**Title**

Utilizing bio-conjugation and protein engineering to develop enhanced cancer therapeutics, viral diagnostics, and biochemical tools

**Permalink**

https://escholarship.org/uc/item/6f066638

**Author**

Elledge, Susanna Kuroda

**Publication Date**

2022

Peer reviewed|Thesis/dissertation

Utilizing bio-conjugation and protein engineering to develop enhanced cancer therapeutics, viral diagnostics, and biochemical tools
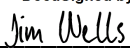
by
Susanna Elledge

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
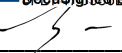DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

_Jim Wells_____     Jim Wells
5F2F4D1A06164C2...                                                                   Chair

_kevan Shokat_____     Kevan Shokat

_____     Zev Gartner
6E99B613A5754DD...

_____

_____
                                                                   Committee Members

*This work is dedicated to my parents, Steve and Mitzi, whose love for science and discovery inspired me from a young age to follow in their footsteps and pursue a career in science. To my brother, Daniel, who has supported me and put up with scientific conversations at the dinner table. And to Jesse, who reminds me every day to find my passion, hold on to it, and nurture it to grow.*

## Acknowledgements

I would not be here today without the mentorship and support that I have received throughout my scientific career. While science may look like an independent working environment, each scientist is made up of individual mentoring efforts from others and the warmth and support of the community around them.

First, I'd like to thank my family, Steve, Mitzi, and Daniel, to whom this work is dedicated. To my parents, thank you for being two wonderful scientific role models that provided a supportive and warm environment for me to grow up in and pursue my passions. Without you two, I wouldn't have a clue what a life in science would look like. To my brother, thank you for your continuing support and for accepting me into the Dr. Elledge club. Second, to my partner Jesse, who supported me through the majority of my time in graduate school, even when I would show up late almost every time after running an experiment in lab. His care and supportive words carried me through the hard times of my PhD, particularly during the pandemic, even when he was 3,000 miles away.

Next, I'd like to thank the UCSF mentors and community that have supported me during my PhD. To Dr. Jim Wells, my PhD mentor – thank you for allowing me to join your lab and work on projects I was passionate about. You were always supportive, no matter good news or bad news and made a wonderful environment to learn and grow in. To Dr. Kevan Shokat and Dr. Zev Gartner, my thesis committee members – thank you for your support and thoughtful questions and suggestions throughout my PhD. I would also like to thank additional members of my qualifying exam committee and rotation mentors, Dr. Arun Witta, Dr. Natalia Jura, and Dr. Tejal Desai.

In addition to the professors, there were many mentors in the Wells lab at UCSF that were essential to my success. Dr. Amy Weeks – my first lab mentor who helped me get started on my first project in the lab; Dr. Sam Pollock – my rotation mentor who took a large amount of his time

I'm also very grateful for all the great collaborators I had over the years at UCSF – Dr. Veronica Steri, Dr. Byron Hann, Dr. Joe Derisi, Sara Sunshine, Dr. Jiuling Yang, Dr. Robert Blelloch, Dr. Bryan Greenhouse, Keirstinne Turcios, Jill Hakim; UC Berkeley - Dr. F. Dean Toste, Dr. Alec Christian, Dr. Helene Wolleb, Dr. Kacper Skakuj, Dr. Chris Chang; CZI Biohub – Dr. Cristina Tato, Maira Phelps, Dr. Paul Lebel, Dr. Diane Wiener, Dr. Ilakkiyan Jeyakumar, Dr. Rafael Sjoberg-Gomez and many others.

I'd be remiss to not thank my scientific mentors before UCSF who helped me get to graduate school in the first place. Thank you to Dr. Matt Francis who allowed me to join his lab as an undergrad and get my first research experience where I fell in love with bioconjugation. A very special thanks to my mentor in the Francis lab, Dr. Jenna Bernard, who taught me how to perform research at the bench and was always willing to sit down with me to plan out a project but also gave me the independence to direct the research how I wanted to. Two other research opportunities in the labs of Dr. Wolfram Goessling and Dr. Takeaki Ozawa opened my eyes to other types of research and helped me further strengthen my research skills.

I'd like to thank my friends, from Boston to Berkeley to San Francisco, that have enriched my life outside of science and have supported my goals to obtain my PhD. Without you all to provide fun and laughter outside of lab, I wouldn't have been able to overcome the many hardships that the PhD road entails.

# Contributions

**Chapter 1:** Adapted version of the following published manuscript: **Susanna K. Elledge**, Hai L. Tran, Alec H. Christian, Veronica Steri, Byron Hann, F. Dean Toste, Christopher J. Chang, James A. Wells. Systematic Identification of engineered methionines and oxaziridines for efficient, stable, and site-specific antibody bioconjugation. *PNAS.* 2020.

**Chapter 2:** Adapted version of the following published manuscript: **Susanna K. Elledge**[*], Xin X. Zhou[*], James R. Byrnes, Alexander J. Martinko, Irene Lui, Katarina Pance, Shion A. Lim, Jeff E. Glasgow, Anum A. Glasgow, Keirstinne Turcios, Nikita S. Iyer, Leonel Torres, Michael J. Peluso, Timothy J. Henrich, Taia T. Wang, Cristina M. Tato, Kevin K. Leung, Bryan Greenhouse & James A. Wells. *Nature Biotechnology.* 2021. [*]Denotes equal contribution to this work.

**Chapter 3:** Adapted version of the following published manuscript: Lisa L. Kirkemo[*], **Susanna K. Elledge**[*], Jiuling Yang, James Byrnes, Jeff Glasgow, Robert Blelloch, James A. Wells. *eLife.* 2022. [*]Denotes equal contribution to this work.

**Utilizing bio-conjugation and protein engineering to develop enhanced cancer therapeutics, viral diagnostics, and biochemical tools**

**Susanna Elledge**

**Abstract**

Bio-conjugation and protein engineering are two powerful biochemical methods to alter proteins to have enhanced functions. Bio-conjugation allows for proteins to be attached to other protein and non-protein molecules in ways that would never appear naturally in the biological world, bestowing new properties onto a protein. Protein engineering allows for manipulation of the linear amino-acid sequence of a protein to alter its biochemical properties. These types of substitutions can change the way a protein folds, how it interacts with partner proteins, and its own chemical reactivity. These methods are currently highly utilized in the development of novel therapeutics and research tools.

For both methods, location is one of the most important considerations for successful applications. Years of bio-conjugation research has revealed that attachment at a controlled and known site improves the reproducibility and properties of a given conjugate. In protein engineering, not all sites on a protein have the same chemical environment and thus even the same amino acid at a different location may react differently. Thus, even with these powerful biochemical methods, careful research must be done to determine the optimal attachment method, mutation location, and site. However, the optimized conjugate or engineered protein has great potential as a novel therapeutic, diagnostic, or research tool.

Here, I describe three approaches using bio-conjugation and protein engineering to develop novel and optimized protein conjugates or engineered protein fusions for various applications. In

Chapter 1, I describe the use of a methionine-based conjugation method to attach cytotoxic cancer drugs on antibodies as selective drug delivery carriers. In Chapter 2, I highlight the development of rapid, facile protein biosensors to detect antibodies to COVID-19. In Chapter 3, I describe the development of an improved cell-surface biotinylation method utilizing novel enzyme conjugates.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

ACE2 – Angiotensin-converting enzyme II

ADC – Antibody-drug conjugate

APEX2 – Ascorbate peroxidase enzyme 2

BLI – Biolayer interferometry

CDR – Complementarity determining region

COVID-19 – Coronavirus disease 2019

DAR – Drug-to-antibody ratio

DBCO – Di-benzyl cyclic octyne

ELISA – Enzyme linked immunosorbent assay

EV – Extracellular vesicles

Fab – Antigen binding fragment

GFP – Green fluorescent protein

GlcNAc – N-Acetylglucosamine

GOCC-PM – Uniprot Gene Ontology Cellular Component Plasma Membrane

HER2 – Human epidermal growth factor receptor 2

HC – Heavy chain

HRP – Horse-radish peroxidase

$K_D$ – Binding affinity

$K_{off}$ – Dissociation rate

$K_{on}$ – Association rate

LC – Light chain

LC-MS/MS – Liquid chromatography – Tandem mass spectrometry

LFA/LFD – Lateral flow assay/device

LFQ – Label-free quantification

MMAE – mono-methyl auristatin E

MMAF – mono-methyl auristatin F

NHS – N-hydroxysuccinimide

NT50 – Neutralization 50%

PCA – Principle component analysis

RBD – Receptor binding domain

ReACT– Redox-activated chemical tagging

RSA – Relative solvent accessibility

SARS-CoV-2 – Severe acute respiratory syndrome coronavirus 2

SASA – Solvent accessible surface area

Spluc – split luciferase assay

SPPLAT – Selective proteomic proximity labeling assay using tyramide

$T_m$ – Unfolding (melting) temperature

vcF – valine-citrulline mono-methyl auristatin F

WGA – Wheat-germ agglutinin

# Chapter 1

**Systematic identification of engineered methionines and oxaziridines for efficient, stable, and site-specific antibody bioconjugation**

**Abstract**

The field of chemical modification of proteins has been dominated by random modification of lysines or more site-specific labeling of cysteines, each with attendant challenges. Recently we have developed oxaziridine chemistry for highly selective modification of methionine called redox-activated chemical tagging (ReACT), but have not broadly tested the molecular parameters for efficient and stable protein modification. Here, we systematically scanned methionines throughout one of the most popular antibody scaffolds, trastuzumab, used for antibody engineering and drug conjugation. We tested the expression, reactivities, and stabilities of 123 single engineered methionines distributed over the surface of the antibody when reacted with oxaziridine. We found uniformly high expression for these mutants and excellent reaction efficiencies with a panel of oxaziridines. Remarkably, the stability to hydrolysis of the sulfimide varied more than ten-fold depending on temperature and the site of the engineered methionine. Interestingly, the most stable and reactive sites were those that were partially buried, presumably because of their reduced access to water. There was also a ten-fold variation in stability depending on the nature of the oxaziridine, which was determined to be inversely correlated with the electrophilic nature of the sulfimide. Importantly, the stabilities of the best analogs were sufficient to support their use as antibody drug conjugates and potent in a breast cancer mouse xenograft model over a month. These studies provide key parameters for broad application of ReACT for efficient, stable, and site-specific antibody and protein bioconjugation to native or engineered methionines.

**Introduction**

Chemical modification of natural amino acids in proteins has a long and storied history but largely limited to modification of thiols and amines with various electrophiles [1]. Recently, a methionine specific chemistry has been developed, called redox-activated chemical tagging (ReACT) (**Fig. 1.1a**)[2]. The ReACT chemistry involves oxidation of methionine to form a sulfimide adduct with an oxaziridine molecule functionalized with an alkyl-azide to allow cargo attachment via click chemistry. Methionine is the second least-abundant residue in proteins after tryptophan[3], making it a potentially ideal target for site-specific conjugation. Most methionine residues are buried and therefore inaccessible, making it an excellent candidate for bioengineered single site chemical conjugation. However, the structural and chemical parameters for reaction of methionines in a folded protein with oxaziradines have not been systematically evaluated for protein bioconjugation.

Antibodies represent the most popular target for bioconjugations because they are among the most versatile tools in biology and medicine[4]. Chemical bioconjugation has been instrumental in expanding the utility of monoclonal antibodies, both as probes and therapeutics, by facilitating covalent attachment of a variety of moieties such as fluorophores[5], metal chelators[6], nucleic acids[7–9], as well as toxins in the form of antibody drug conjugates (ADCs)[10–13]. ADCs can be an improvement over standard chemotherapy treatment by simultaneously increasing targeting efficiency and reducing off-target toxicity[14–17]. There are currently five FDA approved ADCs and more than 100 clinical trials to develop new ADC therapies[14,15].

Ideally the chemical modification should be efficient, stable, reproducible, and site-selective for homogeneity[15]. Researchers have typically targeted lysine or cysteine residues for chemical conjugation because of high reactivity and yield to form stable adducts via amide bonds

with lysines or thioether linkages to cysteines[14–17]. Three out of the five FDA approved ADCs target lysines for conjugation[17]. However, antibodies typically have about 40 surface exposed lysine residues per IgG that can result in more than one million different ADC species and drug antibody ratios (DAR) varying between one to eight[15] which can lead to aggregation, immunogenicity, faster clearance rates and differences in the pharmacodynamic properties of the conjugate[14,18]. Cysteine is becoming more commonly used as it is far less abundant than lysine and affords greater site-selectivity. However, this approach usually involves random reduction of the interchain disulfide bonds of the antibody that can still lead to heterogenous mixtures[14, 19, 207]. Recently, researchers have systematically introduced single cysteine residues into the therapeutic antibody, trastuzumab, to identify sites for stable and specific conjugation[21,22]. Other conjugation strategies are being developed such as enzymatic conjugation, glycan modification, and un-natural amino acid incorporation[14,17]. These strategies can result in homogenous conjugates but can be limited in terms of DAR and involve introducing scars, either by peptide motifs or altered natural glycosylation[17].

Here, we methodically explore ReACT to determine how the methionine site and oxaziridine compound properties affect yield and stability in a folded protein, a therapeutic antibody. We systematically scanned 123 single methionine residues, that vary in accessibility in the trastuzumab scaffold, to identify sites and compounds for optimal conjugation yield and stability. We identified systematic factors that affect stability by more than 10-fold that we believe are portable to other antibodies and proteins. We found sites of high stability, and produced potent ADCs with DARs of two or four that kill tumor cells *in vitro* and *in vivo* over a month. These studies show that the methionine oxaziridine reaction is a promising approach for site-specific, high yielding, and stable protein bioconjugations.

**Results**

*High-throughput scan of the top 95 most accessible sites on the trastuzumab scaffold*

We chose trastuzumab as the antibody scaffold of choice for our studies for a number of reasons. The trastuzumab framework is popular for humanization due to its high stability, high expression in mammalian cells, high developability, and broad use that is now utilized in parts of three different approved antibody drugs (trastuzumab, bevacizumab, and omalizumab) and the TDM-1 anti-Her2 ADC (ado-trastuzumab emtansine). Synthetic complementarity-determining region (CDR) libraries have been constructed on the trastuzumab scaffold[23] and used by the Recombinant Antibody Network for industrialized recombinant antibody generation to over 500 protein targets[24]. The Fab arms in trastuzumab contain three methionines that are buried **(Fig1.2).** Indeed previous studies from our group showed these buried methionines to be unreactive to ReACT but when we attached a single Met to the C-terminus of the light chain we found it could be labeled quantitatively with a simple oxaziridine reagent and conjugated with a fluorophore[2]. While this site can be labelled quantitatively and could be useful for short-term *in vitro* studies, we found it becomes extensively (>80%) hydrolyzed over three days at 37°C **(Fig. 1.3)** and thus is not suitable for long-term studies or ADC development.

To expand the use of ReACT for antibody bioconjugations we sought to systematically determine how methionine mutation, site of labeling, and compound nature affects expression, labeling efficiency, binding affinity, and stability of the antibody (**Fig 1.1a**). We first focused on exposed sites on a well characterized αGFP antibody built on the trastuzumab scaffold as a model for ease of assay[24]. We calculated the surface accessibility of the methionine sulfur for all possible surface methionine substitutions. We mutated the top 95 most accessible sites to methionine (**Fig. 1.1b, Table 1.1**) and expressed each individual mutant in the αGFP Fab without mutating the three

intrinsic and unreactive buried methionines. Remarkably, of those 95 sites, 93 methionine mutants expressed with high yield in *E. coli* (3-18mg/L). All 93 retained high binding affinity for GFP after conjugation with oxaziridine and sulfo-DBCO-NH2, and 92 of those retained high thermostability as measured by differential scanning fluorimetry (DSF). When tested for labeling with 5 equivalents of the oxaziridine reagent (oxaziridine **1**) for 2 hours, 57 mutants labeled to greater than 90% (**Fig. 1.1c, Fig. 1.4**). This could potentially be improved with higher equivalents of oxaziridine. All mutants labeled stoichiometrically and specifically at the mutated methionine residue, as determined by whole protein mass spectrometry (**Fig. 1.1d; Table 1.1**). These data suggest tremendous flexibility in generating site-specific methionine conjugations.

While these sites are likely useful for short-term studies such as immunofluorescence or other *in vitro* studies we wanted to test their suitability for longer term *in vivo* applications . Of the 57 highly labeled sites, we chose 12 representative sites to test conjugation stability as a function of location and temperature (**Fig. 1.5**). The 12 candidate sites spanned both the heavy and light chain, as well as the variable and constant domains of the Fab arm. We incubated each methionine-oxaziridine conjugate at 4°C, 25°C, and 37°C for 3 days and measured the remaining conjugate by whole protein MS (**Fig. 1.5b**). We found a strong temperature dependence for hydrolysis from 4°C, 25°C, and 37°C. There was considerable variation among the sites, but all sites had less than 60% remaining conjugate after 3 days at 37°C. The product had a +16 mass shift consistent with hydrolysis of the sulfimide to a sulfoxide product, which has also been previously reported[25]. Since ADCs can have circulation times up to weeks in the body, it is essential that the linkage is stable for an extended period of time at biological temperatures to retain ADC potency and to eliminate off-target toxicity due to free drug release. Although these

stabilities are sufficient for the many *in vitro* uses for antibody conjugation, we sought to extend the stability of the antibody conjugate for ADCs.

*Enhancing stability of oxaziridine conjugates*

We took two approaches to improve conjugation stability: (1) test different substituted oxaziridine analogs to improve linkage stability and (2) test more buried sites on the Fab scaffold that we hypothesized could better shield the sulfimide from hydrolysis. We obtained 15 different oxaziridine molecules with various functionalities appended to the urea group to determine if the resulting sulfimide bond could be further stabilized (**Fig. 1.5c**). We chose one representative site, LC.T20M, that showed moderate stability at 37°C for oxaziridine **1**. All compounds were conjugated to LC.T20M site on the model αGFP Fab and stability of the sulfimide linkage was measured at 37°C over 3 days. There was considerable variation in stability from 40-90% retained; nonetheless, two of oxaziridines (compound **5** and **8**) provided stability over 80% (**Fig. 1.5d**). In a recent parallel study, it was shown that conjugate stability to isolated methionine was related to the electron density around the carbonyl as measured by the carbonyl stretching frequency[25]. Indeed, we found a strong inverse correlation between carbonyl stretching frequency and the measured stabilities on the Fab (**Fig. 1.5e**) as was also seen with isolated methionine. Interestingly, we also saw the same trend in compound stability for another site on the Fab scaffold (data not shown) and thus we believe these compounds will be more stable for all sites. We synthesized a new azide containing oxaziridine derivative, based on the more stable piperidine-derived oxaziridine **8**, to enable copper-free click chemistry for ADC conjugation (**Fig. 1.5f**).

We next investigated how lowering site accessibility may shield the resulting sulfimide linkage from hydrolysis. We knew that fully buried sites are unreactive, and therefore chose 23

sites that had intermediate degrees of accessibility (**Fig. 1.6a, Table 1.2**) most of which were located on structured β-sheet regions. Remarkably, 19 of the 23 single methionine substitutions at these partially buried sites expressed at high levels in *E. coli* (3-50mg/L); 18 retained high affinity to GFP, and 17 retained high thermostability (**Fig. 1.6b**). These less accessible sites were also less reactive, and thus we increased the labeling reaction to 20 equivalents of oxaziridine to better drive the reactivity. We found four mutants that had greater than 85% stability when labeled with the oxaziridine azide **8** and incubated at 37°C for 3 days (**Fig. 1.6b, Fig. 1.6c**). There was a slight inverse correlation between site accessibility and long-term stability (**Fig. 1.6d**) but the lack of a strong correlation suggests that additional factors are partly responsible. Overall, we found the combination of probing different oxaziridine derivatives and different site accessibility produced highly stable conjugates that were candidates for ADC production.

We next incorporated these mutations into a trastuzumab Fab and tested the ADC conjugates for killing of breast cancer cell lines. However, we noticed that the wild-type trastuzumab Fab labeled 25% with the oxaziridine reagent when reacted at 20 equivalents, which was a necessary concentration of oxaziridine to label the less accessible sites (**Fig. 1.7**). We hypothesized this additional and undesirable labeling was due to labeling of the methionine at position HC.M107 in the CDR H3 of trastuzumab. Simply mutating HC.M107 to an unreactive leucine eliminated labeling at this site (**Fig. 1.7a**). The HC.M107L mutation did not affect binding to HER2 on SKBR3 cells (**Fig. 1.7b**).

We chose our two most stable sites, LC.R66M and LC.T74M, and incorporated methionine into the corresponding sites on trastuzumab αHER2 Fab antibody to use in cellular toxicity and serum stability assays. Both labeled to greater than 80% when reacted with 20 equivalents of oxaziridine-azide **8** (**Table 1.2**). The two stable sites were individually converted to methionines

on the trastuzumab Fab scaffold and then labeled with oxaziridine azide **8**, followed by strain promoted click chemistry with DBCO-PEG4-valine-citrulline-MMAF to be used in a cellular toxicity assay. We chose to use the cathepsin B cleavable linker valine-citrulline for its improved effect over a non-cleavable linker (data not shown). We picked the microtubule inhibitor MMAF as the toxic payload due to its previously characterized strong potency in ADC formats and improved solubility compared to MMAE[26]. Both ADCs showed high potency in a HER2-positive breast cancer cell line, BT474-M1, compared to either trastuzumab alone or an αGFP Fab isotype control (**Fig. 1.6e**). The ADC conjugates were 10-100-fold more potent than the free MMAF reflecting their capacity as a drug chaperone. Interestingly, the ADC derived from the LC.R66M was about 10-fold less active than LC.T74M due to a modest loss in affinity when conjugated with drug (**Fig. 1.8a**). Fortunately, upon conversion to a full IgG, the loss in affinity was greatly restored due to the higher avidity of the IgG and much lower off-rates (**Fig. 1.8b**). Both sites were also tested for their stability in human serum and showed similar levels compared to their stability measured in buffer (**Fig. 1.6f**). Thus, the two sites in the Fab arms are promising candidates for ADC formation.

*Labeling and stability at homologous sites on the Fc domain*

To explore more flexibility in labeling sites for methionine antibody conjugates, we probed for suitable labeling sites on the Fc domain of the IgG. We found there are two endogenous methionines on the Fc (HC.M252 and HC.M428) that are surface exposed and one of which readily reacts with the oxaziridine azide **8** (**Fig. 1.9, Fig. 1.10a**). Also, it is known that these methionines sit directly at the FcRn binding site and that even oxidation at these sites can disrupt FcRn binding[27]. In fact, labeling these methionines with oxaziridine ablated FcRn binding **(Fig. 1.10b).**

Thus, we scrubbed these methionines by mutation to leucine as above and found these had little to no effect on overall protein stability or FcRn binding ability (**Fig. 1.10c**). We also incorporated an N297G mutation to prevent glycosylation of the Fc to simplify our mass spectrometry analysis. We then used this Fc mutant as a template to search for more stable methionine conjugation sites.

To simplify our quest for new methionine sites in the Fc we took advantage of the high structural similarity between the Fc and Fab arms. We used PyMol to align the five most stable conjugation sites from the Fab arm studies above to sites in the Fc domain (**Fig. 1.9a, Table 1.3**). An example alignment is shown between LC.K149 and HC.E382 (**Fig. 1.9b**). We introduced single methionine mutants into these sites in the native methionine-scrubbed Fc, expressed the variants in Expi293 mammalian cells, and tested them for their labeling efficiency and stability. Interestingly, two of the engineered sites (HC.T307M, HC.T437M) did not label at all and thus could not be tested for their stability. The other three sites labelled to over 50%, and site HC.V262M showed greater than 80% labeling efficiency with virtually no hydrolysis after a three day incubation at 37°C (**Fig. 1.9c**).


*Functional activity of methionine oxaziridine ADCs on breast cancer cell lines and in vivo efficacy in a breast cancer xenograft model*

We then tested how each of the three stables sites (LC.R66M, LC.T74M, and HC.V262M) performed as ADCs in an IgG format on HER2-positive breast cancer cell lines (**Fig. 1.11a, Fig. 1.11b**). On both SKBR3 and BT474-M1 cell lines, all three sites were almost equally effective at reducing cell growth (IC$_{50}$ ~100-1000pM). All three were 20-50-fold more potent than trastuzumab alone. When compared to one of the previously reported optimal engineered cysteine sites LC.V205C[20], we saw comparable cell killing at sites LC.T74M and

HC.V262M (**Fig. 1.11c**). We also saw improved efficacy of LC.T74M ADC compared to T-DM1 in the BT474-M1 cell line (**Fig. 1.12a**) and similar efficacy to T-DM1 in the SKBR-3 cell line (**Fig. 1.12b**). We also tested how these conjugates performed by size exclusion chromatography (SEC) as a test for antibody aggregates and a proxy for good pharmacokinetics. ADCs produced at sites LC.T74M and HC.V262M showed a single symmetrical elution peak comparable to trastuzumab, while site LC.R66M formed three broad peaks (**Fig. 1.13**). Thus, we decided not to use site LC.R66M *in vivo*. We also discovered that after reintroducing the wildtype N297 residue and corresponding glycosylation, we were not able to label site HC.V262M, suggesting the glycosylation blocks labeling (**Fig. 1.14**). Therefore, we nominated LC.T74M as our lead candidate for *in vivo* studies. We conjugated trastuzumab IgG to the valine-citrulline cleavable MMAF with an average DAR of 1.9 (**Fig. 1.15**) and performed a dose-response study in a mouse xenograft BT474-M1 breast cancer model (**Fig. 1.11d**). We saw dose-response efficacy and with the highest dose of 6mg/kg saw inhibition of tumor growth compared to PBS control across 5 weeks. All mice maintained healthy body weights during the study (**Fig. 1.16**)

With these promising data, we sought to increase our efficacy by creating a DAR of four ADC with our Met sites. Ideally, we wanted a second site with not only high stability, but also a site spatially apart from the LC.T74M site to reduce the steric hindrance and potential for hydrophobic MMAF interactions between sites. Thus, we chose the partially buried stable site HC.S21M, which is almost directly opposite of the LC.T74M. After incorporation of both methionines into the HER2 IgG, we were able to obtain 80% labeling of the HC.S21M site with oxaziridine-azide 8 at 20 equivalences over 30 minutes. Using the double mutant we were able to obtain quantitative labeling of both sites with DBCO-PEG4-valcit-MMAF with an average DAR

of 3.6 (**Fig. 1.17**). We evaluated the conjugate by SEC and saw a monodispersed peak, suggesting there were no aggregates (**Fig. 1.18**). The DAR of four derivative showed an increase in *in vitro* efficacy on the BT474-M1 cell line compared to DAR of two derivative from the LC.T74M IgG (**Fig. 1.11e**). There was a slight decrease in cell binding with the DAR of four ADC compared to DAR of two ADC (**Fig 1.19**) possibly due to the increase in hydrophobicity. We then tested our DAR of four ADC (10mg/kg) in the BT474-M1 xenograft model and saw efficacy in reducing tumor growth over 5 weeks (**Fig. 1.11f**). The efficacy was greater than Trastuzumab IgG control (10mg/kg), showing the use of these sites in an ADC model. Healthy body weights were maintained for all mice except one mouse in the DAR of four treatment group (**Fig. 1.20**). This data shows the proof of concept application of using these engineered sites in antibody drug conjugates and the potential for this method in future stable protein bioconjugation applications.

**Discussion**

We layout a systematic and general approach for identifying efficient, stoichiometric, and stable methionine labeling sites for antibodies using ReACT that preserves antibody function and stability for stable protein bioconjugations such as ADCs. We explored a number of variables and addressed potential pitfalls to find optimal labeling sites. Surprisingly, almost all of the single methionine mutants were tolerated in the context of the trastuzumab scaffold. Of the 95 highly accessible methionine sites, 93 were expressed at wild-type levels, and 92 retained a Tm greater than 77°C. Even for the 23 partially buried sites, 19 were expressed at wild-type levels and 17 maintained a high Tm. We did not detect methionine oxidation for the purified recombinant antibodies expressed either as Fabs in *E.coli* or as IgGs expressed from mammalian cells. This obviated the need to chemically reduce prior to conjugation with oxaziridine. This is a substantial

advantage to cysteine labeling which typically requires reduction and reoxidation prior to thiol-conjugation. The conjugation to the oxaziridine was done rapidly (30 min at 5-30 fold excess) at room temperature in aqueous conditions and consistently produced high yields of the bioconjugate. For example, of the 92 accessible methionine sites expressed, 57 were labelled over 90%. Even for the 23 expressible partially buried methionine sites, 11 were labeled to over 80%.

One can tolerate, manage, or exploit endogenous methionines for antibody conjugations. In our αGFP trastuzumab Fab there are three buried methionines. We found them to be unreactive and thus preserved them throughout our experiments. Once we switched to the wild type trastuzumab, there was a reactive methionine in CDR H3, but was benignly replaced with a leucine and did not affect the affinity of the antibody. Moreover, methionines are routinely mutated out of CDRs in therapeutic antibodies to avoid oxidation upon long-term storage or treatment[28]. We also identified two endogenous methionines in the Fc and these were readily mutated to leucine without significant impact on expression or binding. In some cases methionine sites have been mutated to non-oxidatively sensitive residues to extend antibody half-life and improve FcRn binding[29].

We found the initial oxaziridine compound did not have the desired stability for long-term studies, but structure-activity analysis identified new compounds with significantly improved stability to hydrolysis. The stability tracked with the electron density surrounding the carbonyl as found in parallel studies on isolated methionine[25]. We believe these new compounds (especially oxaziridine azide **8**) will find general utility for ReACT applications for other protein bioconjugations.

We found significant variation in stability depending on the site of modification. There is an inverse trend between accessibility and site stability. We expect this may be because the sulfimide is shielded from water and hindered from being hydrolyzed. The hydrolysis reaction of

the sulfimide is expected to go through a tetrasubstituted intermediate[30] and neighboring sites will likely impact the stability of this intermediate based on the chemical environment. Further mechanistic and computational work may help to further dissect these factors. Interestingly, the stability, and therefore therapeutic effectivity, of cysteine conjugates also varies depending on the conjugation sites[22,31].

We believe site-specific modification of methionine by ReACT has great potential for antibody and protein bioconjugations. The expression of the methionine mutants is robust and multiple methionines can easily be introduced. The conjugation procedure is rapid, simple, and does not require pre-reduction. There is good flexibility with site selection and the resulting linkage can be stable at biological temperatures. The sites described here will provide candidates for other antibody scaffolds. In fact, the discovery of the stable Fc site did not require a complete methionine surface scan, but rather simple homology modeling was sufficient to identify useful sites. Site-specific methionine labeling by ReACT offers more homogeneity of modification compared to lysine modification. It produced highly stable conjugates, and robust ADC activity in a BT474-M1 mouse xenograft model. While there is still much to do to validate their clinical use, we believe this modification will be useful for many other antibody and protein bioconjugation applications such as for fluorescence, affinity labels, DNA barcoding, and protein-protein bioconjugation. The general parameters studied and optimized here will expand the use of ReACT bioconjugation on many other biomolecules.

**Methods**

*Selection of accessible conjugation sites*

To estimate the relative solvent accessibility (RSA) of engineered methionines on a Fab, a computational methionine scan was performed with MODELLER using PDB structure 1FVE as a template[32]. MODELLER generates homology models for comparative structure analysis by satisfaction of spatial restraints[33]. Single methionine mutations were systematically modeled across the entire structure of the Fab including an additional model with a methionine appended at the end of the light chain for a total of 439 individual models generated. The solvent accessible surface area (SASA) of the engineered methionine sulfur atom was determined using the "get_area" function (dot_solvent = 1, dot_density = 4, solvent_radius = 1.4) in PyMol. Due to the stochasticity of the S-methyl group placement, the group was removed prior to SASA calculations and was found to reduce variability. The RSA was calculated by taking the SASA values and dividing by the maximum SASA value observed in the set. Positions were rank ordered and the top 95 sites with the highest RSA (excluding CDR positions, prolines and cysteines) were selected for bioconjugation.

*Preparation and characterization of αGFP Fab methionine mutants*

All methionine mutants were made using QuikChange to introduce single codon mutations onto the αGFP Fab. Fabs were expressed and purified by an optimized auto-induction protocol previously described[24]. In brief, C43 (DE3) Pro +*E. coli* containing expression plasmids were grown in TB auto-induction media at 37 ˚C for 6 hours, then cooled to 30 ˚C for 16–18 hr. Cells were harvested by centrifugation and Fabs were purified by Protein A affinity chromatography. Fab purity and integrity was assessed by SDS-PAGE and intact protein mass spectrometry using

a Xevo G2-XS Mass Spectrometer (Waters) equipped with a LockSpray (ESI) source and Acquity Protein BEH C4 column (2.1 mm inner diameter, 50 mm length, 300 Å pore size, 1.7 µm particle size) connected to an Acquity I-class liquid chromatography system (Waters). Deconvolution of mass spectra was performed using the maximum entropy (MaxEnt) algorithm in MassLynx 4.1 (Waters).

### Labeling of αGFP Fab methionine mutants with oxaziridine and Sulfo-DBCO-NH2

Fabs were prepared at 30uM in PBS and labeled with 5 equivalents of the original oxaziridine azide reagent. The reaction proceeded for 2 hours at room temperature before being quenched with 500mM methionine. Sulfo-DBCO-NH2 was added at a final concentration of 625uM and incubated at room temperature overnight. Labeling was analyzed by intact protein mass spectrometry using a Xevo G2-XS Mass Spectrometer as previously described.

### Single-point kinetic screen

To determine if binding was perturbed by conjugation, a single-point kinetic screen was performed by bio-layer interferometry on a ForteBio Octet RED384. Biotinylated-GFP was captured by streptavidin biosensors and the remaining biotin binding sites were saturated with free biotin. Association of 10 nM unlabeled or labeled Fab was measured for 15 min followed by dissociation for 30 min. $K_D$ values of all unlabeled and labelled Fabs were estimated to be sub-0.5nM. Binding affinity for FcRn was performed in a similar manner but at pH 6.0 to mimic binding in the acidic endosome. Biotinylated FcRn (Acro Biosystems) was used as the loading ligand.

*Protein stability Differential Scanning Fluorimetry (DSF) assay*

Stability was measured by a Sypro Orange based DSF assay. In brief, Fabs (2μM) were incubated with 4x Sypro Orange Protein Gel Stain (ThermoFischer) in PBS. Fluorescence scanning was performed from 25°C- 95°C at a rate of 1°C/min using a Lightcycler 480 Instrument (Roche Life Scientific). Melting temperatures were calculated from the inflection point in the first-derivative curve.

*Synthesis of Compounds*

All oxaziridine compounds were previously synthesized and reported in Christian *et al*[25]. Synthesis of the azide-piperdine oxaziridine (oxaziridine azide **8**) can be found in the supplementary methods.

*Parameter Derivation*

A conformational search on the respective ureas and carbamates was performed using the MacroModel suite from Schrödinger[34] using an OPLS_2005 force field without solvent corrections. A Monte-Carlo molecular mechanics method was employed. The output was restricted to structures within 1.30 kcal/mol (5 kJ/mol) of the lowest energy conformer. Conformers were submitted to a geometry optimization in Gaussian 09 using the def2-TZVP basis set and M06-2x functional[35]. A triple zeta potential basis set was chosen along with the M06-2x functional, as these generally lead to quantitative correlations[36]. Using a cutoff limit of 2.5 kcal/mol, the parameters of each low energy conformer were weighted using the Boltzmann distribution (equations 1 and 2) where the energy of a given conformer is calculated relative to the lowest energy conformation.

$$c_i = \frac{e^{-Ec_i/RT}}{e^{-Ec_1/RT} + e^{-Ec_2/RT} + \cdots + e^{-Ec_n/RT}} \qquad (1)$$

$$p_w = c_1 p_1 + c_2 p_2 + \cdots + c_n p_n \qquad (2)$$

### *PyMol homology alignment*

To determine analogous stable sites on the Fc, the alignment function was used in PyMol, using the PDB structure 1FVE (Fab) and 1H3X (Fc). Stable sites on light chain or heavy chain were aligned to either the CH2 or CH3 domains in the Fc. Corresponding positions were chosen on the Fc to mutate to methionine.

### *Expression of IgG single methionine mutants*

IgGs containing the engineered methionines were expressed and purified from Expi293 BirA cells according to established protocol from the manufacturer. Briefly, 30μg of pFUSE (InvivoGen) vector was transiently transfected into 75 million Expi293 BirA cells using the Expifectamine kit. Enhancer was added 20 hours after transfection. Cells were incubated for a total of 6 days at 37 ˚C in a 5% $CO_2$ environment before the supernatants were harvested by centrifugation. Protein was purified by Protein A affinity chromatography and assessed for quality and integrity by SDS-PAGE.

*Conjugation of engineered methionine Fabs and IgGs with oxaziridine and DBCO-PEG4-valcit-MMAF*

For Fab ADCs, endotoxins were removed prior to conjugation using Pierce endotoxin removal kits (ThermoFischer Scientific). For conjugation, Fabs were incubated at 50µM with 15 molar equivalents of compound 8 azide oxaziridine for 30 minutes at room temperature in PBS. For IgGs, IgGs were incubated at 10µM with 30 molar equivalents of compound 8 azide oxaziridine per methionine for 1 hour at room temperature in PBS. For both, the reaction was quenched by the addition of methionine and antibody was buffered exchanged into PBS using a 0.5mL Zeba 7kDa desalting column (ThermoFischer Scientific). Then 10 molar equivalents of DBCO-PEG4-valcit-MMAF (Levena Biosciences) was added and the click reaction proceeded overnight at room temperature. The conjugate was desalted twice using two 0.5mL Zeba 7kDa columns to remove excess unconjugated drug. Full conjugation was monitored by intact protein mass spectrometry using a Xevo G2-XS Mass Spectrometer (Waters).

*Stability measurements of oxaziridine labeled Fabs in PBS and serum*

Fabs were labeled as previously described, placed in either buffer or human AB serum (Valley Biomedical), and incubated at 37°C. PBS samples were measured directly by intact protein mass spectrometry. Serum samples were able to be purified with a Ni-NTA column due to the presence of a His-tag on the Fab. The sample was then consequently measured by intact protein mass spectrometry to determine proportion of labeled Fab.

*Conjugation of engineered cysteine ADCs for comparison*

Engineered cysteine conjugation was performed as previously reported[37]. In brief, after purification of the LC.V205C mutant αHer2 IgG (see IgG expression), the IgG (10μM) was buffer exchanged into 50mM Tris-HCl, pH 7.5, 2mM EDTA. DTT was added at 40-fold molar excess and incubated at room temperature for 16 hours. Desalting into PBS proceeded with 0.5mL Zeba 7kDa columns. DHAA was added in 15-fold molar excess to reoxidize the interchain disulfides for 3 hours at room temperature. Maleimide-valcit-MMAF (BOC Sciences) was added at 3-fold molar excess and conjugation was monitored by mass spectrometry. Excess drug was removed by two 0.5mL Zeba desalting columns.

*Cell culture of HER2-positive breast cancer cells*

The BT474-M1 cell line was provide by the Preclinical Therapeutics Core at the UCSF Helen Diller Cancer Center. These cells were maintained in DMEM media supplemented 10% FBS and 1X Pen/Strep. The SKBR3 cells were purchased from the UCSF Cell Culture Facility. They were maintained in McCoy 5a media supplemented with 10% FBS and 1X Pen/Strep. Cell line identities were authenticated by morphological inspection. The SKBR3 cell line identity was validated by UCSF Cell Culture Facility. Symptoms for mycoplasma contamination were not observed and thus no test for mycoplasma contamination was performed. All cell lines that were received as gifts were previously authenticated.

*ADC cell killing assay in vitro*

Antibody drug conjugate cell killing assays were performed using an MTT modified assay to measure cell viability. In brief, 10000 BT474-M1 or SKBR3 cells were plated in each well of a

96-well plate on day 0. On day 1, Fab/IgG was added in a 10-fold dilution series. Cells were incubated for 120 hr at 37°C under 5% $CO_2$. On day 6, 40uL of 2.5mg/mL of Thiazolyl Blue Tetrazolium Bromide (Sigma Aldrich) was added to each well and incubated at 37°C under 5% $CO_2$ for 4 hours. Following, 100µL of 10% SDS 0.01M HCl was added to lyse the cells to release the MTT product. After 4 hours, absorbance at 600nm was quantified using an Infinite M200 PRO plate reader (Tecan). Data points were plotted using GraphPad Prism (version 8.2) and curves were generated by using non-linear regression with Sigmoidal 4PL parameters.

### *ADC study in mouse xenograft model in vivo*

The xenograft was performed with 6-8 week old nude female mice (NCR, nu/nu) purchased from Taconic Labs (n=3 per group for the dose response, n=8 for the DAR=4 study). Prior to tumor cell engraftment, mice were implanted subcutaneously with Estradiol pellet (0.36mg, 60 day release, Innovative Research). BT474-M1 xenografts were then established by bilateral subcutaneous injection into the right and left flanks of mice with BT474-M1 tumor cells ($5x10^6$ cells in 100 µl of serum free medium mixed 1:1 with Matrigel). When BT474-M1 xenografts reached average volume of 200mm$^3$ (measured as width x width x length x 0.52), mice were dosed intravenously weekly for 3 weeks with PBS, drug alone, antibody alone and ADCs. Tumor size and body weight were monitored biweekly for 5 weeks total. Data was plotted in GraphPad Prism and SEM for the 6 tumors across 3 mice in each group was determined for the first study For the second study, data was plotted and SEM was determined for 7 mice in the PBS group, 8 mice in the Trastuzumab control group, and 7 mice in the DAR of 4 ADC group (one mouse is not shown due to early sacrifice due to low body weight). All experiments were performed in accordance with

relevant guidelines, regulations, and in full accordance with UCSF Institutional Animal Care and Use Committee (IACUC).

**Supplementary Detailed Methods**

*Preparation and characterization of αGFP Fab methionine mutants*

All methionine mutants were made using QuikChange to introduce single codon mutations onto the αGFP Fab. Fabs were expressed and purified by an optimized auto-induction protocol previously described[24]. In brief, C43 (DE3) Pro +*E. coli* containing expression plasmids were grown in TB auto-induction media at 37 ˚C for 6 hours, then cooled to 30 ˚C for 16–18 hr. Cells were harvested by centrifugation and Fabs were purified by Protein A affinity chromatography. Fab purity and integrity was assessed by SDS-PAGE and intact protein mass spectrometry using a Xevo G2-XS Mass Spectrometer (Waters) equipped with a LockSpray (ESI) source and Acquity Protein BEH C4 column (2.1 mm inner diameter, 50 mm length, 300 Å pore size, 1.7 μm particle size) connected to an Acquity I-class liquid chromatography system (Waters). Deconvolution of mass spectra was performed using the maximum entropy (MaxEnt) algorithm in MassLynx 4.1 (Waters). Endotoxins were removed using ThermoFischer's Pierce High Capacity Endotoxin Removal Columns.

*Labeling of αGFP Fab methionine mutants with oxaziridine and Sulfo-DBCO-NH$_2$*

Fabs were prepared at 30uM in PBS and labeled with 5 equivalents of the original oxaziridine azide reagent. The reaction proceeded for 2 hours at room temperature before being quenched with 500mM methionine. Sulfo-DBCO-NH$_2$ was added at a final concentration of 625uM and incubated

at room temperature overnight. Labeling was analyzed by intact protein mass spectrometry using a Xevo G2-XS Mass Spectrometer as previously described.

***Single-point kinetic screen***

To determine if binding was perturbed by conjugation, a single-point kinetic screen was performed by bio-layer interferometry on a ForteBio Octet RED384. Biotinylated-GFP was expressed and purified as previously described[24]. Biotinylated-GFP was captured by streptavidin biosensors and the remaining biotin binding sites were saturated with free biotin. Association of 10 nM unlabeled or labeled Fab was measured for 15 min followed by dissociation for 30 min. $K_D$ values of all unlabeled and labelled Fabs were estimated to be sub-0.5nM. Binding affinity for FcRn was performed in a similar manner but at pH 6.0 to mimic binding in the acidic endosome. Biotinylated FcRn (Acro Biosystems) was used as the loading ligand.

***Protein stability Differential Scanning Fluorimetry (DSF) assay***

Stability was measured by a Sypro Orange based DSF assay. In brief, Fabs (2µM) were incubated with 4x Sypro Orange Protein Gel Stain (ThermoFischer) in PBS. Fluorescence scanning was performed from 25°C- 95°C at a rate of 1°C/min using a Lightcycler 480 Instrument (Roche Life Scientific). Melting temperatures were calculated from the inflection point in the first-derivative curve.

***Parameter Derivation***

A conformational search on the respective ureas and carbamates was performed using the MacroModel suite from Schrödinger[34] using an OPLS_2005 force field without solvent

corrections. A Monte-Carlo molecular mechanics method was employed. The output was restricted to structures within 1.30 kcal/mol (5 kJ/mol) of the lowest energy conformer. Conformers were submitted to a geometry optimization in Gaussian 09 using the def2-TZVP basis set and M06-2x functional[35]. A triple zeta potential basis set was chosen along with the M06-2x functional, as these generally lead to quantitative correlations[36]. Using a cutoff limit of 2.5 kcal/mol, the parameters of each low energy conformer were weighted using the Boltzmann distribution (equations 1 and 2) where the energy of a given conformer is calculated relative to the lowest energy conformation.

$$c_i = \frac{e^{-E_{c_i}/RT}}{e^{-E_{c_1}/RT} + e^{-E_{c_2}/RT} + \cdots + e^{-E_{c_n}/RT}} \qquad (1)$$

$$p_w = c_1 p_1 + c_2 p_2 + \cdots + c_n p_n \qquad (2)$$

### *PyMol homology alignment*

To determine analogous stable sites on the Fc, the alignment function was used in PyMol, using the PDB structure 1FVE (Fab) and 1H3X (Fc). Stable sites on light chain or heavy chain were aligned to either the CH2 or CH3 domains in the Fc. Corresponding positions were chosen on the Fc to mutate to methionine.

### *Expression of IgG single methionine mutants*

IgGs containing the engineered methionines were expressed and purified from Expi293 BirA cells according to established protocol from the manufacturer. Briefly, 30μg of pFUSE (InvivoGen) vector was transiently transfected into 75 million Expi293 BirA cells using the Expifectamine kit. Enhancer was added 20 hours after transfection. Cells were incubated for a total of 6 days at 37 °C

in a 5% $CO_2$ environment before the supernatants were harvested by centrifugation. Protein was purified by Protein A affinity chromatography and assessed for quality and integrity by SDS-PAGE. Clinical grade T-DM1 was a generous gift from Dr. Mark Moasser.

*Flow cytometry*

Cells were lifted with Versene (0.04% EDTA, PBS pH 7.4 Mg/Ca free), and subsequently blocked with flow cytometry buffer (PBS, pH 7.4, 3% BSA). Fabs or IgGs were added to cells for 30 min on ice. Antibodies were detected with addition of Alexa Fluor 647 AffiniPure F(ab')$_2$ conjugate (Jackson Immuno Researech, 1:50). Cells were extensively washed and fluorescence in the APC channel was quantified using a FACSCanto II (BD Biosciences). All flow cytometry data analysis was performed using FlowJo software.

*Size Exclusion Chromatography*

SEC analysis was performed using an Agilent HPLC 1260 Infinity II LC System using an AdvanceBio SEC column (300Å, 2.7μm, Agilent). Each analyte was injected at 5-10μM and run with a constant mobile phase of 0.15 M sodium phosphate for 15 minutes. Fluorescence (excitation 285nm, emission 340nm) and absorbance was measured.
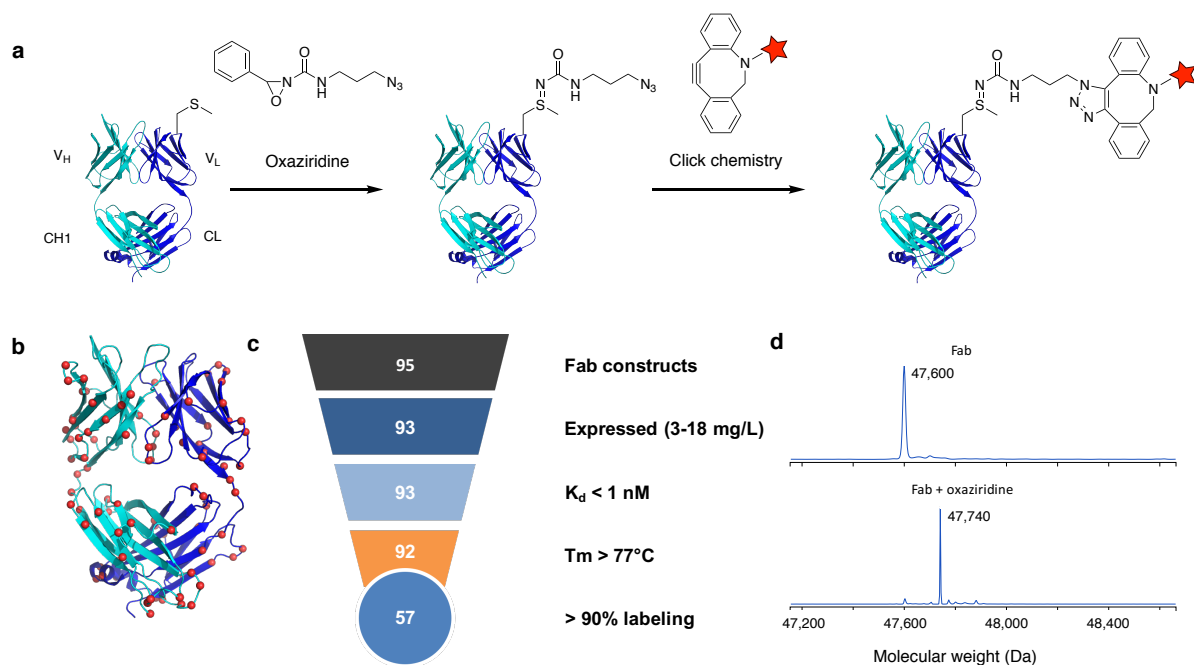
**Figure 1.1: Oxaziridine labeling of most accessible sites on model αGFP-Fab in the trastuzumab scaffold.** (a) Scheme for oxaziridine labeling on trastuzumab Fab. The Fab light chain is shown in dark blue and the heavy chain is shown in cyan. After conjugation with oxaziridine, different functionalities can be clicked on with a DBCO reagent. (b) The top 95 calculated accessible sites on the Fab scaffold are shown as red spheres. (c) Triage of the 95 most accessible mutants is shown. Each site was engineered to methionine on a model αGFP-Fab in the trastuzumab scaffold. Sites were then assessed for expression, affinity, structural stability, and labeling percentage. (d) Representative ESI mass spectra of one labeling reaction of αGFP Fab.Cmet with oxaziridine, shown by a mass shift of 140 (expected: 140).
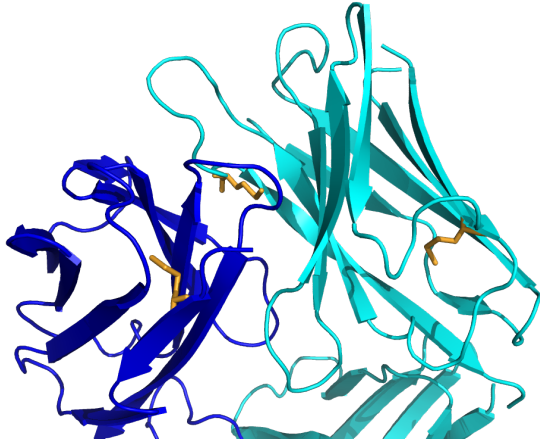
**Figure 1.2: Native methionines in the trastuzumab Fab.**

Methionines present in the trastuzumab Fab are shown in orange. One methionine (LC.M04) is present in the variable domain of the light chain (dark blue) and two (HC.M86, HC.M107) are present on the heavy chain (cyan), with HC.M107 present on the H3 loop of the CDR binding domain.
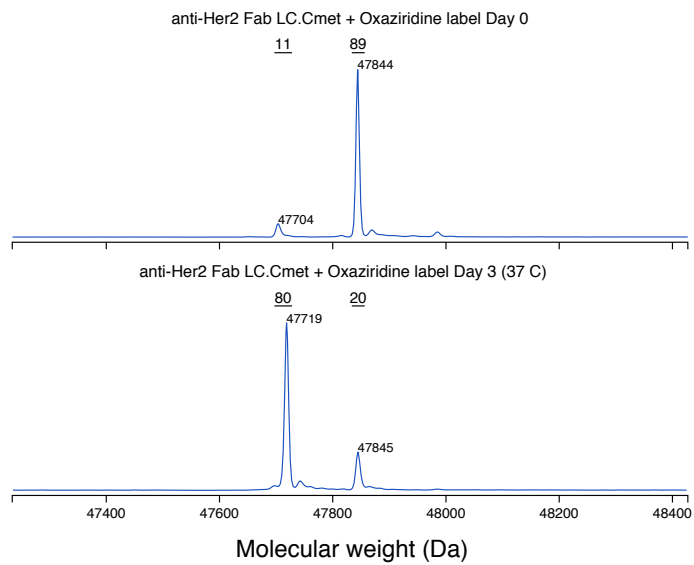
**Figure 1.3: Stability of the light chain C-terminal methionine at 37°C.**

Incubation of the C-terminal methionine oxaziridine conjugate results in 80% hydrolysis over three days at the biological temperature of 37°C.
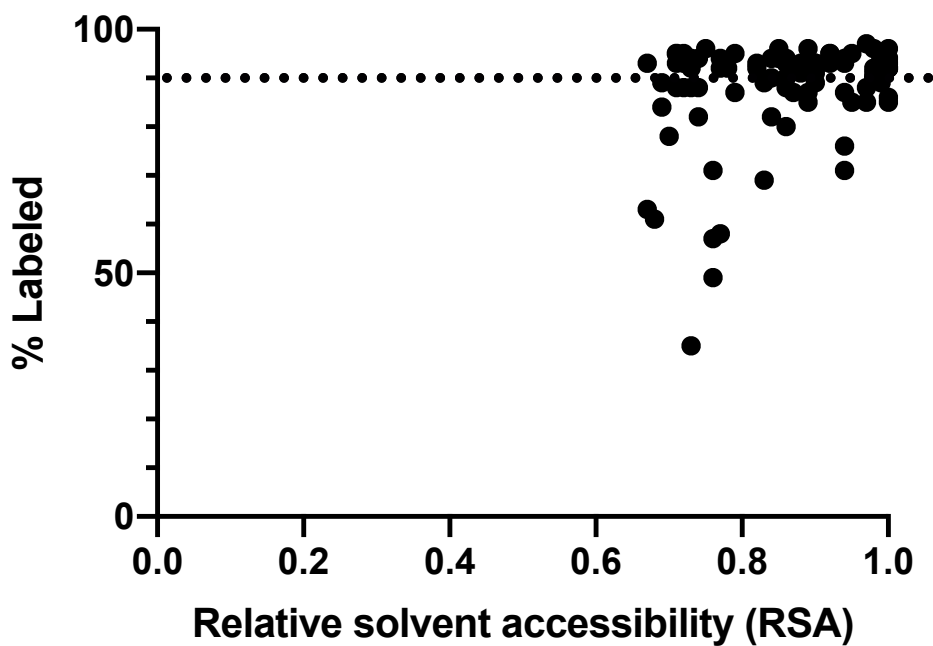
**Figure 1.4: Correlation of labeling and estimated relative solvent accessibility of the 95 mutants in the αGFP Fab.**

57 of 95 methionine mutants label greater than 90% with oxaziridine reagent (dotted line represents 90%). Other sites that do not appear to label well may have improved labeling with higher equivalents of oxaziridine. For the few sites that do not appear to label well with oxaziridine despite having high estimated RSA, there may be other chemical environment factors that govern reactivity.
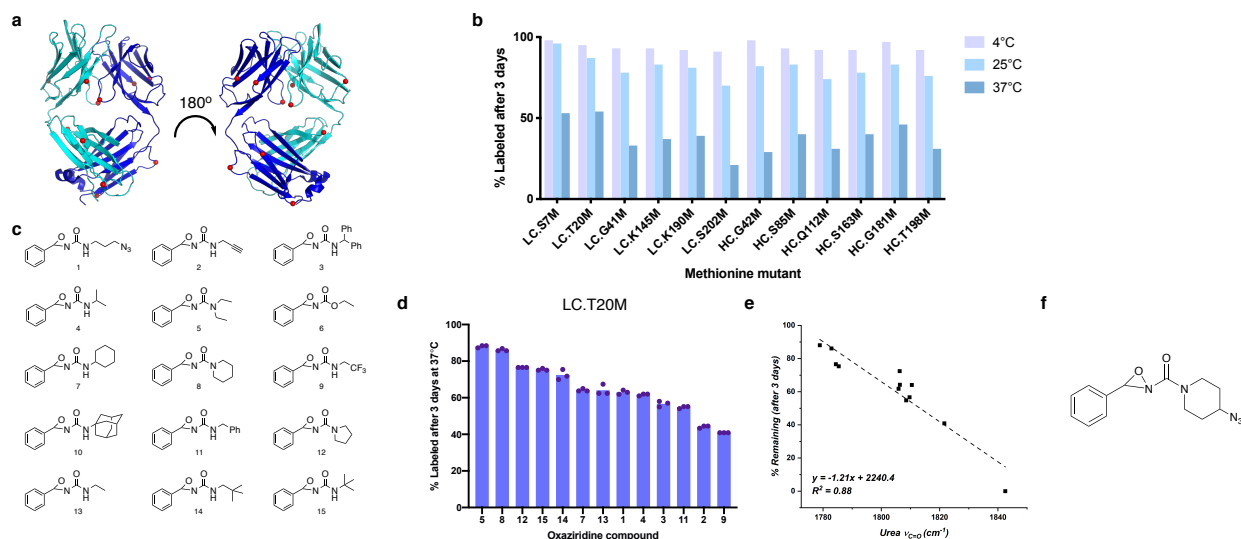
**Figure 1.5: Labeling and stability of top 12 accessible sites with different oxaziridine compounds.** (a) Ribbon diagram of the trastuzumab Fab where the the top 12 labeled sites (all >95% labeling yield) are shown as red spheres. (b) Stability of top 12 sites labeled with oxaziridine at 4°C, 25°C, and 37°C over three days varies across sites and temperatures. All sites show a significant decrease in stability to hydrolysis at elevated temperatures. Each data point represents one measurement from one sample at that temperature. (c) Panel of oxaziridine compounds tested for stability. (d) Conjugate stability at site LC.T20M over three days at 37°C when reacted with each oxaziridine. Oxaziridine **6** is not shown because it showed 0% conjugation. Oxaziridine **10** is not shown because no initial labeling could be detected. Each bar height represents the average from three biological replicates. (e) Correlation of compound conjugate stability of LC.T20M and carbonyl stretching frequency ($\nu_{C=O}$). The more electron rich the substitution on the oxaziridine, the less electrophilic the sulfimide becomes, thus increasing conjugate stability on the protein. Each data point represents the mean of three individual samples. (f) The structure of the azide derivative of the piperdine-derived oxaziridine **8**.
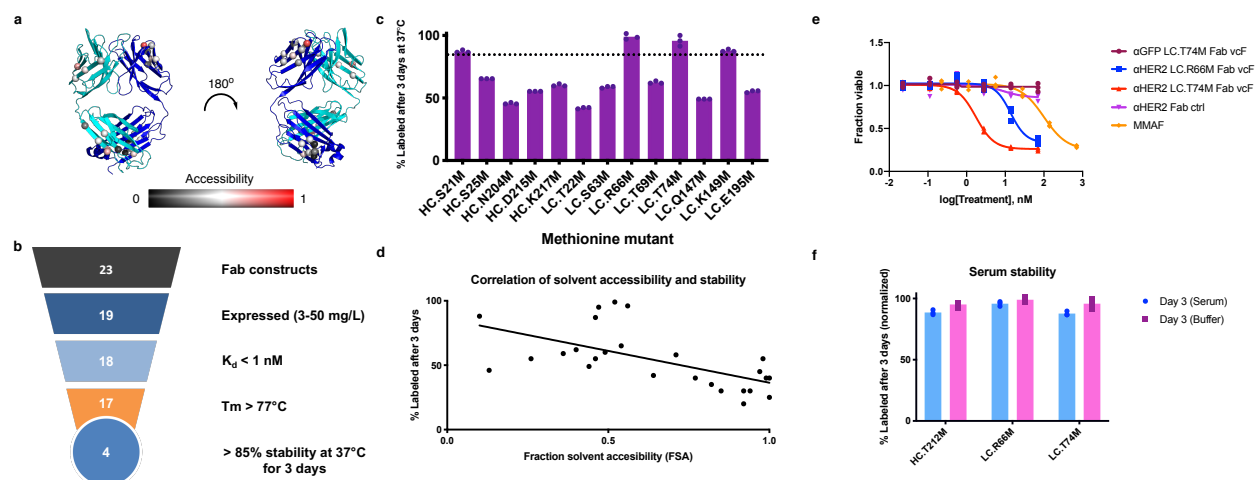
**Figure 1.6: Labeling, stability, and activity of partially buried sites on αGFP-Fab and Trastuzumab Fab.** (a) Structure of the trastuzumab Fab showing 23 partially buried sites (spheres) chosen to mutate to methionine. Scale represents calculated relative fractional surface accessibility from 0 (black) to 1 (red). (b) Triage of the 23 individual mutants on the αGFP-Fab after testing expression, binding to GFP, structural stability, and oxaziridine conjugate stability for 3 days at 37°C. (c) Conjugate stability of 14 sites after incubation at 37°C for 3 days. The dotted line indicates 85% stability. Three biological replicate data points are shown. (d) Correlation between measured stability at 37°C and calculated accessibility for the 23 partially buried sites. Linear regression analysis using GraphPad Prism was used to calculate the coefficient of determination ($R^2$ =0.32) to determine correlation. (e) *In vitro* potency of two stable sites on trastuzumab Fab on the BT474-M1 breast cancer cell line. Three biological replicate data points are shown. (f) Stability measured in human serum for oxaziridine conjugates at the three top stable sites over 3 days at 37°C compared to stability measured in buffer. Three biological replicate data points are shown.

a

Trastuzumab Fab + 20eq Oxaziridine

47572

47739

Trastuzumab Fab M107L + 20eq Oxaziridine

47554

47200    47400    47600    47800    48000    48200

Molecular weight (Da)

b
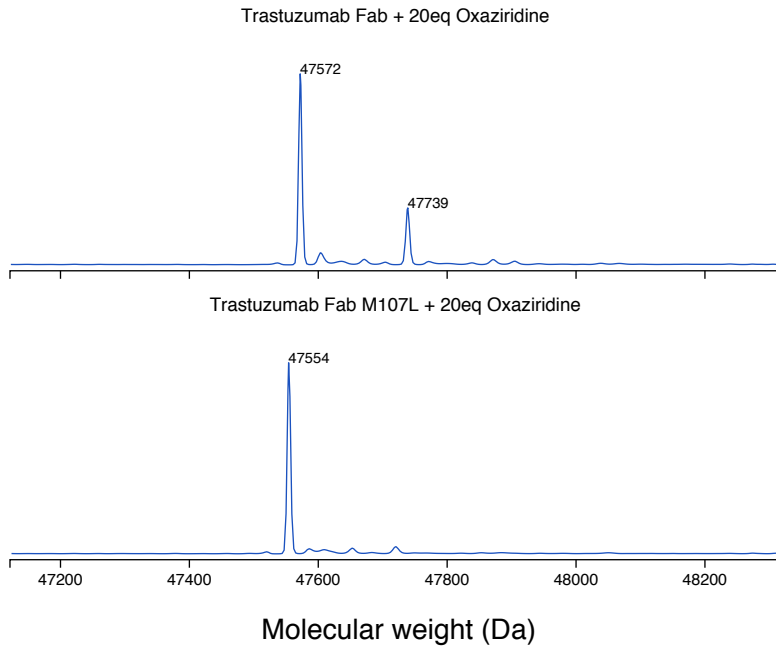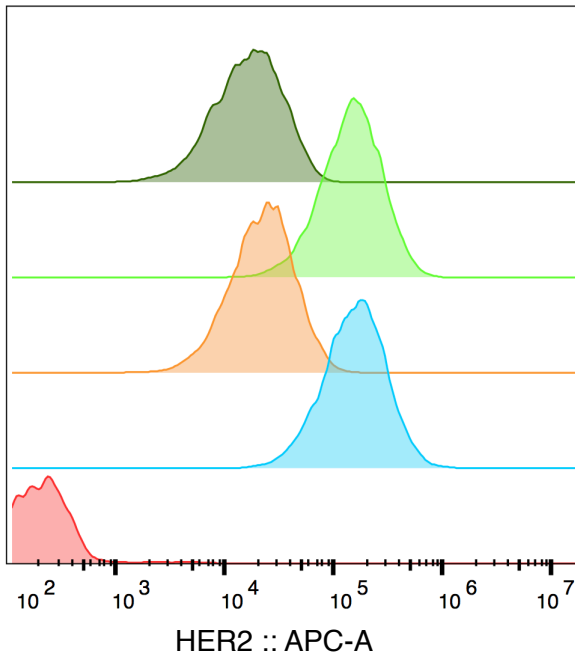


| Sample name | Conc. (µg/mL) | Cell count |
|---|---|---|
| M107L IgG | 0.1 | 9367 |
| M107L IgG | 1 | 10610 |
| IgG only | 0.1 | 10873 |
| IgG only | 1 | 10566 |
| Cell only | - | 10452 |

10² 10³ 10⁴ 10⁵ 10⁶ 10⁷

HER2 :: APC-A
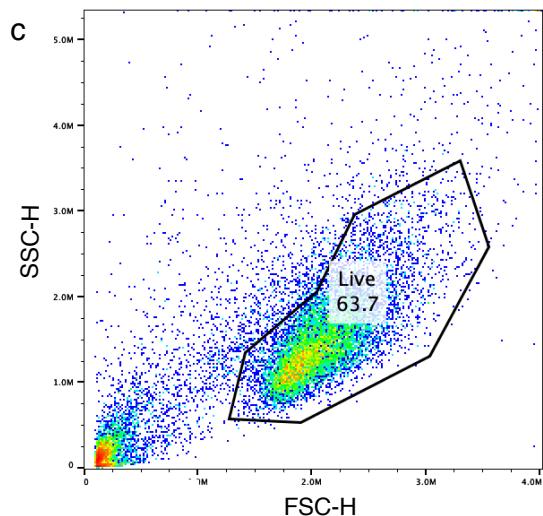
**Figure 1.7: Labeling of trastuzumab Fab HC M107 and mutation to leucine prevents further labeling and minimal change in affinity for HER2.**

(a) Native trastuzumab Fab begins to be labeled with oxaziridine at equivalents of 20 or higher. However, upon mutation of M107 to leucine, this labeling is abolished.
(b) When in the IgG format, there is no detectable difference in affinity for HER2 between trastuzumab and M107L trastuzumab mutant, as detected by flow cytometry on SKBR3 cells.
(c) Flow cytometry gating of live cells by FSC-H and SSC-H is shown here. This applies to all further flow cytometry data.

**a**

| Sample name | Cell count |
|---|---|
| LC.T74M Fab + dbco-vc-MMAF | 10948 |
| LC.R66M Fab + dbco-vc-MMAF | 11222 |
| Fab only | 7723 |
| Cell only | 10452 |

**b**

| Sample name | Conc. (µg/mL) | Cell count |
|---|---|---|
| HC.V262M + dbco-vc-MMAF | 0.1 | 10720 |
| HC.V262M + dbco-vc-MMAF | 1 | 10827 |
| LC.T74M + dbco-vc-MMAF | 0.1 | 10615 |
| LC.T74M + dbco-vc-MMAF | 1 | 10431 |
| LC.R66M + dbco-vc-MMAF | 0.1 | 10673 |
| LC.R66M + dbco-vc-MMAF | 1 | 10700 |
| IgG only | 0.1 | 10509 |
| IgG only | 1 | 10720 |
| Cell only | - | 10476 |
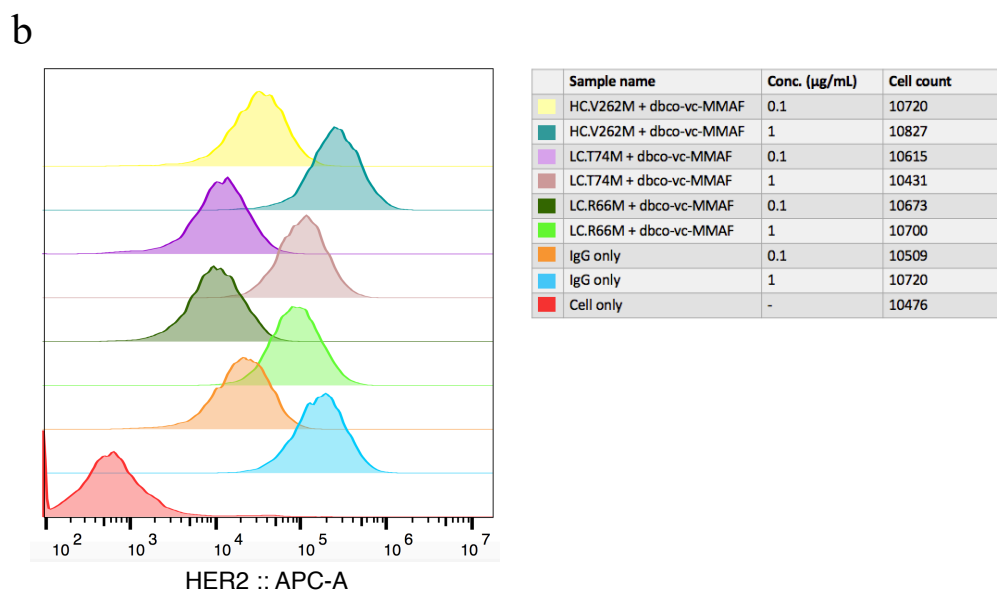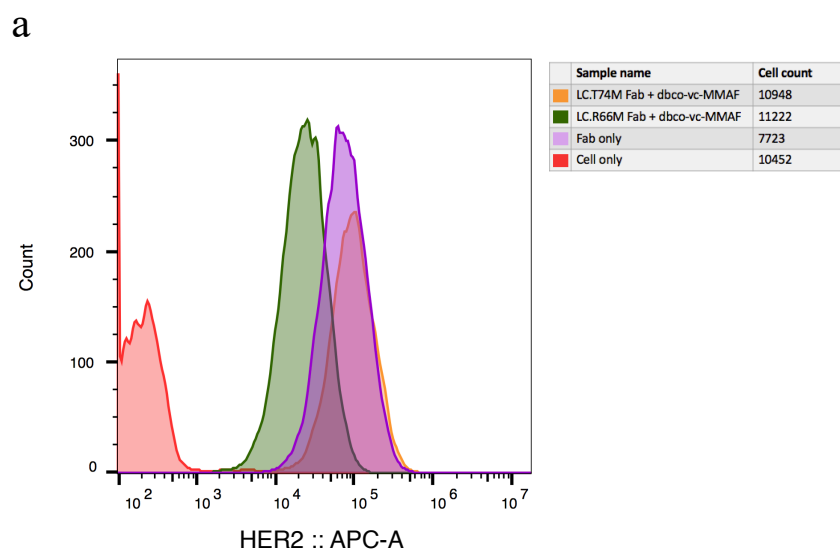
**Figure 1.8: Affinity for HER2 of antibody drug conjugates in Fab and IgG format.**

(a) Flow cytometry was performed on BT474-M1 cells to confirm binding of Fab-drug conjugates at different methionine sites. LC.T74M shows similar affinity to HER2 as the unmodified trastuzumab Fab, while LC.R66M takes about a 10-fold hit in affinity. (b) Upon conversion to IgG and labeling with drug, both sites along with HC.V262M show high affinity for HER2 at two different concentrations, as determined by flow cytometry. All have roughly similar binding shifts as unlabeled IgG. All IgGs contain the HC.M252L and HC.M428L mutations in the Fc domain.
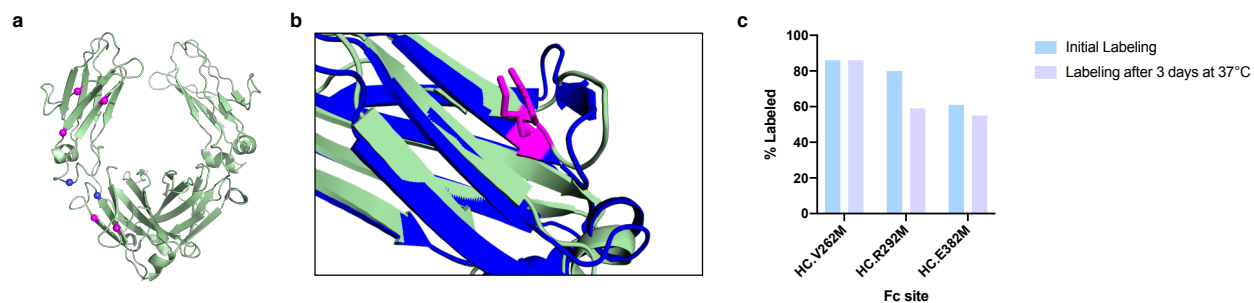
**Figure 1.9: Labeling, stability, and activity of homologous Fc sites.** (a) Structure of IgG1 Fc domain (PDB: 1H3X) showing five sites chosen to individually mutate to methionine (dark purple). The two endogenous methionines are shown in magenta. (b) Example alignment of part of the Fc domain with part of the Fab light chain to show the structural homology between site LC.K149 and HC.E383. (c) Stability and labeling measurements for three top sites on the Fc region. Two sites (HC.T307M, HC.T437M) are not shown because they did produce viable conjugates with oxaziridine. Each data point represents a stability measurement from a single sample.

a

IgG HC

51081

51243

51404

IgG HC + oxaziridine

51248 51412

51574

50500    51000    51500    52000    52500

b

Binding (nm)

0.6

0.4

0.2

0.0

0    500    1000    1500    2000    2500

Time (s)

• IgG (250nM)
■ IgG (125nM)
▲ IgG Met labeled (250nM)
▼ IgG Met labeled (125nM)

c
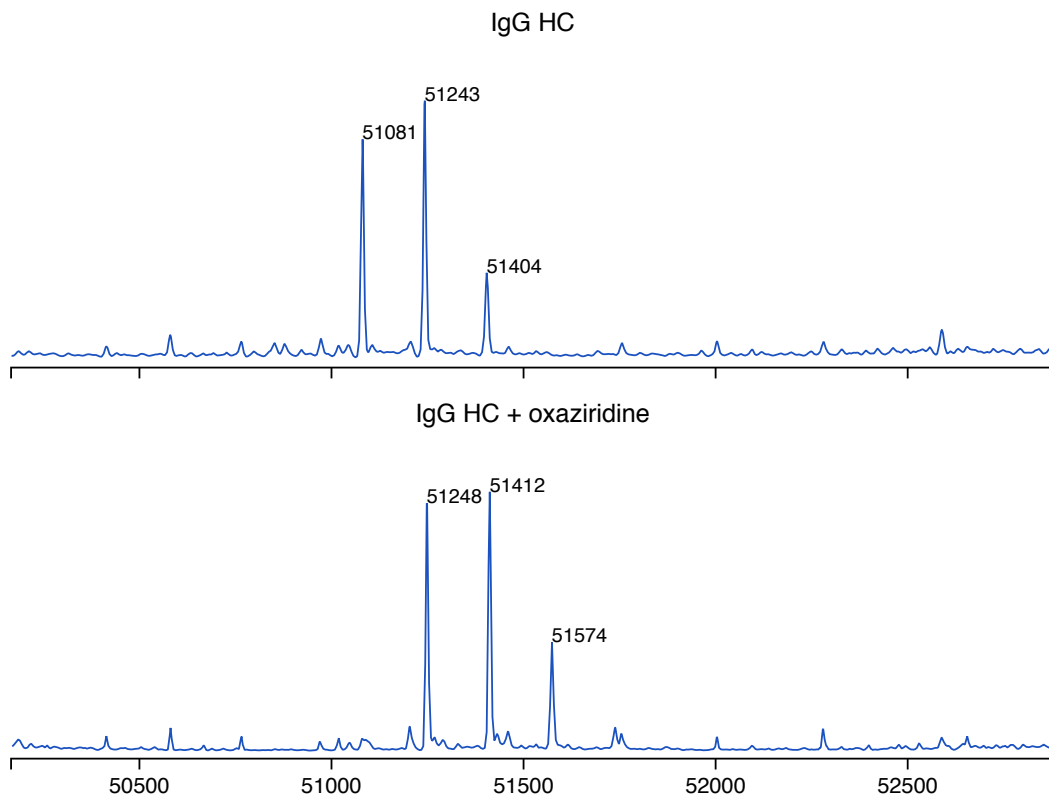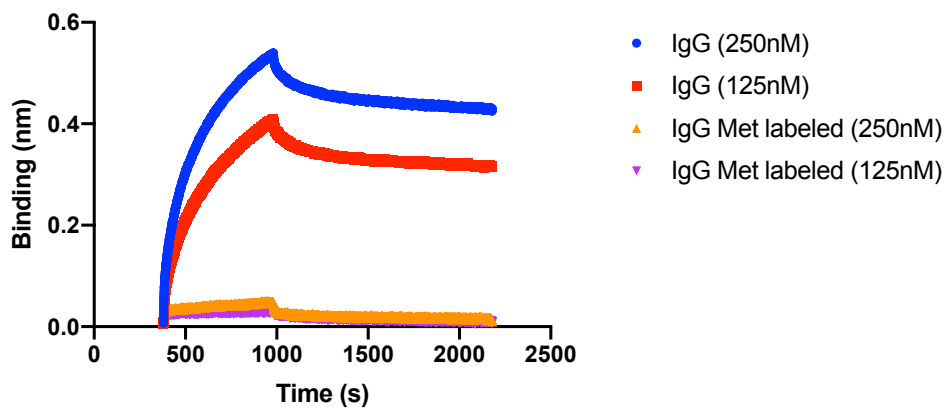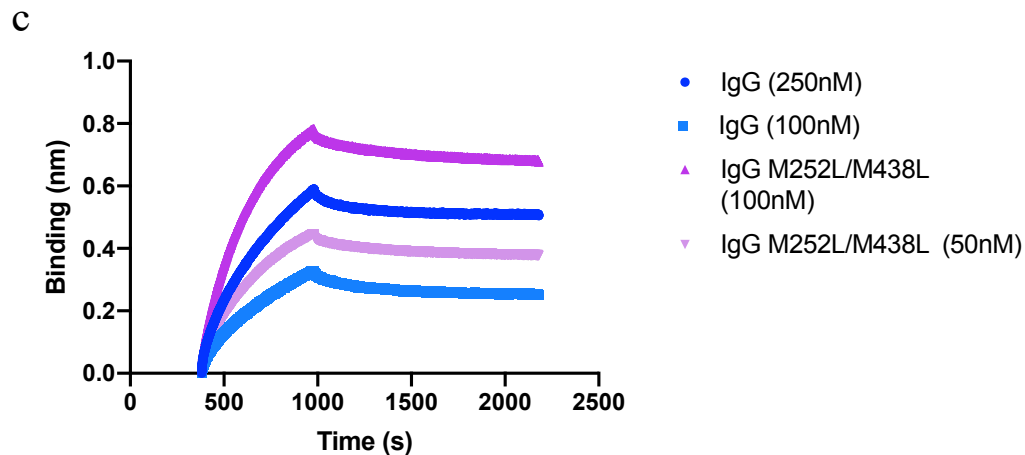


**Figure 1.10: Stability of endogenous Fc methionines and mutation to leucine does not affect**

**FcRn binding.**

(a) Rapid labeling of one methionine on the native Fc with 30 equivalents of oxaziridine. The three peaks are different glycosylation forms of the antibody. All peaks shift 167 demonstrating reaction with the oxaziridine azide **8**.

(b) BLI analysis demonstrates that Fc methionines have undetectable binding to FcRn after being labeled with oxaziridine (orange and purple), while unlabeled IgG shows good binding to FcRn (red and blue).

(c) BLI analysis demonstrates a very similar binding affinity for FcRn of the native Fc of the IgG (blue) and M252L/M438L Fc of the IgG (purple), as noted by the similar shift by octet.
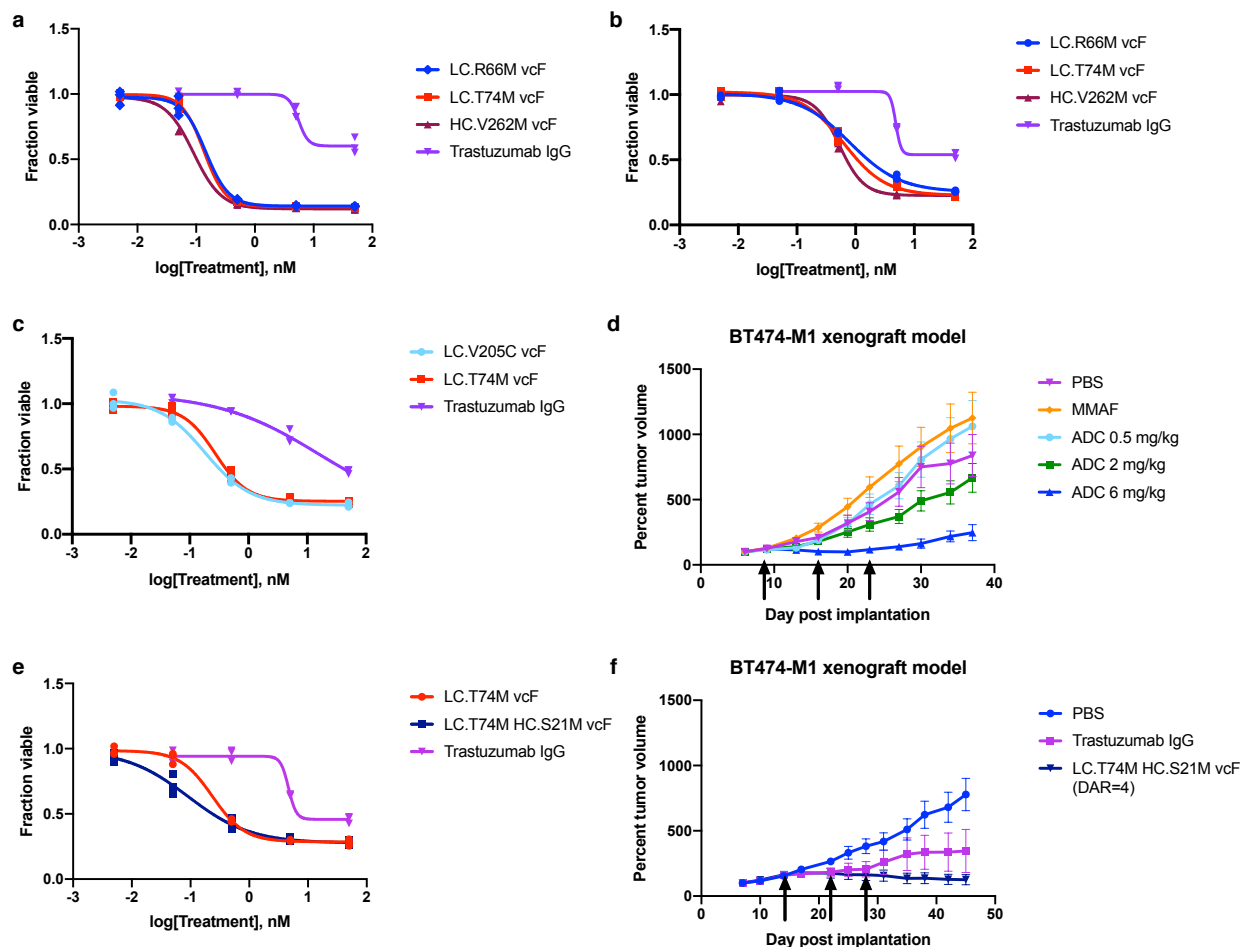
**Figure 1.11: In vitro and in vivo potencies of IgG based ADCs in a breast cancer model.** (a,b) *In vitro* potency of three sites (LC.R66M, LC.T74M, and HC.V262M) on two HER2 positive breast cancer cell lines, SKBR3 (a) and BT474-M1 (b) against Trastuzumab unlabeled IgG control. Three biological replicate data points are shown. (c) Comparison of site LC.T74M and stable cysteine site LC.V205C[20]. (d) *In vivo* potency of site LC.T74M ADC in a breast cancer xenograft model in nude female mice, demonstrating clear dose response of the ADC. Arrows show intravenous administration schemes of controls and ADCs. Lines indicate the average for each group and error bars represent the SEM. (e) In vitro potency of DAR of two and DAR of four ADCs on the BT474-M1 cell line. Increasing to a DAR of four decreases the IC50 of the ADC from 0.24nM (DAR=2, red) to 0.091nM (DAR=4, dark blue). (f) In vivo potency of the DAR of four ADC with sites LC.T74M and HC.S21M in a BT474-M1 xenograft model in nude female mice (PBS: n=7, Herceptin IgG: n=8, ADC DAR of four: n=7). Arrows show intravenous administration schemes of trastuzumab (10mg/kg) and ADC (10mg/kg). Lines indicate the average for each group and error bars represent the SEM.

a



b
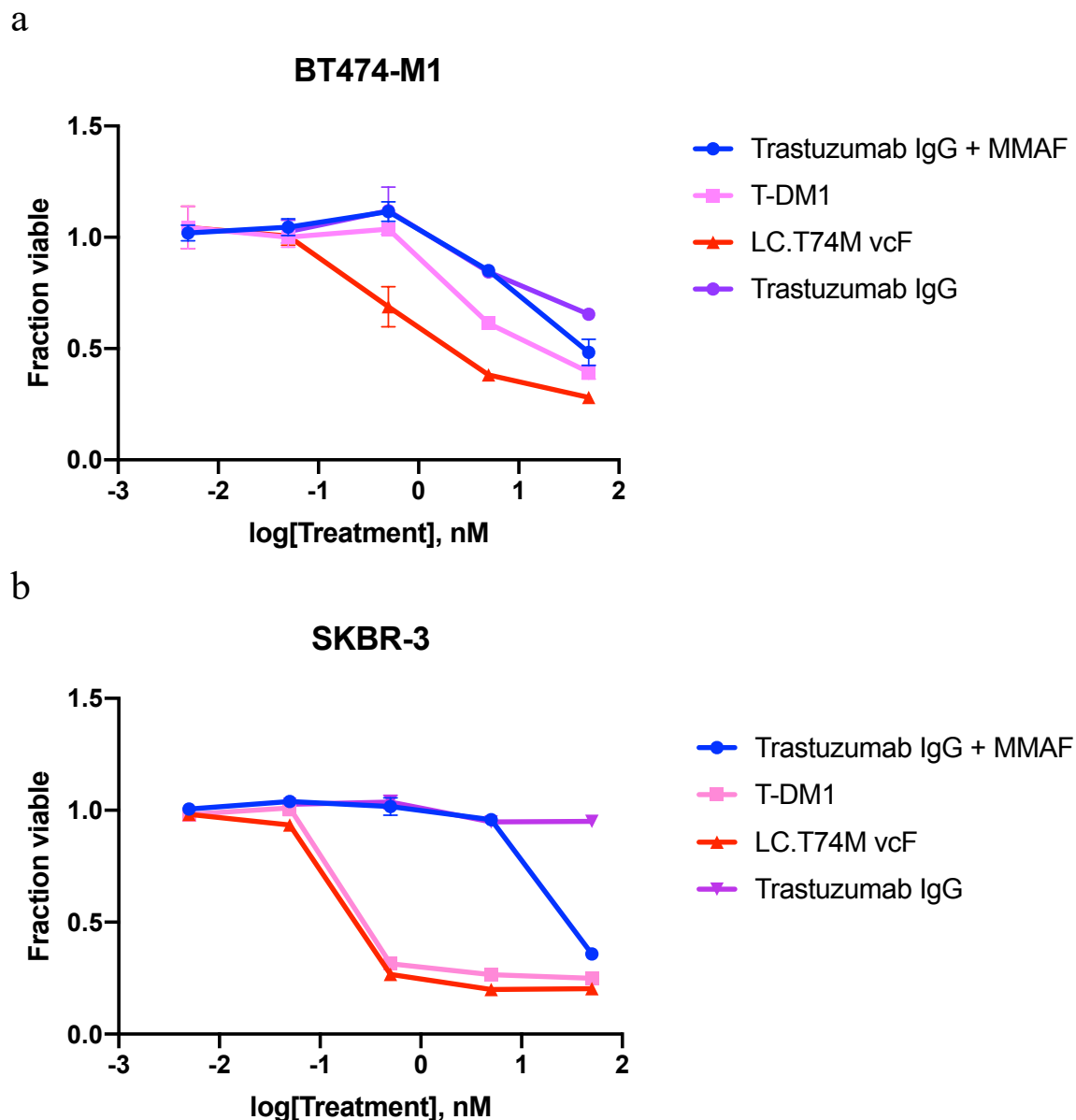


**Figure 1.12: Comparison of LC.T74M vcF ADC with trastuzumab + free MMAF and T-DM1.**

(a) *In vitro* cell killing on the HER2+ BT474-M1 cell line confirmed that conjugation of trastuzumab IgG to MMAF at site LC.T74M was ~100 fold more potent than treating with the IgG and unconjugated free drug. Additionally, the LC.T74M ADC had increased efficacy over the FDA approved drug T-DM1, which may be because maytansine is less toxic than MMAF in this cell line. The error bars represent the standard deviation from data of three biological replicates. The lines simply represent connection between the data points.

(b) *In vitro* cell killing on the HER2+ SKBR-3 cell line also showed 100 fold increased potency of the LC.T74M ADC over IgG and free drug alone. In this cell line, we saw similar efficacy of LC.T74M ADC and T-DM1. The error bars represent the standard deviation from data of three biological replicates. The lines simply represent connection between the data points.
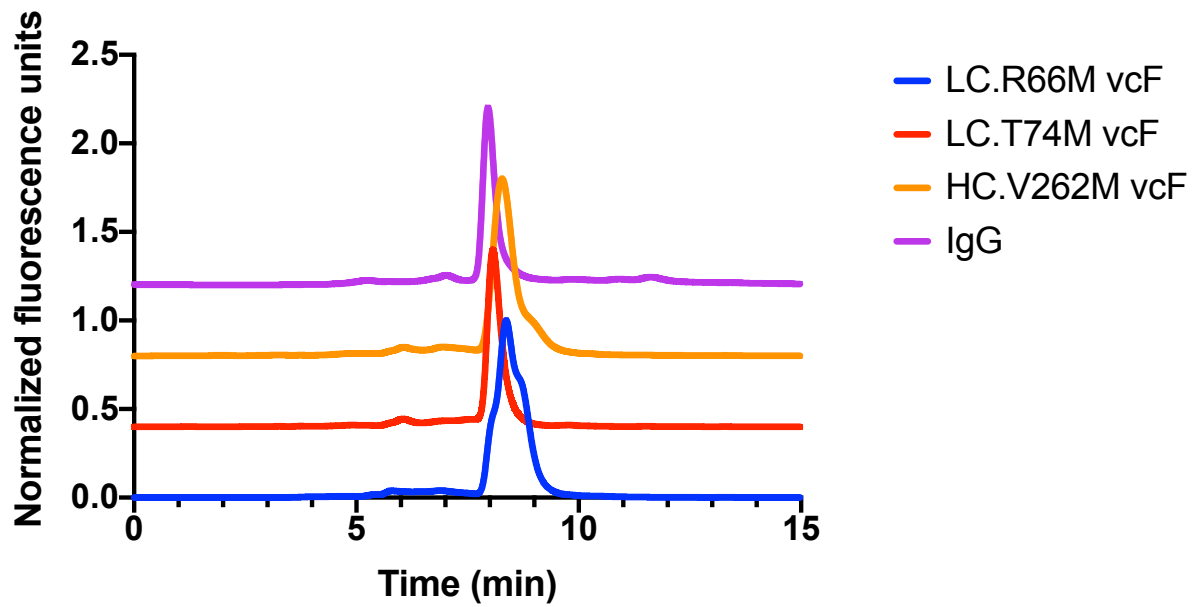
**Figure 1.13: Size exclusion chromatography (SEC) of drug labeled IgG ADCs.**

SEC analysis across three sites of drug labeled IgGs compared to an unlabeled trastuzumab IgG. Drug labeling at sites LC.T74M and HC.V262M both show similar SEC profiles to unlabeled IgG, but site LC.R66M shows a drastically different SEC profile shape.
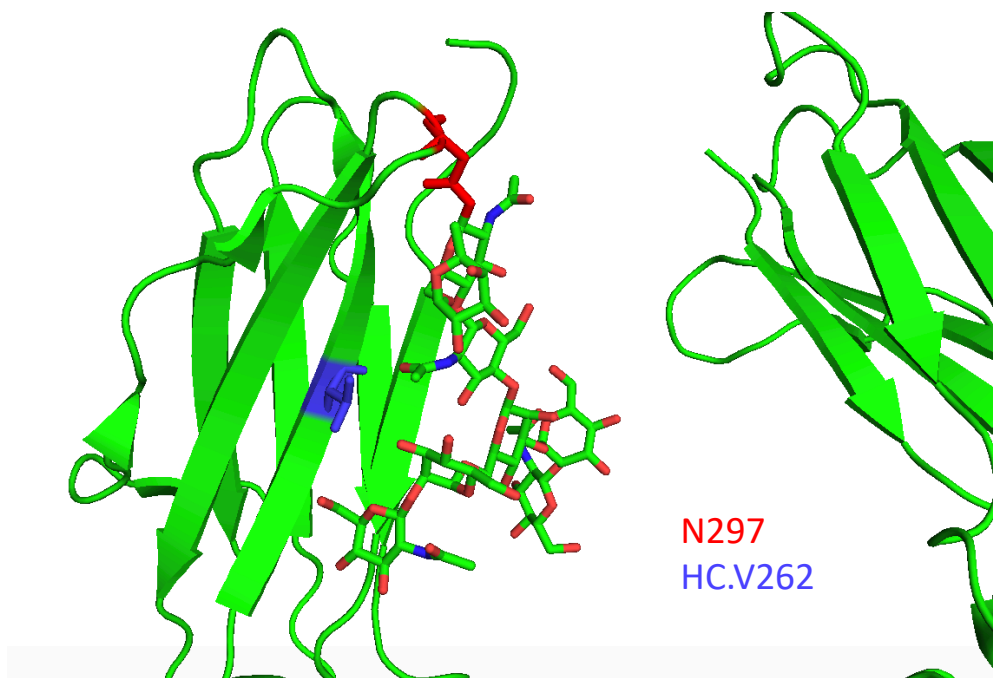
**Figure 1.14: Crystal structure showing the position of Fc glycosylation and site HC.V262.**

Crystal structure of a glycosylated Fc (PDB: 5XJF) demonstrates that the HC.V262 site (shown in purple) sits right beneath the glycans. Thus, it is likely that presence of these glycans inhibits oxaziridine conjugation to HC.V262M.
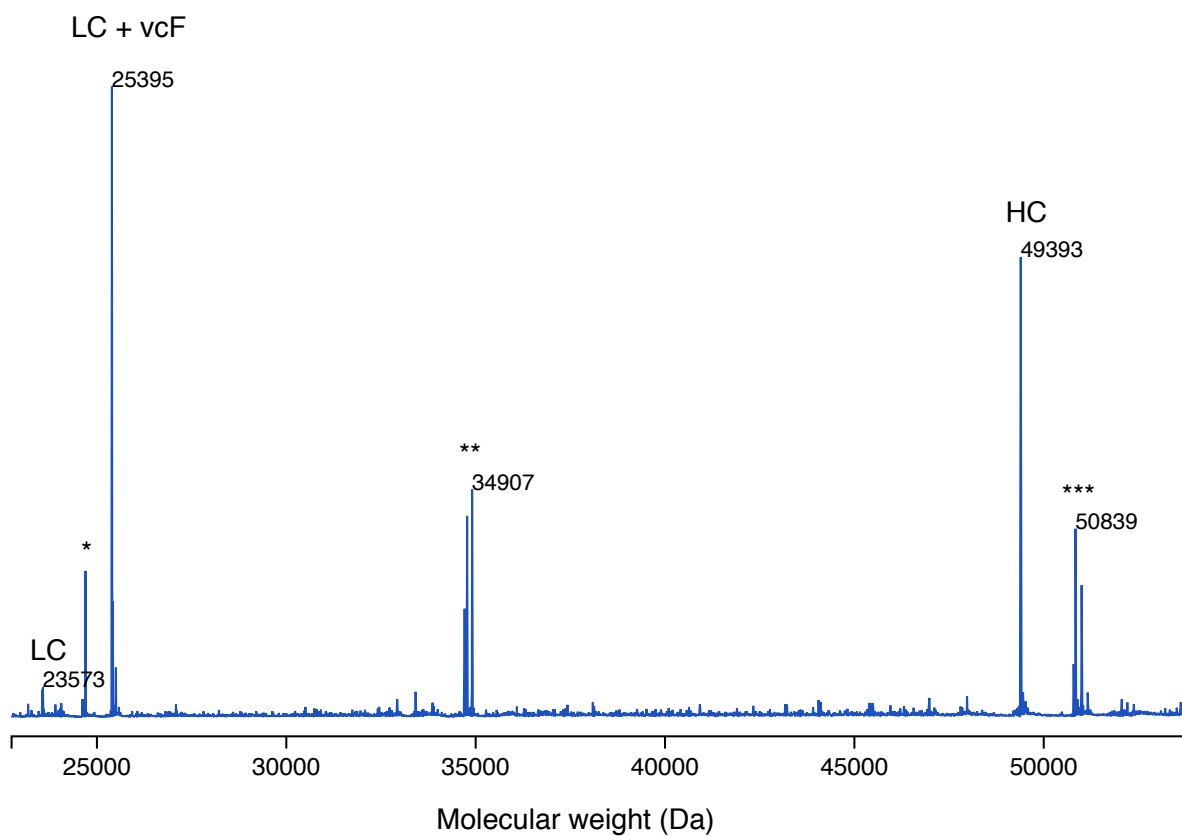
**Figure 1.15: Reduced mass spectrum of final LC. T74M IgG with DBCO-PEG4-vc-MMAF.** The spectrum shows high percentage labeling of the LC (MW: 23,573 Da) with oxaziridine and DBCO-PEG-vc-MMAF (MW: 25,395 Da). The deglycosylated HC is shown unlabeled (MW: 49,393 Da). The calculated average DAR for this prepared ADC was 1.9. The final yield was 30% of the original unlabeled IgG. *Denotes an artifact from MaxEnt, corresponding to half the molecular weight of the HC. **Denotes PNGase F added to remove glycosylation of the HC. ***Denotes not fully deglycosylated HC.
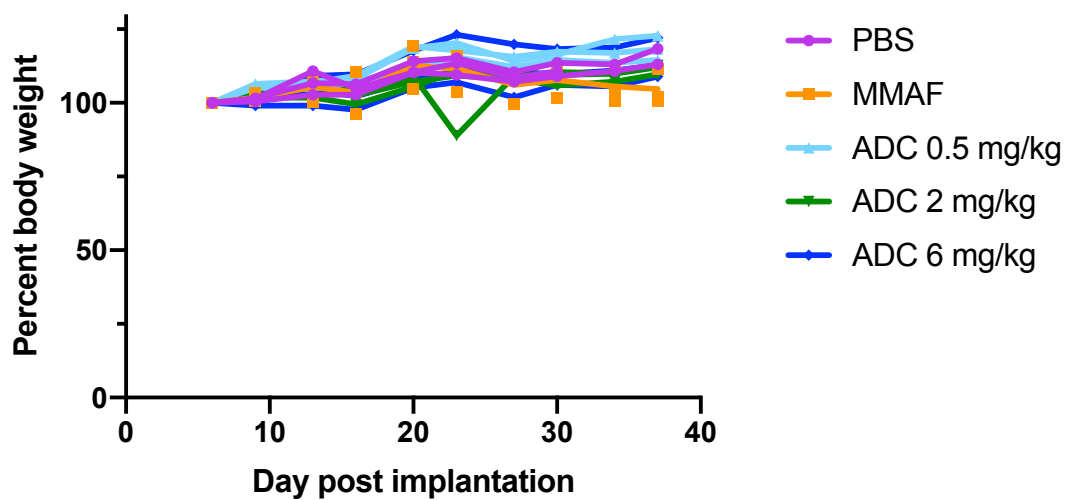
**Figure 1.16: Percent body weight of mice in the initial mouse xenograft study.**

Percent body weight is plotted over time for each mouse in the 5 groups over the course of the study. All mice maintained appropriate body weight, suggesting no toxicity was encountered.
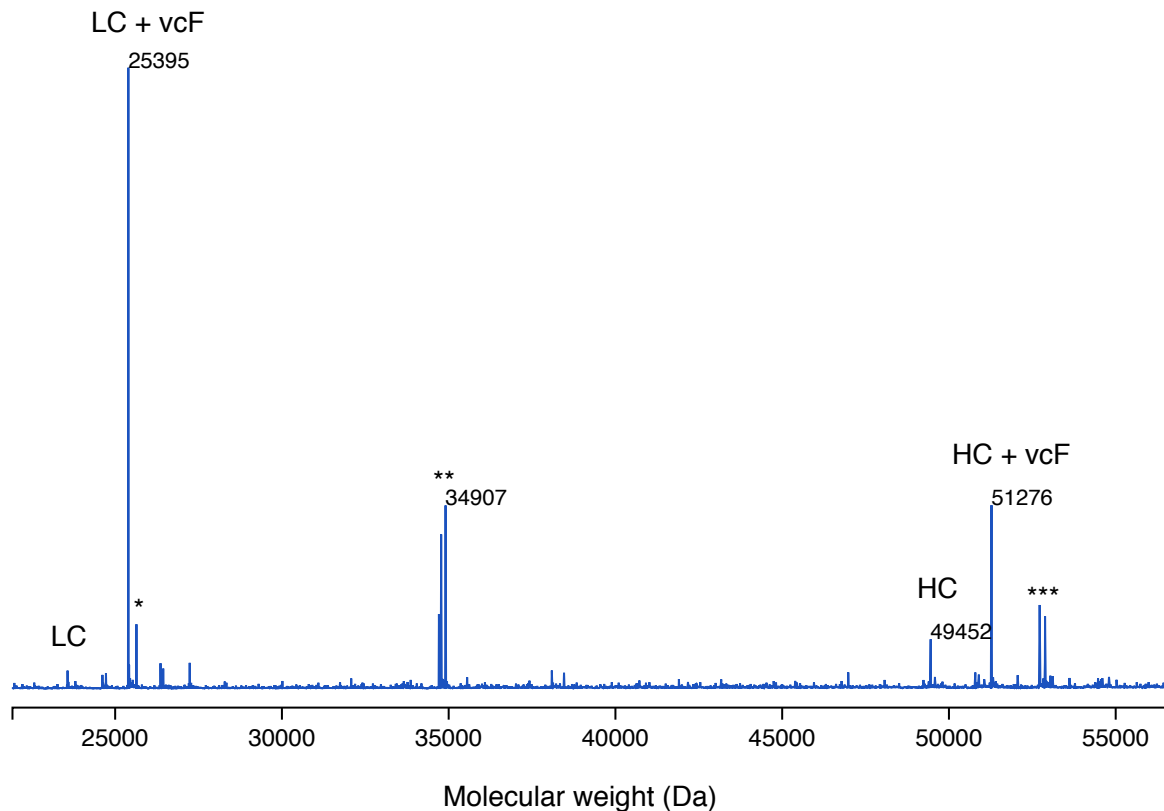
**Figure 1.17: Reduced mass spectrum of final DAR4 HC.S21M LC. T74M IgG with DBCO-**

**PEG4-vc-MMAF.**

The spectrum shows high percentage labeling of the LC (MW: 23,573 Da) with oxaziridine and DBCO-PEG4-vc-MMAF (MW: 25,395 Da). The deglycosylated HC is shown unlabeled (MW: 49,452 Da) and labeled with oxaziridine and DBCO-PEG4-vc-MMAF (MW: 51,276 Da). The calculated average DAR for this prepared ADC was 3.6. The final yield was 20% of the original unlabeled IgG. *Denotes an artifact from MaxEnt, corresponding to half the molecular weight of the labeled HC. **Denotes PNGase F added to remove glycosylation of the HC. ***Denotes not fully deglycosylated HC.

**LC.T74M HC.S21M vcF**
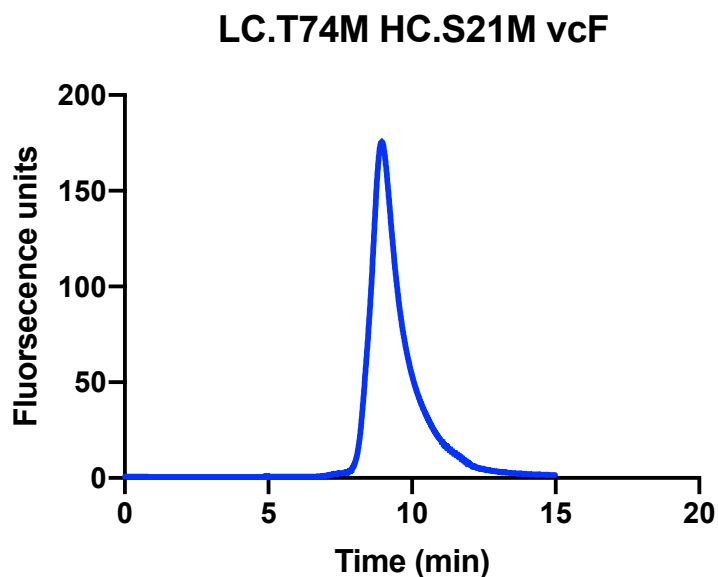


**Figure 1.18: SEC profile of the DAR4 IgG.**

SEC analysis shows a monodispersed peak for the DAR of four IgG and elution at a similar time to unlabeled IgG (Supplementary Fig. 6), suggesting no aggregation is present.
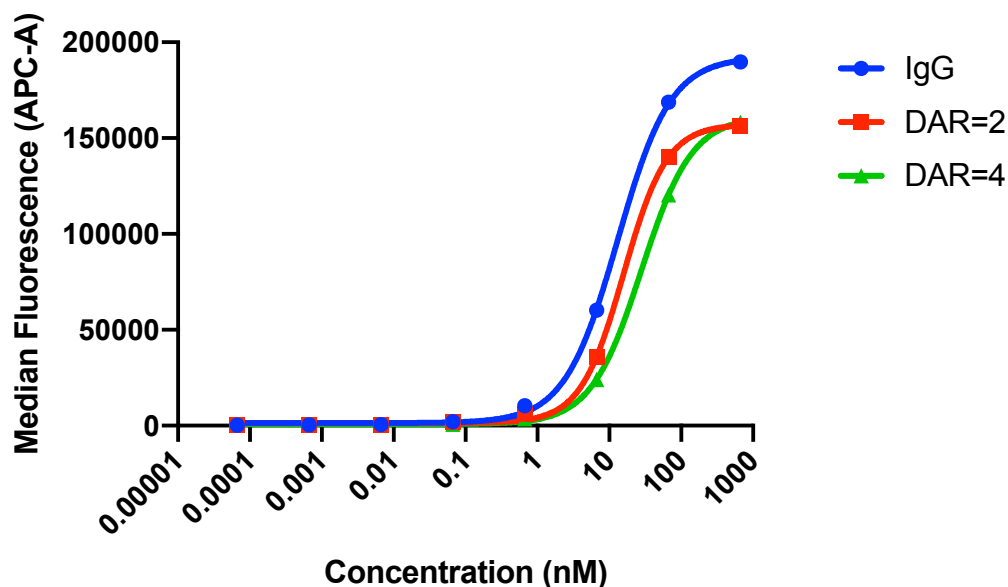


**Figure 1.19: On cell binding on SKBR3 cells for DAR2 and DAR4 ADCs.**

On cell binding was determined by serial dilutions of the IgG on HER2+ SKBR3 cells, followed by detection with a fluorescent anti-Fab antibody conjugate as previously described in the methods. IgG refers to Trastuzumab IgG (unlabeled IgG). EC50 values were determined (Trastuzumab (IgG): 13.3nM , DAR2: 16.0nM, DAR4: 28nM).
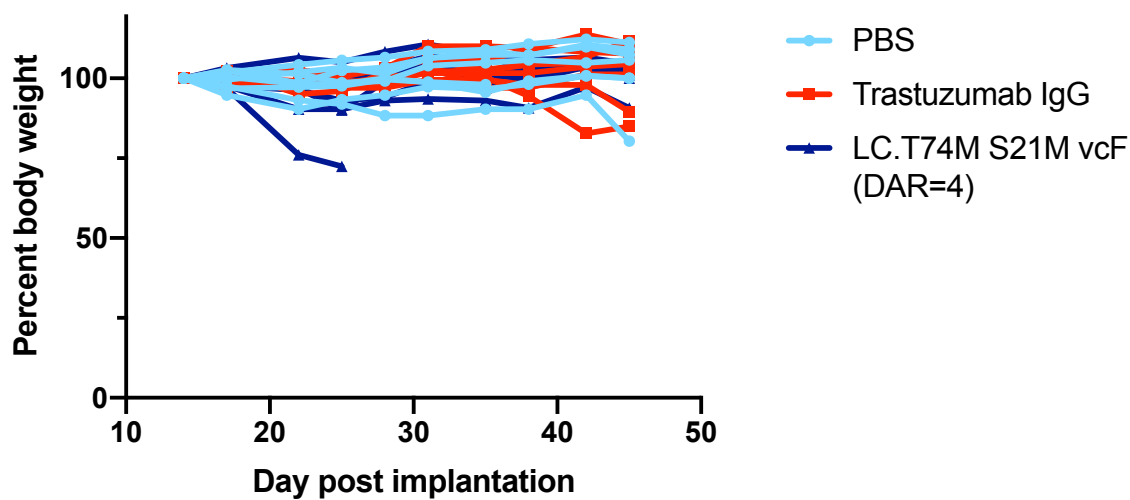
**Figure 1.20: Percent body weight of mice in the DAR4 ADC study.**

Percent body weight is plotted over time for each mouse in the 3 groups over the course of the study. All mice maintained healthy body weight except one mouse in the DAR4 group that was sacrificed at day 25.

**Table 1.1: Labeling, stability, and affinity of 95 methionine substitutions in model aGFP-Fab (rAB1001) in trastuzumab scaffold.**

| | Residue | Kabat | RSA | Yield (mg/L) | % Labeled | $K_d$ (pM) | | Thermostability (°C) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Pre | Post | Pre | Post |
| rAB1001 | - | - | 0 | 8 | 0 | 88 | 197 | 82.5 | 82.6 |
| LC-Cmet | - | - | 0.67 | 7.2 | 93 | 129 | 183 | 82.4 | 82.2 |
| LC001 | D | 1 | 0.94 | 7.1 | 76 | 118 | 293 | 83.2 | 82.5 |
| LC003 | Q | 3 | 0.95 | 0 | 85 | 138 | 58 | 82.5 | 82.2 |
| LC007 | S | 7 | 0.98 | 5.4 | 96 | 111 | 150 | 82.6 | 81.5 |
| LC009 | S | 9 | 0.86 | 16.9 | 93 | 96 | 339 | 81.9 | 81.5 |
| LC010 | S | 10 | 0.73 | 7.2 | 94 | 50 | 189 | 82.9 | 82.2 |
| LC012 | S | 12 | 0.67 | 8.3 | 63 | 115 | 169 | 81.7 | 80.3 |
| LC014 | S | 14 | 0.9 | 6.4 | 93 | 112 | 162 | 82.3 | 81.8 |
| LC018 | R | 18 | 0.86 | 10.9 | 80 | 80 | 246 | 82 | 81.2 |
| LC020 | T | 20 | 0.71 | 9.6 | 95 | 80 | 183 | 81.8 | 80.6 |
| LC041 | G | 41 | 0.92 | 7.3 | 95 | 132 | 175 | 81.7 | 82.4 |
| LC042 | K | 42 | 0.79 | 7.8 | 87 | 101 | 155 | 81.8 | 81.2 |
| LC045 | K | 45 | 0.73 | 5.9 | 93 | 167 | 216 | 80.9 | 79.9 |
| LC057 | G | 57 | 0.94 | 9.1 | 93 | 123 | 252 | 80.3 | 79.9 |
| LC060 | S | 60 | 1 | 6.3 | 92 | 125 | 267 | 81.7 | 81 |
| LC065 | S | 65 | 0.9 | 11.8 | 92 | 82 | 285 | 81.6 | 81.5 |
| LC067 | S | 67 | 1 | 6.6 | 94 | 120 | 205 | 82.2 | 81.9 |
| LC068 | G | 68 | 0.9 | 6.5 | 91 | 88 | 191 | 80.2 | 79.6 |
| LC069 | T | 69 | 0.83 | 11.4 | 69 | 103 | 170 | 82.3 | 81.7 |
| LC076 | S | 76 | 0.74 | 6.3 | 82 | 135 | 175 | 81.9 | 80.7 |
| LC081 | E | 81 | 0.69 | 6.4 | 84 | 116 | 213 | 79.9 | 78.8 |
| LC085 | T | 85 | 0.68 | 8.3 | 61 | 118 | 147 | 81.6 | 80.2 |
| LC100 | Q | 100 | 0.86 | 5.2 | 90 | 116 | 74 | 81.1 | 80.9 |
| LC107 | K | 107 | 0.84 | 5.1 | 90 | 148 | 149 | 81.5 | 80.1 |
| LC108 | R | 108 | 0.8 | 0 | - | - | - | - | - |

| | Residue | Kabat | RSA | Yield (mg/L) | % Labeled | $K_d$ (pM) | | Thermostability (°C) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Pre | Post | Pre | Post |
| LC112 | A | 112 | 0.73 | 6.4 | 88 | 94 | 201 | 80.2 | 77.7 |
| LC114 | S | 114 | 0.83 | 4.7 | 89 | 165 | 207 | 82.5 | 81.3 |
| LC123 | E | 123 | 0.76 | 8.1 | 57 | 147 | 183 | 82.3 | 81.8 |
| LC126 | K | 126 | 0.87 | 7.7 | 87 | 97 | 198 | 82.3 | 82.3 |
| LC128 | G | 128 | 0.98 | 6.5 | 90 | 94 | 177 | 81.6 | 81.3 |
| LC143 | E | 143 | 0.94 | 7.6 | 87 | 129 | 138 | 81.8 | 81.5 |
| LC145 | K | 145 | 0.82 | 6.3 | 93 | 93 | 190 | 82.4 | 82.2 |
| LC152 | N | 152 | 0.71 | 4.3 | 88 | 95 | 192 | 82 | 82.1 |
| LC156 | S | 156 | 0.99 | 9.8 | 89 | 124 | 190 | 82.6 | 82.3 |
| LC157 | G | 157 | 1 | 7.4 | 86 | 116 | 223 | 82 | 82 |
| LC160 | Q | 160 | 0.77 | 4.8 | 58 | 94 | 169 | 82.7 | 80.6 |
| LC168 | S | 168 | 0.92 | 0 | - | - | - | - | - |
| LC169 | K | 169 | 1 | 5.1 | 92 | 264 | 179 | 82 | 81.9 |
| LC182 | S | 182 | 0.9 | 8.1 | 89 | 92 | 168 | 82.3 | 81.6 |
| LC184 | A | 184 | 0.94 | 3.6 | 71 | 110 | 148 | 82.3 | 82.2 |
| LC190 | K | 190 | 1 | 4.4 | 94 | 227 | 202 | 82.5 | 82.3 |
| LC202 | S | 202 | 0.92 | 6.8 | 93 | 103 | 237 | 82.3 | 82.2 |
| LC203 | S | 203 | 1 | 6.3 | 92 | 211 | 152 | 82.3 | 82.2 |
| LC205 | V | 205 | 0.76 | 7.7 | 71 | 97 | 190 | 82.1 | 80.8 |
| LC210 | N | 210 | 0.72 | 7.6 | 88 | 79 | 178 | 82.6 | 82.4 |
| HC001 | E | 1 | 0.78 | 6.3 | 92 | 101 | 141 | 82.3 | 82.2 |
| HC003 | Q | 4 | 0.85 | 4.5 | 94 | 94 | 149 | 82.5 | 82.6 |
| HC005 | V | 5 | 0.89 | 4.9 | 96 | 139 | 111 | 82.4 | 82.5 |
| HC007 | S | 7 | 0.95 | 7.4 | 95 | 128 | 156 | 82.3 | 82.6 |
| HC008 | G | 8 | 0.72 | 5.2 | 95 | 141 | 189 | 81.2 | 81 |
| HC010 | G | 10 | 0.97 | 7.2 | 85 | 113 | 219 | 81.5 | 81.1 |
| HC013 | Q | 13 | 0.69 | 4.1 | 89 | 133 | 138 | 82.1 | 81.7 |
| HC015 | G | 15 | 0.98 | 4.9 | 92 | 114 | 170 | 81.8 | 81.6 |
| HC017 | S | 17 | 0.75 | 5.2 | 96 | 84 | 137 | 81.9 | 82 |
| HC019 | R | 19 | 0.86 | 4.3 | 88 | 83 | 47 | 82.3 | 81.9 |
| HC023 | A | 23 | 0.86 | 5.2 | 94 | 54 | 153 | 82.6 | 82.6 |

| | Residue | Kabat | RSA | Yield (mg/L) | % Labeled | $K_d$ (pM) | | Thermostability (°C) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Pre | Post | Pre | Post |
| HC042 | G | 42 | 1 | 5.4 | 96 | 143 | 183 | 81.8 | 82.2 |
| HC043 | K | 43 | 0.89 | 6.3 | 87 | 122 | 122 | 82.3 | 81.5 |
| HC044 | G | 44 | 0.72 | 7.9 | 94 | 72 | 165 | 80 | 78.7 |
| HC046 | E | 46 | 0.89 | 6.3 | 85 | 93 | 169 | 82 | 81.3 |
| HC066 | G | 65 | 1 | 5.8 | 92 | 216 | 155 | 82 | 81.8 |
| HC069 | T | 68 | 0.84 | 7.3 | 94 | 95 | 187 | 82.2 | 82.3 |
| HC071 | S | 70 | 0.84 | 6 | 82 | 130 | 197 | 82.2 | 81.8 |
| HC075 | S | 74 | 0.9 | 7 | 93 | 126 | 177 | 82.2 | 82.2 |
| HC076 | K | 75 | 0.88 | 10.2 | 93 | 129 | 224 | 82 | 82.2 |
| HC077 | N | 76 | 0.74 | 5.4 | 94 | 64 | 139 | 81.2 | 80.9 |
| HC084 | N | 82A | 0.79 | 6.7 | 95 | 110 | 151 | 81.9 | 82 |
| HC085 | S | 82B | 0.77 | 6.9 | 94 | 86 | 134 | 82.3 | 82.3 |
| HC087 | R | 83 | 0.71 | 7 | 93 | 96 | 193 | 82.2 | 82 |
| HC112 | Q | 105 | 0.85 | 8.8 | 96 | 76 | 168 | 82 | 82.1 |
| HC119 | S | 112 | 0.72 | 4.9 | 93 | 91 | 136 | 79.5 | 79.2 |
| HC120 | S | 113 | 0.7 | 6 | 78 | 105 | 137 | 82.4 | 82.3 |
| HC122 | S | 115 | 1 | 6.2 | 92 | 90 | 155 | 82 | 82.1 |
| HC124 | K | 117 | 0.82 | 4.4 | 92 | 108 | 217 | 81.8 | 81.6 |
| HC127 | S | 120 | 0.74 | 7.2 | 88 | 78 | 214 | 82.3 | 81.9 |
| HC137 | S | 130 | 0.97 | 7.2 | 88 | 72 | 188 | 82.2 | 81.2 |
| HC139 | S | 134 | 1 | 7.2 | 93 | 84 | 163 | 82.3 | 82 |
| HC141 | G | 136 | 0.92 | 5 | 94 | 114 | 177 | 82.3 | 82.3 |
| HC142 | T | 137 | 0.92 | 8 | 95 | 91 | 148 | 82.4 | 81.4 |
| HC151 | D | 146 | 0.76 | 3 | 49 | 94 | 138 | 68.8 | 69.9 |
| HC158 | T | 153 | 0.71 | 5.5 | 95 | 97 | 136 | 82.2 | 81.7 |
| HC163 | S | 163 | 0.99 | 17.7 | 94 | 74 | 177 | 82 | 82.4 |
| HC164 | G | 164 | 0.99 | 4.4 | 94 | 106 | 144 | 80.9 | 80.3 |
| HC168 | S | 168 | 1 | 4.6 | 93 | 125 | 94 | 82.2 | 82.2 |
| HC179 | S | 180 | 0.98 | 2.8 | 91 | 117 | 125 | 82 | 81.8 |
| HC180 | S | 182 | 0.73 | 2.8 | 92 | 161 | 158 | 80.9 | 80.3 |
| HC181 | G | 183 | 0.97 | 3.7 | 97 | 126 | 160 | 81.5 | 82.4 |

| | Residue | Kabat | RSA | Yield (mg/L) | % Labeled | $K_d$ (pM) | | Thermostability (°C) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Pre | Post | Pre | Post |
| HC194 | S | 196 | 0.98 | 4.6 | 90 | 107 | 90 | 82.2 | 82 |
| HC197 | G | 199 | 0.77 | 4.2 | 92 | 166 | 191 | 82.5 | 82.4 |
| HC198 | T | 200 | 0.94 | 4 | 94 | 122 | 134 | 82.2 | 81.9 |
| HC202 | I | 207 | 0.73 | 5 | 35 | 72 | 157 | 81.9 | 81.2 |
| HC206 | N | 211 | 0.92 | 5.7 | 94 | 110 | 144 | 82.5 | 81.8 |
| HC211 | N | 216 | 1 | 10.1 | 85 | 74 | 192 | 82.2 | 82.3 |
| HC213 | K | 218 | 0.99 | 6.7 | 91 | 89 | 126 | 82.3 | 82.1 |
| HC222 | S | 229 | 0.88 | 7.8 | 91 | 113 | 128 | 82.5 | 82.5 |

**Table 1.2: Labeling, stability, and affinity of 23 less accessible methionine substitutions in model αGFP-Fab in trastuzumab scaffold.**

| Site | Residue | Kabat | RSA | Yield (mg/L) | $K_d$ (pM) | Thermo-stability°C | % Labeling at 5x | % Labeling at 20x | % Stability at 37°C |
|------|---------|-------|------|------|------|------|------|------|------|
| LC022 | T | 22 | 0.36 | 11.9 | 770 | 82.3 | 92 | - | 42 |
| LC063 | S | 63 | 0.52 | 10.8 | 190 | 81.9 | 76 | - | 59 |
| **LC066** | **R** | **66** | **0.40** | **6.9** | **780** | **82.7** | **68** | **91** | **99** |
| LC069 | T | 69 | 0.83 | 21.8 | 450 | 82.7 | 55 | 84 | 62 |
| LC070 | D | 70 | 0.40 | 0.4 | 230 | 82.8 | - | - | - |
| LC072 | T | 72 | 0.56 | 0.5 | 270 | 82.2 | - | - | - |
| **LC074** | **T** | **74** | **0.44** | **31.2** | **820** | **81.4** | **47** | **83** | **96** |
| LC147 | Q | 147 | 0.48 | 25.5 | 490 | 82.5 | 61 | 85 | 49 |
| LC148 | W | 148 | 0.10 | 2.0 | 630 | 75.6 | 0 | - | - |
| LC149 | K | 149 | 0.27 | 3.2 | 520 | 82.3 | 55 | - | 88 |
| LC192 | Y | 192 | 0.12 | 0.5 | 290 | 78.2 | - | - | - |
| LC193 | A | 193 | 0.46 | 13.3 | 430 | 81.8 | 0 | - | - |
| LC195 | E | 195 | 0.49 | 49.7 | 430 | 82.4 | 70 | 87 | 55 |
| LC208 | S | 208 | 0.57 | 5.9 | 380 | 82.4 | 0 | - | - |
| HC018 | L | 18 | 0.46 | 33.2 | 650 | 82.6 | 0 | - | - |
| HC021 | S | 21 | 0.54 | 36.7 | 530 | 82.3 | 52 | 78 | 87 |
| HC025 | S | 25 | 0.42 | 17.2 | 800 | 82.7 | 81 | - | 65 |
| HC082 | Q | 81 | 0.67 | 0.0 | - | - | - | - | - |
| HC201 | Y | 206 | 0.13 | 22.4 | 720 | 81.0 | 0 | - | - |
| HC204 | N | 209 | 0.47 | 5.1 | 730 | 81.7 | 84 | - | 46 |
| HC212 | T | 217 | 0.26 | 23.7 | 18200 | 81.8 | 62 | 83 | 95 |
| HC215 | D | 220 | 0.49 | 21.3 | 550 | 80.8 | 85 | - | 55 |
| HC217 | K | 222 | 0.64 | 20.5 | 530 | 82.3 | 62 | 84 | 60 |

**Table 1.3: Alignment, labeling, and stability of chosen Fc sites.**

| Site | Residue | Kabat | Aligned Fab site | % labeling | % stability |
|------|---------|-------|------------------|-----------|-------------|
| HC262 | V | 275 | HC.S21 | 86 | 100 |
| HC292 | R | 309 | LC.R66 | 80 | 74 |
| HC307 | T | 326 | LC.T74 | 0 | - |
| HC382 | E | 407 | LC.K149 | 61 | 90 |
| HC437 | T | 468 | HC.T212 | 0 | - |

# Chapter 2

## Engineering luminescent biosensors for point-of-care SARS-CoV-2 antibody detection

**Abstract**

Current serology tests for SARS-CoV-2 antibodies mainly take the form of enzyme-linked immunosorbent assays or lateral flow assays, with the former being laborious and the latter being expensive and often lacking sufficient sensitivity and scalability. Here we present the development and validation of a rapid, low-cost solution-based assay to detect antibodies in serum, plasma, whole blood, and saliva, using rationally designed split luciferase (spLUC) antibody biosensors. The ability to lyophilize assay components also opens up the possibility of applications of this sensor format in low-resource settings. This new assay, which generates quantitative results in 30 minutes, substantially reduces the complexity and improves the scalability of COVID-19 antibody tests for potential point-of-care and broad population testing.

**Introduction**

As the COVD-19 pandemic continues worldwide, broad testing for SARS-CoV-2 infection still faces severe limitations. While nucleic acid testing is critical to detecting the virus, serological antibody tests are vital tools for monitoring the dynamic human humoral response to SARS-CoV-2 viral infection and vaccines response to primary virus and variants[38]. Antibody tests serve as a complement or an alternative to nucleic acid diagnostics for patients with a low viral load or for low-resource areas where expensive RT-PCR testing is difficult to access[39–41]. Serological tests also support therapeutic development either through identification of individuals who could serve as donors for convalescent serum therapeutics[42], or patients with potentially strong neutralizing antibodies that can be produced *in vitro* as new antivirals and prophylactics[43,44]. Importantly, as a vaccine is developed, population-scale, longitudinal evaluation of antibody responses is needed to determine the response to vaccination and the strength and duration of immunity to the primary virus and variants thereof. This would be greatly accelerated with an assay that is simple, rapid, and high-throughput without sacrificing accuracy and sensitivity[45–50].

Traditional serological assays are not optimal in the face of this broad pandemic. The most widely used laboratory serological tests take the form of enzyme-linked immunosorbent assays (ELISA)[46,51–53], which usually entail a >2-hour protocol involving several steps of protein incubation and washes, and is not readily amenable to deployment outside of a laboratory. A faster but significantly more expensive approach is a lateral flow assay[54,55]. However, lateral flow assays can produce less reliable results depending on the quality of the lateral flow device and different evaluation criteria[55,56]. In addition, lateral flow tests poorly capture the magnitude of a patient's antibody response as the test is qualitative and not quantitative. Here we provide a next-generation, simple, and rapid assay to meet the mounting needs for broad antibody testing in the face of the

ongoing pandemic and eventual vaccine deployment. The assay, which is compatible with serum, plasma, whole blood, and saliva samples, utilizes a simple split luciferase (spLUC) antibody sensor to generate quantitative serological data in 30 minutes. Testing of over 150 patient serum/plasma samples across three validation cohorts demonstrates that the spLUC assay has a sensitivity of 89% for detecting anti-S protein antibodies and 98% for anti-N protein antibodies, and a specificity of > 99% for both.

**Results**

*Engineering split luminescent (spLUC) biosensors for SARS-CoV-2 antibody detection*

When envisioning a next-generation serological assay, we hypothesized that sensitive biosensors for anti-SARS-CoV-2 antibodies could be utilized to greatly enhance the speed and simplicity of serological testing[57]. We constructed anti-SARS-CoV-2 antibody biosensors by fusing split Nanoluciferase (NanoLuc) fragments SmBiT and LgBiT[57] to SARS-CoV-2 viral protein antigens (**Fig. 2.1a**). Since an antibody has two Fragment Antigen Binding (Fab) arms, incubating serum with 1:1 mixed SmBiT and LgBiT biosensors will result in half of the anti-viral antibodies binding LgBiT with one Fab arm, and SmBiT with the other Fab arm. This hetero-bivalent interaction localizes the LgBiT and SmBiT fragments in close proximity, resulting in reconstitution of an intact, active NanoLuc enzyme for luminescence-based detection of reactive antibodies.

We chose to develop S and N sensors for SARS-CoV-2 antibody tests because COVID-19 patient antibodies are predominantly directed against epitopes on the viral S protein, which interacts with the host receptor angiotensin-converting enzyme 2 (ACE2) and mediates viral

entry[58], and the N protein, which packages the viral genome into a ribonucleocapsid[59]. These two viral proteins are the primary antigens used in the current COVID-19 serological tests[46,51,60–63].

The S sensors were constructed by fusing NanoLuc fragments to its receptor binding domain (S-RBD), which is the primary target of neutralizing antibodies (**Fig. 2.2a, Fig. 2.2b**)[46,51,63,64]. We modeled S-RBD binding to two antibodies, C105[43,65] , an ACE2-competing binder, and CR3022[66], an ACE2 non-competing binder, to determine linker lengths (**Supplementary text, Fig. 2.2c**). Based on the models, we constructed SmBiT fusions to the S-RBD C-terminus with 15 or 25 residue Glycine/Serine (GS) linkers (S15 and S25), and LgBiT fusions to S-RBD C-terminus with 5, 15, or 25 residue GS linkers (L5, L15 and L25). These variants varied in expression yields (**Fig. 2.2e**). Using recombinantly expressed S-RBD antibodies and ACE2 variants[67], we determined the optimal linker variant, enzyme concentration, buffer conditions, and impact of antibody-antigen binding affinity to signal strength (**Methods, Fig. 2.2, Fig 2.3, Fig 2.4**). The (L15 + S25) sensor pair at 1 nM enzyme concentration was identified as optimal, and these conditions were used for all subsequent assays.

In further characterizing the relationship between assay signal strength and antibody concentration/binding affinity, we performed ordinary differential equation modeling in R (**Methods and Fig. 2.5**). The modeling predicted a linear relationship between antibody concentration and luciferase signal (**Fig. 2.5b**), consistent with our experimental data (**Figure 2.1b**). In addition, the results highlighted that the sensors at 1 nM are more sensitive to an antibody binder with a $K_D = 1$ nM (**Fig 2.5b,c**). This threshold is equivalent to the median affinity reported for polyclonal antibody repertoires[68,69].

To construct the N sensors, we used the N-terminal sequence because aa 44–257 are found to be more immunogenic than the C-terminal dimerization domain (aa 258–419) (**Fig. 2.6a**)[70]. In

addition, dimerization promoted by the C-terminal domain may lead to high basal NanoLuc reconstitution levels. The atomic structures of N (aa 44–180)[59] showed the N and C termini are not in close proximity and therefore fusion at the N or C terminus may result in different sensor sensitivity (**Fig 2.6b**). Given this knowledge, three fusion sensor pairs were designed: (a) LN+SN: L/S-N(aa 44–257), (b) LC+SC: N(aa 44–180)-L/S (LC+SC), and (c) LC2+SC2: N(aa 44–257)-L/S, where L and S represent LgBiT/SmBiT, C represents C-terminal fusion, and N represents N-terminal fusion (**Fig 2.6c**). Testing on a commercial polyclonal anti-N protein antibody revealed that the LC + SC and LC2 + SC2 sensors generated stronger signals over LN + SN (**Fig 2.6d**). The LC + SC sensors generated linear, dose-dependent signals with commercial anti-N protein antibody (**Fig 2.1c**).

We next designed a simple and rapid protocol to assay a pilot set of serum samples from convalescent SARS-CoV-2 patients (**Fig. 2.1d**). Two healthy control sera collected before the emergence of SARS-CoV-2 virus were also tested. Serial dilutions (1:12.5, 1:25, and 1:50) of heat-inactivated sera were measured using S or N sensors. Robust, dose-dependent luminescence signal was observed across all serum concentrations tested, with the 12.5-fold dilution showing the highest signal (**Fig. 2.1e, f**). The S (L15+S25) sensors generated signal for all five patients tested (**Fig. 2.1e**). The N (LC+SC) sensors detected patient antibodies from all four patients tested (**Fig. 2.1f**). However, the N (LN+SN) sensors only detected antibodies from two patient sera samples that had the strongest seropositivity (**Fig. 2.6e**), which further confirmed a C-terminal fusion enhances NanoLuc reconstitution relative to the N-terminal fusion. Interestingly, much higher signals were generated with the N (LC+SC) sensor compared to the S sensor. This could be due to differing levels of anti-N or S antibodies in the patients[71], or a favored geometry for split enzyme reconstitution when anti-N antibodies bind to the N (LC+SC) sensors.

*Competitive spLUC assay to profile epitope-classes of antibodies*

In addition to a test to determine total binding antibodies, an assay that allows profiling of epitope classes of antibodies can be highly valuable. In this regard, competitive ELISA assays developed by us and others have enabled characterization of percentage of ACE2-competitive antibodies[63,72]. However, S-RBD is known to have multiple additional neutralization epitopes outside of the ACE2-binding site. An assay that allows rapid, unbiased profiling of those alternative epitopes could unveil further details of a patient's humoral response to neutralize SARS-CoV-2.

We first showed that spLUC assay can detect antibodies binding to various S-RBD epitopes (**Fig. 2.1g**). We expressed and tested four reported neutralizing antibodies against S-RBD. This includes: C004 and C105[43], which are ACE2-competitive binders; CR3022[66], which binds at a cryptic site outside of the ACE2-binding site; and C135[43], which does not compete with C004, C105, CR3022 or ACE2-Fc, representing a third binding epitope on S-RBD (**Fig. 2.7**). All four IgG antibodies generated dose-dependent luminescence signals at $\geq 0.1$ nM concentrations (**Fig. 2.1h**).

We then designed a competitive spLUC assay to determine presence of a specific epitope class of antibodies (**Fig. 2.1i**). Out of the four antibodies tested, C135, represents an unconventional and less understood epitope class. It neutralizes very potently (IC50 = 17 ng/ml) and may be used in combination with other ACE2-competitive binders as a cocktail therapy[43]. We converted C135 IgG to a single binding arm Fab binder, and pre-incubated 1 μM of C135 Fab with the S sensors to generate "blocked sensors". We then determined how much signal from a patient's sample corresponds to antibodies with a C135 epitope by comparing signals between the original and the "epitope masked" sensors (**Fig. 2.1i**). We assayed 12 patient serum samples with

representative high, medium, and low anti-S-RBD antibody levels at a 1:25 dilution of serum. IgG C135 served as a control for competition with Fab C135. As expected, the luminescence signal of IgG C135 was reduced by ~90% with the blocked sensors. Sera 7, 42, 98 showed >50% decreases in luminescence signals, indicating a large portion of the antibodies in these samples are C135-competitive (**Fig. 2.1j**). These results suggested that antibodies recognizing this unconventional, neutralizing S-RBD epitope are present in a significant proportion of patient samples. Performing this competitive spLUC assay with different competitive Fab antibodies in an expanded patient cohort could further our understanding of the distribution of epitopes on S-RBD and the correlation between epitopes and clinical outcomes.

*Characterization of larger cohorts of serum/plasma samples using the spLUC assay*

We next applied the spLUC assay in an expanded number of patients (**Fig. 2.8**). First, to determine assay cutoff values and specificity, which reflects how well an assay performs in a group of disease-negative individuals, we performed the tests on three cohorts of negative control samples (Total n = 144), which included mainly healthy individual samples, 12 seasonal coronavirus patient samples, and 20 flu vaccine pre- and post-vaccination samples. All controls were collected before the SARS-CoV-2 pandemic. These controls generated significantly lower luminescent signals than the COVID-19 patient sera samples (**Fig. 2.8a, b**). The range, median, mean and standard deviation values were calculated, and stringent cutoff values were determined by calculating the mean plus three standard deviations (**Table 2.1**). With these determined cutoffs, we calculated the specificity of the S sensors (1:12.5 serum dilution) to be 100% (56/56), and the N sensors (1:12.5 serum dilution) to be 99.2% (119/120).

We then used the spLUC assay to study three additional cohorts of patient samples (**Fig. 2.8a, b**). Cohort 1 is an outpatient cohort recruited at the Rockefeller University Hospital[43]. The samples were collected from individuals free of COVID-19 symptoms for ≥ 14 days. The S sensors showed 84.2% (48/57) sensitivity, and the N sensors showed 100% (56/56) sensitivity. Cohort 2 samples consisted of remnant sera from COVID-19 patients within Kaiser Permanente Hospitals of Northern California. These samples were drawn in any phase of infection, including the early acute phase. A subset of these patients, who may have not fully seroconverted at the time of sampling, had lower S sensor or N sensor signals compared to others in the spLUC assays. The sensitivities of the assays were 89% (49/55) for S sensors and 98% (46/47) for N sensors. Cohort 3 patients were part of the LIINC (Long-term Impact of Infection with Novel Coronavirus) study from San Francisco General Hospital and included plasma of a mixture of outpatient and inpatient samples drawn in the convalescent phase of the disease. With the S sensors, we detected antibodies in 94% (44/47) of outpatient samples and 100% (9/9) of inpatient samples. With the N sensors, we detected antibodies in 96% (45/47) of outpatient samples and 100% (9/9) of inpatient samples. For all cohorts, the S and N signals show a strong correlation (**Fig. 2.8c, Fig. 2.9**). Consistent with previous findings, we observed varying degrees of anti-S and N antibody seropositivity between patients (**Fig. 2.8a, b**), which reflects a wide range of patient humoral response to this virus[39,45].

Importantly, we observed strong correlation of spLUC assay results to anti-Fab and anti-IgG S-RBD ELISA signals (**Fig. 2.10 a-c**, R = 0.43-0.91). A base-10 logarithmic scale conversion was applied to the spLUC assay signals for the correlation analysis to ELISA signals. This non-linear correlation between the spLUC and ELISA assays is likely due to signal compression in ELISAs at high antibody concentrations (Abcam, ELISA guide). For all cohorts, the S sensor seronegative samples also had very low signals in S-RBD ELISA assays (**Fig. 2.10d-f**), which

confirmed the presence of low levels of anti-S-RBD antibodies in these sub-cohorts of patients. Interestingly, the correlations to IgM signals were much weaker (**Fig. 2.10a-c**). It is possible that IgM was not sensitively detected by the spLUC assay due to the weaker affinities of the individual binding arms in IgMs[73], or that the IgG response dominated the signal in many of the tested patients.

One of the key uses of a highly sensitive serology assay is to grade the quality of convalescent sera to neutralize virus[38]. In cohort 1, our analysis showed the S sensor signals correlated with the half-maximal neutralizing titers (NT50s) reported by Robbiani et al (**Fig 2.8d, left panel**), consistent with previous studies[43,47,51,74]. Moreover, we found that the N sensor signals showed a similar correlation with NT50 (**Fig. 2.8d, right panel**). Our results indicate determining either anti-S or anti-N seropositivity is a general means to assess the neutralization potential of sera samples.

We further analyzed our data in the context of demographic and clinical features. First, the degree of seropositivity for inpatient samples was significantly higher than that of outpatient samples (**Fig. 2.8a, b, e**). Disease severity scores and fever were also associated with a stronger antibody response (**Fig. 2.8f, g**). These results indicated a direct correlation of disease severity and adaptive immune response consistent with previous studies[39,41,43,45,47,75,76]. In addition, no statistical differences were observed in antibody titers between males and females (**Fig. Fig 2.6b 2.8h, Fig. 2.11a**) or for different age groups (**Fig. 2.8i, Fig 2.11b**). We expect that analyzing a larger number of samples can further unveil if infected older adults develop more antibodies than younger individuals as suggested in previous studies[55,56]. These results highlight that clinical features are correlated with the antibody response of COVID-19 patients.

Collectively, our assay showed high sensitivity and specificity for all three representative cohorts of serum/plasma samples (inpatient, outpatient, acute phase, convalescent phase), with an overall specificity of 100% (S sensor) and 99% (N sensor), and sensitivity of 89% (S sensor) and 98% (N sensor). These values are comparable or superior to reported values for laboratory ELISA and lateral flow tests[55,56].

Additionally, we sought to determine the inter-day and intra-day assay variability. To further reduce labor work, we used a liquid handler system to dispense plasma, sensors, and substrate. We characterized a set of 46 convalescent plasma samples multiple times over two days to determine assay variation (**Fig. 2.12a**). We observed average intra-assay coefficient of variations (CVs) of 6% for both S and N sensors, intra-day CVs of 7% for S sensors and 9% for N sensors, and inter-day CVs of 8% for S sensors and 9% for N sensors (**Fig 2.12b**).

The simplicity of the assay can greatly improve laboratory antibody test throughput. From our experience, one individual can run 4 plates of 96 samples (in total 384 assays) per hour by hand with multichannel pipettes. The automation of the assay is also straightforward as no wash step is involved. We showed that individual steps can be assisted with liquid dispenser instruments (**Fig. 2.12a**). We further set up a workflow for full automation of this assay using the UCSF Antibiome Center robotics platform (**Fig. 2.13a, b**)[24], which will further improve assay throughput. A simulated run for 40 plates (3840 assays) was estimated to take 3 hours (**Fig. 2.13c**). This robotic platform opens up the opportunity for characterizing hundreds of thousands of samples for population-wide epidemiology and longitudinal studies.

*Adapting the assay for low-resource settings and expanded sample types*

Lastly, we adapted our assays to begin to meet the clinical needs in remote and low-resources settings and for point-of-care deployment. A target assay profile was created based on a WHO consensus meeting report on high-priority target product profiles for new tuberculosis diagnostics (**Table 2.2**)[77]. While the current properties of the assay meet most of the assay profile requirements, we tested to see if the reaction time (30 minutes), reagent format (frozen aliquots of sensors), luminescence instrument type (plate-based luminometer), and sample type (serum/plasma) could be further optimized.

We first tested if our initial reaction times (20-minute sensor incubation and 10-minute substrate incubation, **Figure 2.1d**) are necessary and optimal. CR3022 (10 nM) was incubated with 1 nM S sensors for 5, 10, 15, and 20 min, followed by luciferase substrate addition and incubation for 0, 2, 4, 6, 8, and 10 minutes (**Fig. 2.14a**). We found all time points resulted in comparable, bright luminescence signal, suggesting the time of the assay protocol can be further shortened.

We then tested if the sensors can be lyophilized for ambient temperature storage and transportation. Although a small quantity (0-30%) of S sensors and N sensors were lost due to the lyophilization process (**Fig. 2.15a**), both the lyophilized S and the N sensors can still robustly detect recombinant IgG or patient antibodies in serum with similar sensitivities seen for the fresh sensors **Fig. 2.15b, c**). Furthermore, we found that vacuum drying of the commercial liquid form of the NanoLuc substrate, furimazine, and lyophilization of the substrate dilution buffer, did not affect the signals of the assay (**Fig. 2.15d**).

As plate readers may be difficult to access at point-of-care locations or in the field, we tested a battery-supported portable luminometer (32526-11 Junior LB9509, Berthold

Technologies), and found it detected the same concentration range of CR3022 compared to a Tecan M200 infinite plate reader (**Fig. 2.16**). The device uses a tube to read samples and therefore requires twice or more of the sample volume than in a 384-well plate, which (~2 μL of serum/plasma) is still compatible with the amount of sample generated from a finger prick. Together our data indicated both the reagents and instruments required for the spLUC assay are well suited for point-of-care or field applications.

Finally, we sought to determine if the spLUC assay could be compatible with other sample types. First, whole blood samples were collected from six convalescent COVID-19 patients and plasma samples were prepared in parallel for comparison (**Fig. 2.14b**). Remarkably, although the overall signals were lower from whole blood samples, all six samples generated N sensor signals and four had S sensor signals above control levels with the lyophilized sensors (**Fig. 2.14a**). In comparison, all six patients generated N sensor signals and five had S sensor signals above cutoff values from the plasma samples. Strong correlations were observed between the whole blood signals and the plasma signals (R > 0.9). Fresh and lyophilized sensors showed very little difference in performance.

Next, we tested the potential of using saliva as an input. To determine conditions, we added varying concentrations of the CR3022 antibody into saliva from a healthy individual (**Fig. 2.17**). We saw a significant reduction in sensitivity for undiluted saliva relative to buffer alone, but remarkably no loss in sensitivity when the saliva was diluted 1:2 in PBS buffer. We then tested 42 saliva samples at 1:2 dilution with the S sensors. We increased the reaction volume from 20 to 100 μl and the luminescence signal integration time from 1000 ms to 5000 ms for better sensitivity, as lower antibody concentrations are expected from saliva samples[78]. Out of the 42 samples, 33 had signals above the two healthy saliva controls, indicating a 79% assay sensitivity (**Fig. 2.14c**). A

moderate correlation of saliva signal with corresponding serum signals was observed (R = 0.66), consistent with recent reports[79]. These results highlight the potential of using lyophilized sensors/substrates and whole blood or saliva samples for rapid and quantitative point-of-care antibody testing.

**Discussion**

As the SARS-CoV-2 virus continues to spread, the need will continue to grow for serology assays to determine not only the scope of infection, but also vaccine efficacy during clinical trials and after large-scale vaccine deployment. We present here spLUC, a simple, sensitive, specific, fast, low-input sample volume, and quantitative solution-phase serological assay to detect antibodies against S and N proteins. We were able to test 159 patient samples across three different cohorts with varying clinical and demographic features. Our results enabled association analysis between these features and antibody titers, demonstrating the promise of this assay to generate large datasets to better understand factors that modulate the humoral response following SARS-CoV-2 infection.

The quantitative and solution-based nature of the spLUC assay allows convenient assay variations. We presented a competitive spLUC assay using epitope masked S sensors and used it to study the prevalence of an unconventional neutralization epitope in the S-RBD domain. This competitive spLUC assay has the potential to serve as a surrogate virus neutralization assay and to unveil details of the interaction of patient antibodies to viral antigens.

Robust ELISA-based assays such as the one developed by Krammer and co-workers have enabled tremendous progress of COVID-19 serological studies[51,61], but these assays are still laborious with multiple wash steps, which limits their broad application for population-scale sero-

surveillance, point-of-care diagnostics, and deployment in countries or remote areas that have limited access to analytical equipment and reagents. The spLUC assays have important features amenable to these applications. We have shown that our reagents are not only compatible with lyophilization for easy transport and storage, but can also readily detect antibodies directly from whole blood samples and saliva samples. With simple pipettes and a battery-supported portable luminometer, the spLUC assay could be readily established at care centers or in the field worldwide, regardless of infrastructure. To this end, we showed a commercially available portable luminometer is compatible with this assay, and we are currently collaborating with bioengineers to further develop luminometers that can be manufactured at lower cost but provide equal or better detection sensitivity. We have also shown that the assay is amenable to automation with standard liquid dispenser instruments, which allows for an opportunity to set up automated stations for self-testing at point-of-care locations.

Another important strength of our approach is the modularity. We expect that our strategy can be readily adapted to develop rapid serological tests for immunity against any infectious or autoimmune disease that elicits an antibody response for which the protein antigen is known. Our work identified important protein engineering details in constructing sensitive biosensors against anti-coronavirus antibodies, such as using the RBD domain of the S protein, removing the dimerization domain from the N protein, and fusing split enzymes to the C terminus of the N protein. These sensor designs are instructive for a rapid response if we have to face a future pandemic caused by a new strain of coronavirus. Moreover, the spLUC constructs are also nimble for engineering variant RBD domains to assess antibody response to new viral variants that are emerging. These should be useful for determining vaccine effectiveness in communities having multiple viral strains. Future development of our spLUC assay includes exploring orthogonal split

enzyme systems to allow multiplexing of assays. For instance, split ß-lactamase, used by Huang and co-workers for detecting herpes simplex virus antibodies[80], can provide an orthogonal readout to luminescence. We envision that such multiplexed assays could be used to develop broad-spectrum serological assays to simultaneously detect immunity against multiple infectious diseases.

In summary, we have taken a structure-based protein engineering approach to design novel split enzyme-fused sensors. These biosensors enable spLUC, a next-generation SARS-CoV-2 antibody test suited for population-scale sero-surveillance, epitope mapping of patient antibody responses, and testing in resource-limited areas. Future efforts will focus on continued evaluation of alternative sample source, collaboration with industry and clinical labs for broad deployment of the assay, and development of similar split enzyme-based serological approaches for a range of infectious diseases.

**Methods**

*Plasmid construction*

Plasmids were constructed by standard molecular biology methods. The DNA fragments of Spike-RBD, N protein, ACE2, and LgBiT were synthesized by IDT Technologies. The SmBiT tag was generated by overlap-extension PCR. The Spike-RBD-5/15/25aa-LgBiT-12xHisTag, Spike-RBD-15/25aa-SmBiT-12xHisTag, N protein(44-180)-10aa-LgBiT-12xHisTag, N protein(44-180)-10aa-SmBiT-12xHisTag, LgBiT-10aa-N protein(44-257)-12xHisTag, and SmBiT-10aa-N protein(44-257)-12xHisTag were generated by subcloning into a pFUSE-12xHisTag vector (adapted from the pFUSE-hIgG1-Fc vector from InvivoGen). The ACE2-Fc fusion plasmids were generated by subcloning the gene fragments of ACE2 and mutant into the pFUSE-hIgG1-Fc vector. The C004, C105, and C135 IgGs LC and HC plasmids were a generous gift from the Nussenzweig lab (Rockefeller University). The CR3022 IgG plasmids were a generous gift from the Kim lab (Stanford) and the Wilson lab (Scripps). The C135 Fab was cloned by removing the Fc domain from the HC plasmid. Snapgene and ApE were used for cloning design and sequence verification.

*Expression and protein purification*

All proteins were expressed and purified from Expi293 BirA cells according to established protocol from the manufacturer (Thermo Fisher Scientific). Briefly, 30 μg of pFUSE (InvivoGen) vector encoding the protein of interest was transiently transfected into 75 million Expi293 BirA cells using the Expifectamine kit (Thermo Fischer Scientific). For the IgG and Fab proteins, 15 μg of each chain was transfected. Enhancer was added 20 h after transfection. Cells were incubated for a total of 3 d at 37 °C in an 8% $CO_2$ environment before the supernatants were harvested by

centrifugation. Fc-fusion proteins were purified by Protein A affinity chromatography and His-tagged proteins were purified by Ni-NTA affinity chromatography. Purity and integrity were assessed by SDS/PAGE. Purified protein was buffer exchanged into PBS and stored at −80 °C in aliquots. Concentration was verified by absorbance on a NanoDrop with Nanodrop 2000c software.

### *Solution serology protocol for in vitro, serum, blood, and saliva samples*

LgBiT and SmBiT sensors for either the Spike or N protein were prepared at a final concentration of each sensor at 2nM in PBS + 0.05% Tween-20 + 0.2% BSA (PBSTB). For in vitro IgGs or ACE2-Fc, the samples were prepared at 1:10 dilutions in PBSTB unless otherwise specified. Serum and blood samples were diluted to 1:12.5 for both the S and N sensor samples in PBSTB unless otherwise specified. Healthy individual saliva was spiked in with CR3022 and used undiluted or diluted 1:2 in PBSTB. 10 μL of the 2 nM sensor mix and 10 μL of the sample were combined in a 384 Lumitrac white plate (Greiner), skipping every other well and row to avoid potential bleedover in signal. The plate was mixed on a plate shaker for 20 minutes. NanoLuc substrate was diluted according to protocol 1:50 in NanoLuc dilution buffer (Promega) and 15 μL was added to each well, followed by a 10-minute incubation period for the signal to stabilize. Luminescence was measured on a Tecan M200 infinite plate reader with an integration time of 1000 ms. Tecan i-control plate reader software was used for data acquisition.

### *Competition serology protocol for in vitro and serum samples*

The competition serology assay was performed similarly to the solution serology assay except that the S sensors were individually preincubated at 4 nM with 4 μM of either C004 Fab, C105 Fab, or

C135 Fab for the in vitro competition assay and C135 Fab only for the serum competition assay. The two sensors + Fab were combined 1:1 to make a 2 nM mix, and 10 μL of this mix was added to the assay as described above.

*Epitope binning experiment*

Biolayer interferometry data was measured using an Octet RED384 (ForteBio). Biotinylated Spike RBD protein was immobilized on the streptavidin (SA) biosensor (ForteBio). After blocking with biotin, the sensor was loaded with one IgG followed by another IgG or ACE2-Fc to determine epitope binning. PBS with 0.05% Tween-20 and 0.2% BSA was used for all diluents and buffers. ForteBio Octet Acquisition software was used for acquiring data, and ForteBio Octet Data analysis software was used for analyzing data.

*Spike protein ELISA assay*

The Spike ELISA assay was performed as previously described. Briefly, 384 Maxisorp plates were coated with 100 μL of 0.5 μg/mL Neutravidin for 1 hr. The plate was washed 3 times with PBS + 0.05% Tween-20 (PBST) followed by incubation with 20nM S-RBD for 30 minutes. Following 3 washes, the plate was blocked with 3% non-fat milk in PBS for 1 hour. The plate was washed 3 times before the addition of 1:50 dilutions of serum in 1% non-fat milk for 1 hour. After 3 washes, secondary anti-human Fab (Jackson ImmunoResearch Laboratories 109-036-097 [1:5,000]), anti-human IgM (Sigma-Aldrich A6907 [1:3,000]), or anti-human IgG (Sigma-Aldrich A0170 [1:3,000]) antibody was added and incubated for 30 minutes before the addition of TMB for 3 minutes. The reaction was quenched with 1 M phosphoric acid and absorbance was read on a Tecan M200 infinite plate reader at 450 nm.

*Inter-day and Intra-day assay with liquid dispensers*

Plasma was diluted 12.5 fold into a 96-well plate and subsequently 10µl was dispensed into an individual flat bottom white plate (Greiner LUMITRAC™ 200 384 well plates) using a Biomek FX$^P$ Automated Workstation (Beckmann Coulter). 10 µl of 1 nM biosensors were dispensed using Thermo Multidrop™ Combi Reagent Dispenser (ThermoFisher Scientific) to assay plates and incubated at room temperature for 20 min. 15 µl of substrate was added using the same reagent dispenser and incubated at room temperature for 10 min. Luminescence was read on a Tecan M200 infinite plate reader with an integration time of 1000 ms.

*Simulated automation of the spLUC assay on a robotic platform*

To facilitate high-throughput screening of serum, a semi-automated approach was developed and simulated using the UCSF Antibiome Center robotics platform[40]. Serum in 96-well plates would be first diluted 12.5-fold and 10µl is dispensed into 4 individual flat bottom white plate (Greiner LUMITRAC™ 200 384 well plates) using a Biomek FX$^P$ Automated Workstation (Beckmann Coulter). Serum containing assay plates would be then transferred to a robotics protocol with dispensing of biosensor and substrate followed by luminescence reading. While one iteration of 96 samples takes 40 mins, each additional iteration takes an additional 3.5 mins limited by luminescence reading (1s per well plus plate transfer). As such, it is estimated that 40 plates (3840 assays) could be run in 3 hours. The robotics run was developed and simulated using Thermo Momentum software (v5.0.6).

## Lyophilization of sensors and dilution buffer

The S and N protein sensors were flash frozen in liquid nitrogen at concentrations between 10-60 µM in 10 µL. The dilution buffer was frozen in liquid nitrogen in 5 mL aliquots. A small hole was poked into the caps of the samples and left on a Benchtop K (VirTis) lyophilizer overnight. The next day the sensors were reconstituted in 10 µL of ddH$_2$O and concentration was verified by nanodrop. The dilution buffer was similarly reconstituted in 5 mL of ddH$_2$O.

## Vacuum dried centrifugation of substrate sample

20 µL of substrate was aliquoted in a dark eppendorf tube and subjected to vacuum centrifugation overnight on a Genevac instrument. The substrate was stored in the dark for two days at room temperature before reconstitution in 20 µL of 100% methanol. The substrate was diluted 1:50 as normal in dilution buffer for the assay.

## Serum, plasma, whole blood, and saliva samples

The initial small patient cohort was a generous gift from the Wilson lab (UCSF) and heat inactivated at 56°C for 1 hour before storage at -80°C. The first (outpatient) sample serum set (cohort 1) was a generous gift for the Wilson lab (UCSF) and Nussenzweig lab (Rockefeller). These samples were heat inactivated at 56°C for 1 hour and stored at 4°C in a 1:1 dilution in 40% glycerol, 40 mM HEPES (pH 7.3), 0.04% NaN$_3$, in PBS. The second (inpatient) sample serum set (cohort 2) was a generous gift from the T. Wang lab (Stanford) and were stored at -80°C as pure serum samples. The third plasma cohort (cohort 3) and blood samples were generous gifts from the Greenhouse lab (UCSF) and Henrich Lab (UCSF) as part of the LIINC study. The plasma samples were stored at 4°C in a 1:1 dilution in 40% glycerol, 40 mM HEPES (pH 7.3), 0.04%

NaN$_3$, in PBS. The whole blood was stored undiluted at 4°C. Healthy blood samples were purchased from Vitalent and stored undiluted at 4°C. The saliva samples were obtained unstimulated, unexpectorated saliva and were stored at -80°C. Before assayed, the samples were thawed and centrifuged at 9,000g to remove any insoluble or coagulated matter. Control saliva from Nov 2019 was purchased from Lee Biosciences, stored at -20°C, and processed similarly.

### *Data and Statistical analysis*

All graphing and statistical analysis was performed in GraphPad Prism or Microsoft Excel. The unpaired multiple t test was performed in Prism to compare conditions in each patient for the competition spLUC assay. The non-parametric Spearman correlation analysis was used in Prism to determine the correlation R value between datasets. When two groups were compared, an unpaired Mann-Whitney test was performed to determine the difference between datasets. For comparison of three or more groups, a Kruskal-Wallis test with Dunn's multiple comparison post hoc testing. A two-tail P value was used to determine statistical significance for all analysis. $P < 0.05$ was considered statistically significant.

### *PBD Datasets*

PBD datasets 6XCN, 6W41, 1N8Z, 5IBO, 5D6D used for modeling were downloaded from https://www.rcsb.org.

### *Code availability statement*

R scripts used for ordinary differential equation modeling are available for download from: https://github.com/alexmartinko/Serological_ODE_Model .

*S sensor engineering and characterization*

*Linker modeling*

We modeled S-RBD binding to two antibodies to determine the optimal linker lengths between the S-RBD domains and the SmBiT/LgBiT fusions. The antibody C105 is an ACE2-competitive binder (**Fig. 2.2c**)[1,2], while the antibody CR3022 does not compete with ACE2 (**Fig. 2.2d**)[3]. Based on the assumption that the wing-span of antigen binding sites between Fab arms on a flexible-hinge region of an Fc are roughly ~117-134 Å apart[4], and residue-to-residue distance in a linker lies between the length of tightly packed alpha-helix residues (1.5 Å) and extended beta-strand residues (3.5 Å), we estimated the total number of linker residues should be ~30-80 amino acids. Antibodies binding to the CR3022 epitope may require a shorter linker for NanoLuc reconstitution (**Fig. 2.2d**) than antibodies competitive with ACE2 (**Fig. 2.2c**). Considering S-RBD has a C-terminal 15-residue loop to function as part of the linker, we constructed SmBiT fusions to S-RBD C-terminus with 15 or 25 residue Glycine/Serine (GS) linkers (S15 and S25), and LgBiT fusions to S-RBD C-terminus with 5, 15, or 25 residue GS linkers (L5, L15 and L25). These linker variants were expressed in Expi293 cells and varied in expression yields (**Fig. 2.2e**). The N-terminal fusions to S-RBD were not designed because the N and C termini localize in close proximity and we hypothesized this alternative fusion design would result in similar sensor performance as the C-terminal fusions (**Fig. 2.2b**). All modeling was performed in PyMoL.

*Optimization of enzyme concentrations, linkers and buffer conditions*

We then determined the optimal enzyme concentration. A three-fold dilution series from 27 to 0.11 nM of the L15 + S25 sensors were mixed with increasing 10-fold dilutions of recombinant CR3022 (**Fig. 2.2f**). After a 20-minute incubation, the NanoLuc substrate was added

and allowed to develop for 10 minutes before luminescence signal was read. High sensor concentrations (27, 9, 3 nM) resulted in stronger background luminescence signal and therefore lower detection sensitivity of CR3022, due to increased basal association of the two split sensors. Meanwhile, low sensor concentrations (0.33 and 0.1 nM) generated overall less signal than 1 nM sensors because fewer sensors are captured on each antibody. As a result, sensors at 1 nM were used in all subsequent assays.

Next, we queried if linker lengths affect detection sensitivity. Sensors with varied linker lengths were mixed with 10-fold dilutions of CR3022 and all resulted in dose-dependent luminescence signals (**Fig. 2.2f**). Little difference in detection sensitivity was observed, except that the (L5 + S15) and (L5 + S25) linker combinations resulted in slightly decreased sensitivity at low antibody concentrations. This result indicated that we had selected a proper range of linker lengths. Based on robust signal and expression yields (**Fig. 2.2e**), we chose the L15 and S25 sensor pair for subsequent assays.

Interestingly, we observed that the regular PBSTB assay buffer (PBS, 0.05% Tween-20, 0.2% m/v BSA, PBSTB) produced a higher background signal (average relative luciferase units (RLU) = 70-80) than in serum samples (RLU = 24.5). We tested if supplementing Fetal Bovine Serum (FBS) can reduce background (**Fig. 2.4**). PBS + 0.05% Tween-20 (PBST) with 4-10 % FBS was found to reduce the signal (mean RLU = 21) to a level that is close to signal from 12.5% serum, and therefore can serve as a proper negative control. Both the recombinant anti-S antibody C004 and the commercial anti-N antibody (Sino biological, Cat#40588-T62-50) produced linear dose-dependent signal in this buffer (**Fig. 2.1b,c**), which can be used to generate standard curves and calibrate the instruments for the spLUC assay.

*Impact of binding affinities*

To determine whether the affinity of the target binding to S-RBD affects signal strength, we turned to two dimeric ACE2 constructs: ACE2-Fc, which is the human ACE2 peptidase domain fused to IgG1 Fc[5], and an engineered ACE2-Fc variant that binds ~10x tighter to S-RBD (**Fig. 2.3**). Overall, signal from wild-type ACE2-Fc ($K_D$ = 10 nM) is weak, with signal that is more than two standard deviations above background only detected at the highest tested ACE2-Fc concentration (10 nM). Conversely, the enhanced-affinity ACE2-Fc variant ($K_D$ = 1 nM) generated a dose-dependent signal from 0.1-10 nM protein concentrations and exhibited 2.6-fold higher signal observed at 10 nM relative to the wild-type ACE2-Fc. These findings indicated the sensors report the presence of not only larger quantities of anti-S-RBD binders but also higher-affinity binders. This property of the sensors suggested spLUC assay may be used to characterize binding affinities of S-RBD antibodies or ACE2 variants for therapeutic applications.

*Thermodynamic sensor model*

In further characterizing the relationship between assay signal strength and antibody concentration/binding affinity, we performed ordinary differential equation modeling in R. We made assumptions such as a sensor can only be bound by one antibody, that antibody binding is non-cooperative, and that there is no detectable basal affinity of LgBiT and SmBiT at the concentrations tested (**Fig. 2.5a**). The modeling predicted a linear relationship between antibody concentration and luciferase signal (**Fig. 2.5b**), consistent with our experimental data (**Fig. 2.1b,c**).

The following set of ordinary differential equations (ODEs) was written to describe the system depicted in **Fig. 2.5a** and generated the curve graphs in **Fig. 2.5b, c**:

$$\frac{d[A]}{dt} = -k_{1f}[C][A] - k_{1f}[D][A] - k_{1f}[E][A] + k_{1r}[D] + k_{1r}[G] + k_{1r}[H]$$

$$\frac{d[B]}{dt} = -k_{1f}[C][B] - k_{1f}[E][B] - k_{1f}[D][B] + k_{1r}[E] + k_{1r}[I] + k_{1r}[H]$$

$$\frac{d[C]}{dt} = -k_{1f}[C][A] - k_{1f}[C][B] + k_{1r}[D] + k_{1r}[E]$$

$$\frac{d[D]}{dt} = -k_{1r}[D] - k_{1f}[D][A] - k_{1f}[D][B] + k_{1f}[C][A] + k_{1r}[G] + k_{1r}[H]$$

$$\frac{d[E]}{dt} = -k_{1r}[E] - k_{1f}[E][A] - k_{1f}[DE][B] + k_{1f}[C][B] + k_{1r}[H] + k_{1r}[I]$$

$$\frac{d[G]}{dt} = -k_{1r}[G] + k_{1f}[D][A]$$

$$\frac{d[H]}{dt} = -k_{1r}[H] - k_{1r}[H] + k_{1f}[D][B] + k_{1f}[E][A]$$

$$\frac{d[I]}{dt} = -k_{1r}[I] + k_{1f}[E][B]$$

Where:

A = LgBiT sensor

B = SmBiT sensor

C = Antibody

D = Antibody/ LgBiT sensor heterodimer

E = Antibody/ SmBiT sensor heterodimer

G = Antibody/ LgBiT sensor/LgBiT sensor trimer

H = Antibody/Active Enzyme trimer (Active Enzyme)

I = Antibody/ SmBiT sensor/SmBiT sensor trimer

$k_{1f}$ = on rate of Antibody binding to Spike

$k_{1r}$ = off rate of Antibody binding to Spike

For simplification, we assumed the following: 1) LgBiT sensor and SmBiT sensor had no measurable interaction, 2) Antibody binding to LgBiT sensor or SmBiT sensor was non-cooperative, and 3) Antibody binding to LgBiT sensor was equivalent in rate to antibody binding to SmBiT sensor. The equations above were solved in R using the deSolve package to find the concentration of each species at equilibrium. In all cases the initial concentrations of D,E,G,H, and I were set to 0. All thermodynamic modeling was performed in R, Rstudio and the deSolve package.

**Figures and Tables**
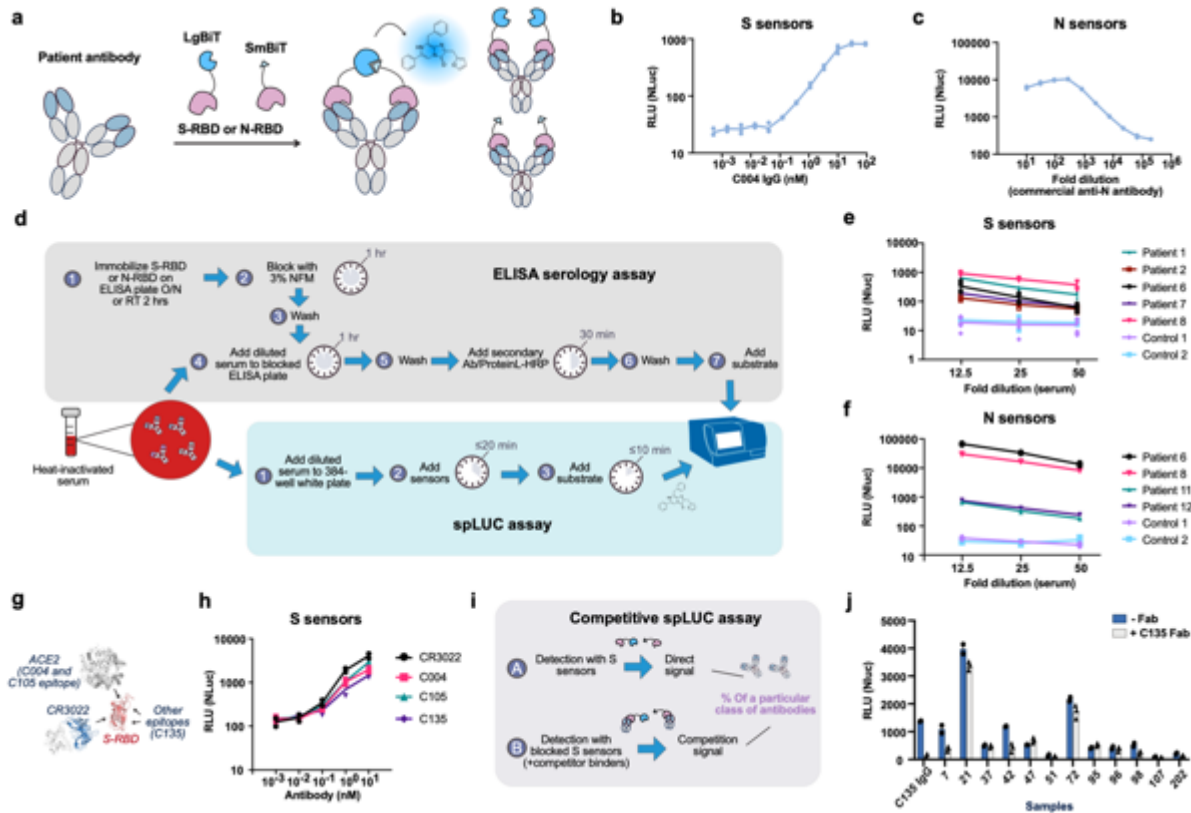


**Figure 2.1: Engineering luminescent biosensors for rapid and quantitative detection of SARS-CoV-2 antibodies. a,** Schematic of the solution-based serology assay. Patient antibodies are incubated with SARS-CoV-2 S or N proteins fused to LgBiT/SmBiT. For the population of antibodies with one arm bound to the LgBiT sensor and the other arm bound to the SmBiT sensor, the NanoBiT luciferase enzyme is reconstituted and thus can produce active luciferase signal. **b,** Dose-dependent spLUC signals for the recombinant anti-S-RBD antibody C004 in PBST + 10% FBS. Two technical replicates are plotted from n=1 independent experiment. **c,** Dose-dependent spLUC signals for an anti-N-RBD antibody (Sino Biological, Cat#40588-T62-50) in PBST + 8% FBS. Two technical replicates are plotted from n=1 independent experiment. **d,** Comparison of assay procedure between the ELISA and the spLUC assay. While ELISA takes > 2 hours and involves multiple wash and incubation steps, the spLUC solution-based assay is simply completed in ≤ 30 minutes without the need for wash steps. **e,** The S (L15+S25) sensors are able to detect antibodies in 5/5 COVID-19 recovered patients. At all dilutions tested, all 5 patients generated signal above the background signal of two control serum samples collected before the pandemic. Each dot represents a technical replicate. N=2 independent experiments with 3 replicates each are plotted for all samples except Patient 1, Patient 7 and Control 2, which have n=1 independent experiment plotted due to limited reagents. **f,** The N (LC+SC) sensors are able to detect antibodies in 4/4 COVID-19 recovered patients. At all dilutions of serum tested, all 4 patients generated signal above the background signal of two control serum samples collected before the pandemic. Two technical replicates are plotted from n=1 independent experiment. **g,** Patient antibodies for SARS-CoV-2 have various epitopes on the S-RBD (red). C004 and C105 have ACE2-competitive

epitopes, while C135 and CR3022 (blue) have non-ACE2 competitive epitopes. **h,** S sensors can detect patient antibodies of various epitopes with similar sensitivity. C004, C105, C135, and CR3022 patient antibodies were incubated with the S sensors at 10-fold antibody dilutions from 10 nM to 0.001 nM. The average of three technical replicates from n=2 independent experiments are plotted **i,** Schematic of antibody epitope competition assay with patient serum samples. Direct signal is compared to signal generated in the presence of the pre-incubated 1 µM Fab +1 nM sensor. **j.** Competition assay performed with C135 Fab on twelve outpatient sera samples and recombinant C135 IgG protein. Samples were incubated with either no Fab (blue) or C135 Fab (off-white). Sera 7, 42, 98 showed >50% decreases in luminescence signal, suggesting the presence of antibodies with the C135 epitope. Each dot represents a technical replicate from n=1 independent experiment. The center of the bar represents the mean of the measurements. Direct signals (-Fab) were measured in technical duplicates and competition signals (+ C135 Fab) samples were done in technical triplicate. For **b,c,e,f,h** the center of the line represents the mean of all measurements. Lines connecting the means of the samples are plotted.

**a**

SARS-CoV-2 Spike protein (protein S)

**b**

S-RBD

**c**

100.9 Å    95.7 Å

C105

**d**

43.5 Å    41.5 Å

CR3022

**e**

| # | Sensor | Yield (mg/L) |
|---|--------|--------------|
| L5 | S(aa 328-533)-5aa-LgBiT | 1.5 |
| L15 | S(aa 328-533)-15aa-LgBiT | 11.1 |
| L25 | S(aa 328-533)-25aa-LgBiT | 2.8 |
| S15 | S(aa 328-533)-15aa-SmBiT | 21.4 |
| S25 | S(aa 328-533)-25aa-SmBiT | 26.5 |

**f** S sensors (L25+S15)

**g**
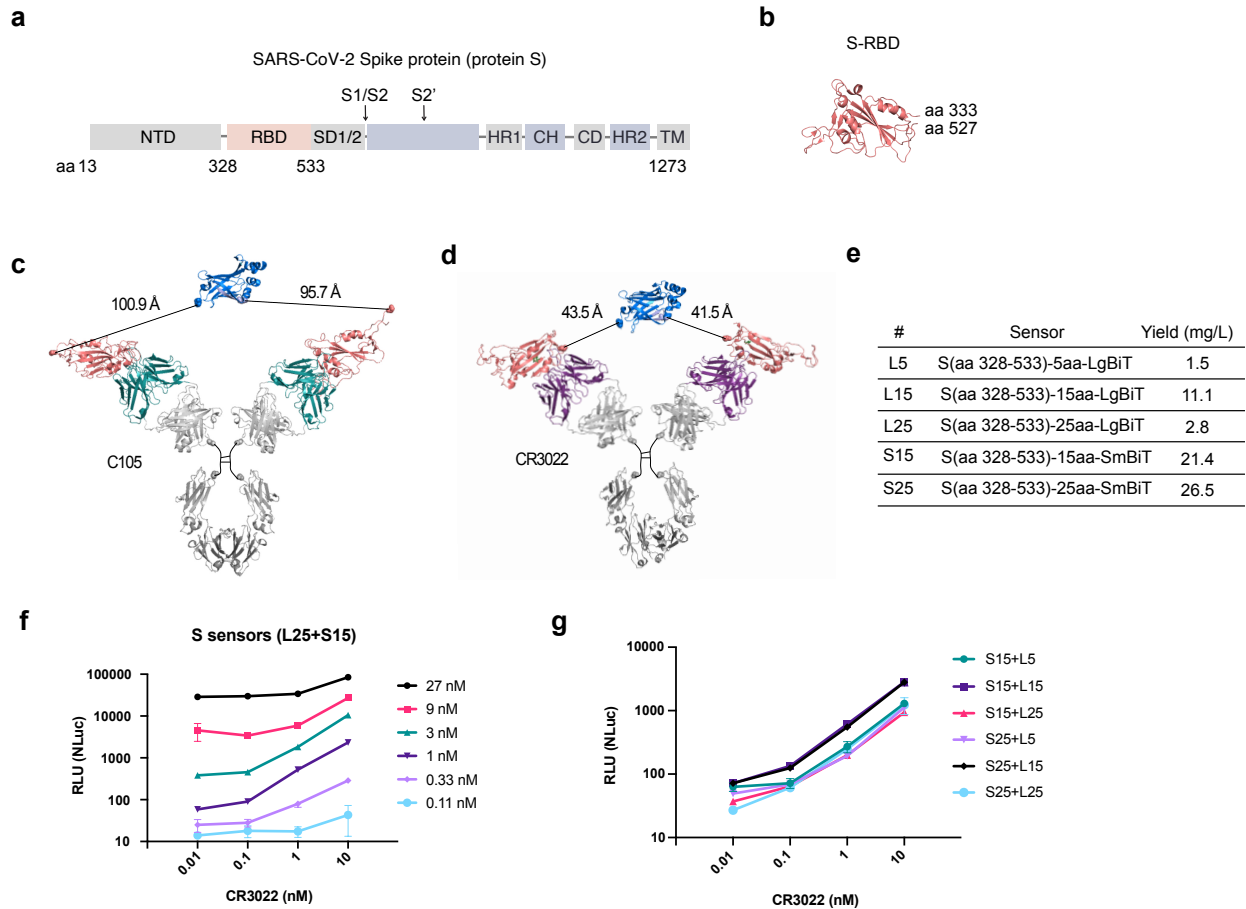
**Figure 2.2: Design and characterization of S sensors. a,** Annotated depiction of the SARS-CoV-2 Spike protein. The S sensors were developed using only the S-RBD domain (aa 328 – 533, PDB: 6W41) shown in pink. **b,** Structure of the S-RBD domain shows the N and C termini locate in close proximity. **c, d,** Modeling of c, ACE2-competitive antibody C105 (PDB: 6XCN) binding to S-RBD-SmBiT/LgBiT sensors, and d, CR3022 (PDB: 6W41) binding to S-RBD-SmBiT/LgBiT sensors. Modeling and distance measurements were performed with PDB 6XCN, 6W41, 1N8Z, 5IBO and 5D6D in PyMOL. **e,** Yield of the 5 Spike-NanoBiT sensor fusions. The Spike LgBiT sensors were made with 5aa, 15aa, and 25aa Glycine-Serine (GS) linkers (L5, L15 and L25). The Spike SmBiT sensors were made with 15aa, and 25aa GS linkers (S15 and S25). Because the N and C termini of the S-RBD domain locate in close proximity, only fusions to the C termini of S-RBD were constructed. **f,** The S sensors are most sensitive at 1 nM for detecting CR3022 in solution compared to higher or lower sensor concentrations. Two technical replicates are plotted from n=2 individual experiments. Lines connecting the means of the samples are plotted. **g,** S sensors with varied linker lengths resulted in very similar signal strength in detecting CR3022. Two technical replicates are plotted from n=1 individual experiment. Lines connecting the means of the samples are plotted.
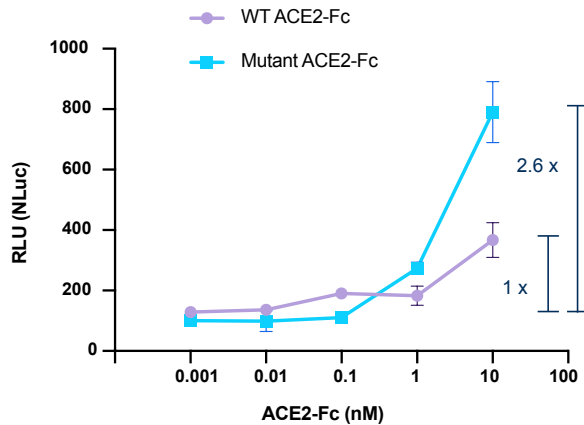
**Figure 2.3: The biosensors are more sensitive to high-affinity binders.** The ACE2-Fc variant which binds 10-fold tighter to S-RBD generated ~3-fold higher signal at 10 nM protein concentration comparing to WT ACE2-Fc. Two technical replicates are plotted from n=1 individual experiment. Lines connecting the means of the samples are plotted.
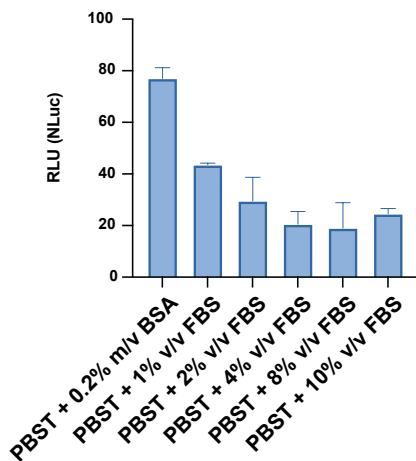


**Figure 2.4: Supplementing FBS reduces background signal in spLUC assays.** PBST with 4-10% FBS can be used as a negative control for serum samples as it shows similar signal suppression. Two technical replicates are plotted from n=1 individual experiment. Center of the bar represents the mean.
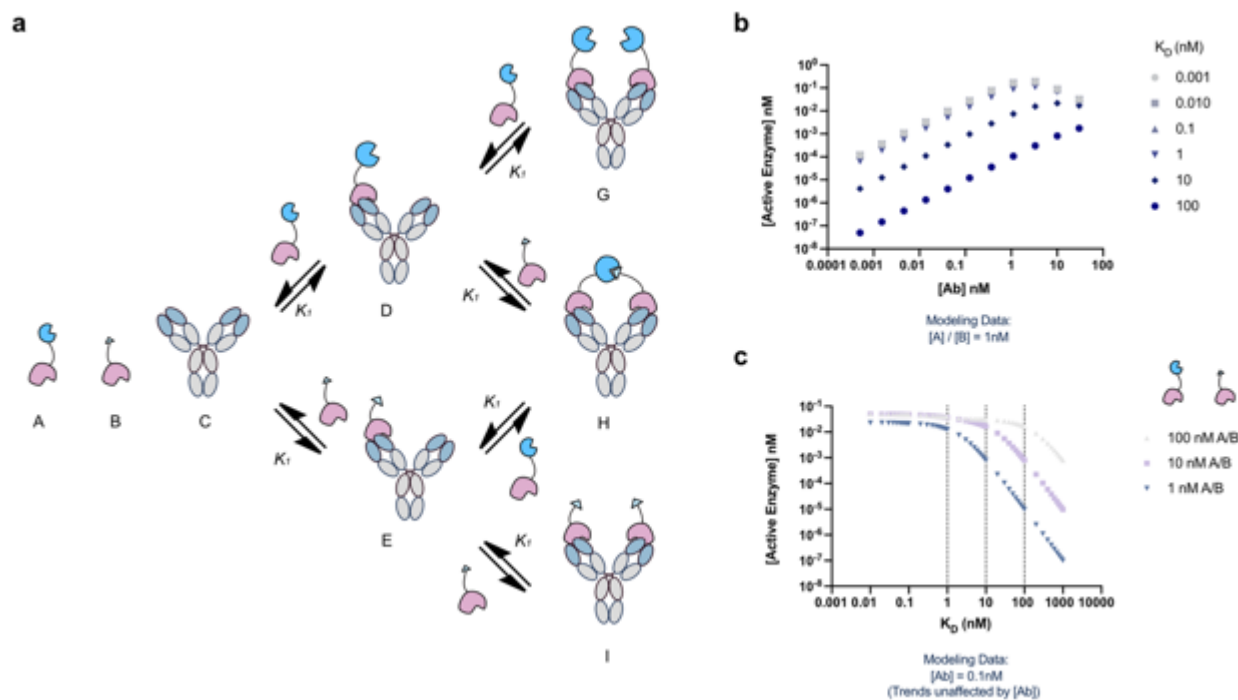
**Figure 2.5: ODE models predict a linear, dose-dependent response and KD dependence of the luminescence signal. a,** Antibody (C) and sensor components (A and B) are in thermodynamic equilibrium with enzymatically inactive (D, E, G, and I) and active (H) sensor bound species. **b,** At 1 nM starting concentration of sensor ([A] and [B]), spLUC assays are predicted to generate signals linearly correlated to a broad range of antibody concentrations ([Ab]). Signal is predicted to be insensitive to antibody concentrations for antibodies with high affinity for the sensor ($\leq$ 1nM), but weaker affinity antibodies ($K_D$ > 1 nM) will result in significantly lower levels of reconstituted enzyme. **c,** At $K_D$ values equivalent or higher than the sensor concentrations, the spLUC signals are predicted to drop significantly.

**a**

SARS-CoV-2 Nucleocapsid protein (protein N)

| RNA-binding domain | | Dimerization domain |

aa 1    44                180          257                419

**b**

N-RBD

aa 44                aa 180

**c**

| # | Sensor | Yield (mg/L) |
|---|--------|--------------|
| LC | N(aa 44-180)-10aa-LgBiT | 6.4 |
| SC | N(aa 44-180)-10aa-SmBiT | 14.2 |
| LC2 | N(aa 44-257)-10aa-LgBiT | 11.2 |
| SC2 | N(aa 44-257)-10aa-SmBiT | 1.8 |
| LN | LgBiT-10aa-N(aa 44-257) | 1.6 |
| SN | SmBiT-10aa-N(aa 44-257) | 6.8 |

**d** N sensors

LC+SC
LC2+SC2
LN+SN

RLU (Nluc)

Fold Dilution
(commercial anti-N antibody)

**e** N sensors (LN+SN)

Patient 6
Patient 8
Patient 11
Patient 12
Control 1
Control 2

RLU (Nluc)

Fold dilution

**Figure 2.6: Design and characterization of N sensors. a,** Annotated depiction of the SARS-CoV-2 Nucleocapsid protein (protein N). All N protein fusions designed included the RNA binding domain (aa 44-180, N-RBD) and excluded the dimerization domain (aa 257-419). **b,** Structure of the N-RBD domain shows the N and C termini locate far from each other and fusion of the split enzyme fragments to N or C termini may result in different detection sensitivity (PDB: 6YI3). **c,** Yield of the six N protein-NanoBiT sensor fusions. **d,** The N-terminal N sensor pair (LN + SN, 44-257) was less sensitive than the LC + SC (44-180) and LC2 + SC2 (44-257) C terminal N sensor pairs when the assay was performed on a rabbit polyclonal anti-N protein antibody (Sino Biological, Cat#: 40588-T62-50). Two technical replicates are plotted from n=1 individual experiment. Lines connecting the means of the samples are plotted. **e,** Additionally, only patient 6 and 8 showed signals above controls in the serological assay performed with LN + SN sensors, while all four patients showed signals with the LC + SC sensors. Each dot represents the average of two technical replicates for a single experiment. N=2 experiments are plotted. Lines connecting the means of the samples are plotted.

**Figure 2.7: Epitope characterizations of CR3022, C004, C105 and C135. a,** Design of a Biolayer interferometry (BLI) experiment to characterize competitive binding of the antibodies with ACE2-Fc and other antibodies. **b,** BLI experiments showed C004 and C105 both competed with ACE2-Fc for binding while C135 does not. **c,** BLI experiments showed C004 competed with C105 for binding while the other antibodies do not compete.
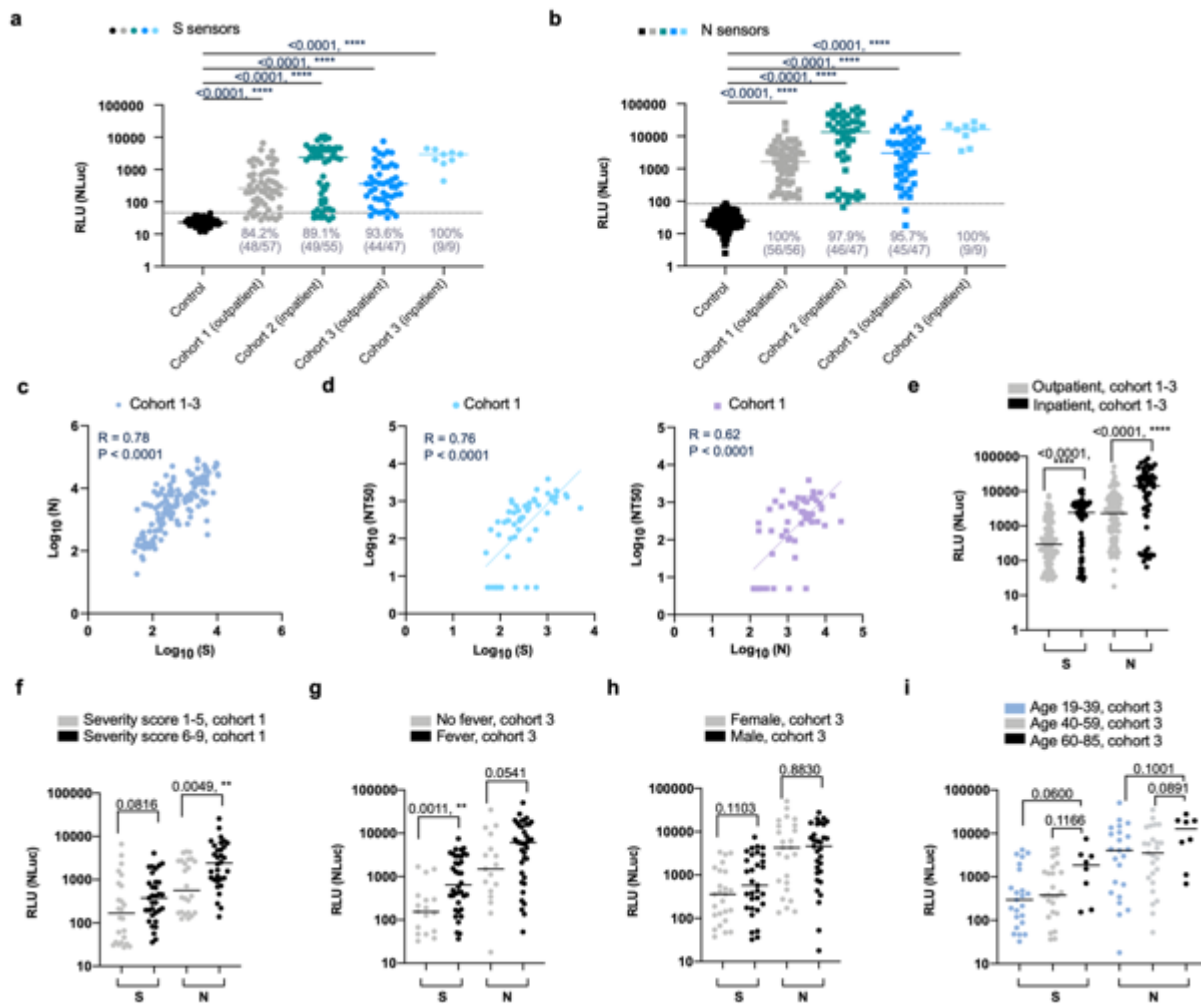
**Figure 2.8: Characterization of outpatient and inpatient serum samples using the spLUC test.** Cohort 1: samples drawn during the convalescent phase of an outpatient group, Cohort 2: samples drawn during the acute phase or the convalescent phase of a hospitalized group, and Cohort 3: samples drawn during the convalescent phase of a mixed inpatient and outpatient group. A 10-base logarithmic scale conversion was applied to all the solution assay signals for the correlation analysis unless otherwise specified. **a,** SpLUC assay tested on expanded COVID-19 patient cohorts with S sensors at 1:12.5 serum dilution. Dots represent the average between two technical duplicates. Lines represent median values. The inpatient samples showed significantly higher antibody titers than the outpatient cohorts. Samples sizes are as indicated in parentheses: Control(56), Cohort 1(57), Cohort 2(55), Cohort 3 outpatient(47), and Cohort 3 inpatient(9). **b,** SpLUC assay tested on expanded COVID-19 patient cohorts with N sensors at 1:12.5 serum dilution. The inpatient samples showed significantly higher antibody titers than the outpatient cohorts. Samples sizes are as indicated in parentheses: Control(120), Cohort 1(56), Cohort 2(47), Cohort 3 outpatient(47), and Cohort 3 inpatient(9). **c,** A positive correlation (R = 0.78) was observed between S sensor signal and N sensor signal in the three cohort samples. All cohorts individually presented a similar trend (**Fig. 2.9**). Line represents linear regression. 159 patient samples are plotted. **d,** Correlation of spLUC signals (cohort 1) to neutralization efficiency[6]. S sensor signal (blue) and N sensor signal (purple) is plotted against 50% maximal neutralization

titer (NT50). Both show positive correlation (R = 0.76 for S and NT50 and R = 0.62 for N and NT50). 57 patient samples are plotted for the S sensor and 56 samples for the N sensor. **e,** Inpatients show significantly higher signal over outpatients in all three cohorts (p < 0.0001). Samples sizes are as indicated in parentheses: S-outpatient(104), S-inpatient(64), N-outpatient(103), and N-inpatient(56). **f,** Patients from cohort 1 that reported higher disease severity (6-10 vs 1-5) had higher antibody titer for both S and N sensors and the difference for N sensors is statistically significant (p = 0.0049). Samples sizes are as indicated in parentheses: S-score 1-5(24), S-score 6-9(33), N-score 1-5 (24), and N-score 6-9(32). **g,** Higher overall antibodies titers were observed in patients that reported fever compared to no fever patients for cohort 3. This difference was statistically significant for the S sensors (p = 0.0011) but not N sensors. 17 patient samples associated with no fever and 38 patient samples associated with a fever are plotted. **h,** Slightly higher overall antibodies titers were observed in females compared to males for cohort 3, although the differences were not statistically significant. There is a similar trend for cohort 1 (**Fig. 2.11a**). The difference was more obvious for S sensors. 25 female patient samples and 30 male patient samples are plotted. **i,** For cohort 3, there is a slightly higher level of antibodies in the 60-85 age group compared to 19-39 and 40-59. There is a similar trend for cohort 1 (**Fig. 2.11b**). The differences were not statistically significant. Samples sizes are as indicated in parentheses: age 19-39 (23), age 40-59 (25), and age 60-85 (8). For **a** and **b**, P value (two-tail) of a Kruskal-Wallis test with Dunn's multiple comparison post hoc testing to compare each clinical cohort with the healthy control group is indicated. For **c-d**, R values and P values (two-tail) of a non-parametric Spearman correlation analysis are labeled in the graphs. For **e-i**, an unpaired Mann-Whitney test is performed and P values (two-tail) for each comparison are labeled on top of the datasets. For all panels, dots represent the average of two technical replicates from n=1 independent experiment. For **a, b, e-i**, horizonal lines represent median values. For **c-d**, lines represent linear regression.
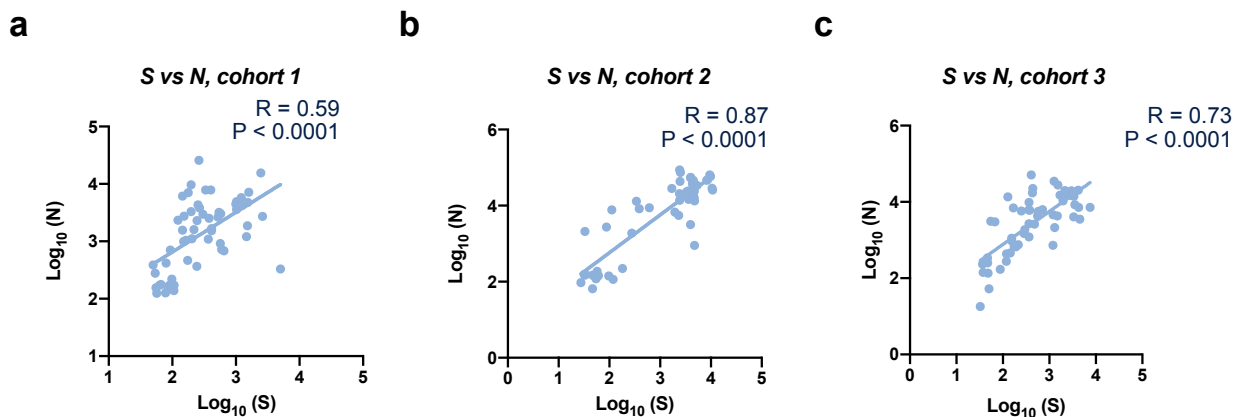


**Figure 2.9: Individual cohorts show good correlation between S and N sensors.** Each cohort shows robust correlation with R = 0.59, 0.87, and 0.73 for **a**, cohort 1 (56 samples), **b**, cohort 2 (47 samples), and **c**, cohort 3 (56 samples), respectively. R values and P values (two-tail) of a non-parametric Spearman correlation analysis are labeled in the graphs. Lines represent linear regression. For all graphs, dots represent the average of two technical replicates from n=1 independent experiment.
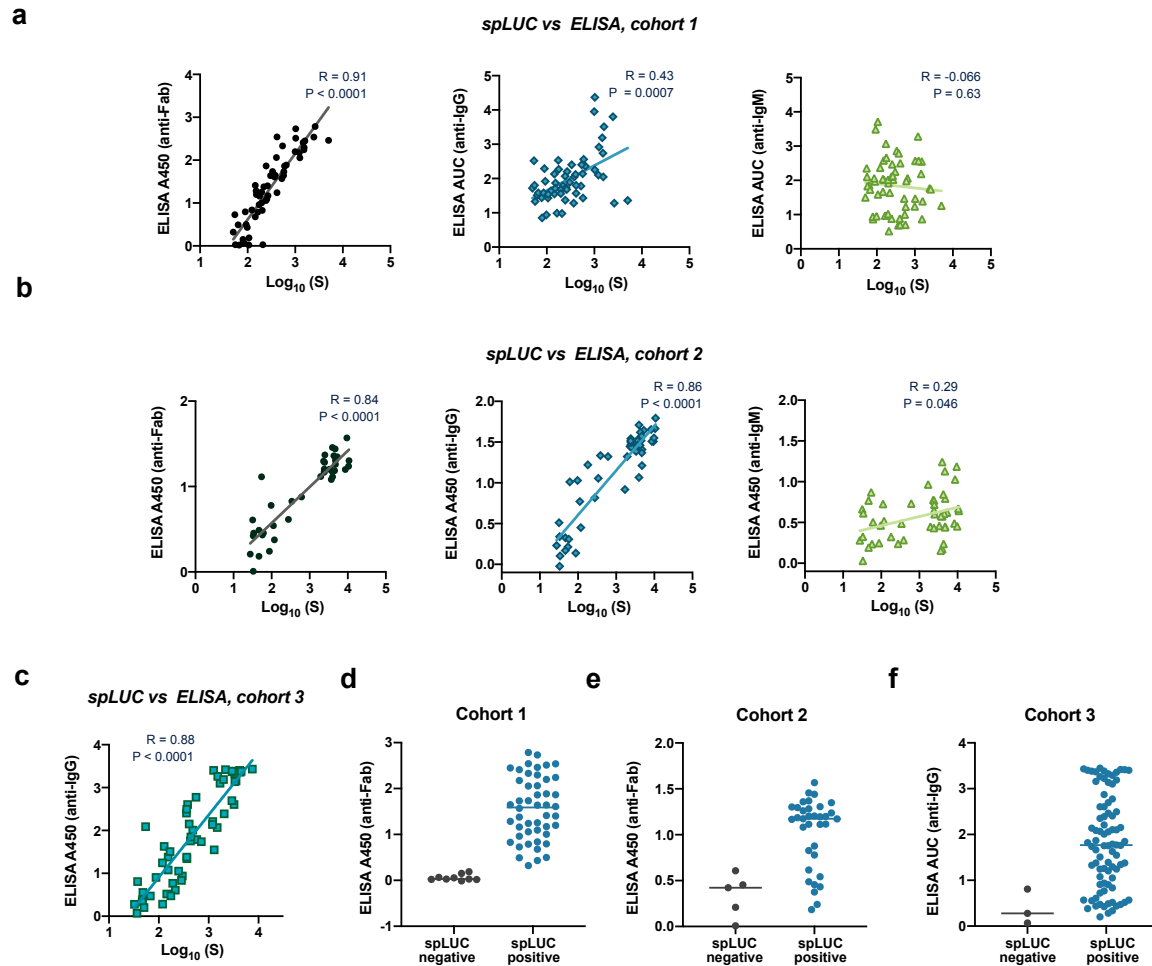
**Figure 2.10: Comparison of the ELISA and the spLUC results. a,** Signals from the S sensor spLUC assay (cohort 1, 57 samples) correlate very well with S-RBD ELISA anti-Fab signals (R = 0.91), moderately well with anti-IgG signals (R = 0.43), and poorly with anti-IgM signals for cohort 1 (R = -0.066). Line represents linear regression. **b,** Signals from the S sensor spLUC assay (cohort 2, 40 samples with anti-Fab detection, 47 samples with anti-IgM/IgG detection) correlate very well with S-RBD ELISA anti-Fab signals (R = 0.84) and with anti-IgG signals (R = 0.86), but poorly with anti-IgM signals for cohort 1 (R = 0.29). Line represents linear regression. **c,** Signals from the S sensor (cohort 3, 56 samples) correlate well with S-RBD ELISA anti-IgG signals (R = 0.88). For a-c, R values and P values (two-tail) of a non-parametric Spearman correlation analysis are labeled in the graphs. Lines represent linear regression. **d, e, f,** The seronegative samples in the anti-S spLUC assay also showed low anti-Fab or anti-IgG signals in ELISA serology tests for **d**, cohort 1 (negative samples (9), positive samples (48)), **e**, cohort 2, (negative samples (5), positive samples (35)), and **f,** cohort 3, (negative samples (3), positive samples (53)). Horizontal lines represent the median value. For all graphs
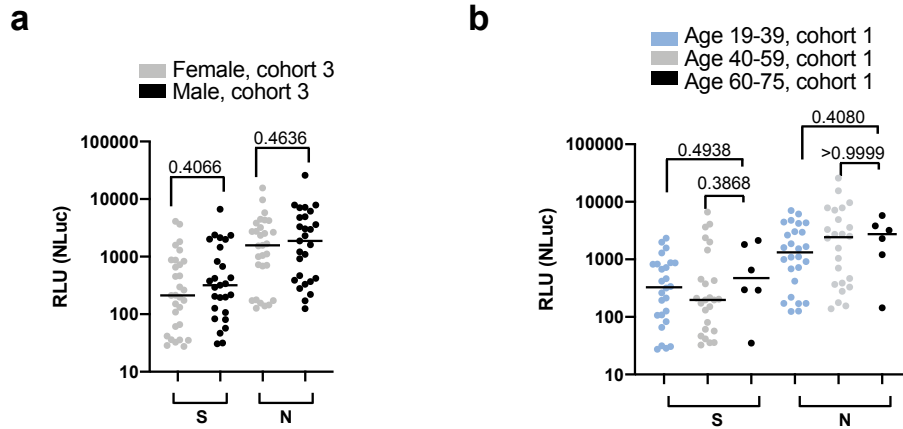
**Figure 2.11: Further correlation of spLUC signal and gender/age. a,** For cohort 3, males show a slightly higher spLUC assay signal compared to females, although this difference is not statistically significant. Samples sizes are as indicated in parentheses: S-female (30), S-male (27), N-female (29), and N-male (27). **b,** Cohort 1 spLUC signal shows no significant difference in signal among age groups. Samples sizes are as indicated in parentheses: age 19-39 (26), age 40-59 (24), and age 60-85 (6). For a and b, an unpaired Mann-Whitney test is performed and P values (two-tail) for each comparison are labeled on top of the datasets. Horizontal lines represent the median value. For both graphs, dots represent the average of two technical replicates from n=1 independent experiment.
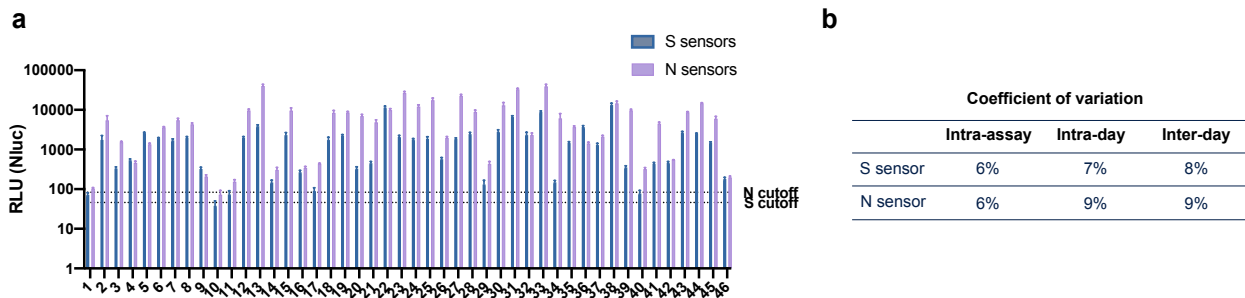


**Figure 2.12: Inter-assay, Inter-day and Intra-day variability of spLUC assay. a,** 46 plasma samples were assayed a total of five times in three independent experiments over two days for each sensor. All five replicates are plotted on the graph, with the bar representing the average. The dotted line represents the cutoff values for positive and negative samples for the S and N sensors. **b,** The coefficient of variation was calculated for intra-assay, intra-day, and inter-day variability. Coefficient of variation is calculated as ratio of standard deviation to the average value.
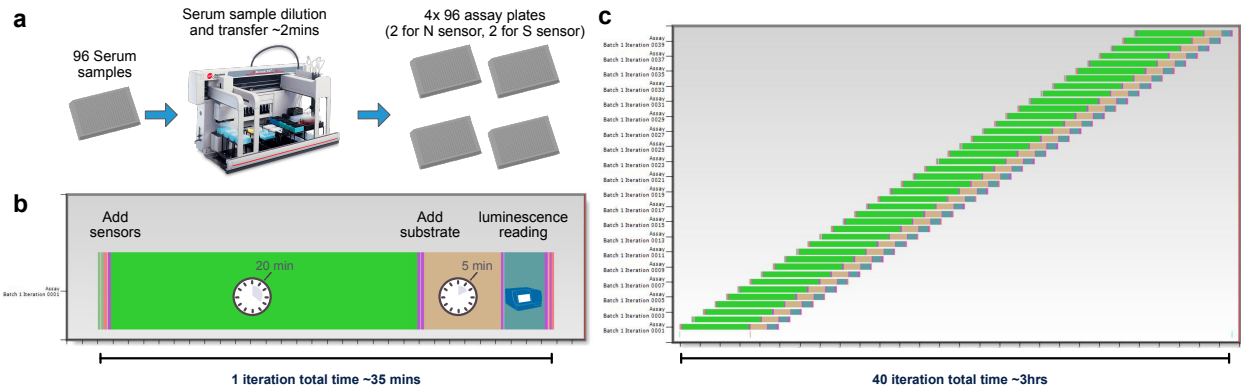
**Figure 2.13: Simulated robotics-assisted spLUC assay. a,** Serum sample transfer to an assay plate using Biomek Fx Automated Workstation in ~2 minutes. **b,** Robotics-assisted dispensing and luminescence reading for one iteration of 96 assays takes ~35 minutes. **c,** Simulated run for 40 iterations (3840 assays) can be completed in 3 hours. Gantt chart generated by simulated run using Thermo Momentum software.
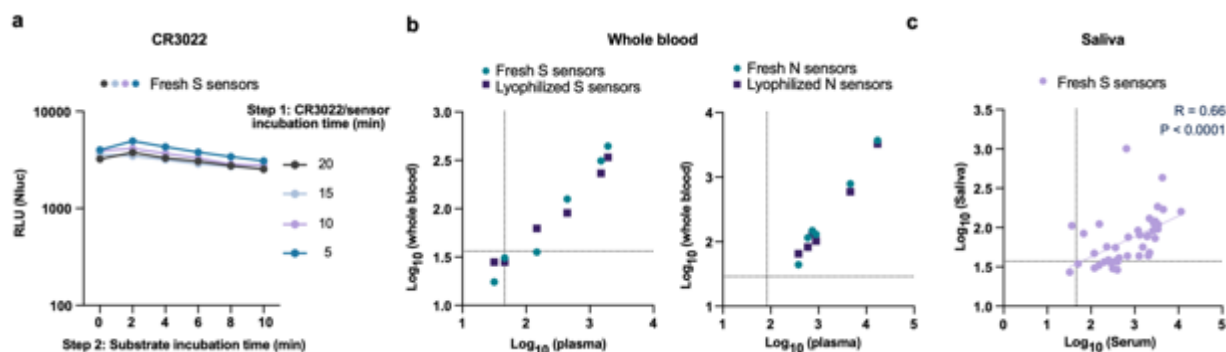
**Figure 2.14: Adapting the assay for whole blood and saliva sample types.** **a,** spLUC assays can be accomplished in as short as 5 minutes. CR3022 (10 nM) was incubated with S sensors for 5, 10, 15, or 20 min. Luciferase substrates were then added and incubated with the reaction mix for 0, 2, 4, 6, 8 or 10 min. All reactions showed bright luminescence signal. Each dot represents the average of two technical duplicates from n=1 independent experiment. Lines connecting the means of the samples are plotted. **b,** The spLUC assay is compatible with whole blood samples and shows similar signal in the corresponding plasma samples with both fresh and lyophilized sensors. Each dot represents the average of two technical replicates from n=1 independent experiment. A non-parametric spearman correlation analysis was performed and R = 0.94 was observed for S sensors, and R = 1 and 0.98 were observed for N sensor fresh and lyophilized sensors, respectively. **c,** Anti-S antibodies were detected in saliva samples with moderate sensitivity (33/42, 79%). The signals from saliva samples positively correlated with corresponding serum samples. Each dot represents the average of two technical replicates from n=1 independent experiment. A non-parametric spearman correlation analysis was performed and the R value (0.66) and P value (<0.0001, two-tail) are labeled in the graphs. Line represents linear regression.

**a**

**Effect of biosensor
lyophilization**

| | | % recovery | % signal |
|---|---|---|---|
| S sensor | SmBiT | 90 | 50 |
| | LgBiT | 100 | |
| N sensor | SmBiT | 70 | 100 |
| | LgBiT | 100 | |

**b**

S sensors,
Fresh substrate

**c**

N sensors,
Fresh substrate

**d**

Substrate,
Fresh S sensors
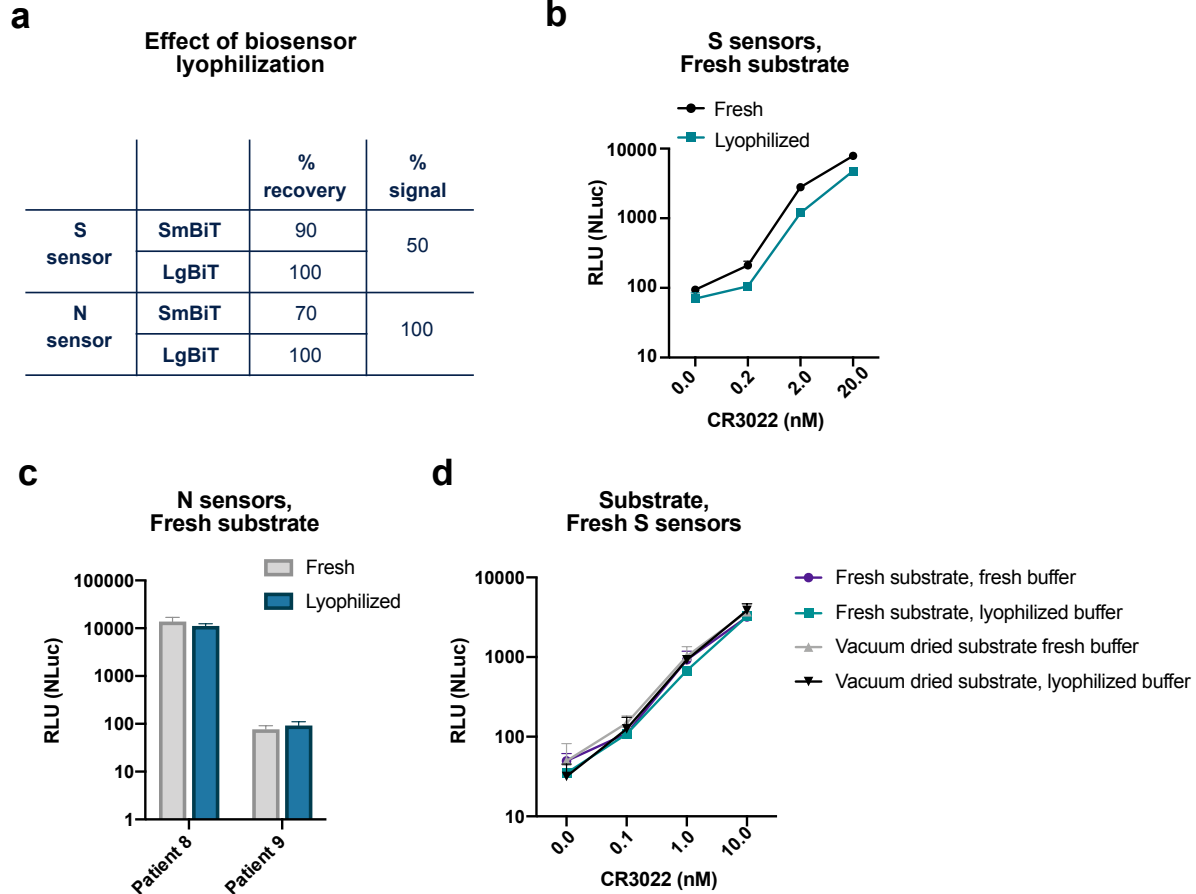
**Figure 2.15: S and N sensors are functional after lyophilization. a,** Both the S and the N sensors can survive lyophilization. The majority of proteins (70-100%) can be reconstituted after lyophilization. The lyophilized S sensors lost 50% of signal. The lyophilized N sensors remain 100% active. **b,** The lyophilized S sensors detected CR3022 at ~50% signal strength compared to fresh sensors. Three technical replicates are plotted from n=1 independent experiment. Lines connecting the means of the samples are plotted. **c,** The lyophilized N sensors detected antibodies from patient sera at similar signal strength compared to fresh sensors. Two technical replicates are plotted from n=1 independent experiment. The bars represent the mean. **d,** Vacuum dried substrate and substrate stored at -20°C (fresh) behave similarly when detecting recombinant CR3022 with S sensors. Lyophilized dilution buffer and dilution buffer stored at -20°C (fresh) also showed similar signal. Three technical replicates are plotted from n=1 independent experiment. Lines connecting the means of the samples are plotted.
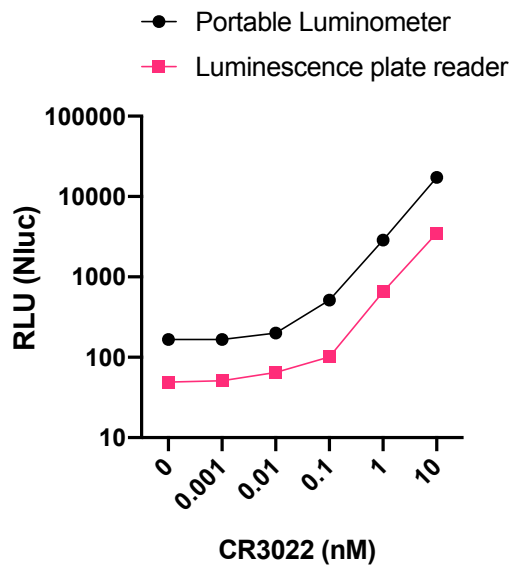
**Figure 2.16: Portable luminometer.** The spLUC assay is also amenable to detection with a Berthold portable luminometer. The handheld luminometer showed similar sensitivity of recombinant CR3022 with S sensors compared to the plate reader. Due to the tube format of the handheld luminometer, the sample volume was doubled and thus the overall signals are higher than for the plate reader samples, but similar sensitivity is maintained. Two technical replicates are plotted from n=1 independent experiment. Lines connecting the means of the samples are plotted.
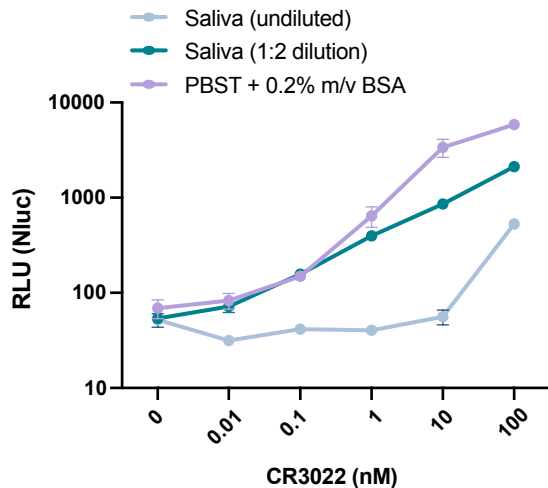
**Figure 2.17: Saliva condition optimization.**  spLUC reactions are compatible with saliva samples. The CR3022 antibody was spiked into healthy individual saliva at 10-fold dilutions from 100 nM to 0.01 nM. While undiluted saliva reduced signal 10-fold and reduced sensitivity, 1:2 dilution of saliva only reduced signal by 3-fold and did not decrease the sensitivity. Two technical replicates are plotted from n=1 independent experiment. Lines connecting the means of the samples are plotted.

**Table 2.1: Determination of assay cutoff values**

|  | S | N |
|---|---|---|
| **SERUM DILUTIONS** | 1:12.5 | 1:12.5 |
| **# SAMPLES** | 56 | 120 |
| **MIN** | 12 | 2.5 |
| **MAX** | 44.5 | 84 |
| **MEDIAN** | 23.2 | 25 |
| **MEAN** | 24.5 | 29.5 |
| **STANDARD DEVIATION (SD)** | 7.1 | 17.8 |
| **DERIVED CUTOFF (MEAN+3XSD)** | 45.9 | 83.1 |

**Table 2.2: A target profile proposed for SARS-CoV-2 serology tests in low resource areas**

| PROPERTY | OPTIMAL GOAL |
|---|---|
| **SENSITIVITY** | >95% |
| **SPECIFICITY** | >95% |
| **QUANTIFICATION** | Quantitative |
| **SAMPLE TYPE** | Whole blood, saliva |
| **SAMPLE PROCESSING** | Not required |
| **REAGENT FORMAT** | RT stable |
| **TIME TO RESULT** | < 30 min |
| **DAILY THROUGHPUT** | High |
| **POWER REQUIREMENT** | Battery |
| **CONSUMABLE COST** | < $10 |
| **DEVICE** | Portable |
| **MAINTENANCE** | None |
| **DATA ANALYSIS** | Integrated |

# Chapter 3

**Cell-surface tethered promiscuous biotinylators enable comparative small-scale surface proteomic analysis of human extracellular vesicles and cells**

**Abstract**

Characterization of cell surface proteome differences between cancer and healthy cells is a valuable approach for the identification of novel diagnostic and therapeutic targets. However, selective sampling of surface proteins for proteomics requires large samples (>10e6 cells) and long labeling times. These limitations preclude analysis of material-limited biological samples or the capture of rapid surface proteomic changes. Here, we present two labeling approaches to tether exogenous peroxidases (APEX2 and HRP) directly to cells, enabling rapid, small-scale cell surface biotinylation without the need to engineer cells. We used a novel lipidated DNA-tethered APEX2 (DNA-APEX2), which upon addition to cells promoted cell agnostic membrane-proximal labeling. Alternatively, we employed horseradish peroxidase (HRP) fused to the glycan binding domain of wheat germ agglutinin (WGA-HRP). This approach yielded a rapid and commercially inexpensive means to directly label cells containing common N-Acetylglucosamine (GlcNAc) and sialic acid glycans on their surface. The facile WGA-HRP method permitted high surface coverage of cellular samples and enabled the first comparative surface proteome characterization of cells and cell-derived small extracellular vesicles (EV), leading to the robust quantification of 953 cell and EV surface annotated proteins. We identified a newly-recognized subset of EV-enriched markers, as well as proteins that are uniquely upregulated on Myc oncogene-transformed prostate cancer EVs. These two cell-tethered enzyme surface biotinylation approaches are highly advantageous for rapidly and directly labeling surface proteins across a range of material-limited sample types.

**Introduction**

The cell surface proteome, termed the surfaceome, serves as the main communication hub between a cell and the extracellular environment[81]. As such, this cellular compartment often reveals the first signs of cellular distress and disease, and is of substantial interest to the medical community for diagnostic and therapeutic development[82]. The precise and comprehensive profiling of the surfaceome, termed surfaceomics, provides critical insights for our overall understanding of human health and can inform drug development efforts. Several strategies have emerged for either selective or comprehensive surfaceomics, including biocytin hydrazide labeling of surface glycoproteins[81], chemical biotinylation of lysines via NHS-ester labeling[83], and promiscuous biotinylator fusion proteins (APEX2, BioID, SPPLAT)[81,84,85]. Membrane protein enrichment is a necessary step in surfaceomics, due to the inherent low abundance of membrane proteins compared to cytosolic proteins, and their identification can be overwhelmed by cytosolic contaminants. While each of these strategies robustly label surface proteins, they: (1) require large sample inputs (biocytin hydrazide), (2) require production of genetically engineered cells (APEX2, BioID), (3) label only partner proteins by binding targeting antibodies fused to APEX2 or HRP (SPPLAT), (4) require extensive sample manipulation (biocytin hydrazide), or (4) exhibit increased nonspecific labeling (NHS-ester)[86–89]. Moreover, many of these methods are not able to capture short and transient changes that occur at the cell surface, such as binding, adhesion, assembly, and signaling[90]. These current methods complicate the direct characterization of small clinical samples such as extracellular vesicles in patient serum. As biological research increasingly depends on animal models and patient-derived samples, the requirement for simple and robust methods amenable to direct labelling of material-limited samples for proteomic analysis will become paramount.

Exosomes and other small extracellular vesicles (EVs) are produced by both healthy and diseased cells[91]. In cancer, these small EVs contribute to tumor growth and metastasis, modulate the immune response, and mediate treatment resistance[92–95]. Consequently, these extracellular vesicles are a focus of intense clinical investigation. Recent studies suggest that small EVs incorporate proteins and RNA from the parent tumor from which they originate[96,97], and certain proteins may be preferentially shuttled into EVs[98]. There is also strong evidence that cancer-derived EVs are unique from the EVs derived from healthy surrounding tissues, and therefore represent a promising target for non-invasive, early-detection diagnostics or EV-focused therapies[94,99,100]. However, strategies for the unbiased profiling of small EV membrane proteomes remain limited. Isolation of high-quality, enriched small EV populations is challenging, requiring numerous centrifugation steps and a final sucrose gradient isolation, precluding the use of current labeling methods for membrane proteome characterization[95,98]. Strategies to characterize the surface proteome of small EVs would propel biomarker discovery and enable the differential characterization of small EVs from that of the parent cell. These important studies could help illuminate mechanisms underlying preferential protein shuttling to different extracellular vesicle populations.

Here, we functionalize the promiscuous biotinylators, APEX2 and HRP, as non-cellularly encoded exogenous membrane tethering reagents for small-scale surfaceomics, requiring <5e5 cells. This method is 10-100 fold more rapid than other existing protocols and requires fewer wash steps with less sample loss. Likewise, due to its selectivity towards tyrosines, it is not hindered by variability in individual protein glycosylation status[82] or by impeding complete tryptic peptide cleavage through modification of lysines[101], like biocytin hydrazide or NHS-biotin methods, respectively. Using this robust new strategy, we performed surfaceomics on cells and

corresponding small EVs from a cellular model of prostate cancer using the prostate epithelial cell line, RWPE-1 with or without oncogenic Myc induction. While certain proteins show increased expression in both parental cell and EV surfaces, a subset of proteins were found to be either pan-EV markers (ITIH4, MFGE8, TF, DSG1, TSPAN14, AZGP1, and IGSF8) or selectively enriched with Myc overexpression in cancer-derived EVs (ANPEP, SLC38A5, FN1, SFRP1, CDH13, THBS1, and CD44). These differentially-regulated proteins pose interesting questions related to preferential protein shuttling, and the proteins upregulated in both cellular and EV contexts reveal candidates for early-stage urine or serum-based detection without invasive surgical intervention. We believe these simple, rapid, and direct surfaceomic labeling tools may be broadly applied to small-scale surfaceomics on primary tissues.

**Results**

*Generation of promiscuous cell-surface tethered peroxidases for exogenous addition to cells*

Both APEX2 and HRP are broadly used promiscuous proximity biotinylators that label nearby tyrosine residues in proteins through a radical intermediate mechanism using a biotin-tyramide reagent **(Fig. 3.1a)**[102,103]. HRP has been targeted to specific cell-surface proteins through antibody conjugation to label target proteins and their binding partners[84]. More recently, HRP was used as a soluble cell surface labeler to identify rapid cell surface proteome changes in response to insulin[104]. Genetically encoded, membrane-targeted APEX2 and HRP have also permitted promiscuous labeling of proteins in specific cellular compartments, but these efforts required cellular engineering[102,105]. We sought to expand the use of these tools to biotinylate surface proteins of cells without the need for cellular engineering, enabling the specific enrichment of surface-resident proteins for mass spectrometry analysis.

The first approach we tested was to tether a DNA-APEX2 conjugate to the cell membrane through a lipidated DNA anchor. Gartner and co-workers have shown lipidated DNA anchors can tether together molecules or even cells[106,107]. Here the lipidated DNA is first added to cells, then hybridized with a complimentary strand of DNA conjugated to APEX2 **(Fig. 3.1b, left panel)**. To conjugate DNA to APEX2, we leveraged the single unpaired cysteine in the protein for site-specific bioconjugation of the complementary DNA. We first reacted APEX2 (**Fig. 3.2**) with DBCO-maleimide, after which the DBCO moiety was readily conjugated with azido-DNA. The kinetics of coupling was monitored using LC-MS and the conjugate was purified by nickel column chromatography, yielding a single conjugated product **(Fig. 3.3a)** that retained full enzymatic function relative to unlabeled APEX2 **(Fig. 3.3b)**. Microscopy was used to observe the colocalization of DNA-conjugated APEX2 to the membrane **(Fig. 3.1c)**. This result was recapitulated using flow cytometry, indicating that this approach results in surface tethering of APEX2, an important step towards the specific labeling of the cell surfaceome **(Fig. 3.3c).**

To avoid the need for bioconjugation, we also tested a commercially available reagent where the promiscuous biotinylator HRP is conjugated to the lectin wheat germ-agglutinin (WGA) **(Fig. 3.1b, right panel)**. WGA-HRP is used regularly in the glycobiology and neuroscience fields to label cell membranes for immuno-histochemistry and live cell imaging[108,109]. This is an inexpensive and widely available tool that only requires the presence of surface protein N-acetylglucosamine (GlcNAc) and sialic acid glycans to localize HRP to the membrane. The successful and rapid colocalization of WGA-HRP to the plasma membrane compared to HRP alone was verified using immunocytochemistry, indicating this approach is a potential alternative for cell surface labeling **(Fig. 3.1d)**. Further testing showed that adding WGA-HRP to cells in the

presence hydrogen peroxide and biotin-tyramide led to robust surface labeling even with no pre-incubation time **(Fig. 3.4)**.

*Cell-tethered biotinylators more effectively label the surfaceome than non-tethered biotinylators and are comparable to biocytin hydrazide*

Next, we set out to optimize labeling conditions for small-scale sample characterization. As APEX2 is kinetically slower than HRP[110], we used APEX2 to establish a suitable concentration range of enzyme for cell surface labeling. We found that 0.5 μM APEX2 produced maximal labeling of cells **(Fig. 3.5a)** and maintained equivalent labeling across a range of cell numbers (2.5e5 – 1e6 cells; **Fig. 3.5b**). Next, we compared the efficiency of DNA-APEX2, WGA-HRP, and their non-tethered counterparts to biotinylate a small sample of 5e5 Expi293 cells. We found a 5- to 10-fold increase in biotin labeling for both tethered DNA-APEX2 and WGA-HRP relative to non-tethered controls as assessed by flow cytometry **(Fig. 3.6a)** and western blotting **(Fig. 3.6b)**. Moreover, tethered DNA-APEX2 and WGA-HRP systems exhibited similar biotinylation efficiency, suggesting either system is suitable for small-scale surfaceomics. Having both systems is useful, as some cells may not widely express glycoproteins recognized by commercially available lectin-HRP conjugates—such as some prokaryotic species—and therefore could require the glycan-agnostic DNA-tethered APEX2 construct[111].

To compare the degree of surface protein enrichment these two systems offer, we enriched biotinylated proteins generated with either approach and compared the resulting enrichments using LC-MS/MS. As an initial efficacy comparison, cell surface labeling with DNA-labeled APEX2 or WGA-HRP was compared using 5e5 cells. In order to eliminate the possibility of suspension cell-specific results, we used a popular cell line model of pancreatic cancer, KP-4. We observed that the WGA-HRP identified slightly more plasma membrane annotated proteins in the Uniprot Gene

Ontology Cellular Component Plasma Membrane (GOCC-PM) database (>2 unique peptides, found in all replicates) relative to DNA-APEX2, totaling 501 and 467, respectively. Notably, the number of IDs for both cell-tethered enzymes was higher than their untethered counterparts, with HRP identifying 389 cell surface proteins and APEX2 identifying 247 **(Fig. 3.6c)**. Importantly, in the upset plot shown, the group with the highest intersection includes all four enzyme contexts, showcasing the reproducibility of labeling through a similar free-radical based mechanism. The cell-tethered biotinylators also showed heightened surface enrichment compared to their untethered counterparts, as illustrated by the higher overall intensities for proteins annotated to the plasma membrane **(Fig. 3.7a,b)**. As equal amounts of total protein is injected on the LC-MS/MS instrument, the higher intensities for plama membrane proteins suggest that localizing the enzyme to the membrane increases labeling of the membrane compartment, which we have previously observed with other enzymatic reactions[112].

As the mode of tethering WGA-HRP involves GlcNAc and sialic acid glycans, we wanted to determine whether there was a bias towards Uniprot annotated 'Glycoprotein' vs 'Non-Glycoprotein' surface proteins identified across the WGA-HRP, APEX2-DNA, APEX2, and HRP labeling methods. We looked specifically at surface annotated proteins found in the SURFY database, which is the most stringent surface protein database and requires that proteins have a predicted transmembrane domain[113]. We performed this analysis by measuring the average MS1 intensity across the top three peptides (LFQ area) for SURFY glycoproteins and non-glycoproteins for each sample and dividing that by the total LFQ area found across all GOCC-PM annotated proteins detected in each sample. We found similar normalized areas of non-glycosylated surface proteins across all samples **(Fig. 3.8)**. If a bias existed towards glycosylated proteins in the WGA-HRP compared to the glycan agnostic APEX2-DNA sample, then we would have seen a larger

percentage of non-glycosylated surface proteins identified in APEX2-DNA over WGA-HRP. Due to the large labeling radius of the HRP enzyme, we find it unsurprising that the WGA-HRP method is able to capture non-glycosylated proteins on the surface to the same degree[84]. There is a slight increase in the area percentage of glycoproteins detected in the WGA-HRP compared to the APEX2-DNA sample, but this is likely due to the fact that a greater number of surface proteins in general are detected with WGA-HRP. As HRP is known to have faster kinetics compared to APEX2, it was anticipated that WGA-HRP would outperform DNA-APEX2 in cell surface protein identifications. The heightened labeling of WGA-HRP was consistent with every cell line tested, including another pancreatic cancer model, PaTu8902, which resulted in more proteins identified in the WGA-HRP sample over DNA-APEX2 for both 1 and 2 minute time points **(Fig. 3.9)**.

To confirm that the improved labeling by WGA-HRP was due to the binding of sugar units on the cell surface, we performed a sugar-blocking experiment with WGA-HRP using N-acetyl-D-glucosamine (GlcNAc) that would block the conjugate from binding to the cell. By pre-incubating WGA-HRP with excess GlcNAc, the ability of WGA-HRP to label the cell surface was markedly lower than WGA-HRP without GlcNAc as observed by microscopy **(Fig. 3.6d)**. A similar effect was also seen by flow cytometry **(Fig. 3.10)**. In addition, we also tested an on-plate protocol for simpler cell surface labeling of adherent KP-4 cells. We showed that cell surface labeling in this manner was comparable to labeling when the cells were in suspension **(Fig. 3.11)**.

As WGA-HRP consistently outperformed DNA-APEX2 by proteomics and represents a more facile method amenable to broad application in the field, we chose to compare the proteomic labeling results of WGA-HRP to other standard cell surface labeling methods (sulfo-NHS-LC-LC-biotin and biocytin hydrazide) on a prostate epithelial cell line, RWPE-1 with and without oncogenic c-Myc overexpression. Sulfo-NHS-LC-LC-biotin reacts with primary amines to form

amide conjugates, but has notoriously high background contamination with intracellular proteins[114]. Biocytin hydrazide labeling is a two-step process that first involves oxidizing vicinal diols on glycoproteins at the cell surface, then reacting the aldehyde byproducts with biocytin hydrazide[87]. Both WGA-HRP and biocytin hydrazide had similar levels of cell surface enrichment on the peptide and protein level when cross-referenced with the SURFY curated database for extracellular surface proteins with a predicted transmembrane domain **(Fig. 3.12a)**. Sulfo-NHS-LC-LC-biotin and whole cell lysis returned the lowest percentage of cell surface enrichment, suggesting a larger portion of the total sulfo-NHS-LC-LC-biotin protein identifications were of intracellular origin, despite the use of the cell-impermeable format. These same enrichment levels were seen when the datasets were searched with the curated GOCC-PM database, as well as Uniprot's entire human proteome database **(Fig. 3.12b)**. Of the proteins quantified across all four conditions, biocytin hydrazide and WGA-HRP returned higher overall intensity values for SURFY-specified proteins than either sulfo-NHS-LC-LC-biotin or whole cell lysis. Importantly, although biocytin hydrazide shows slightly higher cell surface enrichment compared to WGA-HRP, we were unable to perform the comparative analysis at 500,000 cells--instead requiring 1.5 million--as the protocol yielded too few cells for analysis.  All three methods were highly reproducible across replicates **(Fig. 3.13a-c)**. Compared to existing methods, WGA-HRP not only labels cells efficiently with much lower input material requirements, it is also able to enrich for cell surface proteins to a similar extent in a fraction of the time.

*WGA-HRP identifies surface markers of Myc-driven prostate cancer in both cells and small EVs*

Prostate cancer remains one of the most common epithelial cancers in the elderly male population, especially in Western nations[115,116]. While metastatic progression of prostate cancer has been linked to many somatic mutations and epigenetic alterations (PTEN, p53, Myc etc.), more recent

work determined that alterations in Myc occurs in some of the earliest phases of disease, i.e. in tumor-initiating cells[117]. This finding promotes the idea that the development of early-stage diagnostic tools that measure these Myc-driven disease manifestations could improve detection and overall patient disease outcomes[117,118]. One mode of early detection that has gained prominence is the use of prostate cancer-derived exosomes in patient serum and urine[119,120]. Small EVs are known to play important roles in the progression of prostate cancer, including increasing tumor progression, angiogenesis, metastasis, and immune evasion, making this subcellular particle an extremely informative prognostic tool for disease progression[121–123].

To elucidate promising targets in Myc induced prostate cancer, we utilized our WGA-HRP method to biotinylate cells from both normal epithelial prostate cells (RWPE-1 Control) and oncogenic Myc-induced prostate cancer cells (RWPE-1 Myc, **Fig. 3.14a**). Importantly, by using an isogenic system, we are able to delineate specific Myc-driven protein expression changes, which could be helpful in the identification of non-invasive, early-detection diagnostics for cancer driven by early Myc induction. In addition to having marked overexpression of c-Myc in the RWPE-1 Myc cells compared to Control, they also grow with a more mesenchymal and elongated morphology compared to their Control cell counterparts **(Fig. 3.14b)**, which would suggest large cell surface changes upon oncogenic Myc induction. We initially used WGA-HRP to quantitatively compare the cell surface profiles of Myc-induced prostate cancer to Control cells and found large and bidirectional variations in their surfaceomes **(Fig. 3.14c,d)**. We have highlighted the 15 most upregulated proteins in each cell type that are annotated as extracellular surface proteins in the GOCC-PM database. Proteins that are also found in the most restrictive SURFY database that requires a predicted transmembrane domain are bolded in the Fig.. Proteins annotated to be secreted (Uniprot) from the cell are italicized[113]. Vimentin, a marker known to be

associated with epithelial-to-mesenchymal transition (EMT) showed heightened expression, in the context of oncogenic Myc, as well as CDH2 (N-Cadherin), another marker of EMT **(Fig. 3.14e)**[124,125]. While vimentin has traditionally been described as an intracellular protein, an extracellular membrane-bound form has been found to be important in the context of cancer[126,127]. ANPEP and fibronectin-1 were also highly upregulated.  Notably, a number of HLA molecules were downregulated in the Myc induced RWPE cells, consistent with prior findings of loss of MHC class I presentation in prostate cancer[128–130]. A subset of these findings were verified by both western blot **(Fig. 3.14f)** and microscopy **(Fig. 3.14g)**, which highlights the robustness of the protein quantification afforded by using this method.

Next, we wanted to use our WGA-HRP method to quantify cell surface proteins on a sucrose-gradient purified population of small EVs derived from both normal epithelial prostate cells (RWPE-1 Control) and oncogenic Myc-induced prostate cancer cells (RWPE-1 Myc, **Fig. 3.15a**). While sucrose gradient centrifugation generally yields a mixture of vesicle populations, we wanted to confirm that our preparation enriched for vesicles originating from multiple vesicular bodies (MVBs), consistent with an exosome-enriched sample[131]. To do so, we prepared EVs from both Control and Myc cells and carried out label-free quantification (LFQ) mass spectrometry on the whole EV lysates. After normalizing for cell number, we found the Myc cells produced nearly 40% more EVs than the corresponding control cells, which is consistent with previous work that has shown Myc overexpression yields higher quantities of EVs[132]. After averaging the intensities between Control and Myc derived EVs, many of the highest intensity proteins (CD9, SDCB1, CD81, LAMP1, LAMP2, ALIX, and CD63) are consistent with MVB-derived vesicle biogenesis, supporting that the sample was likely enriched in EVs rather than other sedimentable particles that can co-isolate during centrifugation **(Fig. 3.15b)**. Due to the complex process and extensive

washing involved in small EV isolation, many standard labeling methods are not amenable for EV surface labeling. Using WGA-HRP, we are able to biotinylate the small EVs before the sucrose gradient purification and isolation steps **(Fig. 3.15c)**. This delineated an important subset of proteins that are differentially expressed under Myc induction, which could serve as interesting targets for early-detection in patient urine or serum **(Fig. 3.15d)**. This subset included fibronectin-1 (FN1) and ANPEP **(Fig. 3.15e)**, which were further validated by quantitative western blotting **(Fig. 3.15f)**. A subset of these targets display similar phenotypic changes to the parent cell, suggesting that they could be biomarker candidates for non-invasive indicators of disease progression. While certain proteins are shuttled to vesicle compartments largely based off of the extent of expression in the parent cell (Control: IFITM3, BST2, HLA-B, Myc: ANPEP, SLC38A5, FN1), remarkably some proteins are singled out for small EV packaging, indicating a pronounced differential shuttling mechanism of the proteome between cells and EVs. This pattern was recapitulated in both the RWPE-1 Control cells and corresponding EVs, as well as the Myc cells and EVs, where the majority of markers were unique to either cellular or EV origin **(Fig. 3.16a,b)**. These protein targets are of extreme interest for not only biomarker discovery, but also understanding the role of small EVs in secondary disease roles, such as interfering with immune function or priming the metastatic niche[133].

Due to the difficulty of proteomic characterization of vesicular populations, our current understanding of EV protein shuttling remains limited. Prior proteomic EV analysis has involved whole EV preparations, which lacks a surface protein enrichment step[134–136]. Not only is whole EV lysate analysis less advantageous for the specific identification of cell surface proteins on EVs, but it makes it impossible to compare cellular and EV samples due to the inherent surface area-to-volume differences between cells and the vesicles they produce[137,138]. Our WGA-HRP method

allows us to compare surface proteins between small EV populations, as well as between small EV and cell samples **(Fig. 3.17a)**. By principle component analysis (PCA), each sample separates by oncogenic status and origin **(Fig. 3.17b)**. Indeed, when performing functional annotation for each gene cluster defined by the PCA, 'extracellular exosome' and 'extracellular vesicle' are the highest ranking annotation features differentiating the EVs from their parent cells **(Fig. 3.17c)**. Through this comparison, we were able to delineate a host of proteins that were upregulated in EVs over their parent cells and vice versa **(Fig. 3.17d)**. Notably, secreted proteins were more highly represented in the EV surface proteome compared to cells. A subset of proteins were highly upregulated in the small EVs compared to parent cell, including ITIH4, MFGE8, TF, DSG1, TSPAN14, AZGP1, and IGSF8, **(Fig. 3.17e)** and a subset of the findings were validated by western blot **(Fig. 3.17f)**. The samples showed good overlap between replicates across all four datasets, with cellular and EV samples clustering by origin and oncogenic status **(Fig. 3.18)**. To our knowledge, this is the first experiment to wholistically characterize the surface proteome of both small EVs and parental cells. These data strongly suggest that protein triage into EVs is a controlled process, enabling only a subset of the cell surface proteome to be shuttled to this important compartment. Our data shows that there are a variety of pan-prostate-EV markers, notably lactadherin (MFGE8), serotransferrin (TF), inter-alpha-trypsin inhibitor (ITIH4), immunoglobulin superfamily 8 (IGSF8), desmoglein-1 (DSG1), tetraspanin-14 (TSPAN14), and zinc-alpha-2-glycoprotein (AZGP1) **(Fig. 3.17d)**, which do not seem to be Myc-specific. Some of the pan-prostate EV targets in our data have previously been linked to cancer-specific contexts, and we show here that they are also found on Control EVs[139–141]. Our work suggests that these markers are more broadly associated with small EVs, regardless of disease status, outlining an expanded set of targets to probe these vital compartments.

**Discussion**

The importance of understanding and characterizing cellular and EV membrane compartments is vital for improving our understanding of vesicle biogenesis. New, improved methodologies amenable to small-scale and rapid surface proteome characterization are essential for continued development in the areas of therapeutics, diagnostics, and basic research. We sought to develop a simple, rapid surface protein labeling approach that was compatible with small sample sizes, while remaining specific to the cell surface. We took advantage of fast peroxidase enzymes and either complementary lipidated DNA technology (DNA-APEX2) or the glycan binding moiety wheat germ agglutinin (WGA-HRP) and demonstrated that tethering was much more effective than soluble addition, with increases in protein identification of between 30-90%. Additionally, we compared WGA-HRP to the existing methods, sulfo-NHS-LC-LC-biotin and biocytin hydrazide. While these alternative methods are robust, they are unable to capture time-sensitive changes, and are either plagued by low selectivity/specificity (NHS-Biotin) or the requirement for large sample inputs (biocytin hydrazide).

There are many advantages of our new methods over the current cell surface labeling technologies. Compared to both sulfo-NHS-LC-LC-biotin and biocytin hydrazide, WGA-HRP experiments require 2 minutes instead of 30 or 120 minutes, respectively. It is also able to enrich cell surface proteins much more efficiently than sulfo-NHS-LC-LC-biotin labeling. Furthermore, NHS peptide isolation and preparation is complicated due to the reactivity of NHS chemistry towards free-amines, which blocks tryptic and LysC cleavages typically used in proteomics[101,142].

The hydrazide method is highly effective for enriching cell surface proteins, but it is challenging for small sample sizes, due to the two-step labeling process and cell loss from the oxidation step and extensive washing. Additionally, neither NHS-biotin nor biocytin hydrazide are

able to capture short time points to encompass dynamic changes at the cell surface. Due to the fast kinetics of peroxidase enzymes (1-2 min), our approaches could enable kinetic experiments to capture rapid post-translational trafficking of surfaces proteins, such as response to insulin, certain drug treatments, T-cell activation and synapse formation, and GPCR activation[104,143,144]. Another disadvantage of the hydrazide method is that it can only enrich for proteins that are glycosylated at the cell surface and it is estimated that 10-15% of cell surface proteins are not glycosylated[145]. Glycosylation patterns also readily change during tumorigenesis, which can alter the quantification of glycan-based labeling methods, such as biocytin hydrazide[146]. While the WGA-HRP method requires glycosylated proteins to be present to bind, it is still able to label non-glycosylated proteins nearby due to its large labeling radius. It is a possibility that certain cells may have low or uneven levels of glycosylation on their surfaces. In these cases, the DNA-APEX2 method can be utilized to obtain effective labeling. However, both these peroxidase-based methods require the presence of tyrosine residues (natural abundance 3.3%) to react with the biotin-tyramide radical, which is not equally abundant in all proteins[3].

With the WGA-HRP method, we were able to compare the surfaceome of small EVs to parental cells for Myc-induced prostate cancer cells and identified proteins that were upregulated in Myc-induced cells and EVs, as well as proteins that were differentially shuttled between EVs and parental cells. We found a number of Myc specific markers in our study, which were enriched in both Myc EV and Myc Cell samples. These include ANPEP, SLC38A5, FN1, CDH13, VIM, and CA12. ANPEP is a membrane-bound ectopeptidase that degrades N-termini with neutral amino acids and was found 140-fold upregulated in the Myc-induced cell compared to the Control cell and 49-fold upregulated in the Myc-induced EV compared to Control EV. This peptidase has been associated with angiogenesis and cancer growth[147,148]. Recent studies have shown ANPEP is

systematically upregulated on isogenic cell lines expressing[149,150] proliferative oncogenes or in tubular sclerosis bladder cancers[151], suggesting it is commonly upregulated in cancers. The second most differentially expressed protein between the Myc and Control samples was SLC38A5 (23 and 73-fold upregulated in cells and EVs, respectively). SLC38A5 is a glutamine co-transporter and has previously been shown to be a downstream target of c-Myc in glutamine-addicted cancers. Moreover, given that SLC38A5-based glutamine transport leads to proton flux and intracellular alkanization, overexpression of SLC38A5 has also been hypothesized to be a strategy for pH regulation in cancer cells that regularly experience intracellular acidification due to high glycolytic flux[152,153]. Additionally, Fibronectin-1 (FN1) was also found to be upregulated in Myc samples over Control samples (5 and 63-fold upregulated in cells and EVs, respectively), and has been shown to drive all stages of tumorigenesis[154]. Importantly, FN1 provides an extracellular scaffold by which other matrix proteins can be deposited. Through these interactions with matrix proteins and cell-associated integrins, FN1 regulates cellular fate decisions, proliferation, and metastasis[155].

While some proteins were present in both the EV and cellular samples, others were only found enriched in Myc EVs. THBS1, also known as thrombospondin-1 was over 10-fold upregulated in Myc EVs over Control EVs. Interestingly, this relationship was not found in the parent cells, which suggests that THBS1 is differentially shuttled into oncogenic EVs. The role of this protein has newly been associated with the growth and metastasis of glioblastoma and a potential serum prognostic factor in myeloid leukemia[156,157]. Moreover, using a model of THBS1 overexpressing breast cancer, recent work has shown that exosomes laden with THBS1 promote cancer cell migration via disruption of the endothelial barrier[158].

Another such target is CD44, which was over 8-fold upregulated in the Myc EVs over Control EVs. CD44 has long been known to drive cancer progression and aberrant cell signaling[159].

Recently, CD44 has also been found to be preferentially loaded into cancer-derived exosomes and has been implicated in driving chemoresistance in a model of doxorubicin-treated breast cancer[157]. Similarly, it has been shown that exosome-mediated transfer of CD44 from cells with high metastatic potential promoted migratory behavior in neighboring cells with low metastatic potential[160]. These targets delineate an important subset of proteins that are triaged into EVs and could play long-range roles in promoting tumorigenesis and downstream metastasis[96,133,161,162].

As research shifts into analyzing native biological samples from extracellular vesicles to xenograft models or patient biopsies, it will become increasingly important to develop sensitive, effective methods to label these small samples sizes. It is our hope that these tools will provide much needed avenues by which to pursue pressing biological questions in the areas of diagnostic and therapeutic development, as well as basic research.

**Methods**

***Large-Scale APEX2 Expression, Purification, and Heme Reconstitution***

APEX2 was expressed using previous methods in BL21(DE3)pLysS cells[163]. Briefly, APEX2 expression plasmid was transfected into competent BL21(DE3)pLysS cells and heat shocked for 45 seconds before being placed on ice. Cells were plated on LB/Carb plates and grown overnight at 37°C. A single colony was isolated and grown in a mixture of 30 ml of 2XYT + Carb overnight at 37°C while shaking. The overnight culture was combined with 3 L of 2XYT with Carb and placed in a 37°C shaking incubator. At an OD600 of 0.6, 100 µg/ml of IPTG was added and the temperature of the incubator was lowered to 30°C. Cells were allowed to incubate for 3.5 hours and were spun down at 6,000xg for 20 minutes. Cell pellet was resuspended in protease inhibitor containing resuspension buffer (5 mM Imidazole, 300 mM NaCl, 20 mM Tris pH=8) and mixed thoroughly. The mixture was sonicated at 50% (5 seconds on:15 seconds off) for 5 minutes on ice

to avoid bubble formation. Lysate was mixed by inversion at 4°C for 15 minutes and spun down at 19,000xg for 20 minutes. The slurry was introduced to 5 ml of washed Nickel resin slurry and allowed to bind by gravity filtration. The beads were washed 3x with wash buffer (30 mM Imidazole, 300 mM NaCl, 20 mM Tris pH=8) and eluted in 5 ml of elution buffer (250 mM Imidazole, 300 mM NaCl, 20 mM Tris pH=8) before undergoing buffer exchange into PBS.

Enzyme underwent heme reconstitution as per previous methods[164]. Briefly, 50 mg of hemin-Cl (Sigma) was diluted in 2.0 mL of 10 mM NaOH. The mixture was thoroughly resuspended, then diluted further using 8.0 mL of 20 mM $KPO_4$, pH 7.0, and vortexed extensively. Mixture was spun down at 4,000xg 2x to eliminate insoluble hemin. APEX2 was diluted 1:2 in 20 mM $KPO_4$. 6 ml of heme stock was added to 2 ml of APEX over 20 minutes and allowed to rotate at 4°C wrapped in tin foil for 3 hours. The mixture was introduced to a column with 20 ml of DEAE Sepharose pre-equilibrated in 20 mM $KPO_4$, pH 7.0 buffer. Enzyme was eluted using 100 mM $KPO_4$ and spin concentrated. To verify complete reconstitution, absorbance was measured at 403 and 280 nm. A403/280 > 2.0 is considered sufficient for reconstitution. The isolated protein was flash frozen and stored at -80°C for long-term storage. Each batch of enzyme was run out on a 4-12% Bis-Tris gel to confirm purity **(Fig. 3.2)**.

### *APEX2 DNA labeling protocol*

APEX2 was incubated at 50 µM with 40 molar equivalents of maleimide-DBCO for 5 hours at room temperature in PBS. The reaction was desalted with Zeba columns (7 kDa cutoff). 2.5 molar equivalents of Azido-DNA was added to the reaction and incubated at 4°C overnight. Successful conjugation was monitored by LC-MS before the mixture was purified by nickel column.

*Cell culture*

Prior to all experiments, cells were tested for the presence of mycoplasma (MycoAlert PLUS, Lonza, LT07-703). Expi293 suspension cells were maintained in Expi293 media (Thermo, A1435101) while rotating at 125 rpm in a 37°C incubator with 8% $CO_2$. Cells were split every 3 days by diluting into new media. Adherent PaTu8902 and KP-4 cells were grown in pre-warmed Iscove's Modified Dulbecco's Media (IMDM) supplemented with 10% FBS (Gemini Bio-Products, 100-106) and 5% Penicillin/Streptomycin (Thermo Fisher Scientific, 15-140-122) at 37°C in a 5% $CO_2$-humidified incubator. Adherent RWPE-1 prostate cells were grown in complete keratinocyte-SFM (Thermo; 17005-042) supplemented with bovine pituitary extract (BPE), recombinant EGF, and 5% penicillin/streptomycin at 37°C in a 5% $CO_2$-humidified incubator. The media was exchanged every two days. For splitting, cells were lifted with 0.05% Trypsin (Life Technologies) and quenched with 5% FBS before spinning down cells to remove residual trypsin and FBS. Cells were then plated in pre-warmed complete keratinocyte-SFM media.

*Microscopy*

Cells were plated at a density of 15,000 cells per well in a 96-well clear bottom plate (Greiner Bio-One, 655090) pre-treated with poly-D-lysine (Thermo Scientific, A3890401). Cells were allowed 48 hours to reattach and grow undisturbed. Cells were washed 3x in cold PBS. For DNA-APEX2, 100 µl of 0.5 µM enzyme solution was combined with anchor and co-anchor at a final concentration of 1 µm. For all other enzymes, enzyme was combined with PBS at a final concentration of 0.5 µM. For sugar blocking studies, 100 µl of diluted enzyme solution (0.5 µM) was combined with 100 mg/ml N-acetyl-D-glucosamine (Sigma Aldrich, A3286-5G). Cells were allowed to sit on ice for 5 minutes to allow WGA to bind fully, as labeling was not altered by

increased incubation time (**Fig. 3.4**). Biotin tyramide (Sigma Aldrich, SML2135-50MG) was added to cells with a final concentration of 500 µM before adding 1 mM of $H_2O_2$. Reaction was allowed to continue for 2 minutes before rinsing cells 3x with 1X quench buffer (10 mM sodium ascorbate + 5 mM Trolox + 1 mM sodium pyruvate). The cells were rinsed 2x with PBS and crosslinked with 4% PFA for 10 minutes at RT. Cells were washed 3x with PBS before introduction to 1:100 primary antibody. Primary antibodies used were: HisTag-650 (Invitrogen, MA1-21315-D650), Streptavidin-488 (Thermo Fisher Scientific, S-11223), biotin-conjugated anti-HRP (Rockland, 200-4638-0100), ANPEP (R&D Systems, AF3815), vimentin (Cell Signaling Technology, 5741S), and HLA-B (ProteinTech, 17260-1-AP). Cells were washed 3x in PBS and imaged on an IN Cell Analyzer 6500. Images were processed in Fiji using the Bio Formats plugin[165,166].

*Cell-tethered APEX2, soluble APEX2, cell-tethered WGA-HRP and soluble HRP cell surface labeling*

Cultured cells were grown for 3 days in tissue culture plates and dissociated by addition of versene (PBS + 0.05% EDTA). Cells were washed 3x in PBS (pH 6.5), resuspended in PBS (pH 6.5) and aliquoted to 500,000 cells per sample. Samples were resuspended in 100 µL of PBS (pH 6.5). For anchored APEX2 samples, lipidated anchor DNA was allowed to bind for 5 minutes at 1 µM on ice, followed by 1 µM of lipidated co-anchor DNA on ice for 5 minutes. 0.5 µM DNA-labeled APEX2 was allowed to bind on cells for 5 minutes before final wash with PBS (pH 6.5). For soluble APEX2, WGA-HRP, and soluble HRP samples, cells were resuspended in 0.5 µM of the corresponding enzyme. WGA-HRP was allowed to bind to cells for 5 minutes on ice. Biotin tyramide was added at a final concentration of 500 µM and mixed thoroughly, before the addition

of 1 mM $H_2O_2$. Cells underwent labeling in a heated shaker (500 rpm) at 37°C for 2 minutes before being quenched with 5 mM Trolox/10 mM Sodium Ascorbate/1 mM Sodium Pyruvate. Cells were washed 2x in quench buffer and spun down. The pellet was either further processed for flow cytometry, western blot, or flash frozen in liquid nitrogen for mass spectrometry.

### *On plate WGA-HRP cell surface labeling*

KP-4 cells were grown on a 6 cm tissue culture treated plate and washed 3x with PBS (pH 6.5). 2 mL of 0.5 µM WGA-HRP in PBS (pH 6.5) was added to the plate, followed by biotin tyramide (0.5mM final concentration) and $H_2O_2$ (1mM final concentration). After a 2 minute incubation at 37°C, the cells were washed 2x with 5 mM Trolox/10 mM Sodium Ascorbate/1 mM Sodium Pyruvate quenching solution. The cells were washed 1x with PBS before being lifted with versene (PBS + 0.05% EDTA). Once lifted, the cells were washed once with PBS and subsequentially processed for flow cytometry analysis.

### *Biocytin hydrazide cell surface labeling*

Cultured cells were grown for 3 days in tissue culture plates and dissociated by addition of versene (PBS + 0.05% EDTA). Cells were washed 3x in PBS (pH 6.5), resuspended in PBS (pH 6.5) and aliquoted to 1.5 million cells per sample. Samples were resuspended in 100 µL of PBS (pH 6.5) and fresh sodium periodate (Sigma Aldrich, 311448, 1 µL of a 160 mM solution) was added to each sample. The samples were mixed, covered in foil, and incubated while rotating at 4°C for 20 minutes. Following three washes with PBS (pH 6.5), the samples were resuspended in 100 µL of PBS (pH 6.5) with the addition of 1 µL of aniline (Sigma Aldrich, 242284, diluted 1:10 in water) and 1 µL of 100 mM biocytin hydrazide (Biotium, 90060). The reaction proceeded while rotating at 4°C for 90 minutes. The samples were then washed 2x with PBS (pH 6.5) and spun down. The

pellet was either further processed for flow cytometry, western blot, or flash frozen in liquid nitrogen for mass spectrometry.

### *Sulfo-NHS-LC-LC-biotin cell surface labeling*

Cultured cells were grown for 3 days in tissue culture plates and dissociated by addition of versene (PBS + 0.05% EDTA). Cells were washed 3x in PBS (pH 7.4), resuspended in PBS (pH 8) and aliquoted to 1.5 million cells per sample. Samples were resuspended in 50 µL of PBS (pH 8). An aliquot of EZ-Link Sulfo-NHS-LC-LC-Biotin (ThermoFisher, 21338) was resuspended in 150 µL of PBS (pH 8). 7.5 µL was added to each cell sample and the reaction proceeded while rotating at 4°C for 30 minutes. The reaction was quenched by the addition of 2.5 µL of 1M Tris (pH 8.0). The samples were washed 2x in PBS (pH 8.0) and spun down. The pellet was either further processed for flow cytometry, western blot, or flash frozen in liquid nitrogen for mass spectrometry.

### *Flow cytometry for cell surface biotinylation*

After labeling and quench washes, cells were washed once with PBS + 2% BSA to inhibit nonspecific binding. Samples were then incubated with 100 µL Streptavidin-Alexa Fluor 647 (Thermo Fischer, 1:100 in PBS + 2% BSA). Following a 30-minute incubation at 4°C while rocking, samples were washed three times with PBS + 2% BSA. Samples were analyzed in the APC channel and quantified using a CytoFLEX (Beckman Coulter). All flow cytometry data analysis was performed using FlowJo software.

### *RWPE-1 small EV isolation and labeling protocol*

RWPE-1 Control and Myc cells were plated at 7 million and 4 million cells per plate, respectively, across 16 x 15 cm$^2$ plates and allowed to grow in normal keratinocyte-SFM media with provided supplements. Small EVs were isolated as previously described[98]. Briefly, two days prior to EV

isolation, media was replaced with 15 milliliters BPE-free keratinocyte-SFM media. For vesicle enrichment, media was isolated after two days in BPE-free media and centrifuged at 300 x g for 10 minutes at RT, followed by 2,000 x g for 20 minutes at 4°C. Large debris was cleared by a 12,000 x g spin for 40 minutes at 4°C. The pre-cleared supernatant was spun a final time at 100,000 x g at 4°C for 1 hr to pellet extracellular vesicles. Isolated extracellular vesicles were brought up in 50 µl of PBS with 0.5 µM of WGA-HRP and the mixture was allowed to bind on ice for 5 minutes. WGA-HRP bound vesicles were placed on a shaker (500 rpm) at 37°C before the addition of biotin tyramide (0.5 mM final concentration) and $H_2O_2$ (1 mM final concentration). Vesicles underwent labeling for 2 minutes before being quenched with 5 mM Trolox/10 mM Sodium Ascorbate/1 mM Sodium Pyruvate. Biotinylated small EVs were purified from other sedimentable particles by further centrifugation on a sucrose gradient (20-60%) for 16 hours at 4°C at 100,000xg. Precisely, the gradient was loaded using 0%, 20%, 40%, and 60% sucrose fractions from top to bottom. The sample was loaded at the bottom in 60% sucrose and the purified small EVs were isolated in the 20-40% sucrose fractions. Differential sucrose centrifugation yielded between 3-5 µg of small EVs.

***Western blot protocol***

Cultured cells were grown in 15 cm$^2$ tissue culture plates and dissociated by addition of versene (PBS + 0.05% EDTA). Cells were washed in PBS (pH 6.5) and resuspended in 100 µl PBS (pH 6.5) at a concentration of 10 million cells/ml in PBS (pH 6.5). Cells were labeled, reaction was quenched with 1X NuPage Loading Buffer, and immediately boiled for 5 minutes. To enable proper addition of lysate to gel wells, the mixture was thinned with addition of nuclease, and the disulfides were reduced with BME. The samples were subjected to electrophoresis in a 4-12% NuPage Gel until the dye front reached the bottom of the gel cast. For cell and EV blots, equal

amounts of protein content quantified by BCA assay were prepared in 1X NuPage Loading Buffer with BME and boiled for 5 minutes. Samples were loaded and subjected to electrophoresis in a 4-12% NuPage Gel until the dye front reached the bottom of the gel cast. Prepared gels were placed in iBlot2 transfer stacks and transferred using the P0 setting on the iBlot 2 Gel Transfer Device. The PVDF membrane was blocked in TBS Odyssey Blocking buffer for 1 hour at RT. Membranes were washed in TBST and incubated with Strepavidin-800 (1:10,000 dilution, Licor, 926-32230) for 30 minutes in TBS Odyssey Blocking buffer + 0.1% Tween 20. Membranes were washed in TBST 3x with a final wash in water. Membranes were visualized using an Odyssey DLx imager. Western blot samples were run and quantified 2-3 times and a representative image was displayed in figures.

For cell and EV blots, equal amounts of protein content quantified by BCA assay was prepared in 1X NuPage Loading Buffer with BME and boiled for 5 minutes. Samples were loaded and subjected to electrophoresis in a 4-12% NuPage Gel until the dye front reached the bottom of the gel cast. Prepared gels were placed in iBlot2 transfer stacks and transferred using the P0 setting on the iBlot 2 Gel Transfer Device. The PVDF membrane was blocked in TBS Odyssey Blocking buffer for 1 hour at RT. Membranes were washed in TBST and incubated overnight in primary antibody at 4°C in TBS Odyssey Blocking buffer + 0.1% Tween 20 while shaking. Primary antibodies used were ANPEP (R&D Systems, AF3815), FN1 (Abcam, ab2413), vimentin (Cell Signaling Technology, 5741S), ITIH4 (Atlas Antibodies, HPA003948), MFGE8 (Thermo Scientific, PA5-82036), IGSF8 (R&D Systems, AF3117-SP). Membranes were washed in 3x TBST before introduction to a 1:10,000 dilution of secondary antibody in TBS Odyssey Blocking buffer + 0.1% Tween 20 for 1 hour at room temperature while shaking. Secondary antibodies used were Goat Anti-Rabbit HRP (Thermo Scientific, 31460) and Rabbit Anti-Sheep HRP (Thermo

Scientific, 31480). Blots were imaged after 5 minutes in the presence of SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, 34577) and imaged using a ChemiDoc XRS+. Western blot samples were run and quantified 2-3 times and a representative image was displayed in figures. EV blot was run once due to limited sample availability. EVs derived for western blotting were cultured and harvested independently of either biological replicate used for mass spectrometry analysis.

### *Proteomic preparation for whole EVs*

Whole EV pellets were previously flash frozen after collection. EVs were processed for LC-MS/MS using a PreOmics iST kit (P.O.00027). Briefly, EV pellets were brought up in 50 µl of provided LYSE solution and boiled with agitation for 10 minutes. The provided enzymes mixture (Trypsin and LysC) were resuspended in 210 µl of RESUSPEND buffer, mixed, and added to the lysed EVs. Samples were allowed to mix at 500 rpm for 1.5 hours at 37°C, before being quenched with 100 µl of STOP solution. Sample was spun in provided C18 spin cartridge and washed 1X with 200µl of WASH 1 and WASH 2. Peptides were eluted with 2X 100 µl of ELUTE, dried, and resuspended with the provided LC-LOAD solution. Peptides were quantified using Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, 23275).

### *Proteomic preparation for surface enriched samples*

Frozen cell and EV pellets were lysed using 2X RIPA buffer (VWR) with protease inhibitor cocktail (Sigma-Aldrich; St. Louis, MO) at 4°C for 30 mins. Cell lysate was then sonicated, clarified, and incubated with 100 µl of neutravidin agarose slurry (Thermo Fisher Scientific, 29204) at 4°C for 1 hr. The bound neutravidin beads were washed in 2 ml Bio-spin column (Bio-Rad, 732-6008) with 5 ml RIPA buffer, 5 ml high salt buffer (1M NaCl, PBS pH 7.5), and 5 ml

urea buffer (2M urea, 50mM ammonium bicarbonate) to remove non-specific proteins. Beads were allowed to fully drain before transferring to a Low-bind Eppendorf Tube (022431081) with 2M Urea. Sample was spun down at 1,000xg and aspirated to remove excess liquid. Samples were brought up in 100 µl of 4M Urea digestion buffer (50 mM Tris pH 8.5, 10 mM TCEP, 20 mM IAA, 4 M Urea) and allowed to reduce and alkylate for 10 mins at 55°C while shaking. After the addition of 2 µg of total reconstituted Trypsin/LysC, the sample was incubated for 2 hours at RT. To activate the trypsin, mixture was diluted with 200 µl of 50 mM Tris pH 8.5 to a final Urea concentration of below 1.5 M. The mixture was covered and allowed to incubate overnight at RT. The mixture was isolated from the beads by centrifugation in a collection column (Pierce; 69725) before being acidified with 10% TFA until pH of 2.0 was reached. During this time, a Pierce C18 spin column (Pierce, 89873) was prepared as per manufacturing instructions. Briefly, C18 resin was washed twice with 200 µl of 50% LC-MS/MS grade ACN. The column was equilibrated with two 200µl washes of 5% ACN/0.5% TFA. The pre-acidified sample was loaded into the C18 column and allowed to fully elute before washing twice with 200 µl washes of 5% ACN/0.5% TFA. One final wash of 200 µl 5% ACN/1% FA was done to remove any residual TFA from the elution. Samples were eluted in 70% ACN, dried, and dissolved in 0.1% formic acid, 2% acetonitrile prior to LC-MS/MS analysis. Peptides were quantified using Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, 23275).

### LC-MS/MS

Liquid chromatography and mass spectrometry was performed as previously described[167]. Briefly, approximately 200 ng of peptides were separate using a nanoElute UHPLC system (Bruker) with a pre-packed 25 cm x 75 µm Aurora Series UHPLC column + CaptiveSpray insert (CSI) column (120 A pore size, IonOpticks, AUR2-25075C18A-CSI) and analyzed on a timsTOF Pro (Bruker)

mass spectrometer. Peptides were separated using a linear gradient of 2-34% solvent B (Solvent A: 2% acetonitrile, 0.1% formic acid, solvent B: acetonitrile, 0.1% formic acid) over 100 mins at 400 nL/min. Data-dependent acquisition was performed with parallel accumulation-serial fragmentation (PASEF) and trapped ion mobility spectrometry (TIMS) enabled with 10 PASEF scans per topN acquisition cycle. The TIMS analyzer was operated at a fixed duty cycle close to 100% using equal accumulation and ramp times of 100 ms each. Singly charged precursors were excluded by their position in the m/z–ion mobility plane, and precursors that reached a target value of 20,000 arbitrary units were dynamically excluded for 0.4 min. The quadrupole isolation width was set to 2 m/z for m/z < 700 and to 3 m/z for m/z > 700 and a mass scan range of 100-1700 m/z. TIMS elution voltages were calibrated linearly to obtain the reduced ion mobility coefficients (1/K0) using three Agilent ESI-L Tuning Mix ions (m/z 622, 922 and 1,222).

### *Data Processing and Analysis*

Briefly, for general database searching, peptides for each individual dataset were searched using PEAKS Online X version 1.5 against both the plasma membrane (PM) annotated human proteome (Swiss-prot GOCC database, August 3, 2017 release) and the entire Swiss-prot Human Proteome (Swiss-prot). We acknowledge the identification of a number of proteins not traditionally annotated to the plasma membrane, which were published in the final Swiss-prot GOCC-PM database used. Additionally, to not miss any key surface markers such as secreted proteins or anchored proteins without a transmembrane domain, we chose to initially avoid searching with a more stringent protein list, such as the curated SURFY database. However, following the analysis we bolded proteins found in the SURFY database and italicized proteins known to be secreted (Uniprot). Enzyme specificity was set to trypsin + LysC with up to two missed cleavages. Cysteine carbamidomethylation was set as the only fixed modification; acetylation (N-term) and methionine

oxidation were set as variable modifications. The precursor mass error tolerance was set to 20 PPM and the fragment mass error tolerance was set to 0.05 Da. Data was filtered at 1% for both protein and peptide FDR and triaged by removing proteins with fewer than 2 unique peptides. All mass spectrometry database searching was based off of two biological replicates. Biological replicates underwent washing, labeling, and downstream LC-MS/MS preparation separately.

For comparative label-free quantification of cellular and EV samples, datasets were searched using PEAKS Online X version 1.5 against the plasma membrane (PM) annotated human proteome (Swiss-prot GOCC database, August 3, 2017 release). Enzyme specificity was set to trypsin + LysC with up to two missed cleavages. Cysteine carbamidomethylation was set as the only fixed modification; acetylation (N-term) and methionine oxidation were set as variable modifications. The precursor mass error tolerance was set to 20 PPM and the fragment mass error tolerance was set to 0.05 Da. Data was filtered at 1% for both protein and peptide FDR and triaged by removing proteins with fewer than 2 unique peptides. Label free quantification of protein was completed by taking the average intensity of the top three most intense peptides for each protein. Data was normalized by total area sum intensity for each sample. Using Perseus, all peak areas were $\log2(x)$ transformed and missing values were imputed separately for each sample using the standard settings (width of 0.3, downshift of 1.8). Significance was based off of a standard unpaired Student t test with unequal variances across all four replicates. Reported peak area values represent the averages of all four replicates--two biological and two technical replicates. For representation of the data in figures, a Z-score was computed and is defined as (LFQ Area - Mean LFQ Area)/Standard Deviation. Protein IDs that were not annotated to be secreted or expressed extracellularly were removed. Further, in the Cell versus EV graph, any proteins that showed a standard deviation (SD) greater than 1.5 between Control and Myc of each sample type (EV or

Cell) were removed to avoid representation of oncogene-specific changes. EVs and cells from different biological replicates were cultured on different days. Desalting, quantification, and LC-MS/MS runs were performed together. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE  partner repository with the dataset identifier PXD028523.

To compare replicates of data in **(Fig. 3.13)**, a simple linear regression was performed on total area sum intensity normalized data. Replicate 1 was graphed against Replicate 2 for biocytin hydrazide, NHS-biotin, and WGA-HRP, and the resulting data was shown with calculated R and p-values as determined using the simple linear regression software suite in Prism. For the supplementary heatmap output, total area normalized LFQ data found in **(Fig. 3.17)** was loaded into Morpheus (software.broadinstitute.org/Morpheus) and data points were clustered by the Pearson correlation between all replicates on both columns and rows. This same data was used in Perseus to produce the principle component analysis (PCA). Distinct gene clusters were further analyzed by functional annotation using the DAVID Bioinformatics Resource 6.8.
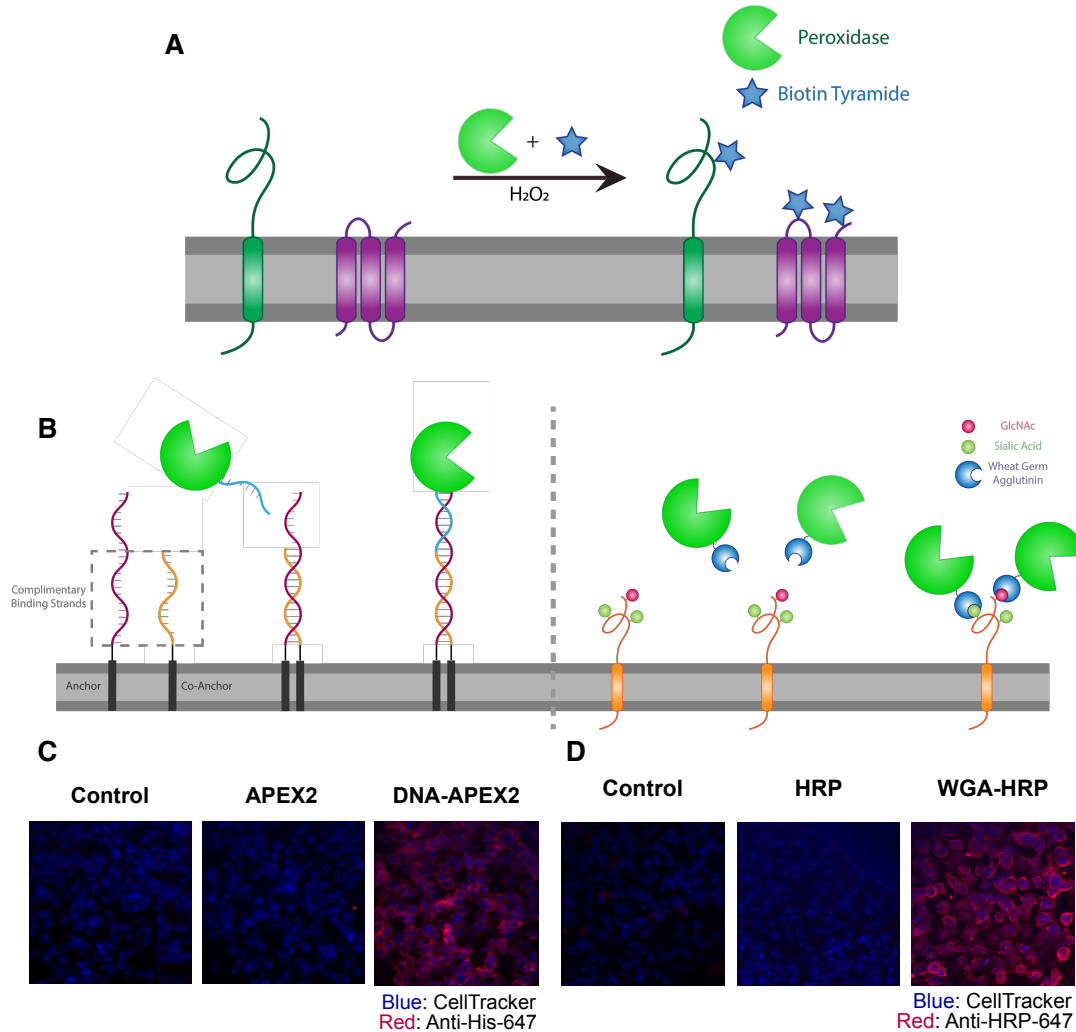
**Figure 3.1: Direct labeling of promiscuous biotinylators to the cell membrane for rapid cell surface proteome characterization of small-scale biological samples.** (A) Outline of enzymatic reaction mechanism. APEX2 and HRP both require biotin tyramide and hydrogen peroxide to produce the biotin-radical intermediate. (B) Tethering either enzyme is completed through differing mechanisms: (i) APEX2 is tethered through bio-conjugation of a single-strand of DNA, which is complementary to an exogenously added sequence of lipidated-DNA attached to the membrane, (ii) Commercially available wheat germ agglutinin (WGA)-HRP associates with native GlcNAc and sialic acid glycan moieties on cell surface proteins. (C) Microscopy images depicting the localization of DNA-APEX2 to the cell surface of KP-4 cells after introduction of the lipidated-DNA complementary strands. (D) Microscopy images depicting the localization of WGA-HRP to the membrane of KP-4 cells. All microscopy images are representative of two biological replicates.
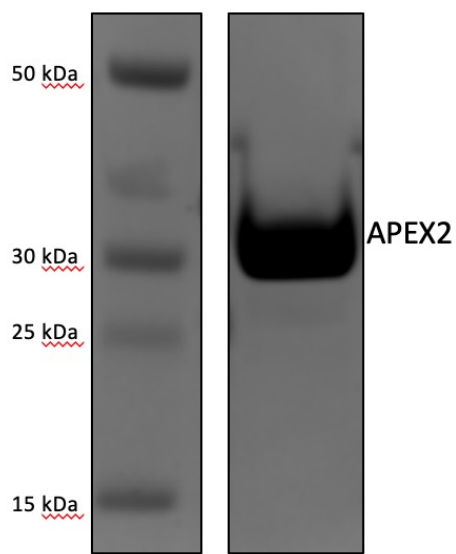
**Figure 3.2: Expression, purification, and validation of APEX2 enzyme.** His-tagged APEX2 was expressed in BL21(DE3)pLysS cells and purified by a nickel column. 10 µg of purified enzyme was run out on a 4-12% Bis-Tris gel to confirm purity.
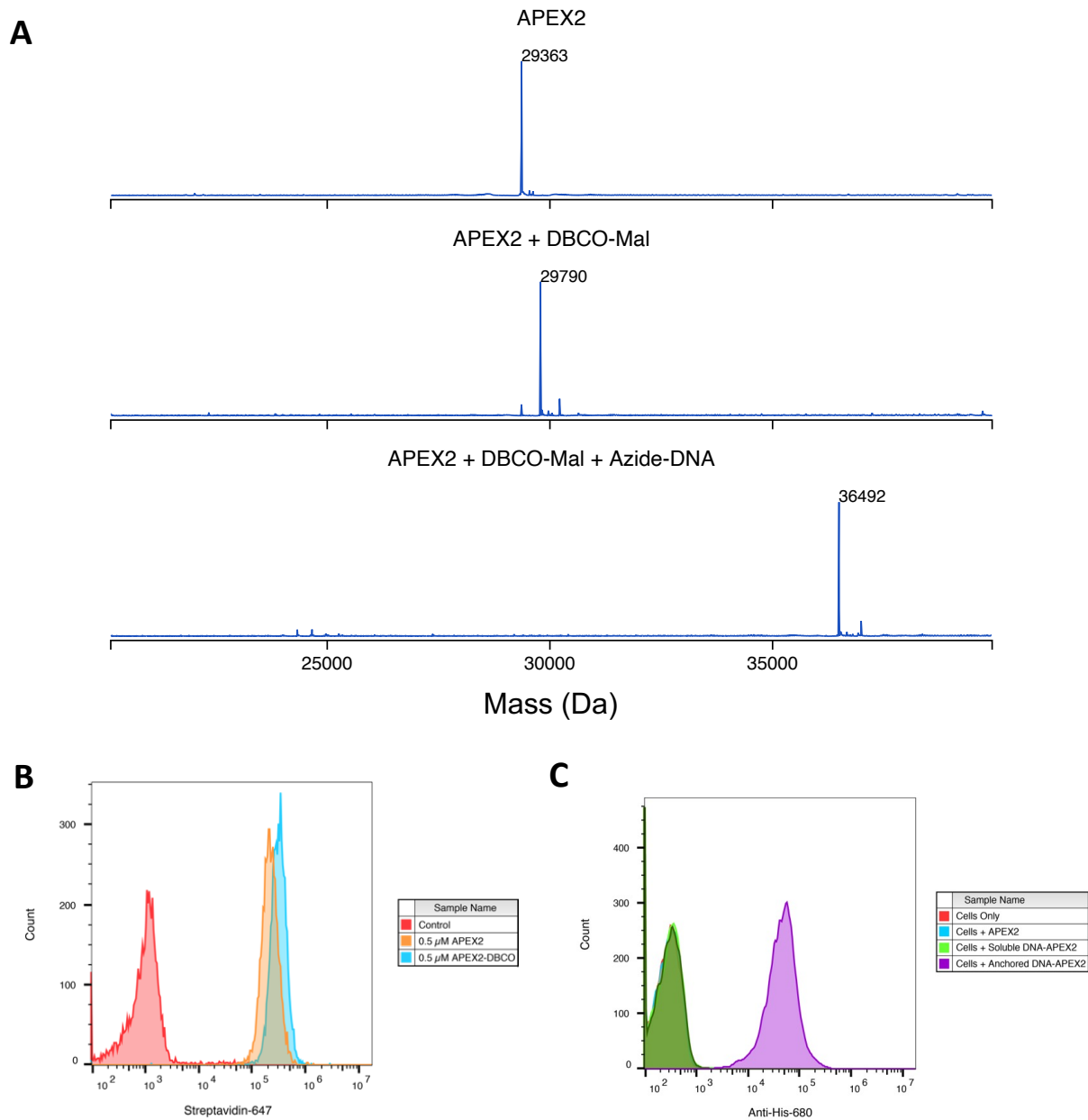
**Figure 3.3: Labeling and efficacy of APEX2 with DNA.** (A) APEX2 was first conjugated with DBCO-Maleimide (DBCO-Mal) reagent at 40 equivalents for 5 hours (80% conversion to the singly labeled product). Following desalting, 3 equivalences of Azide-DNA was added to the conjugate and purified by a Nickel column. Both reactions were monitored by LC-MS as shown. (B) 500,000 Expi293 cells were labeled with 0.5 µM purified APEX2 and DBCO-labeled APEX2 for 2 min. Extent of biotinylation of target cells was quantified by flow cytometry staining with streptavidin-647. (C) The DNA-APEX2 conjugate was shown to be tethered in the presence of the lipidated DNA (purple) and not in the absence (green), as detected by an Anti-His 680 antibody. Unlabeled APEX2 (blue) additionally did not result in a signal shift. Flow cytometry images are representative of one biological replicate.
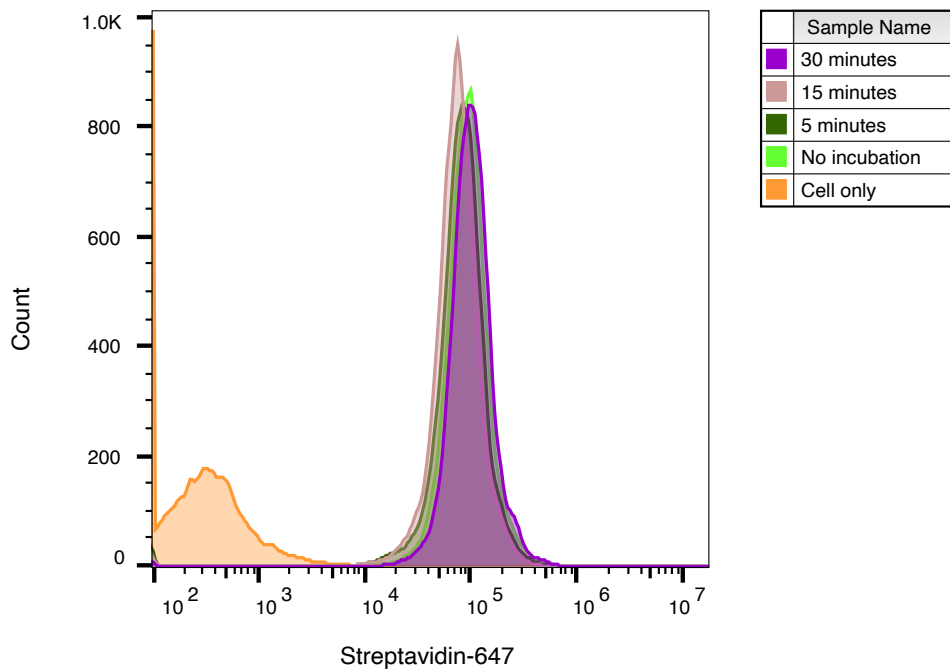
**Figure 3.4: WGA-HRP pre-incubation time on cells has no effect on labeling efficiency.**
WGA-HRP was incubated on Expi293 cells for 0-30 min to determine optimal incubation time on ice before labeling. All tested times resulted in similar cell surface biotinylation efficiencies and signified that no incubation time was needed. Flow cytometry images are representative of one biological replicate.
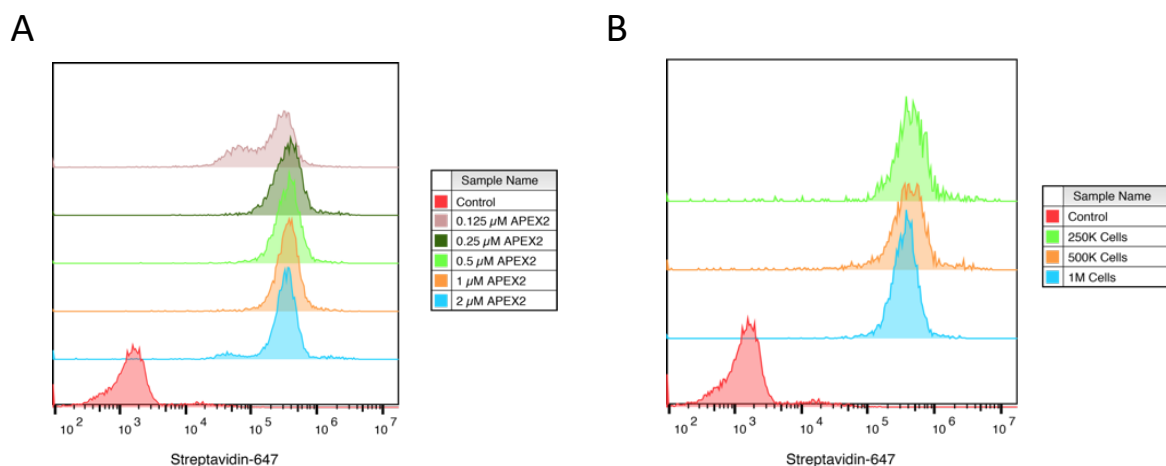
**Figure 3.5: Optimization of APEX2 concentrations on cell by flow cytometry.** (A) 500,000 Expi293 cells were labeled for 2 min with increasing amounts of purified APEX2 enzyme and extent of labeling was quantified by flow cytometry staining with streptavidin-647. (B) Varying numbers of Expi293 cells were labeled for 2 min with 0.5 µM APEX2 to test range of cell numbers for labeling. Flow cytometry images are representative of one biological replicate.
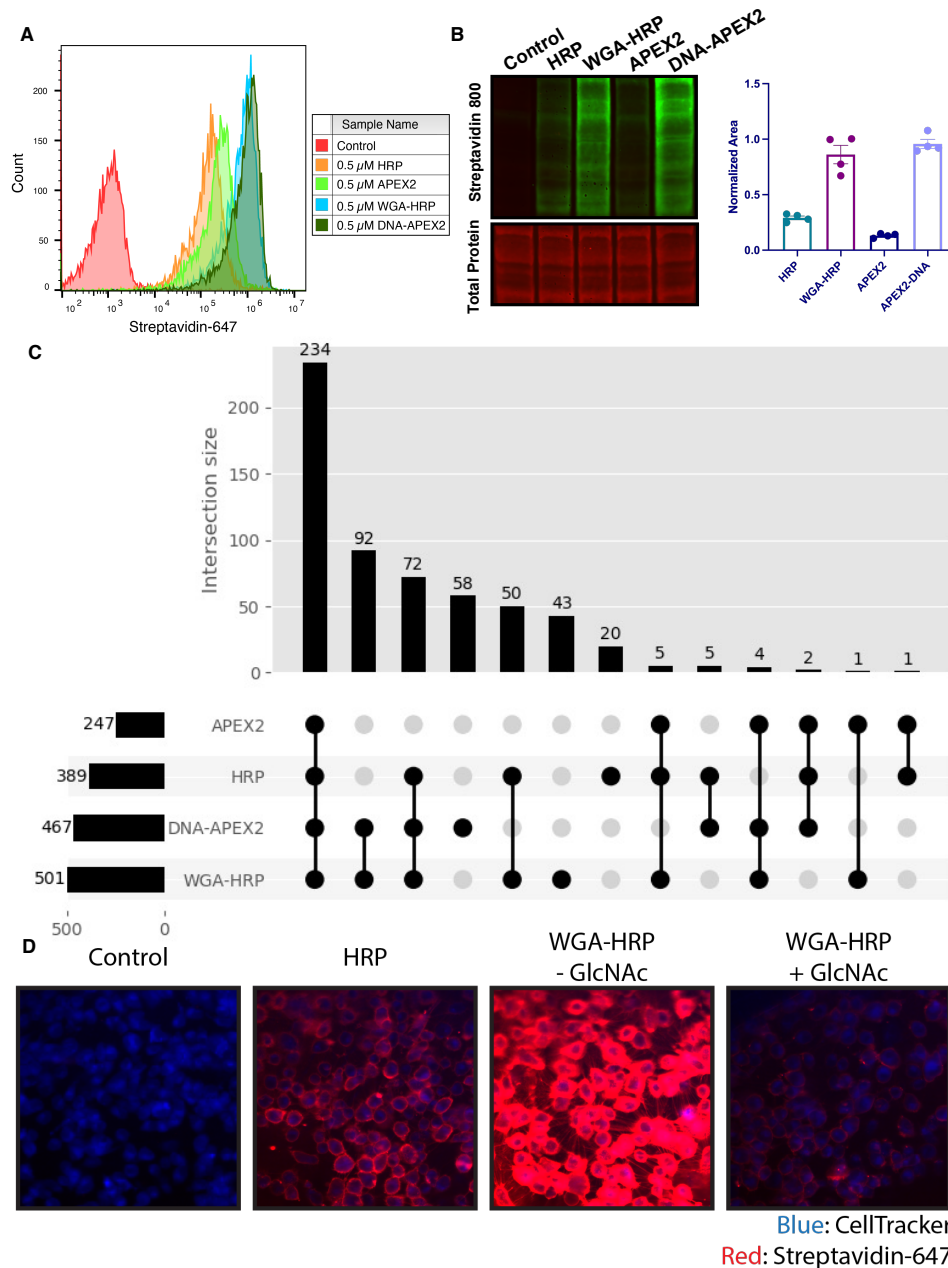
**Figure 3.6: Membrane-localized peroxidases increases membrane proteome biotinylation compared to non-tethered counterparts.** (A) Biotinylation of Expi293 cells treated with free enzyme (APEX2 or HRP) or cell-tethered enzyme (DNA-APEX2 or WGA-HRP) shown by flow cytometry. Signal is read out using Streptavidin-AlexaFluor-647. (B) Comparison of cell labeling with either free enzyme (APEX2 or HRP) or cell-tethered enzyme (DNA-APEX2 or WGA-HRP) shown by Streptavidin-800 western blot and total protein stain. Normalized area is plotted to the right. (C) Number of cell membrane proteins identified by mass spectrometry (>2 unique peptides, <1% FDR, found in both biological replicates) after treating 500,000 KP-4 pancreatic cancer cells with either free enzyme (APEX2 or HRP) or cell-tethered enzyme (DNA-APEX2 or WGA-HRP). (D) Microscopy images depicting extent of labeling with free HRP compared to WGA-HRP with and without the blocking sugar GlcNAc. All western blot images,

microscopy images, mass spectrometry data, and flow cytometry data are representative of two biological replicates.
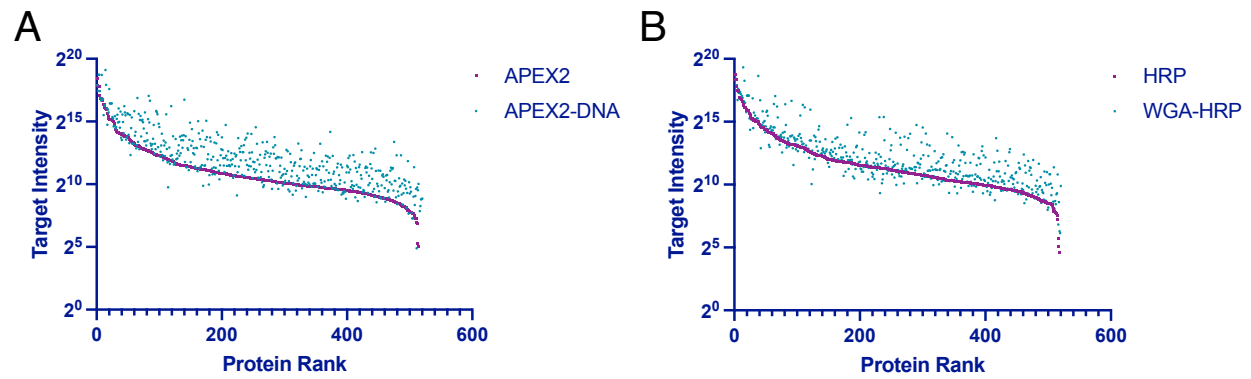
**Figure 3.7: Rank ordered intensities for surface annotated proteins detected in tethered and untethered enzyme samples.** (A) A rank ordered list of MS1 intensities (area) for surface annotated proteins found in APEX2 and DNA-APEX2 samples. (B) A rank ordered list of MS1 intensities (area) for surface annotated proteins found in HRP and WGA-HRP samples. The average of two biological replicates are plotted.
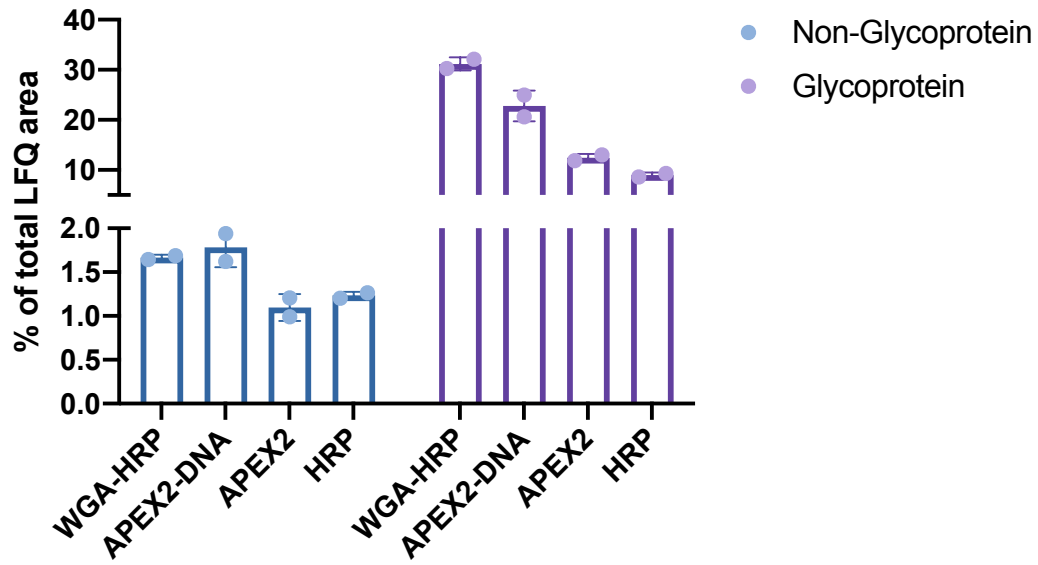
**Figure 3.8: Comparison of enrichment for glycosylated and non-glycosylated proteins.**
Total MS1 intensities (LFQ area) of Uniprot annotated glycoproteins compared to non-glycoproteins in the SURFY database for each labeling method compared to the total area sum intensity.
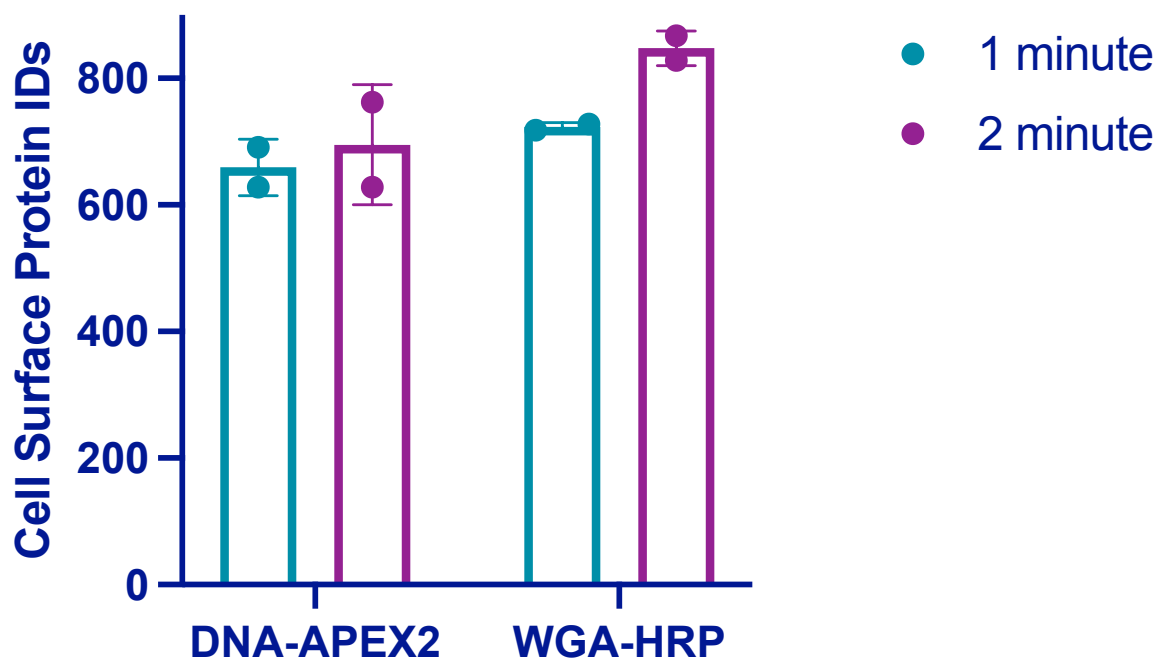
**Figure 3.9: Target intensities and total plasma membrane protein identifications for DNA-APEX2 and WGA-HRP labeling experiments as a function of time.** 500,000 PaTu8902 pancreatic cancer cells were labeled with either 0.5 µM DNA-APEX2 or 0.5 µM WGA-HRP for 1 or 2 minutes at 37°C. After cell surface enrichment and mass spectrometry analysis, the plasma membrane derived protein identifications were totaled.  Two biological replicates are plotted.
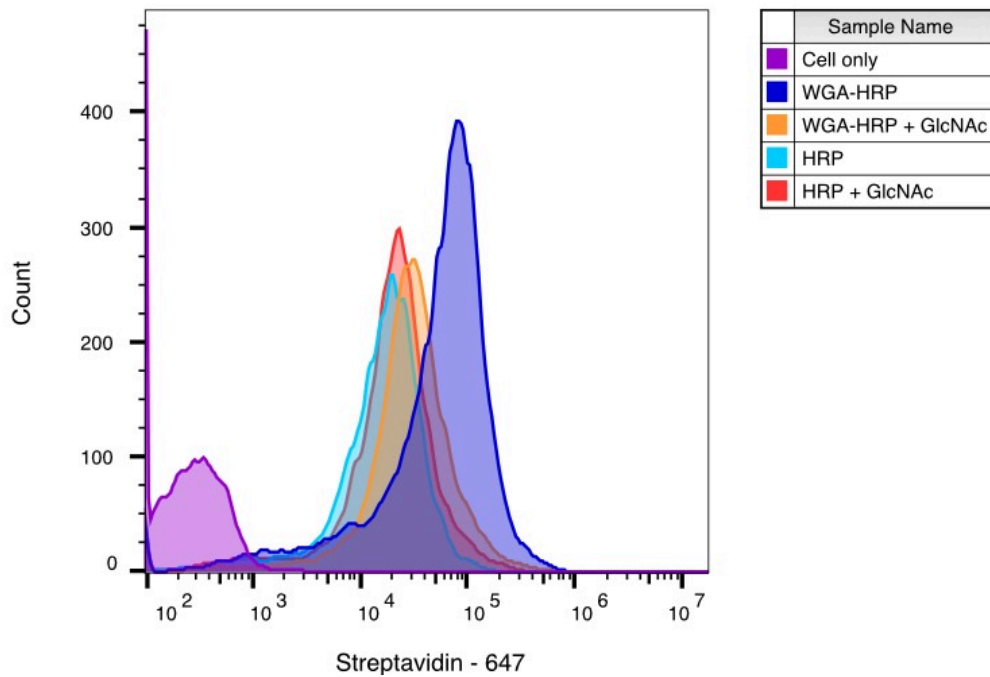
**Figure 3.10: WGA-HRP labeling is N-acetyl-D-glucosamine (GlcNAc) dependent.**
Biotinylation of RWPE-1 Myc cells with WGA-HRP was determined with (orange) and without (dark blue) 100 mg/mL GlcNAc. There is a significant leftward shift in the degree of labeling in the absence of competing GlcNAc, demonstrating that the enhanced labeling by WGA-HRP is GlcNAc dependent. The degree of labeling is similar to soluble HRP, as shown in light blue. Importantly, presence of GlcNAc in solution did not generally affect HRP labeling as seen by the control in red. The data presented is from one biological replicate.
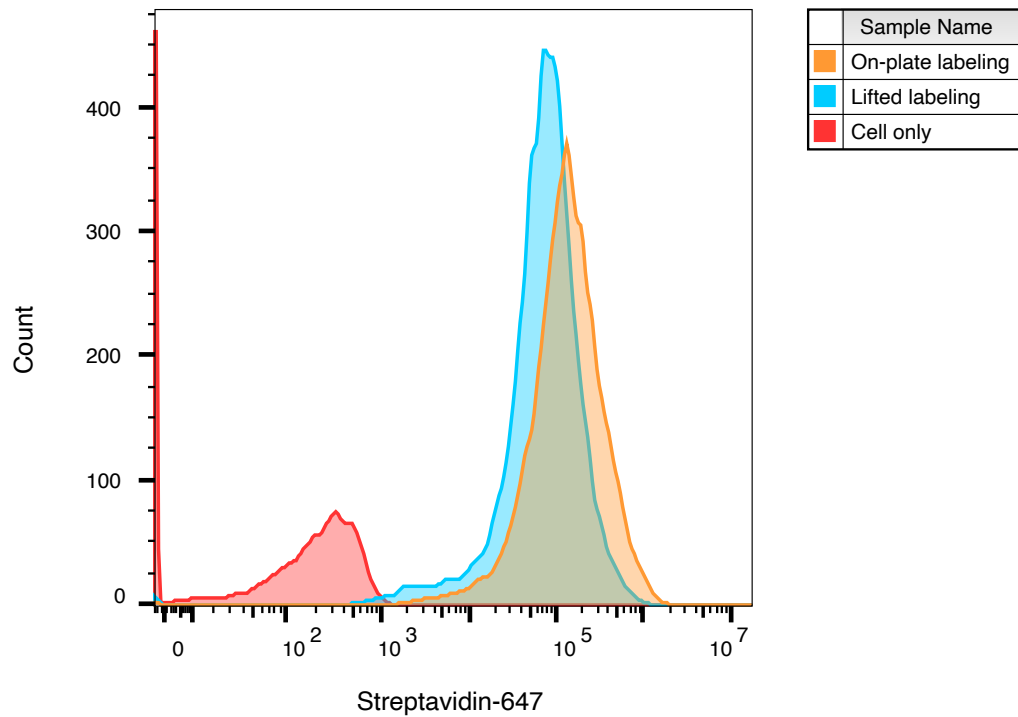
**Figure 3.11: WGA-HRP can be used to label adherent cells on-plate.** Cell surface labeling was compared between labeling adherent cells on a tissue culture plate vs. lifting cells and then performing labeling. Cell surface biotinylation was detected by streptavidin-Alexa Fluor 647. The data presented is from one biological replicate.
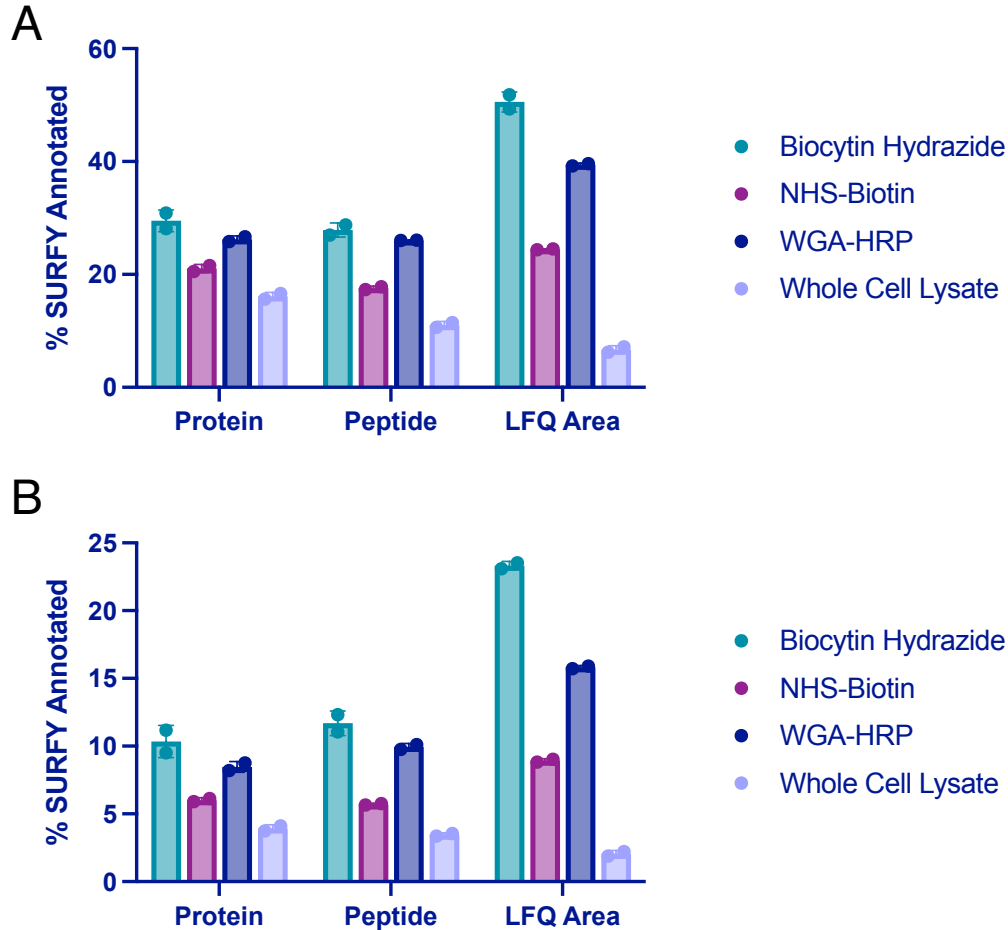
**Figure 3.12: Comparison of surface enrichment between replicates for different mass spectrometry methods.** (A)The top three methods (NHS-Biotin, Biocytin Hydrazide, and WGA-HRP) were compared for their ability to enrich cell surface proteins on 1.5 M RWPE-1 Control cells by LC-MS/MS after being searched with the Uniprot GOCC Plasma Membrane database. Shown are enrichment levels on the protein, peptide, and average MS1 intensity of top three peptides (LFQ area) levels. (B) The top three methods (NHS-Biotin, Biocytin Hydrazide, and WGA-HRP) were compared for their ability to enrich cell surface proteins on 1.5 M RWPE-1 Control cells by LC-MS/MS after being searched with the entire human Uniprot database. Shown are enrichment levels on the protein, peptide, and average MS1 intensity of top three peptides (LFQ area) levels. Proteins or peptides detected from cell surface annotated proteins (determined by the SURFY database) were divided by the total number of proteins or peptides detected. LFQ areas corresponding to cell surface annotated proteins (SURFY) were divided by the total area sum intensity for each sample. The corresponding percentages for two biological replicates were plotted.
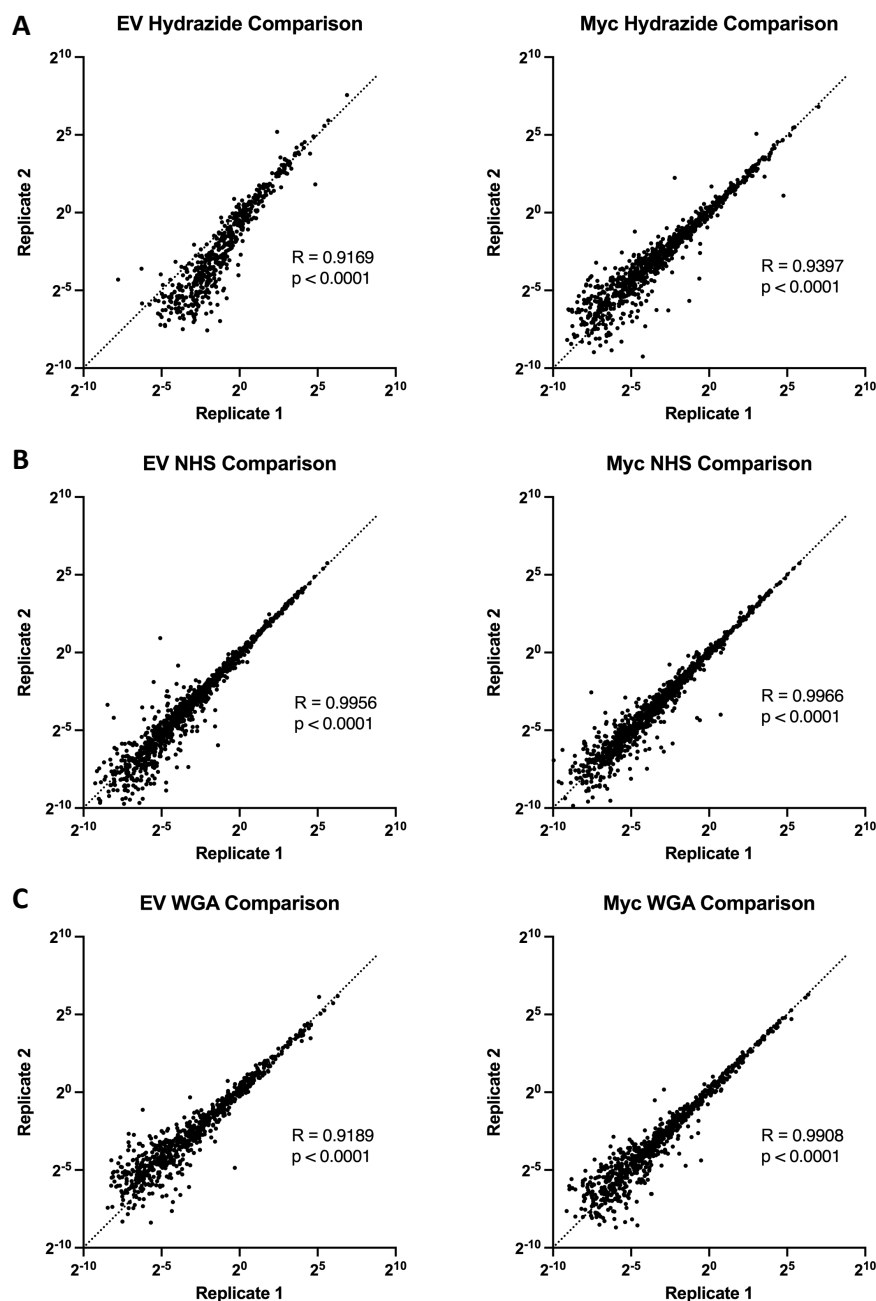
**Figure 3.13: Comparison of replicates for different mass spectrometry methods show that WGA-HRP has comparable reproducibility to NHS-Biotin and Hydrazide labeling.** (A) Spearman correlations of total area sum intensity normalized data from replicates of Hydrazide Control and Myc cells. (B) Spearman correlations of total area sum intensity normalized data from replicates of NHS Control and Myc cells. (C) Spearman correlations of total area sum intensity normalized data from replicates of WGA Control and Myc cells.

**Figure 3.14: WGA-HRP identifies a number of enriched markers on Myc-driven prostate cancer cells.** (A) Scheme for biotin labeling, and label-free quantification (LFQ) by LC-MS/MS for RWPE-1 Control and Myc over-expression cells. (B) Microscopy image depicting morphological changes between RWPE-1 Control and RWPE-1 Myc cells after 3 days in culture. (C) Volcano plot depicting the LFQ comparison of RWPE-1 Control and Myc la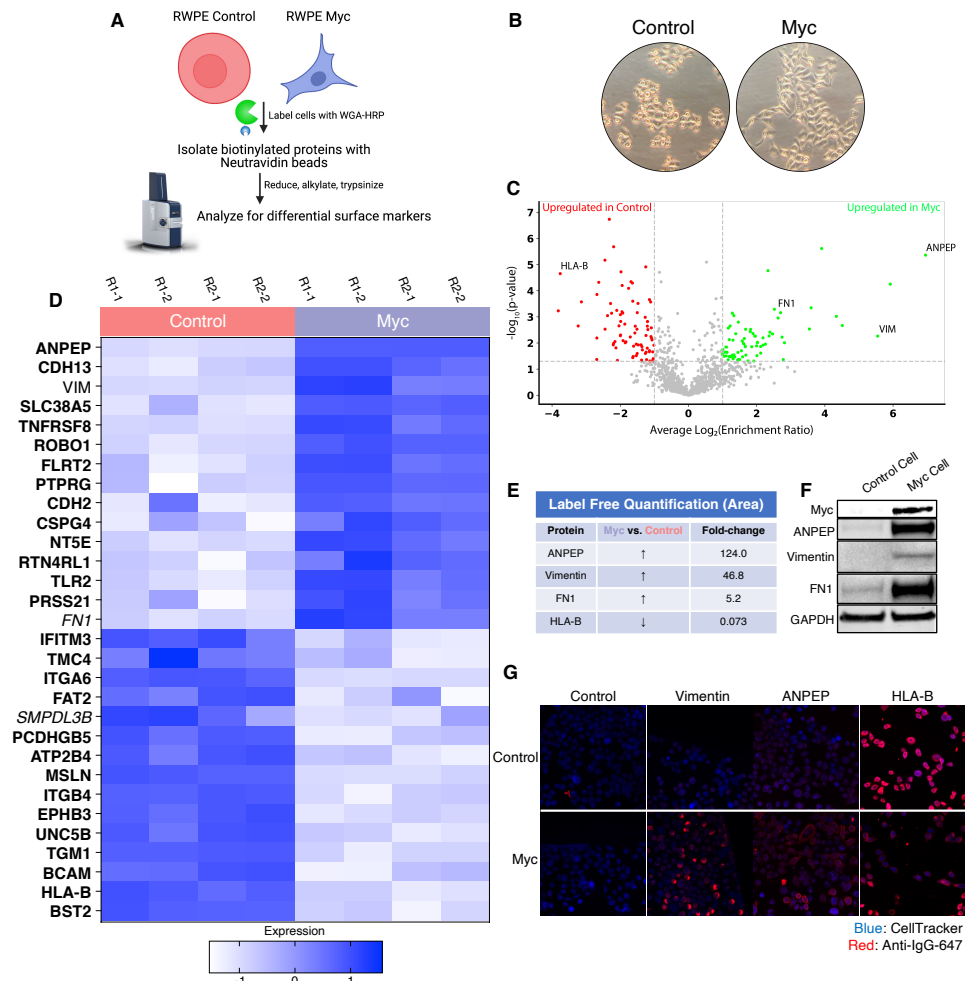beled cells. Red labels upregulated proteins in the RWPE-1 Control cells over Myc cells and green labels upregulated proteins in the RWPE-1 Myc cells over Control cells. All colored proteins are 2-fold enriched in either dataset between four replicates (two technical, two biological, p<0.05). (D) Heatmap of the 30 most upregulated transmembrane (bold) or secreted (non-bold) proteins in either RWPE-1 Control or Myc cells. Scale indicates intensity, defined as (LFQ Area - Mean LFQ Area)/Standard Deviation. (E) Table indicating fold-change of most differentially regulated proteins by LC-MS/MS for RWPE-1 Control and Myc cells. (F) Upregulated proteins in RWPE-1 Myc cells (Myc, ANPEP, Vimentin, and FN1) are confirmed by western blot. (G) Upregulated surface proteins in RWPE-1 Myc cells (Vimentin, ANPEP, FN1) are detected by immunofluorescence microscopy. The downregulated protein HLA-B by Myc over-expression was also detected by immunofluorescence microscopy. All western blot images and microscopy images are representative of two biological replicates. Mass spectrometry data is based on two biological and two technical replicates (N = 4).
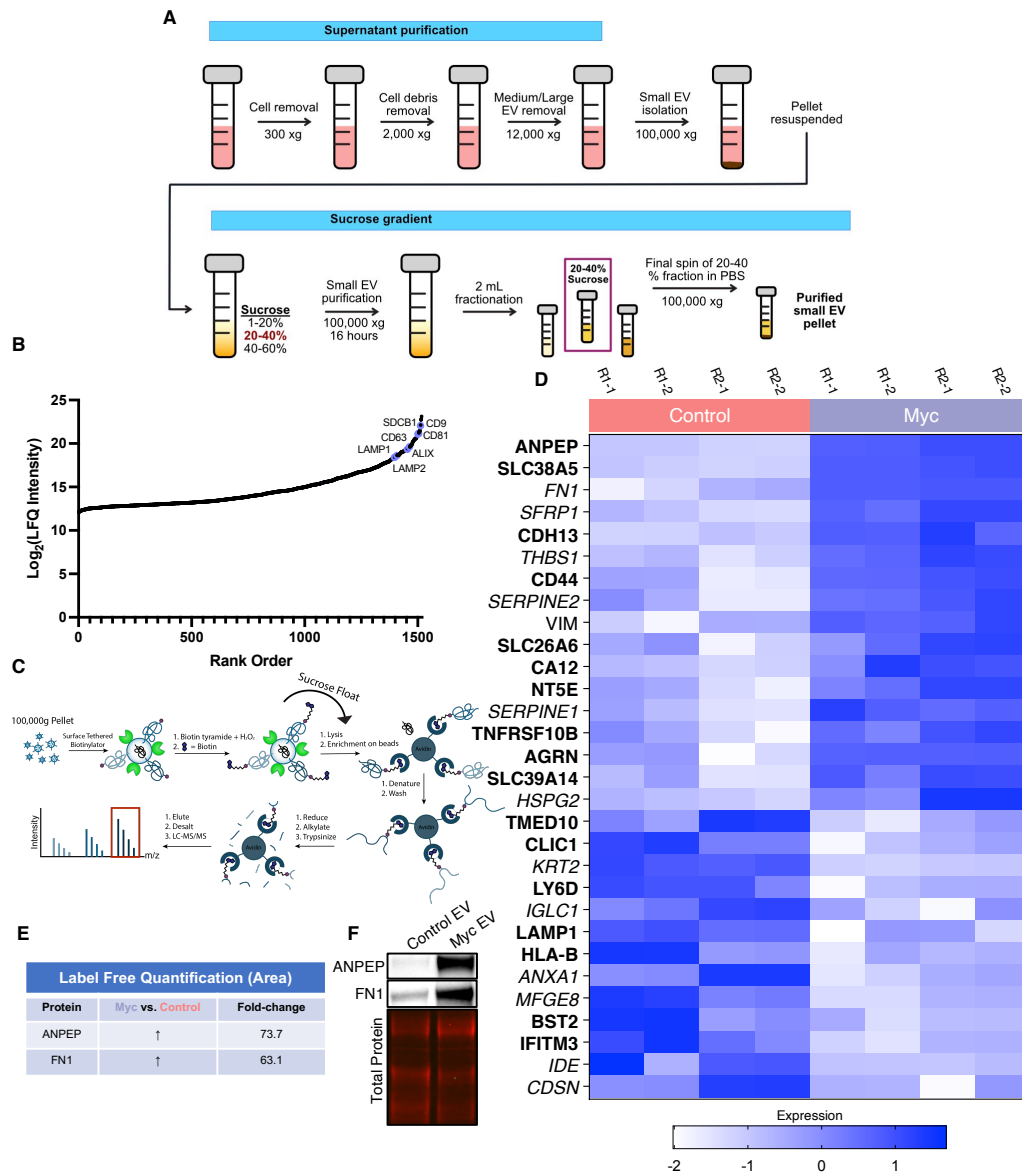
**Figure 3.15: WGA-HRP identifies a number of enriched markers on Myc-driven prostate cancer EVs.** (A) Workflow for small EV isolation from cultured cells. (B) Labeled proteins indicating canonical exosome markers (ExoCarta Top 100 List) detected after performing label-free quantification (LFQ) from whole EV lysate. The LFQ intensities were averaged for Control and Myc EVs, and the resulting protein list is graphed from least abundant to most abundant. (C) Workflow of EV labeling and preparation for mass spectrometry. (D) Heatmap of the 30 most upregulated proteins in either RWPE-1 Control or Myc EVs. Scale indicates intensity, defined as (LFQ Area - Mean LFQ Area)/Standard Deviation. (E) Table indicating fold-change of most differentially regulated proteins by LC-MS/MS for RWPE-1 Control and Myc cells. (F) Upregulated proteins in RWPE-1 Myc EVs (ANPEP and FN1) are confirmed by western blot. Mass spectrometry data is based on two biological and two technical replicates (N = 4). Due to limited sample yield, one replicate was performed for the EV western blot.
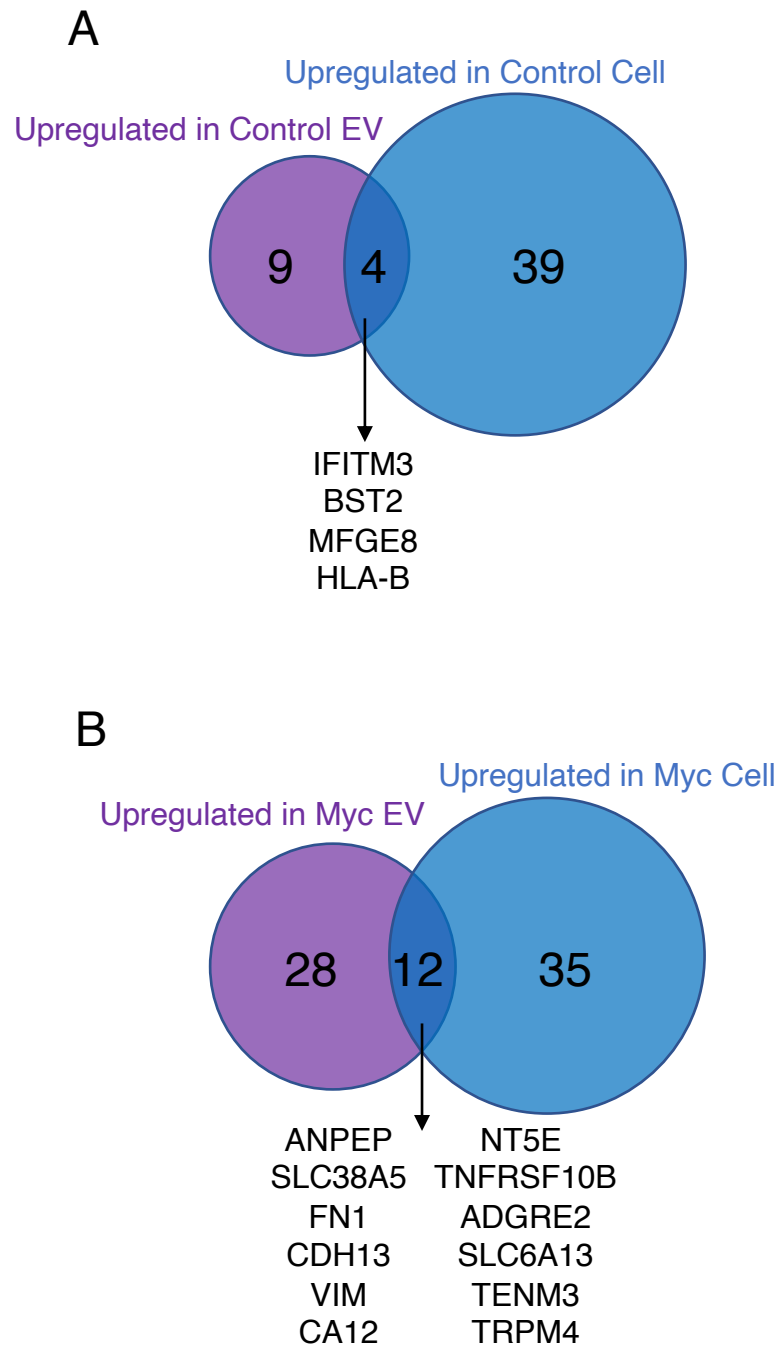
**Figure 3.16: Venn diagram comparing enriched targets (>2-fold) in Cells and EVs.** (A) Surface and secreted targets that were found enriched in the Control EVs (purple) and Control cells (blue) when each is separately compared to Myc EVs and Myc cells, respectively. The 4 overlapping enriched targets in common between Control cells and Control EVs are listed in the center. (B) Surface and secreted targets that were found enriched in the Myc EVs (purple) and Myc cells (blue) when each is separately compared to Control EVs and Control cells, respectively. The 12 overlapping enriched targets in common between Myc cells and Myc EVs are listed in the center.
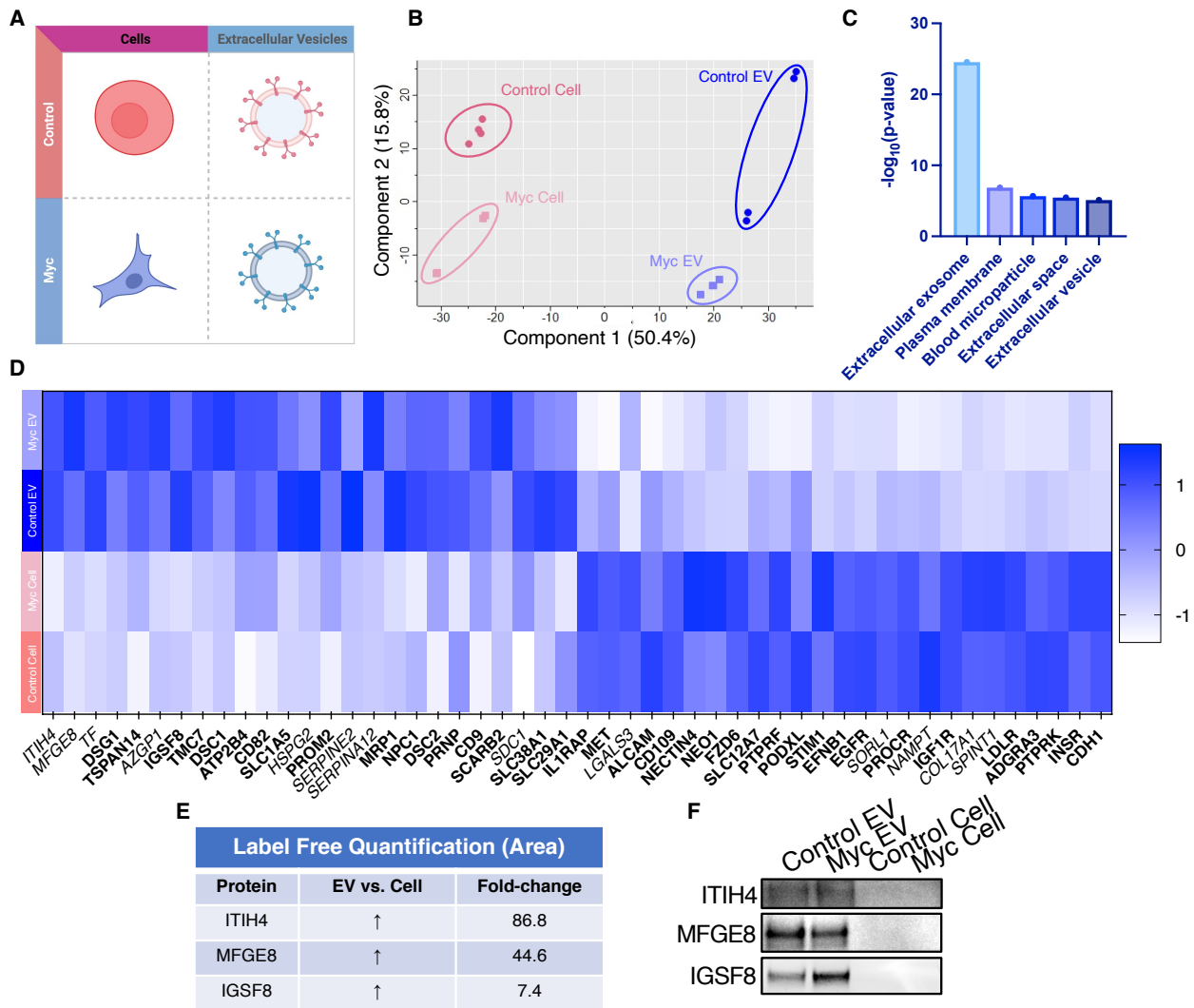
**Figure 3.17: WGA-HRP identifies a number of EV-specific markers that are present regardless of oncogene status.** (A) Matrix depicting samples analyzed during LFQ comparison--Control and Myc cells, as well as Control and Myc EVs. (B) Principle component analysis (PCA) of all four groups analyzed by LFQ. Component 1 (50.4%) and component 2 (15.8%) are graphed. (C) Functional annotation was performed for each gene cluster using DAVID Bioinformatics Resource 6.8 and the highest ranking annotation features for the EV-specific gene cluster are shown. (D) Heatmap of the 50 most upregulated proteins in either RWPE-1 cells or EVs. Proteins are listed in decreasing order of expression with the most highly expressed proteins in EVs on the far left and the most highly expressed proteins in cells on the far right. Averages from all four replicates of each sample type are graphed. Scale indicates intensity, defined as (LFQ Area - Mean LFQ Area)/Standard Deviation. Extracellular proteins with annotated transmembrane domains are bolded and annotated secreted proteins are italicized. (E) Table indicating fold-change of most differentially regulated proteins by LC-MS/MS for RWPE-1 EVs compared to parent cells. (F) Western blot showing the EV specific marker ITIH4, IGSF8, and MFGE8.Mass spectrometry data is based on two biological and two technical replicates (N = 4). Due to limited sample yield, one replicate was performed for the EV western blot.
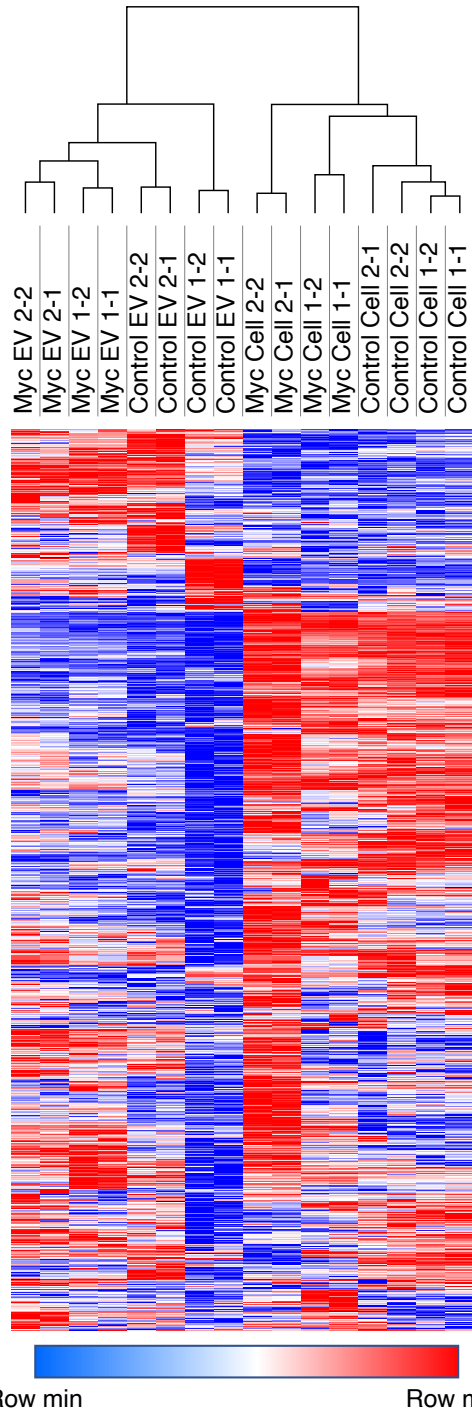
**Figure 3.18: Heatmap comparison of biological and technical replicates of RWPE-1 Control/Myc cells and EVs.** Biological and technical replicates cluster together based on both oncogene status and compartment for EV or cell surface. Proteins with no area values were assigned an imputed value using Perseus. Heatmap clustering is based off of the Pearson correlation between all replicates on both columns and rows. Heatmap was produced using Morpheus, https://software.broadinstitute.org/Morpheus. The first number following the sample name denotes the biological replicated and second number denotes the technical replicate.

**References:**

1.    Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 1–14 (2014).

2.    Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science (80-. ).* **602**, 597–602 (2017).

3.    Dyer, K. F. The Quiet Revolution: A New Synthesis of Biological Knowledge. *J. Biol. Educ.* **5**, 15–24 (1971).

4.    Adams, G. P. & Weiner, L. M. Monoclonal antibody therapy of cancer. *Nat. Biotechnol.* **23**, 1147–1157 (2005).

5.    Mao, S.-Y. & Mullins, J. M. Conjugation of fluorochromes to antibodies. *Methods Mol. Biol.* **588**, 43–8 (2010).

6.    Meares, C. F. *et al.* Conjugation of antibodies with bifunctional chelating agents: isothiocyanate and bromoacetamide reagents, methods of analysis, and subsequent addition of metal ions. *Anal. Biochem.* **142**, 68–78 (1984).

7.    Darmanis, S. *et al.* Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells. *Cell Rep.* **14**, 380–389 (2016).

8.    Stoeckius, M. *et al.* Large-scale simultaneous measurement of epitopes and transcriptomes in single cells. (2017).

9.    Tushir-Singh, J. Antibody-siRNA conjugates: drugging the undruggable for anti-leukemic therapy. *Expert Opin. Biol. Ther.* **17**, 325–338 (2017).

10.   Lewis Phillips, G. D. *et al.* Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Res.* **68**, 9280–90 (2008).

11.   Younes, A. *et al.* Brentuximab vedotin (SGN-35) for relapsed CD30-positive lymphomas.

*N. Engl. J. Med.* **363**, 1812–1821 (2010).

12.     DiJoseph, J. F. *et al.* Antibody-targeted chemotherapy with CMC-544: a CD22-targeted immunoconjugate of calicheamicin for the treatment of B-lymphoid malignancies. *Blood* **103**, 1807–14 (2004).

13.     Egan, P. C. & Reagan, J. L. The return of gemtuzumab ozogamicin: a humanized anti-CD33 monoclonal antibody-drug conjugate for the treatment of newly diagnosed acute myeloid leukemia. *Onco. Targets. Ther.* **11**, 8265–8272 (2018).

14.     Beck, A., Goetsch, L., Dumontet, C. & Corvaïa, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nature Reviews Drug Discovery* vol. 16 315–337 (2017).

15.     Panowski, S., Bhakta, S., Raab, H., Polakis, P. & Junutula, J. R. Site-specific antibody drug conjugates for cancer therapy. *MAbs* **6**, 34–45 (2014).

16.     Chalouni, C. & Doll, S. Fate of Antibody-Drug Conjugates in Cancer Cells. *J. Exp. Clin. Cancer Res.* **37**, 1–12 (2018).

17.     Tsuchikama, K. & An, Z. Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein Cell* **9**, 33–46 (2018).

18.     Hamblett, K. J. *et al.* Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. *Clin. Cancer Res.* **10**, 1–9 (2013).

19.     Bryant, P. *et al.* In vitro and in vivo evaluation of cysteine rebridged trastuzumab-MMAE antibody drug conjugates with defined drug-to-antibody ratios. *Mol. Pharm.* **12**, 1872–1879 (2015).

20.     Liu-Shin, L. P. Y., Fung, A., Malhotra, A. & Ratnaswamy, G. Evidence of disulfide bond scrambling during production of an antibody-drug conjugate. *MAbs* **10**, 1190–1199

(2018).

21.    Junutula, J. R. *et al.* Rapid identification of reactive cysteine residues for site-specific labeling of antibody-Fabs. *J. Immunol. Methods* **332**, 41–52 (2008).

22.    Ohri, R. *et al.* High-Throughput Cysteine Scanning to Identify Stable Antibody Conjugation Sites for Maleimide- and Disulfide-Based Linkers. *Bioconjug. Chem.* **29**, 473–485 (2018).

23.    Persson, H. *et al.* CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J. Mol. Biol.* **425**, 803–811 (2013).

24.    Hornsby, M. *et al.* A High Through-put Platform for Recombinant Antibodies to Folded Proteins. *Mol. Cell. Proteomics* **14**, 2833–47 (2015).

25.    Christian, A. H. *et al.* A Physical Organic Approach to Tuning Reagents for Selective and Stable Methionine Bioconjugation. *J. Am. Chem. Soc.* **141**, 12657–12662 (2019).

26.    Mendelsohn, B. A. *et al.* Investigation of Hydrophilic Auristatin Derivatives for Use in Antibody Drug Conjugates. *Bioconjug. Chem.* **28**, 371–381 (2017).

27.    Gao, X. *et al.* Effect of individual Fc methionine oxidation on FcRn binding: Met252 oxidation impairs FcRn binding more profoundly than Met428 oxidation. *J. Pharm. Sci.* **104**, 368–77 (2015).

28.    Haberger, M. *et al.* Assessment of chemical modifications of sites in the CDRs of recombinant antibodies: Susceptibility vs. functionality of critical quality attributes. *MAbs* **6**, 327–339 (2014).

29.    Kuo, T. T. & Aveson, V. G. Neonatal Fc receptor and IgG-based therapeutics. *MAbs* **3**, 422–430 (2011).

30.    Pyun, S. Y., Kim, T. R., Lee, C. R. & Kim, W. G. Kinetics studies on the mechanism of

hydrolysis of S-phenyl-S-vinyl-N-p-tosylsulfilimine derivatives. *Bull. Korean Chem. Soc.* **24**, 306–310 (2003).

31. Shen, B. Q. *et al.* Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nat. Biotechnol.* **30**, 184–189 (2012).

32. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.* **229**, 969–995 (1993).

33. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **2016**, 5.6.1-5.6.37 (2016).

34. Schrödinger Release 2014-3: MacroModel, Schrödinger, LLC, New York, NY, 2014. Pymol.

35. Tubert-Brohman, I., Sherman, W., Repasky, M. & Beuming, T. Improved docking of polypeptides with Glide. *J. Chem. Inf. Model.* **53**, 1689–99 (2013).

36. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **120**, 215–241 (2008).

37. Sassoon, I. & Blanc, V. Antibody-drug conjugate (ADC) clinical pipeline: A review. *Methods Mol. Biol.* **1045**, 1–27 (2013).

38. Krammer, F. & Simon, V. Serology assays to manage COVID-19. *Science (80-. ).* **368**, 1060–1061 (2020).

39. Long, Q. X. *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med. 2020 266* **26**, 845–848 (2020).

40. To, K. K. W. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* **20**, 565–574 (2020).

41. Zhao, J. *et al.* Antibody Responses to SARS-CoV-2 in Patients With Novel Coronavirus Disease 2019. *Clin. Infect. Dis.* **71**, 2027–2034 (2020).

42. Casadevall, A. & Pirofski, L. A. The convalescent sera option for containing COVID-19. *J. Clin. Invest.* **130**, 1545–1548 (2020).

43. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nat. 2020 5847821* **584**, 437–442 (2020).

44. Rogers, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**, 956–963 (2020).

45. Lynch, K. L. *et al.* Magnitude and Kinetics of Anti-Severe Acute Respiratory Syndrome Coronavirus 2 Antibody Responses and Their Relationship to Disease Severity. *Clin. Infect. Dis.* **72**, 301–308 (2021).

46. 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 (2020).

47. Seow, J. *et al.* Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection. *Nat. Microbiol.* **5**, 1598 (2020).

48. Smith, T. R. F. *et al.* Immunogenicity of a DNA vaccine candidate for COVID-19. *Nat. Commun. 2020 111* **11**, 1–13 (2020).

49. Yu, J. *et al.* DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* **369**, 806–811 (2020).

50. van Doremalen, N. *et al.* ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nat. 2020 5867830* **586**, 578–582 (2020).

51. Amanat, F. *et al.* A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med. 2020 267* **26**, 1033–1036 (2020).

52. Tan, W. *et al.* Viral Kinetics and Antibody Responses in Patients with COVID-19. *medRxiv* 2020.03.24.20042382 (2020) doi:10.1101/2020.03.24.20042382.

53. Xiang, J. *et al.* Evaluation of Enzyme-Linked Immunoassay and Colloidal Gold-Immunochromatographic Assay Kit for Detection of Novel Coronavirus (SARS-Cov-2) Causing an Outbreak of Pneumonia (COVID-19). *medRxiv* 2020.02.27.20028787 (2020) doi:10.1101/2020.02.27.20028787.

54. Li, Z. *et al.* Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J. Med. Virol.* **92**, 1518–1524 (2020).

55. Lassaunière, R. *et al.* Evaluation of nine commercial SARS-CoV-2 immunoassays. *medRxiv* 2020.04.09.20056325 (2020) doi:10.1101/2020.04.09.20056325.

56. Whitman, J. D. *et al.* Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. *Nat. Biotechnol. 2020 3810* **38**, 1174–1183 (2020).

57. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **11**, 400–408 (2016).

58. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).

59. Kang, S. *et al.* Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* **10**, 1228–1238

(2020).

60.     Qu, J. *et al.* Profile of Immunoglobulin G and IgM Antibodies Against Severe Acute
        Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Clin. Infect. Dis.* **71**, 2255–2258
        (2020).

61.     Stadlbauer, D. *et al.* SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a
        Serological Assay, Antigen Production, and Test Setup. *Curr. Protoc. Microbiol.* **57**,
        (2020).

62.     Zhao, J. *et al.* Comparison of Immunoglobulin G Responses to the Spike and
        Nucleocapsid Proteins of Severe Acute Respiratory Syndrome (SARS) Coronavirus in
        Patients with SARS. *Clin. Vaccine Immunol.* **14**, 839 (2007).

63.     Byrnes, J. R. *et al.* Competitive SARS-CoV-2 Serology Reveals Most Antibodies
        Targeting the Spike Receptor-Binding Domain Compete for ACE2 Binding. *mSphere* **5**,
        (2020).

64.     Rosado, J. *et al.* Serological signatures of SARS-CoV-2 infection: Implications for
        antibody-based diagnostics. *medRxiv* 2020.05.07.20093963 (2020)
        doi:10.1101/2020.05.07.20093963.

65.     Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal
        Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842.e16 (2020).

66.     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 (2020).

67.     Glasgow, A. *et al.* Engineered ACE2 receptor traps potently neutralize SARS-CoV-2.
        *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28046–28055 (2020).

68.     Poulsen, T. R., Meijer, P.-J., Jensen, A., Nielsen, L. S. & Andersen, P. S. Kinetic, affinity,

and diversity limits of human polyclonal antibody responses against tetanus toxoid. *J. Immunol.* **179**, 3841–3850 (2007).

69. Reddy, S. B. *et al.* Differences in affinity of monoclonal and naturally acquired polyclonal antibodies against Plasmodium falciparum merozoite antigens. *BMC Microbiol.* **15**, 1–11 (2015).

70. Zamecnik, C. R. *et al.* ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. *Cell reports. Med.* **1**, (2020).

71. Burbelo, P. D. *et al.* Sensitivity in Detection of Antibodies to Nucleocapsid and Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus 2 in Patients With Coronavirus Disease 2019. *J. Infect. Dis.* **222**, 206–213 (2020).

72. Tan, C. W. *et al.* A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction. *Nat. Biotechnol. 2020 389* **38**, 1073–1078 (2020).

73. Mäkelä, O., Rouslahti, E. & Seppälä, I. J. T. Affinity of IgM and IgG antibodies. *Immunochemistry* **7**, 917–932 (1970).

74. Wajnberg, A. *et al.* Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. *Science* **370**, 1227–1230 (2020).

75. Cervia, C. *et al.* Systemic and mucosal antibody responses specific to SARS-CoV-2 during mild versus severe COVID-19. *J. Allergy Clin. Immunol.* **147**, 545-557.e9 (2021).

76. Klein, S. L. *et al.* Sex, age, and hospitalization drive antibody responses in a COVID-19 convalescent plasma donor population. *J. Clin. Invest.* **130**, 6141–6150 (2020).

77. Denkinger, C. M. *et al.* Guidance for the Evaluation of Tuberculosis Diagnostics That Meet the World Health Organization (WHO) Target Product Profiles: An Introduction to

WHO Process and Study Design Principles. *J. Infect. Dis.* **220**, S91–S98 (2019).

78.  Pisanic, N. *et al.* COVID-19 serology at population scale: SARS-CoV-2-specific antibody responses in saliva. *J. Clin. Microbiol.* **59**, (2021).

79.  Faustini, S. E. *et al.* Detection of antibodies to the SARS-CoV-2 spike glycoprotein in both serum and saliva enhances detection of infection. *medRxiv* 2020.06.16.20133025 (2020) doi:10.1101/2020.06.16.20133025.

80.  Fry, S. R. *et al.* Detection of HSV type-1 and type-2 IgG using an in vitro PCA based homogeneous immunoassay. *Biochem. Biophys. Res. Commun.* **372**, 542–546 (2008).

81.  Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol. 2009 274* **27**, 378–386 (2009).

82.  Leth-Larsen, R., Lund, R. R. & Ditzel, H. J. Plasma Membrane Proteomics and Its Application in Clinical Cancer Biomarker Discovery *. *Mol. Cell. Proteomics* **9**, 1369–1382 (2010).

83.  Huang, G. Biotinylation of Cell Surface Proteins. *BIO-PROTOCOL* **2**, (2012).

84.  Rees, J. S., Li, X. W., Perrett, S., Lilley, K. S. & Jackson, A. P. Selective proteomic proximity labeling assay using tyramide (SPPLAT): A quantitative method for the proteomic analysis of localized membrane-bound protein clusters. *Curr. Protoc. Protein Sci.* **2015**, 19.27.1-19.27.18 (2015).

85.  Sears, R. M., May, D. G. & Roux, K. J. BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods Mol. Biol.* **2012**, 299–313 (2019).

86.  Bausch-Fluck, D., Hofmann, A. & Wollscheid, B. Cell Surface Capturing Technologies for the Surfaceome Discovery of Hepatocytes. *Methods Mol. Biol.* **909**, 1–16 (2012).

87.  Elschenbroich, S., Kim, Y., Medin, J. A. & Kislinger, T. Isolation of cell surface proteins

for mass spectrometry-based proteomics. *Expert Review of Proteomics* vol. 7 141–154 (2010).

88. Griffin, N. M. & Schnitzer, J. E. Overcoming Key Technological Challenges in Using Mass Spectrometry for Mapping Cell Surfaces in Tissues *. *Mol. Cell. Proteomics* **10**, S1–S14 (2011).

89. Kuhlmann, L., Cummins, E., Samudio, I. & Kislinger, T. Cell-surface proteomics for the identification of novel therapeutic targets in cancer. *https://doi.org/10.1080/14789450.2018.1429924* **15**, 259–275 (2018).

90. Kalxdorf, M., Gade, S., Eberl, H. C. & Bantscheff, M. Monitoring cell-surface N-glycoproteome dynamics by quantitative proteomics reveals mechanistic insights into macrophage differentiation. *Mol. Cell. Proteomics* **16**, 770–785 (2017).

91. Colombo, M., Raposo, G. & Théry, C. Biogenesis, Secretion, and Intercellular Interactions of Exosomes and Other Extracellular Vesicles. *http://dx.doi.org/10.1146/annurev-cellbio-101512-122326* **30**, 255–289 (2014).

92. Al-Nedawi, K. *et al.* Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat. Cell Biol. 2008 105* **10**, 619–624 (2008).

93. Edgar, J. R. Q & A: What are exosomes, exactly? *BMC Biol.* **14**, 1–7 (2016).

94. Kalluri, R. & LeBleu, V. S. The biology, function, and biomedical applications of exosomes. *Science (80-. )*. **367**, (2020).

95. Shurtleff, M. J., Temoche-Diaz, M. M. & Schekman, R. Extracellular Vesicles and Cancer: Caveat Lector. *https://doi.org/10.1146/annurev-cancerbio-030617-050519* **2**, 395–411 (2018).

96. Hoshino, A. *et al.* Tumour exosome integrins determine organotropic metastasis. *Nat.*

*2015 5277578* **527**, 329–335 (2015).

97.     Soung, Y. H., Ford, S., Zhang, V. & Chung, J. Exosomes in Cancer Diagnostics. *Cancers 2017, Vol. 9, Page 8* **9**, 8 (2017).

98.     Poggio, M. *et al.* Suppression of Exosomal PD-L1 Induces Systemic Anti-tumor Immunity and Memory. *Cell* **177**, 414-427.e13 (2019).

99.     Skog, J. *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol. 2008 1012* **10**, 1470–1476 (2008).

100.    Zhou, B. *et al.* Application of exosomes as liquid biopsy in clinical diagnosis. *Signal Transduct. Target. Ther. 2020 51* **5**, 1–14 (2020).

101.    Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem. 2017 912* **9**, 1181–1190 (2017).

102.    Hung, V. *et al.* Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **11**, 456–75 (2016).

103.    Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein–protein interactions and sensitive visualization of synapses. *Nat. Biotechnol. 2016 347* **34**, 774–780 (2016).

104.    Ye, M. *et al.* Rapid enzyme-mediated biotinylation for cell surface proteome profiling. *Anal. Chem.* **93**, 4542–4551 (2021).

105.    Li, J. *et al.* Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators. *Cell* **180**, 373-386.e15 (2020).

106.    Mcginnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. doi:10.1038/s41592-019-0433-8.

107.  Weber, R. J., Liang, S. I., Selden, N. S., Desai, T. A. & Gartner, Z. J. Efficient targeting of fatty-acid modified oligonucleotides to live cell membranes through stepwise assembly. *Biomacromolecules* **15**, 4621–4626 (2014).

108.  Mathiasen, M. L., Dillingham, C. M., Kinnavane, L., Powell, A. L. & Aggleton, J. P. Asymmetric cross-hemispheric connections link the rat anterior thalamic nuclei with the cortex and hippocampal formation. *Neuroscience* **349**, 128–143 (2017).

109.  Wang, T. & Miller, K. E. Characterization of glutamatergic neurons in the rat atrial intrinsic cardiac ganglia that project to the cardiac ventricular wall. *Neuroscience* **329**, 134–150 (2016).

110.  Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2015).

111.  Schäffer, C. & Messner, P. Emerging facets of prokaryotic glycosylation. *FEMS Microbiol. Rev.* **41**, 49–91 (2017).

112.  Weeks, A. M., Byrnes, J. R., Lui, I. & Wells, J. A. Mapping proteolytic neo-N termini at the surface of living cells. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

113.  Bausch-Fluck, D. *et al.* The in silico human surfaceome. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E10988–E10997 (2018).

114.  Weekes, M. P. *et al.* Comparative Analysis of Techniques to Purify Plasma Membrane Proteins. *J. Biomol. Tech.* **21**, 108 (2010).

115.  Litwin, M. S. & Tan, H. J. The Diagnosis and Treatment of Prostate Cancer: A Review. *JAMA* **317**, 2532–2542 (2017).

116.  Rawla, P. Epidemiology of Prostate Cancer. *World J. Oncol.* **10**, 63–89 (2019).

117.  Koh, C. M. *et al.* MYC and prostate cancer. *Genes and Cancer* **1**, 617–628 (2010).

118. Rebello, R. J., Pearson, R. B., Hannan, R. D. & Furic, L. Therapeutic Approaches Targeting MYC-Driven Prostate Cancer. *Genes 2017, Vol. 8, Page 71* **8**, 71 (2017).

119. Duijvesz, D. *et al.* Proteomic Profiling of Exosomes Leads to the Identification of Novel Biomarkers for Prostate Cancer. *PLoS One* **8**, e82589 (2013).

120. McKiernan, J. *et al.* A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy. *JAMA Oncol.* **2**, 882–889 (2016).

121. Akoto, T. & Saini, S. Role of Exosomes in Prostate Cancer Metastasis. *Int. J. Mol. Sci. 2021, Vol. 22, Page 3528* **22**, 3528 (2021).

122. Lorenc, T. *et al.* Exosomes in Prostate Cancer Diagnosis, Prognosis and Therapy. *Int. J. Mol. Sci. 2020, Vol. 21, Page 2118* **21**, 2118 (2020).

123. Saber, S. H. *et al.* Exosomes are the Driving Force in Preparing the Soil for the Metastatic Seeds: Lessons from the Prostate Cancer. *Cells 2020, Vol. 9, Page 564* **9**, 564 (2020).

124. Liu, C.-Y. *et al.* Vimentin contributes to epithelial-mesenchymal transition cancer cell mechanics by mediating cytoskeletal organization and focal adhesion maturation. *Oncotarget* **6**, 15966–15983 (2015).

125. Nakajima, S. *et al.* N-Cadherin Expression and Epithelial-Mesenchymal Transition in Pancreatic Carcinoma. *Clin. Cancer Res.* **10**, 4125–4133 (2004).

126. Mitra, A. *et al.* Cell-surface Vimentin: A mislocalized protein for isolating csVimentin(+) CD133(-) novel stem-like hepatocellular carcinoma cells expressing EMT markers. *Int. J. cancer* **137**, 491–496 (2015).

127. Noh, H. *et al.* Discovery of cell surface vimentin targeting mAb for direct disruption of GBM tumor initiating cells. *Oncotarget* **7**, 72021 (2016).

128. Blades, R. A., Keating, P. J., McWilliam, L. J., George, N. J. R. & Stern, P. L. Loss of

HLA class I expression in prostate cancer: Implications for immunotherapy. *Urology* **46**, 681–687 (1995).

129. Cornel, A. M., Mimpen, I. L. & Nierkens, S. MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. *Cancers 2020, Vol. 12, Page 1760* **12**, 1760 (2020).

130. Dhatchinamoorthy, K., Colbert, J. D. & Rock, K. L. Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front. Immunol.* **12**, 469 (2021).

131. Mathieu, M. *et al.* Specificities of exosome versus small ectosome secretion revealed by live intracellular tracking of CD63 and CD9. *Nat. Commun. 2021 121* **12**, 1–18 (2021).

132. Kilinc, S. *et al.* Oncogene-regulated release of extracellular vesicles ll Oncogene-regulated release of extracellular vesicles. *Dev. Cell* **56**, 1989-2006.e6 (2021).

133. Costa-Silva, B. *et al.* Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat. Cell Biol. 2014 176* **17**, 816–826 (2015).

134. Bandu, R., Oh, J. W. & Kim, K. P. Mass spectrometry-based proteome profiling of extracellular vesicles and their roles in cancer biology. *Exp. Mol. Med. 2019 513* **51**, 1–10 (2019).

135. Bilen, M. A. *et al.* Proteomics Profiling of Exosomes from Primary Mouse Osteoblasts under Proliferation versus Mineralization Conditions and Characterization of Their Uptake into Prostate Cancer Cells. *J. Proteome Res.* **16**, 2709–2728 (2017).

136. Hosseini-Beheshti, E., Pham, S., Adomat, H., Li, N. & Tomlinson Guns, E. S. Exosomes as biomarker enriched microvesicles: Characterization of exosomal proteins derived from a panel of prostate cell lines with distinct AR phenotypes. *Mol. Cell. Proteomics* **11**, 863–885 (2012).

137. Doyle, L. M. & Wang, M. Z. Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells 2019, Vol. 8, Page 727* **8**, 727 (2019).

138. Santucci, L. *et al.* Biological surface properties in extracellular vesicles and their effect on cargo proteins. *Sci. Reports 2019 91* **9**, 1–12 (2019).

139. Philley, J. V. *et al.* Exosome secretome and mediated signaling in breast cancer patients with nontuberculous mycobacterial disease. *Oncotarget* **8**, 18070–18081 (2017).

140. Shimagaki, T. *et al.* Serum milk fat globule-EGF factor 8 (MFG-E8) as a diagnostic and prognostic biomarker in patients with hepatocellular carcinoma. *Sci. Reports 2019 91* **9**, 1–12 (2019).

141. Tutanov, O. *et al.* Proteomic Analysis of Blood Exosomes from Healthy Females and Breast Cancer Patients Reveals an Association between Different Exosomal Bioactivity on Non-tumorigenic Epithelial Cell and Breast Cancer Cell Migration in Vitro. *Biomol. 2020, Vol. 10, Page 495* **10**, 495 (2020).

142. Chandler, K. B. & Costello, C. E. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis* **37**, 1407–1419 (2016).

143. Gupte, T. M., Ritt, M., Dysthe, M., Malik, R. U. & Sivaramakrishnan, S. Minute-scale persistence of a GPCR conformation state triggered by non-cognate G protein interactions primes signaling. *Nat. Commun. 2019 101* **10**, 1–15 (2019).

144. Valitutti, S., Coombs, D. & Dupré, L. The space and time frames of T cell activation at the immunological synapse. *FEBS Lett.* **584**, 4851–4857 (2010).

145. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as

deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta - Gen. Subj.* **1473**, 4–8 (1999).

146. Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol. 2019 156* **15**, 346–366 (2019).

147. Guzman-Rojas, L. *et al.* Cooperative effects of aminopeptidase N (CD13) expressed by nonmalignant and cancer cells within the tumor microenvironment. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1637–1642 (2012).

148. Wickström, M., Larsson, R., Nygren, P. & Gullbo, J. Aminopeptidase N (CD13) as a target for cancer chemotherapy. *Cancer Science* vol. 102 501–508 (2011).

149. Leung, K. K. *et al.* Broad and thematic remodeling of the surfaceome and glycoproteome on isogenic cells transformed with driving proliferative oncogenes. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7764–7775 (2020).

150. Martinko, A. J. *et al.* Targeting RAS-driven human cancer cells with antibodies to upregulated and essential cell-surface proteins. *Elife* **7**, (2018).

151. Wei, J. *et al.* Profiling the surfaceome identifies therapeutic targets for cells with hyperactive mTORC1 signaling. *Mol. Cell. Proteomics* **19**, 297–307 (2020).

152. Bhutia, Y. D. & Ganapathy, V. Glutamine transporters in mammalian cells and their functions in physiology and cancer. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 2531–2539 (2016).

153. Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl. Acad. Sci.* **105**, 18782–18787 (2008).

154. Wang, J. P. & Hielscher, A. Fibronectin: How its aberrant expression in tumors may

improve therapeutic targeting. *J. Cancer* **8**, 674–682 (2017).

155.    Efthymiou, G. *et al.* Shaping Up the Tumor Microenvironment With Cellular Fibronectin. *Front. Oncol.* **10**, 641 (2020).

156.    Daubon, T. *et al.* Deciphering the complex role of thrombospondin-1 in glioblastoma development. *Nat. Commun. 2019 101* **10**, 1–15 (2019).

157.    Zhu, L. *et al.* THBS1 Is a Novel Serum Prognostic Factors of Acute Myeloid Leukemia. *Front. Oncol.* **9**, 1567 (2020).

158.    Cen, J. *et al.* Exosomal Thrombospondin-1 Disrupts the Integrity of Endothelial Intercellular Junctions to Facilitate Breast Cancer Cell Metastasis. *Cancers 2019, Vol. 11, Page 1946* **11**, 1946 (2019).

159.    Chen, C., Zhao, S., Karnad, A. & Freeman, J. W. The biology and role of CD44 in cancer progression: therapeutic implications. *J. Hematol. Oncol. 2018 111* **11**, 1–23 (2018).

160.    Shen, X. *et al.* Exosome-mediated transfer of CD44 from high-metastatic ovarian cancer cells promotes migration and invasion of low-metastatic ovarian cancer cells. *J. Ovarian Res.* **14**, 1–11 (2021).

161.    Beckler, M. D. *et al.* Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. *Mol. Cell. Proteomics* **12**, 343–355 (2013).

162.    Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat. Med. 2012 186* **18**, 883–891 (2012).

163.    Howarth, M. & Ting, A. Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc. 2008 33* **3**, 534–545 (2008).

164.    Cheek, J., Mandelman, D., Poulos, T. L. & Dawson, J. H. A study of the K+-site mutant

of ascorbate peroxidase: mutations of protein residues on the proximal side of the heme cause changes in iron ligation on the distal side. *JBIC J. Biol. Inorg. Chem. 1999 41* **4**, 64–72 (1999).

165. Linkert, M. *et al.* Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777–782 (2010).

166. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods 2012 97* **9**, 676–682 (2012).

167. Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat. Methods 2020 1712* **17**, 1229–1236 (2020).

**Publishing Agreement**

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution.  UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

*Susanna k. Elledge*

9037F75F403E4DA...          Author Signature

3/6/2022

Date