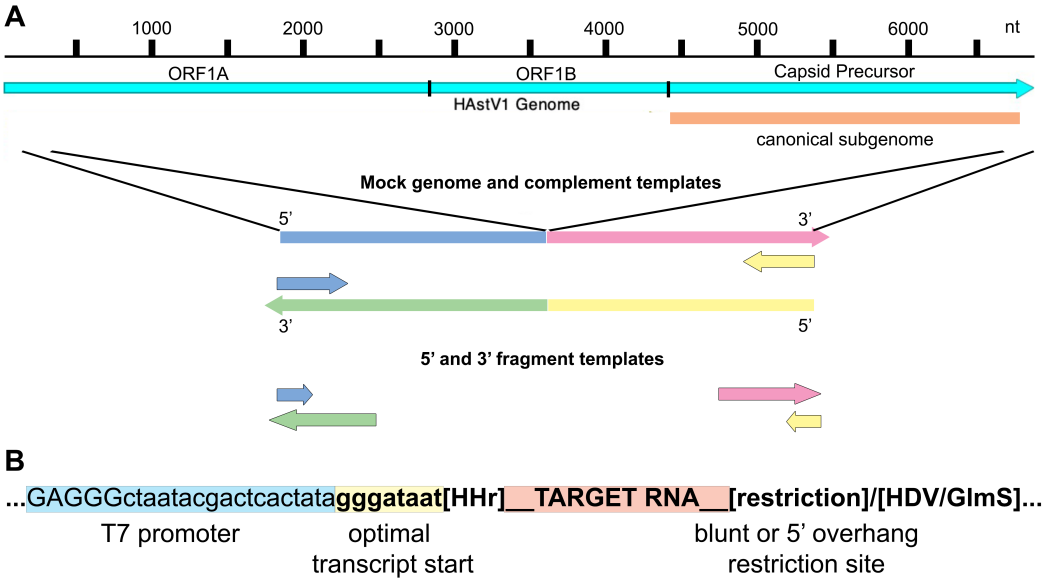


Figure 3.10 Proposed RNA templates for HAstV RdRp activity assessment and in vitro transcription DNA design. (A) Candidate RNA templates for HAstV RdRp include: a mock genome and its complement containing a fusion of the first and last several hundred nucleotides of the HAstV genome; shorter fragments of either end of the genome. In addition to the 5' and 3' ends of the astroviral genome, there is a known subgenomic fragment containing the capsid precursor protein that may bear an additional preferred transcription initiation site. (B) Gene schematic for T7 transcribed RNA sequences. Here 'target RNA' refers to a DNA sequence that would produce the desired RNA when transcribed. Shorter target RNAs can be built via primer overlap and reverse transcription, while larger targets (or high yield goals) warrant cloning into a high copy plasmid followed by e.g. maxi-prep purification. HHr = hammerhead ribozyme; HDV = hepatitis delta ribozyme; GlmS = an inducible, self cleaving riboswitch [32]. RNA affinity tags like the MS2 system would be placed 3' to an inducible, self-cleaving ribozyme like GlmS (detailed methods available in Batey et al. [1])

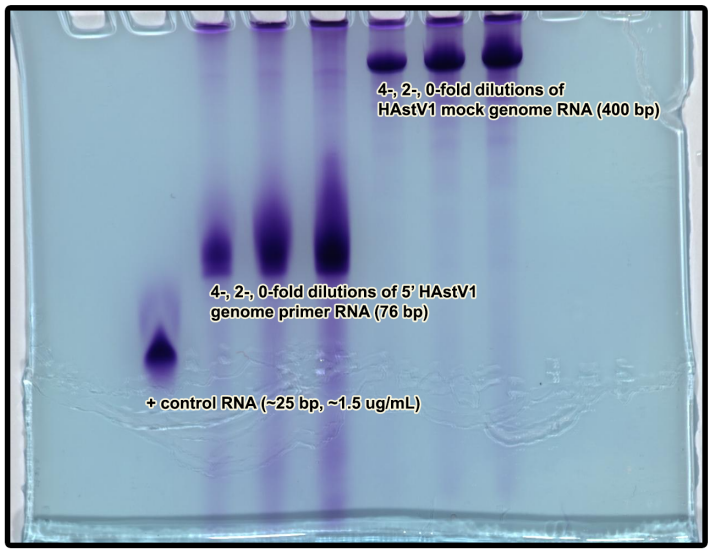


Figure 3.11 Urea-PAGE gel showing successful transcription of HAstV1 5' template and mock-genome RNA. The T7 transcribed RNA fragments were run in dilution series compared to a known good small RNA in an 8% Urea-PAGE gel with TBE buffer.

Urea-PAGE analysis demonstrated relatively good purity, with evidence of presence of some alternative transcripts in the mock-genome sample (See Fig. 3.11). This sets the stage for additional RdRp transcript production needed for activity assay development and potential RdRp-RNA binding studies.

Future work: Biochemical characterization of HAstV1 RNA-dependent RNA Polymerase

Little is known about the requirements and activity of the astrovirus RdRp. Aside from structural characterization of HAstV RdRp, the other primary goal of these studies is to develop a polymerase assay that could be used to screen for RdRp inhibitors that could become anti-viral agents (as well as uncovering the biology of HAstV replication). Many experiments remain to be performed: what are the co-factor requirements (metals, Vpg)?; is HAstV RdRp capable of de novo RNA

synthesis or is it primer dependent?; are there common features or structures in the 5', 3', and canonical subgenome that the RdRp recognizes?; what part of the RdRp catalyzes Vpg uridylation?

Sapovirus RdRp, one of the closest putative relatives structurally speaking to HAstV polymerase is able to initiate synthesis with poly(A) and poly(C) homopolymer templates in a primer dependent fashion.[12] Sapovirus RdRp was also able to catalyze polymerasition with T7 transcribed, polyA-tailed anti-subgenomic RNA (i.e. the negative strand of portion of the Sapovirus genome).[12] In light of these experiments, there is some suggestion that HAstV RdRp may also be receptive to a variety of template and primer configurations.

For inhibitor discovery, an assay developed for a more distant relative of astrovirus could be a good model. A high-throughput, fluorescence based polymerase assay was developed to screen for RdRp inhibiting anti-virals of Zika virus polymerase (Flaviviridae).[31] The Zika RdRp is able to synthesize RNA de novo from homopolymer template independent of any priming, which can be detected with a fluorescent dye by a plate reader. While astrovirus may need a specific template rather than the simple homopolymer used for this Zika virus assay, the concept of using fluorescence intensity via measurement of a double-stranded RNA selective dye is an exciting approach that could allow large scale inhibitor screening with relatively small reaction volumes.

3.4 Methods

3.4.1 Expression of RdRp

```
Use either freshly transformed BL21 variant (T7 express was the primary strain
    for the trials discussed here) colony or glycerol stock made from freshly
    transformed colony to make starter culture.  Although not conclusively
    shown, transformant age seems important for strong expression based on
    early trials. Currently prefer pET28-MHL-10H-TEV-HAstV1-RdRp lineage of
    constructs (KanR).

Ideally use rich broth (ZYM 505 / TB) for starter culture at least (supports
    higher ODs before growth stalls, good for overnight cultures).  I typically
     aim for 20-50 mL starter per liter of media to be inoculated.For
    production have had good success with either rich media (ZYM5052) or LB
    supplemented with MgCl2 (1 mM) and Fe (1 mM).

Have seen good results inducing with IPTG between 0.6-1.2 OD600, though its OK
    to induce at higher OD in rich media where the exponential phase occurs
    over a broader OD600. Overnight at low temperature seems better than 4h at
    30C, but controlled comparisons have not been performed.

Lysis and purification:
Lysis buffer:           for 1 L:
25 mM HEPES pH 8        50 mL 1M stock
600 mM NaCl             150 mL 4M stock
2.5% glycerol           50 mL
10 mM imidazole         680 mg
5 mM B-ME (add just prior to use, short half-life at pH 8)

Wash buffer:          for 1 L:
25 mM HEPES pH 8        50 mL 1M stock
300 mM NaCl             75 mL 4M stock
2.5% glycerol           50 mL
10 mM imidazole         680 mg
5 mM B-ME

Elution buffer:
Wash buffer with 300 mM imidazole

Cell culture and purification protocol is somewhat flexible with reducing agent
     present. Typically, an overnight culture grown in ZYM505 media was used to
     innoculate 1L of Mg-supplemented LB/Kan broth in 2.8L baffled culture
    flasks, starting at 37C and shaking at 200rpm.  Expression was induced
    between 0.6-1.0 OD600 with 500 uL 1M IPTG per Liter of media. At induction,
     incubation temperatures were reduced to 18C and fermentation allowed to
    proceed overnight.

Cells were harvested by centrifugation at 6000xg prior to resuspension in lysis
     buffer with fresh DTT and protease inhibitor. Resuspended lysate was
    sonicated for 15-25 cycles of 10s on, 50s off at 60% intensity.  Insoluble
    material was removed from sonicated lysate by high speed centrifugation
    (19,000 rpm with a JA-20 rotor) followed by 0.45 micron filtration.
    Purified RdRp was obtained from the filtered lysate using either Ni-NTA or
    TALON resin, with TALON generally providing better purity at the expense of
     yield. After loading lysate onto resin, the column was washed with 10
    column volumes of wash buffer, followed by a gradient elution over 5 column
```

```
    volumes from 0->100% elution buffer.

If the his tag is to be removed with TEV-protease prior to SEC chromatography,
    eluate should be dialyzed back into wash buffer so any remaining tagged
    protein can be removed by a second affinity chromatography step. For size
    exclusion chromatography, use of DTT or TCEP is recommended as a
    replacement reducing agent over beta-mercaptoethanol.
```

3.4.2 Expression of Vpg

Histidine-tagged VPg was expressed and purified using the same protocol as described for HAstV RdRp, except without the inclusion of reducing agent. Yields appear marginally better with 'batch' purification over FPLC.

3.4.3 Expression of TEV-protease

```
The His tagged TEV protease is on a plasmid confers KanR; T7 based expression.
    Works fine with standard BL21 expression using LB/Kan; 18C overnight post
    induction.

Lysis buffer
15mM HEPES pH 8
1M NaCl
5% glycerol
2 mM beta-mercaptoethanol
30 mM imidazole
Protease inhibitors
Benzonase

Wash/elution
15mM HEPES pH 8
300mM NaCl
5% glycerol
2 mM beta-mercaptoethanol
30/300 mM imidazole

Storage
15mM HEPES pH 8
300mM NaCl
5% glycerol
2 mM beta-mercaptoethanol

##
Made a glycerol stock with transformed T7 express cells (variant of BL21) from
    a single colony. (added 250uL 80% glycerol to 750uL ~1 OD600 starter
    culture).

For the below prep, ZYM 505 media was used.  Started with a 20mL overnight
    culture from glycerol stock. The next morning, inoculated two 1L cultures
    in baffled 2.8L flasks.  Grew to ~ 2.5 OD600 at 37C with 200RPM shaking
```

```
    before induction with 200uL 1M IPTG stock (ZYM505 supports much higher OD
    before saturation than LB, could have induced even later).  Switched to 18C
     after induction and grew overnight.  Harvested cells by centrifugation at
    6000xg.

Resuspend pellets in cold lysis buffer, sonicate for >10 cycles at 50% power,
    10s on 50s off.  Centrifuge at high speed (>20000 x g) to pellet insoluble
    fraction, 0.45 or 0.22 micron filter.  Expression can go above 10 mg/L, so
    best to use a 5 mL Talon or Ni-NTA column.  Load filtered lysate onto
    column, wash with 10 CV wash buffer, then elute (gradient elution over 6 CV
     done here via AKTA).

Combined all but first several elutions (to minimize contaminants).  Got ~30mL
    at 1.4 mg/mL prior to concentration and dialysis.  Concentrated to ~5 mL at
     6.5 mg/mL and dialyzed into storage buffer overnight at 4C. After dialysis
    , filtered sample and further concentrated to ~10 mg/mL before flash
    freezing.  Yield after concentration and filtering was ~27 mg from 2L ZYM
    media.
```

3.4.4 T7 polymerase

T7 RNA polymerase P266L was expressed as described by the AddGene depositor (AddGene Plasmid #174866 [6]).

3.4.5 RNA template preparation

The in vitro transcription reaction was based on a number of sources including Rio et al, Conrad et al, and Kartje et al [7, 19, 30].

```
## Target concentrations for high yield reaction

20 mM HEPES pH 8
10 mM DTT
0.01\% triton x-100
2 mM spermidine
4 - 20 mM combined rNTPs  (lit suggests 10-20 mM produces best yield, i.e.
    2.5-5mM ea rNTP)
32 mM MgSO4 or MgCl2 (should be 6 mM more than combined rNTP concentration -
    RNA lab manual)
1U/mL inorganic phosphatase
1-2 uM template [note this is only feasible with small templates, ~=300 ug of
    5000bp template = 1uM in 100uL]
5-10uM (500-1000 ug/mL) T7 polymerase

# Stocks
100 mM rNTPs
10X HEPES-Triton (200 mM HEPES pH 8; Triton 0.1\%)
    >>> add DTT fresh (degrades quickly) 15.4 mg/mL for 10X stock
2 M MgCl2
1 M spermidine
100U/mL inorganic phosphatase (commercial stock)

Incubate near 37 C for >2 h.  Literature suggests ~2h for longer transcripts
    and more like 6h for short (less than 100bp) transcripts. Overnight
    reportedly OK.

Typically DNAse treat to minimize contaminating DNA after purification.

LiCl precipitation works well -- bring reaction solution up to ~1M LiCl using
    an 8M stock. Chill in -20C freezer for 30 minutes, spin at 16000g for >20
    min to pellet RNA.  Aspirate supernatant, air dry for 10 min before
    resuspending with water. Consider adding RNAse inhibitor. Aliquot if
    desired and freeze at -80C (or use RNA directly). The Trizol branded phenol
    -chloroform extraction protocol is also effective for RNA purification post
    -transcription.
```

3.4.6 Dendrogram and Heatmap

The structural similarity dendrogram and heatmap were generated by performing a pairwise comparison of the biological assemblies of each polymerase structure listed in Table 3.3 with the USalign program.[46] A bash script, included below, automated the comparison process, producing a list of TMscore values (a measure of topological similarity) which could then be converted into a distance matrix. TMscores range from 0 to 1 (where 1 indicates an identical structure) and are normalized by a protein's size in amino acids, so an A-to-B comparison does not necessarily equal B-to-A if the proteins being compared are different in length.

To create a distance matrix, the complement of the average TMscore for each RdRp pair was used (1-(a:b+b:a)/2). Using the TMscore complement as a quasi-distance, the RdRp structures were clustered by the Ward method within the SciPy Python package's dendrogram module. The heatmap was produced by re-ordering the TMscore matrix in dendrogram order and plotting each TMscore.

```
Bash script that automates calls to the USalign program.

#!/usr/local/bin/bash
# Takes a list of pdb files (no extensions) and generates pairwise list
# USalign tm-score values. because the score is asymmetric when the
# compared structures are different sizes, the sum of the two scores
# are stored (ie a-b + b-a). This captures scenarios where there is high
# similarity, but one of the proteins has more (solved) residues/domains

echo "USalign tm-score value generator"
f=($(cat $1))
output3_file="total.out"

printf "pairwise structure comparison with USalign, sum TM-score %s\n" >
    $output3_file


for ((i = 0; i < ${#f[@]}; i++)); do
      for ((j = i + 1; j < ${#f[@]}; j++)); do

          ./USalign ${f[i]}.pdb1 ${f[j]}.pdb1 | sed -n 15,16p > temp.out;
          tmscores=($(grep -o "TM-score= [0-9]*.[0-9]*" temp.out | sed 's/TM-
    score= //g'))

          tmtotal=$(echo "${tmscores[0]} + ${tmscores[1]}" | bc)
          echo "${f[i]},${f[j]},$tmtotal"
          printf %s "${f[i]},${f[j]},$tmtotal" >> $output3_file;

          printf "%s\n" >> $output3_file;
      done;
  done
```

Listing 3.1 Bash script to automate pairwise calls to the USalign program.

PDB figures were generated by superimposing RdRp structures with either PyMol's 'super' function (more similar structures) or USalign's output (more disparate structures) then rendered using PyMol's raytracing function.

3.4.7 Supplementary Information

PDB ID	Family	Genus	Description
2J7N	sordariaceae	neurospora	neurospora crassa
2PGG	birnaviridae	avibirnavirus	birnavirus
2PUS	birnaviridae	avibirnavirus	bursal disease virus
2YI8	birnaviridae	aquabirnavirus	pancreatic necrosis virus
3UQS	caliciviridae	norovirus	murine norovirus 1
3BSO	caliciviridae	norovirus	norwalk virus
1KHV	caliciviridae	lagovirus	Rabbit Hemorrhagic Disease Virus
2CKW	caliciviridae	sapovirus	Sapporo virus
1HHS	cystoviridae	cystovirus	dsRNA bacteriophage phi6
2D41	flaviviridae	hepacivirus	hepatitis c virus
4HDH	flaviviridae	flavivirus	japanese encephalitis virus
2J7U	flaviviridae	flavivirus	dengue
1S48	flaviviridae	pestivirus	bovine viral diarrhea virus
5WZ3	flaviviridae	flavivirus	zika virus
5AMQ	peribunyaviridae	orthobunyavirus	la crosse orthbunyavirus
4XHA	permutotetraviridae	alphapermutotetravirus	thosea asigna
1U09	picornaviridae	apthovirus	foot-and-mouth disease virus
2F8E	picornaviridae	apthovirus	foot-and-mouth disease virus
4NZ0	picornaviridae	cardiovirus	encephalomyocarditis virus (EMCV)
4WFZ	picornaviridae	enterovirus	Coxsackievirus B3
3CDW	picornaviridae	enterovirus	Coxsackievirus B3
3N6L	picornaviridae	enterovirus	enterovirus A71
4IKA	picornaviridae	enterovirus	enterovirus A71
3OL6	picornaviridae	enterovirus	human poliovirus
1XR7	picornaviridae	enterovirus	human rhinovirus
2IJF	picornaviridae	enterovirus	poliovirus 1 mahoney
6R1I	picornaviridae	kobuvirus	porcine kubovirus
6QWT	picornaviridae	sicinivirus	Chicken sicinivirus JSY
2R7T	reoviridae	rotavirus	simian rotavirus
3JA4	reoviridae	cypovirus	cypovirus
1MUK	reoviridae	reovirus	reovirus
7F0S	togaviridae	alphavirus	ross river virus
7VW5	togaviridae	alphavirus	sinbis virus

Table 3.3 List of PDB files used in Figure 3.1. In most cases, viruses with multiple entries represent structures with and without Vpg or in alternate conformational states.

3.4.8 Sequences

HAstV1 RdRp

Listed here are the HAstV1 RdRp amino acid and nucleotide sequences used for the expressions described above.

```
>HAstV1 RdRp (aa 934-1436 in uniprot Q67726 ORF1AB)
GAQNYHSLDAWKLLLEPPRERRCVPANFPLLGH
LPINRPIFDDKKPRDDLLGLLPEPTWHAFEEYGPTTWGPQ
AFIKSFDKFFYAEPIDFFSEYPQLCAFADWATYREFRYLE
DTRVIHITATEKNTDSTPAYPKMNYFDTEENYLEAHGWAP
YIREFTRVYKGDKPEVLWYLFLKKEIIKEEKIRNSDIRQI
VCADPIYTRIGACLEAHQNALMKQHTDTSVGQCGWSPMEG
GFKKTMQRLVNKGNKHFIEFDWTRYDGTIPPALFKHIKEI
RWNFINKDQREKYRHVHEWYVNNLLNRHVLLPSGEVTLQT
RGNPSGQFSTTMDNNMVNFWLQAFEFAYFNGPDRDLWKTY
DTVVYGDDRLSTTPSVPDDYEERVITMYRDIFGMWVKPGK
VICRDSIVGLSFCGFTVNENLEPVPTSPEKLMASLLKPYK
ILPDLESLHGKLLCYQLLAAFMAEDHPFKVYVEHCLSRTA
KQLRDSGLPARLTEEQLHRIWRGGPKKCD
```

```
>HAstV1 RdRp E. coli optimized DNA sequence
atgGGCGCGCAGAACTACCATAGCCTGGATGCGTGGAAACTGCTGCTGGAACCGCCGCGTGAACGTCGTTGCGTGCCGG
CAAACTTTCCGCTGCTGGGTCATCTGCCGATTAACCGCCCGATTTTTGATGATAAAAAACCGCGCGATGATCTGCTGGG
TCTGCTGCCGGAACCGACCTGGCATGCCTTTGAAGAATACGGTCCGACCACCTGGGGTCCGCAGGCGTTTATTAAAAGC
TTTGATAAATTTTTTTATGCCGAACCGATTGATTTTTTCTCTGAATATCCGCAGCTGTGCGCGTTTGCCGATTGGGCTA
CCTATCGTGAATTTCGCTATCTGGAAGATACCCGCGTGATTCATATTACCGCCACCGAAAAAAACACCGATAGCACGCC
GGCTTATCCGAAAATGAACTATTTTGATACCGAAGAAAACTATCTGGAAGCGCACGGCTGGGCTCCGTATATTCGTGAA
TTTACCCGCGTGTATAAAGGCGATAAACCGGAAGTGCTGTGGTATCTGTTTCTGAAAAAAGAAATTATCAAAGAAGAAA
AAATTCGCAACAGCGATATTCGCCAGATTGTGTGCGCGGATCCGATTTATACCCGTATTGGCGCGTGCCTGGAAGCGCA
TCAGAACGCGCTGATGAAACAGCATACCGATACCAGCGTCGGCCAGTGCGGCTGGAGCCCGATGGAAGGCGGTTTTAAA
AAAACCATGCAGCGTCTGGTGAACAAAGGCAACAAACATTTTATTGAATTTGATTGGACGCGTTATGATGGCACCATTC
CGCCGGCGCTGTTTAAACATATCAAAGAAATTCGCTGGAACTTTATTAACAAAGATCAGCGTGAAAAGTATCGCCATGT
TCATGAATGGTATGTTAACAACCTGCTGAACCGCCATGTGCTGCTGCCGAGCGGCGAAGTGACGCTGCAGACCCGCGGT
AACCCGAGCGGCCAGTTTAGCACCACCATGGATAACAACATGGTTAACTTTTGGCTGCAGGCGTTTGAATTTGCCTATT
TTAACGGTCCGGATCGCGATCTGTGGAAAACCTATGATACCGTGGTTTATGGCGATGATCGTCTGAGCACCACCCCGAG
CGTTCCGGATGATTATGAAGAACGCGTGATTACCATGTATCGCGATATTTTTGGCATGTGGGTGAAACCGGGTAAAGTG
ATTTGCCGTGATAGCATTGTCGGTCTGTCGTTTTGCGGTTTTACCGTTAACGAAAACCTGGAACCGGTTCCGACCAGCC
CGGAAAAACTGATGGCGAGCCTGCTGAAACCGTATAAAATTCTGCCGGATCTGGAAAGCCTGCACGGTAAACTGCTGTG
CTATCAGCTGCTGGCGGCGTTTATGGCGGAAGATCATCCGTTTAAAGTGTATGTCGAACATTGCCTGAGCCGTACCGCG
AAACAGCTGCGTGATAGCGGCCTGCCGGCGCGTCTGACCGAAGAACAGCTGCATCGCATTTGGCGTGGCGGTCCGAAAA
AATGCGATGGC
```

```
Catalytically dead mutants:
>HAstV1 RdRp D360A D361A
GAQNYHSLDAWKLLLEPPRERRCVPANFPLLGH
LPINRPIFDDKKPRDDLLGLLPEPTWHAFEEYGPTTWGPQ
AFIKSFDKFFYAEPIDFFSEYPQLCAFADWATYREFRYLE
DTRVIHITATEKNTDSTPAYPKMNYFDTEENYLEAHGWAP
YIREFTRVYKGDKPEVLWYLFLKKEIIKEEKIRNSDIRQI
VCADPIYTRIGACLEAHQNALMKQHTDTSVGQCGWSPMEG
GFKKTMQRLVNKGNKHFIEFDWTRYDGTIPPALFKHIKEI
RWNFINKDQREKYRHVHEWYVNNLLNRHVLLPSGEVTLQT
RGNPSGQFSTTMDNNMVNFWLQAFEFAYFNGPDRDLWKTY
DTVVYGDDRLSTTPSVPDDYEERVITMYRDIFGMWVKPGK
VICRDSIVGLSFCGFTVNENLEPVPTSPEKLMASLLKPYK
ILPDLESLHGKLLCYQLLAAFMAEDHPFKVYVEHCLSRTA
KQLRDSGLPARLTEEQLHRIWRGGPKKCDG
```

HAstV1 VPg

Listed here are the HAstV1 VPg amino acid and nucleotide sequences used for the expressions described above.

```
>sp|Q67726|665-755 (Uniprot)
KKKGKTKHGRGRVRRNLRKGVKLLTEEEYRELLEKGLDRETFLDLIDR
IIGERSGYPDYDDEDYYDEDDDGWGMVGDDVEFDYTEVINFDQ
Supposedly poor half life due to n-terminal lysines,
so prefer n-terminal tag, (or add Gly, other non-
destabilizing aa)

91aa, 10.8 kDa, 4.59 pI as is without tag
```

```
>HAstV1-VPg\_ml tool optimize seq
AAAAAGAAAGGGAAAACCAAGCACGGCCGTGGCCGTGTCCGTCGCAATC
TTCGTAAAGGGGTCAAACTGCTGACGGAGGAAGAATATCGTGAACTTTT
GGAAAAGGGATTGGACCGTGAAACTTTCTTGGACTTGATTGATCGCATT
ATCGGGGAACGCTCGGGCTACCCAGATTATGACGATGAAGACTACTACG
ATGAAGACGACGATGGTTGGGGGATGGTGGGCGACGACGTTGAGTTTGA
CTATACTGAGGTAATCAACTTCGATCAA
```

3.4.9 DNA encoding RNA sequences

Listed below are the DNA sequences designed to encode various RNA sequences. These sequences were cloned into the pRAV23 vector to act as template or primers in HAstV RdRp experiments after transcription by T7 polymerase. Each are from the HAstV1 genome with accession number MK059949.

```
>HAstV1-MK059949-subgenome ~410bp
CCAAGAGGGGGGTGGTGATTGGCCTTTGGCTTATCAGTGTATAGATAACAT
TTCTCTGCTCGTTTACCATACAATTAACAACAAGATGGCACACGGTGAGCC
ATACTATAGCTCTAAACCTGACAAAGATTTCAATTTTGGAAGCACAATGGC
ACGTAGGCAGATGACACCTACTATGGTAATAAAGCTTCCCAAATTTGTTAG
GAATTCGTCTCGGGGGTAAGCCCAGGCAATGCATGGTCTCATGCGTGTGAA
GAGGCACGCAAAGCAGTAGGGGAAATCAATCCCTGCACATCTGGAAGCCGC
GGCCACGCCGAGTAGGAACGAGGGTACAGCTTCCTTCTTTTCTGTCTCTGT
TTAGATTATTTTAATCACCATTTAAAATTGATTTAATCAGAAGC

>HAstV1-MK059949-subgenome-reverse-complement
GCTTCTGATTAAATCAATTTTAAATGGTGATTAAAATAATCTAAACAGAGA
CAGAAAAGAAGGAAGCTGTACCCTCGTTCCTACTCGGCGTGGCCGCGGCTT
CCAGATGTGCAGGGATTGATTTCCCCTACTGCTTTGCGTGCCTCTTCACAC
GCATGAGACCATGCATTGCCTGGGCTTACCCCCGAGACGAATTCCTAACAA
ATTTGGGAAGCTTTATTACCATAGTAGGTGTCATCTGCCTACGTGCCATTG
TGCTTCCAAAATTGAAATCTTTGTCAGGTTTAGAGCTATAGTATGGCTCAC
CGTGTGCCATCTTGTTGTTAATTGTATGGTAAACGAGCAGAGAAATGTTAT
CTATACACTGATAAGCCAAAGGCCAATCACCACCCCCCTCTTGG

>HAstV1 5' short 1-76
CCAAGAGGGGGGTGGTGATTGGCCTTTGGCTTATCAGTGTATAGATAACATTTCTCTGCTCGTTTACCATACAATT

>HAstV1 5' short 1-76 (reverse complement)
AATTGTATGGTAAACGAGCAGAGAAATGTTATCTATACACTGATAAGCCAAAGGCCAATCACCACCCCCCTCTTGG

>HAstV1 3' short 6709-6771
TTCTTTTCTGTCTCTGTTTAGATTATTTTAATCACCATTTAAAATTGATTTAATCAGAAGC

>HAstV1 3' short 6709-6771 (reverse complement)
GCTTCTGATTAAATCAATTTTAAATGGTGATTAAAATAATCTAAACAGAGACAGAAAAGAA
```

Other sequences

T7 polymerase P266L from AddGene Plasmid #174866 [6].

Histidine tagged MS2 binding motif gene from AddGene Plasmid #67717 [1].

pRAV23 plasmid with RNA template with C-terminal GlmS riboswitch and MS2 motif from AddGene Plasmid #67715 [1].

APPENDIX

Appendix A

Sample Information Tables

Internal ID	*Patient ID*	*Sample SKU*	*Aliquot Serial*	*Gender*	*Age*	*Race*	*Ethnicity*	*Region*	*Sample Draw date*
DLS-1	110033775	BBP0500-A5110033775072514P0	173754487	Male	49	Black	Non-Hispanic/Latino	Huntsville, AL, USA	7/25/14
DLS-2	110032786	BBP0500-A5110032786072313P0	150952714	Male	59	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/23/13
DLS-3	110000224	BBP0500-A5110000224072313P0	156150353	Male	66	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/23/13
DLS-4	110027059	BBP0500-A5110027059110813P0	156151663	Female	35	White	Non-Hispanic/Latino	Huntsville, AL, USA	11/8/13
DLS-5	110000219	BBP0500-A5110000219110713P0	156151819	Male	65	Black	Non-Hispanic/Latino	Huntsville, AL, USA	11/7/13
DLS-6	110035404	BBP0500-A5110035404080613P0	150954519	Male	70	White	Non-Hispanic/Latino	Huntsville, AL, USA	8/6/13
DLS-7	110033582	BBP0500-A5110033582012714P0	159302610	Male	38	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/27/14
DLS-8	110000291	BBP0500-A5110000291012214P0	158368339	Female	56	White	Hispanic/Latino	Huntsville, AL, USA	1/22/14
DLS-9	110027066	BBP0500-A5110027066110513P0	145973823	Male	32	White	Non-Hispanic/Latino	Huntsville, AL, USA	11/5/13
DLS-10	110033642	BBP0500-A5110033642020514P0	159302017	Female	50	White	Non-Hispanic/Latino	Huntsville, AL, USA	2/5/14
DLS-11	110033707	BBP0500-A5110033707021414P0	162046387	Female	36	White	Hispanic/Latino	Huntsville, AL, USA	2/14/14
DLS-12	110033646	BBP0500-A5110033646021414P0	158386097	Male	39	White	Non-Hispanic/Latino	Huntsville, AL, USA	2/14/14
DLS-13	110027070	BBP0500-A5110027070080812P0	128358537	Male	40	Black	Non-Hispanic/Latino	Huntsville, AL, USA	8/8/12
DLS-14	110035403	BBP0500-A5110035403081613P0	156827302	Male	65	White	Hispanic/Latino	Huntsville, AL, USA	8/16/13
DLS-15	110033711	BBP0500-A5110033711030514P0	162047215	Female	45	White	Non-Hispanic/Latino	Huntsville, AL, USA	3/5/14
DLS-16	110035415	BBP0500-A5110035415081313P0	152418158	Male	68	White	Non-Hispanic/Latino	Huntsville, AL, USA	8/13/13
DLS-17	110032789	BBP0500-A5110032789111513P0	156152366	Female	23	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/31/14
DLS-18	110035416	BBP0500-A5110035416100713SH	141531749	Male	72	White	Non-Hispanic/Latino	Huntsville, AL, USA	10/7/13
DLS-19	110029613	BBP0500-A5110029613112113P0	141502563	Male	56	Black	Non-Hispanic/Latino	Huntsville, AL, USA	11/21/13
DLS-20	110033565	BBP0500-A5110033565112213P0	141504119	Male	30	Asian	Non-Hispanic/Latino	Huntsville, AL, USA	11/22/13
DLS-21	110033572	BBP0500-A5110033572112213P0	141504164	Female	22	White	Hispanic/Latino	Huntsville, AL, USA	11/22/13
DLS-22	110033569	BBP0500-A5110033569112213P0	141504138	Male	35	White	Hispanic/Latino	Huntsville, AL, USA	11/22/13
DLS-23	110000296	BBP0500-A5110000296071913P0	149460809	Male	53	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/19/13
DLS-24	110029648	BBP0500-A5110029648110613P0	145974097	Male	66	Black	Non-Hispanic/Latino	Huntsville, AL, USA	11/6/13
DLS-25	110033618	BBP0500-A5110033618022014P0	162046318	Female	28	Other	Non-Hispanic/Latino	Huntsville, AL, USA	2/20/14
DLS-26	110033587	BBP0500-A5110033587121313P0	158367340	Female	28	Black	Non-Hispanic/Latino	Huntsville, AL, USA	12/13/13
DLS-27	110033761	BBP0500-A5110033761061914P0	174556350	Male	25	Black	Non-Hispanic/Latino	Huntsville, AL, USA	6/19/14
DLS-28	110033765	BBP0500-A5110033765062314P0	174555855	Female	34	White	Non-Hispanic/Latino	Huntsville, AL, USA	6/23/14
DLS-29	110033735	BBP0500-A5110033735041414P0	162034611	Female	41	Black	Non-Hispanic/Latino	Huntsville, AL, USA	4/14/14
DLS-30	110033918	BBP0500-A5110033918092214P0	176336325	Female	39	Asian	Non-Hispanic/Latino	Huntsville, AL, USA	9/22/14
DLS-31	110029632	BBP0500-A5110029632071713P0	149460542	Male	78	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/17/13
DLS-32	110027067	BBP0500-A5110027067101716P0	219566034	Male	50	White	Non-Hispanic/Latino	Huntsville, AL, USA	10/17/16
DLS-33	110036064	BBP0500-A5110036064112315P0	202006509	Male	20	Black	Non-Hispanic/Latino	Huntsville, AL, USA	11/23/15
DLS-34	110036542	BBP0500-Z1110036542020916SH	202065896	Female	22	White	Non-Hispanic/Latino	Huntsville, AL, USA	2/9/16
DLS-35	110000239	BBP0500-A5110000239111113P0	145973691	Female	63	White	Non-Hispanic/Latino	Huntsville, AL, USA	11/11/13
DLS-36	110033246	BBP0500-A5110033246072513P0	150953001	Male	52	White	Hispanic/Latino	Huntsville, AL, USA	7/25/13
DLS-37	110000267	BBP0500-A5110000267080912P0	128359003	Male	63	White	Unknown	Huntsville, AL, USA	8/9/12
DLS-38	110033586	BBP0500-A5110033586121113P0	158366754	Female	30	White	Non-Hispanic/Latino	Huntsville, AL, USA	12/11/13
DLS-39	110033594	BBP0500-A5110033594121113P0	158366732	Female	22	White	Non-Hispanic/Latino	Huntsville, AL, USA	12/11/13
DLS-40	110029622	BBP0500-A5110029622011314P0	159304314	Female	56	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/13/14
DLS-41	110033624	BBP0500-A5110033624011714P0	159301735	Male	55	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/17/14
DLS-42	110033395	BBP0500-A5110033395111313P0	145973555	Female	29	Black	Non-Hispanic/Latino	Huntsville, AL, USA	11/13/13
DLS-43	110033584	BBP0500-A5110033584120613P0	158366873	Female	28	Black	Non-Hispanic/Latino	Huntsville, AL, USA	12/6/13
DLS-44	110033643	BBP0500-A5110033643020414P0	159302402	Male	22	White	Hispanic/Latino	Huntsville, AL, USA	2/4/14
DLS-45	110033590	BBP0500-A5110033590010814P0	158386007	Female	25	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/8/14
DLS-46	110000293	BBP0500-A5110000293011014P0	159303362	Male	29	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/10/14
DLS-47	110033585	BBP0500-A5110033585010314P0	158369210	Male	20	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/3/14
DLS-48	110033622	BBP0500-A5110033622010214P0	158366769	Male	29	Black	Non-Hispanic/Latino	Huntsville, AL, USA	1/2/14
DLS-49	110033640	BBP0500-A5110033640012914P0	159303038	Male	43	Asian	Non-Hispanic/Latino	Huntsville, AL, USA	1/29/14
DLS-50	110033626	BBP0500-A5110033626123013P0	158385555	Male	19	White	Non-Hispanic/Latino	Huntsville, AL, USA	12/30/13
DLS-51	110033628	BBP0500-A5110033628123113P0	158369324	Female	25	White	Non-Hispanic/Latino	Huntsville, AL, USA	12/31/13
DLS-52	110033570	BBP0500-A5110033570010314P0	158369292	Female	26	White	Non-Hispanic/Latino	Huntsville, AL, USA	1/3/14
DLS-53	110033632	BBP0500-A5110033632011514P0	159301330	Female	30	American Indian	Non-Hispanic/Latino	Huntsville, AL, USA	1/15/14
DLS-54	110033235	BBP0500-A5110033235121913P0	158385538	Female	21	White	Non-Hispanic/Latino	Huntsville, AL, USA	12/19/13
DLS-55	110033621	BBP0500-A5110033621122713P0	158386170	Female	26	American Indian	Non-Hispanic/Latino	Huntsville, AL, USA	12/27/13
DLS-56	110033770	BBP0500-A5110033770062514P0	174555254	Female	33	White	Non-Hispanic/Latino	Huntsville, AL, USA	6/25/14
DLS-57	110033747	BBP0500-A5110033747050514P0	158382857	Female	48	Black	Non-Hispanic/Latino	Huntsville, AL, USA	5/5/14
DLS-58	110033767	BBP0500-A5110033767070714P0	174555619	Female	44	Black	Non-Hispanic/Latino	Huntsville, AL, USA	7/7/14
DLS-59	110033243	BBP0500-A5110033243052214P0	162047964	Female	38	Black	Non-Hispanic/Latino	Huntsville, AL, USA	5/22/14
DLS-60	110033759	BBP0500-A5110033759061714P0	174556297	Female	40	Black	Non-Hispanic/Latino	Huntsville, AL, USA	6/17/14
DLS-61	110033644	BBP0500-A5110033644061214P0	173753188	Female	32	Black	Non-Hispanic/Latino	Huntsville, AL, USA	6/12/14
DLS-62	110033721	BBP0500-A5110033721043014P0	162047783	Female	27	Black	Non-Hispanic/Latino	Huntsville, AL, USA	4/30/14
DLS-63	110033915	BBP0500-A5110033915082814P0	176337760	Male	43	Black	Non-Hispanic/Latino	Huntsville, AL, USA	8/28/14
DLS-64	110033236	BBP0500-A5110033236072913P0	156150835	Male	48	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/29/13
DLS-65	110036096	BBP0500-A5110036096031416P0	202067167	Male	45	White	Non-Hispanic/Latino	Huntsville, AL, USA	3/14/16
DLS-66	110032771	BBP0500-A5110032771110713P0	156151803	Female	47	White	Non-Hispanic/Latino	Huntsville, AL, USA	11/7/13
DLS-67	110033549	BBP0500-A5110033549110713P0	156151743	Female	29	White	Hispanic/Latino	Huntsville, AL, USA	11/7/13
DLS-68	110035411	BBP0500-A5110035411080913P0	152418774	Male	67	White	Non-Hispanic/Latino	Huntsville, AL, USA	8/9/13
DLS-69	110029656	BBP0500-A5110029656072914P0	173752559	Female	42	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/29/14
DLS-70	110033739	BBP0500-A5110033739072814P0	173754507	Female	65	White	Non-Hispanic/Latino	Huntsville, AL, USA	7/28/14
DLS-71	110029614	BBP0500-A5110029614111313P0	145973737	Male	68	Black	Non-Hispanic/Latino	Huntsville, AL, USA	11/13/13
DLS-72	110000208	BBP0500-A5110000208111313P0	145973706	Female	61	White	Non-Hispanic/Latino	Huntsville, AL, USA	11/13/13
DLS-73	110035420	BBP0500-A5110035420081313P0	152418602	Male	76	White	Non-Hispanic/Latino	Huntsville, AL, USA	8/13/13

Table A.1 Pre-December 2019 samples from Discovery Life Sciences

INTERNAL ID	*Patient ID*	*Sample SKU*	*Gender*	*Age*	*Race*	*Region*	*Sample Draw date*
LDBB-515	515	LD000515	Male	27	Hispanic or Latino	Long Island, NY, USA	2017-April
LDBB-523	523	LD000523	Male	30	Asian	Long Island, NY, USA	Pre 2018
LDBB-526	526	LD000526	Male	59	White	Long Island, NY, USA	2017-April
LDBB-538	538	LD000538	Female	49	Hispanic or Latino	Long Island, NY, USA	2017-May
LDBB-540	540	LD000540	Female	31	Hispanic	Long Island, NY, USA	Pre 2018
LDBB-561	561	LD000561	Male	58	White	Long Island, NY, USA	Pre 2018
LDBB-587	587	LD000587	Male	26	Hispanic	Long Island, NY, USA	Pre 2018
LDBB-610	610	LD000610	Male	65	White	Long Island, NY, USA	2017-June
LDBB-611	611	LD000611	Female	53	White	Long Island, NY, USA	2017-June
LDBB-629	614	LD000614	Male	58	White	Long Island, NY, USA	Pre 2018
LDBB-640	640	LD000640	Female	49	White	Long Island, NY, USA	2017-July
LDBB-657	657	LD000657	Female	52	Hispanic	Long Island, NY, USA	Pre 2018
LDBB-658	658	LD000658	Male	15	White	Long Island, NY, USA	Pre 2018
LDBB-659	659	LD000659	Male	53	White	Long Island, NY, USA	Pre 2018
LDBB-662	662	LD000662	Male	84	White	Long Island, NY, USA	Pre 2018
LDBB-663	663	LD000663	Male	43	Hispanic or Latino	Long Island, NY, USA	2017-August
LDBB-664	664	LD000664	Female	36	Hispanic or Latino	Long Island, NY, USA	2017-August
LDBB-668	668	LD000668	Male	30	White	Long Island, NY, USA	Pre 2018
LDBB-674	674	LD000674	Male	36	White	Long Island, NY, USA	2017-August
LDBB-677	677	LD000677	Female	36	Hispanic or Latino	Long Island, NY, USA	2017-October
LDBB-682	673	LD000682	Male	65	White	Long Island, NY, USA	2017-November

Table A.2 PRE-DECEMBER 2019 SAMPLES FROM THE LYME DISEASE BIOBANK

TABLE 8. HIV reactive Pre-December 2019 samples from Discovery Life Sciences

Internal ID	*Patient ID*	*Sample SKU*	*Gender*	*Age*	*Race/ Ethnicity*	*Region*	*Sample Draw date*	Test 1	Measure	Unit of Measure	Detail 1	Test 2	Test Data 2
DHIV-1	900060022	CIT0000-CS9000600220721111DD	Female	24	Unknown	Spain	7/21/11	Abbott Prism	126	S/Co	HIV	Immunoblot	gp120(4+) gp41(4+) 2ENV(-) p31(4+) p24(4+)
DHIV-2	900066327	CIT0000-CS9000663270507120DD	Female	24	Unknown	Spain	5/7/12	Abbott Prism	121	S/Co	HIV	Immunoblot	gp120(4+) gp41(4+) 2ENV(+/-) p31(2+) p24(4+)
DHIV-3	900066328	CIT0000-CS9000663280531112DD	Male	27	Unknown	Spain	5/31/12	Abbott Prism	131	S/Co	HIV	Immunoblot	gp120(2+) gp41(4+) 2ENV(-) p31(4+) p24(2+)
DHIV-4	900066329	CIT0000-CS9000663290607120DD	Female	35	Unknown	Spain	6/7/12	Abbott Prism	119	S/Co	HIV	Immunoblot	Gp120(3+) gp41(4+) 2ENV(-) p31(3+) p24(3+)
DHIV-5	900066331	CIT0000-CS9000663311111012DD	Male	19	Unknown	Spain	11/10/12	Abbott Prism	146.5	S/Co	HIV	Immunoblot	gp120(2+) gp41(4+) p31(3+) p24(2+) p17(3+)

Table A.3 HIV reactive Pre-December 2019 samples from Discovery Life Sciences

Internal ID	*Patient ID*	*Sample SKU*	*Gender*	*Age*	*Race*	*Region*	*Sample Draw date*	Test 1	Measure	Unit of Measure	Detail 1	Test 2	Test Data 2
DHCV-1	900103674	EDP0000-AI9001036741015 18DD	Female	65	Black	USA	10/15/18	Abbott Architect	14.21 +	N/A	HCV Ab	N/A	N/A
DHCV-2	900062063	EDP0000-AX900062063070215DD	Male	37	Unknown	USA	7/2/15	Roche cobas	3821825	IU/mL	HCV	N/A	N/A
DHCV-3	900066950	EDP0000-AX900066950080615DD	Female	57	Unknown	USA	8/6/15	Roche cobas	1444340	IU/mL	HCV	Roche cobas TaqMan HCV test, v2.0	1,259,324 U/mL
DHCV-4	900105493	EDP0000-AX900105493080615DD	Male	53	Unknown	USA	8/6/15	Roche cobas	340068	IU/mL	HCV	N/A	N/A
DHCV-5	900103675	EDP0000-AI9001036751008 18DD	Female	46	American Indian	USA	10/8/18	Abbott Architect	7.25 +	N/A	HCV Ab	N/A	N/A

Table A.4 HCV reactive Pre-December 2019 samples from Discovery Life Sciences

Appendix B

Detailed Methods

B.1 Code snippets

Functions used to do kinetics fitting of BLI-data. Not shown is the data-wrangling code to prepare the time-series data for the 'plot_curve_fits()' function.

```
import numpy as np
import matplotlib.pyplot as plt
import math
import scipy.signal as scipy
from scipy.optimize import curve_fit
from functools import partial

def association_function(t,Rmax,ka,kd,C):
    return (C*ka*Rmax*(1-math.e**(-((C*ka+kd)*t))))/(C*ka+kd)

def dissociation_function(t,kd,D):
    return D*math.e**(-kd*t)

def steady_state_function(C, Rmax, kD):
    return (C*Rmax)/(C+kD)

def fit_association(x_data,y_data,C,kd=None):
    ka_guess = 10.0**5  # ka in M-1s-1
    kd_guess = 10.0**-2 # kd in s-1
    Rmax = np.max(y_data)
    ka_bounds = [0.0,10**9]
    kd_bounds = [0.0,10.0]
    Rmax_bounds = [Rmax-0.5,Rmax+0.5]

    if kd is not None:
        bounds = ([Rmax_bounds[0], ka_bounds[0]], [Rmax_bounds[1], ka_bounds
    [1]])
        initial_guess = (Rmax,ka_guess)
        association_func = partial(association_function,kd=kd,C=C)
        params, covariance = curve_fit(association_func, x_data, y_data, p0=
    initial_guess, bounds = bounds)
        Rmax,ka = params
    else:
```

```
        bounds = ([Rmax_bounds[0], ka_bounds[0], kd_bounds[0]], [Rmax_bounds[1], ka_bounds[1], kd_bounds[1]])
        initial_guess = (Rmax,ka_guess,kd_guess)
        association_func = partial(association_function,C=C)

        params, covariance = curve_fit(association_func, x_data, y_data, p0=initial_guess, bounds = bounds)
        Rmax,ka,kd = params
    y_fit = association_function(x_data,Rmax,ka,kd,C)

    return y_fit, Rmax,ka,kd

def fit_dissociation(x_data, y_data):
    kd_guess = 10.0**-2 # kd in s-1
    D_guess = y_data[1]

    kd_bounds = [0.0,10.0]
    D_bounds = [D_guess-0.1,D_guess+0.1]
    bounds = ([kd_bounds[0],D_bounds[0]],[kd_bounds[1],D_bounds[1]])
    initial_guess = (kd_guess, D_guess)
    params, covariance = curve_fit(dissociation_function, x_data, y_data, p0=initial_guess, bounds=bounds )
    kd, D = params
    y_fit = dissociation_function(x_data, kd, D)
    return y_fit, kd, D

def plot_curve_fits(x,y,label,plot,restrict_kd=False):
    ### do curve fits
    #initialize
    C=0
    association_stop = 300
    dissociation_stop = 900


    if is_number(label):
        C = float(label)*10**(-9)

    ## dissociation
    t0 = int(association_stop / timestep) - 1
    x_dissoc = x[t0:]
    y_dissoc = y[t0:]
    y_fit_dissoc, kd2, D = fit_dissociation(x_dissoc-association_stop,y_dissoc)
    plot.plot(x_dissoc, y_fit_dissoc, linewidth = 0.6, color = 'red', label = label)

    # association
    t0 = int(association_stop / timestep) - 1
    x_assoc = x[:t0]
    y_assoc = y[:t0]
    if restrict_kd:
        #kd1=kd2
        y_fit_assoc, Rmax,ka,kd1 = fit_association(x_assoc,y_assoc,C,kd=kd2)
    else:
        y_fit_assoc, Rmax,ka,kd1 = fit_association(x_assoc,y_assoc,C)
    plot.plot(x_assoc, y_fit_assoc, linewidth = 0.6, color = 'red', label = label)

    KD1 = kd1/ka
    KD2 = kd2/ka
    fit_values = [label, Rmax, ka, kd1, D, kd2, KD1, KD2]
```

```
    return fit_values
```

Listing B.1 Fitting BLI data with scipy.curve_fit

The code below was used for our BLI-ISA paper to perform AUC analysis via non-linear regression and integration with Python for the ELISA assays in that study.

```
# regression_plotter.py
# plots and fits curve to x, y data (eg ELISA/ other assay), can do AUC calcs
# Nicholas Lorig-Roach
# 2020-06-16
#
# reference websites:
# https://stackoverflow.com/questions/46674870/least-squares-function-and-4-parameter-logistics-function-not-working
# https://www.biolegend.com/en-us/blog/curve-fitting-for-immunoassays-legendplex
# scipy curve fit page

import sys
import matplotlib.pyplot as plt
import matplotlib.patches as mplpatches
import numpy as np
import pandas as pd
from scipy.optimize import curve_fit
import scipy.integrate as integrate
from scipy import stats

def func(x, a, b, c):
    result = a * np.exp(-b * x) + c
    return result

def linear(x, a, b):
    return a * x + b

def fourPL(x, a, b, c, d):
    #print('1: ', (a-d))
    #print('2: ', (x/c))
    #print('2: ', (1.0 + (x / c) **b))
    return d + (a-d) / (1.0 + (x / c)**b)

def fourPL_regression_plot(x, y, ax, label = '', \
                legend = True, eq = True, title = True, color = 'black', mark = 'x'):
    """
        performs four paramater logistic regression using supplied xy values
    """
    # define four parameter function and initial values for its coefficients
    function = fourPL
    guess = [0, -0.5, 0.5, 1]

    # do the curve fit, then calculate the error and R2 based on the covariance matrix (pcov)
    popt, pcov = curve_fit(function, x, y, guess)
    perr = np.sqrt(np.diag(pcov))
    R2 = np.sum((fourPL(x, *popt)-y.mean())**2) / np.sum((y-y.mean())**2)
```

```
46
47      # do the plotting
48      x_min, x_max = np.amin(x), np.amax(x)
49      xs = np.linspace (x_min, x_max, 1000)  # needed for smooth curve
50      ax.plot(xs, fourPL(xs, *popt), '--', linewidth = 0.6, c = color, \
51              label = '{} fit: a = {:.2f}, b = {:.2f}, c = {:.2f}, d = {:.2f},
    $R^2$ = {:.4f}'.format(label, *popt, R2))
52
53      ax.scatter(x,y, s = 8,  c = color, marker = mark, linewidths = 0.6)
54      ### AUC calculations
55      # set upper and lower bound of integration function, then do integral
56      lower, upper = x_min, x_max
57      auc = integrate.quad(function, lower, upper, args=(*popt,))
58
59      # if eq flag is True, print equation with coefficients on graph
60      if eq:
61          ax.text(2000, 1.9, r'$y(x) = d + \frac{a - d}{1 + \left( \frac{x}{c}\
    right)^b}$', va = 'bottom', ha = 'left', fontsize = 7)
62
63      # double curly braces to 'escape' braces intended only for latex parser
64          ax.text(2000, 1.3, 'AUC = $\int_{{{}}}^{{{}}} y(x) = {{{:.1f}}}$'.
    format(lower, upper, auc[0]), va = 'bottom', ha = 'left', fontsize = 7)
65
66      ## finishing up figure
67      ax.set_xlabel('Reciprocal Dilution')
68      ax.set_ylabel('$OD_{490}$')
69      if title:
70          ax.set_title(label)
71      if legend:
72          ax.legend()
73      return auc, popt, R2
```

Listing B.2 Functions for fitting curves to e.g. ELISA data and performing area-under-the-curve (AUC) calculations.

REFERENCES

Note: references from article reprints are printed at the end of their respective sections. Listed here are references from introductory sections or previously unpublished work.

[1] Robert T. Batey and Jeffrey S. Kieft. "Improved native affinity purification of RNA". In: *RNA* 13.8 (Aug. 2007), pp. 1384–1389. DOI: 10.1261/rna.528007.

[2] Jules J. Berman. "Group IV Viruses". In: *Taxonomic Guide to Infectious Diseases* (2012), pp. 237–246. DOI: 10.1016/B978-0-12-415895-5.00042-8.

[3] Broad Institute TCGA Genome Data Analysis Center. *Analysis-ready standardized TCGA data from Broad GDAC Firehose 2016_01_28 run*. 2016.

[4] Sarah J. Butcher et al. "A mechanism for initiating RNA-dependent RNA polymerization". In: *Nature* 410.6825 (Mar. 2001), pp. 235–240. DOI: 10.1038/35065653.

[5] Cheng Chen et al. "Crystal Structure of Enterovirus 71 RNA-Dependent RNA Polymerase Complexed with Its Protein Primer VPg: Implication for a trans Mechanism of VPg Uridylylation". In: *Journal of Virology* 87.10 (May 15, 2013), pp. 5755–5768. DOI: 10.1128/JVI.02733-12.

[6] Isabel Chillón et al. "Native Purification and Analysis of Long RNAs". In: *Methods in Enzymology* 558 (2015), pp. 3–37. DOI: 10.1016/bs.mie.2015.01.008.

[7] Thomas Conrad et al. "Maximizing transcription of nucleic acids with efficient T7 promoters". In: *Communications Biology* 3.1 (Aug. 14, 2020), pp. 1–8. DOI: 10.1038/s42003-020-01167-x.

[8] Marc Damelin et al. "A PTK7-targeted antibody-drug conjugate reduces tumor-initiating cells and induces sustained tumor regressions". en. In: *Science Translational Medicine* 9.372 (Jan. 2017), eaag2611. DOI: 10.1126/scitranslmed.aag2611.

[9] Cristina Ferrer-Orta et al. "The structure of a protein primer–polymerase complex in the initiation of genome replication". In: *The EMBO Journal* 25.4 (Feb. 22, 2006), pp. 880–888. DOI: 10.1038/sj.emboj.7600971.

[10] Hongguang Fu et al. "Codon optimization with deep learning to enhance protein expression". In: *Scientific Reports* 10.1 (Oct. 19, 2020), p. 17617. DOI: 10.1038/s41598-020-74091-z.

[11] Cristina Fuentes et al. "Identification of Human Astrovirus Genome-Linked Protein (VPg) Essential for Virus Infectivity". In: *Journal of Virology* 86.18 (Sept. 15, 2012), pp. 10070–10078. DOI: 10.1128/JVI.00797-12.

[12] Stephen W. B. Fullerton et al. "Structural and Functional Characterization of Sapovirus RNA-Dependent RNA Polymerase". In: *Journal of Virology* (Feb. 2007). DOI: 10.1128/JVI.01462-06.

[13] Peng Gong and Olve B. Peersen. "Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase". In: *Proceedings of the National Academy of Sciences* 107.52 (Dec. 28, 2010), pp. 22505–22510. DOI: 10.1073/pnas.1007626107.

[14] Kristian A. Gray et al. "Genenames.org: the HGNC resources in 2015". eng. In: *Nucleic Acids Research* 43.Database issue (Jan. 2015), pp. D1079–1085. DOI: 10.1093/nar/gku1071.

[15] Arnaud Gruez et al. "The Crystal Structure of Coxsackievirus B3 RNA-Dependent RNA Polymerase in Complex with Its Protein Primer VPg Confirms the Existence of a Second VPg Binding Site on Picornaviridae Polymerases". In: *Journal of Virology* 82.19 (Oct. 2008), pp. 9577–9590. DOI: 10.1128/JVI.00631-08.

[16] Remco van Horssen, Timo L. M. ten Hagen, and Alexander M. M. Eggermont. "TNF-alpha in Cancer Treatment: Molecular Insights, Antitumor Effects, and Clinical Utility". en. In: *The Oncologist* 11.4 (Apr. 2006), pp. 397–408. DOI: 10.1634/theoncologist.11-4-397.

[17] B Jiang et al. "RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis." In: *Proceedings of the National Academy of Sciences* 90.22 (Nov. 15, 1993), pp. 10539–10543. DOI: 10.1073/pnas.90.22.10539.

[18] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[19] Zachary J. Kartje et al. "Revisiting T7 RNA polymerase transcription in vitro with the Broccoli RNA aptamer as a simplified real-time fluorescent reporter". In: *Journal of Biological Chemistry* 296 (Jan. 1, 2021), p. 100175. DOI: 10.1074/jbc.RA120.014553.

[20] Ji-Hye Lee et al. "Insight Into the Interaction Between RNA Polymerase and VPg for Murine Norovirus Replication". In: *Frontiers in Microbiology* 9 (2018).

[21] Mark A. Lemmon and Joseph Schlessinger. "Cell signaling by receptor-tyrosine kinases". In: *Cell* 141.7 (June 2010), pp. 1117–1134. DOI: 10.1016/j.cell.2010.06.011.

[22] Bo Li and Colin N Dewey. "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome". en. In: *BMC Bioinformatics* 12.1 (2011), p. 323. DOI: 10.1186/1471-2105-12-323.

[23] Michael I. Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15 (2014), p. 550. DOI: 10.1186/s13059-014-0550-8.

[24] Asher Mullard. "2015 FDA drug approvals". en. In: *Nature Reviews Drug Discovery* 15.2 (Feb. 2016), pp. 73–76. DOI: 10.1038/nrd.2016.15.

[25] Badr Al-Mutairy et al. "Genome Prediction of Putative Genome-Linked Viral Protein (VPg) of Astroviruses". In: *Virus Genes* 31.1 (Aug. 1, 2005), pp. 21–30. DOI: 10.1007/s11262-004-2196-1.

[26] Aaron L. Nelson, Eugen Dhimolea, and Janice M. Reichert. "Development trends for human monoclonal antibody therapeutics". en. In: *Nature Reviews Drug Discovery* 9.10 (Oct. 2010), pp. 767–774. DOI: 10.1038/nrd3229.

[27] Kenneth K. S. Ng et al. "Crystal Structures of Active and Inactive Conformations of a Caliciviral RNA-dependent RNA Polymerase*". In: *Journal of Biological Chemistry* 277.2 (Jan. 11, 2002), pp. 1381–1387. DOI: 10.1074/jbc.M109261200.

[28] D. J. Oshannessy et al. "Determination of Rate and Equilibrium Binding Constants for Macromolecular Interactions Using Surface Plasmon Resonance: Use of Nonlinear Least Squares Analysis Methods". In: *Analytical Biochemistry* 212.2 (Aug. 1, 1993), pp. 457–468. DOI: 10.1006/abio.1993.1355.

[29] Joshua B. Plotkin and Grzegorz Kudla. "Synonymous but not the same: the causes and consequences of codon bias". In: *Nature Reviews Genetics* 12.1 (Jan. 2011), pp. 32–42. DOI: 10.1038/nrg2899.

[30] Donald Charles Rio et al. *RNA: a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press, 2011. 586 pp.

[31] Yanira Sáez-Álvarez et al. "Development of a fluorescence-based method for the rapid determination of Zika virus polymerase activity and the screening of antiviral drugs". In: *Scientific Reports* 9.1 (Apr. 1, 2019), p. 5397. DOI: 10.1038/s41598-019-41998-1.

[32] Andrew Savinov and Steven M. Block. "Self-cleavage of the glmS ribozyme core is controlled by a fragile folding element". In: *Proceedings of the National Academy of Sciences* 115.47 (Nov. 20, 2018), pp. 11976–11981. DOI: 10.1073/pnas.1812122115.

[33] Andrew M. Scott, Jedd D. Wolchok, and Lloyd J. Old. "Antibody therapy of cancer". en. In: *Nature Reviews Cancer* 12.4 (Apr. 2012), pp. 278–287. DOI: 10.1038/nrc3236.

[34] Ian Shieh, Danielle Leiske, and Ankit Patel. "Dynamics of interfacial adsorption and self-association of antibody therapeutics." In: *Abstracts of Papers, 247th ACS National Meeting and Exposition (Dallas, TX, United States)* (Mar. 2014).

[35] Bernhard B. Singer et al. "Deregulation of the CEACAM Expression Pattern Causes Undifferentiated Cell Growth in Human Lung Adenocarcinoma Cells". In: *PLOS ONE* 5.1 (Jan. 2010), e8747. DOI: 10.1371/journal.pone.0008747.

[36] Hong Tan et al. "Fiber-optic assay apparatus based on phase-shift interferometry". U.S. pat. 7319525B2. Fortebio Inc. Jan. 15, 2008.

[37] THE GTEX CONSORTIUM et al. "The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans". In: *Science* 348.6235 (May 8, 2015), pp. 648–660. DOI: 10.1126/science.1262110.

[38] Aartjan J. W. te Velthuis. "Common and unique features of viral RNA-dependent polymerases". In: *Cellular and Molecular Life Sciences* 71.22 (Nov. 1, 2014), pp. 4403–4420. DOI: 10.1007/s00018-014-1695-z.

[39] Sangita Venkataraman, Burra V L S Prasad, and Ramasamy Selvarajan. "RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution". In: *Viruses* 10.2 (Feb. 10, 2018), p. 76. DOI: 10.3390/v10020076.

[40] John Vivian et al. "Toil enables reproducible, open source, big biomedical data analyses". en. In: *Nature Biotechnology* 35.4 (Apr. 2017), pp. 314–316. DOI: 10.1038/nbt.3772.

[41] Bert Vogelstein et al. "Cancer Genome Landscapes". en. In: *Science* 339.6127 (Mar. 2013), pp. 1546–1558. DOI: 10.1126/science.1235122.

[42] Dennis Wat. "The common cold: a review of the literature". In: *European Journal of Internal Medicine* 15.2 (Apr. 2004), pp. 79–88. DOI: 10.1016/j.ejim.2004.01.006.

[43] George J. Weiner. "Building better monoclonal antibody-based therapeutics". en. In: *Nature Reviews Cancer* 15.6 (June 2015), pp. 361–370. DOI: 10.1038/nrc3930.

[44] John N. Weinstein et al. "The Cancer Genome Atlas Pan-Cancer Analysis Project". In: *Nature genetics* 45.10 (Oct. 2013), pp. 1113–1120. DOI: 10.1038/ng.2764.

[45] Yuri I. Wolf et al. "Origins and Evolution of the Global RNA Virome". In: *mBio* 9.6 (Nov. 27, 2018), e02329–18. DOI: 10.1128/mBio.02329-18.

[46] Chengxin Zhang et al. "US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes". In: *Nature Methods* 19.9 (Sept. 2022), pp. 1109–1115. DOI: 10.1038/s41592-022-01585-1.