

TOBIAS HUNDERTMARK

VALIDATION OF A DPCR-BASED DETECTION
PRINCIPLE FOR PROTEIN INTERACTION
COUPLING

VALIDATION OF A DPCR-BASED DETECTION
PRINCIPLE FOR PROTEIN INTERACTION
COUPLING

universität freiburg

Submitted in fulfillment of the requirements for the degree of
Doctor rerum naturalium
at the Faculty of Engineering of the University of Freiburg

TOBIAS HUNDERTMARK
(NÉ LANGE)



Laboratory for MEMS Applications
Department of Microsystems Engineering - IMTEK
Faculty of Engineering
University of Freiburg

DEAN:

Prof. Dr. Frank Balle

SUPERVISORS:

Prof. Dr. Roland Zengerle
Dr. Can Dincer

LOCATION:

Freiburg im Breisgau, Germany

DATE:

July 2024

Tobias Hundertmark: *Validation of a dPCR-based detection principle for Protein Interaction Coupling*, ©, July 2024

Ich arbeite ausschliiiiiießlich mit dem $G_E h^i R_n$.

— Sido & Helge Schneider

Dedicated to the grasshopper.

2024

DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources, and that I have explicitly marked all material, which has been quoted either literally or by content from the used sources.

Freiburg im Breisgau, Germany, July 2024



Tobias Hundertmark

PREFACE

The research presented in this thesis would not have been possible without the pioneering work of Dr. Csaba Jeney. His inventions are the foundations for the advancements described herein.

In 2016, Dr. Jeney filed the first basic patent for Protein Interaction Coupling (PICO) technology (formerly known as Emulsion Coupling). PICO is a homogeneous immunoassay with digital readout. It was thus soon clear that PICO holds great potential in the fields of proteomics and interactomics. Consequently, Dr. Jeney co-founded [Actome](#) with serial entrepreneur Dr. Peter Koltay. Their collaboration marked a pivotal moment in the commercialization and further development of the PICO technology.

After acquiring two publicly funded projects (MONOGRAM, MET-ID 55, Baden-Württemberg Stiftung and DINAMIK, 7533-7-11.10-6, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg), two PhD students, Tobias Gross and myself, could be hired. This opened up new possibilities for the practical implementation of the PICO technology as well as to prove its applicability to various research questions. A strategic partnership with QIAGEN enabled the use of the QIAcuity, a chamber-based dPCR system, as the standard dPCR system for PICO.

Through dedicated research and development, Actome has successfully transitioned the PICO technology from a novel concept into a commercially viable product. The company now offers specialized kits and reagents for PICO, making this advanced technology accessible to researchers worldwide.

In principle, PICO can be divided into two parts. The first part deals with the calculation of the basic detection unit – complexes – from dPCR data. The second part uses this information for absolute quantification of proteoforms or protein-protein-interactions.

Within this thesis, I present my contribution to the advancement of the PICO technology. My focus was on the development of a mathematical model for the calculation of the number of complexes. I used simulated and experimental data for the validation of the model and concluded my thesis with an application example for the simultaneous detection of a protein and one of its phosphorylations. On the other hand, Tobias Gross focused on the second part of the technology and demonstrated how to turn the number of complexes into absolute values. Together, we co-authored a manuscript entitled “Lossless Single-Molecule Counting To Absolute Quantify Proteoforms”, which is currently available from [bioRxiv](#) (Gross et al., 2024) and in a submission process.

The collaborative efforts and innovative spirit between Actome, Tobias Gross and me, brought PICO to where it is today and yet this is not the end of this journey. There is more to come, paving the way for new possibilities in the field of precision medicine.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Prof. Dr. Roland Zengerle for financial support, provision of laboratory infrastructure and supervision of this thesis. Second, I would like to thank Dr. Can Dincer for being the second supervisor.

I would like to thank the Team of Actome and Dr. Csaba Jeney for providing me with the necessary and always helpful information on the PICO technology. In this regard, I would like to thank Tobias Gross. Based on his recommendation, I had the chance to draw up this PhD thesis in this inspiring environment. Furthermore, *Tobias* you were always a source of motivation and your dedication is admirable. Thank you for all the fruitful discussions and the work together. Special thanks go to my master student Paula Dewes for initial work on the 4EBP1 PICO assay and to Prof. Dr. Heiko Becker and Asmaa Abdel Aziz Elrakaybi for input on clinical relevance of the PICO assay. I would also like to thank Dr. Peter Koltay for constant support throughout this thesis and proofreading of the final manuscript.

Dr. Mareike Zinram and Paul Detwiler are deserving thanks for valuable input and review of mathematical calculations, and proofreading.

I wish to thank Dr. Martin Stoffel very much for his input on statistics, R programming, and thesis structure, for his critical thinking and understanding of the problems of a PhD student, and for proofreading this thesis. In fact, *Martin*, you helped me a lot throughout my entire career so far, and I am very grateful for your friendship. In this regard, I also want to thank the members of our tour group Marvin Kather, Matthias Zinram, Martin Wolff and Emma Block for constant support.

Furthermore, I would like to thank Gui Miotto for being a great help with python programming, for deciphering code, and for great Friday evenings on the rooftop or playing Nintendo. Similarly, I would like to thank my colleagues Yu-Kai Lai, Moritz Bösenberg, Daniel Frejek and Viktoria Zieger. Additionally, I would like to thank Zeba Khan for the good, encouraging, and always entertaining conversations. I greatly value your friendship.

This thesis would not have been possible without the support from my family, friends and colleagues; and my wife *Rieke*: I want to thank you from the bottom of my heart for your loving support, when I needed it the most. I would also like to thank my parents for constant support and understanding.

*Last but not least, I wanna thank me
I wanna thank me for believing in me
I wanna thank me for doing all this hard work
I wanna thank me for having no days off
I wanna thank me for, for never quitting
I wanna thank me for always being a giver
And tryna give more than I receive
I wanna thank me for tryna do more right than wrong
I wanna thank me for just being me at all times*

— Snoop Dogg

ABSTRACT

Precision medicine relies on accurately measuring biomarkers in individuals. Biomarkers can be DNA, RNA, or proteins. The invention of PCR allowed to quantify DNA and RNA, but these are only proxy biomarkers, which predict function only with high uncertainty. Proteoforms, on the other hand, represent all molecular forms of a gene's product (i.e. all possible modifications of a protein) and are the main effectors inside the cells. Their quantification delivers a higher degree of information as the fraction of modified protein enables to assess functional activity. Hence, there is an unmet need for the ability to precisely and absolutely quantify proteins and their modifications simultaneously. Protein-to-nucleic acid transformation is a technique, where two DNA-labeled antibodies are used to transform proteoforms into DNA and by this capture the diversity of the proteoforms, which can then be detected by PCR methods.

The Protein Interaction Coupling (PICO) technology was invented and patented by Dr. Csaba Jeney. PICO is homogeneous, digital immunoassay, which uses two DNA-labeled antibodies for protein-to-nucleic acid transformation and dPCR for signal detection. For PICO to work well, it is crucial to reliably and accurately quantify the basic detection unit: complexes.

Within this thesis, after a comprehensive description of the PICO technology, a mathematical model for the calculation of the number of complexes from dPCR data is established. Using a dPCR simulation created for this purpose, the mathematical model is first numerically validated. Then, experimental data with inherently higher levels of noise are used for further validation. Finally, the PICO technology is applied to a challenging research question; it is used for the simultaneous detection and quantification of a protein and one of its modifications.

It could be shown that the derived mathematical model exhibits an accuracy of $\sim 95\%$. The model is valid for the range $0 < \lambda \leq 0.25$, where λ is the average number of DNA-labeled antibodies per partition in the dPCR. The usage of dPCR for readout renders PICO a highly sensitive homogeneous immunoassay and places it among the most sensitive immunoassays known. PICO exhibits a sensitivity of 4 amol L^{-1} , which corresponds to 3 target molecules per μL .

After the successful validation of the mathematical model, a PICO assay for the detection of 4EBP1 was conceived. 4EBP1 is a component of the PI3K-Akt-mTOR pathway, which is involved in tumor proliferation through mediating protein synthesis. 10 antibodies against the 4EBP1 protein and its T37T46-phosphorylation were tested and

compared. The major limitation in this were the antibodies themselves because assay specificity is highly dependent on the quality of the antibodies. The most suitable antibody combination was applied to model system for acute myeloid leukemia (AML) with treatment-resistant and non-resistant cells lines. However, independent of the resistance state, the cell lines exhibited hyperphosphorylated 4EBP1, which is an indication for constantly active PI3K-Akt-mTOR signaling, which in turn is favors tumor proliferation.

Taken together, this thesis lays the foundation for reference-free absolute quantification of proteoform using PICO. The advantage of reference-free absolute quantification is that no reference standard is needed, which often differs from the actual analyte. This increases meaningfulness and amplifies the common understanding of cellular processes for a better development of medicines and treatment regimes within precision medicine.

ZUSAMMENFASSUNG

Die Präzisionsmedizin beruht auf der genauen Messung von Biomarkern bei einzelnen Patient*innen. Biomarker können DNA, RNA oder Proteine sein. Die Erfindung der PCR ermöglichte es, DNA und RNA zu quantifizieren. Dies sind jedoch nur stellvertretende Biomarker, die die tatsächliche Funktion nur mit großer Unsicherheit vorhersagen können. Proteoformen hingegen repräsentieren alle molekularen Formen eines Genprodukts (d. h. alle möglichen Modifikationen eines Proteins) und sind die wichtigsten Effektoren innerhalb der Zellen. Ihre Quantifizierung liefert ein höheres Maß an Informationen, da der Anteil der modifizierten Proteine eine Bewertung der funktionellen Aktivität ermöglicht. Daher besteht ein ungedeckter Bedarf an der Fähigkeit, Proteine und ihre Modifikationen gleichzeitig, präzise und absolut zu quantifizieren. Die Protein-zu-Nukleinsäure-Transformation ist eine Technik, bei der DNA-markierte Antikörper verwendet werden, um Proteoformen in DNA umzuwandeln. Die umgewandelten Proteoformen können mittels PCR-Methoden nachgewiesen werden und ermöglichen es dadurch die Vielfalt der Proteoformen zu erfassen.

Die Protein Interaction Coupling (PICO) Technologie wurde von Dr. Csaba Jeney erfunden und patentiert. PICO ist ein homogener, digitaler Immunoassay, der zwei DNA-markierte Antikörper für die Umwandlung von Proteinen in Nukleinsäuren und dPCR für die Signaldetektion verwendet. Damit PICO gut funktioniert, ist eine zuverlässige und genaue Quantifizierung der grundlegenden Nachweisbarkeit, der Complexes, unerlässlich.

Im Rahmen dieser Arbeit wird nach der umfassenden Beschreibung der PICO-Technologie ein mathematisches Modell zur Berechnung der Anzahl der Complexes aus dPCR-Daten aufgestellt. Mit Hilfe einer zu diesem Zweck erstellten dPCR-Simulation wird das mathematische Modell zunächst numerisch validiert. Dann werden experimentelle Daten mit einem inhärent höheren Rauschen zur weiteren Validierung verwendet. Schließlich wird die PICO-Technologie auf eine anspruchsvolle Forschungsfrage angewandt; sie wird für den gleichzeitigen Nachweis und die Quantifizierung eines Proteins und einer seiner Modifikationen verwendet.

Es konnte gezeigt werden, dass das abgeleitete mathematische Modell eine Genauigkeit von $\sim 95\%$ aufweist. Das Modell ist für den Bereich $0 < \lambda \leq 0.25$ gültig, wobei λ die durchschnittliche Anzahl der DNA-markierten Antikörper pro Partition in der dPCR ist. Die Verwendung von dPCR für das Auslesen macht PICO zu einem hochempfindlichen homogenen Immunoassay und damit zu einem der empfindlichsten bekannten Immunoassays. PICO weist eine Empfindlichkeit von 4 amol L^{-1} auf, was 3 Zielmolekülen pro μL entspricht.

Nach der erfolgreichen Validierung des mathematischen Modells wurde ein PICO-Assay zum Nachweis von 4EBP1 konzipiert. 4EBP1 ist ein Protein des PI3K-Akt-mTOR Signalwegs, der durch die Proteinsynthese an der Tumorproliferation beteiligt ist. Es wurden 10 Antikörper gegen das 4EBP1-Protein und seine T37T46-Phosphorylierung getestet und verglichen. Die größte Einschränkung waren dabei die Antikörper selbst, da die Spezifität des Tests stark von der Qualität der Antikörper abhängt. Die am besten geeignete Antikörperkombination wurde auf ein Modellsystem für akute myeloische Leukämie (AML) mit behandlungsresistenten und nicht-resistenten Zelllinien angewendet. Unabhängig vom Resistenzstatus wiesen die Zelllinien jedoch hyperphosphoryliertes 4EBP1 auf, was ein Hinweis auf einen konstant aktiven PI3K-Akt-mTOR Signalweg ist, der wiederum die Tumorproliferation begünstigt.

Insgesamt legt diese Arbeit den Grundstein für die referenzfreie absolute Quantifizierung von Proteoformen mit PICO. Der Vorteil der referenzfreien absoluten Quantifizierung liegt darin, dass kein Referenzstandard benötigt wird, der sich oft vom eigentlichen Analyten unterscheidet. Dies erhöht die Aussagekraft und erweitert das allgemeine Verständnis zellulärer Prozesse für eine bessere Entwicklung von Medikamenten und Behandlungsregimen im Rahmen der Präzisionsmedizin.

PUBLICATIONS AND CONFERENCE POSTERS

Part of the work presented in this thesis appeared earlier in the following preprints or conference posters. The corresponding data is highlighted accordingly.

Gross, T, **Hundertmark, Tobias**, Csiszar, V, Sulyok, AA, Gross, N, Breiden, M, Kitschen, N, Von Groll, U, Niemöller, C, Sanchez-Martin, P, Hein, A, Gopffert, J, Szoradi, T, Luebbert, P, Koltay, P, Porschewski, P, Von Stetten, F, Zengerle, R, & Jeney, C. (2024). *Lossless Single-Molecule Counting To Absolute Quantify Proteoforms* (preprint). Molecular Biology. <https://doi.org/10.1101/2024.03.19.585761>

Niemöller, C, Groß, T, **Hundertmark, Tobias**, Elrakaybi, A, Becker, H, & Jeney, C. (2023). Development of scPICO—A high-sensitivity single cell antibody assay with dPCR readout for quantitative analysis of 4EBP1 protein phosphorylation in drug-perturbed cellular populations. *Single Cell Analyses*, Cold Spring Harbor Laboratory.

Lange, Tobias, Elrakaybi, A, Dewes, P, Uhl, D, Jeney, Á, Becker, H, Zimmermann, S, & Jeney, C. (2023). Evaluation of 4EBP1 as a potential biomarker for secondary hypomethylating agents resistance in acute myeloid leukemia using Protein Interaction Coupling technology. *Acute Leukemias XVIII*, Verein für Leukämieforschung und -Therapie e.V.

Lange, Tobias, Groß, T, Jeney, Á, Scherzinger, J, Sinkala, E, Niemöller, C, Zimmermann, S, Koltay, P, Stetten, Fv, Zengerle, R, & Jeney, C. (2022). *Validation of scRNA-seq by scRT-ddPCR using the example of ErbB2 in MCF7 cells* (preprint). Molecular Biology. <https://doi.org/10.1101/2022.05.31.494164>

Lange, Tobias, Groß, T, Scherzinger, J, Sinkala, E, Niemöller, C, Zimmermann, S, Koltay, P, von Stetten, F, Zengerle, R, & Jeney, C. (2022). Validation of scRNA-seq fold changes by single-cell reverse transcription digital PCR. *Single-Cell Genomics - When Stochasticity Meets Precision in a Single Cell*, Gordon Research Conference.

CONTENTS

Introduction

1	Motivation and background	3
1.1	Precision medicine	3
1.2	Molecular biomarkers	4
1.2.1	Genomic and transcriptomic biomarkers	6
1.2.2	Proteomic biomarkers	8
2	State of the art: protein detection and quantification	11
2.1	Total protein detection	11
2.2	Mass spectrometry	13
2.3	Immunoassays	13
2.3.1	General principle	14
2.3.2	Antibodies	16
2.3.3	Protein-to-nucleic acid transformation	18
2.3.4	dPCR for proteoform detection	21
3	Objectives of this thesis	25

Protein Interaction Coupling (PICO)

4	Protein Interaction Coupling (PICO)	31
4.1	Workflow	31
4.2	DPCR double positive cluster segregation (dDPCS) model	38
4.2.1	Mathematical background	38
4.2.2	Derivation of the model	41
4.2.3	Implementation	48

Results and Discussion

5	Creation of a dPCR simulation	53
5.1	Considerations and setup	53
5.1.1	Ideal conditions	53
5.1.2	Biased conditions	56
5.2	Validation	58
5.2.1	Ideal conditions	58
5.2.2	Biased conditions	58
5.3	Summary	62
6	Validation of the dDPCS model	65
6.1	Using simulated data	65
6.1.1	Ideal conditions	65
6.1.2	Biased conditions	76
6.1.3	Summary	82
6.2	Using experimental data	83
6.2.1	Qualification of dPCR system and DNA-labels	83
6.2.2	Based on DNA	93
6.2.3	Based on recombinant protein	95

6.3	Evaluation of dDPCS model performance	102
6.4	Design rules	105
7	Application example: biomarker in leukemia	109
7.1	Background	109
7.2	PICO assay development	112
7.2.1	Validation of antibodies against β EBP1	112
7.2.2	Validation of antibodies against phosphorylated β EBP1	116
7.2.3	Summary	126
7.3	Detection of phosphorylated β EBP1 in AML cell lines	128
7.4	Summary	130
8	Conclusion	131
8.1	Achievements	131
8.2	Limitations	133
8.3	Future perspectives	135
Experimental Procedure		
9	Methods	141
9.1	Cell culture	141
9.1.1	Cells and culture conditions	141
9.1.2	Inhibitor treatment of cells for Western blotting	141
9.1.3	Dactolisib treatment of cells for PICO	142
9.1.4	Cell harvest and counting	142
9.1.5	Cell harvest and lysis for Western blot	142
9.1.6	Cell lysis for PICO	142
9.1.7	Cell lysis for PPase treatment	143
9.2	Molecular biology	144
9.2.1	SDS-PAGE and Western blot	144
9.2.2	ELISA	145
9.2.3	Antibody labeling	145
9.2.4	Protein electrophoresis	145
9.2.5	Determination of antibody labeling efficiency	146
9.2.6	PPase treatment or recombinant proteins	146
9.2.7	Digital PCR	147
9.3	PICO	148
9.3.1	DNA-only PICO	148
9.3.2	Preparation of antibody mix	149
9.3.3	Preparation of binding reaction	149
9.3.4	Preparation for dPCR	149
9.4	Computational methods	149
9.4.1	Programming	149
9.4.2	Statistical analyses	150
10	Materials	153
Bibliography		161

List of Figures 187

List of Tables 191

Listings 192

Appendix

A Protein Interaction Coupling (PICO) A.2-1

 A.1 Comparison of models for complex calculation A.2-1

 A.2 Reconstruction of 2-dimensional raw dPCR data A.2-3

B DPCR simulation B.0-1

C dDPCS model validation C.0-2

D Application example: biomarker in leukemia D.0-1

E Experimental procedure E.0-1

ACRONYMS

4EBP1	eukaryotic translation initiation factor 4E-binding protein 1 9 , 25–27 , 102 , 104 , 109 , 111–130 , 133 , 134 , 144–146 , 149 , 155 , 157 , <i>Glossary: 4EBP1</i>
ABC	antibody binding control 37 , 41 , 42 , 44 , 47 , 48 , 65 , 66 , 69–71 , 76 , 91–93 , 97 , 99–104 , 107 , 115 , 118 , 149 , <i>Glossary: ABC</i>
ABX	antibody mix 32 , 35 , 36 , 99 , 118 , 149 , <i>Glossary: ABX</i>
AML	acute myeloid leukemia 25–27 , 50 , 109–111 , 120 , 128–130 , 133 , <i>Glossary: AML</i>
CB	control buffer 37 , 149 , <i>Glossary: CB</i>
CI	confidence interval 58 , 60–63 , 67 , 69–72 , 75 , 81 , 88 , 90–92 , 94 , 95 , 97 , 98 , 100 , 103 , 104 , 113–115 , 117–122 , 125 , 127–129 , 150 , 152 , <i>Glossary: CI</i>
CV	coefficient of variation 58 , 59 , 84 , 86 , <i>Glossary: CV</i>
ddELISA	droplet digital ELISA 20 , 24 , <i>Glossary: ddELISA</i>
ddIA	digital droplet immunoassay 20 , 21 , 24 , 34 , <i>Glossary: ddIA</i>
dDPCS	dPCR double positive cluster segregation 25 , 26 , 32 , 34 , 35 , 38 , 41–43 , 47 , 48 , 50 , 53–56 , 58 , 62 , 65–69 , 71–73 , 75–80 , 82 , 83 , 91 , 93 , 95 , 97 , 102–107 , 109 , 130–132 , 136 , 149 , 150 , <i>Glossary: dDPCS</i>
dPLA	digital PLA 20–24 , 34 , <i>Glossary: dPLA</i>
dSimoa	dropcast Simoa 20 , 24 , <i>Glossary: dSimoa</i>
DX	compartment decoupling 38 , 48 , 132 , <i>Glossary: DX</i>
eIF4E	eukaryotic translation initiation factor 4E 9 , 111 , 112 , 120 , 123 , 133 , 156 , 157 , <i>Glossary: eIF4E</i>
ELISA	enzyme-linked immunosorbent assay 14 , 15 , 18 , 120 , <i>Glossary: ELISA</i>
Fab	fragment antigen binding 14 , 16 , 17 , <i>Glossary: Fab</i>
FACS	fluorescent-activated cell sorting 20 , 21 , 24 , 89 , 96 , 115 , <i>Glossary: FACS</i>
Fc	fragment crystallizable 14–17 , <i>Glossary: Fc</i>
FRET	Förster resonance energy transfer 19 , <i>Glossary: FRET</i>
HER2	human epidermal growth factor receptor 2 16 , 95–99 , 101 , 102 , 112 , 149 , 156 , 157 , <i>Glossary: HER2</i>

HMA	hypomethylating agent 26, 27, 50, 109, 110, 112, 128–130, 133, <i>Glossary: HMA</i>
LB	lysis buffer 117, 118, 142, 143, 149, <i>Glossary: LB</i>
LC-MS/MS	liquid chromatography with tandem mass spectrometry 13, 20, 132, <i>Glossary: LC-MS/MS</i>
LE	labeling efficiency xxxv, 32, 97, 99, 145, 146, 157, <i>Glossary: LE</i>
LOD	limit of detection 76, 91, 94, 95, 101, 102, 105, 107, 132, 152
LOQ	limit of quantification 76, 94, 95, 101, 102, 105, 107, 132, 152
MGB	minor groove binder 19, 85, 87–89, 158, <i>Glossary: MGB</i>
MOSAIC	molecular on-bead signal amplification for individual counting 20, 24, <i>Glossary: MOSAIC</i>
MS	mass spectrometry 11, 13, 14, 135, 136
mTOR	mechanistic target of rapamycin 111, 116, 123–125, 128, <i>Glossary: mTOR</i>
NGS	next generation sequencing 3, 6, 18, 20, 24, 135
PEA	proximity extension assay 18, 20, 22, 24, 103, 132, 134, <i>Glossary: PEA</i>
PICO	Protein Interaction Coupling 25, 31, 131
PLA	proximity ligation assay 18, 19, 22–24, 31, 36, 103, 132, <i>Glossary: PLA</i>
PLB	passive lysis buffer 117, 118, <i>Glossary: PLB</i>
PPase	protein phosphatase 117–122, 125, 126, 143, 144, 146, <i>Glossary: PPase</i>
PPI	protein-protein interaction 3, 7–9, 12, 18, 23, 31–34, 36, 38, 42, 96, 103, 120, 133, 137
PSC	partition-specific competition 56, 57, 62, 63, 76, 81–83, 88, 89, 106, <i>Glossary: PSC</i>
PTM	post translational modification 7–9, 16, 25, 33, 36, 133, 134, <i>Glossary: PTM</i>
PTZ	pertuzumab 16, 95–98, 102, 112, 115, 130, 133, 135, 156, 157, <i>Glossary: PTZ</i>
rc-overlap	random-couple overlap 45, 48, 55, 56, 71, 73–75, <i>Glossary: rc-overlap</i>
RCA	rolling circle amplification 18, 24, <i>Glossary: RCA</i>
RFU	relative fluorescence unit 89
RIPA	radioimmunoprecipitation assay 117, 118, 143, 144, <i>Glossary: RIPA</i>

S6K1	ribosomal protein S6 kinase beta-1 101 , 111 , <i>Glossary: S6K1</i>
SDS	sodium dodecyl sulfate polyacrylamide gel electrophoresis 12 , 14 , 124 , 144
SDS	sodium dodecyl sulfate 12 , 118 , 143 , 144
Simoa	single molecule array 20 , 24 , 103 , 132 , 134 , <i>Glossary: Simoa</i>
TTZ	trastuzumab 16 , 95–98 , 102 , 112 , 115 , 130 , 133 , 135 , 156 , 157 , <i>Glossary: TTZ</i>
URAB	Unrelated antibody 33 , 37 , 100–102 , 156 , 157 , <i>Glossary: URAB</i>

GLOSSARY

4EBP1	The eukaryotic translation initiation factor 4E-binding protein 1 (4EBP1) is a primary phosphorylation target of mTOR (Qin et al., 2016). Upon phosphorylation, 4EBP1 releases eIF4E, which can thus initiate 5' cap-dependent mRNA translation. Non-phosphorylated, 4EBP1 sequesters eIF4E preventing translation. 9
ABC	An antibody binding control (ABC) is a PICO experiment, where the sample is the CB, which is a buffer containing only dissolved bovine serum albumin. Thus, the sample is lacking the target of interest. After combining the CB with the ABX in the binding reaction, no complexes can be formed. An ABC PICO experiment can be considered a blank sample. 37
ABX	An antibody mix (ABX) is a solution with an equimolar concentration of at least two antibodies against (different) target(s) of interest. In PICO assays, the ABX is combined with the sample in the binding reaction. Additionally, the ABX may contain URABs. 32
AML	Acute myeloid leukemia (AML) is a proliferative disorder of the hematopoietic system. It is characterized by an uncontrolled proliferation of immature myeloblasts. 25
antibody pair	An antibody pair describes two antibodies of all antibodies in the ABX, whose 2-dimensional raw dPCR data is used by the dDPCS model for the calculation of the number of complexes. 36
AQ model	The ternary-equilibrium-based absolute quantification model (AQ model) is a mathematical framework that uses antibody and complex concentrations as well as K_d values to determine the absolute number of targets without the need for a reference standard (Gross et al., 2024). 37, 97
binding reaction	The binding reaction is the combination of the sample containing the target of interest and the ABX. During its incubation time, the antibody-target binding takes place. The binding reaction is incubated long enough so that it can reach chemical equilibrium. 32

CB	The control buffer (CB) is proprietary buffer from Actome used in ABC PICO experiments. It contains bovine serum albumin dissolved in LB. 37
CI	The confidence interval (CI) gives an estimate for the precision of the calculation of a distribution parameter such as the mean. For a given confidence level (most often 95 %), the true distribution mean falls within this interval with a 95 % certainty. The CI can be calculated as follows: $\bar{x} \pm 1.96 \frac{SD}{\sqrt{n}}$, where \bar{x} is the distribution mean, 1.96 is the corresponding value of the Z-distribution for the 95 % CI, SD is the standard deviation of the distribution and n is the sample size. The SD, on the other hand, describes the dispersion of the distribution. 58
colorpairs	A colorpair is a combination of two fluorescent channels of a dPCR system to generate the required 2-dimensional raw data for the dDPCS model for the calculation of the number of complexes. 96
complexes	Complexes are the basic detection of the PICO technology. They consist of two DNA-labeled antibodies and their target. A target can be a protein, a PTM or PPI (Figure 4.2). 32
CV	The coefficient of variation (CV) is defined as the ratio of the SD and the mean of a distribution and is often expressed in %. It thus, describes the dispersion of a distribution relative to its mean. 58
ddELISA	The droplet digital ELISA (ddELISA) is a digital, bead-based immunoassay. The antibody-coated beads capture the target. A detection antibody carrying a β -galactosidase subsequently binds the target, too. Unbound targets and detection antibodies are washed away. Together with the substrate for the β -galactosidase, the beads are partitioned in droplets. The β -galactosidase converts its substrate into a fluorescent product, rendering the droplet fluorescent, too. The number of fluorescent droplets can be used to infer the target concentration (Cohen et al., 2020). 20

ddIA	The digital droplet immunoassay (ddIA) is PLA-based immunoassay in an array of polydisperse droplets. Like in conventional PLA, two DNA-labeled antibodies bind to the same target in solution. This solution is partitioned in polydisperse droplets. After partitioning, the DNA-labels are ligated inside the droplets using a ligation oligonucleotide and a DNA-ligase, when in proximity (see Figure 2.4A). This first, generated DNA template can be amplified and fluorescence is generated. The number of fluorescent droplets can be used to infer the target concentration (Byrnes et al., 2020). 20
dDPCS	The dPCR double positive cluster segregation (dDPCS) model is used to calculate the number of complexes from 2-dimensional raw dPCR data (see also Section 4.2). 25
dPLA	In a digital PLA (dPLA), two DNA-labeled antibodies bind their target in solution. In contrast to ddIA, the first DNA template is generated before partitioning. After partitioning, the DNA template is amplified and fluorescence is generated. The number of fluorescent droplets can be used to infer the target concentration (Abasianik et al., 2020 ; Albayrak et al., 2016). 20
dSimoa	The dropcast Simoa (dSimoa) is a variant of conventional Simoa. Antibody-coated beads capture the target. A detection antibody carrying a circular DNA fragment binds the target, too. Unbound targets and detection antibodies are washed away. Using RCA, fluorescence is generated on the beads. The bead suspension is then dropcasted onto a microscopic slide and allowed to dry to form a monolayer. Then, the number of fluorescent beads can be used to infer the target concentration (Wu et al., 2020). 20
DX	The compartment decoupling (DX) model is another model for the calculation of the number of complexes from 2-dimensional raw dPCR data (see also Gross et al. (2024)). In contrast to the dDPCS model, which is numerically solved, the DX model is a closed-form expression based on the methods of moments (Gross et al., 2024). A comparison of the two models for complex calculation can be found in Appendix A . 38
eIF4E	The eukaryotic translation initiation factor 4E (eIF4E) is a translation initiation factor directing the ribosome to the 5' cap of the mRNA favoring protein synthesis. EIF4E is sequestered by non-phosphorylated 4EBP1. 9

ELISA	The enzyme-linked immunosorbant assay (ELISA) is a frequently used immunoassay for the detection of proteins (see Section 2.3.1). There is a great variety of assay modes, but in general the principle is that labeled antibodies transform the proteoform into a detectable signal (Aydin, 2015). 14
Fab	The fragment antigen binding (Fab) is a region of the antibody that contains the hypervariable regions (paratopes), which are responsible for antigen binding (Figure 2.2). 14
FACS	Fluorescent-activated cell sorting (FACS) is a specialized type of flow cytometry. It sorts a heterogeneous mixture of cells into distinct populations based on specific light scattering and fluorescent characteristics. Cells are labeled with fluorescent markers, passed through a laser beam, and separated using an electric charge based on their emitted fluorescence. 20
Fc	The fragment crystallizable (Fc) is the constant region of the antibody, which mediates the communication with the immune cells to eliminate the recognized, foreign structures. 14
FRET	The Förster resonance energy transfer (FRET) is a principle describing the energy transfer of two light sensitive molecules, for instance fluorophors (Förster, 1948). FRET is used in PCRs to generate fluorescence with sequence-specific hydrolysis probes. Probes are conjugated with a 5'-dye and a 3'-quencher. During amplification the polymerase cleaves the probe, quencher and dye lose spatial proximity and the fluorescence of the dye can be detected (see also Figure 2.3). 19
HER2	The human epidermal growth factor receptor 2 (HER2) is a receptor tyrosine kinase and an oncogene, which is overrepresented in ~ 30 % of all breast cancer cases. It is associated with cell growth and tumorigenesis. Homodimerization activates multiple pathways, which favor proliferation and prevent apoptosis (Iqbal & Iqbal, 2014). 16

HMA	Hypomethylating agents (HMA) like decitabine and azacitidine are cytosine analogs, which are incorporated into the DNA during replication. DNA methyltransferases, trying to methylate HMAs, get trapped by covalent binding and are subsequently degraded (Stresemann & Lyko, 2008). This results in a genome-wide hypomethylation, which re-induces the expression of silenced tumor suppressors (Daskalakis et al., 2002). HMAs are used for palliative treatment of elderly AML patients. 26
LB	The lysis buffer (LB) is a proprietary buffer from Acytome, which is used for cell lysis and sample preparation in PICO assays. Furthermore, it shall ensure effective antibody-target binding conditions and suppress unspecific bindings or cross-reactions. 117
LC-MS/MS	Liquid chromatography with tandem mass spectrometry (LC-MS/MS) is the standard tool for large-scale proteomics analyzes (Cox & Mann, 2011). 13
LE	The labeling efficiency (LE) is a measure of the fraction of an antibody solution that carries a DNA-label. It can be calculated using Equation 9.1. xxxv
MGB	The minor groove binder (MGB) is a chemical group that enhances hybridization of DNA strands by non-covalently binding to the minor groove. This improves stability of the hybridization. Sequence-specific hydrolysis probes often carry a MGB enabling shorter probes. This results in a smaller distance between 5'-dye and a 3'-quencher rendering FRET more efficient (Kutyavin, 2000). 19
MOSAIC	The molecular on-bead signal amplification for individual counting (MOSAIC) assay is an adaption of ddELISA. However, in the MOSAIC assay, the detection antibody carries a circular DNA fragment. Then, using RCA, fluorescence is generated on the beads. Fluorescent beads are detected by FACS and their number can be used to infer the target concentration (Wu et al., 2022). 20
mTOR	Mechanistic target of rapamycin (mTOR) is a protein kinase pivotal in many cellular processes such as growth, proliferation, survival, transcription and more (Saxton & Sabatini, 2017). 4EBP1 is a phosphorylation target of mTOR (Figure 7.1). 111

PEA	The proximity extension assay (PEA) is a variant of PLA. It also uses two DNA-labeled antibodies directed against the same target. However, the DNA-labels are designed in such a way that through binding, they can immediately hybridize. This eliminates the need to DNA ligation and a DNA polymerase is used to generate the first DNA template, which is then amplified and fluorescence is generated. Depending on the fluorescence intensity, the target concentration can be inferred (Lundberg, Eriksson, et al., 2011). 18
PLA	The proximity ligation assay (PLA) is an immunoassay, which uses two DNA-labeled antibodies directed against the same target. Through binding, the DNA-labels come in proximity and are ligated using a ligation oligonucleotide and a DNA-ligase. This first, generated DNA template can be amplified and fluorescence is generated. Depending on the fluorescence intensity, the target concentration can be inferred (Fredriksson et al., 2002). 18
PLB	The passive lysis buffer (PLB) is a proprietary buffer for cell lysis from Promega. 117
PPase	A protein phosphatase (PPase) is an enzyme that cleaves phosphate-groups from serine, threonine or tyrosine residues. 117
PSC	Partition-specific competition (PSC) is a phenomenon that occurs in dPCR (Whale et al., 2016). Each partition only contains a limited amount of resources (e.g. nucleotides, enzymes, cofactors). All DNA molecules in this partition compete for these resources. This can lead to the preferred amplification of one DNA molecule. The other DNA molecule might not be sufficiently amplified to become detectable. 56
PTM	A post translational modification (PTM) is a modification of a protein after it was translated from the mRNA. The most common PTMs are acetylation, acylation, glycosylation, methylation, phosphorylation, and ubiquitination. 7
PTZ	Pertuzumab (PTZ) is a therapeutic antibody directed against HER2. It used to treat HER2-positive breast cancer (Nahta et al., 2004). PTZ and TTZ can bind to HER2 simultaneously to exploit synergies for higher treatment success (Nami et al., 2018; Sharma et al., 2021; Tsao et al., 2022). 16

RCA	The rolling circle amplification (RCA) is an isothermal DNA amplification principle for the amplification of circular DNA fragments. RCA is used by some immunoassays for signal generation. 18
RIPA	The radioimmunoprecipitation assay (RIPA) buffer is a strong, denaturing cell lysis buffer because it contains SDS (Table 9.1). It disrupts cell membranes and allows for protein extraction from all cell compartments. It is frequently used in protein detection and quantification methods. 117
S6K1	The ribosomal protein S6 kinase beta-1 (S6K1) is a primary phosphorylation target of mTOR. Upon phosphorylation, it is activated and phosphorylates further targets, which collectively enhance mRNA translation (Figure 7.1). 101
Simoa	The single molecule array (Simoa) is bead-based immunoassay. Antibody-coated beads capture the target. A detection antibody carrying a β -galactosidase subsequently binds the target, too. Unbound targets and detection antibodies are washed away. Together with the substrate for the β -galactosidase, the beads are loaded into femtoliter-sized wells. The β -galactosidase converts its substrate into a fluorescent product, rendering the well fluorescent, too. The number of fluorescent wells can be used to infer the target concentration (Chang et al., 2012 ; Rissin et al., 2010). 20
TTZ	Trastuzumab (TTZ) is a therapeutic antibody directed against HER2. It used to treat HER2-positive breast cancer (Nahta et al., 2004). PTZ and TTZ can bind to HER2 simultaneously to exploit synergies for higher treatment success (Nami et al., 2018 ; Sharma et al., 2021 ; Tsao et al., 2022). 16
URAB	An unrelated antibody (URAB) is an antibody in the ABX that has no affinity to the target of interest. Thus, in all possible combinations with the other antibodies of the ABX, no complex can be formed. The URAB serves as a binding control. Gross et al. (2024) call such an antibody bystander antibody (BAB). 33

SYMBOLS

Symbol	Description	Unit
AUC	area under the curve	-
A	absorbance of a material (<i>Beer-Lambert-Law</i>)	-
c	concentration of absorbing material (<i>Beer-Lambert-Law</i>)	mol L^{-1}
c_m	concentration of target molecules of interest in partitioned volume of a dPCR	copies/ μL
f_a	fraction of antibodies a that are bound to the target, by default its value is 1	-
f_b	fraction of antibodies b that are bound to the target, by default its value is 1	-
k	number of detectable target molecules per given partition of a dPCR	molecules per given partition
K_d	equilibrium dissociation constant, defines the affinity of an antibody to its antigen	mol L^{-1}
l	path length of light (<i>Beer-Lambert-Law</i>)	cm
m	true number of target molecules in a dPCR	molecules
\hat{m}	observed number of target molecules in a dPCR	molecules
m_a	number of DNA-labeled antibodies a in partitioned volume	-
m_b	number of DNA-labeled antibodies b in partitioned volume	-
m_A	number of antibodies a that are bound to the target	-
m_B	number of antibodies b that are bound to the target	-
m_{AB}	number of targets bound by both antibodies a and b, i.e. number of complexes	-
$m_{\text{target}_{\text{free}}}$	number of free targets in a sample of a PICO assay	-
n	true number of partitions of a dPCR	partitions
\hat{n}	observed number of partitions of a dPCR	partitions
\hat{n}_+	observed number of positive partitions in a dPCR	partitions

Symbol	Description	Unit
\hat{n}_-	observed number of negative partitions in a dPCR	partitions
\hat{n}_{a^+}	observed number of partitions containing DNA-labeled antibody a	partitions
\hat{n}_{b^+}	observed number of partitions containing DNA-labeled antibody b	partitions
$\hat{n}_{a^+b^+}$	observed number of partitions containing DNA-labeled antibody a and containing DNA-labeled antibody b	partitions
$\hat{n}_{a^+b^-}$	observed number of partitions only containing DNA-labeled antibody a but not containing DNA-labeled antibody b	partitions
$\hat{n}_{a^-b^+}$	observed number of partitions only containing DNA-labeled antibody b but not containing DNA-labeled antibody a	partitions
$\hat{n}_{a^-b^-}$	observed number of partitions not containing DNA-labeled antibody a and not containing DNA-labeled antibody b	partitions
n_{AB^+}	true number of partitions containing a complex, independent of other detectable molecules	partitions
\hat{n}_{AB^+}	observed number of partitions containing a complex, independent of other detectable molecules	partitions
$n_{AB^+a^+b^+}$	true number of partitions containing both unbound antibodies a and b and complexes	partitions
$\hat{n}_{AB^+a^+b^+}$	observed number of partitions containing both unbound antibodies a and b and complexes	partitions
$\hat{n}_{AB^+a^+b^-}$	observed number of partitions containing unbound antibodies a and complexes	partitions
$\hat{n}_{AB^+a^-b^+}$	observed number of partitions containing unbound antibodies b and complexes	partitions
$n_{AB^-a^+b^+}$	true number of partitions containing only both unbound antibodies a and b	partitions
$\hat{n}_{AB^-a^+b^+}$	observed number of partitions containing only both unbound antibodies a and b	partitions
$\hat{n}_{AB^-a^+b^-}$	observed number of partitions containing only unbound antibodies a	partitions
$\hat{n}_{AB^-a^-b^+}$	observed number of partitions containing only unbound antibodies b	partitions

Symbol	Description	Unit
$\hat{n}_{AB^{-}a^{-}b^{-}}$	observed number of partitions containing no detectable molecule	partitions
\hat{n}_{d_1}	observed number of partitions containing only both unbound antibodies a and b, see $\hat{n}_{AB^{-}a^{+}b^{+}}$	partitions
\hat{n}_{d_2}	observed number of partitions containing both unbound antibodies a and b and complexes, see $\hat{n}_{AB^{+}a^{+}b^{+}}$	partitions
\hat{n}_{d_3}	observed number of partitions containing only complexes	partitions
\hat{n}_{d_4}	observed number of partitions containing unbound antibodies a and complexes, see $\hat{n}_{AB^{+}a^{+}b^{-}}$	partitions
\hat{n}_{d_5}	observed number of partitions containing unbound antibodies b and complexes, see $\hat{n}_{AB^{+}a^{-}b^{+}}$	partitions
$n_{D_{calc}}$	number of double positive partitions calculated by the dDPCS model	partitions
$\hat{n}_{D_{obs}}$	observed number of partitions containing antibodies a and b, a complex, or a combination of these, i.e. observed number of double positive partitions	partitions
$\hat{n}_{D_{obs, ABC}}$	observed number of double positive partitions from an ABC PICO experiment	partitions
r_{ab}	ratio between all positive partitions in one channel divided by all positive partitions in a second channel (Equation 6.1)	-
T_m	melting temperature of a DNA oligonucleotide	°C
V	partitioned volume of a dPCR	µL
V_p	average volume of a partition	µL
w	number of antibodies in an ABX	-
X_a	true number of partitions containing DNA-labeled antibody a	partitions
X_b	true number of partitions containing DNA-labeled antibody b	partitions
X_{AB}	true number of partitions containing complexes	partitions
X_i	number of target molecules per partition i	molecules per partition

Symbol	Description	Unit
ϵ	molar absorption coefficient	$\text{L mol}^{-1} \text{cm}^{-1}$
η	see labeling efficiency (LE)	%
λ	mean of the Poission distribution; can be interpreted as the average number of detectable target molecules per partition in a dPCR	molecules per partition
λ_a	average number DNA-labeled antibodies a per partition in a dPCR	molecules per partition
λ_{AB}	average number complexes per partition in a dPCR	complexes per partition
λ_b	average number DNA-labeled antibodies b per partition in a dPCR	molecules per partition
μ_c	bifurcation point: value of a parameter from which on a system shows a distinct behavior	-

INTRODUCTION

MOTIVATION AND BACKGROUND

*It is far more important
to know what sort of person
the disease has than what
sort of disease the person has.*

— Hippocrates, (460 BC - 377 BC)

1.1 PRECISION MEDICINE

One of the consequences of *one-size fits-all* medicine (Figure 1.1) is that an estimated 2 200 000 individuals die each year from severe adverse drug reactions (Le Louët & Pitts, 2023) ranking this among the top 5 causes for deaths worldwide. To put this into perspective, 1 190 000 individuals die in car accidents each year (WHO). Yet, individuals that do not suffer from adverse drug reactions are not immediately cured because up to 70 % of patients experience no benefit at all (Schork, 2015). One reason is that the diversity of the clinical study cohort does not reflect the diversity of the disease prevalence in the population (Dekker et al., 2021; Schork, 2015). Through ongoing technological advancements like *next generation sequencing* (NGS), the ultimate and long-pursued goal of physicians comes into reach: *precision medicine* (Figure 1.1).

Precision medicine is patient-centered and aims to administer the appropriate drug to the right patient at the right time (Collins & Varmus, 2015). This needs comprehensive characterization of each patient (*personalized diagnostics*) (Gimeno et al., 2023; Vargas & Harris, 2016). The information required can be divided into two categories: metadata and molecular data. Metadata can contain information about individual health or disease history, lifestyle, and environmental factors. Molecular data delivers information about the individual molecular landscape of a patient, e.g. genome, epigenome, transcriptome, proteome, interactome (Bludau & Aebersold, 2020; Vargas & Harris, 2016). This is achieved through the use of *molecular biomarkers*, like specific genes, epigenetic signatures, mRNAs, proteins or *PPIs*. Molecular biomarkers are considered as the fundamentals of precision medicine (Chaffey & Silmon, 2016; Vargas & Harris, 2016). Therefore, with sufficient knowledge, an individualized treatment regimen can be devised, increasing desirable effects while simultaneously reducing adverse drug reactions and improving treatment efficacy (Figure 1.1). Eventually, precision medicine is a step towards improved public health.

Synonyms for precision medicine are personalized medicine, P4 medicine, or individualized medicine (Vargas & Harris, 2016).

The interactome describes all *PPIs* in a cell (Bludau & Aebersold, 2020).

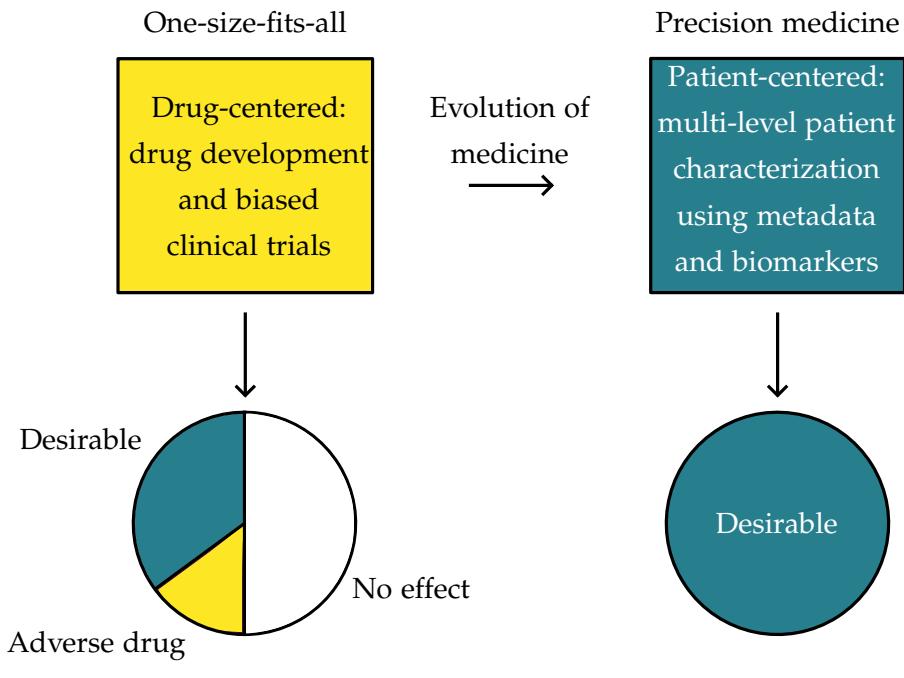


Figure 1.1: Transition from *one-size-fits-all* to precision medicine. One-size-fits-all medicine is drug-centered, meaning that the right patient for the available treatments needs to be identified. Furthermore, drugs are not developed to meet all patients needs and clinical study cohorts are often biased towards white, male participants (Nowogrodzki, 2017; Schork, 2015). This results in a large proportion of patients experiencing no effect, and a significant proportion of patients suffering from adverse drug reactions (pie chart not to scale). Precision medicine, on the other hand, focuses on the patient and aims to gather multi-level information, which consist of different metadata like health history, lifestyle and environmental factors and a molecular characterization enabled by molecular biomarkers. Eventually, this helps to find the appropriate treatment for the right patient at the right time resulting in high healing rates.

1.2 MOLECULAR BIOMARKERS

Biomarkers are indispensable indicators of physiological and pathogenic processes and required for advancing precision medicine (Biomarkers Definitions Working Group, 2001). According to BEST (Biomarkers, EndpointS, and other Tools) resource biomarkers can be categorized into seven groups, namely predictive, prognostic, diagnostic, risk, monitoring, safety and response biomarkers (Figure 1.2) (FDA-NIH Biomarker Working Group, 2016). Classical examples of biomarkers are heart rate, body temperature and blood pressure. While abnormalities in such classical biomarkers can be of various origin, the usage of molecular biomarkers enables differential, patient- and disease-specific characterizations. Molecular biomarkers encompass biomolecules such

as DNA, RNA or proteins, derived from different sources of the human body, ideally in a non-harmful or minimal invasive manner like liquid biopsy (Di Meo et al., 2017; Shah et al., 2020).

A bottleneck in precision medicine is the discovery of new biomarkers (Chaffey & Silmon, 2016), which depends on their accessibility, technologies for isolation, detection and quantification and their druggability. Furthermore, it is desirable to use tools that enable the detection of multiple biomarkers simultaneously; for instance, different biomolecules from different sampling locations, because causes for diseases are usually multifactorial (Chaffey & Silmon, 2016; Schork, 2015).

Druggability is the ability of a biomolecule to serve as a target for a drug and thus be modulated by it (Owens, 2007).

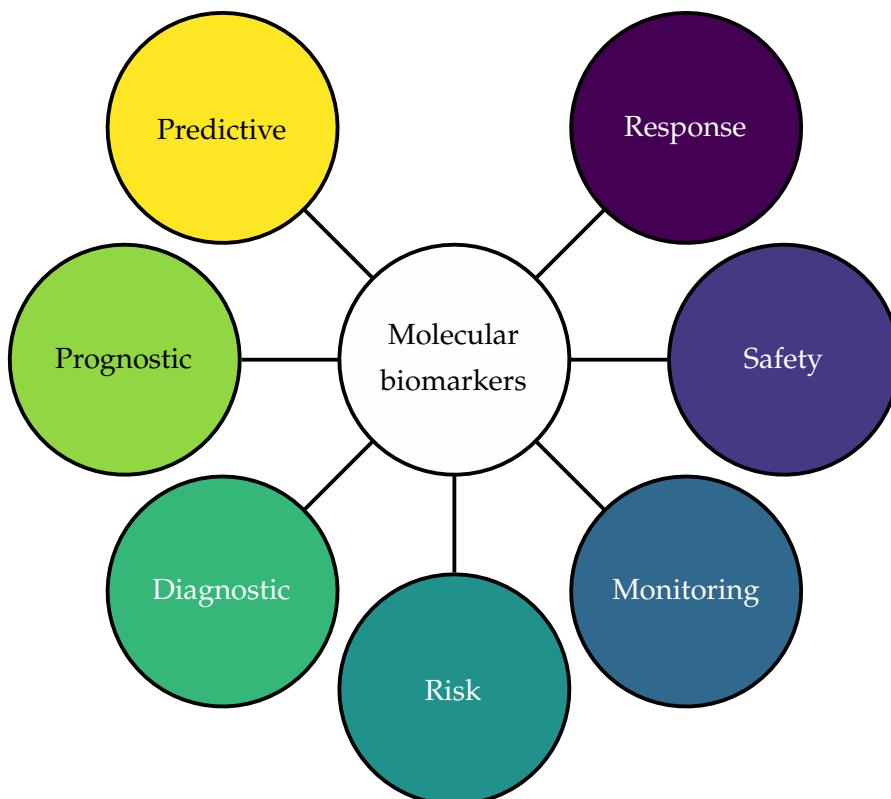


Figure 1.2: Molecular biomarkers according to FDA-NIH Biomarker Working Group (2016). The definitions of the biomarkers can be obtained from BEST (Biomarkers, EndpointS, and other Tools) resource.

The following sections focus on predictive biomarkers in precision oncology. A predictive biomarker allows clinicians to judge if a patient will benefit from a certain treatment or not by facilitating a deeper understanding of the pathophysiological processes (FDA-NIH Biomarker Working Group, 2016). Hence, the presence or absence of a predictive biomarker helps to identify promising drug candidates for designing individualized treatment regimen. This avoids the prevalent *treat-and-observe* strategies (Chaffey & Silmon, 2016). For instance, the chromosomal translocation leading to the Philadelphia chromosome and thus to chronic myeloid leukemia is a clear indication for a suc-

A driver mutation is a somatic mutation that contributes to oncogenesis, while a passenger mutation does not contribute to oncogenesis (Greenman et al., 2007).

cessful treatment of these patients with imatinib (Capdeville et al., 2002). Imatinib is a tyrosine kinase inhibitor, developed to selectively inhibit the BCR-ABL kinase, which is a fusion gene product of the Philadelphia chromosome (Deininger et al., 1997). Unfortunately, such a clear indication like in the case of the Philadelphia chromosome is rather unusual. Cancers typically exhibit large intra- and intertumoral heterogeneity, so that their molecular profiling and characterization is inevitable for the design of personalized treatments (Le Tourneau et al., 2016; Schwartzberg et al., 2017). Deciphering their heterogeneity enables identifying pre-existing, resistant subpopulations of cancer cells. Ill-considered treatments would apply evolutionary pressure and select for these resistant populations that eventually form metastasis (Le Tourneau et al., 2016).

Sequencing the first human genome (Venter et al., 2001) was a milestone in precision medicine, and the first published sequence of a cancer genome was a landmark in precision oncology (Sjöblom et al., 2006). This enabled the rapid discovery of new driver mutations, which got accelerated by novel sequencing technologies such as NGS (Mardis, 2019).

1.2.1 Genomic and transcriptomic biomarkers

Genomic (DNA) and transcriptomic (RNA) biomarkers (Figure 1.3) are preferred for their ease of use in terms of isolation, detection and throughput in analysis (Vargas & Harris, 2016). DNA and RNA¹ consist of 4 nucleotides with complementary base pairing, which makes them easy to amplify and detect via hybridization assays, PCR or NGS (Khodakov et al., 2016; Zhao et al., 2021). DNA biomarkers allow for identifying genomic aberrations like mutations (e.g. single nucleotide polymorphism), copy number variations and translocations resulting in fusion genes like *BCR-ABL* (Malone et al., 2020; Shah et al., 2020). RNA biomarkers are applied to quantify gene expression strength and to identify new RNA isoforms because of gene fusions or alternative splicing events, which potentially alter treatment response (Nattestad et al., 2018; Shah et al., 2020). In general, RNA biomarkers promise a higher resolution for the identification of cancer subtypes (Vargas & Harris, 2016), while a combination of genomic and transcriptomic data can be superior. For instance, before the treatment of non-small cell lung cancer (NSCLC) patients, it is suggested to screen for *EGFR* mutations, *ALK* and *ROS1* rearrangements and PD-L1 expression (Schwartzberg et al., 2017). This information can then be used to treat the patient with the appropriate drug. Thus, while DNA and RNA are reliable biomarkers, when analyzed individually they are not ideal to inform a personalized treatment regime.

¹ The first step in RNA analyzes is reverse transcription, where a reverse transcriptase from a retrovirus is employed to synthesize a DNA on an RNA template (Figure 1.3).

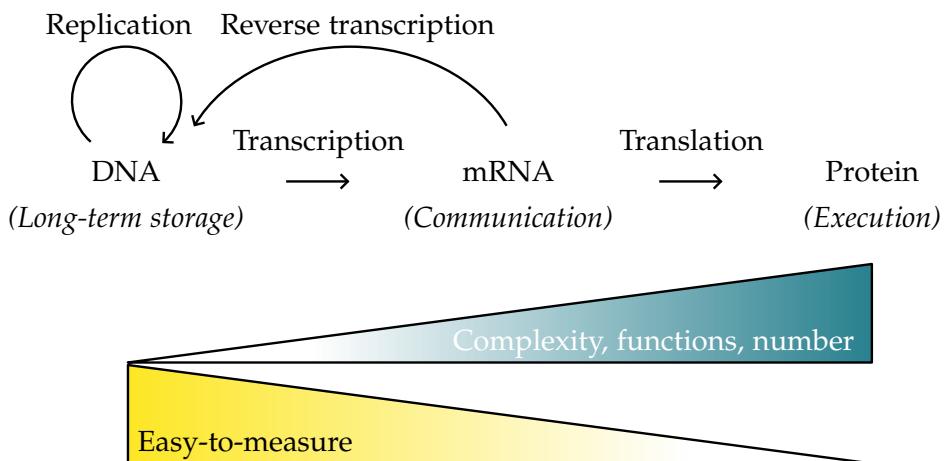


Figure 1.3: Central dogma of biology according to Crick (1970). The dogma describes the flow information between DNA, RNA and protein. DNA is meant for long-term storage of information. During mitosis, it is replicated and when necessary transcribed into mRNA, which mediates the unilateral communication to the proteins. Other RNA types can have regulatory functions (Shah et al., 2020). Retroviruses possess an RNA-dependent DNA-polymerase (reverse transcriptase), which enables them to transcribe their RNA genome into DNA. Proteins can be regulated by post translational modification (PTM) such as phosphorylation and can interact with each other (i.e. PPI) for different purposes (Jeong et al., 2001). DNA and RNA consist of 4 distinct nucleotides with base pair complementary, which render them easy-to-measure. Proteins consist of up to 21 amino acids with varying physicochemical properties, which can change upon PTM. This means that complexity, functions and numbers increase from DNA to protein and measurement is more difficult.

For example, RNA is significantly more dynamic and unstable than DNA (Shah et al., 2020) and its expression changes upon external stimuli such as tumor biopsies (Grizzle et al., 2016; Srinivasan et al., 2002). This distorts the meaningfulness of a freshly collected sample. On the other hand, the presence of a mutation in the DNA does not immediately render it a driver mutation. Given the heterogeneity of cancer, it can be challenging to identify the type of mutation (driver vs. passenger) (Hess et al., 2019; Le Tourneau et al., 2016).

Further analytical issues include poor correlations between mRNA expression and protein abundance or genotype and phenotype (Gygi et al., 1999; Payne, 2015; Snyder et al., 2009; D Wang et al., 2019). Even in single cell approaches with parallel analysis of a specific mRNA and its corresponding protein, the correlation coefficient is below 0.50 (Albayrak et al., 2016; Genshaft et al., 2016; Lin et al., 2019). Protein abundance but not mRNA abundance might be altered by translational repression through micro RNAs (Baek et al., 2008). Additionally, decay rates between mRNA and protein vary by orders of magnitude: the half life of a mRNA is several minutes, while it is hours to years

for proteins (Schwanhäusser et al., 2011). DNA and RNA may thus provide valuable insights into the genomic and transcriptomic landscape of cancer cells, but they are limited in unraveling true functional activity (Bludau & Aebersold, 2020; Duffy, 2023). Drawing conclusions about a patient's drug response based solely on genomic and transcriptomic data entails significant extrapolation and introduces a considerable level of uncertainty, which is antithetical to a system of personalized treatment. DNA and RNA can thus only be considered surrogate markers (Temple, 1999). Personalized drug treatment requires additional information from the main effectors in the cells: proteins (Figure 1.3).

1.2.2 Proteomic biomarkers

Proteins consist of up to 21 amino acids with varying physicochemical properties, which can be modified by phosphorylation, methylation, glycosylation, i.e. PTMs (Aebersold et al., 2018). *Proteoform* is a collective term for all distinct molecular forms of the product of a single gene independent of their cause of distinction. This includes a protein's canonical sequence, endogenous proteolysis, RNA splicing, coding single nucleotide polymorphisms, other mutations, errors in translation, and PTMs (Aebersold et al., 2018; LM Smith & Kelleher, 2018; LM Smith et al., 2013). The discovery of the variability of products from a single gene changed the view on the central dogma of biology (Crick, 1970) because these variations should be considered (Schlüter et al., 2009).

It is estimated that the \sim 20 000 human genes translate into \sim 1 000 000 proteoforms (Aebersold et al., 2018; Muñoz & Heck, 2014) and that a single cell has 30 000 times as many protein molecules as mRNA molecules (MacCoss et al., 2023). PTMs impact a protein's charge, hydrophobicity or hydrophilicity and may affect its degradation, activity or interaction with other proteins (PPI). In summary, proteoforms are highly variable and more complex than DNA and RNA, perform the main actions in the cell and are numerically more abundant than DNA and RNA molecules together (Aebersold et al., 2018; Bludau & Aebersold, 2020; MacCoss et al., 2023; Muñoz & Heck, 2014). In contrast to proteoforms, the architecture of DNA and RNA is relatively simple. The main building blocks are 4 distinct nucleotides with similar physicochemical properties and complementary paring, which renders them easy-to-measure. Consequently, proteoform analysis comes at higher costs regarding throughput and generalization of workflows. Their analysis complements the information obtained from DNA and RNA analyzes confirming preliminary hypotheses. From the assessment of proteoforms, statements on cellular functions are less prone to errors from extrapolation of mRNA abundances.

However, the contribution of external stimuli like cell-cell-interactions is not covered by such analyzes.

The mere presence of a mutation is not sufficient to judge if a dysfunctional proteoform causes a disease. For instance, GAPDH is involved both in physiological metabolism and homeostasis as well as in pathologic processes depending on its PTMs and PPIs (Lazarev et al., 2020; Sirover, 1999). Another example is eukaryotic translation initiation factor 4E-binding protein 1 (4EBP1), which is a regulator of translational activity (Qin et al., 2016). If phosphorylated, 4EBP1 sequesters eukaryotic translation initiation factor 4E (eIF4E) in the nucleus, while 70 % of 4EBP1 is located in the cytosol with yet unknown function (Rong et al., 2008). These examples once again demonstrate, that genomic or transcriptomic information does not necessarily provide insights into the true functional activity of the resulting proteoforms.

*A single human cell has about
1 × 10⁸ ribonucleotides,
1 × 10¹⁰ desoxyribonucleotides
but 1 × 10¹² amino acids.*

— Nikolai Slavov on [the proteome scale](#).

2

STATE OF THE ART: PROTEIN DETECTION AND QUANTIFICATION

As analytical chemists, the highest resolution measurement one can make is at the single molecule level; it just does not get any better than that.

— David R. Walt (Walt, 2013).

The field of protein detection and quantification encompasses a great variety of methods, ranging from non-specific, total protein detection to highly specific immunoassays and high-throughput methods such as [mass spectrometry \(MS\)](#). Some of these methods allow distinguishing present proteoforms and thus offer a higher degree of information. In the following, the major differences regarding the general approaches for protein detection and signal generation will be described and compared.

2.1 TOTAL PROTEIN DETECTION

Methods for total protein detection and quantification is a group of methods detecting all proteins in a sample. Such methods cannot distinguish between distinct proteins, modified proteins or interacting proteins because they rely on general interactions of the detection agents with the amino acids. Thus, their sensitivity greatly depends on the presence of certain amino acids in the primary structure of proteins. These methods were among the first ones to be developed for the detection of proteins in biological samples, with their inventions dating back to the mid-20th century.

The simplest methods are photometric methods, which exploit the inherent feature of proteins to absorb ultraviolet light. This does not require further reagents. The absorbance (A) can then be transformed into the concentration (c) using the molar absorption coefficient of the protein (ϵ) and the path length of the light (l). This relation is called *Beer-Lambert-Law*:

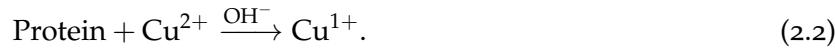
$$c = \frac{A}{\epsilon l} \quad (2.1)$$

The absorbance of proteins can be measured at 205 nm, where the peptide bond contributes the most to the absorbance (Anthis & Clore, 2013). At 280 nm, the aromatic amino acids tryptophan (W) and tyrosine (Y) as well as disulfide bonds (S-S) contribute the most to the

The molar absorption coefficient can be calculated with the approach from Pace et al. (1995): $\epsilon_{280} = 5500n_W + 1490n_Y + 125n_{S-S}$, where n_W denotes the number of tryptophan amino acids, n_Y denotes the number of tyrosine amino acids and n_{S-S} denotes the number of disulfide bonds.

absorbance (Stoscheck, 1990). The molar absorption coefficient needs to be adapted (ϵ_{205} and ϵ_{280}) and necessarily depends on the presence of certain amino acids in the protein (Grimsley & Pace, 2003). This implies that the signal varies between proteins because of their primary structure; for example, proteins with few aromatic amino acids or short proteins might be hardly detectable. Furthermore, detection of proteins in crude cell lysates is complicated because nucleic acids also absorb light of similar wavelengths.

The BCA assay is more specific to proteins and relies on the *Biuret reaction* (P Smith et al., 1985). Under alkaline conditions, proteins reduce copper ions



Reduced copper then forms a complex with the BCA, whose absorbance can be measured at 562 nm. The protein concentration can be calculated with the appropriate ϵ_{562} using Equation 2.1. Acidification or chelators of copper such as ethylenediaminetetraacetic acid (EDTA) interfere with the reaction (Stoscheck, 1990).

In a heterogeneous mixture of different proteins, the results of the first two methods will be inaccurate, while in a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), all proteins of a mixture can be distinguished with high accuracy. First, proteins in the mixture are denatured and unfolded by sodium dodecyl sulfate (SDS). Disulfide bonds are disrupted by reducing agents such as dithiotreitol. This results in a uniform, negative charge of all proteins rendering size the predominant difference. The sample is then loaded onto a polyacrylamide gel. After applying an electric field, the negatively charged proteins travel through the gel mesh to the anode and are separated by size due to differential migration rates through the gel. The velocity of protein travel is inversely proportional to their size (Laemmli, 1970). After separation, proteins can be detected using stains such as Coomassie Brilliant Blue, which interacts with hydrophobic amino acids (Diezel et al., 1972; Stoscheck, 1990). Using a reference standard of known concentration, the proteins can be quantified. However, very large or very small proteins cannot be detected, and under denaturing and reducing conditions PPIs cannot be detected, either. To distinguish proteins of the same size a 2-dimensional SDS can be used (Klose, 1975; O'Farrell, 1975). In the first dimension, proteins are separated by isoelectric point using isoelectric focusing (Bjellqvist et al., 1982). This depends on the presence of charged amino acids in the primary structure. Proteins are then separated by size and detected using stains as described above.

2.2 MASS SPECTROMETRY

Similar to the total protein detection methods described in (Section 2.1), **MS** represents an untargeted and thus data-driven approach. However, it enables the distinction of proteins in a mixture, e.g. crude cell lysate. Liquid chromatography with tandem mass spectrometry (**liquid chromatography with tandem mass spectrometry (LC-MS/MS)**) is the most widespread method for proteomic analysis (Cox & Mann, 2011), among a vast number of variations and similar methods (Da Silva Araújo & Lima Tavares Machado, 2020). **MS**-based methods aim for large-scale proteomic studies, for instance, to draft the human proteome (Wilhelm et al., 2014), to relate the transcriptome to the proteome (D Wang et al., 2019) or to understand global splicing events and proteoforms (Sinitcyn et al., 2023).

In general, two main approaches can be distinguished: *top-down* and *bottom-up* proteomics. In bottom-up proteomics, the proteoforms are fragmented, separated by size in a high-performance liquid chromatography, ionized and then subjected to the mass analyzer measuring the *mass-over-charge* ratio. Typically, this procedure is followed by a database search for the identification of the analyzed fragments (Cassidy et al., 2023; Domon & Aebersold, 2006). One drawback is that through the fragmentation, information about the proteoforms gets lost. In contrast, top-down proteomics lack this fragmentation process, which maintains the proteoform for the mass analyzer, but this comes at the cost of lower throughput (Da Silva Araújo & Lima Tavares Machado, 2020). In both cases, however, reference standards are required for quantification, which either interfere with the biological system (e.g. heavy isotope labeling) or differ from the analyte (Calderón-Celis et al., 2018; Lindemann et al., 2017; Shuford et al., 2017). This adds an unknown fraction of uncertainty to the quantification. Furthermore, strong biases originate from highly abundant proteins, which can mask the signal of low abundant proteins (Domon & Aebersold, 2006). Approaches for improvement combine **MS** with immunoassays for sample enrichment (RW Nelson et al., 1995) or enhanced precision (Kumar et al., 2018).

2.3 IMMUNOASSAYS

In contrast to total protein detection and quantification (Section 2.1) and **MS**-based methods (Section 2.2), immunoassays are targeted methods, specific for one or multiple proteins or proteoforms of interest. This means that low abundant proteoforms can be detected and quantified, even if outnumbered by other, irrelevant biomolecules. This implies that immunoassay are hypothesis-driven methods, which require prior knowledge, potentially gained from data-driven approaches

such as [MS](#) (Cox & Mann, 2011). However, the superior specificity and sensitivity of immunoassays comes at the cost of lower throughput.

2.3.1 General principle

Biorecognition can also be mediated by antibody fragments such as fragment antigen binding (Fab), single chain variable fragments or half antibodies (Hermanson, 2013).

Antibodies ([Section 2.3.2](#)) are used as biorecognition elements in immunoassays (Hermanson, 2013). This enables the high specificity and sensitivity of immunoassays for the detection and quantification of specific proteoforms. The labeled antibodies transform the proteoform level into a detectable signal by using:

- enzymes, which convert a colorless substrate into a bright product (Aydin, 2015);
- radioactivity (Yalow & Berson, 1959);
- fluorophors (Coons et al., 1941); or
- DNA, which can be detected by PCR (Sano et al., 1992)¹.

Using a reference standard with known concentration, the signal can be used for absolute quantification.

Immunoassays can be performed in two different modes: homogeneous and heterogeneous. The majority of immunoassay are heterogeneous (Aydin, 2015). This involves capturing the antigen by binding it to a solid phase, either directly or indirectly via capture antibodies ([Figure 2.1](#)). After antibody-antigen binding, unbound molecules are removed by washing. This renders such assays laborious, but it increases sensitivity and reduces matrix effects (Selby, 1999; Wood, 1991). Typical examples are:

- Western blot, detection of proteins of interest after [SDS](#) ([Section 2.1](#));
- Immunohistochemistry, detection of proteins *in situ* in formalin-fixed paraffin-embedded tissue sections (Hrycaj, 2023); and
- Enzyme-linked immunosorbent assay (enzyme-linked immunosorbent assay (ELISA)) (Aydin, 2015).

Furthermore, the choice of assay format depends on the application. The simplest format employs one antibody, which carries the label ([Figure 2.1A](#)). This format is rarely used because of high background signals and low specificity. In Western blots and immunohistochemistry, the standard format is shown in [Figure 2.1B](#), where the secondary antibody carries the label and was raised against the [fragment crystallizable \(Fc\)](#)-region of the primary antibody. It has been shown that the usage of two antibodies enhances specificity (Y Liu et al., 2011;

¹ In this regard, Klebes et al. (2024) coined the term *protein-to-nucleic acid transformation*.

Padhan et al., 2017). Such sandwich setups are frequently used in ELISAs (Figure 2.1C and D).

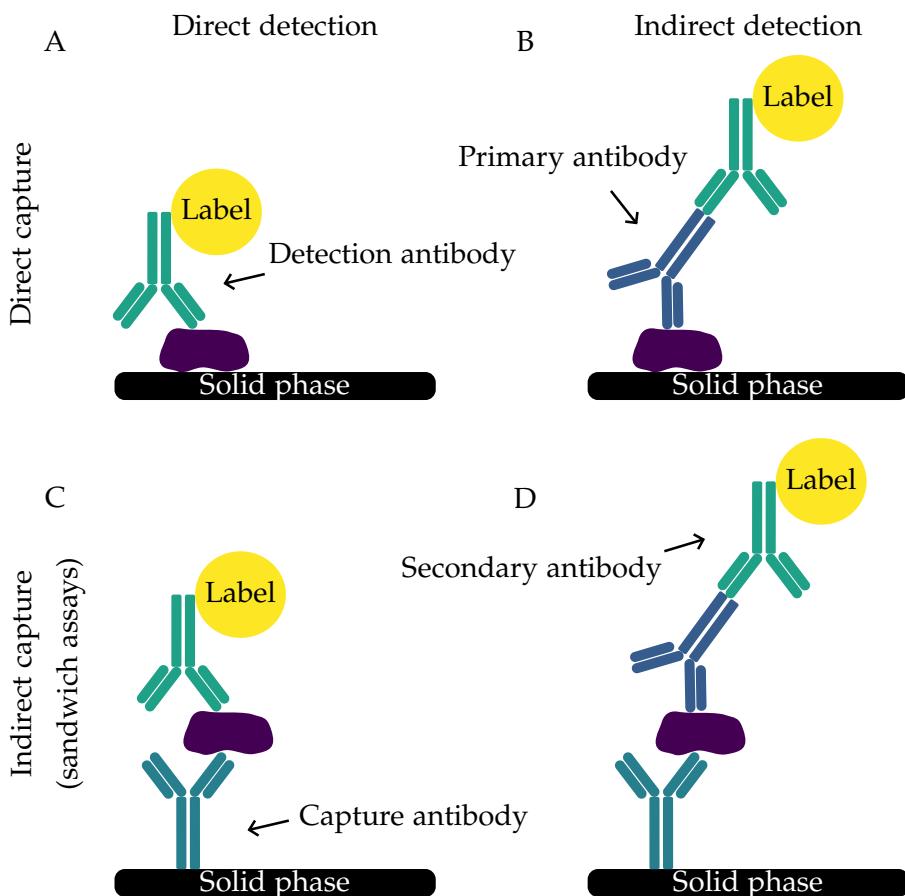


Figure 2.1: Detection and capture formats of heterogeneous immunoassays.

A: Direct assay format.

B: Direct capture-indirect detection format. The secondary antibody carries the label and recognizes the Fc-region of the primary antibody. This format is typically used in Western blot and immunohistochemistry.

C: Sandwich assay format with direct detection. The antigen is recognized by two antibodies. This enhances specificity (Y Liu et al., 2011; Padhan et al., 2017). This format is also applied in homogeneous immunoassays. Then, no solid phase is involved, and usually both antibodies are labeled.

D: Sandwich assay format with indirect detection.

In contrast, homogeneous immunoassays (Jenkins, 1992; Khanna, 1991) are less laborious because they do not require washing steps, although this increases the potential for non-specific interferences potentially impairing sensitivity and specificity (Jenkins, 1992; Selby, 1999; Wood, 1991). All reactants are available at any time during the reaction, which permits application of equilibrium binding equations (Douglass et al., 2013; Han, 2020; Yang & Hlavacek, 2011), eventually allowing

the reference-free absolute quantification of proteoforms (Gross et al., 2024). Homogeneous assays are mainly conducted in a sandwich format (Figure 2.1C)² and are considered to have less antibody cross-reactivity (Fredriksson et al., 2007).

2.3.2 Antibodies

Glycoproteins are proteins that were modified by attachment of sugar chains (see PTM).

Antibodies are most frequently used as biorecognition elements in immunoassays and originate from the adaptive immune system. They are glycoproteins consisting of two light chains and two heavy chains linked by disulfide bonds (Figure 2.2A). The Fab mediates the antigen binding and the Fc is constant and mediates the communication with immune cells.

In response to foreign molecular structures such as proteins, peptides, polysaccharides, lipids or nucleic acids, lymphocytes (plasma cells) produce antibodies. Because these foreign structures trigger the production of antibodies, they are called “antibody generators”, short antigens (Murphy & Weaver, 2018). Antibody generation is random. Lymphocytes undergo somatic recombination in a process called variable-diversity-joining rearrangement (Tonegawa, 1983). In this way, myriads of lymphocytes emerge, each producing a different antibody. A cell that by chance produces the antibody against the antigen is selectively replicated and so the production of this antibody is increased.

The hypervariable region of the antibody (paratope) located in the Fab-region mediates the binding to the epitope on the antigen (Figure 2.2). This region thus defines an antibody’s specificity. In a living organism, the binding of the antibody to the antigen triggers one of the following immune responses: opsonization, neutralization or agglutination. Either way, the foreign molecular structure is prevented from performing harmful actions. Simultaneously, foreign molecules are marked by the antibody, which is then recognized at its Fc-region by phagocytes. These cells ingest and digest the foreign structure (Murphy & Weaver, 2018). The immune response can be mediated by different classes of antibodies, e.g. IgA, IgD, IgE, IgG or IgM. In the blood, the primary class is IgG (Murphy & Weaver, 2018). IgGs are mainly used as biorecognition molecules in immunoassays and are of high interest for cancer therapy (Scott et al., 2012). Paragons are trastuzumab (TTZ) and pertuzumab (PTZ) for the treatment of human epidermal growth factor receptor 2 (HER2)-positive breast cancer (Nahta et al., 2004; Nami et al., 2018; Sharma et al., 2021; Tsao et al., 2022).

Lymphocytes also present immunoglobulins as receptors on their surface. The term antibody refers to secreted immunoglobulins.

² Homogeneous sandwich assays do not employ solid phases, obviously.

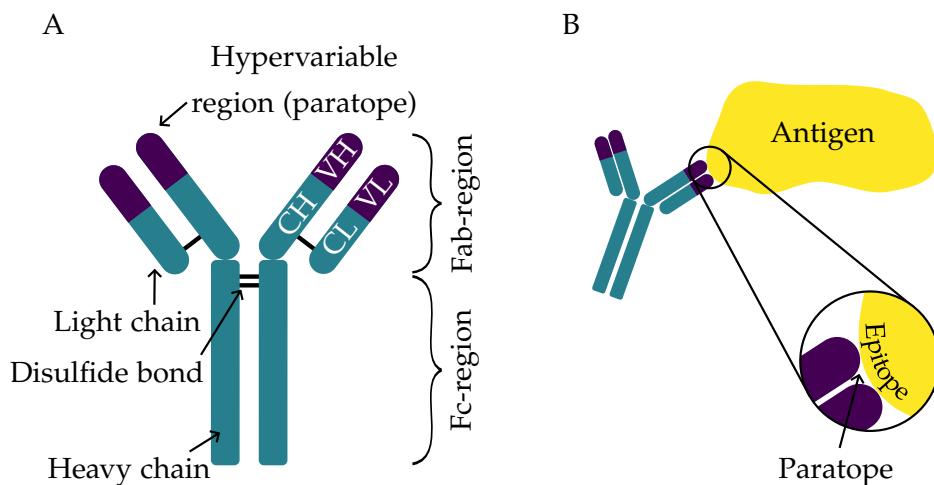


Figure 2.2: Monoclonal antibody.

A: Characteristic Y-shaped depiction of a monoclonal antibody. The two light chains and heavy chains are linked by disulfide bonds. The hypervariable region for antigen binding consists of the variable domain of the heavy chain (VH) and the variable domain of the light chain (VL) and is located in the Fab-region. The constant domain of the heavy chain (CH) and the constant domain of the light chain (CL) as well as the Fc-region are constant. Besides, the Fc-region mediates the communication with immune cells.

B: The hypervariable region of the antibody, the paratope, recognizes its counterpart, the epitope, on the antigen.

PRODUCTION Immunoassays and cancer therapies require the controlled production of antibodies against the desired antigens. In general, three techniques for antibody production with distinct outcomes can be distinguished. First, animals such as rabbits, mice or rats are injected with the antigen of interest. This triggers an immune response eliciting antibodies, which can be extracted as an antibody cocktail from the blood serum. These cocktails are called polyclonal antibodies because they all recognize the same antigen but at different epitopes (Murphy & Weaver, 2018). The second technique is the hybridoma technique, where lymphocytes of immunized mice are fused with myeloma cells resulting in a stable cell line producing a monoclonal antibody against one specific epitope of an antigen (Köhler & Milstein, 1975). Third, antibodies can be recombinantly produced. The genetic information for the hypervariable region of the antibody is selected by phage-display (Neri et al., 1995). Then, Chinese hamster ovary (CHO) cells are correspondingly genetically engineered and cultured (Kunert & Reinhart, 2016). Afterwards, the antibody can be purified from the cell culture supernatant. This process allows for great flexibility for the choice of antigens. Therefore, antibodies against proteoforms can be produced (Hattori & Koide, 2018; Hattori et al., 2013). Recombinant antibodies are monoclonal because they are produced by monoclonal cell lines.

CONJUGATION For cancer therapies, antibodies are often conjugated with cytotoxic agents (antibody-drug conjugates), so that the tumor cell ingesting the antibody induces its death (Hermanson, 2013).

For immunoassays, antibodies need to be conjugated with a label to become detectable. Thus, this form of conjugation is often referred to as *antibody labeling*. In this technique, it is critical that the functionality of the antibody regarding specificity and affinity is preserved (Chio & Bane, 2020; Gong et al., 2016; Hermanson, 2013).

2.3.3 Protein-to-nucleic acid transformation

In the following, the focus will be on immunoassays that employ DNA-labeled antibodies for detection and quantification of proteoforms, i.e. protein-to-nucleic acid transformation (Klebes et al., 2024). In contrast to the central dogma of biology (Figure 1.3), this inverts the flow of information. The rationale behind this is to take advantage of the simplicity of DNA for amplification and detection, while preserving the rich information from the proteoforms (Janssen et al., 2013; Nong et al., 2012). Amplification of DNA simultaneously generates fluorescence, either by sequence-specific hydrolysis probes or by DNA-intercalating dyes (Figure 2.3), which can then be detected by an optical system (Biassoni & Raso, 2014). The infrastructure for DNA amplification and detection is vast, e.g. quantitative PCR, digital PCR and NGS (Khodakov et al., 2016; Zhao et al., 2021).

The first immunoassay employing this idea was *immuno-PCR*, which showed an improvement of sensitivity by five orders of magnitude compared to conventional ELISA (Sano et al., 1992). The very first version of immuno-PCR used an agarose-gel and a DNA-intercalating dye to detect the amplified DNA (Sano et al., 1992). This is considered an analog detection principle.

An amplicon is the section of a DNA template that is flanked by the binding sites of the primers and is thus amplified during PCR.

Antigens become detectable only upon binding of DNA-labeled antibodies.

ANALOG DETECTION PRINCIPLES In analog detection, the signal of a mixture of multiple DNA-labeled antibodies (bulk signal) is measured by some type of optical system (Duffy, 2023). The DNA-labeled antibodies are combined with the sample in bulk. Like immuno-PCR, *in situ proximity ligation assay (PLA)*, homogeneous and heterogeneous PLA and *proximity extension assay (PEA)* are representatives of analog immunoassays (Table 2.1). In PLA and PEA, proteoforms or PPIs are detected by two DNA-labeled antibodies (sandwich format, similar to Figure 2.1C), which perform the protein-to-nucleic acid transformation (Fredriksson et al., 2002; Lundberg, Thorsen, et al., 2011). *In situ PLA* uses *rolling circle amplification (RCA)* for amplification and fluorescence generation, while PLA and PEA use qPCR. The difference between PLA and PEA is the generation of the first DNA template. In both methods, the binding of two antibodies to proximal epitopes (not necessarily on the same antigen) brings their DNA-labels

in proximity. In **PLA**, a connector oligonucleotide hybridizes to both labels and thus enables ligation (Figure 2.4A).

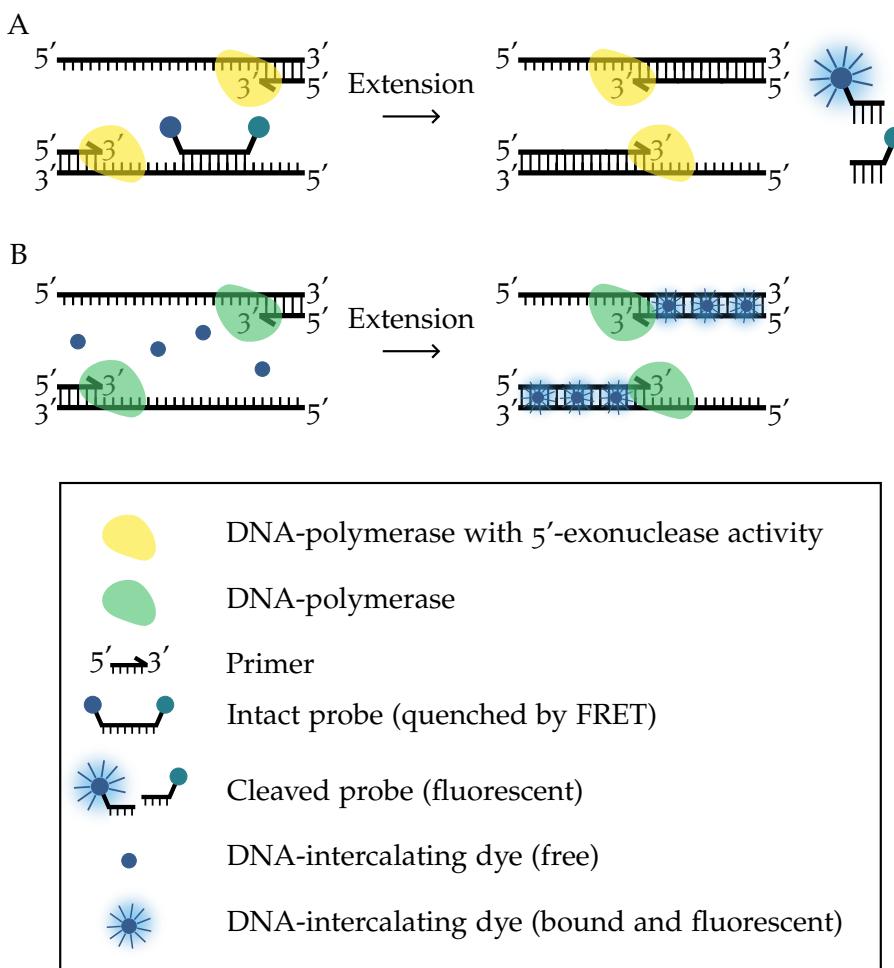


Figure 2.3: DNA detection chemistry (inspired by Basu (2017)).

A: Hydrolysis probes (e.g. TaqMan probes) are complementary to a short DNA sequence within the amplicon and anneal during primer annealing. During extension, a DNA-polymerase with 5'-exonuclease activity (e.g. *Taq* DNA-polymerase) cleaves the probe. 5'-dye and 3'-quencher thus lose spatial proximity and fluorescence can be excited with light of a wavelength matching with the dye (Jothikumar et al., 2009). This process is called **Förster resonance energy transfer (FRET)** (Förster, 1948). For enhanced sensitivity, **minor groove binder (MGB)** quenchers can be used (Kutyavin, 2000).

B: DNA-intercalating dyes intercalate into double-stranded DNA. Only then, the dye can be excited with light of a matching wavelength to emit fluorescence. This process is independent of the DNA sequence but fluorescence intensity depends on the length of the amplicon.

This construct is then amplified and detected by qPCR using sequence-specific hydrolysis probes, which bind only to the newly formed DNA construct (Figure 2.3A). For increased sensitivity and specificity,

the connector oligonucleotide can also be conjugated to a third antibody, which also binds in the proximity of the first two antibodies (Schallmeiner et al., 2007). In PEA, the DNA-labels are designed in such a way that they have a complementary overlap of nucleotides, which hybridizes when in proximity (Figure 2.4B). The overlap is then extended by a DNA-polymerase eliminating the need for the more fragile DNA-ligases (Lundberg, Eriksson, et al., 2011; Lundberg, Thorsen, et al., 2011). This newly formed construct is then amplified and detected sequence-specific hydrolysis probes (Figure 2.3A) in a qPCR. In addition, both methods allow for amplification and detection by NGS (Darmanis et al., 2011; Wik et al., 2021). Although fundamentally different, PEA showed good complementarity to LC-MS/MS (Petrera et al., 2021).

DIGITAL DETECTION PRINCIPLES In digital detection, the signal of single DNA-labeled antibody (single molecule resolution) is measured by some type of optical system. Like in analog detection, DNA-labeled antibodies and antigens bind in bulk. However, single DNA-labeled antibodies are then obtained by partitioning (Basu, 2017; Duffy, 2023). This means the bulk solution with DNA-labeled antibodies bound to antigens, measured in its entirety via analog detection (see above), is divided into numerous minute reaction partitions, each of which contains less than a single molecule on average. In each partition, DNA is amplified separately and fluorescence for DNA detection is generated independently (Figure 2.3). Then, the fluorescent partitions are counted and the number of DNA-labeled antibodies can be determined (see Section 4.2.1 for DNA quantification in dPCR) from which the number of proteoforms can be inferred. Single molecule resolution is the ultimate resolution with the best sensitivity (Walt, 2013). Representatives of digital immunoassays are digital PLA (dPLA), digital droplet immunoassay (ddIA) (see below Section 2.3.4) and single molecule array (Simoa) and related assays (Table 2.1). Of note, not all of these digital immunoassays require DNA-labeled antibodies, e.g. Simoa and droplet digital ELISA (ddELISA) employ enzyme-labeled antibodies (Table 2.1)³.

³ Simoas are heterogeneous, bead-based assays. Magnetic beads are coated with a capture antibody. The beads are mixed with the sample and the antigens bind to the capture antibodies. Then, detection antibodies labeled with either DNA or an enzyme are added and bind to the antigen, too. Partitioning happens by loading the beads into femtoliter-sized wells and the fluorescence is generated (either enzymatically or by DNA amplification). Then, fluorescent partitions can be counted and the number of proteins can be inferred (Chang et al., 2012; Rissin et al., 2010). Variants of the conventional Simoa are dropcast Simoa (dSimoa) (Wu et al., 2020), ddELISA (Cohen et al., 2020) and molecular on-bead signal amplification for individual counting (MOSAIC) (Wu et al., 2022). The MOSAIC assay counts positive beads similarly to the counting of positive droplets by the dPCR system of Bio-Rad in a fluorescent-activated cell sorting (FACS)-like manner (BJ Hindson et al., 2011).

2.3.4 dPCR for proteoform detection

dPLA and ddIA employ dPCR for partitioning, DNA amplification and fluorescence detection. The chemistry for DNA amplification and simultaneous fluorescence generation in dPCR is the same as for qPCR in analog detection (Figure 2.3). However, prior to thermocycling, the reaction volume is partitioned into numerous, minute reaction partitions. These can either be droplets (water-in-oil emulsion, e.g. droplet digital PCR) or physical chambers (i.e. chamber digital PCR) (Basu, 2017; Quan et al., 2018; Tan et al., 2023). DNA-labeled antigens⁴ are thus separated from each other into individual partitions and then amplified using primarily sequence-specific probes for fluorescence generation (Figure 2.3A) (Jothikumar et al., 2009). After amplification, each partition can either be negative or positive for the DNA-labeled antigen. Because of this binary signal, like in computer sciences, Vogelstein and Kinzler (1999) coined the term *digitalPCR* (Morley, 2014). Positive partitions can be simply counted and by using the Poisson distribution, the absolute number of DNA molecules can be calculated⁵ (Basu, 2017; Moon et al., 2011; Quan et al., 2018).

DPCR has several advantages over analog DNA detection by qPCR. Partitioning increases the effective target concentration enabling single molecule detection (BJ Hindson et al., 2011; CM Hindson et al., 2013). It also enriches the target compared to other substances increasing the signal-to-noise ratio by reducing background signals from matrix effects (Quan et al., 2018; Wood, 1991). Therefore, dPCR has a better amplification efficiency, precision and linearity, simply because single molecules are counted rather than a mixture (Basu, 2017). This enables to resolve fold-difference down to 1.16-fold between two DNA targets (CM Hindson et al., 2013), while qPCR can only distinguish 2-fold differences because DNA is doubled with each cycle. Furthermore, dPCR is an endpoint analysis with a binary signal (partition positive or negative for a target). This makes it less dependent on the properties of the detector and the assay chemistry (Basu, 2017). Current dPCR systems use two distinct approaches for partition status identification. The Droplet Digital PCR system from Bio-Rad aspirates the droplets (water-in-oil) from the reaction tube after thermocycling, separates them by injecting a spacer fluid and transports them to a fluorescence detector, where their status is detected in single file as they pass the detector (BJ Hindson et al., 2011). This process is comparable to FACS. The naica system from Stilla (Madic et al., 2016) and the QIAcuity system from QIAGEN use an automated fluorescence microscope to image a monolayer of droplets or chambers. The dPCR devices can detect up to six different fluorescent dyes. DPCR can also be

The number of DNA-labeled antigens per partition depends on their concentration prior to partitioning, the number of partitions, and their volume. The average of DNA-labeled antigens per partition is defined as λ (see Section 4.2.1).

⁴ After protein-to-nucleic acid transformation using DNA-labeled antibodies, antigens are technically DNA-labeled.

⁵ The mathematical background of the Poisson distribution for dPCR evaluation is described in Section 4.2.1.

performed in “home-made” devices (Lai et al., 2023; Schulz et al., 2020, 2021). An overview of the current, commercially-available dPCR systems is provided in Tan et al. (2023).

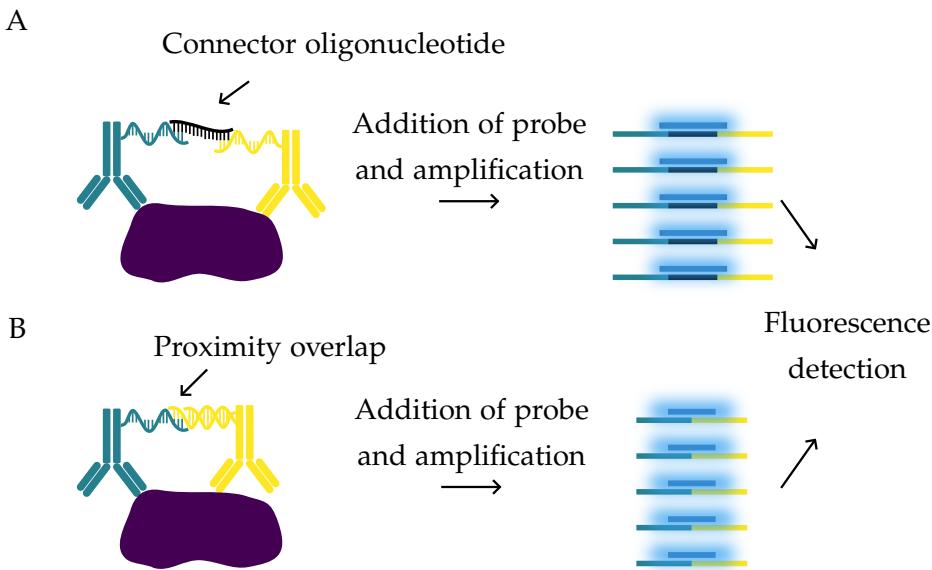


Figure 2.4: Generation of first DNA template for amplification PLA and PEA. After generation, the DNA template amplified and detected by qPCR using a probe specific to the just generated DNA template (Figure 2.3A).

A: In PLA, two antibodies bind to proximal epitopes bringing the DNA-labels in proximity. A connector oligonucleotide hybridizes to both DNA-labels. This connection is stabilized by DNA ligation.
B: In PEA, two antibodies bind to proximal epitopes bringing the DNA-labels in proximity. The labels were designed in such a way that they have a complementary overlap of several nucleotides. This overlap is extended by a DNA-polymerase.

The research group of Savaş Tay invented dPLA, which uses dPCR for the detection and quantification of proteins (Abasiyanik et al., 2020; Albayrak et al., 2016; Lin et al., 2019; Vistain et al., 2022). The DNA-labeled antibodies bind their antigens in bulk, so that the DNA-labels come in proximity and can be ligated using a connector oligonucleotide and a DNA-ligase. Then, the reaction is partitioned, and the DNA is amplified. Fluorescence is only generated when both antibodies are bound. This means a positive partition equals one protein detected by both antibodies. In this way, homogeneous dPLA leverages the sensitivity of dPCR to improve assay performance as homogeneous assays usually have lower sensitivities compared to heterogeneous assays (Table 2.1 and Section 2.3.1). The formation of simultaneous binding of both antibodies is secured in bulk by ligation, which is only possible when the DNA-labels can “find” each other. However, Vistain et al. (2022) claim that the DNA-labels are long enough to span the size of average proteins. Worth considering for the specificity of

dPLA is the efficiency of preventing spontaneous ligation resulting in false-positive signals. In contrast to that, the approach from Byrnes et al. (2020) features proximity ligation after partitioning. Still two DNA-labeled antibodies can be partitioned in the same partition and thus the issue of spontaneous ligation remains or is even exacerbated. In general, PLA relies on the proximity of the antibodies defined by the proximity of the epitopes, which might not be given in the case of PPI detection. Furthermore, for signal generation the ligation needs to successfully generate the first DNA template for amplification and Lundberg, Thorsen, et al. (2011) experienced that DNA-ligases do not work reliably in all buffer systems (Lundberg, Eriksson, et al., 2011; Lundberg, Thorsen, et al., 2011). Consequently, what is required is a homogeneous immunoassay with a digital detection principle without the need for proximity of DNA-labels and the ability to distinguish false-positive signals. Eventually, this enables the immunoassay to inherit the ability for reference-free, absolute quantification from dPCR to be able to absolutely quantify proteoforms.

The approach investigated in this thesis aims for both of these goals: proximity-independent amplification and absolute quantification of the proteoform.

Table 2.1: Comparison of immunoassay employing DNA-labeled antibodies. All methods apply sandwich capture and detection.

Method	Mode	Maximum targets	Sensitivity (fmol L ⁻¹)	Detection principle	Reference
<i>in situ</i> PLA PLA	heterogeneous (surface) homogeneous	1 6	n/a ~ 1	Analog (<i>in situ</i> by RCA) Analog (qPCR)	Söderberg et al. (2006), Blazek et al. (2013), Blazek et al. (2015) Fredriksson et al. (2002), Gustafsdottir et al. (2005), Fredriksson et al. (2007), Schallmeiner et al. (2007)
PLA	heterogeneous (beads)	35	~ 5	Analog (NGS)	Darmanis et al. (2011)
PEA ^a	homogeneous	1500	~ 9	Analog (qPCR, NGS)	Lundberg, Eriksson, et al. (2011), Genshaft et al. (2016), Wik et al. (2021)
dPLA	homogeneous	741	~ 0.3	Digital and analog (dPCR, NGS)	Albayrak et al. (2016), Lin et al. (2019), Abasiyanik et al. (2020), Vistain et al. (2022)
ddIA	homogeneous	1	~ 7	Digital (polydisperse dPCR)	Byrnes et al. (2020)
SiMoA ^b	heterogeneous (beads)	12 ^c	~ 0.0001	Digital (enzyme-generated fluorescence ^d in femtoliter-sized wells)	Rissin et al. (2010), Rissin et al. (2011), Chang et al. (2012), Rissin et al. (2013), Wilson et al. (2016)
dSiMoA	heterogeneous (beads)	1	~ 0.02	Digital (RCA in dropcasted monolayer film (Kaliyara) Selva Kumar et al., 2020))	Wu et al. (2020)
ddELISA	heterogeneous (beads)	1	~ 0.02	Digital (enzyme-generated fluorescence ^d in droplets)	Cohen et al. (2020)
MOSAIC	heterogeneous (beads)	8 ^e	~ 0.006	Digital (RCA-generated fluorescence detected by FACS)	Wu et al. (2022)

^a Commercialized by Olink.

^b Commercialized by Quanterix.

^c Gilboa et al. (2021) describe a technique for sequential multiplexing, which extends the maximal number of targets significantly.

^d These methods employ enzyme-labeled antibodies and the fluorescence is generated by converting a colorless substrate into a bright product.

^e By using different combinations of bead sizes and fluorescence intensities, this value can be significantly improved (Wu et al., 2022).

3

OBJECTIVES OF THIS THESIS

The overall objectives of this thesis comprise four main parts addressing the presented shortcomings of existing digital immunoassays. First, the Protein Interaction Coupling (PICO) technology (Actome GmbH, 2024; Gross et al., 2024) is described and a novel detection principle (**dPCR double positive cluster segregation (dDPCS)** model) is numerically devised. The model is required to calculate the basic detection unit of the PICO technology – complexes – and is based on the mathematical principles of dPCR. Second, a new dPCR simulation with the ability to simulate PICO experiments is created and validated. Third, the **dDPCS** model is validated using simulated data and experimental data. Finally, a PICO assay for the detection of **4EBP1** and one of its crucial **PTMs** is designed and validated. The assay is tested in the context of therapy resistance in **acute myeloid leukemia (AML)**. Each main part contains more specific aims:

PROTEIN INTERACTION COUPLING Chapter 4 comprehensively describes the PICO technology including defining relevant terms, stating underlying assumptions and describing the general workflow. PICO uses dPCR for the amplification and detection of DNA-labeled antibodies. The mathematical background of dPCR is presented and used to devise the **dDPCS** model for calculating the number of complexes, the basic detection unit of the PICO technology, from a dPCR signal. Finally, a model implementation using the programming language **R** is presented.

CREATION OF A DPCR SIMULATION Chapter 5 presents a novel dPCR simulation written in the programming language **R** with the ability to simulate PICO experiments. The dPCR simulation features functionalities to simulate biases that can occur during real dPCR experiments, for instance, false-negative or false-positive clustering biases. Finally, the simulation is validated.

VALIDATION OF THE DDPCS MODEL Chapter 6 is the most extensive chapter of this thesis. It contains the experimental validation of the **dDPCS** model, which is the foundation of the PICO technology. The model calculates the number of *complexes* from a raw dPCR signal. Complexes are ternary complexes that are formed by binding of at least two differently labeled antibodies to a target. Complexes represent the basic detection unit of the PICO technology (Actome GmbH, 2024). The model is validated in two phases with each two

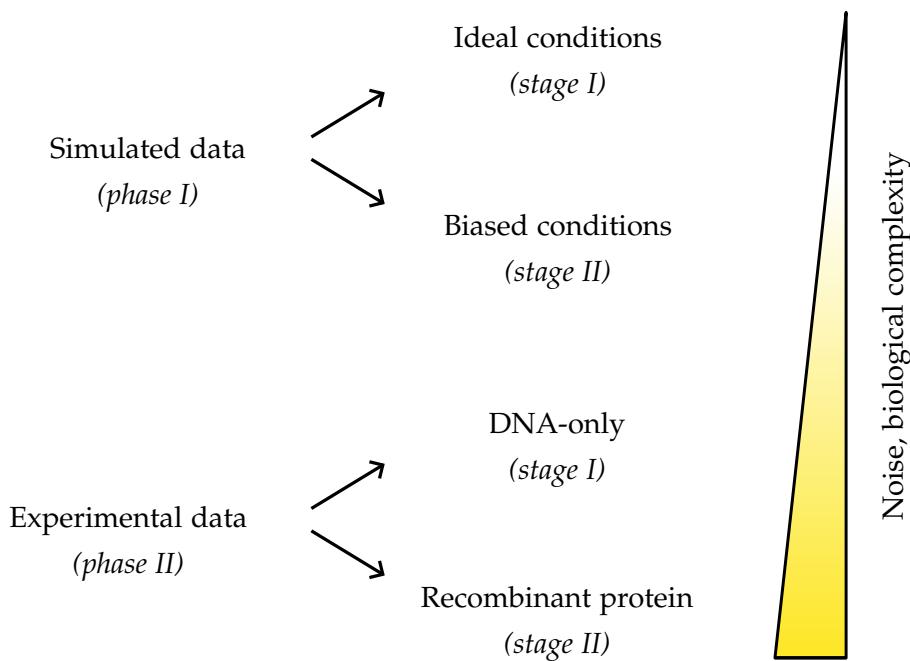


Figure 3.1: Multilevel validation of the dDPCS model.

stages and with an overall increasing amount of noise, i.e. biological complexity (Figure 3.1). In phase I, the model is applied to ideal and biased simulated data from the dPCR simulation created in Chapter 5 to estimate the theoretical, maximal performance, while in phase II the model is applied to experimental data, which inherently contains more noise. In phase II stage I, DNA-only PICO experiments¹ are performed. These prevent artifacts that can originate from protein-to-nucleic acid transformation (e.g. unspecific antibody binding) and thus enable the calculation of expectation values. In phase II stage II, DNA-labeled antibodies are used to transform a recombinant protein into DNA. Analytical performance parameters of the PICO technology are summarized across the phases and stages of the validation procedure. Finally, this chapter conceives design rules for the development of PICO assays.

APPLICATION EXAMPLE: BIOMARKER IN LEUKEMIA Chapter 7 applies the PICO technology to the detection of *4EBP1*, a potentially relevant protein in the context of therapy resistance in *AML*. *4EBP1* represents a putative, predictive biomarker for *hypomethylating agent (HMA)* treatment. Its T37T46-phosphorylation plays a pivotal role in mediating protein synthesis through regulation of mRNA translation. Thus, antibodies against *4EBP1* and its T37T46-phosphorylation are selected, compared and validated using recombinant and cellular

¹ DNA-only PICO experiments only contain the DNA-labels and no antibodies. A synthetic joint of both labels imitates a complex. Other than that the experimental workflow as well as the evaluation remains the same.

⁴EBP1. Anti-phosphorylation antibodies are further validated by controlled enzymatic or mechanistic dephosphorylation of their antigen. Finally, *⁴EBP1* abundance and the degree of T37T46-phosphorylation are compared between relevant *AML* cell lines with and without *HMA* resistance.

PROTEIN INTERACTION COUPLING (PICO)

4

PROTEIN INTERACTION COUPLING (PICO)

PICO is a technology for the detection and quantification of proteoforms and PPIs (Actome GmbH, 2024; Gross et al., 2024). It was invented and patented by Dr. Csaba Jeney (Jeney, 2016, 2023b; Jeney & Koltay, 2020), who is now CEO of [Actome](#), a startup commercializing the PICO technology. It has been applied to confirm CRISPR/Cas9-mediated RANKL knock-outs in mesenchymal stem cells on the protein level (Gross et al., 2021). PICO also contributed to a breakthrough in influenza virus research. While it was previously assumed that influenza viruses enter mammalian cells via binding to sialic acids of glycoproteins or -lipids (Dou et al., 2018), PICO proved in agreement with [PLA](#) that mammalian major histocompatibility complex (MHC) class II surface proteins mediate the entry of bat influenza viruses (Karakus et al., 2019). In the following, the working steps of the PICO technology and the workflow of a PICO assay with its underlying mathematical principles are described and discussed.

4.1 WORKFLOW

PICO is a digital immunoassay based on DNA-labeled antibodies for protein-to-nucleic acids transformation and dPCR. Monoclonal antibodies are preferred for protein-to-nucleic acid transformation (Jeney, 2016). In the PICO technology, two DNA-labeled antibodies are indispensable for signal generation and evaluation ([Figure 4.3](#) and [Figure 4.8](#)). This renders PICO a sandwich immunoassay ([Figure 2.1](#)).¹ Depending on the purpose of the PICO assay (detection of a specific proteoform or [PPI](#)), two antibodies with affinity to the same target of interest but with different epitopes ([Figure 2.2B](#) and [Figure 4.2A](#)) are selected and labeled with distinct single-stranded DNA-labels ([Section 9.2.3](#)). The design of the labeling process pre-determines the stoichiometric ratio between antibody and DNA-label, as each deglycosylated heavy chain can only be covalently linked to a single dibenzozacyclooctyne-conjugated DNA-label via click-chemistry (Chio & Bane, 2020; Gong et al., 2016; Jeney, 2023a). Of note, both antibodies may target the same protein at different epitopes,

¹ The terms *antigen* and *target* can be used interchangeably. However, the term *antigen* refers to a macromolecular structure recognized by an antibody, while the term *target* is PICO-specific and refers to a macromolecular structure recognized by more than one antibody ([Figure 4.2A](#)). This term is borrowed from dPCR, where a *target* is defined as a DNA molecule that is selectively amplified and detected. After antibody-antigen binding, the antigens can be detected by dPCR because the antibodies carry DNA-labels.

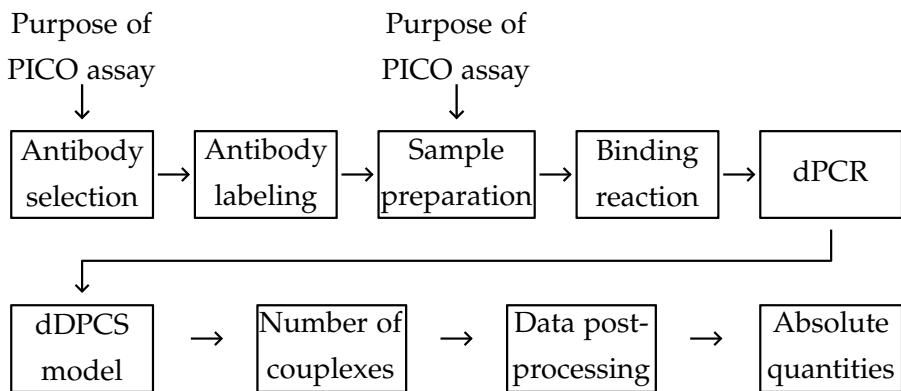


Figure 4.1: Working steps of the PICO technology. The purpose of the PICO assay has substantial effect on the antibody selection and the sample preparation. The data post-processing is extensively described in Tobias Gross' PhD thesis, while this thesis focuses on the dPCR and the [dDPCS](#) model.

or one antibody may target the protein and the other one may target a modification of the same protein, or one antibody may target one protein and the other antibody may target a putative interaction partner (Figure 4.2A). After antibody labeling, the [LE](#) is determined (Section 9.2.5). Then, the antibodies are combined to an equimolar antibody mix ([ABX](#)) (Section 9.3.2 and Figure 4.3A).

The purpose of the PICO assay predetermines not only the selection of antibodies but also the sample preparation. For the detection of cellular proteoforms, the cells need to be lysed and the lysate needs to be homogenized (Section 9.1.6). The detection of [PPI](#) might necessitate chemical crosslinking (Tang & Bruce, 2009), while the detection of recombinant proteins needs thorough solubilization. Then, [ABX](#) and the sample are combined in the [binding reaction](#) (Section 9.3.3 and Figure 4.3B). In the binding reaction, the two DNA-labeled antibodies bind to their target. This step is also referred to as the formation of ternary complexes (Douglass et al., 2013; Han, 2020; Yang & Hlavacek, 2011) (Figure 4.4). In the PICO terminology, ternary complexes are called [couplexes](#) (Figure 4.2B) and represent its basic detection unit (Actome GmbH, 2024; Gross et al., 2024). The binding reaction is incubated overnight to allow it to reach chemical equilibrium. This implies that some targets as well as some antibodies remain unbound, but remain in the solution. This renders PICO a homogeneous immunoassay, meaning that all reactants (Figure 4.4) are available at any time during the reaction. In general, a homogeneous immunoassay is an assay in which both bound and unbound antibodies are not separated by capturing and washing steps (Jenkins, 1992; Khanna, 1991).

A chemical equilibrium is a state of a chemical reaction, where the concentrations of the reactants and products do not change upon prolonged incubation time. It can only be reached in a homogeneous assay.

Couplexes can also be described as antibody-target-antibody sandwiches.

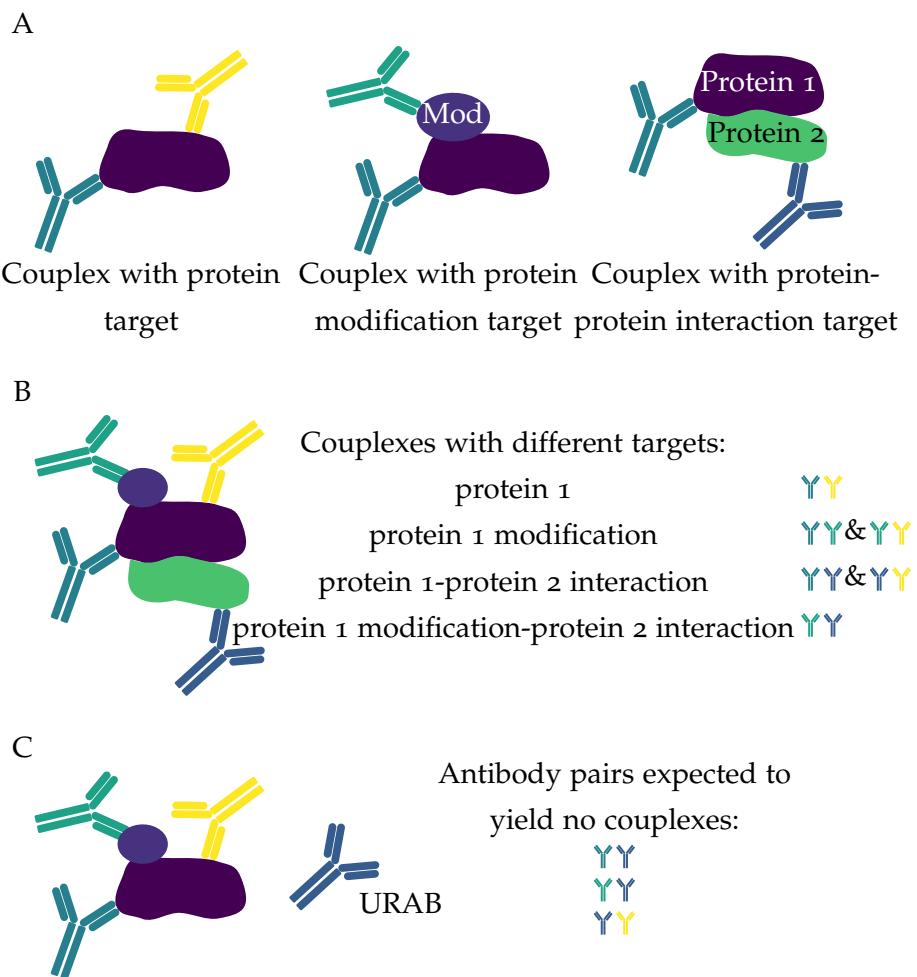


Figure 4.2: Definition of PICO-relevant terms. All antibodies carry DNA-labels, which are not shown in this depiction.

A: Possible complexes depending on the choice of antibodies. Complexes are formed during the binding reaction through the simultaneous binding of two antibodies. Complexes can also be described as antibody-target-antibody sandwiches or ternary complexes. A target can be a protein, a PTM or a PPI. This allows distinguishing proteoforms and understanding PPI of these.

B: Multiplex PICO assay for the assessment of distinct features of the same protein in the same reaction. Although more than 2 antibodies are present in the reaction, the readout is antibody pairwise. The number of antibody pairs and thus the number of different complex kinds depends on the number of antibodies w and is given by $\binom{w}{2} = \frac{w(w-1)}{2}$. A specific antibody might be evaluated with respect to different features of the same target. In this case, the yellow antibody is involved in protein 1 modification and protein 1-protein 2 interaction detection.

C: Unrelated antibody (URAB) control in PICO assays. The URAB has no affinity to the target of interest and was ideally raised against an artificial epitope like His-tag to minimize cross-reactions. Antibody pairs employing the URAB are not expected to form complexes.

After incubation, the binding reaction is diluted² and subjected to the dPCR. During the first step of the dPCR, unbound antibodies, complexes as well as unbound targets and other molecular species are separated into individual partitions, i.e. partitioning. The information of a binding event between antibodies and targets in form of a complex is encapsulated and thus preserved (Figure 4.3C). From this point on, the dissociation of complexes becomes irrelevant for their detection, while a dissociation before partitioning is detrimental. After partitioning, the DNA-labels are amplified and fluorescence is generated using DNA-label-sequence-specific hydrolysis probes (Figure 2.3A). dPLA also applies dPCR for the detection of DNA-labels, but the information of antibody-target binding is preserved by proximity ligation instead of partitioning (Abasianik et al., 2020; Albayrak et al., 2016; Gross et al., 2024; Lin et al., 2019; Vistain et al., 2022). This requires spatial proximity of the DNA-labels to be ligated. After amplification, the partitions are imaged and counted in the detection channels corresponding to the dyes of the hydrolysis probes for detection of the DNA-labels. Afterwards, the partitions are assigned to one of the following groups in a process called clustering: negative (no signal in both channels), single positive (signal in one of the channels), double positive (signal in both channels) (Figure 4.3C). The couplex positive partitions are part of the double positive cluster, because of the use of two DNA-labeled antibodies. Yet, not all double positive partitions contain complexes only but may also contain two unbound antibodies (Figure 4.3C and Figure 4.8). This is the results of random colocalization of both DNA-labeled antibodies, which depends on the number of antibodies in the reaction volume prior to partitioning. Based on this 2-dimensional raw dPCR data, the number of complexes can be calculated using the dDPCS model (Section 4.2). This is another difference when compared to dPLA and ddIA. In those assays, two DNA-labeled antibodies generate only one fluorescence signal (1-dimensional raw dPCR data).

Because some targets remain unbound (Figure 4.3C and Figure 4.4), complexes are not equal to the number of targets of interest in the sample, but are a requirement for calculating the absolute number of proteoforms or PPIs without reference (Gross et al., 2024) (see box **absolute quantification of proteoforms with PICO**). However, under certain assumptions, complexes can be directly compared even if they do not represent all targets of interest in the sample (Gross et al., 2024).

² The dilution is required to maintain evaluability of the dPCR (Section 4.2.1). It would become obsolete in a dPCR with an infinite number of partitions. This would have the positive side effects of reduced partitioning and subsampling errors (Basu, 2017).

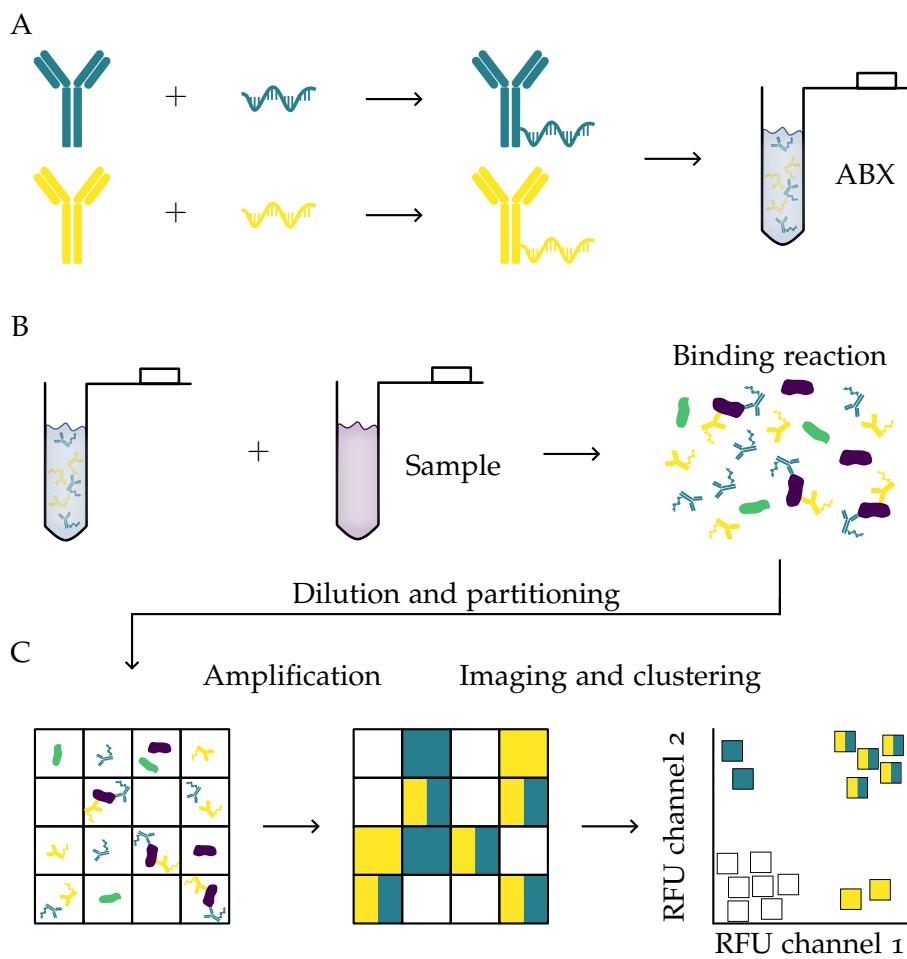


Figure 4.3: Workflow of PICO assays. This example shows a PICO assay with two antibodies. In multiplex PICO assays, there are more antibodies (Figure 4.2B), but the readout remains pairwise.

A: Antibodies against the same target are labeled with distinct single-stranded DNA-labels. The labeled antibodies are then combined to an equimolar ABX.

B: The ABX and the sample containing the target of interest are combined to the *binding reaction*, where the antibodies bind their targets and form complexes (Figure 4.2A). The chemical processes in the binding reaction are illustrated in Figure 4.4. The binding reaction is incubated overnight until chemical equilibrium is reached. Unbound antibodies and targets are not removed by washing. This renders PICO a homogeneous immunoassay.

C: After incubation, the binding reaction is diluted and subjected to the dPCR. After partitioning, and thus encapsulation of binding information, the DNA-labels are amplified and fluorescence is generated using hydrolysis probes (Figure 2.3A). The partitions are imaged and assigned to one of the following clusters based on their fluorescence: negative (no signal in both channels), single positive (signal in one of the channels), double positive (signal in both channels). Using this 2-dimensional raw dPCR data, the dDPGS model calculates the number of complexes.

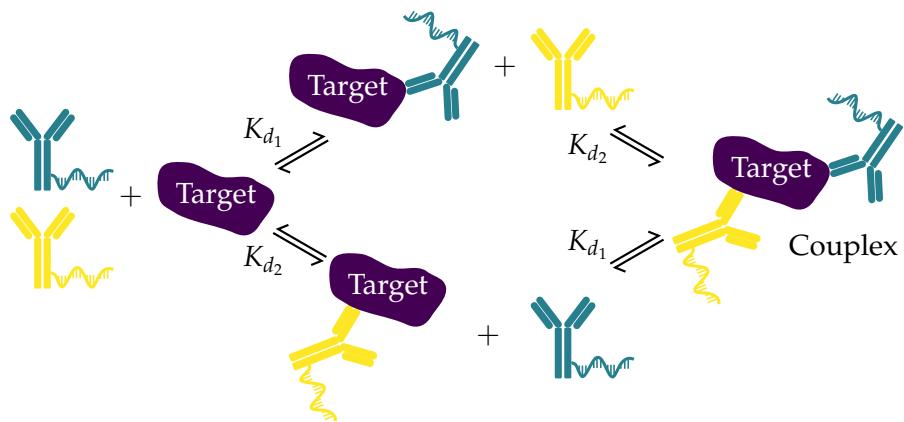


Figure 4.4: Formation of ternary complexes, i.e. *coupling*, during the binding reaction following Yang and Hlavacek (2011), Douglass et al. (2013), Han (2020) and Gross et al. (2024). The selected and DNA-labeled antibodies have an inherent affinity to the target of interest. During incubation the antibodies bind to the target. Some targets remain unbound because the reaction reaches a chemical equilibrium. The models describe two possible routes for the formation of complexes in the binding reaction. The spontaneous formation of a complex as described by Han (2020) with K_{d_3} is considered insignificant. The models assume that the binding of the first antibody does not affect the binding affinity of the second antibody, i.e. no augmentation or diminution effects ($K_{d_1} = \text{const.}$ and $K_{d_2} = \text{const.}$). The underlying mathematics are comprehensively described in the literature cited above, and specifically for the PICO assay in Tobias Gross' PhD thesis.

MULTIPLEX PICO ASSAYS Speaking in dPCR terminology, PICO assays are by definition duplex assays (Whale et al., 2016). For PLA, the usage of three antibodies improves specificity (Schallmeiner et al., 2007). But in contrast to Schallmeiner et al. (2007), additional antibodies in PICO cannot only be used to improve specificity, but also to gain more information on features of the same target or its network. Thus, such a multiplex assay enables information gain on protein abundance, degree of PTM (especially phosphorylation), and number of PPIs (Figure 4.2B). This information is considered to be directly comparable. However, increasing the number of antibodies increases the potential for unspecific cross-reactions (Nong et al., 2012). Independent of the number of antibodies in the ABX, the readout is always pairwise. Thus, for w antibodies, $\binom{w}{2} = \frac{w(w-1)}{2}$ antibody pairs can be evaluated and the number of complexes can be calculated from the corresponding 2-dimensional raw dPCR data. Multiplex PICO assays can have an inherent redundancy as some antibody pairs deliver the same information (Figure 4.2B). These factors lead to questions like:

- How many protein complexes are in the sample?
- How many of the protein complexes are modified?

- How many of the protein complexes interact with other protein complexes?
- How many of the modified protein complexes interact with other protein complexes?

Absolute quantification of proteoforms with PICO

During the binding reaction, antibodies, targets and complexes reach a chemical equilibrium, which is defined by the dissociation constants (K_d values) and the concentrations (Figure 4.4). The model for absolute quantification of the target (i.e. AQ model) relies on reaching a chemical equilibrium and requires the concentrations of all measurable reactants (antibodies and complexes) as well as the K_d values as inputs. Based upon that, the absolute number of targets in the sample can be quantified, even those that are unbound and thus undetectable. Existing models described the concentration of ternary complexes based on the concentration of the targets using the antibody concentrations and their K_d values (Douglass et al., 2013; Han, 2020; Yang & Hlavacek, 2011). In the PICO technology, however, this relationship is inverted and the number of targets becomes calculable from the number of detected complexes (Gross et al., 2024; Jeney, 2023b; Jeney & Koltay, 2020)^a. In contrast to the mentioned publications, the PICO technology provides means to determine the number of ternary complexes, i.e. complexes (Section 4.2).

^a A comprehensive description of the model calculating the number of targets based on the complexes is found in Tobias Gross' PhD thesis. The present thesis, however, provides a framework for reliable complex detection and quantification.

CONTROLS IN PICO ASSAYS An antibody binding control (ABC) is a PICO experiment, where the sample lacks the target of interest and therefore complexes cannot be formed during the binding reaction (Figure 4.3B). The sample in this case is the control buffer (CB) (Actome GmbH, 2024). An ABC PICO experiment can thus be considered as a blank. In ABC PICO experiments, the DNA-labeled antibodies distribute independently in the partitions. Double positive partitions can only be a result of the random colocalization of both DNA-labeled antibodies in the same partition (Figure 4.6 and Figure 4.8). Similarly, an URAB may be added³ as a different control. This antibody has no affinity to the target of interest and was ideally raised against an artificial epitope like a His-tag. Antibody pairs employing this antibody are not expected to form complexes (Figure 4.2C).

A blank is a solution lacking the analyte of interest (according to evisa).

³ Gross et al. (2024) introduced this antibody as bystander antibody (BAB).

4.2 DPCR DOUBLE POSITIVE CLUSTER SEGREGATION (dDPCS) MODEL

For absolute quantification of proteoforms or PPIs, the number of complexes formed during the binding reaction is required (see box [absolute quantification of proteoforms with PICO](#)), which can be calculated from the 2-dimension raw dPCR data ([Figure 4.8](#)). For the purpose of calculating the number of complexes, a novel model, called [dDPCS](#) model, is devised⁴. The [dDPCS](#) model assumes the following:

1. Targets are molecularly dispersed and freely accessible for binding.
2. The antibodies are not chemically selective, which means they can bind the target in any conformational state/proteoform.
3. The labeling of antibodies does not impair their functionality.
4. Both antibodies bind the target simultaneously and show no augmentation or diminution effects.
5. Within the incubation time of the binding reaction, a chemical equilibrium is reached. The law of mass conservation applies.
6. The buffer ensures effective antibody-target binding and suppresses unspecific bindings or cross-reactions.
7. The formed complexes do not dissociate until the partitioning step in the dPCR.
8. Partitioning and subsampling errors are insignificant.
9. Unbound antibodies, targets and complexes distribute independently and freely in all partitions.
10. All DNA-labels are sufficiently amplified so that positive partitions can be clearly distinguished from negative partitions.
11. The assignment of partitions to the corresponding cluster is correct.

4.2.1 Mathematical background

PICO is based on dPCR and inherits its mathematical fundaments, which are established in the following. In a dPCR, m target molecules⁵ distribute perfectly random in n partitions ([Figure 4.5A](#)). It holds that $m \in \mathbb{N}_0$ and $n \in \mathbb{N}$. Let X_i be the number of molecules per partition i

⁴ The [compartment decoupling \(DX\)](#) model presented by Gross et al. (2024) is based on the same assumptions but applies a different implementation ([Section A.1](#)).

⁵ Regarding PICO, the term molecule means DNA-labeled antibody. These expressions can thus be used interchangeably.

($i = 1, 2, \dots, n$). Thus, the number of molecules per partition can be described by a binomial distribution with n and $p = \frac{m}{n}$ (Equation 4.1).

$$X_i \sim B(n, \frac{m}{n}) \quad (4.1)$$

For large n and reasonably small p , X_i can be approximated by a Poisson distribution (Equation 4.2). For dPCR, this approximation is common practice (Basu, 2017; Moon et al., 2011; Quan et al., 2018). The number of molecules for any single partition, say X , follows

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (4.2)$$

where λ is the mean of the Poisson distribution and can be interpreted as the average number of detectable molecules of interest per partition (Equation 4.3) and k is the number of successes, i.e. the number of molecules in a given partition (Basu, 2017). Thus, the Poisson distribution provides an estimate of the probability for a partition to contain a given number of molecules.

$$\lambda = \frac{m}{n} \quad (4.3)$$

In a dPCR, there is no possibility to directly count the number of molecules per partition (Figure 4.5B), while it is possible to observe if a partition contains no molecule ($X = 0$) or at least 1 molecule ($X \geq 1$). If a partition contains at least 1 molecule, it is called *positive*, if it is empty, it is called *negative*. This is the origin of the name *digital* PCR, because this is a binary signal like in computer sciences: 0 or 1 (Basu, 2017; Vogelstein & Kinzler, 1999). However, in real experiments this distinction is not sharp. Some partitions may have an intermediate fluorescence and cannot be unambiguously categorized negative or positive. These partitions are called “rain” (The dMIQE Group et al., 2020). As spontaneous fluorescence is unlikely, they are considered to contain a molecule of interest, which suffered from poor amplification because of various reasons (Lievens et al., 2016). A threshold set by the dPCR system is used to discriminate positive from negative partitions. There are elaborate tools for thresholding (Brink et al., 2018; De Falco et al., 2023). However, for this thesis, the threshold set by the dPCR systems used is sufficient.

After amplification, the number of positive partitions \hat{n}_+ , as well as the number of negative partitions \hat{n}_- can be counted. It holds that $\hat{n} = \hat{n}_+ + \hat{n}_-$ and that $0 \leq \hat{n}_+ < \hat{n}$. The goal is to estimate the number of molecules \hat{m} from these observations. To do so, let Y be a random variable having such two *digital* states, positive and negative. Y can be expressed in the terms of the Poisson distributed measure (Equation 4.2)

$$P(Y = \text{negative}) = P(X = 0) = e^{-\lambda} \quad (4.4)$$

$$P(Y = \text{positive}) = P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda}. \quad (4.5)$$

Usually $n \geq 100$ and $p \leq 0.1$ are considered appropriate to approximate the binomial distribution by a Poisson distribution.

This calculator is a nice tool to get a feeling for the Poisson distribution.

The term negative always refers to negative for a certain molecule. A partition may still be positive for a different molecule and may contain unlabeled and thus undetectable molecules.

The hat over a variable refers to an observed value.

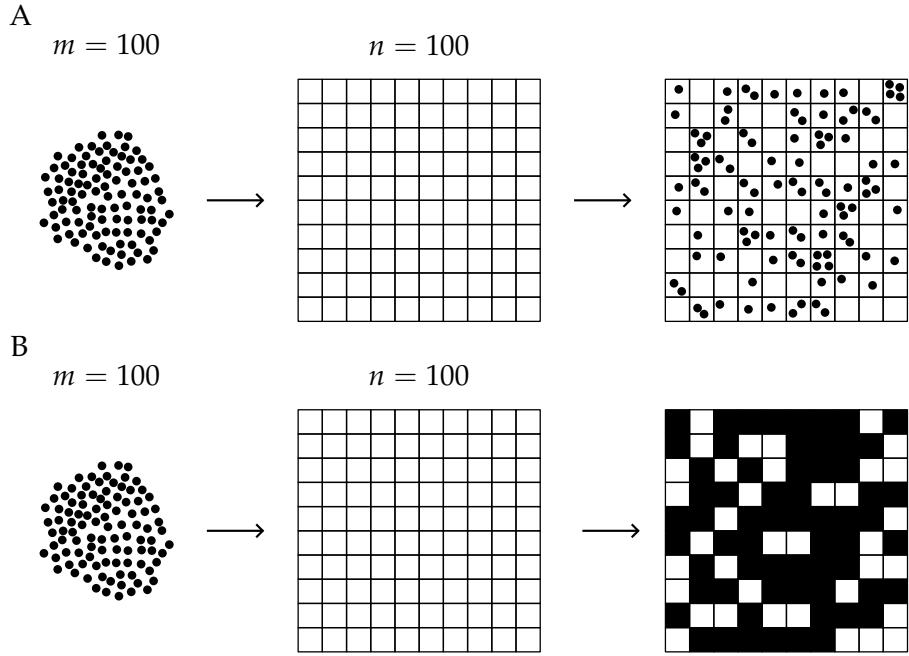


Figure 4.5: Exemplary, random distribution process of $m = 100$ molecules in $n = 100$ partitions. This corresponds to $\lambda = 1$ (Equation 4.3).

A: 100 molecules distribute randomly in 100 partitions. 37 partitions are empty ($X = 0$), 37 partitions contain 1 molecule ($X = 1$), 18 partitions contain 2 molecules ($X = 2$), 6 partitions contain 3 molecules ($X = 3$) and 2 partitions contain 4 molecules ($X = 4$).
B: In a real dPCR experiment, the only information available is the number of negative partitions ($\hat{n}_- = 37$; $X = 0$, white) and positive partitions ($\hat{n}_+ = 63$; $X \geq 1$, black). Plugging the values into Equation 4.7 yields an estimator for $\hat{m} = 99.4$ molecules.

Approximations of expected values are denoted with a \cong sign.

To estimate \hat{m} from the observation, Equation 4.3 can be used in the expectation that \hat{n}_+ approximates $nP(Y = \text{positive})$ well enough. This yields

$$\begin{aligned}\hat{n}_+ &\cong nP(Y = \text{positive}) \\ \hat{n}_+ &\cong \hat{n}(1 - e^{-\frac{\hat{m}}{\hat{n}}}).\end{aligned}\tag{4.6}$$

Solving this for \hat{m} , gives

$$\hat{m} \cong \hat{n}(\ln \hat{n} - \ln(\hat{n} - \hat{n}_+)),\tag{4.7}$$

which is a proof for the absolute quantification of DNA molecules by dPCR. Negative partitions are indispensable, otherwise $\hat{n} - \hat{n}_+ = 0$, which would lead to $\ln(0)$, which is not defined. Also, λ can be calculated from the observed number of positive partitions plugging Equation 4.3 into Equation 4.7.

$$\lambda \cong \ln \hat{n} - \ln(\hat{n} - \hat{n}_+)\tag{4.8}$$

From that, the concentration of the molecule of interest c_m in the partitioned volume $V = \hat{n}V_p$ can be calculated using

$$c_m = \frac{\hat{m}}{V} = \frac{\hat{n}\lambda}{\hat{n}V_p} = \frac{\lambda}{V_p}, \quad (4.9)$$

where V_p is the average partition volume.

In standard dPCR evaluation, this information is enough for the analysis of copy number variations, the detection of circulating tumor DNA after liquid biopsy or the quantification of rare viral or bacterial sequences in infectious diseases (Basu, 2017; Morley, 2014).

4.2.2 Derivation of the model

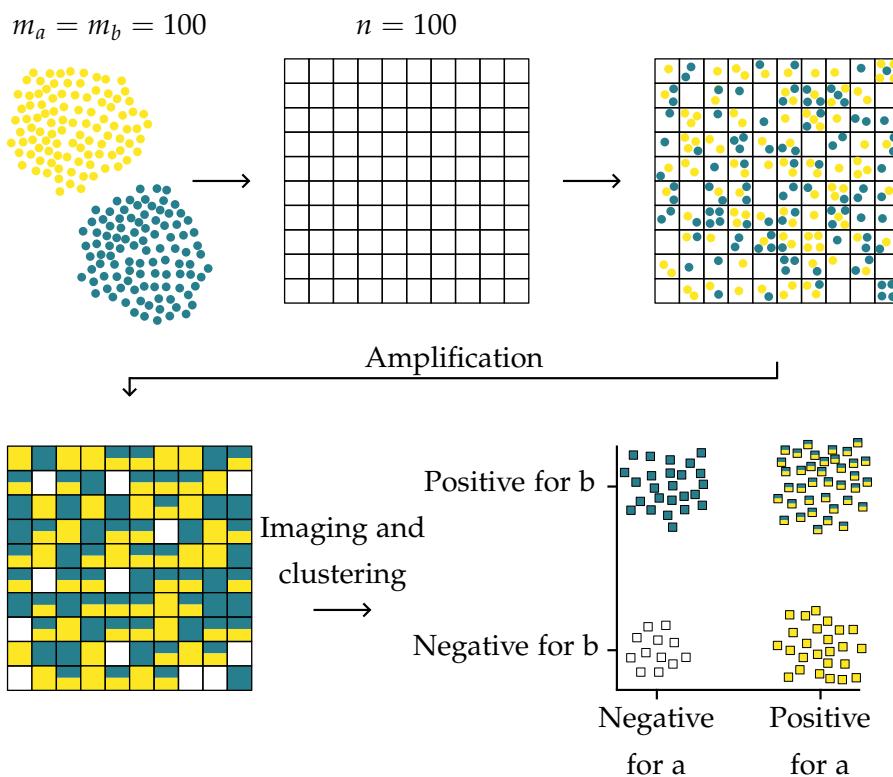


Figure 4.6: Exemplary ABC PICO experiment.

This experiment is similar to Figure 4.5 but now with two molecules a and b (DNA-labeled antibodies). Both molecules have a count of $m_a = m_b = 100$ and randomly and independently distribute in $n = 100$ partitions ($\lambda_a = \lambda_b = 1$). After amplification, imaging and clustering $\hat{n}_{a+b+} = 38$ partitions contain at least one molecule a and one molecule b ($X_a \geq 1 \wedge X_b \geq 1$), $\hat{n}_{a+} = 63$ partitions contain at least one molecule a ($X_a \geq 1$), $\hat{n}_{b+} = 63$ partitions contain at least one molecule b ($X_b \geq 1$) and $\hat{n}_{a-b-} = 12$ partitions contain neither molecule ($X_a = 0 \wedge X_b = 0$). The dDPCS models calculated 2 couples for this example, which is reasonably close to the expectation of 0 couples.

PICO is based upon dPCR and extends its applications towards the detection and quantification of proteoforms and PPIs. The fundamental, mathematical principles of dPCR (Section 4.2.1) can be used for calculating the number of complexes from the double positive cluster. To do so, a new model, called **dDPCS** model, was devised.

For the detection of any target (Figure 4.2A), a PICO assay requires two antibodies (Section 4.1). For the simplest case (ABC PICO experiment), Figure 4.5 can be adjusted by adding a second molecule (now a and b). Let m_a and m_b be the number of molecules or DNA-labeled antibodies a and b. Again, both molecules distribute randomly and independently in n partitions (Figure 4.6). \hat{n}_{a^+} and \hat{n}_{b^+} are then the numbers of positive partitions for molecules a or b, independent of the other molecule. It holds that $\hat{n}_{a+b^-} = \hat{n}_{a^+} - \hat{n}_{a+b^+}$ (and similarly for \hat{n}_{a-b^+}), where \hat{n}_{a+b^+} is the number of double positive partitions. Hence, the total number of partitions n can be written as

$$\hat{n} = \hat{n}_{a^+} + \hat{n}_{b^+} + \hat{n}_{a-b^-} - \hat{n}_{a+b^+}. \quad (4.10)$$

Let Y be the random variable with now four different states: negative for a and b, positive for a, positive for b, and positive for a and b. Then, due to independence

$$\begin{aligned} P(Y = \text{negative for a and b}) &= P(X_a = 0)P(X_b = 0) \\ &= e^{-\lambda_a}e^{-\lambda_b} \end{aligned} \quad (4.11)$$

$$\begin{aligned} P(Y = \text{positive for a}) &= P(X_a \geq 1) \\ &= (1 - e^{-\lambda_a}) \end{aligned} \quad (4.12)$$

$$\begin{aligned} P(Y = \text{positive for b}) &= P(X_b \geq 1) \\ &= (1 - e^{-\lambda_b}) \end{aligned} \quad (4.13)$$

$$\begin{aligned} P(Y = \text{positive for a and b}) &= P(X_a \geq 1)P(X_b \geq 1) \\ &= (1 - e^{-\lambda_a})(1 - e^{-\lambda_b}). \end{aligned} \quad (4.14)$$

Thus, in analogy to Equation 4.6,

$$\hat{n}_{a-b^-} \cong nP(X_a = 0 \wedge X_b = 0) \quad (4.15)$$

$$\hat{n}_{a^+} \cong nP(X_a \geq 1) \quad (4.16)$$

$$\hat{n}_{b^+} \cong nP(X_b \geq 1) \quad (4.17)$$

$$\hat{n}_{a+b^+} \cong nP(X_a \geq 1 \wedge X_b \geq 1). \quad (4.18)$$

After amplification and clustering, the number of positive partitions \hat{n}_{a+b^+} , \hat{n}_{a^+} , \hat{n}_{b^+} and the number of negative partitions \hat{n}_{a-b^-} can be determined by counting (Figure 4.6). In expectation, that the observation \hat{n}_{a+b^+} can be used to approximate $nP(Y = \text{positive for a and b})$ (Equation 4.18), Equation 4.14 can be rewritten:

$$P(Y = \text{positive for a and b}) \cong \frac{\hat{n}_{a+b^+}}{\hat{n}} = \frac{\hat{n}_{a^+}\hat{n}_{b^+}}{\hat{n}^2}. \quad (4.19)$$

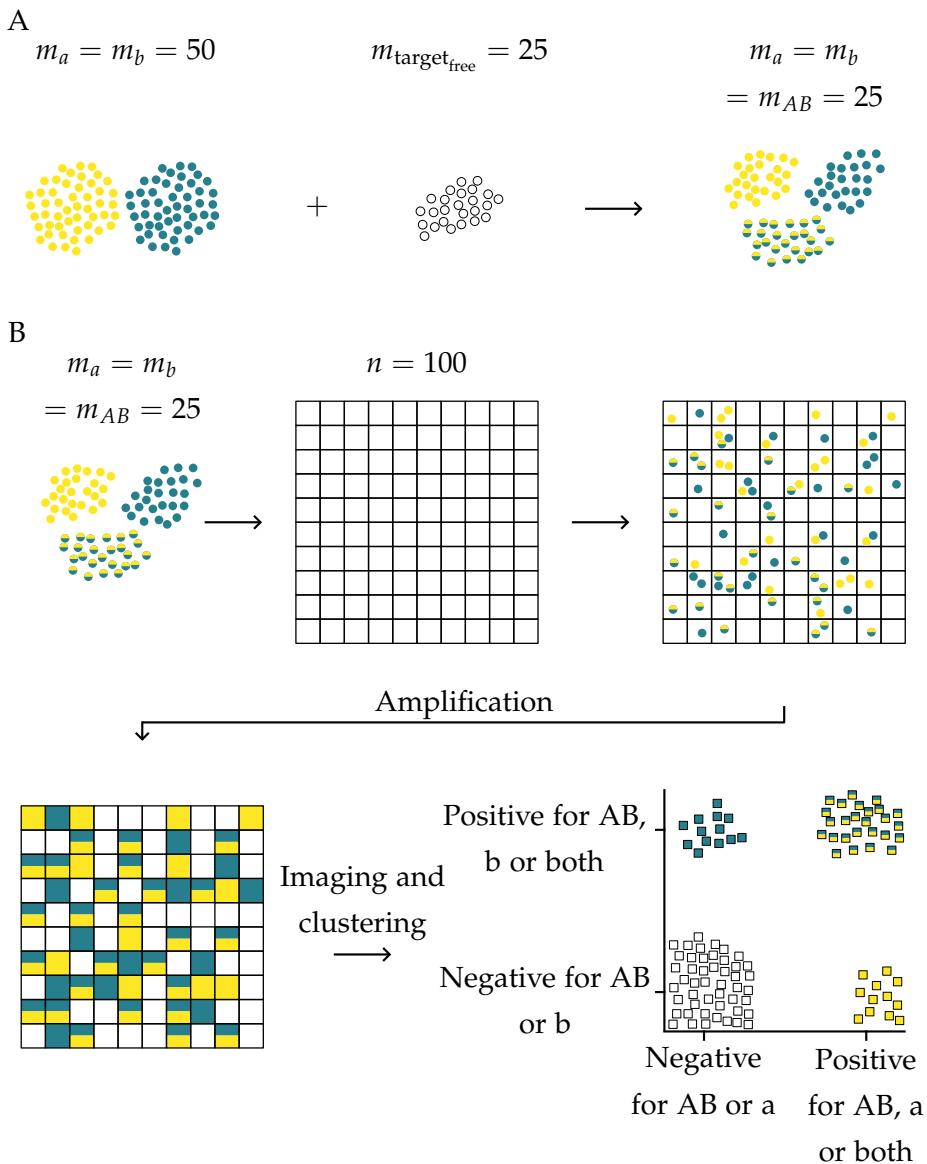


Figure 4.7: Exemplary PICO experiment.

A: In the binding reaction (Figure 4.3B and Figure 4.4), the DNA-labeled antibodies bind their targets. The free targets $m_{\text{target}_{\text{free}}}$ are bound by both antibodies a and b . Assuming non-residue binding ($f_a = f_b = 1$), the numbers of unbound antibodies are reduced by the number of targets (Equation 4.20 and Equation 4.21). Upon binding, the undetectable targets become detectable (protein-to-nucleic acid transformation). They are now complexes. In this example, there are $m_a = m_b = m_{AB} = 25$ molecules prior to partitioning.

B: All molecules distribute randomly and independently in $n = 100$ partitions. After amplification, clustering and counting, $\hat{n}_{D_{\text{obs}}} = 26$ partitions contain at least 1 of each unbound antibody ($X_a \geq 1 \wedge X_b \geq 1$) or at least 1 couplex ($X_{AB} \geq 1$) or a combination of all ($X_a \geq 1 \wedge X_b \geq 1 \wedge X_{AB} \geq 1$). 38 partitions contain at least 1 antibody a ($X_a \geq 1$) or at least 1 couplex ($X_{AB} \geq 1$), 39 partitions contain at least 1 antibody b ($X_b \geq 1$) or at least 1 couplex ($X_{AB} \geq 1$) and $\hat{n}_{AB-a-b^-} = 49$ contain no detectable molecule. The dDPCS model calculated 27 couplexes for this example, which is reasonably close to the expectation of 25 couplexes.

Note that, by increasing the number of molecules m_a and m_b , their probability for random colocalization increases, which leads to a higher number of double positive partitions \hat{n}_{a+b+} .

In contrast to an ABC PICO experiment, where the target of interest is absent, the sample in a PICO experiment contains an unknown number of targets $m_{\text{target free}}$ (Figure 4.7A). Let now m_{AB} be the number of targets that are bound by both DNA-labeled antibodies, i.e. the number of complexes (Figure 4.4 and Figure 4.7A). The goal is to determine m_{AB} . From now on, the notation will be that small letters indicate unbound antibodies, while capital letters indicate bound antibodies. $m_{\text{target free}} \geq m_a$ or $m_{\text{target free}} \geq m_b$ is prevented by choosing appropriate experimental conditions so that $m_a \gg m_{\text{target free}}$ and $m_b \gg m_{\text{target free}}$ ⁶. Of note, m_A cannot be distinguished from m_a and similarly for m_B (Figure 4.4). The number of unbound antibodies is reduced by the number of targets that have bound the corresponding antibody (Figure 4.7A):

$$m_a \leftarrow m_a - f_a m_{\text{target free}} \quad (4.20)$$

$$m_b \leftarrow m_b - f_b m_{\text{target free}}. \quad (4.21)$$

Changing the values of f_a and f_b allows considering imperfect binding of antibodies *a* and *b*.

Yet, this is a simplified approach as the biological processes are known to be more complex. These cases are not considered in this thesis.

Let f_a and f_b be two factors for antibodies *a* and *b*, which define the fraction of $m_{\text{target free}}$ that is bound by the corresponding antibody. Assuming $f_a = f_b = 1$ and the previously mentioned excess of antibodies over targets, i.e. $m_a \gg m_{\text{target free}}$ and $m_b \gg m_{\text{target free}}$, it is justifiable to consider that after binding (Equation 4.20 and Equation 4.21) $m_{\text{target free}} = 0$ (Figure 4.7A). It holds that after binding $m_a \gg 0$ and that $m_b \gg 0$. Consequently, this non-residue binding means that the number of complexes equals the number of targets. This also means that complexes can be compared to each other without further manipulation.

m_{AB} , m_a and m_b randomly and independently distribute in n partitions (Figure 4.7B). The random variable Y has the same four states as before (Equation 4.11, Equation 4.12, Equation 4.13 and Equation 4.14), where Equation 4.14 describes the state $P(Y = \text{positive for } a \text{ and } b)$, which is the random colocalization of both antibodies in the same partition. However, a complex renders a partition positive for both antibodies, too. But in this case, the antibodies are bound to the target and the notation with small letters indicates unbound antibodies. Thus, a comprehensive description for this cluster is now $P(Y = \text{double positive})$ as it can be reached by multiple different scenarios (Figure 4.8). In contrast to the ABC experiment (Figure 4.6), the cluster of double positive partitions can no longer be explained by the random colocalization of both antibodies in the same partition. A double positive partition can now also contain a complex or a random colocalization of both DNA-labeled antibodies or any combination of

⁶ Gross et al. (2024) refer to this as saturating conditions.

these Figure 4.7B. The resulting clusters can be described as follows (Figure 4.8A):

- \hat{n}_{AB-a-b^-} : observed number of partitions containing no detectable molecule;
- $\hat{n}_{AB-a^+b^-}$: observed number of partitions containing only unbound antibodies a;
- $\hat{n}_{AB-a^-b^+}$: observed number of partitions containing only unbound antibodies b;
- $\hat{n}_{AB-a^+b^+}$: observed number of partitions containing only both unbound antibodies a and b;
- $\hat{n}_{AB+a^-b^-}$: observed number of partitions containing unbound antibodies a and complexes;
- $\hat{n}_{AB+a^-b^+}$: observed number of partitions containing unbound antibodies b and complexes;
- $\hat{n}_{AB+a^+b^-}$: observed number of partitions containing both unbound antibodies a and b and complexes;
- $\hat{n}_{D_{obs}}$: observed number of partitions containing antibodies a and b, a complex, or a combination of these, i.e. observed number of double positive partitions (see Figure 4.8B), and
- \hat{n}_{AB^+} : observed number of partitions containing a complex, independent of other detectable molecules. This is to be determined to calculate m_{AB} using Equation 4.7.

The number of double positive partitions $\hat{n}_{D_{obs}}$ can be described as the sum of all different possible states of a double positive partition (Figure 4.8B) under the assumption that $\hat{n}_{D_{obs}}$ approximates $nP(Y = \text{double positive})$ well enough:

$$\hat{n}_{D_{obs}} = \hat{n}_{d_1} + \hat{n}_{d_2} + \hat{n}_{d_3} + \hat{n}_{d_4} + \hat{n}_{d_5} \quad (4.22)$$

The individual summands can be described as follows.

$$\hat{n}_{d_1} \cong nP(X_a \geq 1 \wedge X_b \geq 1 \wedge X_{AB} = 0) \quad (4.23)$$

$$\hat{n}_{d_2} \cong nP(X_a \geq 1 \wedge X_b \geq 1 \wedge X_{AB} \geq 1) \quad (4.24)$$

$$\hat{n}_{d_3} \cong nP(X_a = 0 \wedge X_b = 0 \wedge X_{AB} \geq 1) \quad (4.25)$$

$$\hat{n}_{d_4} \cong nP(X_a \geq 1 \wedge X_b = 0 \wedge X_{AB} \geq 1) \quad (4.26)$$

$$\hat{n}_{d_5} \cong nP(X_a = 0 \wedge X_b \geq 1 \wedge X_{AB} \geq 1) \quad (4.27)$$

Alternatively, $\hat{n}_{d_1} + \hat{n}_{d_2}$ can be described by

$$\hat{n}_{d_1} + \hat{n}_{d_2} \cong nP(X_a \geq 1 \wedge X_b \geq 1), \quad (4.28)$$

In the course of this thesis \hat{n}_{d_1} will be referred to as random double positive partitions, \hat{n}_{d_2} will be referred to as rc-overlap double positive partitions and \hat{n}_{d_3} will be referred to as complex double positive partitions.

irrespective of X_{AB} . Similarly, $\hat{n}_{d_2} + \hat{n}_{d_3} + \hat{n}_{d_4} + \hat{n}_{d_5}$ can be written as

$$\hat{n}_{d_2} + \hat{n}_{d_3} + \hat{n}_{d_4} + \hat{n}_{d_5} \cong \hat{n}P(X_{AB} \geq 1), \quad (4.29)$$

irrespective of X_a and X_b . Plugging Equation 4.24, Equation 4.28 and Equation 4.29 in Equation 4.22 yields

$$\begin{aligned} \hat{n}_{D_{obs}} &\cong nP(X_a \geq 1 \wedge X_b \geq 1) + \hat{n}P(X_{AB} \geq 1) - \hat{n}_{d_2} \\ &\cong nP(X_a \geq 1 \wedge X_b \geq 1) + nP(X_{AB} \geq 1) \\ &\quad - nP(X_a \geq 1 \wedge X_b \geq 1 \wedge X_{AB} \geq 1). \end{aligned} \quad (4.30)$$

In analogy to Equation 4.19, Equation 4.30⁷ can be rewritten using the number of positive partitions from Figure 4.8A.

$$\begin{aligned} nP(Y = \text{double positive}) &\cong \hat{n}_{D_{obs}} \\ &\cong \frac{(\hat{n}_{AB^+a^+b^-})(\hat{n}_{AB^+a^-b^+})}{\hat{n}} \\ &\quad + (\hat{n}_{AB^+}) \\ &\quad - \frac{(\hat{n}_{AB^+a^+b^-})(\hat{n}_{AB^+a^-b^+})(\hat{n}_{AB^+})}{\hat{n}^2} \end{aligned} \quad (4.32)$$

Equation 4.32 describes the cluster of double positive partitions using the available information from Figure 4.8A. It holds that $\hat{n} > \hat{n}_{AB^-a^+b^-}$, $\hat{n} > \hat{n}_{AB^-a^-b^+}$ and that $\hat{n} > \hat{n}_{AB^+}$. Furthermore, it applies that $\{\hat{n}_{AB^-a^+b^-}, \hat{n}_{AB^-a^-b^+}, \hat{n}_{AB^+}\} \in \mathbb{N}_0$ and that $\hat{n} \in \mathbb{N}$. Equation 4.32 can be used to calculate \hat{n}_{AB^+} (Section 4.2.3), from which the unknown m_{AB} is immediate (Equation 4.7). For the considered case of saturating conditions, i.e. $m_a \gg m_{\text{target free}}$ and $m_b \gg m_{\text{target free}}$, it holds that $m_{AB} = m_{\text{target free}}$ (Gross et al., 2024).

⁷ Equation 4.30 can also be expressed in the Poisson distributed measure with the union of two non-mutually exclusive sets:

$$\begin{aligned} nP(Y = \text{double positive}) &= n(P((X_a \geq 1 \cap X_b \geq 1) \cup X_{AB} \geq 1)) \\ &= n(P(X_a \geq 1 \cap X_b \geq 1) + P(X_{AB} \geq 1) \\ &\quad - P(X_a \geq 1 \cap X_b \geq 1 \cap X_{AB} \geq 1)) \\ &= n((1 - e^{-\lambda_a})(1 - e^{-\lambda_b}) + (1 - e^{-\lambda_{AB}}) \\ &\quad - (1 - e^{-\lambda_a})(1 - e^{-\lambda_b})(1 - e^{-\lambda_{AB}})). \end{aligned} \quad (4.31)$$

However, λ_a and λ_b cannot be obtained from the experiment but can be approximated by $\frac{\hat{n}_{AB^+a^+b^-}}{\hat{n}}$ and $\frac{\hat{n}_{AB^+a^-b^+}}{\hat{n}}$ (Equation 4.3), respectively. Likewise, λ_{AB} can be approximated by $\frac{\hat{n}_{AB^+}}{\hat{n}}$, but this cannot be obtained from the experiment either. This population is contained in the number of double positive partitions $\hat{n}_{D_{obs}}$ (Figure 4.8B). Rewriting Equation 4.31 with this information again results in Equation 4.32. X_{AB}

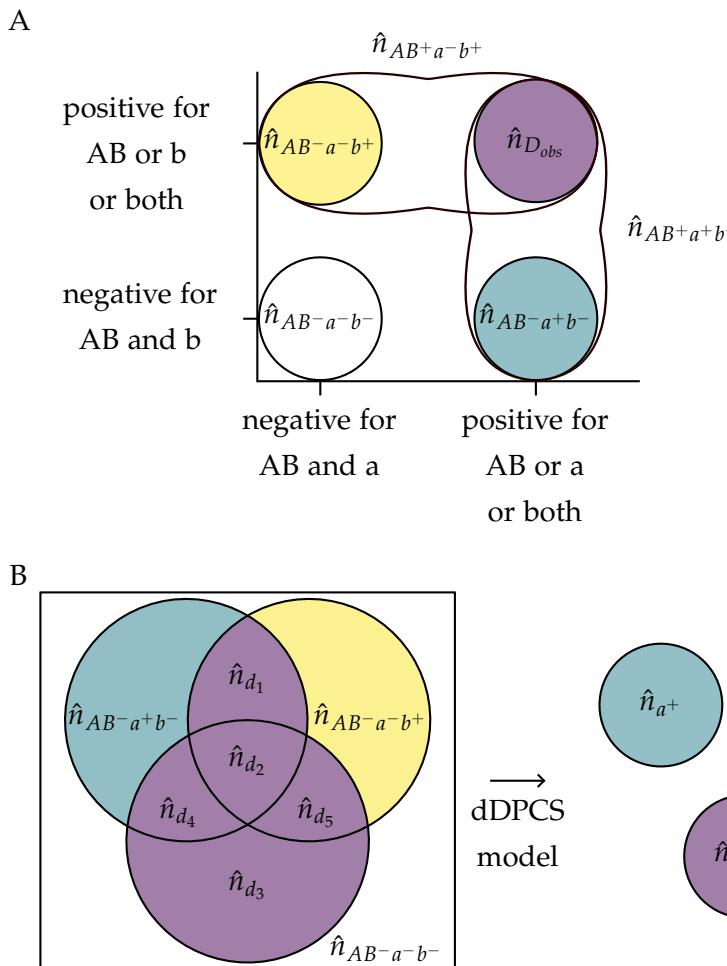


Figure 4.8: 2-dimensional raw dPCR data and composition of double positive cluster.

A: After amplification and imaging, the partitions are assigned to one of the clusters based on their fluorescence, which depends on the molecule(s) contained in the partition (Figure 4.7B). \hat{n}_{AB-a-b^-} is the number of empty partitions (no detectable molecule contained), $\hat{n}_{AB-a^+b^-}$ is the number of partitions containing only unbound antibodies a, \hat{n}_{AB-a-b^+} is the number of partitions containing only unbound antibodies b and $\hat{n}_{D_{obs}}$ is the number of double positive partitions, which can have multiple origins (Equation 4.22). \hat{n}_{AB+a+b^-} and \hat{n}_{AB+a-b^+} are total numbers of partitions containing antibodies a and b, respectively.

B: Compared to an ABC PICO experiment, in a PICO experiment the antibodies colocalize more often in the same partition than expected based on their numbers m_a and m_b (Poisson distribution) because of binding to the target (Figure 4.4). $\hat{n}_{D_{obs}}$ can be expressed as the sum of different double positive states (Equation 4.22). It holds that $\hat{n}_{AB^+} < \hat{n}_{D_{obs}}$. The segregation of the populations (mint, yellow and purple) is the core of the dDPCS model.

4.2.3 Implementation

Finally, the **dDPCS** model is implemented using the programming language **R** with vectorized functions. The goal is to identify a triple of $\{\hat{n}_{AB+a+b^-}, \hat{n}_{AB+a-b^+}, \hat{n}_{AB^+}\}$ for that $n_{D_{calc}} \approx \hat{n}_{D_{obs}}$. To do so, the vector n_{AB} is created containing the possible numbers of partitions positive for a complex. It starts from 0 and is increased by 1 until $\hat{n}_{D_{obs}}$ ([Listing 4.1 Line 13](#)). The rationale is that there cannot be more complexes than double positive partitions ([Figure 4.8B](#)). In contrast to the **DX** (Gross et al., [2024](#)), this implementation therefore avoids negative numbers of complexes ([Section A.1](#)). This is a constraint from the biology itself as the number of molecules in sample can either be 0 or ≥ 1 . na and nb are vectors containing the number of partitions positive for antibodies a or b , respectively. na and nb range from the corresponding observed number of positive partitions \hat{n}_{AB+a+b^-} or \hat{n}_{AB+a-b^+} to $\hat{n}_{AB+a+b^-} - \hat{n}_{D_{obs}}$ and $\hat{n}_{AB+a-b^+} - \hat{n}_{D_{obs}}$ in steps of 1 ([Listing 4.1 Line 15](#) and [Line 17](#)). The number of observed double positive partitions n_{D_obs} and the total number of partitions n are constant but need to be vectorized to match the length of the other vectors ([Listing 4.1 Line 19](#) and [Line 21](#)). For each index of the vectors, the number of random double positive partitions nR (\hat{n}_{AB-a+b^+} , [Listing 4.1 Line 25](#)) and the number of **rc-overlap** double positive partitions nRC (\hat{n}_{AB+a+b^+} , [Listing 4.1 Line 28](#)) are calculated. Next, the number of double positive partitions n_{D_calc} is calculated for each index of the vectors using [Equation 4.32](#).

The initial assumption for the first index ($i = 1$) of the vectors is that all double positive partitions $\hat{n}_{D_{obs}}$ can solely be explained by the random colocalization of both DNA-labeled antibodies in the same partition. This would mean that this experiment is an **ABC** ([Equation 4.19](#)). The second index calculates $n_{D_{calc}}$ for $\hat{n}_{AB^+} = i - 1$, and so on. As 1 complex requires 1 antibody a and 1 antibody b , their number is reduced by 1 for 1 complex ([Figure 4.4](#), [Equation 4.20](#) and [Equation 4.21](#)). For all results in the vector n_{D_calc} , the difference to the observation n_{D_obs} is calculated ([Listing 4.1 Line 33](#)) and the minimal difference is identified ([Listing 4.1 Line 36](#)). A high difference means that even the best triple found cannot explain all variability of the observation, potentially because of violations of at least one model assumption. The corresponding number of complex positive partitions (\hat{n}_{AB^+}) is considered to be the most likely explanation of the observation (maximum-likelihood approach). This number is converted into the total number of complexes per reaction using [Equation 4.7](#). The number of complex positive partitions, the number of complexes and the difference between observation and calculation are returned from the function. If the difference is larger than 1 % of $\hat{n}_{D_{obs}}$, the corresponding calculated number of complexes is not considered for further analyses, because of stronger statistical variability than expected.

Listing 4.1: Implementation of couplex detection and calculation in R.

```

1  #' @param n total number of partitions
2  #' @param nABA observed number of partitions positive for
   molecule a or a couplex
3  #' @param nABB observed number of partitions positive for
   molecule b or a couplex
4  #' @param nD_obs observed number of double positive
   partitions
5  #' @returns couplex_positives: number of couplex positive
   partitions; couplexes: total number of couplexes,
   difference_to_observation: difference from calculated
   number of double positive partitions (nD_calc) to
   observed number of double positive partitions (nD_obs)
6  calculate_couplexes <- function(
7      n,
8      nABA,
9      nABB,
10     nD_obs) {
11     # generate vectors for required variables
12     # generate sequence of partitions positive for couplex
       starting from 0 to nD_obs
13     nAB <- seq(0, nD_obs, by = 1)
14     # generate sequence of partitions positive for molecule a
15     na <- rep(nABA, length(nAB)) - nAB
16     # generate sequence of partitions positive for molecule b
17     nb <- rep(nABB, length(nAB)) - nAB
18     # generate sequence of double positive partitions
19     nD_obs <- rep(nD_obs, length(nAB))
20     # generate sequence of total number of partitions
21     n <- rep(n, length(nAB))

22     # calculate the number of partitions positive for
       molecule a and b
23     # random double positive partitions
24     nR <- na * nb / n
25     # calculate the number of partitions positive for
       molecule a and b and couplex
26     # rc-overlap
27     nRC <- na * nb * nAB / n^2
28     # calculate the number of double positive partitions
29     nD_calc <- nR + nAB - nRC
30
31     # calculate the difference from the observation nD_obs to
       nD_calc
32     diff <- round((nD_obs - nD_calc)^2)

33     # identify minimal distance to observation
34     i_minimal_diff <- which.min(diff)
35
36     return(
37         list(
38
39

```

```

40             # return the number of partitions positive for a
41             # complex
42             couplex_positives = nAB[i_minimal_diff],
43             # return the number of complexes
44             complexes = round(n[i_minimal_diff] * (log(n[i_
45                 minimal_diff]) - log(n[i_minimal_diff] - nAB[
46                     i_minimal_diff]))),
47             # return the difference to the observation
48             difference_to_observation = diff[i_minimal_diff]
49         )
50     )
51 }

```

In the following chapters, the **dDPCS** model⁸ will be validated using dPCR simulations and experimental data (Figure 3.1 and Chapter 6) and will be applied for the detection of a potentially relevant biomarker in AML regarding HMA therapy resistance (Chapter 7).

⁸ Depending on the dPCR system used, the output data might not directly compatible with the here shown implementations of the calculation of the number of complexes. For instance, when working with more than two antibodies, the raw data output file (*MultipleOccupancy* file) from the QIAcuity dPCR system (QIAGEN) does not immediately provide the counts per cluster $\hat{n}_{AB^+a^+b^-}$, $\hat{n}_{AB^+a^-b^+}$, $\hat{n}_{D_{obs}}$ and $\hat{n}_{AB^-a^-b^-}$, so that the raw data needs corresponding formatting (Section A.2) to be used for the here shown implementation of the **dDPCS** model.

RESULTS AND DISCUSSION

5

CREATION OF A DPCR SIMULATION

The second objective of this thesis comprises the creation and validation of a dPCR simulation with the ability to simulate PICO experiments. The simulated data will then be used in phase I of **dDPCS** model validation (Figure 3.1 and Section 6.1). This chapter focuses on the integral considerations for the creation of a dPCR simulation, its setup and finally its validation.

5.1 CONSIDERATIONS AND SETUP

Inspired by dPCR simulations written for different purposes (dPCR quality assessment or data evaluation) (Burdukiewicz et al., 2016; Lievens et al., 2016), a novel dPCR simulation was developed specifically for simulating PICO experiments. The complete code is available upon request from <https://github.com/LangeTo>. Core elements are shown in this chapter and in Appendix B. In general, the simulation makes the following assumptions and simplifications:

- The dPCR simulation cannot simulate multiplex PICO assays (Figure 4.2B). The minimum and maximum of antibodies is two.
- The antibodies bind all targets available without residue (Equation 4.20 and Equation 4.21). Thus, the number of complexes directly equals the number of targets.
- The total binding reaction is subjected to the dPCR, there is no prior dilution step (Figure 4.3) and no dead volume.
- The total number of partitions is constant at $n = 25\,000$.
- All DNA-labels are perfectly amplified and detected in each partition.
- λ is the same for both antibodies.

The dead volume in a dPCR is the residual volume of the total volume that was not partitioned.

5.1.1 Ideal conditions

The simulation mainly consists of the `dpcr_sim()` function, which has the basic arguments `number_of_targets`, `lambda_of_antibodya`, `lambda_of_antibodyb` and `total_partitions`, all of which can be either single values or vectors of the same length. This allows to simulate arbitrary many dPCRs. In the first step, `dpcr_sim()` simulates antigen-target binding, which is the formation of ternary complexes or complexes

in the binding reaction (Figure 4.3B and Figure 4.4). The bound antibodies can no longer distribute independently, and thus their `lambda` is corrected by `lambda_of_complex` (Listing B.1). Then, partitioning is simulated (Listing 5.1).

Listing 5.1: Simulation of partitioning process independently for antibodya, antibodyb and complexes. Each row is a single partition (25 000 in total), which can contain ≥ 0 antibodiesa and/or ≥ 0 antibodiesb and/or ≥ 0 complexes.

```

1 # distribution of antibodies and complexes according to
  Poisson distribution in total_partitions
2 partitions <- data.frame(
3   partition = 1:total_partitions,
4   antibodya = rpois(total_partitions, lambda_of_antibodya),
5   antibodyb = rpois(total_partitions, lambda_of_antibodyb),
6   complex = rpois(total_partitions, lambda_of_complex)
7 )

```

Separately and thus independently, the distribution of antibodies a and b and complexes in all partitions is simulated using `rpois(n, lambda)`. This generates three vectors containing non-negative integers at the length of `total_partitions`. The average of each vector equals the corresponding `lambda`. Then, the vectors are assembled to one data frame, called `partitions` (Listing 5.1), where each row represents a single partition. Thus, the columns `antibodyA`, `antibodyB` and `complex` contain the number of corresponding molecules in each partition. However, the information on the actual number of molecules per partition is not available in real experiments (Figure 4.5). The only information available is the status of each partition, i.e. positive or negative for a certain molecule. Amplification of DNA-labels is not considered in this simulation, which means that a partition containing a certain molecule is immediately positive for it. Depending on which molecules are present in a partition, it is assigned to the corresponding cluster (Figure 4.8A: negative, single positive or double positive) using the logic from Listing 5.2.

The dPCR simulation by Burdukiewicz et al. (2016) simulates a fluorescence value for each partition between 0 and 1, which is closer to the real experiments because of partitions with intermediate fluorescence, called “rain” (The dMIQE Group et al., 2020). This means that the signal is not strictly binary, and a threshold is used to decide if a partition is positive or negative for a certain molecule. Such a feature might be desirable for the next iteration of this dPCR simulation and may include clustering algorithms from Brink et al. (2018) or De Falco et al. (2023) for more elaborate thresholding. For phase I of the dDPCS model validation (Section 6.1), the strictly binary signal shall be sufficient. However, the here presented simulation comprises a functionality to simulate clustering biases (see below and Listing B.4 and Listing B.5). The variability of the fluorescence value and cluster-

ing biases both impair the decision to which cluster a certain partition belongs. Finally, the number of partitions per cluster is calculated ([Listing B.2](#)), which is the required input for the `dDPCS` model.

Listing 5.2: Simulation of clustering. Depending on which molecules are present in a partition, it is assigned to the corresponding cluster ([Figure 4.8A](#): negative, single positive or double positive). A partition is double positive when it contains antibody a and antibody b or a couplex. A partition is positive for antibody a when it contains antibody a or a couplex. Similarly, a partition is positive for antibody b when it contains antibody b or a couplex. A partition is negative when it does not contain any detectable molecules.

```

1 library(dplyr)
2
3 partitions <- partitions %>%
4   mutate(
5     double_positive = ifelse(
6       (antibodya > 0 & antibodyb > 0) |
7       couplex > 0,
8       TRUE,
9       FALSE
10    ),
11    antibody1_positive = ifelse(
12      antibodya > 0 | couplex > 0,
13      TRUE,
14      FALSE
15    ),
16    antibody2_positive = ifelse(
17      antibodyb > 0 | couplex > 0,
18      TRUE,
19      FALSE
20    ),
21    negative = ifelse(
22      (antibodya == 0 & antibodyb == 0 & couplex == 0),
23      TRUE,
24      FALSE
25    )
26  )

```

Furthermore and in contrast to real experiments, the dPCR simulation can count the actual number of each molecule per partition ([Figure 4.5A](#)). These ground truth values will be used in [Section 6.1.1](#) for the validation of the `dDPCS` model. The number of ground truth random double positive partitions $n_{AB-a+b+}$ and the number of ground truth `rc-overlap` double positive partitions $n_{AB+a+b+}$ can be determined easily ([Listing B.3](#)).

Finally, `dpcr_sim()` returns a data frame with the following information. The number of rows depends on the length of the initial vectors `number_of_targets`, `lambda_of_antibodya`, `lambda_of_antibodyb` and `total_partitions`. Each row is one dPCR simulation.

- Input information
 1. Total number of partitions (n),
 2. λ of antibody a (λ_a),
 3. λ of antibody b (λ_b),
 4. Total number of targets/couplexes ($m_{\text{target}_{\text{free}}}$).
- Ground truth values
 1. Number of complex double positive partitions (n_{AB^+}),
 2. Number of random double positive ($n_{AB^-a^+b^+}$),
 3. Number of rc-overlap double positive ($n_{AB^+a^+b^+}$).
- 2-dimensional raw dPCR data
 1. Number of observed double positive partitions ($\hat{n}_{D_{\text{obs}}}$),
 2. Number of observed partitions containing unbound antibodies a and couplexes ($\hat{n}_{AB^+a^+b^-}$),
 3. Number of observed partitions containing unbound antibodies b and couplexes ($\hat{n}_{AB^+a^-b^+}$),
 4. Number of observed negative partitions ($\hat{n}_{AB^-a^-b^-}$).
- dDPCS output
 1. Number of calculated complex double positive partitions (\hat{n}_{AB^+}),
 2. Number of calculated random double positive partitions ($\hat{n}_{AB^-a^+b^+}$),
 3. Number of calculated rc-overlap double positive partitions ($\hat{n}_{AB^+a^+b^+}$),
 4. Calculated number of couplexes,
 5. Difference between observed and calculated number of double positive partitions.

5.1.2 Biased conditions

Real experiments are disturbed by numerous issues and a dPCR simulation considering some of these may deliver valuable insights for the design of experiments. Two of these issues will be examined in detail because they are considered to impact results of PICO experiments significantly: clustering biases and partition-specific competition (PSC) (Aralar et al., 2020; Whale et al., 2016). Clustering biases impair the classification of a certain partition. A false-negative clustering bias means that the partition is classified negative, although the target is present, a false-positive clustering bias means that the partition is classified positive, although the target is absent. Clustering biases most often occur for partitions with intermediate fluorescence, i.e. “rain”

(Aralar et al., 2020; Lievens et al., 2016; The dMIQE Group et al., 2020). Additionally, a false-negative clustering bias can be a consequence of poor imaging or can be induced by a molecular dropout, which is an event, where the target is present but fails to amplify (Whale et al., 2013). False-positive clustering bias can be a consequence of poor imaging. A false-positive clustering bias will lead to an overprediction of the number of targets, while a false-negative clustering bias and PSC will lead to an underprediction¹.

The PSC describes the fact that each partition only contains a limited amount of resources (e.g. nucleotides, enzymes, cofactors), which need to be shared by all DNA molecules to be amplified (Whale et al., 2016). Furthermore, some DNA molecules might be amplified preferably because of their sequence (Benita et al., 2003; Veal et al., 2012). This results in imbalanced amplification efficiencies of the amplicons in the same partition and will distort the classification of that partition. For standard dPCRs, PSC depends on λ , whereas in multiplex PICO assays (Figure 4.2B) multiple DNA-labeled antibodies are forced into the same partition because the antibodies are bound to the same target. This can lead to PSC independent of λ resulting in an underprediction of the number of targets.

With some changes, `dpcr_sim()` can simulate clustering biases and PSC separately. In reality these issues might occur simultaneously, but this needs substantial input from real experiments for adequate estimation of the parameters.

To simulate PSC, the additional parameter `dropout_threshold` defines how many molecules can be successfully amplified in a single partition. Based on the total number of molecules per partition and the `dropout_threshold`, the number of `surplus_molecules` is calculated (Listing B.6 Line 10). This value is randomly split into three parts (Listing B.6 Line 17) using Listing B.7. Then, each part is subtracted correspondingly from the number antibody a, b or the number of complexes (Listing B.6 Line 28). This simulates random dropouts of DNA molecules in one partition without preference for a specific sequence (Benita et al., 2003; Veal et al., 2012). The preference for a specific sequence to drop out could be considered in a simulation, which considers the amplification step and has knowledge about the sequence including primer annealing sites and GC-content, which is beyond the scope of this thesis.

For the simulation of clustering biases, `dpcr_sim()` gets two additional parameters `bias` and `clustering_bias`. The parameter `bias` defines the type of clustering bias and can have the values `c("none", "false-negative", "false-positive")`. For a false-negative clustering bias, Listing 5.2 is adjusted by replacing `TRUE` in Line 8 with `runif(n()) >`

For the detection of a complex, the DNA-labels of both antibodies need to be successfully amplified in the same partition Figure 4.7.

¹ Imagine in Figure 4.5B, there are accidentally more positive or negative partitions than to be expected after the partitioning step in A. This impairs the precision of the dPCR by distorting the Poisson distribution.

`clustering_bias` Listing B.4. Similarly, false-positive clustering bias is simulated by replacing `FALSE` in Listing 5.2 Line 9 with `runif(n()) > clustering_bias` (Listing B.5). `runif(n())` generates a random number between 0 and 1, if the number is lower than the `clustering_bias`, this partition will falsely be classified. The parameter `clustering_bias` can have different values for each cluster and can be applied to each cluster separately or to an arbitrary combination of these (Figure 4.8).

5.2 VALIDATION

5.2.1 Ideal conditions

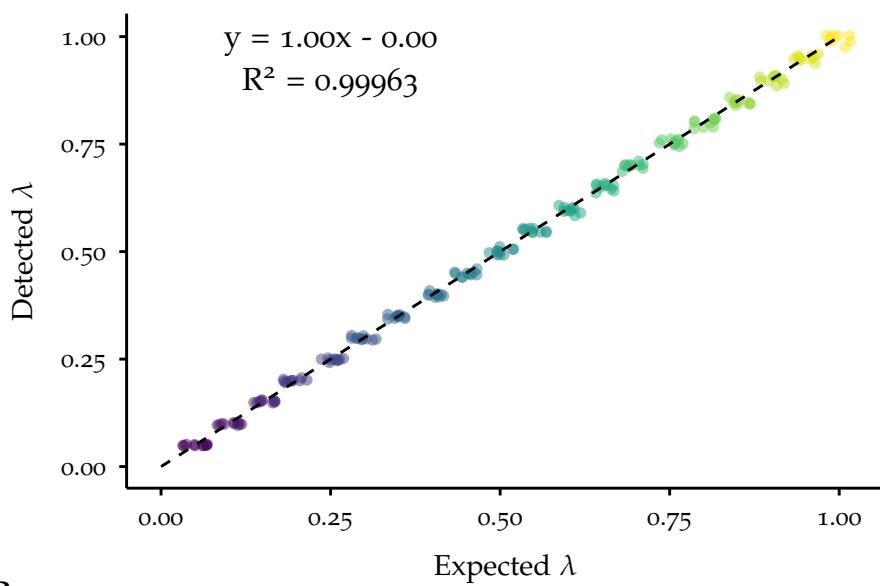
Before using the dPCR simulation for the validation of the dDPCS model, its validity for simulating a standard dPCR was verified. For this purpose, only antibody a was considered. Parameters `number_of_targets` and `lambda_of_antibodyb` of `dpcr_sim()` were set to 0. λ of antibody a was varied from 0.05 to 1.00 in steps of 0.05 and each value was simulated 10 000 times.

Expected and detected λ values exhibited a perfect correlation ($R^2 = 1.00$, Figure 5.1A) with an estimated slope of 1.00 (95% confidence interval (CI) from 1.00 to 1.00). This indicates that the detection of molecules using Listing 5.2 works as intended and the initial number of antibodies distributed in the partitions can be retrieved. The variability of the differences between expected and detected λ (normalized to expected λ) decreased with the λ . The coefficient of variation (CV) of the detected λ decreased with λ , too (Figure 5.1B). This is expected behavior as the calculation of λ depends on the number of negative partitions (Equation 4.8). In border cases, where almost all partitions are positive or negative, the confidence of the λ calculation is lower. On the other hand, Quan et al. (2018) and Majumdar et al. (2015) collectively report the highest confidence at $\lambda = 1.6$.

5.2.2 Biased conditions

CLUSTERING BIAS Next, the impact of false-negative and false-positive clustering biases at different levels was assessed. Lievens et al. (2016) claim that 2.5% partitions misclassified during clustering is a tolerable threshold and Aralar et al. (2020) found that 3% to 10% of the positive partitions are “rain”. Thus, the clustering bias was varied in $\{0, 1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$. λ was varied from 0.05 to 0.50 in steps of 0.05. Each combination of clustering bias and λ was simulated 550 times. Again only antibody a was considered and the parameters `number_of_targets` and `lambda_of_antibodyb` of `dpcr_sim()` were set to 0.

A



B

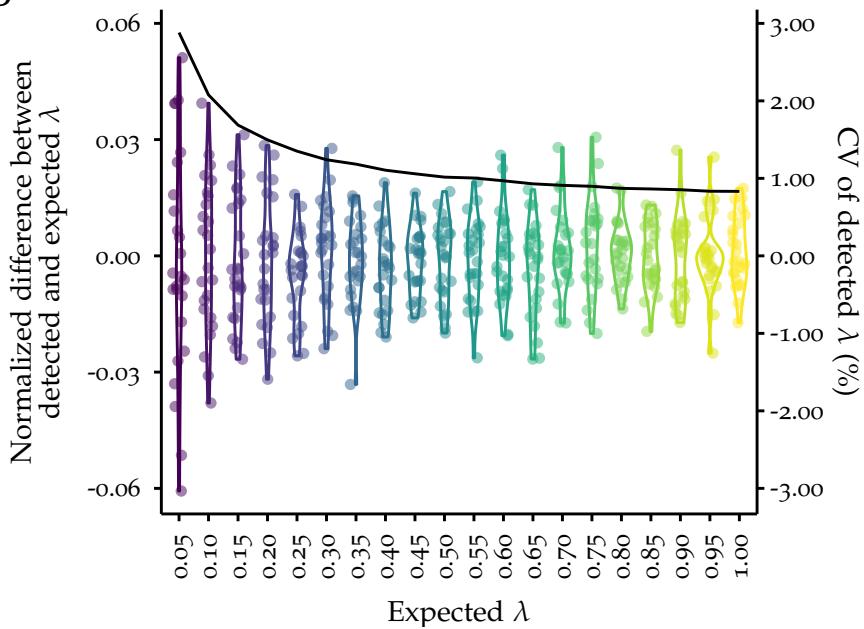


Figure 5.1: Comparison of expected and detected λ values for one DNA-labeled antibody after dPCR simulation. λ was varied from 0.05 to 1.00 in steps of 0.05 (color-coded). Each value of λ was simulated 10 000 times. Both plots show a representative subset of all simulations. The dots were jittered to avoid overplotting.

A: Scatterplot of detected λ against expected λ with results of the linear model `lm(detected_lambda ~ expected_lambda)`.

B: Violin plots of normalized difference between detected and expected λ . The CV of the detected λ (black line) is depicted on the right y-axis.

Statistics within this thesis

The statistical considerations of this thesis align with the recommendations from Nakagawa and Cuthill (2007).

The estimate of an effect size gives an impression of the magnitude of an effect and the corresponding CI gives an impression of the precision of this estimate. Finally, the p-value tells if the observed estimate is significant or not. Together, these measures enable a better judgement of differences between experimental groups, while p-values alone, as in classical statistics, allow for dichotomous decisions only. For instance, the slope of a linear model can be an estimate of the effect size (Nakagawa & Cuthill, 2007). Further details can be obtained from Section 9.4.2.

Detected λ was estimated to change by 0.13 (95 % CI from 0.13 to 0.13) (Figure 5.2A) and 0.14 (95 % CI from 0.14 to 0.14) (Figure 5.2B) with 1 SD of expected λ for false-negative and false-positive clustering bias, respectively (Figure 5.2C). The standardized estimates can be viewed as effect size and the CI as the associated uncertainty (see box [statistical considerations within this thesis](#) and [Section 9.4.2](#)). This means that the effect of expected λ on detected λ was similar for both clustering biases. However, false-negative and false-positive clustering biases had distinct effects on detected λ . Detected λ was reduced by 0.03 (95 % CI from 0.03 to 0.03) with a 1 SD increase of false-negative clustering bias. In false-positive clustering, detected λ was increased by 0.09 (95 % CI from 0.09 to 0.09) per 1 SD of the clustering bias. The false-negative clustering bias reduced the detected λ because initially positive partitions were now classified negative. This effect is enhanced by the interaction of expected λ and the clustering bias (Figure 5.2C, indicated by the * symbol) as with higher λ values the number of positive partitions increases and with that the absolute number of partitions that can be misclassified. On the other hand, false-positive clustering increased detected λ by a constant number depending only on the clustering bias. Listing B.5 was designed in such a way that new positive partitions are added which were initially negative. This is independent of expected λ . As shown by Majumdar et al. (2015), false-negative clustering has a relatively stronger impact at high λ values. False-positive clustering biases constant for the value of clustering bias. This means that the relative impact of false-positive clustering is higher at low λ values.

Generally, to reduce the impact of false-negative clustering, λ should be chosen reasonably low. Avoiding false-positive clustering is not straight-forward as it largely depends on the properties of the optical system, e.g. imaging conditions and interferences like dust particles.

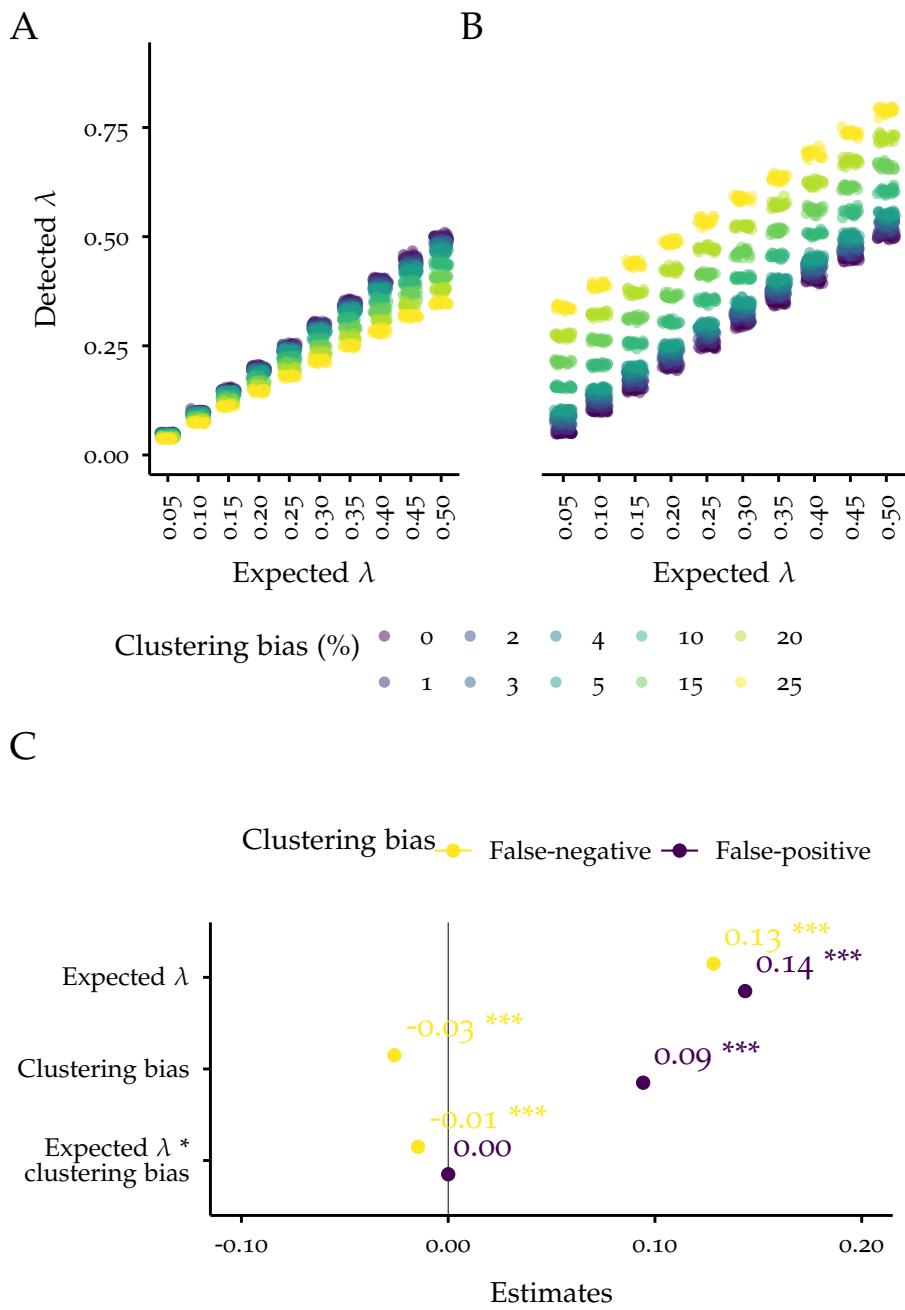


Figure 5.2: Comparison of expected and detected λ values for one DNA-labeled antibody after dPCR simulation with different clustering biases. λ was varied from 0.05 to 0.50 in steps of 0.05. The clustering bias was varied in $\{0, 1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$. Each combination of clustering bias and λ was simulated 550 times.

A: and B: Scatterplot of detected λ against expected λ for false-negative and false-positive clustering bias, respectively. The plot shows a representative subset of all simulations. The dots were jittered to avoid overplotting.

C: Forest plot of estimates with 95 % CI from the linear models of A (yellow) and B (purple) (`lm(detected_lambda ~ expected_lambda * clustering_bias)`). The predictors were standardized using Equation 9.2. The * symbol indicates the interaction of two input variables. The estimate is the change in the response variable for a 1 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for both models was 0.999.

PARTITION-SPECIFIC COMPETITION For the assessment of the impact of **PSC** on the dPCR, the `dropout_threshold` was varied in $\{2, 3, 4, 5, \infty\}$ molecules². λ was varied from 0.05 to 1.00 in steps of 0.05. Each combination of clustering bias and λ was simulated 550 times. Again only antibody a was considered and the parameters `number_of_targets` and `lambda_of_antibodyb` of `dpcr_sim()` were set to 0.

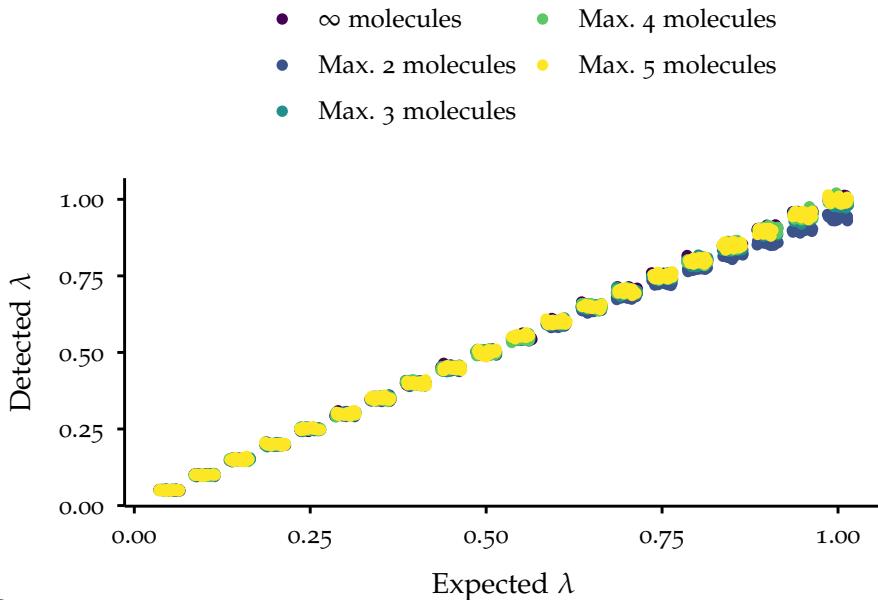
Detected λ changed by 1.01 (95 % **CI** from 1.01 to 1.01) with 1 SD of the expected λ (**Figure 5.3**). A maximal number of 2 or 3 successfully amplifiable molecules per partitions had an effect on detected λ (-0.05 (95 % **CI** from -0.05 to -0.05) and -0.01 (95 % **CI** from -0.01 to -0.01), respectively), yet significantly lower compared to the contribution of expected λ (**Figure 5.3B**). The interaction of expected λ and maximal number of amplifiable molecules changed the expected λ stronger additionally with the same effect. This is reasonable as with increasing λ more molecules localize in the same partition and thus, the propensity for **PSC** increases. The Poisson distribution dictates that even at $\lambda = 1$, the probability for a partition to contain more ≥ 2 molecules is $\sim 26\%$, while it is $\sim 8\%$ for ≥ 3 molecules per partition and $\sim 2\%$ for ≥ 4 molecules per partition. This explains why higher maximal numbers of molecules per partition did not have any impact for the simulated λ range. This is expected to change when increasing λ . Like the false-negative clustering bias, **PSC** can be prevented by reducing λ and by careful design of DNA-labels regarding, for instance, GC-content and primer binding sites (Benita et al., 2003; Veal et al., 2012).

5.3 SUMMARY

Under ideal conditions, the dPCR simulation can properly retrieve the number of molecules initially partitioned (**Figure 5.1A**), which underlines the readiness of this simulation for the validation of the **dDPCS** model. Moreover, the dPCR simulation demonstrated to be useful for the assessment of impairing factors like clustering biases and **PSC**. As a guideline for real experiments, λ should be thus chosen low enough to avoid strong impacts of false-negative clustering bias and **PSC** but high enough to avoid the regions with high uncertainty levels. A suitable range is $0.05 \leq \lambda \leq 0.45$.

² Simulating a `dropout_threshold` of 1 is not meaningful in the context of PICO as the signal of two DNA-labeled antibodies per partitions is required ([Chapter 4](#)).

A



B

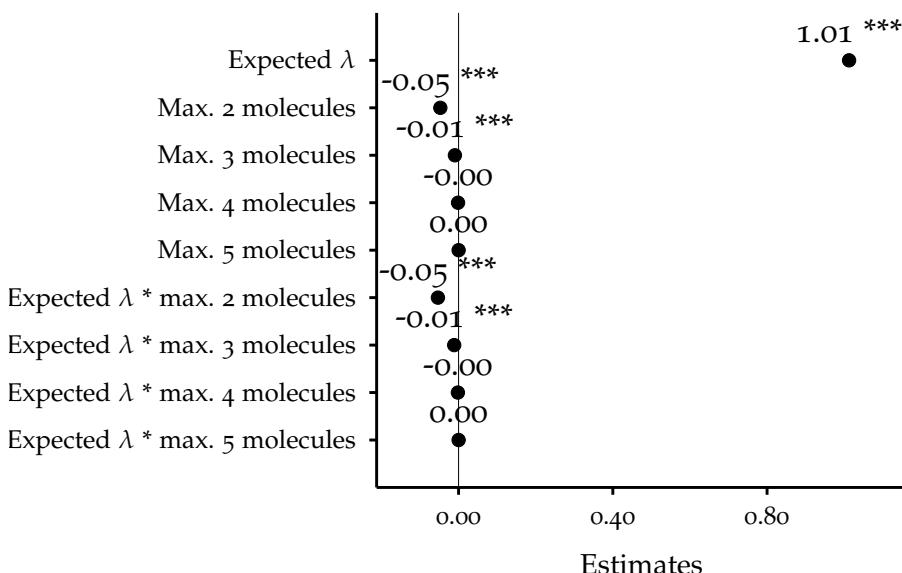


Figure 5.3: Comparison of expected and detected λ values for one antibody after dPCR simulation with PSC. λ was varied from 0.05 to 1.00 in steps of 0.05. The maximal number of amplifiable molecules (`dropout_threshold`) was varied in $\{0, 2, 3, 4, 5\}$. Each combination of maximal number of amplifiable molecules and λ was simulated 550 times.

A: Scatterplot of detected λ against expected λ . The plot shows a representative subset of all simulations. The dots were jittered to avoid overplotting.

B: Forest plot of estimates with 95 % CI from the linear model `lm(detected_lambda ~ expected_lambda * dropout_threshold)` of A. The predictors were standardized using Equation 9.2. The * symbol indicates the interaction of two input variables. The estimate is the change in the response variable for a 1 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for both models was 0.999.

6

VALIDATION OF THE DDPCS MODEL

Within this chapter, the **ddPCS** model for calculating of the number of complexes from 2-dimensional raw dPCR data (Figure 4.8) will be explored, and its applicability will be validated following Figure 3.1. Complexes are the basic detection unit of the PICO technology, their calculation is thus inevitable for successful PICO assays such as the application example in Chapter 7. The **ddPCS** model will be validated using simulated data (phase I, see Section 6.1) and experimental data (phase II, see Section 6.2). Moreover, the **ddPCS** is characterized regarding key performance parameters for the validation of analytical procedures (European Medicines Agency, 2023), which are summarized at the end of this chapter (Table 6.2) and the impact of the model is evaluated (Section 6.3). Finally, design rules guiding PICO assay design are established (Section 6.4).

6.1 USING SIMULATED DATA

In phase I of the model validation, the **ddPCS** model was applied to calculate the number of complexes from simulated data under ideal conditions (stage I, see Section 6.1.1) and bias conditions (stage II, see Section 6.1.2). This allows to control the amount of noise and enables a distinction of sources of error.

6.1.1 *Ideal conditions*

ABC PICO EXPERIMENTS First, the **ddPCS** model was applied to a simulated dataset from **ABC** PICO experiments (Figure 4.6). Herein, the cluster of double positive partitions must be explainable by the random colocalization of both DNA-labeled antibodies in the same partition (random double positives). This means that $\hat{n}_{D_{obs}} = \hat{n}_{AB^-a^+b^+}$ (Equation 4.14 and Equation 4.19). The dPCR simulation was set up in such a way that λ values of both antibodies were varied simultaneously from 0.05 to 1.00 in steps of 0.05. Each λ value was simulated 10 000 times. The parameter `number_of_targets` of the `dpcr_sim()` function was set to 0, thus the expected value for complexes is 0.

The distributions of calculated complexes for each simulated value of λ could be categorized into three groups with distinct shape (Figure 6.1). In the *narrow range* ($0.05 \leq \lambda \leq 0.25$), the majority of values fell in the first bin [0, 50] of the histogram and the mean value of the distributions were not significantly different from 0 (Figure 6.2). In the *tailing range* ($0.30 \leq \lambda \leq 0.55$), the complex distributions de-

veloped a tail (positive skewness), while the means of the first two distributions from $\lambda = \{0.30, 0.35\}$ were not significantly different from 0 (Figure 6.2). Eventually, in the *satellite range* ($\lambda \geq 0.6$), the tail transformed into a satellite population, which separated further from the origin with increasing λ (Figure 6.1). This implies that within this range, all mean values were different from 0 and constantly increased (Figure 6.2). A zoom of at least one representative distribution per range is shown in Figure C.1.

The **ddPCS** model allows to plot the maximum-likelihood function (difference between the observed and the calculated number of double positive partitions (Listing 4.1 Line 33) for all possible values of nAB (Listing 4.1 Line 13)). A representative image for the maximum-likelihood function for each simulated λ can be obtained from Figure C.2 and Figure C.3.

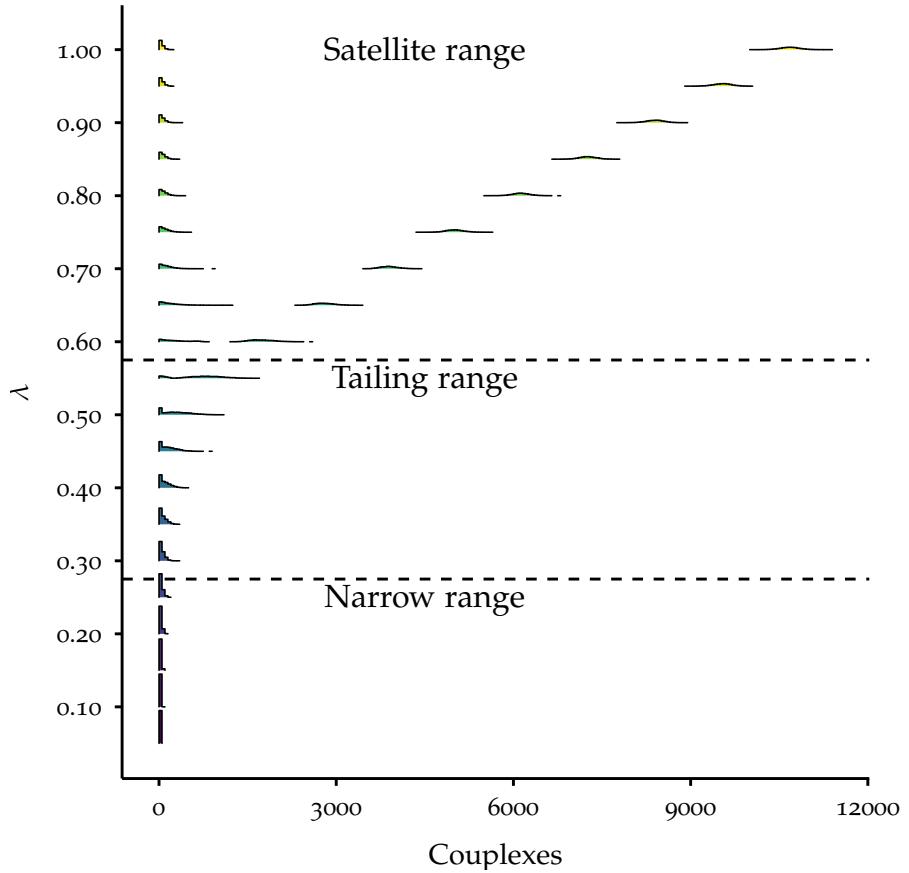


Figure 6.1: Ridgeline plot using histograms of complex distributions calculated from simulated ABC PICO experiments using the **ddPCS** model. λ was varied from 0.05 to 1.00 in steps of 0.05 (color-coded). Each value of λ was simulated 10 000 times. The bin width is 50 complexes. Based on their shape, the histograms are categorized into three groups: *narrow range* ($0.05 \leq \lambda \leq 0.25$), *tailing range* ($0.30 \leq \lambda \leq 0.55$), *satellite range* ($\lambda \geq 0.6$).

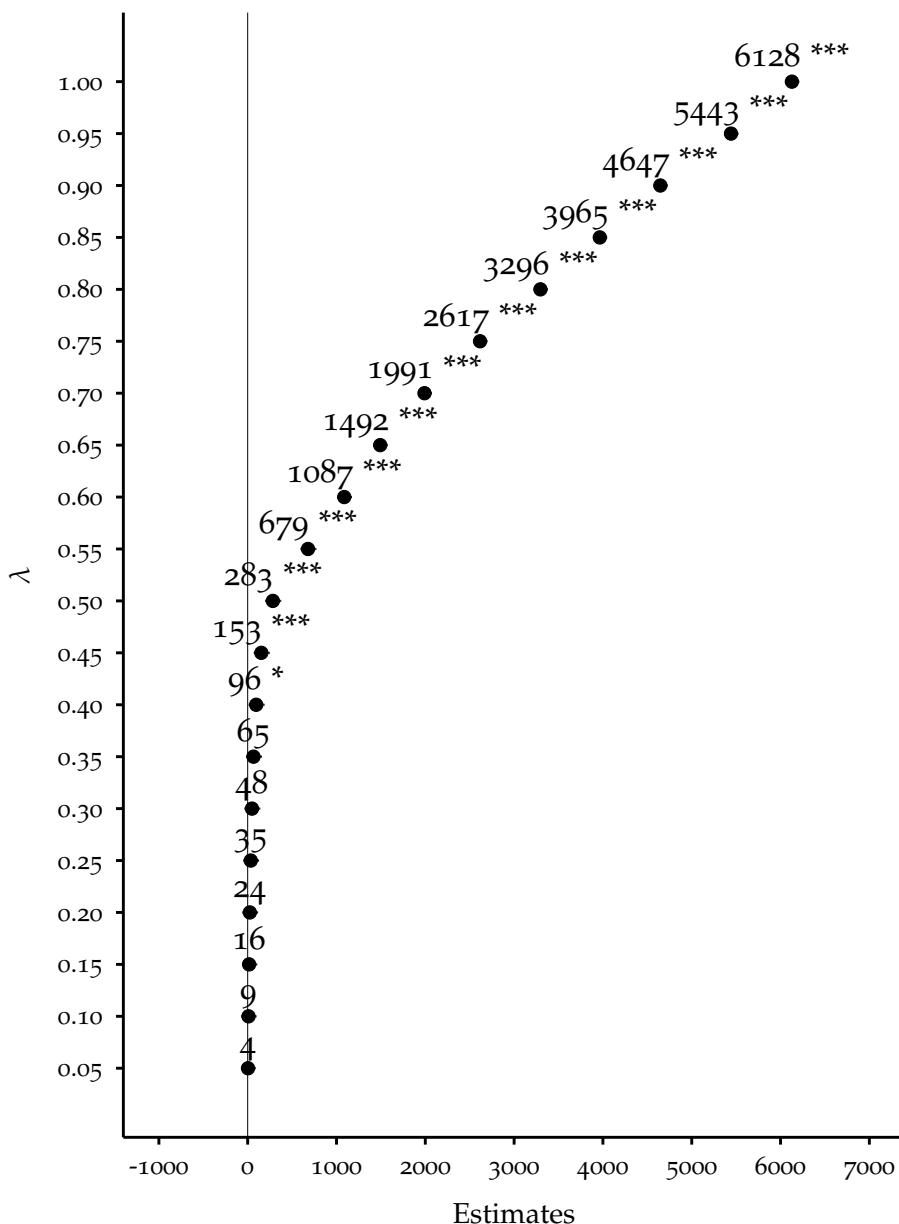


Figure 6.2: Forest plot of estimates with 95 % CI from the comparison of complex distributions with the expected mean of 0 couplexes. The linear model `lm(calculated_complexes ~ as.factor(lambda))` was used to compare the determined complex distributions using the `dDPCS` model with the expected reference. The adjusted R^2 is 0.412. The predictor was not standardized. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the CI is the associated uncertainty.

For $0.05 \leq \lambda \leq 0.55$, the maximum-likelihood functions showed a monotonic behavior. Beyond $\lambda = 0.6$, the function loses its monotonicity. Because of this change, the identification of the minimal difference

Monotonicity describes the behavior of a function. It means that this function is only increasing or decreasing.

becomes ambiguous. The maximum-likelihood function has more than one minimum and thus the **dDPCS** model finds more than one possible explanation for the observation, i.e. multiple different numbers of couplex positive partitions \hat{n}_{AB^+} are calculated for the same number of $\hat{n}_{D_{obs}}$. For instance, for $\lambda = 0.15$, the function was minimal for 0 or 1 couplex positive partitions (Figure 6.3A). The occurrence of two values can be explained by rounding imprecision (Listing 4.1 Line 33). However, for $\lambda = 0.65$, maximum-likelihood function was minimal at two strikingly different positions $\hat{n}_{AB^+} \approx 35$ and $\hat{n}_{AB^+} \approx 2462$ (Figure 6.3B). Again because of rounding imprecision, the model detected multiple values around the two mentioned values of \hat{n}_{AB^+} . With increasing λ , the tendency for ambiguity of the minimum detection increases, because of the change in the monotonic behavior. This results in the emergence of satellite populations (Figure 6.1).

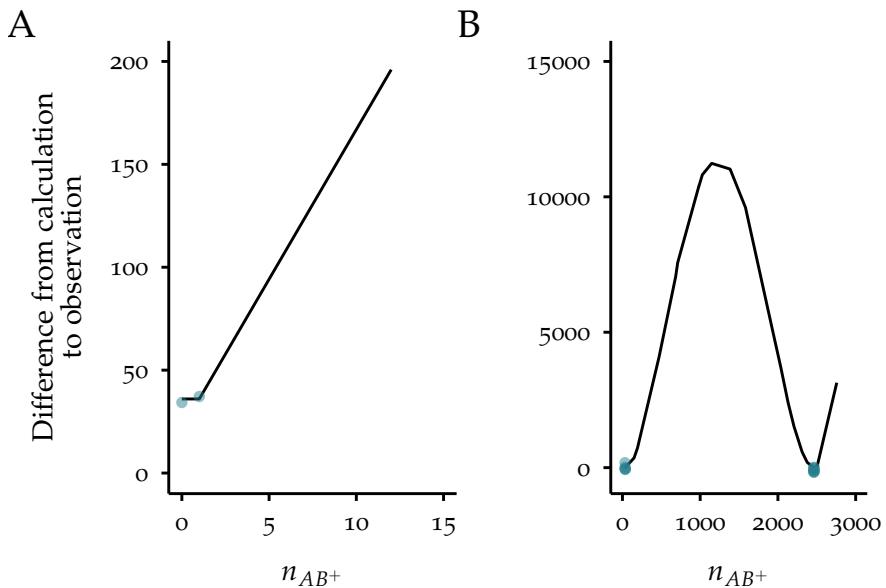


Figure 6.3: Ambiguity of the maximum-likelihood function (difference between the observed and the calculated number of double positive partitions (Listing 4.1 Line 33) for all possible values of n_{AB^+} (Listing 4.1 Line 13)). The colored dots indicate the determined minima.
A: Maximum-likelihood function for $\lambda = 0.15$.
B: Maximum-likelihood function for $\lambda = 0.65$.

Through the use of the dPCR simulation, the true number of random double positive partitions $n_{AB-a+b+}$ (ground truth) becomes available (Listing B.3). The number of random double positive partitions $\hat{n}_{AB-a+b+}$ calculated by the **dDPCS** model can be compared herewith (Figure 6.4). After $\lambda = 0.55$, which defines the onset of the satellite range (Figure 6.1) and which marks the change from monotonic to non-monotonic behavior (Figure C.2 and Figure C.3), the number of calculated random double positive partitions branched (Figure 6.4A).

For the same ground truth, the **dDPCS** model calculated two distinct function values. In mathematics, this behavior is described as bifurcation, where a system qualitatively changes its behavior upon the change of a parameter μ . μ_c is defined as the bifurcation point, from which on the system shows distinct behavior (Zhou, 2013). $\lambda = 0.55$ can thus be described as the bifurcation point $\mu_c = \lambda_{0.55}$ of this system. After $\lambda_{0.55}$, the fraction of wrong values (values in the lower branch of Figure 6.4A), was around 0.5 (Figure C.5). Before the bifurcation point, the calculated values were in line with the ground truth and the fraction of wrong values was 0.

The impact of the bifurcation could also be quantified by calculating an iterative linear model for Figure 6.4A. Iterative, in this case means that the range of the predictor λ is extended step-by-step, i.e. in Figure 6.4B, an x-value of $x = 0.25$ means that $0.05 \leq \lambda \leq 0.25$ and $x = 1.00$ means that all values from Figure 6.4A were included in the calculation of the corresponding estimate. The estimate (i.e. slope of the linear model) for $\lambda \leq 0.05$ was the lowest (0.08 with 95 % CI from 0.07 to 0.08 per unit of the ground truth) because it only contained the calculated number of double positive partitions from $\lambda = 0.05$. This means these values were replicates of the same dPCR simulation and correlation is thus not meaningful. Upon including the second simulated value of λ , the estimate jumped to values > 0.90 per unit of the ground truth and stayed until $\lambda = 0.55$ (Figure 6.4B) with a high coefficient of determination of $R^2 \geq 0.98$ (Figure C.5). Upon including larger λ values, the estimate decreased monotonically and reached a minimum at $\lambda = 1.00$ with 0.54 per unit of the ground truth (95 % CI from 0.54 to 0.55). $\lambda = 0.30$ had the highest estimate 0.97 (95 % CI from 0.97 to 0.97). However, the estimate was always different from 1, which indicates that the **dDPCS** model has a minimal offset of 3 %. When the estimate dropped below 0.90 after $\lambda = 0.55$, R^2 decreased, too (Figure C.5). This implies that the variability of the calculated number of random double positive partitions cannot be explained completely by the variability of the ground truth. While the fraction of wrong values remained constant after the bifurcation point, R^2 continued to decrease, which can be explained by increasing differences between the branches with increasing ground truth values (Figure 6.4A).

Taken together, the **ABC** PICO simulations for the **dDPCS** model validation revealed high precision for $\lambda \leq 0.30$. In general below $\lambda_{0.55}$, the differences between ground truths and calculated values is tolerable. Thus, the specificity of the model is given until $\lambda_{0.55}$ with a maximal offset of 3 %.

PICO EXPERIMENTS Next, the **dDPCS** model was applied to a simulated dataset of a PICO experiments. In contrast to the **ABC** simulation, the cluster of double positive partitions can no longer be

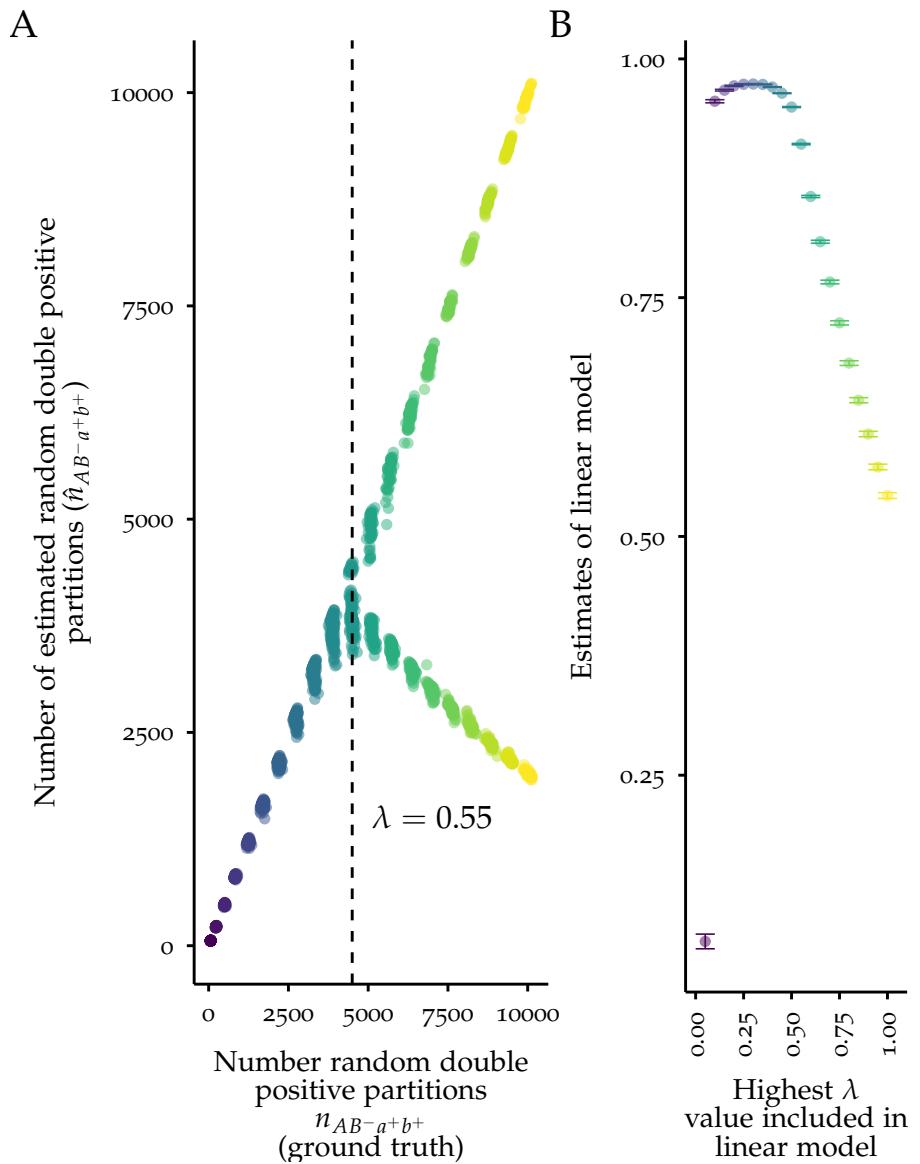


Figure 6.4: Comparison of calculated and ground truth random double positive partitions from simulated ABC PICO experiments. λ was varied from 0.05 to 1.00 in steps of 0.05 (color-coded). Each value of λ was simulated 10 000 times.

A: Scatterplot of calculated against ground truth random double positive partitions. The vertical dashed line depicts the bifurcation point $\lambda_{0.55}$. The plot shows a representative subset of all simulations. The dots were *not* jittered and so depict actual variability.

B: Scatterplot of estimates of the linear model from A (`lm(calculated_random_doubles ~ ground_truth_random_doubles)`) with 95 % CI as error bars. The x-axis denotes the highest λ included in the range for the linear model, e.g. $x = 0.25$ means that $0.05 \leq \lambda \leq 0.25$. The corresponding adjusted R^2 values can be obtained from Figure C.5. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the CI is the associated uncertainty.

explained by the random double positives because of the presence of targets, which coerce both antibodies to colocalization in the same partition ([Figure 4.7](#) and [Figure 4.8](#)). Hence, $\hat{n}_{D_{obs}} > \hat{n}_{AB-a+b+}$. For the simulation of a PICO experiment, the argument `number_of_targets` of the `dpcr_sim()` function was varied from 0 to 1000 in steps of 25, while `lambda_of_antibodya` and `lambda_of_antibodyb` were varied from 0.05 to 0.55 in steps of 0.05 and corrected by the `lambda_of_complex` ([Listing B.1](#)). Each combination of `number_of_targets`, `lambda_of_antibodya` and `lambda_of_antibodyb` was simulated 50 times.

[Figure 6.5A](#) shows exemplarily the histograms for 100 (left panel) and 1000 (right panel) simulated targets for all simulated λ . Further histograms can be obtained from [Figure C.4](#). The vertical dashed lines indicate the expected mean value of the distributions. In general, for $\lambda \leq 0.15$, the complex distributions were in the range of the expected mean for all numbers of simulated targets. The differences to the expected mean increased with λ but also with the expected number of targets ([Figure 6.5A](#) and [Figure C.4](#)). Furthermore, the SD of the distributions increased with λ ([Figure 6.5B](#)). For the quantification of these differences, a simplified linear relationship between λ and the expected number of targets and the moments mean and SD of the calculated complex distributions was assumed. The mean changed by 0.71 (95 % CI from 0.68 to 0.75) with 1 SD of the expected number of targets and by 0.58 (95 % CI from 0.55 to 0.62) with 1 SD of λ . The number of targets had a bigger impact compared to λ . Their combination led to an overprediction of the number of complexes ($0.71 + 0.58 = 1.30$). When restricting the λ range to values ≤ 0.30 , the mean changed by 0.99 (95 % CI from 0.98 to 1.00) with 1 SD of the expected number of targets ([Figure C.6](#)). Also, the impact of λ was reduced to 0.12 (95 % CI from 0.11 to 0.13) per 1 SD of λ . So that the overprediction was reduced by 15 % ($0.99 + 0.12 = 1.11$). Additionally, the SD of the complex distribution changed by 0.87 (95 % CI from 0.82 to 0.91) with 1 SD of λ and by -0.14 (95 % CI from -0.18 to -0.09) with 1 SD of the expected number of targets ([Figure C.6](#)). Thus, λ has a strong impact on the SD of the complex distributions, which flatten for higher values of λ . A possible explanation for the increasing deviation from the expected value with λ and the expected number of targets lies in the bifurcation of the complex distribution at higher values of λ ([Figure 6.1](#) and [Figure 6.4](#)). In the absence of targets, the mean value of the complex distribution already increases with λ ([Figure 6.2](#)). Restricting the λ range increased R^2 for the calculated linear models ([Figure 6.5B](#) and [Figure C.6](#)).

Similar to the simulation of ABC PICO experiments, the ground truth of couple double positive n_{AB+} , random double positive $n_{AB-a+b+}$ and rc-overlap double positive $n_{AB+a+b+}$ partitions can be obtained ([Listing B.3](#)) and compared to the calculated numbers from the dDPCS model ([Listing 4.1](#)). [Figure 6.6](#) shows representative scatterplots for

A moment is a quantitative characteristic of a distribution. The moments are mean, variance, skewness and kurtosis.

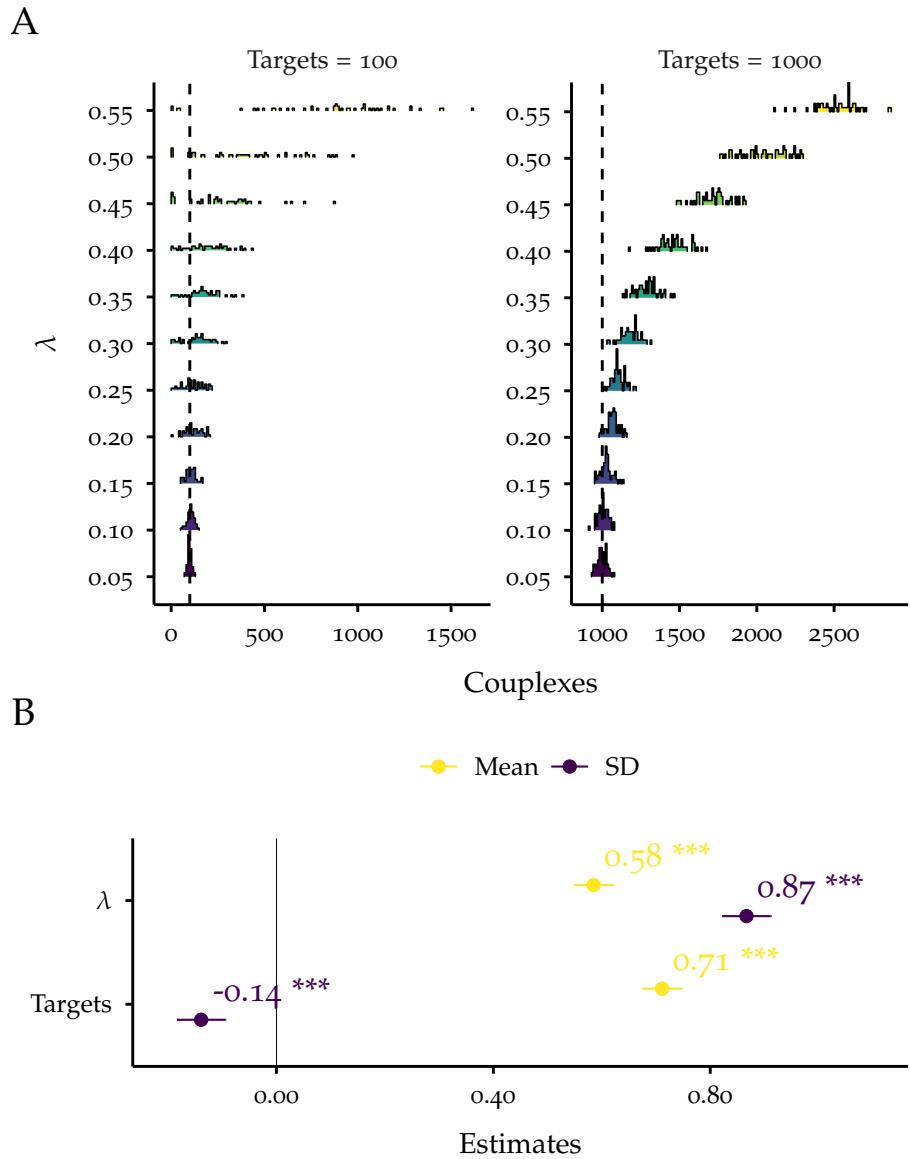


Figure 6.5: Evaluation of **ddPCS** model performance using simulated data with targets. λ was varied from 0.05 to 0.55 in steps of 0.05 (color-coded). The number of targets was varied from 0 to 1000 in steps of 25. Each combination of λ and number of targets was simulated 50 times.

A: Two representative ridgeline plots using histograms of couplex distributions for two expected numbers of targets (100 (left panel) and 1000 (right panel), indicated by the vertical dashed lines in each panel). The bin width is 10 couplexes.

B: Forest plot of estimates with 95 % CI from the linear models `lm(calculated_couplexes ~ expected_lambda + targets)` (yellow) and `lm(calculated_SD_couplexes ~ expected_lambda + targets)` (purple). The predictors were standardized using [Equation 9.2](#). The adjusted R^2 for the first model (yellow) is 0.848 and for the second model (purple) 0.770. The estimate is the change in the response variable for a 1 SD change in the predictor variable and the CI is the associated uncertainty.

$\lambda = \{0.05, 0.15, 0.30, 0.55\}$ of calculated number of double positive partitions against ground truth number of double positive partitions. Of note, for each λ the number of targets was varied from 0 to 1000 in steps of 25. Thus, independent of λ the number of couplex double positive partitions n_{AB^+} varies from 0 to 1000, which can also be seen in [Figure 6.6](#) (cyan dots). On the other hand, the number of random double positive partitions $n_{AB^-a+b^+}$ increased with λ as a consequence of a larger overlap between the Poisson distributions of both DNA-labeled antibodies (mint dots). Consequently, the number of [rc-overlap](#) positive partitions $n_{AB^+a+b^+}$ increased with λ and the number of targets (yellow dots), since this is the product of their Poisson distributions ([Figure 4.8B](#) and [Equation 4.31](#))¹. Beyond $\lambda = 0.20$, there were more random double positive partitions than couplex double positive partitions. Because of this change in the relation between couplex double positive and random double positive partitions, λ can also be regarded as the background from which the complexes need to be differentiated. With increasing λ , the background increases and the precision of the couplex calculation is reduced ([Figure 6.5](#) and [Figure 6.6](#)).

For the λ range of interest (≤ 0.55) and the three populations (\hat{n}_{AB^+} , $\hat{n}_{AB^-a+b^+}$ and $\hat{n}_{AB^+a+b^+}$), the relationship between calculated and ground truth number of partitions can be quantified using a linear model. An estimate of 1 means that the number of detected partitions is the same as the corresponding ground truth, while an estimate > 1 , indicates an overprediction of the positive partitions by the [dDPCS](#) model, and conversely, an estimate < 1 means underprediction. In general, the estimates for random double positive partitions and couplex double positive partitions for $\lambda \leq 0.2$ were within $1 \pm 10\%$, which can be considered as a tolerable variability. At $\lambda = 0.15$, the difference between the estimates and the expectation was the smallest ([Figure 6.7A](#)). However, the estimates of the [rc-overlap](#) were significantly different from 1 in the range $\lambda \leq 0.2$. The smallest difference to 1 for the [rc-overlap](#) was around $\lambda = 0.35$ to 0.40. Interestingly, the estimates for the [rc-overlap](#) did not follow the same pattern as the estimates of the random and couplex positive partitions ([Figure 6.7A](#)). Across the entire range of λ , R^2 decreased for random double positive and couplex double positive partitions, while in the range of $\lambda \leq 0.3$, it was above 0.8 and 0.95, respectively. R^2 for [rc-overlap](#) double positive partitions increased with λ until 0.35 and then started decreasing. At $\lambda = 0.25$, all three values of R^2 are the closest to 1.

¹ Assuming 25 000 partitions and $\lambda = 0.15$ for both DNA-labeled antibodies, $n_{AB^-a+b^+} \approx 563$ is expected without targets. Adding 1000 targets, $n_{AB^-a+b^+}$ is expected to change to 303 partitions because of [Equation 4.20](#) and [Equation 4.21](#). Then, $n_{AB^+a+b^+}$ is expected to have a value of 12.

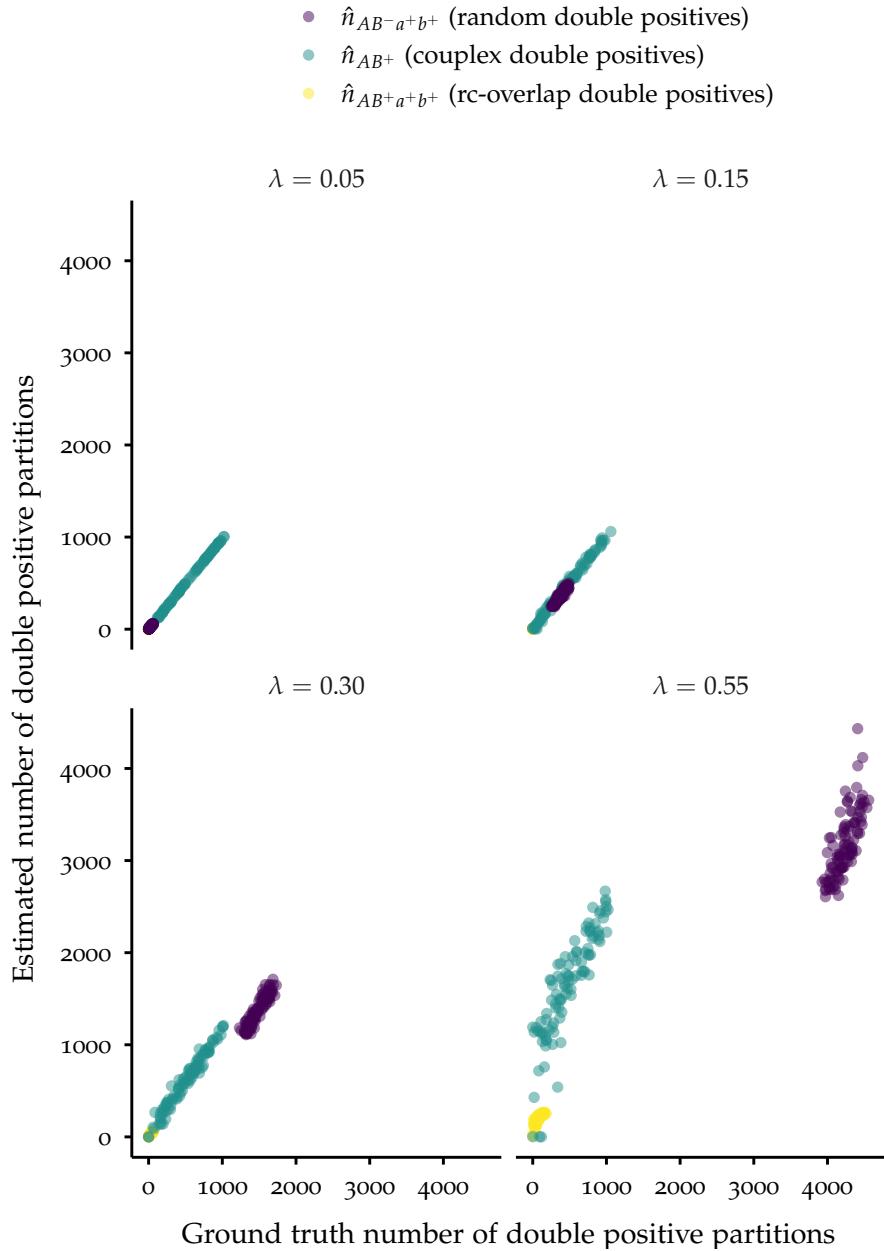
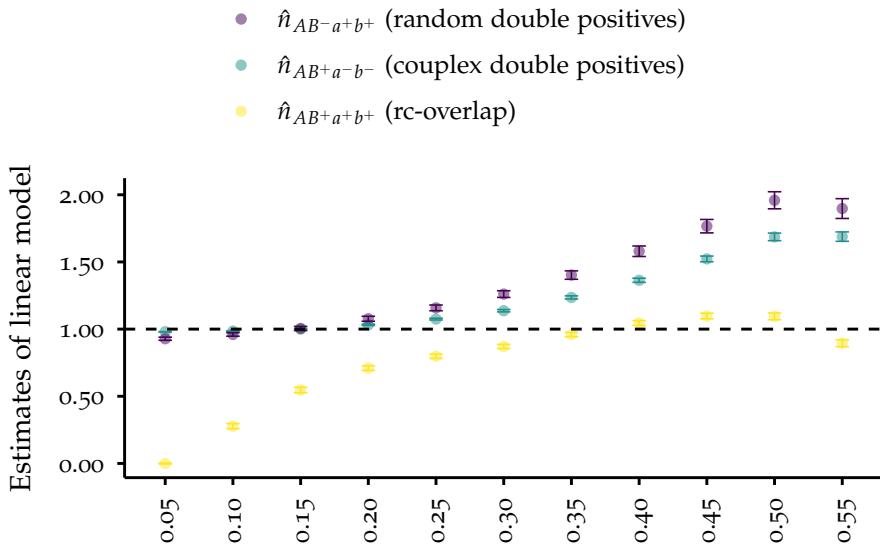


Figure 6.6: Representative scatterplots of calculated number of double positive partitions against ground truth double positive partitions for couplex (\hat{n}_{AB^+}), random ($\hat{n}_{AB-a+b+}$) and rc-overlap (\hat{n}_{AB+a+b^+}) double positive partitions. In the presence of targets, the cluster of double positive partitions $\hat{n}_{D_{obs}}$ can have multiple origins (Figure 4.8B). For $\lambda = \{0.05, 0.15, 0.30, 0.55\}$, the calculated values are plotted against the ground truth. Each plot shows a representative subset of the data. The dots were *not* jittered and so depict actual variability.

A



B

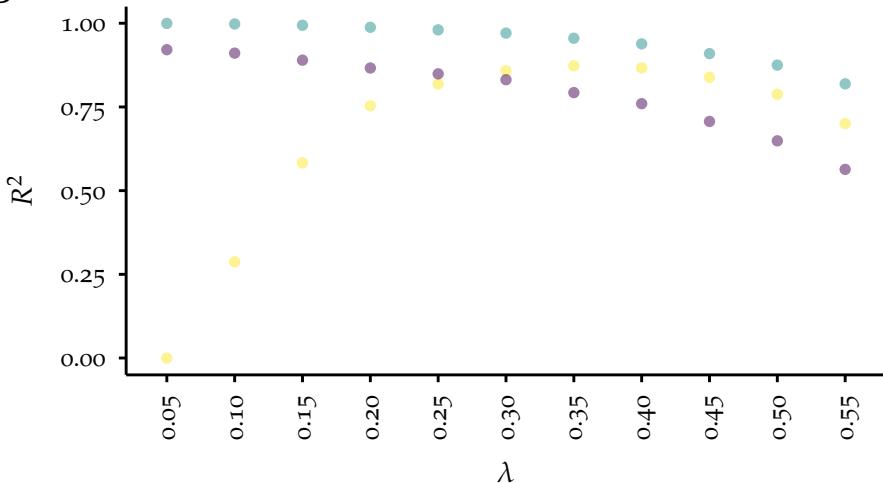


Figure 6.7: Comparison of calculated values and ground truth values for complex, random and rc-overlap double positive partitions using a linear model (`lm(calculated_double_positives ~ ground_truth_double_positives)`). The model was calculated for all simulated values of λ as well as complex double positive, random double positive and rc-overlap double positive partitions separately. The predictors were not standardized.

A: Scatterplot of the estimates from the linear models with 95 % CI as error bars comparing calculated and ground truth double positive partitions. An estimate of 1 means that the `dDPCS` model calculated the expected number of partitions for the specified type. An estimate smaller than 1 means that the `dDPCS` model underpredicts the number of double positive partitions. Conversely, an estimate larger than 1 means that the `dDPCS` model overpredicts the number of double positive partitions.

B: Scatterplot of the coefficients of determinations R^2 from all calculated linear models from A.

One sample in a dPCR is costs around 3.5 times as much as a sample in a qPCR (Basu, 2017).

Taken together, this underlines the need for a λ threshold, which needs to be lower than 0.55 (Figure 6.4) and ideally lower than 0.30 (Figure 6.7) to avoid imprecision of the ddPCS model due to increased variability and bifurcation. However, the variability introduced by manual pipetting can be effectively captured by the high precision of dPCR, which might result in exceeding a too stringent threshold. While this generally enhances the precision of a PICO assay, it could also significantly increase costs due to the exclusion of a potentially large number of experiments.

Furthermore, limit of detection (LOD) and limit of quantification (LOQ) were calculated according to the recommendations from the European Medicines Agency (2023) (Equation 9.4 and Equation 9.5) with the estimates from Figure 6.7 and the complexes calculated from the ABC PICO simulations. The LOD thus varies between 62 and 68 complexes and the LOQ between 187 and 206.

6.1.2 Biased conditions

In phase I stage II of the ddPCS model validation using simulated data, the robustness of the model towards dPCR issues like clustering biases and PSC (Section 5.1.2) (Aralar et al., 2020; Whale et al., 2016) will be evaluated.

CLUSTERING BIAS The same assumptions for clustering biases in the dPCR simulations also apply for clustering biases in PICO simulations (Section 5.1.2). However, in the dPCR simulation, the detection of positive partitions from only one antibody was biased, while PICO experiments are inherently duplex reactions. This results in multiple combinations of clusters, which can be biased in their detection. The four clusters shown in Figure 4.8A can be biased separately or in arbitrary combinations (Table 6.1). False-positive, false-negative clustering bias and PSC were regarded separately. Of note, if a clustering bias is applied on the number of negative partitions \hat{n}_{AB-a-b^-} , this can be considered equal to a biased detection of the total number of partitions \hat{n} (Figure 4.8). For simplicity, it is assumed, that if multiple clusters are affected, the clustering biases are equal. This assumes that the biases are constant in the same dPCR run, which is a reasonable approximation of the reality. The simulation `dpcr_sim()` was set up in such a way that the argument `number_of_targets` was varied from 0 to 1000 in steps of 100, λ for both antibodies was varied from 0.05 to 0.30 in steps of 0.05, the clustering biases were varied in $\{0, 1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$ and the combinations of biased clusters were varied from `c1` to `c10` (Table 6.1). Each combination of `number_of_targets`, λ , `clustering_bias` and combination of biased clusters was simulated 50 times.

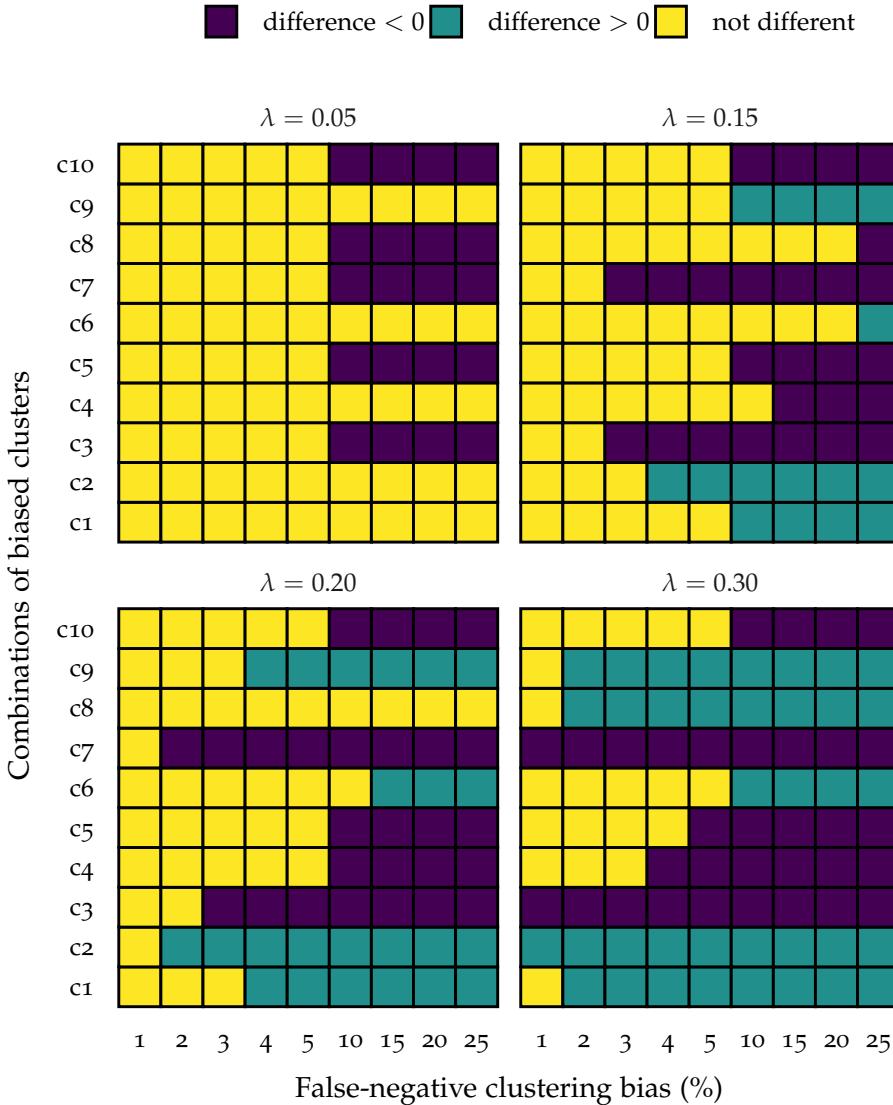


Figure 6.8: Qualitative impact of false-negative clustering bias on the `dDPCS` model. The clusters $\hat{n}_{AB^+a^+b^-}$, $\hat{n}_{AB^+a^-b^+}$, $\hat{n}_{D_{obs}}$ and $\hat{n}_{AB^-a^-b^-}$ (Figure 4.8A) were biased in the combinations mentioned in Table 6.1 (c1 to c10). The value of the clustering bias was varied in $\{0, 1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$. The tiles of each heatmap (for four chosen λ values), indicate whether the number of complexes after applying different combinations of clustering biases is smaller, larger or equal to unbiased conditions ($\text{clustering_bias} = 0\%$) using the linear model `lm(calculated_complexes ~ as.factor(clustering_bias))`. The quantitative differences can be obtained from Figure C.7.

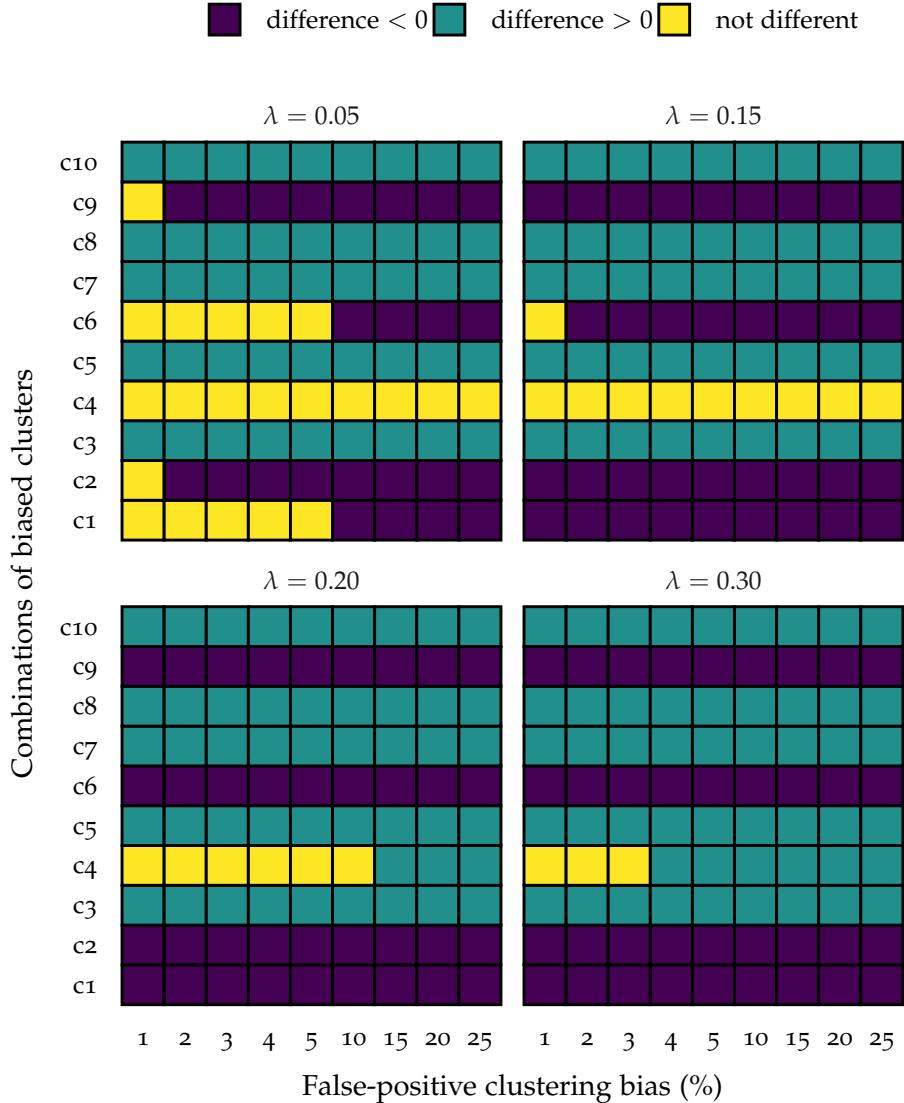


Figure 6.9: Qualitative impact of false-positive clustering bias on the dDPCS model. The clusters $\hat{n}_{AB+a^+b^-}$, $\hat{n}_{AB+a^-b^+}$, $\hat{n}_{D_{obs}}$ and $\hat{n}_{AB-a^-b^-}$ (Figure 4.8A) were biased in the combinations mentioned in Table 6.1 (c1 to c10). The value of the clustering bias was varied in $\{0, 1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$. The tiles of each heatmap (for four chosen λ values), indicate whether the number of complexes after applying different combinations of clustering biases is smaller, larger or equal to unbiased conditions (clustering_bias = 0%). using the linear model `lm(calculated_complexes ~ as.factor(clustering_bias))`. The quantitative differences can be obtained from Figure C.8.

Table 6.1: The clusters from Figure 4.8A can be biased separately or in arbitrary combinations. In case of multiple affected clusters, the clustering bias was the same for all clusters. The clustering biases were varied in $\{0, 1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$.

Bias combination	\hat{n}_{AB+a+b^-}	\hat{n}_{AB+a-b^+}	$\hat{n}_{D_{obs}}$	\hat{n}_{AB-a-b^-}
c1	x	-	-	-
c2	x	x	-	-
c3	-	-	x	-
c4	-	-	-	x
c5	x	-	x	-
c6	x	-	-	x
c7	-	-	x	x
c8	x	x	x	-
c9	x	x	-	x
c10	x	x	x	x

For the assessment of the impact of both false-negative and false-positive clustering biases, the number of complexes calculated under biased conditions were compared to the number of complexes calculated under ideal conditions, i.e. reference level was `clustering_bias = 0%`. The qualitative results (different < 0 , $= 0$ or > 0) were summarized in a heatmap (Figure 6.8 and Figure 6.9) and quantitative results (estimates, i.e. differences to ideal conditions) were summarized in a scatterplot (Figure C.7 and Figure C.8).

In general, for a false-negative clustering bias, the absolute difference to ideal conditions increases with λ (minimal difference at $\lambda = 0.05$, biased combination c7 and clustering bias 25 % was -142 complexes and maximal difference at $\lambda = 0.30$, biased combination c2 and clustering bias 25 % was 977 complexes (Figure C.7)) and with the value of the clustering bias (Figure C.7). This was inline with the results for the biased dPCR simulation (Figure 5.2A and C), because the clustering bias is defined as a proportion of the total number of positive partitions (Listing B.4) and thus its absolute value increases with λ (Majumdar et al., 2015). With increasing λ , the amount of random double positive partitions increases (Figure 6.6), thus the background from which the same number of complexes needs to be distinguished increases. For $\lambda = 0.05$, false-negative clustering biases below 5 % did not have a significant impact (Figure 6.8). Beyond 5 %, combinations of biases involving double positive partitions $\hat{n}_{D_{obs}}$ (c3, c5, c7, c8, c10) were significantly lower. In this λ range and for the number of targets simulated (0 to 1000), the double positive cluster contained mainly complex double positive partitions (Figure 6.6), which is a possible explanation for the observed behavior. The dDPCS model calculates the

number of random double positive partitions $\hat{n}_{AB-a+b+}$ based on the number of partitions positive for antibody a $\hat{n}_{AB+a+b-}$ and b $\hat{n}_{AB+a-b+}$ (Equation 4.32). Thus, a false-negative clustering of the antibodies but not the double positives (biased combinations c1, c2, c6 and c9) led to an overprediction of the number of complexes because there were more double positive partitions than the **DDPCS** model calculated based on the abundance of the antibodies. Conversely, a false-negative clustering of the number of double positive partitions and/or negative partitions (biased combination c3, c4 and c7) reduced the number of complexes in comparison to ideal conditions. The **DDPCS** model assumed more double positive partitions than were observed. Thus, it calculated less complexes. Interestingly, biased combination c8 became more insensitive towards clustering biases with λ . For $\lambda = 0.05$ the difference was negative but at $\lambda = 0.30$, the differences to ideal conditions were significantly larger (Figure 6.8). The increased tolerance towards clustering biases might be attributed to compensatory effect of λ . A too strong increase of λ led to an overcompensation (Figure 6.8). Biased combinations c5 and c10 were constant for the λ values examined. It is assumed that for a certain clustering bias (maximal 5 %), the errors cancel out, while a stronger bias eventually distorts the system.

Next, the impact false-positive clustering biases was examined. As for the biased simulation of the dPCR (Figure 5.2B and C), false-positive clustering biases were mostly independent of λ (Figure 6.9). Biased combinations c1 and c6 were not different from ideal conditions for $\lambda = 0.05$ and clustering biases below 5 %, but changed fast, while the c4 (clustering bias of negative partitions (Table 6.1)) required larger values of λ to experience a significant effect. The reason is that biasing the negative partitions changes the total number of partitions, which has an influence on the clustering bias. In general, the bias combinations c1, c2, c6 and c9 had a negative impact on the number of complexes. Thereby the number of antibodies is increased, while the number of double positive partitions remains constant. Thus, the **DDPCS** model expected more random double positive partitions than actually present and thus reduced the number of complexes (Figure 6.9 and Figure C.8). On the other hand, the bias combinations c3, c4, c5, c7, c8 and c10 had a positive impact on the number of complexes because the total number of double positive partitions was increased beyond what the **DDPCS** model expects based on the antibodies. Thus, it determined more complexes than under ideal conditions (Figure 6.9 and Figure C.8). Quantitatively, the bias combinations c8 and c10 (biased detection of antibodies and double positive partitions (Table 6.1)) monotonically increased with the clustering bias for all λ values (Figure C.8). In c5 and c7, the detection of double positive partitions and antibody a or negative partitions was biased. Interestingly, the bias combinations c5 and c7, increased monotonically until a certain clus-

tering bias and then reached a plateau. The level of the plateau and its onset depends on λ (Figure C.8). This means that increasing the number of partitions or the number of partitions positive for one antibody increases the number of determined complexes in a λ -dependent manner.

Overall, the clusters that are negatively or positively impacted by false-negative clustering bias are changed exactly into the opposite direction when applying false-positive clustering bias. However, as false-negative clustering bias is λ -dependent, there are some finer nuances when comparing the clustering biases in detail.

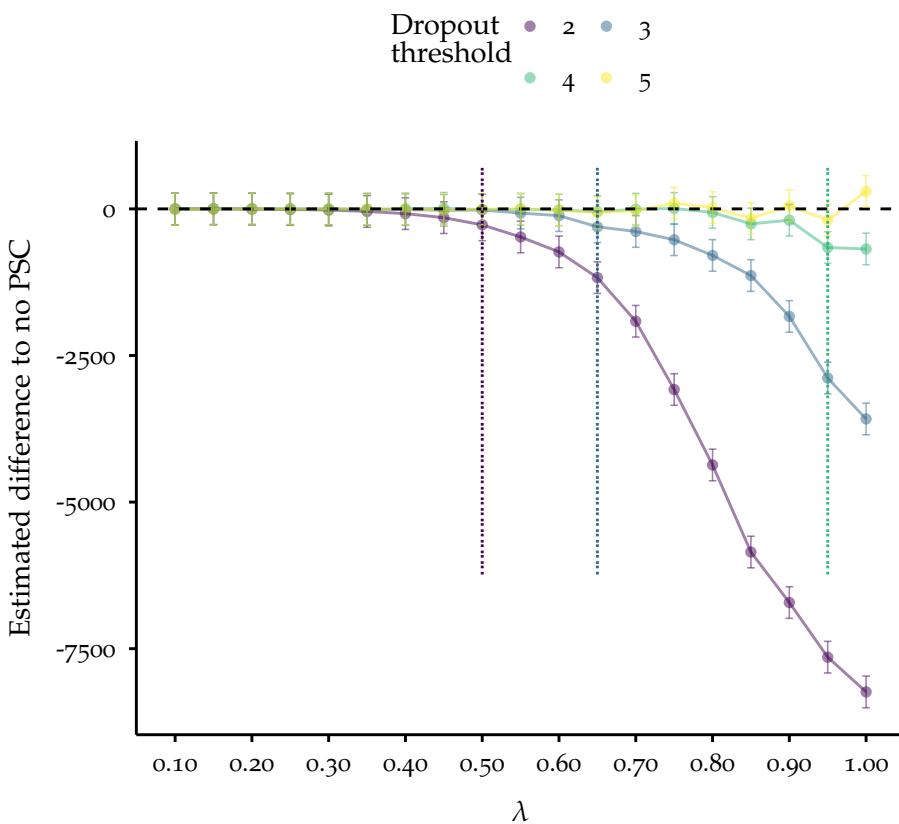


Figure 6.10: Scatterplot of the estimated impact (with 95 % CI as error bars) of dropout threshold on the number of calculated complexes. The linear model `lm(calculated_complexes ~ dropout_threshold * as.factor(lambda))` was used to estimate the difference of the interaction of dropout threshold and λ on the detection of complexes compared to an infinitely large dropout threshold, i.e. no PSC. The dropout threshold was varied in $\{2, 3, 4, 5\}$. The vertical dotted lines indicate the value of λ for which the estimate starts to be significantly different from 0.

PARTITION-SPECIFIC COMPETITION For the simulation of PICO experiments, the impact of PSC was assessed similarly to the simu-

lation of the dPCR ([Figure 5.3](#)). This effect is considered relevant in PICO experiments, since the signal generation, i.e. the formation of complexes ([Figure 4.3](#) and [Figure 4.4](#)), results in at least two DNA-labels per partition or more in the case of multiplex PICO assays ([Figure 4.2B](#)). The simulation `dpcr_sim()` was set up in such a way that the argument `number_of_targets` was varied from 0 to 1000 in steps of 100, λ for both antibodies was varied from 0.05 to 1.00 in steps of 0.05 and the `dropout_threshold` was varied in $\{2, 3, 4, 5, \infty\}$ successfully amplified target molecules per partition, i.e. DNA-labeled antibodies. Each combination of `number_of_targets`, λ and `dropout_threshold` was simulated 50 times.

The linear model `lm(calculated_complexes ~ dropout_threshold * as.factor(lambda))` was used to estimate the difference of the interaction of dropout threshold and λ on the detection of complexes compared to an infinitely large dropout threshold, i.e. no **PSC**. The impact of the dropout threshold alone was not significant and thus neglected (data not shown). Similar to the simulation of **PSC** in a dPCR ([Figure 5.3](#)), the estimated difference to ideal conditions increased with λ ([Figure 6.10](#)). The lower the dropout threshold, the lower the λ from which on the difference was significant ([Figure 6.10](#)). For $\lambda < 0.50$, the difference was not different from 0 for all dropout thresholds. At $\lambda = 0.50$, the difference for a dropout threshold of maximal two successfully amplified molecules per partition became significant. For a dropout threshold of three successfully amplified molecules per partition, the difference was significant at $\lambda = 0.65$ and for a dropout threshold of four successfully amplified molecules per partition, the differences was significant for $\lambda = 0.95$. A dropout threshold of five successfully amplified molecules per partition did not show a change in the complexes detection for the simulated values of λ . If a complex is contained in a partition, this partitions immediately contains two DNA-labels ([Figure 4.3](#)). Thus, the λ for which the estimated difference is significant is lower compared to a dPCR, which only relies on single DNA molecules per partition ([Figure 5.3](#)). Regarding multiplex PICO assays ([Figure 4.2B](#)), the effect is enforced because multiple DNA-labeled antibodies bind to the same target and are coerced to localize in the same partition. Thus, for multiplex PICO assays the described effect is decoupled of λ . In general, **PSC** has a negative impact on the determined number of complexes since, simply because of the loss of information, but can easily be avoided by lower values of λ through stronger dilution or larger number of partitions ([Whale et al., 2016](#)).

6.1.3 Summary

Using simulated data, it was shown that the **ddPCS** model exhibits a strong λ dependence, which defines constraints for experimental

design. Values of $\lambda > 0.55$ should be avoided, because this defines the onset of the bifurcation (Figure 6.4). The number of the determined complexes and the number of random double positive partitions are significantly higher than the corresponding ground truth for $\lambda > 0.25$ (Figure 6.6). Thus, a suitable λ range for reliable determination of complexes using the **dDPCS** model is $\lambda \leq 0.25$. For this range, the accuracy based on the estimates of Figure 6.7 varies between 92 % and 98 %. Furthermore, a false-negative clustering bias with a maximal clustering bias of 5 % does not have a significant impact within this λ range (Figure 6.8). Similarly, **PSC** can be regarded as insignificant for the relevant range of λ (Figure 6.10). The false-positive clustering bias cannot be controlled via λ (Figure 6.9) as it rather depends on the properties of the optical system for partition imaging. To achieve a λ in the desired range, there are two parameters that can be adjusted: the dilution prior to partitioning or the number of partitions. Of note, when increasing the number of partitions, there is no lower limit for λ . However, technical limitations currently prevent partitioning at large scale and available systems exhibit total partition numbers of about ten thousands of partitions (Basu, 2017; Lai et al., 2023; Schulz et al., 2020, 2021; Tan et al., 2023). In contrast, achieving low λ values by dilution is easily achieved at hardly any costs but risks the loss of complexes because of dilution effects.

6.2 USING EXPERIMENTAL DATA

Phase II of the **dDPCS** model validation first required the qualification of the dPCR system and the DNA-labels. Then, the **dDPCS** model was validated using DNA-only PICO experiments (stage I) and recombinant protein (stage II).

6.2.1 Qualification of dPCR system and DNA-labels

dPCR is capable of detecting single molecules and by this resolve fold differences down to 1.16 (Basu, 2017; CM Hindson et al., 2013). Therefore, dPCR can be considered by its principle a highly sensitive detection system for DNA. For a reliable setup of the validation experiments, variances from the dPCR device or the DNA-labels needed to be identified to possibly account for them.

DNA-LABELS The schematic structure of the DNA-labels (Table 10.1) used in Section 6.2.2 is shown in Figure 6.11. The P8 and BLA label 2 have distinct primer and probe binding sites (Figure 6.11A). As DNA-only PICO experiments do not comprise a binding reaction because of the absence of antibodies (Figure 6.16), complexes are mimicked by a synthesized joint oligonucleotide from the P8 and the BLA label 2 (Figure 6.11B). This synthetic joint is called *artificial*

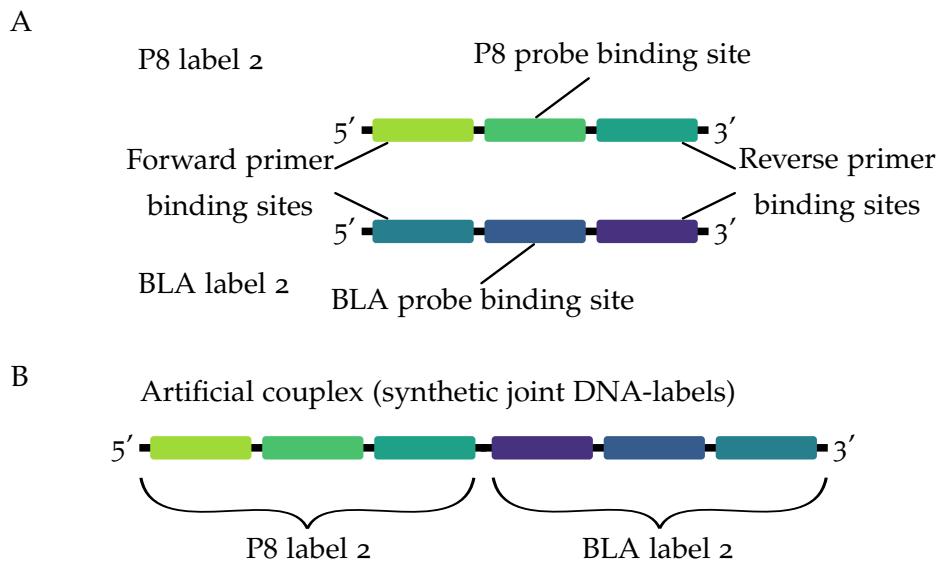


Figure 6.11: Schematic depiction of DNA-labels and *artificial couplex* for DNA-only PICO experiments.

A: The P8 label 2 (51 bp) and the BLA label 2 (57 bp) (Table 10.1) each have distinct binding sites for forward and reverse primers (Table 10.8) and probes (Table 10.7).

B: The dimer (128 bp) is a synthesized oligonucleotide of both labels from A. It thus imitates a couplex, as partitions containing it will become double positive. It can also be called *artificial couplex*.

couplex, which is expected to render a partition double positive because it contains the binding sites for both probes. The DNA-labels and the artificial couplex are characterized using dPCR regarding amplification efficiency and concentration to ensure a reliable setup of the DNA-only PICO experiments (Section 6.2.2).

The stock concentration of the DNA-labels and the artificial couplex was monitored using dPCR over the time-course of ~ 10 months (Figure 6.12). Across the entire time range, the P8 label 2 had an average concentration of $54.74 \mu\text{mol L}^{-1}$ with a CV of 38 %. The BLA label 2 had an average concentration of $38.46 \mu\text{mol L}^{-1}$ with a CV of 59 %. The artificial couplex had an average concentration of $2.97 \mu\text{mol L}^{-1}$ with a CV of 63 %². In all cases, this was a deviation from the expected concentration stated by the manufacturer, which is $100.00 \mu\text{mol L}^{-1}$ for the labels and $10.00 \mu\text{mol L}^{-1}$ for the artificial couplex. Additionally, the concentration varied significantly over time. The difference between the maximal measured concentration and the minimal concentration normalized to the overall mean was 1.23-fold, 1.91-fold and 2.44-fold for P8 label 2, BLA label 2 and the artificial couplex, respectively. Of note, the manufacturer determined the concentration via optical density. This approach detects all nucleic acids in the

² The concentration of the artificial couplex was determined as the average of the fluorescent signal from both probes (BLA and P8).

sample, while determining the concentration via dPCR with highly specific MGB-hydrolysis-probes (Kutyavin, 2000) only detects correct amplicons. Thus, the difference in these values can be explained by the fraction of labels that cannot be amplified or cannot be detected because the binding sites for primers and/or probes contain sequence errors. This is crucial information for the setup of the DNA-only PICO experiments (Section 6.2.2) because the expected number of complexes will be calculated based on the concentration of the artificial complex.

Additionally, the ability of the artificial complex to render a partition double positive needed validation. It was thus analyzed by dPCR. This oligonucleotide contains binding sites for the P8 and BLA probe (Figure 6.11B), however, a significant amount of the positive partitions was single positive ($\hat{n}_{AB-a^+b^-} = 880$ (green channel, P8 fluorescent signal) and $\hat{n}_{AB-a^-b^+} = 772$ (yellow channel, BLA fluorescent signal)), and 1187 partitions were double positive ($\hat{n}_{D_{obs}}$) (Figure C.9A). This indicates that a defined volume of the solution of the artificial complex contains three populations:

1. molecules, which generate double positive partitions;
2. molecules, which generate single positive partitions in the green channel, or
3. molecules, which generate single positive partitions in the yellow channel.

To quantify their composition, the artificial complex was analyzed in a dilution series (Figure 6.13). The fractions of double and single positive partitions from all positive partitions was constant until $\lambda \approx 0.15$. Then, the fraction of double positive partitions increased, while the fractions of single positive partitions decreased (Figure 6.13). This observation is related to the increased colocalization of DNA-labels in the same partition, which correlates with λ (Section 4.2.2). Thus, the colocalization of molecules from type two and three (see enumeration directly above, Item 2 and Item 3) in the same partition results in the same outcome as a molecule of type one (see enumeration directly above, Item 1). Considering $\lambda \leq 0.15$ (51 values of originally 74), where the probability of a partition to contain ≤ 1 molecule is high, on average 41.1 % of the positive partitions were double positive partitions ($\hat{n}_{D_{obs}}$), while 31.1 % were single positive in the green channel ($\hat{n}_{AB-a^+b^-}$) and 27.8 % were single positive in the yellow channel ($\hat{n}_{AB-a^-b^+}$). Furthermore, the ratio r_{ab} (Equation 6.1) between all partitions with a signal in the green channel ($\hat{n}_{AB+a^+b^-}$) and all partitions with a signal in the yellow channel ($\hat{n}_{AB+a^-b^+}$) can be calculated to identify potentially uneven amplification or probe binding issues. The expected value is 1.

The probability for more than 1 molecule per partition is ~1 % at $\lambda = 0.15$.

$$r_{ab} = \frac{\hat{n}_{AB^+a^+b^-}}{\hat{n}_{AB^+a^-b^+}} \quad (6.1)$$

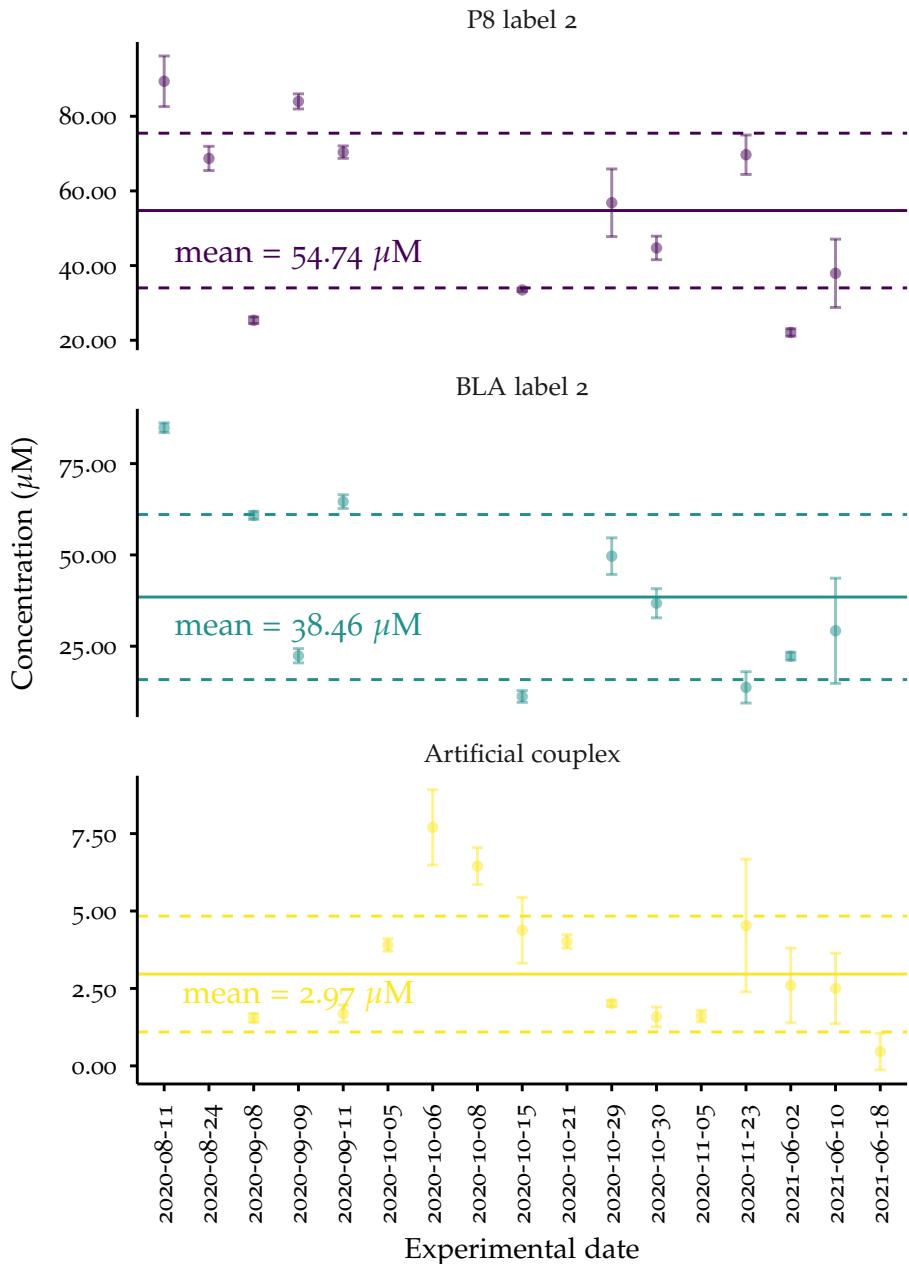


Figure 6.12: Time-course quality control of oligonucleotide concentration. The concentration of the P8 label 2, the BLA label 2 and the artificial couplex was monitored over the course of ~ 10 months. The P8 label 2 has an average concentration of $54.74 \mu\text{mol L}^{-1}$ with a CV of 38 %, the BLA label 2 has an average concentration of $38.46 \mu\text{mol L}^{-1}$ with a CV 59 % and the dimer has an average concentration of $2.97 \mu\text{mol L}^{-1}$ with a CV 63 %. The maximal concentration difference for each molecule is $67.30 \mu\text{mol L}^{-1}$ (P8 label 2), $73.60 \mu\text{mol L}^{-1}$ (BLA label 2) and $7.24 \mu\text{mol L}^{-1}$ (artificial couplex). The sample size per date is ≥ 3 .

r_{ab} based on the fractions on positive partitions for $\lambda \leq 0.15$ was 1.05 (Figure 6.13), while it was 1.07, when considering all λ values (Figure C.9B). The variability of r_{ab} decreased with higher values of λ . Of note, r_{ab} can only be calculated for DNA molecules with two probe binding sites like the artificial couple.

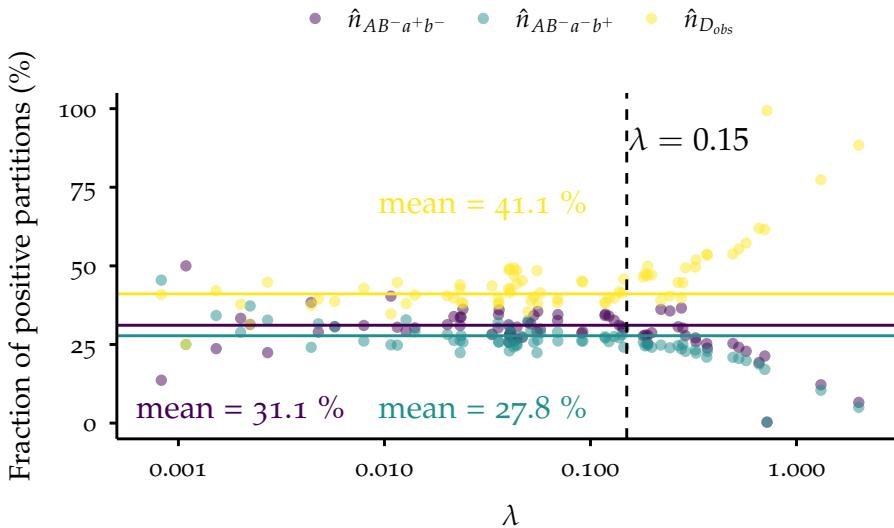


Figure 6.13: Scatterplot of fractions of single and double positive partitions ($\hat{n}_{AB-a^+b^-}$, $\hat{n}_{AB-a^-b^+}$ and $\hat{n}_{D_{obs}}$) from all positive partitions for the artificial couple only. Fractions of positive partitions were calculated for different λ values of the artificial couple. The average fraction of single and double positives was calculated for $\lambda \leq 0.15$ because the probability of more than 1 molecule per partition is then $\sim 1\%$. This excluded 23 values from the calculation. On average 41.1 % of the positive partitions are double positive partitions, while 31.1 % are single positive in the green channel ($\hat{n}_{AB-a^+b^-}$) and 27.8 % are single positive in the yellow channel ($\hat{n}_{AB-a^-b^+}$). This results in $r_{ab} = 1.05$. Data was from 13 independent experiments.

The artificial couple was synthesized, which means that the orientation of the synthesis is inverse to nature (3' → 5'). Furthermore, during the chemical synthesis the probability for a wrong base increases with the length of the sequence (Ma et al., 2012). One would thus expect that the P8 probe binding site has a higher sequence variability than the BLA probe binding site (Figure 6.11B) and that the binding of the highly specific MGB-probes is affected hereby (Kutyavin, 2000). This would mean $r_{ab} < 1$. However, there might be a second effect that overlaps. A successful amplification of a DNA fragment depends on the primer pair. The primers for the amplification of the BLA label 2 have a difference in melting temperature (T_m) of 4.7 °C (Table 10.8), while the maximal recommended difference is 2.0 °C (Álvarez-Fernández, 2013).

Thus, the P8 label 2 might be preferably amplified. The MGB-probe is more efficiently binding to the BLA probe binding site because of the potential sequence errors. This might also explain why r_{ab} is close to 1. Both effects are present and counteract each other, yet are strong enough to generate almost equal numbers of single positive partitions when the other effect is absent.

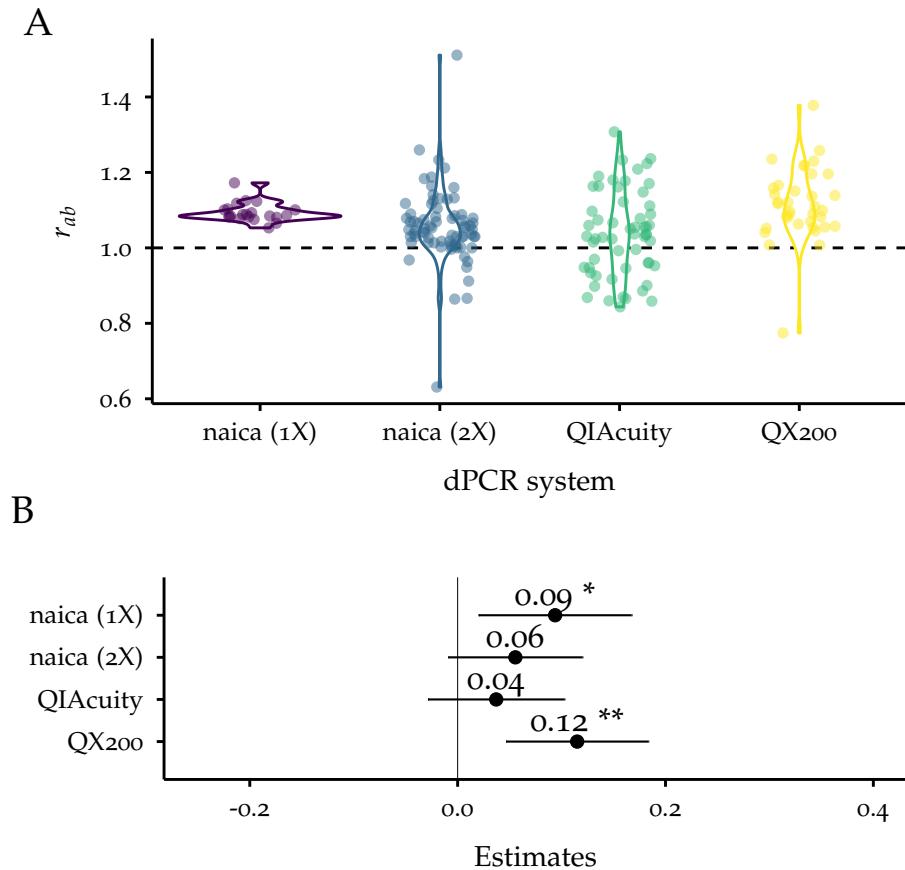


Figure 6.14: Comparison of dPCR systems using the artificial couple.

A: Violin plots of r_{ab} for the dPCR systems. The expected value is 1 (dashed line). DPCR in the naica system was performed with 1X and 2X ToughMix concentration to potentially reduce PSC (Whale et al., 2016). Sample sizes were as follows: $n_{\text{naica (1X)}} = 21$, $n_{\text{naica (2X)}} = 74$, $n_{\text{QIAcuity}} = 57$, $n_{\text{QX200}} = 36$. The dots were jittered to avoid overplotting.

B: Forest plot of estimates with 95 % CI from the linear model `lm(ratio ~ dPCR_system)` compared to the expected value. The adjusted R^2 for the model was 0.079. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the CI is the associated uncertainty.

Overall, the potentially better amplification of P8 label 2 slightly prevails by 5 % (because $r_{ab} = 1.05$). A possible solution to circumvent these issues as well as to identify the causes for this is an approach called “amplicon-size multiplexing” (McDermott et al., 2013; Whale

et al., 2016), where a DNA-intercalating dye such as EvaGreen is used to detect double-stranded DNA (Figure 2.3B). The fluorescent intensity depends on the amplicon length and by designing primers in such a way that amplicons of different size are generated. This allows multiplexing in the same fluorescent channel (McDermott et al., 2013), which is necessary for the detection of the artificial couple. The advantage is then that the fluorescence generation is not sequence-specific, reducing effects of potential sequence errors detected by the MGB-probes.

DPCR SYSTEMS The artificial couple can also be used to qualify the available dPCR systems, which differ in the partitioning (droplets vs. chambers) and imaging steps (FACS-like vs. high-resolution imaging) and further subtle differences (Tan et al., 2023). r_{ab} was used to compare the dPCR systems. The dPCR system with the least deviation from the expected value of 1 will be used for subsequent PICO experiments. The QX200 system from Bio-Rad showed the highest value of $r_{ab} = 1.12$, which was significantly different from 1 (Figure 6.14). Overall, the naica system with 1X ToughMix showed the lowest SD, but r_{ab} was significantly higher than 1 (1.09 ± 0.03). In multiplex dPCRs, it is expected that PSC is elevated so that increasing the resources is a considerable approach (Markoulatos et al., 2002; Whale et al., 2016). Thus, the concentration of the ToughMix was doubled for dPCRs on the naica system. This did not change the number of generated droplets (Figure 6.15A), but reduced the ratio to 1.06 ± 0.10 at the cost of a higher SD (Figure 6.14). The QIAcuity systems showed the lowest r_{ab} with a value of 1.04 ± 0.11 . Both, the QIAcuity and the naica system with 2X ToughMix were not significantly different from 1 (Figure 6.14B). This suggests usage of one of these two dPCR systems for PICO experiments.

There are further practical considerations for the choice of the dPCR system. The naica system has six fluorescent channels and allows for a customized spillover compensation (Madic et al., 2016), while the QIAcuity system has four usable channels³ without spillover compensation.

The QIAcuity dPCR system accounts for variances in partition size using an approach similar to Majumdar et al. (2017) and by this improves the precision of the calculation (Košir et al., 2017). For the naica and the QX200 dPCR systems no such correction is known.

³ In fact the QIAcuity has five fluorescent channels, but only four can be used for PICO experiments. The population of negative partitions in the crimson channel has a high background fluorescence, possibly because of spillovers from the reference channel (fluorescence channel for the detection of the partitions themselves). Thus, the difference between the populations of positive and the negative partitions is blurred, which complicates clustering (Figure C.10). However, usage of different fluorescent dyes and a manual spillover compensation based on the available relative fluorescence unit (RFU) files might help to circumvent this issue in future research.

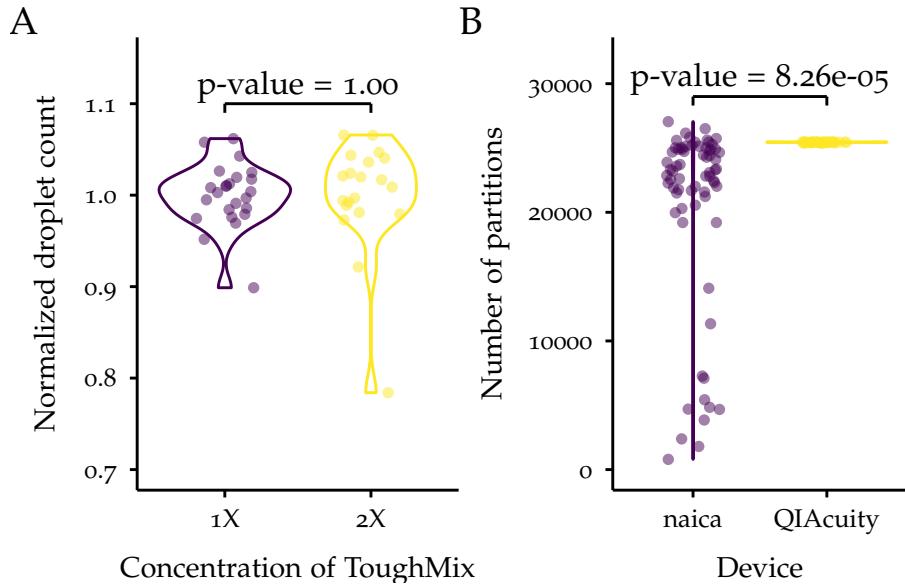


Figure 6.15: Comparison of number of partitions.

A: Violin plots of normalized droplet count in naica dPCR system using 1X or 2X ToughMix. The number of droplets generated may vary depending on age and batch of the Sapphire chips as well as on atmospheric circumstances favoring electrocoalescence (personal communication with Stilla Technologies), thus droplet counts were normalized to the mean droplet count per run. The dots were jittered to avoid overplotting. Data was from 4 independent experiments. Droplet counts were compared using `lm(droplet_count ~ mastermix_conc)`.

B: Comparison of number of evaluable partitions between naica and QIAcuity dPCR systems. On average, a dPCR in the QIAcuity dPCR system has $25\,439 \pm 32$ partitions and in the naica dPCR system $20\,497 \pm 7215$ partitions. The QIAcuity dPCR system has an estimated number of 4942 partitions more than the naica system (95 % CI from 2551 to 7332). The dots were jittered to avoid overplotting. Data was from 3 independent experiments. Number of partitions were compared using `lm(number_of_partitions ~ dPCR_system)`.

Resource efficiency is an aspect that puts financial constraints on the choice of dPCR system. Considering a partition volume of $V_{p_{\text{QIAcuity}}} = 0.78 \text{ nL}^4$ and $V_{p_{\text{naica}}} = 0.59 \text{ nL}$ and the total reaction volumes of $40.00 \mu\text{L}$ and $20.00 \mu\text{L}$ for QIAcuity and naica dPCR systems, respectively, gives a dead volume of 50.39 % for the QIAcuity dPCR system and 39.53 % for the naica dPCR system. Financial constraints are also the reason why the QIAcuity system was not operated with higher concentrations of QIAcuity Probe PCR Kit. Furthermore, the effect of the increased

⁴ This value was determined using Equation 4.9 and the output of the QIAcuity Software Suite. It is known that the actual V_p differs from the information provided by the manufacturer (Košir et al., 2017).

QIAcuity Probe PCR Kit concentration on the fluidic properties is unknown and might impair successful partition filling.

A final aspect influencing the choice of the dPCR system is the stability of the partitions and the variance of the number of partitions. It is apparent that in this regard chamber digital PCR systems are preferred because issues like coalescence of droplets are non-existent. Indeed, the QIAcuity dPCR system partitioned the reaction volume on average in $25\,439 \pm 32$ partitions, while the naica systems partitioned the reaction volume in $20\,497 \pm 7215$ partitions. This had an estimated difference of 4942 partitions (95 % CI from 2551 to 7332) (Figure 6.15B). In contrast, droplet digital PCR systems usually enable the recovery of the partitioned volume for further analyzes.

The precision of a dPCR increases with the number of partitions (Basu, 2017).

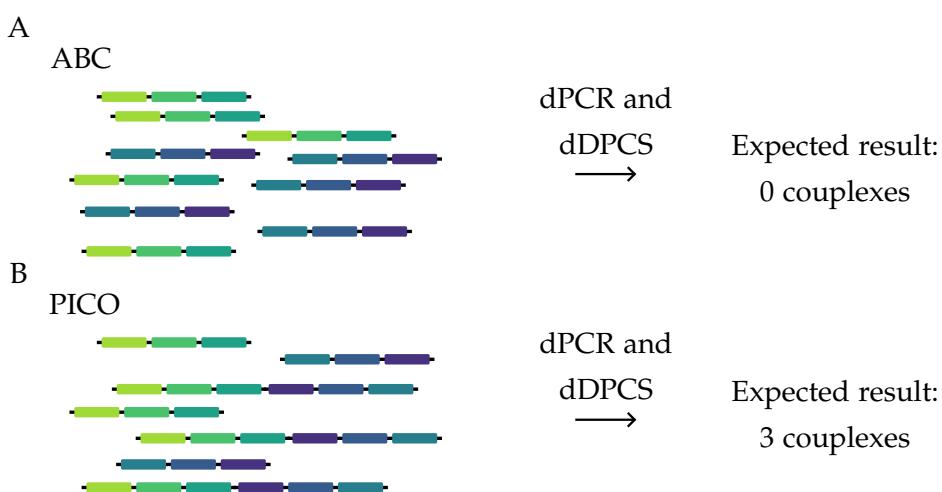


Figure 6.16: Exemplary setup of DNA-only PICO experiments.

A: Schematic setup of a DNA-only ABC PICO experiment (no target present, see Figure 4.6). A mixture with equimolar concentrations of the DNA-labels (Figure 6.11) are subjected to a dPCR and from the 2-dimensional raw data the number of complexes is calculated using the dDPCS model. The expected number of complexes is 0, since $\hat{n}_{D_{obs}}$ can only consist of \hat{n}_{d_1} (Figure 4.8B). B: Schematic setup of a DNA-only PICO experiment (Figure 4.7). Instead of 6 DNA-labels of each type (like in A), only 3 DNA-labels of each type and 3 artificial complexes are composing the mixture. The total number of probe binding sites and thus detectable signals remains constant and $\hat{n}_{D_{obs}}$ can now be explained by multiple states (Equation 4.22 and Figure 4.8B). The expected number of complexes is 3. This schematic depicts an exemplary DNA-only PICO experiment. In real PICO experiments, multiple ratios between DNA-labels and artificial complexes were analyzed.

SUMMARY The DNA-labels as well as the artificial complex showed a high concentration variance over time (Figure 6.12). This has significant impact on the DNA-only PICO (Section 6.2.2), which has the intention to identify the LOD of PICO assays. As the expected number

of complexes is calculated based on the oligonucleotide concentrations, this value needs to be reliable. It is therefore suggested to run an in-process concentration control to calculate the actual oligonucleotide concentration. Furthermore, only 41.1 % of the artificial complexes added, generate a double positive partition (Figure 6.13). The required degree of multiplexing and the need for recovering the partitioned volume influences the choice of dPCR system. Among the here compared dPCR systems, chamber-based systems provided more reliable partitioning and higher numbers of evaluable partitions (Figure 6.15). However, when using the QIAcuity, it is not recommended using the crimson channel for fluorescence detection (Figure C.10).

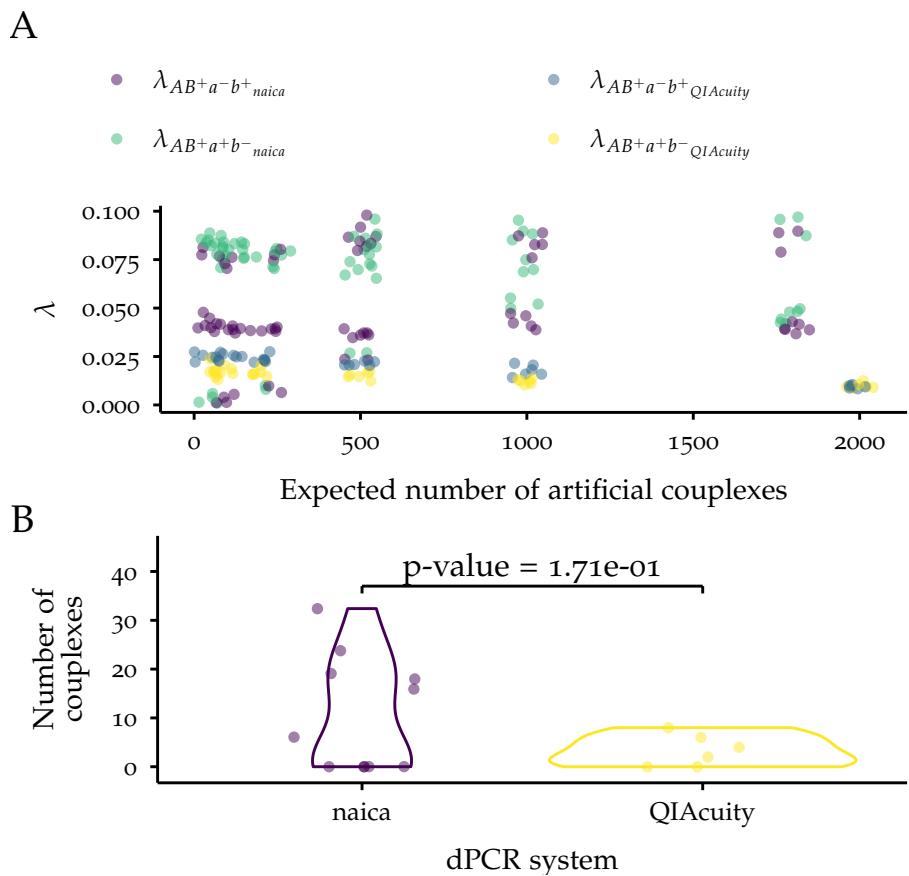


Figure 6.17: Controls from naica and QIAcuity DNA-only PICO experiments.

A: Scatterplot of λ values against the expected number of artificial complexes. The dots were jittered to avoid overplotting. Data was from 3 independent experiments.

B: Violin plots of detected complexes from DNA-only ABC PICO experiments. The mean number of complexes detected in the naica dPCR system is 10.5 ± 11.8 ($n = 11$) and in QIAcuity dPCR system 3.3 ± 3.3 ($n = 6$) with SD as error. The estimated difference between the two systems is -7.14 complexes (95 % CI from -17.70 to 3.46) using the linear model `lm(calculated_complexes ~ dPCR_system)`.

6.2.2 Based on DNA

With the knowledge from [Section 6.2.1](#), the **dDPCS** model was applied in this section to calculate the number of complexes from DNA-only PICO experiments. The only molecules in such PICO experiments are DNA-labels and the artificial complex, which imitates a real couplex ([Figure 6.11](#)). The binding reaction is thus avoided ([Figure 4.3B](#)), which might introduce variations between observed and expected number of complexes in real PICO experiments originating from antibodies issues, e.g. specificity (Weller, 2016).

Different mixtures of DNA-labels and artificial complexes were prepared and analyzed in the dPCR ([Section 9.3.1](#)), starting with a mixture without artificial complexes imitating an ABC. Of note, only 41.1 % of the added artificial complexes truly render a partition double positive, while the other 58.9 % render a partition single positive ([Figure 6.13](#)). This factor is considered when calculating the expectation value. Because of the day-to-day variance in the concentration of the DNA-labels and the artificial couplex ([Figure 6.12](#)), in-process concentration controls were added, which allowed to refine the expectation values. In general, the aim was to maintain a constant signal from the P8 and BLA probe independent of the origin (label vs artificial couplex). Another point-of-view is to keep the number of probe binding sites constant. For example, the addition of 1 artificial couplex entails that the number of P8 label 2 and BLA label 2 needs to be reduced by ~ 1 ([Figure 6.16](#))⁵. Furthermore and in contrast to the simulated PICO experiments, the number of partitions may vary ([Figure 6.15](#)). Thus, λ will be used as an indication for the constant number of probe binding sites as this is independent of the total number of partitions n ([Equation 4.8](#)). Then, in analogy to [Figure 4.8](#), the λ values can be written as $\lambda_{AB+a^+b^-}$ (green channel detecting the fluorescent signal from the P8 label) and $\lambda_{AB+a^-b^+}$ (yellow channel detecting the fluorescent signal from the BLA label). In general, experiments were designed to ensure $\lambda \leq 0.25$ ([Section 6.1](#)). These experiments were conducted with the naica (with 2X ToughMix) and the QIAcuity dPCR systems.

As intended, λ was constant for the different systems and the different mixtures of DNA-labels and artificial complexes ([Figure 6.17A](#)). This indicates that the overall number of probe binding sites was constant. In a PICO experiment with antibodies, their number and thus the number of DNA-labels is constant. The number of targets may change depending on their concentration in the sample and therewith the number of formed couplexes. Imitation of this behavior was thus achieved by this experimental design.

The number of couplexes in the mixture without the artificial couplex calculated by the **dDPCS** model is expected to be 0, since the

⁵ Cave: The actual calculation is more elaborated since only 41.1 % of the added artificial couplexes render a partition double positive.

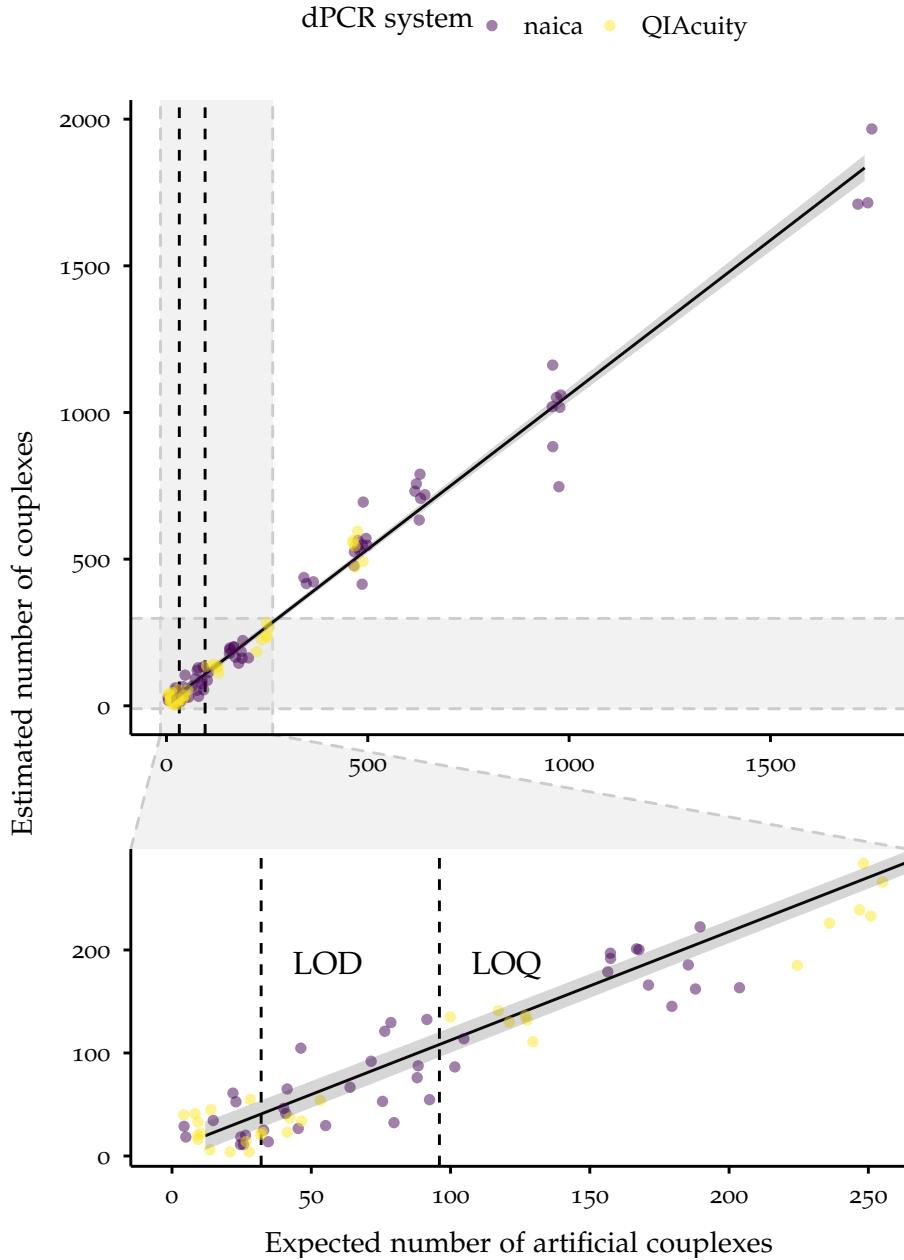


Figure 6.18: Scatterplot of calculated complexes against expected artificial complexes with zoom. The expectation was compensated by the fraction of functional artificial complexes (Figure 6.13) and the variability of the concentration (Figure 6.12). The linear model `lm(calculated_complexes ~ expected_complexes)` estimates a change of 1.05 with 95 % CI from 1.02 to 1.08 for the calculated complexes with the expected artificial complexes (black line with CI ribbon). The choice of dPCR system is not significant (data not shown). The vertical dashed lines indicate LOD and LOQ according to the European Medicines Agency (2023). The dots were jittered to avoid overplotting. Data was from 3 independent experiments. The corresponding λ values are shown in Figure 6.17A. This data was also already presented in the manuscript by Gross et al. (2024).

number of double positive partitions can be explained by the random colocalization of the DNA-labels only (Figure 6.16 and Figure 4.8). Indeed, the mean number of calculated complexes (\pm SD) was 10.5 ± 11.8 and 3.3 ± 3.3 for the naica and the QIAcuity dPCR systems, respectively. The difference between these two dPCR systems was not significant (Figure 6.17B). As these results were not significantly different from 0, there is no arbitrary offset in samples without a target.

For samples containing the artificial complex, the **dDPCS** model calculated a number of complexes significantly different from 0. The coefficient of determination R^2 between calculated and expected complexes was 0.98. The intercept of the linear regression was 7.05 (95 % CI from -6.73 to 20.80). Together with Figure 6.17B, this means that the **dDPCS** model does not exhibit an arbitrary offset. The estimated slope was 1.05 (95 % CI from 1.02 to 1.08). The accuracy can thus be considered 95 %. This means that the **dDPCS** model overpredicts the number of complexes by 5 %. This is along with findings from the Section 6.1, where the **dDPCS** showed a tendency to overpredict the number of complexes for high values of λ . Again this underlines the need for a λ threshold.

Additionally, **LOD** and **LOQ** were calculated according to the recommendations from the European Medicines Agency (2023) and amount to 32 and 96 complexes, respectively. In a total reaction volume of $40\text{ }\mu\text{L}$, this corresponds to a sensitivity of 4 amol L^{-1} .

SUMMARY Using DNA-only PICO experiments, the **dDPCS** model could be validated in highly controlled manner, using only DNA-labels and artificial complexes in the dPCR. This enables identifying sources of error in the later stages of the validation procedure, where antibodies are used. In summary, the **dDPCS** model showed high specificity and accuracy with attomolar sensitivity.

6.2.3 Based on recombinant protein

In this final stage of the validation, the **dDPCS** model was applied to calculate the number of complexes from PICO experiments using two DNA-labeled antibodies and recombinant protein as target. In contrast to Section 6.2.2, this entails higher biological complexity, because the protein-to-nucleic acid transformation via DNA-labeled antibodies (Janssen et al., 2013; Klebes et al., 2024; Nong et al., 2012) is now included (Figure 4.3B and Figure 4.4).

The protein used was a soluble recombinant **HER2** comprising the extracellular domain of native **HER2**. The rationale behind the usage of recombinant **HER2** is that well-characterized, high quality antibodies **TTZ**⁶ and **PTZ**⁷ are available (Tsao et al., 2022). This avoids the usage

⁶ Extensive information of TTZ.

⁷ Extensive information of PTZ.

of research antibodies and their potential quality issues (Weller, 2016). Therefore, variability originating from unspecific antibody binding was reduced to a minimum. Moreover, it is known, that **TTZ** and **PTZ** can simultaneously bind the extracellular domain of **HER2** (Nahta et al., 2004; Nami et al., 2018; Sharma et al., 2021), which is a requirement for a successful PICO assay (Figure 4.4). This holds promise for a reliable antibody-target binding step in the binding reaction.

The experiments conducted so far used the green and the yellow fluorescent channel of the dPCR systems for the detection of the DNA-labels (Section 6.2.2). For multiplex PICO assays (Figure 4.2B), however, it is crucial that independent of the fluorescent channels for detection of the DNA-labels, the signal is the same for the same input. This ensures that the observed differences can be used to make conjectures on true differences in proteoform and/or **PPI** abundance. The QIAcuity dPCR system exhibited poor quality for the detection of targets in the crimson channel (Figure C.10 and Section 6.2.1), which was thus not used for these experiments. Prior to PICO experiments, **TTZ** and **PTZ** were labeled with DNA-labels (Section 9.2.3) (Chio & Bane, 2020; Gong et al., 2016; Jeney, 2023a). Vistain et al. (2022) showed by **FACS** that such a method for antibody labeling does not impair their binding ability. This is another requirement for the PICO technology. The DNA-labels (Table 10.2) were designed in such a way that they can be amplified with the same pair of primers (Table 10.8) and sequence differences were limited to the probe binding site (Table 10.7) and the barcode⁸. The use of the same primer pair for amplification of different DNA-labels reduces potential amplification biases (Álvarez-Fernández, 2013). **TTZ** and **PTZ** were labeled in such a way that all possible combinations of fluorescent channels (green, yellow, orange and red channel) for readout can be compared (Table 10.6)⁹. This results in six combinations, the so called *colorpairs*: green-yellow (GY), green-orange (GO), green-red (GR), yellow-orange (YO), yellow-red (YR) and orange-red (OR) (Table C.1). These acronyms will be used in the following to refer to the combination of fluorescent channels for readout. The labeling efficiency (Section 9.2.5) varied between 65.1 % and 99.2 % (Table 10.6).

The used recombinant **HER2** has a purity of 100 % by size exclusion chromatography with multiangle light scattering (SEC-MALS) (Gross et al., 2024). This underlines another advantage of validation experiments with recombinant proteins, because matrix effects are reduced (Selby, 1999; Wood, 1991), and the sample is a homogeneous solution of the target of interest only. For future studies, it is conceivable to

⁸ The barcode is a stretch of 10 random nucleotides, which allows for distinguishing individual molecules after sequencing, i.e. unique molecular identifier.

⁹ This implies that the P8 label is always detected with a FAM-probe in the green channel, the BLA label with a HEX-probe in the yellow channel, the NOS6 label with an Atto550-probe in the orange channel and the ORC7 label with a TxRed-probe in the red channel (Figure C.14).

add BSA or FBS imitating a complex matrix. In contrast to previous validation steps for the **dDPCS** model (Section 6.1 and Section 6.2.2), the calculation of an expectation values is not possible. While the absolute number of targets in the sample solution can be determined by, for instance, total protein detection methods (Section 2.1), the formed number of complexes is hard to calculate as this requires knowledge about the K_d values (Figure 4.4 and [absolute quantification of proteoforms with PICO](#)). Furthermore, this requires usage of the **AQ model** from Gross et al. (2024), which is beyond the scope of this thesis.

The objective of the first experiment using the two DNA-labeled antibodies **TTZ** and **PTZ** with different labels in different combinations (Table C.1) was to determine the consistency of the couple counts detected by the different color pairs. As the antibodies are always the same, the different color pairs should yield the same results.

The values of all groups together were used as a reference for the linear model because of the lack of an appropriate expectation value. The colorpairs GY, GO, GR and YO show estimated differences between –17 and 54 complexes, which equal 5 % and 15 % of the reference mean (354 complexes). These estimated differences are not considered significant (Figure 6.19B). The estimated difference of YR to reference was 56 complexes (95 % CI from 2 complexes to 111 complexes) was considered significant. However, the relative deviation from the reference mean is 16 %. Strikingly, the estimated difference of the colorpair OR to the reference was the largest (–221 complexes with 95 % CI from –292 complexes to –150 complexes). This is an indication that the detection and calculation of the number of complexes with the colorpair OR is impaired and might not be reliable.

As said, the calculation of a reference value requires the application of the **AQ model** (Gross et al., 2024), which is beyond the scope of this thesis. Therefore, the determination of the accuracy is not possible. However, the expectation that all measurements should result in the same number of complexes still holds true because the same antibodies and the same concentration of recombinant **HER2** was used. This gives rise to several hypotheses which can be examined. The differences can originate from:

1. different **LE**;
2. varying background signal (**ABC**, i.e. blank);
3. spillovers between fluorescent channels;
4. labeling of the antibodies, which changed their affinity to their target;
5. amplification biases of the DNA-labels, or
6. clustering issues.

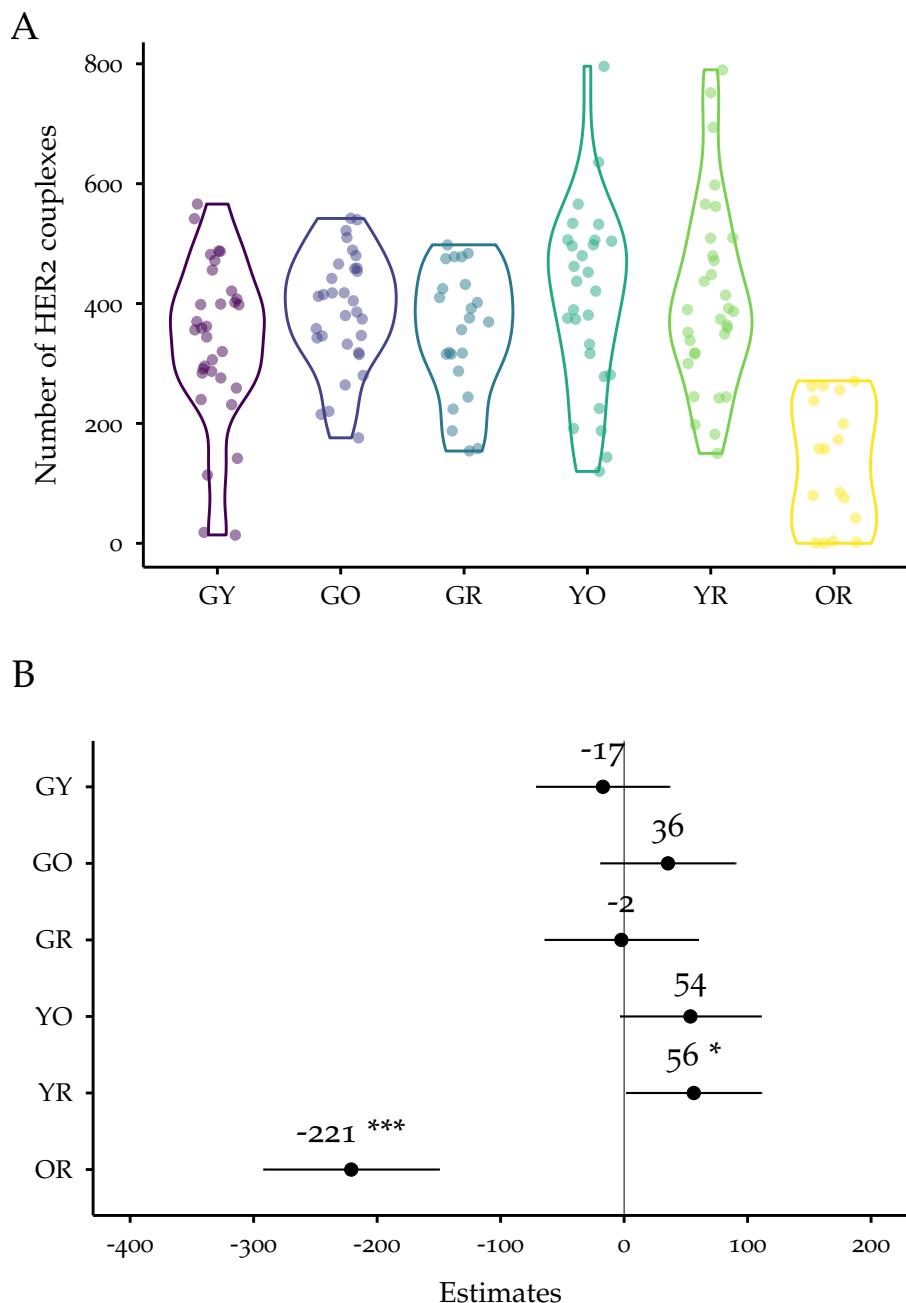


Figure 6.19: Detection of recombinant HER₂ complexes using TTZ and PTZ with distinct DNA-labels for detection in all possible combinations of fluorescent channels (colorpairs). Data was from 2 independent experiments. Data was filtered for $\lambda \leq 0.25$.

A: Violin plots of numbers of HER₂ complexes per colorpair. Sample sizes were $n_{GY} = 32$, $n_{GO} = 31$, $n_{GR} = 23$, $n_{YO} = 28$, $n_{YR} = 31$ and $n_{OR} = 17$. The dots were jittered to avoid overplotting.

B: Forest plot of estimates with 95 % CI from the linear model `lm(calculated_complexes ~ colorpair)` with values from all groups together as reference. The adjusted R^2 for the model was 0.127. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the CI is the associated uncertainty.

The first hypothesis, differences between the LEs (Item 1), can be verified by correcting the obtained number of complexes with the LE. The rationale behind this is that only two successfully labeled antibodies render a formed complex detectable. The approach suggested by Gross et al. (2024) was used to correct the data of Figure 6.19A. However, this increased the differences (Figure C.16). Potential explanations can be that the independently determined LE is erroneous or that the suggested correction is not correct. However, there might be further underlying factors and interactions of these, which make it difficult to judge the root of this difference. The LE-corrected complexes inversely correlate with the LE (Figure C.17). This indicates an overprediction of the number of complexes through an underprediction of the LE and thus challenges this correction. Furthermore, the setup of the ABX (Section 9.3.2) considers already only the labeled fraction of the antibodies. Applying the LE correction therefore results in a double consideration of this factor.

For the verification of the second hypothesis, i.e. background signals (Item 2), the obtained number of complexes were corrected by the number of complexes from the corresponding ABC (Figure 6.20) using the equation suggested by Gross et al. (2024). However, this correction did not reduce the differences to the reference value (data not shown).

Verification of the third hypothesis, spillovers between fluorescent channels (Item 3), requires knowledge about the filter bandwidth of the fluorescent channels of the QIAcuity dPCR system and about excitation and emission spectra of the used dyes (Figure C.14). The bandwidths of the excitation filters fit well to the used dyes, while not all emission spectra of the dyes aligned with the bandwidths of the emission filters. For FAM, HEX and TxRed dyes, the filters fit well to the emission spectra. However, the emission filter for the orange channel, detects a significant proportion of the fluorescence generated by TxRed. The relationship between the fluorescence emissions can be quantified by the ratio of the AUCs for the dyes within the bandwidth of the emission filter, the so-called *spillover coefficient* (Ortolani, 2022). The spillover coefficient of TxRed into the orange channel for Atto550 detection was 75 %, which means the detected fluorescence did not primarily originate from the desired dye. The status of the partitions is no longer unambiguous, which is a violation of correct assignment of a certain to partition to the correct cluster (Item 11). This is related to the fact that the peak of Atto550 is outside the emission filter bandwidth (Figure C.14). This spillover most likely contributed to the low signal of HER2 complexes detected by the colorpair OR (Figure 6.19). The importance of this is based on the fact that PICO requires a signal in two channels and if there is a spillover between these, the consequences are immediate. The combination of fluorescent channels and used dyes should thus be carefully considered. For instance, using NovaFluor Yellow 590 (Thermo Scientific) instead of Atto550 and Alexa Fluor

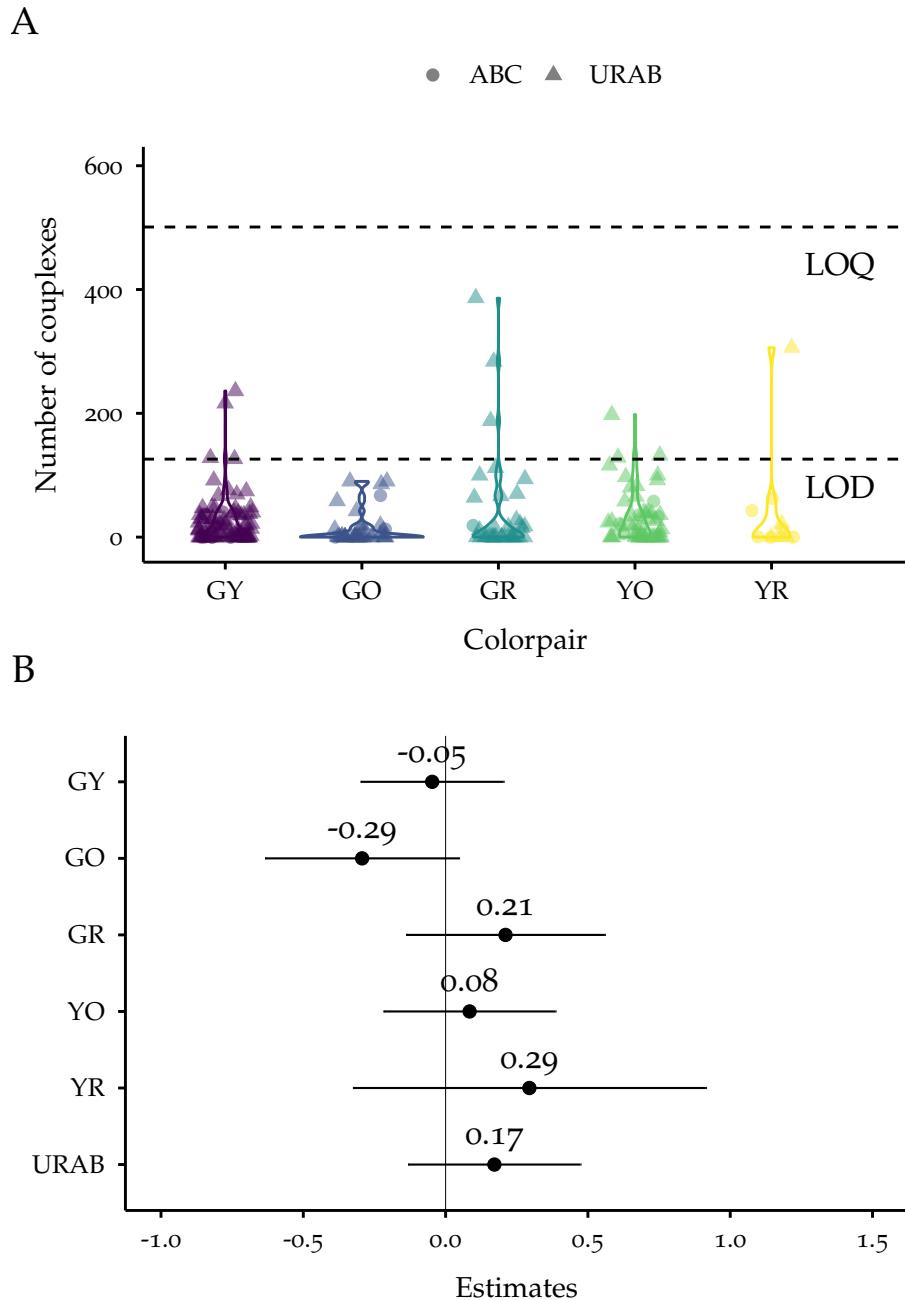


Figure 6.20: ABC PICO experiments with the same antibodies as in Figure 6.19 and results from antibody pairs with URABs. Data is from 2 independent experiments. Data was filtered for $\lambda \leq 0.25$. A: Violin plots of number of complexes detected in ABC (dot) and URAB (triangle) PICO experiments per colorpair. Sample sizes were $n_{GY} = 84$, $n_{GO} = 39$, $n_{GR} = 37$, $n_{YO} = 52$ and $n_{YR} = 11$. The mean values are $\text{mean}_{GY} = 27$ with 95 % CI from 18 to 36, $\text{mean}_{GO} = 14$ with 95 % CI from 6 to 22, $\text{mean}_{GR} = 41$ with 95 % CI from 14 to 68, $\text{mean}_{YO} = 34$ with 95 % CI from 22 to 46 and $\text{mean}_{YR} = 41$ with 95 % CI from -12 to 94. The dots were jittered to avoid overplotting.

B: Forest plot of estimates with 95 % CI from the linear model `lm(calculated_complexes ~ colorpair + experiment_type)` with the colorpair GY as reference. The predictors were standardized using Equation 9.3 (Gelman, 2008). The estimate is the change in the response variable for a 2 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for the model was 0.009.

594 (Thermo Scientific) instead of TxRed would decrease the contribution of unintended fluorescence detection in the orange channel by > 2-fold to 35 % (Figure C.15). The impact on the PICO experiments, however, needs further investigation in follow-up studies. After improvement of the dyes, the remaining spillovers can be compensated by different means (Ortolani, 2022). At the moment, a customizable spillover compensation is not implemented for the QIAcuity dPCR system.

Fourth hypothesis, altered antibody binding affinity because of labeling (Item 4): Although, Vistain et al. (2022) showed that for a similar method of antibody labeling, the antibody binding affinity did not change, the here used antibodies should be characterized similarly to ensure that the observed differences do not originate from impaired binding behavior. This hypothesis represents a violation of the assumption that the labeling process should not change the antibody binding affinity (Item 3). Hypothesis five (Item 5) and six (Item 6) refer to PICO assumptions that all DNA-labels have to be sufficiently amplified (Item 10) and that the assignment of a given partition to a cluster is correct (Item 11). However, they remain to be explored. Because their assessment is complex, such investigations are beyond the scope of this work. A good indicator for high quality data might be the calculation of the peak resolution according to (Lievens et al., 2016).

In general, it is most likely that the observed behavior is a mixture of the described effects. Unraveling these and identifying their contributions is challenging. However, by thorough optimization of the fluorescent dyes, spillover compensations and label amplifications, the full potential of the five detection channels of the QIAcuity comes into reach.

The labeled antibodies used for the detection of recombinant HER2 were used in ABC PICO experiments to access the specificity of the PICO assay (Figure 6.20A). Furthermore, some experiments from above (Figure 6.19) contained URABs¹⁰. The so detected complexes can also be used for the assessment of the specificity. The complexes detected with a URAB were higher but not significantly (Figure 6.20B). In general, the estimated differences between the colorpairs were not significant (Figure 6.20B), while all numbers of complexes were significantly higher than 0 except for the colorpair YR (Figure 6.20). From these values, LOD and LOQ can be calculated using the approach from the European Medicines Agency (2023). Because of the issues for calculating an expectation value, a correlation between expected and detected complexes cannot be calculated. Thus, the slope from Figure 6.18 will be used for the calculation of LOD and LOQ. Across

An antibody pair with an URAB cannot be considered a blank sample in the conventional sense as the target for one antibody is present. However, the detection unit of PICO assay is a complex, which requires two antibodies and a target for both, which is not present.

¹⁰ An URAB is an antibody, which has no affinity to the target. The antibodies used here target the proteins ribosomal protein S6 kinase beta-1 (S6K1) and interleukin 6 receptor or artificial His-tag (Table 10.5) and were also present in the (ABC) PICO experiments shown in Figure 6.19 and Figure 6.20.

all colorpairs and independent of **ABC** or **URAB**, the **LOD** is 165 complexes and the **LOQ** is 501 complexes in 40 μL , which corresponds to a sensitivity between 7 amol L^{-1} and 21 amol L^{-1} .

SUMMARY In this final stage of model validation, the **ddPCS** model was applied for the determination of **HER2** complexes formed with the therapeutic antibodies **TTZ** and **PTZ**. Compared to DNA-only PICO experiments (Section 6.2.2), this stage included the crucial antibody-target binding (Figure 4.3B and Figure 4.4) and was expanded to test all possible combinations of the fluorescence channels for the detection of complexes. With the instruments, fluorescent dyes and settings used in this thesis, the colorpair OR could not be used for complex detection because of fluorescence spillovers. Using the **ddPCS** model for the calculation of complexes from PICO experiments with antibodies, reduced its specificity compared to DNA-only PICO experiments because of background signals in **ABC** PICO experiments. These might be related to the variability introduced by antibody-target binding during the binding reaction. Determining the accuracy was not possible because of lacking expectation values. However, through empirical evidence the impact of fluorescence spillovers could be quantified and an improvement for dyes was suggested (Figure C.15). Finally, **LOD** and **LOQ** were determined and still show attomolar sensitivity of the PICO technology.

6.3 EVALUATION OF DDPCS MODEL PERFORMANCE

The overall performance of the **ddPCS** model can be evaluated by comparing the number of complex double positive partitions (\hat{n}_{AB^+} , i.e. **ddPCS** result) to the total number of observed double positive partitions ($\hat{n}_{D_{\text{obs}}}$, i.e. raw data) and to the number of double positive partitions subtracted by the number of double positive partitions from the corresponding **ABC** ($\hat{n}_{D_{\text{obs}}} - \hat{n}_{D_{\text{obs}}, \text{ABC}}$, i.e. educated guess) using data from Section 6.1.1 (simulated data), Section 6.2.3 (recombinant protein) and Section 7.3 (4EBP1 PICO assay). The rationale behind subtracting $\hat{n}_{D_{\text{obs}}, \text{ABC}}$ from $\hat{n}_{D_{\text{obs}}}$ is the assumption that the random colocalization of both antibodies in the same partition is constant independent on the presence of targets. The hypothesis is that the presence of the target increases $\hat{n}_{D_{\text{obs}}}$ by its count and thus subtracting the blank (**ABC**) results in the number of complexes. The simulation dataset (Section 6.1.1) furthermore allows for the comparison to the ground truth of complex double positive partitions (Figure 6.6).

The number of complexes double positive partitions calculated by the **ddPCS** model needs to be significantly different from $\hat{n}_{D_{\text{obs}}}$ and $\hat{n}_{D_{\text{obs}}} - \hat{n}_{D_{\text{obs}}, \text{ABC}}$ to be considered relevant. If the values were too similar, the approaches mentioned above for complex calculation would be equally good and there was no need for the **ddPCS** model. Data was filtered

for $\lambda \leq 0.25$ and for the large differences between observed and calculated double positive partitions. Complexes values were removed, when the difference was larger than 1 % of $\hat{n}_{D_{obs}}$ (Section 4.2.3)¹¹. The latter threshold was not optimized during this thesis but holds potential for improving the dDDPCS model in future investigations. This topic can be addressed together with bifurcation issue (Figure 6.4).

For the used λ range, the determined number of couplex double positive partitions was not significantly different from the ground truth of the simulated data (Figure 6.21, Figure C.11 and Figure 6.7A). Increasing λ from 0.05 to 0.25, increases $\hat{n}_{D_{obs}}$ (Figure 6.6 and Figure C.11) because of an increased probability of colocalization of both antibodies in the same partition (Section 4.2.2). Attempting to determine the number of couplex positive partitions by subtracting ABC double positive partitions from the observed double positive partitions, works for $\lambda = 0.05$ (Figure C.11), while in general, this approach underpredicts the number of couplex double positive partitions by -148 partitions (95 % CI from -166 partitions to -133 partitions) (Figure 6.21).

For the other datasets, the estimated difference of this approach is in the same magnitude as the difference to the ground truth, but these differences are considered significant (-32 partitions with 95 % CI from -59 partitions to -5 partitions for data from Section 6.2.3 and -30 partitions with 95 % CI from -46 partitions to -14 partitions for data from Section 7.3). This is because within this λ range and the used sample dilutions the contribution of Equation 4.28 to $\hat{n}_{D_{obs}}$ is smaller compared to the contribution of Equation 4.29. However, $\hat{n}_{D_{obs}}$ is never a good approximation for the number of couplex double positive partitions (Figure 6.21, Figure C.11, Figure C.12 and Figure C.13).

Furthermore, the simulations and experiments shown in Chapter 5 and Chapter 6 prove the attomolar sensitivity of the PICO technology. With this analytical performance, the PICO technology belongs to the most sensitive immunoassays (Figure 6.22). PICO has the second best sensitivity among the regarded assays, only Simoa has a better sensitivity (Figure 6.22). Simoa relatives are in the same range, which are all heterogeneous assays. Interestingly, the sensitivity of homogeneous assays like PLA and PEA is lower by orders of magnitude. PICO, however, does not rely on the proximity of the antibodies for signal generation. Thus, it can leverage the full potential of dPCR for detection of proteoforms and PPIs. With the here discussed hypotheses, the sensitivity of the PICO technology could be further improved.

¹¹ Depending on the dataset, further filtering steps may be applied (colorpair, cell line, ...).

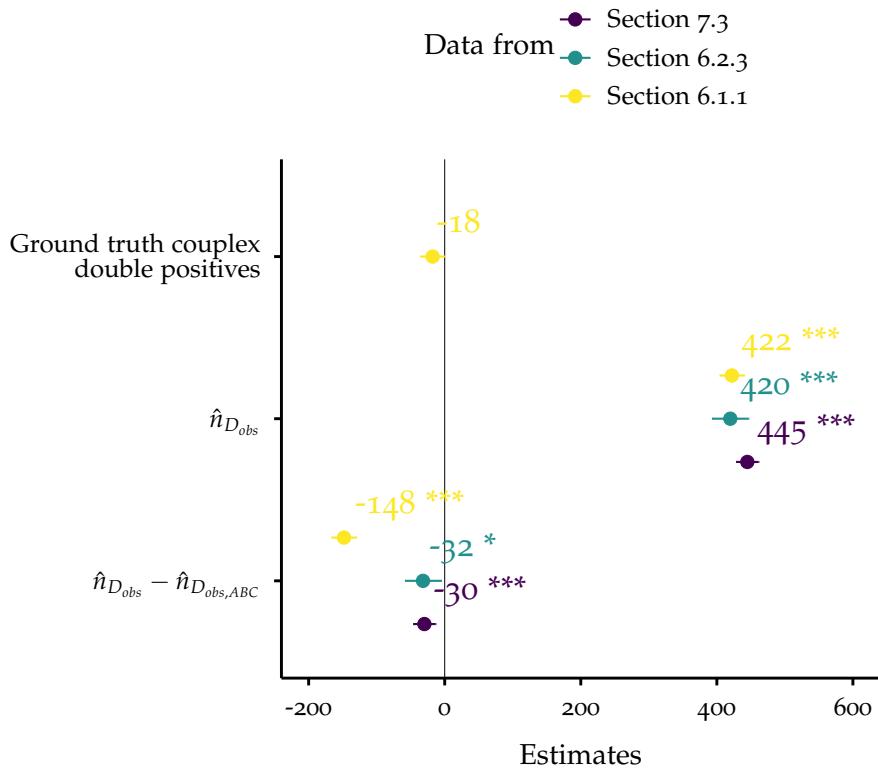


Figure 6.21: Forest plot of estimates between different complexes calculation approaches with 95 % CI as error bar from the linear model `lm(calculated_complexes ~ calculation_approach)` with the complexes from the dDPCS model as reference. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the CI is the associated uncertainty. The model was calculated separately for all datasets. The number of couplex double positive partitions from the dDPCS model was compared to the number of observed double positive partitions from the dPCR and to the observed number of double positive partitions subtracted by the number of observed double positive partitions form the corresponding ABC. The simulation dataset (Section 6.1.1) additionally allows for the comparison to ground truths (Figure 6.6). The sample size of the simulation data set was 2500 (Section 6.1.1), the sample size of the recombinant protein dataset was 360 (Section 6.2.3) and the sample size for the 4EBP1 dataset was 1408 (Section 7.3). The adjusted R^2 for the models were 0.304, 0.563 and 0.493, respectively.

The data of this chapter was used to determine performance parameters (Table 6.2) suggested by the *ICH Q2(R2) Guideline on validation of analytical procedures* from the European Medicines Agency (2023). The primary specificity parameter is highly dependent on the choice

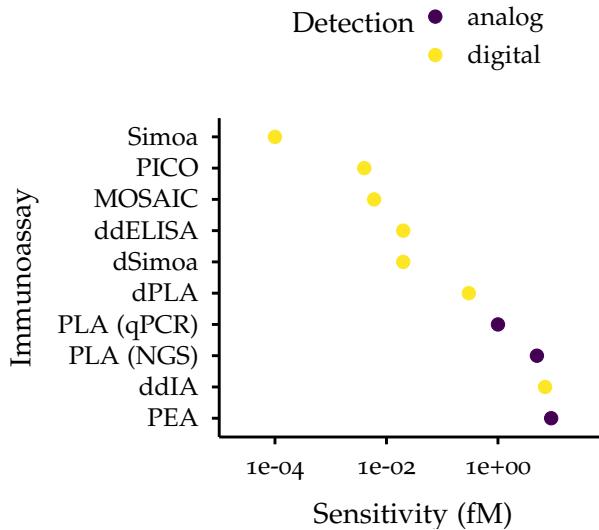


Figure 6.22: Comparison of sensitivities of immunoassays from Table 2.1.

of antibodies and thus needs to be validated for each assay. This can also change LOD, LOQ and intermediate precision¹².

6.4 DESIGN RULES

Based on the data for the model validation of the **dDPCS** model, design rules for the application of the **dDPCS** model to PICO assays can be derived. These rules address the following:

1. Antibodies

- Selectivity/specificity should be validated prior to each PICO assay because assay specificity depends hereon.
- Simultaneous binding should be validated prior to each PICO assay.

2. λ

- Must be kept below $\mu_c = \lambda_{0.55}$ and the recommended range is $\lambda \leq 0.25$ with an optimum at $\lambda = 0.15$.
- The desired λ value can be achieved by dilution prior to dPCR or by increasing the total number of partitions in the dPCR.

¹² LOD and LOQ might also be substantially different for different dPCR systems. Mainly, the number of partitions plays a crucial role in this context. Therefore, it should be pointed out that the values provided for LOD and LOQ are only valid for systems with $\sim 25\,000$ partitions. For lower or higher partition numbers, the analytical performance might be inferior or superior.

- A low value of λ in general reduces several sources of error like PSC or clustering biases and improves the accuracy of the dDPCS model.

3. dPCR

- A dPCR system with chambers instead of droplets is preferred to achieve a stable partitioning and a consistent number of partitions.
- A high number of total partitions is preferred.
- Adapting the fluorescent dyes to the detection channels of the dPCR system is required for usage of all channels and thus full multiplexing capacity. Furthermore, this reduces unwanted signals that distort the actual signals.

Table 6.2: Validation parameters for analytical methods according to European Medicines Agency (2023) applied to the dDPCS model for the calculation of complexes from 2-dimensional raw dPCR data.

Specificity	Range	Accuracy	LOD ¹	LOQ ¹	Sensitivity (amol L ⁻¹)	Intermediate precision
Simulation (Section 6.1)	Given until $\lambda_c = 0.55$ complexes considering 25 000	$0 < \lambda \leq 0.25$, maximal 6250 complexes	92 % to 98 % ³	62 complexes to 68 complexes ⁴	187 complexes to 206 complexes ⁴	n/a
DNA-only (Section 6.2.2) ⁵	100 %, no significant background signal	n/a	95 %	32 complexes ⁶	96 complexes ⁶	n/a
Recombinant protein (Section 6.2.3)	ABC signal significantly higher than 0, but primarily lower than 50 complexes	n/a	n/a	165 complexes ⁷	501 complexes ⁷	7

¹ LOD and LOQ were calculated using the approach by the European Medicines Agency (2023) and represent the number of molecules in 40 µL dPCR master mix. The number of complexes is not necessarily equal to the number of targets in the sample (Gross et al., 2024).

² Increasing the total number of partitions through the usage of other dPCR systems, the maximal number of complexes can be increased.
³ The accuracy depends on λ . These are the values for the minimal slope at $\lambda = 0.05$ and the maximal slope at $\lambda = 0.25$ in the given range (Figure 6.7).

⁴ LOD and LOQ were calculated using the minimal slope at $\lambda = 0.05$ and the maximal slope at $\lambda = 0.25$ in the given range (Figure 6.7).
⁵ The λ range for these experiments was $\lambda < 0.10$.

⁶ LOD and LOQ were lower in DNA-only PICO experiments, because the λ range was lower than in simulation experiments (see comments before).
⁷ These values were calculated using the equation from the European Medicines Agency (2023) with the slope from Figure 6.18.

APPLICATION EXAMPLE: BIOMARKER IN LEUKEMIA

Within this chapter, the validated dDPCS model will be applied for the determination of complexes from a PICO assay for the detection of *4EBP1* proteoforms. The experiments were designed according to Section 6.4. The aim of this chapter was to exploit the ability of the PICO technology for simultaneous detection of proteoforms in a multiplex *4EBP1* PICO assay to gain insights into the suitability of *4EBP1* as a potential predictive biomarker (FDA-NIH Biomarker Working Group, 2016) for HMA treatment strategies in AML.

7.1 BACKGROUND

Leukemia, lymphoma and myeloma are the main hematological malignancies (blood cancers). Leukemia is a collective term for life-threatening and aggressive proliferative disorders of the hematopoietic system (National Cancer Institute, 2023). Depending on the proliferation rate, leukemia can be distinguished into acute (rapid disease progression) and chronic (slow disease progression) and together with the affected cell lineage, leukemia can be categorized as (Arber et al., 2016; Döhner et al., 2017; Juliusson & Hough, 2016):

- AML;
- acute lymphoid leukemia;
- chronic myeloid leukemia, and
- chronic lymphoid leukemia.

In general, leukemia is characterized by an accumulation of immature cells in the bone marrow, which impair physiological functions. In AML, the myeloblast is the malignant cell. While during normal hematopoiesis it differentiates into granulocytes or monocytes, it proliferates uncontrolled in AML (Hoffbrand & Moss, 2016).

ACUTE MYELOID LEUKEMIA The incidence rate for AML is ~ 4.1 per 100 000 individuals representing 1 % of all new cancer cases in the USA (National Cancer Institute, 2023). The 5-year overall survival rate was ~ 32 % in the period from 2013 to 2019 (National Cancer Institute, 2023), while already ~ 84 % of the patients die within the first year after diagnosis (Meyers et al., 2013). The median age of disease onset is 69 years (National Cancer Institute, 2023).

Methylation is the most important epigenetic regulation. In cancer, oncogenes become hypomethylated and are thus activated. Tumor suppressor genes become hypermethylated and are thus silenced (Shames et al., 2007).

In contrast to other cancers, the number of mutations needed to trigger AML is low (Australian Pancreatic Cancer Genome Initiative et al., 2013), but key mutations are observed in ~ 97 % of all AML patients (Patel et al., 2012). However, the mutation rate of cells from the hematopoietic system is low: one protein-coding mutation per ten years per cell (Jaiswal & Ebert, 2019). A significant subset of the mutations affect genes involved in epigenetic regulation (Döhner et al., 2015, 2017). Furthermore, it was shown that the expression of transcription factors for the differentiation of myeloblasts is reduced by hypermethylation (Agrawal et al., 2007). RNA and especially DNA biomarkers (Section 1.2.1) are thus surrogate markers (Temple, 1999), and making conjecture on treatment success is accompanied by a high degree of uncertainty. Epigenetic regulation such as methylation is a reversible process and thus of high interest for therapeutic intervention. For instance, it offers the possibility of reprogramming tumor cells (Heerboth et al., 2014). Moreover, the impact of epigenetic regulation can be observed on the protein level (Gibney & Nolan, 2010).

TREATMENT WITH HYPOMETHYLATING AGENTS In general, treatment possibilities range from hematopoietic stem-cell transplantation, antibody-drug conjugates, tyrosine kinase inhibitors, HMAs to combinations of each, depending on age, genetic predisposition and comorbidities of each patient (Al-Ali et al., 2014; Ferrara & Musto, 2011; Luskin et al., 2016). Hematopoietic stem-cell transplantation, being the only curative therapy, is limited to a small subset of patients because it is an intensive therapy and not appropriate for the majority of AML patients because of the high age of disease onset (Al-Ali et al., 2014; Döhner et al., 2015; National Cancer Institute, 2023). Thus, the use of HMAs is being evaluated for treatment of elderly patients (Al-Ali et al., 2014; Ferrara & Musto, 2011; Lübbert et al., 2023).

The rationale behind HMA treatment is the reactivation of silenced tumor suppressors (Daskalakis et al., 2002; Szmigelska-Kaplon & Robak, 2011). Currently used HMAs are decitabine and azacitidine, which are cytosine analogs that incorporate into the DNA during replication (Stresemann & Lyko, 2008). Once incorporated into the DNA, HMAs trap DNA methyltransferases to the DNA, which are subsequently degraded resulting in a genome-wide hypomethylation (Jüttermann et al., 1994). However, frequent treatment resistance impairs application of this approach (Stomper et al., 2021) and to date there is no suitable predictive biomarker.

EUKARYOTIC TRANSLATION INITIATION FACTOR 4E-BINDING PROTEIN 1 (4EBP1) The PI3K-Akt-mTOR pathway, which is constitutively active in ~ 60 % of the AML patients (Nepstad et al., 2020; Xu et al., 2005), is involved in tumor proliferation (Zou et al., 2020) by mediating protein synthesis. In this regard, primary phosphorylation

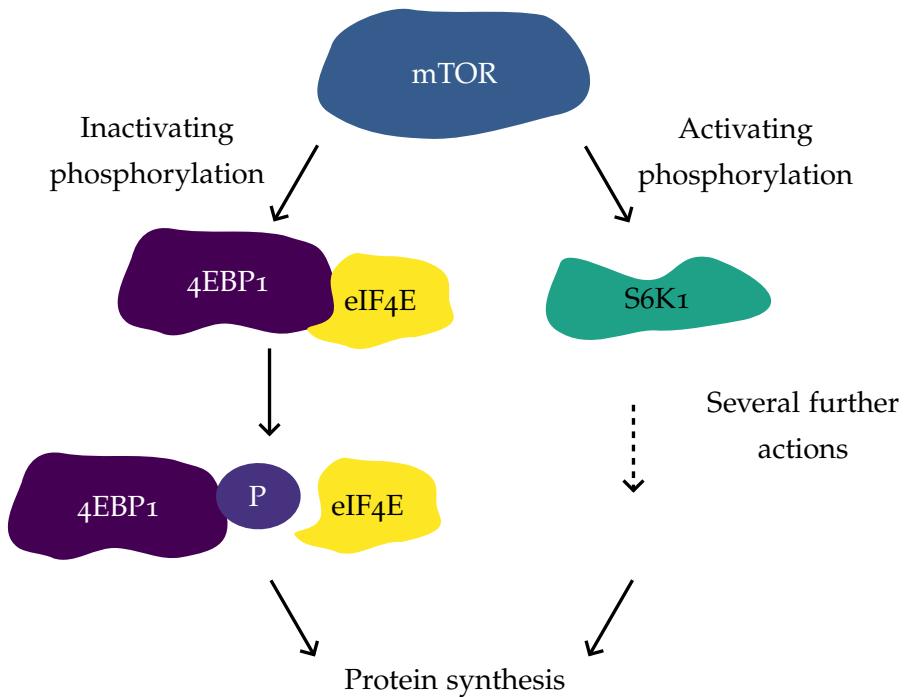


Figure 7.1: Primary downstream targets of mechanistic target of rapamycin (mTOR) for protein synthesis. After activation of mTOR, it phosphorylates its downstream targets: 4EBP1 and S6K1. S6K1 becomes activated and in turn phosphorylates further proteins to enhance mRNA translation and thus protein synthesis. 4EBP1 is inactivated upon phosphorylation and releases eIF4E, which can then form the translation initiation complex for 5'-cap-dependent mRNA translation. Figure inspired by Saxton and Sabatini (2017) and Qin et al. (2016).

targets of mTOR are 4EBP1 and S6K1 (Saxton & Sabatini, 2017). Upon phosphorylation by mTOR, S6K1 is activated and phosphorylates further downstream targets, which eventually leads to enhanced translation efficiency of mRNAs (Figure 7.1). On the other hand, mTOR phosphorylates 4EBP1 on T37 and T46 as a priming event, which leads to further phosphorylations, eventually resulting in the release of sequestered eIF4E (Gingras et al., 1999; Gingras et al., 2001; Qin et al., 2016). Free eIF4E then assembles with further proteins to form the translation initiation complex, which enables 5' cap-dependent mRNA translation (Saxton & Sabatini, 2017). The phosphorylations on T37 and T46 are key drivers for blocking the sequestration of eIF4E (Livingstone & Bidinosti, 2012; Livingstone et al., 2009). Interestingly, it was shown that 4EBP1 is hyperphosphorylated in AML (Xu et al., 2005) and that 4EBP1 is involved in mediating drug resistance in other cancers such as prostate or colorectal cancer (Hsieh et al., 2015; Zhang & Zheng, 2012).

4EBP1 represents a suitable target for the development of a PICO assay because it plays a crucial role in cancer and the degree of phos-

phorylation may provide valuable insights into tumor cell proliferation and survival after HMA treatment. Furthermore, it can be used to demonstrate the versatility of the PICO technology in a multiplex PICO assay (Figure 4.2B), because *4EBP1* proteoforms as well as its interaction with *eIF4E* can be analyzed¹.

7.2 PICO ASSAY DEVELOPMENT

The development of a PICO assay for the detection of *4EBP1* comprises several steps. As the PICO technology requires two antibodies for the detection of a protein, their compatibility needs to be ensured. The largest unknown in the development is the specificity of the antibodies, which directly influences the specificity of the PICO assay (Section 6.3). To account for this, the antibodies are tested on recombinant *4EBP1* in parallel. This opens up the possibility to understand unspecific antibody binding that originates from matrix effects or assay design (Selby, 1999; Wood, 1991).

7.2.1 Validation of antibodies against *4EBP1*

The therapeutic antibodies *TTZ* and *PTZ* are well-characterized and known to simultaneously bind *HER2* (Nahta et al., 2004; Nami et al., 2018; Sharma et al., 2021), which is a prognostic, predictive and drugable biomarker for breast cancer (Iqbal & Iqbal, 2014). However, for *4EBP1* there are no therapeutic antibodies available to date. Several research antibodies were thus purchased and tested in different combinations for the detection of recombinant *4EBP1*² and cellular *4EBP1* from U937 cells (AML cell line). Simultaneous and specific binding needs to be validated for each antibody pair.

A pitfall during the development of PICO assays using recombinant protein is the assumption that the antibody-target binding is the same when using the cellular protein. Recombinant proteins may sometimes represent only a fraction of the full-length protein, may have a different conformation or may be modified with a protein tag, which can change its physicochemical properties (Bernier et al., 2018). The here used recombinant *4EBP1* has full length but contains an N-terminal His-tag, which was used for purification after production in *E. coli*. His-tags may – depending on the protein – negatively impact the solubility of the protein (Woestenenk et al., 2004). This can affect antibody-target binding.

¹ The PICO assay for the detection of *eIF4E* complexes was not (yet) successful (Figure D.2) and needs further investigation.

² Prior to PICO experiments, the recombinant *4EBP1* was characterized using Agilent's Bioanalyzer (Section 9.2.4) and showed a single peak at the expected size (Figure D.1).

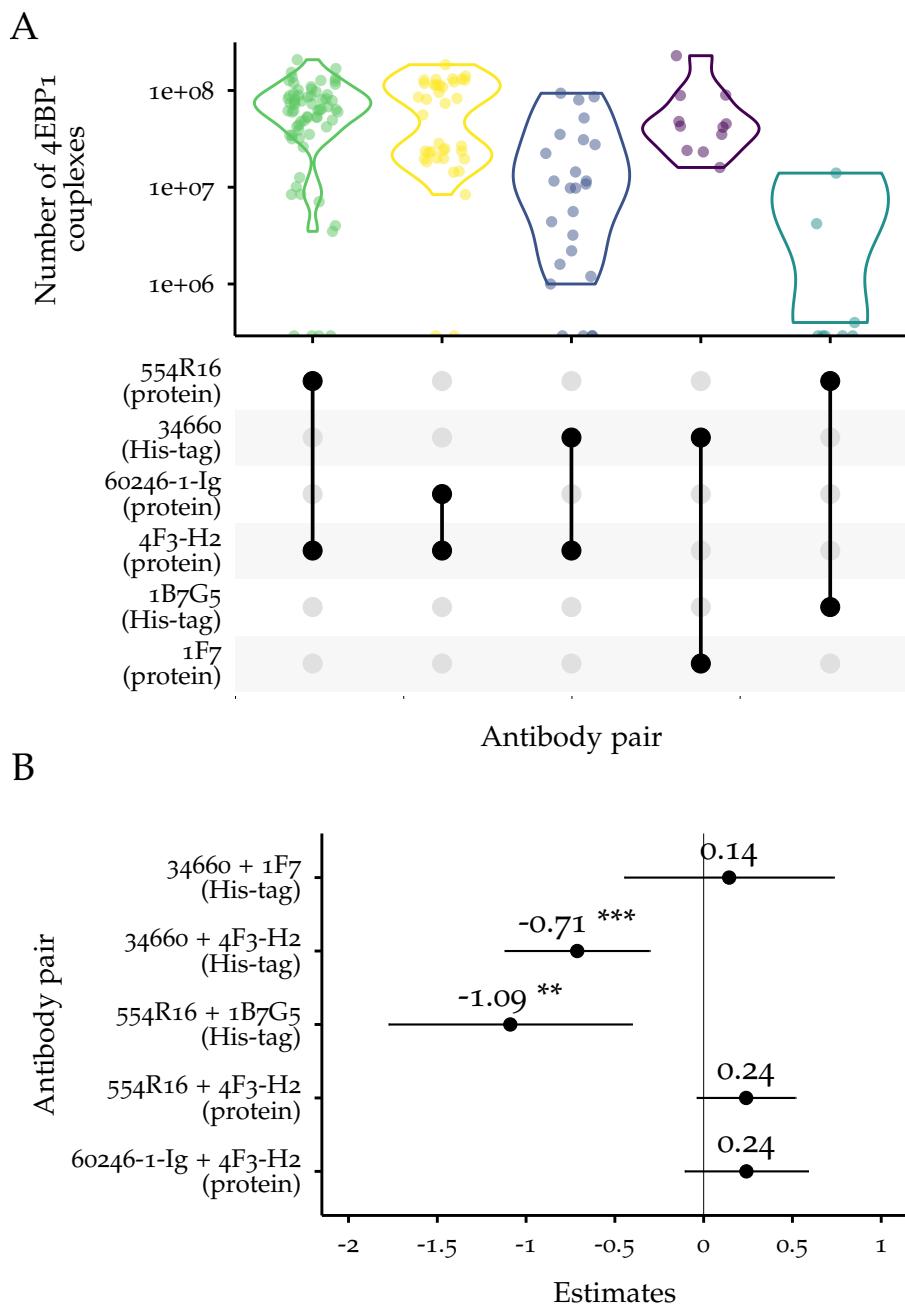


Figure 7.2: Comparison of antibody pairs for the detection of recombinant 4EBP₁ complexes. Data is from 3 independent experiments.

A: UpSet plot of the number of recombinant 4EBP₁ complexes per reaction for different antibody pairs. Clones 34660 and 1B7G5 are anti-His-tag antibodies. Sample sizes were: 67, 37, 25, 11, 8 (from left to right). Low sample sizes do not necessarily mean that experiments were not thoroughly repeated but that this plot depicts already filtered results. The dots were jittered to avoid overplotting. The y-axis was log₁₀-transformed for better visibility.

B: Forest plot of estimates with 95 %CI from the linear model `lm (calculated_complexes ~ antibody_pair)` with antibody pair 60246-1-Ig + 4F3-H2 as reference. The predictors were standardized (Equation 9.2). The adjusted R² for the model was 0.081. The estimate is the change in the response variable for a 1 SD increase in the predictor variable and the CI is the associated uncertainty.

The chosen monoclonal antibodies were raised in mice using full-length recombinant *4EBP1* from *E. coli* (Table 10.4). The sequence of these antibodies and their epitopes is unknown, which is a common issue with the majority of research antibodies (Weller, 2016). Thus, two of these antibodies have a certain chance of sharing (partially) equal epitopes. *4EBP1* has a length of 118 amino acids, considering the average length of epitopes 7 to 9 amino acids (Buus et al., 2012), this results in 13 to 16 continuous, non-overlapping epitopes. However, proteins usually have a tertiary structure and most epitopes are discontinuous (Barlow et al., 1986). Steric hindrance might also prevent simultaneous binding, even when the epitopes are generally accessible. Additionally, without knowledge of antibody sequence and/or paratope and epitope sequences, machine learning approaches (Bryant et al., 2022; Huang et al., 2022; Ye et al., 2022) are not feasible. The remaining strategy is trial-and-error for the identification of a suitable antibody pair. Of note, the recombinant *4EBP1* can be detected using His-tag antibodies (Table 10.5).

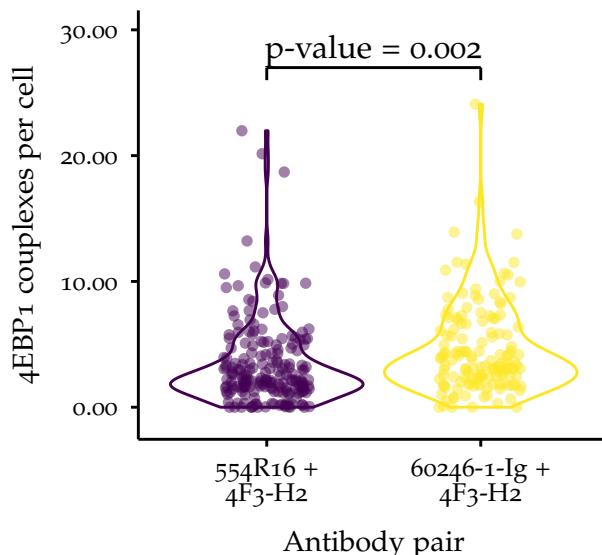


Figure 7.3: Comparison of antibody pairs for the detection of cellular *4EBP1* complexes. The number of complexes from the antibody pairs was compared using the linear model `lm(calculated_complexes_per_cell ~ antibody_pair)`. The antibody pair 60246-1-Ig + 4F3-H2 detected 1.17 more complexes per cell (95 % CI from 0.43 to 1.92). The adjusted R^2 for the model was 0.025. Sample sizes were 196 and 143 (from left to right). Data was from at least 2 independent experiments. The dots were jittered to avoid overplotting.

Four monoclonal antibodies against *4EBP1* (clones 1F7, 4F3-H2, 554R16 and 60246-1-Ig (Table 10.4)) and two monoclonal antibodies against the His-tag (clones 34660 and 1B7G5 (Table 10.5)) were chosen and labeled in such a way that multiple antibody combinations can be tested (Table 10.6). The antibody pairs 554R16 + 4F3-H2 and 60246-1-Ig

+ 4F3-H2 yielded the highest number of recombinant *4EBP1* complexes (Figure 7.2A) and were not different from the reference³ (standardized estimates 0.24 with 95 % CI from –0.04 to 0.52 and 0.24 with 95 % CI from –0.11 to 0.59, respectively). However, 60246-1-Ig + 4F3-H2 detected more *4EBP1* complexes from U937 cell lysate than 554R16 + 4F3-H2 (Figure 7.3) (estimate 1.17 complexes per cell (95 % CI from 0.43 to 1.92)). Also, the antibody pair 34660 + 1F7 was not significantly different from the reference group (Figure 7.2) demonstrating the detection of recombinant *4EBP1* complexes via its His-tag (standardized estimate 0.14 with 95 % CI from –0.45 to 0.73). The antibody pair 34660 + 4F3-H2 resulted in a significantly lower number of recombinant *4EBP1* complexes (Figure 7.2A). This underlines that the specificity and also the accuracy of a PICO assay depends on the choice of antibodies. 554R16 + 1B7G5 detected no recombinant *4EBP1* complexes (Figure 7.3A) and shows the strongest deviation from the reference (standardized estimate –1.09 with 95 % CI from –1.77 to –0.40). These findings imply that key to establishing a successful PICO assay is choosing the right combination of antibodies as they may only work in certain combinations. Identification of antibody compatibility *a priori* is desirable because of better resource efficiency.

The most likely explanation for non-performing antibodies is that they cannot bind simultaneously. Yet these antibodies are monoclonal, but the full-length recombinant protein was used as immunogen for antibody elicitation (Table 10.4). Therefore, the possibility of overlapping epitopes is not excluded. This hypothesis could be further validated by, for instance, competitive binding experiments and epitope binning using FACS (Chan et al., 2018).

As explained in Section 6.2.3, the calculation of an expectation value for these experiments requires usage of the *ternary-equilibrium-based absolute quantification model (AQ model)* form Gross et al. (2024) and unlike for TTZ and PTZ, the K_d values for the here used antibodies are unknown. Their determination by surface plasmon resonance spectroscopy (Bakhtiar, 2013) lies beyond the scope of this thesis.

Therefore, the antibody pair with the highest signal is considered the preferred pair under assumed saturating antibody conditions (Section 9.3.2 and Section 9.3.3) (Gross et al., 2024). Additionally, the antibody pairs 554R16 + 4F3-H2 and 554R16 + 1F7 showed the highest signal in the ABC PICO experiments (Figure D.3), which might indicate some unspecificity of the antibodies. Also, the antibody pairs employing 34460 resulted in elevated numbers of complexes in ABC PICO experiments, while the antibody pair 60246-1-Ig + 4F3-H2 detected a low number of complexes indicating low unspecificity (Figure D.3). Thus, this antibody pair seems to be reliable and was used for the detection of *4EBP1* complexes in the following.

³ The numbers of complexes from all groups were used as reference group.

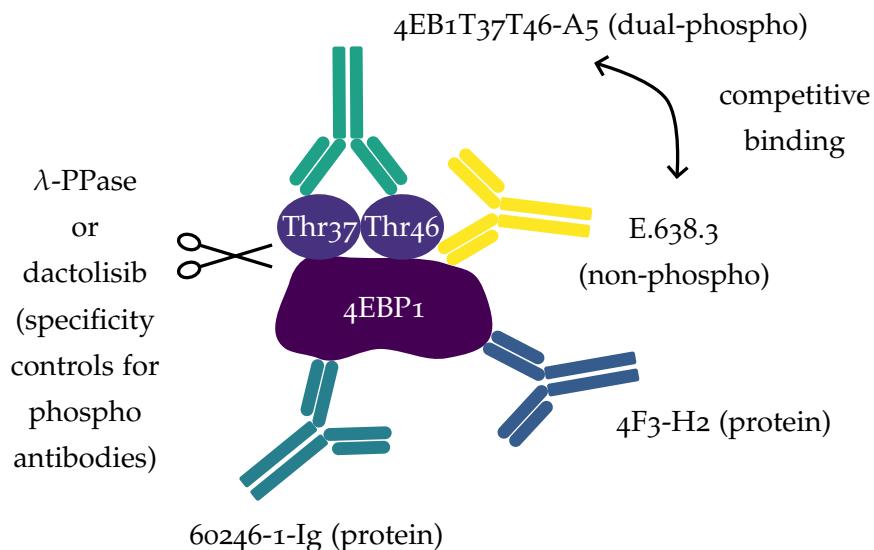


Figure 7.4: PICO assay for the detection of *4EBP1* using antibodies *4EB1T37T46-A5*, *E.638.3*, *60246-1-Ig* and *4F3-H2*. The antibodies *4EB1T37T46-A5* and *E.638.3* remain to be validated within Section 7.2.2. The antibody pairs *4EB1T37T46-A5* + *4F3-H2* and *4EB1T37T46-A5* + *60246-1-Ig* both detect the T37T46-phosphorylated *4EBP1* and can thus be considered as redundant measurements. Ideally, they result in the same number of complexes.

7.2.2 Validation of antibodies against phosphorylated *4EBP1*

The best antibody pair under the tested conditions for the detection of *4EBP1* complexes, *60246-1-Ig* + *4F3-H2*, was completed with two further antibodies targeting the phosphorylation site and the same site but non-phosphorylated. This allows for the discrimination of phosphorylated and non-phosphorylated *4EBP1*. The first antibody was raised in rabbit against a synthetic T37T46-phosphorylated *4EBP1* peptide (clone *4EB1T37T46-A5*) and the second antibody was raised in rabbit against a synthetic non-T46-phosphorylated *4EBP1* peptide (clone *E.638.3*) (Table 10.4 and Figure 7.4). These antibodies compete (almost) for the same epitope and cannot bind simultaneously, although they are not perfect competitors as *E.638.3* only detects the non-T46-phosphorylated residue, while *4EB1T37T46-A5* requires the phosphorylation of both residues, T37 and T46. However, there is strong evidence that, if phosphorylated, *4EBP1* is always phosphorylated on both sites, T37 and T46 (Gingras et al., 1999; Qin et al., 2016). This allows for deciphering the degree of *4EBP1* T37T46-phosphorylation or non-T46-phosphorylation. Eventually, conjectures can be made on the strength of protein synthesis through mTOR signaling (Section 7.1 and Figure 7.1). For the validation of the specificity of anti-phospho antibodies, a controlled dephosphorylation is necessary (Mandell, 2003). In this section, two approaches for dephos-

phorylation were applied: enzymatic using λ -protein phosphatase (PPase) (Zhuo et al., 1993) and mechanistic using the **4EBP1** kinase inhibitor dactolisib (Shi et al., 2018).

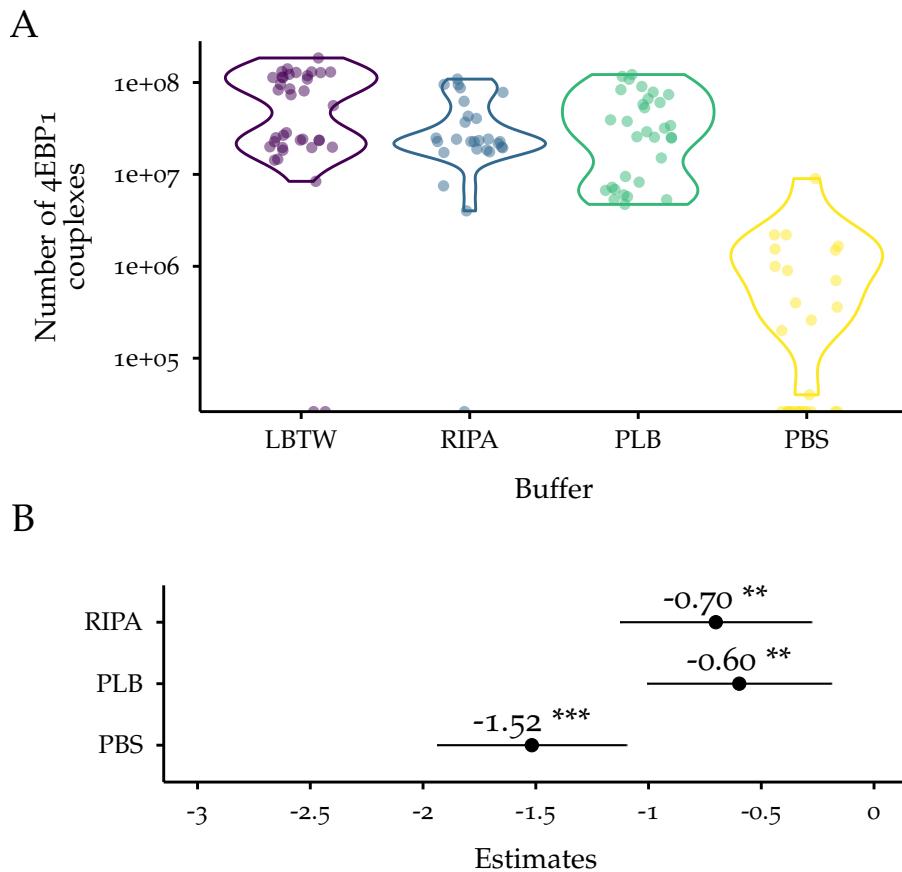


Figure 7.5: Comparison of buffers for the detection of recombinant **4EBP1** using the antibody pair 60246-1-Ig + 4F3-H2. Data was from at least 2 independent experiments.

A: Violin plots of couples for lysis buffer (LB), radioimmunoprecipitation assay (RIPA), passive lysis buffer (PLB) and PBS buffers. Sample sizes were 37, 27, 31, 28 (from left to right). The dots were jittered to avoid overplotting. The y-axis was \log_{10} -transformed for better visibility.

B: Forest plot of estimates with 95 % CI from the linear model `lm(calculated_couples ~ buffer)` with LB as reference. The predictors were standardized (Equation 9.2). The adjusted R^2 for the model was 0.285. The estimate is the change in the response variable for a 1 SD increase in the predictor variable and the CI is the associated uncertainty.

ENZYMATIC DEPHOSPHORYLATION Recombinant and cellular **4EBP1** can be dephosphorylated enzymatically by using λ -PPase, which cleaves all phosphate groups from serine (S), threonine (T) and tyrosine (Y) residues of proteins (Zhuo et al., 1993). This treatment requires a buffer without phosphatase inhibitors such as vanadate

(Gordon, 1991). In general, the choice of the buffer is crucial for any immunoassays as it can influence the binding affinity of the antibody to the antigen (Petzold, 2019; Selby, 1999; Wood, 1991). The RIPA buffer (Table 9.1 and Table 9.2) and the PLB⁴ were compared to the LB from Actome (Actome GmbH, 2024) and to PBS (negative control)⁵. Recombinant 4EBP1 was dissolved in LB, RIPA buffer, PLB. Also, the ABXs containing the antibody pair 60246-1-Ig + 4F3-H2 were prepared in the corresponding buffers.

All tested buffers detected a significantly lower number of recombinant 4EBP1 complexes compared to the LB (Figure 7.5B). The number of complexes with PBS was the lowest (standardized estimate -1.52 with 95 % CI from -1.94 to -1.10), while RIPA and PLB were similarly different (standardized estimate -0.70 with 95 % CI from -1.13 to -0.28 and -0.60 with 95 % CI from -1.01 to -0.19 , respectively) (Figure 7.5B). In addition, the number of complexes form ABC PICO experiments delivers information about the antibody specificity. Interestingly, the number of complexes with PLB was significantly higher compared to LB, while PBS and RIPA buffer were not significantly different (Figure D.5) (standardize estimate 1.63 with 95 % CI from 1.10 to 2.16). This indicates that PLB did not establish the buffer conditions required for specific antibody binding in PICO, i.e. ensuring effective antibody-target binding and suppressing unspecific bindings. Thus, RIPA buffer was used to replace LB in experiments with λ -PPase. Of note, SDS was added to the RIPA buffer after λ -PPase treatment (Section 9.1.7) because it inhibits the activity of phosphatases (Hou et al., 2020).

Additionally, the capability of the RIPA buffer for cell lysis needed to be compared to LB to validate its suitability for λ -PPase treatments of cell lysates. Using the RIPA buffer, the number of 4EBP1 complexes per cell changed by -0.60 (95 % CI from -1.50 to 0.30) compared to LB, which is not considered significant (Figure 7.6). RIPA can thus be used to replace LB for cell lysis in experiments employing λ -PPase. After these preparations, recombinant 4EBP1 and 4EBP1 from U937 cell lysate was analyzed using the PICO assay shown in Figure 7.4.

Recombinant 4EBP1 was treated with λ -PPase or H₂O for mock (Section 9.1.7). Other than that, the workflows were the same. Strikingly, λ -PPase treatment increased the number of 4EBP1 complexes significantly (standardized estimate 0.69 with 95 % CI from 0.26 to 1.13) (Figure 7.7). It is possible, yet unlikely, that recombinant 4EBP1 produced in *E. coli* is phosphorylated (Schastnaya et al., 2021) and if at different sites than the canonical phosphorylations. The rationale behind this is that upon λ -PPase treatment recombinant 4EBP1 under-

⁴ This buffer was used for the PICO assay shown in Karakus et al. (2019). PICO was termed Emulsion Coupling in this publication.

⁵ Tris-based buffers with and without NaCl, which are commonly used in molecular biology were tested, too, but did not show the performance needed, i.e. similar complex numbers as PBS (data not shown).

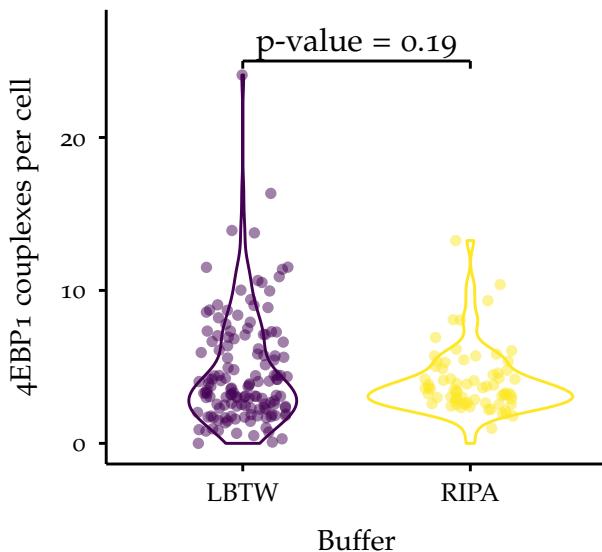


Figure 7.6: Comparison of buffers for the detection of native *4EBP1* from cell lysate of U937 cells using the antibody pair 60246-1-Ig + 4F3-H2. The linear model `lm(calculated_complexes ~ buffer)` estimated a difference of -0.60 complexes per cell with 95 % CI from -1.50 to 0.30 . The adjusted R^2 for the model is 0.025 . Sample sizes are from left to right: 143, 72. Data is at least from 4 independent experiments. The dots were jittered to avoid overplotting.

goes a conformational change, which enhances antibody binding and thus increases the number of complexes. Compared to the reference antibody pair 60246-1-Ig + 4F3-H2 (detection of the *4EBP1* protein), the antibody pairs 4EB1T37T46-A5 + 60246-1-Ig and 4EB1T37T46-A5 + 4F3-H2 for the detection of the T37T46 phosphorylation (Figure 7.4), detected significantly lower numbers of complexes (standardized estimates -1.14 with 95 % CI from -2.05 to -0.23 and -1.12 with 95 % CI from -1.85 to -0.39 , respectively) independent of λ -PPase treatment (Figure 7.7B). The antibody pair 60246-1-Ig + E.638.3 detecting only non-phosphorylated *4EBP1* was significantly lower than the reference (standardized estimate -1.13 with 95 % CI from -2.23 to -0.04 ⁶). Together with the absence of the phosphorylations, this suggests that the E.638.3 antibody did not work as intended with the antibody 60246-1-Ig, which in turn worked with 4F3-H2 for the detection of recombinant *4EBP1* (Figure 7.7A). Finally, the competitive antibody pair 4EB1T37T46-A5 + E.638.3 detected similar numbers of recombinant *4EBP1* complexes to the phosphorylation measurements (Figure 7.7A). However, based on this data, it is difficult to judge if the E.638.3 antibody is functional in PICO assays. These results indicate that the 4EB1T37T46-A5 antibody did not detect non-phosphorylated,

⁶ The antibody pair E.638.3 + 4F3-H2 was detected by the colorpair OR and could thus not be used for data analysis (Section 6.2.3).

recombinant 4EBP_1 or that recombinant 4EBP_1 is not phosphorylated at T37T46. The positive control for antibody specificity of antibodies 4EB1T37T46-A5 and E.638.3 will be validated by applying the 4EBP_1 PICO assay (Figure 7.4) to U937 cell lysate for the detection of cellular 4EBP_1 in the presence of phosphorylations and potential matrix effects. 4EBP_1 is hyperphosphorylated in AML (Xu et al., 2005).

After dephosphorylation of U937 cell lysate with λ -PPase (Section 9.1.7), 4EBP_1 phosphorylation was assessed with the PICO assay (Figure 7.4). Unlike, after dephosphorylation of recombinant 4EBP_1 , λ -PPase treatment did not change the number of detected 4EBP_1 complexes (Figure 7.8) (standardized estimate 0.11 with 95 % CI from -0.16 to 0.37). The antibody pair 4EB1T37T46-A5 + 4F3-H2 detected the same number of T37T46-phosphorylated 4EBP_1 complexes as the 60246-1-Ig + 4F3-H2 antibody pair (standardized estimate 0.19 with 95 % CI from -0.17 to 0.55), indicating that all T37T46 residues on 4EBP_1 are phosphorylated (Figure 7.8). Upon λ -PPase treatment, this fraction was significantly reduced (standardized estimate -1.31 with 95 % CI from -1.82 to -0.80). Interestingly, the 4EB1T37T46-A5 + 60246-1-Ig detected a lower degree of 4EBP_1 phosphorylation than the other antibody pair detecting the phosphorylation (Figure 7.8A), which also got reduced upon λ -PPase treatment (Figure 7.8B). All three antibodies 4EB1T37T46-A5, 60246-1-Ig and 4F3-H2 were available in the same binding reaction, meaning that the biological material for binding was identical. The difference in phosphorylation detection might be a consequence of one antibody's selectivity or affinity for the target. This means that a higher phosphorylation signal than the actual protein should not be overemphasized as it might actually be lower than the protein signal after absolute quantification, which considers the K_d values of the antibodies (Gross et al., 2024). Furthermore, in two ELISA assays for the detection of 4EBP_1 and T37T46-phosphorylated 4EBP_1 , the phosphorylation was significantly higher than the protein (Figure D.4)⁷. Another fact to consider, is that the PPI of 4EBP_1 and eIF4E might impair the phosphorylation measurement. However, detection of eIF4E was not successful (Figure D.2) to make conjectures on this issue. Antibody pairs employing the E.638.3 antibody again did not detect any complexes (Figure 7.8 and Figure 7.7). This is growing evidence that at least with the other antibodies used in this assay (Figure 7.4), the E.638.3 antibody cannot form complexes. As this antibody is generally of high interest as a competitor for 4EB1T37T46-A5, a deeper investigation in future research is desirable.

⁷ The manufacturer did not provide information on the antibodies.

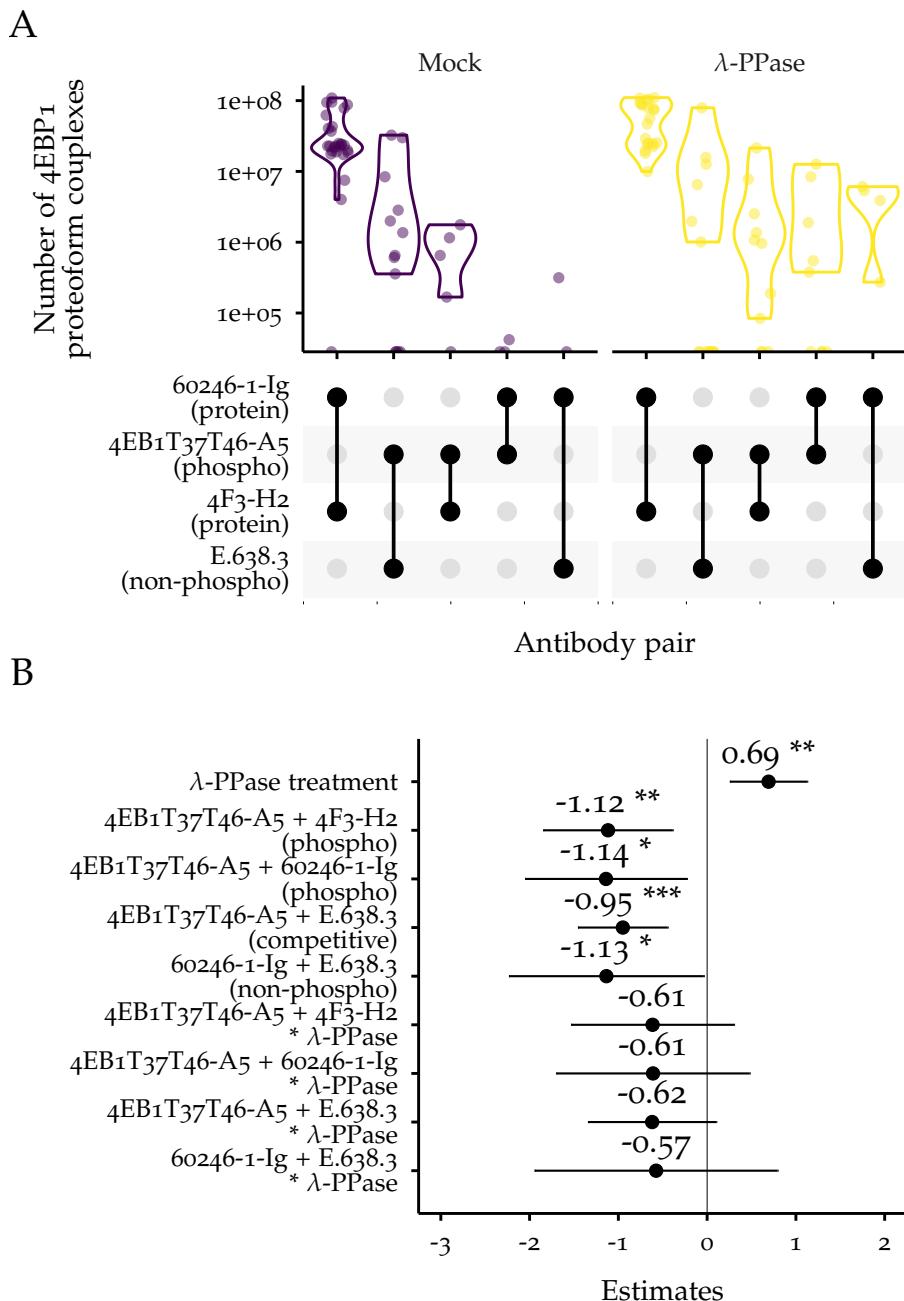


Figure 7.7: Comparison of phosphorylation antibody specificity using recombinant 4EBP₁ before and after λ -PPase treatment with the 4EBP₁ PICO assay shown in Figure 7.4. Data was from at least 2 independent experiments.

A: UpSet plot of the number of couples for mock and λ -PPase treated recombinant 4EBP₁. Sample sizes were 27, 13, 5, 3, 2, 21, 14, 11, 9, and 4 (from left to right). Low sample sizes do not necessarily mean that experiments were not thoroughly repeated but that this plot depicts filtered results. The dots were jittered to avoid overplotting. The y-axis was \log_{10} -transformed for better visibility.

B: Forest plot of estimates with 95 % CI from the linear model `lm` (`calculated_couples ~ antibody_pair * treatment`) with mock treatment and antibody pair 60246-1-Ig + 4F3-H2 as references. The predictors were standardized using Equation 9.3 (Gelman, 2008). The * symbol indicates the interaction of two input variables. The estimate is the change in the response variable for a 2 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for the model was 0.430.

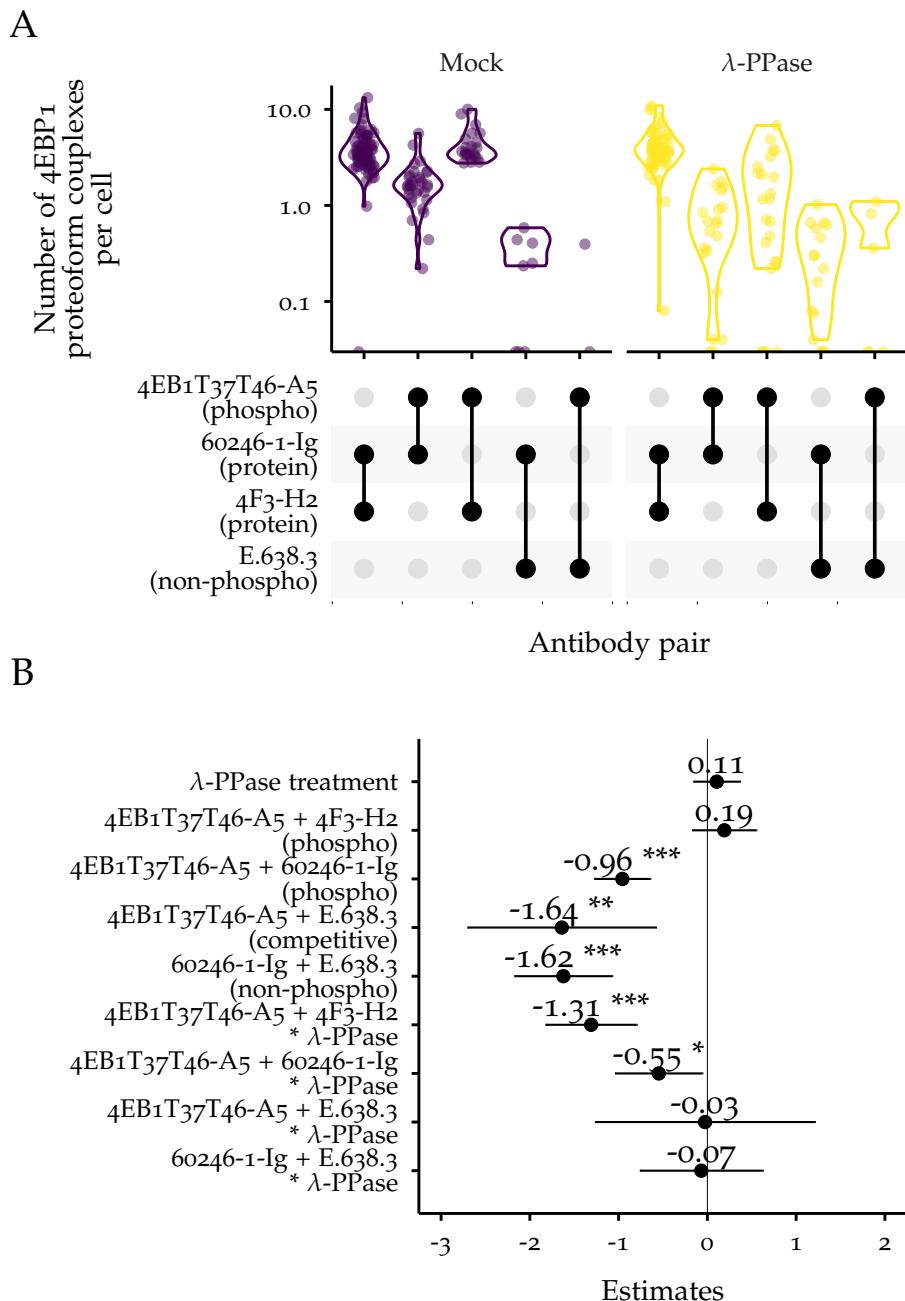


Figure 7.8: Comparison of 4EBP₁ phosphorylation between λ -PPase treated U937 cell lysate and mock treated U937 cell lysate using the 4EBP₁ PICO assay shown in Figure 7.4. Data was from at least 2 independent experiments.

A: UpSet plot of the number of complexes per cell detected in λ -PPase and mock treated cell lysates. Sample size were 72, 32, 22, 8, 2, 58, 21, 23, 16, 6 (from left to right). Low sample sizes do not necessarily mean that experiments were not thoroughly repeated but that this plot depicts already filtered results. The dots were jittered to avoid overplotting. The y-axis was \log_{10} -transformed for better visibility.

B: Forest plot of estimates with 95 % CI from the linear model `lm (calculated_complexes ~ antibody_pair * treatment)` with mock treatment and antibody pair 60246-1-Ig + 4F3-H2 as reference. The predictors were standardized using Equation 9.3 (Gelman, 2008). The * symbol indicates the interaction of two input variables. The estimate is the change in the response variable for a 2 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for the model is 0.435.

In attempt to explain the large variability of 4EBP_1 and 4EBP_1 phosphorylation detection, the biological heterogeneity of cell samples can be adduced. Biological samples will always contain a mixture of 4EBP_1 with varying degrees of phosphorylations because this correlates with the cell-cycle phase (Heesom et al., 2001; Sun et al., 2019). To circumvent this, the cells could be stratified by cell-cycle phase prior to lysis (Ligasová et al., 2023; Pozarowski & Darzynkiewicz, 2004), cells from different phases could be synchronized (RC Wang & Wang, 2022) or the 4EBP_1 assay could be performed on single cells⁸. However, this does not explain the high variability in the detection of recombinant 4EBP_1 , which has homogeneous conformation. Thus, the majority of variability may be attributed to the antibodies themselves. Interestingly, it was found that 4EBP_1 is completely unstructured in free and unbound states (Fletcher & Wagner, 1998), which means that upon binding either by physiological partners or antibodies for detection, 4EBP_1 can change its conformation and thus epitope availability might change. Furthermore, Rong et al. (2008) found that 30 % of the 4EBP_1 is located in the nucleus sequestering eIF4E , while the residual 70 % are located in the cytosol with unknown function and spatial colocalization of 4EBP_1 and eIF4E is inversely proportional to T37T46 dephosphorylation (Livingstone et al., 2009). Based upon this, farther-reaching research questions can be asked:

1. Does the cell lysis efficiently lyse the nucleus?
2. How will the results change after separating the 4EBP_1 proteins from the nucleus and the ones from the cytosol?
3. How to maintain spatial information to understand correlated phosphorylation?
4. What is the degree of 4EBP_1 interacting with eIF4E ?⁹

MECHANISTIC DEPHOSPHORYLATION In addition to the enzymatic and unspecific cleavage of phosphate groups, the kinases responsible for 4EBP_1 phosphorylation can be inhibited by treating cells with a suitable inhibitor prior to lysis. Residues T37 and T46 on 4EBP_1 are primarily phosphorylated by mTOR (Gingras et al., 1999) and potentially further kinases (Qin et al., 2016). Different inhibitors of mTOR and phosphoinositide 3-kinase (PI3K) were compared and tested for efficacy: rapamycin and analogs everolimus and temsirolimus (Zheng & Jiang, 2015), ATP-competitive torin1 (Thoreen et al., 2009), dual PI3K/mTOR inhibitor dactolisib (BEZ235) (Shi et al., 2018) and PI3K inhibitor alpelisib (Wilhoit et al., 2020).

⁸ Preliminary experiments were conducted (data not shown).

⁹ An attempt to answer this question was the design of an eIF4E PICO assay, which was not successful (Figure D.2) as the number of complexes were 3 orders of magnitude lower compared to 4EBP_1 .

4EBP1 migrates in three different bands on the SDS. The α form carries the least number of phosphorylations (hypophosphorylated), while the β and the γ forms are hyperphosphorylated (Gorr & Vogel, 2015; Jiang et al., 2001). The degree of phosphorylation is the reason for the difference in electrophoretic mobility.

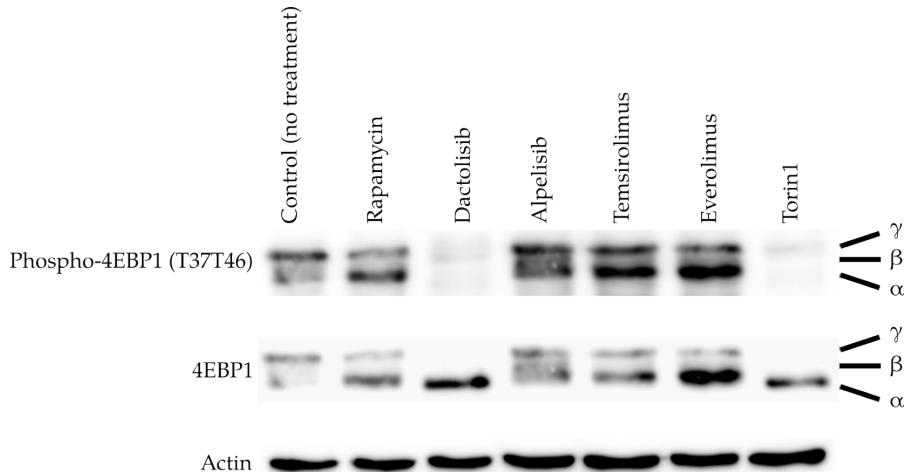


Figure 7.9: Western blot comparison of inhibitors for *4EBP1*-phosphorylating kinases in U937 cells. U937 cells were treated for 24 h with mTOR inhibitors rapamycin, everolimus, temisorlimus, torin1, dual PI3K/mTOR inhibitor dactolisib (BEZ235) or PI3K inhibitor alpelisib (Shi et al., 2018; Thoreen et al., 2009; Wilhoit et al., 2020; Zheng & Jiang, 2015), then lysed and prepared for Western blot (Section 9.2.1). Total *4EBP1* was detected using the 554R16 antibody and T37T46-phosphorylated *4EBP1* was detected using the 4EB1T37T46-A5 antibody (Table 10.4). Experimental procedure was performed by Asmaa Abdel Aziz Elrakaybi (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany) within the framework of the MONOGRAM project (MET-ID 55).

For Western blots, the 554R16 antibody was used for detection of total *4EBP1* and the 4EB1T37T46-A5 antibody was used for the detection of the T37T46-phosphorylated *4EBP1* (Table 10.4). The β form is hardly visible in any of the shown blots (Figure 7.9). In the control, the γ band is the most prominent in both detections using 4EB1T37T46-A5 and 554R16 antibodies, which means that *4EBP1* is hyperphosphorylated (Figure 7.9) corroborating previous findings (Figure 7.8) (Xu et al., 2005). For rapamycin and its analogs temsirolimus and everolimus, there was an increase in the α form detected by both antibodies, which means that these inhibitors decreased phosphorylation of *4EBP1* but not at the T37T46 residue because the 4EB1T37T46-A5 detected the protein (Figure 7.9). This is in agreement with published data, which demonstrated that the T37T46 phosphorylation is rapamycin insensitive (Livingstone & Bidinosti, 2012; Thoreen et al., 2009). For the

inhibitors dactolisib (BEZ235) and torin1, the 4EB1T37T46-A5 antibody could not detect any band, which indicates that the T37T46 phosphorylation was removed. Sole inhibition of PI3K using alpelisib (Wilhoit et al., 2020) could not change 4EBP1 phosphorylation. Furthermore, the 554R16 antibody could only detect the α form when using dactolisib or torin1 (Figure 7.9). This means that these inhibitors induced complete dephosphorylation of 4EBP1. Interestingly, they have different modes of action (dual PI3K/mTOR inhibition vs. ATP-competitive mTOR inhibition), while the detected outcome is similar. For a better understanding in future studies, a broader picture is necessary, which captures changes of other proteins along the PI3K/AKT/mTOR pathway (Nepstad et al., 2020) using either Western blot or PICO. In the following experiments, dactolisib was used to induce mechanistic dephosphorylation of 4EBP1.

U937 cells were treated for 4 h with dactolisib, then harvested and prepared for the 4EBP1 PICO assay (Figure 7.4). Compared to mock treatment, dactolisib treatment did not significantly change the number of detected 4EBP1 complexes per cell (standardized estimate -0.35 with 95 % CI from -0.70 to 0.01) (Figure 7.11). As this treatment interferes with the signaling of the cells, it is not granted that the abundance of 4EBP1 itself is not affected. This might change upon prolonged exposure as regulatory mechanisms will try to reduce the impact of the drug on the cell. The antibody pair 4EB1T37T46-A5 + 4F3-H2 detected a significantly higher number of T37T46-phosphorylated 4EBP1 complexes than actual 4EBP1 complexes (standardized estimate 0.45 with 95 % CI from 0.26 to 0.64), while the other antibody pair against the phosphorylation (4EB1T37T46-A5 + 60246-1-Ig) detected a significantly lower number of T37T46-phosphorylated 4EBP1 complexes (standardized estimate -0.20 with 95 % CI from -0.39 to -0.01) (Figure 7.11). However, after dactolisib treatment, the number of T37T46-phosphorylated 4EBP1 complexes was reduced in both antibody pairs detecting the phosphorylation (standardized estimates -1.08 with 95 % CI from -1.69 to -0.47 and -0.47 with 95 % CI from -1.25 to 0.31 , respectively) as expected from the Western blot results (Figure 7.9). Although, the reduction detected with the antibody pair 4EB1T37T46-A5 + 60246-1-Ig was not significant (from -0.20 before treatment to -0.47 after treatment (Figure 7.11B)). Again the antibody pairs employing the E.638.3 antibody did not result in the detection of complexes and after treatment the number of complexes from antibody pairs employing this antibody did not increase (Figure 7.11). This is now strong evidence that the E.638.3 antibody is not functional under these conditions. Overall, dactolisib treatment seems to be more efficient for the dephosphorylation of 4EBP1 as the numbers of T37T46-phosphorylated 4EBP1 complexes with both antibody pairs are around 0.1 , while they were around 1.0 for λ -PPase dephosphorylation (Figure 7.8A and Figure 7.11A).

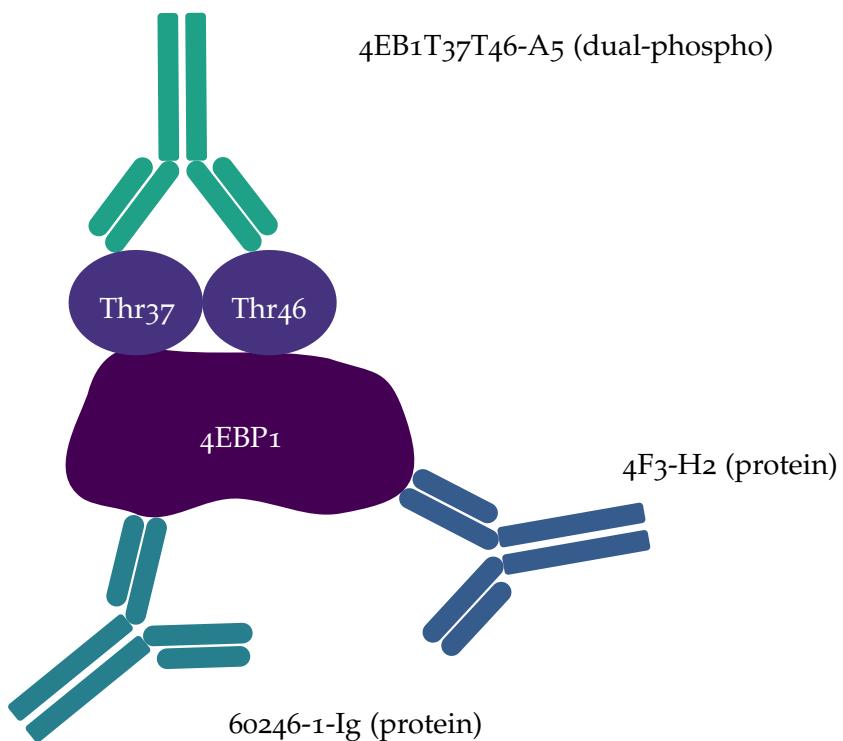


Figure 7.10: Validated PICO assay for the detection of 4EBP_1 using antibodies 4EB1T37T46-A5 , $60246-1\text{-Ig}$ and 4F3-H2 .

7.2.3 Summary

Within this section, two antibodies $60246-1\text{-Ig}$ and 4F3-H2 for the detection of 4EBP_1 were identified using recombinant and cellular 4EBP_1 . Although, information about the exact location of the binding epitopes is not available, the repeated detection of complexes is a strong indication for simultaneous binding. Furthermore, these antibodies were completed with the 4EB1T37T46-A5 anti-T37T46-phosphorylated 4EBP_1 (Figure 7.10). This antibody showed no unspecificity using recombinant protein (Figure 7.7) and its binding was reduced upon dephosphorylation with either $\lambda\text{-PPase}$ (Figure 7.8) or dactolisib (Figure 7.11) indicating high specificity. Its counterpart, the E.638.3 non-T46-phosphorylated 4EBP_1 antibody did not work within these settings. However, an additional non-phospho antibody would improve the meaningfulness of the PICO assay in future studies.

Nevertheless, in some cases the phosphorylation antibody pairs detected more phosphorylation than actual 4EBP_1 . This issue might be a consequence of chemically-selective antibodies as seen in the Western blots (Figure 7.9). Depending on the degree of phosphorylation, the intensity of some bands increased and partial dephosphorylation also changes the affinity of antibodies to the protein. Chemically-selective antibodies are not desirable for use in PICO because they might complicate stoichiometric calculations as shown here. A reasonable

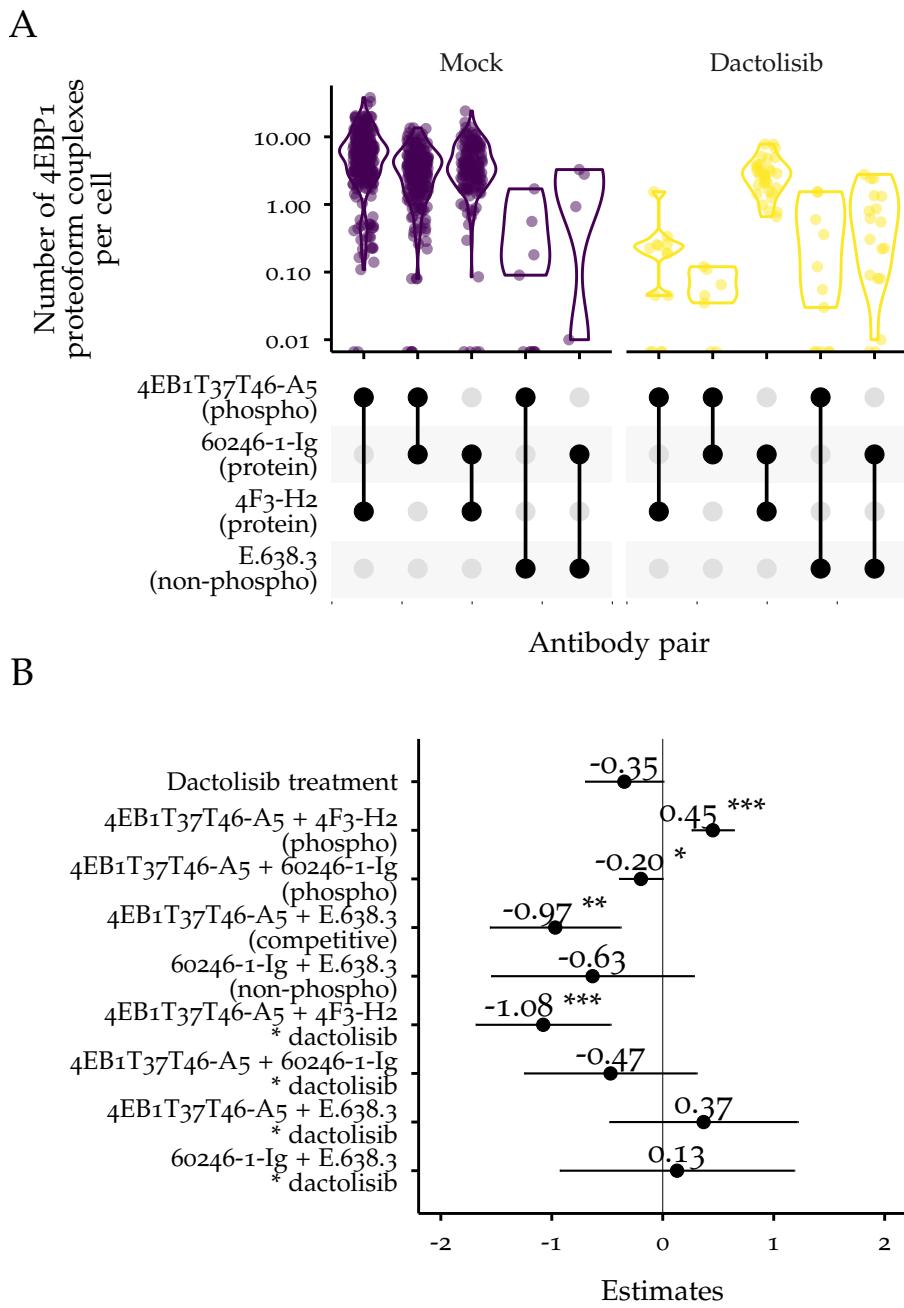


Figure 7.11: Comparison of 4EBP₁ phosphorylation between dactolisib and mock treated U937 cells using the 4EBP₁ PICO assay shown in Figure 7.4. Data was from at least 2 independent experiments.

A: UpSet plot of the number of complexes per cell detected from dactolisib and mock treated U937 cell lysates. Sample size were 241, 206, 143, 10, 4, 14, 7, 32, 12, 18 (from left to right). Low sample sizes do not necessarily mean that experiments were not thoroughly repeated but that this plot depicts already filtered results. The dots were jittered to avoid overplotting. The y-axis was log₁₀-transformed for better visibility.

B: Forest plot of estimates with 95 % CI from the linear model `lm (calculated_complexes ~ antibody_pair * treatment)` with mock treatment and antibody pair 60246-1-Ig + 4F₃-H₂ as reference. The predictors were standardized using Equation 9.3 (Gelman, 2008). The * symbol indicates the interaction of two input variables. The estimate is the change in the response variable for a 2 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R² for the model is 0.158.

explanation for more phosphorylation than actual protein can be the fact that the antibodies did not detect every target in every state. This delivers these seemingly conflicting statements. However, it requires methods with comparable sample preparation for the assessment of each antibody's selectivity and specificity. This means that the here suggested antibodies (Figure 7.10) require deeper characterization regarding chemical selectivity and location of binding epitopes.

7.3 DETECTION OF PHOSPHORYLATED 4EBP1 IN AML CELL LINES

The validated 4EBP1 PICO assay (Figure 7.10) was then used within this section to reveal potential differences between U937 and MOLM13 cell lines. Both are AML cells lines and were rendered resistant against the HMA decitabine by David Uhl (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany) in the framework of the MONOGRAM project (MET-ID 55) (Section 9.1.1). Both U937 and MOLM13 are monocytic AML cell lines and as such part of the LL-100 panel for blood cancer studies (Quentmeier et al., 2019). The major difference lies in the fact that MOLM13 cells have a FMS-like tyrosine kinase-3 (*FLT3*) internal tandem duplication (ITD) mutation. *FLT3* is located upstream of mTOR (Nepstad et al., 2020).

Overall, the resistant cell lines are not different from their native counterpart (standardized estimates -0.14 with 95 % CI from -0.68 to 0.40 and -0.17 with 95 % CI from -1.03 to 0.69 , for U937 and MOLM13 cell lines respectively) (Figure 7.12 and Figure D.6). Interestingly, for both cell lines the antibody pair 4EB1T37T46-A5 + 4F3-H2 detected a higher number of complexes than the 4EB1T37T46-A5 + 60246-1-Ig antibody pair. However, in MOLM13 cells the differences to the reference antibody pair 60246-1-Ig + 4F3-H2 was lower compared to U937 cells (Figure 7.12), where the 4EB1T37T46-A5 + 4F3-H2 antibody pair detected more phosphorylation than protein (see also Figure 7.8 and Figure 7.11). The estimated difference between the native and the resistant version of the cell lines was not significant (Figure 7.12). Thus, based on this data native and HMA-resistant AML cells cannot be distinguished. To establish a more sophisticated data basis, the here shown experiments need more repetitions, more cell lines need to be included (for instance form the LL-100 panel (Quentmeier et al., 2019)) and wild-type and AML patient samples need to be analyzed. However, before such an in-depth analysis for the potential of 4EBP1 as a predictive biomarker for HMA-resistance in AML, physicians should be approached and consulted to answer questions concerning the drugability (Owens, 2007) of 4EBP1, the meaning of a protein far downstream in a pathway and potential issues of non-surface markers.

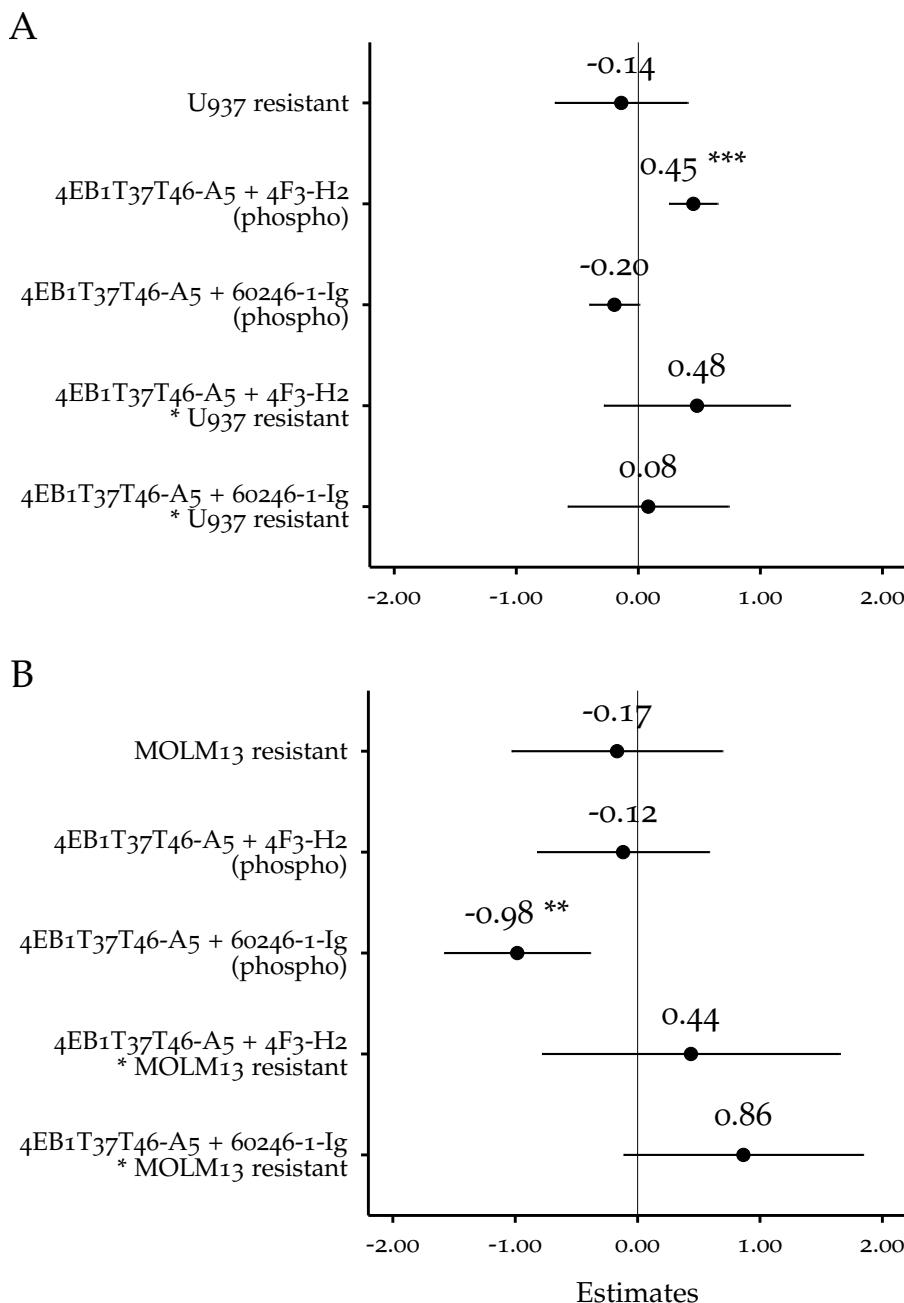


Figure 7.12: Forest plots of estimates with 95 % CI for the comparison of (HMA-resistant) AML cell lines using the validated 4EBP1 PICO assay (Figure 7.4). Data was from at least 2 independent experiments. The predictors were standardized using Equation 9.3 (Gelman, 2008). The * symbol indicates the interaction of two input variables. The estimate is the change in the response variable for a 2 SD increase in the predictor variable and the CI is the associated uncertainty.

A: Forest plot from the linear model `lm(calculated_complexes ~ antibody_pair * cell_line_version)` with the native U937 cells and antibody pair 60246-1-Ig + 4F3-H2 as reference. The adjusted R^2 for the model was 0.086.

B: Forest plot from the linear model `lm(calculated_complexes ~ antibody_pair * cell_line_version)` with the native MOLM13 cells and antibody pair 60246-1-Ig + 4F3-H2 as reference. The adjusted R^2 for the model was 0.121.

7.4 SUMMARY

It requires more information on 4EBP1 to use it as a predictive biomarker for the identification of **HMA** resistance in **AML** patients (e.g. reliable non-phosphorylation antibody, more cell lines and patient samples). However, hyperphosphorylation of 4EBP1 independent of the resistance state could be shown. This is a common case for cancers (Hsieh et al., 2015; Xu et al., 2005; Zhang & Zheng, 2012) as this indicates constantly active PI3K-Akt-mTOR signaling, which in turn is favors cell proliferation.

The here described steps, i.e. comparison of different antibodies for total protein detection and controlled dephosphorylation in case of anti-phospho antibodies, shed light on the work required to validate a PICO assay. Yet, the steps could not rule out all variability, and also the fact that multiplex PICO assays detected more protein modification than protein remains. The combined usage of recombinant and cellular 4EBP1 and Western blot analyzes allowed identifying that these issues partially originate from chemically-selective antibodies or the protein itself (e.g. α , β and γ proteoform).

Although the **dDPCS** model for the determination of complexes was validated ([Chapter 6](#)), the results from [Chapter 7](#) underline the dependence of the PICO assay on the quality of the antibodies. High quality and well-documented antibodies with proven ability for simultaneous binding like **TTZ** and **PTZ** are highly reliable as demonstrated in [Section 6.2.3](#). With better antibodies (Weller, 2016, 2018), the PICO assays studied in this thesis would achieve a higher resolution and the shallow conclusion on 4EBP1 phosphorylation upon different treatments might be revised. This requires further investigations. Thus, this chapter can be seen as a proof of concept for PICO to detect 4EBP1 in a relevant framework, while it cannot yet deliver conclusive findings on the suitability of 4EBP1 as a biomarker for **HMA** treatment in **AML**.

8

CONCLUSION

Within this chapter, the main findings of this thesis are summarized and limitations and derived future perspectives are discussed. In general, this thesis

- provided the first comprehensive description of the **PICO** technology;
- derived the mathematical model for couple complex calculation based on the mathematics of dPCR and independent distribution processes, forming the **dDPCS** model;
- provided a dPCR simulation for the numerical validation of the **dDPCS** model;
- achieved empirical validation of the **dDPCS** model using experimental data, demonstrating reliability despite inherent biological variability, and
- provided an application of the PICO technology to a challenging research question: the simultaneous detection of different proteoforms of the same protein, a task previously difficult to achieve with existing immunoassays.

8.1 ACHIEVEMENTS

To sum this up, the **PICO** technology comprises the following steps:

1. Selection of antibodies against the target of interest and labeling of them with DNA-labels.
2. Incubation of the target of interest together with the labeled antibodies in a homogeneous solution for the protein-to-nucleic acid transformation.
3. Partitioning, amplification and detection of DNA-labels by dPCR.
4. Calculation of the number of complexes from the 2-dimensional raw dPCR signal using the **dDPCS** model.
5. Conversion of complexes into absolute number of targets of interest¹.

¹ This part was not an objective of this thesis (Chapter 3) and is described in Tobias Gross' PhD thesis and in Gross et al. (2024)

PICO uses antibodies to specifically transform proteoforms into DNA. This combines the high degree of information from proteoforms with highly sensitive and easy-to-measure methods for DNA detection and absolute quantification (Figure 1.3). The usage of DNA-labeled antibodies enables utilization of the vast infrastructure for nucleic acid analysis (Zhao et al., 2021). Unlike competitor methods such as Simoa or LC-MS/MS, PICO requires no special equipment other than standard laboratory equipment and a dPCR system. Adopting dPCR, PICO constitutes a digital immunoassay with attomolar sensitivity. Other immunoassays like PLA or PEA exhibit sensitivities inferior to PICO (Figure 6.22), underlining that PICO leverages the sensitivity of dPCR for the detection and quantification of proteoforms. Simoa shows the highest sensitivity among the immunoassays compared here, which are heterogeneous assays lacking reference-free absolute quantification. In contrast, PICO is a homogeneous assay allowing calculation of absolute number of proteoforms (Gross et al., 2024). From a 2-dimensional raw dPCR signal, the dDPCS model presented in this thesis calculates the number of complexes. Under certain assumptions, the complexes can be directly compared with each other and allow for the calculation of ratios between different experimental groups. Further on, the number of complexes can be used together with the antibody concentrations and their dissociation constants (K_d values) (Figure 4.4) for calculating the absolute number of targets of interest in the sample (Gross et al., 2024).

For the calculation of the number of complexes, the dDPCS model was devised and validated within this thesis². It is based on the mathematical fundaments of dPCR and exploits the difference between distributions of DNA-labeled antibodies in the partitions when bound to their target or free. The model was validated in two phases with two stages each (Figure 3.1). The focus of the first phase was the application of the model to simulated data. Thus, a dPCR simulation needed to be created and was successfully validated first (Chapter 5). The focus of the second phase was the application of the model to experimental data. Using simulated data, the dDPCS model demonstrated high accuracy and specificity and that the majority of tested biases could be avoided by low λ values. However, including the protein-to-nucleic acid transformation mediated by the DNA-labeled antibodies reduced specificity and both LOD and LOQ (Table 6.2). In general, the dDPCS model proved to be in agreement with ground truth values and improved educated-guess couple calculations by orders of magnitudes (Figure 6.21).

Furthermore, through the usage of multiple antibodies, the PICO technology allows for multiplexing (Figure 7.10). This increased the

² This model is based on the same assumptions as the DX model from Gross et al. (2024) but uses a different implementation. A comparison of the models can be obtained from Section A.1.

information gained from the same sample, which only depends on the number of chosen antibodies and the fluorescent detection channels of the dPCR system. Thus, the PICO technology holds the potential for the simultaneous detection of proteoforms and PPIs without major changes to the workflow. This is an unique asset, since PPIs are often controlled by PTMs (e.g. *4EBP1* and *eIF4E*). It could be shown that *4EBP1* is hyperphosphorylated in HMA- and non-resistant AML cell lines, which is an indication for constantly active PI₃K-Akt-mTOR signaling, which in turn is favors cell proliferation.

Taken together, this thesis provides the fundamental advancement for the absolute quantification of proteoforms by providing the number of complexes through a novel detection principle of ternary complexes (complexes) from 2-dimensional raw dPCR data. However, complexes on their own can be used to calculate ratios between different experimental groups. Thus, PICO can provide invaluable insights into the stoichiometry-based characteristics of complex biochemical processes, including the dynamic behavior of the proteome. This paves the way for the analysis of cellular signaling and pathways, which ultimately provides a deeper the understanding of cellular behavior.

8.2 LIMITATIONS

Yet, the introduction and description of a novel method never comes without limitations, which need to be highlighted and addressed in follow-up studies. In general, the development of any immunoassay is time- and money-consuming, because of the need for validated antibodies. In case of a PICO assay, two antibodies are required that simultaneously bind without affecting each other and without cross-reactivity. Within this thesis, 10 antibodies were tested and compared³ for the development of the *4EBP1* PICO assay as depicted in Figure 7.4. The costs for research antibodies range from 240 € to 660 € (antibodies.com, 2023). This means that on average 4500 € were spent just on antibodies. Together with the consumables for the dPCR and antibody labeling, the costs can easily reach up 10 000 €. Then, the assay requires validation before it can be finally used to address scientific questions. Yet, the question remains, what is a sufficient endpoint for assay validation? The *4EBP1* PICO assay in this thesis is not completely free of ambiguity, which is most likely related to the antibodies. Usually research antibodies have a poor documentation, which makes it hard to judge if a given pair of antibodies will work together (Weller, 2016, 2018). Therapeutic antibodies, on the other hand, with known ability for simultaneous binding like TTZ and PTZ (Nahta et al., 2004; Nami et al., 2018; Sharma et al., 2021) are preferred and worked out of

³ Not all combinations of antibodies were tested because of incompatible fluorescence dyes of the probes for DNA-label detection. This would have required 16 probes for each of the 4 labels with all 4 dyes. Not all tested combinations were shown.

the box as demonstrated here (Section 6.2.3). The following questions arise when it comes to antibody selection:

- Are antibodies available against the PTM of interest?
- How well can the antibodies differentiate between (slightly) different proteoforms?
- Can polyclonal antibodies⁴ reduce the selectivity of monoclonal antibodies for a specific proteoform? If yes, how would they work in a PICO assay?
- How much does the cross-reactivity of the antibodies increase, when adding more antibodies?
- How many antibodies against the same protein or protein complex can be used without steric hindrances?

The addition of multiple antibodies allows for multiplexing, which increases the information from a single sample. However, the number of calculated complexes might not be directly comparable (Section 7.3). Although the signal originates from the same source material, the $4\text{EBP}1$ PICO assay detected more phosphorylation than protein in some cases. There are three apparent possible hypotheses to explain this finding:

- The chosen antibodies are chemically-selective for a given proteoform of $4\text{EBP}1$ (Figure 7.9).
- The K_d values of the antibodies vary strikingly.
- The antibodies affect each other's binding to the target.

The first hypothesis can be checked by Western blot, although this requires employing similar binding conditions as in the binding reaction of the PICO assay. The second hypothesis can be addressed by calculating the absolute quantities considering the K_d values, which requires knowledge about these. The final hypothesis requires sequential assay design and thorough knowledge about the antibodies. This might limit the direct comparability of complexes. For other immunoassays (Table 2.1), these issues might not be so obvious. The sensitivity of PICO is in the attomolar range (Figure 6.22) but the affinities of antibodies are generally lower (Abcam, n.d.). In Simoa, for instance, several antibodies are coated on a bead, while in PICO a single antibody detects the protein because single complexes are partitioned in the dPCR. Thus, PICO reveals binding issues of single antibodies with previously unforeseen resolution. This eventually sheds light onto the heterogeneity of seemingly monoclonal antibody populations.

⁴ Polyclonal antibodies are considered to have higher affinities and were already used in PEA (Lundberg, Eriksson, et al., 2011; Selby, 1999)

In practice, the degree of multiplexing is limited by the number of fluorescent channels of the dPCR system for the detection of DNA-labeled antibodies. Current dPCR systems have up to 6 channels (Tan et al., 2023), which allow detection of up to 15 different antibody pairs because of the combinations of the fluorescent channels (Figure 4.2B). Yet, not all combinations of fluorescent channels enable the detection of antibody pairs because of spillovers (Section 6.2.3) and a certain fraction of antibody pairs deliver the same information (inherent redundancy). Therefore, this requires careful tuning of fluorescent dye combinations and fluorescent channels of the dPCR system. However, by designing DNA-labels in such a way that two assemble to one during amplification in a partition by overlap-extension PCR (MD Nelson & Fitch, 2012) with subsequent recovery of amplified DNA-labels and sequencing by NGS, the degree of multiplexing can be significantly enhanced (Section 8.3). This renders the detection of DNA-labeled antibodies independent of the number of fluorescent channels. In this case, the availability of monoclonal antibodies against the targets of interest and their simultaneous binding ability becomes the limiting factor for multiplex PICO assays.

Through the use of antibodies for the transformation of proteins to nucleic acids, immunoassays in general are targeted approaches and do not allow for *de novo* discoveries. It is therefore debatable to use such technologies for biomarker discovery (Nong et al., 2012), while it might still be useful for the validation of putative biomarkers. Biomarker discovery is best done with data-driven approaches, instead of hypothesis-driven approaches (Chaffey & Silmon, 2016), because they allow broader perspectives. In contrast to immunoassays, MS-based approaches demonstrated their utility for drafting the human proteome (Wilhelm et al., 2014), for studying relationships between transcriptome and proteome (D Wang et al., 2019) and for understanding global splicing events and describing proteoforms (Sinitcyn et al., 2023). However, there are also some protein sequencing methods based on protein-to-nucleic acid transformation (Darmanis et al., 2011; Hong et al., 2022; Vistain et al., 2022; Wik et al., 2021), and recently, the spotlight was put on nanopores for protein sequencing (Singh, 2023).

8.3 FUTURE PERSPECTIVES

The PICO technology holds great promise for research in the post-genomic era and the opportunities are numerous. For instance, as PICO relies on the simultaneous binding of two antibodies, this ability can be exploited for the validation of two therapeutics, which require simultaneous binding or whose simultaneous binding enhances therapies as in the case of TTZ and PTZ (Nahta et al., 2004). It is also imaginable that PICO could allow for the detection and quantification of protein-RNA interactions with suitable sample preparation.

This could generate high impact because the interaction of these two types of biomolecules is of great relevance to physiological processes such as mRNA translation (S Liu et al., 2020). Indeed, PICO is not a high-throughput method and is thus probably not suitable for the discovery of biomarkers (see above). However, it can be used for the confirmation of putative biomarkers after screening with MS-based approaches because it comprises a higher sensitivity (Table 6.2). The combination of an immunoassay such as PICO for target enrichment and MS-based approaches for high-throughput analysis and deciphering of proteoform diversity might also be a fruitful combination (RW Nelson et al., 1995).

Furthermore, the information gathered here (e.g. dependence of assay specificity on antibodies and λ) can be used to refine the dDPCS model and to improve the calculation of the number of complexes. Another approach would be the incorporation of technical and biological variability in the dDPCS model. This can improve the yield of the dDPCS model in terms of more successful calculations with differences between calculation and observation lower than 1% of $\hat{n}_{D_{obs}}$ positive partitions (Section 4.2.3). In this thesis, clustering for the decision of a partition's status was not investigated. This could also be addressed in a refinement of the dDPCS model using inspiration from Brink et al. (2018), De Falco et al. (2023), and Lievens et al. (2016). The use of artificial intelligence for the identification of complex double positive partitions based on existing approaches, for instance from Lee et al. (2023), is imaginable. However, the workload for generation of training data, the extent of the training data and the training of the model is difficult to estimate. Finally, based on the findings and limitations discussed here, explicit enhancements and extensions of the PICO technology can be conceived (ordered by relevance):

- To overcome the issues of research antibodies (Weller, 2016, 2018), fully defined and chemically synthesizable aptamers might be used to replace antibodies as binding agents because the PICO technology does not rely on antibodies *per se* (Jeney, 2016). The great advantage of aptamers is that they are cheap and quickly produced (compared to antibodies) and act as both binding agent and detection agents (Janssen et al., 2013; Nong et al., 2012). Protein-to-nucleic acid transformation would thus not require the labeling process. However, the identification of suitable binders requires systematic evolution of ligands by exponential enrichment (SELEX) processes. SomaLogic is a company that has developed an aptamer panel with now 11 000 aptamers against human proteins (Candia et al., 2022; Gold et al., 2012).
- The multiplexing capabilities of the PICO assay can be expanded by using monochrome multiplexing approaches like “reporter emission multiplexing” (Calabrese et al., 2023; Wadle et al., 2014)

or “selective photobleaching” (Schlenker et al., 2021). The advantage of these approaches is that dPCR remains the technology for partitioning, amplification and detection, although monochrome multiplexing approaches require fine-tuning of fluorescent dyes and data evaluation is more complex.

- So far, information on the formation of a ternary complex (complex) was secured by encapsulation of the DNA-labels in partitions. If the PCR for amplification was designed in such a way that 2 DNA-labels are joined together by overlap-extension PCR (MD Nelson & Fitch, 2012), the binding information would be secured even after recovery of the amplified DNA-labels from the partitions, which could then be sequenced. Then the number of complexes could be determined from the number of joined DNA-labels. However, this requires a dPCR system with the ability to recover the reaction mix after thermocycling, as is possible in devices from Schulz et al. (2020), Schulz et al. (2021) or Lai et al. (2023) and the integration of unique molecular identifier into the DNA-labels. This enhances multiplexing capabilities significantly, but the resources inside a single partition will become the limiting factors for the number of joined DNA-labels that can be efficiently amplified in the same partition. Furthermore, this poses major challenges for data analysis, as the number of possible joined DNA-labels scales exponentially with the number of antibodies in a single partition.
- Bispecific antibodies (Kontermann & Brinkmann, 2015) demonstrated utility in the detection of modified proteins and PPIs (van Dieck et al., 2014). They might be of great relevance for multiplex PICO assays because the same information can be gained with fewer antibodies and thus expanding multiplexing capabilities. However, data analysis might become more complex because bispecific antibodies do not necessarily have to bind both targets simultaneously.

EXPERIMENTAL PROCEDURE

METHODS

9.1 CELL CULTURE

All centrifugation steps with cells were performed at 400 g for 5 min in a Heraeus Multifuge.

9.1.1 *Cells and culture conditions*

U937 (ACC 5, DSMZ) and MOLM13 (ACC 554, DSMZ) cell lines in a native and decitabine-resistant variant were a kind gift from Heiko Becker (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany). The IC₅₀ of decitabine was 275.4 nmol L⁻¹ and 133 840.0 nmol L⁻¹ in native and decitabine-resistant U937 cells, respectively. The IC₅₀ of decitabine was 75.4 nmol L⁻¹ and 1010.0 nmol L⁻¹ in native and decitabine-resistant MOLM13 cells, respectively. The cells were rendered decitabine-resistant by David Uhl (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany) within the framework of the MONOGRAM project (MET-ID 55).

All cell lines were cultured in RPMI Medium 1640 (1X) + GlutaMAX-I (61670-010, Gibco) supplemented with 10 % fetal bovine serum (10270106, Gibco) and 1 % penicillin-streptomycin (15140-122, Gibco) in a 5 % CO₂ atmosphere at 37 °C. Depending on the required cell numbers, cells were incubated in non-treated T25 (CC7672-4825, CytoOne) or non-treated T75 flasks (CC7672-4875, CytoOne).

9.1.2 *Inhibitor treatment of cells for Western blotting*

Inhibitor treatment of cells was performed by Asmaa Abdel Aziz Elrakaybi (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany) within the framework of the MONOGRAM project (MET-ID 55). U937 cells were treated during culture with 1 µmol L⁻¹ rapamycin, alpelisib, everolimus, temsirolismus or torin1 or 2.5 µmol L⁻¹ dactolisib for 24 h and then harvested for Western blot ([Section 9.1.5](#)).

9.1.3 Dactolisib treatment of cells for PICO

5 mg of dactolisib (HY-50673, MedChemExpress) were dissolved in DMSO (D8418, Sigma-Aldrich) resulting in a 10 mmol L^{-1} solution. The solution was sonicated in an ultrasonic bath (Sonorex) for 5 min at full power. Afterwards, the solution was heated for 15 min at 60°C . The dactolisib solution was stored at -20°C until use.

Prior to treatment, the dactolisib solution was diluted to 1 mmol L^{-1} using cell culture medium. DMSO was diluted in the same way for mock treatment. After harvesting and counting (Section 9.1.4), 4×10^6 cells were transferred to a new T25 culture flask with 5 mL fresh medium. 25 μL diluted DMSO and 25 μL diluted dactolisib solution (final concentration of $\sim 5.6 \mu\text{mol L}^{-1}$) were added to each one flask. Afterwards, the cells were cultured for 4 h and then harvested (Section 9.1.4 and Section 9.1.6).

9.1.4 Cell harvest and counting

The cell suspension was aspirated with a serological pipette and transferred to a 15 mL conical tube. After centrifugation the supernatant was discarded and the cells were washed with 10 mL PBS and centrifuged again. The supernatant was removed. Then, the cell pellet was dissolved in 1 mL PBS and transferred into a fresh 1.5 mL reaction tube (0030120086, eppendorf). In another, new tube, 15 μL of this suspension were combined with the same volume of trypan blue stain (T10282, Invitrogen) for live/dead staining. The sample was counted using Countess Cell Counting Chamber Slides (C10228, Invitrogen) with the Countess 3 Automated Cell Counter (Invitrogen).

9.1.5 Cell harvest and lysis for Western blot

Cell harvest and lysis for Western blot was performed by Asmaa Abdel Aziz Elrakaybi (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany) within the framework of the MONOGRAM project (MET-ID 55). After treatment of cells with inhibitors (Section 9.1.2), they were pelleted and dissolved in 80 μL to 100 μL Tris-based lysis buffer and incubated for 20 min on ice. The supernatant was cleared by centrifugation for 20 min at 4°C at full speed.

9.1.6 Cell lysis for PICO

After harvest and counting (Section 9.1.4), cells were pelleted, and the supernatant was removed. Then, LB (PICO-000010, Actome) lysis buffer was added to dissolve the cell pellet and to adjust the cell

concentration to 1×10^4 cells/ μL , if not stated differently. The lysis reaction was incubated for 3 h at 4 °C. Afterwards, the lysate was sonicated in an ultrasonic bath (Sonorex) for 5 min at full power. The sonicated lysate was homogenized using a QIAshredder spin column (79656, QIAGEN) at 20 000 g for 2 min. If not stated differently, cell lysates were diluted by 10- and (2000 cells/ μL) 20-fold (1000 cells/ μL) using LB.

9.1.7 Cell lysis for PPase treatment

After cell harvest and counting (Section 9.1.4), 2×10^6 cells were aliquoted for λ -PPase and mock (H_2O) treatment. The cells were pelleted by centrifugation. The pellet was dissolved in 175 μL modified RIPA buffer (Table 9.2) and incubated for 1.5 h at 4 °C. Afterwards, the lysate was sonicated in an ultrasonic bath (Sonorex) for 5 min at full power. The sonicated lysate was homogenized using a QIAshredder spin column (79656, QIAGEN) at 20 000 g for 2 min. This solution was transferred to a 0.5 mL Protein LoBind tube (0030108094, eppendorf) and λ -PPase or H_2O were added according to Table 9.2. The reaction was incubated for 1 h at room temperature. Then, SDS was added according to Table 9.2. The reaction was incubated for 1 h at 4 °C. Cell lysates were diluted by 10- and (2000 cells/ μL) 20-fold (1000 cells/ μL) using RIPA buffer (89900, Thermo Scientific), similar to Section 9.1.6.

Table 9.1: Recipe for preparation of 2X RIPA buffer.

Ingredient	Final concentration	Manufacturer	Catalog number
Tris	100 mmol L ⁻¹	Sigma-Aldrich	252859
Sodium deoxycholate	1 %	Sigma-Aldrich	D6750
NaCl	300 mmol L ⁻¹	Roth	3957.1
Triton X-100	2 %	Sigma-Aldrich	X100-100ml
cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail	2X	Roche	04693159001
DNase-/RNase-free, distilled H ₂ O	(ad desired volume)	Invitrogen	10977035

Table 9.2: Recipe for the preparation of modified RIPA buffer for λ -PPase treatment.

Ingredient	Final concentration	Manufacturer	Catalog number
2X RIPA buffer	1X	Table 9.1	n/a
10X NEBuffer for Protein MetalloPhosphatases (PMP)	1X	NEB	P0753S
10 mmol L ⁻¹ MnCl ₂	1 mmol L ⁻¹	NEB	P0753S
DNase-/RNase-free, distilled H ₂ O	17.5 %	Invitrogen	10977035
This component is added after the first incubation step (1.5 h).			
λ -PPase	2.5 %	NEB	P0753S
This component is added after the second incubation step (1 h).			
1 % SDS	0.1 %	Roth	2326.2

9.2 MOLECULAR BIOLOGY

If not stated differently, all incubation steps were performed in a C1000 Touch Thermal Cycler (Bio-Rad).

9.2.1 SDS-PAGE and Western blot

SDS and Western blot were performed by Asmaa Abdel Aziz El-rakaybi (Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany) within the framework of the MONOGRAM project (MET-ID 55). In brief, the protein concentration from the samples prepared in Section 9.1.5 was determined using Bradford Assay (Stoscheck, 1990) with a BSA standard. For each sample, 45 µg total protein were combined with 5X SDS loading buffer and boiled for 5 min at 95 °C. After the gel run, the proteins were transferred to a membrane using 1X methanol transfer buffer. The membrane was blocked with 5 % milk in TBS with 0.1 % Tween20. The membrane was then incubated with the primary antibodies at 4 °C overnight. The anti-actin antibody¹ and the 4EBP1 Monoclonal Antibody (554R16) (AHO1382, Invitrogen) were used in a 1000-fold dilution. The Phospho-4EBP1 (Thr37, Thr46) Recombinant Rabbit Monoclonal Antibody (4EB1T37T46-A5) (MA5-36935, Invitrogen) was used in a 500-fold dilution. After incubation, the membrane was washed three times with TBS with 0.1 % Tween20. The secondary antibody against primary rabbit antibodies was used in 6000-fold dilution, and

¹ No detailed information provided.

the secondary antibody against primary mouse antibodies was used in a 3000-fold dilution. The membrane was incubated for 1 h with the secondary antibodies. The membrane was washed three times with TBS with 0.1% Tween20 and once with TBS. After developing, the membrane was imaged.

9.2.2 *ELISA*

Total 4EBP_1 and T37T46-phosphorylated 4EBP_1 from U937 cells was quantified using PathScan Total 4E-BP_1 Sandwich ELISA Kit (#7179, Cell Signaling) and PathScan Phospho- 4E-BP_1 (Thr37/Thr46) Sandwich ELISA Kit (#7216, Cell Signaling) according to manufacturer's instructions. Absorbance at 450 nm was measured using a SPARK 20M microplate reader (Tecan).

9.2.3 *Antibody labeling*

Antibodies were labeled using the PICOglue Antibody Labeling Kit (PICO-000110, Actome) (Jeney, 2023a). All centrifugation steps were performed at 14 000 g. In brief, 100 µg antibody was rebuffered in 1X Ultrafiltration Buffer using 100K Ultrafiltration Columns. The re-buffered antibody was recovered by inverting the 100K Ultrafiltration Column, placing it in a new 1.5 mL reaction tube and centrifuging it for 20 s at ~ 1000 g. On average, the recovered volume was between 15 µL and 40 µL, of which 1 µL was saved as a control. Next, the antibody was deglycosylated using 3 µL PICOzyme Buffer and 3 µL PICOzyme. The reaction was incubated for 1 h at 37 °C. The deglycosylated antibody was added to the PICOTransferase Substrate and PICOTransferase Buffers A and B were added as described in the protocol. 1.3 µL PICOTransferase was added, and the reaction was incubated at 30 °C overnight for azide attachment. The next day, the antibody was re-buffered as described above. Then, 4 µL PICOglue Label (Table 10.2) was added, and the reaction was incubated at 30 °C overnight. Finally, the antibody was rebuffered as described above and recovered in a 0.5 mL Protein LoBind tube (0030108094, eppendorf). 1 µL sample was saved for determining the LE (Section 9.2.5). A matching amount of PICOglue Antibody Storage Buffer was added to the antibody containing preservatives and stabilizers. The concentration of the antibody was determined by dPCR (Section 9.2.7) according to manufacturer's instructions and was refined after each PICO experiment (Section 9.3).

9.2.4 *Protein electrophoresis*

After antibody labeling (Section 9.2.3), the taken samples were analyzed using the 2100 Bioanalyzer System (Agilent) with the Protein 230 Kit (5067-1517, Agilent). In brief, the samples were filled up to

4 µL with PBS and 2 µL Denaturing Solution containing dithiotreitol was added. The samples were boiled at 95 °C for 5 min. After cooling, 84 µL deionized water (10977035, Invitrogen) was added. The Protein Chip was prepared, the samples were loaded and the 2100 Bioanalyzer was started as described in the manufacturer's instructions.

9.2.5 Determination of antibody labeling efficiency

Rebuffered antibodies show two prominent peaks on the electropherogram, originating from the light chain at (26 ± 3) kDa and the heavy chain at (54 ± 4) kDa (Figure E.1 and Table E.1). Labeled antibodies show three prominent peaks on the electropherogram, originating from the light chain, the deglycosylated but unlabeled heavy chain at (50 ± 4) kDa and the labeled heavy chain at (75 ± 7) kDa (Figure E.2 and Table E.1). After peak integration, the AUC can be used to calculate the LE (η) using Equation 9.1. The light chain is not considered in this calculation because it cannot carry a label (Jeney, 2023a).

$$\eta = 100 \left(1 - \left(\frac{\text{AUC}_{\text{unlabeled heavy chain}}}{\text{AUC}_{\text{labeled heavy chain}} + \text{AUC}_{\text{unlabeled heavy chain}}} \right)^2 \right) \quad (9.1)$$

9.2.6 PPase treatment or recombinant proteins

Recombinant 4EBP1 (NBP1-30218, Novus Biologicals) was 10-fold diluted using PBS (0.1 µg µL⁻¹). 10 µL of this dilution were combined with 30 µL PBS, 5 µL 10X NEBuffer for Protein MetalloPhosphatases (PMP) (P0753S, NEB), 5 µL 10 mmol L⁻¹ MnCl₂ (P0753S, NEB), and 2.5 µL λ-PPase (P0753S, NEB) in a 0.5 mL Protein LoBind tube (0030108094, eppendorf). The reaction was incubated for 1 h at room temperature.

9.2.7 Digital PCR

Table 9.3: Cycling conditions for the used dPCR systems (Section 9.2.7).

Step	Temperature (°C)	Duration
Cycling conditions for QX200 dPCR system (Bio-Rad).¹		
1 Initial denaturation	95	10 min
2 Denaturation	95	30 s
3 Annealing and Extension	58	1 min
4 Back to step 2 (40 cycles)		
5 Enzyme deactivation	98	10 min
Cycling conditions for naica dPCR system (Stilla).¹		
1 Partitioning	40	(predefined)
2 Initial denaturation	95	10 min
3 Denaturation	95	15 s
4 Annealing and Extension	58	30 s
5 Back to step 3 (45 cycles)		
6 Pressure release	Room temperature	(predefined)
Cycling conditions for QIAcuity dPCR system (QIAGEN).		
1 Partitioning	Room temperature	(predefined)
2 Initial denaturation	95	2 min
3 Denaturation	95	15 s
4 Annealing and Extension	58	30 s
5 Back to step 3 (40 cycles)		
6 Imaging	Room temperature	Depends on the number of channels to image.

¹ Imaging with these dPCR systems is done in separate devices.

DROPLET DIGITAL PCR DPCR using Bio-Rad's QX200 Droplet Digital PCR system (BJ Hindson et al., 2011) was performed according to manufacturer's instructions. Master mix was prepared according to Table 10.3. The master mix was pipetted in a 96-well plate and 1 µL sample was added. Using a multichannel pipette the master mix was mixed and 20 µL was loaded to the DG8 Cartridges (1864008, Bio-Rad) and emulsified using 70 µL of Droplet Generation Oil for Probes (1863005, Bio-Rad) in the Droplet Generator (1864002, Bio-Rad).

A C1000 Touch Thermal Cycler (Bio-Rad) was used for thermal cycling ([Table 9.3](#)). Droplets were analyzed using the Droplet Reader (1864003, Bio-Rad).

DPCR using Stilla's naica system (Madic et al., [2016](#)) was performed according to manufacturer's instructions but with reduced master mix volume as shown before (Lange et al., [2022](#)) ([Table 10.3](#)). The master mix was pipetted in a 96-well plate and 1 μ L sample was added. Using a multichannel pipette the master mix was mixed and 20 μ L master mix was loaded to each chamber of the Sapphire Chips (Stilla). The chips were placed in the Geode (Stilla) and cycled according to [Table 9.3](#). The chips were analyzed in the Prism3 reader (Stilla) with the following imaging settings: 65 ms and 150 ms exposure time for blue and green channel at 82 mm focus.

CHAMBER DIGITAL PCR DPCR using QIAGEN's QIAcuity Digital PCR system was performed according to manufacturer's instructions. Master mix was prepared according to [Table 10.3](#). The master mix was pipetted in a 96-well plate and 1 μ L sample was added. Using a multichannel pipette the master mix was mixed and 40 μ L master mix was loaded to the wells of a 26k 24-well Nanoplate (250001, QIAGEN) and cycled ([Table 9.3](#)). The Nanoplates were imaged with the following conditions: green channel (500 ms exposure time, gain 6), yellow (400 ms exposure time, gain 6), orange (400 ms exposure time, gain 6), and red (300 ms exposure time, gain 4).

9.3 PICO

9.3.1 DNA-only PICO

For DNA-only PICO experiments, sequences different from the ones used in antibody labeling ([Section 9.2.3](#)) were used ([Table 10.1](#)). The concentrations of P8 label 2, BLA label 2 and the artificial complex ([Table 10.1](#)) were determined prior to each experiment and within each experiment using either naica or QIAcuity dPCR systems ([Section 9.2.7](#)). The aim was to keep the total number of probe binding sites per P8 and BLA probe constant in mixtures containing different concentrations of P8 label 2, BLA label 2 and artificial complex. Given a constant number of partitions n , this means that independent of the composition of the mixture, the number of positive partitions \hat{n}_{AB+a+b^-} and \hat{n}_{AB+a-b^+} is constant. The number of double positive partitions $\hat{n}_{D_{obs}}$ is expected to increase with increasing concentrations of the artificial complex. Upon that, the volume of labels and the artificial complex were calculated under the consideration of the total, constant sample volume ($V_{sample} = 1 \mu\text{L}$). The volumes were dispensed into the dPCR master mix ([Table 10.3](#)) using the I.DOT (Dispendix) liquid handling system with I.DOT PURE plates 90 μm orifice (Dispendix).

9.3.2 Preparation of antibody mix

For PICO experiments, the DNA-labeled antibodies (Table 10.6) were combined in an equimolar ABX (Figure 4.3A) with a concentration of 4×10^{-11} mol L⁻¹ or 5×10^{-10} mol L⁻¹² in LB (PICO-000010, Actome). The concentration of the antibodies was determined by dPCR after labeling (Section 9.2.3), thus the concentration only considers labeled antibodies, while unlabeled antibodies could not be detected.

9.3.3 Preparation of binding reaction

In general, samples (recombinant HER2 (10126-ER-050, R&D Systems), recombinant 4EBP1 (NBP1-30218, Novus Biologicals), or cell lysates (Section 9.1.6)) were diluted to a concentration range, where target concentrations are low enough to avoid the signals beyond the peak of the curve, which relates complex and target concentration to each other³. For the binding of antibodies and targets (Figure 4.3B), 2 µL ABX and 2 µL sample were combined in a PCR-plate (AB0648, Thermo Scientific), quickly centrifuged in a mini-centrifuge (Biozym) and incubated overnight at 4 °C. In case of ABC PICO experiments, the ABX was combined with the CB, which is a buffer only containing dissolved bovine serum albumin (Actome GmbH, 2024).

9.3.4 Preparation for dPCR

The binding reaction was diluted with PBS in such a way that the concentration of antibodies was ≤ 321 molecules/µL in the dPCR master mix corresponds to $\lambda \leq 0.25$ for the QIAcuity dPCR system assuming a partition volume $V_p \approx 0.000\,78$ µL.

9.4 COMPUTATIONAL METHODS

9.4.1 Programming

Data from the PICO experiments was evaluated using the development version of Actome's AMULATOR software⁴. The software was developed under python 3.9.15 and uses the *MultipleOccupancy* output from the QIAcuity Software Suite 2.1.7.182 (Section A.2) for the calculation of the number of complexes with the dDPCS model (Section 4.2.3). The dPCR simulation for PICO (Chapter 5) was developed under R

² This antibody concentration is considered to saturate all epitopes (Gross et al., 2024).

³ The explanation of the concept that relates the complex concentration to the target concentration is discussed in Tobias Gross' PhD thesis and in Gross et al. (2024). In this regard the terms *low-side* and *high-side* are relevant.

⁴ The software is available upon request from github.com/LangeTo.

4.3.1⁵. All plots were generated using R 4.3.1 with a customized theme for `ggplot2` (`themeFreecastle`). To avoid overplotting, data points were jittered. This is indicated at the respective passages. Packages used for `python` and `R` are listed below. All code was written in Visual Studio Code 1.86.2 (Microsoft) and is available upon request.

9.4.2 Statistical analyses

DATA FILTERING Before statistical analyzes, data was filtered. The `dDPCS` model calculates a difference between the detected and the observed number of double positive partitions $((\hat{n}_{D_{obs}} - n_{D_{calc}})^2)$. If the difference is larger than 1 % of $\hat{n}_{D_{obs}}$, the corresponding calculated number of complexes is not considered for further analyzes, because of stronger statistical variability than expected. Further filtering is indicated at the respective passages.

It is the absence of accessible recommendations and systematic guidelines for effect size presentation, as much as an ignorance of the issues, which has hindered the spread of good statistical practice in the biological literature.

— Shinichi Nakagawa
(Nakagawa & Cuthill, 2007).

ESTIMATES OF EFFECT SIZE In the pursuit of transparent and comprehensible statistics in biological sciences, Nakagawa and Cuthill (2007) advocate the use of effect size and **CI** based statistical reporting. The estimate of an effect size gives an impression of the magnitude of an effect and the corresponding **CI** gives an impression of the precision of this estimate. Finally, the p-value tells if the observed estimate is significant or not. Together, these measures enable a better judgement of differences between experimental groups, while p-values alone, as in classical statistics, allow for dichotomous decisions only (Nakagawa & Cuthill, 2007).

The slope of a linear model can be an estimate of the effect size (Nakagawa & Cuthill, 2007). Linear models are the main statistical tool of this thesis and enable to calculate the effect of multiple (continuous or discrete) input variables on the response variable as well as their interactions. The slope can be interpreted as the change of the response variable for a unit change in the predictor compared to the reference level (Nakagawa & Cuthill, 2007). The estimate of one predictor is calculated for the scenario when all other predictors are held constant at their average level. Generally, linear models are robust regarding violations of distributional assumptions (Schielzeth et al., 2020). Linear models were calculated using `lm()` from the `stats`-package in R 4.3.1⁶. This blog post gives a short yet comprehensive overview of the implementation of linear models in R and how to interpret their results.

The standard way of depicting estimates and their **CIs** from linear models are forest plots. The R package `sjPlot` comprises an implementation of **forest plots**. They show the estimates and **CIs** on the x-axis and the predictors on the y-axis, where an * symbol in-

⁵ The software is available upon request from github.com/LangeTo.

⁶ The code is available upon request from github.com/LangeTo.

Table 9.4: R and python packages used for data analysis and data visualization.

Package	Version
R packages.	
arm	1.13-1
cowplot	1.1.3
dplyr	1.1.2
furrr	0.3.1
future	1.33.0
ggforce	0.4.1
ggplot2	3.4.4
ggpubr	0.6.0
ggeffects	1.3.1
gggridges	0.5.4
ggupset	0.3.0
gridExtra	2.3
purrr	1.0.1
readr	2.1.4
readxl	1.4.3
scales	1.2.1
see	0.8.2
sjPlot	2.8.15
stringr	1.5.0
superb	0.95.7
themeFreecastle	0.3.3
tidyverse	1.3.0
tikzDevice	0.12.5
viridis	0.6.4
viridisLite	0.4.2
python packages.	
matplotlib	3.7.1
mpmath	1.2.1
numpy	1.23.5
pandas	1.5.2
seaborn	0.12.2
sympy	1.11.1

dicates the interaction of two input variables. When the CI of an estimate crosses $x = 0$, this effect is not considered to be significant. On the other hand, when the CI of the estimate does not cross $x = 0$, this effect is considered to be statistical significant. The calculated linear models are shown in the captions of the forest plots in the following style: `lm(response_variable ~ predictor_variable)`. However, some linear models used here calculated an estimate for two predictor variables: `lm(response_variable ~ predictor_variable1 + predictor_variable2)`; and some other linear models calculated an estimate for the interaction of two predictor variables: `lm(response_variable ~ predictor_variable1 * predictor_variable2)`.

Significance levels were indicated as follows: *** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05. No asterisks indicated an insignificant estimate.

STANDARDIZATION The estimate of an effect size depends on the scale of the predictor. For linear models with multiple predictors, the estimates might thus not be comparable. Standardization (z-transformation) is a method to put the predictors on the same scale and thus to maintain comparability (Schielzeth, 2010). This blog post gives an overview of how standardization can be implemented in R and how estimates can be interpreted because interpretation of standardized estimates is more complex (Nakagawa & Cuthill, 2007). The estimate now defines the change of the response variable for a 1 SD change of the predictor variable.

$$z = \frac{x - \bar{x}}{\text{SD}}, \quad (9.2)$$

where \bar{x} is the mean of the group.

In the presence of binary predictors, continuous predictors were standardized by dividing by two SDs (Gelman, 2008). Here, the estimate defines the change of the response variable for a 2 SD change of the predictor variable.

$$z_{2\text{SD}} = \frac{x - \bar{x}}{2\text{SD}}, \quad (9.3)$$

Standardization of predictors is indicated at the respective passages.

CALCULATION OF LIMIT OF DETECTION AND QUANTIFICATION OF ANALYTICAL METHODS LOD and LOQ were calculated using the equations suggested by European Medicines Agency (2023):

$$\text{LOD} = \frac{3.3\text{SD}}{\text{slope of linear fit}} \quad (9.4)$$

$$\text{LOQ} = \frac{10\text{SD}}{\text{slope of linear fit}} \quad (9.5)$$

10

MATERIALS

Table 10.1: Sequences of labels for DNA-only PICO. P8 and BLA label 2 were purchased from biomers.net. The artificial complex was purchased from metabion¹.

Name	Sequence (5'→3')
P8 label 2	GGTGACGATCCCGAAAAGCGGCCTTAACCTCCCT GCAAGCCTCAGCGACC
BLA label 2	CGCGCCACATAGCAGAACTTAAAAGTGCTCATCA TTGGAAAACGTTCTCGGGCGACCTCTACAATGG
Artificial couplex	CGCGCCACATAGCAGAACTTAAAAGTGCTCATCA TTGGAAAACGTTCTCGGGCGACCTCTACAATGG ATGCGCAGGTGACGATCCCGAAAAGCGGCCTTA ACTCCCTGCAAGCCTCAGCGACC

¹ The artificial couplex was 2-fold HPLC-purified. The same molecule was purchased from biomers.net without 2-fold HPLC-purification. Results were not reliable (data not shown).

Table 10.2: Sequences of labels for antibody labeling (Section 9.2.3). The labels were purchased from biomers.net. N is a placeholder for any nucleotide. The stretch of Ns is a barcode.

Name	Sequence (5'→3')
P8	TTTTGGTGACGATCCCGAAAAGNNNNNNNNNNCG GCCTTAACTCCCTGCAAGCCTCAGCGACC
BLA	TTTTGGTGACGATCCCGAAAATNNNNNNNNNNCC AATGATGAGCACTTTTGCAAGCCTCAGCGACC
NOS6	TTTTGGTGACGATCCCGAAAATNNNNNNNNNNNCA CCTACCGGCCCTGCAAGCCTCAGCGACC
ORC7	TTTTGGTGACGATCCCGAAAATNNNNNNNNNNCC CTCCTAGTCCCCTGCAAGCCTCAGCGACC

Table 10.3: Recipes for the preparation of master mix for the used dPCR systems (Section 9.2.7).

Ingredient	Final concentration	Manufacturer	Catalog number
Recipe for QX200 dPCR system (Bio-Rad).			
ddPCR Supermix for Probes (No dUTP)	1X	Bio-Rad	1863024
Forward and reverse primers	800 nmol L ⁻¹	Table 10.8	
Probes	250 nmol L ⁻¹	Table 10.7	
Sample	1 µL		
DNase-/RNase-free, distilled H ₂ O	ad 23 µL	Invitrogen	10977035
Recipe for naica dPCR system (Stillा).			
PerfeCTa Multiplex qPCR ToughMix	1X or 2X	Quantabio	95147-01K
Forward and reverse primers	800 nmol L ⁻¹	Table 10.8	
Probes	250 nmol L ⁻¹	Table 10.7	
100X fluorescein (prepared according to Stillा's instructions)	1X	Sigma-Aldrich	46955
Sample	1 µL		
DNase-/RNase-free, distilled H ₂ O	ad 23 µL	Invitrogen	10977035
Recipe for QIAcuity dPCR system (QIAGEN).			
QIAcuity Probe PCR Kit	1X	QIAGEN	250101
Forward and reverse primers	800 nmol L ⁻¹	Table 10.8	
Probes	400 nmol L ⁻¹	Table 10.7	
Sample	1 µL		
DNase-/RNase-free, distilled H ₂ O	ad 42 µL	Invitrogen	10977035

Table 10.4: Monoclonal antibodies against 4EBP₁ protoforms.

Antibody name	Clone	Antigen	Immunogen	Host	Manufacturer	Catalog number
4EBP ₁ monoclonal antibody	554R16	4EBP ₁	Full-length protein produced in <i>E. coli</i>	Mouse	Invitrogen	AHO1382
4EBP ₁ monoclonal antibody	60246-1-Ig	4EBP ₁	Full-length protein produced in <i>E. coli</i>	Mouse	Proteintech	60246-1-Ig
EIF4EBP ₁ monoclonal antibody	4F3-H2	4EBP ₁	Full-length protein produced in <i>E. coli</i>	Mouse	Abnova	H00001978-M01
Monoclonal anti-EIF4EBP ₁ antibody produced in mouse	1F7	4EBP ₁	Full-length protein produced in <i>E. coli</i>	Mouse	Sigma-Aldrich	SAB1403761
Nonphospho-4EBP ₁ (Thr46) monoclonal antibody	E.638.3	4EBP ₁ -non-phospho-T46	Synthetic peptide around Thr46	Rabbit	Invitrogen	MA5-15032
Phospho-4EBP ₁ (Thr37, Thr46) recombinant rabbit monoclonal antibody	4EB1T37T46-A5	4EBP ₁ -phospho-T37T46	Synthetic phospho-peptide around Thr37/Thr46	Rabbit	Invitrogen	MA5-36935

Table 10.5: Monoclonal antibodies against various further targets. The TTZ, PTZ and tocilizumab antibodies were a kind gift from the pharmacy of the Medical Center – University of Freiburg.

Antibody name	Clone	Antigen	Manufacturer	Catalog number
Antibodies against His-tag.				
6*His, his-tag monoclonal antibody	1B7G5	6*His	Proteintech	66005-1-Ig
Penta-his antibody, BSA-free	34660	5*His	QIAGEN	34660
Antibodies against HER2.				
Pertuzumab (Perjeta®)	n/a	HER2	Roche	n/a
Trastuzumab (Herceptin®)	n/a	HER2	Roche	n/a
Antibodies against different targets, i.e. URABs.				
Tocilizumab (RoActemra®)	n/a	IL6R	Roche	n/a
Anti-RPS6KB1 antibody produced in mouse	zC2	S6K1	Sigma-Aldrich	SAB1412617
Antibodies against eIF4E.				
Anti-eIF4E antibody [Y449]	Y449	eIF4E	Abcam	ab33768
eIF4E Antibody (299910) [Unconjugated]	299910	eIF4E	R&D Systems	MAB3228
eIF4E Antibody (P-2)	P-2	eIF4E	Santa Cruz Biotechnology	sc-9976
eIF4E Monoclonal Antibody (5D11)	5D11	eIF4E	Invitrogen	MA1-089
eIF4E Mouse anti-Human	PCRP-EIF4E-1D3	eIF4E	Neo-Biotechnologies	NBT-1977-MSM1-P1ABX

Table 10.6: Labeled antibodies use in this thesis. The concentration was determined by dPCR after labeling (Section 9.2.3), thus only the labeled fraction could be detected.

Clone	Label	LE (%)	Concentration (molecules/ μL)
Antibodies against 4EBP1.			
1F7	ORC7	95.6	5.96×10^{12}
4EB1T37T46-A5	P8	65.5	8.76×10^{11}
4F3-H2	NOS6	76.9	7.15×10^{12}
4F3-H2	ORC7	90.6	9.98×10^{12}
554R16	BLA	52.1	7.40×10^{12}
554R16	P8	31.4	5.00×10^{12}
60246-1-Ig	BLA	63.1	8.24×10^{12}
E.638.3	NOS6	66.6	7.79×10^{10}
Antibodies against His-tag.			
34660	BLA	86.7	1.02×10^{11}
1B7G5	ORC7	84.2	8.73×10^{11}
Antibodies against HER2.			
PTZ	NOS6	88.7	1.48×10^{10}
PTZ	ORC7	99.2	5.84×10^{11}
PTZ	P8	65.7	8.12×10^{10}
TTZ	BLA	91.6	5.84×10^{10}
TTZ	NOS6	87.0	7.17×10^{11}
TTZ	ORC7	65.1	3.19×10^{11}
Antibodies against other targets, i.e. URABs.			
2C2	BLA	n/a	4.93×10^{11}
2C2	NOS6	n/a	8.96×10^{10}
Tocilizumab	P8	n/a	2.37×10^{11}
Antibodies against eIF4E.			
Y449	BLA	76.9	6.80×10^{11}
299910	P8	30.0	1.19×10^{13}
P-2	ORC7	37.0	2.95×10^{11}
5D11	P8	69.9	2.12×10^{11}
PCRP-EIF4E-1D3	BLA	85.6	6.08×10^{12}

Table 10.7: Sequences of probes and their modifications, purchased from Eurofins Genomics.

Name	Sequence (5'→3')	Purpose	5' Dye	3' Modifi- cation
P8 probe	CGGCCTTT AACTCC	Detection of P8 label, P8 label 2 and artificial couplex	FAM	MGB- Eclipse
BLA probe	CAATGAT GAGCACT TTT	Detection of BLA label, BLA label 2 and artificial couplex	HEX	MGB- Eclipse
NOS6 probe	ACCTACC GGCCTCC	Detection of NOS6 label	Atto550	MGB- Eclipse
ORC7 probe	CCTCCTA GTTCCCC	Detection of ORC7 label	TxRed	MGB- Eclipse

Table 10.8: Sequences of forward (fwd) and reverse (rev) primers for the amplification of the labels (Table 10.1 and Table 10.2). All primers were purchased from biomers.net. Melting temperature T_m was estimated using OligoCalc (Kibbe, 2007).

Name	Sequence (5' → 3')	Purpose	Length	GC-content (%)	T_m (°C)
Fwd primer 1	GGTGACGATCCCCGCAAAA	Amplification of P8, P8 label 2, BLA, NOS6, ORC7 label and artificial complex	18	56	50.3
Rev primer 1	GGTCGCTGAGGCTTGCA	Amplification of P8, P8 label 2, BLA, NOS6, ORC7 label and artificial complex	17	65	51.9
Fwd primer 2	CGCCCCGAAGAACGTTT	Amplification of BLA label 2 and artificial complex	17	59	49.5
Rev primer 2	CGGCCACATAGCAGAACT T	Amplification of BLA label 2 and artificial complex	20	55	53.8

BIBLIOGRAPHY

- Abasianik, MF, Wolfe, K, Van Phan, H, Lin, J, Laxman, B, White, SR, Verhoef, PA, Mutlu, GM, Patel, B, & Tay, S. (2020). Ultrasensitive digital quantification of cytokines and bacteria predicts septic shock outcomes. *Nature Communications*, 11(1), 2607. <https://doi.org/10.1038/s41467-020-16124-9>
- Abcam. (n.d.). Kd value: A quantitative measurement of antibody affinity. Retrieved May 13, 2024, from <https://www.abcam.com/primary-antibodies/kd-value-a-quantitative-measurement-of-antibody-affinity>
- Actome GmbH. (2024). The PICO Handbook. Retrieved January 29, 2024, from https://www.actome.de/downloads/PICO_Handbook.pdf
- Aebersold, R, Agar, JN, Amster, IJ, Baker, MS, Bertozzi, CR, Boja, ES, Costello, CE, Cravatt, BF, Fenselau, C, Garcia, BA, Ge, Y, Gunawardena, J, Hendrickson, RC, Hergenrother, PJ, Huber, CG, Ivanov, AR, Jensen, ON, Jewett, MC, Kelleher, NL, ... Zhang, B. (2018). How many human proteoforms are there? *Nature Chemical Biology*, 14(3), 206–214. <https://doi.org/10.1038/nchembio.2576>
- Agrawal, S, Hofmann, WK, Tidow, N, Ehrlich, M, Boom, DVD, Koschmieder, S, Berdel, WE, Serve, H, & Müller-Tidow, C. (2007). The C/EBP δ tumor suppressor is silenced by hypermethylation in acute myeloid leukemia. *Blood*, 109(9), 3895–3905. <https://doi.org/10.1182/blood-2006-08-040147>
- Al-Ali, HK, Jaekel, N, & Niederwieser, D. (2014). The role of hypomethylating agents in the treatment of elderly patients with AML. *Journal of Geriatric Oncology*, 5(1), 89–105. <https://doi.org/10.1016/j.jgo.2013.08.004>
- Albayrak, C, Jordi, CA, Zechner, C, Lin, J, Bichsel, CA, Khammash, M, & Tay, S. (2016). Digital Quantification of Proteins and mRNA in Single Mammalian Cells. *Molecular Cell*, 61(6), 914–924. <https://doi.org/10.1016/j.molcel.2016.02.030>
- Álvarez-Fernández, R. (2013). Explanatory chapter: PCR primer design. In *Methods in Enzymology* (pp. 1–21, Vol. 529). Elsevier. <https://doi.org/10.1016/B978-0-12-418687-3.00001-X>
- Anthis, NJ, & Clore, GM. (2013). Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science*, 22(6), 851–858. <https://doi.org/10.1002/pro.2253>
- antibodies.com. (2023). Price Comparison. Retrieved May 13, 2024, from <https://www.antibodies.com/price-comparison>

- Aralar, A, Yuan, Y, Chen, K, Geng, Y, Ortiz Velez, D, Sinha, M, Lawrence, SM, & Fraley, SI. (2020). Improving Quantitative Power in Digital PCR through Digital High-Resolution Melting (MJ Loeffelholz, Ed.). *Journal of Clinical Microbiology*, 58(6), e00325–20. <https://doi.org/10.1128/JCM.00325-20>
- Arber, DA, Orazi, A, Hasserjian, R, Thiele, J, Borowitz, MJ, Le Beau, MM, Bloomfield, CD, Cazzola, M, & Vardiman, JW. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, 127(20), 2391–2405. <https://doi.org/10.1182/blood-2016-03-643544>
- Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC Ped-Brain, Alexandrov, LB, Nik-Zainal, S, Wedge, DC, Aparicio, SAJR, Behjati, S, Biankin, AV, Bignell, GR, Bolli, N, Borg, A, Børresen-Dale, AL, Boyault, S, Burkhardt, B, Butler, AP, Caldas, C, Davies, HR, ... Stratton, MR. (2013). Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415–421. <https://doi.org/10.1038/nature12477>
- Aydin, S. (2015). A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides*, 72, 4–15. <https://doi.org/10.1016/j.peptides.2015.04.012>
- Baek, D, Villén, J, Shin, C, Camargo, FD, Gygi, SP, & Bartel, DP. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209), 64–71. <https://doi.org/10.1038/nature07242>
- Bakhtiar, R. (2013). Surface Plasmon Resonance Spectroscopy: A Versatile Technique in a Biochemist's Toolbox. *Journal of Chemical Education*, 90(2), 203–209. <https://doi.org/10.1021/ed200549g>
- Barlow, DJ, Edwards, MS, & Thornton, JM. (1986). Continuous and discontinuous protein antigenic determinants. *Nature*, 322(6081), 747–748. <https://doi.org/10.1038/322747ao>
- Basu, AS. (2017). Digital Assays Part I: Partitioning Statistics and Digital PCR. *SLAS Technology*, 22(4), 369–386. <https://doi.org/10.1177/2472630317705680>
- Benita, Y, Oosting, RS, Lok, MC, Wise, MJ, & Humphery-Smith, I. (2003). Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Research*, 31(16), 99e–99. <https://doi.org/10.1093/nar/gng101>
- Bernier, SC, Cantin, L, & Salesse, C. (2018). Systematic analysis of the expression, solubility and purification of a passenger protein in fusion with different tags. *Protein Expression and Purification*, 152, 92–106. <https://doi.org/10.1016/j.pep.2018.07.007>
- Biassoni, R, & Raso, A (Eds.). (2014). *Quantitative Real-Time PCR: Methods and Protocols* (Vol. 1160). Springer New York. <https://doi.org/10.1007/978-1-4939-0733-5>

- Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3), 89–95. <https://doi.org/10.1067/mcp.2001.113989>
- Bjellqvist, B, Ek, K, Giorgio Righetti, P, Gianazza, E, Görg, A, Westermeier, R, & Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*, 6(4), 317–339. [https://doi.org/10.1016/0165-022X\(82\)90013-6](https://doi.org/10.1016/0165-022X(82)90013-6)
- Blazek, M, Betz, C, Hall, MN, Reth, M, Zengerle, R, & Meier, M. (2013). Proximity Ligation Assay for High-content Profiling of Cell Signaling Pathways on a Microfluidic Chip. *Molecular & Cellular Proteomics*, 12(12), 3898–3907. <https://doi.org/10.1074/mcp.M113.032821>
- Blazek, M, Santisteban, TS, Zengerle, R, & Meier, M. (2015). Analysis of fast protein phosphorylation kinetics in single cells on a microfluidic chip. *Lab on a Chip*, 15(3), 726–734. <https://doi.org/10.1039/C4LC00797B>
- Bludau, I, & Aebersold, R. (2020). Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nature Reviews Molecular Cell Biology*, 21(6), 327–340. <https://doi.org/10.1038/s41580-020-0231-2>
- Brink, BG, Meskas, J, & Brinkman, RR. (2018). ddPCRclust: An R package and Shiny app for automated analysis of multiplexed ddPCR data (J Wren, Ed.). *Bioinformatics*, 34(15), 2687–2689. <https://doi.org/10.1093/bioinformatics/bty136>
- Bryant, P, Pozzati, G, & Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, 13(1), 1265. <https://doi.org/10.1038/s41467-022-28865-w>
- Burdukiewicz, M, Rödiger, S, Sobczyk, P, Menschikowski, M, Schierack, P, & Mackiewicz, P. (2016). Methods for comparing multiple digital PCR experiments. *Biomolecular Detection and Quantification*, 9, 14–19. <https://doi.org/10.1016/j.bdq.2016.06.004>
- Buus, S, Rockberg, J, Forsström, B, Nilsson, P, Uhlen, M, & Schafer-Nielsen, C. (2012). High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays. *Molecular & Cellular Proteomics*, 11(12), 1790–1800. <https://doi.org/10.1074/mcp.M112.020800>
- Byrnes, SA, Huynh, T, Chang, TC, Anderson, CE, McDermott, JJ, Oncina, CI, Weigl, BH, & Nichols, KP. (2020). Wash-Free, Digital Immunoassay in Polydisperse Droplets. *Analytical Chemistry*, 92(5), 3535–3543. <https://doi.org/10.1021/acs.analchem.9b02526>
- Calabrese, S, Markl, AM, Neugebauer, M, Krauth, SJ, Borst, N, Von Stetten, F, & Lehnert, M. (2023). Reporter emission multiplex-

- ing in digital PCRs (REM-dPCRs): Direct quantification of multiple target sequences per detection channel by population specific reporters. *The Analyst*, 148(20), 5243–5254. <https://doi.org/10.1039/D3AN00191A>
- Calderón-Celis, F, Encinar, JR, & Sanz-Medel, A. (2018). Standardization approaches in absolute quantitative proteomics with mass spectrometry. *Mass Spectrometry Reviews*, 37(6), 715–737. <https://doi.org/10.1002/mas.21542>
- Candia, J, Daya, GN, Tanaka, T, Ferrucci, L, & Walker, KA. (2022). Assessment of variability in the plasma 7k SomaScan proteomics assay. *Scientific Reports*, 12(1), 17147. <https://doi.org/10.1038/s41598-022-22116-o>
- Capdeville, R, Silberman, S, & Dimitrijevic, S. (2002). Imatinib: The first 3 years. *European Journal of Cancer*, 38, S77–S82. [https://doi.org/10.1016/S0959-8049\(02\)80607-4](https://doi.org/10.1016/S0959-8049(02)80607-4)
- Cassidy, L, Kaulich, PT, & Tholey, A. (2023). Proteoforms expand the world of microproteins and short open reading frame-encoded peptides. *iScience*, 26(2), 106069. <https://doi.org/10.1016/j.isci.2023.106069>
- Chaffey, B, & Silmon, A. (2016). Biomarkers in personalized medicine: Discovery and delivery. *The Biochemist*, 38(1), 43–47. <https://doi.org/10.1042/BIO03801043>
- Chan, BM, Badh, A, Berry, KA, Grauer, SA, & King, CT. (2018). Flow Cytometry-Based Epitope Binning Using Competitive Binding Profiles for the Characterization of Monoclonal Antibodies against Cellular and Soluble Protein Targets. *SLAS Discovery*, 23(7), 613–623. <https://doi.org/10.1177/2472555218774334>
- Chang, L, Rissin, DM, Fournier, DR, Piech, T, Patel, PP, Wilson, DH, & Duffy, DC. (2012). Single molecule enzyme-linked immunosorbent assays: Theoretical considerations. *Journal of Immunological Methods*, 378(1-2), 102–115. <https://doi.org/10.1016/j.jim.2012.02.011>
- Chio, TI, & Bane, SL. (2020). Click Chemistry Conjugations [Series Title: Methods in Molecular Biology]. In LN Tumei (Ed.), *Antibody Drug Conjugates* (pp. 83–97, Vol. 2078). Springer US. https://doi.org/10.1007/978-1-4939-9929-3_6
- Cohen, L, Cui, N, Cai, Y, Garden, PM, Li, X, Weitz, DA, & Walt, DR. (2020). Single Molecule Protein Detection with Attomolar Sensitivity Using Droplet Digital Enzyme-Linked Immunosorbent Assay. *ACS Nano*, 14(8), 9491–9501. <https://doi.org/10.1021/acsnano.ac02378>
- Collins, FS, & Varmus, H. (2015). A New Initiative on Precision Medicine. *New England Journal of Medicine*, 372(9), 793–795. <https://doi.org/10.1056/NEJMmp1500523>
- Coons, AH, Creech, HJ, & Jones, RN. (1941). Immunological Properties of an Antibody Containing a Fluorescent Group. *Experimental*

- Biology and Medicine*, 47(2), 200–202. <https://doi.org/10.3181/00379727-47-13084P>
- Cox, J., & Mann, M. (2011). Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annual Review of Biochemistry*, 80(1), 273–299. <https://doi.org/10.1146/annurev-biochem-061308-093216>
- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227(5258), 561–563. <https://doi.org/10.1038/227561ao>
- Da Silva Araújo, J., & Lima Tavares Machado, O. (2020). Proteoforms: General Concepts and Methodological Process for Identification. In X Zhan (Ed.), *Proteoforms - Concept and Applications in Medical Sciences*. IntechOpen. <https://doi.org/10.5772/intechopen.89914>
- Darmanis, S, Nong, RY, Vänelid, J, Siegbahn, A, Ericsson, O, Fredriksson, S, Bäcklin, C, Gut, M, Heath, S, Gut, IG, Wallentin, L, Gustafsson, MG, Kamali-Moghaddam, M, & Landegren, U. (2011). ProteinSeq: High-Performance Proteomic Analyses by Proximity Ligation and Next Generation Sequencing (J Rao, Ed.). *PLoS ONE*, 6(9), e25583. <https://doi.org/10.1371/journal.pone.0025583>
- Daskalakis, M, Nguyen, TT, Nguyen, C, Guldberg, P, Köhler, G, Wijermans, P, Jones, PA, & Lübbert, M. (2002). Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2'-deoxycytidine (decitabine) treatment. *Blood*, 100(8), 2957–2964. <https://doi.org/10.1182/blood.V100.8.2957>
- De Falco, A, Olinger, CM, Klink, B, Mittelbronn, M, & Stieber, D. (2023). Digital PCR cluster predictor: A universal R-package and shiny app for the automated analysis of multiplex digital PCR data (J Wren, Ed.). *Bioinformatics*, 39(5), btad282. <https://doi.org/10.1093/bioinformatics/btad282>
- Deininger, MW, Goldman, JM, Lydon, N, & Melo, JV. (1997). The Tyrosine Kinase Inhibitor CGP57148B Selectively Inhibits the Growth of BCR-ABL-Positive Cells. *Blood*, 90(9), 3691–3698. <https://doi.org/10.1182/blood.V90.9.3691>
- Dekker, MJHJ, De Vries, ST, Versantvoort, CHM, Drost-van Velze, EGE, Bhatt, M, Van Meer, PJK, Havinga, IK, Gispen-de Wied, CC, & Mol, PGM. (2021). Sex Proportionality in Pre-clinical and Clinical Trials: An Evaluation of 22 Marketing Authorization Application Dossiers Submitted to the European Medicines Agency. *Frontiers in Medicine*, 8, 643028. <https://doi.org/10.3389/fmed.2021.643028>
- Di Meo, A, Bartlett, J, Cheng, Y, Pasic, MD, & Yousef, GM. (2017). Liquid biopsy: A step forward towards precision medicine in urologic malignancies. *Molecular Cancer*, 16(1), 80. <https://doi.org/10.1186/s12943-017-0644-5>

- Diezel, W, Kopperschläger, G, & Hofmann, E. (1972). An improved procedure for protein staining in polyacrylamide gels with a new type of Coomassie Brilliant Blue. *Analytical Biochemistry*, 48(2), 617–620. [https://doi.org/10.1016/0003-2697\(72\)90117-0](https://doi.org/10.1016/0003-2697(72)90117-0)
- Döhner, H, Estey, E, Grimwade, D, Amadori, S, Appelbaum, FR, Büchner, T, Dombret, H, Ebert, BL, Fenaux, P, Larson, RA, Levine, RL, Lo-Coco, F, Naoe, T, Niederwieser, D, Ossenkoppele, GJ, Sanz, M, Sierra, J, Tallman, MS, Tien, HF, ... Bloomfield, CD. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*, 129(4), 424–447. <https://doi.org/10.1182/blood-2016-08-733196>
- Döhner, H, Weisdorf, DJ, & Bloomfield, CD. (2015). Acute Myeloid Leukemia (DL Longo, Ed.). *New England Journal of Medicine*, 373(12), 1136–1152. <https://doi.org/10.1056/NEJMra1406184>
- Domon, B, & Aebersold, R. (2006). Mass Spectrometry and Protein Analysis. *Science*, 312(5771), 212–217. <https://doi.org/10.1126/science.1124619>
- Dou, D, Revol, R, Östbye, H, Wang, H, & Daniels, R. (2018). Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology*, 9, 1581. <https://doi.org/10.3389/fimmu.2018.01581>
- Douglass, EF, Miller, CJ, Sparer, G, Shapiro, H, & Spiegel, DA. (2013). A Comprehensive Mathematical Model for Three-Body Binding Equilibria. *Journal of the American Chemical Society*, 135(16), 6092–6099. <https://doi.org/10.1021/ja311795d>
- Duffy, DC. (2023). Digital detection of proteins. *Lab on a Chip*, 23(5), 818–847. <https://doi.org/10.1039/D2LC00783E>
- European Medicines Agency. (2023). ICH Q2(R2) Guideline on validation of analytical procedures.
- FDA-NIH Biomarker Working Group. (2016). *BEST (Biomarkers, End-pointS, and other Tools) Resource*. Food; Drug Administration (US). Retrieved December 3, 2023, from <http://www.ncbi.nlm.nih.gov/books/NBK326791/>
- Ferrara, F, & Musto, P. (2011). Hypomethylating agents for the treatment of acute myeloid leukemia in the elderly: For all, none, or which patients? *Cancer*, 117(17), 3879–3881. <https://doi.org/10.1002/cncr.25934>
- Fletcher, CM, & Wagner, G. (1998). The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Science*, 7(7), 1639–1642. <https://doi.org/10.1002/pro.5560070720>
- Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437(1-2), 55–75. <https://doi.org/10.1002/andp.19484370105>

- Fredriksson, S, Dixon, W, Ji, H, Koong, AC, Mindrinos, M, & Davis, RW. (2007). Multiplexed protein detection by proximity ligation for cancer biomarker validation. *Nature Methods*, 4(4), 327–329. <https://doi.org/10.1038/nmeth1020>
- Fredriksson, S, Gullberg, M, Jarvius, J, Olsson, C, Pietras, K, Gústafsdóttir, SM, Östman, A, & Landegren, U. (2002). Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*, 20(5), 473–477. <https://doi.org/10.1038/nbt0502-473>
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27(15), 2865–2873. <https://doi.org/10.1002/sim.3107>
- Genshaft, AS, Li, S, Gallant, CJ, Darmanis, S, Prakadan, SM, Ziegler, CGK, Lundberg, M, Fredriksson, S, Hong, J, Regev, A, Livak, KJ, Landegren, U, & Shalek, AK. (2016). Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biology*, 17(1), 188. <https://doi.org/10.1186/s13059-016-1045-6>
- Gibney, ER, & Nolan, CM. (2010). Epigenetics and gene expression. *Heredity*, 105(1), 4–13. <https://doi.org/10.1038/hdy.2010.54>
- Gilboa, T, Maley, AM, Ogata, AF, Wu, C, & Walt, DR. (2021). Sequential Protein Capture in Multiplex Single Molecule Arrays: A Strategy for Eliminating Assay Cross-Reactivity. *Advanced Healthcare Materials*, 10(4), 2001111. <https://doi.org/10.1002/adhm.202001111>
- Gimeno, M, Sada Del Real, K, & Rubio, A. (2023). Precision oncology: A review to assess interpretability in several explainable methods. *Briefings in Bioinformatics*, 24(4), bbad200. <https://doi.org/10.1093/bib/bbad200>
- Gingras, AC, Gygi, SP, Raught, B, Polakiewicz, RD, Abraham, RT, Hoekstra, MF, Aebersold, R, & Sonenberg, N. (1999). Regulation of 4E-BP1 phosphorylation: A novel two-step mechanism. *Genes & Development*, 13(11), 1422–1437. <https://doi.org/10.1101/gad.13.11.1422>
- Gingras, AC, Raught, B, Gygi, SP, Niedzwiecka, A, Miron, M, Burley, SK, Polakiewicz, RD, Wyslouch-Cieszynska, A, Aebersold, R, & Sonenberg, N. (2001). Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes & Development*, 15(21), 2852–2864. <https://doi.org/10.1101/gad.912401>
- Gold, L, Walker, JJ, Wilcox, SK, & Williams, S. (2012). Advances in human proteomics at high scale with the SOMAscan proteomics platform. *New Biotechnology*, 29(5), 543–549. <https://doi.org/10.1016/j.nbt.2011.11.016>
- Gong, H, Holcomb, I, Ooi, A, Wang, X, Majonis, D, Unger, MA, & Ramakrishnan, R. (2016). Simple Method To Prepare Oligonucleotide-Conjugated Antibodies and Its Application in Multiplex Pro-

- tein Detection in Single Cells. *Bioconjugate Chemistry*, 27(1), 217–225. <https://doi.org/10.1021/acs.bioconjchem.5b00613>
- Gordon, JA. (1991). [41] Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. In *Methods in Enzymology* (pp. 477–482, Vol. 201). Elsevier. [https://doi.org/10.1016/0076-6879\(91\)01043-2](https://doi.org/10.1016/0076-6879(91)01043-2)
- Gorr, TA, & Vogel, J. (2015). Western blotting revisited: Critical perusal of underappreciated technical issues. *PROTEOMICS – Clinical Applications*, 9(3-4), 396–405. <https://doi.org/10.1002/prca.201400118>
- Greenman, C, Stephens, P, Smith, R, Dalgliesh, GL, Hunter, C, Bignell, G, Davies, H, Teague, J, Butler, A, Stevens, C, Edkins, S, O'Meara, S, Vastrik, I, Schmidt, EE, Avis, T, Barthorpe, S, Bhamra, G, Buck, G, Choudhury, B, ... Stratton, MR. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132), 153–158. <https://doi.org/10.1038/nature05610>
- Grimsley, GR, & Pace, CN. (2003). Spectrophotometric Determination of Protein Concentration. *Current Protocols in Protein Science*, 33(1). <https://doi.org/10.1002/0471140864.ps0301s33>
- Grizzle, WE, Otali, D, Sexton, KC, & Atherton, DS. (2016). Effects of Cold Ischemia on Gene Expression: A Review and Commentary. *Biopreservation and Biobanking*, 14(6), 548–558. <https://doi.org/10.1089/bio.2016.0013>
- Gross, T, Hundertmark, T, Csiszar, V, Sulyok, AA, Gross, N, Breiden, M, Kitschen, N, Von Groll, U, Niemoller, C, Sanchez-Martin, P, Hein, A, Gopffert, J, Szoradi, T, Luebbert, P, Koltay, P, Porschewski, P, Von Stetten, F, Zengerle, R, & Jeney, C. (2024). *Lossless Single-Molecule Counting To Absolute Quantify Proteoforms* (preprint). Molecular Biology. <https://doi.org/10.1101/2024.03.19.585761>
- Gross, T, Jeney, C, Halm, D, Finkenzeller, G, Stark, GB, Zengerle, R, Koltay, P, & Zimmermann, S. (2021). Characterization of CRISPR/Cas9 RANKL knockout mesenchymal stem cell clones based on single-cell printing technology and Emulsion Coupling assay as a low-cellularity workflow for single-cell cloning (G Papaccio, Ed.). *PLOS ONE*, 16(3), e0238330. <https://doi.org/10.1371/journal.pone.0238330>
- Gustafsdottir, SM, Schallmeiner, E, Fredriksson, S, Gullberg, M, Söderberg, O, Jarvius, M, Jarvius, J, Howell, M, & Landegren, U. (2005). Proximity ligation assays for sensitive and specific protein analyses. *Analytical Biochemistry*, 345(1), 2–9. <https://doi.org/10.1016/j.ab.2005.01.018>
- Gygi, SP, Rochon, Y, Franz, BR, & Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. *Molecular and Cellular Biology*, 19(3), 1720–1730. <https://doi.org/10.1128/MCB.19.3.1720>

- Han, B. (2020). A suite of mathematical solutions to describe ternary complex formation and their application to targeted protein degradation by heterobifunctional ligands. *Journal of Biological Chemistry*, 295(45), 15280–15291. <https://doi.org/10.1074/jbc.RA120.014715>
- Hattori, T, & Koide, S. (2018). Next-generation antibodies for post-translational modifications. *Current Opinion in Structural Biology*, 51, 141–148. <https://doi.org/10.1016/j.sbi.2018.04.006>
- Hattori, T, Taft, JM, Swist, KM, Luo, H, Witt, H, Slattery, M, Koide, A, Ruthenburg, AJ, Krajewski, K, Strahl, BD, White, KP, Farnham, PJ, Zhao, Y, & Koide, S. (2013). Recombinant antibodies to histone post-translational modifications. *Nature Methods*, 10(10), 992–995. <https://doi.org/10.1038/nmeth.2605>
- Heerboth, S, Lapinska, K, Snyder, N, Leary, M, Rollinson, S, & Sarkar, S. (2014). Use of Epigenetic Drugs in Disease: An Overview. *Genetics & Epigenetics*, 6, GEG.S12270. <https://doi.org/10.4137/GEG.S12270>
- Heesom, KJ, Gampel, A, Mellor, H, & Denton, RM. (2001). Cell cycle-dependent phosphorylation of the translational repressor eIF-4E binding protein-1 (4E-BP1). *Current Biology*, 11(17), 1374–1379. [https://doi.org/10.1016/S0960-9822\(01\)00422-5](https://doi.org/10.1016/S0960-9822(01)00422-5)
- Hermanson, GT. (2013). Antibody Modification and Conjugation. In *Bioconjugate Techniques* (pp. 867–920). Elsevier. <https://doi.org/10.1016/B978-0-12-382239-0.00020-0>
- Hess, JM, Bernards, A, Kim, J, Miller, M, Taylor-Weiner, A, Haradhvala, NJ, Lawrence, MS, & Getz, G. (2019). Passenger Hotspot Mutations in Cancer. *Cancer Cell*, 36(3), 288–301.e14. <https://doi.org/10.1016/j.ccr.2019.08.002>
- Hindson, BJ, Ness, KD, Masquelier, DA, Belgrader, P, Heredia, NJ, Makarewicz, AJ, Bright, IJ, Lucero, MY, Hiddessen, AL, Legler, TC, Kitano, TK, Hodel, MR, Petersen, JF, Wyatt, PW, Steenblock, ER, Shah, PH, Bousse, LJ, Troup, CB, Mellen, JC, ... Colston, BW. (2011). High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry*, 83(22), 8604–8610. <https://doi.org/10.1021/ac202028g>
- Hindson, CM, Chevillet, JR, Briggs, HA, Galichotte, EN, Ruf, IK, Hindson, BJ, Vessella, RL, & Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nature Methods*, 10(10), 1003–1005. <https://doi.org/10.1038/nmeth.2633>
- Hoffbrand, AV, & Moss, PAH. (2016). *Hoffbrand's essential haematology* (Seventh edition) [Section: ix, 369 pages : illustrations (chiefly color) ; 28 cm.]. Wiley Blackwell.
- Hong, JM, Gibbons, M, Bashir, A, Wu, D, Shao, S, Cutts, Z, Chavarha, M, Chen, Y, Schiff, L, Foster, M, Church, VA, Ching, L, Ahadi,

- S, Hieu-Thao Le, A, Tran, A, Dimon, M, Coram, M, Williams, B, Jess, P, ... Pawlosky, A. (2022). ProtSeq: Toward high-throughput, single-molecule protein sequencing via amino acid conversion into DNA barcodes. *iScience*, 25(1), 103586. <https://doi.org/10.1016/j.isci.2021.103586>
- Hou, H, He, H, & Wang, Y. (2020). Effects of SDS on the activity and conformation of protein tyrosine phosphatase from thermus thermophilus HB27. *Scientific Reports*, 10(1), 3195. <https://doi.org/10.1038/s41598-020-60263-4>
- Hrycaj, S. (2023). Immunohistochemistry: Origins, Tips, and a Look to the Future. Retrieved April 28, 2024, from <https://www.the-scientist.com/immunohistochemistry-origins-tips-and-a-look-to-the-future-71439>
- Hsieh, AC, Nguyen, HG, Wen, L, Edlind, MP, Carroll, PR, Kim, W, & Ruggero, D. (2015). Cell type-specific abundance of 4EBP1 primes prostate cancer sensitivity or resistance to PI3K pathway inhibitors. *Science Signaling*, 8(403). <https://doi.org/10.1126/scisignal.aad5111>
- Huang, Y, Zhang, Z, & Zhou, Y. (2022). AbAgIntPre: A deep learning method for predicting antibody-antigen interactions based on sequence information. *Frontiers in Immunology*, 13, 1053617. <https://doi.org/10.3389/fimmu.2022.1053617>
- Iqbal, N, & Iqbal, N. (2014). Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Molecular Biology International*, 2014, 1–9. <https://doi.org/10.1155/2014/852748>
- Jaiswal, S, & Ebert, BL. (2019). Clonal hematopoiesis in human aging and disease. *Science*, 366(6465), eaan4673. <https://doi.org/10.1126/science.aan4673>
- Janssen, K, Knez, K, Spasic, D, & Lammertyn, J. (2013). Nucleic Acids for Ultra-Sensitive Protein Detection. *Sensors*, 13(1), 1353–1384. <https://doi.org/10.3390/s130101353>
- Jeney, C. (2016). Method (WO2016/083793A1). Pilisszentlászló.
- Jeney, C. (2023a). High-efficacy labeling of antibodies with oligonucleotide labels (WO2023/084070A1). Freiburg.
- Jeney, C. (2023b). Quantitative detection method and corresponding device (EP4242656A1). Müllheim.
- Jeney, C, & Koltay, P. (2020). Method and kit for measuring of analytes in bi-component systems and uses therof (WO2020/260277A1). Freiburg.
- Jenkins, SH. (1992). Homogeneous enzyme immunoassay. *Journal of Immunological Methods*, 150(1-2), 91–97. [https://doi.org/10.1016/0022-1759\(92\)90067-4](https://doi.org/10.1016/0022-1759(92)90067-4)
- Jeong, H, Mason, SP, Barabási, AL, & Oltvai, ZN. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833), 41–42. <https://doi.org/10.1038/35075138>

- Jiang, YP, Ballou, LM, & Lin, RZ. (2001). Rapamycin-insensitive Regulation of 4E-BP1 in Regenerating Rat Liver. *Journal of Biological Chemistry*, 276(14), 10943–10951. <https://doi.org/10.1074/jbc.M007758200>
- Jothikumar, P, Hill, V, & Narayanan, J. (2009). Design of FRET-TaqMan probes for multiplex real-time PCR using an internal positive control. *BioTechniques*, 46(7), 519–524. <https://doi.org/10.2144/000113127>
- Juliusson, G, & Hough, R. (2016). Leukemia. In D Stark & G Vassal (Eds.), *Progress in Tumor Research* (pp. 87–100, Vol. 43). S. Karger AG. <https://doi.org/10.1159/000447076>
- Jüttermann, R, Li, E, & Jaenisch, R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proceedings of the National Academy of Sciences*, 91(25), 11797–11801. <https://doi.org/10.1073/pnas.91.25.11797>
- Kaliyaraj Selva Kumar, A, Zhang, Y, Li, D, & Compton, RG. (2020). A mini-review: How reliable is the drop casting technique? *Electrochemistry Communications*, 121, 106867. <https://doi.org/10.1016/j.elecom.2020.106867>
- Karakus, U, Thamamongood, T, Ciminski, K, Ran, W, Günther, SC, Pohl, MO, Eletto, D, Jeney, C, Hoffmann, D, Reiche, S, Schinköthe, J, Ulrich, R, Wiener, J, Hayes, MGB, Chang, MW, Hunziker, A, Yángüez, E, Aydillo, T, Krammer, F, ... Stertz, S. (2019). MHC class II proteins mediate cross-species entry of bat influenza viruses. *Nature*, 567(7746), 109–112. <https://doi.org/10.1038/s41586-019-0955-3>
- Khanna, P. (1991). Homogeneous enzyme immunoassay. In CP Price & DJ Newman (Eds.), *Principles and Practice of Immunoassay* (pp. 326–364). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-11234-0_12
- Khodakov, D, Wang, C, & Zhang, DY. (2016). Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Advanced Drug Delivery Reviews*, 105, 3–19. <https://doi.org/10.1016/j.addr.2016.04.005>
- Kibbe, WA. (2007). OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Research*, 35(Web Server), W43–W46. <https://doi.org/10.1093/nar/gkm234>
- Klebes, A, Ates, HC, Verboket, RD, Urban, GA, Von Stetten, F, Dincer, C, & Früh, SM. (2024). Emerging multianalyte biosensors for the simultaneous detection of protein and nucleic acid biomarkers. *Biosensors and Bioelectronics*, 244, 115800. <https://doi.org/10.1016/j.bios.2023.115800>
- Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues: A novel approach to testing

- for induced point mutations in mammals. *Humangenetik*, 26(3), 231–243. <https://doi.org/10.1007/BF00281458>
- Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, 256(5517), 495–497. <https://doi.org/10.1038/256495a0>
- Kontermann, RE, & Brinkmann, U. (2015). Bispecific antibodies. *Drug Discovery Today*, 20(7), 838–847. <https://doi.org/10.1016/j.drudis.2015.02.008>
- Košir, AB, Divieto, C, Pavšič, J, Pavarelli, S, Dobnik, D, Dreо, T, Bel-lotti, R, Sassi, MP, & Žel, J. (2017). Droplet volume variability as a critical factor for accuracy of absolute quantification using droplet digital PCR. *Analytical and Bioanalytical Chemistry*, 409(28), 6689–6697. <https://doi.org/10.1007/s00216-017-0625-y>
- Kumar, M, Joseph, SR, Augsburg, M, Bogdanova, A, Drechsel, D, Vastenhout, NL, Buchholz, F, Gentzel, M, & Shevchenko, A. (2018). MS Western, a Method of Multiplexed Absolute Protein Quantification is a Practical Alternative to Western Blotting. *Molecular & Cellular Proteomics*, 17(2), 384–396. <https://doi.org/10.1074/mcp.O117.067082>
- Kunert, R, & Reinhart, D. (2016). Advances in recombinant antibody manufacturing. *Applied Microbiology and Biotechnology*, 100(8), 3451–3461. <https://doi.org/10.1007/s00253-016-7388-9>
- Kutyavin, IV. (2000). 3'-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Research*, 28(2), 655–661. <https://doi.org/10.1093/nar/28.2.655>
- Laemmli, UK. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), 680–685. <https://doi.org/10.1038/227680ao>
- Lai, YK, Kao, YT, Hess, JF, Calabrese, S, Von Stetten, F, & Paust, N. (2023). Interfacing centrifugal microfluidics with linear-oriented 8-tube strips and multichannel pipettes for increased throughput of digital assays. *Lab on a Chip*, 23(11), 2623–2632. <https://doi.org/10.1039/D3LC00339F>
- Lange, T, Groß, T, Jeney, Á, Scherzinger, J, Sinkala, E, Niemöller, C, Zimmermann, S, Koltay, P, Stetten, Fv, Zengerle, R, & Jeney, C. (2022). Validation of scRNA-seq by scRT-ddPCR using the example of ErbB2 in MCF7 cells (preprint). *Molecular Biology*. <https://doi.org/10.1101/2022.05.31.494164>
- Lazarev, VF, Guzhova, IV, & Margulis, BA. (2020). Glyceraldehyde-3-phosphate Dehydrogenase Is a Multifaceted Therapeutic Target. *Pharmaceutics*, 12(5), 416. <https://doi.org/10.3390/pharmaceutics12050416>
- Le Louët, H, & Pitts, PJ. (2023). Twenty-First Century Global ADR Management: A Need for Clarification, Redesign, and Coordi-

- nated Action. *Therapeutic Innovation & Regulatory Science*, 57(1), 100–103. <https://doi.org/10.1007/s43441-022-00443-8>
- Le Tourneau, C., Kamal, M., Tsimberidou, AM., Bedard, P., Pierron, G., Callens, C., Rouleau, E., Vincent-Salomon, A., Servant, N., Alt, M., Rouzier, R., Paoletti, X., Delattre, O., & Bièche, I. (2016). Treatment Algorithms Based on Tumor Molecular Profiling: The Essence of Precision Medicine Trials. *Journal of the National Cancer Institute*, 108(4), djv362. <https://doi.org/10.1093/jnci/djv362>
- Lee, YS, Choi, JW, Kang, T, & Chung, BG. (2023). Deep Learning-Assisted Droplet Digital PCR for Quantitative Detection of Human Coronavirus. *BioChip Journal*, 17(1), 112–119. <https://doi.org/10.1007/s13206-023-00095-2>
- Lievens, A., Jacchia, S., Kagkli, D., Savini, C., & Querci, M. (2016). Measuring Digital PCR Quality: Performance Parameters and Their Optimization (KYK Chan, Ed.). *PLOS ONE*, 11(5), e0153317. <https://doi.org/10.1371/journal.pone.0153317>
- Ligasová, A., Frydrych, I., & Koberna, K. (2023). Basic Methods of Cell Cycle Analysis. *International Journal of Molecular Sciences*, 24(4), 3674. <https://doi.org/10.3390/ijms24043674>
- Lin, J., Jordi, C., Son, M., Van Phan, H., Drayman, N., Abasiyanik, MF., Vistain, L., Tu, HL, & Tay, S. (2019). Ultra-sensitive digital quantification of proteins and mRNA in single cells. *Nature Communications*, 10(1), 3544. <https://doi.org/10.1038/s41467-019-11531-z>
- Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, HE., Wolters, D., & Marcus, K. (2017). Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biological Chemistry*, 398(5-6), 687–699. <https://doi.org/10.1515/hsz-2017-0104>
- Liu, S., Li, B., Liang, Q., Liu, A., Qu, L., & Yang, J. (2020). Classification and function of RNA-protein interactions. *WIREs RNA*, 11(6), e1601. <https://doi.org/10.1002/wrna.1601>
- Liu, Y., Gu, J., Hagner-McWhirter, Å., Sathiyanarayanan, P., Gullberg, M., Söderberg, O., Johansson, J., Hammond, M., Ivansson, D., & Landegren, U. (2011). Western Blotting via Proximity Ligation for High Performance Protein Analysis. *Molecular & Cellular Proteomics*, 10(11), O111.011031. <https://doi.org/10.1074/mcp.O111.011031>
- Livingstone, M., & Bidinosti, M. (2012). Rapamycin-insensitive mTORC1 activity controls eIF4E:4E-BP1 binding. *F1000Research*, 1, 4. <https://doi.org/10.12688/f1000research.1-4.v1>
- Livingstone, M., Larsson, O., Sukarieh, R., Pelletier, J., & Sonenberg, N. (2009). A Chemical Genetic Screen for mTOR Pathway Inhibitors Based on 4E-BP-Dependent Nuclear Accumulation

- of eIF4E. *Chemistry & Biology*, 16(12), 1240–1249. <https://doi.org/10.1016/j.chembiol.2009.11.010>
- Lübbert, M, Schmoor, C, Grishina, O, Rummelt, C, Thomas, J, Heuser, M, Döhner, K, Döhner, H, Becker, H, & Wäsch, R. (2023). Decider-2: Prospective Randomized Multicenter Phase III Trial of Decitabine and Venetoclax Administered in Combination with All-*Trans* Retinoic Acid (ATRA) or Placebo in Patients with Acute Myeloid Leukemia Who Are Ineligible for Induction Chemotherapy. *Blood*, 142(Supplement 1), 1495–1495. <https://doi.org/10.1182/blood-2023-181871>
- Lundberg, M, Eriksson, A, Tran, B, Assarsson, E, & Fredriksson, S. (2011). Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Research*, 39(15), e102–e102. <https://doi.org/10.1093/nar/gkr424>
- Lundberg, M, Thorsen, SB, Assarsson, E, Villalblanca, A, Tran, B, Gee, N, Knowles, M, Nielsen, BS, González Couto, E, Martin, R, Nilsson, O, Fermer, C, Schlingemann, J, Christensen, IJ, Nielsen, HJ, Ekström, B, Andersson, C, Gustafsson, M, Brunner, N, ... Fredriksson, S. (2011). Multiplexed Homogeneous Proximity Ligation Assays for High-throughput Protein Biomarker Research in Serological Material. *Molecular & Cellular Proteomics*, 10(4), M110.004978. <https://doi.org/10.1074/mcp.M110.004978>
- Luskin, MR, Gimotty, PA, Smith, C, Loren, AW, Figueiroa, ME, Harrison, J, Sun, Z, Tallman, MS, Paietta, EM, Litzow, MR, Melnick, AM, Levine, RL, Fernandez, HF, Luger, SM, Carroll, M, Master, SR, & Wertheim, GB. (2016). A clinical measure of DNA methylation predicts outcome in de novo acute myeloid leukemia. *JCI Insight*, 1(9). <https://doi.org/10.1172/jci.insight.87323>
- Ma, S, Saaem, I, & Tian, J. (2012). Error correction in gene synthesis technology. *Trends in Biotechnology*, 30(3), 147–154. <https://doi.org/10.1016/j.tibtech.2011.10.002>
- MacCoss, MJ, Alfaro, JA, Faivre, DA, Wu, CC, Wanunu, M, & Slavov, N. (2023). Sampling the proteome by emerging single-molecule and mass spectrometry methods. *Nature Methods*, 20(3), 339–346. <https://doi.org/10.1038/s41592-023-01802-5>
- Madic, J, Zocevic, A, Senlis, V, Fradet, E, Andre, B, Muller, S, Dangla, R, & Droniou, M. (2016). Three-color crystal digital PCR. *Biomolecular Detection and Quantification*, 10, 34–46. <https://doi.org/10.1016/j.bdq.2016.10.002>
- Majumdar, N, Banerjee, S, Pallas, M, Wessel, T, & Hegerich, P. (2017). Poisson Plus Quantification for Digital PCR Systems. *Scientific Reports*, 7(1), 9617. <https://doi.org/10.1038/s41598-017-09183-4>

- Majumdar, N, Wessel, T, & Marks, J. (2015). Digital PCR Modeling for Maximal Sensitivity, Dynamic Range and Measurement Precision (R Margis, Ed.). *PLOS ONE*, 10(3), e0118833. <https://doi.org/10.1371/journal.pone.0118833>
- Malone, ER, Oliva, M, Sabatini, PJB, Stockley, TL, & Siu, LL. (2020). Molecular profiling for precision cancer therapies. *Genome Medicine*, 12(1), 8. <https://doi.org/10.1186/s13073-019-0703-1>
- Mandell, JW. (2003). Phosphorylation State-Specific Antibodies. *The American Journal of Pathology*, 163(5), 1687–1698. [https://doi.org/10.1016/S0002-9440\(10\)63525-0](https://doi.org/10.1016/S0002-9440(10)63525-0)
- Mardis, ER. (2019). The Impact of Next-Generation Sequencing on Cancer Genomics: From Discovery to Clinic. *Cold Spring Harbor Perspectives in Medicine*, 9(9), a036269. <https://doi.org/10.1101/cshperspect.a036269>
- Markoulatos, P, Siafakas, N, & Moncany, M. (2002). Multiplex polymerase chain reaction: A practical approach. *Journal of Clinical Laboratory Analysis*, 16(1), 47–51. <https://doi.org/10.1002/jcla.2058>
- McDermott, GP, Do, D, Litterst, CM, Maar, D, Hindson, CM, Steenblock, ER, Legler, TC, Jouvenot, Y, Marrs, SH, Bemis, A, Shah, P, Wong, J, Wang, S, Sally, D, Javier, L, Dinio, T, Han, C, Brackbill, TP, Hodges, SP, ... Lowe, AJ. (2013). Multiplexed Target Detection Using DNA-Binding Dye Chemistry in Droplet Digital PCR. *Analytical Chemistry*, 85(23), 11619–11627. <https://doi.org/10.1021/ac403061n>
- Meyers, J, Yu, Y, Kaye, JA, & Davis, KL. (2013). Medicare Fee-for-Service Enrollees with Primary Acute Myeloid Leukemia: An Analysis of Treatment Patterns, Survival, and Healthcare Resource Utilization and Costs. *Applied Health Economics and Health Policy*, 11(3), 275–286. <https://doi.org/10.1007/s40258-013-0032-2>
- Moon, S, Ceyhan, E, Gurkan, UA, & Demirci, U. (2011). Statistical Modeling of Single Target Cell Encapsulation (CV Borlongan, Ed.). *PLoS ONE*, 6(7), e21580. <https://doi.org/10.1371/journal.pone.0021580>
- Morley, AA. (2014). Digital PCR: A brief history. *Biomolecular Detection and Quantification*, 1(1), 1–2. <https://doi.org/10.1016/j.bdq.2014.06.001>
- Muñoz, J, & Heck, AJR. (2014). From the Human Genome to the Human Proteome. *Angewandte Chemie International Edition*, 53(41), 10864–10866. <https://doi.org/10.1002/anie.201406545>
- Murphy, K, & Weaver, C. (2018). *Janeway Immunologie*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-56004-4>
- Nahta, R, Hung, MC, & Esteva, FJ. (2004). The HER-2-Targeting Antibodies Trastuzumab and Pertuzumab Synergistically Inhibit

- the Survival of Breast Cancer Cells. *Cancer Research*, 64(7), 2343–2346. <https://doi.org/10.1158/0008-5472.CAN-03-3856>
- Nakagawa, S., & Cuthill, IC. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, 82(4), 591–605. <https://doi.org/10.1111/j.1469-185X.2007.00027.x>
- Nami, B, Maadi, H, & Wang, Z. (2018). Mechanisms Underlying the Action and Synergism of Trastuzumab and Pertuzumab in Targeting HER2-Positive Breast Cancer. *Cancers*, 10(10), 342. <https://doi.org/10.3390/cancers10100342>
- National Cancer Institute. (2023). Cancer Stat Facts: Leukemia — Acute Myeloid Leukemia (AML). Retrieved April 16, 2024, from <https://seer.cancer.gov/statfacts/html/amyl.html>
- Nattestad, M, Goodwin, S, Ng, K, Baslan, T, Sedlazeck, FJ, Rescheneder, P, Garvin, T, Fang, H, Gurtowski, J, Hutton, E, Tseng, E, Chin, CS, Beck, T, Sundaravadanam, Y, Kramer, M, Antoniou, E, McPherson, JD, Hicks, J, McCombie, WR, & Schatz, MC. (2018). Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Research*, 28(8), 1126–1135. <https://doi.org/10.1101/gr.231100.117>
- Nelson, MD, & Fitch, DHA. (2012). Overlap Extension PCR: An Efficient Method for Transgene Construction [Series Title: Methods in Molecular Biology]. In V Orgogozo & MV Rockman (Eds.), *Molecular Methods for Evolutionary Genetics* (pp. 459–470, Vol. 772). Humana Press. https://doi.org/10.1007/978-1-61779-228-1_27
- Nelson, RW, Krone, JR, Bieber, AL, & Williams, P. (1995). Mass Spectrometric Immunoassay. *Analytical Chemistry*, 67(7), 1153–1158. <https://doi.org/10.1021/ac00103a003>
- Nepstad, I, Hatfield, KJ, Grønningssæter, IS, & Reikvam, H. (2020). The PI3K-Akt-mTOR Signaling Pathway in Human Acute Myeloid Leukemia (AML) Cells. *International Journal of Molecular Sciences*, 21(8), 2907. <https://doi.org/10.3390/ijms21082907>
- Neri, D, Petrul, H, & Roncucci, G. (1995). Engineering recombinant antibodies for immunotherapy. *Cell Biophysics*, 27(1), 47–61. <https://doi.org/10.1007/BF02822526>
- Nong, RY, Gu, J, Darmanis, S, Kamali-Moghaddam, M, & Landegren, U. (2012). DNA-assisted protein detection technologies. *Expert Review of Proteomics*, 9(1), 21–32. <https://doi.org/10.1586/epr.11.78>
- Nowogrodzki, A. (2017). Inequality in medicine. *Nature*, 550(7674), S18–S19. <https://doi.org/10.1038/550S18a>
- O'Farrell, P. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, 250(10), 4007–4021. [https://doi.org/10.1016/S0021-9258\(19\)41496-8](https://doi.org/10.1016/S0021-9258(19)41496-8)

- Ortolani, C. (2022). *Flow Cytometry Today: Everything You Need to Know about Flow Cytometry*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-10836-5>
- Owens, J. (2007). Determining druggability. *Nature Reviews Drug Discovery*, 6(3), 187–187. <https://doi.org/10.1038/nrd2275>
- Pace, CN, Vajdos, F, Fee, L, Grimsley, G, & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science*, 4(11), 2411–2423. <https://doi.org/10.1002/pro.5560041120>
- Padhan, N, Yan, J, Boge, A, Scrivener, E, Birgisson, H, Zieba, A, Gullberg, M, Kamali-Moghaddam, M, Claesson-Welsh, L, & Landegren, U. (2017). Highly sensitive and specific protein detection via combined capillary isoelectric focusing and proximity ligation. *Scientific Reports*, 7(1), 1490. <https://doi.org/10.1038/s41598-017-01516-7>
- Patel, JP, Gönen, M, Figueroa, ME, Fernandez, H, Sun, Z, Racevskis, J, Van Vlierberghe, P, Dolgalev, I, Thomas, S, Aminova, O, Huberman, K, Cheng, J, Viale, A, Soccia, ND, Heguy, A, Cherry, A, Vance, G, Higgins, RR, Ketterling, RP, … Levine, RL. (2012). Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *New England Journal of Medicine*, 366(12), 1079–1089. <https://doi.org/10.1056/NEJMoa1112304>
- Payne, SH. (2015). The utility of protein and mRNA correlation. *Trends in Biochemical Sciences*, 40(1), 1–3. <https://doi.org/10.1016/j.tibs.2014.10.010>
- Petrera, A, Von Toerne, C, Behler, J, Huth, C, Thorand, B, Hilgendorff, A, & Hauck, SM. (2021). Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. *Journal of Proteome Research*, 20(1), 751–762. <https://doi.org/10.1021/acs.jproteome.0c00641>
- Petzold, A. (2019). The relevance of buffer system ionic strength in immunoassay development. *Journal of Immunological Methods*, 465, 27–30. <https://doi.org/10.1016/j.jim.2018.11.013>
- Pozarowski, P, & Darzynkiewicz, Z. (2004). Analysis of Cell Cycle by Flow Cytometry. In *Checkpoint Controls and Cancer* (pp. 301–312, Vol. 281). Humana Press. <https://doi.org/10.1385/1-59259-811-0:301>
- Qin, X, Jiang, B, & Zhang, Y. (2016). 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle*, 15(6), 781–786. <https://doi.org/10.1080/15384101.2016.1151581>
- Quan, PL, Sauzade, M, & Brouzes, E. (2018). dPCR: A Technology Review. *Sensors*, 18(4), 1271. <https://doi.org/10.3390/s18041271>
- Quentmeier, H, Pommerenke, C, Dirks, WG, Eberth, S, Koeppel, M, MacLeod, RAF, Nagel, S, Steube, K, Uphoff, CC, & Drexler, HG. (2019). The LL-100 panel: 100 cell lines for blood cancer studies.

- Scientific Reports*, 9(1), 8218. <https://doi.org/10.1038/s41598-019-44491-x>
- Rissin, DM, Fournier, DR, Piech, T, Kan, CW, Campbell, TG, Song, L, Chang, L, Rivnak, AJ, Patel, PP, Provuncher, GK, Ferrell, EP, Howes, SC, Pink, BA, Minnehan, KA, Wilson, DH, & Duffy, DC. (2011). Simultaneous Detection of Single Molecules and Singulated Ensembles of Molecules Enables Immunoassays with Broad Dynamic Range. *Analytical Chemistry*, 83(6), 2279–2285. <https://doi.org/10.1021/ac103161b>
- Rissin, DM, Kan, CW, Campbell, TG, Howes, SC, Fournier, DR, Song, L, Piech, T, Patel, PP, Chang, L, Rivnak, AJ, Ferrell, EP, Randall, JD, Provuncher, GK, Walt, DR, & Duffy, DC. (2010). Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nature Biotechnology*, 28(6), 595–599. <https://doi.org/10.1038/nbt.1641>
- Rissin, DM, Kan, CW, Song, L, Rivnak, AJ, Fishburn, MW, Shao, Q, Piech, T, Ferrell, EP, Meyer, RE, Campbell, TG, Fournier, DR, & Duffy, DC. (2013). Multiplexed single molecule immunoassays. *Lab on a Chip*, 13(15), 2902. <https://doi.org/10.1039/c3lc50416f>
- Rong, L, Livingstone, M, Sukarieh, R, Petroulakis, E, Gingras, AC, Crosby, K, Smith, B, Polakiewicz, RD, Pelletier, J, Ferraiuolo, MA, & Sonenberg, N. (2008). Control of eIF4E cellular localization by eIF4E-binding proteins, 4E-BPs. *RNA*, 14(7), 1318–1327. <https://doi.org/10.1261/rna.950608>
- Sano, T, Smith, CL, & Cantor, CR. (1992). Immuno-PCR: Very Sensitive Antigen Detection by Means of Specific Antibody-DNA Conjugates. *Science*, 258(5079), 120–122. <https://doi.org/10.1126/science.1439758>
- Saxton, RA, & Sabatini, DM. (2017). mTOR Signaling in Growth, Metabolism, and Disease. *Cell*, 168(6), 960–976. <https://doi.org/10.1016/j.cell.2017.02.004>
- Schallmeiner, E, Oksanen, E, Ericsson, O, Spångberg, L, Eriksson, S, Stenman, UH, Pettersson, K, & Landegren, U. (2007). Sensitive protein detection via triple-binder proximity ligation assays. *Nature Methods*, 4(2), 135–137. <https://doi.org/10.1038/nmeth974>
- Schastnaya, E, Raguz Nakic, Z, Gruber, CH, Doubleday, PF, Krishnan, A, Johns, NI, Park, J, Wang, HH, & Sauer, U. (2021). Extensive regulation of enzyme activity by phosphorylation in Escherichia coli. *Nature Communications*, 12(1), 5650. <https://doi.org/10.1038/s41467-021-25988-4>
- Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2), 103–113. <https://doi.org/10.1111/j.2041-210X.2010.00012.x>
- Schielzeth, H, Dingemanse, NJ, Nakagawa, S, Westneat, DF, Allegue, H, Teplitsky, C, Réale, D, Dochtermann, NA, Garamszegi,

- LZ, & Araya-Ajoy, YG. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions (C Sutherland, Ed.). *Methods in Ecology and Evolution*, 11(9), 1141–1152. <https://doi.org/10.1111/2041-210X.13434>
- Schlenker, F, Kipf, E, Borst, N, Hutzenlaub, T, Zengerle, R, Von Stetten, F, & Juelg, P. (2021). Virtual Fluorescence Color Channels by Selective Photobleaching in Digital PCR Applied to the Quantification of KRAS Point Mutations. *Analytical Chemistry*, 93(30), 10538–10545. <https://doi.org/10.1021/acs.analchem.1c01488>
- Schlüter, H, Apweiler, R, Holzhütter, HG, & Jungblut, PR. (2009). Finding one's way in proteomics: A protein species nomenclature. *Chemistry Central Journal*, 3(1), 11. <https://doi.org/10.1186/1752-153X-3-11>
- Schork, NJ. (2015). Personalized medicine: Time for one-person trials. *Nature*, 520(7549), 609–611. <https://doi.org/10.1038/520609a>
- Schulz, M, Probst, S, Calabrese, S, R. Homann, A, Borst, N, Weiss, M, Von Stetten, F, Zengerle, R, & Paust, N. (2020). Versatile Tool for Droplet Generation in Standard Reaction Tubes by Centrifugal Step Emulsification. *Molecules*, 25(8), 1914. <https://doi.org/10.3390/molecules25081914>
- Schulz, M, Ruediger, J, Landmann, E, Bakheit, M, Frischmann, S, Rassler, D, Homann, AR, Von Stetten, F, Zengerle, R, & Paust, N. (2021). High Dynamic Range Digital Assay Enabled by Dual-Volume Centrifugal Step Emulsification. *Analytical Chemistry*, 93(5), 2854–2860. <https://doi.org/10.1021/acs.analchem.0c04182>
- Schwanhäusser, B, Busse, D, Li, N, Dittmar, G, Schuchhardt, J, Wolf, J, Chen, W, & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347), 337–342. <https://doi.org/10.1038/nature10098>
- Schwartzberg, L, Kim, ES, Liu, D, & Schrag, D. (2017). Precision Oncology: Who, How, What, When, and When Not? *American Society of Clinical Oncology Educational Book*, (37), 160–169. https://doi.org/10.1200/EDBK_174176
- Scott, AM, Wolchok, JD, & Old, LJ. (2012). Antibody therapy of cancer. *Nature Reviews Cancer*, 12(4), 278–287. <https://doi.org/10.1038/nrc3236>
- Selby, C. (1999). Interference in Immunoassay. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, 36(6), 704–721. <https://doi.org/10.1177/000456329903600603>
- Shah, A, Grimberg, DC, & Inman, BA. (2020). Classification of Molecular Biomarkers. *Société Internationale d'Urologie Journal*, 1(1), 8–15. <https://doi.org/10.48083/AKUI6936>

- Shames, D, Minna, J, & Gazdar, A. (2007). DNA Methylation in Health, Disease, and Cancer. *Current Molecular Medicine*, 7(1), 85–102. <https://doi.org/10.2174/156652407779940413>
- Sharma, R, Kumbhakar, M, & Mukherjee, A. (2021). Toward Understanding the Binding Synergy of Trastuzumab and Pertuzumab to Human Epidermal Growth Factor Receptor 2. *Molecular Pharmaceutics*, 18(12), 4553–4563. <https://doi.org/10.1021/acs.molpharmaceut.1co00775>
- Shi, F, Zhang, J, Liu, H, Wu, L, Jiang, H, Wu, Q, Liu, T, Lou, M, & Wu, H. (2018). The dual PI₃K/mTOR inhibitor dactolisib elicits anti-tumor activity *in vitro* and *in vivo*. *Oncotarget*, 9(1), 706–717. <https://doi.org/10.18632/oncotarget.23091>
- Shuford, CM, Walters, JJ, Holland, PM, Sreenivasan, U, Askari, N, Ray, K, & Grant, RP. (2017). Absolute Protein Quantification by Mass Spectrometry: Not as Simple as Advertised. *Analytical Chemistry*, 89(14), 7406–7415. <https://doi.org/10.1021/acs.analchem.7b00858>
- Singh, A. (2023). Nanopores for sequencing proteins. *Nature Methods*, 20(12), 1870–1870. <https://doi.org/10.1038/s41592-023-02108-2>
- Sinitcyn, P, Richards, AL, Weatheritt, RJ, Brademan, DR, Marx, H, Shishkova, E, Meyer, JG, Hebert, AS, Westphall, MS, Blencowe, BJ, Cox, J, & Coon, JJ. (2023). Global detection of human variants and isoforms by deep proteome sequencing. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-023-01714-x>
- Sirover, MA. (1999). New insights into an old protein: The functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1432(2), 159–184. [https://doi.org/10.1016/S0167-4838\(99\)00119-3](https://doi.org/10.1016/S0167-4838(99)00119-3)
- Sjöblom, T, Jones, S, Wood, LD, Parsons, DW, Lin, J, Barber, TD, Mandelker, D, Leary, RJ, Ptak, J, Silliman, N, Szabo, S, Buckhaults, P, Farrell, C, Meeh, P, Markowitz, SD, Willis, J, Dawson, D, Willson, JKV, Gazdar, AF, . . . Velculescu, VE. (2006). The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science*, 314(5797), 268–274. <https://doi.org/10.1126/science.1133427>
- Smith, LM, & Kelleher, NL. (2018). Proteoforms as the next proteomics currency. *Science*, 359(6380), 1106–1107. <https://doi.org/10.1126/science.aat1884>
- Smith, LM, Kelleher, NL, & The Consortium for Top Down Proteomics. (2013). Proteoform: A single term describing protein complexity. *Nature Methods*, 10(3), 186–187. <https://doi.org/10.1038/nmeth.2369>
- Smith, P, Krohn, R, Hermanson, G, Mallia, A, Gartner, F, Provenzano, M, Fujimoto, E, Goeke, N, Olson, B, & Klenk, D. (1985).

- Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, 150(1), 76–85. [https://doi.org/10.1016/0003-2697\(85\)90442-7](https://doi.org/10.1016/0003-2697(85)90442-7)
- Snyder, M., Weissman, S., & Gerstein, M. (2009). Personal phenotypes to go with personal genomes. *Molecular Systems Biology*, 5(1), 273. <https://doi.org/10.1038/msb.2009.32>
- Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, KJ., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, LG., & Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature Methods*, 3(12), 995–1000. <https://doi.org/10.1038/nmeth947>
- Srinivasan, M., Sedmak, D., & Jewell, S. (2002). Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. *The American Journal of Pathology*, 161(6), 1961–1971. [https://doi.org/10.1016/S0002-9440\(10\)64472-0](https://doi.org/10.1016/S0002-9440(10)64472-0)
- Stomper, J., Rotondo, JC., Greve, G., & Lübbert, M. (2021). Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: Mechanisms of resistance and novel HMA-based therapies. *Leukemia*, 35(7), 1873–1889. <https://doi.org/10.1038/s41375-021-01218-o>
- Stoscheck, CM. (1990). [6] Quantitation of protein. In *Methods in Enzymology* (pp. 50–68, Vol. 182). Elsevier. [https://doi.org/10.1016/0076-6879\(90\)82008-P](https://doi.org/10.1016/0076-6879(90)82008-P)
- Stresemann, C., & Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer*, 123(1), 8–13. <https://doi.org/10.1002/ijc.23607>
- Sun, R., Cheng, E., Velásquez, C., Chang, Y., & Moore, PS. (2019). Mitosis-related phosphorylation of the eukaryotic translation suppressor 4E-BP1 and its interaction with eukaryotic translation initiation factor 4E (eIF4E). *Journal of Biological Chemistry*, 294(31), 11840–11852. <https://doi.org/10.1074/jbc.RA119.008512>
- Szmigelska-Kaplon, A., & Robak, T. (2011). Hypomethylating Agents in the Treatment of Myelodysplastic Syndromes and Myeloid Leukemia. *Current Cancer Drug Targets*, 11(7), 837–848. <https://doi.org/10.2174/156800911796798940>
- Tan, LL., Loganathan, N., Agarwalla, S., Yang, C., Yuan, W., Zeng, J., Wu, R., Wang, W., & Duraiswamy, S. (2023). Current commercial dPCR platforms: Technology and market review. *Critical Reviews in Biotechnology*, 43(3), 433–464. <https://doi.org/10.1080/07388551.2022.2037503>
- Tang, X., & Bruce, JE. (2009). Chemical Cross-Linking for Protein–Protein Interaction Studies [Series Title: Methods In Molecular Biology]. In JM Walker, MS Lipton, & L Paša-Tolic (Eds.), *Mass*

- Spectrometry of Proteins and Peptides* (pp. 283–293, Vol. 492). Humana Press. https://doi.org/10.1007/978-1-59745-493-3_17
- Temple, R. (1999). Are Surrogate Markers Adequate to Assess Cardiovascular Disease Drugs? *JAMA*, 282(8), 790. <https://doi.org/10.1001/jama.282.8.790>
- The dMIQE Group, Whale, AS, De Spiegelaere, W, Trypsteen, W, Nour, AA, Bae, YK, Benes, V, Burke, D, Cleveland, M, Corbisier, P, Devonshire, AS, Dong, L, Drandi, D, Foy, CA, Garson, JA, He, HJ, Hellemans, J, Kubista, M, Lievens, A, ... Huggett, JF. (2020). The Digital MIQE Guidelines Update: Minimum Information for Publication of Quantitative Digital PCR Experiments for 2020. *Clinical Chemistry*, 66(8), 1012–1029. <https://doi.org/10.1093/clinchem/hvaa125>
- Thoreen, CC, Kang, SA, Chang, JW, Liu, Q, Zhang, J, Gao, Y, Reichling, LJ, Sim, T, Sabatini, DM, & Gray, NS. (2009). An ATP-competitive Mammalian Target of Rapamycin Inhibitor Reveals Rapamycin-resistant Functions of mTORC1. *Journal of Biological Chemistry*, 284(12), 8023–8032. <https://doi.org/10.1074/jbc.M900301200>
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, 302(5909), 575–581. <https://doi.org/10.1038/302575a0>
- Tsao, LC, Crosby, EJ, Trotter, TN, Wei, J, Wang, T, Yang, X, Summers, AN, Lei, G, Rabiola, CA, Chodosh, LA, Muller, WJ, Lyerly, HK, & Hartman, ZC. (2022). Trastuzumab/pertuzumab combination therapy stimulates antitumor responses through complement-dependent cytotoxicity and phagocytosis. *JCI Insight*, 7(6), e155636. <https://doi.org/10.1172/jci.insight.155636>
- van Dieck, J, Schmid, V, Heindl, D, Dziadek, S, Schraeml, M, Gerg, M, Massoner, P, Engel, AM, Tiefenthaler, G, Vural, S, Stritt, S, Tetzlaff, F, Soukupova, M, Kopetzki, E, Bossenmaier, B, Thomas, M, Klein, C, Mertens, A, Heller, A, & Tacke, M. (2014). Development of Bispecific Molecules for the In Situ Detection of Protein-Protein Interactions and Protein Phosphorylation. *Chemistry & Biology*, 21(3), 357–368. <https://doi.org/10.1016/j.chembiol.2013.12.018>
- Vargas, AJ, & Harris, CC. (2016). Biomarker development in the precision medicine era: Lung cancer as a case study. *Nature Reviews Cancer*, 16(8), 525–537. <https://doi.org/10.1038/nrc.2016.56>
- Veal, CD, Freeman, PJ, Jacobs, K, Lancaster, O, Jamain, S, Leboyer, M, Albanes, D, Vaghela, RR, Gut, I, Chanock, SJ, & Brookes, AJ. (2012). A mechanistic basis for amplification differences between samples and between genome regions. *BMC Genomics*, 13(1), 455. <https://doi.org/10.1186/1471-2164-13-455>
- Venter, JC, Adams, MD, Myers, EW, Li, PW, Mural, RJ, Sutton, GG, Smith, HO, Yandell, M, Evans, CA, Holt, RA, Gocayne, JD, Amanatides, P, Ballew, RM, Huson, DH, Wortman, JR, Zhang,

- Q, Kodira, CD, Zheng, XH, Chen, L, ... Zhu, X. (2001). The Sequence of the Human Genome. *Science*, 291(5507), 1304–1351. <https://doi.org/10.1126/science.1058040>
- Vistain, L, Van Phan, H, Keisham, B, Jordi, C, Chen, M, Reddy, ST, & Tay, S. (2022). Quantification of extracellular proteins, protein complexes and mRNAs in single cells by proximity sequencing. *Nature Methods*, 19(12), 1578–1589. <https://doi.org/10.1038/s41592-022-01684-z>
- Vogelstein, B, & Kinzler, KW. (1999). Digital PCR. *Proceedings of the National Academy of Sciences*, 96(16), 9236–9241. <https://doi.org/10.1073/pnas.96.16.9236>
- Wadle, S, Rubenwolf, S, Lehnert, M, Faltin, B, Weidmann, M, Hufert, F, Zengerle, R, & Von Stetten, F. (2014). Mediator Probe PCR: Detection of Real-Time PCR by Label-Free Probes and a Universal Fluorogenic Reporter [Series Title: Methods in Molecular Biology]. In R Biassoni & A Raso (Eds.), *Quantitative Real-Time PCR* (pp. 55–73, Vol. 1160). Springer New York. https://doi.org/10.1007/978-1-4939-0733-5_6
- Walt, DR. (2013). Optical Methods for Single Molecule Detection and Analysis. *Analytical Chemistry*, 85(3), 1258–1263. <https://doi.org/10.1021/ac3027178>
- Wang, D, Eraslan, B, Wieland, T, Hallström, B, Hopf, T, Zolg, DP, Zecha, J, Asplund, A, Li, Lh, Meng, C, Frejno, M, Schmidt, T, Schnatbaum, K, Wilhelm, M, Ponten, F, Uhlen, M, Gagneur, J, Hahne, H, & Kuster, B. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular Systems Biology*, 15(2), e8503. <https://doi.org/10.15252/msb.20188503>
- Wang, RC, & Wang, Z. (2022). Synchronization of Cultured Cells to G₁, S, G₂, and M Phases by Double Thymidine Block [Series Title: Methods in Molecular Biology]. In Z Wang (Ed.), *Cell-Cycle Synchronization* (pp. 61–71, Vol. 2579). Springer US. https://doi.org/10.1007/978-1-0716-2736-5_5
- Weller, MG. (2016). Quality Issues of Research Antibodies. *Analytical Chemistry Insights*, 11, ACI.S31614. <https://doi.org/10.4137/ACI.S31614>
- Weller, MG. (2018). Ten Basic Rules of Antibody Validation. *Analytical Chemistry Insights*, 13, 117739011875746. <https://doi.org/10.1177/1177390118757462>
- Whale, AS, Cowen, S, Foy, CA, & Huggett, JF. (2013). Methods for Applying Accurate Digital PCR Analysis on Low Copy DNA Samples (G Maga, Ed.). *PLoS ONE*, 8(3), e58177. <https://doi.org/10.1371/journal.pone.0058177>
- Whale, AS, Huggett, JF, & Tzonev, S. (2016). Fundamentals of multiplexing with digital PCR. *Biomolecular Detection and Quantification*, 10, 15–23. <https://doi.org/10.1016/j.bdq.2016.05.002>

- Wik, L, Nordberg, N, Broberg, J, Björkosten, J, Assarsson, E, Henriksson, S, Grundberg, I, Pettersson, E, Westerberg, C, Liljeroth, E, Falck, A, & Lundberg, M. (2021). Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Molecular & Cellular Proteomics*, 20, 100168. <https://doi.org/10.1016/j.mcpro.2021.100168>
- Wilhelm, M, Schlegl, J, Hahne, H, Gholami, AM, Lieberenz, M, Savitski, MM, Ziegler, E, Butzmann, L, Gessulat, S, Marx, H, Mathieson, T, Lemeer, S, Schnatbaum, K, Reimer, U, Wenschuh, H, Moltenhauer, M, Slotta-Huspenina, J, Boese, JH, Bantscheff, M, ... Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502), 582–587. <https://doi.org/10.1038/nature13319>
- Wilhoit, T, Patric, JM, & May, MB. (2020). Alpelisib: A Novel Therapy for Patients With PIK3CA-Mutated Metastatic Breast Cancer. *Journal of the Advanced Practitioner in Oncology*, 11(7). <https://doi.org/10.6004/jadpro.2020.11.7.9>
- Wilson, DH, Rissin, DM, Kan, CW, Fournier, DR, Piech, T, Campbell, TG, Meyer, RE, Fishburn, MW, Cabrera, C, Patel, PP, Frew, E, Chen, Y, Chang, L, Ferrell, EP, von Einem, V, McGuigan, W, Reinhardt, M, Sayer, H, Vielsack, C, & Duffy, DC. (2016). The Simoa HD-1 Analyzer: A Novel Fully Automated Digital Immunoassay Analyzer with Single-Molecule Sensitivity and Multiplexing. *SLAS Technology*, 21(4), 533–547. <https://doi.org/10.1177/2211068215589580>
- Woestenenk, EA, Hammarström, M, Van Den Berg, S, Härd, T, & Berglund, H. (2004). His tag effect on solubility of human proteins produced in Escherichia coli: A comparison between four expression vectors. *Journal of Structural and Functional Genomics*, 5(3), 217–229. <https://doi.org/10.1023/B:jsfg.0000031965.37625.0e>
- Wood, WG. (1991). “Matrix Effects” in Immunoassays. *Scandinavian Journal of Clinical and Laboratory Investigation*, 51(sup205), 105–112. <https://doi.org/10.3109/00365519109104608>
- Wu, C, Dougan, TJ, & Walt, DR. (2022). High-Throughput, High-Multiplex Digital Protein Detection with Attomolar Sensitivity. *ACS Nano*, 16(1), 1025–1035. <https://doi.org/10.1021/acsnano.1c08675>
- Wu, C, Garden, PM, & Walt, DR. (2020). Ultrasensitive Detection of Attomolar Protein Concentrations by Dropcast Single Molecule Assays. *Journal of the American Chemical Society*, 142(28), 12314–12323. <https://doi.org/10.1021/jacs.0c04331>
- Xu, Q, Thompson, JE, & Carroll, M. (2005). mTOR regulates cell survival after etoposide treatment in primary AML cells. *Blood*,

- 106(13), 4261–4268. <https://doi.org/10.1182/blood-2004-11-4468>
- Yalow, RS, & Berson, SA. (1959). Assay of Plasma Insulin in Human Subjects by Immunological Methods. *Nature*, 184(4699), 1648–1649. <https://doi.org/10.1038/1841648bo>
- Yang, J, & Hlavacek, WS. (2011). Scaffold-mediated nucleation of protein signaling complexes: Elementary principles. *Mathematical Biosciences*, 232(2), 164–173. <https://doi.org/10.1016/j.mbs.2011.06.003>
- Ye, C, Hu, W, & Gaeta, B. (2022). Prediction of Antibody-Antigen Binding via Machine Learning: Development of Data Sets and Evaluation of Methods. *JMIR Bioinformatics and Biotechnology*, 3(1), e29404. <https://doi.org/10.2196/29404>
- Zhang, Y, & Zheng, XS. (2012). mTOR-independent 4E-BP1 phosphorylation is associated with cancer resistance to mTOR kinase inhibitors. *Cell Cycle*, 11(3), 594–603. <https://doi.org/10.4161/cc.11.3.19096>
- Zhao, Y, Zuo, X, Li, Q, Chen, F, Chen, YR, Deng, J, Han, D, Hao, C, Huang, F, Huang, Y, Ke, G, Kuang, H, Li, F, Li, J, Li, M, Li, N, Lin, Z, Liu, D, Liu, J, ... Fan, C. (2021). Nucleic Acids Analysis. *Science China Chemistry*, 64(2), 171–203. <https://doi.org/10.1007/s11426-020-9864-7>
- Zheng, Y, & Jiang, Y. (2015). mTOR Inhibitors at a Glance. *Molecular and Cellular Pharmacology*, 7(2), 15–20.
- Zhou, T. (2013). Bifurcation. In W Dubitzky, O Wolkenhauer, KH Cho, & H Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 79–86). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_500
- Zhuo, S, Clemens, J, Hakes, D, Barford, D, & Dixon, J. (1993). Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase. *Journal of Biological Chemistry*, 268(24), 17754–17761. [https://doi.org/10.1016/S0021-9258\(17\)46769-X](https://doi.org/10.1016/S0021-9258(17)46769-X)
- Zou, Z, Tao, T, Li, H, & Zhu, X. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: Progress and challenges. *Cell & Bioscience*, 10(1), 31. <https://doi.org/10.1186/s13578-020-00396-1>

LIST OF FIGURES

Figure 1.1	Transition from <i>one-size-fits-all</i> to precision medicine.	4
Figure 1.2	Molecular biomarkers according to FDA-NIH Biomarker Working Group (2016).	5
Figure 1.3	Central dogma of biology according to Crick (1970).	7
Figure 2.1	Detection and capture formats of heterogeneous immunoassays.	15
Figure 2.2	Monoclonal antibody.	17
Figure 2.3	DNA detection chemistry.	19
Figure 2.4	Schematic depiction of the generation of first DNA template for amplification in PLA and PEA.	22
Figure 3.1	Multilevel validation of the dDPCS model.	26
Figure 4.1	Working steps of the PICO technology.	32
Figure 4.2	Definition of PICO-relevant terms.	33
Figure 4.3	Workflow of PICO assays.	35
Figure 4.4	Formation of ternary complexes, i.e. <i>couplexes</i> , during the binding reaction.	36
Figure 4.5	Exemplary, random distribution process of $m = 100$ molecules in $n = 100$ partitions.	40
Figure 4.6	Exemplary ABC PICO experiment.	41
Figure 4.7	Exemplary PICO experiment.	43
Figure 4.8	2-dimensional raw dPCR data and composition of double positive cluster.	47
Figure 5.1	Comparison of expected and detected λ values for one DNA-labeled antibody after dPCR simulation.	59
Figure 5.2	Comparison of expected and detected λ values for one DNA-labeled antibody after dPCR simulation with different clustering biases.	61
Figure 5.3	Comparison of expected and detected λ values for one antibody after dPCR simulation with PSC	63
Figure 6.1	Ridgeline plot using histograms of couplex distributions calculated from simulated ABC PICO experiments using the dDPCS model.	66
Figure 6.2	Forest plot of estimates with 95 % CI from the comparison of couplex distributions with the expected mean of 0 couplexes.	67
Figure 6.3	Ambiguity of the maximum-likelihood function.	68

- Figure 6.4 Comparison of calculated and ground truth random double positive partitions from simulated ABC PICO experiments. 70
- Figure 6.5 Evaluation of dDPCS model performance using simulated data with targets. 72
- Figure 6.6 Representative scatterplots of calculated number of double positive partitions against ground truth double positive partitions. 74
- Figure 6.7 Comparison of calculated values and ground truth values for complex, random and rc-overlap double positive partitions. 75
- Figure 6.8 Qualitative impact of false-negative clustering bias on the dDPCS model. 77
- Figure 6.9 Qualitative impact of false-positive clustering bias on the dDPCS model. 78
- Figure 6.10 Scatterplot of the estimated impact of dropout threshold on the number of calculated complexes. 81
- Figure 6.11 Schematic depiction of DNA-labels and *artificial couplex* for DNA-only PICO experiments. 84
- Figure 6.12 Time-course quality control of oligonucleotide concentration. 86
- Figure 6.13 Scatterplot of fractions of single and double positive partitions from all positive partitions for the artificial couplex only. 87
- Figure 6.14 Comparison of dPCR systems using the artificial couplex. 88
- Figure 6.15 Comparison of number of partitions. 90
- Figure 6.16 Exemplary setup of DNA-only PICO experiments. 91
- Figure 6.17 Controls from naica and QIAcuity DNA-only PICO experiments. 92
- Figure 6.18 Scatterplot of calculated complexes against expected artificial complexes. 94
- Figure 6.19 Detection of recombinant HER2 complexes using TTZ and PTZ. 98
- Figure 6.20 ABC PICO experiments with the same antibodies as in Figure 6.19 and results from antibody pairs with URABs. 100
- Figure 6.21 Forest plot of estimates between different complexes calculation approaches. 104
- Figure 6.22 Comparison of sensitivities of immunoassays. 105
- Figure 7.1 Primary downstream targets of mTOR for protein synthesis. 111
- Figure 7.2 Comparison of antibody pairs for the detection of recombinant 4EBP1 complexes. 113

- Figure 7.3 Comparison of antibody pairs for the detection of cellular 4EBP1 complexes. [114](#)
- Figure 7.4 PICO assay for the detection of 4EBP1 using antibodies 4EB1T37T46-A5, E.638.3, 60246-1-Ig and 4F3-H2. [116](#)
- Figure 7.5 Comparison of buffers for the detection of recombinant 4EBP1. [117](#)
- Figure 7.6 Comparison of buffers for the detection of native 4EBP1 from cell lysate of U937 cells. [119](#)
- Figure 7.7 Comparison of phosphorylation antibody specificity using recombinant 4EBP1 before and after λ -PPase treatment. [121](#)
- Figure 7.8 Comparison of 4EBP1 phosphorylation between λ -PPase treated U937 cell lysate and mock treated U937 cell lysate. [122](#)
- Figure 7.9 Western blot comparison of inhibitors for 4EBP1-phosphorylating kinases in U937 cells. [124](#)
- Figure 7.10 Validated PICO assay for the detection of 4EBP1 using antibodies 4EB1T37T46-A5, 60246-1-Ig and 4F3-H2. [126](#)
- Figure 7.11 Comparison of 4EBP1 phosphorylation between dactolisib and mock treated U937 cells. [127](#)
- Figure 7.12 Forest plots of estimates with 95 % CI for the comparison of (HMA-resistant) AML cell lines using the validated 4EBP1 PICO assay ([Figure 7.4](#)). [129](#)
- Figure A.1 Comparison of models for complex calculation. [A.2-2](#)
- Figure A.2 Fractions of complex values in the categories > 0 , $= 0$ and < 0 . [A.2-3](#)
- Figure A.3 Forest plots of estimates from linear models comparing the dDPCS and the DX model. [A.2-4](#)
- Figure C.1 Selected histograms of simulated complexes from ABC PICO experiments. [C.0-2](#)
- Figure C.2 Maximum-likelihood function for simulated λ values for the calculation of the number of complexes using the dDPCS model (part 1). [C.0-3](#)
- Figure C.3 Maximum-likelihood function for simulated λ values for the calculation of the number of complexes using the dDPCS model (part 2). [C.0-4](#)
- Figure C.4 Ridgeline plot using histograms of complex distributions calculated from simulated PICO experiments using the dDPCS model. [C.0-5](#)
- Figure C.5 Characterization of the ground truth comparisons. [C.0-6](#)
- Figure C.6 Forest plot of estimates with 95 % CI from the comparison of complex distributions. [C.0-7](#)

- Figure C.7 Estimated difference between false-negative clustering bias and ideal conditions for different λ . [C.o-8](#)
- Figure C.8 Estimated difference between false-positive clustering bias and ideal conditions for different λ . [C.o-9](#)
- Figure C.9 Analysis of the artificial couplex. [C.o-10](#)
- Figure C.10 Scatterplots of RFU for partitions detected in the fluorescent channels of the QIAcuity dPCR system. [C.o-11](#)
- Figure C.11 Impact of the dDPCS model using simulated data ([Section 6.1.1](#)) for different values of λ . [C.o-12](#)
- Figure C.12 Impact of the dDPCS model using recombinant protein detected in different colorpairs ([Section 6.2.3](#)). [C.o-13](#)
- Figure C.13 Impact of the dDPCS model using data from the 4EBP1 PICO assay on U937 cell lysate ([Section 7.3](#)). [C.o-14](#)
- Figure C.14 Fluorescent channels of the QIAcuity dPCR system and used fluorescent dyes with excitation and emission wavelengths. [C.o-15](#)
- Figure C.15 Fluorescent channels of the QIAcuity dPCR system and optimized fluorescent dyes with excitation and emission wavelengths. [C.o-16](#)
- Figure C.16 Detection of recombinant HER2 complexes using TTZ and PTZ in different colorpairs with LE correction according to (Gross et al., 2024). [C.o-17](#)
- Figure C.17 Scatterplot of number of LE-corrected complexes against the LE. [C.o-18](#)
- Figure D.1 Representative electropherogram of recombinant 4EBP1. [D.o-1](#)
- Figure D.2 Detection of eIF4E complex using different antibodies. [D.o-1](#)
- Figure D.3 UpSet plot of number of ABC complexes using different 4EBP1 antibodies. [D.o-2](#)
- Figure D.4 Violin plots for comparison of total 4EBP1 and T37T46-phosphorylated 4EBP1 using ELISA. [D.o-2](#)
- Figure D.5 Comparison of ABCs from different buffers. [D.o-3](#)
- Figure D.6 Violin plots for comparison of (HMA-resistant) AML cell lines using the validated 4EBP1 PICO assay ([Figure 7.4](#)). [D.o-4](#)
- Figure E.1 Representative electropherogram from unlabeled antibody (clone 1F7 ([Table 10.4](#))). [E.o-1](#)
- Figure E.2 Representative electropherogram from labeled antibody (clone 1F7, compare to [Figure E.1](#)). [E.o-1](#)

LIST OF TABLES

Table 2.1	Comparison of immunoassay employing DNA-labeled antibodies. 24
Table 6.1	The clusters from Figure 4.8A can be biased separately or in arbitrary combinations. In case of multiple affected clusters, the clustering bias was the same for all clusters. The clustering biases were varied in {0, 1, 2, 3, 4, 5, 10, 15, 20, 25}%. 79
Table 6.2	Validation parameters for analytical methods according to European Medicines Agency (2023) applied to the dDPCS model. 107
Table 9.1	Recipe for preparation of 2X RIPA buffer. 143
Table 9.2	Recipe for the preparation of modified RIPA buffer for λ -PPase-treatment. 144
Table 9.3	Cycling conditions for the used dPCR systems (Section 9.2.7). 147
Table 9.4	R and python packages used for data analysis and data visualization. 151
Table 10.1	Sequences of labels for DNA-only PICO. 153
Table 10.2	Sequences of labels for antibody labeling (Section 9.2.3). 153
Table 10.3	Recipes for the preparation of master mix for the used dPCR systems (Section 9.2.7). 154
Table 10.4	Monoclonal antibodies against 4EBP1 proteoforms. 155
Table 10.5	Monoclonal antibodies against various further targets. 156
Table 10.6	Labeled antibodies use in this thesis. 157
Table 10.7	Sequences of probes. 158
Table 10.8	Sequences of forward (fwd) and reverse (rev) primers. 159
Table A.1	Excerpt from QIAcuity dPCR <i>MultipleOccupancy</i> file (Software Suite version 2.1.7.182). A.2-5
Table C.1	Combinations of TTZ and PTZ for colorpair comparison. C.0-18
Table E.1	Area under the curve after peak integration before and after labeling. E.0-2

LISTINGS

Listing 4.1	Implementation of couplex detection and calculation in R.	49
Listing 5.1	Simulation of partitioning process independently for antibodya, antibodyb and complexes.	54
Listing 5.2	Simulation of clustering.	55
Listing B.1	Simulation of antibody binding to a target.	B.o-1
Listing B.2	Counting of the number of partitions per cluster.	B.o-1
Listing B.3	Determination of ground truth double positive partitions from a simulated dPCR.	B.o-2
Listing B.4	Simulation of false-negative clustering.	B.o-3
Listing B.5	Simulation of false-positive clustering.	B.o-4
Listing B.6	Simulation of molecular dropouts.	B.o-5
Listing B.7	Randomly splitting a number into three parts.	B.o-6

APPENDIX

A

PROTEIN INTERACTION COUPLING (PICO)

A.1 COMPARISON OF MODELS FOR COMPLEX CALCULATION

Apart from **dDPCS** model (Section 4.2), the **DX** model (Gross et al., 2024) is another model for complex calculation. It uses the method of moments to find an estimator for m_{AB} from Equation 4.31¹:

$$\text{Number of complexes} = -\hat{n} \ln \left(\frac{(\hat{n}_{AB^+a^+b^-})(\hat{n}_{AB^+a^-b^+})}{\hat{n}(\hat{n}_{AB^-a^-b^-})} \right) \quad (\text{A.1})$$

The **dDPCS** and the **DX** model were compared using three datasets from Section 6.1.1, Section 6.2.3 and Section 7.3. The **dDPCS** model excludes the calculation of negative numbers of complexes by limiting the possible solution for \hat{n}_{AB^+} to a sequence from 0 to $\hat{n}_{D_{obs}}$ in steps of 1 (Listing 4.1). Thus, for all datasets, the numbers of complexes were ≥ 0 , while for the **DX** models a significant proportion (depending on the dataset) of the complex values was < 0 (Figure A.1 and Figure A.2). Using the **dDPCS** model these values were projected to 0. The fraction of values < 0 from the **DX** model and the fraction of values $= 0$ from the **dDPCS** model were the same (Figure A.2). However, values > 0 remain constant independent of the calculation (indicated by the parallel lines in Figure A.1). This differences between the model for all three datasets were not significant (Figure A.3). Negative values are mathematically possible in **DX** model but do not convey biological meaning. An attempt to explain this behavior is that the model does not capture biological variability or artifacts well enough. The assumptions for the model might be violated, and the result is no longer reliable. This was underlined by the fact that only a small fraction of values was < 0 for the simulated dataset (Section 6.1.1), which does not contain any biological variability (Figure A.1A). As mentioned, the **dDPCS** model project values < 0 to 0, however, they can still be differentiated from true 0 values by large differences between calculated and observed number of double positive partitions (Listing 4.1 Line 33). Values with a difference_to_observation larger than 1 % of $\hat{n}_{D_{obs}}$ ² positive partitions were removed from the results of both model. This reduced the differences between the model regarding the calculated number of complexes for all three datasets (Figure A.3).

¹ The variable names were adjusted to the here used nomenclature (Section 4.2)

² This threshold was used throughout this thesis to remove data points that did not fulfill the required assumptions for the **dDPCS** model (Section 4.2).

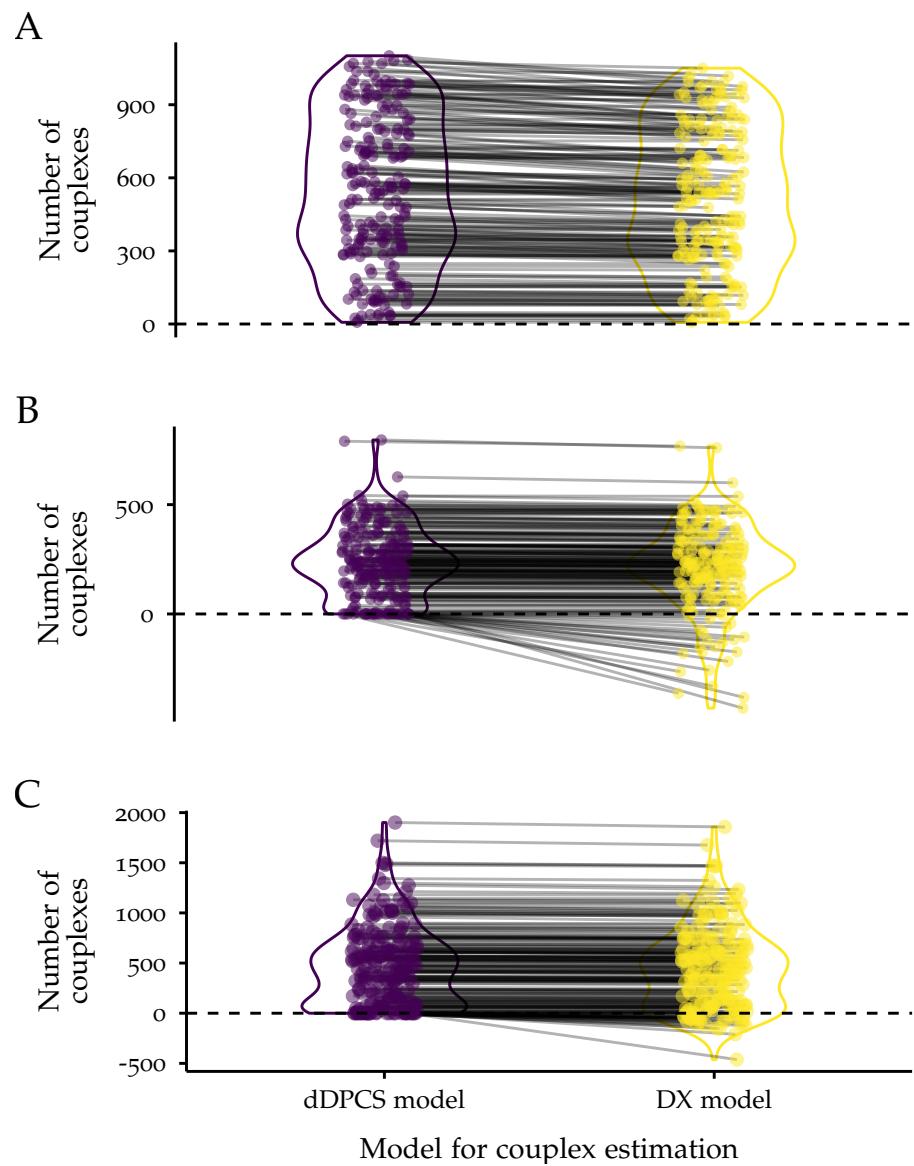


Figure A.1: Comparison of models for complex calculation. The [dDPCS](#) model ([Section 4.2](#)) and the [DX](#) (Gross et al., 2024) were applied to three datasets from [Section 6.1.1](#), [Section 6.2.3](#) and [Section 7.3](#). The [dDPCS](#) model uses a maximum-likelihood approach to calculate number of complexes, while the [DX](#) model is a closed-form expression. The gray lines connect couple values from the [dDPCS](#) and the [DX](#) model using the same input data. Each plot shows a representative subset of all data. The dots were jittered to avoid overplotting. According to [Chapter 6](#), values with $\lambda > 0.25$ were removed. [ABC](#) PICO experiments and [URABs](#) were removed, too.

A: Comparison of couple calculations using a simulated dataset from [Section 6.1.1](#).

B: Comparison of couple calculations using a dataset from [Section 6.2.3](#). Only antibodies [TTZ](#) and [PTZ](#) were included.

C: Comparison of couple calculations using a dataset from [Section 7.3](#). Only samples with antibodies [4EB1T37T46-A5](#), [60246-1-Ig](#), [E.638.3](#) and [4F3-H2](#), mock treatment, [LB](#) and [U937](#) cells were included.

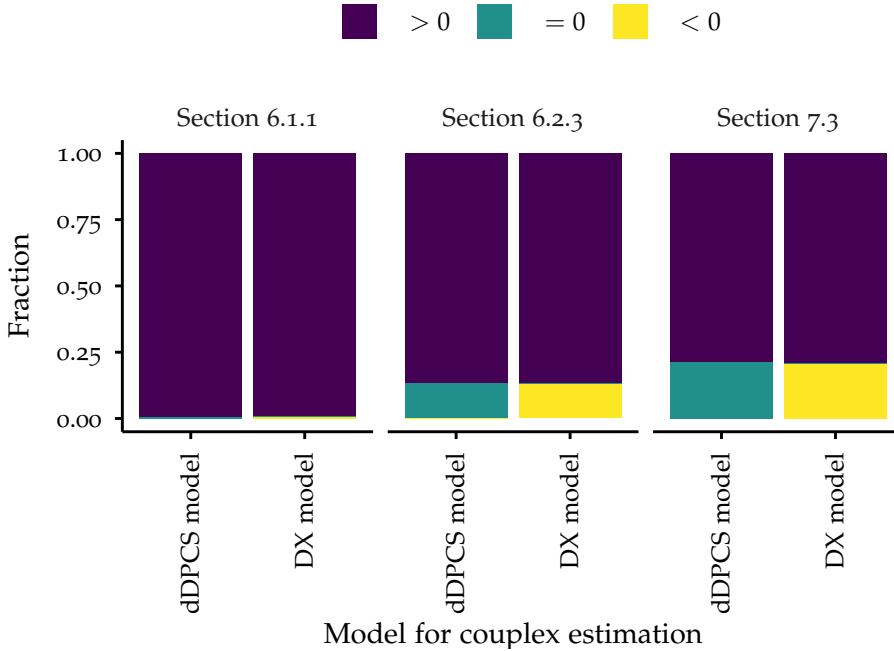


Figure A.2: Fractions of complex values in the categories > 0 , $= 0$ and < 0 .

The fraction of complex values above, equal or below 0 proportional to all values was calculated for **dDPCS** and the **DX** model using three datasets (Figure A.1)

A.2 RECONSTRUCTION OF 2-DIMENSIONAL RAW DPCR DATA

Table A.1 shows an excerpt of the *MultipleOccupancy* file from the QIAcuity dPCR Software Suite (version 2.1.7.182). The table shows the number of positive partitions for the well *B1*. The data originates from a multiplex PICO assay with four antibodies. For the identification of the number partitions for a given antibody pair, the columns *Categories*, *Group*, *Count categories* and *Valid partitions* are relevant. The + sign in the column *Group* indicates whether the corresponding fluorescent channel in the column *Categories* was considered for counting, for instance, $++--$ contains the number of partitions positive for green and yellow. Considering now that the antibody pair is detected with the fluorescent colors green and yellow (colorpair GY), the corresponding clusters can be calculated as follows

$$\hat{n}_{D_{obs}} = \hat{n}_{++++} + \hat{n}_{+++-} + \hat{n}_{+-+-} + \hat{n}_{+---} \quad (\text{A.2})$$

$$\hat{n}_{AB^+a^-b^-} = \hat{n}_{+-+-} + \hat{n}_{+---} + \hat{n}_{-++-} + \hat{n}_{-+--} \quad (\text{A.3})$$

$$\hat{n}_{AB^+a^-b^+} = \hat{n}_{-++-} + \hat{n}_{+---} + \hat{n}_{-+--} + \hat{n}_{-+--} \quad (\text{A.4})$$

$$\hat{n}_{AB^-a^-b^-} = \hat{n}_{---+} + \hat{n}_{-+--} + \hat{n}_{--+-} + \hat{n}_{----}. \quad (\text{A.5})$$

These values are subsequently fed into **dDPCS** model (Listing 4.1) for complex calculation.

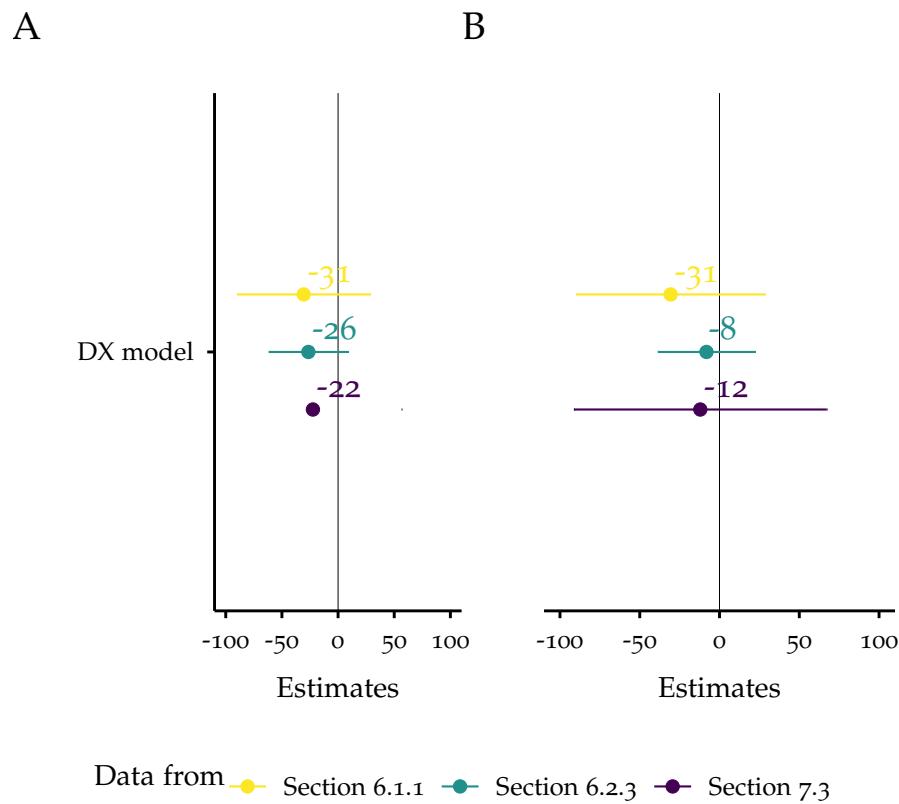


Figure A.3: Forest plots of estimates from linear models comparing the **dDPCS** and the **DX** model with 95 % **CI** as error bars. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the **CI** is the associated uncertainty.
A: Comparison of the complete datasets from Figure A.1.
B: Comparison of the datasets from Figure A.1 with a difference_to_observation smaller than 1 % of $\hat{n}_{D_{obs}}$.

Table A.1: Excerpt from QIAcuity dPCR *MultipleOccupancy* file (Software Suite version 2.1.7.182).

Sample name	Target names	Categories	Group	Count categories	Valid partitions	Volume per well (µL)
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+++	0	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+++-	11	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+-+-+	13	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	++--	100	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+-+-+	12	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+-+-+	96	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+-+-+	119	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	+-+-+	1170	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	-+++	15	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	-+ + -	133	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	-+ - +	180	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	-+ - -	1405	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	- - + +	207	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	- - + -	1637	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	- - - +	2167	25282	19.683
ABX1-ABC	Green,Yellow,Orange,Red	GREEN-YELLOW-ORANGE-RED	- - - -	18017	25282	19.683

B

DPCR SIMULATION

This appendix contains code fragments of the dPCR simulation for ideal and biased conditions. For a detail description, see [Chapter 5](#).

Listing B.1: Simulation of antibody binding to a target according to [Equation 4.20](#) and [Equation 4.21](#) under the assumption that $f_a = f_b = 1$. This means that the number of couplelexes equals the number of targets ([Line 3](#)).

```
1 # convert the number of targets into lambda of couplex
2 lambda_of_target <- number_of_targets / total_partitions
3 lambda_of_complex <- lambda_of_target * fa * fb
4
5 # subtract the complex lambda from the antibody lambda
6 lambda_of_antibodya <- lambda_of_antibodya - lambda_of_
    complex
7 lambda_of_antibodyb <- lambda_of_antibodyb - lambda_of_
    complex
```

Listing B.2: Counting of the number of partitions per cluster. `sum` is used to count the sum of all `TRUE` values in a vector. `doubles` contains the number of observed double positive partitions $\hat{n}_{D_{obs}}$. `positives_aba` contains the observed number of partitions positive for antibody a $\hat{n}_{AB^+a^+b^-}$. Similarly, `positives_abb` contains the observed number of partitions positive for antibody b $\hat{n}_{AB^+a^-b^+}$. `negatives` contains the number of observed empty partitions $\hat{n}_{AB^-a^-b^-}$.

```
1 # count the number of double positive partitions
2 doubles <- sum(partitions$double_positive)
3
4 # count the number of antibody a positive partitions (single
    positive)
5 positives_aba <- sum(partitions$antibodya_positive)
6
7 # count the number of antibody b positive partitions (single
    positive)
8 positives_abb <- sum(partitions$antibodyb_positive)
9
10 # count the number of empty partitions
11 negatives <- sum(partitions$negatives)
```

Listing B.3: Determination of ground truth double positive partitions from a simulated dPCR. In contrast to real experiments, the dPCR simulation has access to the actual number of molecules per partition (Figure 4.5A). Depending on which molecules are present in a partition, it is assigned to the corresponding cluster. A partition is random double positive when it contains antibody a and antibody b. A partition is *rc-overlap* double positive when it contains antibody a, antibody b and a complex. `sum` is used to count the sum of all `TRUE` values in the vector. `random_doubles` contains the number of random double positive partitions and `rc_overlap_doubles` contains the number of *rc-overlap* double positive. These results can be compared to the output of the dDPCS model.

```

1  library(dplyr)
2
3  partitions <- partitions %>%
4    mutate(
5      # determine if a partition only contains both antibodies
6      random_double = ifelse(
7        (antibodya > 0 & antibodyb > 0)
8        TRUE,
9        FALSE
10      ),
11      # determine if a partition contains both antibodies and
12      # complexes
13      rc_overlap = ifelse(
14        antibodyb > 0 & antibodya > 0 & couplex > 0,
15        TRUE,
16        FALSE
17      )
18
19      # count the number random double positive partitions
20      random_doubles <- sum(partitions$random_double)
21
22      # count the number of rc_overlap double positive partitions
23      rc_overlap_doubles <- sum(partitions$rc_overlap)

```

Listing B.4: Simulation of false-negative clustering. Similar to Listing 5.2, partitions are assigned to the corresponding cluster depending on the molecules present. However, `runif(n())` generates a number between 0 and 1, which needs to be higher than the `clustering_bias` for a correct positive decision. `clustering_biases` can be defined for each individual cluster and simulated separately or simultaneously.

```

1 library(dplyr)
2
3 partitions <- partitions %>%
4   mutate(
5     double_positive = ifelse(
6       (antibodya > 0 & antibodyb > 0) |
7       couplex > 0,
8       runif(n()) > clustering_bias_doubles,
9       FALSE
10    ),
11    antibodya_positive = ifelse(
12      antibodya > 0 | couplex > 0,
13      runif(n()) > clustering_bias_aba,
14      FALSE
15    ),
16    antibodyb_positive = ifelse(
17      antibodyb > 0 | couplex > 0,
18      runif(n()) > clustering_bias_abb,
19      FALSE
20    ),
21    negative = ifelse(
22      (antibodya == 0 & antibodyb == 0 & couplex == 0),
23      runif(n()) > clustering_bias_negatives,
24      FALSE
25    )
26  )

```

Listing B.5: Simulation of false-positive clustering. Similar to Listing 5.2, partitions are assigned to the corresponding cluster depending on the molecules present. However, `runif(n())` generates a number between 0 and 1, which needs to be higher than the `clustering_bias` for a correct negative decision. `clustering_biases` can be defined for each individual cluster and simulated separately or simultaneously.

```

1  library(dplyr)
2
3  partitions <- partitions %>%
4    mutate(
5      double_positive = ifelse(
6        (antibodya > 0 & antibodyb > 0) |
7          couplex > 0,
8        TRUE,
9        runif(n()) > clustering_bias_doubles
10    ),
11    antibodya_positive = ifelse(
12      antibodya > 0 | couplex > 0,
13      TRUE,
14      runif(n()) > clustering_bias_aba
15    ),
16    antibodyb_positive = ifelse(
17      antibodyb > 0 | couplex > 0,
18      TRUE,
19      runif(n()) > clustering_bias_abb
20    ),
21    negative = ifelse(
22      (antibodya == 0 & antibodyb == 0 & couplex == 0),
23      TRUE,
24      runif(n()) > clustering_bias_negatives
25    )
26  )

```

Listing B.6: Simulation of molecular dropouts. `dropout_threshold` defines the maximal number of amplifiable molecules per partition. The `surplus_molecules` are calculated, split randomly into three parts ([Listing B.7](#)) and subtracted from antibodya, antibodyb and complexes. Finally, it is checked if there are negative molecule counts, if yes, there are artificially set to zero. This is an oversimplification as negative number of molecules per partition are not possible. After the subtraction, the regular clustering is performed ([Listing 5.2](#)).

```

1  library(dplyr)
2  library(tidyr)
3
4  if (dropout_threshold != 0) {
5      partitions <- partitions %>%
6          mutate(
7              # calculate sum of molecules per partition
8              total_molecules = antibodya + antibodyab +
9                  couplex,
10             # calculate number of surplus molecules
11             surplus_molecules = ifelse(
12                 molecules_total > dropout_threshold,
13                 molecules_total - dropout_threshold,
14                 0
15             ),
16             # split the surplus molecules into three parts
17             # to subtract them from antibodya, antibodyb and
18                 couplex
19             subtract_values = pmap(
20                 list(surplus_molecules),
21                 random_split
22             )
23         ) %>%
24             # random_split retruns as list
25             # unnest_wider places the elements in separate
26                 columns
27             unnest_wider(subtract_values) %>%
28                 # subtract surplus molecules from the respective
29                 molecules
30                 # if the subtraction results in negative values,
31                 # the number of molecules is reset to zero
32             mutate(
33                 antibodya = ifelse(
34                     antibodya < subtracta,
35                     0,
36                     antibodya - subtracta
37                 ),
38                 antibodyb = ifelse(
39                     antibodyb < subtractb,
40                     0,
41                     antibodyb - subtractb
42                 ),
43                 couplex = ifelse(
44

```

```

40                  couplex < subtractc,
41                  0,
42                  couplex - subtractc
43              )
44          )
45      }

```

Listing B.7: Randomly splitting a number into three parts. The surplus molecules determined in [Listing B.6](#) are divided into three parts, which will be subtracted from the number of antibodies a, b or complexes per partition.

```

1  random_split <- function(value) {
2      parta <- sample(0:value, 1)
3      partb <- sample(0:(value - parta), 1)
4      partc <- value - parta - partb
5      return(
6          list(
7              subtracta = parta,
8              subtractb = partb,
9              subtractc = partc
10         )
11     )
12 }

```

C

DDPCS MODEL VALIDATION

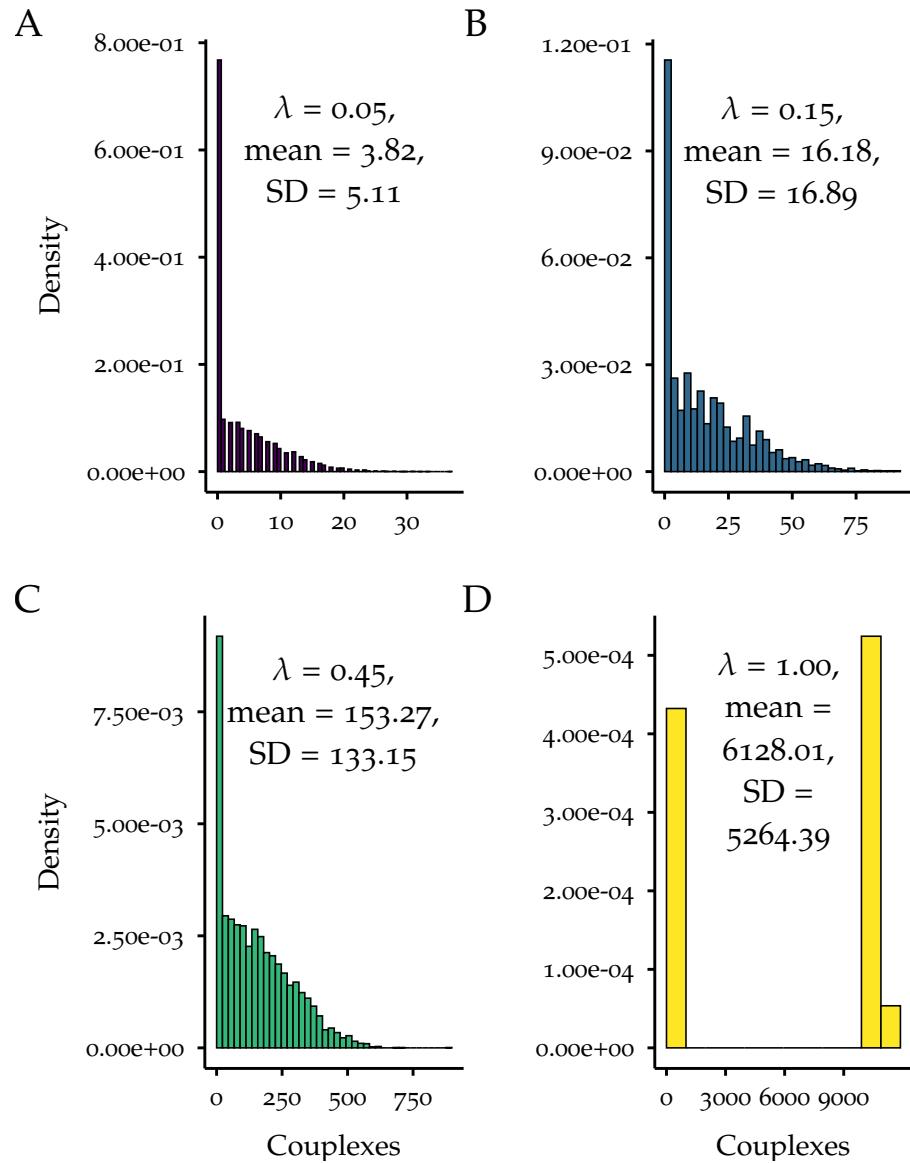


Figure C.1: Selected histograms of simulated complexes from ABC PICO experiments for $\lambda = \{0.05, 0.15, 0.45, 1.00\}$ with mean and SD.
 A: Histogram of calculated complexes from $\lambda = 0.05$. According to its shape, this distribution is in the *narrow range* (Figure 6.1).
 B: Histogram of calculated complexes from $\lambda = 0.15$. According to its shape, this distribution is in the *narrow range* (Figure 6.1).
 C: Histogram of calculated complexes from $\lambda = 0.45$. According to its shape, this distribution is in the *tailing range* (Figure 6.1).
 D: Histogram of calculated complexes from $\lambda = 1.00$. According to its shape, this distribution is in the *satellite range* (Figure 6.1).

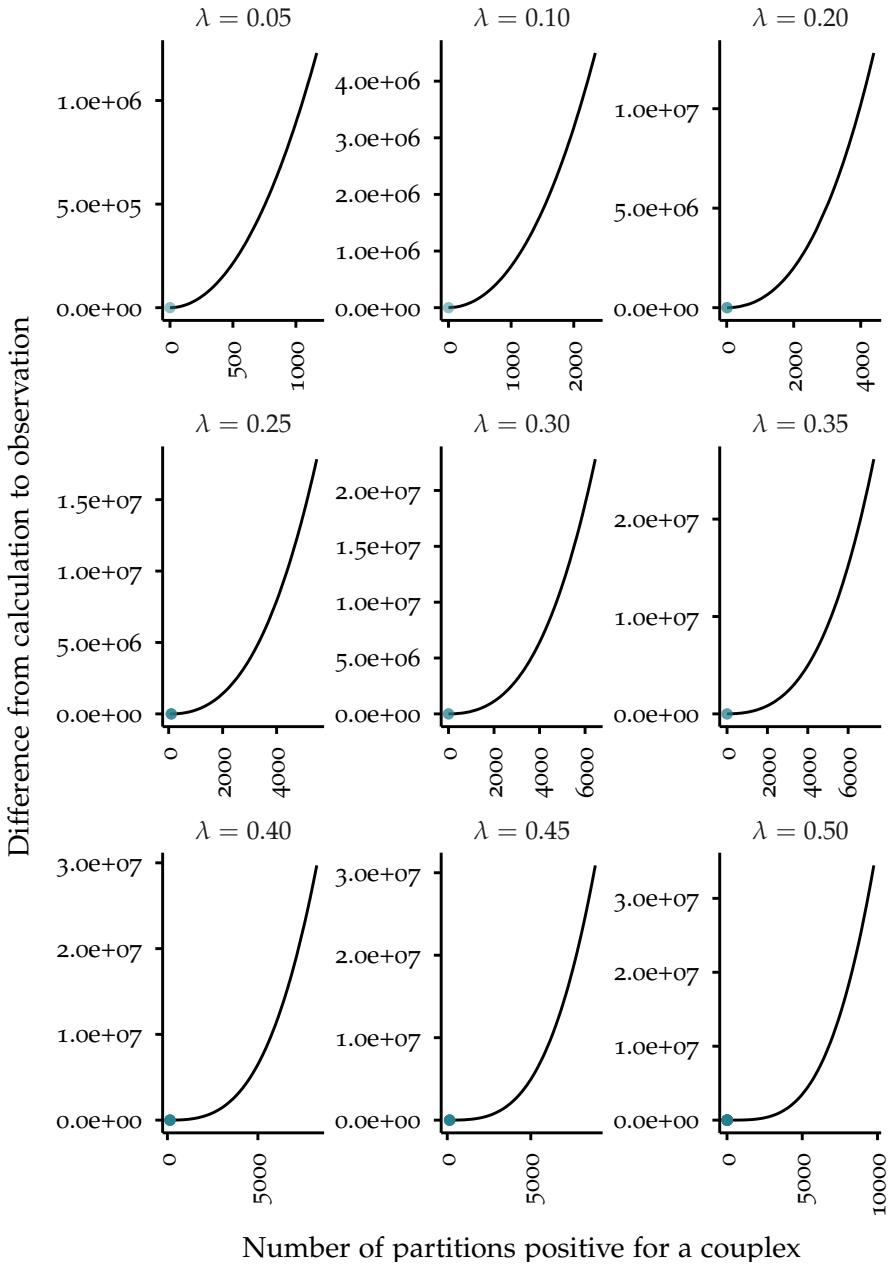


Figure C.2: Maximum-likelihood function for simulated λ values for the calculation of the number of complexes using the [dDPCS](#) model (part 1). The maximum-likelihood function calculates the difference between the calculated and the observed number of double positive partitions ($(\hat{n}_{D_{obs}} - n_{D_{calc}})^2$) for all $\hat{n}_{AB+a^-b^-}$ in the range from 0 to $\hat{n}_{D_{obs}}$ in steps of 1 partition. The turquoise dots indicate the minimum/minima.

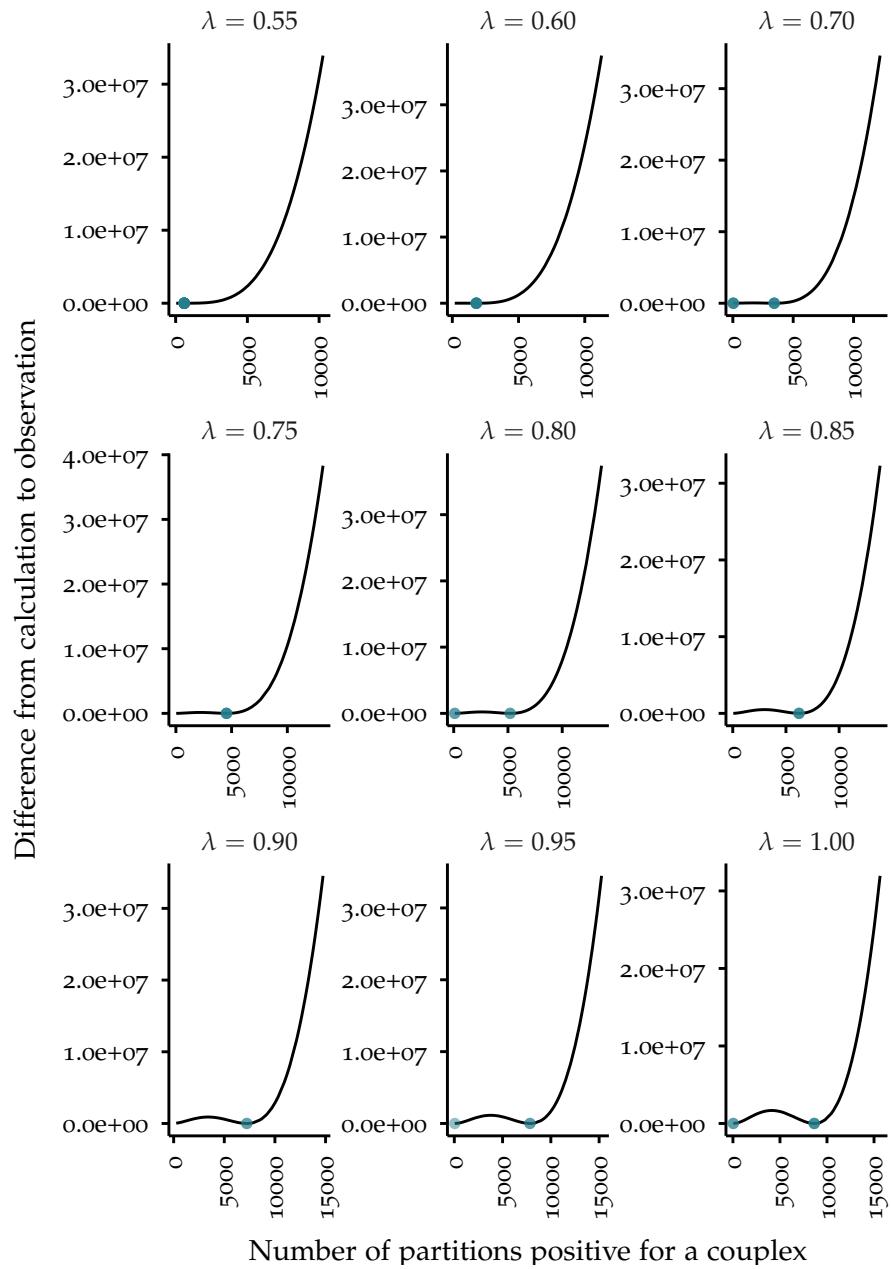


Figure C.3: Maximum-likelihood function for simulated λ values for the calculation of the number of complexes using the [dDPCS](#) model (part 2). The maximum-likelihood function calculates the difference between the calculated and the observed number of double positive partitions ($(\hat{n}_{D_{obs}} - n_{D_{calc}})^2$) for all \hat{n}_{AB+a-b^-} in the range from 0 to $\hat{n}_{D_{obs}}$ in steps of 1 partition. The turquoise dots indicate the minimum/minima.

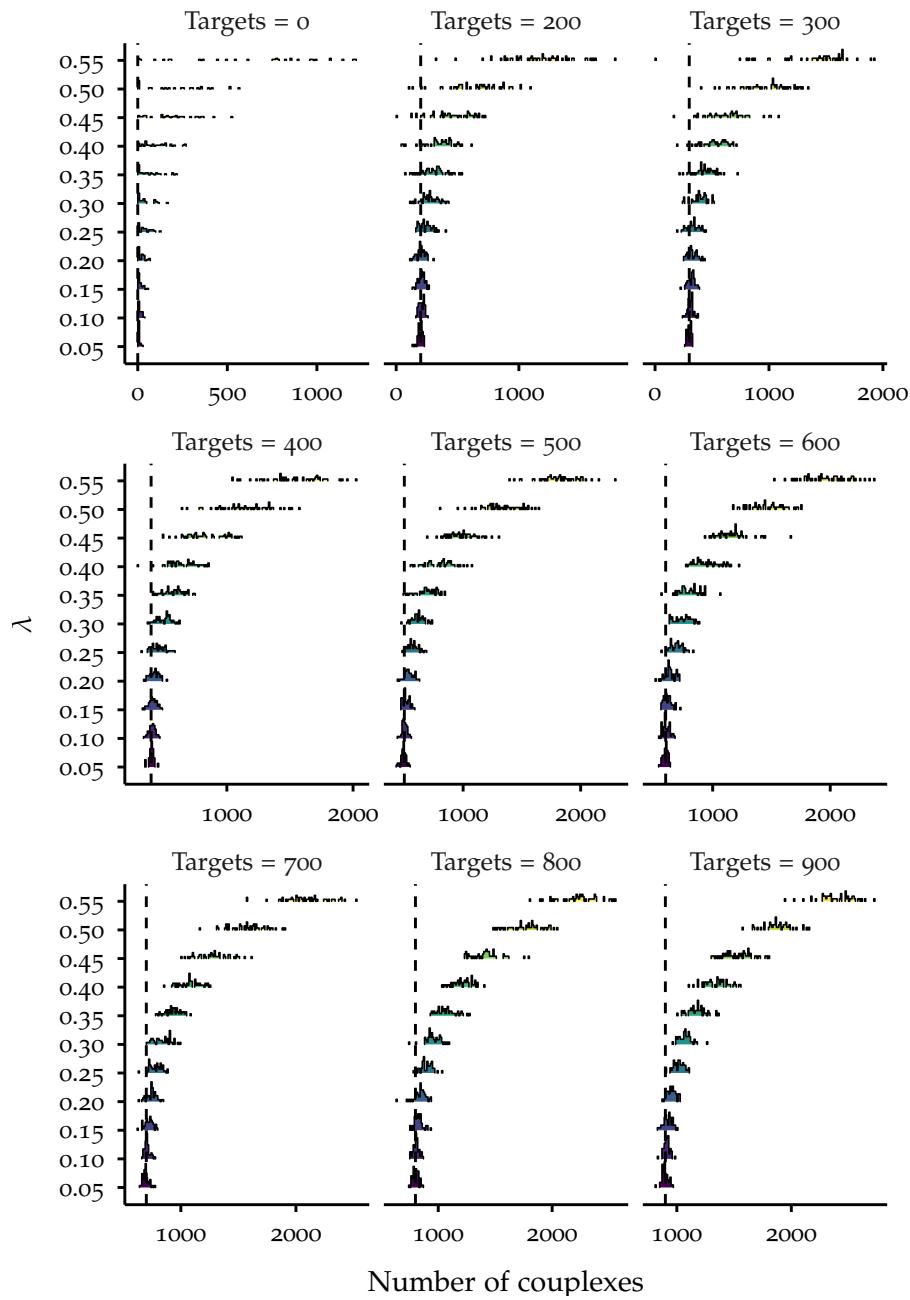


Figure C.4: Ridgeline plot using histograms of complex distributions calculated from simulated PICO experiments using the dDPCS model. λ was varied from 0.05 to 0.55 in steps of 0.05 (color-coded). The number of targets was varied in $\{0, 200, 300, 400, 500, 600, 700, 800, 900\}$ (indicated by the vertical dashed lines in each panel). Each combination of λ and number of targets was simulated 50 times. The bin width is 10 couplexes.

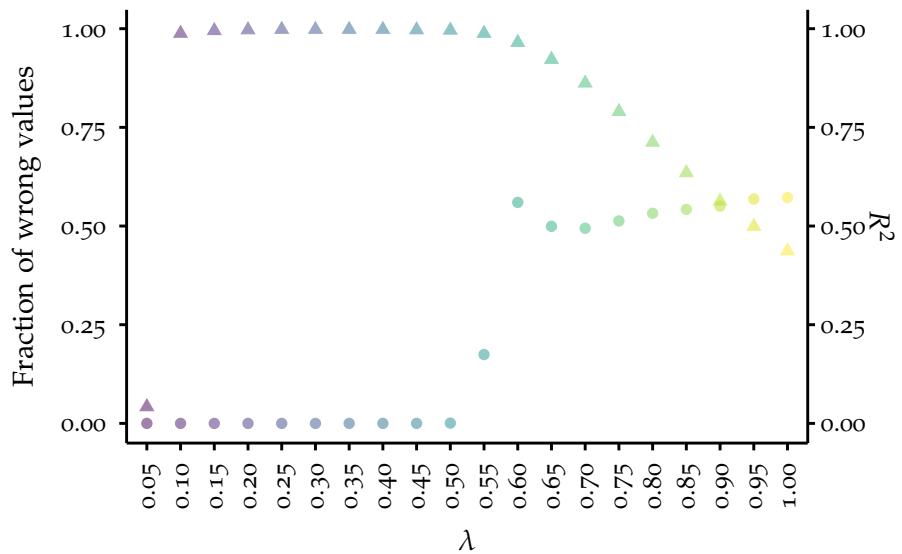


Figure C.5: Characterization of the ground truth comparisons. The fraction of wrong values (values in the lower branch of Figure 6.4A) is shown (dots) per value of λ (color-coded). The second y-axis depicts the R^2 (triangles) of the iterative linear model from Figure 6.4B per value of λ (color-coded).

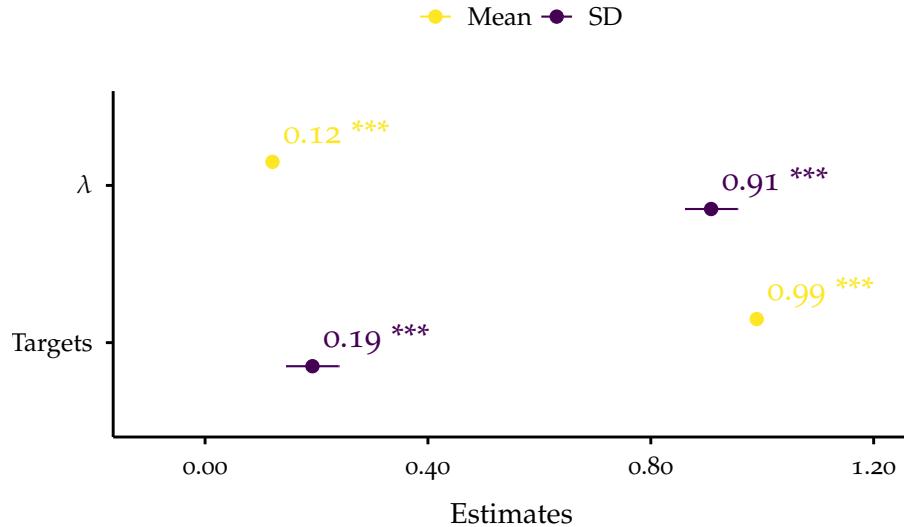


Figure C.6: Forest plot of estimates with 95 % CI from the comparison of complex distributions. The linear models `lm(calculated_complexes ~ expected_lambda + targets)` (yellow) and `lm(calculated_SD_complexes ~ expected_lambda + targets)` (purple) were calculated for the λ range $0.05 \leq \lambda \leq 0.25$. The predictors were standardized using [Equation 9.2](#). The estimate is the change in the response variable for a 1 SD increase in the predictor variable and the [CI](#) is the associated uncertainty. The adjusted R^2 for the first model was 0.995 and for the second model 0.861.

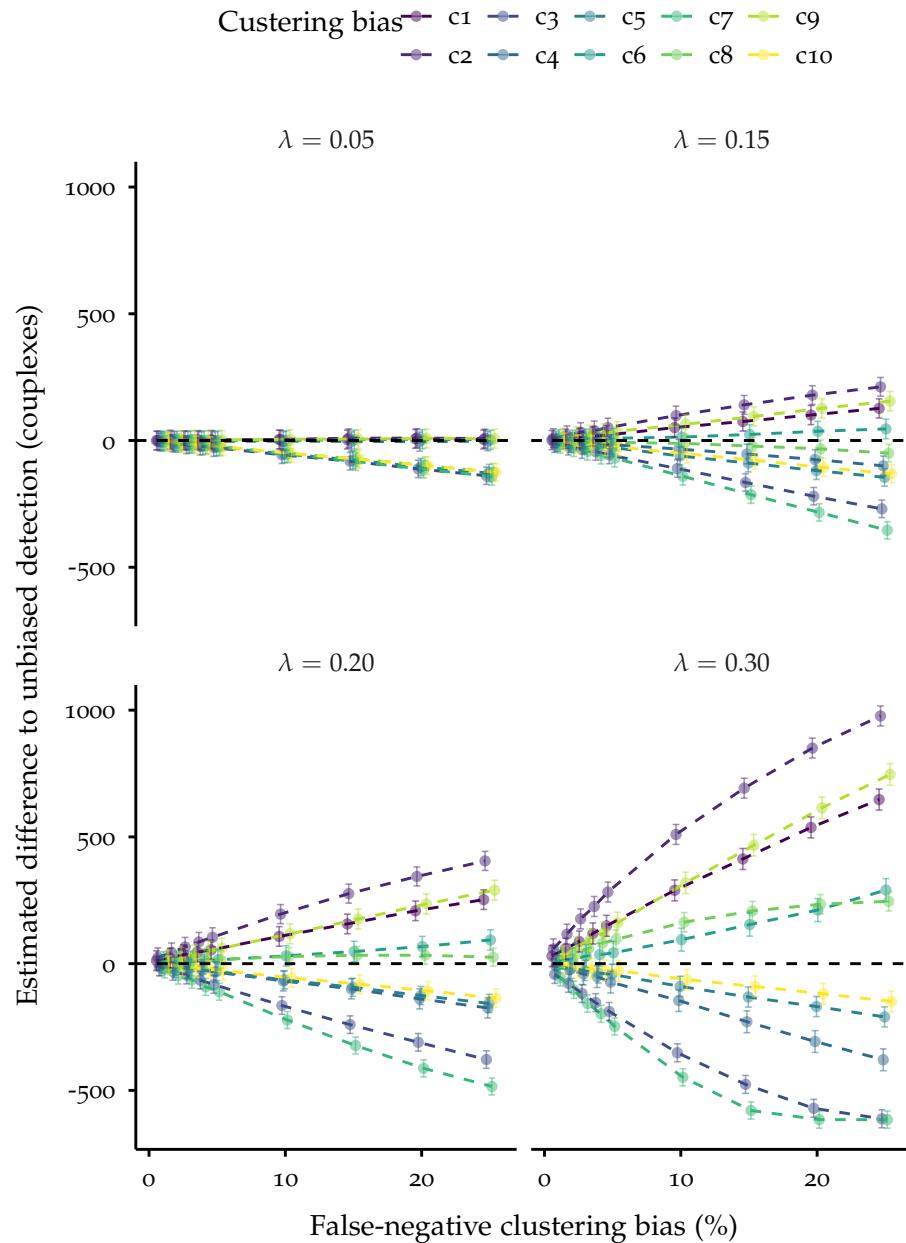


Figure C.7: Estimated difference between false-negative clustering bias and ideal conditions for $\lambda = \{0.05, 0.15, 0.20, 0.30\}$ form the linear model `lm(calculated_couplexes ~ as.factor(detection_error))` with 95 % CI as error bars. The clusters \hat{n}_{AB+a+b^-} , \hat{n}_{AB+a-b^+} , $\hat{n}_{D_{obs}}$ and \hat{n}_{AB-a-b^-} (Figure 4.8A) were biased in the combinations c1 to c10 (Table 6.1). The strength of the clustering bias was varied in $\{1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$. The dots, lines and error bars were jittered to avoid overplotting.

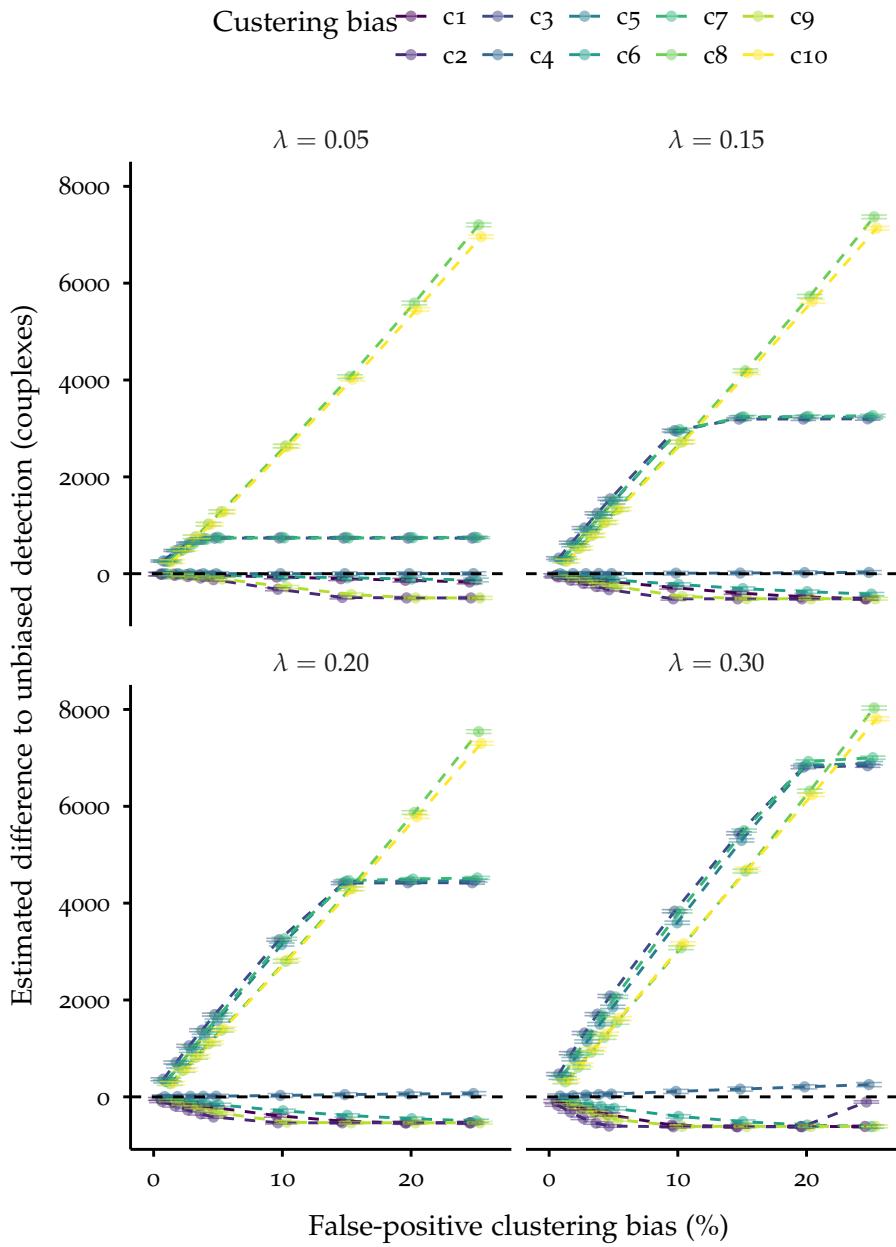


Figure C.8: Estimated difference between false-positive clustering bias and ideal conditions for $\lambda = \{0.05, 0.15, 0.20, 0.30\}$ form the linear model `lm(calculated_couplexes ~ as.factor(detection_error))` with 95 % CI as error bars. The clusters $\hat{n}_{AB+a+b-}$, $\hat{n}_{AB+a-b+}$, $\hat{n}_{D_{obs}}$ and $\hat{n}_{AB-a-b-}$ (Figure 4.8A) were biased in the combinations c1 to c10 (Table 6.1). The strength of the clustering bias was varied in $\{1, 2, 3, 4, 5, 10, 15, 20, 25\}\%$. The dots, lines and error bars were jittered to avoid overplotting.

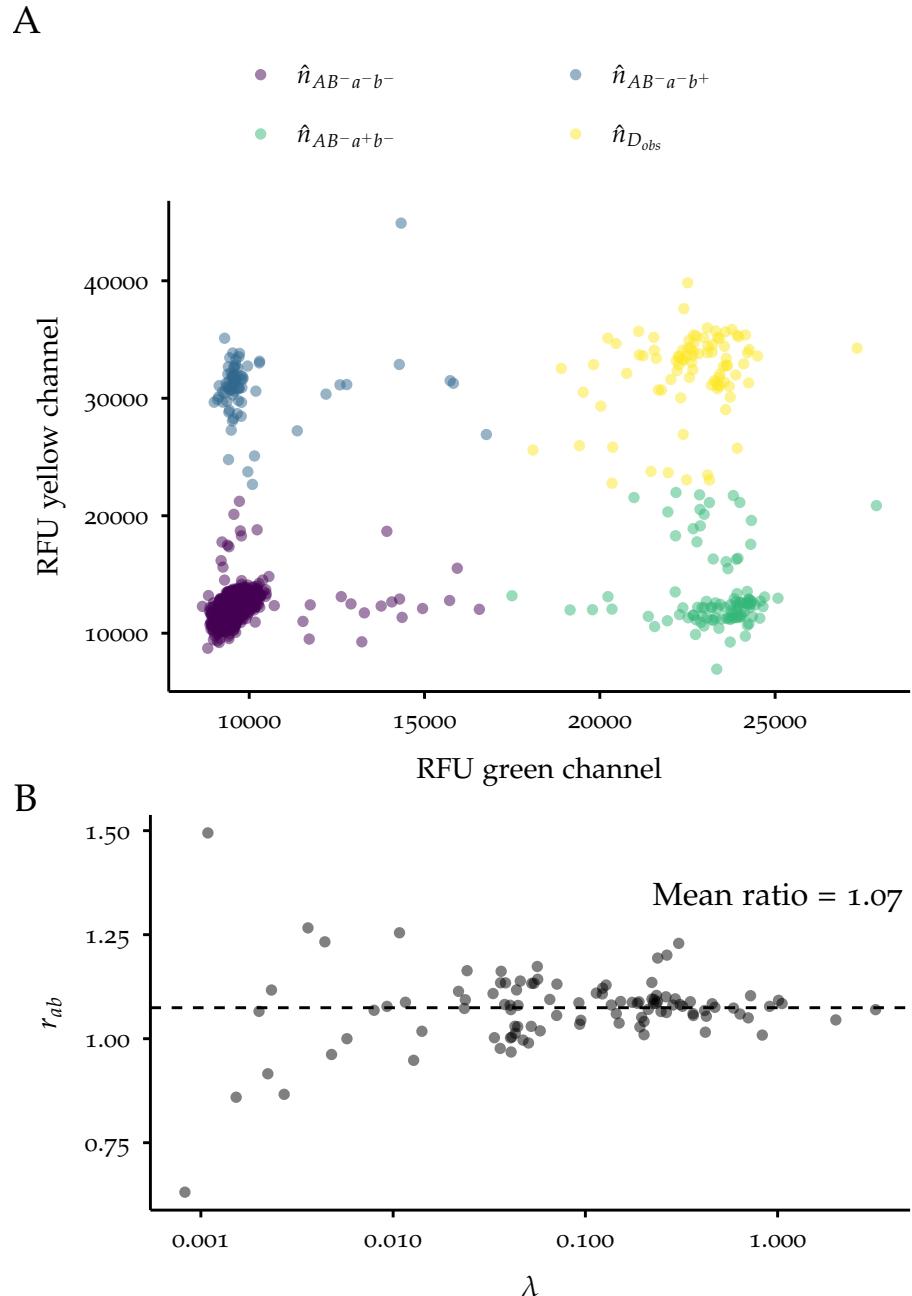


Figure C.9: Analysis of the artificial couplex.

A: Representative scatterplot of artificial couplex with $\lambda \approx 0.15$. Each dot represents a partition with its corresponding RFU value in the green and the yellow channel. This scatterplot shows a representative subset from originally 21 092 partitions with 18 253 empty partitions (\hat{n}_{AB-a-b^-}), 1187 double positive partitions ($\hat{n}_{D_{obs}}$), 880 single positive partitions in the green channel (\hat{n}_{AB-a+b^-}) and 772 single positive partitions in the yellow channel (\hat{n}_{AB-a-b^+}).

B: Scatterplot of r_{ab} (Equation 6.1) for different values of λ of the artificial couplex. Data is from experiments performed in naica and QIAcuity dPCR systems.

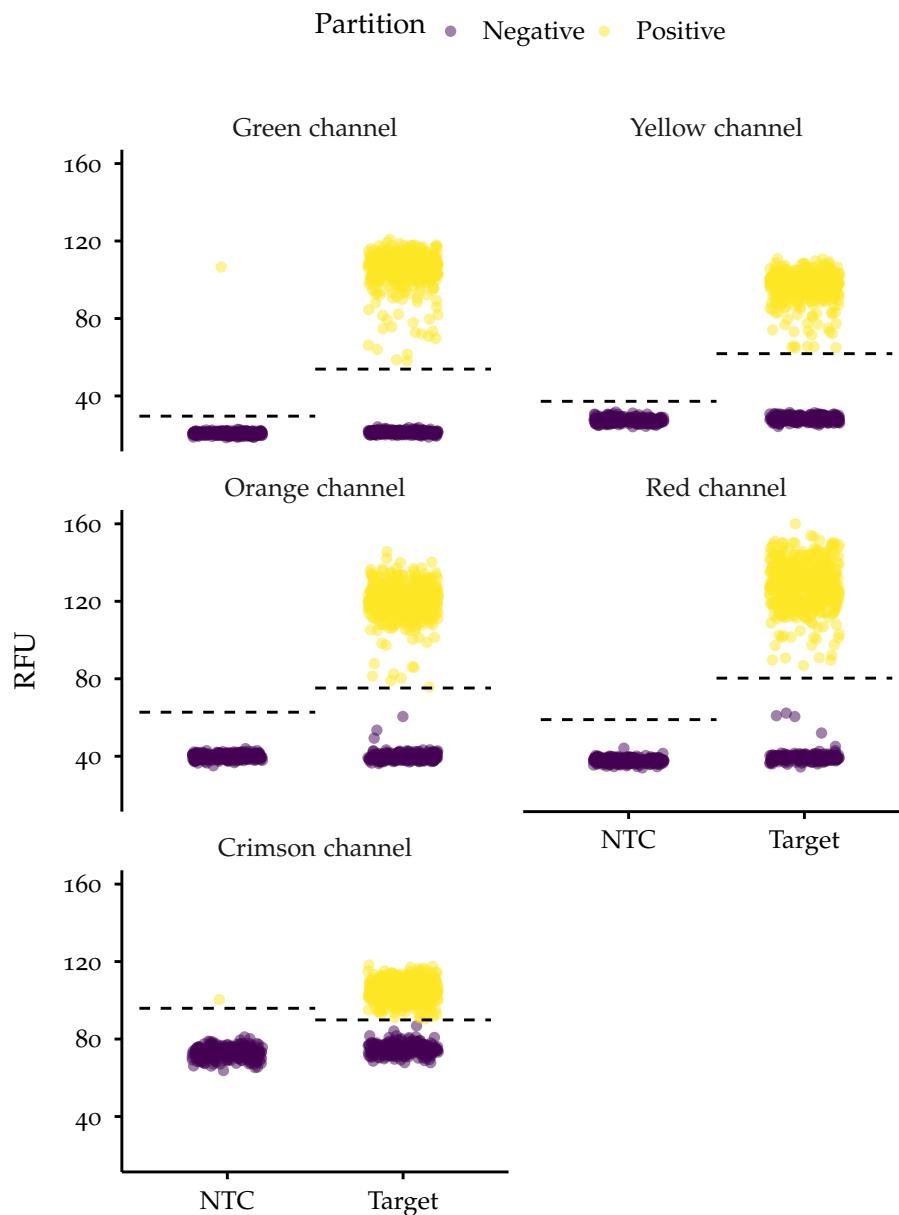


Figure C.10: Scatterplots of RFU for partitions detected in the fluorescent channels of the QIAcuity dPCR system. The fluorescent dyes were FAM (green channel), HEX (yellow channel), Atto550 (orange channel), TxRed (red channel) and Cy5 (crimson channel). For each channel a representative scatterplot with a subset of the total partitions shows the detection of a target and a non-template control (NTC). The dots were jittered to avoid overplotting.

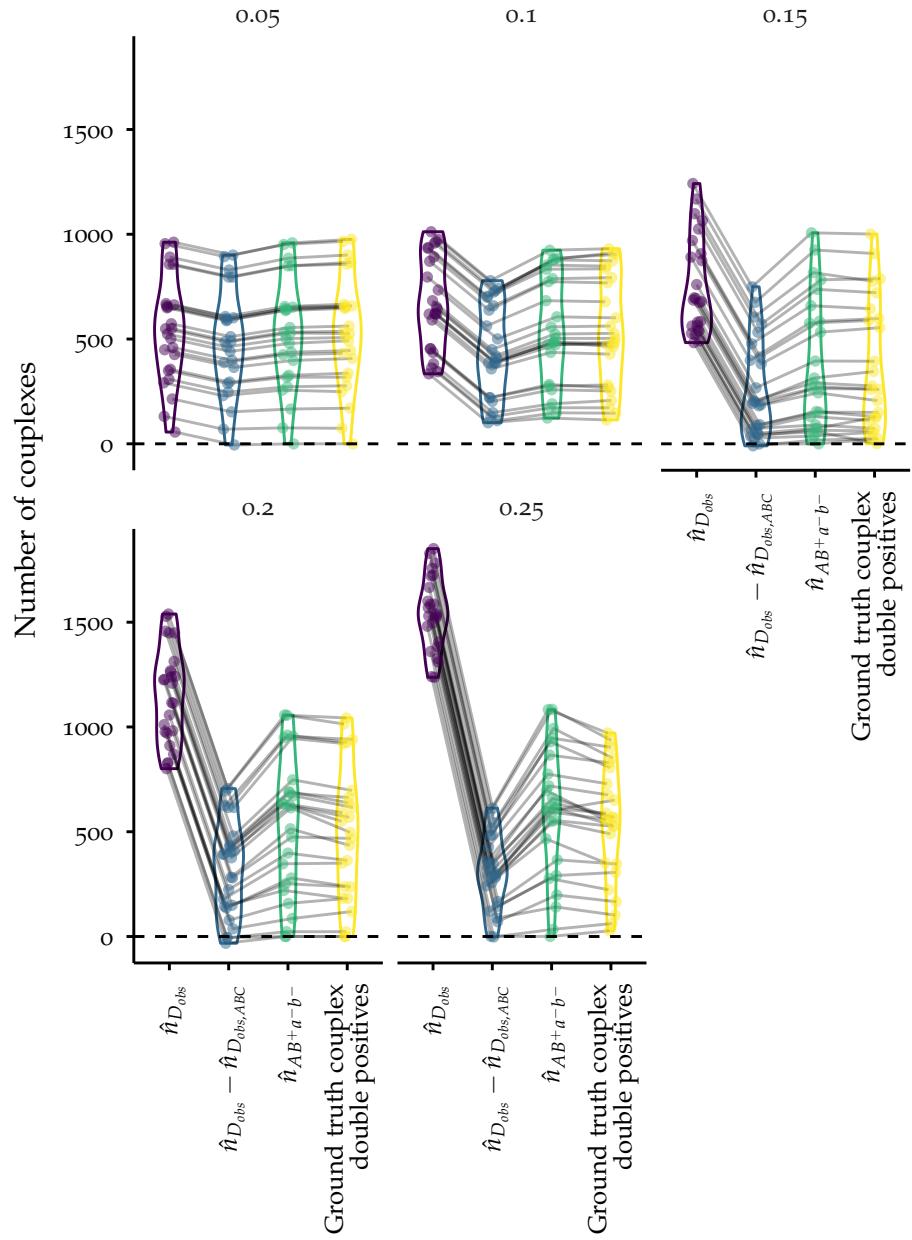


Figure C.11: Impact of the dDPCS model using simulated data (Section 6.1.1) for different values of λ . The number of complex double positive partitions from the dDPCS model ($\hat{n}_{AB^+a^-b^-}$) was compared to the number of observed double positive partitions from the dPCR ($\hat{n}_{D_{obs}}$), to the observed number of double positive partitions subtracted by the number of observed double positive partitions from the corresponding ABC ($\hat{n}_{D_{obs}} - \hat{n}_{D_{obs,ABC}}$) and to the ground truths (Figure 6.6). The plot shows a representative subset of all data. The dots were jittered to avoid overplotting. Grey lines connect the same value after the different processing steps.

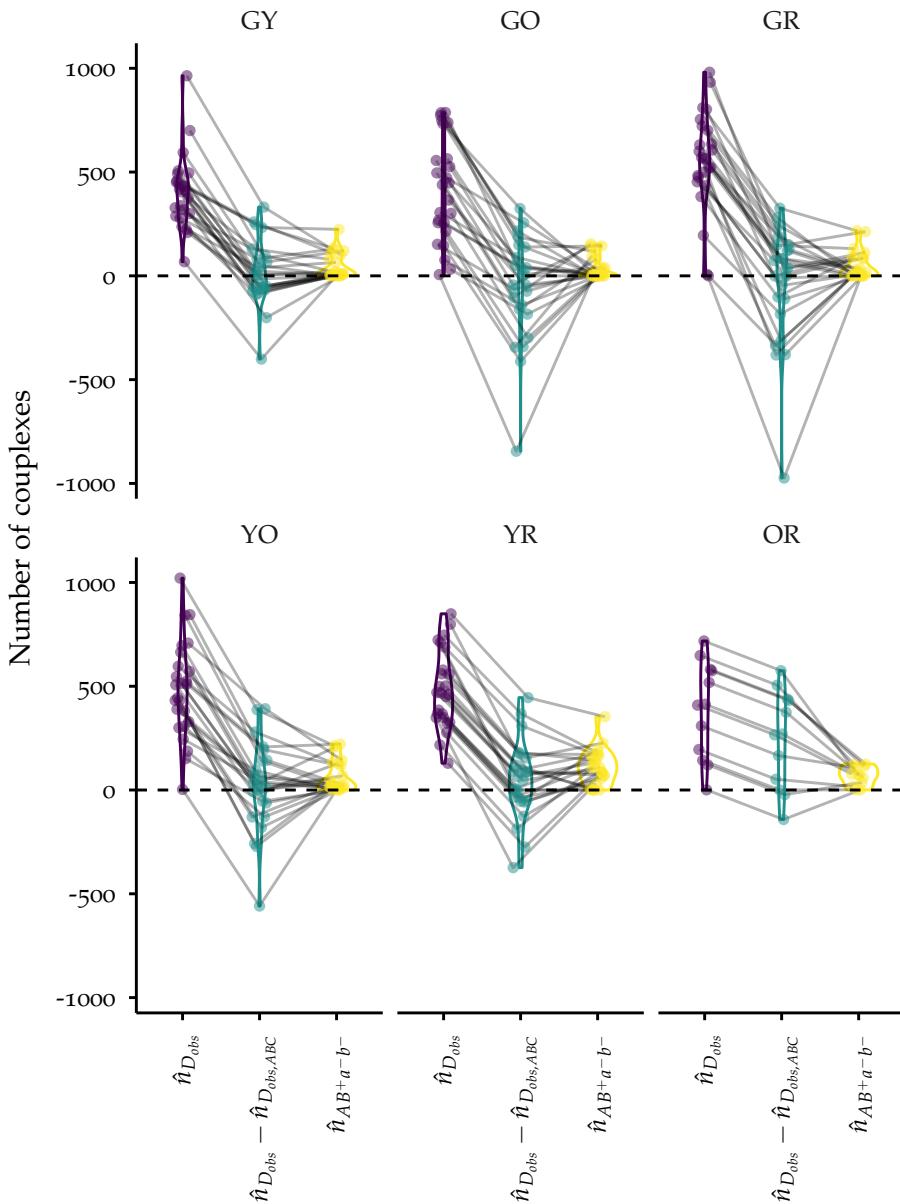


Figure C.12: Impact of the dDPCS model using recombinant protein detected in different colorpairs (Section 6.2.3). The number of complex double positive partitions from the dDPCS model ($\hat{n}_{AB^+a^-b^-}$) was compared to the number of observed double positive partitions from the dPCR ($\hat{n}_{D_{obs}}$), to the observed number of double positive partitions subtracted by the number of observed double positive partitions form the corresponding ABC ($\hat{n}_{D_{obs}} - \hat{n}_{D_{obs,ABC}}$) and to the ground truths (Figure 6.6). The plot shows a representative subset of all data. The dots were jittered to avoid overplotting. Grey lines connect the same value after the different processing steps.

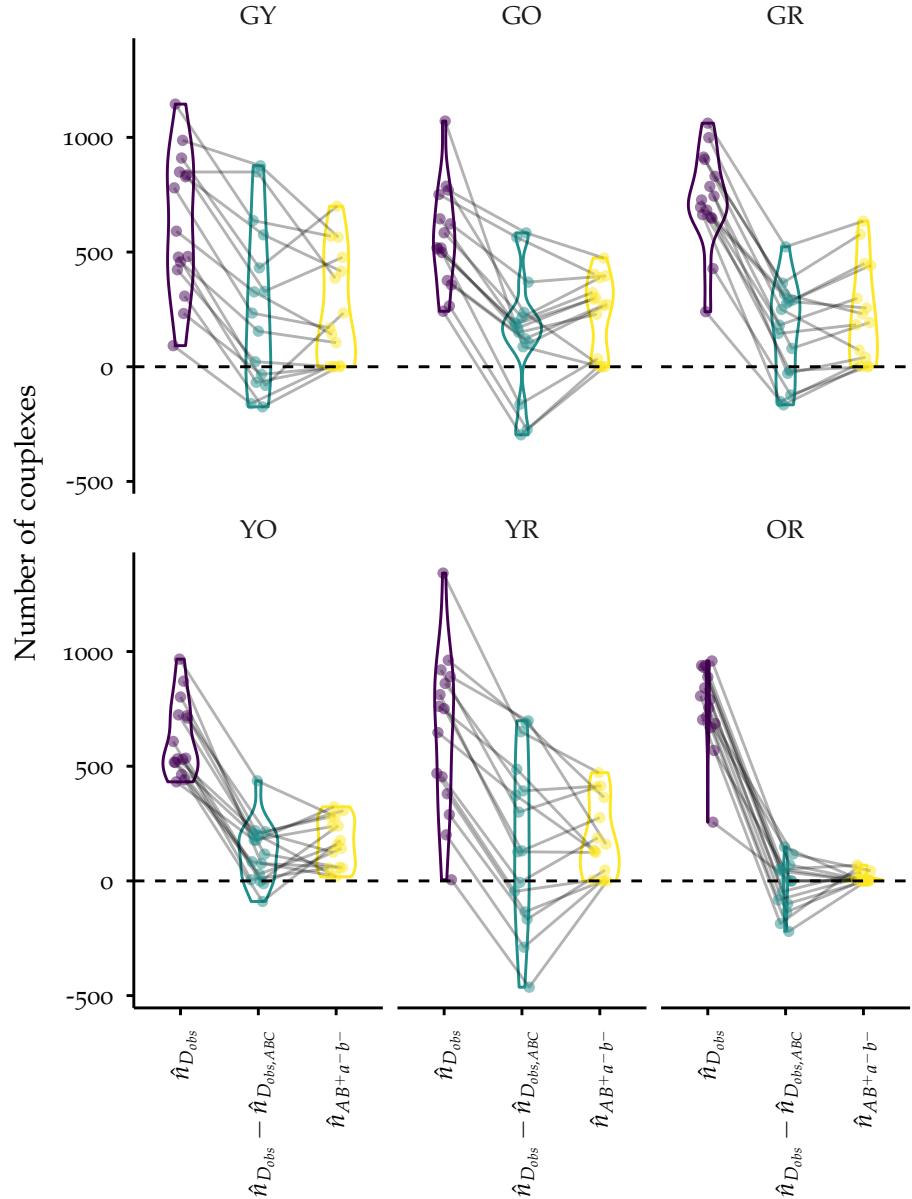


Figure C.13: Impact of the **dDPCS** model using data from the **4EBP1** PICO assay on U937 cell lysate (Section 7.3). The number of complex double positive partitions from the **dDPCS** model ($\hat{n}_{AB^+a^-b^-}$) was compared to the number of observed double positive partitions from the dPCR ($\hat{n}_{D_{obs}}$), to the observed number of double positive partitions subtracted by the number of observed double positive partitions from the corresponding ABC ($\hat{n}_{D_{obs}} - \hat{n}_{D_{obs,ABC}}$) and to the ground truths (Figure 6.6). The plot shows a representative subset of all data. The dots were jittered to avoid overplotting. Grey lines connect the same value after the different processing steps.

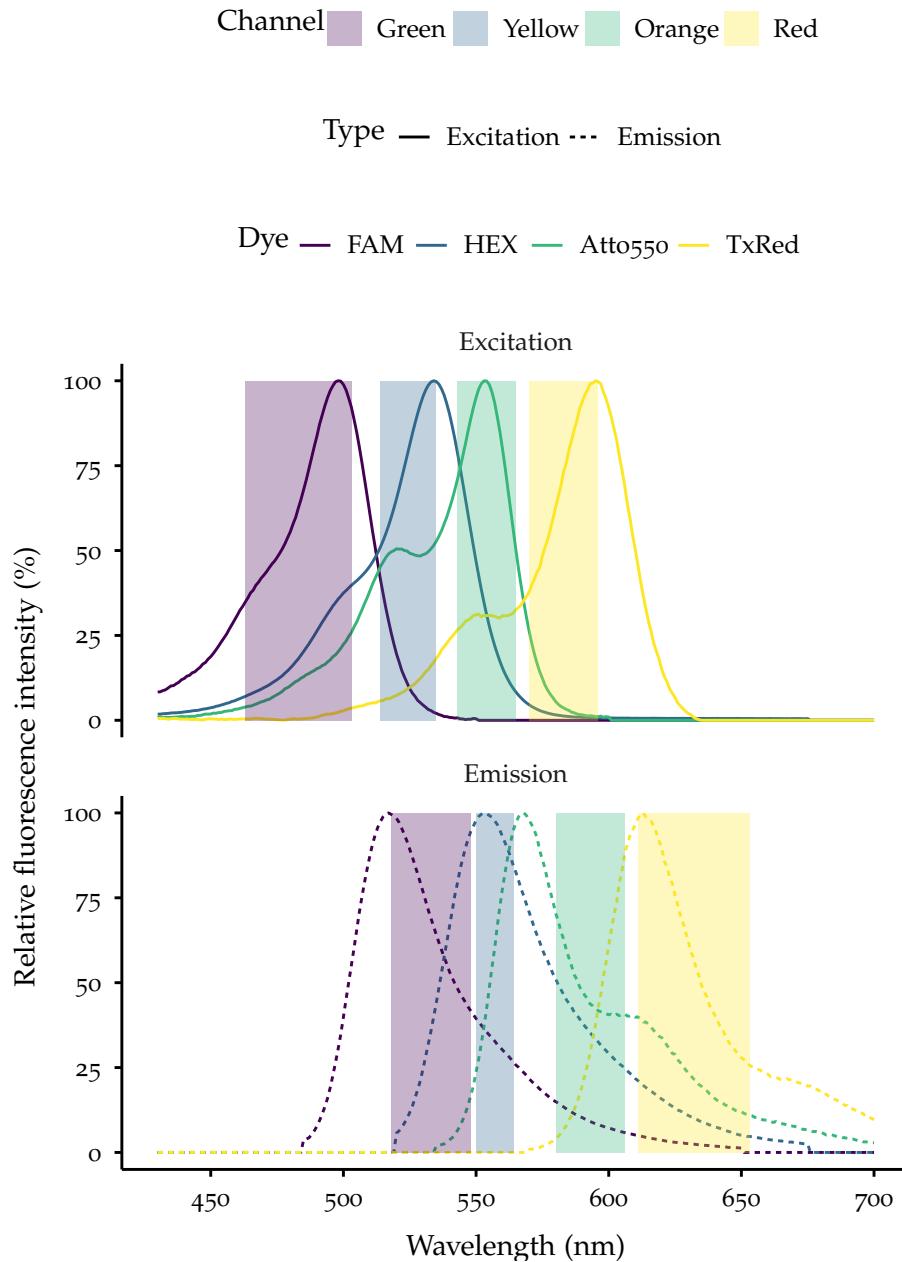


Figure C.14: Fluorescent channels of the QIAcuity dPCR system and used fluorescent dyes with excitation and emission wavelengths. The spillover coefficient for the HEX dye and the Atto550 dye in the bandwidth of the yellow emission filter is 63 %, and spillover coefficient for the Atto550 dye and the TxRed dye in the bandwidth of the orange emission filter is 75 %.

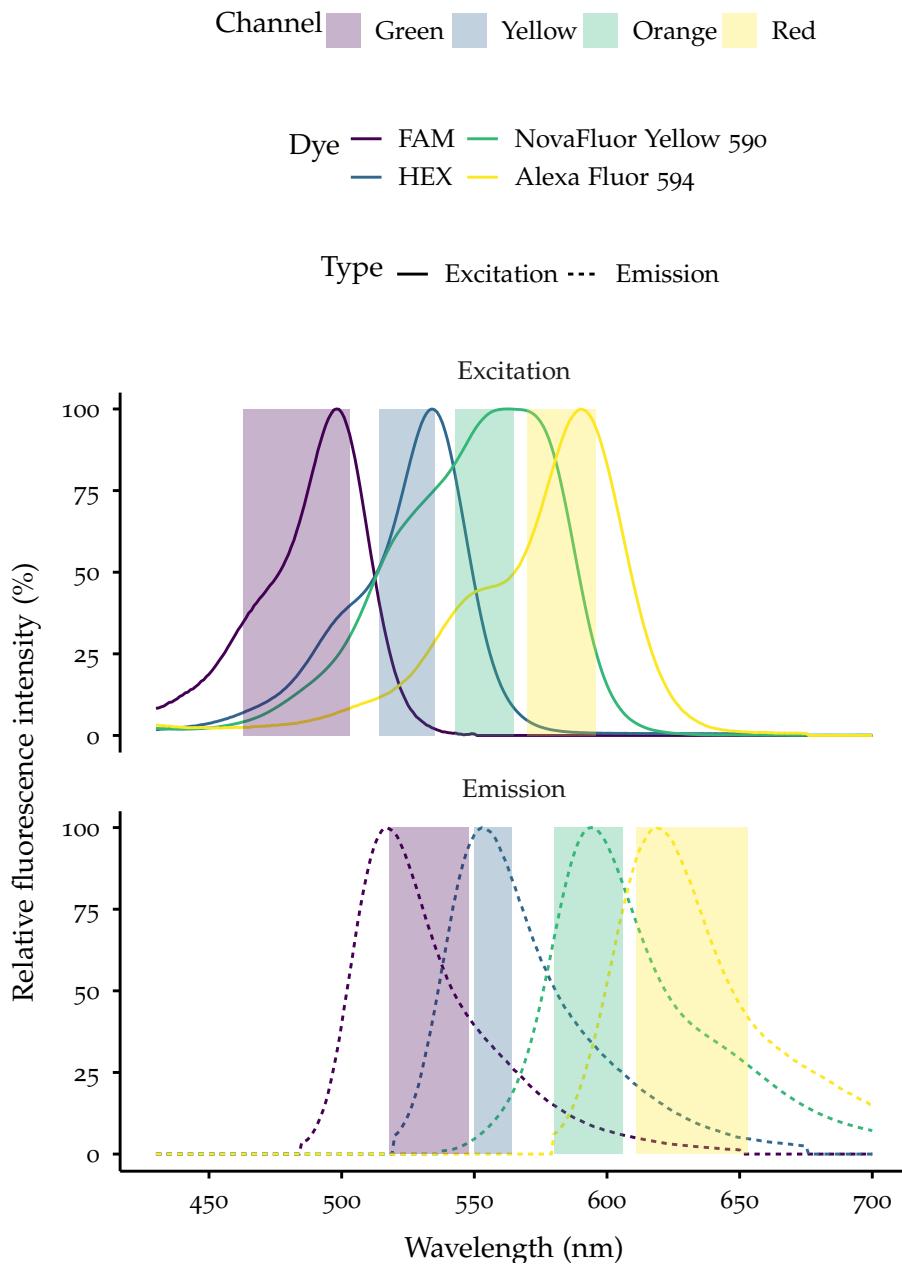


Figure C.15: Fluorescent channels of the QIAcuity dPCR system and optimized fluorescent dyes with excitation and emission wavelengths. The spillover coefficient for the HEX dye and the NovaFluor Yellow 590 dye in the bandwidth of the yellow emission filter is 11 %, and the spillover coefficient for the NovaFluor Yellow 590 dye and the AlexaFluor 594 dye in the bandwidth of the orange emission filter is 35 %. This corresponds to 5.73-fold and 2.14-fold improvements for the yellow and the orange channels, respectively (Figure C.14).

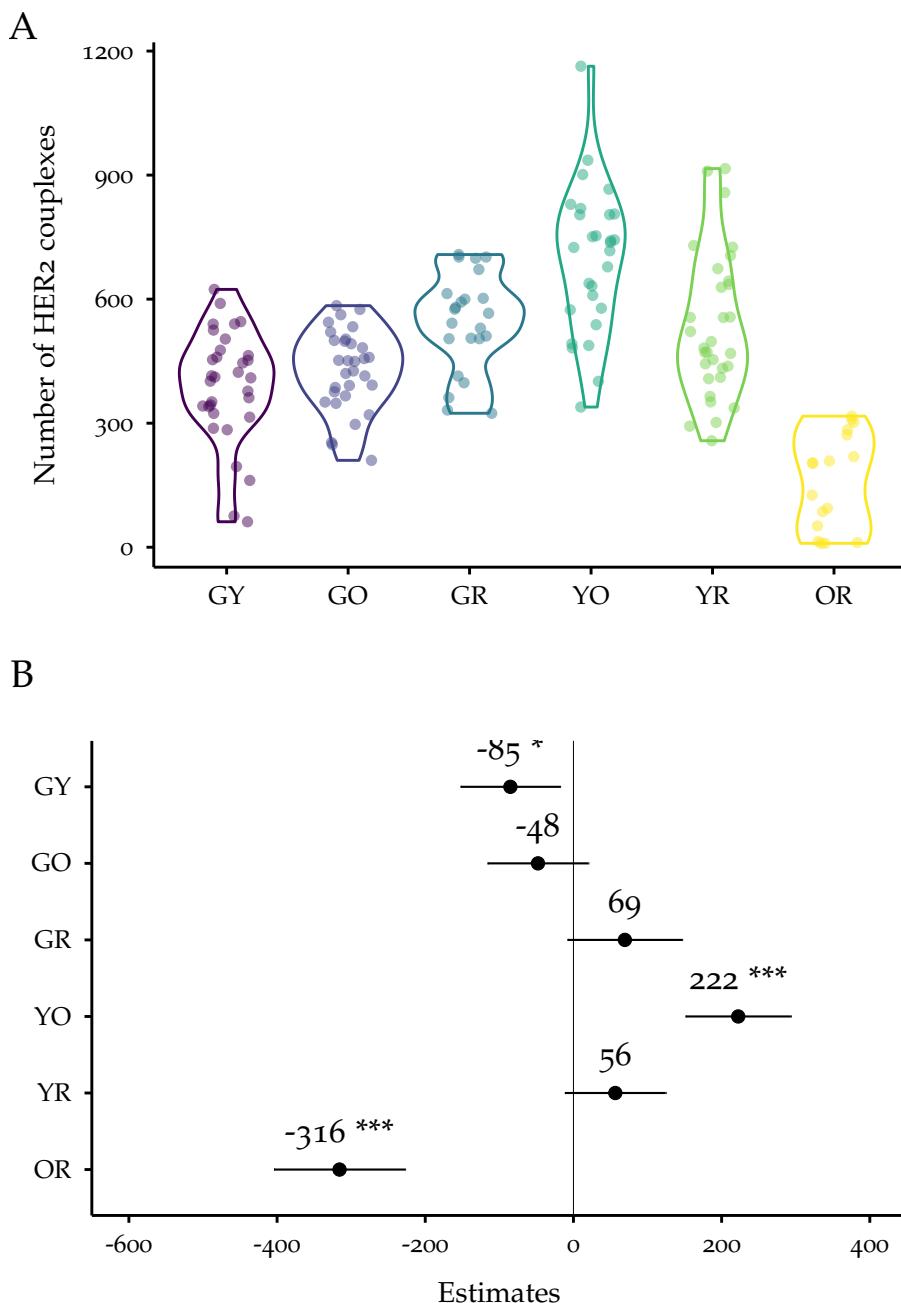


Figure C.16: Detection of recombinant HER2 complexes using TTZ and PTZ in different colorpairs with LE correction according to (Gross et al., 2024). Data is from 2 independent experiments. Data was filtered for $\lambda \leq 0.25$.

A: Violin plot of number of LE-corrected complexes per colorpair, sample sizes were $n_{GY} = 32$, $n_{GO} = 31$, $n_{GR} = 23$, $n_{YO} = 28$, $n_{YR} = 31$ and $n_{OR} = 17$. The dots were jittered to avoid overplotting.

B: Forest plot of estimates with 95 % CI from the linear model `lm` (`calculated_complexes ~ colorpair`) with the overall average as the reference. The estimate is the change in the response variable for a 1 unit increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for the model was 0.253.

Table C.1: Combinations of TTZ and PTZ for colorpair comparison.

Colorpair	Antibody a	DNA-label a	Antibody b	DNA-label b	Total LE (%)
GY	PTZ	P8	TTZ	BLA	60.2
GO	PTZ	P8	TTZ	NOS6	57.2
GR	PTZ	P8	TTZ	ORC7	42.8
YO	TTZ	BLA	PTZ	NOS6	81.2
YR	TTZ	BLA	PTZ	ORC7	90.9
OR	PTZ	NOS6	TTZ	ORC7	57.7

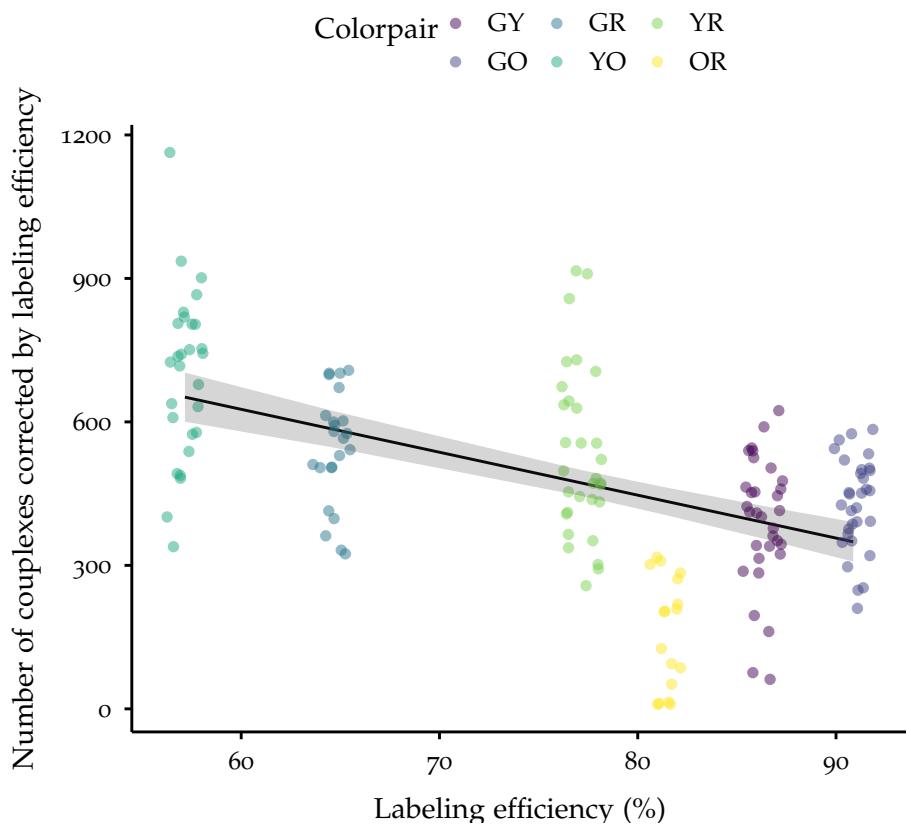


Figure C.17: Scatterplot of number of LE-corrected complexes against the LE. The linear model `lm(calculated_complexes.LE_corrected ~ LE)` estimates a change of -8.97 (95 % CI from -11.20 to -6.74) for the LE-corrected complexes with the LE. The dots were jittered to avoid overplotting. Data was from 2 independent experiments. The sample sizes per colorpair can be obtained from Figure C.16.

D

APPLICATION EXAMPLE: BIOMARKER IN LEUKEMIA

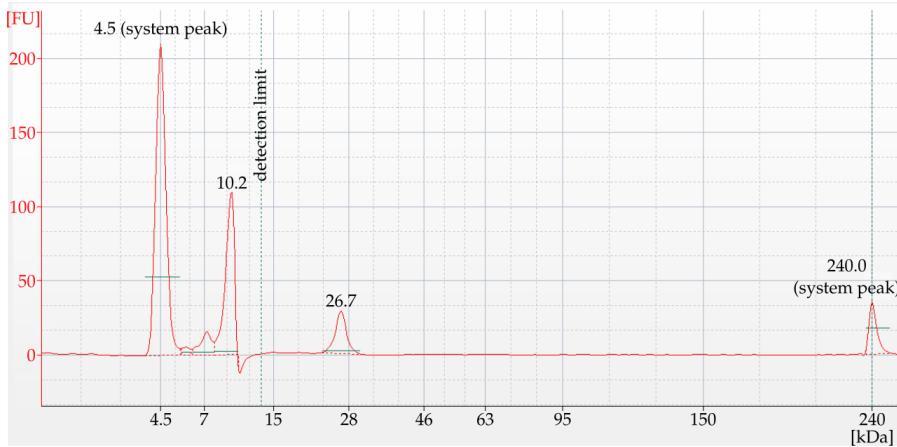


Figure D.1: Representative electropherogram of recombinant 4EBP1. The electrophoresis was run under denaturing and reducing conditions (Section 9.2.4). Apart from the system peaks, there is only one further peak visible at 26.7 kDa.

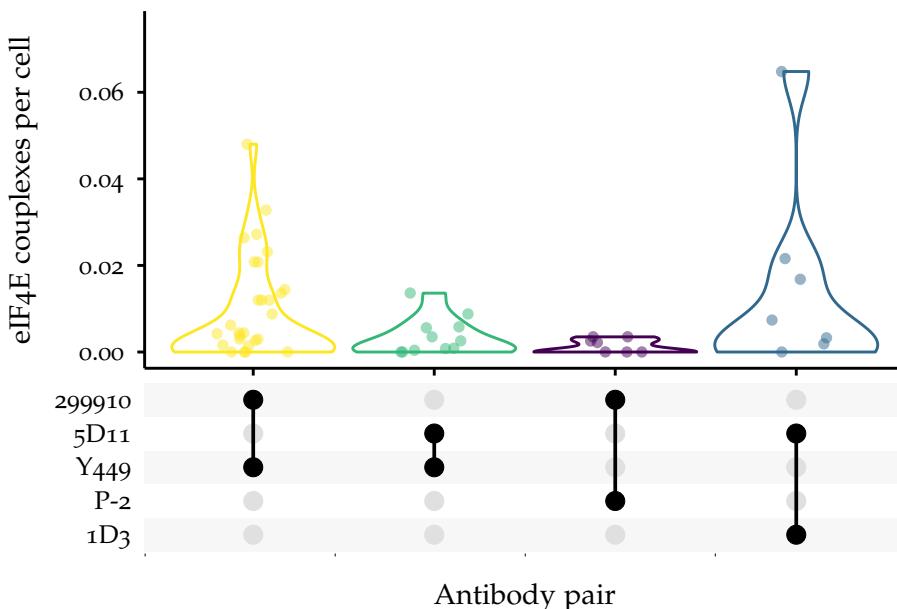


Figure D.2: Detection of eIF4E complex using different antibodies.

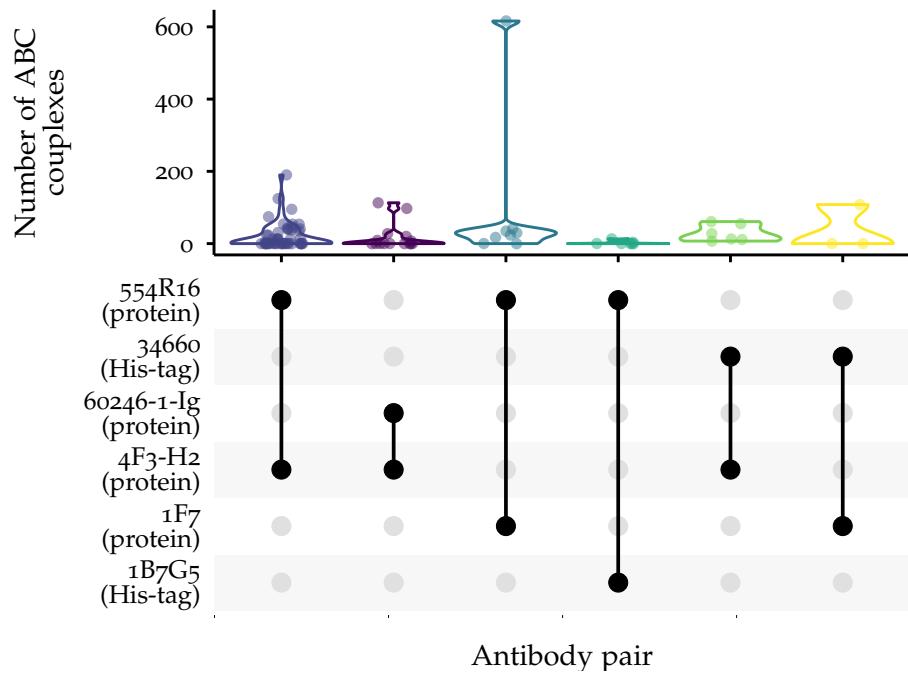


Figure D.3: UpSet plot of number of ABC complexes using different 4EBP_1 antibodies. Sample sizes were 48, 14, 8, 7, 7, 6, 3 (from left to right).

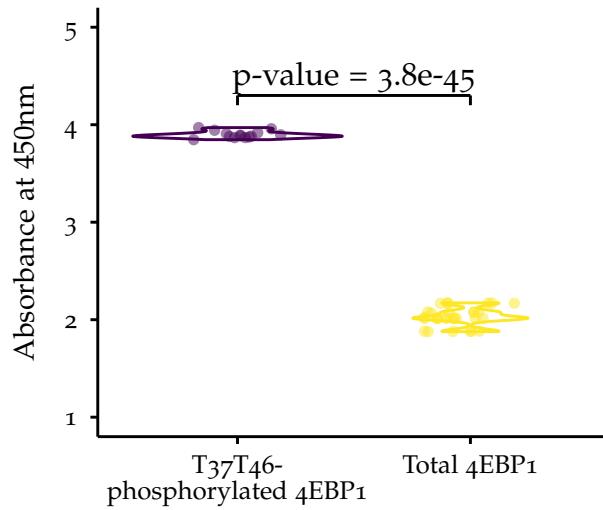


Figure D.4: Violin plots for comparison of total 4EBP_1 and T₃₇T₄₆-phosphorylated 4EBP_1 using ELISA. The linear model `lm(` absorbance ~ 4EBP_1 _assay) calculated a difference of 1.87 units of absorbance with 95 % CI = 1.81 to 1.92. Sample sizes were 30 and 14 from left to right.

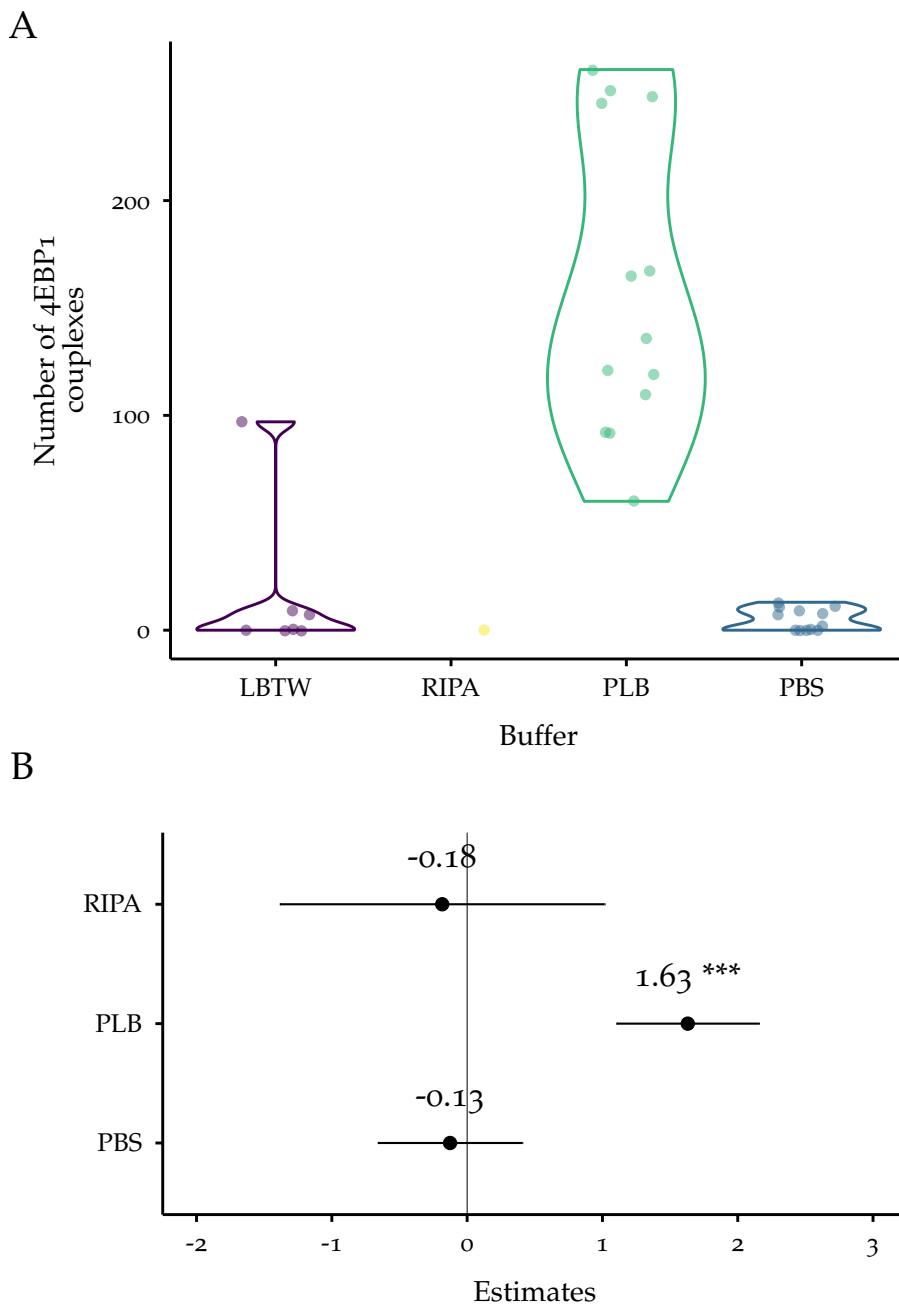


Figure D.5: Comparison of ABCs from different buffers using the antibody pair 60246-1-Ig + 4F3-H2. Sample sizes were 7, 1, 13, 12 (from left to right).

A: Violin plots of number of ABC complexes for different buffers.
 B: Forest plot of the estimates with 95 % CI from the linear model `lm(couplexes ~ buffer)` with LB as reference. Data was standardized using [Equation 9.2](#). The estimate is the change in the response variable for a 1 SD increase in the predictor variable and the CI is the associated uncertainty. The adjusted R^2 for the model was 0.699.

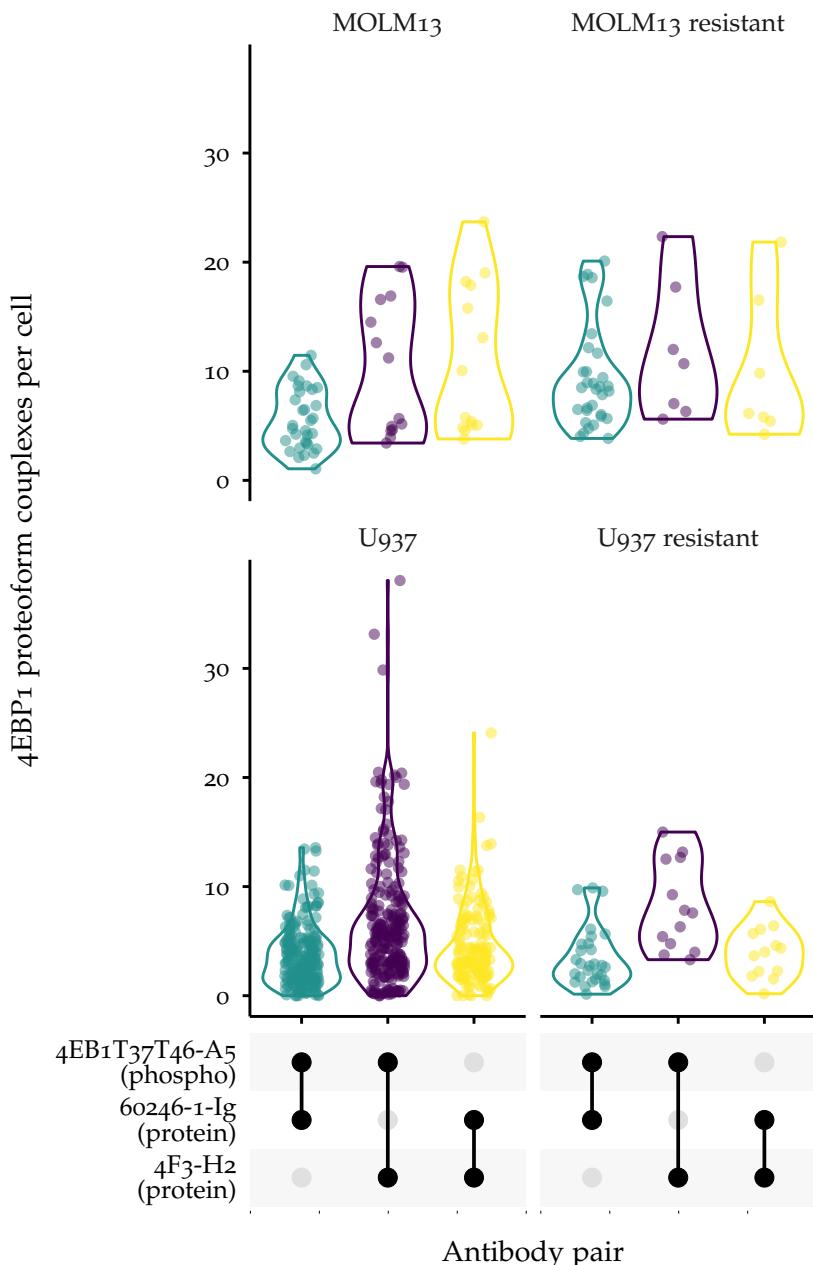


Figure D.6: Violin plots for comparison of (HMA-resistant) AML cell lines using the validated 4EBP_1 PICO assay (Figure 7.4). Data was from at least 2 independent experiments. Sample sizes were (from left to right): MOLM13: 32, 14, 14; MOLM13 resistant: 32, 7, 7; U937: 206, 241, 143; U937 resistant: 29, 13, 13. Low sample sizes do not necessarily mean that experiments were not thoroughly repeated but that this plot depicts already filtered results. The dots were jittered to avoid overplotting.

EXPERIMENTAL PROCEDURE

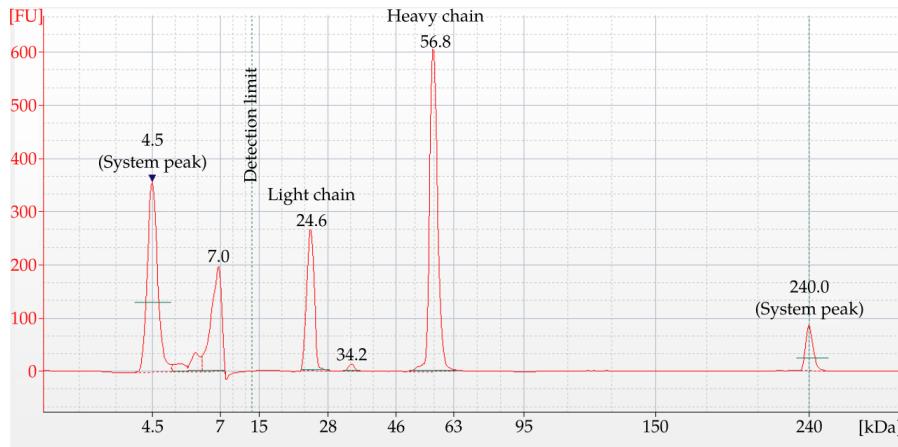


Figure E.1: Representative electropherogram from unlabeled antibody (clone 1F7 (Table 10.4)). The electrophoresis was run under denaturing and reducing conditions.

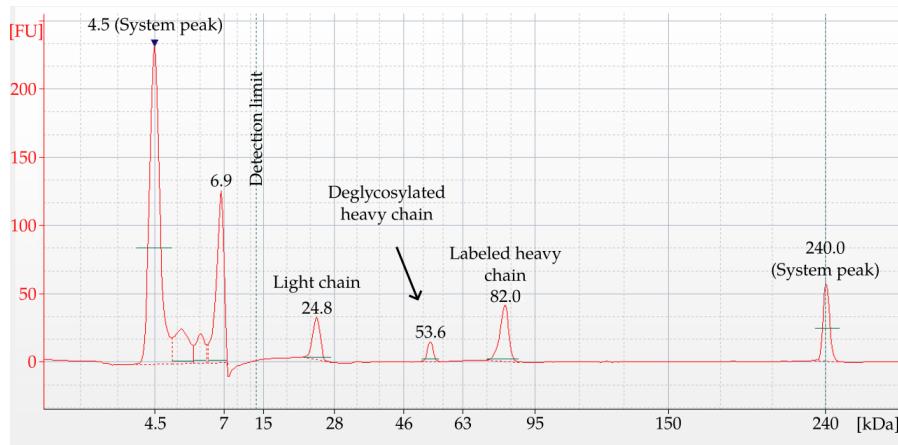


Figure E.2: Representative electropherogram from labeled antibody (clone 1F7, compare to (Figure E.1)). The electrophoresis was run under denaturing and reducing conditions. The attachment of the DNA-label induces a shift in size of the heavy chain (Gong et al., 2016). The unlabeled heavy chain is smaller than the one of the unlabeled antibody (Figure E.1) because of deglycosylation (53.6 kDa vs. 56.8 kDa). Using the values from Table E.1 in Equation 9.1, the LE is 95.6 %.

Table E.1: Area under the curve ([AUC](#)) after peak integration before and after labeling.

Peak sizes and AUCs before labeling (Figure E.1).			
Peak identity	Peak size (kDa)	AUC (ng μL^{-1})	% of total AUC
Light chain	24.6	339.9	33.1
Heavy chain	56.8	675.5	65.7
Peak sizes and AUCs after labeling (Figure E.2).			
Light chain	24.8	68.2	40.3
Deglycosylated heavy chain	53.6	21.2	12.5
Labeled heavy chain	82.0	80.0	47.2

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both `LATEX` and `LyX`:

<https://bitbucket.org/amiede/classicthesis/>

Final Version as of June 24, 2024 (classicthesis v4.6).