Scalable Affinity-Proteomics on Microparticles

Milad Dagher

Department of Biomedical Engineering
McGill University
Montreal, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Doctorate of Philosophy

December 2018

To *Najah*.

ACKNOWLEDGEMENTS

I would like to thank everyone that contributed to this thesis. First and foremost, I would like to thank Prof. David Juncker for the opportunity to develop this work within his lab. When I first showed up unannounced at your office, you told me about the problems in protein analysis; those problems became my challenges, and I hope we're one step closer to solving them today. Thank you for the guidance, feedback, and support David.

I also want to thank my PhD committee members, Prof. Hanadi Sleiman, Prof. Dusica Maysinger, and Prof. Brandon Xia for their scientific guidance and advice. Prof. Sleiman, I always look forward to your clever suggestions and ideas. Prof Maysinger, thank you for challenging me to learn more about the biology.

I am quite indebted to have had Dr. Andy Ng's scientific support during the early days of this project. Andy was always available to brainstorm new ideas, answer my (often basic) biochemistry questions, offer feedback on reports and manuscripts, or even help around the lab. Thank you Andy, our successes in this project owe a lot to your expertise.

I would like to acknowledge the support from the Natural Sciences and Engineering Research Council of Canada (NSERC) which funded part of the research presented in this thesis. I would also like to acknowledge the Fonds de Recherche du Québec – Nature et technologies (FQRNT) Doctoral fellowship, the NSERC CREATE program for Integrated Sensors and Systems (ISS), and McGill's Biomedical Engineering department for the excellence awards which all helped financially support me during the course of my PhD.

I want to thank everyone who helped with the administrative matters during my PhD: Pina Sorrini, Pammy Lo, and Trang Tran.

Thank you for the good times, DJ Group. The past years would have not been the same without you. To the 3010 crew: Grant, Ayo, Kate, and Philippe: sharing an inappropriately small conference room that first year was almost enjoyable thanks to you. To Jeff and Arya: some excellent rips in there, thank you for all the brainstorming sessions. Abhishek, Anne, Aless, Alex, Donald, Fred, Gina, Grace, Huiyan, MC, Oriol, Pammy, Philip, Philippe, Roozbeh, Seb B, and Veronique – it was a pleasure working alongside you all. To everyone else who I may have missed, or who briefly passed through DJ Group – thank you.

I am also grateful to all the outstanding undergraduate students that accepted to work with me and taught me how to become a better mentor. Special thanks to Farah Mousharbash, Joy Que, Raphi Zonis, Jill Laurin, Ilias Hurley, Nila Wu, and Zijie Jin.

I also want to thank all my friends, who helped me keep my head above water while making sure that I stayed grounded: Ferdinand, Arya, Tarek, Gilbert, Lea, Hind, Bashir, Tania, and Nila.

Last but not least, I want to thank my Dad and my brothers: Joseph and Jihad. Without their relentless backing and support, I could not be writing this thesis today.

ABSTRACT

The immunoassay is a core method for basic and clinical research, as well as clinical diagnostics. In particular, owing to the requirements of dual-antibody binding, the sandwich immunoassay provides exquisite specificity, sensitivity, and cost-efficiency in the measurement of low-abundance proteins within serum or plasma. Fluorescently-barcoded microparticles are a popular format for immunoassays as they allow multiplexing with high sample throughput and rapid read-out by cytometry. However, this widely used format suffers from two major shortcomings that limit its scalability and prohibit its use in large-scale proteomic studies. First and foremost, the specificity and sensitivity of multiplexed sandwich assays (MSAs) is severely hindered with increased multiplexing due to the well-known cross-reactivity of the detection antibodies. This reagent-driven cross-reactivity (rCR) occurs in all MSAs that apply the detection antibodies as a mixture. While rCR has been mitigated in planar and solution-phase assays, it continues to be the main obstacle towards increased multiplexing in microparticle-based MSAs. The second challenge is the generation of large numbers of spectrally-distinct microparticles for use in barcoding. Spectral barcoding of microparticles calls for their functionalization with distinct proportions of multicolour dyes to create the desired and unique barcode; however, multicolour Förster resonance energy transfer (mFRET) leads to hitherto unpredictable shifts in emission intensities thus confounding the barcodes. Thus, current methods for spectral barcoding do not tolerate mFRET and are constrained to using specialized dyes and cytometers that minimize the spectral overlap, and employing lengthy trial-and-error optimization protocols to experimentally determine the barcode-specific dye ratios.

Here, we introduce a platform for (i) rCR-free multiplexing of sandwich assays on microparticles with (ii) mFRET-tolerating high-capacity barcoding. Building on the work of Förster, we first introduce an ensemble multicolour FRET (emFRET) model that can accurately predict the ensemble spectral profiles of any combination of dyes. This method was developed hand-in-hand with a facile, proportional microparticle labelling method that enables accurate and robust generation of spectral barcodes. To achieve proportional labeling, DNA oligonucleotides were conjugated to the dyes and served as chemically-homogeneous cross-linkers. The emFRET model enabled *in silico* design of 580 distinguishable barcodes using dyes with strong spectral overlap (FAM, Cy3, Cy5, Cy5.5), and with inter-dye FRET efficiency reaching up to 0.76, and which were tested experimentally. The emFRET also allowed robust and fully-automated decoding without need for manual calibration. Taken together, the results established a platform for rapid, high-capacity microparticle barcoding using common dyes, and fully automated decoding using commonly used cytometers, and which can be used for immunoassays.

Next, we developed a cross-reactivity-free MSA on microparticles termed "colocalization-by-

linkage assay on microparticles" (CLAMP). In our approach both members of an antibody pair are pre-immobilized to the surface of micron-sized microparticles, with one of the members tethered via a cleavable DNA linker. Different microparticle sets can be prepared, each containing both affinity reagents for specific detection of their target analyte. Importantly, only complete protein-antibody sandwich complexes on each bead are labeled using a DNA-displacement strategy, prior to rapid read-out of CLAMPs by flow cytometry. The CLAMP assay was optimized by tuning the valency and surface density of antibody conjugates, and was shown to eliminate rCR from microparticle-based MSAs. The CLAMP assay holds potential as the first scalable MSA on microparticles, and paves the way for simple, rapid, and high-throughput MSAs. By pre-assembling antibody pairs, this immunoassay concept represents a departure from typical immunoassays, and offers the possibility to design the multivalent sensors *a priori*, which could enable higher sensitivity assays.

In this dissertation, we have presented an integrated platform that addresses two of the most important and highly sought-for scaling challenges for MSAs. These results set the foundation for the next-generation of protein assays that can be multiplexed while maintaining the excellent sensitivity and specificity of the single-plex ELISAs. Importantly, the approach taken in this work is cost-efficient and can be immediately deployed, enabling application in large population-wide studies and promising to meet the increasing demands in precision medicine.

RESUMÉ

Le dosage immunologique est une méthode de référence pour la recherche fondamentale et clinique, ainsi que pour le diagnostic clinique. En particulier, les dosages immunologiques en sandwich, basés sur une double liaison des anticorps, offrent une excellente spécificité et sensibilité de détection ainsi qu'un rapport coût-efficacité élevé, permettant la mesure de faibles concentrations de protéines dans le sérum ou le plasma. Les microparticules à codes-barres fluorescents sont un format populaire pour les dosages immunologiques car elles permettent le multiplexage des analyses avec un rendement élevé et une lecture rapide par cytométrie. Cependant, ce format largement utilisé souffre de deux restrictions majeures qui limitent l'évolutivité et l'utilisation des analyses à base de microparticules. En premier lieu, la spécificité et la sensibilité des dosages multiplexés en sandwich (MSA) sont fortement entravées par l'augmentation du multiplexage en raison de la réactivité croisée, communément rencontrée avec les anticorps de détection. Cette réactivité croisée induite par les réactifs (CRr), a lieu dans tous les MSA qui appliquent les anticorps de détection sous forme de mélange. Alors que CRr peut être significativement réduite dans les dosages plans ou en phase liquide, ce phénomène reste le principal obstacle à l'augmentation du multiplexage dans les MSA à base de microparticules. Le second défi est la génération d'un grand nombre de microparticules spectralement distinctes nécessaires au codage à barres. Typiquement, des codes-barres spectraux uniques sont générés en modulant les proportions de sondes multicolores à la surface des microparticules. Cependant, les cascades de transfert d'énergie de résonance de Förster multicolores (mFRET) entraînent des variations imprévisibles des intensités d'émission, réduisant alors la résolution des codes-barres. Ainsi, le mFRET est une limitation majeure des méthodes actuelles de codage spectral qui les contraint à utiliser des colorants et des cytomètres spécialisés pour minimiser le chevauchement spectral, et à employer de longues approches d'optimisation par tâtonnement (essai-erreur) pour déterminer expérimentalement les ratios de colorants spécifiques des différents codes-barres.

Nous présentons ici une plateforme pour (i) le multiplexage sans rCR des dosages en sandwich sur microparticules ainsi (ii) qu'un codage à barres à haute capacité de multiplexage tolérant le mFRET. En s'appuyant sur les travaux de Förster, nous introduisons d'abord un modèle d'ensembles multicolores FRET (emFRET) capable de prédire avec précision les profils spectraux totaux de toutes combinaisons de colorants. Ce modèle a été développé en parallèle d'une nouvelle méthode de marquage de microparticules facile et proportionnelle, qui permet la génération précise et robuste de codes-barres spectraux. Le marquage proportionnel est obtenu au moyen d'oligonucléotides d'ADN conjugués aux colorants et utilisés comme agents de réticulation chimiquement homogènes. Le modèle emFRET a permis la conception in silico de 580 codes-barres distincts utilisant des

colorants à fort recouvrement spectral (FAM, Cy3, Cy5, Cy5.5) dont l'efficacité FRET inter-colorée atteignait 0,76 et qui ont été ensuite validés expérimentalement. L'emFRET a également permis un décodage robuste et entièrement automatisé sans nécessiter de calibrage manuel. Ensemble, ces résultats ont conduit à l'établissement d'une plateforme de codage à microparticules rapide et à forte capacité de multiplexage, utilisant des colorants classiques et un décodage entièrement automatisé à l'aide de cytomètres standards, et pouvant être appliquée pour des dosages immunologiques.

Ensuite, nous avons développé un MSA sans CRr sur des microparticules appelés "test de colocalisation par liaison sur microparticules" (CLAMP). Dans notre approche, les deux membres d'une paire d'anticorps sont pré-immobilisés à la surface de particules micrométriques, l'un des membres étant attaché via un brin d'ADN clivable. Différentes populations de microparticules peuvent être créées, chacune contenant les deux réactifs d'affinité pour une détection spécifique de leur analyte cible. Il est important de noter que seuls les complexes protéine-anticorps en sandwich complets sur chaque particule sont marqués en utilisant une stratégie de déplacement de l'ADN, avant une lecture rapide par cytométrie en flux. Le CLAMP a été optimisé en ajustant la valence et la densité de surface des anticorps conjugués, permettant d'éliminer la rCR des MSA à base de microparticules. Le CLAMP est potentiellement le premier MSA évolutif sur les microparticules, ouvrant la voie à d'autres MSA simples, rapides et à haut débit. En pré-assemblant des paires d'anticorps, ce concept constitue le point de départ des dosages immunologiques typiques et offre la possibilité de concevoir les capteurs multivalents a priori, et ainsi permettre des tests à plus haute sensibilité.

Dans cette thèse, nous avons présenté une plateforme intégrée qui répond à deux des défis de dimensionnement les plus importants et les plus étudiés pour les MSA. Ces résultats servent de fondements pour la prochaine génération de tests protéiques pouvant être multiplexés tout en maintenant l'excellente sensibilité et spécificité des tests ELISA à analyte unique. Finalement, il est important de souligner que l'approche adoptée dans ce travail est économique et peut être immédiatement déployée, ce qui permet une application dans des études à grande échelle et promet de répondre aux demandes croissantes en médecine de précision.

CONTRIBUTION OF AUTHORS

In accordance with McGill University's "Guidelines for Thesis Preparation" policy, this thesis is presented as a collection of manuscripts written by the PhD candidate with the collaboration of co-authors, as detailed below. The manuscripts and thesis presented here are based on experiments designed and conducted by the candidate, in response to original research questions raised by the candidate. Additionally, the candidate analyzed and interpreted the experimental datasets, organized the results, and composed the manuscripts. Reflecting my role in the research, I am first author on all manuscripts included in this thesis.

The PhD supervisor (Prof. David Juncker) appears as a co-author on all manuscripts to reflect his supervisory role in the projects and his contribution to data interpretation and manuscript preparation. Likewise, Dr. Andy Ng, a research associate in Prof. Juncker's group, appears as a co-author on all manuscripts to reflect his contribution to research design, the interpretation of datasets, and to the editing of the manuscripts presented here. Acknowledgements to additional parties have been made in the preface to each chapter – these researchers and institutions contributed specific expertise, equipment or funding to the research presented in the chapter.

The thesis has been organized in five separate chapters, and short prefaces have been added to serve as transitions between each chapter, where appropriate.

Chapter 1 serves as an synopsis, outlining the background, motivation, hypotheses and research goals of the dissertation. The key contributions of the work and overall conclusions are also summarized.

Chapter 2 provides a broad overview to the field of proteomics. We first discuss the needs for protein analysis for faster discovery of protein biomarkers, then list the currently available proteomic tools for measuring proteins in blood, focusing on their capabilities and limitations and how these have influenced current practice in biomarker discovery.

Chapter 3 presents our manuscript in which we describe our spectral barcoding platform for microparticles. This manuscript includes a theoretical development of stochastic energy transfer between multicolour fluorescent dyes, and introduces the equations that describe this transfer. This model was used to guide the encoding and decoding of multicolour microparticles. Michael Kleinman, an undergraduate student I supervised, is included as a co-author on the manuscript to reflect his contributions towards writing the decoding algorithm. This chapter is formated as a manuscript, and has been published as an article in *Nature Nanotechnology*.

Chapter 4 presents a manuscript in which we describe a multiplexed immunoassay format that

mitigates cross-reactivity between antibody reagents. This chapter is formatted as a manuscript, and will be submitted shortly.

Chapter 5 includes a summary of the work carried out, a discussion of the main drawbacks of our current research, and an outlook section in which we discuss the potential future directions of the present work.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

PREFACE TO CHAPTER I

Microparticle-based sandwich assays provide distinct advantages over other assay platforms. However, this format remains limited in the number of proteins that can be simulatneously detected (*i.e.* multiplexing) due to two major limitations: reagent-driven cross-reactivity, and microparticle barcoding. The present thesis introduces a platform that addresses both limitations.

This first, synopsis chapter, aims to provide an overview of the present thesis. First, we begin with a brief historical account of the development of multiplexed sandwich immunoassays on microparticles. We then succinctly describe the aforementionned limitations, highlight key advances in each area, present the objective of the present thesis, describe the approach taken, and summarize the key contributions.

Chapter 1

SYNOPSIS

1.1 Multiplexed sandwich assays on microparticles

The development of the sandwich immunoassay in the 1970s [1–3] contributed to the rise of the immunodiagnostics industry [4, 5]. To this day, the sandwich immunoassay remains the gold standard for the detection of protein biomarkers [6, 7]. In its solid-phase format, a surface-bound antibody captures a target protein from a complex mixture such as serum or plasma, whereas the detection antibody recognizes and binds to a different epitope on the protein to signal detection (**Figure 1.1a**) [8]. This dual-recognition requirement in sandwich assays is a proofreading measure that ensures that a single non-specific binding event does not result in a false-positive, and affords the sandwich immunoassay its high sensitivity and specificity.

In a single-plex assay such as ELISA, capture antibodies are typically coated on the inner walls of a microtiter-plate wells. The use of antibody-coated microspots [9] or microparticles [10] as small, discrete capture 'zones' was demonstrated before the advantages of this miniaturization were understood. In 1989, Ekins *et al.* demonstrated that miniaturization typically ensures that only a negligible fraction of the targets are captured at equilibrium. Assay operation within this so-called 'ambient analyte regime' generates a signal that is only dependent on the analyte concentration, and thus significantly improves the quantitative robustness of the assay [11]. Furthermore, Ekins *et al.* noted that miniaturization using microspots is amenable for running multiple assays in parallel, *i.e.* multiplexing [12]. Indeed, based on this principle, DNA microarrays were implemented a few years later [13], followed by protein microarrays several years after [14].

In comparison to planar microarrays, assays on microparticles suspended in solution afford many advantages in terms of flexibility, ease-of-handling, batch functionalization, improved mass transport, and statistical robustness [15–18]. However, as opposed to planar microarrays, the multiplexing of microparticle-based assays requires a means to track and distinguish every reaction from solution-that is, microparticles must be somehow encoded and then decoded after read-out. The advent and rapid development of multiparametric flow cytometers provided a practical and rapid mean to optically interrogate microparticles [19], and they were quickly applied as assay carriers [10]. Early multiplexed immunoassays on microparticles utilized up to 4 different particle sizes as barcodes, which were distinguishable in the scattering channel on the cytometer [20]. Later, detection antibodies labeled with different dyes were also used to detect proteins in multiplex form [21]; however, this method is ill-suited for barcoding as the intensity of the code is directly dependent on

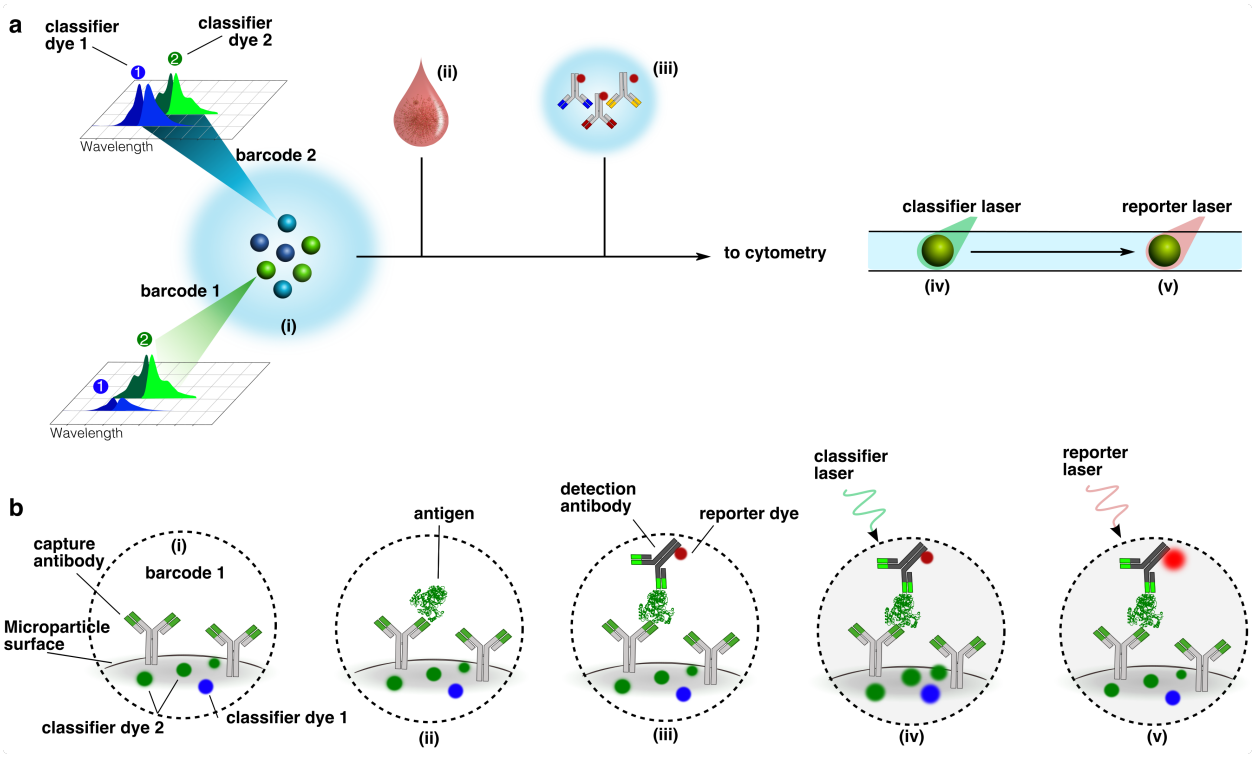


Figure 1.1: **The multiplexed sandwich immunoassay on spectrally-encoded microparticles.** (**a**) Antibody-coated microparticles are spectrally-encoded using unique proportions of different classifier dyes. The microparticle mixture (i) is incubated with the bio-fluid sample (ii), followed by incubation with a detection antibody cocktail (iii) and read-out by cytometry using a classifier laser (iv) and a reporter laser (v). (**b**) Schematic showing the microparticle surface during the stages of the multiplexed sandwich assay (i)-(v).

the concentration of the analyte.

The introduction of spectral barcoding in the late 1990s (**Figure 1.1a**), where microparticles of the same size are dyed *a priori* so as to yield a spectrally distinct and measureable signature irrespective of the assay signal (**Figure 1.1b**) , was a quantum leap in microparticle barcoding [22]. In the first demonstration of this kind, Fulton *et al.* impregnated 5.5 μm microparticles with two dyes (termed classifier dyes) in precise proportions so as to yield a unique overall or 'ensemble' fluorescent profile–a barcode–that can be distinguished and measured by a cytometer. The assay would then use a differently-colored reporter dye for assay signal, which can be independently interrogated by the cytometer (**Figure 1.1c**), and employ an additional scatter detector that provides size-discrimination to exclusively analyze single-bead events (so-called 'gating'). This approach provided the conceptual basis for the xMAP™ platform, and continues to be the predominant method for microparticle barcoding [23, 24].

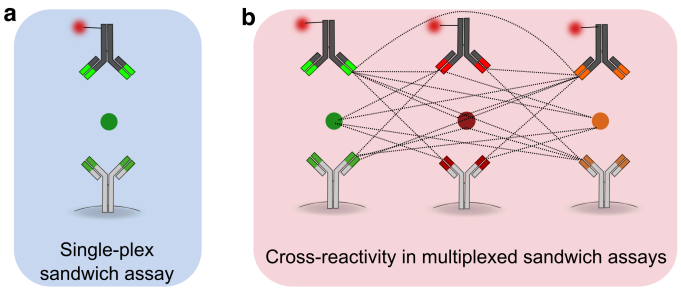


Figure 1.2: **Reagent-driven cross-reactivity in MSAs.** (**a**) The single-plex sandwich immunoassay provides exquisite sensitivity and specificity due to the dual-binding requirement. (**b**) The dual-binding requirement is lost in multiplexed sandwich assays due to the mixing of reagents, which creates vulnerability to reagent-driven cross-reactivity (lines) and generates false-positivies.

1.2 Multiplexed sandwich assays on microparticles: two shortcomings

As we review in **Chapter 2**, sandwich immunoassays on microparticles provide the sensitivity, speed, cost-efficiency, and sample-throughput to measure low-abundance proteins from a large number of samples. However, while it would be desirable to be able to measure all proteins for which antibody pairs are available in a single assay, the largest multiplex immunoassays on microparticles to date are limited to 50-60 proteins. As we discuss in **Chapter 2**, this goal has been stymied by two technical challenges: (i) mitigating the incompatibility and cross-reactivity of antibody reagents when used in a multiplexed format, and (ii) developing a robust, high-capacity barcoding platform that is compatible with high-throughput analysis by cytometry.

1.2.1 Cross-reactivity in multiplexed immunoassays: advances and limitations

Since their development, multiplexed sandwich assays (MSAs) promised running hundreds of ELISAs in parallel. However, they have been limited to a few tens of targets because conventionally, detection antibodies are added as a cocktail mixture, which gives rise to a new form of cross-reactivity between the reagents (**Figure 1.2**). As we review in Chapter 2, this reagent-driven cross-reactivity (rCR) is typically avoided by assay manufacturers through combinatorial screens that aim to optimize each panel by removing and replacing cross-reacting antibodies [25]. Importantly, these combinatorial screens are lengthy, scale quickly, and are virtually impossible to perform beyond 40-50 targets, rendering rCR the current bottleneck to multiplexing.

Two strategies to mitigate rCR have been implemented for planar and solution-phase assays respectively [26]. The solution-phase assays proposed by Landegren *et al.* is a clever solution that employs DNA oligonucleotide-antibody conjugates that are designed to hybridize and transduce a signal only when two matched antibodies bind their target, therefore discriminating against rCR

(**Section 2.2.3**) [27, 28]. One of their variations on this assay, termed proximity extension assay (PEA), is now commercialized as Proseek™ by Olink Biosciences which offers 14 different 92-plex panels covering 981 proteins overall (www.olink.com, [29]). However, the scalability of the PEA assay is limited by the need for separately reading out every target by qPCR, necessitating a microfluidic equipment to fractionate each sample into 92 separate qPCR reactions for read-out.

On the other hand, the antibody-colocalization approach taken by our group eliminated rCR from occurring in the first place [30]. As mentioned, rCR occurs because reagents, specifically detection antibodies, are allowed to interact in solution. The antibody colocalization microarray (ACM), eliminates mixing through precise delivery of each detection antibody set atop its cognate capture antibody spot–this was performed by our group using double printing [30] and snap-chip devices [31, 32], while others used two-phase aqueous systems [33]. However, the adoption of this method has been hindered by the need for precise fluidic delivery equipment, and the limited robustness and sample throughput of antibody microarrays.

In contrast, current microparticle-based platforms have failed to address rCR and as a consequence are severely limited in the number of proteins that can be measured in multiplex, ranging between 1-50 proteins, and in most commercial assays limited to less than 20 [23]. Despite this severe limitation in multiplexing, microparticle-based MSAs continue to be the predominant format, owing to its flexibility, robustness, and high sample-throughput.

1.2.2 Microparticle barcoding: advances and limitations

While microparicles are the predominant multiplexing format for MSAs, barcoding remains a significant challenge for use in assays. Indeed, despite the availability and widespread use of multiparametric flow cytometers with expanding numbers of lasers and a wide range of suitable dyes [34, 35], virtually all microparticle-based assays in-use today are spectrally encoded using proprietary dyes and readout on dedicated equipment.

In theory, the use of 4 classifier dyes encoded with 10 intensity levels each would yield $10^4 = 10,000$ barcodes [36]. However, due to the wide spectral response of fluorescent dyes, spectral overlap becomes unavoidable as the number of dyes increases beyond 2 or 3 (**Figure 1.3a**), limiting the orthogonality of the classifier dyes. Furthermore, owing to dense packing of fluorophores on the microparticles, interactions such as Förster resonance energy transfer (FRET) arise between all the classifier dyes, obfuscating the intended barcode response (**Figure 1.3b**). Increasing the barcoding capacity beyond a few dozen barcodes is typically met with rapidly intensifying multicolour FRET (mFRET), entangling the desired barcodes and leading to indistinguishable (*i.e.* non-functional) barcodes. In the absence of a theoretical model to predict mFRET and compensate for it, the ensemble fluorescence signal remains unpredictable and intractable. As a result, barcodes need to

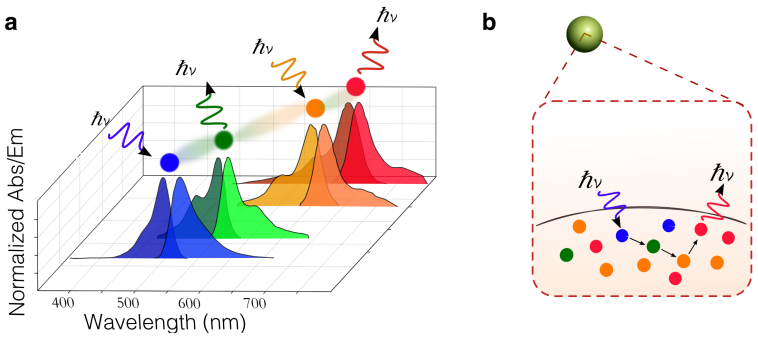


Figure 1.3: **Multicolour Förster resonance energy transfer (mFRET) between classifier dyes.** (**a**) Spectral overlap between classifier dyes results in (**b**) mFRET and mFRET cascades.

be painstakingly optimized in iterative trial-and-error experiments, which need to be repeated from scratch for every modification as is discussed in **Chapter 3**.

Approaches for barcoding in the past have focused on mitigating FRET. The enabling feature in the seminal work by Fulton *et al.* in 1997 [37] was the use of dyes that do not engage in FRET; this is achieved using proprietary dyes with large Stokes shifts and rationally designed spectra. To generate 500 barcodes, the current xMAP platform uses three dyes extending in the NIR region, which are all excitable by a single red laser [38]. However, further expansion of the barcoding capacity using the same approach would require infrared-emitting dyes with a substantially large Stokes shift. As we discuss in **Chapter 2**, many workarounds have been proposed, but despite 20 years of development, there continues to be no clear alternative for organic dyes as classifiers, which are ubiquitous in life sciences, nor a clear solution to mitigate mFRET to expand the barcoding capacity without compromising sample throughput.

1.3 Objective and rationale of the present dissertation

The overarching aim of the present thesis is to implement a microparticle-based sandwich immunoassay platform with scalable multiplexing and throughput. To this end, our specific aims are to develop (i) a multiplexed assay format that mitigates rCR in MSAs, and (ii) a barcoding method hand-in-hand with a theoretical model for mFRET to guide the encoding and decoding process and increase the barcoding capacity.

1.4 Summary of the thesis structure and contributions

In **Chapter 2**, we describe the needs for protein quantification in biology and medicine and review the established and emerging proteomic technologies, focusing on protein quantification in the context of bio-fluids such as plasma. We then detail and compare the technical attributes of different technologies. Finally, we review and discuss how the current technology landscape

may have negatively impacted outcomes in biomarker and drug discovery by enforcing unwanted trade-offs.

In **Chapter 3**, we describe a scalable, open-platform spectral barcoding method using common dyes. As opposed to avoiding mFRET, we aimed to model the interaction in order to guide the encoding and decoding. To this end, we describe an ensemble multicolour FRET (emFRET) model with an accessible analytical solution. Building on the emFRET model, we derive a generalized fluorescence model that links the ensemble fluorescence measurements on the cytometer to the dye proportions on the surface of the microparticles. Prediction of the ensemble fluorescence was thus contingent on accurately controlled dye proportions–which we achieved via a one-pot multicolour encoding strategy using DNA oligonucleotides as homogeneous cross-linkers. Performing cytometric analysis of multicoloured microparticles, we validated the model experimentally which enabled understanding, prediction, and design of stochastic mFRET cascades. We applied the emFRET model to perform multicolour barcoding and automated decoding in regimes with extreme FRET levels. This approach enabled rapid design of 580 barcodes with precision, demonstrating accurate fluorescent responses, and enabling fully automated decoding. We utilized our platform to perform a 35-plex immunoassay that demonstrated the extent of rCR in MSAs. Altogether, these advancements enable (i) expansion of the barcoding capacity to previously unattainable high FRET regimes, (ii) the use of common, spectrally overlapping dyes for barcoding, and open-platform cytometers for read-out.

In **Chapter 4**, we developed a rCR-free MSA, termed "Colocalization-by-Linkage Assays on MicroParticles" (CLAMP) that avoids mixing of different antibodies by pre-immobilizing the antibodies on the microparticles. In CLAMP, both members of an antibody pair are attached to the surface of micron-sized beads, with one of the antibodies tethered via a cleavable DNA linker. We also developed a robust signal transduction mechanism that ensures that only complete protein-antibody sandwich complexes on each bead are labeled and result in a signal, which can be readily measured by flow cytometry. We optimized the CLAMP assay by tuning the valency and the surface density of antibody oligonucleotide conjugates, and demonstrate rCR-free multiplexing. By pre-assembling antibody pairs, this immunoassay concept represents a departure from typical immunoassays, and offers the possibility to design the multivalent sensor a priori, which could enable higher sensitivity assays.

In **Chapter 5**, we discuss the significance of the present work as well as its limitations, and discuss possible next steps towards scaling up the CLAMP platform.

References

1. Wide, L., Bennich, H. & Johansson, S. Diagnosis of Allergy By an in-Vitro Test for Allergen Antibodies. *The Lancet* **290,** 1105–1107 (1967).
2. Salmon, S. E., Mackey, G. & Fudenberg, H. "Sandwich" Solid Phase Radioimmunoassay for the Quantitative Deteriminaton of Human Immunoglobulins. *The Journal of Immunology* **103,** 129–135 (1969).
3. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry* **8,** 871–874 (1971).
4. Wide, L. Inventions leading to the development of the diagnostic test kit industry - From the modern pregnancy test to the sandwich assays. *Upsala Journal of Medical Sciences* **110,** 193–216 (2005).
5. Borrebaeck, C. A. Antibodies in diagnostics - From immunoassays to protein chips. *Immunology Today* **21,** 379–382 (2000).
6. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome. *Molecular & Cellular Proteomics* **1,** 845–867 (2002).
7. Anderson, N. L. The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry* **56,** 177–185 (2010).
8. Wild, D. *The Immunoassay Handbook* (ed Wild, D.) 48–57 (Oxford, 2013).
9. Feinberg, J. G. & Wheeler, A. W. Detection of auto-immune antibody and tissue antigens by the 'microspot' technique. *Journal of Clinical Pathology* **16,** 282–284 (1963).
10. Horan, P. K. & Wheeless, L. L. Quantitative Single Cell Analysis and Sorting. *Science* **198,** 149–157 (1977).
11. Ekins, R. P. Multi-analyte immunoassay. *Journal of pharmaceutical and biomedical analysis* **7,** 155–168 (1989).
12. Ekins, R., Chu, F. & Biggart, E. Development of microspot multi-analyte ratiometric immunoassay using dual fluorescent-labelled antibodies. *Analytica Chimica Acta* **227,** 73–96 (1989).
13. Pease, A. C. *et al.* Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 5022–6 (1994).
14. MacBeath, G. Protein microarrays and proteomics. *Nature Genetics* **32,** 526–532 (2002).
15. Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature reviews. Drug discovery* **5,** 310–20 (2006).
16. Chandra, H., Reddy, P. J. & Srivastava, S. Protein microarrays and novel detection platforms. *Expert review of proteomics* **8,** 61–79 (2011).
17. Nolan, J. P. & Sklar, L. a. Suspension array technology: evolution of the flat-array paradigm. *Trends in biotechnology* **20,** 9–12 (2002).

18. Wilson, R., Cossins, A. R. & Spiller, D. G. Encoded microcarriers for high-throughput multiplexed detection. *Angewandte Chemie International Edition* **45,** 6104–6117 (2006).

19. Shapiro, H. M. *Practical Flow Cytometry* (Wiley, Hoboken, 2003).

20. McHugh, T. M. Flow Microsphere Immunoassay for the Quantitative and Simultaneous Detection of Multiple Soluble Analytes. *Methods in Cell Biology* **42,** 575–595 (1994).

21. Collins, D. P., Luebering, B. J. & Shaut, D. M. T-lymphocyte functionality assessed by analysis of cytokine receptor expression, intracellular cytokine expression, and femtomolar detection of cytokine secretion by quantitative flow cytometry. *Cytometry* **33,** 249–255 (1998).

22. Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J. & Kettman, J. R. Advanced multiplex analysis with the FlowMetrix system. *Clinical Chemistry* **43,** 1749–1756 (1997).

23. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplex era: Potentials and pitfalls. *Proteomics - Clinical Applications* **9,** 406–422 (2015).

24. Lin, G., Baker, M. A., Hong, M. & Jin, D. The Quest for Optical Multiplexing in Biodiscoveries. *Chem* **4,** 997–1021 (2018).

25. Neiman, M. & Högskolan, K. T. *Bead based protein profiling in blood* PhD thesis (2013).

26. Juncker, D., Bergeron, S., Laforte, V. & Li, H. Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. *Current Opinion in Chemical Biology* **18,** 29–37 (2014).

27. Lundberg, M. *et al.* Multiplexed homogeneous proximity ligation assays for high-throughput protein biomarker research in serological material. *Molecular & cellular proteomics : MCP* **10,** M110.004978 (2011).

28. Nong, R. Y., Gu, J., Darmanis, S., Kamali-Moghaddam, M. & Landegren, U. DNA-assisted protein detection technologies. *Expert review of proteomics* **9,** 21–32 (2012).

29. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS ONE* **9** (2014).

30. Pla-Roca, M. *et al.* Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & cellular proteomics : MCP* **11** (2012).

31. Li, H., Bergeron, S. & Juncker, D. Microarray-to-Microarray Transfer of Reagents by Snapping of Two Chips for Cross-Reactivity-Free Multiplex Immunoassays. *Analytical chemistry* **84,** 4776–83 (2012).

32. Li, H., Munzar, J. D., Ng, A. & Juncker, D. A versatile snap chip for high-density subnanoliter chip-to-chip reagent transfer. *Scientific Reports* **5,** 1–12 (2015).

33. Frampton, J. P. *et al.* Aqueous two-phase system patterning of detection antibody solutions for cross-reaction-free multiplex ELISA. *Scientific reports* **4,** 4878 (2014).

34. Telford, W. G. Near infrared lasers in flow cytometry. *Methods* **82,** 12–20 (2015).

35. Wilson, A. C. *et al.* A Scalable Pipeline for High-Throughput Flow Cytometry. *SLAS Discovery* **23,** 708–718 (2018).

36. Chan, B. M., Badh, A., Berry, K. A., Grauer, S. A. & King, C. T. Flow Cytometry-Based Epitope Binning Using Competitive Binding Profiles for the Characterization of Monoclonal Antibodies against Cellular and Soluble Protein Targets. *SLAS Discovery,* 1–11 (2018).

37. Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J. & Kettman, J. R. Advanced multiplexed analysis with the FlowMetrix(TM) system. *Clinical Chemistry* **43,** 1749–1756 (1997).

38. *Luminex Kit Finder* `https://kitfinder.luminexcorp.com/`. Accessed: 2018-08-10.

PREFACE TO CHAPTER II

In the previous chapter, we provided a brief overview of the present thesis, highlighting the technical challenges in microparticle-based MSAs, conveying our overall rational, and summarizing our contributions.

In this chapter, we look to shed light on the analysis of proteins from bio-fluids to try and situate the challenges in multiplexed immunoassays within the field as a whole to identify the remaining obstacles necessary for bridging the gap between fundamental proteomics and clinical diagnostics or drug development. We first discuss the motivation and need for protein analysis in bio-fluids and plasma (Section 2.1), then review the established and emerging tools in bio-fluid based proteomics (Section 2.2), and finally detail how their limitations have enforced unwanted trade-offs in applicaitons such as biomarker detection (Section 2.3).

Chapter 2

PROTEOMICS: TECHNOLOGIES, APPLICATIONS, AND FUTURE DIRECTIONS

2.1 Proteomics for Precision Medicine: Promise and Status

2.1.1 Data-driven biology

Modern molecular biology views the paths from genes to proteins as complex interwoven networks of molecular pathways and systems that collectively explain the emergent behavior of cells, tissues, organs, and organisms [1]. Indeed, biological systems are multiparametric, complex, and dynamic. In the face of such complexity, the classical approach to the analytical interrogation of biological questions calls for dividing complex intractable problems until they are small, measurable, and tractable [2]. This so-called "reductionist" methodology focuses on finding mechanistic models to the biological elements, and is essential for applications such as mechanism-based drug discovery [3]. While the reductionist approach is a proven methodology, technological advances in the last decade have converged to allow for a complementary approach whereby biological systems can be probed in a more comprehensive manner. For instance, several emerging disciplines aim to target all the molecules pertaining to a certain class (*i.e.* omics). For the study of nucleic acids, the combination of polymerase chain reaction (PCR), DNA synthesis, and micro- and nano-engineering enabled scientists to shift the focus from individual genes (genetics) to entire genomes, spawning the fields of genomics and transcriptomics [4, 5]. Similarly, mass spectrometry (MS) has matured to allow the analysis of a substantial portion of the human proteome and metabolome, giving rise to proteomics [6, 7] and metabolomics [8], respectively. Information technologies are continuously making staggering improvements in storage and computation, allowing adequate analysis of extremely large datasets; notably, advanced machine learning algorithms now provide analytic tools to probe large data sets without a priori assumptions [9–11]. Together, these advancements are inspiring new strategies for biological inquiry, and several new approaches have been gaining momentum. For example, "systems biology" is a discipline that calls for analysis of all the molecular components of a certain biological system to uncover emergent properties through the combination of mathematics, computation, and empiric observation [2, 12, 13]. On the medical side, personalized and precision medicine argue that maximal effectiveness in disease prevention and treatment is best reached when considering the precise molecular, environmental, and behavioral variability between individuals [14–17]. In fundamental biology, researchers are now able to incorporate and analyze -omics and multi-omics within their methodology [18–20]. Despite having different end-goals, these endeavors possess a common denominator that calls for the capture and analysis

of comprehensive and multi-layered datasets in the aim of refining, and perhaps re-defining, our understanding of wellness and disease.

2.1.2 Genomics: a case in point

Omics approaches necessitate the generation of a large amount of data to uncover insights. To this end, the development of data-efficient technologies is critical. For instance, genomics' rapid adoption by biologists and clinicians has been largely fuelled by the advent of cost-efficient, high-throughput technologies that lowered entry barriers and enabled new applications, some of which saw quick translation to the clinic. Furthermore, adoption and implementation of genomic tools for new applications was often concurrent with an additional increase in efficiency. In a little over a decade, DNA hybridization microarrays were scaled from 256 probes [21] to 6.5 million probes per chip, prompting their application towards large-scale RNA expression analysis [22]. This platform also made it possible to detect single-nucleotide polymorphisms (SNPs) and later proved instrumental for genome-wide association studies (GWAS) that have insofar linked SNPs and other genetic variants to over 1800 diseases and traits, uncovering thousands of associations with previously unknown physiology [23–25].

Similarly, DNA sequencing famously experienced rapid adoption that was driven by monumental progress in sequencing technology that surpassed the original description of Sanger sequencing [26]. Whereas first-generation sequencing-that is, Sanger sequencing-was based on capillary electrophoresis of long, individual, labelled strands which proved difficult to scale, next-generation sequencing (NGS) platforms allowed parallel read-out of millions of DNA molecules simultaneously and resulted in a sharp rise in throughput and efficiency [27] (**Figure 2.1**). Indeed, the scaling brought by NGS was faster than the scaling trends seen in the semiconductor industry, widely known as "Moore's Law". This shift in the basic sequencing mechanism was a watershed moment that ultimately led to a 50,000-fold reduction in cost, up to 10 orders of magnitude increase in output capacity, spawned a variety of new commercial instruments [24, 28], and enabled inspiring projects such as the "1000 Genomes Project" [29]. The improvement in data-efficiency encouraged the adoption of NGS by mainstream research community, inspired new applications, and generated new biological insights [27]. Exome and genome sequencing have already been demonstrated to aid in diagnosing disease [30, 31], establishing predictive models for disease [32, 33], and understanding how genetic differences impact health and disease [28]. The use of tumour sequencing has also brought insights-for example, by providing a look at the mutations acquired from somatic cell replication in tumourigenic lesions [34]. Furthermore, sequencing has allowed looking at shorter nucleic acid fragments such as cell-free circulating tumour DNA or microRNAs, which are being considered in the early diagnosis and prognosis of cancer [35]. Transcriptomics has benefited directly from NGS, which has given researchers the ability to sequence RNA with

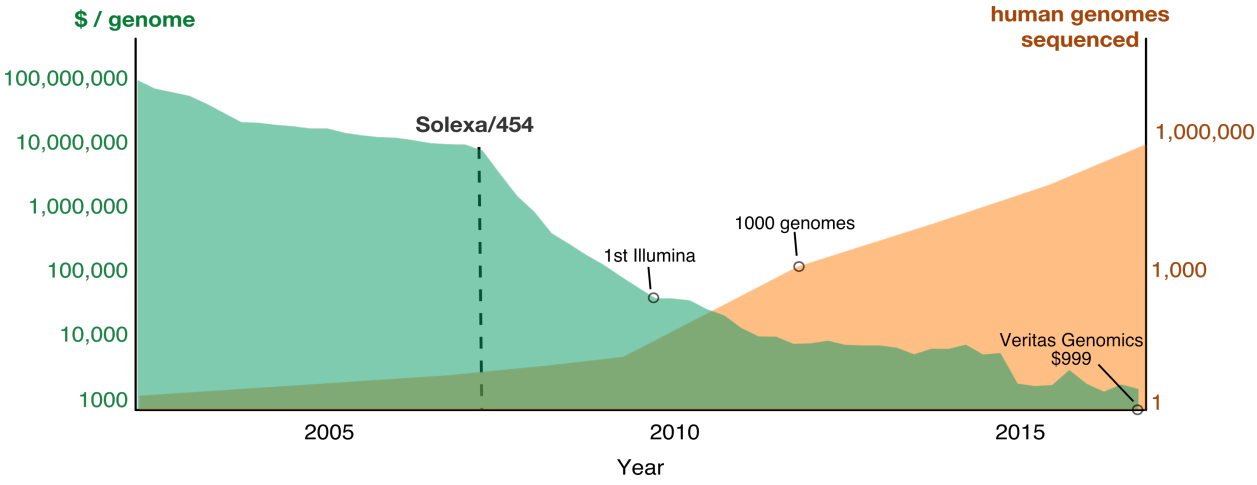


Figure 2.1: **Correlation between cost of sequencing and the number of genomes sequenced.** The cost of sequencing a human genome [39] and the number of human genomes sequences since the Human Genome Project [40]. Whereas the sequencing of the first human genome was completed in 2001, the technology used was costly and saw minimal adoption. NGS, spawned from technological innovations by Solexa (later bought by Illumina) and 454 in 2007, allowed massively parallel sequencing and brought down the effective cost of sequencing, kick-starting widespread adoption and the complete sequencing of 1000 human genomes [29].

high sensitivity as well as with single-cell resolution [36–38].

In the last decade, NGS has seen further development and evolution, affording 2-3 orders of magnitude increase in read capacity, with some newer technologies allowing increasingly long read lengths [24]. Finally, the cost of sequencing entire genomes has finally reached the long-sought $1,000 mark [1] prompting implementation in the consumer market as well as clinical applications. Today, a single technician is capable of sequencing the entire human genome for about a thousand dollars in a few days.

2.1.3 Protein analysis is indispensable to biology and medicine

DNA is, for the most part, a static information database that is blind to the complex interplay found downstream at the protein level, which are often the sole determinants for health and disease. Indeed, large-scale genotyping studies in the last decade have shown that (i) common diseases are not well explained by unique genetic variants, (ii) the same genetic variants can result in different outcomes due to various downstream factors, and (iii) even heritable diseases can only be partially explained by genetic variations [4]. Downstream from DNA, many researchers have turned to analyzing messenger RNA (mRNA) and transcription levels to understand the more dynamic cellular responses and their regulatory pathways. Whilst mRNA transcripts have been utilized as a proxy for proteins, in most situations the levels of mRNA transcripts do not correlate

with protein abundance [41, 42]. Further highlighting the importance of proteins, strong evidence has pointed that the level of protein orthologues, not transcript orthologues, is conserved across large evolutionary distances [42]. Furthermore, the gene-centric view of the "central dogma of molecular biology", has somewhat changed since its description more than 50 years ago; cleavage and alternative splicing of mRNA transcripts, post-translational modification (PTMs) of proteins, and interaction between protein and DNA or RNA all play a role in modifying the activity of a protein from what would be predicted by its sequence or transcription level. As such, it has become widely accepted that the activity and function of a protein cannot accurately be predicted from DNA or RNA [43]. In the last decade, proteomics has become a sought-after component of multi-omics analyses, with many studies using it to complement the genetic or transcriptomic data and to identify associations between genetic variants and protein levels [16, 18–20, 33, 44–46].

Protein analysis is fundamental to various applications in biology and medicine (**Figure 2.2a**). In clinical settings, proteins are the most common class of biomarkers, contributing to clinical decision making through disease diagnosis and confirmation, risk prediction, and evaluation of treatment effectiveness [47] . Indeed, 40% of blood-based laboratory-tests are run on proteins [48] (**Figure 2.2b**). For example, proteins such as troponin I and B-type natriuretic peptide are used to diagnose myocardial infarction and congestive heart failure [49], respectively. In addition, PSA and CA-125 are detected from serum for prostate, ovarian, and pancreatic cancer monitoring [47]. Diagnostic methods also rely heavily on protein detection in tissues other than blood. For example, protein markers are detected in tissue biopsies using immunohistochemistry (IHC) to support cancer diagnosis, classify neurodegenerative disease, and diagnose muscular disorders. Cell-based immunoassays measured by flow cytometry are also regularly used for diagnosing and staging hematological neoplasm, HIV infection, and chemotherapy monitoring [50].

Proteins have an especially important relevance in the pharmaceutical (pharma) and biotech industry as they are developing drugs that are either targeting proteins or are proteins themselves [51–53]. In drug discovery, the requirements of finding drugs with high efficacy and low toxicity is more demanding than biomarker discovery. This task necessitates an understanding of signalling networks underlying normal or aberrant function. Protein analysis thus underpins many stages of the drug discovery pipeline, such as (i) identification of targets, (ii) understanding mechanism of action and construction of signaling pathway, (iii) stratification of compounds, (iv) understanding drug efficacy and toxicity, (vi) repositioning drugs, and (vii) uncovering markers with prognostic value [52, 54]. For example, new monoclonal antibodies are routinely being approved for treatment of autoimmune diseases, allergies, and cancer, by targeting specific cytokine pathways within the cell; unfortunately, these drugs are prone to causing unwanted and deadly inflammatory response known as 'cytokine storm', and therefore the ability to monitor the levels of cytokines more efficiently in

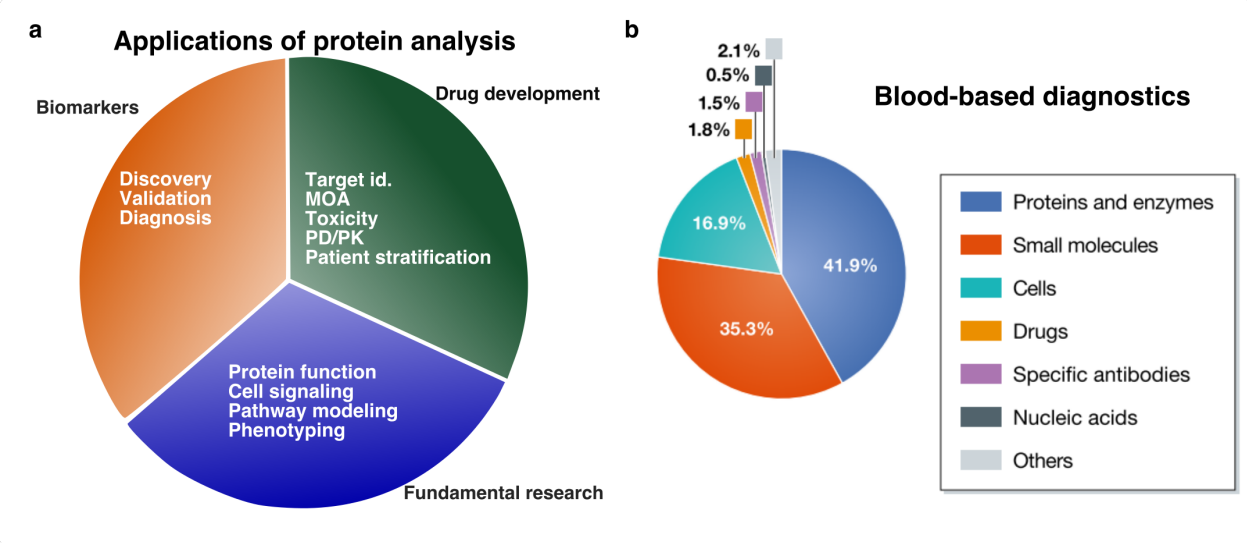


Figure 2.2: **Protein analysis: from bench to clinic.** (a) Common applications of proteins analysis divided among the three main axes of research in the life sciences (MOA: Mechanism of action, PD/PK: pharmacodynamics/pharmacokinetic). (b) The distribution of molecule types detected from blood in hospitals, clinics, and centralized laboratories (adopted from [48], licensed under CC BY 4.0).

preclinical and clinical stages is important [55].

Altogether, these applications demonstrate the importance of protein analysis for fundamental biology, drug development, and biomarker discovery. Should one have the capability to monitor every functional, structural, and interactional aspect of every protein in an organism, then it would be possible to exactly define the current health status of the organism, characterize disease networks, and identify therapeutic targets and useful biomarkers [1]. In this context, proteins are often the most relevant molecular component to characterize, which is why protein analysis underpins many applications in life sciences. However, as will be discussed in **Section 2.2.2**, proteins are complex molecules and their exhaustive characterization necessitates different approaches that are fraught with various technical challenges.

2.1.4 The elusive potential of plasma proteomics

Recent estimates place the number of protein-coding genes in the human genome at around 21,000 [56]. By taking into account alternative splicing isoforms, PTMs such as proteolytic cleavage, phosphorylation, methylation, acetylation, and glycosylation, the estimated number of unique protein variants, or proteoforms, can be as high as 1 million [57]. However, the fraction of the human proteome that ends up in circulation remains undefined today. While some argue that, in principle, all proteins end up in the blood at one point or another during the life span of an individual

[58], empirical evidence for the number of unique proteins in blood is strongly method-dependent. For instance, recent estimates using mass spectrometry place the number of canonical proteins in blood around 5400 but that does not include at least 831 proteins that are only detectable using antibodies [59]; therefore, the number of proteins in plasma is likely much higher.

Regardless, blood plays a chief role in clinical decision making and is a practical tissue for interrogation, particularly due to its accessibility and potential for repeated sampling. In addition to containing proteins that have specific functions in plasma (e.g. signalling, transport, growth, repair, and defence against infection), blood is in contact with all internal organs and hence should contain early, telltale proteomic signs of aberrant function and/or disease. Researchers have long argued that plasma has the potential to be used broadly as a liquid biopsy that can detect and diagnose disease long before clinical manifestations, and, more generally, provide systemic information about an individual at the time of blood draw [58, 60, 61]. In addition to early diagnosis, differential diagnosis, disease staging, determination of disease prognosis, patient selection in drug trials, and surrogate endpoints are all applications that could reasonably be met in the coming decades [49].

The success of the current immunodiagnostic industry ($20B market size) has motivated the search for more biomarkers and, with the advent of -omics technologies such as proteomics, this vision has attracted major efforts and investments into this space (**Figure 2.3a**). However, despite the massive efforts by the proteomics community [62], the strong potential, and the hundreds of millions of dollars invested in the last 15 years [63], the number of established protein biomarkers is only a modest subset of the pathways related to disease, and of the plasma proteome in general (**Figure 2.3b**). Today, there are approximately 115 unique proteins (excluding autoantibodies and PTMs of said proteins) targeted by FDA-cleared assays in serum or plasma in the United States, with 80% of them measured using an immunoassay [48, 64, 65]. An additional 96 unique proteins are approved as Laboratory Developed Tests (LDTs) [47, 66]. Further, most of these biomarkers have been established decades ago, and the rate of introduction of new protein biomarkers to the clinic has stagnated at an average of 1.5 new biomarkers per year [47, 64]. Multi-biomarker studies are known to have a higher power of classification [67], especially when the markers are correlated [68]. In fact, there are already two historical examples of biomarkers that rely on ratios of protein abundance, namely the ASAT/ALAT ratio for liver disease differentiation, and SFlt-1/PlGF for preeclampsia diagnosis [48]. Despite the recent push for multi-parametric biomarker studies, only the OVA1 and OVA2 ovarian cancer tests that uses multiple serum biomarkers has received 510(k) FDA clearance [67, 69].

The number of validated and approved protein biomarkers represents only 1% of the total human proteome (**Figure 2.4a**). Many researchers believe it is likely that many more proteins than currently discovered and used could be implemented to improve diagnostics [65, 70]. As a result, the current

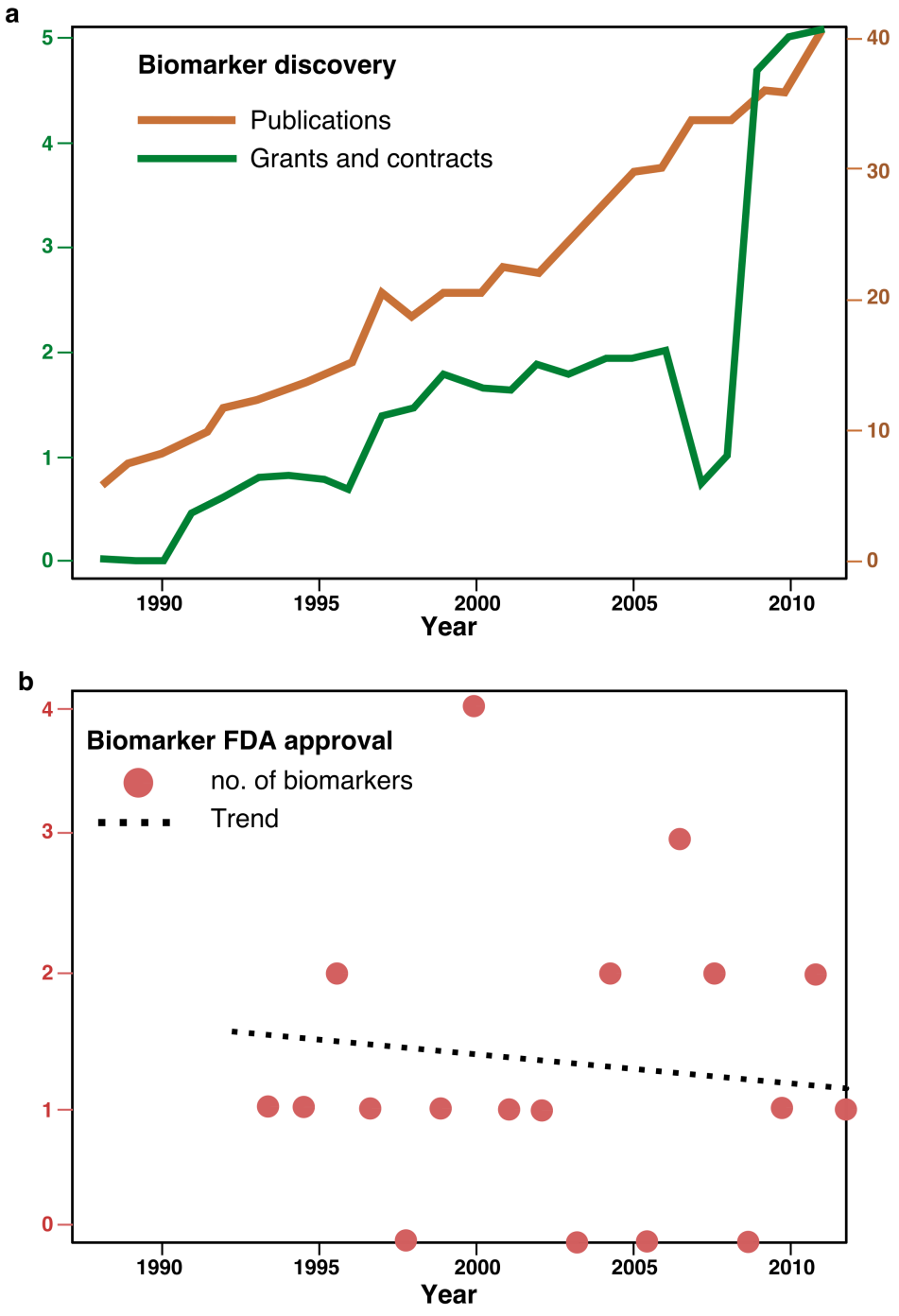


Figure 2.3: **Biomarker discovery timeline.** (a) The number of grants and publications performing biomarker discovery in the period spanning 1988-2012 [64]. (b) FDA clearance of protein-based assays between 1993-2012 [64].

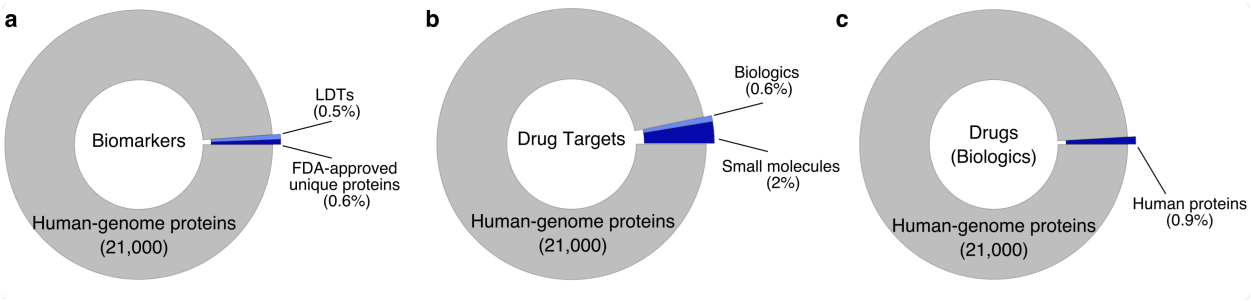


Figure 2.4: **The fraction of proteins used in diagnostics and therapeutics.** (a) The fraction proteins used as biomarkers. (b) The fraction of proteins used as drug target by small molecules or other proteins (biologics). (c) The fraction of proteins used as drugs [53].

landscape of diagnostic tests is clearly growing too slowly to meet the increasing demands from health care. Meanwhile, diseases and conditions that would benefit from early diagnostic markers such as Alzheimer disease, any and all cardiovascular malfunctions, chronic obstructive pulmonary disease, strokes, and cancer, continue to lack adequate diagnostic tests.

This large discordance between the efforts and outcomes in biomarker discovery has been a focus for some time [70, 71]. Small sample sizes, regulatory hurdles, inter-individual variability and the focus on single biomarkers are but some of the aspects of current biomarker discovery pipelines that have been called into question [48, 62, 65, 70, 72]. However, an exploration of the potential limitations is incomplete without a discussion of the technologies used, which will be presented in **Section 2.2**.

In drug development, proteins have always been a preferred drug target, and represent the majority of current drug targets [53]. In the last decade, proteins (including antibodies) have risen to prominence as therapeutics themselves. However, the number of approved drugs has not increased commensurate to R&D spending in pharma, and the number of unique proteins used as drug targets or drugs themselves is 2.6% and 0.9% of total human proteome, respectively [53] (**Figure 2.4b-c**).

Again, our shortcoming in understanding and applying proteomics in plasma may be better understood in light of the available technologies. In the following sections, we will outline the most common contemporary proteomics methods used in plasma and other bio-fluids, focusing on the limitations and trade-offs of each technology; this will be followed by a perspective on how the limitations enforced non-ideal studies and discovery pipelines, and how emerging methods may address these challenges.

2.2 Proteomic Analysis Toolbox

2.2.1 Branches of proteomics and scope

Proteomics presents formidable challenges because proteins function in multiple dimensions consisting of structural, post-translational, interactional, and spatial. Hence, arriving to an in-depth understanding of the protein function requires probing each dimension, ultimately resulting in multi-dimensional datasets [7]. This heterogeneity of protein properties has resulted in several major and overlapping branches of study and many combinations thereof, namely structural proteomics, expression proteomics, and functional proteomics. Briefly, structural proteomics aims to resolve protein structure to understand and predict function, this discipline typically relies on methods such as X-ray crystallography and nuclear magnetic resonance and has become paramount for drug development. Functional proteomics' aim is to understand protein interactions with its molecular environment-methods used include yeast two-hybrid systems that reveal binary interactions to understand a protein's function within regulatory pathways. Finally, expression proteomics, which is a major portion of proteomic analyses, focuses on measuring the abundance or expression levels of proteins within specific tissue such as the blood proteome, cell lysates, or cell membranes. The abundance of proteins can be tightly controlled in a biological system and therefore expression levels can be associated with alternative states such as health and disease.

The remainder of this chapter will be focused on expression proteomics. In particular, we will focus on conventional and emerging strategies for protein analysis in bio-fluids in the context of fundamental research, clinical research and clinical diagnostics. The tools and techniques will be first briefly described and compared. The practical capability and importance of each method will be elucidated by focusing and discussing metrics such as analytical sensitivity, analytical specificity, and assay throughput. Strictly speaking, analytical sensitivity is the ability to discriminate between different concentrations (and indeed, the presence or the absence) of the analyte; however, as is common, we will use the term *sensitivity* to refer to the lowest concentration an analyte may be detected with confidence. Importantly, the discussion will be limited to established techniques and tools that have been validated either through successful commercial development, approval by a regulatory body for clinical use, or adoption by Contract Research Organizations (CROs).

2.2.2 Challenges of protein measurement

Protein structure. Proteins are composed of polypeptide chains that self-assemble into complex and functional 3D structures, held together by several types of bonds including covalent and non-covalent (*i.e.* hydrogen bonds, ionic bonds, and hydrophobic interaction). Strictly speaking, the folding of a protein is the direct result of the thermodynamic free-energy that is gained and is highly dependent on the environment such as the solvent, the pH, ionic strength, and temperature. Changes in these parameters often result in a change of a protein's tertiary structure and could lead to altered

function or complete denaturation. Furthermore, different splice variants of the same protein-encoding gene as well as different PTMs of that same protein can result in profoundly different functions which may be important to consider and differentiate clinically [73]. Depending on the method used to detect and measure proteins, these structural properties could pose an important challenge to accurate quantification, as will be discussed later. Finally, analysis of proteins is at a disadvantage compared to nucleic acids because of the lack of a singular rule, akin to the Watson-Crick base-pairing, that can be used for detection.

The challenging concentration range of the plasma proteome. Measurement of proteins from plasma is perhaps most challenging because of the matrix's constitution (**Figure 2.5**). Plasma proteins can be divided into three classes which are well stratified by their concentrations in blood. The first class of proteins are the functional plasma proteins; these proteins perform their function in circulation and represent a large majority of the plasma proteome, including proteins such as serum albumin, various immunoglobulins of the innate immune system, and proteins of the coagulation cascade. Indeed, only a handful of the most abundant functional proteins represent about 85% of the total protein mass in plasma. The second class is leakage proteins without a known function in circulation (*e.g.* enzymes used to diagnose liver disease) this class contains the largest number of unique proteins circulating in plasma. The third and last known class of proteins in circulation are hormones and signalling proteins such as cytokines, which are typically present at very low concentrations at steady state and upregulated when needed [58]. This results in immense differences in protein concentrations, with an estimated total concentration range greater than 10^{12} [74]. To detect proteins present in low abundance, which are expected to provide meaningful biological information from a diagnostic stand-point [49, 75], and distinguish them from the complex mixture, highly sensitive, specific, and reproducible technologies are required. To put this into perspective, finding a low-abundance protein (*e.g.* troponin) among albumin molecules can be compared to finding a specific individual while searching the entire human population [76]. Furthermore, the difficulty of detecting low-abundance proteins is exacerbated by the lack of a protein amplification method akin to PCR for DNA.

Variability amongst individuals. Another challenge for measuring protein abundance is biological variability. The abundance of certain proteins in blood varies significantly between individuals [78]. Within individuals, proteins have been shown to vary depending on time-of-day, and physiological and psychological conditions [78], but are relatively stable after accounting for these factors [45]. This phenomenon has motivated the notion of personalized cut-offs or individual baselines, which calls for robust longitudinal studies to uncover.

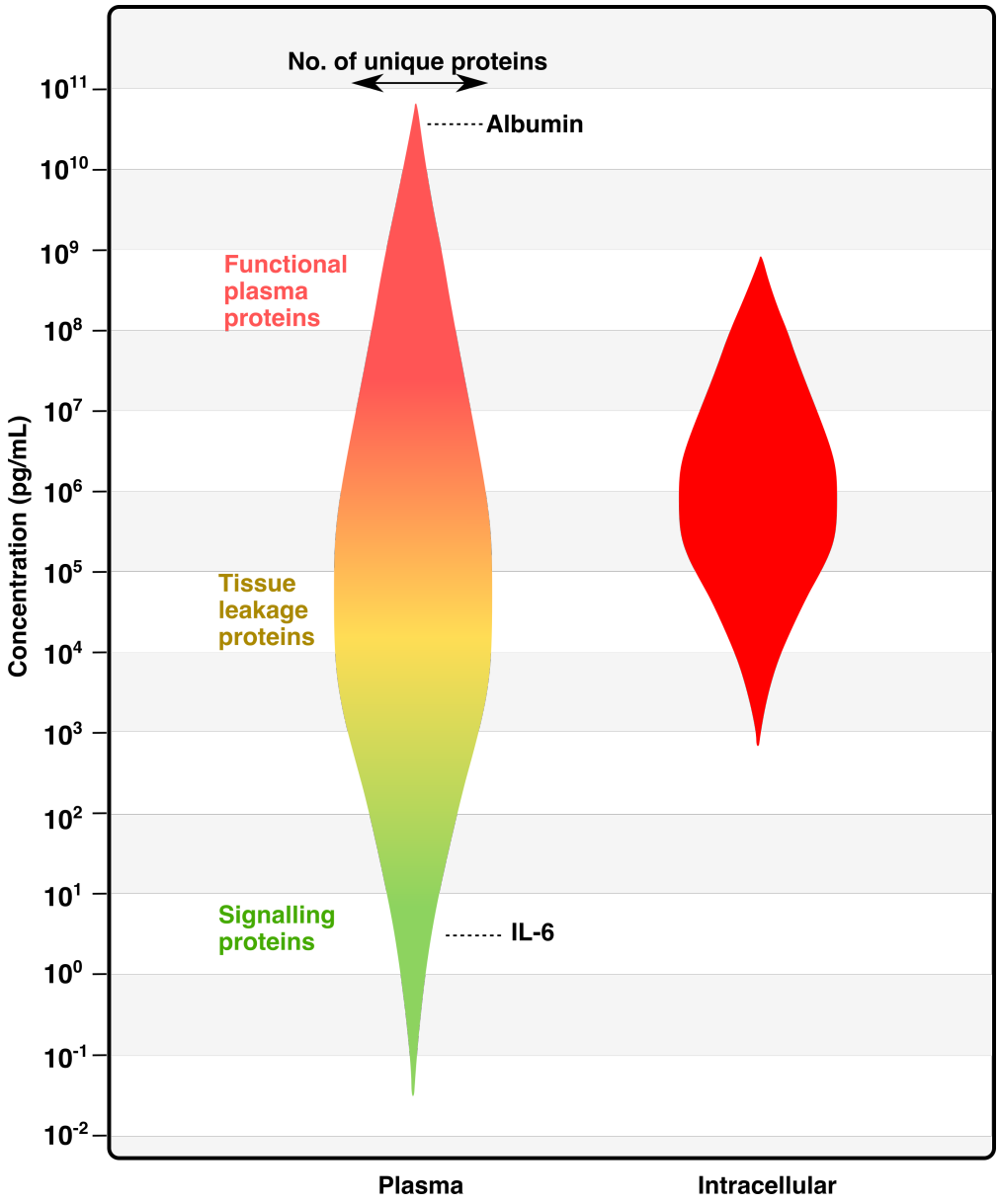


Figure 2.5: **Dynamic range of protein concentrations in human plasma and cells.** Estimated distribution of protein abundance in plasma (left) and cells (right) [74, 77].

2.2.3 Mass spectrometry-based techniques

Separation to reduce sample complexity. In the face of complex biological matrices, many proteomic workflows have historically relied on separating proteins before analysis, based on charge, size, or hydrophobicity. One of the most common methods is separation by mass using size-exclusion liquid chromatography (LC). Separation can itself be a standalone strategy to inspect and compare biological samples. Notably, 2-D electrophoresis performs sequential, gel-based separation of proteins according to their mass and charge resulting in 2-D separation. Since its introduction in 1975, 2-D electrophoresis was applied to plasma samples [58, 79], allowing the resolution of over 1000 proteins, and has proven useful in recognizing genetic variants between samples such as proteolytic cleavages [80]. However, 2-D electrophoresis has remained for the most part a fundamental research tool as it is limited in throughput, sensitivity, and resolution.

Basis of mass spectrometry. The advent of MS solved the challenge of protein identification in 2-D electrophoresis. Since its introduction, MS has seen drastic improvements and has developed into a core tool for proteomics [6]. Many technology variations exist with broad range of capabilities and characteristics; they all rely on the general principle of protein digestion into peptide fragments, volatilization and ionization of the peptides, and finally separation and detection according to their mass-to-charge-ratio (m/z). The separation of ionized gaseous mixtures by their m/z occurs through acceleration by electromagnetic fields, and detection of the separated fractions by ion or electron photomultipliers. Therefore, the strength of MS is its unbiased ability to identify most peptides based on their separated m/z signature.

Shotgun vs. Targeted MS. A broad range of workflows have been developed for different types of proteomic samples but most are outside the scope of this discussion [6, 7]. For identification and quantification of proteins in plasma, two workflows are used most often, namely "shotgun-MS" and "targeted-MS". As the names suggest, shotgun-MS is used for untargeted, hypothesis-free, analyses of biological samples, whereas targeted-MS aims to detect and quantify specific proteins. For both methods, the general workflow consists of fractionating the digested proteins (*i.e.* peptides) via LC, ionizing and introducing the eluted product into a first mass analyzer (MS-1) for a m/z measurement, followed by fragmentation of a selected MS-1 fraction and its m/z measurement in a second, tandem mass-analyze (MS-2) for identification (**Figure 2.6**). In shotgun-MS, the most abundant MS-1 fractions are typically chosen for further identification in the second mass analyzer (MS-2); this is also called data-dependent acquisition (**Figure 2.6b**). In contrast, targeted-MS workflows identify a priori the LC elution times and MS-1 profiles of the target proteins, and hence it is only these targeted fractions that are analyzed by MS-2 (**Figure 2.6c**).

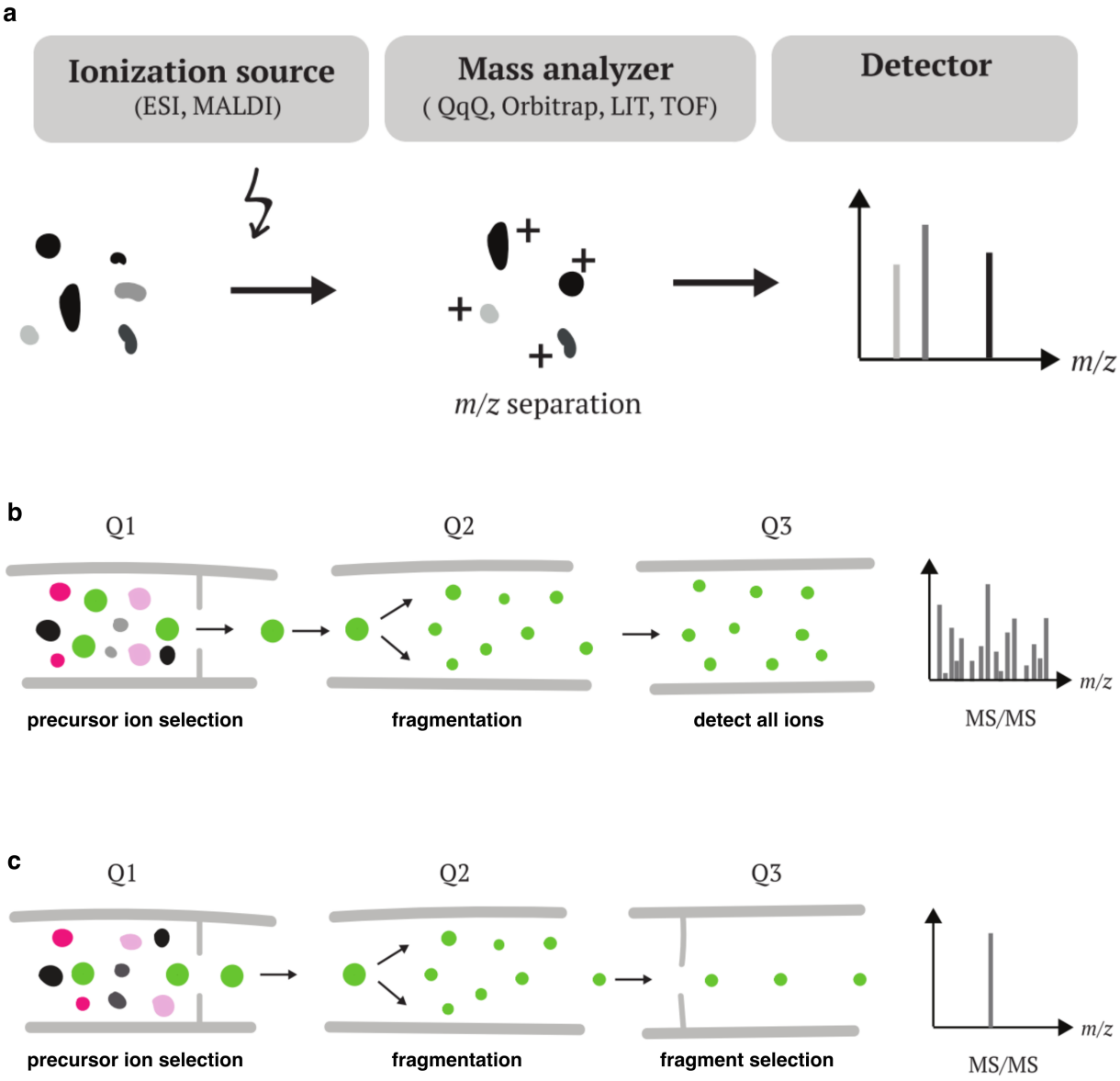


Figure 2.6: **Shotgun MS and targeted MS**. (a) The basic workflow of MS involves ionization of intact or digested proteins by, for example, electrospray ionization (ESI), followed by separation and detection based on m/z ratios. (b) In shotgun-MS, the highest abundance fractions (also called precursor ions) eluted from the LC at every moment are fragmented and all of the ions are detected downstream. (c) In targeted-MS, pre-defined precursor ions are fragmented, and pre-defined fragment ions are detected. Adapted from Sanna Byström [81] with permission.

Fractionation and depletion. Through the years, researchers have catalogued the resulting spectra from different peptides, and using robust informatics are now able to identify over 50% of proteins present in an unknown sample using shotgun-MS [6, 48]. However, the massive dynamic range of proteins in samples such as plasma is a severe obstacle for MS-based analyses because high-abundance proteins typically mask the lower abundance fractions and severely limits sensitivity. High-resolution fractionation by LC-either at the protein or the peptide level-is a preferred strategy, but it is chosen at the cost of stymieing throughput. For example, it is now possible to identify 5300 proteins using shotgun proteomics; however, the throughput of this method is limited to 1 sample per day with a considerable cost [82]. The throughput challenge can be partially mitigated by depleting the highly abundant proteins in the sample using affinity chromatography or other intensive sample depletion methods, enabling quicker LC steps, but creating quantitative biases and introducing variability [48]. Targeted-MS addresses the dynamic range challenge through a priori knowledge of the MS-1 fractions that must be fragmented and analyzed, which allows avoiding the unwanted and masking proteins. This approach of targeted-MS results in an improved sensitivity and throughput over shotgun-MS. Also known as Multiple Reaction Monitoring (MRM), targeted-MS can be used to target a few dozen proteins simultaneously [83].

Accuracy and quantification. MS is heralded for its analytical specificity. However, identification of the peptides originating from the right target protein with absolute selectivity requires stringent statistical identification criteria that is not always applied [7, 48, 83]. For instance, the number of identified proteins from a human plasma sample can widely depend on the stringency of the bioinformatic algorithms [77]. Further, due to the data-dependent acquisition in shotgun-MS and often co-eluted fractions, repeated measurements of the same sample may result in different subsets chosen for fragmentation, impacting the reproducibility of these assays [84]. In addition, shotgun-MS provides only a rough estimate of protein concentrations due to matrix interference and variability introduced during protein digestion, and is often used for gross differences in protein expression levels (>5-fold); on the other hand, MS may be performed using isotope-labeled internal standards, imparting robustness against matrix interference and improving quantitative accuracy [83]. For instance, isobaric tags for relative an absolute quantitation (iTRAQ) are chemical groups that attach to the primary amines on proteins and degrade into reporter ions that are visible in MS-2 [85]. Asides from imparting more accurate quantification in both shotgun and targeted MS, this method allow sample multiplexing by pooling together samples that are pre-labeled with different tags.

Sensitivity and throughput. MS-based analyses can yield comprehensive information about protein abundance in bio-fluids, but are typically lacking in sensitivity and sample throughput.

The necessity for fractionation results in a sample preparation procedure that is often complex and lengthy. Even without fractionation, data acquisition in shotgun-MS takes weeks for a large study with over a few hundred samples. Indeed, the sensitivity of MS-based analysis does not enable probing signalling proteins-even the most sensitive targeted-MS workflows fail to detect proteins in below the low ng/mL range [48]. The multiplexing capability of targeted MS has been scaled to target more than 100 proteins simultaneously, but at the cost of sensitivity [86, 87]. The majority of MS-based assays are developed in-house using commercial MS instruments. However, targeted-MS is in general a highly versatile method particularly appreciated for analytical performances and reproducibility and is more commonly used by CROs and in clinical diagnostics. In general, CROs and academic-based facilities offer custom MS-based analyses as services to academia and pharmaceutical companies, with some offering multiplexed MRM ready-to-use kits (*e.g.* PeptiQuant from MRM Proteomics, www.mrmproteomics.com).

PTMs. A strong advantage of MS-based analysis is the suitability to comprehensively profile some forms of PTMs, a feature that is particularly interesting for drug discovery [52, 88]. For instance, separation columns such as ion-exchange or hydrophilic-interaction chromatography can be used to enrich certain modifications such as phosphorylated peptides before analysis by MS [89]. Other modifications, however, may require affinity chromatography using anti-PTM antibodies [52], which will be discussed later.

2.2.4 Affinity-based detection of a single analyte

Rise of affinity-based detection. Proteins are complex molecules whose function is often directly linked to their conformation. Immune systems have evolved antibodies with highly variable epitopes that can be used to recognize any protein (be it pathogen-related or endogenous) through complementary shapes and chemistries. Therefore, it comes as no surprise that for as long as 90 years ago, researchers have exploited the natural immune system to develop antibodies against different antigens to repurpose them for assays [90]. However, the immunoassays were only analytically well developed in 1960-1980 [91–93], contributing to the rise of the diagnostics industry and spawning many products that are still heavily relied on today, from single dipstick tests to large automated analyzers, for applications ranging from diagnostic tests to patient monitoring that together target various clinical indications [94]. Interestingly, the underlying assay formats have conceptually remained the same since their development, and the immunoglobulin G (IgG) antibody continues to be the most widely used type of affinity-binder today, with over 3,000,000 antibodies available commercially, targeting 94% of the protein-encoding genes [77].

Immunoassay formats. Affinity-based assays can be seen as sensors for a protein target. Therefore, virtually all immunoassays workflows are composed of three elementary steps, namely target recognition, signal transduction, and signal generation, and hence can be classified accordingly. Firstly, target recognition can be classified based on the number of affinity-binders used to simultaneously bind to and recognize the target, which influences attributes such as sensitivity, specificity, and the time it takes to complete the assay. The most common formats are single-binder assays (SBAs) and sandwich assays (SAs) which, as the names entail, use a single-binder and two-binders to respectively detect an analyte of interest (**Figure 2.7c-d**). To simultaneously bind to the analyte, the binders in SAs must bind to different epitopes on the target protein. Hence, polyclonal antibodies are useful to build matched (*i.e.* non-competing) antibody pairs for SAs.
Secondly, the signal transduction mechanism is almost always specific to the number of affinity binders. In SAs, a monoclonal antibody is most often immobilized on a solid-phase and used as a capture antibody (cAb), whereas a polyclonal antibody is applied as a detection antibody (dAb). The dAb is often labeled to signal the presence of the analyte. This signal transduction mechanism is typically referred to as 'reagent-excess' due to the addition of a molar excess amount of labeled dAbs [94, 95]. On the other hand, single-analyte SBAs are typically competitive assays, wherein the analyte competes for antibody binding against a labeled analyte standard.
Lastly, affinity-based assays may also be classified based on their signal generation scheme. The radioimmunoassay, which was one of the first quantitative SBAs developed and can still be found in some clinical tests today [91], conjugates the protein to a stable radioactive label. To curtail the use of radioactive materials, the enyzme-linked immunosorbent assay (ELISA) employs antibody-conjugated enzymes to generate colorimetric signals and benefits from their catalytic capabilities [96]. The ELISA is perhaps the most widely known assay, although the use of the term to refer to SAs is a common misnomer. Other signal generation mechanisms for single-binder assays include chemiluminscence, fluorescence, and electrochemiluminescence [94].

The challenge of antibody promiscuity. In general, employing affinity interactions for target recognition allows for rapid assays with high sensitivity, but is met with a panoply of specificity issues. There are two types of non-specific interactions: cross-reactivity (CR) and non-specific binding (NSB). CR (or specific non-specificity) is the result of binding of the affinity-binder to an off-target molecule that may be structurally related or similar to the analyte. While the high-affinity antibody-antigen interaction is considered highly specific, antibodies have been known to engage in rather promiscuous low-affinity CR with off-target proteins [97–99]. This may be a limitation inherent to all affinity-type detection of proteins, with some arguing that non-specific, low-affinity interactions should be considered as the primary source of biological noise [100, 101]. While there is no clear consensus regarding the rate or probability of CR, it is estimated that 5-20% of

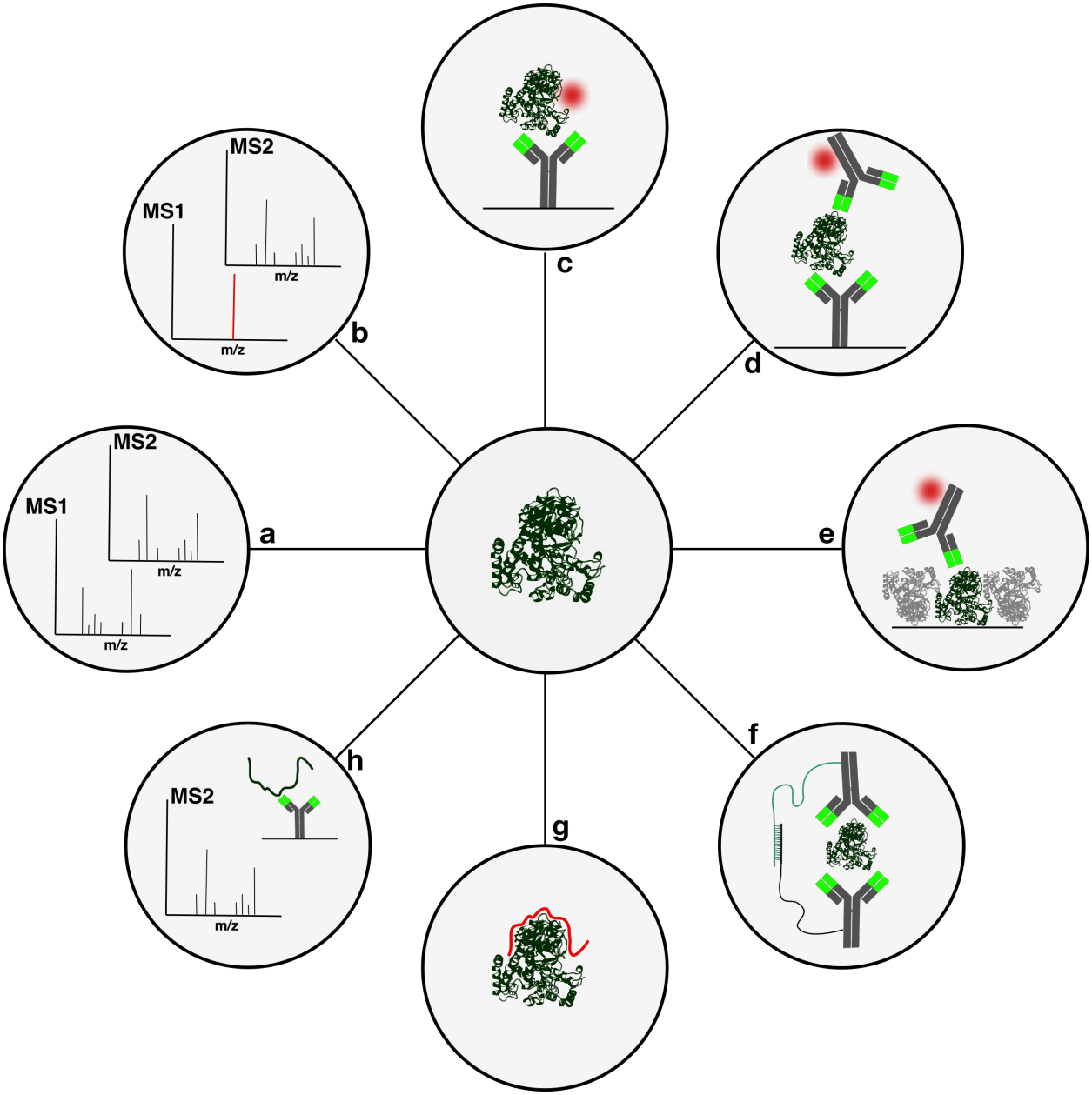


Figure 2.7: **Established and emerging methods in proteomics for bio-fluids.** (a) Shotgun-MS, (b) targeted-MS, (c) single-binder assay (SBA) or multiplexed SBA (MSBA), (d) sandwich assay (SA) or multiplexed SA (MSA), (e) reverse-phase assay (RPA), (f) proximity extension assay (PEA), (g) slow off-rate modified aptamers scan (SOMAscan), (h) stable isotope standards and capture by anti-peptide antibodies (SISCAPA).

antibodies will exhibit CR to a significant degree [74, 75, 102, 103]. Specificity is therefore a substantial problem in affinity-based assays, and specificity assurance of antibodies is a topic that recently has sparked renewed interest and discussion within the proteomics community [104–106].

The challenge of non-specific binding. NSB includes all other sources of binding that impact the precision and accuracy of affinity-based assays. For example, several types of endogenous plasma proteins such as heterophilic antibodies, soluble receptors, and immune complexes may interfere with immunoassays by binding to the protein targets or the affinity binders [107]. Interference is a major source of problem in quantitative assays that requires careful optimization with appropriate buffers and controls [108]. NSB can be assay-configuration specific, such as the binding of labelled affinity reagent to surfaces [74], or the adsorption of proteins onto surfaces [109]. As a result, NSB is a significant concern for immunoassays and draws significant attention and effort for its avoidance.

SBAs vs SAs: specificity and sensitivity. In practice, assays that exhibit reduced specificity result in increased background signal, which in turn reduces the sensitivity [110]. Whilst competitive SBAs are tolerant to some scenarios of CR, off-targets with sufficient CR affinity and/or concentration may bind to the affinity binder immediately resulting in a change in signal. Even though affinity-binders have a dissociation constant (K_d) in the fM range, that is insufficient to achieve a commensurate Limit of Detection (LOD).
On the other hand, the dual-requirement of the single-analyte SA tolerates most CR scenarios and results in a significantly more robust, sensitive, and specific assay–only the unlikely event of a dual-CR (dCR) results in a false-positive signal [74, 110, 111]. The four steps in a SA including two washing steps can be seen as proofreading measures that challenge dCR and act to reduce the impact of the false-positive signal (**Figure 2.8**). As a result, SAs such as ELISAs are typically capable of detecting analytes present in complex samples at low pg/mL concentrations, with a dynamic range of 3 orders of magnitude, which is also higher than typical SBAs that are limited by a higher background signal.

Throughput and commercial solutions. Since their discovery, both SBAs and SAs found extensive use both in research and clinical settings. SBAs are rapid and cost-efficient and allow for the measurement of small molecules such as hormones and vitamins which cannot effectively bound by two affinity-binders [94]. SAs were initially developed as a sensitive immunoglobulin test for allergies [112] but soon thereafter found application in specific antigen detection [93]. The exquisite sensitivity of SAs uniquely afford the detection and measurement of the low-abundance signalling protein such as IL-6 and the p24 HIV antigen [94], which explains their widespread use in clinical

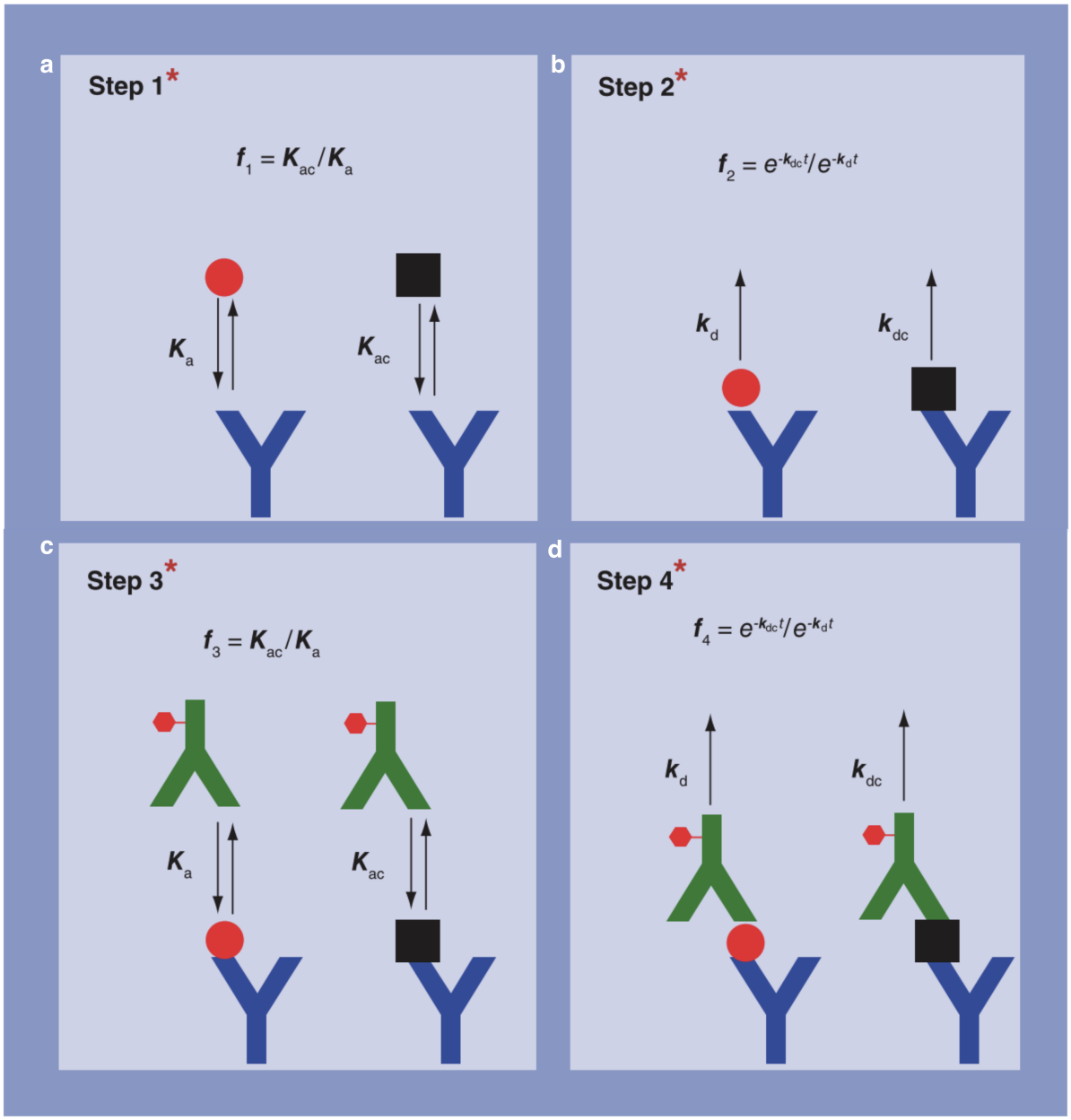


Figure 2.8: **Analytical specificity of a sandwich immunoassay, and dual cross-reactivity (dCR)**. The four steps of a sandwich immunoassay, namely (**a**) capture, (**b**) sample washing, (**c**) detection, (**d**) and detection antibody washing, wherein the association and dissociation constants of targets (K_a and K_d, red circles) and off-targets (K_{ac} and K_{dc}, black square) are used to calculate the fraction (f) of the non-specifically bound molecule at every step. Assuming equal concentrations of target and off-target molecules, and association/dissociation constants that are 2 orders of magnitude worse for the off-target molecule at every step, only 1 in 18,000 molecules detected at the end of the assay will correspond to that of the off-target molecule. Adapted from [110], licensed under CC BY 3.0.

diagnostics. Unfortunately, commercially-available matched antibody pairs (*i.e.* validated to be non-competing in binding the same target) are less common, and only seem to be more commonly available against secreted proteins. Given an antibody, immunoassays are relatively simple assays to develop in-house; still, many manufacturers provide user-friendly solutions for researchers and clinicians, in the form of antibody-coated well plates, solution-based reagents, or fully-automated instruments with disposable cartridges [94]. The sample throughput for immunoassays can be high, as they can be easily automated and performed in 96- or 384-well plates. These assays can be rapidly deployed as they do not require sophisticated instruments and the cost per sample ranges between $1-10. In addition, the sensitivity, speed, and ease-of-use of the immunoassay explain its continued use in fundamental and clinical research. For all these reasons, the clinical diagnostics industry has established a large infrastructure of measurement tools, bio-banking facilities, supply chains, and a trained workforce that employs immunoassays for disease testing, diagnosis, and prognosis.

2.2.5 Multiplexing platforms: planar and microparticle-based platforms

While multiplexed immunoassays were proposed and demonstrated as early as 1989 by Ekins *et al.* [113], the potential for multiplexed assays in general was in DNA microarrays first [49]. One obvious and immediate advantage of running assays in parallel is a reduced number of experiments and reduced sample consumption. In addition, as discussed previously, multiplexed analysis enables new lines of investigation because many proteins such as cytokines typically act in networks and pathways and often exert their effect in concert; therefore targeting multiple proteins in the same experiment is likely to reveal much more information.

Planar and microparticle-based multiplexing. Many multiplexing platforms have been developed in the last two decades to run affinity-based assays in parallel. Most of these platforms fall into two categories, namely planar microarrays and microparticle-based assays (also called suspension-based assays). While planar platforms separate and run the assays on discrete micro-spots on a surface, microparticle-based assays use sets of encoded microparticles suspended in solution (**Figure 2.9**). In the planar format, each antibody's identity is encoded in the 2-D location on the planar substrate, which could be a glass slide or a microtiter plate bottom [114]. While multiplexing and read-out are straightforward tasks in a planar microarray, multiplexing microparticle-based assays requires a means to track and distinguish every reaction—that is, microparticles must be encoded and then decoded after read-out.

Several multiplexing platforms have been developed and offered commercially with varying features and limitations. In general, suspension-based assays afford many advantages over their planar counterpart in terms of flexibility, ease-of-handling, batch functionalization, improved mass

transport, and statistical robustness [49, 115–117]. The microparticle-based, spectrally-encoded xMAP platform from Luminex is the prevailing multiplexing platform today, with the technical capability of running up to 500 parallel assays that are readout using a dedicated cytometer [118] (www.luminexcorp.com). Typically, microparticle-based assays can be run in 96 or 384 well-plates and are amenable for high sample-throughput and automation [119]. Conversely, planar formats possess an inherent trade-off between multiplexing and sample volume needed, wherein high capacity multiplexing requires more microarray spots and hence larger overall footprints and sample volumes. The increase in sample volumes is also typically associated with a decreased operational throughput. For example, the Meso Scale Discovery (MSD) platform (www.mesoscale.com) immobilizes arrays of antibodies on the bottom of a 96-well microtiter plate, which limits multiplexing to 10 targets (*i.e.* 10-plex) but is amenable to large sample throughput [120]; on the other hand, the 1000-plex microarray platform of Raybiotech can only run 4 samples per slide (www.raybiotech.com).

Microparticle barcoding. While microparticle-based assays provide substantial advantages in comparison to planar-based arrays, barcoding constitutes a substantial challenge. Many encoding strategies have been developed for MPs through the years including spectral [121], graphical [122–124], and chemical approaches [125]. Of these, spectral encoding remains the most popular method [49, 117, 126], as it allows the use of small MPs with better reproducibility, higher sensitivities, and are amenable to high-throughput optical read-out. Typically, spectrally barcoded microspheres are impregnated [118, 121], or surface-labeled [127], with different fluorescent molecules or nanoparticles in precise proportions to yield a unique overall or 'ensemble' spectral response.

Theoretically, spectral barcoding can be used to yield millions of barcodes. In an ideal scenario where N classifier dyes are used, the barcoding capacity is simply $L_1 \times L_2 \times ...L_N$ barcodes where L_f is the number of intensity levels that may be encoded by every distinct classifier dye f. In practice, the limited spectral bandwidth available (350 - 750 nm), and the wide spectral response of common organic dyes greatly constrain the total number of dyes that can be used. Indeed, as N increases beyond 2 or 3, spectral overlap between the dyes becomes unavoidable which, due to the short distance between fluorophores, possible interactions such as Förster resonance energy transfer (FRET) entangle the encoding and decoding of the barcodes, as will be discussed and demonstrated in **Chapter 3**. The enabling feature in the seminal work by Fulton *et al.* in 1997 [118] which propelled the xMAP platform to market dominance, was the use of dyes that do not engage in FRET. To achieve this, the dyes used possess overlapping absorption spectra (580 nm) and increasingly large Stokes-shift such that their emission profiles are not spectrally overlapping with absorption profiles-this enables simultaneous excitation and mitigates FRET (**Figure 2.10a**). The first generation used 2 dyes (orange and red) encoding a total of 100 barcodes; the second

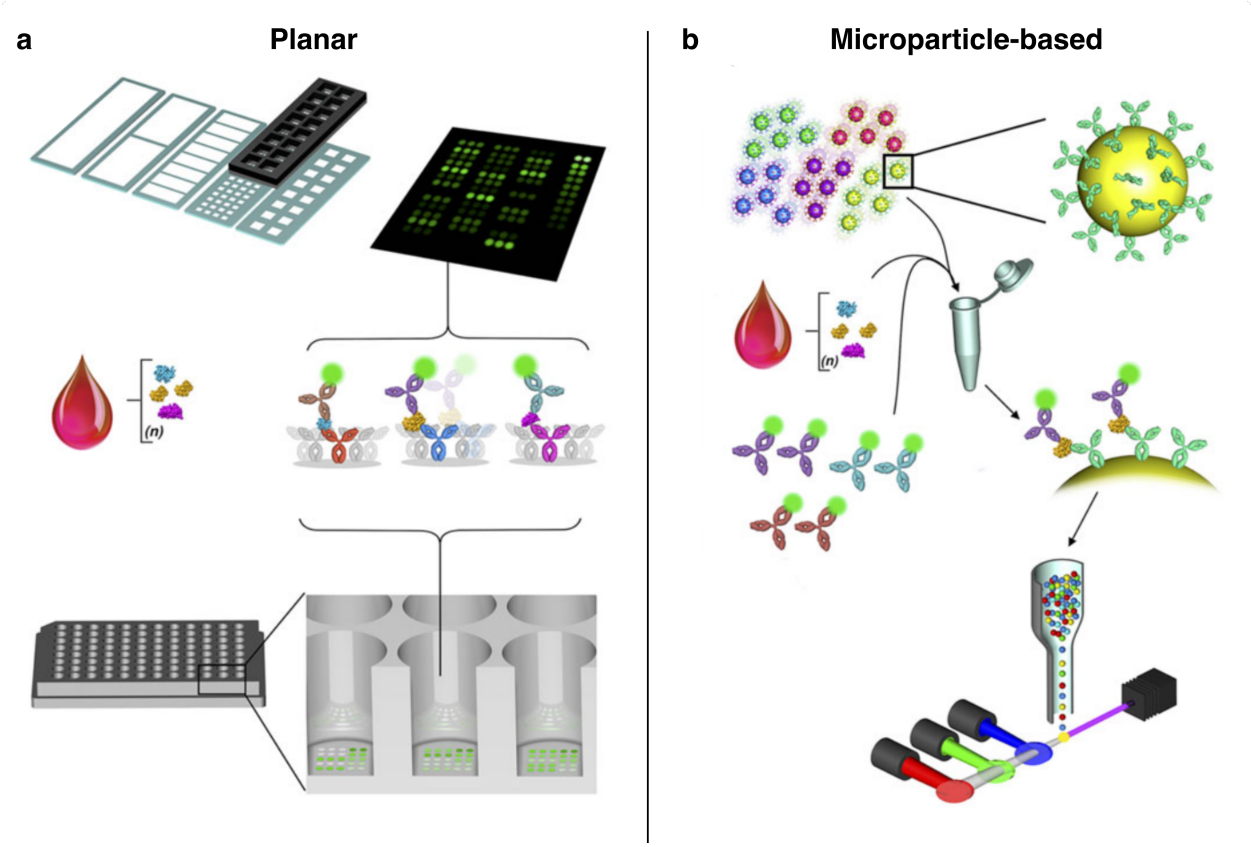


Figure 2.9: **Conventional multiplexing platforms for sandwich assays.** (a) Planar multiplexed immunoassays may be conducted on glass slides (top) or in microtiter plates (bottom) by immobilizing the different capture antibodies on discrete and separate spots on the surface. (b) Microparticle-based multiplexed immunoassays are most often conducted on spectrally-encoded microparticles that are coated with target-specific capture antibodies. Adapted from Taghe2015 from [110], licensed under CC BY 4.0.

generation used a third dye that emits in the near infrared (NIR) region (> 680 nm) which increased the barcoding capacity to 500 [128]. While the spectral profile of the constituent dyes is not publicly available (and indeed, their physical structure, to the best of my knowledge, remains a trade secret), the NIR dye most likely has a Stokes shift greater than 150 nm. To further expand the barcoding capacity while maintaining the same approach would require infrared-emitting dyes with very large Stokes shift, which is not anticipated to occur soon.

To increase the capacity of barcoded MPs that can be read-out rapidly by a cytometer, Fridjtof *et al.* used a brute-force method using common organic dyes. Using five commercial dyes, Lund-Johansen and colleagues labeled the surface of microparticles while constraining their respective surface densities to avoid regimes with appreciable FRET [127, 129]. This approach yielded a record of 1728 fluorescent barcodes that were read out using four interrogating lasers [130]. However,

even though FRET was minimized, the barcode response was unpredictable and required labour-intensive experimental iterations to obtain distinguishable barcodes (personal communication), may not be repeatable, and requires manual calibration to perform BMP-to-barcode assignment before automated read-out is possible [131]. Similarly, further expansion of the capacity using this strategy would impose unavoidable FRET that would require increasingly painstaking fabrication and decoding.

Because of the wide spectra of common organic dyes that lead to high propensity for FRET, other approaches aimed to eliminate the use organic dyes altogether [126]. Quantum dots (QDs) were envisioned as suitable classifier dyes as their size-tunable spectra and wide-bandwidth absorption allows single laser excitation of different QDs [132, 133]. In practice, the barcoding capacity of QD-laden BMPs has only reached 105 [134] due to challenges such as electronic and long-range energy transfer that rapidly increases at higher QD concentrations [126, 135–139]. As a result, there continues to be no demonstrations of highly-multiplexed QD-based barcodes read-out by cytometry. Lanthanides constitute another attractive candidate for barcoding owing to their large Stokes shift and narrow emission spectra [140, 141]. However, lanthanides are not compatible with common flow cytometers due to the their long fluorescence lifetimes and need for an uncommon 285 nm excitation source; as such, fluorescence microscopy must be used for read-out, severely limiting throughput.

Another approach to mitigate energy transfer calls for the spatial separation between dyes. This approach has been investigated by separating different dyes within multi-layer microparticles; however, these approaches require careful optimization and do not completely limit FRET [135, 138]. In a clever approach, dyes were separated by immobilizing the dyes within well-separated regions in rod-shaped hydrogel microparticles (**Figure 2.10b**). This large separation completely avoids FRET and, additionally, the barcoding regions are registered in the cytometer as two events (*i.e.* $N = 2$). This approach is commercialized as FireFly™ by Abcam [142] and is theoretically scalable, but would require increasingly larger particles to avoid FRET; the platform is currently limited to 70-plex. Importantly, these codes can be rapidly adapted to commercial cytometers, but requires customization by the end-user.

As a result, and despite 20 years of development, there continues to be no clear alternative to using cytometry with common organic dyes, which are ubiquitous in life sciences, and no clear solution to increase the multiplexing capacity without compromising the scalability, sample throughput, or both. Addressing these issues is a central focus of the present work (**Chapter 3**)

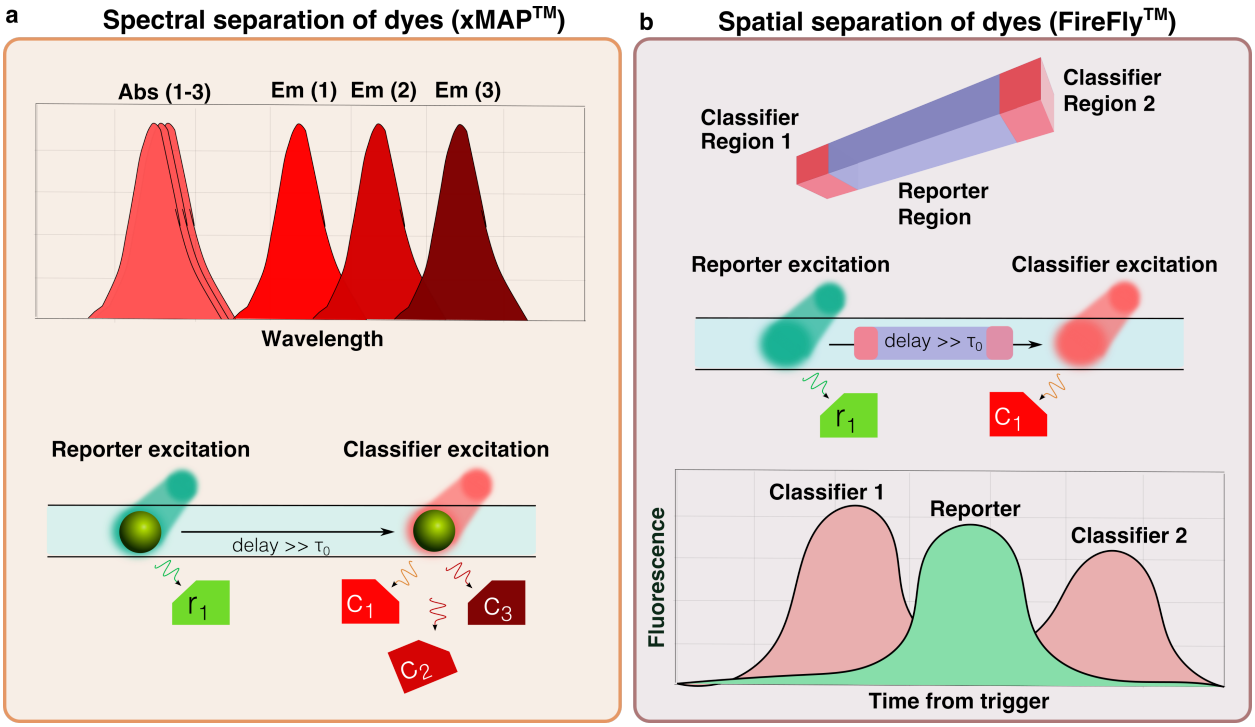


Figure 2.10: **Spectral encoding methods with multiple orthogonal classifiers.** (a) In the approach used in the xMAP platform, classifier dyes with the same absorption spectra and staggered emission (top)-and hence, increasing Stokes shift-are used to allow orthogonal (*i.e.* FRET-free) barcoding, as well as single-laser excitation (bottom). (b) In the approach used in the FireFly platform [142], the dyes are immobilized at the ends of a hydrogel rod, and act as independent, non-interacting classifiers. In this approach, the regions are analyzed sequentially in the flow cell, and appear as three separate events (bottom).

2.2.6 Multiplexed immunoassays

Multiplexed assay formats: single-binder and sandwich assays. Operationally speaking, SBAs are more practical for multiplexing than SAs. Two types of multiplexed single-binder formats have emerged, either consisting of immobilizing affinity binders into discrete arrays (multiplexed SBAs - MSBAs), or incubating affinity binders onto arrayed biological samples (reverse-phase arrays - RPAs) [49]. MSBAs can be implemented on planar [69, 75] or microparticle-based systems [127, 143] by labeling the biological sample entirely before the assay, for example using a fluorescent dye or an intermediate label such as biotin [115]. However, sample labeling has the potential to disrupt the binding of the target analyte [75] and can be avoided by employing label-free detection systems. In RPAs, cell or tissue lysates are most commonly spotted on slides before detection with specific antibodies [102]; despite removing any spatial features from these samples, this method offers several alternatives over tissue staining (IHC) such as high sample throughput, lower sample consumption, and good reproducibility. RPAs are attractive as different tissue and cell types can be

interrogated in the same experiment [61]. Similar to the workflow of SAs, in multiplexed sandwich assays (MSAs) the sample is incubated with the antibody arrays (be it planar or microparticle-based), followed by washing and detection using a pre-mixed dAb cocktail targeting the analyte panel. As described below, the application of dAbs against different targets as a mixture leads to a new form of CR.

Vulnerability of MSBAs and RPAs to CR and NSB. In general, multiplexing increases the vulnerability of affinity-based assays to CR and NSB. For instance, MSBAs are significantly less specific when the entire sample is labeled; indeed, every NSB or CR by an off-target protein contributes to assay background or potentially leads to a false-positive (**Figure 2.11c**). Increased NSB in MSBAs significantly impedes the sensitivity by 2-3 orders of magnitude [144], resulting in the failure to detect many of the low-abundance signalling proteins [110]. This problem has implications that reach beyond a reduced sensitivity: a detectable signal originating from a CR is indistinguishable from a specific binding [100, 111]. In addition, the incidence of a false-positive signal depends on both the affinity of the cross-reactive event as well as the concentration of the off-target molecule. In other words, even low-affinity CR with a high-abundance off-target protein in MSBAs will interfere significantly and lead to a false-positive. This effect poses a significant challenge in plasma because of the potential large concentration difference between target and off-target molecules [100, 102]. As a result, it is estimated that as little as 20% of the data in MSBAs may be accurate [145]. The strong susceptibility of MSBAs and RPAs to NSB and false-positives impacted their reliability, and in turn limited their commercial development. Indeed, most of the MSBAs in the literature are developed in-house [69, 127] by independent research groups with only a few commercially available platforms mostly as assay services [146]. On the other hand, the large multiplexing afforded by MSBAs and RPAs coupled with their cost effectiveness and high sample throughput has motivated their use for semi-quantitative screens in large biomarker studies, and has even been implemented in large clinical trials as will be discussed later.

Sensitivity of MSAs vs. SBAs. Whereas SBAs are immediately susceptible to any CR or NSB from the sample (**Figure 2.11a,c**), MSAs are significantly more robust against this type of CR due to the dual-recognition requirement (**Figure 2.11b,d**). Similar to SAs, an off-target protein can only generate a false-positive in the unlikely case of dCR (**Figure 2.11e**). As a result, MSAs command strong advantages in sensitivity and specificity over MSBAs [74, 94, 110, 111]. Indeed, a great number of MSAs are regularly developed (both in-house and commercially) using the aforementioned multiplexed platforms [114, 147]; notably, xMAP is licensed to over 70 partners (*e.g.* BioRad, ThermoFisher) who produce and sell pre-optimized MSA panels [148, 149].

MSAs introduce a new type of cross-reactivity. On the other hand, the MSA presents a unique underlying analytical challenge. While MSAs are indeed impervious against CR from off-target proteins (**Figure 2.11d**), the application of detection antibodies in cocktail format creates an environment that is highly susceptible to CR between the reagents themselves and the targeted protein panel (**Figure 2.11f**). This new form of CR arising in MSAs is called reagent-driven CR (rCR) [111, 150]. As a result, the dual-recognition requirement afforded by SAs is lost in multiplexing as a single rCR can lead to a false-positive result [75, 151]. Moreover, the number of potential scenarios leading to rCR, termed 'liability scenarios' [150], scales as $4N^2$ with the number of targets N (*i.e.* the size of the panel), quickly increasing the number of false-positive signals, and significantly eroding the specificity and reliability of the MSAs with increased number of targets.

Multiplexing limitations of MSAs due to rCR. Despite the underlying rCR, MSAs using planar microarrays or barcoded microparticles are widely used in academia and pharma today, but continue to be limited by it [49, 55, 114]. The impact of rCR on the development of MSAs is three-fold. First, to cope with rCR and build fit-for-purpose assays, development requires brute-force optimization that involves combinatorial specificity screening to remove and replace cross-reacting antibodies [152–155]. This method of avoiding rCR is not scalable–not at least because the amount of optimization necessary also scales quadratically. Second, the multiplexing levels of MSAs have been effectively capped to 50 proteins per assay because optimization is prohibitive or impossible beyond this multiplexing level [150]. Third, ineffective optimizations by manufacturers have negatively impacted the reliability of MSAs, with many users reporting poor reproducibility and/or poor correlation with ELISA [156–158]. Indeed, despite this issue being long known [100, 159] many users of commercial MSA kits are often unaware of this underlying problem and conflate the issue with the poor specificity of antibody reagents [49, 160]. Furthermore, some researchers mistakenly believe that the limited multiplexing of MSAs is due to the limited availability of antibody pairs [110]. Despite these misconceptions, rCR is the current bottleneck for the development and use of MSAs [111], and addressing it is a central focus of the present work (**Chapter 4**).

MSAs and commercial kits. As a direct result of rCR, MSA kits offered on the market today have been mostly limited in the 4-25plex range, with only a few reaching above 40-plex [155]. These multiplexed kits typically target cytokine/chemokines, but also cancer-related, cardiovascular, neurological, and infectious disease markers [161]. The xMAP platform was FDA-approved in 2007 which enabled its adoption in immunodiagnostics [114], drug development [119, 155] and preclinical/clinical trials [55]. Well developed and optimized MSAs match ELISA's sensitivity (pg/mL), reproducibility (CV<10%), and high-throughput capabilities and robustness [155, 162,

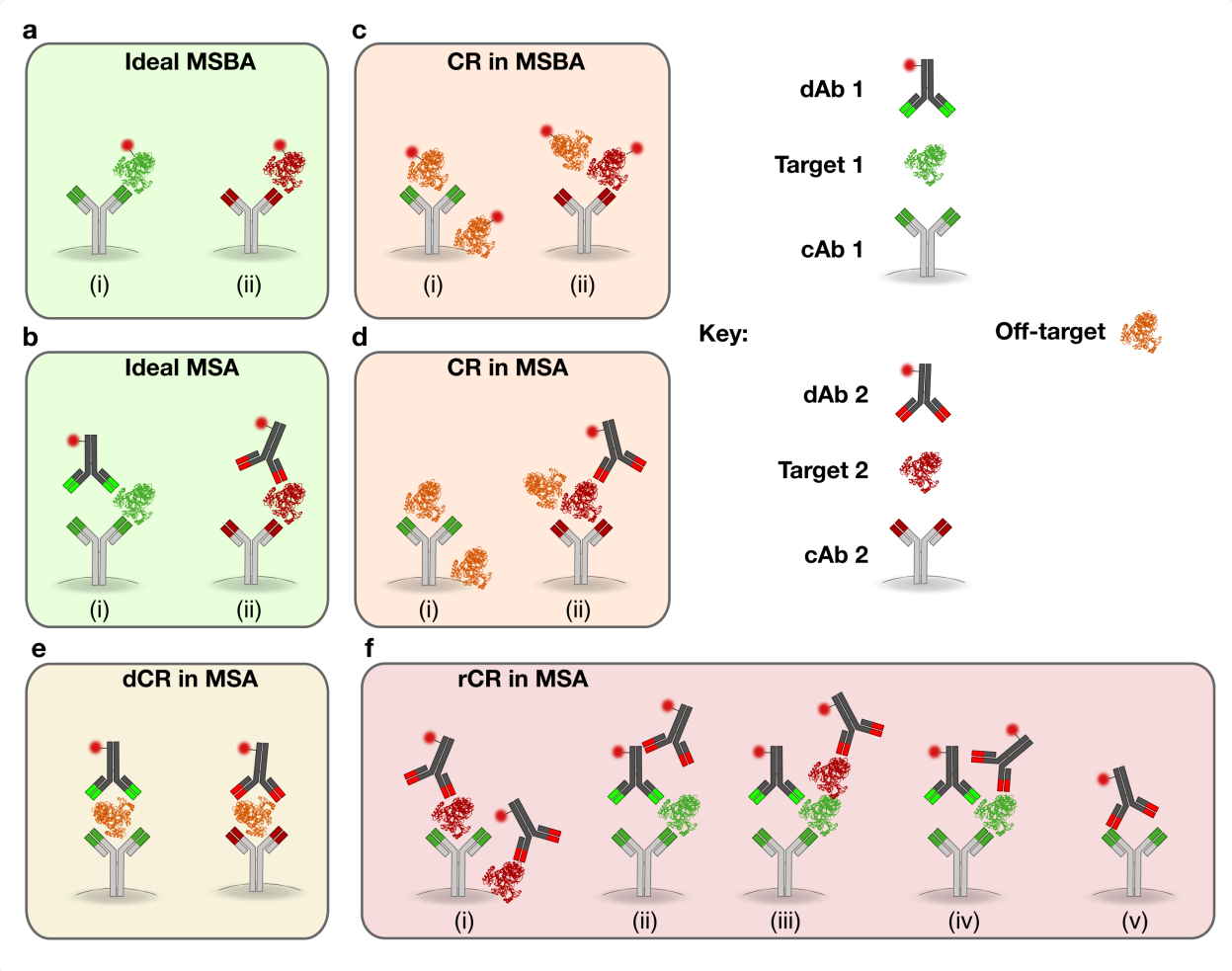


Figure 2.11: **CR and NSB in multiplexed immunoassays.** (a,b) Ideal 2-plex in (a) MSBA and (b) MSA, showing targets 1 and 2 captured by their cognate antibodies, (i) and (ii), respectively. (c,d) Off-target protein (i) CR and NSB with capture spot 1, and (ii) CR with target protein 2 in a (c) MSBA and (d) MSA, respectively. (e) dCR of an off-target protein with both capture and detection antibodies. (f) rCR scenarios in MSAs: (i) Target 2 binds to capture spot 1, (ii) dAb 2 binds to dAb 1, (iii) target 2 forms a complex with target 1, (iv) dAb 2 binds to target 1, and (v) dAb 2 binds to cAb 1.

163]. In general, these assays allow saving precious sample volume, reducing sample freeze/thaw cycles, and other costs associated with repeated analysis.

Antibody availability and dynamic range in multiplexing. An other challenge for all multiplexed immunoassays is that of using antibodies with different affinities in the same assay [155]. While, ideally, every antibody (or antibody pair) would have the affinity needed to quantify their protein of interest in physiological concentrations, many antibodies have higher (or lower) affinity which requires more (or less) dilution of the sample [102, 155]. This is further aggravated by that large variations of endogenous proteins, which can vary up to 4 orders of magnitude depending on the expression levels [110]. The availability and compatibility of antibodies used is another concern. An example of a known issue is that many antibodies that were raised against denatured proteins may not be able to recognize the native state of the antigen [130]. The availability of matched antibody pairs is an issue for MSAs where non-competing antibody pairs must be utilized; however this is not the bottleneck as there are at least 1200 antibody pairs that are used in various MSAs [128]. In addition, different proteoforms are challenging to detect using affinity-binding, and when unaccounted for impact the precision of assays, especially when different PTMs of the same proteins are targeted in the same assay.

2.2.7 Emerging proteomic methods with improved sensitivity and multiplexing

Due to the longstanding analytical and scaling challenges in proteomics, significant focus has been placed to address these technological shortcomings, with many efforts recently leading to fruition, as will be discussed here.

Advances in MS and immuno-MS. Efforts to improve the sensitivity and throughput of MS-based workflows are continuous. Geyer *et al.* recently identified and showcased a shotgun-MS workflow that can detect 285 proteins from 1 μL of plasma with a reported low ng/mL sensitivity in as little over 3 hours [164]. This workflow was later used in a longitudinal study for sustained weight loss, measuring a total of 1300 samples. Despite the impressive improvement in throughput, data acquisition in this study still took 10 weeks [165].

Despite the promise of targeted-MS in increasing the sensitivity and throughput, its reliance on columns and the impact of matrix effects such as ion suppression on the quality of the data limit its potential use. To avoid some of these issues, targeted-MS has been combined with antibody-based enrichment [161, 166]. Notably, in "stable isotope standards and capture by anti-peptide antibodies" (SISCAPA), Anderson *et al.* enrich the target peptides along with spiked-in labeled standards, which is followed by targeted-MS [87, 167]. This method **(Figure 2.12a)** allow multiplexed detection of targets with sensitivity reaching low ng/mL in reasonable plasma volumes (< 100 μL). Pre-digestion

enrichment (*i.e.* protein enrichment) using antibodies may further decrease the sensitivity to 100 pg/mL [83, 166]-however, these assays narrow the advantage compared to purely antibody-based methods as they require generation of specific antibodies. Importantly, immune-enrichment can be used to mitigate the need for LC which simplifies the sample prep and increases throughput. Post-digestion enrichment is particularly promising for bypassing cross-reactivity and interference from endogenous antibodies [161].

Proximity-dependent dual-recognition assays. Landegren's group in Uppsala University in Sweden tackled several problems in MSAs by designing DNA-assisted assays that increase specificity and sensitivity (**Figure 2.12b**). By conjugating target-specific antibody pairs to DNA probes containing complementary sequences, the equilibrium of a solution-phase assay is designed such that DNA probes only form a stable hybrid when the antibodies are brought within proximity by the antigen. The stable hybrid can then be used to perform polymerase-based amplification and read-out by qPCR. Effectively, this 'proximity-based' assay adds an additional proof-reading step that stifles rCR from generating non-specific signals. Subsequently, several assays were developed [168–170]. One format, termed "Proximity extension assay" (PEA), was proven amenable to multiplexing and has been commercialized by Olink Biosciences (www.olink.com, **Figure 2.13**) which offers several 92-plex panels that can be measured from 1µL of sample [171]. The assay allows sensitive detection of proteins present in low pg/mL and across 5 orders of magnitude with high precision. A caveat of proximity-based assays in general is that, while protein recognition is performed in multiplexed format (*i.e.* in the same sample volume), DNA amplification by PCR is performed for every probe separately, in separate volumes. Indeed, as PCR amplification is not specific enough in multiplex, the assay calls for sample fractionation into isolated PCR reactions. In the commercialized version of the assay, this is performed using microfluidic qPCR chips, strongly limiting the efficiency and scalability of this technique. To minimize annealing of the probes in the absence of antigen, and hence minimize signal background, the complementary sequence between the matching probes is typically 5-bp. Furthermore, because rCR still occurs in these assays, as in other MSAs, DNA primer design must ensure no cross-hybridization between non-matched probes, which is a challenging task given the small hybrid size. Indeed, while there are over 1024 potential 5-bp sequences, it's unclear how many orthogonal primers may be designed in the future. In effect, while this approach allows simultaneous recognition of 96 proteins with little-to-no rCR, signal transduction via qPCR will pose significant challenges towards further scaling.

Modified aptamers: SOMAmers. The second successful approach, taken by Gold *et al.*, is to dispose of antibodies altogether. Their rationale is that cross-reactivity between affinity-reagents, and the long and costly development times for antibodies are the biggest challenges facing increased

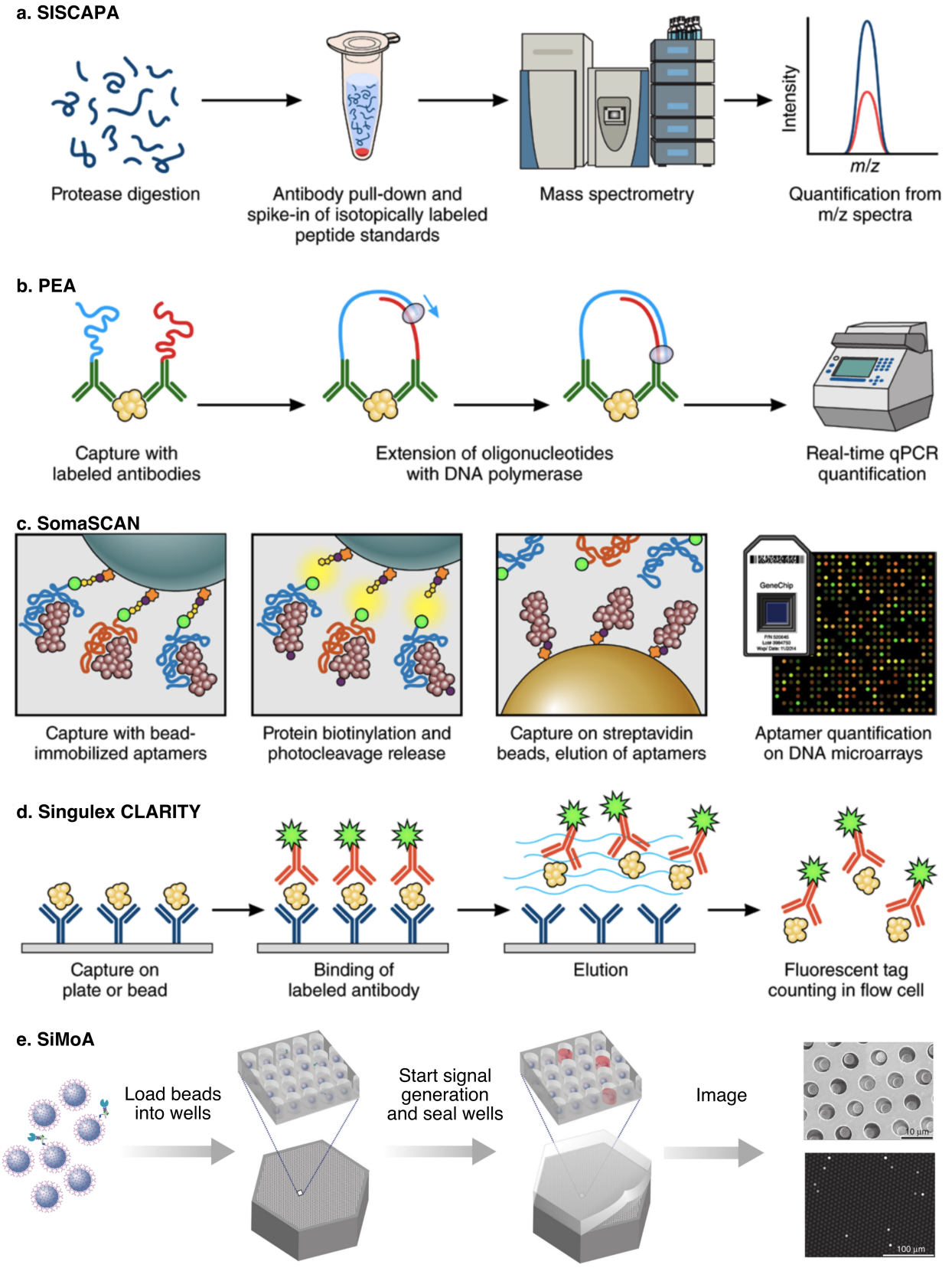


Figure 2.12: **Emerging proteomics technologies for clinical research and diagnostics. (Continued on next page.)**

Figure 2.12: (**a**) SISCAPA improves the sensitivity of MS via antibody-based enrichement of digested peptides. (**b**) PEA improves the specificity of MSAs using target-specific, complementary DNA oligonucleotides and read-out by qPCR. (**c**) SomaSCAN uses modified aptamer reagents and strignent washing protocols to improve the specificity and sensitivity of MSBAs. (**d**) Singulex's CLARITY and (**e**) SiMoA use single-molecule counting methods to improve the sensitivity of sandwich assays.

multiplexing and throughput in protein analysis [100]. To this end, this group has developed modified aptamers with amino-acid like internal bases to increase the diversity of binding motifs typically available for pure DNA aptamers. These molecules are then selected for high affinity, low dissociation rate constants, and selective binding against their targets [172]. To mitigate non-specific reactions in this single-binder assay, bound complexes are sequentially captured, released, and re-captured on two different sets of microparticles, as well as challenged with an anionic competitor (**Figure 2.12c**) [173]. The improved sensitivity and specificity of these slow off-rate modified aptamers (SOMAmers) has enabled their use as affinity reagents for multiplexed assays in single-binder formats, which forgoes the need for using detection reagents which are prone for rCR [173]. A highly multiplexed platform based on SOMAmers is commercialized by SomaLogic. This platform has gained popularity as a discovery tool and has recently allowed screening of up to 3622 proteins in parallel [20]. The specificity improvement that SOMAmers provide over antibodies is largely explained by the larger surface area in binding. However, the use of a single binder per target is expected to reduce the resolution with which similar proteins can be distinguished. Recently, a proteomic heart-disease study that used the SOMAlogic platform received much interest. The study intially found a link between GDF11 and age-related heart problems [174], whereas subseqent studies shown that the aptamer cannot discriminate between GDF11 and GDF8, and that some of the study's conclusions may have been unfounded as a result [175].

Antibody colocalization. As mentioned, rCR occurs because reagents, specifically dAbs, are allowed to interact in solution. Therefore, rCR in MSAs can be completely eliminated by eliminating interaction between dAbs and reverting to the same conditions in single-binder SAs. This approach was first demonstrated by our group on antibody microarrays through the precise delivery of each dAb atop its cognate cAb spot using precise microarray spotters [150]. This method, termed "antibody-colocalization microarray" (ACM) was shown to scale without the need for lengthy optimization of antibodies. The second-generation ACM, termed "snap-chip", performed the delivery through snapping of glass slides together, transferring nL droplet arrays containing each dAbs onto their cognate spot [176, 177]. The snap-chip provided the advantage of pre-spotting and storing the arrays ahead of the experiment, which eliminated the need for large microarray spotters. Others also implemented antibody colocalization, albeit in larger scales, through manual

pipetting [178]. While the ACM mitigates rCR, the need for precise fluidic delivery tools and the dependence on the planar microarray format has limited the adoption of this method. On the other hand, as we discuss further in **Chapter 4**, antibody-colocalization allows, conceptually, a return to ELISA-like environments at the nano-scale, and hence it is a potentially powerful concept that deserves further attention.

Ultrasensitive sandwich assays via single-molecule counting. The classic sandwich immunoassay provides exquisite sensitivity for most clinical applications; for example, ELISA-based assays are capable in detecting Troponin down to pg/mL concentrations to allow the investigation of myocardial infarctions [179]. However, many cytokines and tissue leakage products, that could provide useful information, are found below these concentrations and below the range of most SAs [58, 75, 180]. Theoretically, conventional SAs should be able to detect proteins with fM sensitivity, but are limited by non-specific binding and other sources of background [74, 110]. Many approaches have been proposed to achieve this so-called ultrasensitive detection [181–183]. However, only few approaches are robust and compatible with current workflows, enabling their adoption by industry [184]. In particular, two popular approaches perform the sandwich immunoassay on microparticles followed by a unique single-molecule counting techniques and have demonstrated 1-3 orders of magnitude improvement in sensitivity compared to a sandwich ELISA. In the first approach, now commercialized by Singulex (www.singulex.com), fluorescently-labeled dAbs are eluted off the microparticles after assay completion, followed by the counting of fluorescence events corresponding to each antibody using a sensitive single photon detector (**Figure 2.12d**) [185]. The other platform, termed 'single molecule assay'(SiMoA) and now commercialized by Quanterix, is based on loading and isolating individual microparticles within femtoliter wells, after completion of the sandwich assay, and are sealed immediately after kicking-off enzymatic signal amplification (**Figure 2.12e**) [186–188]. This assay is designed to operate in regimes where the average number of proteins per microparticles is less than 1; hence, by counting a sufficient number of positive hits so as to be above the Poisson noise limit [189], digital 'counting' of the captured molecules may be performed [187]. Together, these methods are uniquely positioned to detect proteins below the typical reach of a sandwich immunoassay, and have thus generated a lot of interest for the detection of low-abundance proteins such as IL-17A [61] that are often undetectable by conventional sandwich assays. Hence, these methods have been established as next-generation ultrasensitive sandwich immunoassays. Unfortunately, the strategies undertaken by these methods are not amenable to multiplexing.

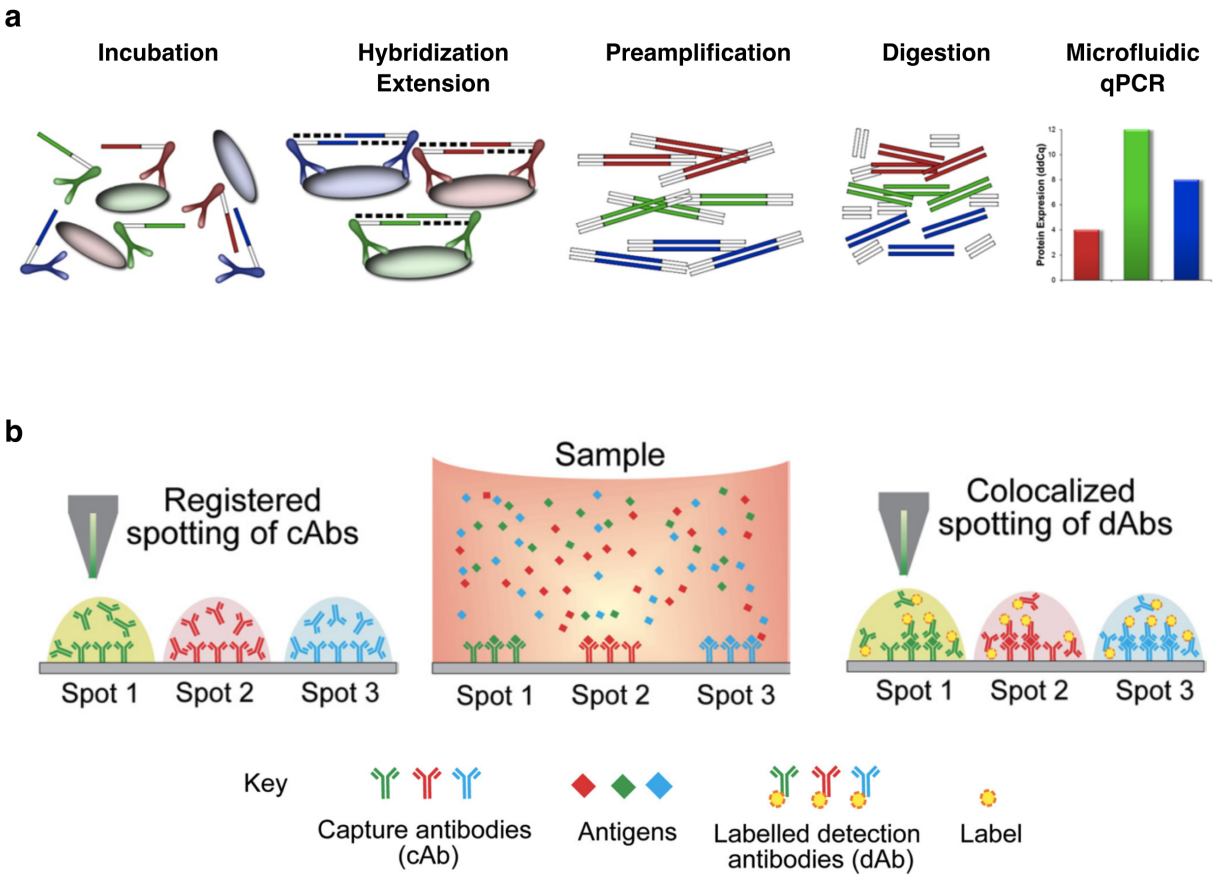


Figure 2.13: **Successful approaches in mitigating rCR in MSAs.** (a) The proximity-extension assay (PEA) conjugates every antibody pair to partially complementary oligonucleotides designed to hybridize only when brought in close-proximity by the antigen. Two PCR amplifications follow, a target independent amplification, and a target specific one. The latter is performed in a microfluidic qPCR. (b) The antibody-colocalization microarray (ACM) performed double-spotting, delivering each dAb atop its cognate cAb to eliminate antibody mixing and hence rCR.

2.2.8 Technology Landscape

Figure 2.14 displays the detection ranges for the aforementioned protein analysis strategies in comparison with the protein distribution in plasma. As discussed, the sandwich immunoassay can indeed be used to detect proteins across a large concentration range, reaching low-abundance proteins. However, it's important to note that despite their sensitivity, the typical dynamic range of a single immunoassay sensor is limited to 2-4 orders of magnitude. On the other hand, ultrasensitive assays were displayed to have an upper limit because they are single-molecule counting assays that rely on the low abundance regimes as explained earlier [189].

Figure 2.15 compares the aforementioned technologies across the 3 dimensions: multiplexing, cost per sample, and sensitivity. We substituted throughput for cost because for large-scale studies, the

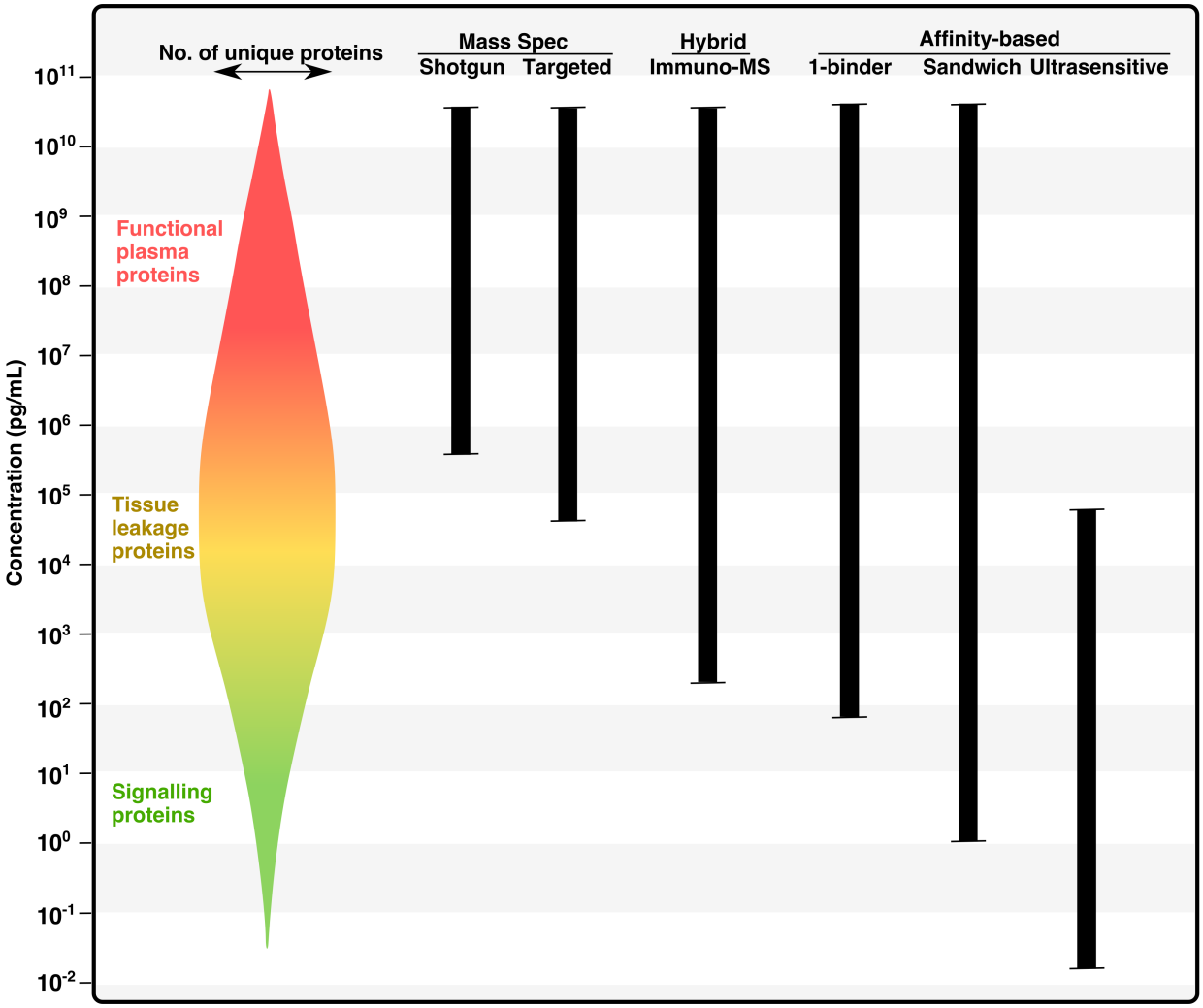


Figure 2.14: **Detection range of different assay formats in plasma, compared to the dynamic range of proteins.**

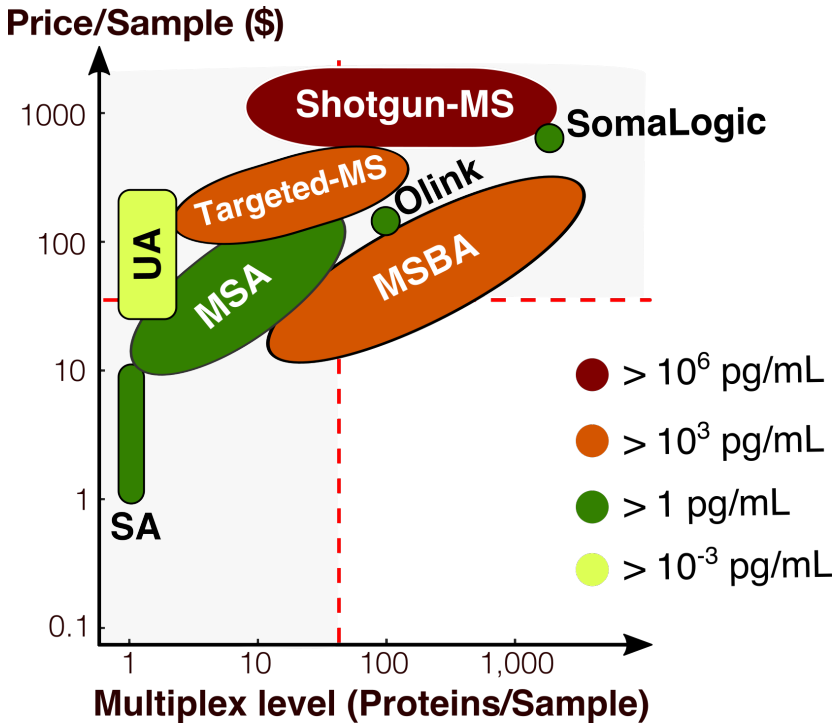


Figure 2.15: **Proteomic technology landscape.** Established and emerging technologies are stratified based on the three important metrics: multiplexing (x-axis), price per sample (y-axis), sensitivity (color-coded).

cost per sample may be the determinant metric which includes operational and technical scalability. In displaying this image, we aimed to find the price per sample as available through resellers, or offered as a service-this latter may the most informative because it also include the cost of labor and other operational costs.

2.3 Proteomics: shortcomings, and future directions

In the previous section, I presented a list of widely used, as well as emerging, proteomic tools for measuring or profiling bio-fluids. In comparing these methods, three attributes or metrics were, namely (i) sensitivity (*i.e.* proteomic depth), (ii) multiplexing capacity (proteomic breadth), and (iii) sample throughput. In this section, the limitations of these technologies will be explored vis-à-vis their impact on current practice in protein analysis across biology and medicine. The goal is to assess how these attributes played a role in the adoption for different applications, and their overall impact, and also discuss the high-priority challenges and how they may be addressed.

2.3.1 The impact of technological trade-offs on current practice

The inherent trade-off in current technologies. The technology landscape shown in **Figure 2.15** highlights the trade-offs that exist between multiplexing, sensitivity, and cost. Notably, one conclusion that may be drawn is that, whereas profiling large fractions of the proteome is feasible in small studies, it is costly to do so at-scale. This trade-off could explain current practice in biology and medicine across different applications and needs, particularly because these typically follow a hypothesis-driven pipeline. Indeed, discovery stage or exploratory studies tend to value the ability to profile thousands of proteins without a priori hypotheses and tend to choose MS-based studies, which limits the sample numbers that can be studied. On the other end, large, population-wide analyses, aiming at achieving statistical significance, tend to be resource intensive and therefore cost-effective technologies are valued, and hence resort to single-marker immunoassays. Technologies with 'intermediate' scales that are able to measure dozens of proteins in dozens of samples are used to bridge the gap between proteome-wide, small-scale studies and single-marker population-wide studies. In this section we provide examples for these typical workflows.

Bridging preclinical and translational stages. To help bridge fundamental proteomics with more clinical research and diagnostics, and to aid in translating discoveries to the clinic, a so-called 'triangular' pipeline has long been advocated [48, 190], shown in **Figure 2.16**. For example, in biomarker discovery pipelines, this multi-stage strategy from discovery to validation and clinical assay development, has been common practice in biomarker discovery pipelines in the past decade [48, 190]. A similar staging can be found in the drug discovery pipeline [36, 52, 191].

Proteome-wide discovery stage. The discovery stage in biology typically consists of a broad and semi-quantitative screen between a few sample sets with the aim to generate a hypothesis. At this stage, shotgun-MS has provided a great tool to support efforts looking for more proteins, allowing unbiased, hypothesis-free lines of questioning, and targeting up to a thousand protein at once with high specificity and dynamic range, and with the ability to profile PTMs [192, 193]. Hence, shotgun-MS has been by far the leading discovery tool in the last decade [36, 194]. The most common workflow at this stage is a semi-quantitative screen using LC/MS-MS along with established software algorithms to identify peptides in database searches [195]. In general, over 1000 studies published per year perform untargeted biomarker discovery in various tissue types [62]–however, only about 1% analyze plasma samples [48]. Indeed, aside from being more proximal to the disease, cell and tissue lysates are less complex to analyze by MS because masking by high-abundance proteins is less pronounced. Importantly, due to the limited throughput of shotgun-MS, the typical sample size of these MS-based biomarker discovery studies is 20 samples [48, 71].

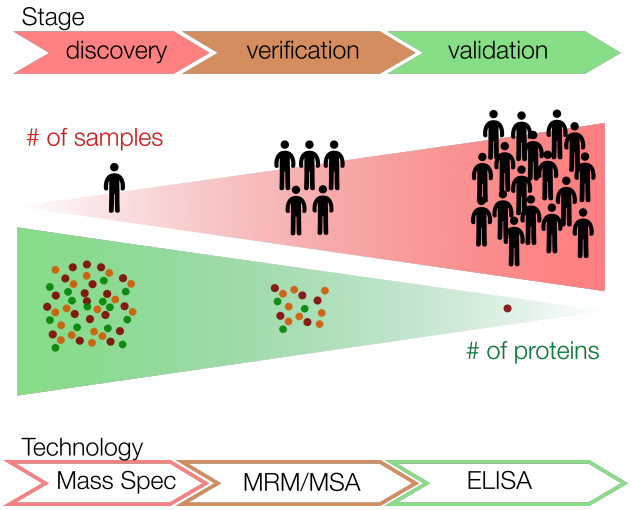


Figure 2.16: **The multi-stage, triangular approach in proteomics**. The typical approach to proteomic biomarker discovery is divided into a multi-stage approach, which starts with a discovery stage that is performed with few samples using MS-based techniques, and ends at the validation stage with ELISA measuring only one protein at-a-time.

Intermediate verification stage. The following stage is typically one that aims to verify the results of the discovery-stage. At this intermediate level, the studies are commonly performed on accessible bio-fluids using more quantitative methods with higher throughput, and typically target a smaller subset of proteins. Appropriate methods that are used here are multiplexed immunoassays and targeted-MS. When larger samples must be analyzed and better quantitation is needed, MSAs [196] and other multiplexed immunoassays, are employed [52]. Typically, 50-200 samples are analyzed for about 10-200 proteins in these studies [49, 52, 61, 102]. It is common to see studies using assay services such as Myriad RBM (www.myriadrbm.com) or Eve Technologies (www.evetechnologies.com). In a typical example, Gurbel *et al.* performed profiling of 100 proteins in 140 samples using the RBM service, uncovering a panel of 14 biomarkers that together were deemed indicative for quiescent coronary artery disease [197]. MRM-based studies are the preferable MS-based analysis at this stage, with some workflows capable of detecting >150 proteins [198]. For instance, Percy *et al.* built fit-for-purpose mouse serum assays targeting as much as 195 proteins [86]; this and other similar ready-to-use proteomic kits are now offered from MRM Proteomics (www.mrmproteomics.com).

Population-wide validation stage. The final phases of the biomarker discovery pipeline aim at biomarker validation in larger cohorts to reach statistical significance. Large sample sizes are crucial to ensure statistical power of the validation and establish clinical sensitivity and specificity against other indications. Therefore, single-plex SAs such as ELISAs are the most common method used at

this stage as they provide the specificity, sensitivity, and throughput needed to perform population-level analyses reaching over 1000-patients cohorts. Low-plex MSAs on robust platforms such as xMAP or MSD have also been implementation in clinical trials [199–201]. RPAs have also proven useful in probing distinct signalling pathways which has been successfully applied in large-scale clinical trials enabling the elucidation of signalling pathways in cancer using several thousand patient samples across multiple diseases [202, 203].

The shortcomings of the triangular model. Unfortunately, despite a 20-year effort that saw hundreds of millions of dollars invested by granting agencies and diagnostic companies [63], and over 10,000 published studies [62], no drugs already in use, or indeed reached clinical trials, as a direct result of pipelines that were MS-driven [62]. Similarly, despite early promising early biomarker findings using the aforementioned proteomic pipelines, the findings were not substantiated in larger trials [71]. As of yet, none of the potential biomarkers found using these studies have reached the clinic. On a more promising note, some biomarkers that were discovered by MS-driven studies are currently being evaluated in clinical trials [62, 194]. This failure has been a recurring subject in the literature and, apart from funding, almost every conceivable reason has been cited or considered [63, 64, 71, 204]. Some of these reasons include but are not limited to: technological variability within/across proteomic platforms, regulatory hurdles, availability of reagents/assays/datasets, lack of data analysis tools, poor experimental design, and last but not least, underpowered studies. The latter refers to the low statistical power of small sample sizes used early in the pipeline, which is perhaps the most cited technical reason for failure [71, 205]. Indeed, the typical shotgun-MS plasma discovery study consists of approximately 20 samples [48]. These studies have generated thousands of 'candidate biomarkers' [206], which as Frantzi *et al.* point out, are typically left for "someone else to study" [62]. Furthermore, because the limitations at the validation side dictate falling-back on single or few-markers, this may have also generally limited the search upstream to single-marker hypotheses.

Protein analysis for clinical diagnostics. Compared to a research setting, protein analysis in a diagnostic setting has to comply to far more parameters, including stringent conditions set by regulatory bodies; hence the technology trade-offs mentioned are not the only driving factor however limiting they may be. In the clinic, immunoassays make up 40% of all blood-based clinical tests [48], and represent an estimated total market size of $20B, making up a big share (30-35%) of the in-vitro diagnostics (IVD) market. Whereas many reference laboratories continue to use SBAs as they value cost-efficiency, speed, and automation (verbal communication with a Canadian Blood Services official), sandwich immunoassays continue to be the gold standard for protein quantification as they provide reliable quantitation down to 1 pg/mL concentrations

[49]. In addition, the sensitivity, speed, and ease-of-use of the immunoassay explain its continued use in fundamental and clinical research. For all these reasons, the clinical diagnostics industry has established a large infrastructure of measurement tools, bio-banking facilities, supply chains, and a trained workforce that employs immunoassays for infectious disease testing, diagnosis, and prognosis [94]. Despite the fact that some small-molecule immunoassays in the clinic (*e.g.* Vitamin D) are being phased out in favor of MS because of the immunoassay's inability to distinguish its various forms [63], MS is largely not being used in the clinic for protein markers. The lack of adoption in diagnostics is in contrast to metabolomics, where mass spectrometry-based assays that measure multiple metabolic biomarkers are now routinely used, replacing single-plex assays.

2.3.2 The potential of emerging proteomic platforms

A shift towards a more "rectangular approach". A decade ago, GWAS were performed in two stages. At the time, the high-costs of genotyping forced researchers to use a smaller sample (around 1000 samples) for genome-wide analysis to discover promising SNPs. Typically, 0.1% of the SNPs were carried forward for a larger study (around 10,000 samples) which meant to reach statistical significance independent of the first stage. However, first, it was found that a joint analysis of both stages is more efficient compared to only considering the second stage [207, 208]. Later, with the declining costs of commercial genotyping, researchers abandoned the two-staged design and instead performed genome-wide analysis in a single-stage on the biggest sample size possible [209]. The importance of well-powered studies has thus been demonstrated in genomics. Again, borrowing from genomics, some proteomics researchers have argued that an analogous shift in concept is already occurring in proteomics [48, 69, 74].

Biomarker discovery using MSBAs pipelines. The large multiplexing afforded by MSBAs coupled with their cost effectiveness has motivated many large-scale biomarker studies. For example, Bachmann *et al.* screened 1000 proteins in 700 samples for cerebral malaria [210]. To improve the specificity of MSBAs, Borrebaeck's group developed and optimized arrays using antibody fragments instead of whole immunoglobulin molecules to improve the stability and function of these affinity binders [211, 212]. This group has been developing this platform for 15 years, tackling various technical and operational aspects inherent in antibody microarrays, and implementing useful bioinformatics concepts such as backwards eliminations which have proven helpful in identifying and discovering potential biomarker panels for breast cancer and pancreatic cancer [69]. Recently, this platform was applied to run a multi-center trial for the early detection of pancreatic ductal adenocarcinoma, using 350 samples to characterize the discriminatory power of 25 biomarkers [213], followed by a studying using 1300 pancreatic carcinoma patients [214]. This highly promising effort is currently headed by Immunovia (www.immunovia.com) who is running

prospective studies in the USA, and had their initial public offering (IPO) in February 2016. This example may demonstrate the value in applying the same platform across all stages and increasing the sample throughput at earlier stages in the pipeline. However, as discussed previously, MSBAs lack the specificity and hence the sensitivity to interrogate low abundance proteins with reliability, and it is thus unclear whether it is an appropriate model to follow.

Applications of PEA. By tackling the problem of rCR in MSAS, PEA rapidly scaled to 92-plex and is perhaps on the road to upending a decade of xMAP prominence. Addressing rCR also afforded improvements in precision and reproducibility, and allowed rapid development of different panels targeting different applications and proteins within the same pathways. Due to the reliance on Fluidigm's chip for qPCR, Olink has established analysis centers, offering these assays as services, which has helped the rapid adoption in the last few years. PEA has been utilized for several applications, including personalized medicine [78, 215] and biomarker discovery and validation in oncology [216–218]. However, despite its application to measure 5,000 samples [20], the cost per sample ($120) is still limiting for widespread population-level analysis. Indeed, the reliance on qPCR microfluidic chips to fractionate every sample to automate the 92 PCR reactions per sample is limiting, and proximity-based assays may ultimately require a different mode of signal transduction for increased scaling.

SOMAmers: ready for prime-time? The SOMAscan platform meets many of the requirements of the ideal affinity-based proteomic platform and the number of binders has grown steadily over the years, which has allowed interrogating fundamental to clinical questions, notably looking for biomarkers in cardiovascular and Alzheimer's disease [219]. For example, it allowed interrogating associations between plasma profiles and genetic data [20, 46]. Recently the platform was used to associate the levels of approximately 3000 plasma proteins with 10 million DNA variants in 3,301 healthy individuals [20]. On the other end, Pfizer used the platform to run a retrospective study to investigate the potential of using proteomic profiling in clinical trials, and indeed, a panel of proteins biomarkers indicative of toxicity was detected much earlier than clinical indications [220]. While the platform is semi-quantitative in nature [221]–presumably because spiking and measurement of thousands of standards may be problematic–its use in large scale studies provides sufficient means for normalization and comparative analyses. On the other hand, as shown in **Figure 2.15**, SOMAscan services can still cost upwards of $700 per sample, which is prohibitively expensive for most. Furthermore, the proprietary binders (and their sequences) are kept as a trade secret, which has also limited widespread adoption in the early stages. To overcome this problem, it appears that the company has changed its business model to an insight-driven one, offering many of its services to large partners (*e.g.* universities, health centers) at a loss in order to capitalize on

the biological and clinical insights gained later down the line [20].

2.3.3 Outlook

No silver bullet. In conclusion, it is evident that technologies for improved protein analysis are still needed. The rapid adoption of emerging platforms with superior attributes is in a way another indication of need and willingness to attempt novel approaches. Furthermore, the recent technological developments may have initiated a resurgence of efforts towards biomarker discovery in plasma in the coming decade. On the other hand, it is not evident that the current technologies are capable in meeting these increased demands in multiplexing and sample throughput. Indeed, in view of **Figure 2.15**, it seems that current proteomics technologies are not yet *data-efficient*.

An all-in-one ELISA? As discussed throughput this chapter, several technical challenges limit the broader and more efficient implementations of many of the technology platforms that exist today. However, one particular challenge has evaded a solution for several decades, namely the multiplexing of immunoassays. Indeed, despite the commercial availability of 1,000-2,000 matched antibody pairs, the vast majority of them are often analyzed in single-plex assays such as ELISA, or concurrently with up to a 20 other markers in MSAs such as microparticle-based assays or microarrays. Hence, the challenge in immunoassays today may be formulated as follows:

Can all of the available antibody pairs be merged into a single, highly-multiplexed assay, while maintaining the high-throughput, cost-effectiveness, and sensitivity of the single-plex format?

Such a formidable task would necessitate a truly scalable MSA platform that must meet at least two key challenges: (i) mitigating the scale-dependent rCR, and (ii) the development and use of an appropriate high-throughput, high-multiplexing, and cost-efficient multiplexing platform. While the first goal may have seemed unsurmountable 10 years ago, the advent of PEA, ACM, and other solutions have nullified that assumption and demonstrated increased multiplexing. However, as discussed, both PEA and ACM are instead limited by the multiplexing platform used, *i.e.* qPCR and antibody microarrays. On the other hand, despite the potential of microparticle-based MSAs for scalable multiplexing, this platform still has not seen a solution for rCR.

In this dissertation, we set out to address the two technical challenges in microparticle-based MSAs, namely (i) interference and energy transfer between fluorescent dyes when used for barcoding (**Chapter 3**), and (ii) the incompatibility and cross-reactivity of the antibody reagents when used in a multiplexed format (**Chapter 4**).

References

1. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422,** 835–847 (2003).

2. Ahn, A. C., Tewari, M., Poon, C. S. & Phillips, R. S. The limits of reductionism in medicine: Could systems biology offer an alternative? *PLoS Medicine* **3,** 0709–0713 (2006).

3. Sorger, P. K. A reductionist's systems biology: Opinion. *Current Opinion in Cell Biology* **17,** 9–11 (2005).

4. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biology* **18,** 1–15 (2017).

5. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nature Reviews Genetics* **19,** 299–310 (2018).

6. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).

7. Larance, M. & Lamond, A. I. Multidimensional proteomics for cell biology. *Nature Reviews Molecular Cell Biology* **16,** 269–280 (2015).

8. Mathew, A. K. & Padmanaban, V. C. Metabolomics: The apogee of the omics trilogy. *International Journal of Pharmacy and Pharmaceutical Sciences* **5,** 45–48 (2013).

9. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16,** 321–332 (2015).

10. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13,** 731–740 (2016).

11. Swan, A. L., Mobasheri, A., Allaway, D., Liddell, S. & Bacardit, J. Application of Machine Learning to Proteomics Data: Classification and Biomarker Identification in Postgenomics Biology. *OMICS: A Journal of Integrative Biology* **17,** 595–610 (2013).

12. Kitano, H. Systems biology: a brief overview. *Science* **295,** 1662–1664 (2002).

13. Hood, L., Heath, J., Phelps, M. & Lin, B. Systems Biology and New Technologies Enable Predictive and Preventive Medicine. *Science* **306,** 640–643 (2004).

14. Precision Medicine Initiative (PMI) Working Group. The precision medicine initiative cohort program – building a research foundation for 21st century medicine. *Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, NIH* **Sept 17,** 1–108 (2015).

15. Collins, F. S. & Varmus, H. A new initiative on precision medicine. *New England Journal of Medicine,* 793–795 (2015).

16. Price, N. D. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nature Biotechnology* **35,** 747–756 (2017).

17. Chambliss, A. B. & Chan, D. W. Precision medicine: From pharmacogenomics to pharmacoproteomics. *Clinical Proteomics* **13,** 1–9 (2016).

18. Federici, G. *et al.* Systems Analysis of the NCI-60 Cancer Cell Lines by Alignment of Protein Pathway Activation Modules with "-OMIC" Data Fields and Therapeutic Response Signatures. *Molecular Cancer Research* **11,** 676–685 (2013).

19. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534,** 55–62 (2016).

20. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558,** 73–79 (2018).

21. Pease, A. C. *et al.* Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 5022–6 (1994).

22. Ragoussis, J. & Elvidge, G. Affymetrix GeneChip® system: Moving from research to the clinic. *Expert Review of Molecular Diagnostics* **6,** 145–152 (2006).

23. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *American Journal of Human Genetics* **90,** 7–24 (2012).

24. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17,** 333–351 (2016).

25. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics* **101,** 5–22 (2017).

26. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74,** 5463–5467 (1977).

27. Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470,** 198–203 (2011).

28. Metzker, M. L. Sequencing technologies the next generation. *Nature Reviews Genetics* **11,** 31–46. arXiv: 20 9 (2010).

29. Altshuler, D. L. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).

30. Worthey, E. A. *et al.* Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genetics in Medicine* **13,** 255–262 (2011).

31. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics* **42,** 30–35. arXiv: 15334406 (2010).

32. Ashley, E. A. *et al.* Clinical assessment incorporating a personal genome. *The Lancet* **375,** 1525–1535 (2010).

33. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148,** 1293–1307 (2012).

34. Leblanc, V. G. & Marra, M. A. Next-generation sequencing approaches in cancer: Where have they brought us and wherewill they take us? *Cancers* **7,** 1925–1958 (2015).

35. Wan, J. C. *et al.* Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nature Reviews Cancer* **17,** 223–238 (2017).

36. Tan, D. S. W. *et al.* Biomarker-driven early clinical trials in oncology: a paradigm shift in drug development. *Cancer journal (Sudbury, Mass.)* **15,** 406–420 (2009).

37. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509,** 371–375 (2014).

38. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine* **9,** 1–12 (2017).

39. *Genome sequencing cost timeline* https://www.genome.gov/27541954/dna-sequencing-costs-data/. Accessed: 2018-08-10.

40. Stephens, Z. D. *et al.* Big data: Astronomical or genomical? *PLoS Biology* **13,** 1–11 (2015).

41. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165,** 535–550 (2016).

42. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13,** 227–232 (2012).

43. Cayer, D. M., Nazor, K. L. & Schork, N. J. Mission critical: The need for proteomics in the era of next-generation sequencing and precision medicine. *Human Molecular Genetics* **25,** R182–R189 (2016).

44. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **3247,** 3247 (2018).

45. Brodin, P. & Davis, M. M. Human immune system variation. *Nature Reviews Immunology* **17,** 21–29 (2017).

46. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nature Communications* **8** (2017).

47. Anderson, N. L. The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry* **56,** 177–185 (2010).

48. Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasmaproteomics. *Molecular Systems Biology* **13,** 942 (2017).

49. Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature reviews. Drug discovery* **5,** 310–20 (2006).

50. Brown, M. & Wittwer, C. Flow cytometry: Principles and clinical applications in hematology. *Clinical Chemistry* **46,** 1221–1229 (2000).

51. Orth, A. P., Batalov, S., Perrone, M. & Chanda, S. K. The promise of genomics to identify novel therapeutic targets. *Expert opinion on therapeutic targets* **8,** 587–596 (2004).

52. Morris, M. K., Chi, A., Melas, I. N. & Alexopoulos, L. G. Phosphoproteomics in drug discovery. *Drug Discovery Today* **19,** 425–432 (2014).

53. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery* **16,** 19–34 (2016).

54. Dias, M. H., Kitano, E. S., Zelanis, A. & Iwai, L. K. Proteomics and drug discovery in cancer. *Drug Discovery Today* **21,** 264–277 (2016).

55. Jager, W. D., Bourcier, K., Rijkers, G. T., Prakken, B. J. & Seyfert-margolis, V. Prerequisites for cytokine measurements in clinical trials with multiplex immunoassays. **11,** 1–11 (2009).

56. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Research* **44,** D710–D716 (2016).

57. Smith, L. M. & Kelleher, N. L. Proteoform: A single term describing protein complexity. *Nature Methods* **10,** 186–187 (2013).

58. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome. *Molecular & Cellular Proteomics* **1,** 845–867 (2002).

59. Schwenk, J. M., Lindberg, J., Sundberg, M., Uhlén, M. & Nilsson, P. Determination of binding specificities in highly multiplexed bead-based assays for antibody proteomics. *Molecular & cellular proteomics : MCP* **6,** 125–32 (2007).

60. Bhat, A., Dakna, M. & Mischak, H. Clinical Proteomics. *JAMA : the journal of the American Medical Association* **1243,** 237–253 (2001).

61. Solier, C. & Langen, H. Antibody-based proteomics and biomarker research-current status and limitations. *Proteomics* **14,** 774–783 (2014).

62. Frantzi, M., Latosisnka, A., Kontostathi, G. & Mischak, H. Clinical proteomics: closing the gap from discovery to implementation. *Proteomics,* 1700463 (2018).

63. Meo, A. D. *et al.* What is wrong with clinical proteomics? *Clinical Chemistry* **60,** 1258–1266 (2014).

64. Anderson, N. L., Ptolemy, A. S. & Rifai, N. The riddle of protein diagnostics: Future bleak or bright? *Clinical Chemistry* **59,** 194–197 (2013).

65. Anderson, L., Razavi, M., Skates, S., Anderson, N. G. & Pearson, T. W. Squeezing more value from the analytes we have: Personal baselines for multiple analytes in serial DBS. *Bioanalysis* **8,** 1539–1542 (2016).

66. *FDA Discussion Paper on Laboratory Developed Tests* https://www.fda.gov/medicaldevices/productsandmedicalprocedures/invitrodiagnostics/laboratorydevelopedtests/default.htm. Accessed: 2018-08-10.

67. Mor, G. *et al.* Serum protein markers for early detection of ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 7677–82 (2005).

68. Demler, O. V., Pencina, M. J. & D'Agostino, R. B. Impact of correlation on predictive ability of biomarkers. *Statistics in Medicine* **32,** 4196–4210 (2013).

69. Borrebaeck, C. A. Precision diagnostics: Moving towards protein biomarker signatures of clinical utility in cancer. *Nature Reviews Cancer* **17,** 199–204 (2017).

70. Landegren, U., Al-Amin, R. A. & Björkesten, J. A myopic perspective on the future of protein diagnostics. *New Biotechnology,* 0–1 (2018).

71. Mischak, H. *et al.* Recommendations for biomarker identification and qualification in clinical proteomics. *Science Translational Medicine* **2,** 1–7 (2010).

72. Dakna, M. *et al.* Addressing the Challenge of Defining Valid Proteomic Biomarkers and Classifiers. *BMC Bioinformatics* **11** (2010).

73. Mahotka, C., Wenzel, M., Springer, E., Gabbert, H. E. & Gerharz, C. D. Advances in Brief Survivin-Ex3 and Survivin-2B : Two Novel Splice Variants of the Apoptosis Inhibitor Survivin with Different Antiapoptotic Properties 1. *Cancer research* **59,** 6097–6102 (1999).

74. Landegren, U. *et al.* Opportunities for sensitive plasma proteome analysis. *Analytical Chemistry* **84,** 1824–1830 (2012).

75. MacBeath, G. Protein microarrays and proteomics. *Nature Genetics* **32,** 526–532 (2002).

76. Smith, J. G. & Gerszten, R. E. Emerging affinity-based proteomic technologies for large-scale plasma profiling in cardiovascular disease. *Circulation* **135,** 1651–1664 (2017).

77. Schwenk, J. M. *et al.* The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *Journal of Proteome Research* **16,** 4299–4310 (2017).

78. Enroth, S., Johansson, A., Enroth, S. B. & Gyllensten, U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nature communications* **5,** 4684 (2014).

79. Anderson, L. & Anderson, N. G. High resolution two-dimensional electrophoresis of human plasma proteins. *Proceedings of the National Academy of Sciences of the United States of America* **74,** 5421–5 (1977).

80. Hughes, G. J. *et al.* Plasma protein map: An update by microsequencing. *Electrophoresis* **13,** 707–714 (1992).

81. Bystrom, S. *Affinity assays for profiling disease-associated proteins in human plasma* PhD thesis (2017).

82. Keshishian, H., Addona, T., Burgess, M., Kuhn, E. & Carr, S. a. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Molecular & cellular proteomics : MCP* **6,** 2212–29 (2007).

83. Carr, S. A. *et al.* Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-based Assay Development Using a Fit-for-Purpose Approach. *Molecular & Cellular Proteomics* **13,** 907–917 (2014).

84. Hu, A., Noble, W. S. & Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research* **5,** 419 (2016).

85. Rauniyar, N. & Yates, J. R. Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of Proteome Research* **13,** 5293–5309 (2014).

86. Percy, A. J., Chambers, A. G., Yang, J. & Borchers, C. H. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics* **13,** 2202–2215 (2013).

87. Whiteaker, J. R., Zhao, L., Anderson, L. & Paulovich, A. G. An Automated and Multiplexed Method for High Throughput Peptide Immunoaffinity Enrichment and Multiple Reaction Monitoring Mass Spectrometry-based Quantification of Protein Biomarkers. *Molecular & Cellular Proteomics* **9,** 184–196 (2010).

88. Zahedi, R. & Sickmann, A. Analysis of Post-translational Modifications. *Proteomics* **13,** 901–903 (2013).

89. Kelstrup, C. D. *et al.* Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field orbitrap mass spectrometer. *Journal of Proteome Research* **13,** 6187–6195 (2014).

90. Heidelberger, M. & Kendall, F. E. A quantitative study of the precipitin reaction between type III Pneumococcus Polysaccharide and purified homologous antibody. *The Journal of experimental medicine* **50,** 809–23 (1929).

91. Yalow, R. S. & Berson, S. A. Immunoassay of endogenous plasma insulin in man. *Journal Clinical Investigation,* 1157–1175 (1960).

92. Lequin, R. M. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clinical Chemistry* **51,** 2415–2418 (2005).

93. Wide, L. Inventions leading to the development of the diagnostic test kit industry - From the modern pregnancy test to the sandwich assays. *Upsala Journal of Medical Sciences* **110,** 193–216 (2005).

94. Wild, D. *The Immunoassay Handbook* (ed Wild, D.) 48–57 (Oxford, 2013).

95. Wilson, A. C. *et al.* A Scalable Pipeline for High-Throughput Flow Cytometry. *SLAS Discovery* **23,** 708–718 (2018).

96. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry* **8,** 871–874 (1971).

97. Kramer, A. *et al.* Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody. *Cell* **91,** 799–809 (1997).

98. James, L. C., Roversi, P. & Tawfik, D. S. Antibody multispecificity mediated by conformational diversity. *Science* **299,** 1362–1367 (2003).

99. Michaud, G. a. *et al.* Analyzing antibody specificity with whole proteome microarrays. *Nature biotechnology* **21,** 1509–12 (2003).

100. Zichi, D., Eaton, B., Singer, B. & Gold, L. Proteomics and diagnostics: Let's Get Specific, again. *Current opinion in chemical biology* **12,** 78–85 (2008).

101. James, L. & Tawfik, D. S. The specificity of cross-reactivity : Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Science,* 2183–2193 (2003).

102. Gulmann, C., Sheehan, K. M., Kay, E. W., Liotta, L. a. & Petricoin, E. F. Array-based proteomics: mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. *The Journal of pathology* **208,** 595–606 (2006).

103. Schwenk, J. M. *et al.* Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays. *Molecular & cellular proteomics : MCP* **9,** 2497–507 (2010).

104. Baker, M. Antibody anarchy: A call to order. *Nature* **527,** 545–551 (2015).

105. Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521,** 274–276 (2015).

106. Uhlen, M. *et al.* A proposal for validation of antibodies. *Nat. Methods* **13,** 823–827 (2016).

107. Selby, C. Interference in Immunoassay. *Annals of clinical biochemistry,* 704–721 (1999).

108. Cummings, J., Ward, T. H. & Dive, C. Fit-for-purpose biomarker method validation in anticancer drug development. *Drug Discovery Today* **15,** 816–825 (2010).

109. Lee, J. W. *et al.* Fit-for-purpose method development and validation for successful biomarker measurement. *Pharmaceutical Research* **23,** 312–328 (2006).

110. Wilson, R. Sensitivity and specificity: twin goals of proteomics assays. Can they be combined? *Expert review of proteomics* **10,** 135–49 (2013).

111. Juncker, D., Bergeron, S., Laforte, V. & Li, H. Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. *Current Opinion in Chemical Biology* **18,** 29–37 (2014).

112. Wide, L., Bennich, H. & Johansson, S. Diagnosis of Allergy By an in-Vitro Test for Allergen Antibodies. *The Lancet* **290,** 1105–1107 (1967).

113. Ekins, R. P. Multi-analyte immunoassay. *Journal of pharmaceutical and biomedical analysis* **7,** 155–168 (1989).

114. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplex era: Potentials and pitfalls. *Proteomics - Clinical Applications* **9,** 406–422 (2015).

115. Chandra, H., Reddy, P. J. & Srivastava, S. Protein microarrays and novel detection platforms. *Expert review of proteomics* **8,** 61–79 (2011).

116. Nolan, J. P. & Sklar, L. a. Suspension array technology: evolution of the flat-array paradigm. *Trends in biotechnology* **20,** 9–12 (2002).

117. Wilson, R., Cossins, A. R. & Spiller, D. G. Encoded microcarriers for high-throughput multiplexed detection. *Angewandte Chemie - International Edition* **45,** 6104–6117 (2006).

118. Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J. & Kettman, J. R. Advanced multiplexed analysis with the FlowMetrix(TM) system. *Clinical Chemistry* **43,** 1749–1756 (1997).

119. Tang, H., Panemangalore, R., Yarde, M., Zhang, L. & Cvijic, M. E. 384-Well Multiplexed Luminex Cytokine Assays for Lead Optimization. *Journal of Biomolecular Screening* **21,** 548–555 (2016).

120. Lin, G., Baker, M. A., Hong, M. & Jin, D. The Quest for Optical Multiplexing in Bio-discoveries. *Chem* **4,** 997–1021 (2018).

121. Han, M., Gao, X., Su, J. Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat. Biotechnol.* **19,** 631–635 (2001).

122. Nicewarner-Pena, S. R. *et al.* Submicrometer metallic barcodes. *Science (New York, N.Y.)* **294,** 137–41 (2001).

123. Lee, H., Kim, J., Kim, H., Kim, J. & Kwon, S. Colour-barcoded magnetic microparticles for multiplexed bioassays. *Nature materials* **9,** 745–749 (2010).

124. Lee, J. *et al.* Universal process-inert encoding architecture for polymer microparticles. *Nature Materials* **13,** 524–529 (2014).

125. Nam, J.-M., Thaxton, C. S. & Mirkin, C. a. Nanoparticle-based bio-bar codes for the ultra-sensitive detection of proteins. *Science (New York, N.Y.)* **301,** 1884–6 (2003).

126. Leng, Y., Sun, K., Chen, X. & Li, W. Suspension arrays based on nanoparticle-encoded microspheres for high-throughput multiplexed detection. *Chem. Soc. Rev.* **25,** 23–25 (2015).

127. Wu, W. *et al.* Antibody array analysis with label-based detection and resolution of protein size. *Molecular & cellular proteomics : MCP* **8,** 245–57 (2009).

128. *Luminex Kit Finder* `https://kitfinder.luminexcorp.com/`. Accessed: 2018-08-10.

129. Stuchlý, J. *et al.* An automated analysis of highly complex flow cytometry-based proteomic data. *Cytometry A* **81A,** 120–129 (2012).

130. Holm, A., Wu, W. & Lund-Johansen, F. Antibody array analysis of labelled proteomes: how should we control specificity? *New biotechnology* **29,** 578–85 (2012).

131. Stuchlý, J. *et al.* An automated analysis of highly complex flow cytometry-based proteomic data. *Cytometry Part A* **81A,** 120–129 (2012).

132. Ji, X.-h. *et al.* On-demand preparation of quantum dot-encoded microparticles using a droplet microfluidic system. *Lab on a chip,* 2561–2568 (2011).

133. Ming, K. *et al.* Integrated Quantum Dot Barcode Smartphone Optical Device for Wireless Multiplexed Diagnosis of Infected Patients. *ACS nano* **9,** 3060–3074 (2015).

134. Fournier-Bidoz, S. *et al.* Facile and Rapid One-Step Mass Preparation of Quantum-Dot Barcodes. *Angewandte Chemie* **120,** 5659–5663 (2008).

135. Vaidya, S. V., Couzis, A. & Maldarelli, C. Reduction in aggregation and energy transfer of quantum dots incorporated in polystyrene beads by kinetic entrapment due to cross-linking during polymerization. *Langmuir* **31,** 3167–3179 (2015).

136. Kagan, C., Murray, C. & Bawendi, M. Long-range resonance transfer of electronic excitations in close-packed CdSe quantum-dot solids. *Physical Review B* **54,** 8633–8643 (1996).

137. Clapp, A. R., Medintz, I. L. & Mattoussi, H. Förster resonance energy transfer investigations using quantum-dot fluorophores. *ChemPhysChem* **7,** 47–57 (2006).

138. Bilan, R. S. *et al.* Engineering of Optically Encoded Microbeads with FRET-Free Spatially Separated Quantum-Dot Layers for Multiplexed Assays. *ChemPhysChem* **18,** 970–979 (2017).

139. Lu, S. *et al.* Three-Dimensional Barcodes with Ultrahigh Encoding Capacities: A Flexible, Accurate, and Reproducible Encoding Strategy for Suspension Arrays. *Chemistry of Materials* **29,** 10398–10408 (2017).

140. Gerver, R. E. *et al.* Programmable microfluidic synthesis of spectrally encoded microspheres. *Lab on a Chip* **12,** 4716 (2012).

141. Nguyen, H. Q. *et al.* Programmable microfluidic synthesis of over one thousand uniquely identifiable spectral codes. *Adv. Opt. Mater.* **5,** 1–6 (2016).

142. Abcam. *Firefly particle technology for multiplex miRNA assays*

143. Slaastad, H. *et al.* Multiplexed immuno-precipitation with 1725 commercially available antibodies to cellular proteins. *Proteomics* **11,** 4578–82 (2011).

144. Ingvarsson, J. *et al.* Design of recombinant antibody microarrays for serum protein profiling: Targeting of complement proteins. *Journal of Proteome Research* **6,** 3527–3536 (2007).

145. Haab, B. B., Dunham, M. J. & Brown, P. O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biology* **2,** 4–1 (2001).

146. Betzen, C. *et al.* Clinical proteomics: Promises, challenges and limitations of affinity arrays. *Proteomics - Clinical Applications* **9,** 342–347 (2015).

147. Jani, I. V., Janossy, G., Brown, D. W. G. & Mandy, F. Multiplexed immunoassays by flow cytometry for diagnosis and surveillance of infectious diseases in resource-poor settings Multiplexed immunoassays for infectious diseases. *The Lancet* **2,** 243–250 (2002).

148. Chowdhury, F., Williams, A. & Johnson, P. Validation and comparison of two multiplex technologies, Luminex®and Mesoscale Discovery, for human cytokine profiling. *Journal of Immunological Methods* **340,** 55–64 (2009).

149. Tighe, P., Negm, O., Todd, I. & Fairclough, L. Utility, reliability and reproducibility of immunoassay multiplex kits. *Methods (San Diego, Calif.)* 1–7 (2013).

150. Pla-Roca, M. *et al.* Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & cellular proteomics : MCP* **11** (2012).

151. Landegren, U. AFFINOMICS and the prospects for large-scale protein analyses. *New Biotechnology* **33,** 491–493 (2016).

152. Gonzalez, R. M. *et al.* Development and Validation of Sandwich ELISA Microarrays with Minimal Assay Interference. *J. Proteome Res.* **7,** 2406–2414 (2008).

153. Perlee, L. *et al.* Development and standardization of multiplexed antibody microarrays for use in quantitative proteomics. *Proteome science* **2,** 9 (2004).

154. Schweitzer, B. *et al.* Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnology* **20,** 359–365 (2002).

155. Jani, D. *et al.* Recommendations for Use and Fit-for-Purpose Validation of Biomarker Multiplex Ligand Binding Assays in Drug Development. *The AAPS Journal* **18,** 1–14 (2016).

156. Butterfield, L. H., Potter, D. M. & Kirkwood, J. M. Multiplex serum biomarker assessments: technical and biostatistical issues. *Journal of translational medicine* **9,** 173 (2011).

157. Djoba Siawaya, J. F. *et al.* An evaluation of commercial fluorescent bead-based luminex cytokine assays. *PLoS ONE* **3,** 1–12 (2008).

158. De Koning, L. *et al.* A multiplex immunoassay gives different results than singleplex immunoassays which may bias epidemiologic associations. *Clinical biochemistry* **45,** 848–51 (2012).

159. Kakabakos, S. E., Christopoulos, T. K. & Diamandis, E. P. Multianalyte immunoassay based on spatially distinct fluorescent areas quantified by laser-excited solid-phase time-resolved fluorometry. *Clinical chemistry* **38,** 338–42 (1992).

160. Fu, Q., Schoenhoff, F. S., Savage, W. J., Zhang, P. & Van Eyk, J. E. Multiplex assays for biomarker research and clinical application: translational science coming of age. *Proteomics. Clinical applications* **4,** 271–84 (2010).

161. Fredolini, C. *et al.* Immunocapture strategies in translational proteomics. *Expert Review of Proteomics* **13,** 83–98 (2016).

162. Ellington, A. a., Kullo, I. J., Bailey, K. R. & Klee, G. G. Measurement and quality control issues in multiplex protein assays: a case study. *Clinical chemistry* **55,** 1092–9 (2009).

163. De Jager, W., Bourcier, K., Rijkers, G. T., Prakken, B. J. & Seyfert-Margolis, V. Prerequisites for cytokine measurements in clinical trials with multiplex immunoassays. *BMC immunology* **10,** 52 (2009).

164. Geyer, P. E. *et al.* Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Molecular Systems Biology* **12,** 901 (2016).

165. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Systems* **2,** 185–195 (2016).

166. Li, H., Popp, R. & Borchers, C. H. Affinity-mass spectrometric technologies for quantitative proteomics in biological fluids. *TrAC - Trends in Analytical Chemistry* **90,** 80–88 (2017).

167. Razavi, M., Leigh Anderson, N., Pope, M. E., Yip, R. & Pearson, T. W. High precision quantification of human plasma proteins using the automated SISCAPA Immuno-MS workflow. *New Biotechnology* **33,** 494–502 (2016).

168. Gullberg, M. *et al.* Cytokine detection by antibody-based proximity ligation. *Proceedings of the National Academy of Sciences* **101,** 8420–8424 (2004).

169. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic acids research* **39** (2011).

170. Lundberg, M. *et al.* Multiplexed homogeneous proximity ligation assays for high-throughput protein biomarker research in serological material. *Molecular & cellular proteomics : MCP* **10,** M110.004978 (2011).

171. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS ONE* **9** (2014).

172. Kraemer, S. *et al.* From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PloS one* **6,** e26332 (2011).

173. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5** (2010).

174. Loffredo, F. S. *et al.* Growth Differentiation Factor 11 Is a Circulating Factor that Reverses Age-Related Cardiac Hypertrophy. *Cell* **153,** 828–839 (2013).

175. Walker, R. G. *et al.* Biochemistry and Biology of GDF11 and Myostatin Similarities, Differences, and Questions for Future Investigation Ryan. *Circulation Research* **118,** 1125–1142 (2016).

176. Li, H., Bergeron, S. & Juncker, D. Microarray-to-Microarray Transfer of Reagents by Snapping of Two Chips for Cross-Reactivity-Free Multiplex Immunoassays. *Analytical chemistry* **84,** 4776–83 (2012).

177. Li, H., Munzar, J. D., Ng, A. & Juncker, D. A versatile snap chip for high-density subnanoliter chip-to-chip reagent transfer. *Scientific Reports* **5,** 1–12 (2015).

178. Frampton, J. P. *et al.* Aqueous two-phase system patterning of detection antibody solutions for cross-reaction-free multiplex ELISA. *Scientific reports* **4,** 4878 (2014).

179. Jarolim, P. High sensitivity cardiac troponin assays in the clinical laboratories. *Clinical Chemistry and Laboratory Medicine* **53,** 635–652 (2015).

180. Smith, M. a. *et al.* Annotation of human cancers with EGFR signaling-associated protein complexes using proximity ligation assays. *Science signaling* **8** (2015).

181. Scrimin, P. & Prins, L. J. Sensing through signal amplification. *Chemical Society Reviews* **40,** 4488 (2011).

182. Spengler, M., Adler, M. & Niemeyer, C. M. Highly sensitive ligand-binding assays in preclinical and clinical applications: immuno-PCR and other emerging techniques. *The Analyst* **140,** 6175–6194 (2015).

183. Tekin, H. C. & Gijs, M. A. Ultrasensitive protein detection: A case for microfluidic magnetic bead-based assays. *Lab on a Chip* **13,** 4711–4739 (2013).

184. Fischer, S. K. *et al.* Emerging Technologies to Increase Ligand Binding Assay Sensitivity. *The AAPS Journal* **17,** 93–101 (2015).

185. Todd, J. *et al.* Ultrasensitive flow-based immunoassays using single-molecule counting. *Clinical Chemistry* **53,** 1990–1995 (2007).

186. Rissin, D. M. *et al.* Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nature biotechnology* **28,** 595–9 (2010).

187. Chang, L. *et al.* Single molecule enzyme-linked immunosorbent assays: Theoretical considerations. *Journal of Immunological Methods* **378,** 102–115 (2012).

188. Rissin, D. M. *et al.* Multiplexed single molecule immunoassays. *Lab on a Chip* **13,** 2902 (2013).

189. Walt, D. R. Optical methods for single molecule detection and analysis. *Analytical Chemistry* **85,** 1258–1263 (2013).

190. Rifai, N., Gillette, M. a. & Carr, S. a. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature biotechnology* **24,** 971–83 (2006).

191. Matthews, H., Hanison, J. & Nirmalan, N. "Omics"-Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives. *Proteomes* **4,** 28 (2016).

192. Gstaiger, M. & Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nature Reviews Genetics* **10,** 617–627 (2009).

193. Azad, N. S. *et al.* Proteomics in Clinical Trials and Practice. *Molecular & Cellular Proteomics* **5,** 1819–1829 (2006).

194. Murray, H. C., Dun, M. D. & Verrills, N. M. Harnessing the power of proteomics for identification of oncogenic, druggable signalling pathways in cancer. *Expert Opinion on Drug Discovery* **12,** 431–447 (2017).

195. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26,** 1367–1372 (2008).

196. Du, J. *et al.* Bead-based profiling of tyrosine kinase phosphorylation identifies SRC as a potential target for glioblastoma therapy. *Nature Biotechnology* **27,** 77–83 (2009).

197. Gurbel, P. A., Kreutz, R. P., Bliden, K. P., DiChiara, J. & Tantry, U. S. Biomarker analysis by fluorokine multianalyte profiling distinguishes patients requiring intervention from patients with long-term quiescent coronary artery disease: A potential approach to identify atherosclerotic disease progression. *American Heart Journal* **155,** 56–61 (2008).

198. Lemoine, J. *et al.* The current status of clinical proteomics and the use of MRM and MRM ³ for biomarker validation. *Expert Review of Molecular Diagnostics* **12,** 333–342 (2012).

199. Dahut, W. L. *et al.* A phase II clinical trial of sorafenib in androgen-independent prostate cancer. *Clinical Cancer Research* **14,** 209–214 (2008).

200. Yurkovetsky, Z. R. *et al.* Multiplex Analysis of Serum Cytokines in Melanoma Patients Treated with Interferon- 2b. *Clinical Cancer Research* **13,** 2422–2428 (2007).

201. Galasko, D. R. *et al.* Antioxidants for Alzheimer disease: A randomized clinical trial with cerebrospinal fluid biomarker measures. *Archives of Neurology* **69,** 836–841 (2012).

202. Akbani, R. *et al.* A pan-cancer proteomic perspective on the cancer genome atlas. *Nature Communications* **5** (2014).

203. Mueller, C., Liotta, L. A. & Espina, V. Reverse phase protein microarrays advance to use in clinical trials. *Molecular Oncology* **4,** 461–481 (2010).

204. Hayes, D. F. *et al.* Breaking a Vicious Cycle. *Science Translational Medicine* **5,** 196cm6–196cm6 (2013).

205. Nedelkov, D. Human proteoforms as new targets for clinical mass spectrometry protein tests. *Expert Review of Proteomics* **14,** 691–699 (2017).

206. Polanski, M. & Anderson, N. L. A List of candiate cancerbiomakers for targeted Proteomics. *Biomarker Insights* **2,** 1–48 (2006).

207. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* **38,** 209–213 (2006).

208. Scherag, A. *et al.* Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genetics* **6,** 2–11 (2010).

209. Thomas, D. C. *et al.* Methodological Issues in Multistage Genome-Wide Association Studies. *Statistical Science* **24,** 414–429 (2009).

210. Bachmann, J. *et al.* Affinity Proteomics Reveals Elevated Muscle Proteins in Plasma of Children with Cerebral Malaria. *PLoS Pathogens* **10** (2014).

211. Borrebaeck, C. A. K. & Wingren, C. Design of high-density antibody microarrays for disease proteomics: Key technological issues. *Journal of Proteomics* **72,** 928–935 (2009).

212. Wingren, C., Ingvarsson, J., Dexlin, L., Szul, D. & Borrebaeck, C. A. K. Design of recombinant antibody microarrays for complex proteome analysis: Choice of sample labeling-tag and solid support. *Proteomics* **7,** 3055–3065 (2007).

213. Gerdtsson, A. S. *et al.* A Multicenter Trial Defining a Serum Protein Signature Associated with Pancreatic Ductal Adenocarcinoma. *International Journal of Proteomics* **2015,** 1–10 (2015).

214. Gerdtsson, A. *et al.* Plasma protein profiling in a stage defined pancreatic cancer cohort - Implications for early diagnosis. *Mol Oncol,* 30061–30068 (2016).

215. Enroth, S., Enroth, S. B., Johansson, Å. & Gyllensten, U. Protein profiling reveals consequences of lifestyle choices on predicted biological aging. *Scientific Reports* **5,** 1–10 (2015).

216. Boylan, K. L. *et al.* A multiplex platform for the identification of ovarian cancer biomarkers. *Clinical Proteomics* **14,** 1–21 (2017).

217. Schneiderova, P. *et al.* Serum protein fingerprinting by PEA immunoassay coupled with a pattern-recognition algorithms distinguishes MGUS and multiple myeloma. *Oncotarget* **5,** 69408–69421 (2016).

218. Mahboob, S. *et al.* A novel multiplexed immunoassay identifies CEA, IL-8 and prolactin as prospective markers for Dukes' stages A-D colorectal cancers. *Clinical Proteomics* **12,** 1–12 (2015).

219. Ngo, D. *et al.* Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. *Circulation* **134,** 270–285 (2016).

220. Williams, S. A. *et al.* Improving Assessment of Drug Safety Through Proteomics: Early Detection and Mechanistic Characterization of the Unforeseen Harmful Effects of Torcetrapib. *Circulation* (2017).

221. Christiansson, L. *et al.* The use of multiplex platforms for absolute and relative protein quantification of clinical material. *EuPA Open Proteomics* **3,** 37–47 (2014).

PREFACE TO CHAPTER III

This chapter reports on the development of a high-capacity, microparticle barcoding platform using common organic dyes for encoding, and widely available cytometers for read-out. To this end, we report on the development of a multicolour-labeling method hand-in-hand with an ensemble multicolour FRET (emFRET) model to understand, predict, and design ensemble fluorescence response. The emFRET model was used to guide the encoding and decoding of microparticles, enabling the generation of high capacity barcodes using dyes with large spectral overlap.

The following manuscript has been published as an *Article* in *Nature Nanotechnology*.

Dagher, M., Kleinman, M., Ng,A. & Juncker, D. Ensemble multicolour FRET model enables barcoding at extreme FRET levels. *Nature Nanotechnology* **24**. doi:101038/s41565-018-0205-0 (2018).

Acknowledgements

We thank Thomas Gervais for helpful discussions, and Jeffrey Munzar for proofreading our manuscript. We thank NSERC and FQRNT for funding. M.D. acknowledges the NSERC-CREATE ISS program for support. The flow cytometry work was performed at two McGill core flow facilities, namely the Microbiology and Immunology (MIMM) department, and the Life Science Complex, which is supported by funding from the Canadian Foundation for Innovation.

Chapter 3

ENSEMBLE MULTICOLOUR FRET MODEL ENABLES BARCODING AT EXTREME FRET LEVELS

Milad Dagher[1,2], Michael Kleinman[1,2], Andy Ng[1,2], David Juncker[1,2,3]

[1] Biomedical Engineering Department, McGill University, 3775 University Street, Montreal, QC, H3A2B4, Canada

[2] McGill University and Genome Quebec Innovation Center, 740 Dr. Penfield Avenue, Montreal, QC, H3A0G1, Canada

[3] Neurology and Neurosurgery Department, McGill University, 3801 University Street, Montreal, QC, H3A2B4, Canada

3.1 Abstract

Quantitative models of Förster resonance energy transfer (FRET)—pioneered by Förster—define our understanding of FRET and underpin its widespread use. However, multicolour FRET (mFRET) that arises between multiple, stochastically distributed fluorophores lacks a mechanistic model and remains intractable. mFRET notably arises in fluorescently barcoded microparticles (BMPs), resulting in a complex, non-orthogonal fluorescence response that impedes their encoding and decoding. Here, we introduce an ensemble mFRET (emFRET) model, and apply it to guide barcoding into regimes with extreme FRET. We further introduce a facile, proportional multicolour labelling method using oligonucleotides as homogeneous linkers. 580 barcodes were rapidly designed and validated using four dyes—with FRET efficiencies reaching 76%—and used for multiplexed immunoassays with cytometric read-out and fully automated decoding. The emFRET model helped expand the barcoding capacity of BMPs using common organic dyes and will benefit other applications subject to stochastic mFRET.

3.2 Introduction

Barcoded microparticles (BMPs) are paramount for multiplexed suspension assays as they allow distinguishing probes from a large mixture [1, 2]. Fluorescent encoding of BMPs using precise proportions of multicolour classifier dyes is the most popular approach as it allows simple and high throughput read-out by flow cytometry [3, 4]. In an ideal BMP system, the concentrations and fluorescence intensities of differently-coloured classifier dyes are orthogonal and may be independently controlled to allow straightforward encoding and decoding [2]. However, spectral overlap

between common dyes—such as organic fluorophores and quantum dots—becomes unavoidable beyond 2 or 3 colours because of the limited spectral bandwidth available ($\sim 350-750$ nm), giving rise to multicolour Förster resonance energy transfer (mFRET) and cascades thereof [4–8]. As a result, efforts to expand the barcoding capacity are met with rapidly intensifying mFRET and an intractable ensemble fluorescence, imposing labour-intensive experimental iterations to obtain distinguishable barcode responses and barring fully-automated decoding [4, 5, 9]. In addition, common microparticle (MP) barcoding methods lead to hard-to-predict, non-proportional multicolour labelling, for example due to competing physical and chemical properties of the different dyes [2, 10–12].

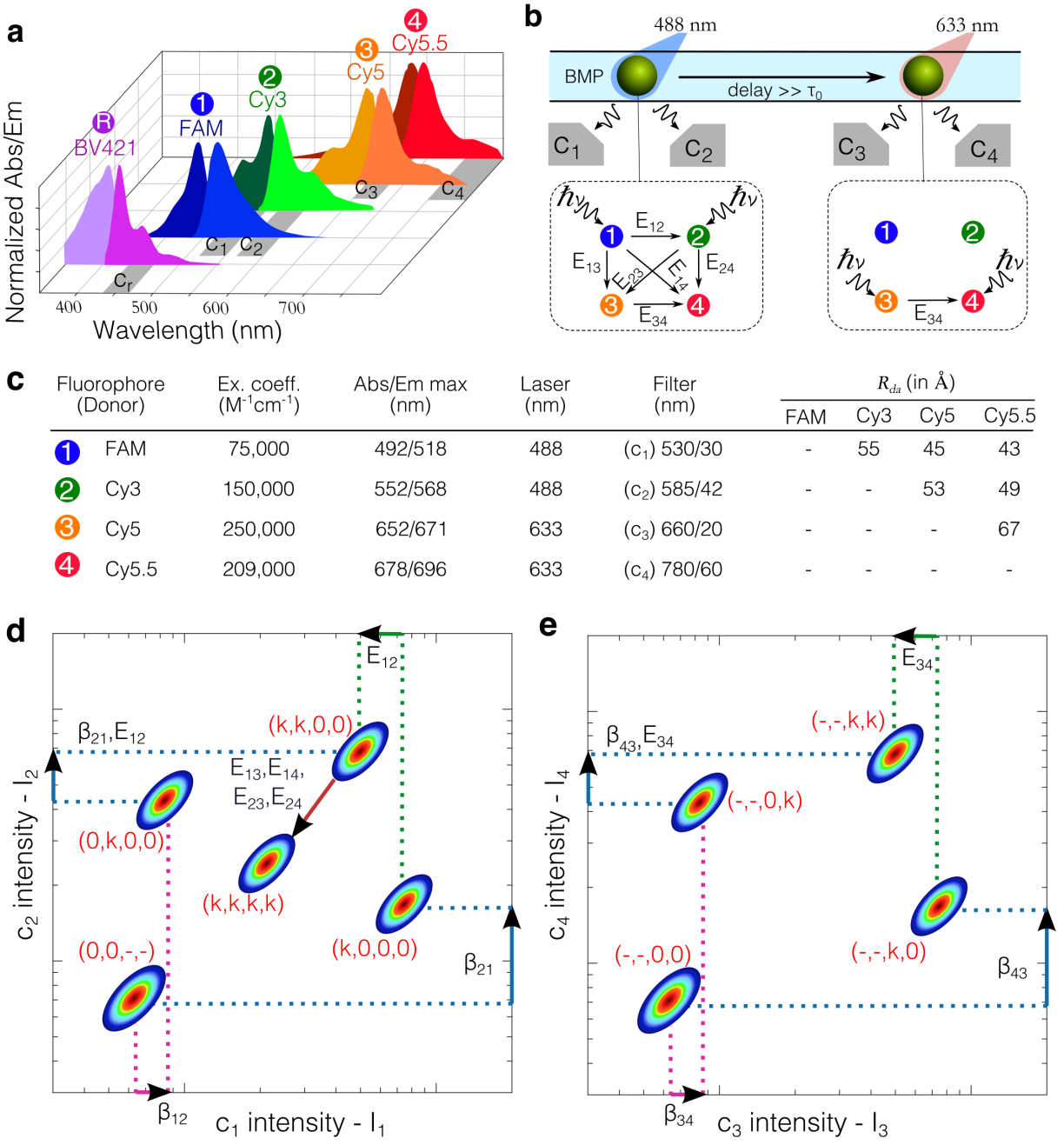
Analytical models for two-colour FRET (2FRET) between stochastically distributed dyes in a volume and on a surface were introduced by Förster [13], and later Wolber and Hudson [14] respectively. In recent years, there has been a growing interest in exploiting mFRET for tracking multiple intermolecular distances [15–19] and assembling energy-harvesting photonic networks [20]. However, mFRET models have been restricted to single-molecules with fixed inter-dye distances [15, 21, 22]. To the best of our knowledge, an analytical model describing mFRET between stochastically distributed dyes, as is the case for BMPs, is still lacking.

This report presents an ensemble mFRET (emFRET) model and incorporates it within a multicolour fluorescence model (MFM) for flow cytometry, which is used together with a precise labelling method employing DNA as a homogeneous crosslinker to accurately predict and control ensemble fluorescence of BMPs. The emFRET model represents an accessible analytical solution and affords quantitative insight into stochastic mFRET cascades, allowing rational design and fine-tuning of the spectral response. This approach enabled effective use of common, spectrally overlapping dyes by extending barcoding into extreme FRET regimes, resulted in accurate *in silico* barcode design and automatic read-out by flow cytometry, and was used for rapid, combinatorial screening of non-specific binding for sandwich immunoassays.

3.3 Results

3.3.1 Barcoding using spectrally overlapping dyes.

To emphasize the problem of mFRET in barcoding, we chose classifier dyes with strong spectral overlap (see Methods). The four classifier dyes in order of increasing wavelength are FAM, Cy3, Cy5, and Cy5.5, referred from here on as dyes **1** to **4** respectively (**Figure 3.1a**). This combination of dyes allows pairwise excitation and readout using common lasers and optical filters (**Figure 3.1b-c**). The dyes' absorption/emission spectra showcase the substantial spectral overlap (**Figure 3.1a**), and the calculated Förster radii R_{da} (where d and a are the donor and acceptor respectively) underline



	Fluorophore (Donor)	Ex. coeff. ($M^{-1}cm^{-1}$)	Abs/Em max (nm)	Laser (nm)	Filter (nm)	R_{da} (in Å)			
						FAM	Cy3	Cy5	Cy5.5
1	FAM	75,000	492/518	488	(c_1) 530/30	-	55	45	43
2	Cy3	150,000	552/568	488	(c_2) 585/42	-	-	53	49
3	Cy5	250,000	652/671	633	(c_3) 660/20	-	-	-	67
4	Cy5.5	209,000	678/696	633	(c_4) 780/60	-	-	-	-

Figure 3.1: **Spectrally overlapping classifier dyes and the impact of multicolour FRET on the fluorescence response of barcoded microparticles.** (**a**) Normalized absorption and emission spectra of the four spectrally overlapping classifier dyes (**1**-**4**), overlaid with the channel-specific emission filters in the FACS CANTO II cytometer ($\mathbf{c_1}$-$\mathbf{c_4}$) used here. For analyte detection, a blue-shifted reporter dye (R, BV-421) that does not interfere with barcode responses was selected. (**b**) Schematic representation of BMP readout by flow cytometry, indicating the lasers used for excitation and their corresponding channels. Direct excitation of dyes as well as potential energy transfer pathways are highlighted in each flow cell to show the propensity for mFRET and mFRET cascades. (**c**) Fluorophore photophysical properties, excitation/readout optics, and calculated inter-dye Förster radii (R_{da}). (Continued on next page)

Figure 3.1: **(d,e)** The effects of spectral overlap on the relative positions of BMP clusters given by their dye proportions $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ within the intensity spaces (**d**) I_1-I_2 and (**e**) I_3-I_4. Bleed-through is quantified by the fraction of dye f fluorescence leaking into channel c (β_{cf}). For example, BMPs $(k, 0, 0, 0)$ and $(-, -, k, 0)$, where k is an arbitrary non-zero number and '-' may take any value, would also be detected by $\mathbf{c_2}$ (panel d) and $\mathbf{c_4}$ (panel e) respectively. FRET, which is quantified by the efficiency of transfer (E_{da}) from donor d to acceptor a, occurs across all dyes in this setup and results in a strongly non-orthogonal response. Note that adding dye (**2**) to a BMP from $(k, 0, 0, 0)$ to $(k, k, 0, 0)$ in (**d**) can result in a decreased I_1 value due to E_{12}. Similarly, going from $(k, k, 0, 0)$ to (k, k, k, k), a decrease in I_1 and I_2 is expected in (**d**) because the presence of dyes **3** and **4** at a significant density will result in energy transfer to these long-wavelengths dyes (**Supplementary Figure 3.8**).

their propensity for energy transfer (**Figure 3.1c**). BV-421 was used as assay reporter dye as it is bright and spectrally isolated with an excitation at 405 nm.

Spectral barcodes are typically generated by modulating the proportions of the dyes' surface density, $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$, to generate well-resolved clusters arising from the intensity scatter plots in each channel pair, thereby allowing unambiguous decoding of the BMPs. Bleed-through and FRET, however, break down the orthogonality between the density (σ_f) and the cognate channel intensity (I_f), for a given dye f, which prevents straightforward barcoding (**Figure 3.1d-e**). Bleed-through is a linear effect at the detector level and can readily be accounted for by solving simultaneous linear equations [23]. FRET, on the other hand, creates a non-linear response to the dyes' surface densities and cannot be deconvolved as easily (**Figure 3.1d-e** and **Supplementary Figures 3.7-3.9**). In this work, laser excitation at 488 nm results in 6 potential inter-dye energy transfers with varying efficiencies, E_{da}, between donor d and acceptor a (**Figure 3.1b**). Hence, barcode responses in the I_1-I_2 intensity space are also dependent on (σ_3, σ_4) through the density-dependent energy transfer pathways E_{13}, E_{14}, E_{23}, and E_{24} (**Figure 3.1d**, **Supplementary Figure 3.8**). Without an accurate model to guide the design process, the non-linear nature of FRET imposes empirical optimizations of (σ_1, σ_2) values for every (σ_3, σ_4) value (**Supplementary Figure 3.9**). Furthermore, the number of optimization steps increases exponentially with every added classifier dye, collectively justifying the current practice that focuses on minimizing spectral overlap and mFRET, albeit at the expense of barcoding capacity.

3.3.2 Proportional labelling of multicolour dyes.

To establish a mechanistic mFRET model for surface immobilized dyes, it is necessary to achieve accurate control over the proportions of dye densities, which is a requisite not met by commonly used labelling techniques [2]. Here, as a foundation for our experiments, we introduce a widely

applicable one-pot microparticle labelling method. All classifier dyes were linked to the 3' end of an identical 21-nt DNA oligonucleotide that, when annealed to its complementary 5' biotinylated strand, served as a linker oligo (LO) for streptavidin-coupled MPs (**Supplementary Figure 3.11**). In this work, DNA is used solely as a homogeneous crosslinker to normalize reactivity and footprint across all classifier dyes. Each classifier dye (**1**-**4**) was conjugated to a LO (LO_1-LO_4). A non-fluorescently-labelled LO (LO_0) was used to balance and conserve the total amount of LOs in solution, and on the MP surface, independent of the particular barcode (**Figure 3.2a**). As a result, the proportions of coloured LOs in solution, given by (n_1, n_2, n_3, n_4) and defining the barcode, are faithfully replicated on the MP after labelling—that is, $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = t \times (n_1, n_2, n_3, n_4)$, where t is a labelling constant (**Supplementary Note 1.2**). BMPs co-labelled with mouse anti-goat immunoglobulin G (IgG) and LOs were characterized by cytometry (see Methods). The bead-to-bead coefficient-of-variation (CV) of the barcode response, which defines the cluster size, was $\approx 10\%$ and barcode-independent (**Figure 3.2b**).

To verify that the labelling reaction conserves dye proportions and affords independent surface densities of differently-coloured LOs, we measured I_3-I_4 for a set of BMPs with constant (n_3, n_4) but varying (n_1, n_2). I_3 and I_4 are neither impacted by bleed-through nor FRET from dyes **1** and **2** due to the spatial and temporal separation of the two excitation cells (**Figure 3.1b**). Hence, any detectable dependence of I_3-I_4 on (n_1, n_2) would unambiguously be ascribed to changes in (σ_3, σ_4) due to surface competition. The intensity of I_3-I_4 was constant for a wide range of n_1 and n_2 in barcodes $(n_1, n_2, 8, 10)$, demonstrating the independence of the dyes' surface densities (**Figure 3.2c**). The conservation of total LO also ensured barcode-independent antibody surface coverage (**Supplementary Figure 3.10**). The stability of the barcode response over long incubation times was investigated for single-colour BMPs and found to be satisfactory for the scope of this work (**Supplementary Figure 3.12**).

3.3.3 Multicolour fluorescence model.

To model barcode responses, and overcome the limitations of empirical barcoding by trial-and-error, we derive a general, platform-independent multicolour fluorescence model (MFM) that links the fluorescence intensities of every channel to the barcode-specific dye proportions. The MFM considers direct (*i.e.* laser) excitation of dyes, sensitization by FRET (including FRET cascades) and bleed-through. Here, the signal in a given channel is registered in response to only one laser. A MFM with a higher degree of generality, also considering channel intensities in response to an arbitrary number of lasers, is derived and presented in detail in **Supplementary Note 1**. Briefly, the signal in a given channel is modeled as the sum of the ensemble fluorescence intensities of N distinct dyes [23–25]. Accordingly, the equation for the intensity of channel c can be expressed as:

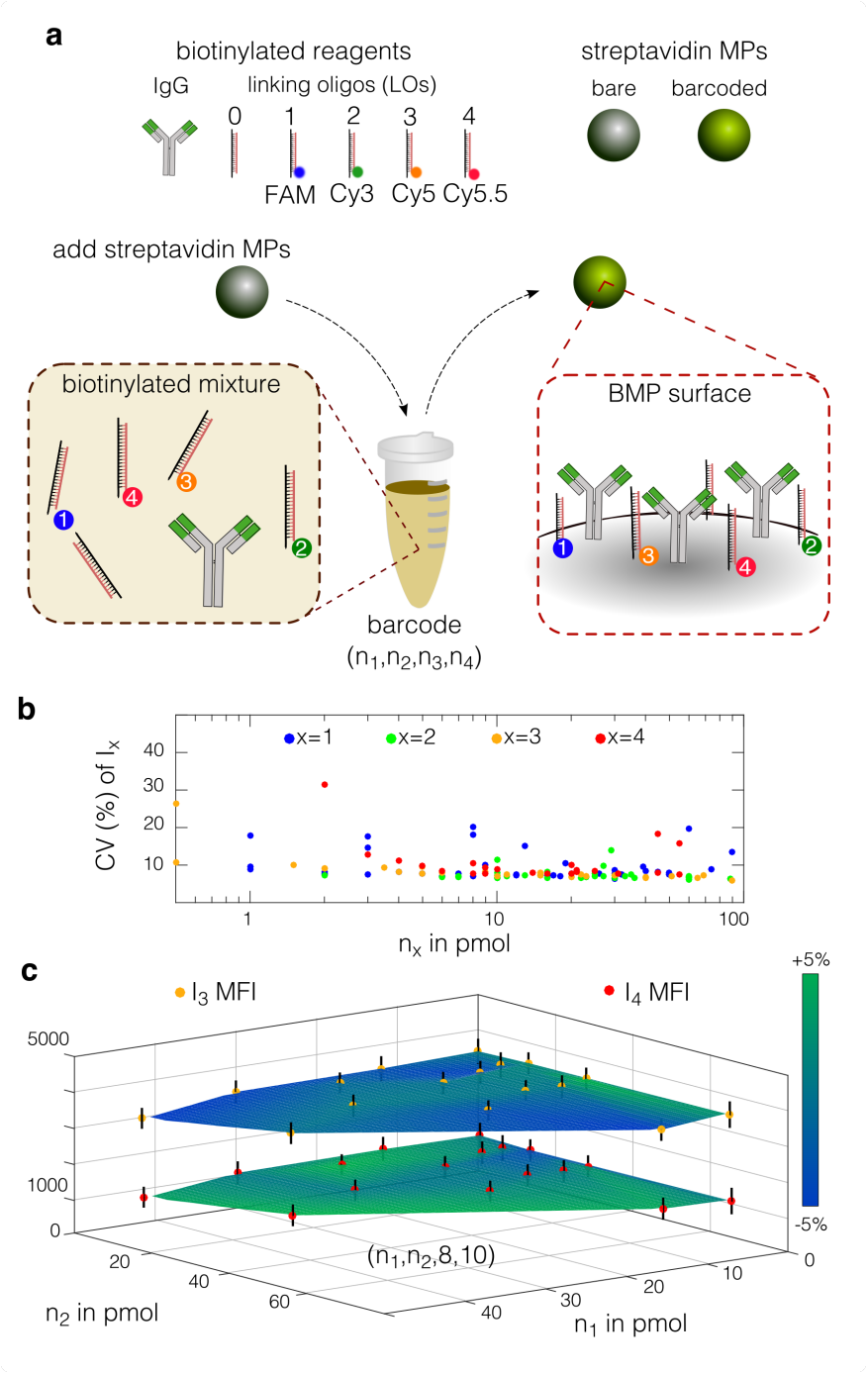


Figure 3.2: **One-pot DNA-assisted microparticle labelling conserves dye proportions.** (**a**) Cartoon illustrating the one-pot reaction of streptavidin-coated microparticles with a biotinylated mixture of LO-linked classifier dyes and antibodies for simultaneous encoding and functionalization. Four dye-conjugated LOs are mixed in precise proportions (n_1, n_2, n_3, n_4), defining the barcode, while a non-fluorescent LO_0 is added to conserve the total amount of LOs across all BMPs. (**b**) Coefficient-of-variation (CV) of the fluorescence intensity for 78 four-colour BMPs reported for each channel and its corresponding dye concentration, yielding a CV of $\approx$ 10% for most BMPs, regardless of dye concentration. (Continued on next page)

Figure 3.2: (**c**) Median fluorescence intensity (MFI) of I_3 and I_4 for BMPs $(n_1, n_2, 8, 10)$ in response to varying n_1 and n_2. The constancy of I_3 and I_4 values confirms that the density of (n_3, n_4) is independent of variations in (n_1, n_2) and that the labelling reaction conserves the solution proportions. Error bars represent the bead-to-bead standard-deviations.

$$I_c = I_c^0 + \sum_{f=1}^{N} \beta_{cf} F_f^e, \quad c = 1, \ldots, N, f = 1, \ldots, N \tag{3.1}$$

where I_c and I_c^0 are, respectively, the signal and background (*i.e.* bare MPs) in channel c when excited by the channel-specific laser, F_f^e is the ensemble fluorescence of dye f, and β_{cf} is the bleed-through ratio in channel c from dye f. To account for FRET cascades, the sensitized fluorescence, F_f^s, denotes the unattenuated ensemble emission (that is, considering only radiative decay) of dye f and is modelled as the sum of direct excitation as well as FRET excitation from all potential acceptors, which is a simplification afforded by the low exciton density [26] (**Supplementary Note 2.4**):

$$F_f^e = F_f^s (1 - \sum_{i=f+1}^{N} E_{fi}^{em}) \tag{3.2a}$$

$$F_f^s = F_f^0 + \sum_{j=1}^{f-1} \alpha_{jf} E_{jf}^{em} F_j^s \tag{3.2b}$$

where E_{da}^{em} is the ensemble-averaged multicolour FRET efficiency, α_{da} is the FRET proportionality constant that depends on the dyes' mutual optical properties and which can be seen as an 'energy exchange rate' [24], and F_f^0 is the basal fluorescence in response to direct excitation. These equations model steady-state FRET cascades whereby excitons may undergo multiple transfers before radiative emission. We consider operation in a regime where the basal fluorescence of dye f is proportional to its surface density (*i.e.* $F_f^0 \propto \sigma_f$). By considering that $\sigma_f = t n_f$, where t is a dye-independent labelling constant, the basal fluorescence may be expressed as $F_f^0 = \mu_f n_f$, where μ_f is the direct-excitation constant (**Supplementary Note 1.2**).

3.3.4 Ensemble multicolour FRET in 2D.

To use the MFM, it is necessary to establish the energy transfer distribution in eqs. (3.2a) and (3.2b) as a function of the dyes' surface densities. Ensemble 2FRET (e2FRET), whereby each donor may transfer excitons to a stochastically distributed array of similar acceptors, is mechanistically

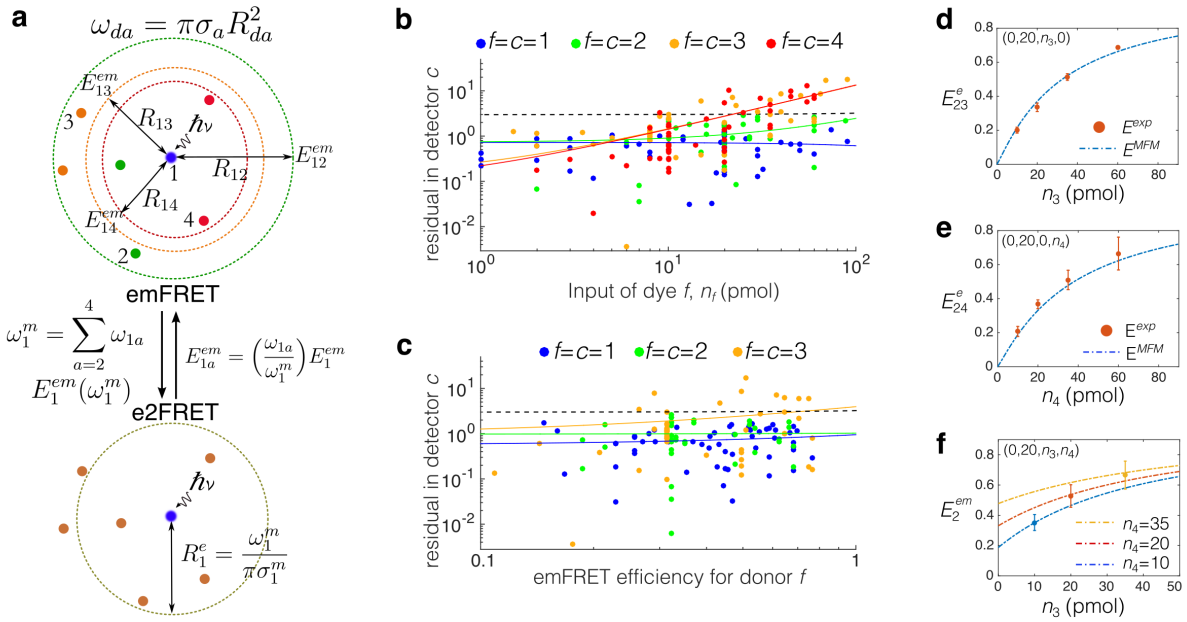


Figure 3.3: **Schematic representation of the ensemble mFRET (emFRET) model and experimental validation of the multicolour fluorescence model (MFM).** (**a**) Schematic representation of the emFRET model and its conversion to e2FRET configuration by means of the effective acceptor transform yielding an effective Förster radius (R_d^e) and a multicolour density σ^m. For every excited donor (illustrated with dye **1** here), the newly introduced multicolour Förster acceptor number (ω_1^m) is computed, and in turn used to calculate the emFRET efficiency (E_1^{em}) and, subsequently, the inter-dye energy distribution E_{1a}^{em} (**Supplementary Note 2.3**). (**b-c**) Normalized residual error, calculated as $|I_c^{exp} - I_c^{MFM}|/s_c$ where I_c^{exp}, I_c^{MFM} , and s_c are the experimental intensity, MFM-predicted intensity (making use of the emFRET model), and standard deviation of the BMP intensities in channel c, respectively, plotted versus (**b**) the input amount of LO_f (n_f) and (**c**) the predicted emFRET efficiency for dye f (E_f^{em}), for four-colour barcodes. Linear fits to data are shown as solid lines while the horizontal dotted line is the residual error of $3 \times s_c$. (**d-f**) Comparison of measured (E^{exp}, dots) and MFM-predicted FRET (E^{MFM}, dashed lines) efficiencies for different dye concentrations for two-colour FRET $(0, 20, n_3, 0)$ and $(0, 20, 0, n_4)$ BMPs (**d** and **e**, respectively), and three-colour FRET $(0, 20, n_3, n_4)$ BMPs (**f**), showing good agreement. Error bars represent the standard deviation between triplicate bead ensembles.

described using an ensemble average of single-donor environments in 3D [13] and 2D [14, 27]. The resulting e2FRET efficiency (E^e_{da}) can be calculated using a power series expansion [14] or a practical closed-form approximation [27] (**Supplementary Note 2**). Importantly, E^e_{da} is only dependent on the average number of acceptor molecules in an R_{da} radius around the donor, ω_{da}, which is a dimensionless number that naturally emerges from the model. Hence, $\omega_{da} = \pi\sigma_a R^2_{da}$ [14] and is referred from hereon as the *Förster acceptor number* given its relation to R_{da} (**Supplementary Note 2**, **Supplementary Figure 3.14**).

Here, we extend the model and derive the ensemble multicolour FRET (emFRET) efficiency between N differently-coloured dyes that are stochastically distributed on a planar surface (2D) (see **Supplementary Note 2** for a detailed derivation). Notably, we find that the emFRET efficiency of a given donor to multicolour acceptors (E^{em}_d) may be found through simple addition of the pairwise Förster acceptor numbers—yielding $\omega^m_d = \sum_a \omega_{da}$, where ω^m_d is termed the *multicolour Förster acceptor number*. This finding is directly equivalent to an e2FRET scenario with a single effective acceptor species and an effective Förster radius (R^e_d), as depicted in **Figure 3.3a**. Unlike the classical definition of Förster radius, R^e_d is dependent on the surface density of acceptors as well as the spectral overlap of the dyes involved. E^{em}_d can then be readily calculated using the closed-form e2FRET expression after substitution with the donor-specific ω^m_d,

$$E^{em}_d = \frac{\left(\omega^m_d/\gamma\right)^\lambda}{1 + \left(\omega^m_d/\gamma\right)^\lambda} \tag{3.3}$$

where γ and λ are the exclusion and fitting constants, respectively [27] (**Supplementary Note 2**). Using equation (3.3), we find that the energy is transferred to the different acceptor species proportionally to $\left(\omega_{da}/\omega^m_d\right)^\lambda$, allowing quick determination of energy transfer distributions and cascades from the surface densities, and imparting mechanistic insight to complex multicolour interactions (**Supplementary Note 2**). Note that the ensemble average calculation is predicated on approximations such as non-saturated exciton density [25, 26] and independent surface densities for the differently-coloured dyes, which are reasonably met by our labelling method (**Supplementary Figure 3.13**, **Supplementary Note 2**).

3.3.5 Model calibration and validation.

The emFRET model constitutes the kernel of the MFM, which can be calculated after determining the values of the photophysical parameters (*i.e.* cytometer and classifier dyes) in a one-time calibration experiment that can be rapidly performed (Methods, **Supplementary Note 1**, **Supplementary Note 3**). The physical parameters extracted via calibration were found to be reproducible

(**Supplementary Table 3.2**), and will be used from here on to model the fluorescent response of any combination of dyes **1-4**.

The accuracy of the MFM was evaluated by comparing the predicted and measured fluorescence for a number of arbitrary four-colour BMPs and calculating, for every channel c, the residual error normalized by the standard deviation (s_c) of the bead intensities. The residual error was typically $< 3 \times s_c$ for most conditions, which is adequate for barcoding applications. The general trend of the error in channel c was plotted against n_f for $f = c$ (**Figure 3.3b**) showing that the residual errors of I_1 and I_2 are independent of n_1 and n_2, whereas the residual error in I_3 and I_4 increases beyond $3 \times s_c$ at high concentrations of dyes **3** and **4** respectively, which points to a breakdown in the linear basal fluorescence approximation and which may be caused by self-quenching in these dyes. Importantly, as plotted in **Figure 3.3c**, the normalized residual error was independent of the calculated emFRET efficiency for the three donor dyes, indicating that the accuracy of the emFRET model is steady from low to extreme FRET levels. Next, we interrogated the accuracy of the MFM-computed FRET efficiency (E^{MFM}) in barcodes that permit FRET efficiency calculation using the donor quenching method (see Methods). Experimental FRET efficiencies (E^{exp}), averaged over bead ensembles, were found to be in good agreement with the emFRET model, as shown in **Figure 3.3(d-f)**.

3.3.6 Barcode design and synthesis.

The barcode positions in intensity-space were designed *in silico* using the calibrated MFM through optimization of dye proportions, which in effect allows anticipating and compensating for emFRET to avoid overlapping clusters. This procedure enabled the generation of 580 barcodes with well-resolved regions (**Figure 3.4a**). The list of barcodes as well as the predicted fluorescence responses are presented in order of increasing (n_3, n_4) in **Supplementary Table 3.3** (**Appendix A**). The inter-dye emFRET efficiency predicted within each barcode highlights the strong energy transfer, which reaches 76% at its maximum (**Figure 3.4b**), and wherein only 67 barcodes incur less than 10% of total FRET efficiency for any given dye (**Supplementary Figure 3.19**). Clusters of synthesized BMP subsets were superimposed over the predicted barcode intensities, showing a very good overall agreement (**Figure 3.4c-e**). The clusters in I_3-I_4 deviated from their predicted values at high concentrations, which is consistent with our observation above (see **Figure 3.3b**). The clusters in I_1-I_2 were in very good agreement with the predictions as seen in **Figure 3.4d** and **Figure 3.4e** for barcodes $(n_1, n_2, 0, 0)$ and $(n_1, n_2, 8, 10)$ respectively. The impact of emFRET within the MFM is illustrated in **Supplementary Figure 3.16**, which highlights the necessity to account for energy transfer for accurate barcode design.

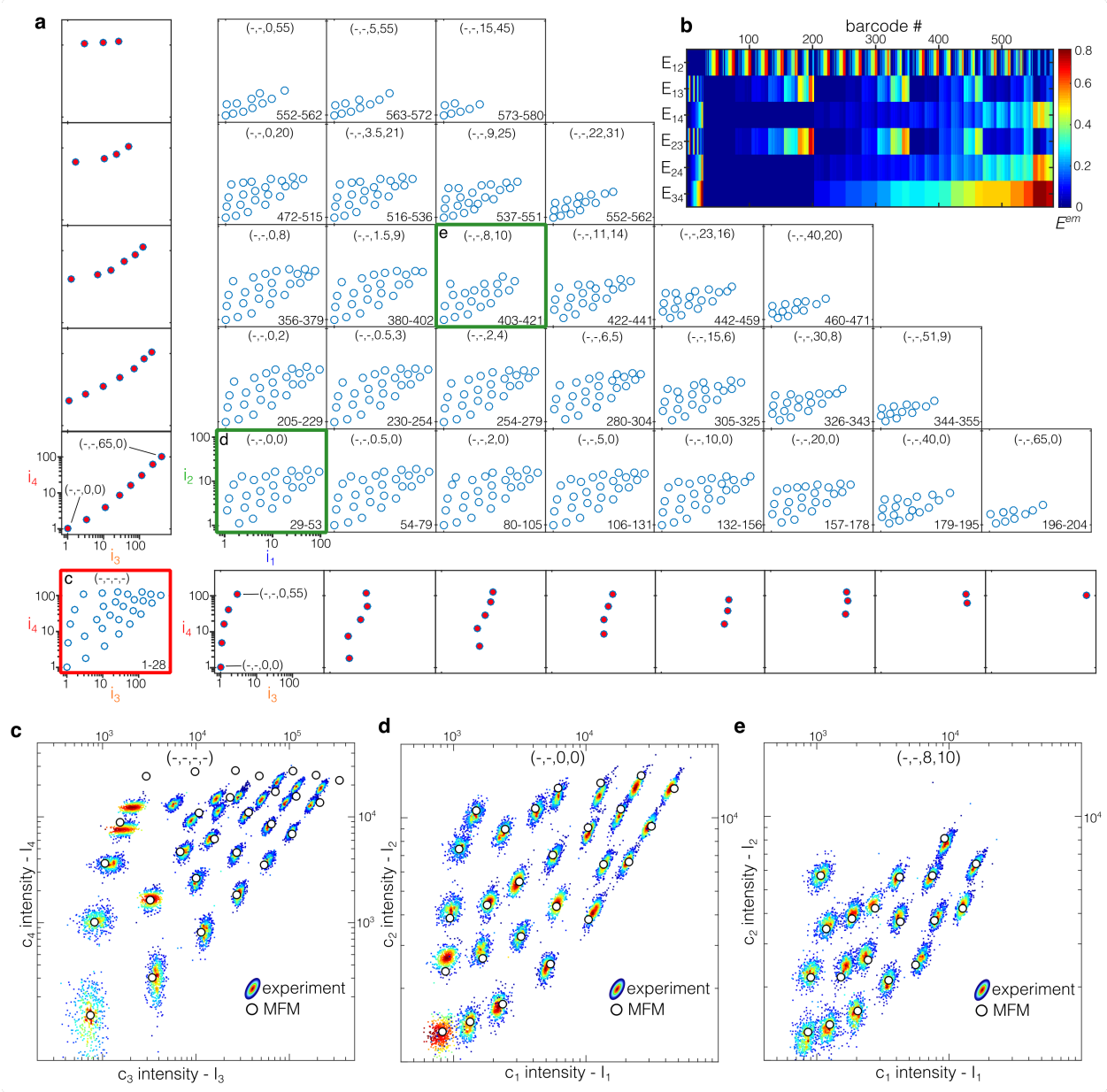


Figure 3.4: ***In silico* design and experimental verification of four-colour barcodes with extreme emFRET**. (**a**) 2D plots of the MFM-designed positions of 580 barcodes in 4D normalized intensity-space (i_1, i_2, i_3, i_4) where $i_c = I_c/I_c^0$. The barcodes are represented and plotted as circular regions with a radius equal to a 35% variation to account for experimental variation of the BMP clusters (See Methods). Graph at the lower left corner shows i_3-i_4 intensity space. Graphs in the central block show i_1-i_2 space for a set value of n_3 and n_4. In general, n_3 increases from left to right, and n_4 from bottom to top. The marginal plots in the bottom and to the left are the i_3-i_4 projections of the subsets plotted in the associated column and row, respectively. The range of barcode numbers for each subset is listed in the bottom right of each sub-plot and the full list of the barcodes is presented in **Supplementary Table S3** (**Appendix A**). (**b**) Breakdown of inter-dye FRET efficiencies between all dye combinations for each barcode showcasing the extreme levels of FRET in some cases. (Continued on next page)

Figure 3.4: (**c-e**) Experimental intensity scatter plots of BMPs are overlaid with their MFM-predicted values from the three sub-plots highlighted in (**a**) to showcase the excellent agreement with the MFM, save for some barcodes with high n_3 and n_4 in (**c**), which we ascribe to the breakdown of the linear response.

3.3.7 Automated decoding.

To benefit from the throughput of flow cytometry and the high barcoding capacity, automated decoding is imperative, but has not been possible to date for barcodes subject to mFRET. Automated decoding entails (i) *clustering* the BMP dataset, (ii) *classifying* the BMPs into the different clusters, and (iii) *assigning* these clusters—and thus the BMPs within—to their cognate barcodes. Whereas (i) and (ii) are straightforward with orthogonal classifier dyes [28, 29], these tasks develop into a multivariate problem in the case of non-orthogonal classifiers, and rapidly become challenging and computationally expensive [30]. Furthermore, without *a priori* knowledge of the relative barcode responses, (iii) is only achievable using manual initialization, even for relatively low FRET levels [31]. We sought to leverage the MFM to fully automate the decoding of BMPs.

Clustering and classification were automated by fitting the fluorescent response to a 2D Gaussian-mixture model (GMM) using expectation maximization (see methods). A digitally-concatenated representative dataset of 45 barcodes (**Supplementary Table S3, Appendix A**) was used to validate the decoding. Without the MFM, the EM-algorithm leads to inadequate clustering, significant misclassification (**Figure 3.5a**) and, compounded by error propagation and cluster mis-assignment, results in unsuccessful decoding (**Figure 3.5c**). In contrast, MFM values can be used to initialize the EM algorithm, yielding deterministic clustering and assignment, and resulting in $> 99\%$ confidence in BMP decoding (**Figure 3.5b-d** and **Supplementary Figure 3.18**).

3.3.8 Binding and specificity profiling.

In multiplex sandwich immunoassays, reagent sets need to be validated individually and combinatorially to ensure selective binding [32, 33] and avoid reagent-driven non-specific binding (NSB) [34, 35]. However, the scale of combinatorial screening experiments required grows rapidly, becoming a financial and operational burden [36, 37]. To illustrate the applicability of our method, specific and NSB (including cross-reactivity), were combinatorially measured in a multiplexed sandwich immunoassay.

Binding and NSB in a 35-plex sandwich assay including 3745 combinatorial interactions (**Figure 3.6, Supplementary Table S5, Appendix A**) were rapidly tested using our BMPs. BMPs were synthesized and functionalized with 35 different capture antibodies (cAbs), as shown in **Figure 3.6a**. To uncover NSB events, pools of BMPs against the 35 targets (cAb-BMPs) were incubated in different

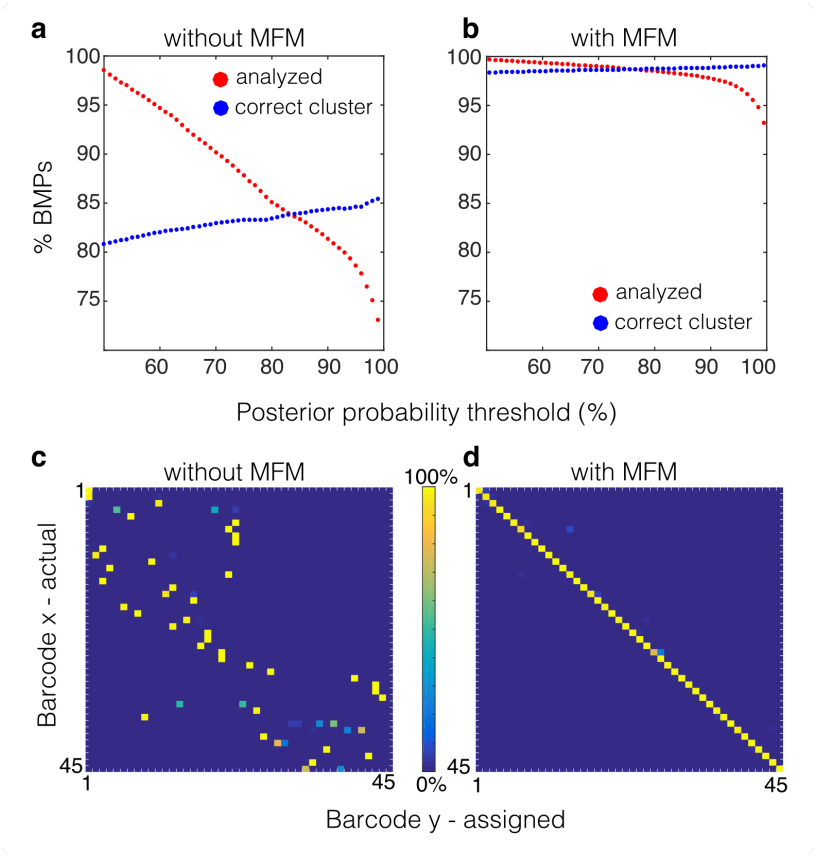


Figure 3.5: **Multicolour fluorescence model enables automated decoding. (a, b)** Fraction (in %) of BMPs classified to the correct cluster with respect to the posterior probability thresholds using the GMM **(a)** without and **(b)** with the initial conditions based on MFM-predicted intensities. **(c, d)** Heatmaps quantifying the fraction (in %) of BMPs from barcode x (actual) that were assigned to barcode y using a EM-based decoding algorithm, performed **(c)** without and **(d)** with MFM-predicted intensities as initial conditions. The use of the MFM greatly increased the accuracy in cluster assignment, and yielded a close to perfect barcode assignment.

wells with different assay reagents, including secondary antibodies (sAbs), detection antibodies (dAbs), and antigens (Ags) (**Figure 3.6b-e**). Cross-reactivity between non-cognate cAb-dAb pairs was detected by incubating the cAb-BMPs with a dAb cocktail followed by sAb, or sAb only, and comparing signals (**Figure 3.6f-g**). Indeed, several instances of cAb-dAb cross-reactivity were identified by incubating cAb-BMPs with individual dAbs (**Figure 3.6h**, **Supplementary Figure 3.20**). To study Ag binding, full assays were conducted by incubating the cAb-BMPs with individual Ag standards at 1 ng.mL^{-1} or 100 ng.mL^{-1}, followed by dAbs and sAbs (**Figure 3.6e**). Out of 35 sandwich assays, 30 met binding quality standards with strong, specific signals (**Figure 3.6i-j**). However, 8 proteins led to pervasive, concentration-dependent NSB at 100 ng.mL^{-1} and, in addition, five capture-BMPs cross-reacted significantly with three or more Ags (**Figure 3.6j**, **Supplementary Figure 3.20**). These results illustrate the utility of BMPs in screening and identifying

for NSB events.

Conclusions

We have presented an emFRET model and a microparticle labelling method that together enable *in silico* design, accurate synthesis, and completely automated decoding of fluorescent barcodes using highly non-orthogonal classifier dyes. We synthesized BMPs using dyes with large spectral overlap, realizing 580 barcodes with up to 76% energy loss by mFRET, and showcasing the predictive accuracy of the MFM, demonstrating it is fit-for-purpose with multiplexed assays, and indicating a potential for large scale multiplexing. Importantly, we have presented an emFRET model that imparts mechanistic insight into multiplexed FRET interactions, enables designing mFRET experiments with increasing complexity [19], and allows rational design of ensemble mFRET interactions. For example, it is expected that a similar model could be derived for homogeneous dye ensembles in 3D, which could have applications in monitoring multiplexed intermolecular interactions at the sub-cellular level. In addition, the ability to rationally design ensemble mFRET interactions may be particularly useful to optimize energy relays in dye-sensitized solar cells [38].

Notably, our methods reduce the technical and economical entry barriers for deploying a high-capacity multiplexing platform through the use of: (i) common organic dyes that are compatible with most flow cytometers, (ii) an easy-to-implement BMP synthesis method using DNA oligos that yields quick, accurate, and precise labelling, and (iii) an MFM model for tailoring barcode responses to different optical set-ups without necessitating modifications to the automated decoding software.

The barcoding capacity may be readily expanded through additional classifier dyes or increased stringency during the barcode design stage. Yet the multiplexing of sandwich immunoassays continues to be limited by the cost and availability of reagents, which indeed also limited our demonstration to 35-plex. More importantly, the scale of sandwich multiplexed immunoassays is limited by the propensity of antibodies to cross-react when mixed, which was confirmed in our experiments, and which remains unresolved for bead-based suspension assays. Multiplexed assays using single antibodies, or nucleic acids, may however immediately benefit from high capacity BMPs enabled by the emFRET model. The emFRET model is general, and may for example guide the optimization of exciton transfer in dye-sensitized solar cells [38].

Finally, we hope that the results shown here will spur the development multiplexed suspension assays using BMPs, as well as inspire new theoretical development and applications of mFRET.

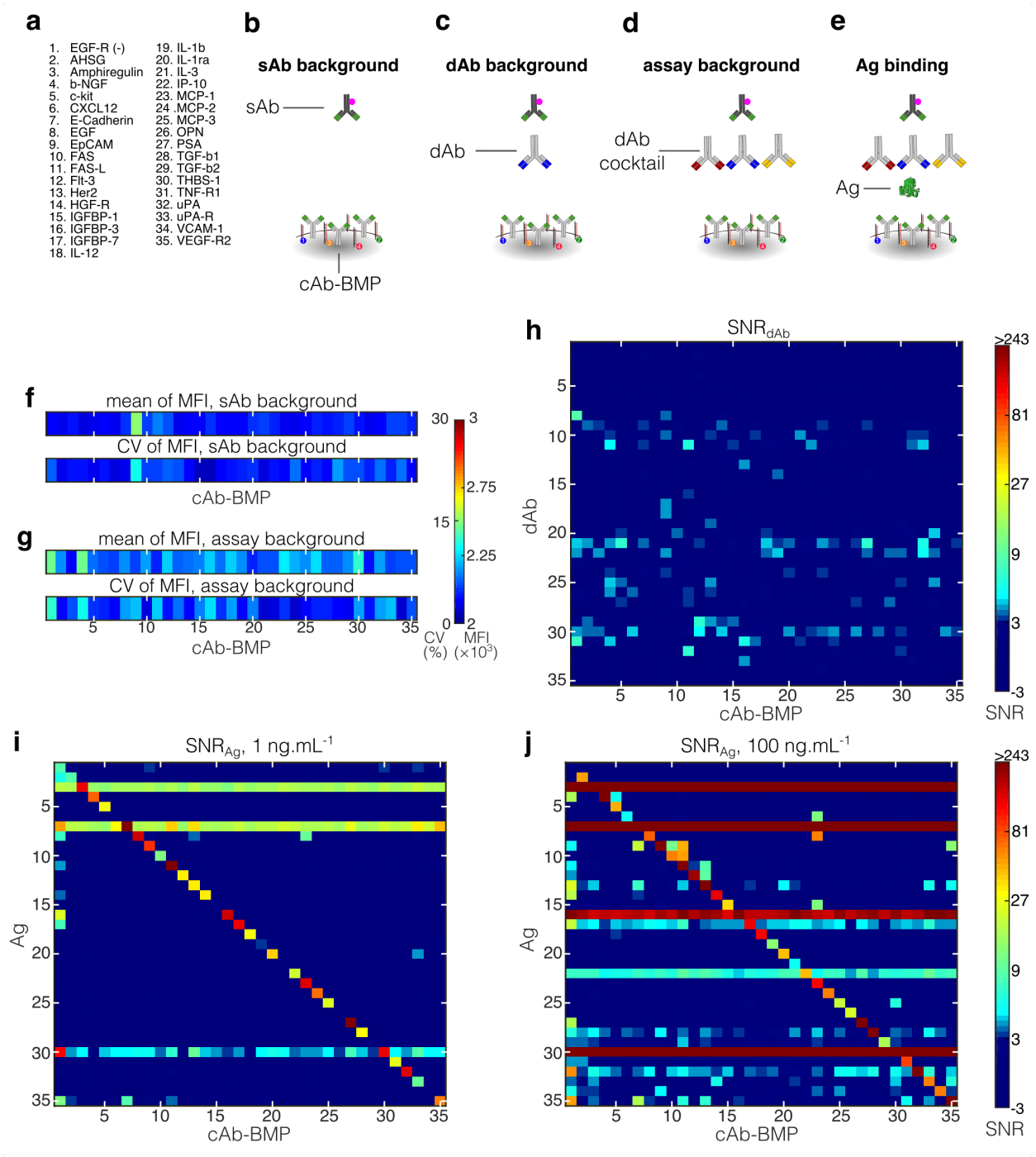


Figure 3.6: **Screening of binding specificities for a 35-plex sandwich immunoassay. (a)** List of antibody pairs and recombinant Ags characterized for specific binding and NSB (including cross-reactivity, **Supplementary Table S5, Appendix A**). **(b-e)** Schematic representation of the controls and assays performed. Negative controls include incubation of cAb-BMP mixture with sAb (panel b), dAb (panel c), and dAb cocktail (panel d), respectively, allowing detection of background signal and NSB. Full assays were performed by incubating the cAb-BMP mixture with individual Ags, followed by dAb cocktail and sAb (panel e). **(f,g)** mean MFI and CV of sAb background and assay background, respectively. The number of well replicates for sAb and assay background was 3 and 6, respectively. (Continued on next page.)

Figure 3.6: **(h)** SNR_{dAb} quantifying cross-reactivity between individual dAbs (rows) and specific cAb-BMPs (columns). SNR_{dAb} is the signal-to-noise ratio calculated relative to the mean sAb background ($n = 3$) and normalized to the global standard deviation ($n = 105$) (Methods, **Supplementary Figure 3.20**). **(i, j)** Specific (diagonal) and non-specific (off-diagonal) assay SNR (SNR_{Ag}) of individual Ags (rows) at **(i)** 1 ng.mL^{-1} or **(j)** 100 ng.mL^{-1} for every cAb-BMP (columns). SNR_{Ag} was calculated relative to the mean assay background ($n = 6$) and normalized to the global standard deviation ($n = 210$) (Methods, **Supplementary Figure 3.20**). SNRs > 3 are visible and highlighted.

3.3.9 Methods

Choice of classifier dyes. Our rationale for the choice of classifier dyes was four-fold. First, using dyes that emit at closely spaced wavelengths capitalizes on the limited spectral bandwidth available, and whereas it leads to substantial spectral overlap, it will help expanding the number of dyes used and, with it, the barcoding capacity. Second, the substantial mFRET that is expected as a consequence will serve as a challenging example to validate the proposed emFRET model. Third, using small organic dyes that may be acquired off-the-shelf with different chemical functionalization allows easier implementation and greater flexibility for downstream applications. Fourth, using an assay reporter dye in the blue-shifted spectrum avoids interference with barcodes through analyte-dependent FRET (which could be significant for high analyte concentration) as it cannot act as an acceptor to any of the classifier dyes.

Design and preparation of LOs. LO_f were formed through hybridization of complementary 21-nt oligos: a 5' biotinylated oligo (BO) and a fluorescent oligo (FO) 3'-labelled with dye f (**Supplementary Figure 3.20**). FO_0 was unlabelled, whereas FO_1-FO_4 where labelled with dyes **1**-**4** where 1≡FAM, 2≡Cy3, 3≡Cy5, and 4≡Cy5.5 respectively. The BO sequence is: 5'Biotin/TTTTTTTTTGTGGCGGCGGTG/3'. The fluorescent oligo sequence is: 5'/CACCGCCGCCACAAAAAAAAA/-f. AOs and FOs were annealed at 10 μM in phosphate buffer saline (PBS) + 300 mM NaCl. All oligonucleotides were acquired already modified from Integrated DNA Technologies (IDT, Coralville, IA, USA). The sequences were tested using *in silico* DNA-analysis tools (*e.g.* Mfold) to ensure minimal secondary structure formation.

BMP synthesis. 25 μL of any given barcode (n_1, n_2, n_3, n_4), were prepared by pre-mixing biotinylated reagents containing 6.7 pmol of IgG (1 μg), the corresponding amounts of LO_f, such that $n_0 + n_1 + n_2 + n_3 + n_4 = 90$ pmol and PBS + 300 mM NaCl. Co-immobilization of the biotinylated reagents was initiated through addition of 3.25×10^6 streptavidin-coated superparamagnetic

microparticles (M270-Streptavidin from Life Technologies, Carlsbad, CA, USA) in 25 μL of PBS + 300 mM NaCl, and the reaction mixture incubated with mixing for one hour, followed by 3 cycles of magnetic aggregation and washing in PBS + 0.1 v/v % Tween-20 . Following synthesis, batches of BMPs were stored separately in the dark at 4 °C and were mixed prior to use in a multiplexed assay. Proof of concept experiments were performed using biotinylated mouse anti-goat IgG (Life Technologies), which were detected using secondary IgGs (goat anti-mouse IgG, BV421, BioLegend) by incubation at 4 μg/mL for 45 minutes.

Flow Cytometry. BMPs were read out using the FACS CANTO II cytometer by BD with blue (488 nm) and red (633 nm) lasers. In the blue-laser flow cell, 530/30 and 585/42 band-pass (BP) filters were used for channels 1 and 2, respectively. In the red-laser flow cell, 660/20 and 780/60 were used for channels 3 and 4, respectively. For reporter dye detection during assays, the violet 405nm laser was used with a 450/40 BP filter. For decoding validation, BMPs were measured separately and concatenated digitally before any subsequent data analysis.

Single-molecule Förster radii. Emission and absorption spectra were used to calculate the overlap integral, $J_{da}(\lambda)$, and subsequently the single-molecule Förster radius R_{da}. The latter was calculated for each donor-acceptor pair using the following expression [39, 40]:

$$R_{da} = 9.78 \times 10^3 \left(\kappa^2 \check{n}^{-4} Q_d J_{da}(\lambda)\right)^{1/6} \tag{3.4}$$

where κ^2 is the dipole-dipole orientation factor, taken to be as 2/3 as per the dynamic isotropic approximation (**Supplementary Note 2**), $\check{n}$ is the medium's refractive index, and Q_d is the fluorescence quantum yield of the donors. Absorption and emission spectra of LOs were measured using a SpectraMax i3x Multi-mode microtiter plate.

emFRET model The FRET efficiency, E_d^m, from a donor to multicolour acceptors stochastically distributed on a 2D surface was calculated using the probabilistic decay function $\rho_d(t)$, which denotes the probability that a donor excited at time $t = 0$ is still excited at t [13, 14],

$$E_d^m = 1 - \frac{1}{\tau_d^0} \int_0^\infty \rho_d(t) dt. \tag{3.5}$$

where τ_d^0 is the unperturbed donor lifetime. For an excited donor molecule, the decay function is governed, as per Förster's theory[13], by the following differential equation

$$-\frac{d}{dt}\rho_d(t) = \left(1 + \sum_{a=d+1}^{N}\sum_{z=1}^{Z_a}\left(\frac{R_{da}}{r_{z,a}}\right)^6\right)\frac{\rho_d(t)}{\tau_d^0} \tag{3.6}$$

where Z_a is the number of acceptors from dye species a in the vicinity of an excited donor d and r_{za} is the distance between donor d and the z-th acceptor of species a. The solution of equation (3.6) is then ensemble-averaged for all donors (*i.e.* for all potential configurations of acceptors) to arrive at a solution for the ensemble multicolour FRET efficiency, E_d^{em}. We show that (see **Supplementary Note 2** for a detailed derivation) the ensemble-averaged decay function of the donor in an emFRET scenario is equivalent to that of the e2FRET using the transformation on the Förster acceptor number $\omega_d \rightarrow \omega_d^m = \sum_{a=d+1}^{N}\omega_{da}$, where ω_d^m may be directly plugged in equation (3.3). Within this transformation, the e2FRET acceptor corresponds to an *effective* acceptor with an *effective* Förster radius (**Supplementary Note 2**)

$$R_d^e = \left(\sum_{a=d+1}^{N}\frac{R_{da}^2\sigma_a}{\sigma_d^m}\right)^{1/2}. \tag{3.7}$$

where σ_d^m is the density sum for all acceptors around donor d.

Model parametrization. In the combined dye and cytometer set-up addressed herein, many parameters take a zero value, greatly simplifying the algebraic equations constituting the MFM (**Supplementary Note 1**). Because of the spatially and temporally separate excitation in a flow cytometer and the spectral properties of the dyes in question, a number of variables in the bleed-through and FRET proportionality matrices (**B** and **A** respectively) are irrelevant and set to zero. For instance, $\beta_{31} = 0$ because dye **1** is not excited during the registration of intensity in $\mathbf{c_3}$. In our case, there are 13 variables representing key physical parameters that need to be extracted (**Supplementary Note 1**), which we completed by measuring 18 judiciously selected barcodes, as discussed in **Supplementary Note 3**. Briefly, single-colour BMPs are measured to fit the values of μ_f and β_{cf} to the response using relative weighting, whereas two-colour BMPs allow calculating the remaining variables (**Supplementary Figure S9**). The values particular to our set-up, which were calibrated on three separate days to ensure reproducibility, are presented in **Supplementary Table S2**. A detailed presentation of the model parametrization and calibration steps are presented in **Supplementary Notes 1 and 3**, respectively.

Analysis of cytometry data for barcoding. All data analysis was performed in MATLAB. To quantitatively compare fluorescence intensities across cytometry experiments, we performed a linear normalization on signal-to-background ratios to account for fluctuations in laser power intensities. Single-beads were distinguished from bead aggregates and other particulates by using forward and side-scatter intensities and gating was automated using a MATLAB script for all data shown in this work (**Supplementary Figure 3.17**). Experimental emFRET was measured using the donor quenching method [25] for select barcodes that allow crosstalk-free measurement of a single donor species (*e.g.* I_1 when $n_2 = 0$ or vice versa). Therefore, by measuring donor-associated channel ($c = d$) for BMPs with and without any acceptor species (I_c^{FRET} and I_c^{noFRET} respectively), the experimental emFRET efficiency can then be calculated using equation (3.10) for $c = d$ where it can be shown that $(E_d^{em})^{exp} = \left(I_c^{noFRET} - I_c^{FRET}\right) / \left(I_c^{noFRET} - I_c^{0}\right)$.

***In silico* design of barcode responses.** The predicted BMP intensities were represented as regions that delimit a 35% variation from their centre, a value that is $\sim 3.5\times$ the measured standard deviation (see **Figure 3.2b**) and thus expected to include $> 99\%$ of the BMPs for a normal distribution. Regions in the i_3-i_4 joint intensity space were designed first as they are only dependent on (n_3, n_4). 28 non-overlapping regions were generated as shown in the bottom left plot of Figure 3.4a. Next, for each of the (n_3, n_4) values, regions of the i_1-i_2 intensity space were optimized and, as expected, were strongly dependent on (n_3, n_4) due to emFRET (**Figure 3.4a**). Aside from a strong dependence on emFRET, the number of regions in i_1-i_2 was also limited by the requirement of a conserved total LO of 90 pmol (*e.g.* $n_1 + n_2 \leq 30$ pmol for barcodes $(n_1, n_2, 5, 55)$).

Automated decoding. To decode a mixture of BMPs based on 4D intensity data (I_1, I_2, I_3, I_4), sequential 2D clustering, classification, and assignment was performed by fitting a Gaussian Mixture Model to sequential 2D channel intensities using the expectation-maximization algorithm [41]. The 4D intensity data (I_1, I_2, I_3, I_4) was initially classified according to the I_3-I_4 intensity scatter plot, and each BMP classified to a cluster with an assigned (n_3, n_4) value. This was followed by independently decoding each of these clusters in the I_1-I_2 intensity space following the same protocol to complete the decoding of the (n_1, n_2, n_3, n_4) value. We use the Gaussian-mixture model (GMM) to model a 2D intensity dataset, $\mathbf{I}$, to the probability distribution function given by

$$p(\mathbf{I}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{I} \mid \boldsymbol{M}_k, \boldsymbol{\Sigma}_k) \tag{3.8}$$

where $\boldsymbol{M}_k$, and $\boldsymbol{\Sigma}_k$ are, respectively, the means and covariances of the k Gaussian given by $\mathcal{N}(\mathbf{I} \mid \boldsymbol{M}_k, \boldsymbol{\Sigma}_k)$, and π_k is the mixing coefficient which is a normalized metric that denotes how well the BMPs fit the k-th Gaussian. The total number of clusters, K, is defined by the number of unique combinations of dye proportions to be decoded in the corresponding space (*e.g.* number of unique (n_3, n_4) when classifying the (I_3, I_4) data). When using the MFM model, the expected intensities for every barcode are used as the initial value of the means in the GMM, that is, $\boldsymbol{M}_k^0 = \mathbf{I}^{MFM}$. Without the MFM, a set of arbitrary intensity values from the experimental dataset are used as the initial means. For both methods, the initial covariance matrix value was a diagonal matrix with 10% CV in each dimension $\boldsymbol{\Sigma}_k^0 = diag(0.1 \times \boldsymbol{M}_k^0)$, in accordance with our measured CV values in **Figure 3.2b**, and the initial cluster probabilities to be homogeneous ($\pi_k^0 = 1/K$).

When $\mathbf{I}$ is an experimental dataset, $p(\mathbf{I})$ is a measure of likelihood that this dataset is fit by the GMM clusters. During the expectation step of the expectation-maximization search[41], the probability that a certain BMP ψ, belongs to a cluster k, also referred to as the posterior probability, is calculated using [41]

$$\phi_{\psi k} = \frac{\pi_k \mathcal{N}(\mathbf{I}_\psi \mid \mathbf{M}_k, \boldsymbol{\Sigma}_k)}{p(\mathbf{I}_\psi)}. \tag{3.9}$$

During the maximization step, the values of the Gaussian components ($\boldsymbol{M}_k$, $\boldsymbol{\Sigma}_k$, and $\boldsymbol{\pi}_k$) are updated to maximize the log-likelihood (*i.e.* $ln(p(\mathbf{I}))$). This process is repeated for up to 5000 iterations or until the condition for convergence $ln(p(\mathbf{I}_i)) - ln(p(\mathbf{I}_{i-1})) < 1e^{-7}$ is reached. Typically, the EM algorithm performs 'soft classification', whereby the ψ-th BMP is assigned to the population for which it has the maximal $\phi_{\psi k}$. To improve the fraction of correctly classified BMPs after the algorithm converged to its solution, a posterior probability threshold was used, and varied from 50 to 99%, thereby rejecting BMPs with lower ϕ. Finally, to perform 2D cluster-to-barcode assignment without use of the MFM, and thus without a priori knowledge of the mean intensities of the barcode clusters, the means of the Gaussian clusters as well as the input dye proportions were sorted and assigned according to their root mean square (*i.e.* $[\boldsymbol{M}_k(1)^2 + \boldsymbol{M}_k(2)^2]^{1/2}$ and $[\boldsymbol{n}_k(1)^2 + \boldsymbol{n}_k(2)^2]^{1/2}$, respectively).

Multiplexed binding and NSB analysis. cAbs, Ags, and dAbs were purchased from R&D Systems (Minneapolis, MN, USA), and stored at -20°C for up to 36 months, see Supplementary Table S5 (Appendix A. All dAbs were mouse IgGs, with the exception of EGF-R which used a goat dAb and was used as negative control. Conversely, none of the capture antibodies were from mouse. sAb (goat anti-mouse IgG, BV421) was purchased from BioLegend (San Diego, CA, USA) and stored at 4°C for up to 6 months. Biotinylated cAbs were coupled to their respective BMP during synthesis as described above. cAb BMP mixtures were combined to a final concentration

of 50 BMPs per μL per barcode. Incubations were performed in a conical bottom 96-well plate at room temperature with horizontal shaking at 950 rpm. Prior to incubation with assay reagents, BMPs were first blocked for one hour with 1% bovine serum albumin in 0.05% Tween-20 in PBS. Incubation with Ags was conducted for 120 min at the specified concentrations. BMPs were incubated with dAbs for 60 min at 5 μg.mL^{-1} and 2 μg.mL^{-1} for individual and cocktail incubation, respectively. BMPs were incubated with sAbs for 45 mins at 4 μg.mL^{-1}. SNR$_{dAb}$ is calculated by subtracting the cAb-specific mean sAb background ($n = 3$) from the MFI signals and normalizing to the global standard-deviation (*i.e.* across all barcodes, $n = 105$) of the sAb background. SNR$_{Ag}$ is calculated by subtracting the cAb-specific mean assay background ($n = 6$) from the MFI signals and normalizing to the global standard-deviation (*i.e.* across all barcodes, $n = 210$) of the assay background.

3.4 Supplementary Information

3.4.1 Supplementary Note 1: Derivation of MFM

Here we present the derivation of a generalized multicolour fluorescence model (MFM) for microparticles labelled with a total of N different fluorophores. The different dye species (or colours) are denoted by f according to increasing wavelengths (*i.e.* fluorophore $f = 1$ has the smallest absorption maxima wavelength in nm). The dyes are excited with L lasers and measured in C channels, respectively. The interactions modelled are direct excitation (laser-induced emission), sensitized emission (FRET-induced emission), and FRET cascades. Please refer to Supplementary Table S1 for variables and definitions. Whereas equations (1-3) in the main text denote the specific scenario where every dye is directly excited through only one laser, the equations below are general and could be used to model cases where dyes may be excited at different time points using different lasers.

Generalized detector intensity expressions

A general expression of the detector intensities in the channels can be found by considering the signal of every channel to be the sum of the ensemble (*i.e.* steady-state total) fluorescence from all potential fluorophores excited with a certain laser. Spectral overlap between the dyes' emission spectra and the optical filter in different channels leads to a platform-dependent bleed-through matrix **B**. The detector intensity in channel c, in response to laser l is then

$$I_{cl} = I_{cl}^{0} + \sum_{f=1}^{N} \beta_{cf} F_{fl}^{e}, \quad c = 1, \ldots, C, f = 1, \ldots, N, l = 1, \ldots, L \tag{3.10}$$

Due to the possibility of close proximity between fluorophores on the BMPs, the ensemble fluores-

cence of a given dye takes into account quenching and sensitized-fluorescence through long-range non-radiative transfer (*i.e.* FRET). The number of excited dyes at steady-state is considered to be a small fraction of the total number, and hence the ensemble fluorescence is taken as a simple sum of laser-induced and FRET-sensitized emission, assuming no competition between those two processes (Fig. 3.13 and Supplementary Note 2). The ensemble fluorescence intensity of dye f, where $1 < f < N$,

$$F^e_{fl} = F^s_{fl}(1 - \sum_{i=f+1}^{N} E^{em}_{fi}) \tag{3.11}$$

$$F^s_{fl} = F^0_{fl} + \sum_{j=1}^{j<f} \alpha_{jf} E^{em}_{jf} F^s_{jl} \tag{3.12}$$

where, in order to model FRET cascades, F^s is the sensitized fluorescence, which accounts all potential sources of excitation (direct or through a FRET), and assumes no de-excitation by FRET (i.e. no acceptors). It can be understood as the potential maximum fluorescence intensity for every ensemble of dyes.

From (3.11), and (3.12) we get,

$$F^e_{fl} = \left(F^0_{fl} + \sum_{j=1}^{j<f} \alpha_{jf} E^{em}_{jf} F^s_{jl} \right) \left(1 - \sum_{i=f+1}^{N} E^{em}_{fi} \right) \tag{3.13}$$

which, when merged with (3.10) gives the general equation

$$I_{cl} = I^0_{cl} + \sum_{f=1}^{N} \beta_{cf} \left(F^0_{fl} + \sum_{j=1}^{j<f} \alpha_{jf} E^{em}_{jf} F^s_{jl} \right) \left(1 - \sum_{i=f+1}^{N} E^{em}_{fi} \right) \tag{3.14}$$

Case of independent classifier dye densities

Using our proposed co-immobilization strategy, the concentration of LOs in solution is constant and balanced out by the non-labelled oligo (n_0), then the total surface density of oligos must also be constant and barcode-independent. Furthermore, the frequency of LOs labeled with one dye species (LO_f) in solution is independent of dye concentrations. This leads to the surface density of LO labeled with dye f only dependent on, and proportional to, the solution amount,

$$\sigma_f = t_f n_f \tag{3.15}$$

where t_f can be considered as the dye-dependent labelling constant that relates to the affinity constant and footprint for each dye f, or more specifically, for each LO_f. Furthermore, as the LOs

are homogeneous in affinity and footprint to the MPs, then the labelling constant is exactly the same for all the dyes, i.e.

$$\sigma_f = tn_f \tag{3.16}$$

For non-saturating conditions (*i.e.* no self-quenching), F^0_{fl} will be linearly proportional to number of dyes f on the microsphere surface, and thus the relation between the input oligo-dye construct (n_f) and the output fluorescence F^0_{fl} can be assumed to be linear and thus is described as:

$$F^0_{fl} = \mu_{fl} n_f \tag{3.17}$$

where μ_{fl} is a dye- and laser- specific constant that is a measure of the its quantum yield and spectral properties of excitation wavelengths and emission filter set-up.

(3.13) becomes

$$F^e_{fl} = \left(\mu_{fl} n_f + \sum_{j=1}^{j<f} \alpha_{jf} F^s_j E^{em}_{jf}\right)\left(1 - \sum_{k=f+1}^{k=N} E^{em}_{fk}\right) \tag{3.18}$$

MFM initialization for optical set-up: BD FACS CANTO II

Here we parametrize the model for the imaging system (cytometry or microscopy) and fluorophores at hand. Here, a set of four fluorophores (i.e. $N = 4$) which are (FAM, Cy3, Cy5, Cy5.5) are labeled on MPs and read-out using 4 channels ($C = 4$) on the FACS CANTO II cytometer using two lasers ($L = 2$) namely, the blue (488 nm) and the red (633 nm) laser, as shown in Figs 1(b-c) of the main text.

Thus, as a response to excitation with laser 1, a value registered for the detectors in channel 1 and 2. Conversely, as a response to excitation with laser 2, a value registered for the detectors in channel 3 and 4 (see Fig. 1b of main text). Hence, the detector values to be computed are $I_{11}, I_{21}, I_{32}, I_{42}$.

To start, from the dyes' absorption spectra one can already assume that $\mu_{12} = \mu_{22} = \mu_{31} = \mu_{41} = 0$. For example, dye 3 will not be (directly) excited by laser 1 (488nm). Similarly, dye 1 will not be excited by laser 2 (633nm).

$$\mu_{fl} = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \\ \mu_{31} & \mu_{32} \\ \mu_{41} & \mu_{42} \end{bmatrix} = \begin{bmatrix} \mu_{11} & 0 \\ \mu_{21} & 0 \\ 0 & \mu_{32} \\ 0 & \mu_{42} \end{bmatrix} \tag{3.19}$$

Next, one can safely make the following assumptions with regards to the bleed-through constants such as $\beta_{13} = \beta_{14} = \beta_{23} = \beta_{24} = \beta_{31} = \beta_{32} = \beta_{41} = \beta_{42} = 0$ since the emission of said dyes

and detector channels are not overlapping and/or the dyes are not excited at the time of intensity registration of said channel.

The resulting **B** matrix looks like

$$\mathbf{B} = B_{cf} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \\ \beta_{31} & \beta_{32} & \beta_{33} & \beta_{34} \\ \beta_{41} & \beta_{42} & \beta_{43} & \beta_{44} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & 0 & 0 \\ \beta_{21} & \beta_{22} & 0 & 0 \\ 0 & 0 & \beta_{33} & \beta_{34} \\ 0 & 0 & \beta_{43} & \beta_{44} \end{bmatrix} \tag{3.20}$$

and as a result,

$$I_{11} = I_{11}^0 + \beta_{11} F_{11}^e + \beta_{12} F_{21}^e \tag{3.21a}$$

$$I_{21} = I_{21}^0 + \beta_{21} F_{11}^e + \beta_{22} F_{21}^e \tag{3.21b}$$

$$I_{32} = I_{32}^0 + \beta_{33} F_{32}^e + \beta_{34} F_{42}^e \tag{3.21c}$$

$$I_{42} = I_{42}^0 + \beta_{43} F_{32}^e + \beta_{44} F_{42}^e \tag{3.21d}$$

Thus

$$F_{11}^e = \mu_{11} n_1 \left(1 - E_{12}^{em} - E_{13}^{em} - E_{14}^{em}\right) \tag{3.22a}$$

$$F_{21}^e = \left(\mu_{21} n_2 + \alpha_{12} E_{12}^{em} \mu_{11} n_1\right) \left(1 - E_{23}^{em} - E_{24}^{em}\right) \tag{3.22b}$$

$$F_{32}^e = \mu_{32} n_3 \left(1 - E_{34}^{em}\right) \tag{3.22c}$$

$$F_{42}^e = \mu_{42} n_4 + \alpha_{34} E_{34}^{em} \mu_{32} n_3 \tag{3.22d}$$

3.4.2 Supplementary Note 2: Ensemble multicolour FRET model

Rate of transfer between two dyes

Förster developed the theory for non-radiative energy transfer using the very-weak coupling formalism, which considers conservation of unperturbed energy states, regards electronic excitation as temporarily completely localized, and considers energy transfer as 'statistical hopping' or step-wise energy transfer as opposed to energy propagation [40]. Förster proposed a dipole-dipole interaction model and derived the energy transfer rate expression through quantum mechanics [40],

$$k_{da}^{t} = \frac{1}{\tau_d^0}\left(\frac{R_{da}}{r}\right)^6 \tag{3.23}$$

where r is the inter-dye distance, τ_d^0 is the unperturbed lifetime of the donor and R_{da} is the Förster radius given by Equation (4) of the main text. As was noted before[42], after considering the equation for the Förster radius, the unperturbed lifetime of the donor is cancelled and one is left with the radiative decay rate for the donor, k_d^r, which is an inherent property and does not depend on the nature or number of acceptors nearby.

2FRET efficiency to acceptors with fixed, known positions

The rate of decay of an excited donor (d) in the presence a single acceptor (a) is given by[43]

$$(\tau_d)^{-1} = k_d^{nr} + k_d^r + k_{da}^t = (\tau_d^0)^{-1} + k_{da}^t \tag{3.24}$$

where k^{nr} is the rate of all non-radiative decay pathways, k^r is the rate of radiative decay (*i.e.* emission), k^t is the rate of energy transfer to the acceptor, and $(\tau_d^0)^{-1}$ is the unperturbed rate of decay of the donor (*i.e.* without an acceptor). The efficiency of transfer between the two dyes may be defined using the rate equation

$$E_{da} = \frac{k_{da}^t}{k_d^{nr} + k_d^r + k_{da}^t} \tag{3.25}$$

which leads to the widely used relation between efficiency and the inter-dye distance

$$E_{da} = \frac{(R_{da})^6}{r^6 + (R_{da})^6} \tag{3.26}$$

When multiple acceptors of the same species are available for energy transfer, the total rate of transfer between the two dyes becomes

$$k_{da}^{t} = \frac{1}{\tau_d^0}\sum_{z=1}^{Z}\left(\frac{R_{da}}{r_z}\right)^6 \tag{3.27}$$

where Z is the total number of acceptors around the donor and r_z is the distance between the donor and acceptor z. Hence,

$$E_{da} = \frac{\sum_{z=1}^{Z} \left(\frac{R_{da}}{r_z}\right)^6}{1 + \sum_{z=1}^{Z} \left(\frac{R_{da}}{r_z}\right)^6} \tag{3.28}$$

Showing that even if all the acceptors were added at the same distance from the donor, the ensemble efficiency of transfer is inherently increased [13, 44].

e2FRET efficiency to stochastically distributed acceptors in 2D

As a result of the increase in transfer rates for every added acceptor, sometimes described as the 'antenna effect' [19], the use of an average distance in equation (3.26) when the acceptors are stochastically distributed leads to an underestimation of ensemble FRET efficiency equation [19, 25, 26, 44]. The analytical solution for ensemble efficiency of transfer between 2 stochastically distributed dyes (e2FRET) in 3D was first derived by Förster using an ensemble average of single-donor environments [13]. Later, using a similar treatment, Wolber and Hudson derived the e2FRET efficiency for dyes stochastically distributed in 2D [14]. We present their derivation in this section.

A calculation radius R_c is considered around the donor, containing Z acceptors of the same species (*i.e.* constant R_{da}) with a surface density $\sigma_a = Z/(\pi(R_c)^2)$. The decay function follows a first-order differential equation given by

$$-\frac{d}{dt}\rho_d(t) = (\tau_d)^{-1}\rho_d(t) \tag{3.29}$$

where

$$(\tau_d)^{-1} = \frac{1}{\tau_d^0}\left(1 + \sum_{z=1}^{Z}\left(\frac{R_{da}}{r_z}\right)^6\right) \tag{3.30}$$

The solution to the differential equation, taking $\rho_d(0) = 1$, is

$$\rho_d(t) = \exp\left[\frac{-t}{\tau_d^0}\left(1 + \sum_{z=1}^{Z}\left(\frac{R_{da}}{r_z}\right)^6\right)\right] \tag{3.31}$$

and hence,

$$\rho_d(t) = e^{-t/\tau_d^0}\Pi_{z=1}^{Z} \exp\left[\frac{-t}{\tau_d^0}\left(\frac{R_{da}}{r_z}\right)^6\right] \tag{3.32}$$

The decay function of a finite calculation disk with radius R_c is then ensemble-averaged over all donors on the BMP to account for all other configurations based on their likelihood.

$$< \rho_d(t) >_Z = e^{-t/\tau_d^0}\Pi_{z=1}^{Z} \int_0^{R_c} \exp\left[\frac{-t}{\tau_d^0}\left(\frac{R_{da}}{r_z}\right)^6\right] W(r_z)dr_z \tag{3.33}$$

where $W(r_z)dr_z$ is the probability of finding acceptor z in the annulus within radii r_z and $r_z + dr_z$. For a completely randomized distributions of acceptors with a certain density, the products within the integral are thus all equal,

$$< \rho_d(t) >= e^{-t/\tau_d^0}[M(t)]^Z \tag{3.34}$$

where

$$M(t) = \int_0^{R_c} \exp\left[\frac{-t}{\tau_d^0}\left(\frac{R_{da}}{r_z}\right)^6\right] W(r)dr \tag{3.35}$$

where $W(r)dr = (2\pi\sigma_a r dr)/Z$ for random distributions in 2D. $M(t)$ may be then integrated by parts, expanded as power series, neglecting small terms [14] to yield

$$M(t) = 1 - \frac{\pi\epsilon\sigma_a(R_{da})^2}{N}\left(\frac{t}{\tau_d^0}\right)^{1/3} \tag{3.36}$$

For a large calculation disk,

$$\lim_{R_c\to\infty} Z = \infty \tag{3.37}$$

and using the relationship

$$\lim_{Z\to\infty}(1 + a/Z)^Z = e^a \tag{3.38}$$

yields

$$< \rho_d(t) >= \exp\left[\frac{-t}{\tau_d^0} - \frac{\epsilon}{\pi}\omega_{da}\left(\frac{t}{\tau_d^0}\right)^{1/3}\right] \tag{3.39}$$

where $\epsilon = \Gamma(2/3)$ and $\omega_{da} = \pi(R_{da})^2\sigma_a$. By expanding $e^{t^{1/3}}$ as a power series, the *ensemble* efficiency may be now calculated through

$$E_{da}^e =< E_{da} >= 1 - \frac{1}{\tau_d^0}\int_0^\infty < \rho_d(t) > dt \tag{3.40}$$

using term-by-term integration to yield

$$E_{da}^e = 1 - \sum_{j=0}^{\infty}\left(-\frac{\epsilon\omega_{da}}{\pi}\right)^j \frac{\Gamma(j/3+1)}{j!} \tag{3.41}$$

where $\Gamma(j)$ is the gamma function.

Koppel et al. undertook a similar derivation with similar assumptions but further considered that there would be minimum distance of approach between two dyes based on the host molecule's size[27]. They found a fit to the integral solution of the ensemble FRET efficiency in function of ω,

$$E_{da}^e = \frac{(\omega_{da}/\gamma)^\lambda}{1 + (\omega_{da}/\gamma)^\lambda} \tag{3.42}$$

where λ and γ are the fitting and exclusion constant respectively. The fit resulted in $\lambda = 1.1$ and $\gamma = 0.62\exp\left(-0.34r_0 + 1.63r_0^2\right)$ with r_0 being a measure of the closest distance of approach between donor and acceptor with respect to the Forster distance $r_0 = (\frac{r_{min}}{R_{da}})^2$ where r_{min} was taken to be 2 nm. In both instances, the FRET efficiency was found to be a function of the natural dimensionless number ω_{da} (which we here call the Förster acceptor number) which denotes the average number of acceptors in a Förster radius around the donor.

emFRET efficiency to stochastically distributed dyes in 2D

Here we extend the 2D e2FRET model to include multicolour energy transfer (*i.e.* emFRET), whereby N different species of dyes are stochastically distributed on a planar surface. Hence, for every donor species d, there is $(N - d)$ number of potential acceptors species.

Decay function with multiple acceptor species For a donor d with an initial lifetime (*i.e.* without acceptors) of τ_d^0, the ensemble-average rate constant is calculated within a radius R_c which contains Z_a acceptors for each separate species a depending on their density σ_a. The rate constant becomes

$$(\tau_d)^{-1} = \frac{1}{\tau_d^0}\left(1 + \sum_{a=d+1}^{N}\sum_{z=1}^{Z_a}\left(\frac{R_{da}}{r_{z,a}}\right)^6\right) \tag{3.43}$$

The solution for the decay function becomes,

$$\rho_d(t) = e^{-t/\tau_d^0}\Pi_{a=d+1}^{N}\left[\Pi_{z=1}^{Z_a}\exp\left(\frac{-t}{\tau_d^0}\left(\frac{R_{da}}{r_{z,a}}\right)^6\right)\right] \tag{3.44}$$

The decay function is then ensemble-averaged over the multiple acceptor species, which may be done separately when the acceptor densities are independent of one another, giving

$$< \rho_d(t) >_{Z_{d+1},Z_{d+2},\ldots,Z_N} = e^{t/\tau_d^0}\Pi_{a=d+1}^{N}[M_a(t)]^{Z_a} \tag{3.45}$$

where, as before,

$$M_a(t) = 1 - \frac{\pi\epsilon\sigma_a R_{da}^2}{N}\left(\frac{t}{\tau_d^0}\right)^{1/3} \tag{3.46}$$

For a large calculation radius, the ensemble-averaged decay function becomes

$$< \rho_d(t) >= \exp\left[\frac{-t}{\tau_d^0} - \frac{\epsilon}{\pi}\left(\sum_{a=d+1}^{N}\omega_{da}\right)\left(\frac{t}{\tau_d^0}\right)^{1/3}\right] \tag{3.47}$$

By comparing equations (3.47) and (3.39), it becomes apparent that the impact of having acceptors of different species on the ensemble FRET efficiency from a donor species d is that the Förster

acceptor number is replaced by the sum of the individual Förster numbers, *i.e.* $\omega_{da} \rightarrow \omega_d^m = \sum_{a=d+1}^{N} \omega_{da}$ where ω_d^m is termed the multicolour Förster acceptor number. Thus, the analytical solution of the ensemble multicolour FRET efficiency (emFRET) from donor d to all potential acceptors becomes

$$E_d^{em} = 1 - \sum_{j=0}^{\infty} \left(-\frac{\epsilon \omega_d^m}{\pi} \right)^j \frac{\Gamma(j/3+1)}{j!} \tag{3.48}$$

Hence, the total emFRET efficiency for a given dye d may be calculated from the e2FRET efficiency relations using the Förster acceptor number transformation ($\omega_{da} \rightarrow \omega_d^m = \sum_{a=d+1}^{N} \omega_{da}$). This transformation is also applied to the closed-form solution from Eq.(3.42), giving

$$E_d^{em} = \frac{\left(\omega_d^m/\gamma\right)^\lambda}{1 + \left(\omega_d^m/\gamma\right)^\lambda} \tag{3.49}$$

Effective Förster radius The solution for the emFRET efficiency is equivalent to the e2FRET scenario, for every donor d, when the different acceptor species are re-defined as one virtual dye species with a multicolour Förster acceptor number ω_d^m, and a multicolour density $\sigma_d^m = \sum_{a=d+1}^{N} \sigma_a$. To conserve the total energy lost from donor d, an effective Förster radius, R_d^e may be determined such that

$$\omega_d^m = \pi (R_d^e)^2 \sigma_d^m = \pi (R_d^e)^2 \sum_{a=d+1}^{N} \sigma_a \tag{3.50}$$

For independent densities of dyes, $\omega_d^m = \sum_{a=d+1}^{N} \omega_{da}$, giving

$$\omega_d^m = \sum_{a=d+1}^{N} \omega_{da} = \pi \sum_{a=d+1}^{N} (R_{da})^2 \sigma_a \tag{3.51}$$

By equating (3.50) and (3.51), a general expression for the effective Förster radius is then

$$R_d^e = \left(\sum_{a=d+1}^{N} \frac{R_{da}^2 \sigma_a}{\sigma_d^m} \right)^{1/2} \tag{3.52}$$

Distribution of energy between multiple acceptors Having found the ensemble multicolour efficiency of transfer for a given donor d (*i.e.* including transfer to all acceptor species), here we calculate the distribution of energy amongst acceptors of different species as a function of their acceptor numbers w_{da}. Using equation (3.25), $N = 4$, and $d = 1$,

$$\frac{E_{12}^m}{E_{12}} = \frac{k_{12}}{1/\tau_1^0 + k_{12} + k_{13} + k_{14}} \frac{1/\tau_1^0 + k_{12}}{k_{12}} = 1 - \frac{k_{13} + k_{14}}{1/\tau_1^0 + k_{12} + k_{13} + k_{14}} = 1 - E_{13}^m - E_{14}^m \tag{3.53}$$

where E^m_{da} is the efficiency of energy transfer between the two dyes when all four dyes are present, and E_{da} is the efficiency of transfer when no other dyes are present (2-colour FRET). Considering that $E^m_1 = E_{12} + E_{13} + E_{14}$, which also applies to the ensemble values. It follows that,

$$E^{em}_{12} = E^e_{12}\left(1 + E^{em}_{12} - E^{em}_1\right) \tag{3.54}$$

and hence,

$$E^{em}_{12} = \frac{E^e_{12}(1 - E^{em}_1)}{1 - E^e_{12}} = (E^{em}_1 - 1)\left(1 + \frac{1}{E^e_{12} - 1}\right) \tag{3.55}$$

Next, using equation (3.42), $a_{da} = \omega_{da}/\gamma$, and the approximation $(\sum_f a_f)^\lambda \approx \sum_f a^\lambda_f$ since $\lambda \approx 1$, to show that,

$$E^{em}_{12} = \left(\frac{a^\lambda_{12} + a^\lambda_{13} + a^\lambda_{14}}{1 + a^\lambda_{12} + a^\lambda_{13} + a^\lambda_{14}} - 1\right)\left(1 + \frac{1}{\frac{a^\lambda_{12}}{a^\lambda_{12}+1} - 1}\right) \tag{3.56}$$

which can be simplified to

$$E^{em}_{12} = \frac{a^\lambda_{12}}{1 + a^\lambda_{12} + a^\lambda_{13} + a^\lambda_{14}} \tag{3.57}$$

and re-written as

$$E^{em}_{12} = \left(\frac{a_{12}}{a_{12} + a_{13} + a_{14}}\right)^\lambda \frac{(a_{12} + a_{13} + a_{14})^\lambda}{1 + (a_{12} + a_{13} + a_{14})^\lambda} = \left(\frac{a_{12}}{a_{12} + a_{13} + a_{14}}\right)^\lambda E^{em}_1 \tag{3.58}$$

It is thus apparent that the energy is distributed amongst acceptors proportional to the ratio $\left(\omega_{da}/\omega^T_d\right)^\lambda$, and hence the individual FRET efficiency is

$$E^{em}_{da} = \left(\frac{\omega_{da}}{\omega^m_d}\right)^\lambda E^{em}_d \tag{3.59}$$

This equality is then plugged in directly into the MFM relations (**Eqs 3.2a and 3.2b**) and thus is paramount to enable calculation of the ensemble fluorescence of each dye in the MFM.

Application of MFM and emFRET to the BMPs in this work

In deriving the emFRET and MFM, several theoretical assumptions were taken. Here, we discuss how they apply to BMPs generated using our labeling strategy.

- **Non-saturated exciton density.** In the MFM, the ensemble fluorescence of dye f is calculated as the sum of laser-sensitized and FRET-sensitized fluorescence. Furthermore, by reducing the energy transfer problem to single-donor environments, the emFRET model

assumes negligible competition between donors to the same acceptors. Therefore, as with previous derivations of this kind [13, 14, 27, 45], exciton-exciton competition was neglected. For this assumption to hold, the concentration of excited dyes after interrogation by the laser at steady-state must be of low enough density to produce negligible donor-donor and laser-donor competition [25, 26]. The exciton density is primarily dependent on the dye density and the laser irradiance values. For the laser irradiance used in this work ($\approx 10^7$ W/m^2), the exciton flux on area of radius R_0 is estimated to be $10 - 10^3$ s^{-1} which is indeed negligible relative to typical fluorescence lifetimes. We further investigated this assumption using the ExiFRET MonteCarlo numerical simulation tool [26], which takes into account donor-donor competition, we calculated the FRET efficiency for varying donor/acceptor densities, and saw no impact from donor-donor competition on the FRET efficiency at the typical laser irradiance (**Figure 3.13**).

- **HomoFRET and self-quenching.** HomoFRET, which occurs between homogeneous dye species [39, 42, 46], is possible for all dyes used in this work due to their small Stokes shift. However, homoFRET does not inherently result in a decreased fluorescence intensity, and was not considered in the emFRET model [39]. On the other hand, self-quenching is commonly considered to be the result of non-fluorescent dye aggregates [47, 48] and only becomes an appreciable effect at short inter-dye distances [47], at which point its effect may be enhanced by FRET resulting in a non-fluorescent 'sink' for excitons [48, 49]. Self-quenching was not modelled either as it was thought to set a practical upper limitation on dye density. However, a breakdown of the linear basal fluorescence, consistent with self-quenching, was observed for Cy5 and Cy5.5 (*i.e.* dyes **3** and **4** respectively) at non-saturating densities, suggesting that one could improve the performance of the model by using a non-linear, experimentally determined fit.

- **Randomly positioned dyes in 2D.** Ensemble averaging, which is performed across all donor concentrations in e2FRET and emFRET, considers homogeneous distributions of dyes on a planar surface. Here, the use of DNA oligos as linkers with homogeneous affinity to the surface, and the one-pot nature of the labelling ensures a homogeneously distributed ensemble of dyes.

- **Independent dye densities.** To perform ensemble averaging of the decay function across all dyes of different species, the densities and positioning of different dyes must be independent. As demonstrated in **Figure 3.2c**, the densities of different LOs are indeed independent. This is a chief consideration since barcoding techniques such as sequential labelling or labelling with inhomogeneous affinities will not yield independent densities.

- **Isotropic orientation of dyes.** The use of a precise average value for the orientation factor and hence the Förster radius between two dyes (R_{da}) is a complex issue [50]. In the case of isotropic dipole orientations, a value of $\kappa^2 = 2/3$ is most commonly to assume an ideal scenario where the re-orientation of dipoles happens rapidly within the lifetime of an exciton [51]. This 'dynamic isotropic' assumption has been shown, in the special case of randomly distributed ensembles in 2D, to be an adequate estimate that does not yield large errors in calculating the Förster radius for modest re-orientation speeds [45, 52, 53]. Furthermore, in the case of a static limit, a geometry-dependent ensemble value for the orientation factor may also be considered [14]. In this work, the relative orientation of the microparticle-linked dyes possess multiple degrees of freedom resulting from (i) the azimuthal angle from the streptavidin orientation and subsequent binding region of biotin, (ii) the normal angle the LO makes with the bead surface, and (iii) the dye orientation. As a result, the orientation can reasonably be considered isotropic; however, a completely dynamic re-orientation cannot be readily assumed here given the tendency of cyanine dyes to interact with DNA [54].

3.4.3 Supplementary Note 3: Calibration and parameter extraction

In (3.21) and (3.22), the physical parameters may now be determined experimentally by measuring a judicious selection of barcode combinations that allow calculation of specific physical parameters. Eq (3.21) is first normalized by background signal to yield the signal-to-background ratios of the four channels,

$$i_{11} = 1 + \beta'_{11}F^e_{11} + \beta'_{12}F^e_{21} \tag{3.60a}$$

$$i_{21} = 1 + \beta'_{21}F^e_{11} + \beta'_{22}F^e_{21} \tag{3.60b}$$

$$i_{32} = 1 + \beta'_{33}F^e_{32} + \beta'_{34}F^e_{42} \tag{3.60c}$$

$$i_{42} = 1 + \beta'_{43}F^e_{32} + \beta'_{44}F^e_{42} \tag{3.60d}$$

where $I_{cl}/I^0_{cl} = i^0_{cl}$ and $\beta_{cf}/I^0_{cl} = \beta_{cf}$. Then, to simplify the equations and reduce the number of unknowns, the parameters may be normalized to the self bleed-through constants using the change of variables, namely $\beta'_{ff}\mu_{cf} = \mu'_{cf}$ for $c = f$, $\beta'_{cf}/\beta'_{ff} = \beta''_{cf}$ for $c \neq f$, and $(\alpha_{cf}\beta'_{ff})/(\beta' cc) = \alpha'_{cf}$, to yield

$$i_{11} = 1 + F^e_{11} + \beta''_{12}F^e_{21} \tag{3.61a}$$
$$i_{21} = 1 + \beta''_{21}F^e_{11} + F^e_{21} \tag{3.61b}$$
$$i_{32} = 1 + F^e_{32} + \beta''_{34}F^e_{42} \tag{3.61c}$$
$$i_{42} = 1 + \beta''_{43}F^e_{32} + F^e_{42} \tag{3.61d}$$

where i_{cl} is the background normalized intensity in channel c.

To determine the parameters, we use the following procedure, which is applied to our BMPs in **Supplementary Figure 3.15**. Briefly, single-colour BMPs are used to determine the direct-excitation and bleed through constants, whereas two-colour BMPs are used to determine the proportionality constant as well as the labelling constant. For notation simplicity, the new variables of the system of equations are hereon renamed as $\beta'' \rightarrow \beta$, $\alpha' \rightarrow \alpha$, and $\mu' \rightarrow \mu$.

1. μ_f were determined using single-colour beads with increasing dye concentrations (Fig. S9b). This can be accomplished by fitting the i_{cl} data to the linear equation $i_c = 1 + \mu_f n_f$ with $f = c$.

2. Using the same single-colour barcodes, the bleed through constants B_{cf} may be found by fitting $i_c = 1 + \beta_{cf}\mu_f n_f$ with $f \neq c$.

3. From (3.61c), (3.61d), (3.22c), and (3.22d), we find α_{34} and E_{34} such that

$$E^{em}_{34} = 1 + \frac{1 - \beta_{34} - i_{32} + \beta_{34}i_{42}}{\mu_{32}n_3 - \beta_{43}\beta_{34}\mu_{32}n_3} \tag{3.62}$$

$$\alpha_{34} = \frac{-1 + \beta_{43} - \beta_{43}i_{32} + i_{42} + \mu_{42}n_4(-1 + \beta_{43}\beta_{34})}{1 - \beta_{34} - i_{32} + \beta_{34}i_{42} + \mu_{32}n_3 - \beta_{43}\beta_{34}\mu_{32}n_3} \tag{3.63}$$

   which allows us to find α_{34} from a BMP containing both Cy5 and Cy5.5 at non-saturating levels.

4. Equation (3.42) can be re-written as

$$\sigma_a = \frac{\gamma}{\pi(R_{da})^2}\left(\frac{E^{em}_{da}}{1 - E^{em}_{da}}\right)^{1/\lambda} \tag{3.64}$$

   which allows us to find σ_4 from (3.62). Then calculate t using $t = n_4/\sigma_4$.

Supplementary Table 3.1 – Parameters and variables

Variable	Variable type	Parameter	Notes
N	Scalar	Number of fluorophores	
f	Scalar	Fluorophore identifier	$1 < f < N$
C	Scalar	Number of channels	
c	Scalar	Channel identifier	$1 < c < C$
L	Scalar	Number of lasers	
l	Scalar	Laser identifier	$1 < l < L$
$\mathbf{I}$	CxL Matrix	where I_{cl} is the intensity of channel c in response to laser l	a function of n_f
I_{cl}	Scalar	Channel 'c' detector intensity in response for laser l	
$\mathbf{I^0}$	CxL Matrix	where I^0_{cl} is intensity of bare MPs in channel c for laser l	
$\mathbf{F^e}$	NxL Matrix	where F_{fl} is ensemble intensity of dye f for laser l	
F^e_{fl}	Scalar	where F_{fl} is ensemble intensity of dye f for laser l	
n_f	Scalar	Input amount of LO_f in coupling reaction (in pmols)	
$\mathbf{B}$	CxN Matrix	where β_{cf} is the bleed-through ratio from dye f into channel c	
β_{cf}	Scalar	bleed-through ratio - fraction of dye f fluorescence that is detected in channel c	
μ_{fl}	Scalar	Direct excitation constant for dye f excited by laser l	
F^s_{fl}	Scalar	Sensitized fluorescence of f = direct excitation by l + FRET from donors (ie. assumes no acceptors)	
F^0_{fl}	Scalar	Basal fluorescence intensity of dye f by l ie. assuming no donors/acceptors	
$\mathbf{A}$	NxN Matrix	Energy exchange ratio / FRET proportionality constant	
α_{da}	Scalar	FRET proportionality for donor d to acceptor a	$d < a$
E_{da}	Scalar	Single-molecule 2-colour FRET efficiency from donor d to acceptor a	$d < a$
E^e_{da}	Scalar	Ensemble-averaged 2-colour FRET efficiency from donor d to acceptor a	$d < a$
E^{em}_{da}	Scalar	Ensemble-averaged multicolour FRET efficiency from donor d to acceptor a	$d < a$
E^{em}_d	Scalar	Ensemble-averaged multicolour FRET efficiency from donor d to all acceptors	$E^{em}_N = 0$
σ_f	Scalar	Surface density of dye f (in m^{-2})	
t	Scalar	Labelling constant	
R_{da}	Scalar	Förster radius between two dyes in nm	
ω_{da}	Scalar	"Förster acceptor number" *i.e.* the number of a acceptors in an R_{da} radius	
$\rho(t)$	Scalar	Instantaneous probability of decay	
τ	Scalar	Lifetime of excited-state decay (in s)	
k^x	Scalar	Rate of process x (in s^{-1})	
r_z	Scalar	Distance between donor and acceptor z	
γ	Scalar	Exclusion constant[27]	
λ	Scalar	Fitting constant[27]	

5. When $n_3 = n_4 = 0$, it follows that $F^e_3 = F^e_4 = 0$ and $E^{em}_1 = E^{em}_{12}$. Furthermore, from (3.61a), (3.61b), (3.22a), and (3.22b) it follows that

$$E^{em}_{12} = \frac{-1 + \beta_{12} + i_{11} - \beta_{12} i_{21} - \mu_{11} n_1 + \beta_{21}\beta_{12} n_1 \mu_{11}}{-\mu_{11} n_1 (1 - \beta_{21}\beta_{12})} \tag{3.65}$$

and

$$\alpha_{12} = \frac{-1 + \beta_{21} - \beta_{21} i_{11} + i_{22} - \mu_{21} n_2 + \beta_{12}\beta_{21}\mu_{21} n_2}{1 - \beta_{12} - i_{11} + \beta_{12} i_{21} + \mu_{11} n_1 - \beta_{21}\beta_{11}\mu_{11} n_1} \tag{3.66}$$

Using the aforementioned steps, the determined physical parameters are shown in **Supplementary Table 3.2**.

Supplementary Table 3.2 – Extracted MFM Parameters

Parameter	Mean (SD)
Direct excitation constant	$\mu = \begin{bmatrix} 0.87\ (0.04) \\ 0.11\ (0.001) \\ 4.154\ (0.068) \\ 2.131\ (0.102) \end{bmatrix}$
Bleed-through matrix	$B = \begin{bmatrix} 1 & 0.065\ (0.004) & 0 & 0 \\ 0.167\ (0.01) & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.013\ (0.002) \\ 0 & 0 & 0.365\ (0.036) & 1 \end{bmatrix}$
FRET proportionality matrix	$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.307\ (0.048) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.482\ (0.041) & 0 \end{bmatrix}$
Labelling constant	$t = 2.16e^{14}\ (2.03e^{13})$

3.4.4 Supplementary Figures

References

1. Nolan, J. P. & Sklar, L. a. Suspension array technology: evolution of the flat-array paradigm. *Trends Biotechnol.* **20,** 9–12 (2002).

2. Wilson, R., Cossins, A. R. & Spiller, D. G. Encoded microcarriers for high-throughput multiplexed detection. *Angewandte Chemie - International Edition* **45,** 6104–6117 (2006).

3. Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J. & Kettman, J. R. Advanced multiplexed analysis with the FlowMetrix(TM) system. *Clinical Chemistry* **43,** 1749–1756 (1997).

4. Wu, W. *et al.* Antibody array analysis with label-based detection and resolution of protein size. *Molecular & cellular proteomics : MCP* **8,** 245–57 (2009).

5. Wang, L. & Tan, W. Multicolor FRET silica nanoparticles by single wavelength excitation. *Nano Letters* **6,** 84–88 (2006).

6. Vaidya, S. V., Couzis, A. & Maldarelli, C. Reduction in aggregation and energy transfer of quantum dots incorporated in polystyrene beads by kinetic entrapment due to cross-linking during polymerization. *Langmuir* **31,** 3167–3179 (2015).

7. Clapp, A. R., Medintz, I. L. & Mattoussi, H. Förster resonance energy transfer investigations using quantum-dot fluorophores. *ChemPhysChem* **7,** 47–57 (2006).

8. Wagh, A. *et al.* Polymeric nanoparticles with sequential and multiple FRET cascade mechanisms for multicolor and multiplexed imaging. *Small* **9,** 2129–2139 (2013).

9. Stuchlý, J. *et al.* An automated analysis of highly complex flow cytometry-based proteomic data. *Cytometry Part A* **81A,** 120–129 (2012).

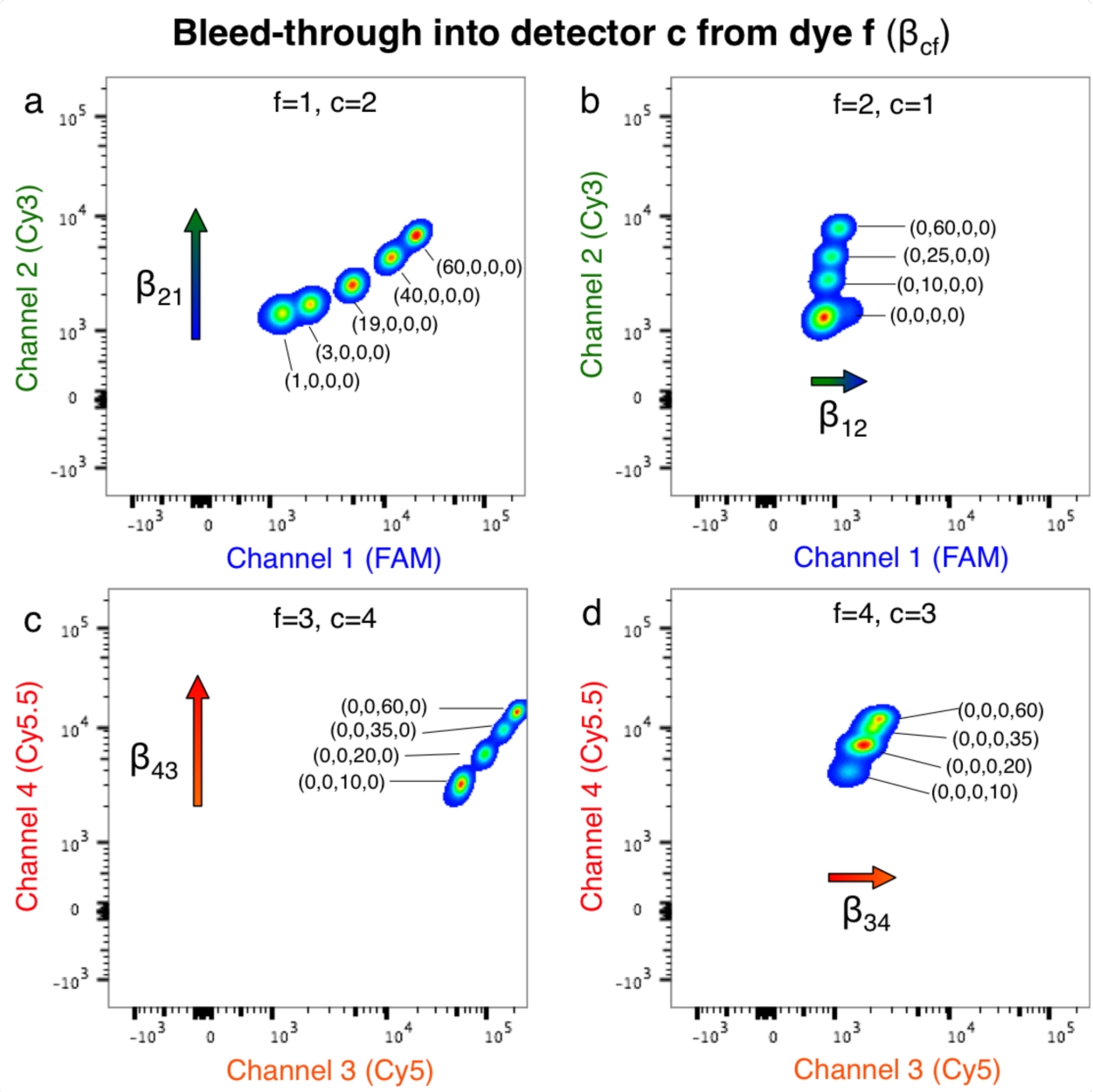


Figure 3.7: **Bleed-through between dyes and off-detectors on the FACS CANTO II platform.** Single-colour BMPs with increasing amounts of dyes used to probe the bleed-through from (a) dye 1 to channel 2 (β_{21}), (b) dye 2 to channel 1 (β_{12}), (c) dye 3 to channel 4 (β_{43}), and (d) dye 4 to channel 3 (β_{43}). As expected from the large spectra overlap between dyes seen in the Figure 3.1a, β_{21} and β_{43} are most prominent.

10. Han, M., Gao, X., Su, J. Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat. Biotechnol.* **19,** 631–635 (2001).

11. Wang, G. *et al.* Highly efficient preparation of multiscaled quantum dot barcodes for multiplexed hepatitis b detection. *ACS Nano* **7,** 471–481 (2013).

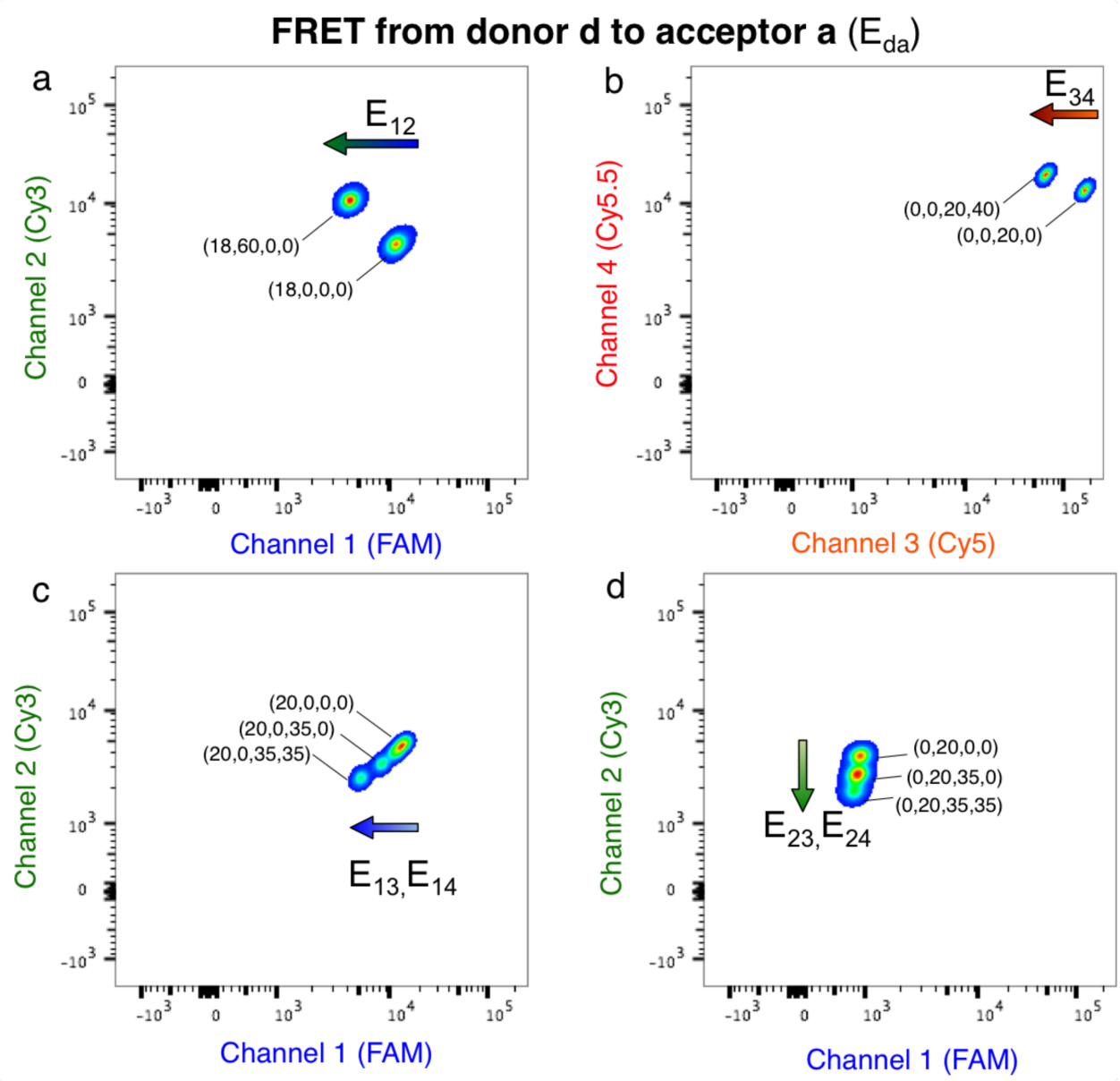


Figure 3.8: **Effect of inter-dye FRET on ensemble BMP fluorescence.** Two-colour BMPs used to probe energy loss through FRET from (a) dye 1 to 2 (E_{12}), (b) dyes 3 to 4 (E_{34}), (c) dyes 1 to 3 and 4 (E_{13} and E_{14}), (d) dyes 2 to 3 and 4 (E_{23} and E_{24}). Note that concentration, and thus surface density, of the donors is constant for all experiments shown here, therefore any change in donor fluorescence can only be ascribed to FRET loss.

12. Ferguson, J. A., Steemers, F. J. & Walt, D. R. High-density fiber-optic DNA random microsphere array. *Anal. Chem.* **72,** 5618–5624 (2000).

13. Forster, T. Expermentelle und theoretische Untersuchung des zwischengmolekularen uebergangs von elektronenanregungsenergie. *Naturforsch. A. Astrophys. Phys. Phys. Chem.* **4a,** 321–327 (1949).

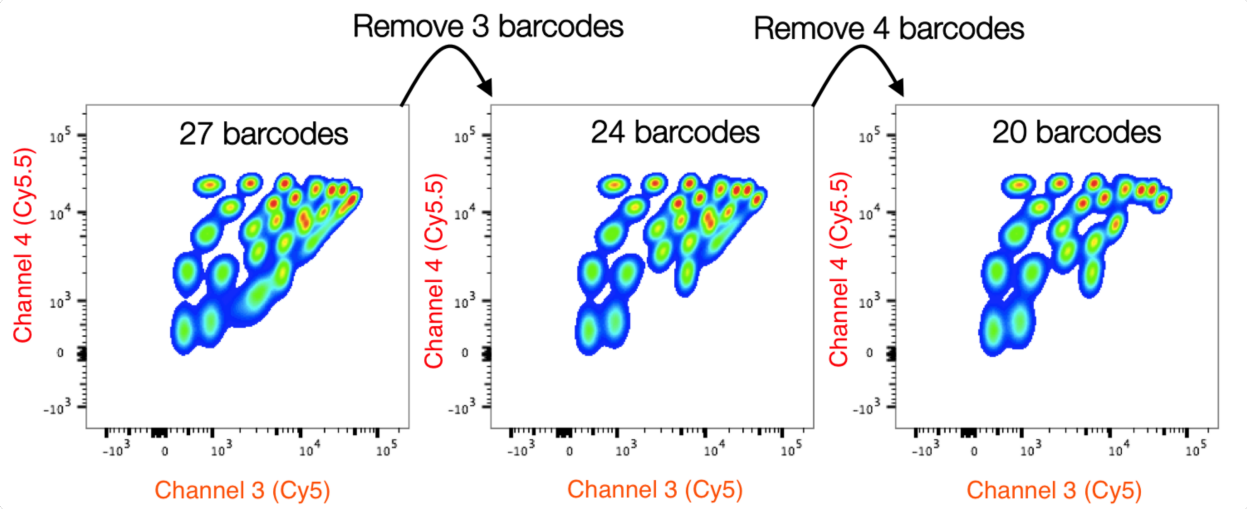


Figure 3.9: **Barcoding at FRET regimes is iterative.** Due to FRET, the position of barcode clusters within the intensity parameter space does not correspond linearly to surface stoichiometry, leading to difficulties in designing the barcodes clusters and leading to overlapping barcodes that have to be removed or optimized. Here, we illustrate the iterative nature of designing codes without a model to predict the response and thus guide the process. In this example, (>3) trial-and-error experiments were required to optimize barcodes in $I_3 - I_4$ intensity space and increase barcoding capacity without risking misidentified barcodes due to cluster overlap. Without a model, the barcoding space in $I_3 - I_4$ was not fully optimized after three iterations, as there were significant empty spaces and overlap. Note that for 4 non-orthogonal classifier dyes, the number of parameter spaces that must be optimized is at least equal to the number of clusters in $I_3 - I_4$ (*i.e.* >20 parameter spaces, and >60 design iterations) and becomes a lengthy and tedious process that scales exponentially with additional classifier dyes.

14. Wolber, P. & Hudson, B. An analytic solution to the Forster energy transfer problem in two dimensions. *Biophys. J.* **28,** 197–210 (1979).

15. Watrob, H. M., Pan, C. P. & Barkley, M. D. Two-step FRET as a structural tool. *J. Am. Chem. Soc.* **125,** 7336–7343 (2003).

16. Fabian, A., Horvath, G., Vamosi, G., Vereb, G. & Szollosi, J. TripleFRET measurements in flow cytometry. *Cytometry A* **83 A,** 375–385 (2013).

17. Lee, J. *et al.* Single-molecule four-color FRET. *Angew. Chem. Int. Ed.* **49,** 9922–9925 (2010).

18. Uphoff, S. *et al.* Monitoring multiple distances within a single molecule using switchable FRET. *Nat. Methods* **7,** 831–836 (2010).

19. Bunt, G. & Wouters, F. S. FRET from single to multiplexed signaling events. *Biophys. Rev.* **9,** 119–129 (2017).

20. Buckhout-White, S. *et al.* Assembling programmable FRET-based photonic networks using designer DNA scaffolds. *Nat. Commun.* **5,** 5615–5630 (2014).

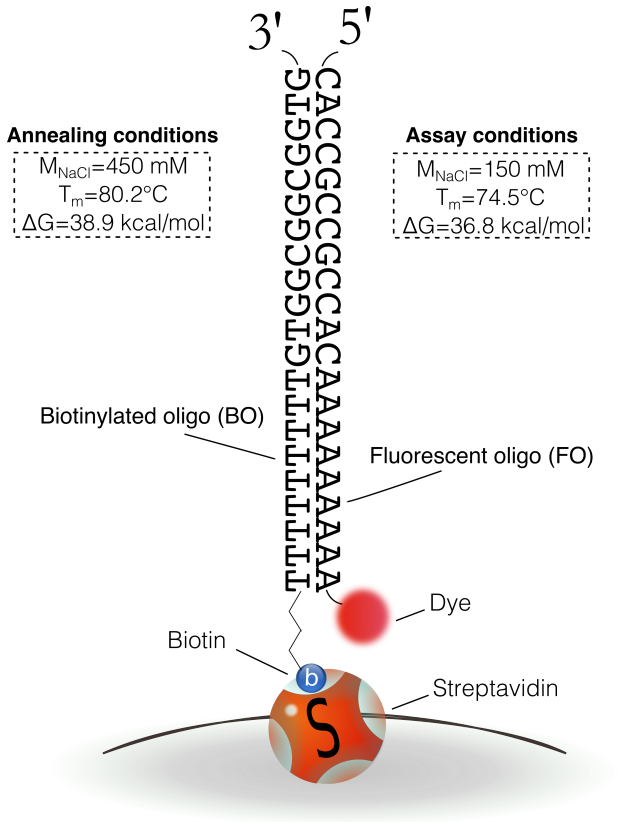


Figure 3.10: **Sequence and structure of the linking oligo (LO).** The linking oligo consists of two complementary 21-nt strands, the biotinylated oligo (BO) and the fluorescent oligo (FO), covalently bound with a biotin and a dye at the 5' and 3' ends, respectively. The biotin was attached to the 5' end using a six-carbon (C6) linker. The sequence of the oligos were designed to avoid G-C bases near the surface to avoid quenching by G bases, while maintaining a high melting temperature of >70 degrees at all times. Using in silico DNA analysis such as Mfold, we verified the absence of major secondary structure formation.

21. Raicu, V. Efficiency of resonance energy transfer in homo-oligomeric complexes of proteins. *Journal of Biological Physics* **33,** 109–127 (2007).

22. Liu, J. & Lu, Y. FRET study of a trifluorophore-labelled DNAzyme. *Journal of the American Chemical Society* **124,** 15208–15216 (2002).

23. Shapiro, H. M. *Practical Flow Cytometry* (Wiley, Hoboken, 2003).

24. Matyus, L. Fluorescence resonance energy transfer measurements on cell surfaces. A spectroscopic tool for determining protein interactions. *J. Photochem. Photobiol.* **12,** 323–37 (1992).

25. Berney, C. & Danuser, G. FRET or no FRET: a quantitative comparison. *Biophys. J.* **84,** 3992–4010 (2003).

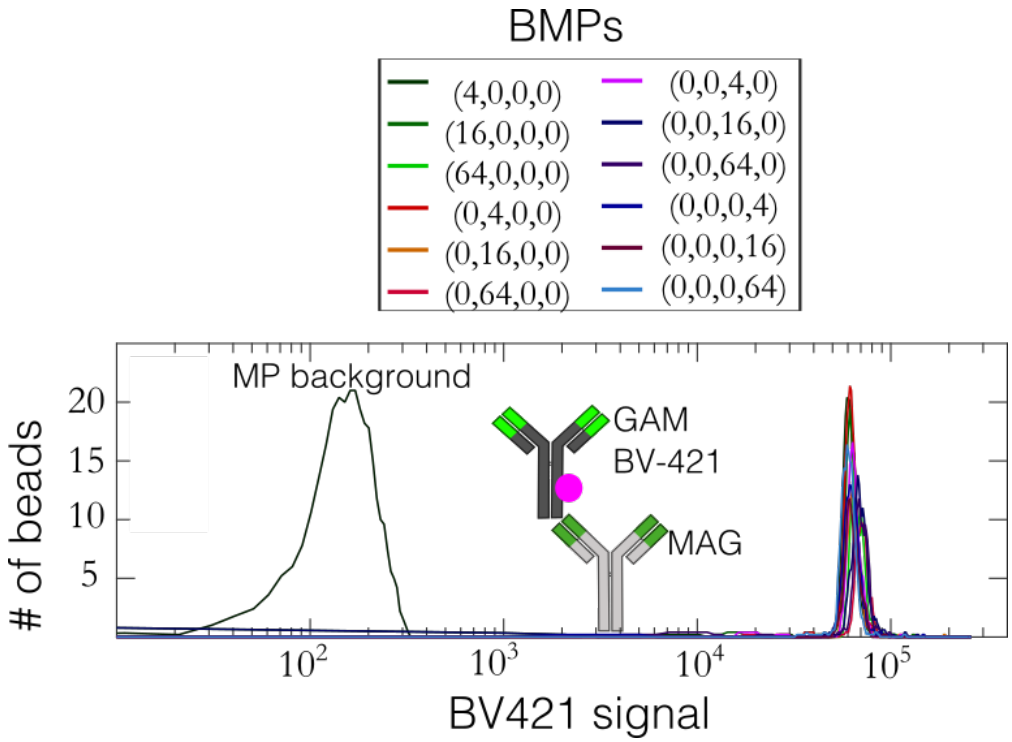


Figure 3.11: **Homogeneous antibody labelling across barcodes.** To test whether the antibody densities were dependent of barcodes and the densities of different LOs, mouse-anti-goat IgG co-labeled on single-color BMPs (barcodes in legend) were labeled with BV421-labeled secondary-Ab (goat anti-mouse). The histogram plots the bead intensities for different barcodes, and compares it with the background signal of the BMPs in the BV421 channel. The CV of BMPs in the BV-421 has been determined to be 5% which is well within the expected range for experiments involving manual pipetting. As a direct result of the one-pot labelling with conserved LO concentration, the surface density of antibodies was barcode-independent.

26. Corry, B., Jayatilaka, D. & Rigby, P. A flexible approach to the calculation of resonance energy transfer efficiency between multiple donors and acceptors in complex geometries. *Biophys. J.* **89,** 3822–36 (2005).

27. Koppel, D. E., Fleming, P. J. & Strittmatter, P. Intramembrane positions of membrane-bound chromophores determined by excitation energy transfer. *Biochemistry* **18,** 5450–5457 (1979).

28. Nguyen, H. Q. *et al.* Programmable microfluidic synthesis of over one thousand uniquely identifiable spectral codes. *Adv. Opt. Mater.* **5,** 1–6 (2016).

29. Lee, J. *et al.* Universal process-inert encoding architecture for polymer microparticles. *Nat. Mater.* **13,** 524–529 (2014).

30. Fielding, A. H. *Cluster and classification techniques for the biosciences* **9,** 1689–1699 (Cambridge University Press, 2007).

31. Stuchlý, J. *et al.* An automated analysis of highly complex flow cytometry-based proteomic data. *Cytometry A* **81A,** 120–129 (2012).

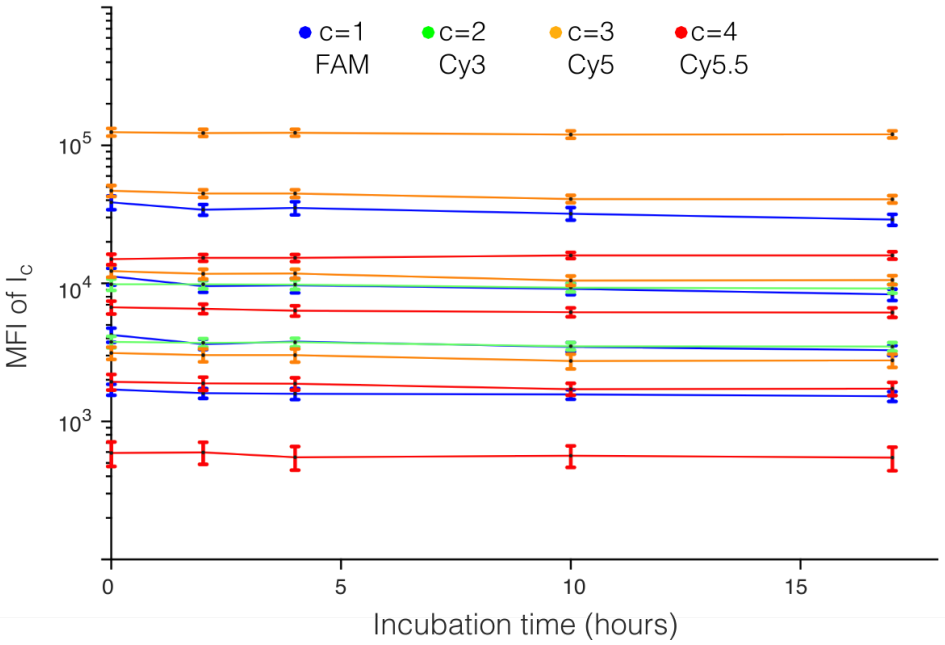


Figure 3.12: **Stability of BMP signal during lengthy incubation steps.** 16 single-colour BMPs with different intensity levels (same barcodes synthesized in Fig. S5) were mixed and incubated for up to 17 hours at room temperature. A sample of the BMPs was measured at different time points, and the fluorescence intensity of the corresponding detector was recorded. The MFI of the cyanine dyes BMPs (cy3, cy5, and cy5.5) remained within 2-10% of initial MFI during entire incubation period. The MFI of FAM labeled BMPs exhibited a larger drop, with a maximal drop of 15% observed, which was expected as FAM is prone to photobleaching and/or chemical bleaching. Interestingly, the majority of the fluorescence drop for FAM occurred within the first 1-2 hours of incubation. For future studies, FAM may be replaced by organic dyes with similar spectra (*e.g.* Cy2 or AlexaFluor488) to improve signal stability.

32. King, L. E. *et al.* Ligand binding assay critical reagents and their stability: recommendations and best practices from the Global Bioanalysis Consortium Harmonization Team. *AAPS J.* **16,** 504–515 (2014).

33. Uhlen, M. *et al.* A proposal for validation of antibodies. *Nat. Methods* **13,** 823–827 (2016).

34. Pla-Roca, M. *et al.* Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & cellular proteomics : MCP* **11** (2012).

35. Juncker, D., Bergeron, S., Laforte, V. & Li, H. Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. *Current Opinion in Chemical Biology* **18,** 29–37 (2014).

36. Schweitzer, B. *et al.* Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnology* **20,** 359–365 (2002).

37. Gonzalez, R. M. *et al.* Development and Validation of Sandwich ELISA Microarrays with Minimal Assay Interference. *J. Proteome Res.* **7,** 2406–2414 (2008).

38. Hardin, B. E. *et al.* Increased light harvesting in dye-sensitized solar cells with energy relay dyes. *Nat. Photonics* **3,** 406–411 (2009).

39. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy* 443–472 (Springer, New York, 2006).

40. Förster, T. in *Modern Quantum Chemistry Part III: Action of Light and Organic Crystals* 93–137 (Academic Press, New York, 1965).

41. Dempster, A., Laird, N. & Donald, R. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Series B* **39,** 1–38 (1977).

42. Jares-Erijman, E. a. & Jovin, T. M. FRET imaging. *Nature biotechnology* **21,** 1387–1395 (2003).

43. Clegg, R. M. Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymology* **211,** 353–388 (1992).

44. Fabian, A. I., Rente, T., Szollosi, J., Matyus, L. & Jenei, A. Strength in Numbers: Effects of Acceptor Abundance on FRET Efficiency. *ChemPhysChem* **11,** 3713–3721 (2010).

45. Snyder, B. & Freire, E. Fluorescence energy transfer in two dimensions. A numeric solution for random and nonrandom distributions. *Biophysical journal* **40,** 137–48 (1982).

46. Chan, F. T. S., Kaminski, C. F. & Schierle, G. S. K. HomoFRET fluorescence anisotropy imaging as a tool to study molecular self-assembly in live cells. *ChemPhysChem* **12,** 500–509 (2011).

47. Tleugabulova, D., Zhang, Z. & Brennan, J. D. Characterization of bodipy dimers formed in a molecularly confined environment. *Journal of Physical Chemistry B* **106,** 13133–13138 (2002).

48. Conroy, E. M., Li, J. J., Kim, H. & Algar, W. R. Self-Quenching, Dimerization, and Homo-FRET in Hetero-FRET Assemblies with Quantum Dot Donors and Multiple Dye Acceptors. *Journal of Physical Chemistry C* **120,** 17817–17828 (2016).

49. Luchowski, R. *et al.* Single Molecule Studies of Multiple-Fluorophore Labeled Antibodies. Effect of Homo-FRET on the Number of Photons Available Before Photobleaching R. *Curr Pharm Biotechnol* **9,** 411–420. arXiv: NIHMS150003 (2008).

50. Medintz, I. & Hildebrandt, N. *FRET - Förster Resonance Energy Transfer* (2013).

51. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature Methods* **5,** 507–516 (2008).

52. Fung, B. K.-K. & Stryer, L. Surface density determination in membranes by fluorescence energy transfer. *Biochemistry* **17,** 5241–5248 (1978).

53. Loura, L. M. S. Simple estimation of forster resonance energy transfer (FRET) orientation factor distribution in membranes. *International Journal of Molecular Sciences* **13,** 15252–15270 (2012).

54. Levitus, M. & Ranjit, S. Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Quarterly reviews of biophysics* **44,** 123–151 (2011).

55. Corry, B., Jayatilaka, D. & Rigby, P. A flexible approach to the calculation of resonance energy transfer efficiency between multiple donors and acceptors in complex geometries. *Biophysical journal* **89,** 3822–36 (2005).

56. Wolber & Hudson. An analytic solution to the Förster energy transfer problem in two dimensions. *Biophysical Journal* **28,** 197–210 (1979).

57. Koppel, D. E., Fleming, P. J. & Strittmatter, P. Intramembrane positions of membrane-bound chromophores determined by excitation energy transfer. *Biochemistry* **18,** 5450–5457 (1979).

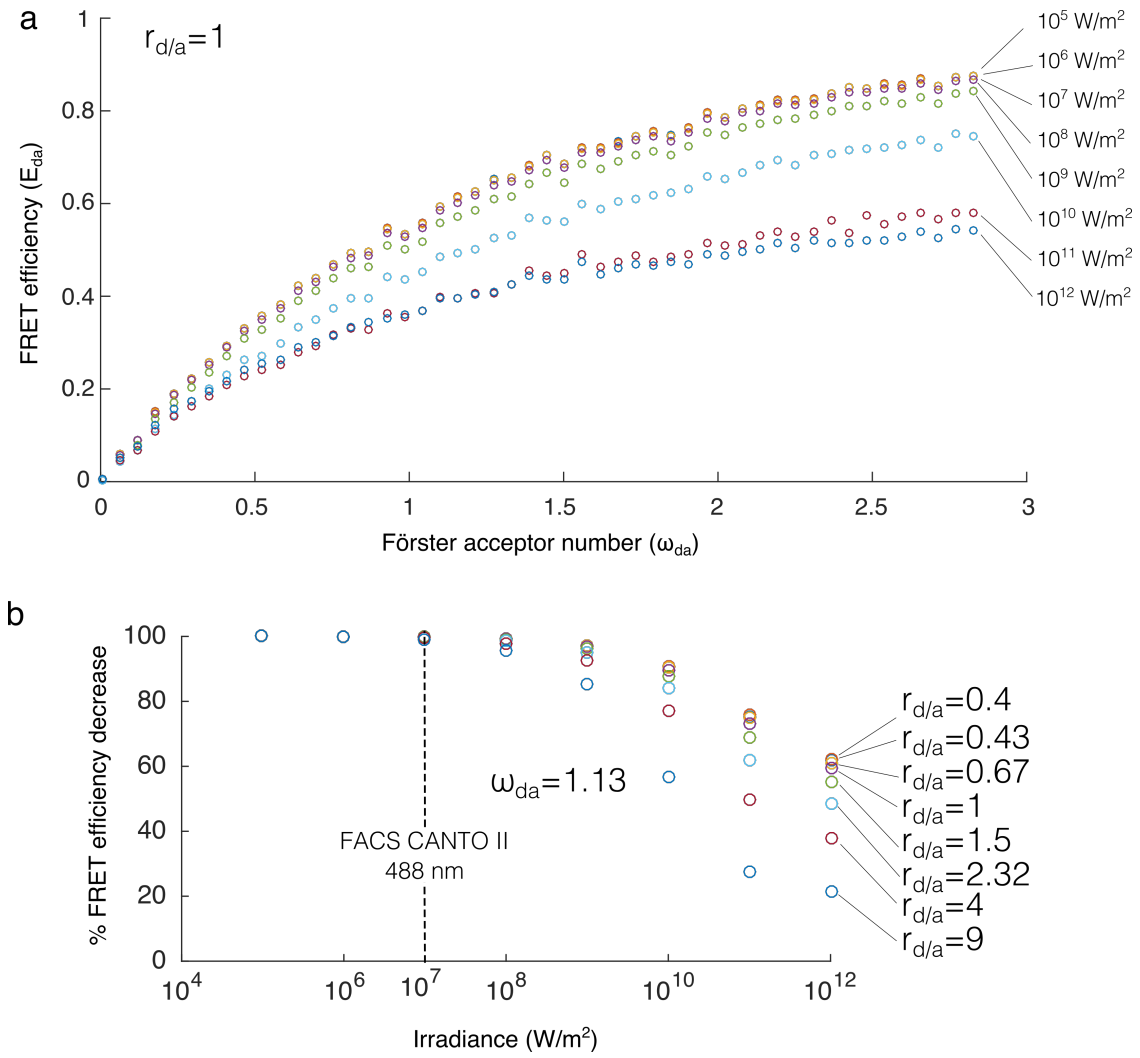


Figure 3.13: **Donor-donor competition threshold: numerical calculation using ExiFRET.** Here we investigate, through the use of ExiFRET, a Monte Carlo-based numerical calculation developed by Corry et al.[55] , the dependence of the FRET efficiency on donor-donor competition (i.e. competition for the same acceptor). (a) The FRET efficiency is calculated at varying levels of laser irradiances with increasing acceptor concentrations. The donor-to-acceptor ratio (rda) is kept at 0.5. FRET becomes dependent on and decreases with laser irradiances only at high levels ($>10^9 W/m^2$). At that point, the availability of unexcited acceptors at any given moment decreases, resulting in increased competition and decreased FRET transfers. (b) The reduction % of FRET efficiency due to donor-donor competition is plotted with respect to at increasing rd/a, highlighting the highest-irradiance laser used in this work (488 nm, FACS CANTO II, $10^7 W/m^2$). As expected, at high levels of irradiances, increasing donors results in decreased energy transfer due to donor-donor competition, with radiative decay for the donor becoming more prominent due unavailable (non-excited) acceptors. However, for the weak irradiances levels considered in this work, donor-donor competition can be safely assumed even at exceedingly large donor-to-acceptor ratios. The simulations' calculation area included 500 host molecules with 2 nm radii (no intra-FRET), and the FRET efficiency was averaged for 2000 configurations with 500 excitons each.

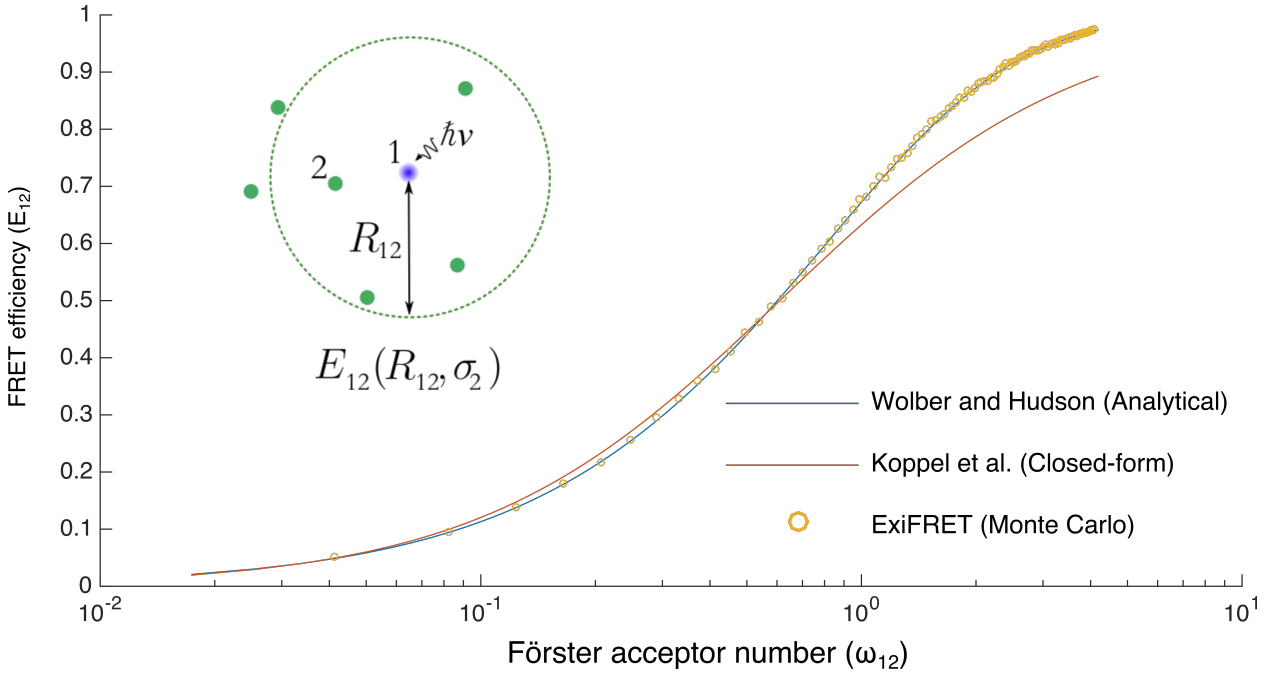


Figure 3.14: **Comparison between analytical, closed-form, and numerical (MonteCarlo) solutions for ensemble two-colour FRET (e2FRET).** (inset) Schematic representation of the e2FRET model, which considers an excited donor with multiple surrounding acceptors of the same type (*i.e.* the same Forster radius). Donor-donor competition is neglected in this model. The plot compares the analytical (Wolber and Hudson[56], closed-form (Koppel et al.[57]), and numerical (ExiFRET MonteCarlo web-module solutions. The closed-form solution agrees well to approximately 60% FRET transfer efficiency, above which there is a small deviation which was acceptable for this study. It is important to note that the closed-form solution affords enhanced speed and little computation power in the computation of the FRET, compared to exiFRET or the power series, which was essential to perform rapid design and optimization of the BMP fluorescence on a laptop computer.

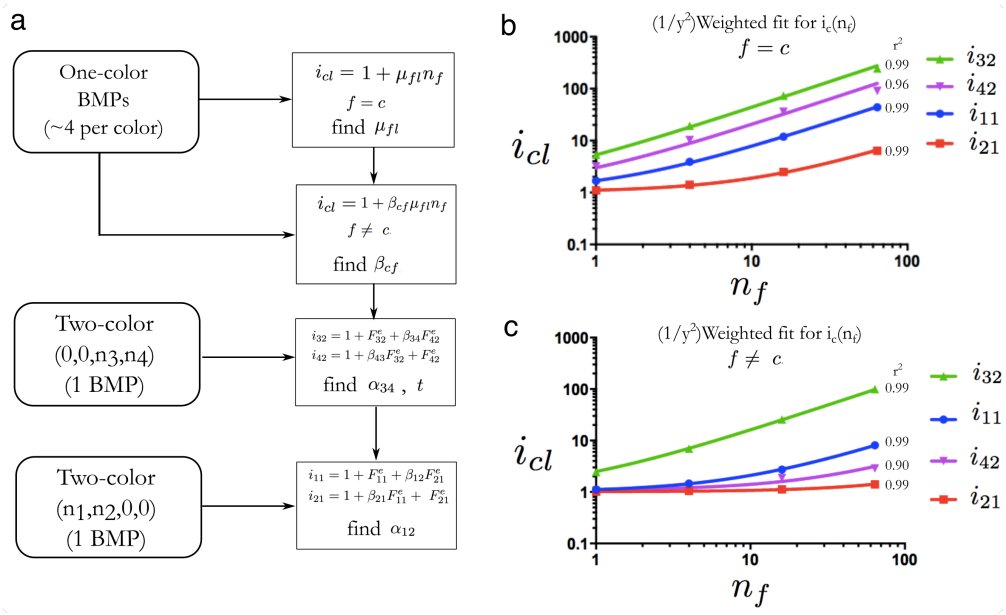


Figure 3.15: **Calibration of the MFM to extract physical parameters.** (a) The parameters within the MFM equations were determined using 18 judiciously selected barcodes via the process flow described here and in Supplementary Note 3. First, one-colour BMPs are used to perform a linear fit of the linear basal fluorescence to input dye amounts using the equations shown, and extract direct-excitation constants. Second, the same one-colour BMPs barcodes can be fit to off-detectors ($f \neq c$) intensities as shown by the equations, to determine the bleed-through constants. Finally, the FRET proportionality and labeling constants are determined using two-colour BMPs. (b,c) Fitting of the one-colour BMPs by linear regression to calculate the (b) direct excitation and (c) bleed-through constants was performed using relative weighting ($1/y^2$) during least-squares minimization which accounts for the heteroscedasticity of the BMP intensity where CV=cte with respect the dye concentrations (Figure 3.2b). The fitted lines are plotted and the R-square values are presented next to each fit.

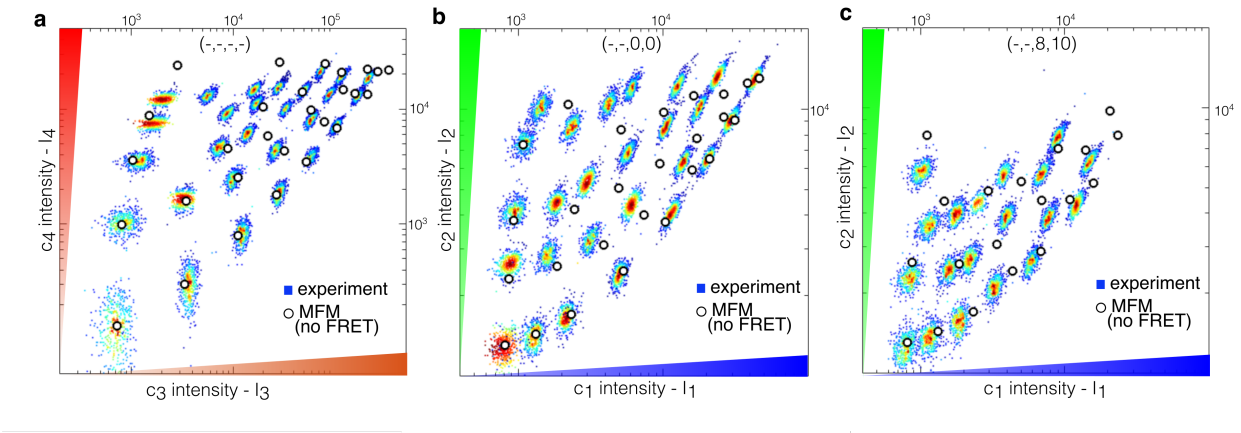


Figure 3.16: **Mismatch between BMP clusters and MFM predictions when neglecting mFRET.** To highlight the impact of the emFRET model within the MFM, the FRET efficiencies were set to 0 for all donors, and the MFM-predicted values were overlaid with the experimental intensity scatter plots from Figure 3.4(c-e). As expected, without taking emFRET into account, the MFM prediction rapidly diverges from the observed intensities of the barcodes which undergo inter-dye FRET. The discrepancy is more pronounced for the clusters in $I_1 - I_2$ in the presence of n_3 and n_4 which are subject to mFRET cascades, as can be seen in panel (c), highlighting the importance of accounting for FRET for predictive barcode design.

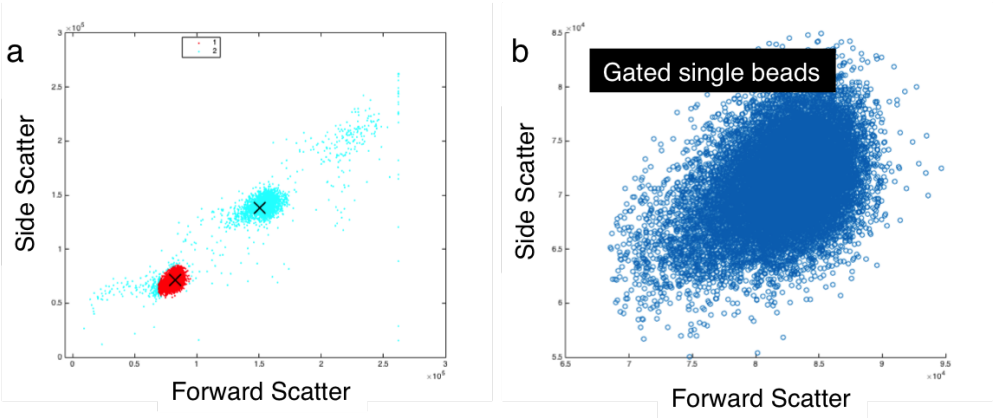


Figure 3.17: **Automated gating of single beads.** (a,b) Single beads (in red) were isolated from bead-aggregates and background particles (in blue) through automated clustering by GMM of the forward and side scatter detector intensities.

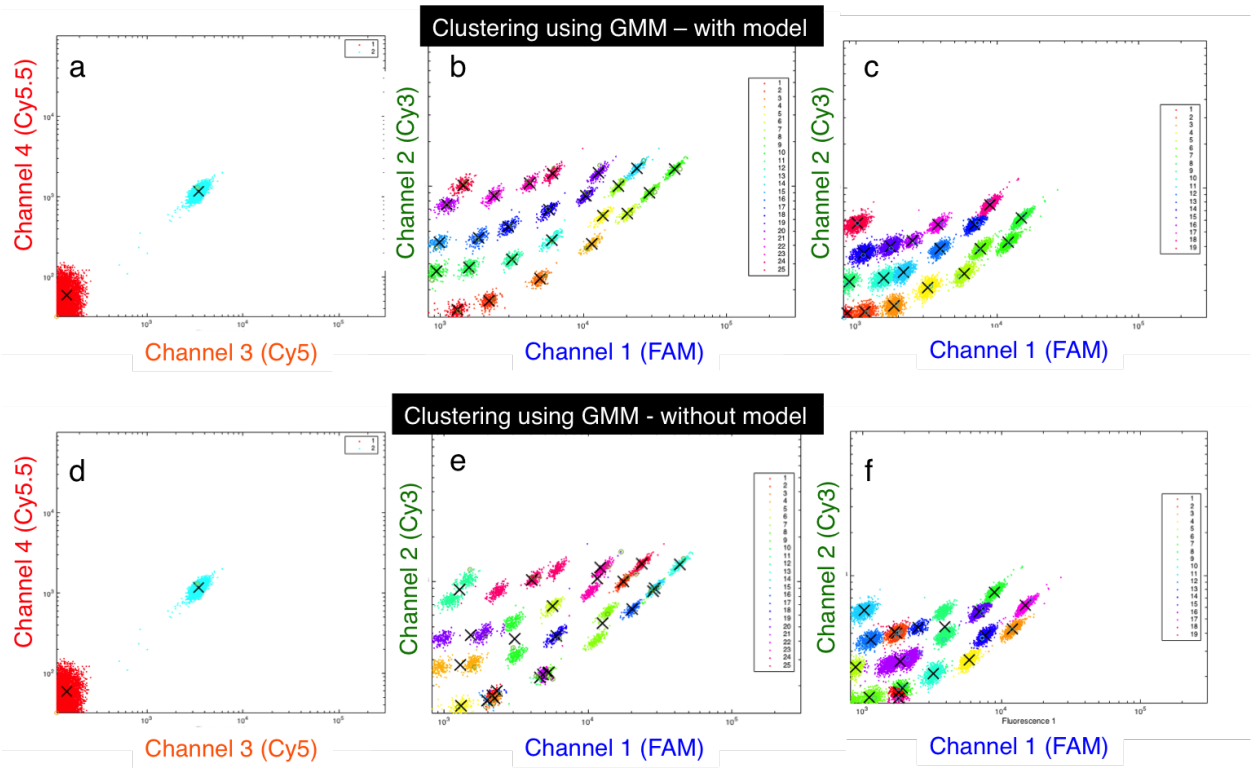


Figure 3.18: **Automated barcode clustering using GMM.** Clustering of data in the I1-I2 space with and without using the model. 45 barcodes were clustered simultaneously. Whereas all the barcodes were clustered correctly using the model output as GMM's initial means (b-c), without the model, the GMM algorithm (with 5000 iterations and a log-likelihood threshold of 10-7) did not detect the clusters of over 10 barcodes and 6 other barcodes were sub-divided into multiple clusters (e-f).

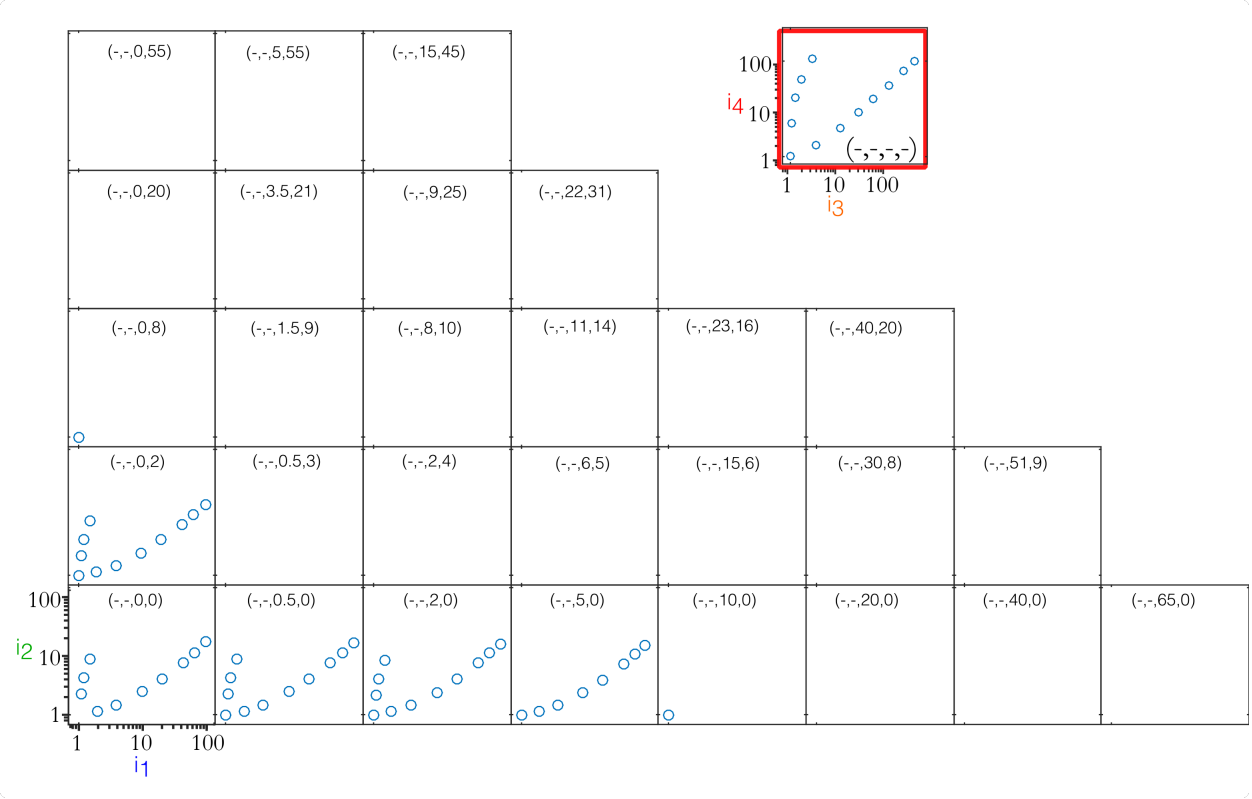


Figure 3.19: **Barcode dataset with restriction to 10% ensemble FRET transfer efficiency for any given donor.** Subset (67 barcodes) of the designed barcodes (presented in Fig. 4a) with $E_d^T < 10\%$ for any given donor d. Similarly, 124 barcodes possessed ETd <20%.

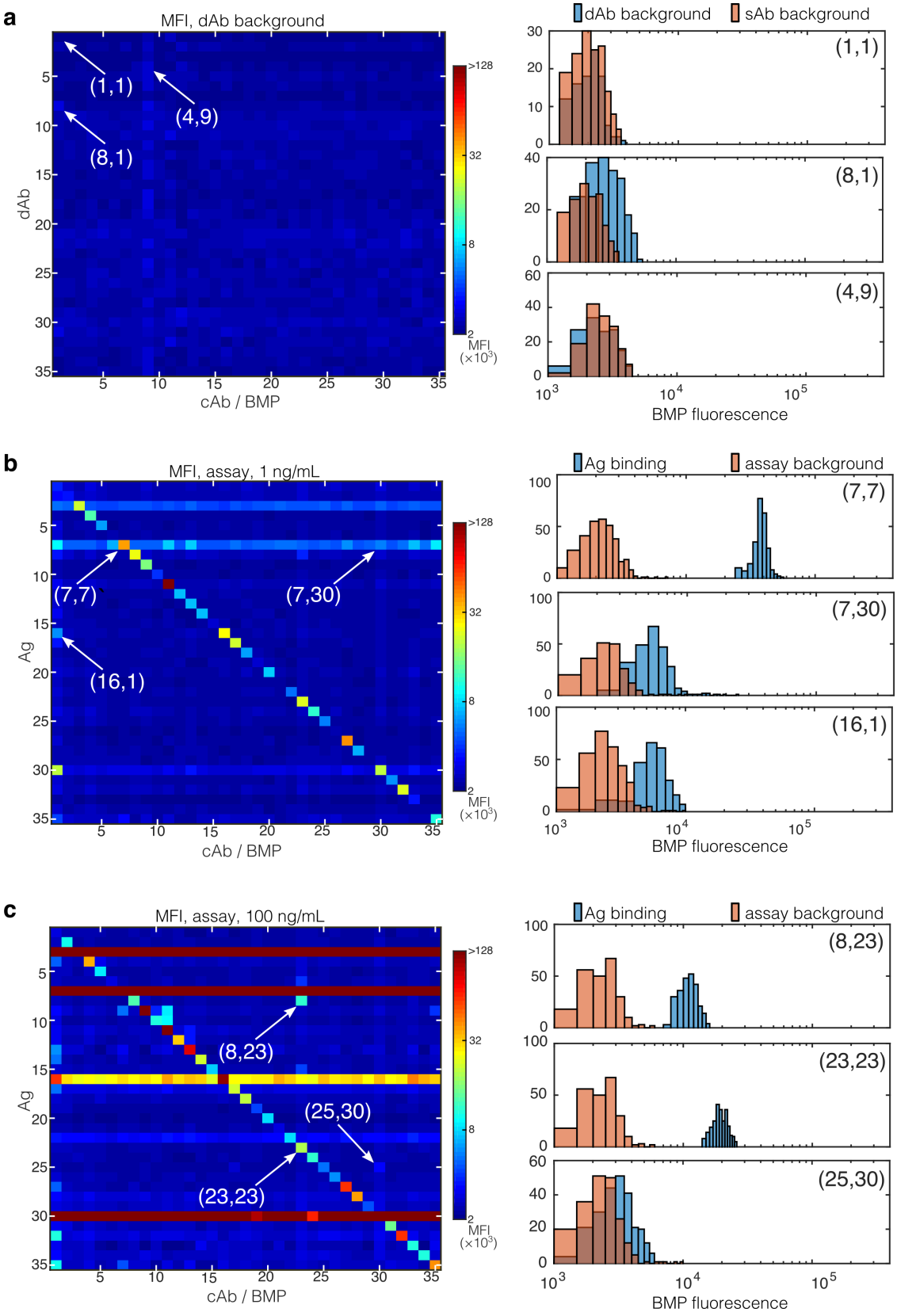


Figure 3.20: (Next page)

Figure 3.20: **Mean Fluorescence Intensity (MFI) of 35-plex specific and non-specific binding.** MFI signals and bead-histograms underlying the signal-to-noise results reported in Figure 6 of the main text. (a) BMP-specific (column) MFI upon addition of a dAb (row), followed by sAb labelling. Examples of the frequency counts for (dAb,cAb) shown on the right. cAb 9 (EpCAM) displayed some reactivity to the sAb as can be seen by the increase in the sAb background, which was also noticeable for this BMP in the assay background (Figure 6f of the main text). Signals above sAb background denote dAb-cAb non-specific binding as is seen in (8,1). (b,c) BMP-specific (column) MFI after incubation with Ag (row), denoting specific (diagonal) and non-specific (off-diagonal) antigen binding. Antigens were incubated separately at (b) 1 ng/mL and (c) 100 ng/mL. Examples of frequency counts for specific and non-specific interactions are shown on the right.

PREFACE TO CHAPTER IV

This chapter introduces the "colocalization-by-linkage assay on microparticle" (CLAMP), a multiplexed sandwich immunoassay format that inherently eliminates reagent-driven cross-reactivity (rCR) by pre-colocalizing antibody pairs on microparticles. The concept and operative principle of the CLAMP are described in detail as well as experimentally demonstrated.

This chapter is in preparation for submission.

Acknowledgements

We thank Jeffrey Munzar for helpful discussions. We thank NSERC and FQRNT for funding. The flow cytometry work was performed at two McGill core flow facilities, namely the Microbiology and Immunology (MIMM) department, and the Life Science Complex, which is supported by funding from the Canadian Foundation for Innovation.

Chapter 4

CROSS-REACTIVITY-FREE MULTIPLEXED IMMUNOASSAYS ON MICROPARTICLES

Milad Dagher[1,2], Andy Ng[1,2], David Juncker[1,2,3]

[1] Biomedical Engineering Department, McGill University, 3775 University Street, Montreal, QC, H3A2B4, Canada

[2] McGill University and Genome Quebec Innovation Center, 740 Dr. Penfield Avenue, Montreal, QC, H3A0G1, Canada

[3] Neurology and Neurosurgery Department, McGill University, 3801 University Street, Montreal, QC, H3A2B4, Canada

4.1 Abstract

Multiplexed sandwich immunoassays (MSAs) are essential for rapid, quantitative measurement of low abundance proteins; in particular, microparticle-based MSAs are widely used because of their precision and high sample-throughput capacity at affordable costs. However, multiplexing levels on microparticle-based MSAs have not scaled during the last 20 years due to reagent-driven cross-reactivity (rCR) that arises when non-matched antibodies are mixed together during the assay. Indeed, methods that mitigate rCR have been developed for planar and solution-phase platforms but could not be applied to microparticle-based MSAs, which continue to be limited by rCR. Here, we introduce a colocalization-by-linkage assay on microparticles (CLAMP) that overcomes rCR by pre-assembling matched antibody pairs on barcoded microparticles, effectively eliminating the mixing of non-matched antibodies. To assemble and colocalize matched antibody pairs, detection antibodies are conjugated to oligonucleotides (dAb-oligos) and tethered to the surface of capture antibody-coated microparticles via reversible oligonucleotide linkages. After target capture and sandwich binding, a fluorescent oligonucleotide strand displaces the linkage and simultaneously labels the dAb-oligos to detect the sandwich-bound fraction. The single-plex CLAMP was optimized by tuning the valency and the surface density of the antibody-oligonucleotide conjugates. A 6-plex CLAMP was assembled and was shown to eliminate rCR using antibody pairs with known incompatibility in a conventional MSA. By pre-assembling antibody pairs, the CLAMP represents a departure from typical immunoassays, and holds potential as the first multiplexed ELISA on microparticles.

4.2 Introduction

The sandwich immunoassay (SA) uses a pair of antibodies to recognize two distinct epitopes on a single target protein (**Figure 4.1a**) and has proven invaluable for the detection of proteins since its development 50 years ago [1, 2]. This dual-recognition affords SAs higher sensitivity and specificity compared to single-binder assay formats. Indeed, SAs are impervious against a single cross-reactivity (CR) because a false-positive is only possible in the case of dual-CR (dCR), where both the capture antibody (cAb) and the detection antibody (dAb) bind to an off-target protein, and which is a more stringent condition. However, this enhanced specificity is offset in multiplexed sandwich assays (MSAs) because dAbs against different targets are mixed and incubated together, leading to combinatorial interactions between all immobilized cAbs, proteins, and the dAbs. As a result, a single CR binding between reagents in MSAs can result in false-positives–a situation termed reagent-driven CR (rCR) (**Figure 4.1b**). rCR has long been recognized as a challenge to the specificity and reliability of MSAs [3–6], and is typically minimized using brute-force combinatorial optimization to identify, remove and replace cross-reacting antibodies [7–11]. However, the risk of rCR grows as $4N^2$ with the number of targets N [5], and the labor involved in optimization also grows quadratically while often no alternatives exist to cross-reacting reagents, preventing their joint use. As a result, rCR has limited the multiplexing of conventional MSAs to between 1 and 60 targets, while the majority of commercial panels is limited to around a dozen or so targets only.

In the last decade, several immunoassay platforms were developed to mitigate rCR [6]. For example, our group avoided rCR on planar microarrays by delivering each set of detection antibodies onto their capture spot, which physically separates the assays and eliminates interaction between non-cognate antibodies [5, 12, 13]. For solution-phase MSAs, Landegren's group developed the proximity-based assay that prevents rCR from generating a signal by conjugating antibody pairs to complementary oligonucleotides that provide an added layer of redundancy and act as target-specific signal-transduction elements [14, 15]. Finally, another approach has been to abandon antibodies altogether, and use so-called "slow off-rate modified aptamers", or SOMAmers, to reduce the amount of non-specific binding and simultaneously detect a large number of proteins [16, 17].

Compared to other MSAs platforms, microparticle (MP)-based assays offer distinct benefits. For example, the ability to rapidly read out a large number of MPs by flow cytometry affords increased precision and sample throughput [11, 18, 19]. In addition, MPs may be functionalized in large batches and are stored, used, and readout while in solution, both of which reduce lot-to-lot variability and enables quantitative analysis [20, 21]. Several commercial MP-based platforms have been developed, the most popular being the xMAP technology by Luminex, and are regularly used by academia and pharma in various stages of drug development and biomarker discovery pipelines [4, 10, 22].

Nevertheless, MSAs on MPs are still limited by rCR [6, 11, 23]. Notably, the size of multiplexed panels available on the market has not scaled in the last two decades [24, 25]. While the proximity-based assay has been implemented on MPs [26], the use of a third antibody immobilized on the MPs undercuts the specificity of the approach by rendering it susceptible to rCR between the cAb and off-target analytes in the panel.

Here, we introduce colocalization-by-linkage assay on MPs (CLAMP), a multiplexed assay designed to inherently eliminate rCR by colocalizing and confining each antibody pair onto a set of barcoded MPs, thereby avoiding interaction between non-cognate antibodies (**Figure 4.1c**). Oligonucleotides (oligos) are used as programmable building blocks to implement the key molecular 'operations' of CLAMP, including (i) flexible linkage of dAbs, (ii) on-demand release of dAbs, (iii) transduction of assay signals, as well as (iv) fluorescent barcoding of MPs. In this report, we detail the conceptual operation and experimental validation and optimization of the CLAMP assay and showcase its efficacy to inherently eliminate rCR using reagents that otherwise strongly cross-react in a conventional MSA.

4.3 Results

4.3.1 Assay architecture

The architecture and operative principle of the CLAMP are illustrated in panels **d** and **e** of **Figure 4.1**, respectively. To colocalize each pair of antibodies, an 82-nt hook oligo (HO) is covalently bound to the detection antibody (dAb) and is partially hybridized via a 21-bp hybrid to a capture oligo (CO) bound to the surface of the capture antibody (cAb)-coated MP. Whereas the cAb is immobile on the surface, the dAb is flexible due to the HO's 61-nt single-stranded domain; this flexibility is necessary to form a tertiary complex with the analyte (**Figure 4.1e**). The confinement of antibody pairs precludes interaction between non-matched antibodies and restores the single-plex assay configuration on every MP, ensuring that single cross-reacting events (*e.g.* a target analyte reacting to a non-cognate cAb) do not lead to sandwich binding. *A priori* colocalization of antibodies allows for rapid dual-recognition of proteins but necessitates a concomitant method for signal transduction and generation. One approach would be to first break the HO-CO linkage-for example, via photo-induced or enzymatic cleavage [27, 28], or toe-hold mediated displacement [29], then, after washing of the released dAb-HO complexes, label the dAbs remaining on the surface to signal sandwich formation. However, unbroken CO-HO linkages would result in labeling of the corresponding dAbs irrespective of the presence of the target analyte, which consequently increases the background signal. For example, 2% dAb coverage on a 3 μm MP corresponds to 1000-5000 dAbs that, if labeled, would result in a large increase in background signal and significantly impede sensitive detection [30–32]. To mitigate this effect in CLAMP assays, we designed a detection scheme to exclusively label 'successfully' released conjugates through the

use of a fluorescently-labeled displacer oligo (DO) that binds to a toe-hold domain on HO, displacing and labeling it simultaneously (**Figure 4.1e**, **elementary Figure 4.4**). Importantly, this 'detection-by-displacement' operates as an AND logical gate, requiring both protein dual-capture and dAb release for a detectable signal (**Figure 4.1f**). CLAMP reagents are assembled on magnetic MPs in two steps, benefitting from the affinity of biotin-streptavidin bond and Watson-Crick base pairing (Methods, **Supplementary Figure 4.5**). In a first step, a mixture of biotinylated oligos and antibodies are co-immobilized to the surface of streptavidin-coated MPs. The one-pot nature of the labeling affords accurate control over the CO surface density (**Supplementary Figure 4.6**), and simultaneously allows MP-encoding via one-pot labeling with multicolour classifier dyes as described elsewhere [23]. In a second step, dAb-HO complexes are pulled-down via HO-CO hybridization to complete the assembly of CLAMPs.

4.3.2 Assay optimization

We first optimized the efficiency of the toe-hold mediated displacement reaction by displacing unconjugated, Cy5-labeled HOs (Supplementary Figure S4). HOs were pulled down on MPs with different CO densities, then released using unlabeled DOs. Increased ionic strengths in the displacement buffer (M_{NaCl} > 500 mM) were helpful for screening the negatively charged oligos and improved the efficacy of DO hybridization to, and release of, the HOs. 98% displacement was reached over a wide range of CO densities with increased ionic strengths and DO concentrations ($M_{NaCl} \sim$ 500 mM and M_{DO} = 1 µM, respectively).

Next, we studied the impact of antibody-oligonucleotide conjugates on assay background by measuring the residual signal on the MPs following a labeled-displacement step in buffer (see Methods). We first conjugated HOs to immunoglobulin-G (IgGs) using a commercial kit (Solulink) leading to $\sim$ 90% antibody conjugation yield and an average of 2 HOs per IgG (*i.e.*$\lambda \sim 2$) (**Supplementary Figure 4.8**). Using these conjugates, the assay background was an order of magnitude greater than the assay background of unconjugated HOs (**Figure 4.2a**). The increase in background signal is due to multivalent HO conjugates, which can result in unreleased dAb-HOs complexes (due to an unbroken HO-CO linkage) that are labeled by hybridization of a DO to at least one of the other HO strands, thereby generating a fluorescent signal in the absence of a sandwich binding with a protein (**Figure 4.2b**). An effective way to minimize multivalent dAb-HO conjugates is to reduce the average conjugation valency; for example, by aiming for λ of 0.1, Poisson statistics indicates that < 5% dAb would be bound to multiple HOs (**Supplementary Figure 4.6**). The trade-off of such a low conjugation valency is a decreased antibody conjugation yield (10%), which leaves 90% of antibodies unreacted. To avoid wasting unreacted antibodies, we developed a conjugation and purification workflow that maintains the native state of unconjugated antibodies and allows their recycling (Methods, **Supplementary Figure 4.10**). The relative concentration of dAb and

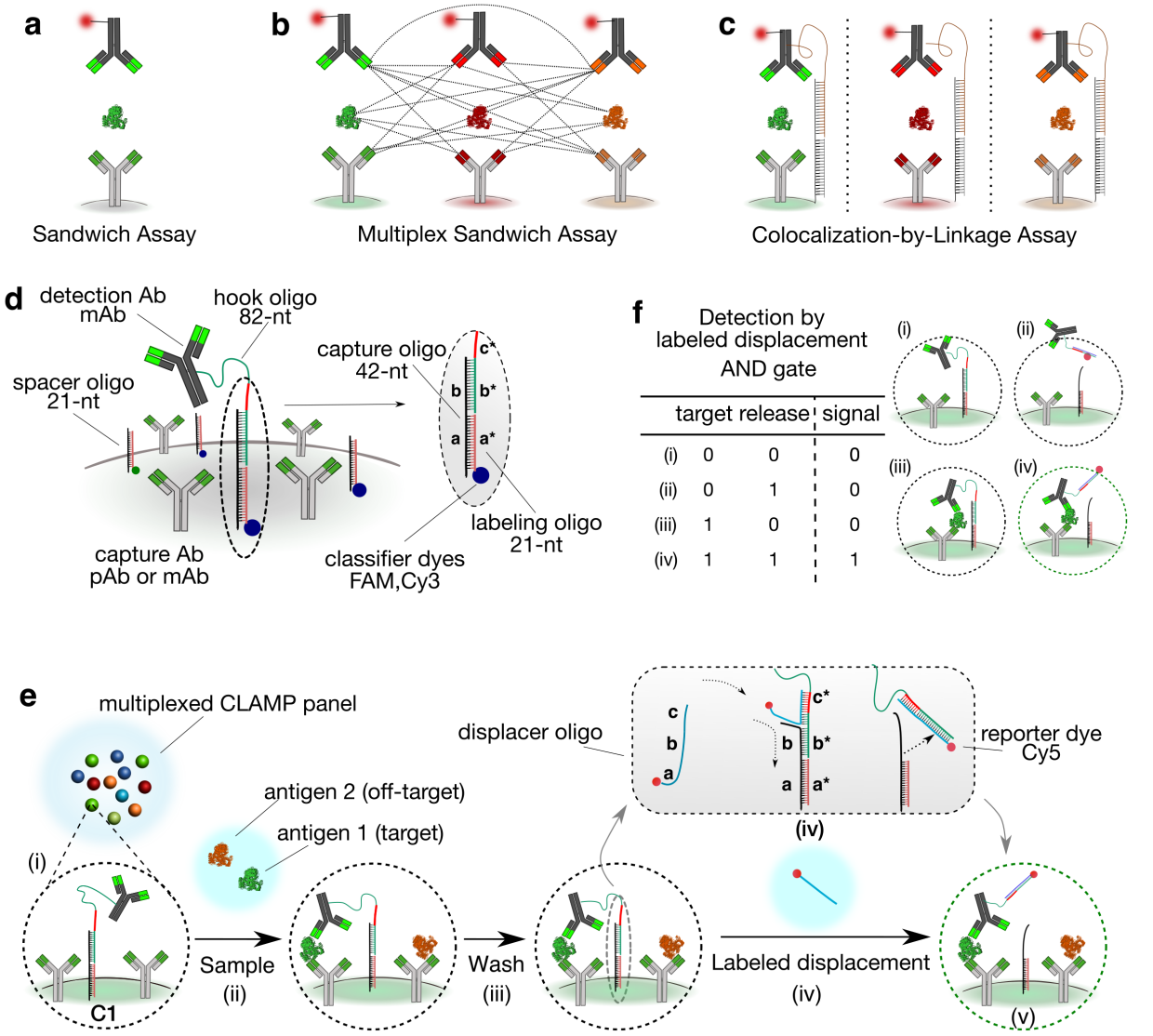


Figure 4.1: **Single-plex and multiplex sandwich assays, and CLAMP.** (a) Single-plex sandwich immunoassay (also known as ELISA), comprising a pair of matched antibodies. (b) MSA with mixing of antibodies is exposed to a large number of interactions between non-matched antibodies and proteins, often resulting in rCR and false-positives. (c) CLAMP with pre-colocalization of antibody pairs using DNA linkage permitting sandwich binding while eliminating interaction between non-cognate antibodies. (d) The dAb is bound to a hook oligo (HO) that is tethered to the surface via partial hybridization with a capture oligo (CO) strand. A spacer oligo (SO) is used to control the density of the COs and dAb-HOs on the surface (**Supplementary Figure 4.5**). (e) A multiplexed CLAMP assay is carried out by (i) mixing barcoded CLAMPs against different targets, (ii) incubating the biological sample with CLAMPs generating sandwich binding in the presence of the target analyte only, (iii) washing, and (iv) displacing and labeling HOs using a fluorescently-labeled displacer oligo (DO) via toe-hold mediated displacement (inset), leading to (v) labeling of the sandwich complexes that remain on surface. (f) An AND (Boolean) logic gate representation of the detection by labeled-displacement step, where detection at the single-molecule level requires both the capture of target and successful HO release.

HO were adjusted and was modulated from 1.25 to 0.1 (**Figure 4.2c**). The dAb-HO conjugates of varying valency were pulled down on MPs with varying CO densities. As expected, lower valency significantly decreased residual assay background, matching the background signal exhibited by unconjugated HOs for $0.1<\lambda<0.2$ (**Figure 4.2d**), leading to low valency conjugates with fewer than 8% of multivalent conjugates. Consistent with a multivalent scenario, increasing CO density amplified the high background signals for higher valency dAb-HOs.

To optimize assay performance, we modulated the dAb-HO density. In CLAMP, adequate local dAb concentrations are key for sensitive and high capacity sandwich binding which, for a set HO length, is chiefly dependent on the surface densities of dAb-HOs and, through hybridization capture, COs. CLAMPs against urokinase plasminogen activator (anti-uPA CLAMP) with varying CO densities were prepared using low valency dAb-HO conjugates with fewer than 8% multivalent conjugates (**Figure 4.2d**, see Methods). anti-uPA CLAMPs were incubated with a serial dilution of recombinant uPA antigen, followed by washing and detection by labeled-displacement. As expected, increasing CO densities modulated the signal-to-noise ratio (SNR) of the assay, revealing that a density greater than 13×10^{14} m^{-2} is necessary for adequate SNRs (**Figure 4.2e**). Densities greater than 39×10^{14} m^{-2}, on the other hand, provided little improvement in SNR as they also resulted in increased background signal. Lastly, to assess the importance of conjugate valency on assay performance, we compared anti-uPA of high valency ($\lambda \sim 2$, Solulink) against low valency conjugates ($\lambda = 0.1$, in-house conjugation protocol, **Figure 4.2f**). The lower valency conjugates resulted in significantly lower background signals (10-folds) and, correspondingly, a 3-fold improvement in detection limits (**Figure 4.2f**). On the other hand, low valency conjugates exhibited a decreased dynamic range of the fluorescence as sandwich-bound dAb-HO complexes are predominantly labeled with a single dye. Taken together, these results highlight the importance of conjugation valency on background signals and assay performance in general.

4.3.3 Multiplexed CLAMP

Finally, to test CLAMP's efficacy in eliminating rCR, we sought to screen the assay specificity of multiplexed CLAMP. In addition, to challenge the CLAMP assay we selected antibody pairs that have been shown to exhibit different types of rCR when used together in a conventional MSA. To this end, antibody pairs against six targets (EpCAM, PSA, E-Cadherin, EGF, uPA and MCP) were short-listed from a 35-protein panel that we previously characterized for specific and non-specific binding in a conventional MSA [23]. For conventional MSAs, the specificity screen consisted of incubating each individual antigen with a pool of cAb-coated barcoded MPs, followed by addition of mixed dAb cocktail and secondary-antibody (sAb) for detection and labeling, respectively (**Figure 4.3a**). Measuring the fluorescence across the different barcodes in response to an antigen concentration of 1 and 100 ng/mL (**Figure 4.3b-c**) uncovered two types of non-specific binding that generate

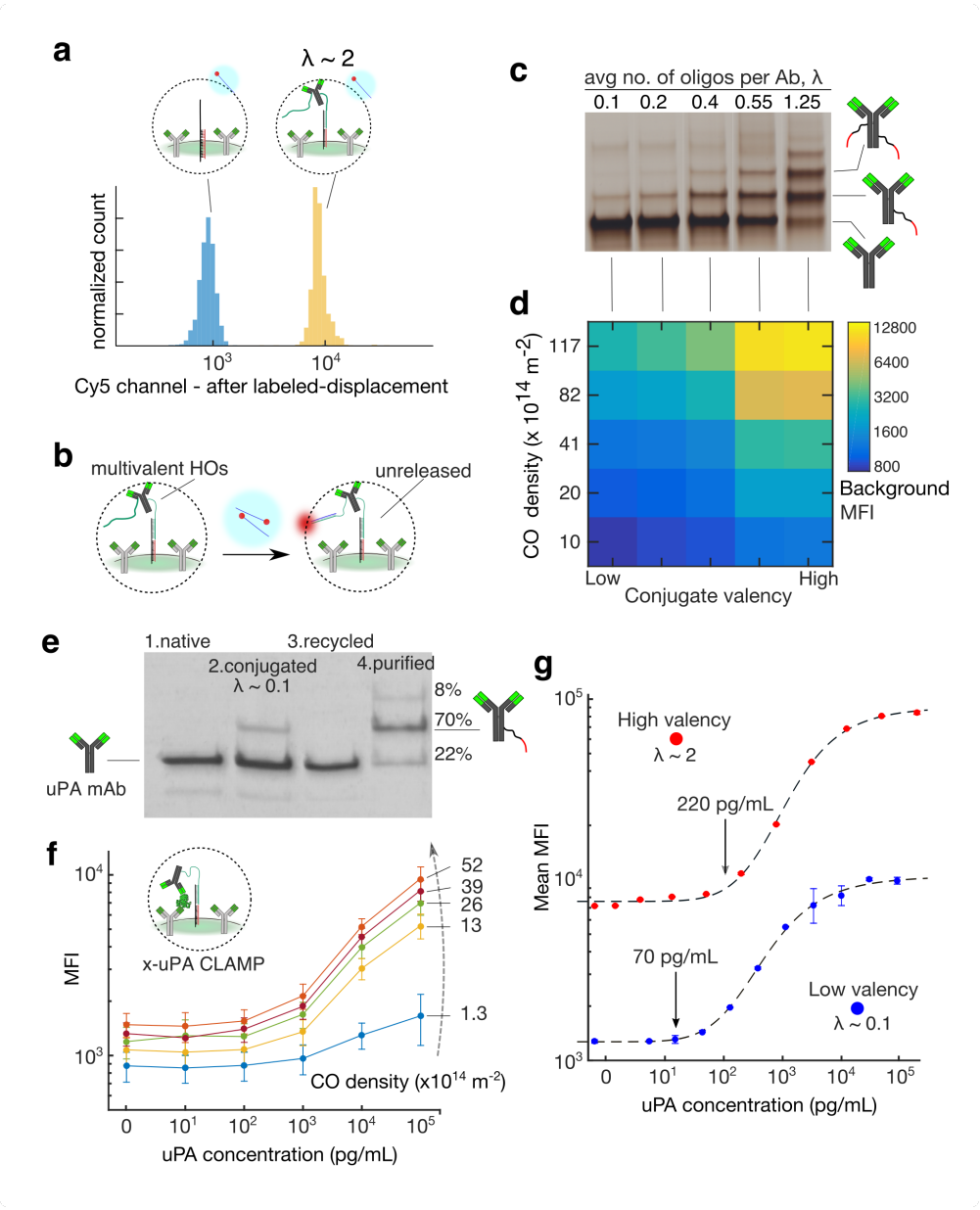


Figure 4.2: **CLAMP optimization by modulating conjugate valency and surface density.** (a) Normalized histograms compare the CLAMP background signal (*i.e.* residual signal after incubation with Cy5-labeled DO) for MPs without HOs (in blue) and with multivalent dAb-HOs conjugates (in yellow) (**Supplementary Figure 4.8**). (b) Illustration depicting how multivalent ($\lambda \sim 2$) dAb-HO conjugates may increase background signal by labeling unsuccessfully displaced dAb-HO complexes despite the absence of sandwich binding with the target analyte. (c) SDS-PAGE of mouse anti-goat IgGs conjugated with HOs with increasing valencies and stained by silver amplification. (d) Assay background MFI plotted with respect to increasing conjugate valency (columns) and increasing CO density (rows). (e) SDS-PAGE of low valency dAb-HO (mouse uPA mAbs) conjugates at different stages of the purification protocol where (1) native dAb, (2) crude conjugation product dAb-HO, (3) retrieved (non-conjugated) dAb, and (4) purified dAb-HO (Methods, **Supplementary Figure 4.10**). (Continued on next page.)

Figure 4.2: (f) MFI assay values for anti-uPA CLAMP assays against standard dilutions of uPA antigen and for varying CO densities. Error bars are standard-deviation of the microparticle signals in Cy5 channel. (g) Mean MFI signals in anti-uPA CLAMP assays using low (blue dots) and high (red dots) valency conjugates. Error bars are standard deviations of MFI signals across wells (n=3). The LODs shown on each curve were calculated as discussed in Methods.

false-positives, namely indiscriminate sticking of antigens (observed for E-Cadherin and uPA) and cross-reactivity between antigens and antibodies. On the other hand, the specificity screen for CLAMP assays was performed by incubating a single antigen at-a-time with multiplexed CLAMPs and running the detection by labeled-displacement (Methods, **Figure 4.3d-f**). All but one of the non-specific signals detected in conventional MSAs were completely eliminated using CLAMP assays. For example, the pervasive, non-specific binding of E-Cadherin, which led to a signal on all off-target beads in conventional MSAs, was not detectable in CLAMP assays. In contrast, cross-reactivity was detectable between MCP-1 antibodies and EGF antigen at 100 ng/mL both in conventional MSA as well as CLAMP. To investigate the source of this false-positive signal, we performed single-plex assays using MCP-1 antibodies only, separately spiking MCP-1 or EGF at 1 or 100 ng/mL (**Figure 4.3g**). The detection of EGF by MCP-1 antibodies in single-plex indicated a dCR. Indeed, this dCR cannot be mitigated by CLAMP nor ELISA, and is indication of poor affinity binders. Overall, this result showcases the strength of CLAMP in eliminating rCR in a multiplexed assay, as well as identifying dCR in multiplexed, combinatorial fashion. Finally, dilution curves of the remaining 5 proteins were generated and their SNRs were plotted as shown in **Figure 4.3h**.

4.4 Conclusion

In summary, we have introduced CLAMP, a homogeneous MSA that uses oligonucleotides to pre-colocalize antibody pairs on MPs. By confining each antibody pair to their respective MPs during sample incubation, CLAMP can be multiplexed while maintaining single-plex assay environments on each MP and, in doing so, eliminates reagent-driven CR. Notably, the pre-colocalization of antibodies in CLAMP represents a departure from the assay workflow in our previous implementation of antibody colocalization [5, 12], and indeed in sandwich immunoassays in general, where matched antibodies are separate at the beginning of the assay. To detect correct sandwich binding, we have shown that a labeled displacer oligo can be used to simultaneously release and label dAb-oligo complexes. We studied and demonstrated the importance of using monovalent antibody-oligo conjugates to avoid labeling unreleased complexes and increasing background signals. We have experimentally validated the assay, both in single-plex and multiplex, and screened the specificity of the assay in multiplex using five antibody pairs pre-selected for CR, demonstrating that CLAMP eliminates all rCR experienced in a conventional MSA.

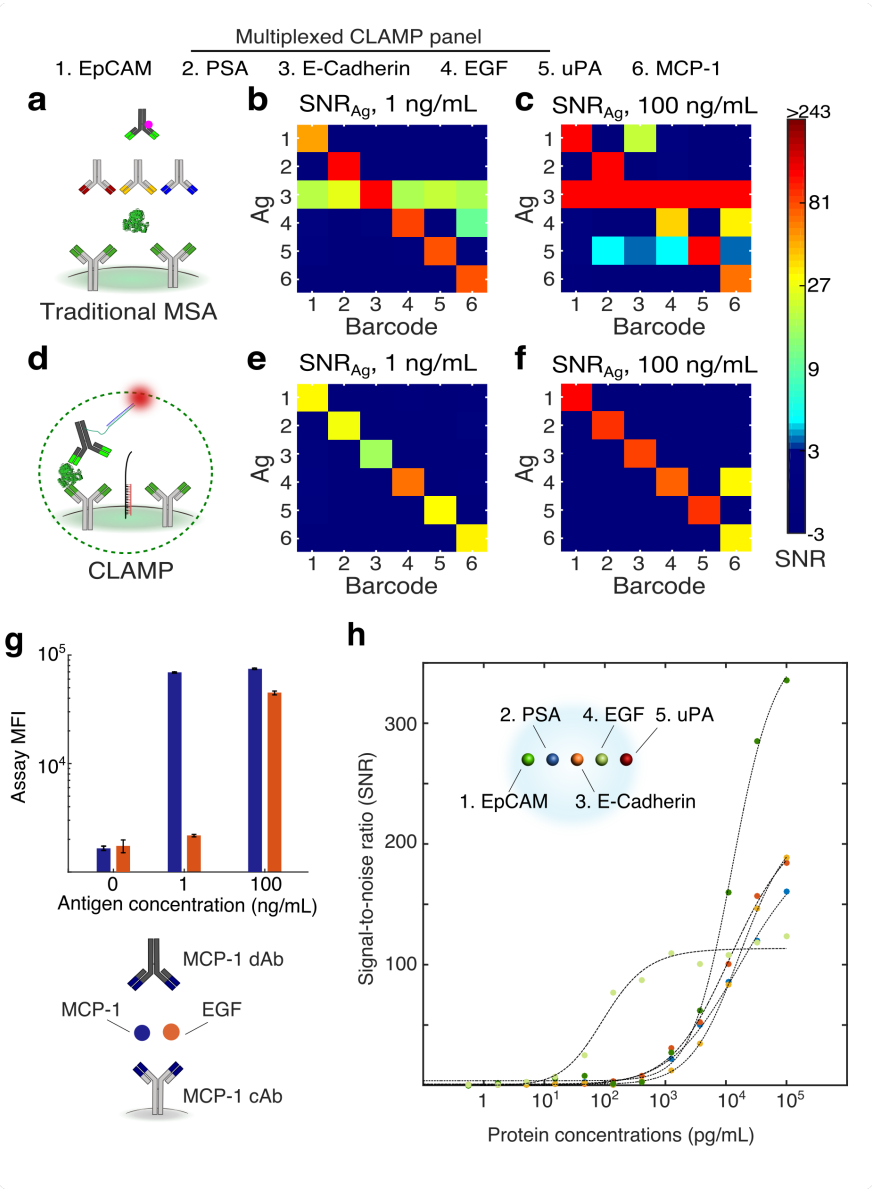


Figure 4.3: **CLAMP optimization by modulating conjugate valency and surface density.** (a) Normalized histograms compare the CLAMP background signal (*i.e.* residual signal after incubation with Cy5-labeled DO) for MPs without HOs (in blue) and with multivalent dAb-HOs conjugates (in yellow) (**Supplementary Figure 4.8**). (b) Illustration depicting how multivalent (λ 2) dAb-HO conjugates may increase background signal by labeling unsuccessfully displaced dAb-HO complexes despite the absence of sandwich binding with the target analyte. (c) SDS-PAGE of mouse anti-goat IgGs conjugated with HOs with increasing valencies and stained by silver amplification. (d) Assay background MFI plotted with respect to increasing conjugate valency (columns) and increasing CO density (rows). (e) SDS-PAGE of low valency dAb-HO (mouse uPA mAbs) conjugates at different stages of the purification protocol where (1) native dAb, (2) crude conjugation product dAb-HO, (3) retrieved (non-conjugated) dAb, and (4) purified dAb-HO (Methods, **Supplementary Figure 4.10**). (Continued on next page.)

Figure 4.3: (f) MFI assay values for anti-uPA CLAMP assays against standard dilutions of uPA antigen and for varying CO densities. Error bars are standard-deviation of the microparticle signals in Cy5 channel. (g) Mean MFI signals in anti-uPA CLAMP assays using low (blue dots) and high (red dots) valency conjugates. Error bars are standard deviations of MFI signals across wells (n=3). The LODs shown on each curve were calculated as discussed in Methods.

CLAMP provides several distinct advantages over currently available MSAs. First, CLAMP is easily deployable as it does not necessitate dedicated equipment for readout or introduce new workflows. Second, CLAMP is a rapid assay as it can be completed in little over three hours. Finally, by eliminating the need to incubate detection antibodies in solution (which is typically done at high concentrations), CLAMP provides significant reductions in reagent consumption. Owing to its highly scalable and highly efficient nature, we believe CLAMP has the potential to evolve toward a truly-scalable multiplexed ELISA platform that meets the increasing demands in biomarker discovery and drug development.

4.5 Methods

Materials and Reagents. HPLC-purified oligonucleotides were purchased from IDT (Coralville, IA, USA); the sequences and modifications are shown in Supplementary Table S1. cAbs, antigens, and dAbs were purchased from RnD Systems (Minneapolis, MN, USA), and stored at -20 °C for up to 36 months. Streptavidin- and Protein-G magnetic MPs (M270) were purchased from Life Technologies (Carlsbad, CA, USA).

Synthesis of CLAMPs. CLAMPs were assembled on streptavidin-coated magnetic MPs with a 2.7 μm diameter (M270-Streptavidin) in two steps. The first step consisted of the immobilization of a biotinylated mixture of antibodies and oligos to functionalize the MPs and simultaneously encode them as described in detail elsewhere [23]. Briefly, 90 pmols biotinylated oligos (COs, and SOs) and a total of 90 pmols of LOs (LO0-LO2) were mixed together in 25 μL of PBS + 0.05% Tween20 + 300 mM NaCl (PBST0.05 + NaCl300). Whereas the proportions of each LO0:LO1:LO2 is designed to generate a unique ensemble fluorescence to define the barcode, the proportion of CO:SO allows tuning of the surface density of pulled dAb-HOs. The mixture is annealed by heating to 80 °C and cooling back to room temperature by removing the mixture from the heat source. Next, 5 μg biotinylated cAb in 17 μL of PBST0.05 + NaCl300 were added to and mixed with the annealed oligonucleotide mixture. The biotinylated reagents are thereafter co-immobilized on the MPs in a single step by adding 3.25M MPs in 10 μL PBST0.05 + NaCl300 and immediately mixing by pipetting. The mixture was incubated for 90 min with end-over-end mixing at room temperature, followed by 3x washing by magnetic aggregation in 150 μL PBST0.1. The barcoded and functionalized MPs were stored at 4 °C until needed. In a second step, 100,000 of the

prepared MPs were mixed with the HO-containing solution (*e.g.* dAb-HOs) diluted in PBST0.05 + NaCl300 for 30 minutes. After pull-down of HOs, the fully-assembled CLAMPs were washed 3x in PBST0.01, and were stored until the time of the assay for up to a week at 4 °C.

Characterization of CLAMPs. To characterize CLAMPs, the immobilization of antibodies and oligos was confirmed by labeling using an anti-goat IgG conjugated with Alexa-Fluor 647 (AF647), or hybridization of a Cy5-labeled oligo (LO) targeting the HOs. The density of COs was estimated by fitting the ensemble fluorescence response of multicolour MPs using a multicolour fluorescence model, as described elsewhere [23]. To determine the expected assay background signal for a particular set of CLAMPs, the MPs were incubated with 1 μM Cy5-labeled DOs in PBST0.05 + NaCl300 for one hour, followed by 3x magnetic washing in PBST0.05, and the residual signal was determined by cytometry.

Antibody oligo conjugation, purification, and characterization. Anti-uPA monoclonal antibodies were conjugated to amine-modified HOs using a hydrazone chemistry (Solulink) followed by purification according to the manufacturer's protocol. Alternatively, monoclonal antibodies were conjugated to thiol-terminated HOs using a heterobifunctional amine/thiol-reactive cross-linker. 40 μL of 30 μM thiol-modified HOs are first reduced in 200 mM DTT in PBST at 37 °C for one hour. The reduced oligos were (i) buffer exchanged into PBS pH 7.0 using a Zeba desalting spin-colum (7K MWCO, Thermo), (ii) activated for 10 min using 8 µL of 9 mM sulfo-succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC) dissolved in 80% PBS pH7.0 and 20% anhydrous dimethyl solfoxide, (iii) buffer exchanged again into PBS pH 7.0 to remove excess sulfo-SMCC, and (iv) a 1-10 μL fraction (depending on the desired) reacted with 10 μL of 1 mg/mL antibodies. The reaction was left at room temperature for 1 hr and incubated overnight at 4 °C thereafter. The conjugates were purified in two purification steps. Firstly, the HOs and dAb-HOs were pulled-down on HO purification MPs coated with capture oligo no. 2 (CO2) permitting recovery of the unreacted antibodies from the supernatant, and displaced using displacer oligo no. 2 (DO2, Supplementary Table S1, **Supplementary Figure 4.10**). Secondly, the dAb-HOs were recovered from the displaced eluate via affinity-purification using Protein-G coated MPs according to the manufacturer's protocol. For characterization by SDS-PAGE, samples were heated to 70 C for 10 min in LDS-containing sample buffer (NuPAGE™, Novex) and ran on a 3-8% NuPAGE Tris-acetate gel electrophoresis in 1X NuPAGE™ Tris-Acetate SDS Running Buffer at 150 V for 60 min. The gel was stained by silver amplification, simultaneously staining proteins and nucleic acids. The average conjugation valency (λ) of the conjugates was determined by quantifying the proportions of different valencies using image analysis (**Supplementary Figure 4.9**). Antibody oligo conjugation, purification, and characterization. Anti-uPA monoclonal anti-

bodies were conjugated to amine-modified HOs using a hydrazone chemistry (Solulink) followed by purification according to the manufacturer's protocol. Alternatively, monoclonal antibodies were conjugated to thiol-terminated HOs using a heterobifunctional amine/thiol-reactive cross-linker. 40 μL of 30 μM thiol-modified HOs are first reduced in 200 mM DTT in PBST at 37 °C for one hour. The reduced oligos were (i) buffer exchanged into PBS pH 7.0 using a Zeba desalting spin-colum (7K MWCO, Thermo), (ii) activated for 10 min using 8 µL of 9 mM sulfo-succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC) dissolved in 80% PBS pH7.0 and 20% anhydrous dimethyl solfoxide, (iii) buffer exchanged again into PBS pH 7.0 to remove excess sulfo-SMCC, and (iv) a 1-10 μL fraction (depending on the desired λ) reacted with 10 μL of 1 mg/mL antibodies. The reaction was left at room temperature for 1 hr and incubated overnight at 4 °C thereafter. The conjugates were purified in two purification steps. Firstly, the HOs and dAb-HOs were pulled-down on HO purification MPs coated with capture oligo no. 2 (CO2) permitting recovery of the unreacted antibodies from the supernatant, and displaced using displacer oligo no. 2 (DO2, Supplementary Table S1, **Supplementary Figure 4.10**). Secondly, the dAb-HOs were recovered from the displaced eluate via affinity-purification using Protein-G coated MPs according to the manufacturer's protocol. For characterization by SDS-PAGE, samples were heated to 70 °C for 10 min in LDS-containing sample buffer (NuPAGE™, Novex) and ran on a 3-8% NuPAGE Tris-acetate gel electrophoresis in 1X NuPAGE™ Tris-Acetate SDS Running Buffer at 150 V for 60 min. The gel was stained by silver amplification, simultaneously staining proteins and nucleic acids. The average conjugation valency (λ) of the conjugates was determined by quantifying the proportions of different valencies using image analysis (**Supplementary Figure 4.9**).

Single-plex and multiplex CLAMP assay. Incubations were performed in a conical bottom 96-well plate at room temperature with horizontal shaking at 950 rpm. CLAMPs were mixed at roughly 80 MPs per barcode per µL and blocked with PBST0.05 + NaCl150 + 0.5% BSA (PBST0.05 + NaCl150 + BSA0.5) for 30 min. A 25 μL aliquot of the blocked, multiplexed CLAMP mixture was added into each well and incubated with 25 μL containing the specificd antigen(s) at 2x the specified concentrations in PBST0.05 + NaCl150 + BSA0.25, the incubation was performed for 3 hr at 950 rpm shaking. Magnetic aggregation and washing with 150 μL of PBST0.1 was repeated 4x in over a total of 30 min. Finally, detection-by-displacement is performed through the addition of 1 μM DO-Cy5 in PBST0.05 + NaCl300 + BSA0.25 and incubation for 1 hr with shaking, followed by 3x washing in PBST0.1.

Conventional MSA. To screen the specificity and non-specific binding in conventional MSA-format, MPs were barcoded and coupled with their respective biotinylated cAbs during synthesis

as described above. MP mixtures were combined to a final concentration of 2,000 MPs per barcode per assay. Incubations were performed in a conical bottom 96-well plate at room temperature with horizontal shaking at 950 rpm. Prior to incubation with assay reagents, MPs were first blocked for one hour with 1% bovine serum albumin in 0.05% Tween-20 in PBS (PBST0.05). Incubation with antigens was conducted for 120 min at the specified concentrations. BMPs were incubated with the dAb cocktail for 60 min at 2 μg/mL, followed by incubation with sAbs for 45 mins at 4 μg/mL. SNRAg is calculated by subtracting the cAb-specific mean assay background (n = 6) from the MFI signals and normalizing to the global standard-deviation (i.e. across all barcodes, n = 210) of the assay background.

Read-out and data analysis. MPs were read out using the FACS CANTO II cytometer by BD with blue (488 nm), red (633 nm), and violet (405 nm) lasers. In blue-laser flow cell, 530/30 and 585/42 band-pass filters were used for FAM and Cy3, respectively. In the red-laser flow cell, 660/20 band-pass filter was used for Cy5/AF647, respectively. The MPs were decoded using an automated algorithm implemented on MATLAB [23]. All data analysis was performed in MATLAB. Single-beads were distinguished from bead aggregates and other particulates by using forward and side-scatter intensities and gating was automated.

Table 4.1: Oligonucleotide design

MP	Strand	Modification	Sequence
CLAMP	Capture Oligo	Biotin	5'-Biotin-TTTTTTTTTGTGGCGGCGGTGATTGGTTATTGAGAGTTTATG
	Spacer Oligo	Biotin	5'-Biotin- TTTTTTTTTGTGGCGGCGGTG
	Encoder Oligo		5'-CACCGCCGCCACAAAAAAAAA
	Encoder Oligo (FAM)	FAM	5'-CACCGCCGCCACAAAAAAAAA-FAM
	Encoder Oligo (Cy3)	Cy3	5'-CACCGCCGCCACAAAAAAAAA-Cy3
	Hook Oligo	Thiol	5'-Thiol-TTTTTTACTTTTCAACCACCACTCAACCATATTCAACTCATTCGCCATAAACTCATTCGCCATAAACTCTCAATAACCAAT
	Hook Oligo (Cy5)	Cy5	5'-Cy5-TTTTTTACTTTTCAACCACCACTCAACCATATTCAACTCATTCGCCATAAACTCATTCGCCATAAACTCTCAATAACCAAT
	Displacer Oligo (Cy5)		5'-ATTGGTTATTGAGAGTTTATGGCGAATGAG-Cy5
	Displacer Oligo		5'-ATTGGTTATTGAGAGTTTATGGCGAATGAG
	Labeler Oligo		5'-GTTGAGTGGTGGTTGATTGGTTGGGATTGA
HOP-MPs	Capture Oligo (2)	Biotin	5'-Biotin-TTTTTTTTTGTGGCGGCGGTGGTTGAGTGGTGGTTGGTGTTGATT
	Displacer Oligo (2)		5'-CTCGTCATAATGTAAACCGGGAATCAACACCAACCACCACTCAAC

References

1. Wide, L., Bennich, H. & Johansson, S. Diagnosis of Allergy By an in-Vitro Test for Allergen Antibodies. *The Lancet* **290,** 1105–1107 (1967).

2. Wide, L. Inventions leading to the development of the diagnostic test kit industry - From the modern pregnancy test to the sandwich assays. *Upsala Journal of Medical Sciences* **110,** 193–216 (2005).

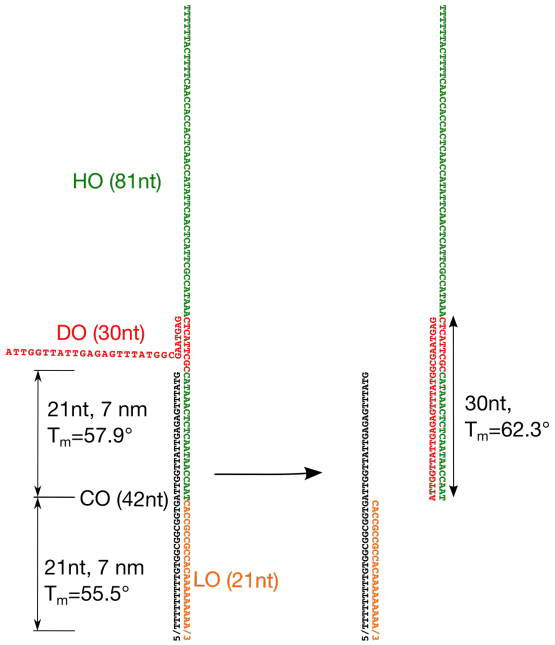


Figure 4.4: **Sequences and melting temperatures of oligonucleotides used in CLAMP.** The capture oligo (CO, 42nt) is bound to the streptavidin surface via 5'biotin, and links both the 3' fluorescent labeling oligo (LO, 21nt) and 5' antibody-conjugated hook oligo (HO, 81nt) to the microparticle surface. A displacer oligo (DO 30nt) binds a 9nt toehold on the hook oligo to displace the HO-CO hybrid.

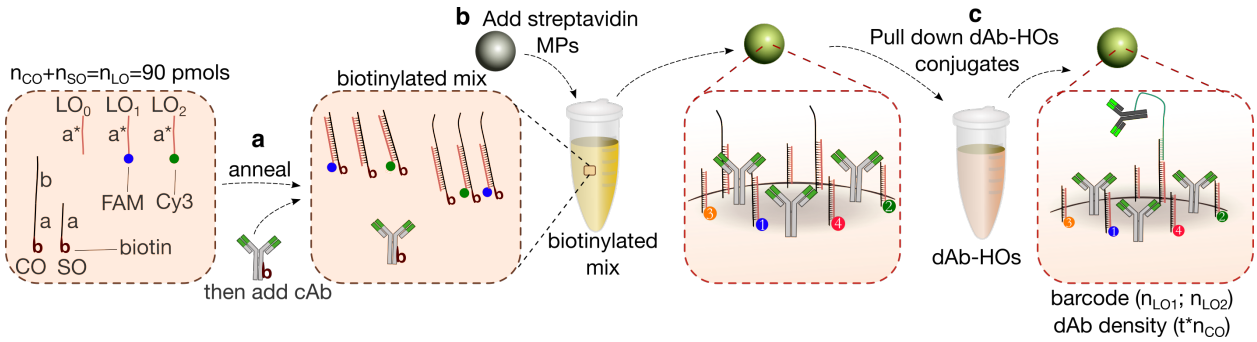


Figure 4.5: **Synthesis of CLAMPs.** Illustration describing the primary steps in the synthesis of the CLAMP. (a) Oligo constructs are pre-annealed and antibodies are added to form the biotinylated mixture of reagents. The mixture of biotinylated oligos includes precisely controlled proportion of CO/SO totalling 90 pmols and defining the CO (and later dAb-HO) surface density; the biotinylated oligos are annealed to a precisely controlled proportion of $LO_0/LO_1/LO_2$, totalling 90 pmols and defining the barcode. (b) Thereafter, streptavidin MPs are added to the biotinylated mix to proportionally and stochastically label them with reagents on the surface of the MPs, wherein the relative densities of the oligo components (*e.g.* LO_1/LO_2) is conserved on the surface. (c) The dAb-HOs are finally pulled down on the surface to complete the synthesis of CLAMPs. The dAb density on the surface is proportional to the CO density, and hence, n_{CO}.

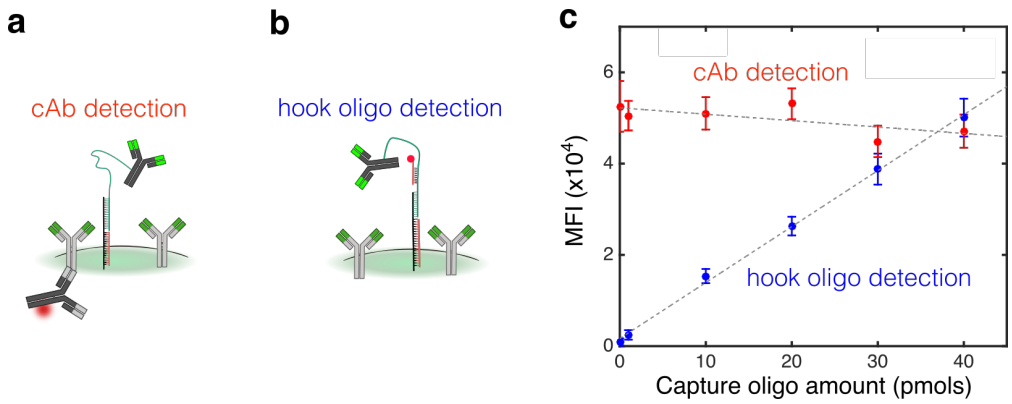


Figure 4.6: **Fine-tuning and control of surface densities.** (a) cAb detection using anti-goat antibody labeled with AF-647. (b) HO detection using a complementary, but non-displacing, oligo labeled with Cy5. Fluorescent intensities of CLAMPs with varying CO reaction amounts, labeled using AF647 x-goat secondary antibody (red) and cy5-labeled oligos targeting the HOs (non-displacing, blue dots). (c) cAb and HO detection (in red and blue, respectively) for CLAMP prepared with increasing amounts of starting CO and decreasing SO such that $n_{CO} + n_{SO} = 90$ pmols. Linear fits to the data are shown in dashed lines, and the error bars plot the standard deviation of the MP fluorescence.

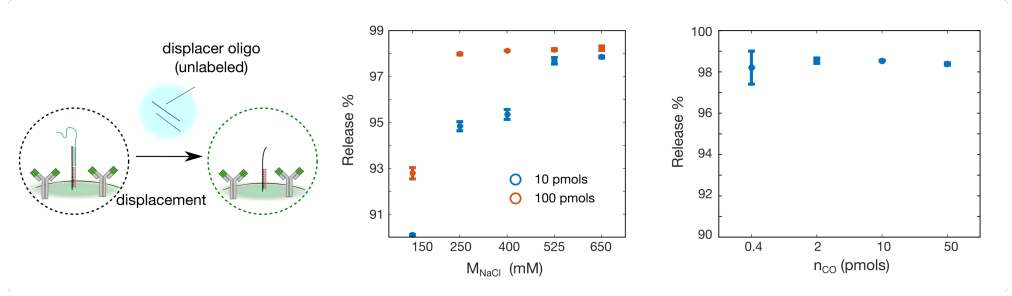


Figure 4.7: **Optimization of toe-hold mediated displacement efficiency.** (a) Illustration of the displacement reaction, wherein Cy5-labeled HOs were displaced using unlabeled DOs. (b) Release efficiency for increasing NaCl concentration (x-axis) for varying CO starting amounts (blue and red for for $n_{CO} = 10$ pmols, $n_{CO} = 100$ pmols, respectively). No SOs were used in this experiment (*i.e.* $n_{SO} = 0$ pmols). The release efficiency was calculated as $(I_0 - I_f)/(I_0 - I_B)$ where I_0, I_f, and I_B are the fluorescence before release, after release, and of the background, respectively. The release efficiency was significantly improved at increased ionic strengths. Increased density of COs led to increased release efficiency, which may be ascribed to reduced fraction of non-specifically bound oligos. (c) Release efficiency with respect to CO density at high salt concentrations (500 mM of NaCl).

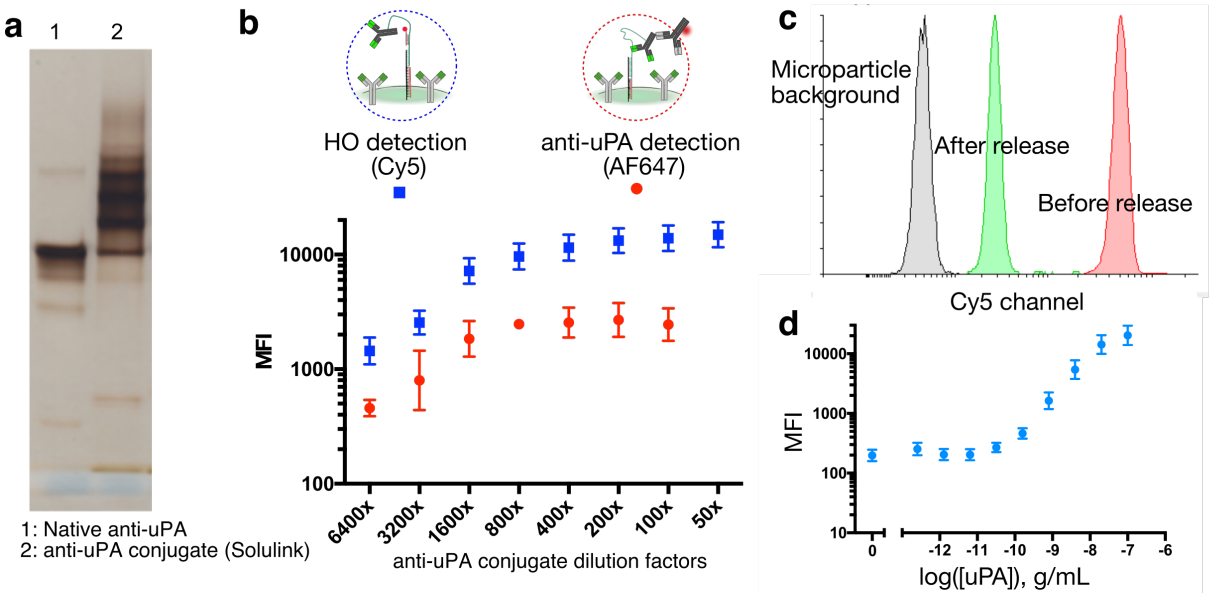


Figure 4.8: **Characterization and validation of anti-uPA conjugates made using a commercial kit (Solulink)**. (a) SDS-PAGE of anti-uPA conjugates, characterized by silver staining. (b) Pull-down and characterization of conjugates through hybridization to HOs (blue) and antibody labeling (red). (c) Histogram of MP fluorescence for CLAMPs pulled down with anti-uPA conjugates and labeled with through detection of HOs (red) before displacement with unlabeled DO, (green) after displacement with unlabeled DO, compared to MP barckground (grey). (d) Binding curve of uPA performed using a convetional (*i.e.* sequential) single-plex sandwich assay, using anti-uPA conjugates as detection reagent.

3. Kakabakos, S. E., Christopoulos, T. K. & Diamandis, E. P. Multianalyte immunoassay based on spatially distinct fluorescent areas quantified by laser-excited solid-phase time-resolved fluorometry. *Clinical chemistry* **38,** 338–42 (1992).

4. Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature reviews. Drug discovery* **5,** 310–20 (2006).

5. Pla-Roca, M. *et al.* Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & cellular proteomics : MCP* **11** (2012).

6. Juncker, D., Bergeron, S., Laforte, V. & Li, H. Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. *Current Opinion in Chemical Biology* **18,** 29–37 (2014).

7. Gonzalez, R. M. *et al.* Development and Validation of Sandwich ELISA Microarrays with Minimal Assay Interference. *J. Proteome Res.* **7,** 2406–2414 (2008).

8. Perlee, L. *et al.* Development and standardization of multiplexed antibody microarrays for use in quantitative proteomics. *Proteome science* **2,** 9 (2004).

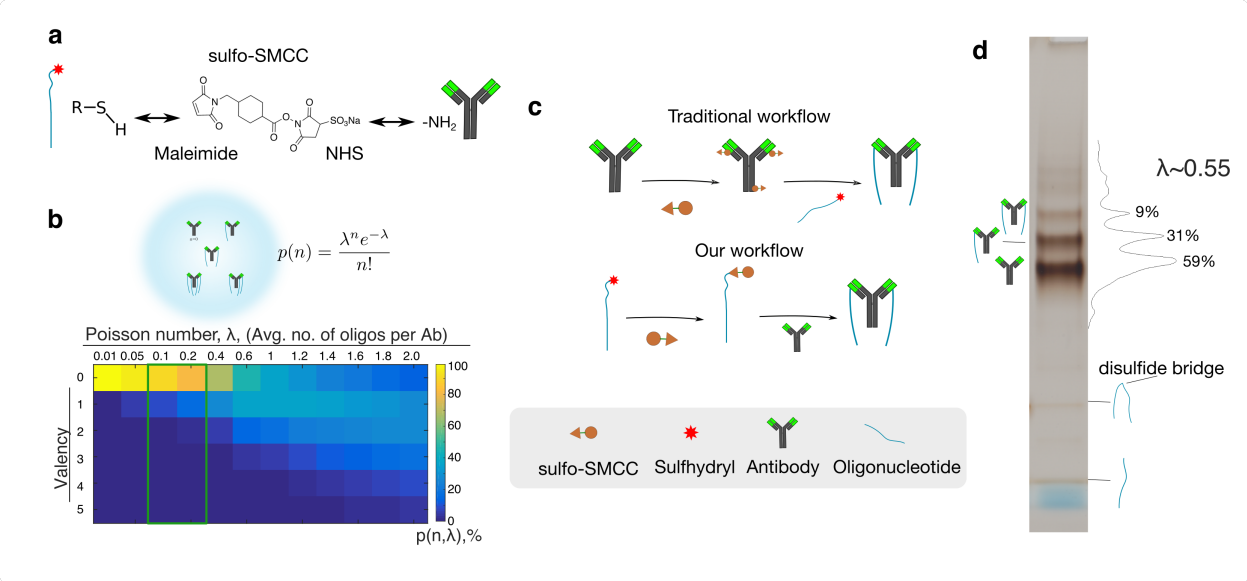


Figure 4.9: **Antibody-oligonucleotide conjugation: chemistry and valency modulation.** (a) Sulfo-SMCC (sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate) is a maleimide-NHS heterobifunctional crosslinker, used to link sulfhydryl-modified oligos with amine groups on the IgG. (b) The wide availability of amine-groups on a typical IgG molecule results in multivalent binding that follows a Poisson distribution, where the Poisson number, λ , corresponds to the average number of oligos per IgG molecule. The green box indicates the targeted range, which maximizes the monovalent conjugates fraction of all conjugated antibodies. (c) Traditional workflows (top) seek to maximize the antibody conjugation yield by activating antibodies first to minimize NHS hydrolysis. However, unconjugated antibodies are fully activated which preclude antibody recycling. We sought a reverse workflow (bottom), wherein oligonucleotides are activated first, thereby ensuring that unconjugated antibodies remain intact. (d) Characterization of conjugation valency using SDS-PAGE followed by silver staining revealing the distribution of conjugate valencies.

9. Schweitzer, B. *et al.* Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnology* **20,** 359–365 (2002).

10. Jani, D. *et al.* Recommendations for Use and Fit-for-Purpose Validation of Biomarker Multiplex Ligand Binding Assays in Drug Development. *The AAPS Journal* **18,** 1–14 (2016).

11. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplex era: Potentials and pitfalls. *Proteomics - Clinical Applications* **9,** 406–422 (2015).

12. Li, H., Bergeron, S. & Juncker, D. Microarray-to-Microarray Transfer of Reagents by Snapping of Two Chips for Cross-Reactivity-Free Multiplex Immunoassays. *Analytical chemistry* **84,** 4776–83 (2012).

13. Li, H., Munzar, J. D., Ng, A. & Juncker, D. A versatile snap chip for high-density subnanoliter chip-to-chip reagent transfer. *Scientific Reports* **5,** 1–12 (2015).

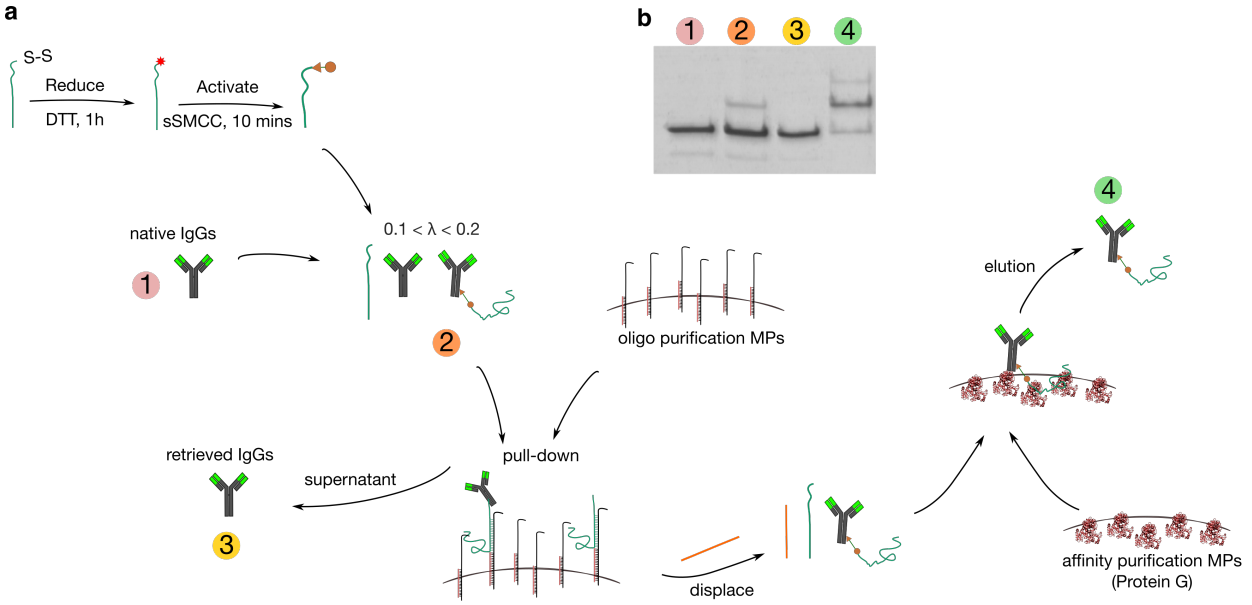


Figure 4.10: **Antibody-oligonucleotide conjugation and purification workflow.** (a) Antibodies (1) are conjugated using the reverse workflow as described in Fig. S6, aiming at a low conjugate valency (2) of lambda 0.1-0.2. Oligo-purification MPs (see Methods) are used in excess to pull-down free and conjugated oligos, leaving native IgGs in supernatant (3). Thereafter, free and conjugated oligos are displaced from the oligo-purification MPs using a toe-hold displacement strand, wherein the displacer binds to the capture oligo on the MPs. Conjugated antibodies (4) are purified from the resulting supernatant using a MP-based affinity purification step. (b) SDS-PAGE gel with silver staining of products (1-4).

14. Gullberg, M. *et al.* Cytokine detection by antibody-based proximity ligation. *Proceedings of the National Academy of Sciences* **101,** 8420–8424 (2004).

15. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic acids research* **39** (2011).

16. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5** (2010).

17. Kraemer, S. *et al.* From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PloS one* **6,** e26332 (2011).

18. Jani, I. V., Janossy, G., Brown, D. W. G. & Mandy, F. Multiplexed immunoassays by flow cytometry for diagnosis and surveillance of infectious diseases in resource-poor settings Multiplexed immunoassays for infectious diseases. *The Lancet* **2,** 243–250 (2002).

19. Krishhan, V. V., Khan, I. H. & Luciw, P. a. Multiplexed microbead immunoassays by flow cytometry for molecular profiling: Basic concepts and proteomics applications. *Critical reviews in biotechnology* **29,** 29–43 (2009).

20. Tighe, P., Negm, O., Todd, I. & Fairclough, L. Utility, reliability and reproducibility of immunoassay multiplex kits. *Methods (San Diego, Calif.)* 1–7 (2013).

21. Fu, Q., Schoenhoff, F. S., Savage, W. J., Zhang, P. & Van Eyk, J. E. Multiplex assays for biomarker research and clinical application: translational science coming of age. *Proteomics. Clinical applications* **4,** 271–84 (2010).

22. De Jager, W., Bourcier, K., Rijkers, G. T., Prakken, B. J. & Seyfert-Margolis, V. Prerequisites for cytokine measurements in clinical trials with multiplex immunoassays. *BMC immunology* **10,** 52 (2009).

23. Dagher, M., Kleinman, M., Ng, A. & Juncker, D. Ensemble multicolour FRET model enables barcoding at extreme FRET levels. *Nature Nanotechnology* **24** (2018).

24. Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J. & Kettman, J. R. Advanced multiplexed analysis with the FlowMetrix(TM) system. *Clinical Chemistry* **43,** 1749–1756 (1997).

25. Surenaud, M. *et al.* Optimization and evaluation of Luminex performance with supernatants of antigen-stimulated peripheral blood mononuclear cells. *BMC Immunology* **17** (2016).

26. Darmanis, S. *et al.* ProteinSeq: High-performance proteomic analyses by proximity ligation and next generation sequencing. *PLoS ONE* **6** (2011).

27. Agasti, S. S., Liong, M., Peterson, V. M., Lee, H. & Weissleder, R. Photocleavable DNA barcode-antibody conjugates allow sensitive and multiplexed protein analysis in single cells. *Journal of the American Chemical Society* **134,** 18499–18502 (2012).

28. Kanaras, A. G. *et al.* Site-specific ligation of DNA-modified gold nanoparticles activated by the restriction enzyme StyI. *Small (Weinheim an der Bergstrasse, Germany)* **3,** 67–70 (2007).

29. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131,** 17303–14 (2009).

30. Landegren, U. *et al.* Opportunities for sensitive plasma proteome analysis. *Analytical Chemistry* **84,** 1824–1830 (2012).

31. Rissin, D. M. *et al.* Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nature biotechnology* **28,** 595–9 (2010).

32. Wilson, R. Sensitivity and specificity: twin goals of proteomics assays. Can they be combined? *Expert review of proteomics* **10,** 135–49 (2013).

Chapter 5

CONCLUSION

Summary. In this dissertation, we presented a platform for scalable multiplexed sandwich assays (MSAs) on microparticles. The microparticle-based platform comprises (i) a method for improved spectral barcoding and (ii) a colocalization-by-linkage assay on microparticle (CLAMP) that mitigates reagent-driven cross-reactivity (rCR); together, these advances afford the integrated platform, henceforth simply called CLAMP, high sample throughput and high multiplexing capacity.

In **Chapter 3**, we developed a multicolour-labeling method hand-in-hand with a FRET model to guide the encoding and decoding of microparticles, enabling generation of high capacity barcodes using dyes with large spectral overlap. To overcome the challenge of energy transfer between classifier dyes, we modeled the energy transfer between densely-packed, stochastically distributed multicolour dyes on the surface of a microparticle by analytically deriving the equations for ensemble mFRET (emFRET) efficiency. The emFRET model was used as the kernel of a generalized multicolour fluorescence model that serves to predict the measured fluorescence intensity on a particular cytometer from the proportions of differently-coloured dyes on the surface of the microparticle. The use of this model was thus contingent on accurately assembled dye proportions on the microparticle surface. For this, we developed a DNA-assisted proportional labelling approach for microparticles. The different fluorescent dyes were conjugated to short, biotinylated DNA oligonucleotides, called linking oligos (LOs), which are co-immobilized along with the biotinylated antibodies onto streptavidin-coated microparticles in a one-pot format. The LOs function as homogeneous linkers owing to their identical affinity and surface footprint and, irrespective of the chemical properties of the dyes, enable faithful reproduction of the solution dye proportions onto the surface, *i.e.* proportional labeling. The inclusion of a non-labeled LO for compensation was key to maintaining a constant total LO density while permitting independent tuning of each dye's density. Using the model and the labeling method, calibration to the FACS Canto II cytometer was performed, allowing validation of the fluorescence model as adequately predictive for the purposes of barcoding, with a mean error < 30%. The model thus afforded the ability to simulate the fluorescence response, enabling *in silico* design of non-overlapping barcode responses, and optimization thereof to increase the maximal barcoding capacity. 580 barcodes were generated in silico for 4 common dyes (FAM; Cy3; Cy5; Cy5.5) with large spectral overlap, with many barcodes extending into regimes with extreme inter-dye FRET reaching 76% efficiency. In addition, the emFRET model also enabled fully-automated decoding with >99%confidence. To illustrate the applicability of the barcoding approach, we conducted a 35-plex combinatorial screening of specificity and non-

specific binding. 3,745 combinatorial interactions were tested quickly using our approach, further highlighting the ease-of-use and rapid workflow. These results demonstrated the applicability of the barcoding platform for scalable implementation in affinity-based proteomics.

In **Chapter 4**, we introduced CLAMP, a multiplexed sandwich immunoassay that tackles reagent-driven cross-reactivity (rCR) by pre-colocalizing antibody pairs on microparticles. For each target, both antibody sets were functionalized to the same microparticle. To allow tertiary or 'sandwich' complex formation with the target analyte, the detection antibody (dAb) was pre-linked to the surface of the microparticle via a DNA linker. The CLAMP assay is assembled using the microparticle labeling approach by replacing LOs with capture oligos (COs) that partially hybridize to hook oligos (HOs) conjugated to dAbs. HOs were 82-nt long, with a substantial single-stranded portion (61-nt) that provided flexibility and reach to scan an extended local region of the surface and form the tertiary complex with an analyte and the surface-immobilized capture antibody (cAb). For signal transduction, we designed and implemented a toe-hold-mediated displacement mechanism. A fluorescently-labeled displacer oligo (DO) binds to HOs preferentially, displacing the HO-CO hybrid, and served to releasing unbound dAb-HOs from the surface, while labeling dAb-HOs bound to their analyte and the cAb via a sandwich. This signal transduction mechanism also served to minimize assay background by exclusively labeling the 'successfully' displaced conjugates. As a result, detection only occurs in the dual-requirement of (i) a tertiary-complex formation and (ii) the successful displacement of the DO. We demonstrate that this mechanism is contingent on dAb-HO conjugates with low valency of HO–indeed we found that multivalent conjugates resulted in a 10-12 -fold increase in assay background above the raw signal of the microparticles. To overcome this problem, conjugates with low valency (*i.e.* low average number of oligos per Ab, λ) were prepared using a specially developed conjugation protocol, restoring the assay background to only 1.1-1.25-fold above the microparticle background. A single-plex CLAMP was prepared against urokinase Plasminogen Activator (uPA) to demonstrate function and optimize the surface densities of dAb-HOs as defined by the CO density. Conjugates with low valency ($\lambda \sim 0.15$, home protocol) exhibited improved sensitivity compared to high valency ($\lambda \sim 2$, conjugation by a commercial kit from Solulink). Thereafter, 6 antibody pairs selected for high non-specific binding (NSB) and rCR were chosen from the 35-plex specificity screening performed in Chapter 3 and implemented on the CLAMP as a 6-plex assay. The CLAMP was found to eliminate the rCR observed previously; in addition, CLAMP identified a dual cross-reactivity event, wherein both MCP-1 antibodies cross-reacted with EGF protein, which was confirmed using a single-plex sandwich assay. These results illustrate the applicability of the CLAMP platform in mitigating rCR for microparticle-based MSAs, and together with barcoding, the possibility of scaling up the degree of multiplexing.

Conclusions. In this dissertation, we have addressed several longstanding technical challenges that have limited the scalability of microparticle-based MSAs. Specifically, our emFRET model is, to our knowledge, the first mechanistic description of energy transfer between multicolour dyes stochastically distributed in space. The emFRET model enabled barcoding using common dyes with high spectral overlap that are typically deemed not suitable for barcoding. Coupled with the proportional labeling approach we developed, this demonstration represents the first 'quantitative' barcoding technique using common dyes, wherein ensemble spectra were accurately designed *in silico*. The CLAMP is the first MSA on microparticles that mitigates rCR, which was until now the current bottleneck to multiplexing. Further, the pre-colocalization of antibodies represents a departure from typical immunoassays, and promises to allow rational design of the bivalent sensor for improved performance. Crucially, all advances were integrated into the CLAMP platform.

The results of the specificity screening using conventional MSAs in **Chapter 3** demonstrated their susceptibility to rCR and NSB, in accordance with many previous studies, including ones from our group [1]. Indeed, out of 35 antibody pairs, only a few combinations of 5-10plex were deemed compatible for multiplexed analysis, whereas some binders were evidently not compatible any other binder pairs. On the other hand, multiplexed CLAMP overcame this rCR without need for optimization, demonstrating its usefulness towards enabling increased multiplexing in microparticle-based assays. Interestingly, the CLAMP stands to be uniquely positioned to enable rapid, multiplexed screening of dCR in sandwich assays.

Owing to the flexible and easily-customizable nature of the barcoding approach, the microparticles can be read-out using common cytometers with fully automated decoding. The CLAMP's ease-of-use and cost-efficiency coupled with the magnetic microparticles' amenability for automation paves the way for assays with increased sample throughput. When using the strongly spectrally overlapping dyes of this work, only 60 out of the 580 barcodes have sufficiently low FRET efficiencies (E < 20 %) to allow straightforward generation and decoding, absent a model.

The use of DNA oligonucleotids was key for both barcoding and CLAMP. For example, in the barcoding approach DNA was used as a homogeneous scaffold to link the dyes and the microparticles together. In the assay, the use of DNA was elevated to functional, programmable, and trigger-addressable elements that mediated protein recognition and signal transduction. Indeed, an overarching strategy of the CLAMP platform is the combination of off-the-shelf reagents including magnetic microparticles, common dyes, commercial antibodies, with rationally designed oligos that link the reagents together to improve performance and integrate new functionalities.

Limitations. Further studies could contribute to the betterment of the presented method by addressing several limitations pertaining to the experimental methods as well as theoretical modeling.

On the experimental side, a covalent immobilization strategy stands to provide some advantages in comparison to the biotin-streptavidin linkage used throughout this dissertation. Despite the very low K_d exhibited by the biotin-streptavidin bond, covalent bonds will provide a stability that is amenable for long-term storage. Importantly, covalent attachment will provide the possibility to store mixtures of multiplexed CLAMPs against different targets, which simplifies the assay preparation steps. Evidently, the ability to mix multiplexed CLAMPs *a priori* also depends on the stability of the DNA hybrid to ensure no antibody exchange occurs during storage. In addition, with increased conjugation scales, the use of liquid chromatography techniques such as fast-protein liquid chromatography (FPLC) could provide improved purification yield and purity in comparison to our current approach. Indeed, the conjugation and purification workflow presented in **Chapter 4** calls for several steps including two purification steps, which contributes to losses and variability of the yield. Furthermore, the purity of the monovalent conjugates will not surpass 90%. An FPLC could potentially allow the purification of the monovalent conjugates from a heterogeneous mixture. Furthermore, as discussed in **Chapter 4**, deploying a signal amplification scheme constitute a key objective towards assays with improved sensitivity. Indeed, the current assay format featured one fluorophore per binding event, which is lower than typical sandwich assays with multivalent fluorophore conjugation, which is expected to limit the assay performance. Finally, the assays presented in this thesis were performed by spiking antigen standards in buffer; it is important that upcoming experiments explore and characterize assay performance in complex samples such as plasma to ensure robustness and applicability to clinical samples. Indeed, it is expected that plasma samples may contain interfering factors such as auto-antibodies, protein-protein interactions, and other matrix effects which may impact the accuracy of the CLAMP read-out. For this reason, experimentation with different buffers and interefence-reducing approaches will be necessary to build fit-for-purpose assays, as reviewer in (**Chapter 2**).

Modeling and barcode design may be developed further to achieve higher capacities and improved accuracy. For example, The current approach in designing barcodes *in silico* may be substantially improved upon implementation of an automated algorithm that maximizes the barcoding capacity. Indeed, as is apparent in **Figure 3.4**, the designed barcodes did not exhaustively utilize the spectral-intensity space available. Current barcodes were designed by trial-and-error *in silico* by manually changing the dye proportions (*i.e.* Supplementary Table S3 - Appendix A) to yield distinguishable barcodes. Thirdly, while the cluster location in the barcoding space or their mean intensities were predicted by the multicolour fluorescence model, the size and spread of the clusters (*i.e.* the dispersion matrix) were estimated empirically and considered to be independent of intensity levels. However, the results have demonstrated that the dispersion is density dependent. Hence, a improved model could also take into account these differences, even if only empirically, which could further help increase barcoding capacitity. Furthermore, as discussed in **Chapter 3**, our model neglects

self-quenching which may be substantial at very high densities, and may have contributed to the reduced accuracy of the model at high intensities in the red channels. The quenching response for every dye might be predictable using theoretical models, or could be determined empirically, and the model further refined by taking it into account for the purpose of predictive barcoding.

Outlook Future work could proceed in several directions, including (i) scaling of the barcoding platform, (ii) scaling of the multiplexing level in the assay, (iii) application and optimization of the CLAMP platform for different biological samples, as well as (iv) generalization of the CLAMP concept to exploring different applications.

The future directions of the present dissertations could focus on several directions including scaling of the barcoding platform, scaling of the multiplexing level in the assay, application and optimization of the CLAMP platform for different biological samples, as well as generalization of the CLAMP concept to exploring different applications. The barcoding capacity of the current platform stands to be increased through three avenues. First, as mentioned above, an algorithm for *in silico* barcode design and optimization may be implemented to yield additional distinguishable clusters in the barcoding space, all while using the same set of dyes and optical setup. Development of such an algorithm is an interesting optimization problem which could be tackled through either linear or evolutionary optimization means, with the singular goal of increasing the number of distinguishable barcodes. Second, with the advent of near infrared lasers in flow cytometers, dyes such as Cy7 and Cy7.5 may be added to generate six-colour microparticles. The high degree of flexibility provided by our synthesis and modeling would permit simple inclusion of additional organic dyes into the LO mixture, and modeling of their multicolour fluorescence to generate barcodes as well as permit their automated decoding. Third, new assay carriers with increased volumetric capacity for immobilization, such as hydrogel microparticles [2], may be explored. Indeed, evolution of the platform from 2D (*i.e.* microparticle surface) to 3D (*i.e.* microparticle matrix) immobilization schemes will provide increased volumetric dye capacity that would increase the number of intensity levels that can be encoded per dye. A challenge of this option would be to find suitable hydrogel carriers with magnetic properties, so as to maintain this advantageous attribute for assays.

In addition, future studies could investigate other potential advantages stemming from the pre-colocalization format of CLAMP. One of the possible benefits of pre-colocalization is an increased avidity for the target, which may resist stringent washing during the sample incubation step and achieve a more complete washing of non-specifically bound analytes. The effect of increased washing cycles, or otherwise the addition of anionic competitors, could be investigated through the use of a labeled plasma sample with a differently colored dye permit kinetic measurements and optimization of the washing step.

In addition, future studies could investigate additional advantages enabled by the pre-colocalization format of CLAMP. One of the potential benefits of pre-colocalization is an increased avidity for the target, which may resist stringent washing during the sample incubation step and achieve a more complete washing of non-specifically bound analytes. The effect of increased washing cycles, or otherwise the addition of anionic competitors, can be investigated through the use of a labeled plasma sample.

The use of a DNA-mediated signal transduction provides an opportunity for deploying DNA-based signal amplification schemes. For instance, the displacer oligo (DO) used in the current format could be used in non-fluorescent form, and may contain a single-stranded sequence to trigger hybridization of larger DNA assemblies through hybridization chain reaction, or non-linear hybridization-based amplification [3].

The continued development of the technologies introduced in this dissertation by the addition of increased number of targets, the increased sensitivity and specificity, or increased miniaturization of the CLAMP could further its biological applications. With additional modifications such as the ones mentioned above, the CLAMP platform has the potential to provide both the multiplexing scalability and large sample throughput to enable cost-efficient, highly multieplexed, population-wide studies that promise to mirror the developments in genomics.

References

1. Pla-Roca, M. *et al.* Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & cellular proteomics : MCP* **11** (2012).
2. Pregibon, D. C., Toner, M. & Doyle, P. S. Multifunctional encoded particles for high-throughput biomolecule analysis. *Science (New York, N.Y.)* **315,** 1393–6 (2007).
3. Dirks, R. M. & Pierce, N. a. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 15275–8 (2004).

Appendix A

SUPPLEMENTARY TABLE 3.3

	Starting stoichiometries in pmoles					Channel intensities (SBR)				emFRET efficienci			
BC#	LO_0	LO_1	LO_2	LO_3	LO_4	i1	i2	i3	i4	E1T	E12	E2T	E34
1	90	0	0	0	0	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
2	90	0	0	0.5	0	1.00	1.00	3.15	1.76	0.01	0.00	0.01	0.00
3	88	0	0	2	0	1.00	1.00	9.59	4.03	0.03	0.00	0.04	0.00
4	85	0	0	5	0	1.00	1.00	22.48	8.58	0.08	0.00	0.11	0.00
5	80	0	0	10	0	1.00	1.00	43.96	16.15	0.16	0.00	0.22	0.00
6	70	0	0	20	0	1.00	1.00	86.92	31.30	0.29	0.00	0.37	0.00
7	50	0	0	40	0	1.00	1.00	#####	61.61	0.47	0.00	0.56	0.00
8	25	0	0	65	0	1.00	1.00	#####	99.49	0.60	0.00	0.68	0.00
9	88	0	0	0	2	1.00	1.00	1.06	4.92	0.03	0.00	0.04	0.07
10	87	0	0	0.5	3	1.00	1.00	3.01	7.70	0.05	0.00	0.07	0.11
11	84	0	0	2	4	1.00	1.00	8.50	12.18	0.09	0.00	0.12	0.14
12	79	0	0	6	5	1.00	1.00	22.44	21.02	0.17	0.00	0.22	0.18
13	69	0	0	15	6	1.00	1.00	52.40	38.80	0.30	0.00	0.37	0.21
14	52	0	0	30	8	1.00	1.00	96.43	70.56	0.45	0.00	0.53	0.26
15	30	0	0	51	9	1.00	1.00	#####	####	0.57	0.00	0.66	0.29
16	82	0	0	0	8	1.00	1.00	1.26	16.69	0.12	0.00	0.15	0.26
17	80	0	0	1.5	9	1.00	1.00	5.88	21.38	0.16	0.00	0.20	0.29
18	72	0	0	8	10	1.00	1.00	13.15	28.01	0.21	0.00	0.26	0.31
19	65	0	0	11	14	1.00	1.00	30.03	49.79	0.33	0.00	0.41	0.40
20	51	0	0	23	16	1.00	1.00	57.78	77.85	0.45	0.00	0.53	0.43
21	30	0	0	40	20	1.00	1.00	89.12	####	0.57	0.00	0.65	0.50
22	70	0	0	0	20	1.00	1.00	1.64	40.22	0.27	0.00	0.33	0.50
23	66	0	0	3.5	21	1.00	1.00	9.13	49.38	0.32	0.00	0.39	0.51
24	56	0	0	9	25	1.00	1.00	19.15	68.99	0.41	0.00	0.48	0.56
25	37	0	0	22	31	1.00	1.00	39.02	####	0.53	0.00	0.61	0.61
26	35	0	0	0	55	1.00	1.00	2.77	####	0.53	0.00	0.60	0.75
27	30	0	0	5	55	1.00	1.00	8.31	####	0.56	0.00	0.63	0.75
28	30	0	0	15	45	1.00	1.00	21.85	####	0.56	0.00	0.63	0.71
29	89	1	0	0	0	1.88	1.14	1.00	1.00	0.00	0.00	0.00	0.00
30	87	3	0	0	0	3.65	1.43	1.00	1.00	0.00	0.00	0.00	0.00
31	81	9	0	0	0	8.96	2.28	1.00	1.00	0.00	0.00	0.00	0.00
32	71	19	0	0	0	17.81	3.70	1.00	1.00	0.00	0.00	0.00	0.00
33	53	30	7	0	0	23.24	6.52	1.00	1.00	0.17	0.17	0.00	0.00
34	50	40	0	0	0	36.38	6.69	1.00	1.00	0.00	0.00	0.00	0.00
35	25	50	15	0	0	31.60	11.29	1.00	1.00	0.32	0.32	0.00	0.00
36	30	60	0	0	0	54.07	9.53	1.00	1.00	0.00	0.00	0.00	0.00
37	0	90	0	0	0	80.61	13.79	1.00	1.00	0.00	0.00	0.00	0.00
38	80	0	10	0	0	1.08	2.10	1.00	1.00	0.23	0.23	0.00	0.00
39	78	2	10	0	0	2.45	2.43	1.00	1.00	0.23	0.23	0.00	0.00
40	74	6	10	0	0	5.20	3.09	1.00	1.00	0.23	0.23	0.00	0.00

41	67	13	10	0	0	10.00	4.24	1.00	1.00	0.23	0.23	0.00	0.00
42	0	74	16	0	0	45.24	15.66	1.00	1.00	0.33	0.33	0.00	0.00
43	65	0	25	0	0	1.20	3.75	1.00	1.00	0.45	0.45	0.00	0.00
44	62	3	25	0	0	2.69	4.31	1.00	1.00	0.45	0.45	0.00	0.00
45	55	8	27	0	0	5.03	5.47	1.00	1.00	0.47	0.47	0.00	0.00
46	46	17	27	0	0	9.33	7.16	1.00	1.00	0.47	0.47	0.00	0.00
47	35	32	23	0	0	17.66	9.40	1.00	1.00	0.43	0.43	0.00	0.00
48	5	50	35	0	0	22.05	14.57	1.00	1.00	0.54	0.54	0.00	0.00
49	30	0	60	0	0	1.48	7.60	1.00	1.00	0.68	0.68	0.00	0.00
50	22	8	60	0	0	3.83	9.26	1.00	1.00	0.68	0.68	0.00	0.00
51	12	18	60	0	0	6.78	11.34	1.00	1.00	0.68	0.68	0.00	0.00
52	0	30	60	0	0	10.31	13.84	1.00	1.00	0.68	0.68	0.00	0.00
53	0	2	88	0	0	2.15	11.11	1.00	1.00	0.76	0.76	0.00	0.00
54	90	0	0	0.5	0	1.00	1.00	3.15	1.76	0.01	0.00	0.01	0.00
55	89	1	0	0.5	0	1.88	1.14	3.15	1.76	0.01	0.00	0.01	0.00
56	87	3	0	0.5	0	3.63	1.42	3.15	1.76	0.01	0.00	0.01	0.00
57	81	9	0	0.5	0	8.90	2.27	3.15	1.76	0.01	0.00	0.01	0.00
58	71	19	0	0.5	0	17.69	3.68	3.15	1.76	0.01	0.00	0.01	0.00
59	53	30	7	0.5	0	23.05	6.46	3.15	1.76	0.17	0.17	0.01	0.00
60	50	40	0	0.5	0	36.13	6.65	3.15	1.76	0.01	0.00	0.01	0.00
61	25	50	15	0.5	0	31.36	11.18	3.15	1.76	0.32	0.32	0.01	0.00
62	30	60	0	0.5	0	53.70	9.47	3.15	1.76	0.01	0.00	0.01	0.00
63	0	90	0	0.5	0	79.61	13.63	3.15	1.76	0.01	0.00	0.01	0.00
64	80	0	10	0.5	0	1.08	2.09	3.15	1.76	0.24	0.23	0.01	0.00
65	78	2	10	0.5	0	2.44	2.41	3.15	1.76	0.24	0.23	0.01	0.00
66	74	6	10	0.5	0	5.16	3.06	3.15	1.76	0.24	0.23	0.01	0.00
67	67	13	10	0.5	0	9.92	4.20	3.15	1.76	0.24	0.23	0.01	0.00
68	0	74	16	0.5	0	44.60	15.41	3.15	1.76	0.34	0.33	0.01	0.00
69	65	0	25	0.5	0	1.20	3.72	3.15	1.76	0.45	0.45	0.01	0.00
70	62	3	25	0.5	0	2.67	4.27	3.15	1.76	0.45	0.45	0.01	0.00
71	55	8	27	0.5	0	5.01	5.42	3.15	1.76	0.47	0.47	0.01	0.00
72	46	17	27	0.5	0	9.27	7.09	3.15	1.76	0.47	0.47	0.01	0.00
73	35	32	23	0.5	0	17.55	9.30	3.15	1.76	0.43	0.42	0.01	0.00
74	4.5	50	35	0.5	0	21.92	14.42	3.15	1.76	0.54	0.54	0.01	0.00
75	30	0	60	0.5	0	1.48	7.53	3.15	1.76	0.68	0.68	0.01	0.00
76	22	8	60	0.5	0	3.82	9.18	3.15	1.76	0.68	0.68	0.01	0.00
77	12	18	60	0.5	0	6.75	11.23	3.15	1.76	0.68	0.68	0.01	0.00
78	0	30	60	0.5	0	10.11	13.60	3.15	1.76	0.68	0.68	0.01	0.00
79	0	2	87.5	0.5	0	2.14	10.96	3.15	1.76	0.76	0.76	0.01	0.00
80	88	0	0	2	0	1.00	1.00	9.59	4.03	0.03	0.00	0.04	0.00
81	87	1	0	2	0	1.86	1.14	9.59	4.03	0.03	0.00	0.04	0.00
82	85	3	0	2	0	3.57	1.41	9.59	4.03	0.03	0.00	0.04	0.00
83	79	9	0	2	0	8.71	2.24	9.59	4.03	0.03	0.00	0.04	0.00

84	69	19	0	2	0	17.28	3.62	9.59	4.03	0.03	0.00	0.04	0.00
85	51	30	7	2	0	22.48	6.29	9.59	4.03	0.20	0.16	0.04	0.00
86	48	40	0	2	0	35.27	6.51	9.59	4.03	0.03	0.00	0.04	0.00
87	23	50	15	2	0	30.66	10.82	9.59	4.03	0.34	0.31	0.04	0.00
88	28	60	0	2	0	52.40	9.26	9.59	4.03	0.03	0.00	0.04	0.00
89	0	88	0	2	0	76.38	13.11	9.59	4.03	0.03	0.00	0.04	0.00
90	78	0	10	2	0	1.08	2.05	9.59	4.03	0.25	0.22	0.04	0.00
91	76	2	10	2	0	2.40	2.37	9.59	4.03	0.25	0.22	0.04	0.00
92	72	6	10	2	0	5.06	2.99	9.59	4.03	0.25	0.22	0.04	0.00
93	65	13	10	2	0	9.70	4.09	9.59	4.03	0.25	0.22	0.04	0.00
94	0	74	14	2	0	45.82	14.57	9.59	4.03	0.32	0.29	0.04	0.00
95	63	0	25	2	0	1.19	3.63	9.59	4.03	0.46	0.44	0.04	0.00
96	60	3	25	2	0	2.64	4.16	9.59	4.03	0.46	0.44	0.04	0.00
97	53	8	27	2	0	4.93	5.27	9.59	4.03	0.48	0.46	0.04	0.00
98	44	17	27	2	0	9.11	6.87	9.59	4.03	0.48	0.46	0.04	0.00
99	33	32	23	2	0	17.20	9.00	9.59	4.03	0.44	0.42	0.04	0.00
100	3	50	35	2	0	21.55	13.93	9.59	4.03	0.55	0.53	0.04	0.00
101	28	0	60	2	0	1.46	7.31	9.59	4.03	0.69	0.67	0.04	0.00
102	20	8	60	2	0	3.77	8.89	9.59	4.03	0.69	0.67	0.04	0.00
103	10	18	60	2	0	6.66	10.87	9.59	4.03	0.69	0.67	0.04	0.00
104	0	28	60	2	0	9.55	12.84	9.59	4.03	0.69	0.67	0.04	0.00
105	0	0.5	87.5	2	0	1.78	10.30	9.59	4.03	0.77	0.76	0.04	0.00
106	85	0	0	5	0	1.00	1.00	22.48	8.58	0.08	0.00	0.11	0.00
107	84	1	0	5	0	1.81	1.13	22.48	8.58	0.08	0.00	0.11	0.00
108	82	3	0	5	0	3.44	1.39	22.48	8.58	0.08	0.00	0.11	0.00
109	76	9	0	5	0	8.31	2.17	22.48	8.58	0.08	0.00	0.11	0.00
110	66	19	0	5	0	16.43	3.48	22.48	8.58	0.08	0.00	0.11	0.00
111	48	30	7	5	0	21.40	5.95	22.48	8.58	0.24	0.16	0.11	0.00
112	45	40	0	5	0	33.48	6.22	22.48	8.58	0.08	0.00	0.11	0.00
113	20	50	15	5	0	29.34	10.14	22.48	8.58	0.37	0.30	0.11	0.00
114	25	60	0	5	0	49.73	8.83	22.48	8.58	0.08	0.00	0.11	0.00
115	0	85	0	5	0	70.03	12.09	22.48	8.58	0.08	0.00	0.11	0.00
116	75	0	10	5	0	1.07	1.98	22.48	8.58	0.29	0.22	0.11	0.00
117	73	2	10	5	0	2.33	2.27	22.48	8.58	0.29	0.22	0.11	0.00
118	69	6	10	5	0	4.86	2.86	22.48	8.58	0.29	0.22	0.11	0.00
119	62	13	10	5	0	9.28	3.89	22.48	8.58	0.29	0.22	0.11	0.00
120	0	72	13	5	0	43.56	13.14	22.48	8.58	0.34	0.27	0.11	0.00
121	60	0	25	5	0	1.18	3.44	22.48	8.58	0.48	0.43	0.11	0.00
122	57	3	25	5	0	2.57	3.93	22.48	8.58	0.48	0.43	0.11	0.00
123	50	8	27	5	0	4.77	4.96	22.48	8.58	0.50	0.45	0.11	0.00
124	41	17	27	5	0	8.80	6.44	22.48	8.58	0.50	0.45	0.11	0.00
125	30	32	23	5	0	16.56	8.42	22.48	8.58	0.46	0.40	0.11	0.00
126	0	50	35	5	0	20.85	12.97	22.48	8.58	0.57	0.52	0.11	0.00

127	25	0	60	5	0	1.43	6.85	22.48	8.58	0.69	0.66	0.11	0.00
128	17	8	60	5	0	3.67	8.31	22.48	8.58	0.69	0.66	0.11	0.00
129	7	18	60	5	0	6.49	10.14	22.48	8.58	0.69	0.66	0.11	0.00
130	0	30	55	5	0	10.33	11.79	22.48	8.58	0.67	0.64	0.11	0.00
131	0	3	82	5	0	2.26	9.56	22.48	8.58	0.76	0.73	0.11	0.00
132	80	0	0	10	0	1.00	1.00	43.96	16.15	0.16	0.00	0.22	0.00
133	79	1	0	10	0	1.74	1.12	43.96	16.15	0.16	0.00	0.22	0.00
134	77	3	0	10	0	3.23	1.36	43.96	16.15	0.16	0.00	0.22	0.00
135	71	9	0	10	0	7.68	2.07	43.96	16.15	0.16	0.00	0.22	0.00
136	61	19	0	10	0	15.11	3.27	43.96	16.15	0.16	0.00	0.22	0.00
137	43	30	7	10	0	19.79	5.46	43.96	16.15	0.30	0.15	0.22	0.00
138	40	40	0	10	0	30.70	5.77	43.96	16.15	0.16	0.00	0.22	0.00
139	15	50	15	10	0	27.35	9.15	43.96	16.15	0.41	0.28	0.22	0.00
140	20	60	0	10	0	45.56	8.16	43.96	16.15	0.16	0.00	0.22	0.00
141	0	80	0	10	0	60.41	10.55	43.96	16.15	0.16	0.00	0.22	0.00
142	70	0	10	10	0	1.06	1.86	43.96	16.15	0.34	0.21	0.22	0.00
143	68	2	10	10	0	2.23	2.13	43.96	16.15	0.34	0.21	0.22	0.00
144	64	6	10	10	0	4.57	2.65	43.96	16.15	0.34	0.21	0.22	0.00
145	57	13	10	10	0	8.65	3.58	43.96	16.15	0.34	0.21	0.22	0.00
146	55	0	25	10	0	1.16	3.16	43.96	16.15	0.51	0.41	0.22	0.00
147	52	3	25	10	0	2.46	3.59	43.96	16.15	0.51	0.41	0.22	0.00
148	45	8	27	10	0	4.54	4.50	43.96	16.15	0.53	0.43	0.22	0.00
149	36	17	27	10	0	8.33	5.82	43.96	16.15	0.53	0.43	0.22	0.00
150	25	32	23	10	0	15.58	7.58	43.96	16.15	0.50	0.38	0.22	0.00
151	0	48	32.5	10	0	19.50	10.92	43.96	16.15	0.57	0.47	0.22	0.00
152	20	0	60	10	0	1.38	6.18	43.96	16.15	0.71	0.63	0.22	0.00
153	12	8	60	10	0	3.53	7.47	43.96	16.15	0.71	0.63	0.22	0.00
154	2	18	60	10	0	6.22	9.08	43.96	16.15	0.71	0.63	0.22	0.00
155	0	29	51	10	0	9.98	9.98	43.96	16.15	0.67	0.59	0.22	0.00
156	0	3	77	10	0	2.16	8.14	43.96	16.15	0.75	0.69	0.22	0.00
157	69	1	0	20	0	1.63	1.10	86.92	31.30	0.29	0.00	0.37	0.00
158	67	3	0	20	0	2.88	1.30	86.92	31.30	0.29	0.00	0.37	0.00
159	61	9	0	20	0	6.65	1.91	86.92	31.30	0.29	0.00	0.37	0.00
160	51	19	0	20	0	12.92	2.92	86.92	31.30	0.29	0.00	0.37	0.00
161	35	30	5	20	0	17.84	4.49	86.92	31.30	0.37	0.10	0.37	0.00
162	30	40	0	20	0	26.10	5.03	86.92	31.30	0.29	0.00	0.37	0.00
163	5	50	15	20	0	24.05	7.63	86.92	31.30	0.48	0.26	0.37	0.00
164	10	60	0	20	0	38.65	7.05	86.92	31.30	0.29	0.00	0.37	0.00
165	60	0	10	20	0	1.05	1.69	86.92	31.30	0.43	0.18	0.37	0.00
166	58	2	10	20	0	2.06	1.91	86.92	31.30	0.43	0.18	0.37	0.00
167	54	6	10	20	0	4.08	2.34	86.92	31.30	0.43	0.18	0.37	0.00
168	47	13	10	20	0	7.62	3.11	86.92	31.30	0.43	0.18	0.37	0.00
169	45	0	25	20	0	1.13	2.73	86.92	31.30	0.57	0.37	0.37	0.00

170	42	3	25	20	0	2.29	3.08	86.92	31.30	0.57	0.37	0.37	0.00
171	35	8	27	20	0	4.14	3.82	86.92	31.30	0.58	0.39	0.37	0.00
172	26	17	27	20	0	7.53	4.88	86.92	31.30	0.58	0.39	0.37	0.00
173	15	32	23	20	0	13.93	6.31	86.92	31.30	0.55	0.35	0.37	0.00
174	10	0	60	20	0	1.30	5.16	86.92	31.30	0.73	0.59	0.37	0.00
175	2	8	60	20	0	3.29	6.18	86.92	31.30	0.73	0.59	0.37	0.00
176	0	14	56	20	0	4.91	6.66	86.92	31.30	0.71	0.58	0.37	0.00
177	0	25	45	20	0	8.57	7.23	86.92	31.30	0.67	0.52	0.37	0.00
178	0	3	67	20	0	2.03	6.03	86.92	31.30	0.75	0.62	0.37	0.00
179	49	1	0	40	0	1.47	1.08	#####	61.61	0.47	0.00	0.56	0.00
180	47	3	0	40	0	2.41	1.23	#####	61.61	0.47	0.00	0.56	0.00
181	44	6	0	40	0	3.83	1.45	#####	61.61	0.47	0.00	0.56	0.00
182	39	11	0	40	0	6.18	1.83	#####	61.61	0.47	0.00	0.56	0.00
183	32	18	0	40	0	9.48	2.36	#####	61.61	0.47	0.00	0.56	0.00
184	20	30	0	40	0	15.13	3.27	#####	61.61	0.47	0.00	0.56	0.00
185	0	50	0	40	0	24.55	4.78	#####	61.61	0.47	0.00	0.56	0.00
186	40	0	10	40	0	1.04	1.49	#####	61.61	0.55	0.15	0.56	0.00
187	38	2	10	40	0	1.83	1.65	#####	61.61	0.55	0.15	0.56	0.00
188	34	8	8	40	0	4.30	2.02	#####	61.61	0.54	0.12	0.56	0.00
189	25	0	25	40	0	1.09	2.22	#####	61.61	0.64	0.31	0.56	0.00
190	22	3	25	40	0	2.04	2.47	#####	61.61	0.64	0.31	0.56	0.00
191	15	8	27	40	0	3.56	2.99	#####	61.61	0.65	0.33	0.56	0.00
192	0	30	20	40	0	11.25	4.44	#####	61.61	0.62	0.26	0.56	0.00
193	0	0	50	40	0	1.18	3.44	#####	61.61	0.74	0.48	0.56	0.00
194	0	4	46	40	0	2.15	3.59	#####	61.61	0.73	0.46	0.56	0.00
195	0	15	35	40	0	5.31	3.99	#####	61.61	0.69	0.39	0.56	0.00
196	0	1	0	65	0	1.35	1.06	#####	99.49	0.60	0.00	0.68	0.00
197	22	3	0	65	0	2.06	1.17	#####	99.49	0.60	0.00	0.68	0.00
198	19	6	0	65	0	3.12	1.34	#####	99.49	0.60	0.00	0.68	0.00
199	13	12	0	65	0	5.25	1.68	#####	99.49	0.60	0.00	0.68	0.00
200	7	18	0	65	0	7.37	2.02	#####	99.49	0.60	0.00	0.68	0.00
201	7	30	0	65	0	11.62	2.71	#####	99.49	0.60	0.00	0.68	0.00
202	5	0	20	65	0	1.05	1.70	#####	99.49	0.69	0.22	0.68	0.00
203	0	7	18	65	0	2.99	2.05	#####	99.49	0.69	0.20	0.68	0.00
204	0	3	20	65	0	1.87	1.88	#####	99.49	0.69	0.22	0.68	0.00
205	88	0	0	0	2	1.00	1.00	1.06	4.92	0.03	0.00	0.04	0.07
206	87	1	0	0	2	1.86	1.14	1.06	4.92	0.03	0.00	0.04	0.07
207	87	3	0	0	2	3.58	1.41	1.06	4.92	0.03	0.00	0.04	0.07
208	79	9	0	0	2	8.73	2.24	1.06	4.92	0.03	0.00	0.04	0.07
209	69	19	0	0	2	17.32	3.62	1.06	4.92	0.03	0.00	0.04	0.07
210	69	30	7	0	2	22.54	6.32	1.06	4.92	0.19	0.16	0.04	0.07
211	48	40	0	0	2	35.37	6.52	1.06	4.92	0.03	0.00	0.04	0.07
212	23	50	15	0	2	30.74	10.88	1.06	4.92	0.34	0.31	0.04	0.07

213	23	60	0	0	2	52.55	9.28	1.06	4.92	0.03	0.00	0.04	0.07
214	88	90	0	0	2	78.33	13.43	1.06	4.92	0.03	0.00	0.04	0.07
215	78	0	10	0	2	1.08	2.06	1.06	4.92	0.25	0.22	0.04	0.07
216	78	2	10	0	2	2.41	2.37	1.06	4.92	0.25	0.22	0.04	0.07
217	72	6	10	0	2	5.07	3.01	1.06	4.92	0.25	0.22	0.04	0.07
218	65	13	10	0	2	9.72	4.11	1.06	4.92	0.25	0.22	0.04	0.07
219	65	74	16	0	2	44.02	15.06	1.06	4.92	0.35	0.33	0.04	0.07
220	63	0	25	0	2	1.19	3.65	1.06	4.92	0.46	0.44	0.04	0.07
221	60	3	25	0	2	2.64	4.18	1.06	4.92	0.46	0.44	0.04	0.07
222	60	8	27	0	2	4.94	5.29	1.06	4.92	0.48	0.46	0.04	0.07
223	44	17	27	0	2	9.13	6.91	1.06	4.92	0.48	0.46	0.04	0.07
224	33	32	23	0	2	17.24	9.05	1.06	4.92	0.44	0.42	0.04	0.07
225	33	50	35	0	2	21.60	14.01	1.06	4.92	0.55	0.53	0.04	0.07
226	28	0	60	0	2	1.46	7.35	1.06	4.92	0.69	0.67	0.04	0.07
227	20	8	60	0	2	3.78	8.94	1.06	4.92	0.69	0.67	0.04	0.07
228	20	18	60	0	2	6.67	10.93	1.06	4.92	0.69	0.67	0.04	0.07
229	0	30	58	0	2	10.33	13.09	1.06	4.92	0.68	0.66	0.04	0.07
230	10	0	0	0.5	3	1.00	1.00	3.01	7.70	0.05	0.00	0.07	0.11
231	0	1	0	0.5	3	1.84	1.13	3.01	7.70	0.05	0.00	0.07	0.11
232	84	3	0	0.5	3	3.51	1.40	3.01	7.70	0.05	0.00	0.07	0.11
233	84	9	0	0.5	3	8.54	2.21	3.01	7.70	0.05	0.00	0.07	0.11
234	68	19	0	0.5	3	16.92	3.56	3.01	7.70	0.05	0.00	0.07	0.11
235	68	30	7	0.5	3	22.02	6.16	3.01	7.70	0.21	0.16	0.07	0.11
236	47	40	0	0.5	3	34.53	6.39	3.01	7.70	0.05	0.00	0.07	0.11
237	47	50	15	0.5	3	30.10	10.55	3.01	7.70	0.35	0.31	0.07	0.11
238	27	60	0	0.5	3	51.29	9.08	3.01	7.70	0.05	0.00	0.07	0.11
239	27	90	0	0.5	3	76.43	13.12	3.01	7.70	0.05	0.00	0.07	0.11
240	77	0	10	0.5	3	1.07	2.02	3.01	7.70	0.27	0.22	0.07	0.11
241	77	2	10	0.5	3	2.37	2.33	3.01	7.70	0.27	0.22	0.07	0.11
242	71	6	10	0.5	3	4.97	2.94	3.01	7.70	0.27	0.22	0.07	0.11
243	71	13	10	0.5	3	9.52	4.01	3.01	7.70	0.27	0.22	0.07	0.11
244	0	71	15.5	0.5	3	41.84	14.02	3.01	7.70	0.36	0.31	0.07	0.11
245	0	0	25	0.5	3	1.19	3.56	3.01	7.70	0.47	0.43	0.07	0.11
246	59	3	25	0.5	3	2.61	4.07	3.01	7.70	0.47	0.43	0.07	0.11
247	59	8	27	0.5	3	4.86	5.15	3.01	7.70	0.49	0.45	0.07	0.11
248	43	17	27	0.5	3	8.98	6.71	3.01	7.70	0.49	0.45	0.07	0.11
249	43	32	23	0.5	3	16.93	8.78	3.01	7.70	0.45	0.41	0.07	0.11
250	1.5	50	35	0.5	3	21.26	13.56	3.01	7.70	0.56	0.52	0.07	0.11
251	1.5	0	60	0.5	3	1.45	7.14	3.01	7.70	0.69	0.67	0.07	0.11
252	19	8	60	0.5	3	3.73	8.68	3.01	7.70	0.69	0.67	0.07	0.11
253	19	18	60	0.5	3	6.59	10.60	3.01	7.70	0.69	0.67	0.07	0.11
254	1	30	56	0.5	3	10.25	12.35	3.01	7.70	0.67	0.65	0.07	0.11
255	84	0	0	2	4	1.00	1.00	8.50	12.18	0.09	0.00	0.12	0.14

256	83	1	0	2	4	1.80	1.13	8.50	12.18	0.09	0.00	0.12	0.14
257	81	3	0	2	4	3.41	1.39	8.50	12.18	0.09	0.00	0.12	0.14
258	75	9	0	2	4	8.22	2.16	8.50	12.18	0.09	0.00	0.12	0.14
259	65	19	0	2	4	16.25	3.45	8.50	12.18	0.09	0.00	0.12	0.14
260	47	30	7	2	4	21.18	5.90	8.50	12.18	0.24	0.16	0.12	0.14
261	44	40	0	2	4	33.10	6.16	8.50	12.18	0.09	0.00	0.12	0.14
262	19	50	15	2	4	29.07	10.02	8.50	12.18	0.37	0.30	0.12	0.14
263	24	60	0	2	4	49.16	8.74	8.50	12.18	0.09	0.00	0.12	0.14
264	0	84	0	2	4	68.42	11.83	8.50	12.18	0.09	0.00	0.12	0.14
265	74	0	10	2	4	1.07	1.97	8.50	12.18	0.30	0.22	0.12	0.14
266	72	2	10	2	4	2.32	2.26	8.50	12.18	0.30	0.22	0.12	0.14
267	68	6	10	2	4	4.82	2.84	8.50	12.18	0.30	0.22	0.12	0.14
268	61	13	10	2	4	9.19	3.85	8.50	12.18	0.30	0.22	0.12	0.14
269	0	70	14	2	4	41.09	12.87	8.50	12.18	0.36	0.28	0.12	0.14
270	59	0	25	2	4	1.18	3.41	8.50	12.18	0.49	0.42	0.12	0.14
271	56	3	25	2	4	2.56	3.90	8.50	12.18	0.49	0.42	0.12	0.14
272	49	8	27	2	4	4.74	4.91	8.50	12.18	0.51	0.44	0.12	0.14
273	40	17	27	2	4	8.74	6.38	8.50	12.18	0.51	0.44	0.12	0.14
274	29	32	23	2	4	16.43	8.33	8.50	12.18	0.47	0.40	0.12	0.14
275	0	49	35	2	4	20.31	12.65	8.50	12.18	0.57	0.51	0.12	0.14
276	24	0	60	2	4	1.42	6.79	8.50	12.18	0.70	0.65	0.12	0.14
277	18	6	60	2	4	3.10	7.87	8.50	12.18	0.70	0.65	0.12	0.14
278	9	15	60	2	4	5.61	9.50	8.50	12.18	0.70	0.65	0.12	0.14
279	0	28	56	2	4	9.59	11.42	8.50	12.18	0.68	0.64	0.12	0.14
280	70	9	0	6	5	7.62	2.06	22.44	21.02	0.17	0.00	0.22	0.18
281	78	1	0	6	5	1.74	1.12	22.44	21.02	0.17	0.00	0.22	0.18
282	76	3	0	6	5	3.21	1.35	22.44	21.02	0.17	0.00	0.22	0.18
283	70	9	0	6	5	7.62	2.06	22.44	21.02	0.17	0.00	0.22	0.18
284	60	19	0	6	5	14.97	3.24	22.44	21.02	0.17	0.00	0.22	0.18
285	42	30	7	6	5	19.63	5.42	22.44	21.02	0.30	0.15	0.22	0.18
286	39	40	0	6	5	30.41	5.73	22.44	21.02	0.17	0.00	0.22	0.18
287	14	50	15	6	5	27.15	9.07	22.44	21.02	0.42	0.28	0.22	0.18
288	19	60	0	6	5	45.11	8.09	22.44	21.02	0.17	0.00	0.22	0.18
289	0	79	0	6	5	59.08	10.33	22.44	21.02	0.17	0.00	0.22	0.18
290	69	0	10	6	5	1.06	1.86	22.44	21.02	0.35	0.20	0.22	0.18
291	67	2	10	6	5	2.22	2.12	22.44	21.02	0.35	0.20	0.22	0.18
292	63	6	10	6	5	4.54	2.64	22.44	21.02	0.35	0.20	0.22	0.18
293	56	13	10	6	5	8.59	3.56	22.44	21.02	0.35	0.20	0.22	0.18
294	0	68	11	6	5	39.59	10.90	22.44	21.02	0.36	0.22	0.22	0.18
295	54	0	25	6	5	1.16	3.14	22.44	21.02	0.52	0.40	0.22	0.18
296	51	3	25	6	5	2.45	3.58	22.44	21.02	0.52	0.40	0.22	0.18
297	44	8	27	6	5	4.51	4.48	22.44	21.02	0.53	0.42	0.22	0.18
298	35	17	27	6	5	8.28	5.78	22.44	21.02	0.53	0.42	0.22	0.18

299	24	32	23	6	5	15.48	7.53	22.44	21.02	0.50	0.38	0.22	0.18
300	0	47	32	6	5	19.33	10.71	22.44	21.02	0.57	0.47	0.22	0.18
301	19	0	60	6	5	1.37	6.15	22.44	21.02	0.71	0.63	0.22	0.18
302	13	6	60	6	5	2.98	7.10	22.44	21.02	0.71	0.63	0.22	0.18
303	4	15	60	6	5	5.39	8.54	22.44	21.02	0.71	0.63	0.22	0.18
304	0	25	54	6	5	8.51	9.57	22.44	21.02	0.69	0.61	0.22	0.18
305	68	1	0	15	6	1.62	1.10	52.40	38.80	0.30	0.00	0.37	0.21
306	66	3	0	15	6	2.87	1.30	52.40	38.80	0.30	0.00	0.37	0.21
307	60	9	0	15	6	6.60	1.90	52.40	38.80	0.30	0.00	0.37	0.21
308	50	19	0	15	6	12.83	2.90	52.40	38.80	0.30	0.00	0.37	0.21
309	32	30	7	15	6	17.05	4.66	52.40	38.80	0.40	0.13	0.37	0.21
310	29	40	0	15	6	25.91	5.00	52.40	38.80	0.30	0.00	0.37	0.21
311	4	50	15	15	6	23.92	7.59	52.40	38.80	0.49	0.25	0.37	0.21
312	9	60	0	15	6	38.36	7.00	52.40	38.80	0.30	0.00	0.37	0.21
313	59	0	10	15	6	1.05	1.69	52.40	38.80	0.43	0.18	0.37	0.21
314	57	2	10	15	6	2.05	1.91	52.40	38.80	0.43	0.18	0.37	0.21
315	53	6	10	15	6	4.06	2.34	52.40	38.80	0.43	0.18	0.37	0.21
316	46	13	10	15	6	7.58	3.09	52.40	38.80	0.43	0.18	0.37	0.21
317	44	0	25	15	6	1.13	2.73	52.40	38.80	0.57	0.37	0.37	0.21
318	41	3	25	15	6	2.28	3.08	52.40	38.80	0.57	0.37	0.37	0.21
319	34	8	27	15	6	4.13	3.81	52.40	38.80	0.58	0.39	0.37	0.21
320	25	17	27	15	6	7.49	4.86	52.40	38.80	0.58	0.39	0.37	0.21
321	9	0	60	15	6	1.30	5.15	52.40	38.80	0.73	0.59	0.37	0.21
322	4	12	53	15	6	4.46	6.18	52.40	38.80	0.71	0.56	0.37	0.21
323	0	20	49	15	6	6.83	6.89	52.40	38.80	0.69	0.54	0.37	0.21
324	18	30	21	15	6	13.45	5.89	52.40	38.80	0.54	0.33	0.37	0.21
325	21	3	45	15	6	2.10	4.48	52.40	38.80	0.68	0.52	0.37	0.21
326	51	1	0	30	8	1.49	1.08	96.43	70.56	0.45	0.00	0.53	0.26
327	49	3	0	30	8	2.46	1.24	96.43	70.56	0.45	0.00	0.53	0.26
328	43	9	0	30	8	5.39	1.71	96.43	70.56	0.45	0.00	0.53	0.26
329	33	19	0	30	8	10.27	2.49	96.43	70.56	0.45	0.00	0.53	0.26
330	15	30	7	30	8	13.89	3.80	96.43	70.56	0.52	0.11	0.53	0.26
331	12	40	0	30	8	20.51	4.14	96.43	70.56	0.45	0.00	0.53	0.26
332	0	52	0	30	8	26.37	5.08	96.43	70.56	0.45	0.00	0.53	0.26
333	42	0	10	30	8	1.04	1.51	96.43	70.56	0.54	0.15	0.53	0.26
334	40	2	10	30	8	1.85	1.68	96.43	70.56	0.54	0.15	0.53	0.26
335	36	6	10	30	8	3.48	2.01	96.43	70.56	0.54	0.15	0.53	0.26
336	29	13	10	30	8	6.33	2.58	96.43	70.56	0.54	0.15	0.53	0.26
337	27	0	25	30	8	1.09	2.28	96.43	70.56	0.64	0.32	0.53	0.26
338	24	3	25	30	8	2.07	2.54	96.43	70.56	0.64	0.32	0.53	0.26
339	17	8	27	30	8	3.63	3.08	96.43	70.56	0.65	0.34	0.53	0.26
340	10	15	27	30	8	5.84	3.70	96.43	70.56	0.65	0.34	0.53	0.26
341	0	0	52	30	8	1.19	3.66	96.43	70.56	0.74	0.50	0.53	0.26

342	0	5	47	30	8	2.42	3.86	96.43	70.56	0.72	0.47	0.53	0.26
343	8	23	21	30	8	8.98	4.05	96.43	70.56	0.61	0.28	0.53	0.26
344	29	0.7	0	51	9	1.26	1.04	#####	####	0.57	0.00	0.66	0.29
345	27	3	0	51	9	2.13	1.18	#####	####	0.57	0.00	0.66	0.29
346	21	9	0	51	9	4.38	1.54	#####	####	0.57	0.00	0.66	0.29
347	11	19	0	51	9	8.15	2.15	#####	####	0.57	0.00	0.66	0.29
348	0	30	0	51	9	12.28	2.81	#####	####	0.57	0.00	0.66	0.29
349	20	0	10	51	9	1.03	1.38	#####	####	0.63	0.13	0.66	0.29
350	19	1.5	10	51	9	1.51	1.47	#####	####	0.63	0.13	0.66	0.29
351	16	5	9	51	9	2.67	1.65	#####	####	0.63	0.12	0.66	0.29
352	6	13	11	51	9	5.19	2.23	#####	####	0.64	0.14	0.66	0.29
353	5	0	25	51	9	1.07	1.94	#####	####	0.70	0.27	0.66	0.29
354	2	3	25	51	9	1.87	2.14	#####	####	0.70	0.27	0.66	0.29
355	0	8	22	51	9	3.29	2.35	#####	####	0.69	0.25	0.66	0.29
356	82	0	0	0	8	1.00	1.00	1.26	16.69	0.12	0.00	0.15	0.26
357	81	1	0	0	8	1.78	1.13	1.26	16.69	0.12	0.00	0.15	0.26
358	79	3	0	0	8	3.34	1.38	1.26	16.69	0.12	0.00	0.15	0.26
359	73	9	0	0	8	8.01	2.13	1.26	16.69	0.12	0.00	0.15	0.26
360	63	19	0	0	8	15.80	3.38	1.26	16.69	0.12	0.00	0.15	0.26
361	45	30	7	0	8	20.63	5.73	1.26	16.69	0.26	0.16	0.15	0.26
362	42	40	0	0	8	32.16	6.01	1.26	16.69	0.12	0.00	0.15	0.26
363	17	50	15	0	8	28.40	9.71	1.26	16.69	0.39	0.29	0.15	0.26
364	22	60	0	0	8	47.75	8.51	1.26	16.69	0.12	0.00	0.15	0.26
365	0	82	0	0	8	64.89	11.27	1.26	16.69	0.12	0.00	0.15	0.26
366	72	0	10	0	8	1.07	1.93	1.26	16.69	0.32	0.21	0.15	0.26
367	70	2	10	0	8	2.29	2.21	1.26	16.69	0.32	0.21	0.15	0.26
368	66	6	10	0	8	4.72	2.77	1.26	16.69	0.32	0.21	0.15	0.26
369	59	13	10	0	8	8.98	3.75	1.26	16.69	0.32	0.21	0.15	0.26
370	0	69	13	0	8	40.39	12.14	1.26	16.69	0.36	0.26	0.15	0.26
371	57	0	25	0	8	1.17	3.33	1.26	16.69	0.50	0.42	0.15	0.26
372	54	3	25	0	8	2.52	3.80	1.26	16.69	0.50	0.42	0.15	0.26
373	47	8	27	0	8	4.66	4.77	1.26	16.69	0.52	0.44	0.15	0.26
374	38	17	27	0	8	8.58	6.19	1.26	16.69	0.52	0.44	0.15	0.26
375	0	47	35	0	8	19.19	11.90	1.26	16.69	0.58	0.51	0.15	0.26
376	22	0	60	0	8	1.41	6.59	1.26	16.69	0.70	0.65	0.15	0.26
377	14	8	60	0	8	3.61	7.98	1.26	16.69	0.70	0.65	0.15	0.26
378	4	18	60	0	8	6.37	9.72	1.26	16.69	0.70	0.65	0.15	0.26
379	0	30	52	0	8	10.43	10.97	1.26	16.69	0.67	0.61	0.15	0.26
380	80	0	0	1.5	9	1.00	1.00	5.88	21.38	0.16	0.00	0.20	0.29
381	79	1	0	1.5	9	1.75	1.12	5.88	21.38	0.16	0.00	0.20	0.29
382	77	3	0	1.5	9	3.24	1.36	5.88	21.38	0.16	0.00	0.20	0.29
383	71	9	0	1.5	9	7.72	2.08	5.88	21.38	0.16	0.00	0.20	0.29
384	61	19	0	1.5	9	15.18	3.28	5.88	21.38	0.16	0.00	0.20	0.29

385	43	30	7	1.5	9	19.88	5.50	5.88	21.38	0.29	0.15	0.20	0.29
386	40	40	0	1.5	9	30.85	5.80	5.88	21.38	0.16	0.00	0.20	0.29
387	15	50	15	1.5	9	27.46	9.25	5.88	21.38	0.41	0.28	0.20	0.29
388	0	80	0	1.5	9	60.33	10.53	5.88	21.38	0.16	0.00	0.20	0.29
389	70	0	10	1.5	9	1.06	1.88	5.88	21.38	0.34	0.21	0.20	0.29
390	68	2	10	1.5	9	2.24	2.15	5.88	21.38	0.34	0.21	0.20	0.29
391	64	6	10	1.5	9	4.58	2.68	5.88	21.38	0.34	0.21	0.20	0.29
392	57	13	10	1.5	9	8.69	3.61	5.88	21.38	0.34	0.21	0.20	0.29
393	0	70	9.5	1.5	9	42.57	11.12	5.88	21.38	0.33	0.20	0.20	0.29
394	55	0	25	1.5	9	1.16	3.20	5.88	21.38	0.51	0.41	0.20	0.29
395	52	3	25	1.5	9	2.47	3.64	5.88	21.38	0.51	0.41	0.20	0.29
396	45	8	27	1.5	9	4.55	4.56	5.88	21.38	0.53	0.43	0.20	0.29
397	36	17	27	1.5	9	8.35	5.90	5.88	21.38	0.53	0.43	0.20	0.29
398	0	45	34.5	1.5	9	18.09	10.92	5.88	21.38	0.58	0.49	0.20	0.29
399	20	0	60	1.5	9	1.38	6.28	5.88	21.38	0.70	0.64	0.20	0.29
400	12	8	60	1.5	9	3.54	7.59	5.88	21.38	0.70	0.64	0.20	0.29
401	1.5	18	60	1.5	9	6.24	9.22	5.88	21.38	0.70	0.64	0.20	0.29
402	0	30	50	1.5	9	10.27	10.12	5.88	21.38	0.67	0.59	0.20	0.29
403	72	0	0	8	10	1.00	1.00	24.99	35.40	0.26	0.00	0.32	0.31
404	71	1	0	8	10	1.66	1.11	24.99	35.40	0.26	0.00	0.32	0.31
405	69	3	0	8	10	2.97	1.32	24.99	35.40	0.26	0.00	0.32	0.31
406	63	7	2	8	10	5.40	1.90	24.99	35.40	0.29	0.04	0.32	0.31
407	60	12	0	8	10	8.89	2.27	24.99	35.40	0.26	0.00	0.32	0.31
408	46	20	6	8	10	12.48	3.68	24.99	35.40	0.35	0.12	0.32	0.31
409	42	30	0	8	10	20.72	4.17	24.99	35.40	0.26	0.00	0.32	0.31
410	20	45	7	8	10	26.25	6.58	24.99	35.40	0.37	0.14	0.32	0.31
411	59	0	13	8	10	1.07	1.97	24.99	35.40	0.44	0.24	0.32	0.31
412	60	2	10	8	10	2.11	1.97	24.99	35.40	0.41	0.19	0.32	0.31
413	56	5	11	8	10	3.64	2.40	24.99	35.40	0.42	0.21	0.32	0.31
414	44	12	16	8	10	6.70	3.62	24.99	35.40	0.47	0.28	0.32	0.31
415	24	26	22	8	10	12.06	5.84	24.99	35.40	0.53	0.35	0.32	0.31
416	41	1	30	8	10	1.53	3.36	24.99	35.40	0.59	0.42	0.32	0.31
417	38	4	30	8	10	2.64	3.74	24.99	35.40	0.59	0.42	0.32	0.31
418	35	8	29	8	10	4.17	4.18	24.99	35.40	0.58	0.42	0.32	0.31
419	20	16	36	8	10	6.63	5.76	24.99	35.40	0.62	0.47	0.32	0.31
420	0	39	33	8	10	15.00	8.48	24.99	35.40	0.61	0.45	0.32	0.31
421	7	0	65	8	10	1.35	5.84	24.99	35.40	0.74	0.63	0.32	0.31
422	64	1	0	11	14	1.59	1.10	30.03	49.79	0.33	0.00	0.41	0.40
423	62	3	0	11	14	2.77	1.29	30.03	49.79	0.33	0.00	0.41	0.40
424	56	9	0	11	14	6.32	1.86	30.03	49.79	0.33	0.00	0.41	0.40
425	46	19	0	11	14	12.24	2.81	30.03	49.79	0.33	0.00	0.41	0.40
426	28	30	7	11	14	16.32	4.46	30.03	49.79	0.43	0.13	0.41	0.40
427	25	40	0	11	14	24.66	4.80	30.03	49.79	0.33	0.00	0.41	0.40

428	0	50	15	11	14	23.00	7.23	30.03	49.79	0.51	0.25	0.41	0.40
429	5	60	0	11	14	36.48	6.70	30.03	49.79	0.33	0.00	0.41	0.40
430	55	0	10	11	14	1.05	1.65	30.03	49.79	0.46	0.18	0.41	0.40
431	53	2	10	11	14	2.01	1.86	30.03	49.79	0.46	0.18	0.41	0.40
432	49	6	10	11	14	3.93	2.27	30.03	49.79	0.46	0.18	0.41	0.40
433	42	13	10	11	14	7.30	2.98	30.03	49.79	0.46	0.18	0.41	0.40
434	40	0	25	11	14	1.12	2.63	30.03	49.79	0.58	0.36	0.41	0.40
435	37	3	25	11	14	2.24	2.96	30.03	49.79	0.58	0.36	0.41	0.40
436	30	8	27	11	14	4.02	3.65	30.03	49.79	0.60	0.37	0.41	0.40
437	21	17	27	11	14	7.27	4.64	30.03	49.79	0.60	0.37	0.41	0.40
438	10	32	23	11	14	13.39	5.99	30.03	49.79	0.57	0.34	0.41	0.40
439	5	0	60	11	14	1.29	4.92	30.03	49.79	0.73	0.58	0.41	0.40
440	0	8	57	11	14	3.26	5.68	30.03	49.79	0.73	0.57	0.41	0.40
441	6	15	44	11	14	5.53	5.62	30.03	49.79	0.68	0.50	0.41	0.40
442	50	1	0	23	16	1.49	1.08	57.78	77.85	0.45	0.00	0.53	0.43
443	48	3	0	23	16	2.46	1.23	57.78	77.85	0.45	0.00	0.53	0.43
444	42	9	0	23	16	5.37	1.70	57.78	77.85	0.45	0.00	0.53	0.43
445	32	19	0	23	16	10.23	2.48	57.78	77.85	0.45	0.00	0.53	0.43
446	14	30	7	23	16	13.84	3.79	57.78	77.85	0.52	0.11	0.53	0.43
447	11	40	0	23	16	20.43	4.12	57.78	77.85	0.45	0.00	0.53	0.43
448	0	51	0	23	16	25.78	4.98	57.78	77.85	0.45	0.00	0.53	0.43
449	41	0	10	23	16	1.04	1.51	57.78	77.85	0.54	0.15	0.53	0.43
450	39	2	10	23	16	1.85	1.68	57.78	77.85	0.54	0.15	0.53	0.43
451	35	6	10	23	16	3.47	2.01	57.78	77.85	0.54	0.15	0.53	0.43
452	28	13	10	23	16	6.31	2.58	57.78	77.85	0.54	0.15	0.53	0.43
453	26	0	25	23	16	1.09	2.28	57.78	77.85	0.64	0.32	0.53	0.43
454	23	3	25	23	16	2.07	2.54	57.78	77.85	0.64	0.32	0.53	0.43
455	16	8	27	23	16	3.62	3.09	57.78	77.85	0.65	0.33	0.53	0.43
456	0	0	51	23	16	1.19	3.62	57.78	77.85	0.73	0.49	0.53	0.43
457	0	3	48	23	16	1.92	3.74	57.78	77.85	0.73	0.48	0.53	0.43
458	0	13	38	23	16	4.72	4.12	57.78	77.85	0.69	0.42	0.53	0.43
459	0	24	27	23	16	8.67	4.49	57.78	77.85	0.65	0.33	0.53	0.43
460	29	1	0	40	20	1.38	1.06	89.12	####	0.57	0.00	0.65	0.50
461	27	3	0	40	20	2.14	1.18	89.12	####	0.57	0.00	0.65	0.50
462	21	9	0	40	20	4.42	1.55	89.12	####	0.57	0.00	0.65	0.50
463	11	19	0	40	20	8.22	2.16	89.12	####	0.57	0.00	0.65	0.50
464	0	30	0	40	20	12.40	2.83	89.12	####	0.57	0.00	0.65	0.50
465	20	0	10	40	20	1.03	1.38	89.12	####	0.63	0.13	0.65	0.50
466	18	2	10	40	20	1.68	1.51	89.12	####	0.63	0.13	0.65	0.50
467	17	6	7	40	20	3.07	1.64	89.12	####	0.61	0.09	0.65	0.50
468	0	13	17	40	20	4.93	2.49	89.12	####	0.66	0.20	0.65	0.50
469	5	0	25	40	20	1.07	1.96	89.12	####	0.70	0.27	0.65	0.50
470	2	3	25	40	20	1.88	2.16	89.12	####	0.70	0.27	0.65	0.50

471	0	8	22	40	20	3.30	2.37	89.12	####	0.68	0.25	0.65	0.50
472	66	0	0	3.5	21	1.00	1.00	9.13	49.38	0.32	0.00	0.39	0.51
473	65	1	0	3.5	21	1.60	1.10	9.13	49.38	0.32	0.00	0.39	0.51
474	63	3	0	3.5	21	2.80	1.29	9.13	49.38	0.32	0.00	0.39	0.51
475	57	9	0	3.5	21	6.41	1.87	9.13	49.38	0.32	0.00	0.39	0.51
476	47	19	0	3.5	21	12.43	2.84	9.13	49.38	0.32	0.00	0.39	0.51
477	29	30	7	3.5	21	16.56	4.54	9.13	49.38	0.42	0.13	0.39	0.51
478	26	40	0	3.5	21	25.06	4.87	9.13	49.38	0.32	0.00	0.39	0.51
479	0.5	50	15	3.5	21	23.30	7.38	9.13	49.38	0.50	0.25	0.39	0.51
480	5.5	60	0	3.5	21	37.09	6.80	9.13	49.38	0.32	0.00	0.39	0.51
481	56	0	10	3.5	21	1.05	1.67	9.13	49.38	0.45	0.18	0.39	0.51
482	54	2	10	3.5	21	2.02	1.88	9.13	49.38	0.45	0.18	0.39	0.51
483	50	6	10	3.5	21	3.98	2.30	9.13	49.38	0.45	0.18	0.39	0.51
484	43	13	10	3.5	21	7.39	3.03	9.13	49.38	0.45	0.18	0.39	0.51
485	41	0	25	3.5	21	1.12	2.68	9.13	49.38	0.58	0.36	0.39	0.51
486	38	3	25	3.5	21	2.25	3.02	9.13	49.38	0.58	0.36	0.39	0.51
487	31	8	27	3.5	21	4.06	3.72	9.13	49.38	0.59	0.38	0.39	0.51
488	25	14	27	3.5	21	6.25	4.41	9.13	49.38	0.59	0.38	0.39	0.51
489	11	32	23	3.5	21	13.55	6.12	9.13	49.38	0.56	0.34	0.39	0.51
490	5.5	0	60	3.5	21	1.29	5.04	9.13	49.38	0.73	0.58	0.39	0.51
491	1.5	4	60	3.5	21	2.27	5.53	9.13	49.38	0.73	0.58	0.39	0.51
492	0	8	57.5	3.5	21	3.28	5.85	9.13	49.38	0.72	0.57	0.39	0.51
493	0	24	41.5	3.5	21	8.40	6.64	9.13	49.38	0.67	0.49	0.39	0.51
494	66	0	0	3.5	21	1.00	1.00	9.13	49.38	0.32	0.00	0.39	0.51
495	65	1	0	3.5	21	1.60	1.10	9.13	49.38	0.32	0.00	0.39	0.51
496	63	3	0	3.5	21	2.80	1.29	9.13	49.38	0.32	0.00	0.39	0.51
497	57	9	0	3.5	21	6.41	1.87	9.13	49.38	0.32	0.00	0.39	0.51
498	47	19	0	3.5	21	12.43	2.84	9.13	49.38	0.32	0.00	0.39	0.51
499	29	30	7	3.5	21	16.56	4.54	9.13	49.38	0.42	0.13	0.39	0.51
500	26	40	0	3.5	21	25.06	4.87	9.13	49.38	0.32	0.00	0.39	0.51
501	0.5	50	15	3.5	21	23.30	7.38	9.13	49.38	0.50	0.25	0.39	0.51
502	5.5	60	0	3.5	21	37.09	6.80	9.13	49.38	0.32	0.00	0.39	0.51
503	56	0	10	3.5	21	1.05	1.67	9.13	49.38	0.45	0.18	0.39	0.51
504	54	2	10	3.5	21	2.02	1.88	9.13	49.38	0.45	0.18	0.39	0.51
505	50	6	10	3.5	21	3.98	2.30	9.13	49.38	0.45	0.18	0.39	0.51
506	43	13	10	3.5	21	7.39	3.03	9.13	49.38	0.45	0.18	0.39	0.51
507	41	0	25	3.5	21	1.12	2.68	9.13	49.38	0.58	0.36	0.39	0.51
508	38	3	25	3.5	21	2.25	3.02	9.13	49.38	0.58	0.36	0.39	0.51
509	31	8	27	3.5	21	4.06	3.72	9.13	49.38	0.59	0.38	0.39	0.51
510	25	14	27	3.5	21	6.25	4.41	9.13	49.38	0.59	0.38	0.39	0.51
511	11	32	23	3.5	21	13.55	6.12	9.13	49.38	0.56	0.34	0.39	0.51
512	5.5	0	60	3.5	21	1.29	5.04	9.13	49.38	0.73	0.58	0.39	0.51
513	1.5	4	60	3.5	21	2.27	5.53	9.13	49.38	0.73	0.58	0.39	0.51

514	0	8	57.5	3.5	21	3.28	5.85	9.13	49.38	0.72	0.57	0.39	0.51
515	0	24	41.5	3.5	21	8.40	6.64	9.13	49.38	0.67	0.49	0.39	0.51
516	56	0	0	9	25	1.00	1.00	19.15	68.99	0.41	0.00	0.48	0.56
517	55	1	0	9	25	1.53	1.08	19.15	68.99	0.41	0.00	0.48	0.56
518	53	3	0	9	25	2.58	1.25	19.15	68.99	0.41	0.00	0.48	0.56
519	50	6	0	9	25	4.15	1.51	19.15	68.99	0.41	0.00	0.48	0.56
520	39	17	0	9	25	9.93	2.44	19.15	68.99	0.41	0.00	0.48	0.56
521	46	0	10	9	25	1.04	1.57	19.15	68.99	0.51	0.16	0.48	0.56
522	44	2	10	9	25	1.91	1.75	19.15	68.99	0.51	0.16	0.48	0.56
523	41	5	10	9	25	3.21	2.02	19.15	68.99	0.51	0.16	0.48	0.56
524	33	13	10	9	25	6.69	2.74	19.15	68.99	0.51	0.16	0.48	0.56
525	31	0	25	9	25	1.10	2.43	19.15	68.99	0.62	0.33	0.48	0.56
526	28	3	25	9	25	2.13	2.71	19.15	68.99	0.62	0.33	0.48	0.56
527	21	8	27	9	25	3.78	3.31	19.15	68.99	0.63	0.35	0.48	0.56
528	22	27	7	9	25	13.41	3.78	19.15	68.99	0.48	0.12	0.48	0.56
529	0	42	14	9	25	18.10	5.64	19.15	68.99	0.54	0.22	0.48	0.56
530	0	0	56	9	25	1.23	4.19	19.15	68.99	0.74	0.53	0.48	0.56
531	43	8	5	9	25	4.83	1.98	19.15	68.99	0.46	0.09	0.48	0.56
532	0	16	40	9	25	5.68	4.88	19.15	68.99	0.69	0.45	0.48	0.56
533	0	24	32	9	25	8.61	5.17	19.15	68.99	0.65	0.39	0.48	0.56
534	0	56	0	9	25	30.43	5.73	19.15	68.99	0.41	0.00	0.48	0.56
535	0	3	53	9	25	1.95	4.33	19.15	68.99	0.73	0.52	0.48	0.56
536	18	38	0	9	25	20.97	4.21	19.15	68.99	0.41	0.00	0.48	0.56
537	37	0	0	22	31	1.00	1.00	39.02	####	0.53	0.00	0.61	0.61
538	36	1	0	22	31	1.42	1.07	39.02	####	0.53	0.00	0.61	0.61
539	34	3	0	22	31	2.25	1.20	39.02	####	0.53	0.00	0.61	0.61
540	31	6	0	22	31	3.49	1.40	39.02	####	0.53	0.00	0.61	0.61
541	27	0	10	22	31	1.03	1.43	39.02	####	0.60	0.14	0.61	0.61
542	25	2	10	22	31	1.74	1.57	39.02	####	0.60	0.14	0.61	0.61
543	19	6	12	22	31	3.10	1.94	39.02	####	0.61	0.16	0.61	0.61
544	16	10	11	22	31	4.52	2.17	39.02	####	0.61	0.15	0.61	0.61
545	12	0	25	22	31	1.08	2.08	39.02	####	0.68	0.29	0.61	0.61
546	9	3	25	22	31	1.95	2.30	39.02	####	0.68	0.29	0.61	0.61
547	2	8	27	22	31	3.34	2.75	39.02	####	0.68	0.30	0.61	0.61
548	3	27	7	22	31	11.03	3.16	39.02	####	0.58	0.10	0.61	0.61
549	22	15	0	22	31	7.23	2.00	39.02	####	0.53	0.00	0.61	0.61
550	0	16	21	22	31	5.93	3.06	39.02	####	0.66	0.25	0.61	0.61
551	0	37	0	22	31	16.36	3.47	39.02	####	0.53	0.00	0.61	0.61
552	35	0	0	0	55	1.00	1.00	2.77	####	0.53	0.00	0.60	0.75
553	34	1	0	0	55	1.42	1.07	2.77	####	0.53	0.00	0.60	0.75
554	32	3	0	0	55	2.25	1.20	2.77	####	0.53	0.00	0.60	0.75
555	29	6	0	0	55	3.49	1.40	2.77	####	0.53	0.00	0.60	0.75
556	24	11	0	0	55	5.57	1.73	2.77	####	0.53	0.00	0.60	0.75

557	17	18	0	0	55	8.48	2.20	2.77	####	0.53	0.00	0.60	0.75
558	0	35	0	0	55	15.55	3.34	2.77	####	0.53	0.00	0.60	0.75
559	20	0	15	0	55	1.05	1.66	2.77	####	0.63	0.19	0.60	0.75
560	18	2	15	0	55	1.71	1.80	2.77	####	0.63	0.19	0.60	0.75
561	12	7	16	0	55	3.33	2.20	2.77	####	0.63	0.20	0.60	0.75
562	0	13	22	0	55	4.98	2.92	2.77	####	0.66	0.26	0.60	0.75
563	30	0	0	5	55	1.00	1.00	8.31	####	0.56	0.00	0.63	0.75
564	29	1	0	5	55	1.39	1.06	8.31	####	0.56	0.00	0.63	0.75
565	27	3	0	5	55	2.18	1.19	8.31	####	0.56	0.00	0.63	0.75
566	24	6	0	5	55	3.36	1.38	8.31	####	0.56	0.00	0.63	0.75
567	19	11	0	5	55	5.32	1.69	8.31	####	0.56	0.00	0.63	0.75
568	12	18	0	5	55	8.07	2.14	8.31	####	0.56	0.00	0.63	0.75
569	0	30	0	5	55	12.78	2.89	8.31	####	0.56	0.00	0.63	0.75
570	15	0	15	5	55	1.04	1.62	8.31	####	0.65	0.19	0.63	0.75
571	13	2	15	5	55	1.67	1.75	8.31	####	0.65	0.19	0.63	0.75
572	7	7	16	5	55	3.22	2.13	8.31	####	0.65	0.20	0.63	0.75
573	30	0	0	15	45	1.00	1.00	21.85	####	0.56	0.00	0.63	0.71
574	29	1	0	15	45	1.39	1.06	21.85	####	0.56	0.00	0.63	0.71
575	27	3	0	15	45	2.17	1.19	21.85	####	0.56	0.00	0.63	0.71
576	24	6	0	15	45	3.33	1.38	21.85	####	0.56	0.00	0.63	0.71
577	19	11	0	15	45	5.28	1.69	21.85	####	0.56	0.00	0.63	0.71
578	15	0	15	15	45	1.04	1.60	21.85	####	0.65	0.19	0.63	0.71
579	13	2	15	15	45	1.67	1.74	21.85	####	0.65	0.19	0.63	0.71
580	9	6	15	15	45	2.92	2.00	21.85	####	0.65	0.19	0.63	0.71
581	2	13	15	15	45	5.10	2.46	21.85	####	0.65	0.19	0.63	0.71

Appendix B

SUPPLEMENTARY TABLE 3.4

Barcodes used for characterization in Figs. 2b, 3(b-c), 4(c-d-e)

	Starting stoichiometries in pmoles				
BC#	LO_0	LO_1	LO_2	LO_3	LO_4
1	90	0	0	0	0
2	89.5	0	0	0.5	0
3	88	0	0	2	0
4	85	0	0	5	0
5	80	0	0	10	0
6	70	0	0	20	0
7	50	0	0	40	0
8	25	0	0	65	0
9	88	0	0	0	2
10	86.5	0	0	0.5	3
11	84	0	0	2	4
12	79	0	0	6	5
13	69	0	0	15	6
14	52	0	0	30	8
15	30	0	0	51	9
16	82	0	0	0	8
17	79.5	0	0	1.5	9
18	72	0	0	8	10
19	65	0	0	11	14
20	51	0	0	23	16
21	30	0	0	40	20
22	70	0	0	0	20
23	65.5	0	0	3.5	21
24	56	0	0	9	25
25	37	0	0	22	31
26	35	0	0	0	55
27	30	0	0	5	55
28	30	0	0	15	45
29	89	1	0	0	0
30	87	3	0	0	0
31	81	9	0	0	0
32	71	19	0	0	0
33	53	30	7	0	0
34	50	40	0	0	0
35	25	50	15	0	0
36	30	60	0	0	0
37	0	90	0	0	0

38	80	0	10	0	0
39	78	2	10	0	0
40	74	6	10	0	0
41	67	13	10	0	0
42	0	74	16	0	0
43	65	0	25	0	0
44	62	3	25	0	0
45	55	8	27	0	0
46	46	17	27	0	0
47	35	32	23	0	0
48	5	50	35	0	0
49	30	0	60	0	0
50	22	8	60	0	0
51	12	18	60	0	0
52	0	30	60	0	0
53	0	2	88	0	0
403	72	0	0	8	10
404	71	1	0	8	10
405	69	3	0	8	10
406	63	7	2	8	10
407	60	12	0	8	10
408	46	20	6	8	10
409	42	30	0	8	10
410	20	45	7	8	10
411	59	0	13	8	10
412	60	2	10	8	10
413	56	5	11	8	10
414	44	12	16	8	10
415	24	26	22	8	10
416	41	1	30	8	10
417	38	4	30	8	10
418	35	8	29	8	10
419	20	16	36	8	10
420	0	39	33	8	10
421	7	0	65	8	10

Barcodes used for independent dyes in Fig. 2c

	Starting stoichiometries in pmoles				
BC#	LO_0	LO_1	LO_2	LO_3	LO_4
403	72	0	0	8	10
404	71	1	0	8	10
405	69	3	0	8	10
406	63	7	2	8	10
407	60	12	0	8	10
408	46	20	6	8	10
409	42	30	0	8	10
410	20	45	7	8	10
411	59	0	13	8	10
412	60	2	10	8	10
413	56	5	11	8	10
414	44	12	16	8	10
415	24	26	22	8	10
416	41	1	30	8	10
417	38	4	30	8	10
418	35	8	29	8	10
419	20	16	36	8	10
420	0	39	33	8	10
421	7	0	65	8	10

Barcodes used forautomated decoding dyes in Fig.5

BC#	Starting stoichiometries in pmoles				
	LO_0	LO_1	LO_2	LO_3	LO_4
29	89	1	0	0	0
30	87	3	0	0	0
31	81	9	0	0	0
32	71	19	0	0	0
33	53	30	7	0	0
34	50	40	0	0	0
35	25	50	15	0	0
36	30	60	0	0	0
37	0	90	0	0	0
38	80	0	10	0	0
39	78	2	10	0	0
40	74	6	10	0	0
41	67	13	10	0	0
42	0	74	16	0	0
43	65	0	25	0	0
44	62	3	25	0	0
45	55	8	27	0	0
46	46	17	27	0	0
47	35	32	23	0	0
48	5	50	35	0	0
49	30	0	60	0	0
50	22	8	60	0	0
51	12	18	60	0	0
52	0	30	60	0	0
53	0	2	88	0	0
403	72	0	0	8	10
404	71	1	0	8	10
405	69	3	0	8	10
406	63	7	2	8	10
407	60	12	0	8	10
408	46	20	6	8	10
409	42	30	0	8	10
410	20	45	7	8	10
411	59	0	13	8	10
412	60	2	10	8	10
413	56	5	11	8	10
414	44	12	16	8	10

415	24	26	22	8	10
416	41	1	30	8	10
417	38	4	30	8	10
418	35	8	29	8	10
419	20	16	36	8	10
420	0	39	33	8	10
421	7	0	65	8	10

Appendix C

SUPPLEMENTARY TABLE 3.5

Repertoire #	Capture #	LO0 n0	LO1 n1	LO2 n2	LO3 n3	LO4 n4
0	1	90	0	0	0	0
5	2	80	0	0	10	0
11	3	84	0	0	2	4
12	4	79	0	0	6	5
16	5	82	0	0	0	8
17	6	79.5	0	0	1.5	9
19	7	65	0	0	11	14
29	8	89	1	0	0	0
31	9	81	9	0	0	0
32	10	71	19	0	0	0
33	11	53	30	7	0	0
38	12	80	0	10	0	0
41	13	67	13	10	0	0
83	14	79	9	0	2	0
84	15	69	19	0	2	0
85	16	51	30	7	2	0
90	17	78	0	10	2	0
92	18	72	6	10	2	0
133	19	79	1	0	10	0
135	20	71	9	0	10	0
137	21	43	30	7	10	0
144	22	70	0	10	10	0
145	23	57	13	10	10	0
256	24	83	1	0	2	4
258	25	75	9	0	2	4
259	26	65	19	0	2	4
260	27	47	30	7	2	4
265	28	74	0	10	2	4
267	29	68	6	10	2	4
268	30	61	13	10	2	4
283	31	69	10	0	6	5
284	32	60	19	0	6	5
285	33	42	30	7	6	5
290	34	69	0	10	6	5
292	35	63	6	10	6	5

Target	capture		
	Supplier	Catalog	Lot
EGF-R	RnD	AUD081104	BAF231
AHSG	RnD	BAF1184	HIF0312091
Amphiregulin	RnD	BAF262	BER051405
b-NGF	RnD	BAF256	AUL051102
c-kit	RnD	BAF332	CAJ0813111
CXCL12	RnD	BAF310	AQL1512121
E-Cadherin	RnD	BAF648	1
EGF	RnD	BAF236	ARA03
EpCAM	RnD	BAF960	FOZ1513111
FAS	RnD	BAF326	1
FAS-L	RnD	BAF126	HEA0312121
Flt-3	RnD	BAF308	ABV0414071
HER2	RnD	BAF1129	ING081305
HGF-R	RnD	BAF358	CLO1113021
IGFBP-1	RnD	BAF871	DOF051203
IGFBP-3	RnD	BAF675	1
IGFBP-7	RnD	BAF1334	HZC011012
IL-12	RnD	BAF219	WR2013021
IL-1b	RnD	BAF-201	UB16
IL-1ra	RnD	BAF280	XK1314081
IL-3	RnD	BAF203	ABZ0412041
IP-10	RnD	BAF266	AAP081011
MCP-1	RnD	BAF279	AKY121207
MCP-2	RnD	BAF281	ANM04140
MCP-3	RnD	BAF282	AUP031201
OPN	RnD	BAF1433	JCY0512071
PSA	RnD		
TGF-b1	RnD	BAF240	WS3214021
TGF-b2	RnD	BAF302	ABB0814032
THBS-1	RnD	BAF3074	81
TNF-R1	RnD	BAF225	BGG061305
uPA	RnD	BAF1310	IRK01
uPA-R	RnD	ALH1313061	100/CF
VCAM-1	RnD	BAF809	DML061303
VEGF-R2	RnD	BAF357	

Target	Antigen Supplier	Catalog	Lot
EGF-R	RnD	1095-ER	
AHSG	RnD	1184-PI-050	1
Amphireguli	RnD	262-AR-	WY1014031
b-NGF	RnD	256-GF	HS3108081
c-kit	RnD	050/CF	1
CXCL12	RnD	350-NS-010/CF	AFV1914061
E-Cadherin	RnD	648-EC-100	1
EGF	RnD	236-EG	ARA051407
EpCAM	RnD	960-EP-050	EHH101304
FAS	RnD	326-FS	AFA011202
FAS-L	RnD	126-FL-	CUI3713021
Flt-3	RnD	005/CF	ZO1213111
HER2	RnD	1129-ER	
HGF-R	RnD	385-MT-	ASR081305
IGFBP-1	RnD	871-B1-025	1
IGFBP-3	RnD	675-B3-025	BUY071301
IGFBP-7	RnD	1334-B7-	IOE0512071
IL-12	RnD	219-IL	PY1414101
IL-1b	RnD	201-LB	
IL-1ra	RnD	280-RA-010	X0112091
IL-3	RnD	203-IL	AF1414041
IP-10	RnD	266-IP	KJ9714081
MCP-1	RnD	279-MC-	EZ290941
MCP-2	RnD	281-CP/CF	
MCP-3	RnD	282-P3-	NA0412012
OPN	RnD	1433-OP	
PSA	RnD		
TGF-b1	RnD	002/CF	AV6614032
TGF-b2	RnD	002/CF	KF1314031
THBS-1	RnD	3074-TH-050	1
TNF-R1	RnD	636-R1-	1
uPA	RnD	1310-SE	HKY011109
uPA-R	RnD	100/CF	ALH1313061
VCAM-1	RnD	809-VR-050	BEJ0513081
VEGF-R2	RnD	357-KD-050	

Target

	Detection Supplier	Catalog	Lot
EGF-R	RnD	AF231	AUC09
AHSG	RnD	MAB11841	CANU02
Amphireguli	RnD	MAB262	BAN0312041
b-NGF	RnD	MAB256	KV0711031
c-kit	RnD	MAB332	IIP0312071
CXCL12	RnD	MAB350	COJ0514022
E-Cadherin	RnD	648-EC-100	1
EGF	RnD	MAB636	AWQ07140
EpCAM	RnD	MAB9601	UTT121307
FAS	RnD	MAB144	BHC011012
FAS-L	RnD		
Flt-3	RnD	MAB608	BBI0113121
HER2	RnD	MAB1129	IBD03
HGF-R	RnD	MAB3581	KXO021431
IGFBP-1	RnD	MAB675	CUP051301
IGFBP-3	RnD	MAB305	1
IGFBP-7	RnD	MAB1334	IHZ0311011
IL-12	RnD	MAB611	LC3014101
IL-1b	RnD	MAB601	GY3111031
IL-1ra	RnD		
IL-3	RnD	MAB603	1
IP-10	RnD	MAB266	ADN031103
MCP-1	RnD	MAB279	BCM03
MCP-2	RnD	MAB281	
MCP-3	RnD	MAB282	F0112071
OPN	RnD	MAB14332	JWE031306
PSA	RnD		
TGF-b1	RnD	MAB240	TS1113121
TGF-b2	RnD	MAB612	
THBS-1	RnD	MAB3074	VYC0212031
TNF-R1	RnD	BAF225	BGG061305
uPA	RnD	MAB1310	JXU0212121
uPA-R	RnD	MAB807	CEP0112111
VCAM-1	RnD	MAB809	BPO051304
VEGF-R2	RnD	MAB3573	