UNIVERSITY OF LATVIA

FACULTY OF BIOLOGY

Kārlis Pleiko

Doctoral Thesis

APTAMERS TARGETING CARCINOMA CELL SURFACE RECEPTORS

Promotion to the degree of Doctor of Biology

Supervisor: Prof. Una Riekstina, PhD

Co-advisor: Prof. Tambet Teesalu, PhD

Rīga, 2023

The work of this doctoral thesis was carried out at the Faculty of Medicine, University of Latvia from 2018 to 2022.

The research was supported by the University of Latvia fundamental research grant "Research of biomarkers and natural substances for acute and chronic diseases' diagnostics and personalized treatment" and ESF grant (No. 8.2.2.0/18/I/006).

Form of the thesis: dissertation, subfield – molecular biology.

Supervisor: Prof. Una Riekstina, PhD
Co-advisor: Prof. Tambet Teesalu, PhD

Reviewers:

1) Dr.hab.biol. Nikolajs Sjakste, University of Latvia
2) Dr.biol. Aija Linē, Latvian Biomedical Research and Study Centre
3) Dr.biol. Eva Baldrich Rubio (Vall d'Hebron Hospital Research Institute, Barcelona, Spain

The thesis will be defended at the public session of the Doctoral Committee of Biology, University of Latvia, at 13:00 on January 27th, 2023 at Latvian Biomedical Research and Study Centre, Ratsupites Str. k-1.

The thesis is available at the Library of the University of Latvia, Kalpaka blv. 4.

This thesis is accepted for the commencement of the degree of Doctor of Biology on October 10th, 2022 by the Doctoral Committee of Biology, University of Latvia.

Chairman of the Doctoral Committee: _____/ Dr. biol., prof. Kaspars Tārs /
Secretary of the Doctoral Committee: _____/ Dr. biol. Daina Eze /

ABSTRACT

Chemotherapeutic medicines are not selective and thus produce substantial side effects in patients. By delivering chemotherapeutic medications precisely to the cancerous site without affecting healthy organs, targeted therapies can lessen side effects and enhance the effectiveness of cancer therapy. Antibodies have already proven their value as therapeutic agents. Aptamers are single-stranded DNA or RNA oligonucleotides that bind in a similar way that antibodies do. Aptamers have various benefits over antibodies, but they have yet to deliver on their promise of being employed as targeted therapeutics.

The aim of the study was to identify aptamers binding to carcinoma cell surface proteins using the cell-SELEX approach. We carried out cell-SELEX using clear cell renal cell carcinoma cell line as the target and kidney epithelial cell line as a negative control. After 11 cell-SELEX cycles, we observed that the library has enriched with carcinoma cell line binding aptamer sequences but did not reach a complete selectivity towards them. The differential binding analysis takes advantage of high throughput sequencing and statistical analysis to characterize selectivity towards target or control cells for each aptamer in the enriched library. The lead aptamers identified using differential binding proved to be carcinoma cell line selective. Target protein identification for the lead aptamer GreenB1 took advantage of proximity labelling. The method can be performed *in vitro* on live cells and reduces the background noise compared to extract-based pull-down methods. Proximity labelling followed by mass spectrometry identified β1-integrin, α3-integrin and CD44 as the most likely GreenB1 targets. Electrophoretic mobility shift assay confirmed the GreenB1 binding to α3β1-integrin with affinity in the low nanomolar range. Fluorescence polarization using β1-integrin as a target protein confirmed that GreenB1 has a high affinity towards it.

Further characterization *in vitro* provided evidence that GreenB1 aptamer is internalized into cells and trafficked to the endolysosomal pathway. When injected into mice carrying carcinoma model, GreenB1 proved to home preferentially to tumour lesions without statistically significant accumulation in control organs. Our results indicate the potential application of GreenB1 for the development of aptamer-guided targeted therapeutical agents.

KOPSAVILKUMS

Tā kā ķīmijterapeitiskie medikamenti nav selektīvi, tie pacientiem rada būtiskas blakusparādības. Piegādājot ķīmijterapeitiskos medikamentus precīzi vēža vietā, neietekmējot veselos orgānus, mērķterapija var mazināt blakusparādības un uzlabot vēža terapijas efektivitāti. Antivielas jau ir pierādījušas savu terapeitisko vērtību. Aptamēri ir vienas virknes DNS vai RNS oligonukleotīdi, kas saistās līdzīgi kā antivielas. Aptamēriem ir dažādas priekšrocības salīdzinājumā ar antivielām, bet tie vēl nav pierādījuši sevi kā mērķterapijas līdzekļus.

Pētījuma mērķis bija identificēt aptamērus, kas saistās ar karcinomas šūnu virsmas proteīniem, izmantojot šūnu-SELEX metodi. Mēs veicām šūnu-SELEX, izmantojot gaišo šūnu nieru šūnu karcinomas šūnu līniju kā mērķi un nieru epitēlija šūnu līniju kā negatīvo kontroli. Pēc 11 šūnu-SELEX cikliem novērojām, ka bibliotēka ir bagātināta ar karcinomas šūnu līniju saistošām aptamēru sekvencēm, bet nav sasniegta pilnīga selektivitāte pret tām. Diferenciālās saistīšanas analīzē izmantoja augstas veiktspējas sekvencēšanu un statistisko analīzi, lai raksturotu katra bagātinātās bibliotēkas aptamēra selektivitāti pret mērķa vai kontroles šūnām. Līderaptamēri, kas tika identificēti, izmantojot diferenciālo saistīšanās metodi, izrādījās selektīvi pret karcinomas šūnu līnijām. GreenB1 līderaptamēra mērķa proteīna identificēšanai tika izmantota tuvuma marķēšanas metode. Šo metodi var veikt *in vitro* izmantojot dzīvas šūnas, un tā samazina fona troksni, salīdzinot ar ekstrakcijā balstītām saistīšanās metodēm. Tuvuma marķēšanas, kam sekoja masas spektrometrija, rezultātā kā visticamākos GreenB1 mērķus identificējām β1-integrīnu, α3-integrīnu un CD44. Elektroforētiskā kustīguma nobīdes tests apstiprināja GreenB1 saistīšanos ar α3β1-integrīnu kompleksu ar afinitāti zemā nanomolāru diapazonā. Fluorescences polarizācija, izmantojot β1-integrīnu kā mērķa proteīnu, apstiprināja, ka GreenB1 ir augsta afinitāte pret to.

Turpmāka raksturošana *in vitro* sniedza pierādījumus tam, ka GreenB1 aptamērs tiek internalizēts šūnās un pārvietojas endolizosomālā ceļā. Injicējot to pelēm, kas bija karcinomas modeļa nēsātājas, izrādījās, ka GreenB1 aptamērs selektīvi nonāk audzēja modeļa šūnās, statistiski nozīmīgi neuzkrājoties kontroles orgānos. Mūsu rezultāti norāda uz GreenB1 potenciālo pielietojumu, lai izstrādātu mērķētus terapeitiskos līdzekļus.

TABLE OF CONTENTS

ABBREVIATIONS

.tsv - tab-separated file
3WjdB - fluorescent dimeric Broccoli with three-way junction
3WjtriB - fluorescent trimeric Broccoli with three-way junction
ALT - alanine aminotransferase
ANOVA - analysis of variance
ASO - antisense oligonucleotides
AST - aspartate aminotransferase
Bcl2 - B-cell lymphoma 2 gene
bp - base pairs
BSA - bovine serum albumin
c-Met - tyrosine-protein kinase mesenchymal-epithelial transition protein
ccRCC - clear cell renal cell carcinoma
CE-SELEX - capillary electrophoresis SELEX
CI-M6PR - cation-independent mannose-6-phosphate receptor
CpG - Unmethylated 2'-deoxycitidine-phosphate-2'-guanine
CPM - counts per million
CRISPR - clustered regularly interspaced short palindromic repeats
CTRL - control
DDM - n-dodecyl-β-D-maltoside
DNA - deoxyribonucleic acid
DNA-PAINT - DNA points accumulation in nanoscale topography
Dtxl -docetaxel
EDTA - ethylenediaminetetraacetic acid
EMA - European Medicines Agency
EMSA - electrophoretic mobility shift assay
ESTA - E-selectin
EU - European Union
FAM -6-Carboxyfluorescein
FBS - fetal bovine serum
FDA - Food and Drug Administration
FLAP - fluorescent light-up aptamer

FP - fluorescence polarization

Glut1 - glucose transporter 1

h - hours

HRP - horseradish peroxidase

HRP-S - horseradish peroxidase conjugate to streptavidin

HRP-S-B-GreenB1 - horseradish peroxidase conjugate to streptavidin that is linked to biotinylated GreenB1 aptamer

HRP-S-B-RND - horseradish peroxidase conjugate to streptavidin that is linked to biotinylated randomized oligonucleotide library

hTfR - human transferrin receptor

HTS - high-throughput sequencing

i.v. - intravenously

IQR - interquartile range

ITGA3 - integrin alpha 3

ITGB1 - integrin beta 1

IVT - intravitreal

K_d - dissociation constant

kDa - kilodalton

LNA - locked nucleic acids

log2CPM - logarithmic value of counts per million

logFC - logarithmic value of fold change

LYTAC - lysosome-targeted chimaeras

M6Pn - serine-O-mannose-6-phosphonate

MFE - minimum free energy

MOE - 2'-O-methoxyethylribose

MS - mass spectrometry

MW - molecular weight

NK cells - natural killer cells

nm - nanometres

nM - nanomolar

nmol - nanomole

NOAEL - no-observed-adverse-effect level

NPs - nanoparticles

NS5B - hepatitis C virus non-structural protein 5B

nt - nucleotide

ORR - objective response rate/overall response rate

PAMP - pathogen-associated molecular pattern

PBS - phosphate buffered saline

PCR - polymerase chain reaction

PEG - polyethene glycol

PfLDH - *Plasmodium falciparum* lactate dehydrogenase

PLGA-b-PEG - poly(D,L-lactic-co-glycolic acid)-block-poly(ethylene glycol)

Plk1 - polo-like kinase 1 gene

pM - picomolar

PSMA - prostate-specific membrane antigen

qPCR - quantitative polymerase chain reaction

RNA - ribonucleic acid

RNAi - RNA interference

RND - randomized oligonucleotide library

ROSALIND - RNA Output Sensors Activated by Ligand

RPM - reads per million

RT - room temperature

SAM - S-adenosyl methionine

SE - standard error

SELEX - Systematic Evolution of Ligands by EXponential Enrichment

siRNA - small interfering RNA/silencing RNA

SLC2A1 - solute carrier family 2 member 1

SOMAmer - slow off-rate aptamer

ssDNA - single-stranded deoxyribonucleic acid

TLR4 - toll-like receptor 4

TNBC - triple-negative breast cancer

TPGS - d-α-tocopheryl polyethene glycol 1000 succinate

US - United States

VEGF - Vascular endothelial growth factor

VM - vascular mimicry

INTRODUCTION

Aptamers are short single-stranded DNA or RNA oligonucleotides that bind specifically to their target molecules with high affinity due to their three-dimensional structure. This binding mechanism makes them more similar to antibodies than to other nucleic acid therapeutic agents like siRNA, miRNA, and shRNA that bind their target in a strand-complementary manner. Aptamers are selected from large randomized sequence libraries of oligonucleotides using the Systematic Evolution of Ligands by EXponential Enrichment (SELEX) method. They have been proven to be a valuable tool in the development of biosensors, as imaging tools in super-resolution microscopy, as an alternative to antibodies in later flow assays and as targeting agents *in vitro* and *in vivo* settings. One of the most promising yet underexplored applications of aptamers is as targeted therapeutic agents that can deliver a payload with high selectivity to cancer. The cell-SELEX method allows performing aptamer selection on live cells *in vitro* to enrich the library with aptamers binding carcinoma cell surface proteins. The advantage of cell-SELEX is the ability to select aptamers against proteins in their native state.

The study aimed to identify carcinoma cell binding aptamers with high affinity using cell-SELEX method.

The tasks for this study:

1. Carry out cell-SELEX using carcinoma cell line as target and epithelial cell line isolated from non-cancerous tissue as a negative control
2. Develop a method to identify carcinoma-binding aptamers from the enriched library using statistical analysis of high-throughput sequencing data
3. Identify the target protein for selected lead aptamer using proximity labelling and mass spectrometry proteomics
4. Characterize lead aptamer interaction with carcinoma cell line in *vitro* and test the ability of aptamer to home to tumorous lesions *in vivo* in mice.

1. LITERATURE OVERVIEW

1.1 Aptamers

Aptamers are short single-stranded DNA or RNA oligonucleotides that bind to target molecules with high selectivity and specificity due to three-dimensional structures rather than strand complementarity. Aptamers are selected from randomized oligonucleotide libraries through an iterative selection process termed SELEX (systematic evolution of ligands by exponential enrichment) that was developed in 1990. Craig Tuerk and Larry Gold (Tuerk and Gold, 1990) introduced the term "SELEX" in a *Science* article in 1990 during the selection of RNA aptamers against T4 DNA polymerase from a library with 8 nucleotide randomized region providing 4^8 or 65,536 different RNA molecules in starting library. *Nature* article in the same year by Andrew D. Ellington and Jack W. Szostak (Ellington and Szostak, 1990) describing the selection of RNA aptamers against organic dyes from a library with a theoretical diversity of ~10^{13} sequences coined the term "aptamer" from Latin word *aptus* ("to fit").

During SELEX, a large oligonucleotide library with a theoretical diversity of up to 10^{15} sequences is incubated with a target molecule. After the incubation, unbound sequences are washed away, and bound sequences are amplified using polymerase chain reaction (PCR) for the next selection round. This cyclic process is repeated until the library is enriched with target-specific oligonucleotide sequences. The enriched sequence pool is then sequenced to determine the nucleotide sequence in oligonucleotides with the most favourable binding properties. SELEX methods have evolved to reduce the time and increase the success rate of selections. The cell-SELEX method using live whole cells as targets for selection was first used to identify Tenascin-C binding aptamers in 2001 (Hicke *et al.*, 2001) and 2003 (Daniels *et al.*, 2003). *In vivo* SELEX was first used in 2009 to identify RNA aptamers able to home to hepatic colon cancer metastasis (Mi *et al.*, 2010). High-throughput sequencing (HTS), coupled together with SELEX to create the HT-SELEX method in 2010, has notably reduced the time necessary for aptamer selection and increased the depth at which it is possible to analyse target-specific sequence enrichment during the SELEX process (Jolma *et al.*, 2010).

Aptamer folding is dependent on hydrogen bonding and π-π stacking. Aptamers have a phosphodiester backbone with six rotatable bonds and a rotatable glycosidic bond, which

further explains the ability of randomized aptamer library to form a large number of three-dimensional conformations (Gelinas, 2016). Aptamers can form different internal secondary structures (Fig. 1) – G-quadruplex, stem, bugle, loop, pseudoknot, and kissing hairpin (Eaton, Gold and Zichi, 1995; Zhou and Rossi, 2017). Multiple secondary structures can further change the three-dimensional structure to interact with target molecules through hydrogen bonding, electrostatic and hydrophobic interactions, base stacking, van der Waals forces and shape complementarity (Hermann and Patel, 2000).

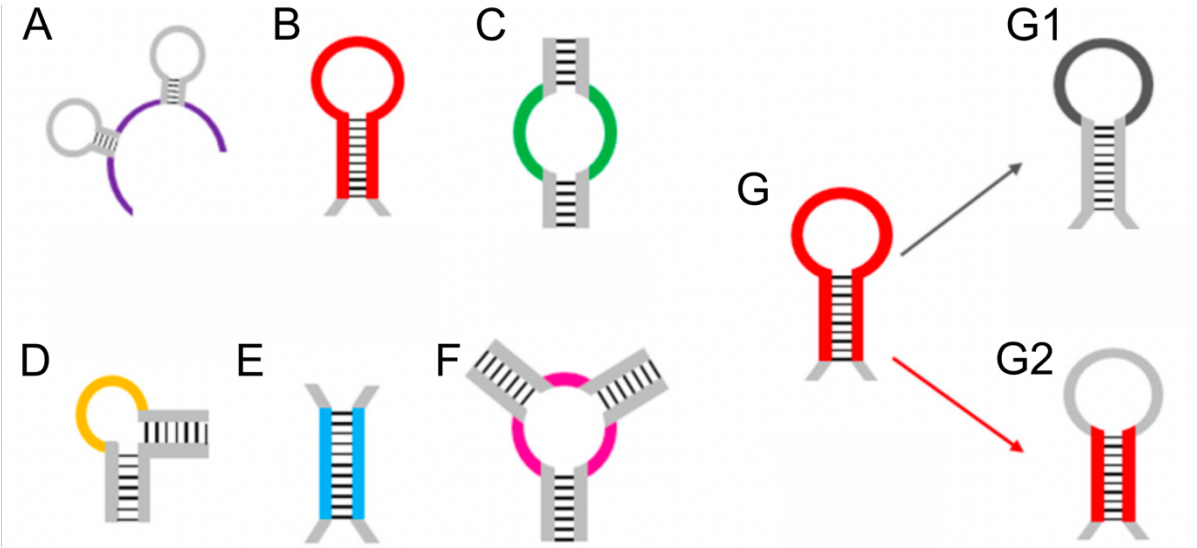

Figure 1. Common functional secondary structures in aptamers. Single-stranded segments can join multiple function structures to each other (A). Hairpin stem/loop (B), internal loop (C), bulge (D), double-stranded duplex (E), multi-branched loop (F). Hairpin stem/loop (G) contains a variable size loop (G1) and variable length stem (G2). Adapted from Sullivan et al. (Sullivan *et al.*, 2019).

1.1.1 Aptamer comparison to antibodies

Aptamers, contrary to other oligonucleotide therapeutics, achieve selective and specific binding to target due to their three-dimensional structure, a mechanism similar to antibody binding to an antigen. Because of this difference from other nucleic acid therapeutics they are sometimes called “chemical antibodies” (Zhou and Rossi, 2017). Aptamers have comparable binding properties to antibodies regarding specificity and affinity. Aptamer size, reliable manufacturing strategies, shelf-life stability, lack of immunogenicity and easily introducible

functional post-selection modifications make them a desirable alternative to antibodies (see Table 1).

Table 1. Aptamer comparison to antibodies. Adapted from Thomas et al. (Thomas, Porciani and Burke, 2022) and Zhou and Rossi (Zhou and Rossi, 2017).

	Aptamers	**Antibodies**
Composition	4 nucleic acids (A, T/U, G, C) if unmodified	20 different amino acids
Molecular weight (MW) and size	MW less than 30 kDa, ~ 3 nm	MW around 150 kDa, ~10-15 nm
Secondary structures	Hairpin, loop, kissing loop, stem, bulge, G-quadruplex	α-helix and β-sheet
Targets	Organic/inorganic molecules, nucleic acids, ions, proteins, peptides, toxins	Immunogenic molecules
Tissue penetration	Faster/higher	Slower/lower
Synthesis	Chemical solid-phase synthesis is completely done in cell-free conditions. Low to no risk of contamination.	Animal-based, from mammalian cells, from transgenic animals or plants.
Batch-to-batch variability	Low or none	Significant during the development stage.
Scalability	Easy	Complicated
Stability (physical and thermal)	Very stable, high shelf-life; DNA aptamers resistant to high temperatures, can be refolded; Can be transported at room temperature if lyophilized.	Susceptible to irreversible denaturation and elevated temperature (room temperature or 37° C. Can be lyophilized for long-term storage (3-5 years).
Cost of production	Low	High due to the use of animal models (except if using phage display)

Shelf life	Months at room temperature, years if frozen.	Months if refrigerated, years if frozen.
Immunogenicity	None or minimal	Less if humanized, otherwise high
Conjugation/post-production modification	Easy and available as needed	Complicated
Specificity	High	High
Affinity	High. Can be modified using multivalent aptamers.	High, depends on identical epitopes on the antigen.
Molecular mechanisms for binding	Three-dimensional interactions (hydrogen bonding, van der Waals forces, electrostatic interactions). Binding can be reversed using complementary aptamers. Surface recognition.	Three-dimensional interactions (Hydrogen bonding, van der Waals forces, electrostatic interactions, salt bridges, aromatic interactions). Binding pocket interactions, epitope recognition.
Internalization via endocytosis	Minimal. Can be modified using post-selection modifications.	Minimal or none
Diagnostic use	Yes	Yes
Half-life	Renal excretion due to the small MW, minutes to hours, is suspectable to nuclease digestion if unmodified.	Weeks, FcRn receptor recycling
Dissociation rate	Fast if monovalent, can be modified by increasing valency	Slow due to bivalency.
Clinical use as therapeutic	Only one FDA-approved drug	Widely used

Aptamer binding to target proteins has been studied using x-ray crystallography, confirming different binding mechanisms than strand-complementarity. Aptamer 2008s was selected to

bind *Plasmodium falciparum* lactate dehydrogenase (PfLDH) with high affinity (K_d = 20 – 50 nM) and specificity. The crystal structure revealed that in its bound state it is arranged in a distorted hairpin structure with an asymmetric internal loop and apical tetraloop (Fig. 2). Aptamer interacts with PfLDH protein mainly through salt bridges. The internal loop forms base interactions, backbone interactions and hydrogen bonds. The apical tetraloop interacts with PfLDH to a lower extent and directly interacts using backbone contacts. (Cheung *et al.*, 2013).

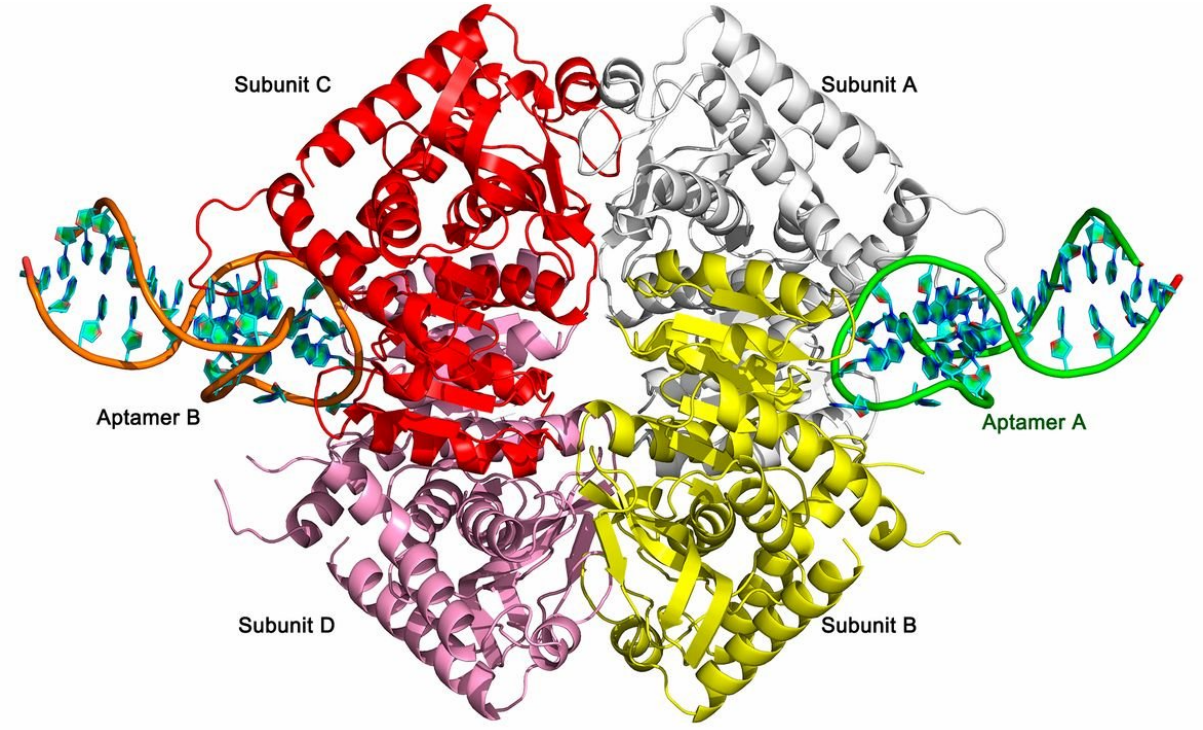

Figure 2. Crystal structure of two 2008s DNA aptamers (Aptamer A and Aptamer B) in complex with tetrameric PfLDH protein (Subunits A, B, C, D). Adapted from Cheung et al. (Cheung *et al.*, 2013)

1.1.2 Challenges in aptamer research for *in vivo* applications/aptamer modifications

Despite having multiple desirable properties compared to antibodies, aptamer research has identified three main drawbacks that need to be overcome to make their use *in vivo* viable – unmodified aptamers are rapidly degraded by nucleases (1), due to the small size aptamers are excreted rapidly (2), toxicity and immunogenicity profiles are understudied and unpredictable (3). There have been multiple strategies developed on how to overcome each of these problems.

Nuclease degradation

Replacing 2'-OH groups with 2'-NH_2 in RNA aptamer pyrimidines (Fig. 3A) has been used to increase the stability of aptamers and to expand the chemical interactions possible for aptamers. Introduction of the 2'-NH_2 group in RNA guarantees resistance to pyrimidine-specific endonucleases. However, 2'-NH_2 modification still leaves RNA vulnerable to purine-specific RNases (Lin *et al.*, 1994). RNA aptamer selection using 2'-F-uridine and 2'-F-cytidine (Fig. 3B), followed by the replacement of 2'-OH-purines with 2'-O-methyl-purines (Fig. 3C) after the selection, has been used to develop a nuclease-resistant version of the aptamer against Vascular Endothelial Growth Factor 165-amino acid long form ($VEGF_{165}$) with picomolar affinity (Ruckman *et al.*, 1998).

Locked Nucleic Acid (LNA) nucleotides containing 2'-O,4'-C-methylene bridge (Fig. 3D) exhibit high nuclease resistance. Importantly, not all polymerases can incorporate modified nucleotides. *KOD Dash* DNA polymerase derived from hyperthermophilic archaeon *pyrococcus kodakaraensis* or *Phusion High Fidelity* DNA can be used to incorporate LNA nucleotides during the PCR amplification. LNA-modified primers and LNA nucleotides were used to select thrombin-binding aptamers using capillary electrophoresis SELEX (CE-SELEX) in the low nanomolar range (Kuwahara and Obika, 2013).

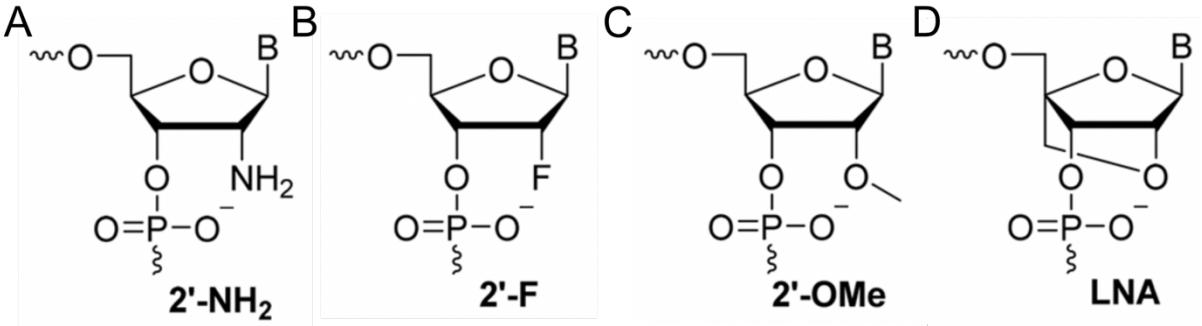

Figure 3. Nucleotide modifications in aptamers to increase nuclease resistance. 2'-OH groups replaced with 2'-NH_2 in RNA aptamer pyrimidines (A), 2'-F-uridine and 2'-F-cytidine (B), 2'-OH-purines replaced with 2'-O-methyl-purines (C), Locked Nucleic Acid (LNA) nucleotides contain 2'-O,4'-C-methylene bridge (D). Adapted from Ni et al. (Ni *et al.*, 2021).

Spiegelmers, mirror-image oligonucleotides that are made from L-nucleotides instead of naturally occurring D-nucleotides, are resistant to nuclease degradation due to chiral-selectivity of nuclease activity (Maasch *et al.*, 2008). Spiegelmer NOX-E36 was identified with protein-SELEX using mirror-image murine monocyte chemoattractant (CLL2) protein composed of D-amino acids as the target. Naturally occurring D-nucleotide library was used for selection to identify mNOX-E36 aptamer with Kd = 157 ± 48 pM (Kulkarni *et al.*, 2007). Importantly, for Spiegelmer selection it is most convenient to have target protein synthesized from D-amino acids since naturally occurring DNA polymerases cannot amplify L-nucleotides used in Spiegelmers. However, the synthesis of mirror-image DNA polymerases (consisting of D-amino acids) that can perform polymerase chain reactions has been reported in recent years (Jiang *et al.*, 2017; Pech *et al.*, 2017) and could make Spiegelmer selection less cumbersome in the future.

Renal filtration

Chemical modifications mentioned in the previous section effectively combat the degradation of aptamers by nucleases. However, folded aptamers are 2-3 nm in diameter (Javier *et al.*, 2008) and are subjected to fast renal clearance, resulting in a very short blood half-life that is not suitable for therapeutic applications due to an inconvenient dosing regimen. To overcome this, aptamers are conjugated to high molecular weight compounds that increase the circulation time *in vivo*.

Conjugation with polyethene glycol (PEG) has shown to be an effective strategy *in vivo* to increase the circulation half-life of aptamers, reduce the elimination rate in urine and distribution to kidneys (Healy *et al.*, 2004). ARC245, a 23 nt long aptamer modified with 2'-O-methyl nucleotides, binds to vascular endothelial growth factor (K_d = 2 nM) and it has a circulation half-life under 1h. When 5'-end was modified with 40 kilodalton (kDa) PEG, the circulation half-life after intravenous injection to CD-1 mice increased to 23 h (Burmeister *et al.*, 2005). PEG conjugation has also been successfully used to increase the circulating half-life of chimeric PSMA-specific aptamer together with cytotoxic siRNA targeting Plk1 and Bcl2 genes from less than 35 min to more than 30 h after intraperitoneal injection in mice (Dassie *et al.*, 2009).

Human coagulation factor IXa binding aptamer 9.3t (apparent Kd = 5.3 ± 1.1 nM) was tested to have a 5-10 min blood half-life in pigs. Conjugation to cholesterol increased the half-life of 9.3t in pigs to 1 – 1.5 h (Rusconi *et al.*, 2004). Nuclease resistant 29 nt long, 2'-F-modified RNA aptamer binding hepatitis C virus non-structural protein 5B (NS5B) was conjugated to cholesterol to increase the circulation time *in vivo*. Aptamer or aptamer-cholesterol conjugate was injected intravenously into male BALB/c mice and samples were taken over 24 h. Conjugation to cholesterol increased the circulating half-life from 5.8 ± 2.1 h to 11 ± 12 h (n=3 in each group; standard deviation indicated after "±") (Lee *et al.*, 2015). The results indicate that different aptamers are affected to a different extent when half-life extending modifications are introduced.

Oligobodies or aptamer-antibody complexes have been shown to increase the blood half-life. VEGF binding aptamer was conjugated to cotinine and further complexed with an anti-cotinine antibody to create an oligobody. Mice (n=5 per group) were injected intraperitoneally with cotinine-aptamer alone or cotinine-aptamer bound to an antibody. The concentration of cotinine was determined from blood samples at different time points. Cotinine-aptamer half-life in blood was determined to be 0.557 h while conjugation to antibody increased the half-life to 14.99 h (Heo *et al.*, 2016).

The pharmacokinetic properties of aptamers can also be enhanced by using conjugation to nanoparticles (NPs). Aptamer functionalized docetaxel loaded poly lactic-co-glycolic acid (PLGA) and d-α-tocopheryl polyethene glycol 1000 succinate (TPGS) copolymer NPs have been shown to have prolonged circulation half-life of 2.246 h after intravenous injection in male Sprague-Dawley rats (Tao *et al.*, 2016). E-selectin (ESTA) binding thioated aptamer was conjugated to PEGylated liposomes and injected into FVB mice (n=3-4) and whole blood was collected to measure the circulating concentration at different time points. When conjugated to liposomes, the determined circulation half-life for aptamer was 24±4 h (Mann *et al.*, 2011).

Immunogenicity and toxicity

Aptamers prepared from unmodified oligonucleotides are well tolerated in preclinical and clinical studies. Contrary to antibodies, aptamers are considered to be non-immunogenic

(Keefe, Pai and Ellington, 2010). However, the modifications that are introduced to increase the stability and circulation half-life, can cause changes in the safety profile.

Pegaptanib (Macugen) is the only aptamer approved for therapeutical use by Food and Drug Administration (FDA) in the US in 2004 and by European Medicines Agency (EMA) in the EU in 2006 for age-related macular degeneration. Pegaptanib contains modified RNA nucleotides (2'-F pyrimidines and 2'-O-Me purines) and it is conjugated to 40 kDa PEG. The lack of newly approved aptamer therapeutics has resulted in limited clinical evidence on the long-term clinical use of aptamers. Preclinical pegaptanib toxicology studies performed for market approval show that intravitreal (IVT) single-dose administration does not cause pegaptanib-related effects in rabbits, monkeys, or rats. Repeated IVT injections (0 mg/mL, 0.1 mg/mL, 0.3 mg/mL and 1mg/mL) in rabbits for 11 weeks resulted in dose-dependent mild cyclitis and presence of vitreal macrophages. No-observed-adverse-effect level (NOAEL) in rabbits was determined to be 1 mg/eye for 6 doses and 2 mg/eye for 2 doses. No treatment-related effects were observed in monkeys and dogs after IVT injections. Intravenous bolus injections (0, 0.1, 1, 10 mg/kg/day) in rats for 13 weeks resulted in chronic progressive nephropathy and increased organ weight connected to vacuolated cells at 10 mg/kg/day. Mild lymphoid depletion was observed in the 1 mg/kg/day group. In rats, NOAEL was determined to be 1 mg/kg. Macrophage infiltrates and vacuolated cells were determined to be related to the clearing of 40 kDa PEG moiety conjugated to pegaptanib. There was no evidence of systemic or local immune response induced by pegaptanib (EMEA, 2006). After the approval, there have been rare case reports of pegaptanib-induced hypersensitivity reactions that are in line with previously observed PEG-induced side effects (Steffensmeier *et al.*, 2007).

The safety of toll-like receptor 4 (TLR4) binding unmodified single-stranded DNA aptamer ApTOLL has been evaluated in a randomized, double-blind phase 1 clinical trial (Clinical-Trial.gov registration number: NCT04742062). ApTOLL was tested at 7 different single-dose levels from 0.7 mg – 70 mg and one multiple-dose group at 21 mg 3 times. The human equivalent dose was from the NOAEL in rats based on body surface area and by applying a safety factor of 10. Half-life in plasma was determined to be 9.3 h. No adverse events and serious adverse events were detected that were attributable to ApTOLL. The total number of treatment-emerged adverse events was 45, the most common being headache, dizziness, blood

creatine phosphokinase increase, nasopharyngitis and diarrhoea. The frequency of adverse events was no more common than in placebo groups (Hernández-Jiménez *et al.*, 2022).

The toxicity of antisense oligonucleotides (ASO) has been studied more extensively than aptamer toxicity. Since both technologies are based on oligonucleotides, modified or unmodified, results from antisense oligonucleotide testing can be also applied to aptamers in some cases. Locked nucleic acids (LNAs) and 2'-O-methoxyethylribose (MOE) containing antisense oligonucleotide *in vivo* studies show target-unrelated hepatoxicity of LNAs. Male Balb/c mice were injected with target-specific or control LNA or MOE ASOs at three different concentrations (0.5 μmol/kg, 1.5 μmol/kg, 4.5 μmol/kg) twice per week for 3 weeks. After 8 days of study, the 4.5 μmol/kg LNA control ASO group showed 46-fold alanine aminotransferase (ALT) and 25-fold aspartate aminotransferase (AST) increase over saline-injected mice. ALT and AST levels were within the normal range in MOE ASOs groups. Non-targeted LNA ASOs showed a 25% liver weight increase over the saline group in a non-dose-dependent manner while MOE ASOs liver weight increase was in the range from 0 – 17% compared to saline control. The hepatotoxic effects of LNAs were also present in rats (Swayze *et al.*, 2007).

Unmethylated 2'-deoxycitidine-phosphate-2'-guanine (CpG) in aptamers can be recognized as a pathogen-associated molecular pattern (PAMP), resulting in innate immune system activation through Toll-like receptors (Krieg, 1999). Immune activation can be tackled by changing phosphate linkage to phosphorothioate, methylating the 5' position on cytosine and introducing 2'-O-methyl or 2'-fluoro modifications (Bruno, 2018). Tumour-targeting aptamers could be able to take advantage of this drawback by not only delivering the payload to a tumorous site but also re-activating the immune response against the tumour (Haßel and Mayer, 2019).

2.1 Overview of aptamer selection strategies

Aptamers are selected using the iterative enrichment method SELEX. Protein-SELEX (Fig. 4) modification where purified proteins-of-interest are used for aptamer selection is the most used variation of SELEX. Over the years, several different SELEX methods, e.g., cell-SELEX (Daniels *et al.*, 2003), *in vivo* SELEX (Mi *et al.*, 2010), capillary electrophoresis or CE-SELEX (Yang and Bowser, 2013), HT-SELEX (Hoinka *et al.*, 2015), microfluidic SELEX

(Dembowski and Bowser, 2017), capture-SELEX (Lyu, Khan and Wang, 2021), structure-switching SELEX (Sanford *et al.*, 2021) and others, have been developed to serve a wide variety of purposes.

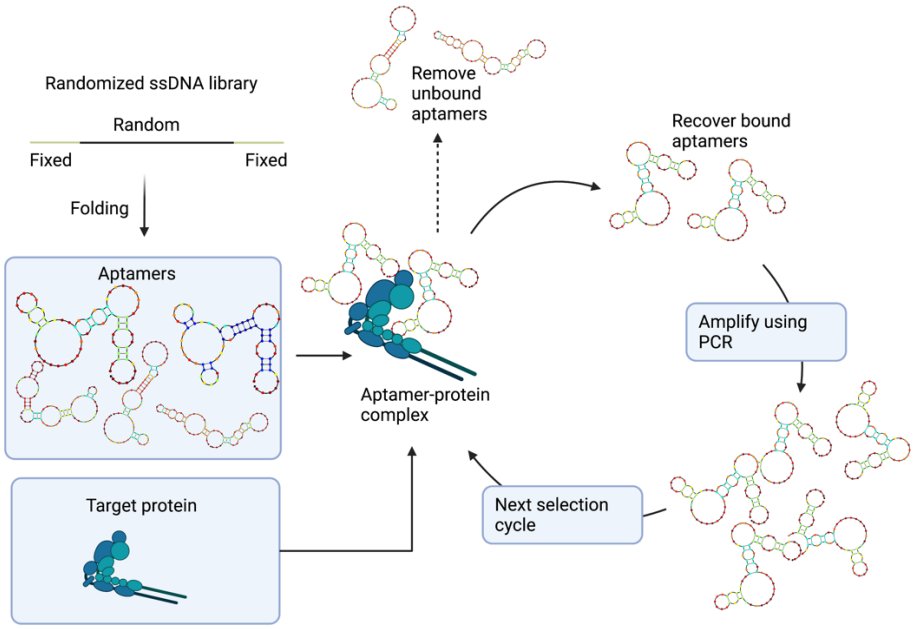

Figure 4. Protein-SELEX method. An initial randomized ssDNA or ssRNA library is folded and incubated with a target protein. After the incubation, unbound sequences are washed away while bound sequences are eluted and amplified for the next selection cycle. The cycle is repeated until the library is enriched with target-binding sequences. Made using BioRender.

2.1.1 cell-SELEX

Cell-SELEX uses *in vitro* grown live cells to identify target cell-specific aptamers (Figure 5). Target cells must represent the pathological state, usually different types of cancers, or cell types that aptamers need to be selected against. Compared to protein-SELEX, it is not necessary to know the exact protein target at the beginning. In cell-SELEX, proteins on the cell surface are in their native state and it increases the likelihood that the aptamers targeting these proteins will be more suitable for applications *in vivo* than aptamers identified using recombinant protein as the target. Negative control cells need to be as closely biologically similar to target cells except for the differences relevant to the pathological state, e.g., primary

cells isolated from cancerous lesion as target cells and adjacent non-cancerous cells as negative control cells (Sefah *et al.*, 2010).

The first cell-SELEX used the U251 glioblastoma cell line as the target cell line and selection was performed without a negative selection step. ssDNA aptamer library with 34 nt randomized region between two primer binding regions (5'-GCCTGTTGTGAGCCTCCT-N34-CGCTTATTCTTGTCTCCC-3') was used for selection. After 21 rounds of cell-SELEX, the enriched library was sequenced to identify aptamer GBI-10 (5'-ggctgttgtgagcctcctCCCAGAGGGAAGACTTTAGGTTCGGTTCACGTCCcgcttattcttactccc-3'). Using affinity purification and mass spectrometry (MS), Tenascin-C was identified as the target protein of GBI-10. K_d at 4° C was determined to be 150 nM but affinity decreased at least 10-fold when binding experiments were done at 22° C or 37° C (Daniels *et al.*, 2003). This can be explained by the fact that selection was done at 4° C to reduce the aptamer degradation, however, it also resulted in the aptamer that changes the conformation when the temperature is elevated which in turn decreases the affinity.

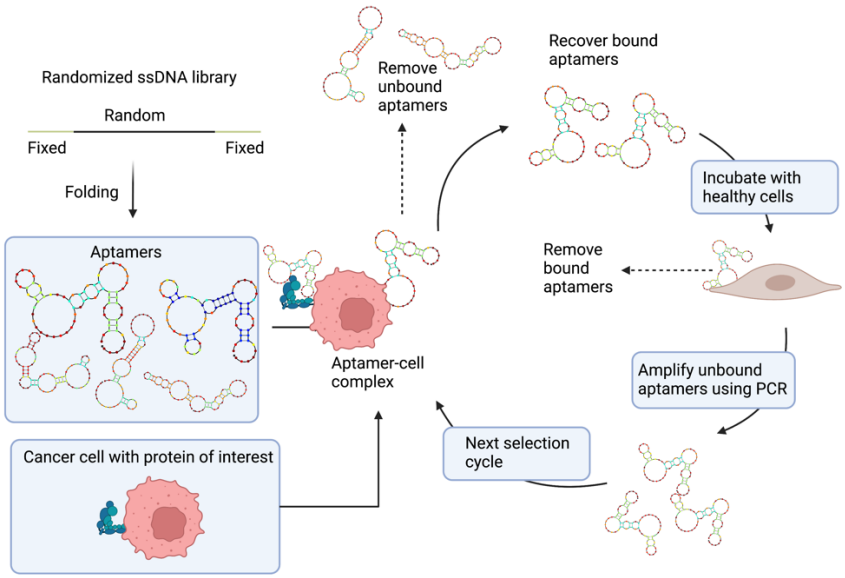

Figure 5. Cell-SELEX. An initial randomized ssDNA or ssRNA library is folded and incubated with a target cells. After the incubation, unbound sequences are washed away while, bound sequences are eluted and incubated with negative control cells. Unbound aptamers are then amplified for the next selection cycle. The cycle is repeated until the library is enriched with target-binding sequences. Made using BioRender.

Toggle cell-SELEX

Toggle-SELEX uses target-of-interest from different species to identify species cross-reactive aptamers (White *et al.*, 2001). Human (hCMEC/D3 cell line) and mouse (bEND3 cell line) brain microvascular endothelial cells were used in toggle cell-SELEX as targets to identify endothelial cell internalizing R11-3 from 2'-fluorinated RNA aptamer library with 40 nt long randomized region. The selection was performed for 13 rounds and only cell-internalizing aptamers were selected by using trypsin-EDTA treatment was used after the incubation with the aptamer library to remove aptamers that do not internalize into cells (Dua *et al.*, 2018). The use of the toggle-SELEX principle allows for increasing the likelihood that the aptamers that show promising results *in vivo* in mice, would achieve a comparable level of targeting when introduced in humans.

Isogenic cell-SELEX

Cell-SELEX method, despite multiple advantages, has a drawback of a high probability of resulting in the enrichment of high-affinity aptamers against disease-unrelated proteins. To combat this, a variation of cell-SELEX, termed isogenic cell-SELEX and developed in 2016 (Takahashi, Sakota and Nakamura, 2016). The initial method used microRNA-mediated silencing but later took advantage of the CRISPR-Cas9 system to knock out the target of interest (Rosch *et al.*, 2020). The resulting cell lines theoretically are different only concerning the target of interest, making negative control selection more effective. Rosch et al. used the CRISPR-Cas9 system to knock out glucose transporter 1 (GLUT1) coding gene SLC2A1 from Caco-2 cells, producing GLUT1-null Caco-2 cells. Further, the GLUT1-null cell line was used for cell-SELEX for negative selection while wild-type Caco-2 cells were used as the target cell line. After ten rounds of cell-SELEX, the authors identified eight lead aptamers. One of the lead aptamers named A5 showed high selectivity towards wild type Caco-2 cell line (K_d = 160 ± 49 nM) compared to GLUT1-null Caco-2 cell line (K_d = 860 ± 53 nM)(Rosch *et al.*, 2020).

2.1.2 *In vivo* SELEX

In vivo SELEX is used to acquire target-binding aptamers using animal disease models (Fig. 6). The resulting aptamers are more suitable for *in vivo* use since off-target vascular binding

has been reduced during the selection. Randomized RNA or DNA library, which is usually modified to be resistant to nuclease-mediated degradation, is injected into a relevant disease-representing animal model. After researcher-defined circulation time, the animal is sacrificed and aptamers homing to part-of-interest are retrieved, amplified using PCR and reinjected for the next selection round (Zhou and Rossi, 2017).

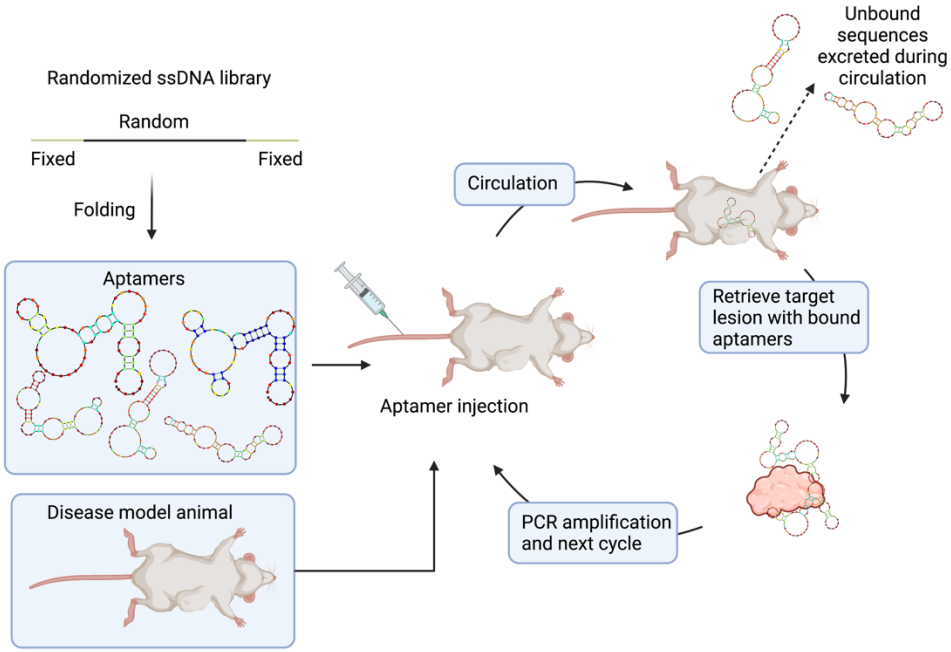

Figure 6. *In vivo* SELEX. An initial randomized ssDNA or ssRNA library is folded and injected to disease model animal. After a certain circulation time, target lesion is retrieved, bound sequences are then amplified for the next selection cycle. The cycle is repeated until the library is enriched with target-binding sequences. Made using BioRender.

The first *in vivo* SELEX study demonstrated aptamer selection from a 2-fluoropyrimidine modified RNA library with 40 nt randomized region (5'-GGGAGGACGATGCGG-N40-CAGACGACTCGCTGAGGATCCGAGA-3') against intrahepatic colorectal cancer metastases. After 14 cycles of selection, sequences of metastasis-binding aptamers were determined using Sanger sequencing and sorted into families containing closely related motifs. Two lead aptamer sequences were then synthesized with Cy3 fluorescent tag and injected into CT26 cell line tumour-bearing mice intravenously. One of these lead aptamers, RNA 14-16 (5'-GGGAGGACGATGCGG-CAGUGCCCAACCGGAACAACAACCACCGGCGGCUCCUGC-CAGACGACTCGCTGAGGATCCGAGA-3'), showed co-localization with CT26

intrahepatic tumours. Affinity purification and MS further identified tumour-upregulated protein p68 RNA helicase (Ddx5) as the target protein for RNA aptamer 14-16 (K_d = 13.8 nM) (Mi *et al.*, 2010).

The usefulness of *in vivo* SELEX has been demonstrated by selecting brain-penetrating aptamers from a library of 2-fluoropyrimidine modified RNA oligonucleotides with 40 nt randomized region. Physiological barriers, such as the brain-blood barrier (BBB), are complicated to model *in vitro*, limiting the use of protein-SELEX or cell-SELEX to identify brain-penetrating aptamers. Starting library was administered through the tail vein and brains were harvested after circulating for 1 to 3 h. Brain-homing oligonucleotides were retrieved, amplified and re-injected for the next selection round. After 22 selection rounds, 90% of the library was represented by 3 motifs. Three identified sequences were then mixed in equal molar ratios with respective scrambled aptamers, injected in mice, and retrieved after the circulation. Aptamer sequence A15 (variable region sequence – 5'-CGTATTGCGCGAGGATTATCCGCTCATCGTTGTTGTTGTG-3') was identified as the most abundant in the brain. A15 ability to reach brain parenchyma was further tested by separating blood vessels from the parenchyma. A15 was found to be most abundant in brain parenchyma fraction (Cheng *et al.*, 2013).

2.1.3 Data analysis tools for aptamer research

The amount of sequence reads has increased from around 100 when using Sanger sequencing to determine best-binding aptamers, to several million sequencing reads acquired with HTS (Jolma *et al.*, 2010; Rosch *et al.*, 2020). This change has not only provided more insight to researchers as to what is happening during the aptamer SELEX but also highlighted the necessity to improve and streamline the data analysis approach. Several different data analysis tools have been developed to address these needs.

FASTAptamer and FASTAptameR 2.0

The initial release of FASTAptamer in 2015 contained several standalone *Perl* scripts that allowed users to analyse HTS data from different types of SELEX and other combinatorial

selection methods that contain FASTA files as an output. FASTAptamer-Count takes the FASTA file from HTS results and counts unique sequences in the file and orders them based on their abundance. The output contains a count of sequencing reads and reads per million (RPM) for each sequence. Output from FASTAptamer-Count can be further used by other tools included in the FASTAptamer kit – FASTAptamer-Compare, FASTAptamer-Cluster, FASTAptamer-Enrich or FASTAptamer-Search. FASTAptamer-Compare can create a tab-separated file (.tsv) from FASTAptamer-Count output to be further compatible with different downstream data analysis tools like *MS Excel* or *RStudio*. FASTAptamer-Enrich compares the abundance of each sequence between two or three different populations. This functionality is used to compare the enrichment of sequences between SELEX cycles, the most common aptamer SELEX analysis approach, or between two parallelly run aptamer selections. FASTAptamer-Cluster creates numerous clusters of closely related aptamers by pooling together sequences with Levenshtein edit distance that is defined by the user (Alam, Chang and Burke, 2015). *FASTAptameR 2.0* (https://fastaptamer2.missouri.edu/) was released in 2022 and is the updated version of the previously developed tool. *FASTAptameR* 2.0 is based on the programming language *R*, available as a web tool and includes additional functionalities. The updated version can translate DNA sequences to peptides/proteins, track and discover sequence motifs over multiple selection rounds, and merge datasets (Kramer *et al.*, 2022).

APTANI and APTANI^2

APTANI is a stand-alone HTS data analysis tool for SELEX, built using Python programming language and its first version was released in 2015. APTANI accepts the FASTQ file as an input and can calculate the frequency of individual aptamers within the library. Further, contrary to other HTS data analysis tools built for SELEX, APTANI predicts the secondary structure of each aptamer that has passed the user-defined frequency filters and extracts the motifs present in these structures (Caroli *et al.*, 2016). APTANI^2, published in 2019, is the second iteration of the tool and includes three additional post-analysis modules. *Evolution Analyzer* pipeline inspects the aptamer secondary structure enrichment over multiple selection rounds. *Motif Fetcher* retrieves motifs of interest from aptamer sequences and *grAPhTANI* visualizes the predicted secondary structures of these motifs (Caroli, Forcato and Bicciato, 2020).

REVERSE

REVERSE (https://www.reverseserver.org) is a relatively recent, released in 2022, web server for HTS data analysis resulting from *in vitro* evolution/selection, including SELEX, experiments. Contrary to other HTS tools developed that can be used for SELEX HTS data analysis, REVERSE includes a pre-processing module that is of particular interest to researchers without any programming knowledge. Raw FASTQ files can be uploaded directly to the server, followed by quality filtering, trimming and production of reverse complement sequence. After the pre-processing, individual sequences or sequence clusters can be analysed in detail. *Analyze Individual Sequences* pipeline can find the most abundant sequences and track the fractional abundance of these sequences, track user-defined motifs over the selection cycles. *Analyze Sequence Clusters* pipeline can identify clusters of closely related sequences. *Conservation Finder* generates a heatmap illustrating the distribution of nucleotides at each position of the identified cluster. *Cluster Peak Tracker* tracks the fractional abundance of each cluster of sequences over multiple rounds of selection (Weiss and DasGupta, 2022).

NUPACK

NUPACK (http://www.nupack.org/) initial release was in 2011 (Zadeh *et al.*, 2011). *NUPACK* is a web-based solution used to predict the secondary structure of RNA or DNA lead aptamers based on minimum free energy (MFE) and visual output of the resulting secondary structure. Secondary structure changes caused by temperature change (melt profile) can also be predicted using *NUPACK*. Additionally to these functions that are offered by other secondary structure prediction tools like *mfold* web server (Zuker, 2003) and ViennaRNA (Lorenz *et al.*, 2011, p. 0), *NUPACK* is also able to accept multiple interacting RNA/DNA strands as input and predict the secondary structure of the resulting complex (Fornace, Porubsky and Pierce, 2020). *Design* module in *NUPACK* accepts the secondary structure using dot-parens-plus or DU+ format as an input and outputs the sequence that would yield this secondary structure – a particularly useful feature if the functional secondary structure of the aptamer is determined but needs to be replaced with more thermodynamically stable for further convenience.

3.1 Overview of aptamer applications

Aptamers can bind their respective target molecules with high specificity and selectivity, guaranteeing a wide range of applications. Over the years, several fields have emerged where aptamers provide a viable alternative to antibodies or are even superior to antibodies. We separately look at aptamer use as diagnostic and therapeutic tools in later chapters, however, aptamers have been used in biosensors (Chiu and Huang, 2009; Liu *et al.*, 2014; Chen and Yang, 2015; Zhao *et al.*, 2015; Bai *et al.*, 2018), to monitor metabolites in live cells (Litke and Jaffrey, 2019), to deliver functional RNAs into live cells (McNamara *et al.*, 2006; Porciani *et al.*, 2018), as imaging agents in super-resolution microscopy (Sunbul *et al.*, 2021) and for other purposes. Several of the most interesting and notable examples of aptamer use are listed further.

Aptamers as biosensors

Biosensors use aptamers as a targeting moiety to detect molecules of interest in complex solutions. Recently, an aptamer-based approach was developed to detect contaminants in water using RNA Output Sensors Activated by Ligand Induction (ROSALIND). ROSALIND uses a mixture of phage RNA polymerase, an allosteric transcription factor that responds to serve as a biosensor and DNA transcription template. When a water sample is added to the freeze-dried ROSALIND mixture, if a target molecule is present, an allosteric transcription factor binds to it, dissociates from the DNA template and initiates transcription of the fluorescent aptamer. Fluorescence can be detected visually using a 3D-printed handheld device. The initial ROSALIND system was able to detect 16 different contaminants, including tetracycline, clarithromycin, benzalkonium chloride, lead (II) and others, in under 1 hour (Jung *et al.*, 2020).

Genetically encoded RNA aptamer-based metabolite biosensors have been effectively used in bacteria (Paige *et al.*, 2012). However, due to the rapid degradation of RNA by nucleases and expression levels in the low nanomolar range (Filonov *et al.*, 2015), the same approach could not be used in mammalian cells. To overcome this, Litke and Jaffrey developed an RNA aptamer expression strategy containing autocatalytic ribozymes. Expressed RNA construct termed pri-racRNA contains modified P3 Twister ribozyme on 5'-end and P1 Twister ribozyme on 3'-end. Self-cleaving ribozymes remove themselves from pri-racRNA, resulting in pre-racRNA with exposed stem sequence and RtcB compatible 5'-OH on the one end and 2',3'-cyclic phosphate on the other end. Endogenous RNA ligase RtcB further ligates both ends,

resulting in circular RNA resistant to intracellular exonuclease digestion. Authors further demonstrate the use of this system for live-imaging by producing folding-induced fluorescent self-circularizing S-adenosyl methionine (SAM) biosensor containing SAM and Broccoli aptamers (Litke and Jaffrey, 2019).

Aptamers for delivery

siRNA delivery to live cells using aptamer-siRNA chimaeras has been demonstrated in 2006 by McNamara II et al (McNamara *et al.*, 2006). However, in 2018 Porciani et al. demonstrated aptamer-guided delivery of large functional RNAs (175 – 250 nt in length) into live cells *in vitro*. Authors demonstrate cell-surface targeting aptamers Waz and C10.36, binding to human transferrin receptor (hTfR) and unidentified receptor on B cell cancer cell lines respectively, linked to 176 nt long fluorescent dimeric Broccoli with three-way junction (3WjdB) or 244 nt long trimeric Broccoli (3WjtriB) through tail-antitail interaction. Both constructs were able to deliver the payload to cancer cells *in vitro* while retaining the functional structure of both targeting aptamer and fluorescent payload aptamer (Porciani *et al.*, 2018). The modular design of the system allows to potentially deliver any functional RNA of similar size to target protein positive cells using cell-targeting aptamers.

Aptamer-based super-resolution imaging

Slow off-rate aptamers (SOMAmers) are small aptamers, modified with protein-like side chains to expand the chemical binding mechanisms of aptamers (Vaught *et al.*, 2010). While conventional affinity ligands used for imaging limit are large from a perspective of super-resolution microscopy (150 kDa for antibodies), SOMAmers have a molecular weight of 6 – 30 kDa. SOMAmer were used to develop DNA points accumulation in nanoscale topography (DNA-PAINT) method that can achieve sub-10-nm imaging in cells with ~3.2 nm average localization precision (Strauss *et al.*, 2018). RhoBAST is a genetically encodable aptamer that can be expressed together with RNA molecules of interest using super-resolution microscopy. RhoBAST can be categorized as a fluorescent light-up aptamer (FLAP) – aptamer that binds fluorophore-quencher conjugate and disrupts the interaction between the fluorophore and the quenches, causing an observable fluorophore signal. FLAPs eliminate photobleaching due to the fast association/dissociation kinetics with the membrane-permeable fluorophore-quencher

conjugate. RhoBAST is 4.6 times brighter than Broccoli fluorescent aptamer and can be used for single-molecule imaging (Sunbul *et al.*, 2021).

3.1.2 Aptamers for cancer therapy

Aptamers have been investigated thoroughly as potential delivery agents for anti-cancer therapeutics in preclinical studies. Bispecific aptamers (Miao *et al.*, 2021; Thomas, Porciani and Burke, 2022), aptamer-guided delivery of therapeutic siRNAs (McNamara *et al.*, 2006; Dassie *et al.*, 2009; Gilboa-Geffen *et al.*, 2015; Kim *et al.*, 2019), aptamer functionalized nanoparticles loaded with chemotherapeutic agents (Farokhzad *et al.*, 2006; Gao *et al.*, 2012; Zeng *et al.*, 2017), immunomodulatory aptamers (Khedri *et al.*, 2015; Thomas, Porciani and Burke, 2022) and other approaches all represent promising future tools for aptamer use as therapeutic agents. Multiple clinical trials have been and are carried out using aptamer functionalized drugs or aptamers together with already established agents as combination therapy.

Immunotherapy and targeted protein degradation are probably two approaches that in the past few years have attracted the attention of researchers the most. Immunotherapy has shown that we can fine-tune the response of the immune system and guide its fight against cancer (Waldman, Fritz and Lenardo, 2020). Targeted protein degradation has gained the interest of pharmaceutical companies in just the last few years and at the end of 2021, four companies were working in this field with a cumulative market capitalization of $ 11 billion (Garber, 2022). Aptamers have also demonstrated that they might have a role to play in these fields.

In vitro preclinical studies with a therapeutic potential

Bispecific aptamers contain two different binding domains that each bind to a different target. Both aptamer domains are present in a single oligonucleotide molecule and together form a construct that can utilize the functional properties of both aptamers. One of the most interesting applications of bispecific aptamers is immunomodulation (Thomas, Porciani and Burke, 2022). Bispecific DNA aptamer construct containing c-Met binding aptamer to target tumour cells and CD16α binding aptamer for specific natural killer (NK) cell recruitment has shown to

induce lysis in Jurkat E6.1 (acute T-cell leukaemia representing cell line) cells in presence of NK cells through antibody-dependent cellular cytotoxicity *in vitro* (Boltz *et al.*, 2011).

Lysosome-targeted chimaeras (LYTACs) consist of one domain that targets the protein of interest on the cell surface or extracellularly and a second domain that binds to cell-surface lysosome-shuttling protein. LYTAC is a targeted protein degradation system first published in 2020 in *Nature* and it can be applied for the therapeutic purpose by acting on proteins that have been considered undruggable. The first LYTAC system was developed by using an antibody as targeting moiety for protein-of-interest and multiple copies of serine-O-mannose-6-phosphonate (M6Pn) as a ligand for the cation-independent mannose-6-phosphate receptor (IGF2R), a protein that is recycled between the cell-surface, endosomes and the Golgi complex and transports proteins with mannose-6-phosphate residues to lysosomes for degradation (Banik *et al.*, 2020). Aptamer chimaeras, containing IGF2R binding aptamer as one degradation inducing domain and aptamers targeting cancer-related proteins mesenchymal-epithelial transition (Met) receptor or protein tyrosine kinase 7 (PTK-7) as protein-of-interest binding domains, have shown promising results that they can be used as LYTACs for targeted protein degradation *in vitro* (Miao *et al.*, 2021).

In vivo therapeutic studies with clinical potential

Therapeutic use of small interfering RNAs (siRNAs), 21- to 23- nucleotides long RNA duplexes that can induce the degradation of complementary mRNA transcripts through RNA interference (RNAi), suffer from tissue specificity and efficacy. Aptamer-siRNA complexes have been developed, containing prostate-specific membrane antigen (PSMA) binding aptamer and siRNA for polo-like kinase 1 (PLK1). The aptamer part selectively delivers the construct to prostate cancer cells *in vitro* and internalizes while siRNA silences PLK1, a survival gene that is overexpressed in tumour cells. This approach has been used *in vivo* in the xenograft mice model of prostate cancer to mediate tumour regression (McNamara *et al.*, 2006).

Aptamer functionalized nanoparticles loaded with chemotherapeutic agents have shown promise to treat different types of cancers *in vivo*. PSMA binding 2'-fluoropyrimidine modified RNA aptamer surface functionalized poly(D,L-lactic-co-glycolic acid)-block-poly(ethylene glycol) (PLGA-b-PEG) nanoparticles (NPs) were loaded with docetaxel (Dtxl), a

chemotherapeutic drug used for prostate cancer. Prostate cancer xenograft-bearing nude mice were treated with aptamer functionalized Dtxl loaded NPs (Dtxl-NP-Apt), non-functionalized Dtxl loaded NPs (Dtxl-NP), Dtxl alone and NPs alone with a single intratumoral injection. All Dtxl-NP-Apt treated mice (n=7 in each group) were alive on day 109, compared to 4 in the Dtxl-NP group, 1 in Dtxl and NP groups and 2 in saline. Complete tumour regression was observed in 5 mice in the Dtxl-NP-Apt group while Dtxl-NP was the only other group where complete regression was observed in 2 mice (Farokhzad *et al.*, 2006).

Clinical studies using aptamer-targeted drug delivery systems

The search in the clinical trials register (https://clinicaltrials.gov/) based on the term "aptamer" returns 34 phase I/II clinical trials that are not currently recruiting only one phase I/II clinical trial currently recruiting participants to study aptamer ApTOLL for the treatment of COVID-19 (ClinicalTrials.gov Identifiers: NCT05293236). Another trial estimates the pharmacokinetic, pharmacodynamic and safety parameters for aptamer AON-D21 (NCT05018403). Completed trials cover broad scope of diseases, age-related macular degeneration being the most prevalent field of study (ClinicalTrials.gov Identifiers: NCT00950638, NCT01089517, NCT00709527, NCT00569140, NCT00021736, NCT02387957, NCT05571267, NCT02214628, NCT02686658, NCT02591914, NCT00215670, NCT00321997). Other clinical trials are aimed to target diseases such as colorectal cancer (NCT03385148), Hippel-Lindau disease (NCT00056199), Von Willebrand Disease (NCT00694785), stroke (NCT04742062), anaemia (NCT02079896), breast cancer (NCT01830244), lymphoma (NCT02780011) and others.

While this approach might give some overview of aptamers under clinical investigation, not all applicants have mentioned the term "aptamer" in their application. DNA aptamer targeting nucleolin, AS1411, has been studied in phase I clinical trial for use against advanced solid tumours (NCT00881244) and in a phase II trial against metastatic renal cell carcinoma (NCT00740441). Factor IXa inhibitory aptamer REG1 has undergone a phase II clinical trial (NCT00715455) for the treatment of coronary artery disease.

Pegylated L-oligoribonucleotide NOX-A12, studied in clinical trials under the brand name olaptesed pegol, is an aptamer developed using Spiegelmer technology that binds and

neutralizes CXCL12 chemokine (stromal cell-derived factor-1). Olaptesed pegol has been studied in phase IIa clinical trial for the treatment of chronic lymphocytic leukaemia in relapsed patients and patients with refractory disease (NCT01486797). In total, 28 patients participated in the trial. The overall response rate (ORR) to combination therapy using olaptesed pegol with bendamustine, a chemotherapeutic drug, and rituximab, an antibody targeting CD20, was 86%. Combination therapy was well tolerated and achieved median progression-free survival of 15.4 months (95% CI 12.2 to 26.2 months) (Steurer *et al.*, 2019). Achieved ORR results are comparable to combination therapy using obinutuzumab, ibrutinib, and venetoclax where 92% ORR (95% CI 62%-100%) was observed in phase 1b clinical trial (Rogers *et al.*, 2018). Olaptesed pegol is also undergoing phase I/II clinical trial in combination with radiotherapy for use in the treatment of partially resected chemotherapy-resistant glioblastoma without MGMT promoter hypermethylation (NCT04121455). Until now, only 10 patients have been enrolled. Olaptesed pegol has shown to be well tolerated and 4 patients have achieved at least partial remission (Giordano *et al.*, 2022).

Despite more and more clinical trials using aptamers being carried out, Macugen is the only aptamer therapeutic agent that has been approved both by FDA and EMA (Curreri *et al.*, 2022). However, with each additional clinical trial being carried out, we also gain more insight into aptamer therapeutics and recalibrate our approach accordingly.

2. MATERIALS AND METHODS

2.2 Oligonucleotides

Biotin or 6-Carboxyfluorescein (FAM) or labelled or unlabelled ssDNA random aptamer library (RND) containing constant primer binding regions and 40 nt randomised region (5'-FAM/Cy5-ATCCAGAGTGACGCAGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGGACACGGTGGCTTAGT-3') and FAM, Cy5 or biotin labelled or unlabelled ssDNA aptamer GreenB1 containing primer binding constant regions and 40 nt sequence in-between (5'-FAM/Cy5-ATCCAGAGTGACGCAGCATGGGGGTAGTGGTGGTTAGGAGTGGAGGCGAGGAGAGCGGTGGACACGGTGGCTTAGT-3′) were purchased from Integrated DNA Technologies (IDT). Oligonucleotides were diluted to 100 μM concentration using DNase and RNase-free water. Aptamers were folded at 10 μM or 1 μM concentration in folding or binding buffer at 95 °C for 5 min, then cooled down to room temperature (RT) for at least 15 min. Binding buffer contained 5 mM $MgCl_2$, 4.5 mg/mL D-glucose, 0.1 mg/mL baker's yeast tRNA and 1 mg/mL bovine serum albumin (BSA) (#A9647, Sigma-Aldrich) in MgCl2 and CaCl2-free Phosphate Buffered Saline (PBS) (#D8537, Sigma-Aldrich, contained K^+ at 4.45 mM, Na^+ at 157 mM concentrations). The folding buffer contained 5 mM MgCl2 in PBS. NUPACK was used to predict the secondary structure of GreenB1 (Zadeh *et al.*, 2011).

2.3 Cell-SELEX

Cell-SELEX protocol was adapted from Sefah et al. (Sefah *et al.*, 2010). The aptamer library was prepared in binding buffer at a 14 μM concentration for the first selection cycle, heated at 95 °C for 5 min, folded on ice for at least 15 min, and added to fully confluent RCC-MF cells in a 100-mm Petri plate (Sarstedt) that were washed 2 times with washing buffer before the addition of the library. The initial library was applied to RCC-MF (Resource Identification Initiative #RRID:CVCL_5884) cells and incubated for 1 hour on ice with RCC-MF (Resource Identification Initiative #RRID:CVCL_5867) cells but not with RC-124 cells in the first selection cycle. After incubation with the oligonucleotide library, the cells were washed with 3 ml of washing buffer for 3 min and collected with a cell scraper after adding 1 ml of DNase-

free water. DNase-free water was used to collect sequences only for the first cycle; in subsequent cycles, the binding buffer was used to retrieve the bound sequences. After collection, the cell suspension was heated at 95 °C for 10 min to remove the bound sequences from the target proteins and centrifuged at 13,000 g; the supernatant containing the selected aptamer sequences was collected.

In subsequent selection cycles, the aptamer library was prepared at a 500 nM concentration and incubated with negative selection cell line RC-124 beforehand. The solution containing unbound sequences was collected and applied to the RCC-MF cell line after washing the cells as described previously. As the selection cycle increased, several modifications were made to the selection procedure: after the 4^{th} selection cycle, 60 mm plates were used instead of 100 mm plates, and an increasing concentration of fetal bovine serum (FBS) (10-20%) was added to the library after folding without changing the final concentration of the aptamer library, the wash volume was increased to 5 ml, the wash time was increased to 5 min and the number of wash times was increased to 3 after incubation.

2.4 Aptamer selection monitoring

In the enriched aptamer pool, a randomized starting library and selected lead aptamers were prepared in binding buffer at 1 µM concentrations, heated to 95 °C for 5 min and then put on ice for at least 15 min. RC-124 and RCC-MF cells were washed with PBS two times and dissociated using a Versene solution (Gibco). Then, 50 µL of the enriched aptamer library, starting library, lead aptamers or binding buffer were added to 50 µL of the cell suspension ($2.5 * 10^5$ cell per sample), followed by the addition of 11 µL of FBS to each sample to a final concentration of 225 nM. The samples were incubated for 35 min on ice. After incubation, the samples were washed two times with 500 µL of binding buffer and resuspended in 500 µL of binding buffer. The samples were passed through a 40 µM cell strainer before flow cytometry analysis. Flow cytometry data were acquired using a Guava EasyCyte 8HT flow cytometer and analysed using the ExpressPro software (Merck Millipore). Flow cytometry data were analysed using FlowJo software, version 10 (FlowJo). 10,000 gated events were acquired for each sample.

Concentration-dependant binding for sequences DB-1, DB-2, DB-3, DB-4 and MB-3 were performed by preparing each sequence in binding buffer at 2 µM, heating at 95 °C for 5 min and folding on ice for at least 15 min. Subsequent manipulations were performed the same way as for a single concentration monitoring except for preparing variable final concentrations (15 nM, 31 nM, 62 nM, 125 nM, 250 nM, 500 nM, 1000 nM) of each sequence in the cell suspension. Flow cytometry data were acquired using Amnis® ImageStream®XMark II (Luminex). Up to 5,000 single-cell gated events were collected for each sample. Data were acquired using the INSPIRE® software and analysed using the IDEAS® software (Luminex).

2.5 Differential binding *in vitro*

Aptamer pools after the 4th and 11th selection cycle were prepared in binding buffer, heated, and folded as described for the cell-SELEX procedure at a 1 ml volume with a final concentration of 500 nM. 500 µL were added to both the RC-124 cells and RCC-MF cells grown on 60 mm plates in appropriate cell culture media up to 95% confluence. The aptamer pools were added to the RC-124 and RCC-MF cells and incubated for 30 min on ice, then the cells were washed two times and collected using a cell scraper, heated immediately at 95 °C for 10 min, and centrifuged for 5 min at 13,000 g. The supernatants containing the bound sequences from both cell lines were frozen at -20 °C. Sequencing was done to compare the differential binding profiles of the enriched oligonucleotide libraries obtained from both cell lines.

The samples for sequencing were prepared by performing two subsequent overlap PCRs as described in the 16S metagenomic sequencing library preparation protocol (Illumina, 2013). The 1st overlap PCR used primers (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-ATCCAGAGTGACGCAGCA-3' and 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-ACTAAGCCACCGTGTCCA-3') that are complementary to constant regions of the randomized oligonucleotide library with an added overhang that includes an Illumina platform-specific sequence. Conditions for the 1st overlap PCR included 12 min of initial activation, followed by 30 sec at 95 °C, 30 sec at 56.3 °C and 3 min at 72 °C. The cycle number was optimized for each sample to reduce the non-specific amplification. Afterwards, PCR products from one sample were pooled together, concentrated using the DNA Clean & Concentrator

(Zymo Research) and run on 3% agarose gel at 110 V for 40 min; the band at 143 bp was cut out and purified using the Zymoclean Gel DNA Recovery kit (Zymo Research).

The second overlap PCR used primers that were partly complementary to the previously added overhang and contained adapters to attach oligonucleotides to the flow cell and i5 and i7 indexes (5'-CAAGCAGAAGACGGCATACGAGAT-[i7 index]-GTCTCGTGGGCTCGG-3' and 5'-AATGATACGGCGACCACCGAGATCTACAC-[i5 index]-TCGTCGGCAGCGTC-3'). Conditions for the second overhang PCR were 12 min at 95 °C, followed by 5 cycles of denaturation at 98 °C for 10 sec, annealing at 63 °C for 30 sec and elongation at 72 °C for 3 min. After PCR products from one sample were pooled together, the mixture was concentrated using DNA Clean & Concentrator (Zymo Research) and run on 3% agarose gel at 110 V for 45 min; the band at 212 bp was cut out and purified using a Zymoclean Gel DNA Recovery kit (Zymo Research). The concentrations for the final products were determined using the NEBNext Library Quant Kit for Illumina (New England BioLabs) by qPCR.

Sequencing was done on the Illumina MiSeq platform using MiSeq 150-cycle Reagent Kit v3 in a single-read mode for 150 cycles. 9% of PhiX was added to the run. Sequencing was done at the Estonian Genome Center, Tartu, Estonia.

2.6 Differential binding sequencing data analysis

Sequencing reads were filtered and demultiplexed. Constant primer binding regions were removed, and sequences that are longer or shorter than 40 nt were discarded using *cutadapt* (Martin, 2011). Counting of recurring sequences was done using *fastaptamer-count*, and matching of the sequences found in replicate samples was done using *fastaptamer-enrich* (Alam, Chang and Burke, 2015).

The differential expression analysis tool *edgeR* (Robinson, McCarthy and Smyth, 2010) was further used for the analysis of sequencing data. Replicate sequencing samples (n=3) from differential binding cell-SELEX experiments after the 4^{th} and 11^{th} selection cycles were combined, and sequences with low abundance (reads per million < 2 and abundant at all in less than 2 sequencing samples) were filtered out. Normalization was performed based on the reads present in each library. Differential binding was estimated using the *edgeR* function to identify

significantly differentially expressed genes using the following parameters: $\log_2$ fold change (log2FC) value > 2, p-value < 0.0001, adjusted for multiple comparisons using the Benjamini & Hochberg (Benjamini and Hochberg, 1995) method.

Enrichment analysis was done separately by using all reads that came from the 4th pool and 11th pool RCC-MF cell binding experiments. We calculated the mean $\log_2$ value of enrichment (mean counts per million (CPM) for a sequence at the 11th cycle divided by the mean CPM for the same sequence at the 4th cycle) for each sequence and kept the sequences that had log2FC > 6 or enrichment between the 4th and 11th cycle.
After these steps, we identified the common sequences in differential binding results and sequence enrichment results to identify the most likely lead aptamer sequences. (*RNotebook* used for 4th cycle differential binding analysis and 11th cycle differential binding analysis, including enrichment analysis, can be found on https://github.com/KarlisPleiko/apta).

2.9 Imaging flow cytometry for monitoring of individual aptamer binding

FAM-RND and FAM-GreenB1 were folded at 1 μM in binding buffer and diluted further (500, 250, 125, 62, 31, 16, 8 nM for MCF-7 and 500, 125, 25, 5 nM for MDA-MB-231). MCF-7 (Resource Identification Initiatiave #RRID:CVCL_0031) and MDA-MB-231 (Resource Identification Initiatiave # RRID:CVCL_0062) cells were cultivated in T75 flask (Sarstedt) until 80% confluence. Cells were washed with PBS and dissociated using non-enzymatic cell-dissociation buffer CellStripper (#25-056-CI, Corning) for 5-9 min, followed by the addition of complete culture medium, centrifugation at 300g for 5 min and removal of the supernatant. Cells were washed twice with binding buffer, split into respective samples, and resuspended with different concentrations of FAM-RND or FAM-GreenB1. Samples were incubated on ice for 1 h, washed twice with washing buffer, resuspended in 40 μL of binding buffer and analysed using Amnis ImageStreamX Mk II imaging flow cytometer and IDEAS software (Luminex).

2.10 Surface β1-integrin availability after GreenB1 binding *in vitro*

FAM-RND or FAM-GreenB1 aptamers were folded and incubated at 0, 50, 100 and 500 nM concentrations with MDA-MB-231 cells grown on a 6-well plate at 37 °C in an incubator for 24 h. Cells were removed from a 6-well plate using a non-enzymatic cell dissociation solution

and incubated on ice with Cy5-GreenB1 aptamer at 100 nM concentration for 1 h. Cells were washed with binding buffer twice, resuspended in 30 μL of binding buffer and analysed using Amnis ImageStreamX Mk II imaging flow cytometer and IDEAS software (Luminex).

2.11 Pulse-chase incubation

MDA-MB-231 cells were cultivated in a 6-well plate until reaching 80% confluence. The Cy5-GreenB1 aptamer was folded in folding buffer at 10 μM concentration and diluted in 1 mL complete growth media to 100 nM before adding to cells. Cells were incubated with Cy5-GreenB1 for 1 h and replaced with complete growth media; afterwards removed for further processing using a non-enzymatic cell dissociation reagent. Cells were analysed 1, 2, 3, 4, and 24 h after adding Cy5-GreenB1. After dissociation, cells were washed twice with PBS/0.1%BSA, resuspended in 100 μL PBS/0.1%BSA and kept on ice. Before imaging flow cytometry, cells were centrifuged at 300 g for 5 min and resuspended in 30 μL of 75 nM LysoTracker green (#L7526, ThermoFisher). Samples were analysed using Amnis ImageStreamX Mk II imaging flow cytometer (Luminex).

2.12 Proximity labelling

Tyramide-AlexaFluor555 working solution was prepared by combining 50 μL of 20X reaction buffer (Component C3 from #B40933),1000 μL purified water, 10 μL of Tyramide-AlexaFluor555 reagent (Component C1 from #B40933) and 10 μL of 0.15% hydrogen peroxide. The tyramide-biotin working solution was prepared by combining 50 μL of 20X reaction buffer (Component C3 from #B40933), 1000 μL purified water, 10 μL of 0.15% hydrogen peroxide and adding Tyramide-Biotin (#LS-3500, Iris Biotech) at 500 μM final concentration.

GreenB1-biotin, unlabelled GreenB1, RND-biotin, and unlabelled RND oligonucleotides were diluted in 500 μL folding buffer to 1 μM concentration and folded as described in the previous sections. Horseradish (HRP)-conjugated streptavidin (Component B, 500 μL) from Alexa Fluor 555 Tyramide SuperBoost Kit (#B40933, ThermoFisher) was added to folded oligonucleotides and incubated at RT for 30 min to create oligonucleotide-biotin-streptavidin-HRP complex. The mixture was transferred to a 100 kDa MWCO Amicon Ultra-4 centrifugal

filter unit (#UFC810008, Merck) centrifuged at 7500g, refilled four times to remove the unbound aptamer and finally concentrated to approximately 100 µL. The resulting complex was diluted to 500 µL and added to cells for performing confocal microscopy or labelled protein pull-down using magnetic streptavidin beads (#65001, ThermoFisher) afterwards.

For confocal microscopy, MDA-MB-231 cells were cultivated in an 8-well culture slide (#354118, Falcon). Cells were washed twice with PBS before applying oligonucleotide-biotin-streptavidin-HRP complexes, followed by incubation at 37 °C for 1 h. The medium was aspirated, and cells were washed 3 times with folding buffer before adding 100 µL of Tyramide-AlexaFluor555 working solution to each well. The reaction was stopped after 2 min by adding 100 µL of 1X stop reagent (100 µL of Component D in DMSO from #B40933 and 1100 µL of PBS) to each well. Cells were washed 3 times with PBS, fixed with 4% formaldehyde at RT for 10 min, and washed twice with PBS. Nuclei were stained with DAPI (#D1306, ThermoFisher) at RT for 5 min and washed once with PBS. Chambers were removed from the slide, and mounting media and coverslip were added to the slide and imaged.

For the pull-down experiment, MDA-MB-231 cells were cultivated in a T75 flask, washed with PBS once and dissociated using non-enzymatic cell-dissociation buffer CellStripper (Corning) for 5-9 min, followed by the addition of a complete culture medium. The cell suspension was split into a necessary number of samples (approximately $1*10^6$ cells per sample), centrifugated for 5 min at 300g and the supernatant was removed. Cells were resuspended in 500 µL of folding buffer and 500 µL of oligonucleotide-biotin-streptavidin-HRP complexes, followed by incubation at RT for 1 h in an end-over-end rotator. Cells were then centrifuged at 300g, washed twice with PBS and resuspended in 100 µL of tyramide-biotin working solution. The reaction was stopped after 2 min by adding 100 µL of 1X stop reagent (100 µL of Component D in DMSO from #B40933 and 1100 µL of PBS) to each sample. Samples were washed with PBS and centrifuged at 300g for 5 min. The samples were either subjected to flow cytometry to confirm biotinylation or lysed for the pull-down of biotinylated proteins. For flow cytometry, samples were incubated with Streptavidin-DyLight-488 (#21832, ThermoFisher) at 20 µg/mL in folding buffer for 10 min, washed 3 times with PBS, fixed with 4% formaldehyde, washed 3 times with PBS, resuspended in 200 µL PBS/0.1%BSA and analysed using flow cytometry (BD Accuri C6 Plus). For protein pull-down, samples were lysed by adding 200 uL of sample

lysis buffer (50 µL of 4X sample buffer, 20 µL of 10% of n-dodecyl-β-D-maltoside (DDM) from NativePage Sample Prep kit, #BN2008, ThermoFisher, and 130 µL of PBS) and pipetting to solubilise the proteins. Lysed samples were centrifuged at >20 000g at 4 °C for 30 min. The supernatant was collected and added to 20 µL of Dynabeads MyOne Streptavidin C1 beads (#65001, ThermoFisher) per sample. The lysate was incubated with magnetic beads at RT on an end-to-end rotator for 30 min. Beads were washed four times with 200 µL of sample lysis buffer. Beads were transferred to a new 1.5 mL centrifuge tube after each wash. Elution was achieved by adding 30 µL of 25 mM biotin in lysis buffer and heating at 95 °C for 5 min. The biotin elution strategy was adapted from Cheah and Yamada (Cheah and Yamada, 2017). Elution was repeated two times, and the supernatant was collected. The third elution step was done by adding 30 µL of reducing sample buffer and heating at 95 °C for 5 min.

Reducing sample-loading buffer (2 µL) was added to 10 µL of each elution from streptavidin beads after proximity labelling. Samples were heated at 95 °C for 5 min and loaded on 12% Mini-PROTEAN TGX precast protein gel (#4561043, Bio-Rad). The gel was run using 1x Tris/Glycine running buffer at 100 V for 90 min. The gel was stained using the SilverQuest Silver staining kit (#LC6070, ThermoFisher). The bands of interest were cut out and sent for mass spectrometry proteomics analysis at the University of Tartu Proteomics core facilities. (https://www.tuit.ut.ee/en/research/proteomics-core-facility). Fig. 22 was prepared using *R*(R Core Team, 2016) in *RStudio*(version 2021.9.2.382) (RStudio Team, 2020) and packages *readxl* (Wickham and Bryan, 2022), *ggplot2* (Wickham, 2016) and *ggrepel* (Slowikowski, 2021).

2.13 Electrophoretic mobility shift assay

GreenB1 or RND aptamers were folded at 1 µM concentration and diluted to 35 nM using a folding buffer with 5% glycerol. α3β1-integrin protein complex (#2840-A3-050, R&D Systems, 20 µL) at 360 nM (100 µg/mL) in PBS was diluted to 180 nM (molar concentrations calculated based on sodium dodecyl-sulfate polyacrylamide gel electrophoresis SDS-PAGE migration of each protein under reducing conditions, 150 kDa for α3-integrin and 125 kDa for β1-integrin) using a folding buffer with 10% glycerol. Further dilutions were prepared using a folding buffer with 5% glycerol. To each 10 µL of α3β1-integrin dilutions, 10 µL of 35 nM of

GreenB1 or RND were added. Final concentration for aptamers was 17.5 nM and α3β1-integrin concentrations were 90, 45, 22.5, 11.25, 5.61, 2.8, 1.4 and 0.7 nM. The mixture was then incubated at RT for 2 h and loaded on 3% agarose gel prepared using 0.5x Tris/Boric acid buffer without ethylenediaminetetraacetic acid (EDTA) and run at 180 V in a cold room (4 °C) using 0.5x Tris/Boric acid as running buffer for 30 min. The gel was stained with SYBR Gold nucleic acid stain (#S11494, ThermoFisher) in 0.5x Tris/Boric acid buffer for 30 min and destained in purified water for 10 min. K_d was calculated from triplicate measurements using Prism 9.3.1 (GraphPad) using one site-specific binding equation (Y = Bmax*X/(Kd + X)).

2.14 Fluorescence polarization

FP reactions (25 μL) were set up in black, non-transparent flat-bottom 384-well microplates (#3821, Corning). Each reaction contained PBS supplemented with 5 mM MgCl2, 0.05% Tween-20, 10 nM of FAM-labelled GreenB1 aptamer or FAM-labelled RND, and varying concentrations of β1-integrin (#10587-H08H1, SinoBiological). Reactions were performed in triplicate. Measurements were recorded on a Hidex Sense microplate reader equipped with 485nm/535nm optical filters, using 100 flashes, medium lamp power and a PMT voltage of 750 V. Titration data were fitted in Prism 9.3.1 (GraphPad) using the following equation:

LR= ((X+Ltot+KD)-SQRT((X+Ltot+KD)^2-4*X*Ltot))/2

Y= BKG + FR*LR

Where:

X - concentration of integrin-β1 (serial 2-fold dilutions)

Ltot - total concentration of aptamer (fixed)

PR - fluorescence ratio, a unitless constant (fitted)

BKG - background polarisation of unbound aptamer (fitted)

KD - dissociation constant (fitted)

Y - fluorescence polarisation (recorded)

2.15 Aptamer homing *in vivo*

Before aptamer homing *in vivo* analysis in the 4T1 cell TNBC orthotopic model, binding of FAM-GreenB1 and FAM-RND to 4T1 cells at 100 nM was confirmed using DB Accuri C6 Plus flow cytometer. Animal experimentation protocols were approved by a Committee of

Animal Experimentation of the Estonian Ministry of Agriculture (Permit #159). FAM-GreenB1 or FAM-RND (2 nmol) were injected i.v. to orthotopic 4T1 tumour-bearing mice (n=3 in each group). The mice were anaesthetised, perfused with 20 ml PBS, and tumour, heart, liver, lung, and kidney were collected after 1 h of circulation. Organs were flash-frozen in liquid nitrogen and kept at -80 °C until further use. Organs were sectioned at 30 μm and fixed in methanol (equilibrated at -20 °C) for 2 min. 5% BSA in PBS with 0.05% Tween-20 was applied to tissue sections for 30 min at RT to block non-specific antibody binding. FAM-GreenB1 or FAM-RND was detected using a 1:200 dilution of rabbit anti-fluorescein antibody (#A-11090, ThermoFisher) at RT for 1 h, followed by 1:400 dilution of secondary anti-rabbit AlexaFluor555 labelled antibody (#A32794, ThermoFisher) at RT for 1 h. Staining with secondary antibody alone was used as a negative control. Five random fields of view were imaged from all organs of all mice. Both primary and secondary antibody labelled, and secondary antibody alone were imaged by fluorescent microscopy. For statistical analysis, mean pixel intensity from primary and secondary antibody labelled organ sections were divided by mean pixel intensity from the corresponding organ from the same animal when stained with secondary antibody alone. Nuclei were stained using Hoechst 33342 (#62249, ThermoFisher) solution at 20 μM concentration. Fluorescence microscopy was done using Nikon C2 microscope (Japan) and images were analysed using Nis-Elements C 4.13 software. Fluorescence confocal microscopy was done using Olympus FV1200MPE, Germany. Whole tumour imaging was performed using slide scanner Aperio VERSA Brightfield, Fluorescence and FISH Digital Pathology Scanner, USA. Statistical analysis was done in Prism 9.3.1 (GraphPad) using 2-way ANOVA followed by Šidák's multiple comparison test. Figure 28B was prepared using *R* (R Core Team, 2016) in *RStudio* (RStudio Team, 2020) and packages *readxl* (Wickham and Bryan, 2022), *ggplot2* (Wickham, 2016), *ggsignif* (Constantin and Patil, 2021).

3. RESULTS

3.1. Aptamer cell-SELEX selection shows library enrichment

To identify carcinoma-binding aptamers, we used the cell-SELEX approach. ccRCC representing cell line RCC-MF that had been isolated from ccRCC (pT2, N1, Mx/ GII-III) metastatic lesion in the lung was used as the target cell line. Non-tumorigenic epithelial cell line RC-124 was used as a negative control cell line to remove aptamers that might bind to healthy cells (Fig 7.)

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cell carcinoma (RCC). While low-stage ccRCC can be successfully treated using surgical resection and ablation therapy, metastatic ccRCC is biologically distinct from primary RCC and almost uniformly lethal (Jonasch, Walker and Rathmell, 2021).

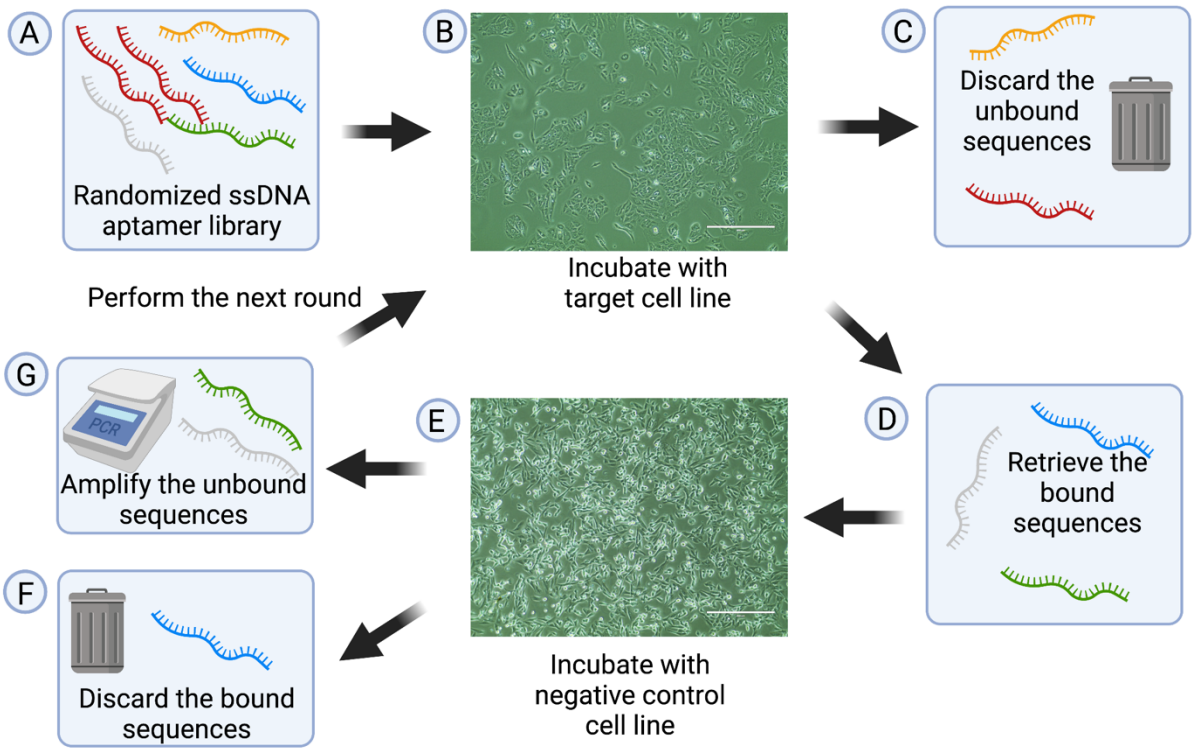

Figure 7. Schematic representation of cell-SELEX. Randomized ssDNA oligonucleotide library (A) is incubated with clear cell renal cell carcinoma cell line RCC-MF *in vitro* (B). After the incubation unbound sequences are discarded by washing (C). Cell-bound oligonucleotide sequences are retrieved (D) and further incubated *in vitro* with non-cancerous kidney epithelial cell line RC-124 (E). After the incubation, oligonucleotide sequences binding

to non-cancerous cells are discarded (F). Unbound sequences are collected, amplified using PCR and applied to the next round of cell-SELEX (G).

ssDNA aptamer selection progress was evaluated using flow cytometry. FAM conjugated oligonucleotide starting library with constant primer binding regions and 40 nt randomized region between them showed minimal (less than 3%) binding to both RC-124 and RCC-MF cell lines (Fig. 8 A-D). Selective binding to the RCC-MF cell line increased up to the 8th selection round when the enriched library was binding to 3.6% of negative control RC-124 cells (Fig. 8E) and around 32% of target RCC-MF cells (Fig. 8F). After we continued the selection up to 11th round, we noticed that the selectivity was no longer observable and enriched oligonucleotide library was binding to around 95% of RC-124 cells (Fig. 8G) and 92.6% of RCC-MF cells (Fig. 8H). The loss of selectivity can be an indication of contamination with non-selectively binding sequences that are readily amplified and have overtaken the selectively binding aptamer sequences.

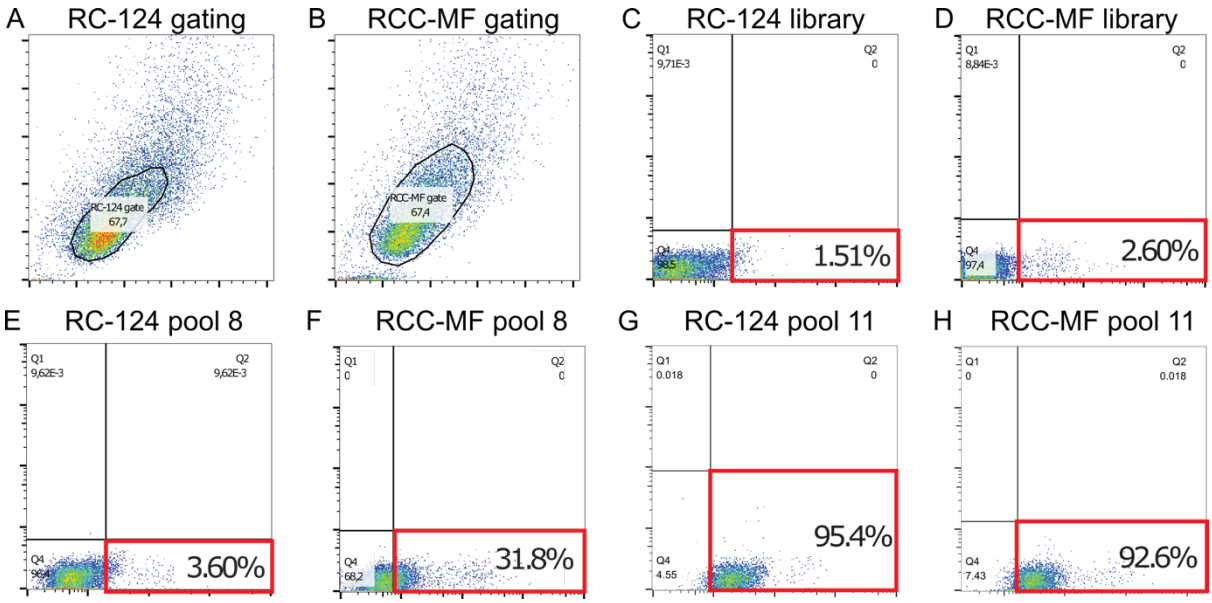

Figure 8. Comparison of aptamer library and selected pools after 8th and 11th cycle binding to kidney epithelial tissue cell line RC-124 and clear cell renal carcinoma cell line RCC-MF. Gating was used for the analysis of RC-124 cells (A) and RCC-MF cells (B). Dot plot of FAM labelled unselected DNA aptamer library binding to RC-124 cells (C) and RCC-MF cells (D). Dot plot of selected aptamer pool after 8th cycle binding to RC-124 cells (E) and RCC-MF cells (F). Dot plot of selected aptamers aptamer pool after 11th cycle binding to RC-124 cells (G) and RCC-MF cells (H).

To test the binding properties of the enriched library, we incubated the 11th round enriched library with RC-124 cells and RCC-MF cells at different concentrations (0-225 nM) and analysed the binding using flow cytometry (Fig. 9).

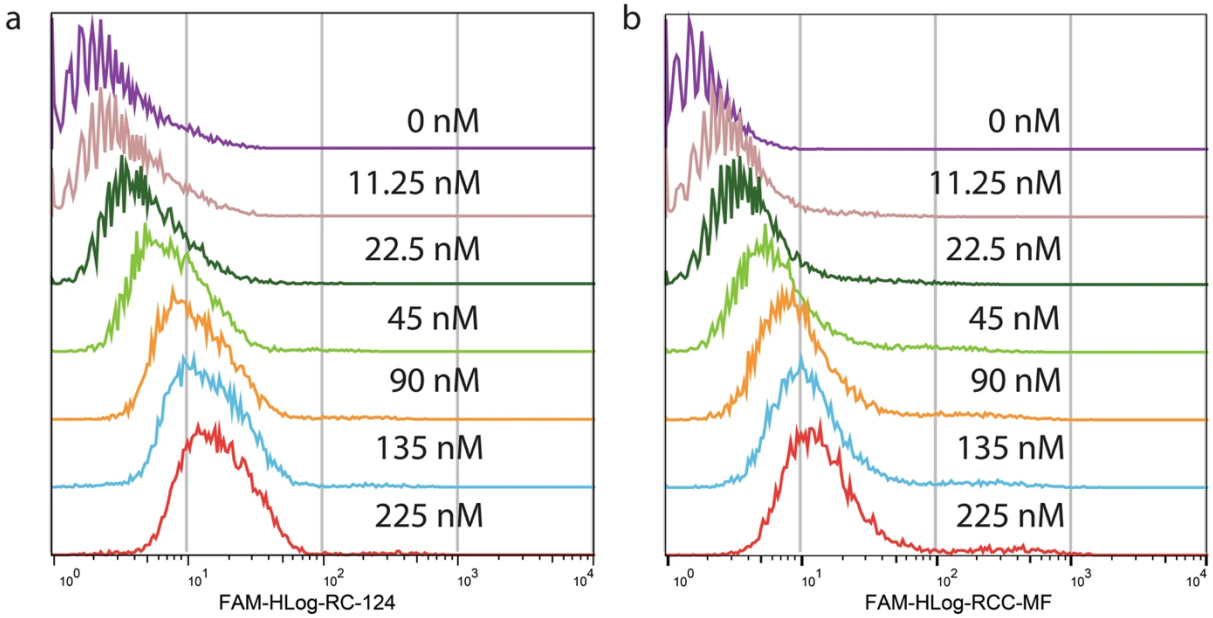

Figure 9. Aptamer binding calculations after the 11th selection cycle. Aptamer binding measurements by flow cytometry using the 11th pool enriched the library at different concentrations on control cells RC-124 (**a**) and clear cell carcinoma cells RCC-MF (**b**).

Based on geometric mean fluorescence intensity at each concentration, we calculated that the affinity of the 11th round library is comparable to both RC-124 cells (apparent Kd = 189 nM) and RCC-MF cells (apparent K_d = 169 nM) (Fig. 10).

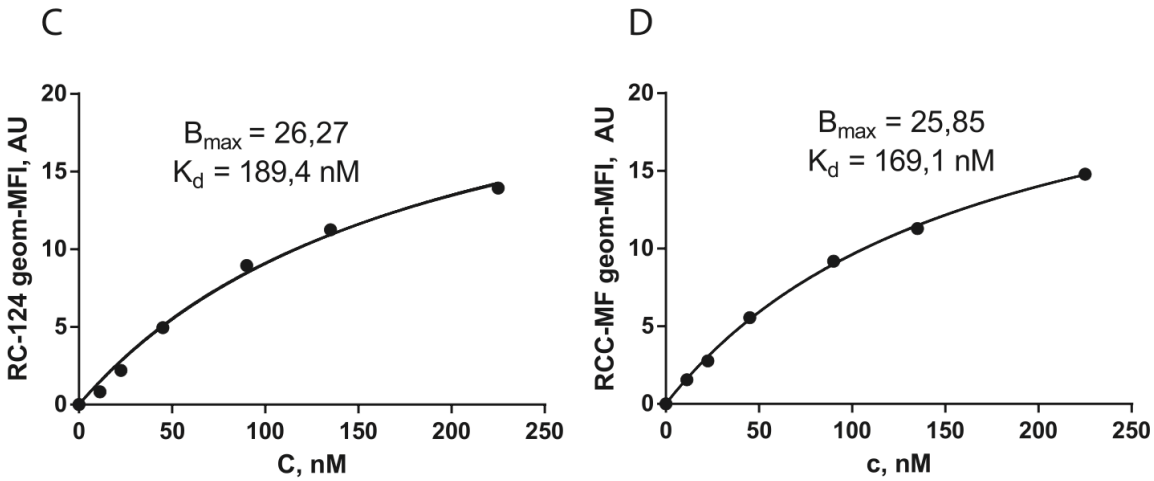

Figure 10. Apparent K_d values of the enriched library to RC-124 and RCC-MF cell line after the 11th cell-SELEX cycle.

3.2. Differential binding cell-SELEX identifies multiple lead aptamers

The presence of possible lead aptamer sequences was interrogated using differential binding cell-SELEX, a method adapted from RNA-sequencing that was developed in our laboratory during this study (Fig. 11).

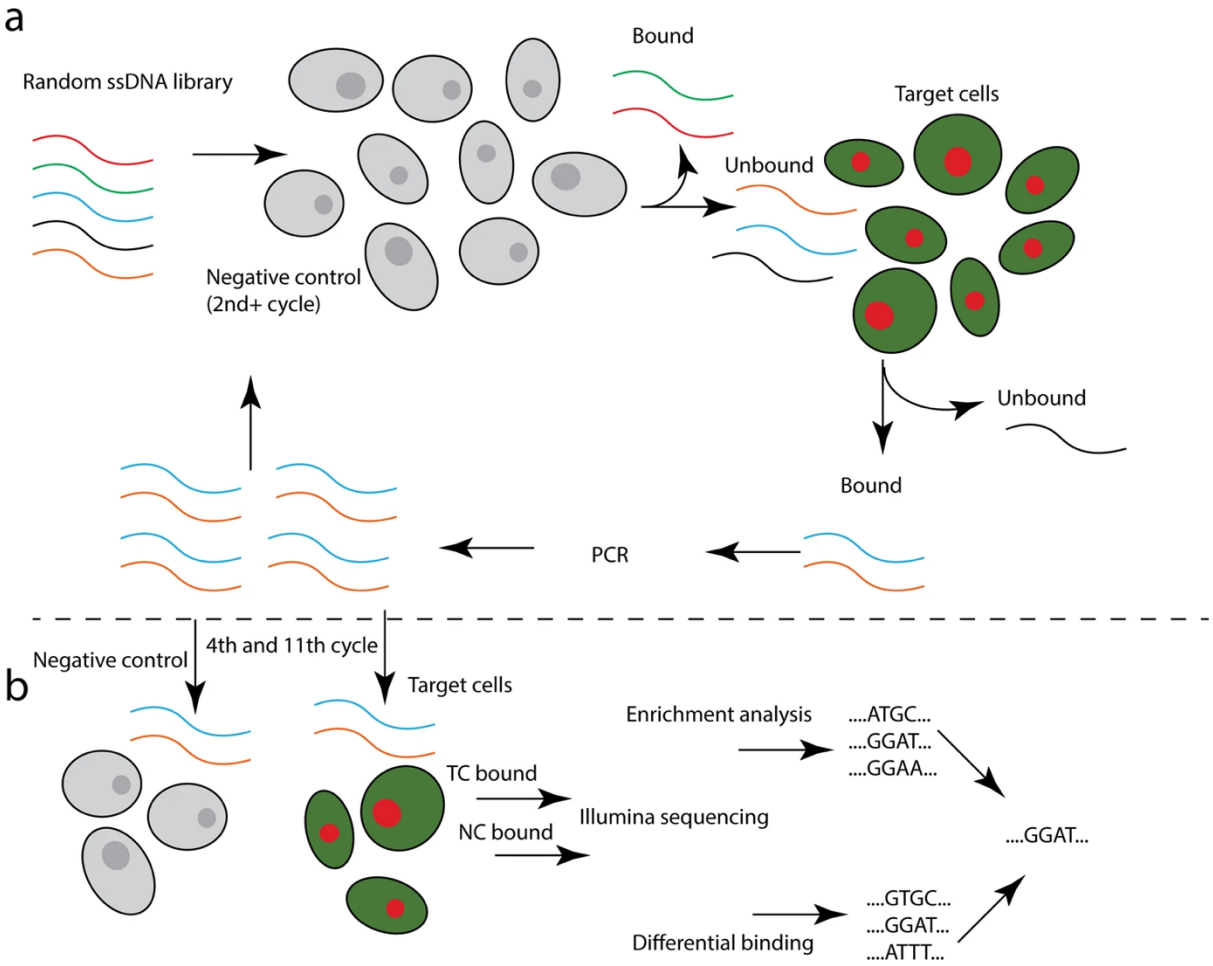

Figure 11. The differential binding cell-SELEX approach combines the regular workflow for cell-SELEX (a) with additional differential binding steps and statistical data analysis to estimate the relative binding differences of each aptamer in an enriched library against target cells and negative control cells (b).

Differential binding analysis was performed after the 4th and 11th selection rounds. Enriched library pools were incubated *in vitro* with either RC-124 cells (n=3) or RCC-MF cells (n=3). After the incubation aptamer sequences were retrieved and custom-made barcodes and Illumina HTS-compatible adapters were attached to constant primer binding regions using overhang PCR. The resulting fragments were run on an agarose gel to confirm the correct size (Fig. 12).

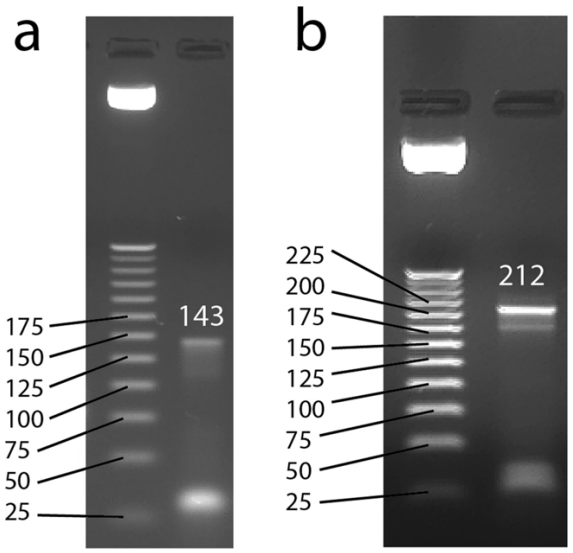

Figure 12. Gel images of differential binding samples after adding Illumina-specific sequencing indexes and adapters. Aptamers after performing 1st overhang PCR have a product length of 143 bp (a) and after 2nd overhang a length of 212 bp (b).

Samples were purified from the gel, quantified using quantitative PCR and sequenced with Illumina MiSeq. The number of reads acquired per sample is between 160 000 and 1 130 000 (Table 2).

Table 2. Differential binding samples for sequencing. Differential binding analysis was performed after the 4th and 11th selection cycles by incubation-enriched libraries with either target RCC-MF cell line (n=3 for each tested cycle) or kidney epithelial RC-124 cell line (n=3 for each tested cycle). Concentration was determined with qPCR, sequencing reads acquired using Illumina MiSeq.

Sample No	**Cell line**	**Selection round**	**Concentration (nM)**	**Sequencing reads**
1	RCC-MF	4	142.6	520′534
2	RCC-MF	4	90.95	781′654
3	RCC-MF	4	185.8	1′130′509
4	RC-124	4	76.23	169′024
5	RC-124	4	67.06	619′635
6	RC-124	4	82.44	548′781
7	RCC-MF	11	15.22	277′070
8	RCC-MF	11	11.99	498′937

9	RCC-MF	11	5.23	326′402
10	RC-124	11	15.58	742′255
11	RC-124	11	28.48	1′142′819
12	RC-124	11	12.88	730′489

Differential binding analysis was done along with enrichment analysis to identify potential lead aptamers (Fig. 13).

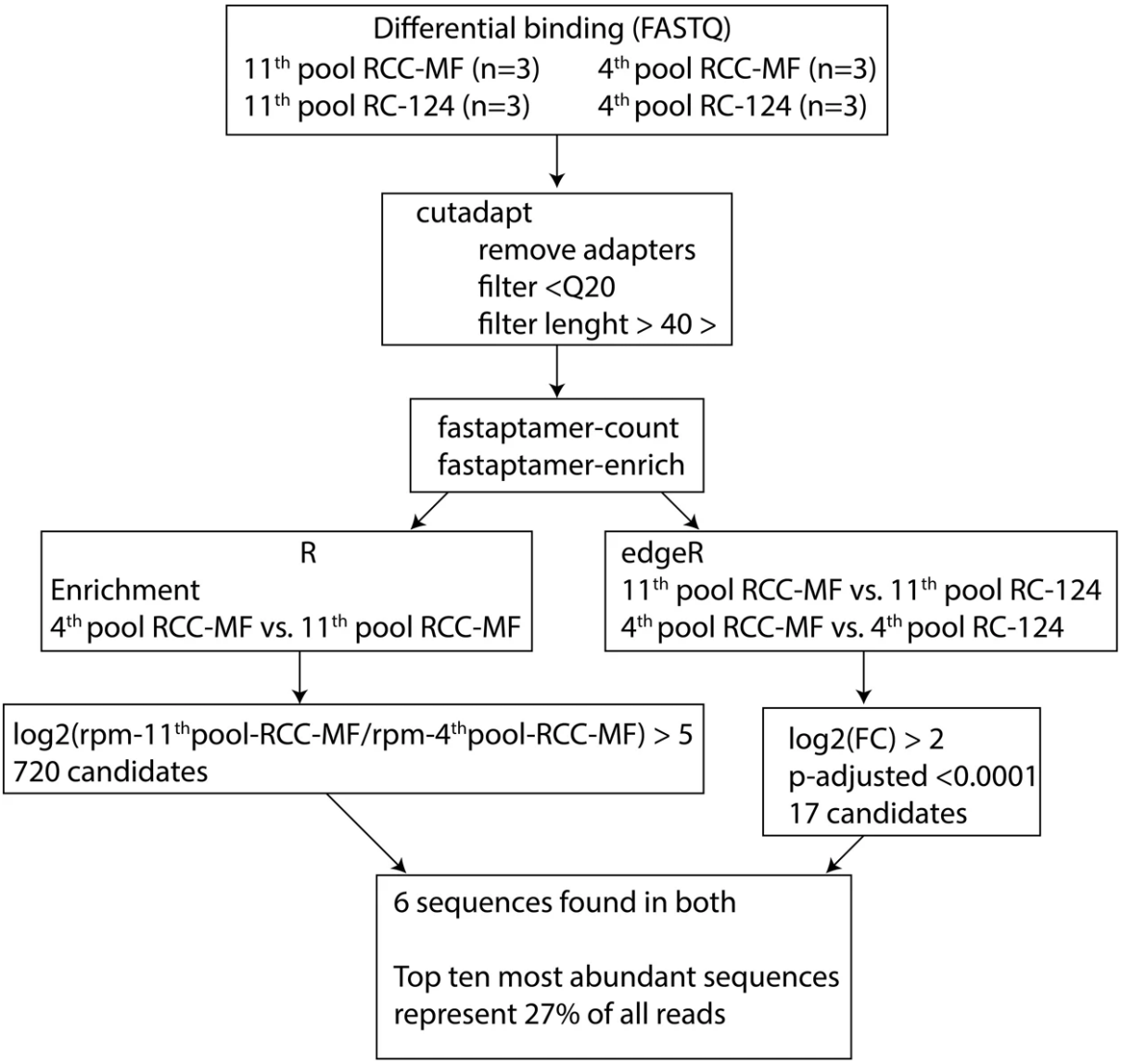

Figure 13. Data analysis workflow for differential binding cell-SELEX. Sequencing reads were trimmed using *cutadapt*, followed by *FASTAptamer* tools *fastaptamer-count* to determine the number of unique sequences, reads per sequence and relative sequence abundance.

fastaptamer-enrich was used to carry out enrichment analysis and *edgeR* was used to calculate the statistical significance of differential binding to each cell line. Enrichment analysis identified 720 sequences with enrichment that had $\log_2 > 5$ enrichment between the 4th and 11th cycles. The differential binding analysis resulted in 17 potential lead aptamers. Six lead aptamers were present in both enrichment analysis and differential binding analysis. The top 10 sequences represent 27% of all sequencing reads after the 11th selection cycle.

Enrichment analysis identified 720 lead aptamers between the 4th round and the 11th round that had log2 enrichment values higher than 5. Differential binding analysis after 11th round yielded 195 aptamers that were statistically differentially binding more to either RC-124 cells or RCC-MF ($p < 0.0001$, logFC > 2 or logFC < -2). From these 195 lead aptamers, 178 aptamers were binding to a higher extent to RC-124 cells than RCC-MF cells (log2CPM < -2), and 17 aptamers were biased towards binding more to RCC-MF cells (log2CPM > 2). Differential binding after 4th round did not identify any lead aptamers binding statistically significantly more ($p<0.0001$, logFC < -2 or logFC > 2) to RC-124 or RCC-MF cell lines (Fig. 14).

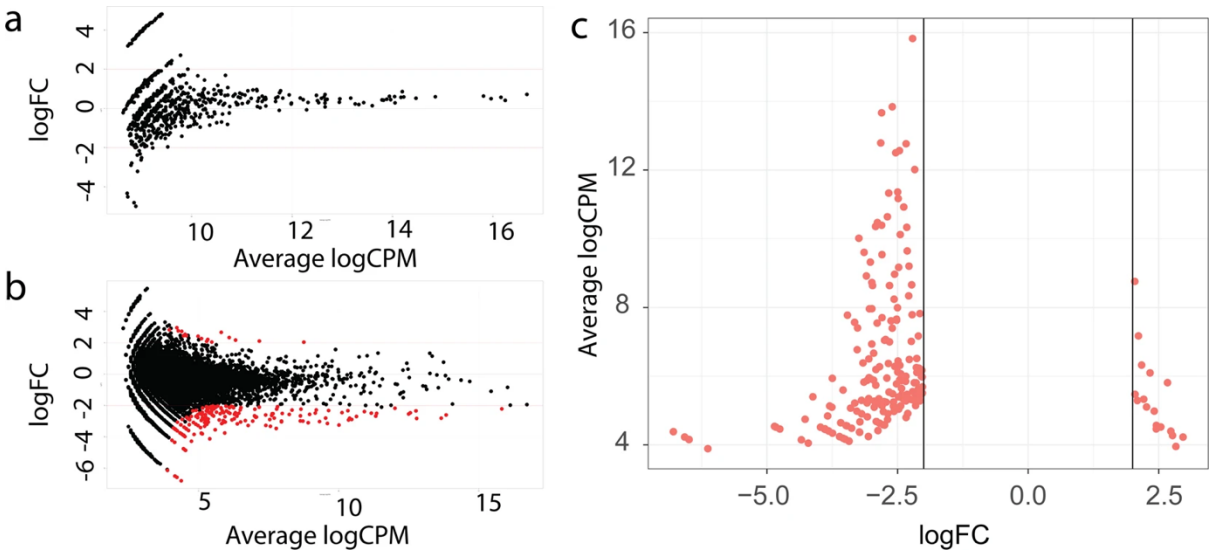

Figure 14. Results from differential binding analysis after 4th (a) and 11th (b) rounds of cell-SELEX. A positive logFC value indicates that the aptamer is more abundant on the RCC-MF cell surface while a negative logFC value indicates higher abundance on RC-124 control cells. Red dots indicate statistically significantly more abundant results on each cell line respectively (multiple comparison adjusted p-value < 0.0001 and logFC > 2 or logFC < -2). Statistically significant results after the 11th cycle (c).

Based on the combined analysis we chose 11 different lead aptamers for further testing (Table 3).

Table 3. Selected lead aptamers for further individual testing. DB-1, DB-2, DB-3, DB-4 and DB-5 were identified using a differential binding approach. EN-1, EN-2 and EN-3 were the most enriched sequences from the 4th cycle to the 11th cycle. MB-1, MB-2 and MB-3 were identified as the most abundant sequences.

Name	**Sequence (the variable region in bold)**
DB-1	5′-ATCCAGAGTGACGCAGCA-**TGCTAGGGTAGGTTGGGCCGGGGGTGGGTGGGTGTGTGAT**-TGGACACGGTGGCTTAGT-3′
DB-2	5′-ATCCAGAGTGACGCAGCA-**AAGAGAAGTATGGGCAGGTTGGGCCGGGGGTGGGTGGGAG**-TGGACACGGTGGCTTAGT-3′
DB-3	(5′-ATCCAGAGTGACGCAGCA-**AGTTGCAGGGTGGGGGTTGGGTGAAGAGCGATGGAGGGGG**-TGGACACGGTGGCTTAGT-3′)
DB-4	(5′-ATCCAGAGTGACGCAGCA-**AGGGGGGGGGGTGGTTTTAGTTGCGTATGGTGGTGGGTGC**-TGGACACGGTGGCTTAGT-3′)
DB-5	(5′-ATCCAGAGTGACGCAGCA-**GGCGGTAGTGGAAAGGGTGGTGTGGGTTGGGACAGGAAAG**-TGGACACGGTGGCTTAGT-3′
EN-1	(5′-ATCCAGAGTGACGCAGCA-**TGACGGGTGGGTGTGGTGGGAAGGGAATTTAGATGCGTGG**-TGGACACGGTGGCTTAGT-3′
EN-2	(5′-ATCCAGAGTGACGCAGCA-**GGGTGGGTTGGTTGGTGTGGTTTGGAGGGTGGGTGAGATG**-TGGACACGGTGGCTTAGT-3′)
EN-3	(5′-ATCCAGAGTGACGCAGCA-**GGTGGCAGTTTTGGGGTTAGGGGTTGATGGGGTTTGGAGG**-TGGACACGGTGGCTTAGT-3′

MB-1	(5′-ATCCAGAGTGACGCAGCA-**GAGTTTGGGGTAAGGGGTTGGGGAGGGACTGTGCGGTTCT**-TGGACACGGTGGCTTAGT-3′)
MB-2	(5′-ATCCAGAGTGACGCAGCA-**GGGAATGTTGGAGGGTGGAGGGGGAGGAGGGTGGTGAGAG**-TGGACACGGTGGCTTAGT-3′)
MB-3	(5′-ATCCAGAGTGACGCAGCA-**TGGGGGTAGTGGTGGTTAGGAGTGGAGGCGAGGAGAGCGG**-TGGACACGGTGGCTTAGT-3′)

From the differential binding analysis, we chose 5 aptamers with the highest logFC value, meaning that they are binding to RCC-MF to a higher extent than to RC-124 (DB-1 to DB-5). From enrichment analysis, we identified 3 aptamers with the highest enrichment value between the 4th round and 11th rounds (EN-1 to EN-3). We also chose 3 aptamers that were the most abundant after the 11th round of cell-SELEX (MB-1 to MB-3).

3.4. Functional testing confirms lead aptamer selectivity

Functional testing was done using flow cytometry. RC-124 or RCC-MF cell lines were incubated with 225 nM concentration of each aptamer (n=3 for each aptamer and each cell line) that had fluorescent FAM label at 5'-end for 1 h and analysed with a flow cytometer. Baseline correction was done by subtracting the fluorescence intensity of the respective cell line after the incubation with a randomized starting library. Aptamers identified using the differential binding approach, DB-1, DB-2, DB-3 and DB-5, show preferential but not statistically significant binding to RCC-MF cells, DB-4 preference for RCC-MF cells is statistically significant (p=0.019). Two aptamers identified using enrichment analysis, EN-1 and EN-3, also show statistically non-significant preference towards the RCC-MF cell line while one lead aptamer, EN-2, binds more to the RCC-MF cell line and the difference is statistically significant. The two most abundant aptamers in the enriched library after the 11th round, MB-1 and MB-2, bind to RC-124 and RCC-MF to a similar extent while aptamer MB-3 preferentially binds to the RC-124 cell line and had the highest fluorescence intensity overall (Fig. 15).

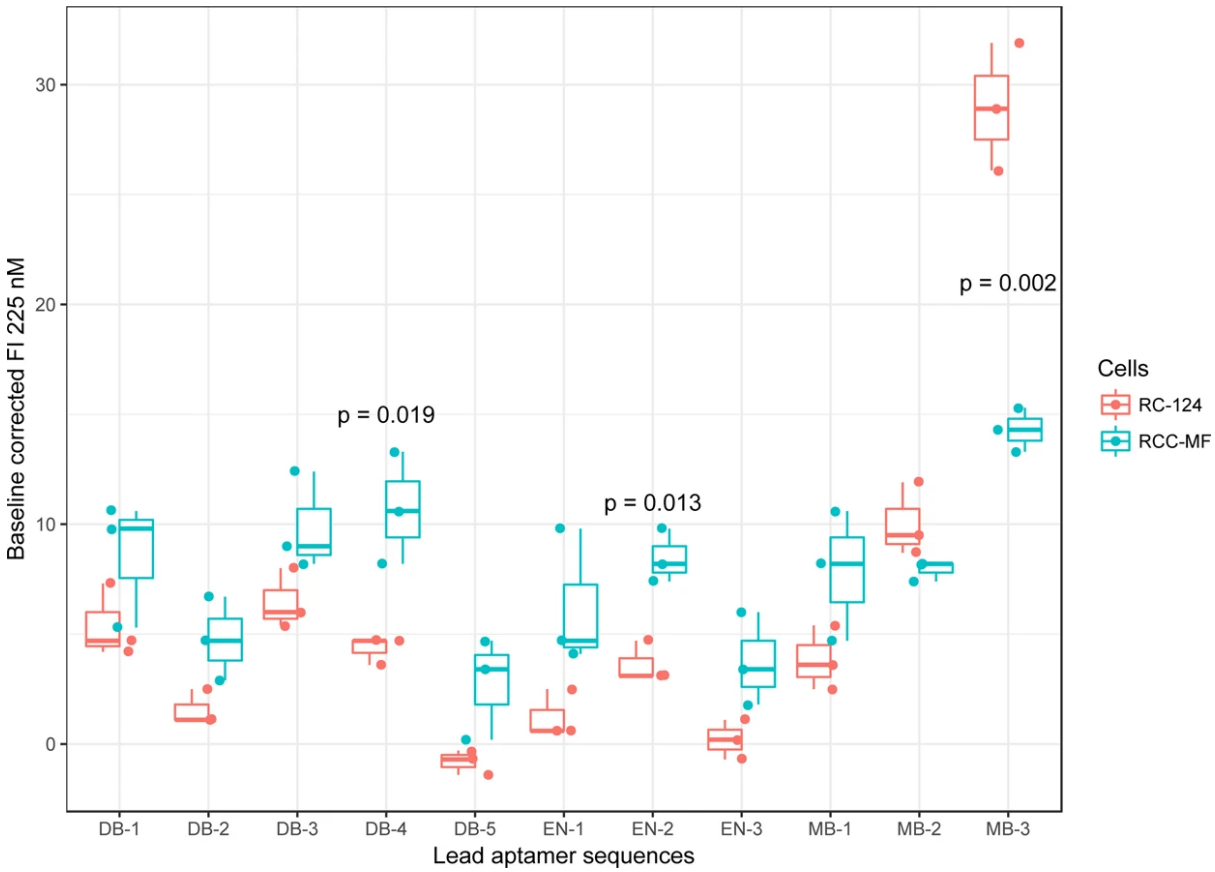

Figure 15. Lead aptamer testing at 225 nM concentration on target RCC-MF and control RC-124 cell lines using flow cytometry. Aptamers DB-1, DB-2, and DB-3 were selected from differential binding identified aptamers based on the highest counts per million. Aptamers DB-3, DB-4, and DB-5 were also identified using a differential binding approach but were selected based on the highest log_2FC values between RCC-MF cells and RC-124 cells. Aptamers EN-1, EN-2 and EN-3 were identified based on the highest enrichment between the 4th and 11th rounds of cell-SELEX. Aptamers MB-1, MB-2 and MB-3 were the most abundant sequences after the 11th round.

Comparison of the mode of fluorescence intensities from the lead sequences binding to RC-124 and RCC-MF cells. The top three sequences were identified by differential binding alone sorting by CPM (DB-1, DB-2, DB-3), differential binding together with enrichment analysis sorting by log2FC (DB-3, DB-4, DB-5), enrichment analysis alone sorting by log2CPM enrichment (EN-1, EN-2, EN-3) or by choosing the most abundant sequences in the sequencing dataset (MB-1, MB-2, MB-3). The upper hinges correspond to the first and third quartiles, and the whiskers mark the 1.5* interquartile range (IQR). Statistical significance was determined with a t-test using the GraphPad Prism software.

We chose four aptamers from differential binding analysis (DB-1, DB-2, DB-3, DB-4) and one aptamer identified as one of the most abundant ones in the enriched library (MB-3) to study concentration-dependant (0 – 1000 nM) binding using imaging flow cytometry.

DB-1 and DB-2 aptamers were binding to a higher extent to RCC-MF cells and the binding to RC-124 at respective concentrations was lower than that of the random library (sample/library fluorescence intensity < 1). DB-3 and DB-4 were binding more to RCC-MF cells in a concentration dependant manner and to a higher extent than DB-1 and DB-2 (sample/library fluorescence ratio higher) but binding to RC-124 cell line was higher than that of randomized oligonucleotide library (Fig. 16).

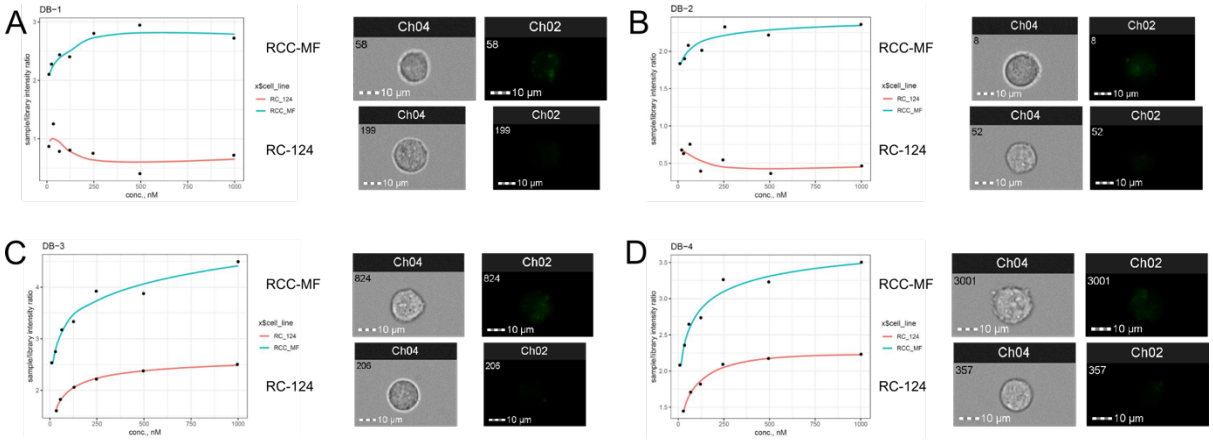

Figure 16. Differential binding aptamers using imaging flow cytometry. Concentration dependant aptamer binding to target RCC-MF cell line and control RC-124 cell line was compared. DB-1 (A) and DB-2 (B) aptamers show preferential binding to the RCC-MF cell line and binding to the RC-124 control cell line is below RND library background binding. DB-3 (C) and DB-4 aptamer (D) show higher binding to the RCC-MF cell line than DB-1 and DB-2 aptamer but the binding to control cell line RC-124 is above the background binding of the RND library. Representative imaging flow cytometry images for each cell line are seen on the right side of each respective graph. The x-axis indicates the concentration used for aptamer binding in nM, the y-axis indicates the fluorescence ratio of tested aptamer versus RND library binding at the respective concentration.

MB-3, aptamer identified as one of the most abundant ones in the enriched 11th round library, was binding to both RCC-MF cells and RC-124 cells. However, the binding intensity was several times higher for RC-124 (Fig. 17).

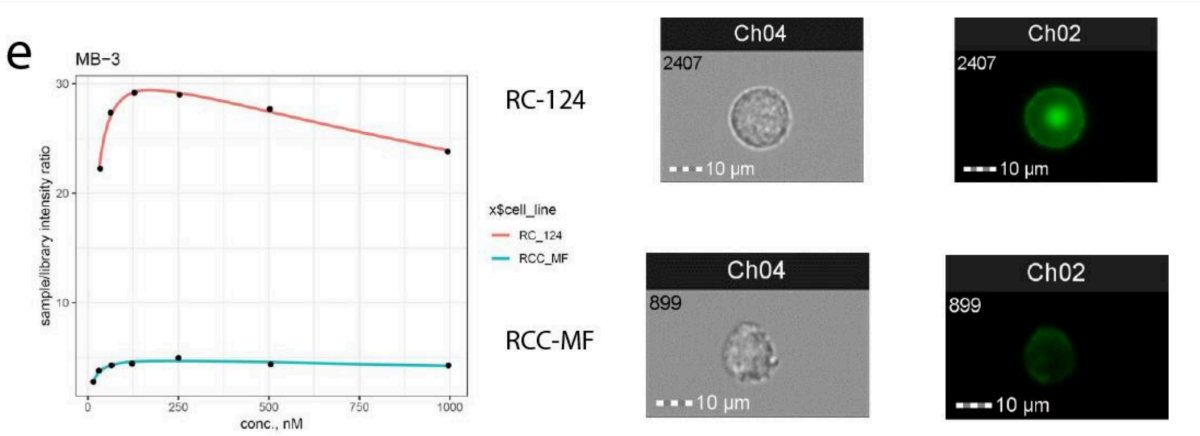

Figure 17. MB-3 binding to RC-124 and RCC-MF cells. MB-3 binds to control cell line RC-124 to a higher extent than to target cell line RCC-MF. Representative imaging flow cytometry images for each cell line are seen on the right side of the graph. The x-axis indicates the concentration used for aptamer binding in nM, the y-axis indicates the fluorescence ratio of tested aptamer versus RND library binding at the respective concentration.

We also predicted secondary structures of aptamers discovered through differential binding. NUPACK (Zadeh *et al.*, 2011) was used to predict structures at 4 °C as the folding temperature, with 5 mM Mg^{2+} and 157 nM Na^{+} concentrations. Surprisingly, DB-1 and DB-2, the aptamers with the highest selectivity for RCC-MF target cells, shared a big stem-loop structure with an 18-19 nt loop connected through the stem with a bulge and several internal loops. Large loops contained two (DB-1) or one (DB-1) smaller stem/loop (Fig. 18).

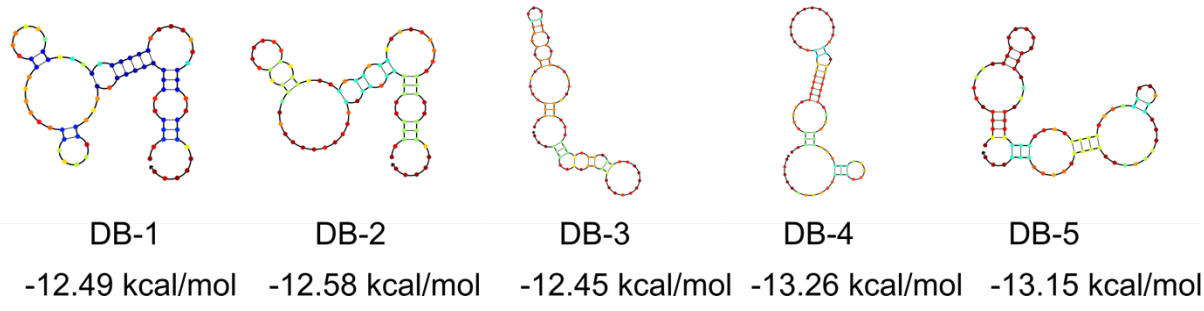

Figure 18. The predicted secondary structures of differential binding identified aptamers. DB-1, DB-2 and DB-3 were top three sequences from differential binding (logFC > 2, p < 0.0001)

when sorted based on CPM. DB-3, DB-4, DB-5 were identified both in enrichment analysis and differential binding analysis and were chosen based on the highest log_2FC values.

3.5. Lead aptamer GreenB1 shows selectivity towards triple-negative breast cancer cells

Due to the high distinctively high binding capacity compared to other aptamers, we tested MB-3, called GreenB1 further in the text, binding to two other carcinoma cell lines – MCF-7 (hormone-dependent estrogen and progesterone receptor-positive breast cancer cell line, has retained differentiated mammary epithelium characteristics) and MDA-MB-231 (poorly differentiated adenocarcinoma, represents triple-negative breast cancer). Both cell lines were incubated with increasing concentrations of either FAM-labelled random oligonucleotide library (RND-FAM) or FAM-labelled GreenB1 aptamer (GreenB1-FAM) and analysed using imaging flow cytometry.

Surprisingly, we noticed that while GreenB1 was binding to almost 95% of cells already at 5 nM concentration, even 1000 nM concentration caused only a minor fluorescent signal increase on MCF-7 cells, indicating a selectivity towards a less differentiated triple-negative breast cancer cell line (Fig. 19).

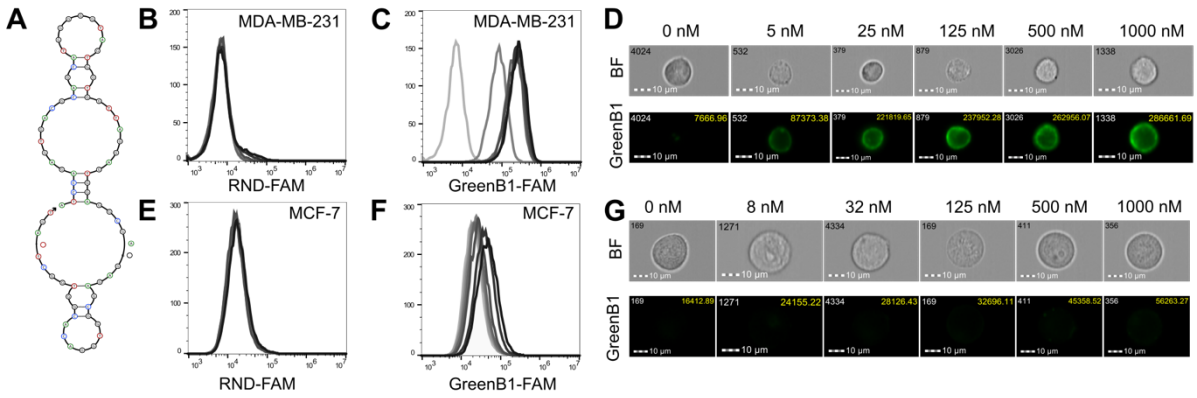

Figure 19. GreenB1 binds to triple-negative breast cancer cells. GreenB1 structure as predicted by NUPACK (A). Random aptamer (B) and GreenB1 (C) binding to MDA-MB-231 triple-negative breast cancer cell line at 0, 5, 25, 125, 250, 500, 1000 nM concentrations. Imaging flow cytometry images of GreenB1 binding to MDA-MB-231 (D) cell line at different concentrations. Random aptamer (E) and GreenB1 (F) binding to MCF-7 progesterone and estrogen receptor-expressing cell line at 0, 8, 16, 32, 62, 125, 500, 1000 nM concentrations.

Imaging flow cytometry images of GreenB1 binding to MCF-7 (G) cell line at different concentrations. The values in the right upper corner of the FAM channel image show the fluorescence intensities of the whole cells.

Imaging flow cytometry files (.rif, .cif, .daf and .cpm formats under "GreenB1-RND-MCF7-MDA-MB-231-different-conc") have been deposited on *BioStudies* (Sarkans *et al.*, 2018) and available here: https://www.ebi.ac.uk/biostudies/studies/S-BSST857.

3.6. Proximity labelling identifies β1-integrin as GreenB1 target protein

Observed selectivity towards the TNBC cell line led us to question the reason for this selectivity. To identify the target protein on the surface of MDA-MB-231 cells, we repurposed proximity labelling, a method extensively used for RNA-protein interaction studies, to be used for aptamer target identification (Fig. 20).

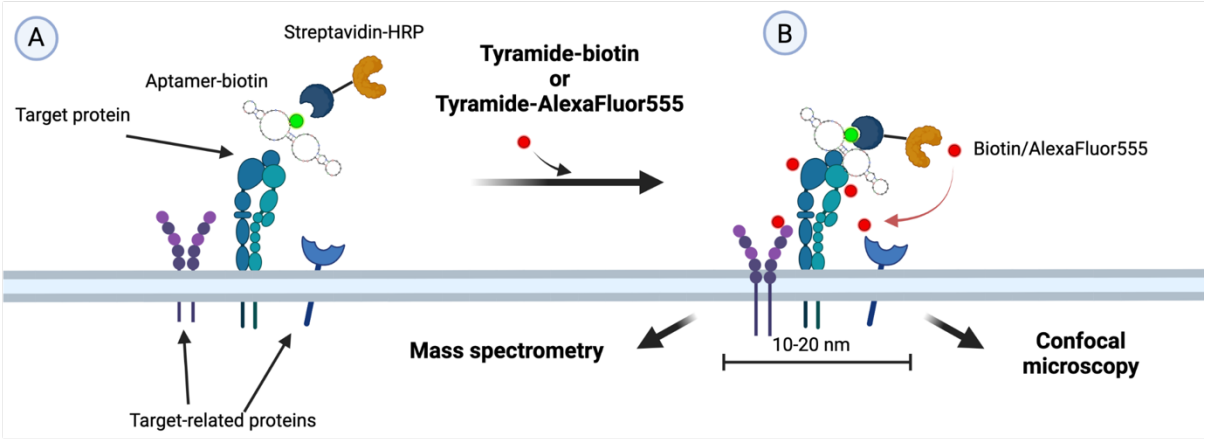

Figure 20. GreenB1 protein target identification using proximity labelling. Biotin labelled GreenB1 or RND was incubated with horseradish peroxidase (HRP) conjugated to streptavidin. Complexes or streptavidin-HRP alone were incubated with live MDA-MB-231 cells for 1 h (A). After washing away the unbound complex, tyramide-biotin or tyramide-AlexaFluor555 with hydrogen peroxide was added to cells for 2 min. HRP, in the presence of hydrogen peroxide, creates a highly reactive tyramide species that labels nearby proteins (B). Fluorescently labelled proteins were further imaged using confocal microscopy. Biotinylated proteins were pulled down using streptavidin-coated magnetic beads, eluted using 25 mM biotin in lysis buffer and heated at 95 °C for 5 min. Eluates were run on the gel and further analysed using mass spectrometry.

Initially, to confirm that the method is suitable for aptamer target identification, we conjugated horseradish peroxidase and streptavidin (HRP-S) to biotin-labelled aptamer GreenB1 or biotin labelled RND library to create complexes with GreenB1 (HRP-S-B-GreenB1) and RND library (HRP-S-B-RND). Prepared complexes and HRP-S were incubated with adherent cells *in vitro*. Tyramide-AlexaFluor555 was used to image the cell labelling with HRP-S-B-GreenB1, HRP-S-B-RND or HRP-S alone. We observed bright staining with HRP-S-B-GreenB1, a few bright spots with HRP-S-B-RND and no signal with unconjugated HRP-S (Fig. 21).

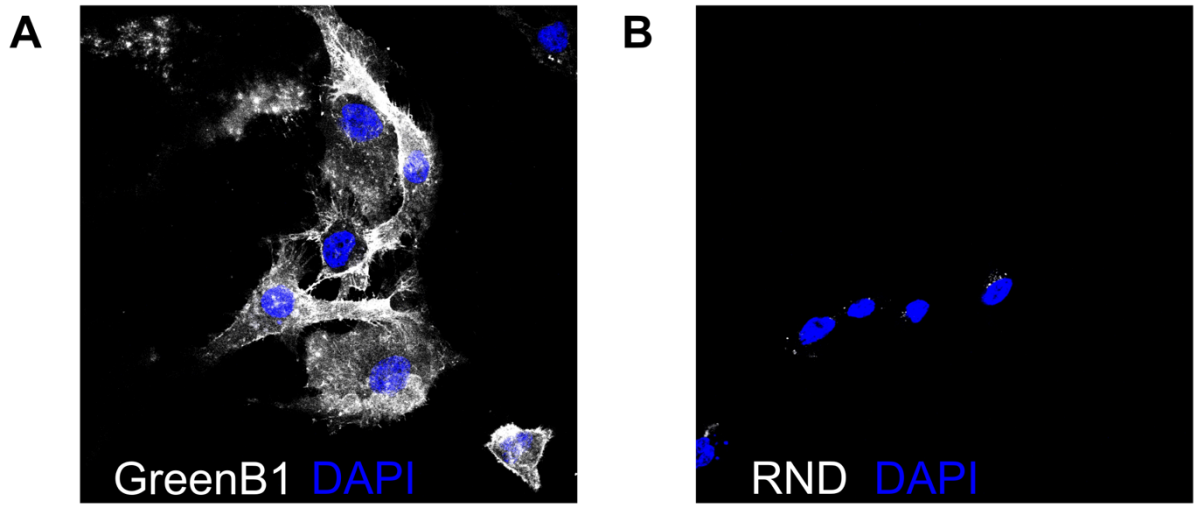

Figure 21. Imaging of proximity labelling reaction. Confocal microscopy images of proximity labelling using Tyramide-AlexaFluor555 with MDA-MB-231 cells after binding of GreenB1-HRP complex (A) or RND-HRP complex (B).

Further, we incubated prepared HRP-S-B-GreenB1, HRP-S-B-RND complexes or HRP-S alone with MDA-MB-231 cells in suspension, performed proximity labelling reaction using tyramide-biotin and pulled-down labelled proteins using streptavidin-conjugated beads. Two initial elution for all samples were done with excess biotin and heat, followed by elution with reducing sample buffer and heat. HRP-S, RND and GreenB1 all had two distinct bands around 40 kDa and 70 kDa in the first elution (Fig 22 bands 2, 5, 8 respectively), most likely coming from HRP protein and streptavidin. However, only the GreenB1 elution sample had another distinguishable band around 130 kDa (Fig 22, band 8, indicated by the arrow). This band of interest was cut out and used for MS proteomics along with the same molecular weight region from the RND sample.

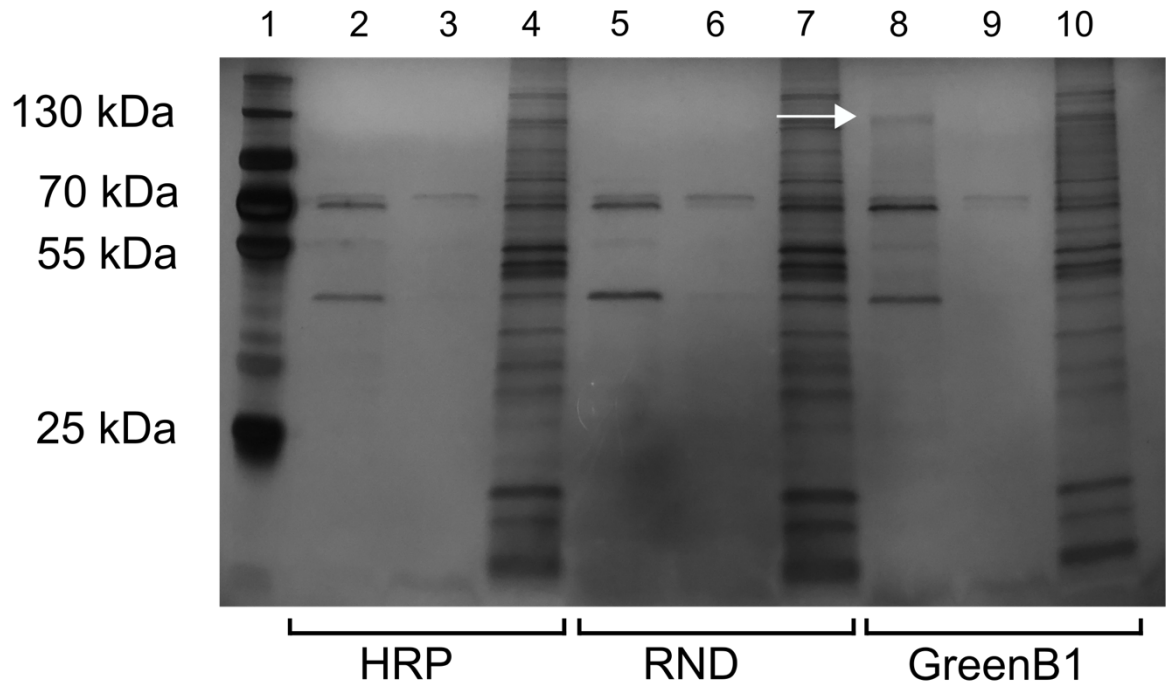

Figure 22. Pull-down results from proximity labelling with tyramide-biotin. Lane 1 contains a marker. Streptavidin-HRP (lanes 2, 3, 4), RND-HRP (lanes 5, 6, 7) and GreenB1-HRP (lanes 8, 9, 10). The first two lanes in each sample were eluted using biotin and heat; the third was eluted using a reducing sample buffer.

After filtering out contaminant proteins, keeping proteins with at least 2 unique peptides and proteins with signal intensity ratio in the GreenB1 sample over the RND sample of at least 10, we were left with a list of 28 proteins (Table 4).

Table 4. MS proteomics after data from proximity labelling experiments using HRP-bound GreenB1 or RND. The first column contains full protein name, the second contains gene symbol for the respective protein. The third column indicates MS intensity ratio between GreenB1 labelled sample and RND labelled sample. The fourth column indicates the fraction signal intensity coming from the respective protein out of total signal intensity within the sample.

Protein names	Gene names	GreenB1/RND	fraction of total GreenB1 intensity*1000
Integrin beta-1	ITGB1	766	22.80
Integrin alpha-3	ITGA3	761650000	4.08
CD44 antigen	CD44	215690000	1.16
Kinectin	KTN1	17	0.99
Actin, cytoplasmic 1	ACTB	15	0.77

Cell surface glycoprotein MUC18	MCAM	134760000	0.72
Isoform 4 of Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	81395000	0.44
Isoform Alpha-6X2A of Integrin alpha-6	ITGA6	73210000	0.39
Histone H4	H4C1	56999000	0.31
Integrin alpha-5	ITGA5	46882000	0.25
Isoform 3 of Integrin alpha-V	ITGAV	43403000	0.23
Protein EVI2B	EVI2B	37945000	0.20
Keratin, type II cytoskeletal 4	KRT4	33353000	0.18
Isoform 2 of Keratin, type II cytoskeletal 80	KRT80	25	0.18
Immunoglobulin heavy constant gamma 1 (Fragment)	IGHG1	30679000	0.16
Activated leukocyte cell adhesion molecule	ALCAM	22385000	0.12
14-3-3 protein sigma	SFN	17335000	0.09
Annexin A1	ANXA1	16258000	0.09
Integrin alpha-2	ITGA2	14651000	0.08
Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1	11468000	0.06
2-phospho-D-glycerate hydro-lyase	ENO1	10733000	0.06
Pyruvate kinase (Fragment)	PKM	9883200	0.05
Immunoglobulin heavy constant alpha 1 (Fragment)	IGHA1	9545600	0.05
Serpin B4	SERPINB4	8388700	0.04
Cytoskeleton-associated protein 4	CKAP4	6733400	0.04
Catalase	CAT	6601700	0.04
Fructose-bisphosphate aldolase	ALDOA	6407500	0.03
Protein-glutamine gamma-glutamyltransferase K	TGM1	3788500	0.02

Table 4. MS proteomics after data filtering.

Integrin beta-1 (ITGB1) was the protein that had the highest signal intensity - 22.8 parts per thousand (ppt) of the total signal intensity from the proximity labelling sample using GreenB1. ITGB1 signal in the GreenB1 sample was 766 times higher than in the RND sample. ITGB1 molecular weight (Bodary and McLean, 1990) also corresponds to the approximate location of the band on the gel. Integrin alpha-3 (ITGA3), the protein with the second highest intensity, had a 5.6 times lower signal. Interestingly, ITGA3 together with ITGB1 forms the most prevalent ITGB1-containing integrin complex ITGA3B1. CD44 antigen was the third protein based on the signal intensity that was 19.7 times lower than that of ITGB1. Signal intensity ratio log2 values between GreenB1 and RND eluate proteins with the highest signal intensity can be seen in Fig. 23.

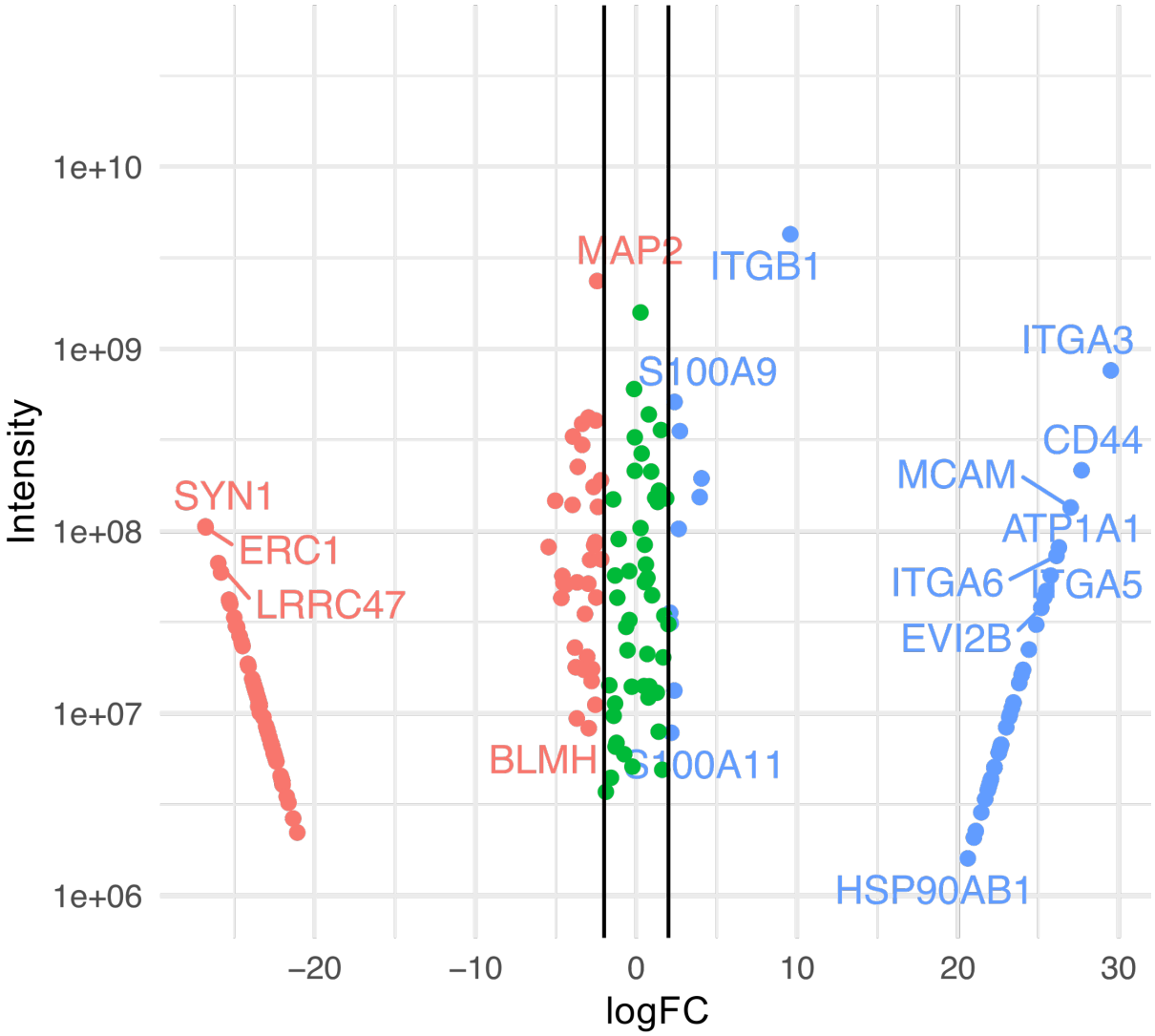

Figure 23. Proximity labelling results for GreenB1 target identification. Mass spectrometry proteomics of control and target bands corresponding to 130 kDa. LogFC on the x-axis. Combined signal intensity from both samples on the y-axis.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch *et al.*, 2017) via the PRIDE (Perez-Riverol *et al.*, 2021) partner repository with the dataset identifier PXD034982 and 10.6019/PXD034982.

3.7. GreenB1 has K_d in the low nanomolar range

We further studied α3β1-integrin complex interaction with GreenB1 using electrophoretic mobility shift assay (EMSA) and β1-integrin interaction with GreenB1 using fluorescence polarization (FP). For EMSA, recombinant α3β1-integrin complex expressed in human cells

was incubated at different concentrations (0 – 90 nM) with either GreenB1 (Fig. 24A) or RND (Fig 24B). After the incubation with GreenB1, it can be observed that the aptamer band that is very distinct when α3β1-integrin is not added, disappears with increasing α3β1-integrin concentration. The second band that would indicate GreenB1-α3β1-integrin complex is not visible due to the very high molecular weight of α3β1-integrin (around 300 kDa) and it is either stuck in the well or does not enter the gel at all. When incubating α3β1-integrin with the RND library, there is no band shift observable, and the oligonucleotide band retains its intensity irrespectively of the α3β1-integrin concentration. Based on the GreenB1 signal reduction with increasing concentrations of α3β1-integrin in EMSA (n=3), we were able to determine the GreenB1 has affinity to α3β1-integrin (K_d = 15 nM, 95% confidence interval 8 – 26 nM) (Fig. 24C).

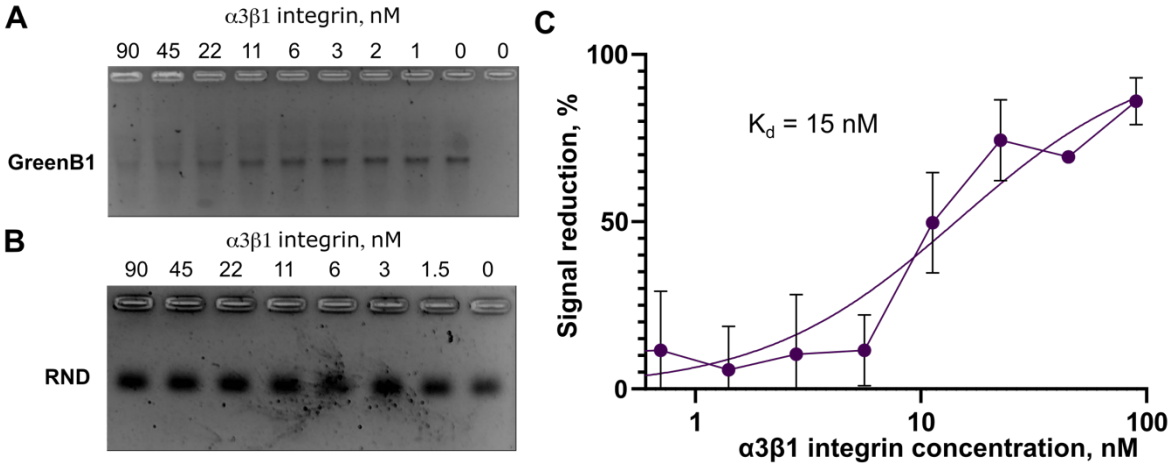

Figure 24. GreenB1 binding to α3β1-integrin complex. Electrophoretic mobility shift assay (EMSA) using increasing concentrations of α3β1-integrin protein at 17.5 nM fixed concentration of GreenB1 (A) or RND (B) aptamers. GreenB1 and α3β1-integrin EMSA results plotted as reduction of GreenB1 signal intensity (Kd = 15 nM, 95% CI 8-26 nM) (C).

FP analysis was performed using β1-integrin instead of α3β1-integrin complex to determine if GreenB1 can bind a single subunit of the complex. Serial dilutions of β1-integrin (0 - 180 nM) were incubated with either FAM labelled GreenB1 or RND library and fluorescence polarization changes were recorded at each concentration (Fig. 25). RND library did not change polarization over the β1-integrin concentration range while with GreenB1 there was observable polarization increase that allowed to calculate the affinity against β1-integrin (K_d = 7 nM, 95% confidence interval 0 - 17 nM).

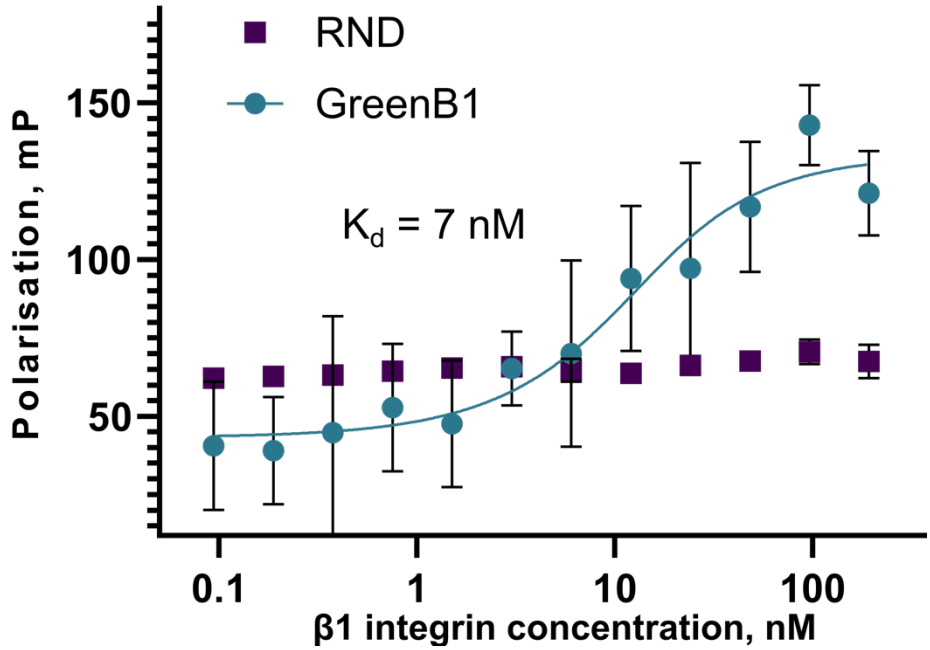

Figure 25. GreenB1 binding to β1-integrin. Fluorescence polarization using 10 nM of FAM-labelled GreenB1 aptamer or FAM-labelled RND and varying concentrations of β1-integrin. (K_d = 7 nM, 95% CI 0 – 17 nM) (D). Plots depict averages of triplicate measurements ± SE and the fitted model.

3.8. GreenB1 internalization after binding to cells *in vitro*

Further, we wanted to determine the fate of GreenB1 upon binding to MDA-MB-231 cells. To do that, we incubated FAM labelled GreenB1 or RND with adherent cells grown in culture in complete growth media at different concentrations (0-500 nM) for 24 h, dissociated afterwards using a non-enzymatic dissociation solution and incubated both FAM-GreenB1 and FAM-RND labelled samples with 100 nM GreenB1-Cy5 for imaging flow cytometry analysis to see if GreenB1 target is still available after the pre-incubation with either GreenB1 or RND. Different concentrations of the FAM-RND library showed no concentration-dependant binding after 24 h (Fig. 26A) and did not change GreenB1 target availability at any of the concentrations (Fig. 26B). FAM-GreenB1 appeared to have concentration-dependant binding during the pre-incubation for 24 h (Fig. 26C) but the target protein availability did not change (Fig. 26D), indicating either that new GreenB1 target protein has been synthesized or GreenB1 aptamer has dissociated from the target protein.

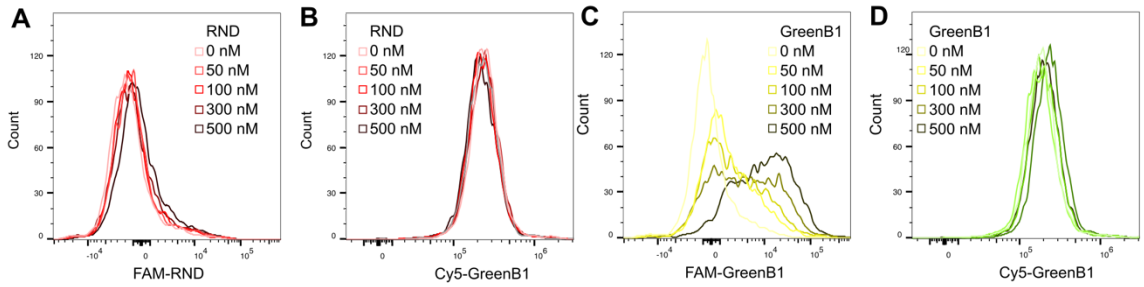

Figure 26. GreenB1 target availability after pre-incubation with GreenB1 or RND. FAM-labelled RND aptamer library pre-incubation with MDA-MB-231 cells on a 6-well plate at different concentrations for 24 h (A) followed by incubation with 100 nM Cy5-GreenB1 at 4°C for 1 h (B) shows no concentration-dependant fluorescence increase. FAM-labelled GreenB1 aptamer pre-incubation for 24 h with MDA-MB-231 cells on a 6-well plate at different concentrations (C), followed by 100 nM Cy5-GreenB1 incubation at 4°C for 1 h (D) results in concentration-dependant fluorescence increase during the pre-incubation but does not affect Cy5-GreenB1 binding afterwards.

Internalization of GreenB1 was characterized by doing a pulse-chase experiment where Cy5-labelled GreenB1 in complete growth media was pre-incubated with adherent cells grown in culture for 1 h and replaced with GreenB1-free complete growth media afterwards. Imaging flow cytometry was performed 1, 2, 3, 4, and 24 h after the removal of Cy5-GreenB1. Before imaging flow cytometry, LysoTracker Green was added to each sample to observe co-localization with acidic compartments in the cells. We observed time-dependent fluorescence increase (Fig. 27A) and Bright Detail Similarity analysis used to determine the co-localization based on the spatial distribution of two fluorescent channels, showed that the highest co-localization of Cy5-GreenB1 and LysoTracker Green can be observed after 1 h and decreases over time (Fig. 27B). Representative images show that LysoTracker GreenB1 signal overlaps with Cy5-GreenB1 after 1 h and dissociates within the cells over time (Fig 27C).

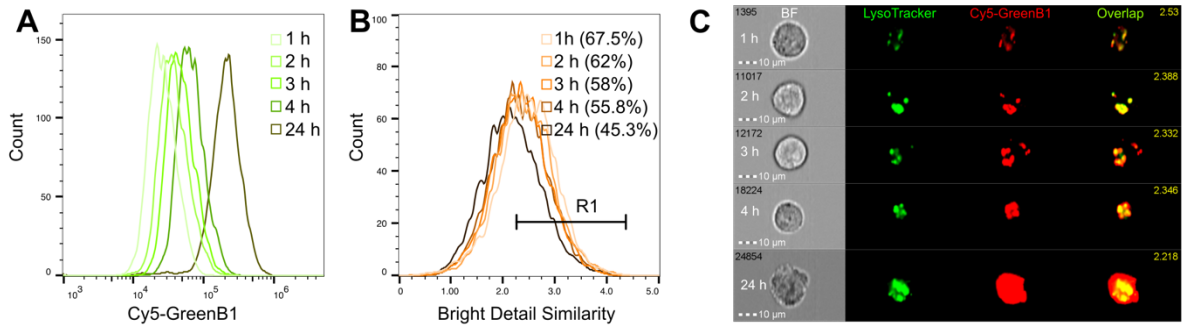

Figure 27. Pulse-chase experiments using LysoTracker. Cy5-labelled GreenB1 aptamer pulse-chase incubation 100 nM with MDA-MB-231 cells in a 6-well plate results in a time-dependent Cy5-fluorescence intensity increase (A). Cy5-GreenB1 co-localization with LysoTracker Green shows the highest co-localization based on bright detail similarity between LysoTracker Green and Cy5-GreenB1 at 1 h after incubation and a slight decrease after 24 h (B). Representative images of Cy5-GreenB1 co-localization with lysosomes after different time points (C).

3.9. Systemically administered GreenB1 homes to β1-integrin positive tumour lesions in mice

GreenB1 ability to home to a β1-positive carcinoma model *in vivo* was explored using orthotopically induced murine 4T1 cell line tumour lesions in mice. GreenB1 ability to bind to 4T1 cells was tested using flow cytometry which confirmed that a fraction of 4T1 cells were bound to FAM-GreenB1 (Fig. 28).

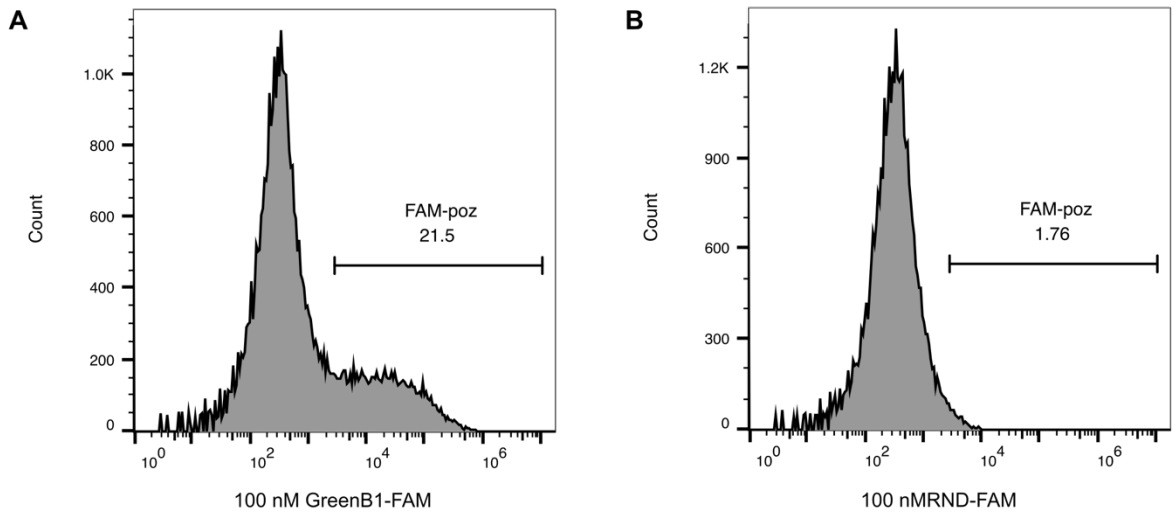

Figure 28. GreenB1 aptamer binding to 4T1 cells. FAM-GreenB1 or FAM-RND were folded and diluted to 100 nM concentration using the binding buffer. 4T1 cells were detached using a non-enzymatic cell dissociation buffer, washed, and incubated with FAM-GreenB1 or FAM-RND for 1 hour on ice. Cells were then washed two times and analysed using BD Accuri C6 Plus. A higher fraction of 4T1 FAM^+ cells in the GreenB1 sample (A) compared to the RND sample (B) confirm GreenB1 binding to 4T1 cells.

Further, $1*10^6$ 4T1 cells were subcutaneously injected in a mammary fat pad into Balb/C mice (n=6). FAM-GreenB1 (2 nmol) or FAM-RND (2 nmol) were injected intravenously into mice with tumour lesions that developed after 7 days. After 1 h circulation time, mice were sacrificed and the tumour along with control organs (heart, kidney, liver, lung) were collected for fluorescent immunohistochemistry imaging (Fig. 29A). We observed no statistically significant binding differences between GreenB1 (n=3) and RND (n=3) in the heart ($p=0.99$), kidney ($p=0.54$), liver ($p=0.99$) or lung ($p=0.99$). However, GreenB1 was binding statistically significantly more ($p<0.0001$) to 4T1 tumour lesions compared to the RND library (Fig. 29B).

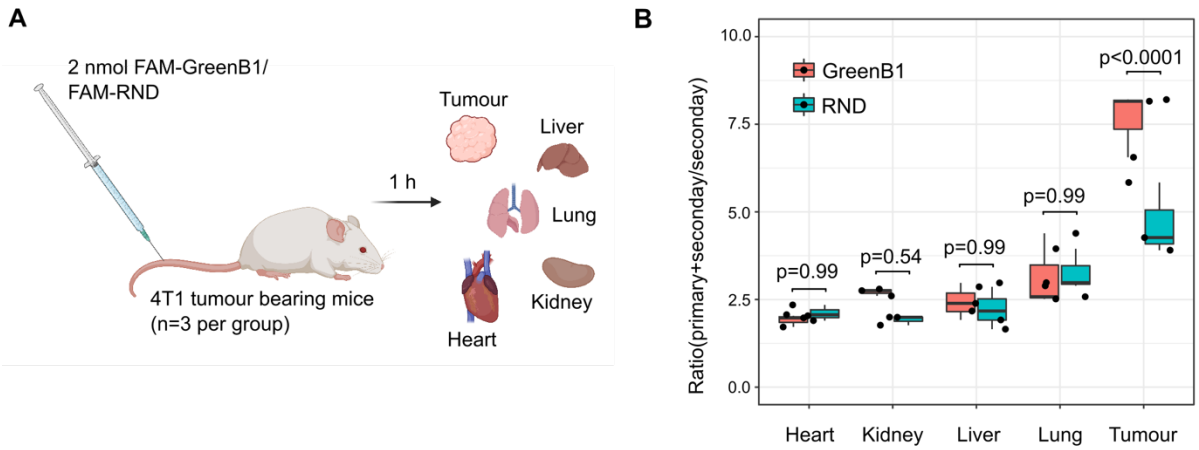

Figure 29. *In vivo* homing of systemically administered GreenB1 aptamer. FAM-GreenB1 or FAM-RND were injected i.v. into Balb/C mice (n=3 per group) carrying orthotopic murine 4T1 triple-negative breast cancer. The tumour and control organs were collected (A). The mean fluorescence intensity from each organ was calculated from five fields of view. A statistically significant fluorescence intensity difference was detected between tumours from mice injected with FAM-GreenB1 compared to those from mice injected with FAM-RND ($p<0.0001$). There were no statistically significant fluorescence differences between control organs (heart, kidney, liver, lung) from FAM-GreenB1 injected mice and FAM-RND injected mice (C).

Confocal imaging of tumour tissue shows a higher accumulation of GreenB1 in tumour lesions compared to the RND library (Fig. 30).

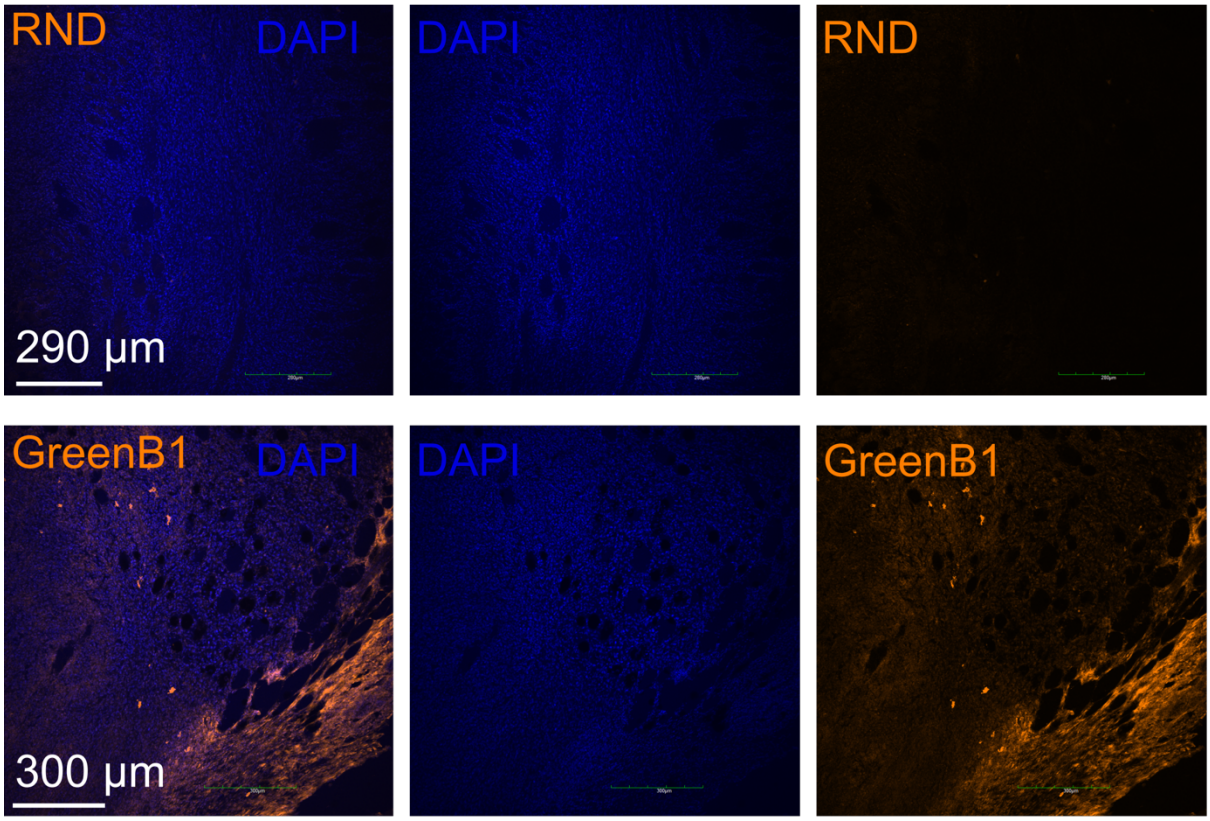

Figure 30. Confocal imaging of tumour lesions. Representative confocal microscopy images of tumours labelled with anti-FAM antibody and secondary antibody with AlexaFluor555 from mice injected with FAM-GreenB1 (second row) or FAM-RND (first row).

Epifluorescence imaging of whole tumour cross-sections using a slide scanner indicates a higher amount of fluorescence within GreenB1 injected mice (Fig. 31A) compared to RND library (Fig. 31B) injected mice. GreenB1 has higher intensity staining at the tumour margins and it can be observed at the tumour core as well.

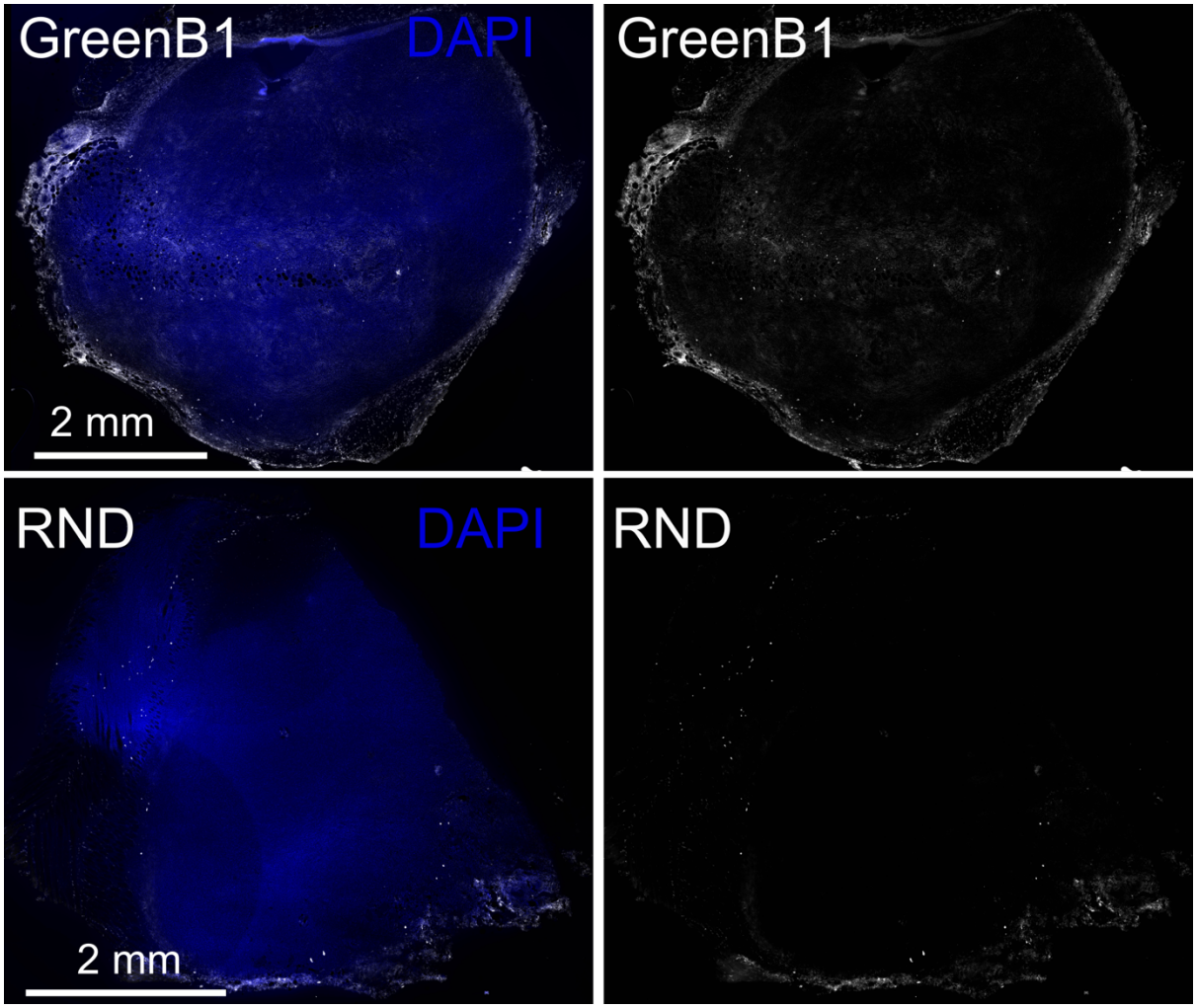

Figure 31. Whole tumour imaging after FAM labelled GreenB1 or RND injection into 4T1 cell line TNBC model bearing mice. The upper row shows tumour collected from GreenB1 injected mouse; the lower row shows tumour from RND injected mouse. Cell nuclei were labelled using DAPI. FAM signal from RND and GreenB1 was amplified using anti-FAM antibody.

Fluorescent microscopy images used for statistical analysis (.nd2 format under "GreenB1-in-vivo-images"), confocal images (.oib format under "GreenB1-confocal-images"), slide scanner images (.scn and .sdf formats under "GreenB1-slide-scanner-images") have been deposited on *BioStudies* (Sarkans *et al.*, 2018) and are available here: https://www.ebi.ac.uk/biostudies/studies/S-BSST857.

4. DISCUSSION

4.1. Aptamer selection and advantages of differential binding SELEX

During the selection of aptamers binding to ccRCC cell line RCC-MF and kidney epithelial cell line RC-124 as a negative control, we observed enrichment of the library and increased binding to the target cell line after 11 rounds. However, we also observed increased binding towards the control cell line. We further hypothesized that it could be possible to isolate individual target cell line selective aptamers from the enriched library by using HTS and developing a differential binding data analysis method.

The main goal for the development of the differential binding cell-SELEX approach was to increase the reliability of identified aptamers by using data analysis methods repurposed from RNA-sequencing. HTS has increased the amount of information extractable from aptamer selections and can be considered one of the biggest advancements in aptamer identification (Ozer, Pagano and Lis, 2014). Already available data analysis tools for aptamer selection are mainly focused on the enrichment of target binding aptamers over multiple selection rounds (Alam, Chang and Burke, 2015; Caroli *et al.*, 2016; Caroli, Forcato and Bicciato, 2020; Kramer *et al.*, 2022). Differential binding SELEX looks at the differences in aptamer binding within a single selection round, making it possible to reduce the number of aptamer selection rounds.

Statistical analysis is the cornerstone of the differential binding approach. Each experiment must be carried out with at least three replicates in target and control groups for the results to be meaningful. Out of several HTS data analysis tools that have been developed to carry out the differential expression analysis for RNA-sequencing (Robinson, McCarthy and Smyth, 2010; Trapnell *et al.*, 2012; Ritchie *et al.*, 2015), we chose *edgeR* due to the compatibility with our existing data analysis workflows that were written using *R* (R Core Team, 2016) programming language. All aptamers that were identified using differential binding analysis (DB-1, DB-2, DB-3, DB-4, DB-5) and were tested individually, showed preferential binding to target cell line RCC-MF compared to control cell line RC-124 (Fig. 15 and 16). Notably, these lead aptamers were neither top hits in enrichment analysis, nor most abundant sequences and it would not be possible to identify these aptamers using previously developed methods for aptamer identification from HTS data. The differential binding approach has been further

adapted for *in vivo* use to identify affinity ligands using a method technically similar to aptamer SELEX – *in vivo* phage display (Põšnograjeva *et al.*, 2022). Differential binding for *in vivo* phage display can not only correctly identify the brain homing peptide that has been used as a spike-in but also provides information about biodistribution between the target organ and several control organs using only HTS data (Pleiko *et al.*, 2021).

Altogether, we demonstrate that even from seemingly unsuccessful cell-SELEX experiments, it is possible to use statistical analysis algorithms coupled with HTS data to identify aptamer sequences that are selective towards the intended target molecule.

4.2. GreenB1 target protein identification

GreenB1 showed, while not selective, the highest binding to the RCC-MF cell line and we thus decided to characterize its interaction further. After screening on available cell lines in our possession, binding analysis to breast cancer cell lines MCF-7 and MDA-MB-231 provided the data that GreenB1 binds to triple-negative breast cancer cell line MDA-MB-231 but not to estrogen and progesterone expressing breast cancer cell line MCF-7 (Fig. 19B-G). The next step that we decided to take was to identify the target protein the cell surface. Most abundant aptamer sequences often are not the best binders (Hoinka *et al.*, 2015) and can be enriched in the library either because they are amplified better in PCR or due to inefficient negative selection. Previously, extract-based protein pull-down methods using whole-cell lysate have been used to identify aptamer target proteins (Bing, Shangguan and Wang, 2015). Proximity labelling has been extensively used to study RNA/DNA interactions with proteins that have RNA/DNA-binding domains (Hentze *et al.*, 2018). We adopted proximity labelling to identify aptamer target protein due to the lower background binding than with extract-based methods. Proximity labelling uses peroxidase or ligase to generate biotin-linked reactive species that covalently label the neighbouring proteins. When linked to a targeting moiety such as aptamer, the enzyme is guided towards the aptamer binding protein and thus labels the target protein and nearby proteins with biotin. Further, the labelled proteins can be pulled down and identified using the mass spectrometry proteomics approach (Qin *et al.*, 2021).

Our results from MS proteomics strongly indicate that the target protein for GreenB1 aptamer is β1-integrin. Not only it had the highest signal intensity, but the MW of β1-integrin, around

130 kDa (Bodary and McLean, 1990), also correspond to the MW of the band that we observed in the GreenB1 proximity labelling sample. Further confirmation that the target protein is β1-integrin came from the fact that the α3-integrin protein had the second highest MS intensity. β1-integrin forms a complex with α3-integrin that binds to laminin (Riopel *et al.*, 2013). This integrin protein complex is a promising anti-cancer target (Subbaram and Dipersio, 2011) and these results led us to the idea to further explore GreenB1 interaction with both α3β1-integrin complex and β1-integrin alone. CD44 was identified as the third most promising protein candidate. While the MW of CD44 did not match that of the excised band, increased expression of CD44 is associated with a more aggressive form of ccRCC (Zanjani *et al.*, 2018).

We used EMSA to measure the GreenB1 affinity to α3β1-integrin. Recombinant α3β1-integrin protein complex was incubated with folded GreenB1 aptamer and run agarose gel to see if we can observe the band shift that would confirm GreenB1 binding to protein. While we saw that the GreenB1 band on the gel became less intense with increased protein concentration, we did not observe the clearly defined second band that would show protein-aptamer complex (Fig. 24A). We did not observe similar aptamer band signal reduction when α3β1-integrin was incubated with RND starting library as a control. This led us to the conclusion that GreenB1 signal reduction is due to the selective binding to protein. The band shift was likely not observed due to the high MW of α3β1-integrin complex (275 kDa) that prohibited the aptamer-protein complex from entering the gel. However, to further confirm GreenB1 binding to β1-integrin, we carried out FP analysis where fluorescently labelled GreenB1 aptamer was incubated with β1-integrin. Both methods determined that GreenB1 binds to α3β1-integrin and β1-integrin alone with low nanomolar affinity.

α4-integrin (CD49d) has been identified as a target protein for aptamer Scg-4c (5'-TTTATCACTTATTCAATTCGAGTGCGGATGCAAACGCCAGACAGGGGGACAGGA GATAAGTGATTT-3'). Sgc-4c has a dissociation constant in the nanomolar range and it was identified using human precursor T cell acute lymphoblastic leukaemia cells (CCRF-CEM) as the target cell line (Shangguan *et al.*, 2007). The aptamer target protein was identified using the pull-down method after binding to cells *in* vitro (Bing, Shangguan and Wang, 2015). Recently, α3-integrin (CD49c) has been identified as a target protein for TNBC cell line

binding aptamer PDGC21T (Wan *et al.*, 2022). These previous studies confirm that integrins can be considered "targetable" using aptamers. GreenB1 is the first aptamer identified to bind β1-integrin.

4.3. Characterization of GreenB1 binding and internalization *in vitro* and *in vivo*

β1-integrin is recycled between the cytosol and cell surface. For some integrins, almost all integrin proteins available on the plasma membrane can be internalized within 30 minutes (Paul, Jacquemet and Caswell, 2015). We aimed to further characterize the interaction with GreenB1 target proteins using *in vitro* cell cultures.

We tested if the GreenB1 target protein is still available for binding after prolonged pre-incubation with GreenB1. After incubation with FAM labelled GreenB1 aptamer or RND oligonucleotide library at different concentrations for 24 h. We dissociated the adherent cells and added 100 nM Cy5 labelled GreenB1 to GreenB1 pre-incubated and RND pre-incubated cells *in vitro* for 1 h. Using imaging flow cytometry, we observed that the FAM-RND library was binding at the background level without any concentration-dependent pattern. However, FAM-GreenB1 increased the fluorescence intensity at each concentration step (50 nM, 100 nM, 300 nM, 500 nM). Results suggest that no receptors saturation is reached at the highest concentration (Figure 26C), something that contradicts the data acquired after the incubation with GreenB1 at 4° C, where we observed only minor fluorescence increase at concentrations higher than 25 nM (Fig. 19C). Cy5-GreenB1 binding at 100 nM did not depend on the concentration of FAM-GreenB1/FAM-RND used for pre-incubation (Figure 26B and Fig. 26D). These results can be explained either by continuous uptake of GreenB1 that could take place if GreenB1 target is being recycled, or synthesis of new target proteins that would provide the same amount of target proteins available for binding. Estimated half-life for β1-integrin is 20 h (Böttcher *et al.*, 2012), suggesting that the recycling of β1-integrin is a viable explanation for not observing the receptor saturation.

Our next goal was to determine the fate of GreenB1 aptamer upon binding to the target protein. We performed a pulse-chase experiment using Cy5 labelled GreenB1. GreenB1 was removed and replaced with GreenB1-free culture media. Cells were dissociated from the flask, LysoTracker Green was added, and cells were analysed using an imaging flow cytometer at

different time points over the next 24 h. The results indicate time-point dependent increase of Cy5-GreenB1 fluorescence intensity without noticeable saturation. Cy5 signal is co-localized with acidic vesicles that are labelled with LysoTracker Green. Co-localization is highest after 1 h and decreases over time. The drop in co-localization can be explained by aptamer being delivered to lysosomes where aptamer itself is degraded while Cy5 fluorescence dissociates within the cytosol. The lack of saturation indicates continuous uptake of aptamer into cells after binding to cell surface proteins and trafficking to acidic vesicles.

The trafficking to lysosomes fits the previously established route of oligonucleotide delivery, according to which, oligonucleotides dissociate from proteins in early endosomes due to the pH change that also affects the three-dimensional structure of the aptamer. Further, while protein is recycled back to the plasma membrane, oligonucleotides are transported to the lysosome for degradation (Juliano, 2016).

4T1 cell line xenograft model is a syngeneic mouse model and it allows researchers to carry out experiments in animals with a functional immune system (Tao *et al.*, 2008). The unreliability of translating aptamer binding observations from protein-SELEX to *in vitro* binding or aptamer binding results from cell-SELEX to *in vivo* setting has been indicated previously and is a relatively well-known problem in the aptamer field (Kelly *et al.*, 2021). Despite GreenB1 being selected using cell-SELEX, we also observed preferential GreenB1 homing *in vivo* to murine triple-negative breast cancer representative cell line 4T1 lesions. Surprisingly, while statistically significantly more GreenB1 accumulation was observed in tumours, the accumulation in the kidney was not statistically different from mice injected with the RND aptamer library (Fig. 29B) despite GreenB1 binding to kidney epithelial cell line RC-124 *in vitro*. Our initial explanation for this observation is that while β1-integrin is expressed on kidney epithelial cells, it is more available for binding from systemic circulation in tumour lesions. Vasculogenic mimicry, a tumour-blood supply mechanism that forms blood-vessel-like structures from cancerous cells, requires β1-integrin expression (Kawahara, Niwa and Simizu, 2018).

4.4. Potential future directions

Modifications to increase GreenB1 suitability for in vivo use

Unmodified aptamers are prone to degradation by nucleases in serum, have a short half-life in circulation and are rapidly excreted through renal filtration (Zhou and Rossi, 2017). However, these characteristics are problematic to a different extent for each aptamer – phase I clinical trial has shown that unmodified ssDNA aptamer ApTOLL has a circulating half-life of 9.3 h (Hernández-Jiménez *et al.*, 2022). Further studies are necessary to determine the pharmacokinetic properties of GreenB1. If necessary, the circulation time can be increased either by creating a multivalent aptamer construct that is larger than the glomerular filtration cut-off of 50-60 kDa (Mallikaratchy *et al.*, 2011) or by conjugation to PEG (Thomas, Porciani and Burke, 2022). Circular aptamers are less suspectable to nuclease degradation and it can be done when aptamer degradation prohibits its use *in vivo* (Litke and Jaffrey, 2019).

Targeted protein degradation using aptamers

GreenB1 co-localization with LysoTracker Green suggests that after being endocytosed it is trafficked to lysosomes for degradation. Lysosome-targeting chimaeras (LYTACs) are taking advantage of endogenous lysosome shuttling proteins and are used to guide extracellular and membrane-bound protein targets for degradation in lysosomes (Banik *et al.*, 2020). Cation-independent mannose-6-phosphate receptor (CI-M6PR) and asialoglycoprotein receptor (ASGPR) have both been used as lysosome-guiding proteins targets that are recycled to develop LYTACs able to degrade proteins-of-interest *in vitro* and *in vivo* (Ahn *et al.*, 2021). Bispecific aptamer constructs can be applied as LYTACs by binding to protein-to-be-degraded and CI-M6PR, the same lysosome-shuttling protein used for the first developed LYTAC, simultaneously (Miao *et al.*, 2021). β1-integrin is continuously recycled within cells (Moreno-Layseca *et al.*, 2019). The observation that GreenB1 is trafficked to acidic vesicles for degradation after binding to the target protein implies that further research into the application of GreenB1 for LYTAC development is warranted.

GreenB1 for in vivo drug delivery

One of the most promising directions where to expand GreenB1 capabilities is the precision delivery of drugs and imaging agents using GreenB1 as a targeting moiety. β1-integrin positive TNBC and other β1-integrin positive solid tumours could be targeted using GreenB1. Role of

β1-integrin has been studied in pancreatic tumour growth (Mia *et al.*, 2021), β1-integrin inactivation shows inhibition in prostate cancer metastatic lesion formation (Lee *et al.*, 2013) and in breast cancer cells α3β1-integrin expression is correlated with invasiveness (Miskin *et al.*, 2021). Integrins have been previously extensively studied as proteins of interest for targeted drug delivery. Antibodies targeting αVβ3/β1/β5 or αV integrins previously have been tested with disappointing results in phase I/II clinical trials and one possible explanation for it could be that the high MW of antibodies makes it complicated to reach poorly vascularized tumour tissue (Bergonzini *et al.*, 2022). Peptide therapeutic agent CEND-1 targeting αVβ3/β5 integrins have noticeably smaller MW and shows promising results in preclinical trials for the treatment of pancreatic ductal adenocarcinoma (Hurtado de Mendoza *et al.*, 2021). CEND-1 has been tested in a first-in-human phase 1 study (ClinicalTrials.gov Identifier: NCT03517176) in combination with nab-paclitaxel and gemcitabine and shows a promising objective response rate (ORR) of 59% (95% CI 39-77) (Dean *et al.*, 2022) compared to nab-paclitaxel plus gemcitabine ORR of 23% (95% CI 19-27) from phase 3 study (Von Hoff *et al.*, 2013). Aptamers have several times smaller MW than antibodies and it would be a reasonable next step to explore GreenB1 capability to access difficult-to-reach tumour vasculature for payload delivery. GreenB1 can be conjugated to liposomes to both increase the circulation half-life of aptamer and deliver the anticancer payload (Simón-Gracia *et al.*, 2021) or to cytotoxic compounds through lysosome sensitive linker that is cleaved and release the drug upon the internalization within lysosomes (Tobi *et al.*, 2020). GreenB1 can be likely modified to achieve endolysosomal escape and thus adapted for payload delivery to the cytosol or other cellular compartments. This approach can be applied to deliver peptide or siRNA payloads with a charged and polar character that are unable to efficiently cross the plasma membrane and enter the cytosol where they would be able to exhibit their activity (Juliano, 2016; Dowdy, 2017).

GreenB1 inherent activity

Receptor tyrosine kinase c-Met replaces the α5-integrin in α5β1-integrin complex in case of glioblastoma resistance to VEGF antibody and breast cancer metastasis. c-Met/β1-integrin complex has a higher affinity to fibronectin than α5β1-integrin complex, allowing to maintain invasive oncologic processes (Jahangiri *et al.*, 2017). It has been shown that in the case of TNBC Talin-1 and β1-integrin interaction blocking can be a potential therapeutic target (Zhang *et al.*, 2022). β1-integrin is necessary to form vasculogenic mimicry (VM), a mechanism where

cancerous cells form blood vessel-like structures to satisfy tumour blood-supply needs (Kawahara, Niwa and Simizu, 2018). GreenB1 has an affinity to β1-integrin in the low nanomolar range and it can be fine-tuned by creating a multimeric aptamer construct. The potential of GreenB1 to disrupt the invasiveness driving interactions between β1-integrin and c-Met or Talin-1 is worth exploring in the future. Additionally, research into GreenB1 ability to target and interact with VM-forming cancerous cells is warranted.

CONCLUSIONS

1. Randomized ssDNA aptamer library can be effectively enriched with carcinoma cell line binding aptamer sequences.

2. Differential binding analysis takes full advantage of high-throughput sequencing data for aptamer identification.

3. Target selective aptamer sequences can be identified using high-throughput sequencing even from seemingly unsuccessful aptamer selection.

4. Aptamer target protein identification can be achieved using a proximity labelling approach and exhibits lower noise compared to extract-based pull-down methods.

5. Fluorescence polarization is a fast and easy method compared to electrophoretic mobility shift assay to measure aptamer binding affinity to target protein if the protein has a molecular weight that is several times bigger than the aptamer.

6. Imaging flow cytometry can be used to study aptamer binding, internalization and subcellular localization using a large number of cells.

7. While GreenB1 aptamer showed a lack of selectivity towards carcinoma cells *in vitro*, results from *in vivo* proved its homing to tumorous lesions. Target proteins can potentially be “hidden” for binding from aptamer in circulation in healthy organs but not in the tumour.

THESIS

1. Differential binding analysis can identify target selective aptamers using statistical analysis of high-throughput sequencing data that would not be identified using enrichment analysis.

2. Proximity labelling followed by mass spectrometry proteomics can identify aptamer target proteins and target-related proteins on the carcinoma cell surface.

3. Aptamer GreenB1 binds to β1-integrin and is internalized to acidic vesicles *in vitro*.

4. GreenB1 preferentially homes to β1-integrin positive triple-negative breast cancer lesions in mice.

PUBLICATIONS

1. **Pleiko, K.,** Saulite, L., Parfejevs, V., Miculis, K., Vjaters, E., & Riekstina, U. (2019). Differential binding cell-SELEX method to identify cell-specific aptamers using high-throughput sequencing. In Scientific Reports (Vol. 9, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-019-44654-w
2. Põšnograjeva, K.*, **Pleiko, K.*,** Haugas, M., & Teesalu, T. (2021). New Tools for Streamlined In Vivo Homing Peptide Identification. In Methods in Molecular Biology (pp. 385–412). Springer US. https://doi.org/10.1007/978-1-0716-1752-6_25 (*contributed equally to this work)
3. **Pleiko, K.,** Haugas, M., Parfejevs, V., Pantelejevs, T., Parisini, E., Teesalu, T., & Riekstina, U. (2022). Targeting triple-negative breast cancer with β1-integrin binding aptamer. *Available on biorxiv.org, submitted to Molecular Therapy – Nucleic Acids, revised version in preparation.* https://doi.org/10.1101/2022.07.28.501822

APPROBATION OF THE RESEARCH

1. **Pleiko K.** NGS for aptamer selection using differential binding cell-SELEX. Oral presentation. FEBS3+ conference of Latvian, Lithuanian and Estonian Biochemical Societies, June 17-19, 2019, Riga.
2. **Pleiko K.,** Parfejevs V., Krims-Davis K., Kunrade L., Jekabsons K., Riekstina U. Internalizing aptamers for the development of bi-functional aptamer constructs. Medicina (Kaunas) Nr.56, Suppl.1: Abstracts of the 78th International Scientific Conference of the University of Latvia Riga, Latvia (2020), p.7. , URL: https://medicina.lsmuni.lt/abstracts-of-the-international-scientific-conference-on-medicine-organized-within-the-frame-of-the-78th-international-scientific-conference-of-the-university-of-latvia/ ISSN 1648-9233.
3. Krims-Dāvis K., **Pleiko K.,** Riekstiņa U. Cell-SELEX-enriched aptamer selectivity screening with flow cell cytometry. Medicina Vol. 56, S.1: Abstracts accepted for the International Scientific Conference on Medicine organized within the frame of the 78th International Scientific Conference of the University of Latvia (2020), P.284. , URL: https://medicina.lsmuni.lt/issue/suppl-1-volume-56/ ISSN 1648-9233.
4. **Pleiko K.,** Teesalu T., Riekstina U. Target protein identification of MDA-MB-231 cell line selective aptamer. Oral presentation. Basic Medical Science & Pharmacy. Medicina Vol. 57, Suppl. 1: Abstracts of the 79th International Scientific Conference of the University of Latvia, Riga, Latvia (2021), p.4. , URL: https://medicina.lsmuni.lt/abstracts-accepted-for-the-international-scientific-conference-on-medicine-organized-within-the-frame-of-the-79th-international-scientific-conference-of-the-university-of-latvia-riga-latvia/ ISSN 1648-9233.
5. **Pleiko K.,** Teesalu T., Riekstina U. Proximity Ligation for *in Vitro* Aptamer Target Protein. Oral presentation. Medicina (Kaunas) Vol. 58, Suppl. 1: Abstracts of the 80th International Scientific Conference of the University of Latvia, Riga, Latvia (2022), p.15. , URL: https://medicina.lsmuni.lt/abstracts-of-the-international-scientific-conference-on-medicine-organized-within-the-frame-of-the-80th-international-scientific-conference-of-the-university-of-latvia/ ISSN 1648-9233.

ACKNOWLEDGEMENTS AND FUNDING

The research was supported by University of Latvia fundamental research grant “Research of biomarkers and natural substances for acute and chronic diseases’ diagnostics and personalized treatment” and ESF grant No. 8.2.2.0/18/I/006. I want to also express my gratitude to University of Latvia Foundation and SIA ‘Mikrotīkls’ for supporting my research activities with PhD research scholarship and through project “The application of novel aptamer technology for the selection of bladder cancer biomarkers”.

I would like to express my deepest gratitude to Professor Una Riekstina, who has supported me throughout all stages of this work, as well as from the very first day in the lab during my undergraduate studies.

I would also like to thank Professor Tambet Teesalu, who introduced me to innovative laboratory techniques.

A huge thank you to my colleagues - Liga, Vadim, Karina, Maarja, Kaspars - who over the years have helped with both small and big tasks in the lab, or just been there when I needed to talk.

Thanks to my partner Linda, without whom this work would be, if not completely impossible, certainly immeasurably more difficult. Thanks also to Muša and Ezis, who not only took me for walks, but also made sure that I kept writing.

Finally, and most importantly, I would like to thank my parents, who have not only always supported my pursuit of science, but have also made me think about my health, something I have not always remembered myself.

REFERENCES

Ahn, G. *et al.* (2021) 'LYTACs that engage the asialoglycoprotein receptor for targeted protein degradation', *Nature Chemical Biology*, 17(9), pp. 937–946. Available at: https://doi.org/10.1038/s41589-021-00770-1.

Alam, K.K., Chang, J.L. and Burke, D.H. (2015) 'FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections', *Molecular Therapy - Nucleic Acids*, 4(3), pp. 1–10. Available at: https://doi.org/10.1038/mtna.2015.4.

Bai, C. *et al.* (2018) 'Aptamer selection and application in multivalent binding-based electrical impedance detection of inactivated H1N1 virus', *Biosensors and Bioelectronics*, 110, pp. 162–167. Available at: https://doi.org/10.1016/j.bios.2018.03.047.

Banik, S.M. *et al.* (2020) 'Lysosome-targeting chimaeras for degradation of extracellular proteins', *Nature*, 584(7820), pp. 291–297. Available at: https://doi.org/10.1038/s41586-020-2545-9.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing', *J. Royal Statist. Soc., Series B*, 57, pp. 289–300.

Bergonzini, C. *et al.* (2022) 'Targeting Integrins for Cancer Therapy - Disappointments and Opportunities', *Frontiers in Cell and Developmental Biology*, 10, p. 479. Available at: https://doi.org/10.3389/fcell.2022.863850.

Bing, T., Shangguan, D. and Wang, Y. (2015) 'Facile Discovery of Cell-Surface Protein Targets of Cancer Cell Aptamers', *Molecular & Cellular Proteomics*, 14(10), pp. 2692–2700. Available at: https://doi.org/10.1074/mcp.M115.051243.

Bodary, S.C. and McLean, J.W. (1990) 'The integrin β1 subunit associates with the vitronectin receptor α(v) subunit to form a novel vitronectin receptor in a human embryonic kidney cell line', *Journal of Biological Chemistry*, 265(11), pp. 5938–5941. Available at: https://doi.org/10.1016/S0021-9258(19)39269-5.

Boltz, A. *et al.* (2011) 'Bi-specific Aptamers Mediating Tumor Cell Lysis', *Journal of Biological Chemistry*, 286(24), pp. 21896–21905. Available at: https://doi.org/10.1074/jbc.M111.238261.

Böttcher, R.T. *et al.* (2012) 'Sorting nexin 17 prevents lysosomal degradation of β1 integrins by binding to the β1-integrin tail', *Nature Cell Biology*, 14(6), pp. 584–592. Available at: https://doi.org/10.1038/ncb2501.

Bruno, J.G. (2018) 'Potential Inherent Stimulation of the Innate Immune System by Nucleic Acid Aptamers and Possible Corrective Approaches', *Pharmaceuticals*, 11(3), p. 62. Available at: https://doi.org/10.3390/ph11030062.

Burmeister, P.E. *et al.* (2005) 'Direct In Vitro Selection of a 2′-O-Methyl Aptamer to VEGF', *Chemistry & Biology*, 12(1), pp. 25–33. Available at: https://doi.org/10.1016/j.chembiol.2004.10.017.

Caroli, J. *et al.* (2016) 'APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data', *Bioinformatics*, 32(2), pp. 161–164. Available at: https://doi.org/10.1093/bioinformatics/btv545.

Caroli, J., Forcato, M. and Bicciato, S. (2020) 'APTANI2: update of aptamer selection through sequence-structure analysis', *Bioinformatics*, 36(7), pp. 2266–2268. Available at: https://doi.org/10.1093/bioinformatics/btz897.

Cheah, J.S. and Yamada, S. (2017) 'A simple elution strategy for biotinylated proteins bound to streptavidin conjugated beads using excess biotin and heat', *Biochemical and Biophysical Research Communications*, 493(4), pp. 1522–1527. Available at: https://doi.org/10.1016/j.bbrc.2017.09.168.

Chen, A. and Yang, S. (2015) 'Replacing antibodies with aptamers in lateral flow immunoassay', *Biosensors and Bioelectronics*, 71, pp. 230–242. Available at: https://doi.org/10.1016/j.bios.2015.04.041.

Cheng, C. *et al.* (2013) 'In vivo SELEX for identification of brain-penetrating aptamers', *Molecular Therapy - Nucleic Acids*, 2(June 2012), pp. 1–9. Available at: https://doi.org/10.1038/mtna.2012.59.

Cheung, Y.-W. *et al.* (2013) 'Structural basis for discriminatory recognition of Plasmodium lactate dehydrogenase by a DNA aptamer', *Proceedings of the National Academy of Sciences*, 110(40), pp. 15967–15972. Available at: https://doi.org/10.1073/pnas.1309538110.

Chiu, T.C. and Huang, C.C. (2009) *Aptamer-functionalized nano-biosensors*, *Sensors*. Available at: https://doi.org/10.3390/s91210356.

Constantin, A.-E. and Patil, I. (2021) '{ggsignif}: R Package for Displaying Significance Brackets for {'ggplot2'}', *PsyArxiv* [Preprint]. Available at: https://doi.org/10.31234/osf.io/7awm6.

Curreri, A. *et al.* (2022) 'RNA therapeutics in the clinic', *Bioengineering & Translational Medicine*, n/a(n/a), p. e10374. Available at: https://doi.org/10.1002/btm2.10374.

Daniels, D.A. *et al.* (2003) 'A tenascin-C aptamer identified by tumor cell SELEX: Systematic evolution of ligands by exponential enrichment', *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), pp. 15416–15421. Available at: https://doi.org/10.1073/pnas.2136683100.

Dassie, J.P. *et al.* (2009) 'Systemic administration of optimized aptamer-siRNA chimeras promotes regression of PSMA-expressing tumors', *Nature Biotechnology*, 27(9), pp. 839–846. Available at: https://doi.org/10.1038/nbt.1560.

Dean, A. *et al.* (2022) 'Dual αV-integrin and neuropilin-1 targeting peptide CEND-1 plus nab-paclitaxel and gemcitabine for the treatment of metastatic pancreatic ductal adenocarcinoma: a first-in-human, open-label, multicentre, phase 1 study', *The Lancet Gastroenterology & Hepatology*, 7(10), pp. 943–951. Available at: https://doi.org/10.1016/S2468-1253(22)00167-4.

Dembowski, S.K. and Bowser, M.T. (2017) 'Microfluidic methods for aptamer selection and characterization', *Analyst*, 143(1), pp. 21–32. Available at: https://doi.org/10.1039/C7AN01046J.

Deutsch, E.W. *et al.* (2017) 'The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition', *Nucleic Acids Research*, 45(D1), pp. D1100–D1106. Available at: https://doi.org/10.1093/nar/gkw936.

Dowdy, S.F. (2017) 'Overcoming cellular barriers for RNA therapeutics', *Nature Biotechnology 2017 35:3*, 35(3), pp. 222–229. Available at: https://doi.org/10.1038/nbt.3802.

Dua, P. *et al.* (2018) 'Cell-SELEX-Based Identification of a Human and Mouse Cross-Reactive Endothelial Cell-Internalizing Aptamer', *Nucleic Acid Therapeutics*, 00(00), p. nat.2017.0711. Available at: https://doi.org/10.1089/nat.2017.0711.

Eaton, B.E., Gold, L. and Zichi, D.A. (1995) 'Let's get specific: the relationship between specificity and affinity', *Chemistry & Biology*, 2(10), pp. 633–638. Available at: https://doi.org/10.1016/1074-5521(95)90023-3.

Ellington, A.D. and Szostak, J.W. (1990) 'In vitro selection of RNA molecules that bind specific ligands', *Nature*, 346, p. 818.

EMEA (2006) *Macugen: Scientific Discussion*. EMEA, p. 39. Available at: https://www.ema.europa.eu/en/documents/scientific-discussion/macugen-epar-scientific-discussion_en.pdf (Accessed: 23 September 2022).

Farokhzad, O.C. *et al.* (2006) 'Targeted nanoparticle-aptamer bioconjugates for cancer chemotherapy in vivo', *Proceedings of the National Academy of Sciences*, 103(16), pp. 6315–6320. Available at: https://doi.org/10.1073/pnas.0601755103.

Filonov, G.S. *et al.* (2015) 'In-Gel Imaging of RNA Processing Using Broccoli Reveals Optimal Aptamer Expression Strategies Article In-Gel Imaging of RNA Processing Using Broccoli Reveals Optimal Aptamer Expression Strategies', *Chemistry & Biology*, 22(5), pp. 649–660. Available at: https://doi.org/10.1016/j.chembiol.2015.04.018.

Fornace, M.E., Porubsky, N.J. and Pierce, N.A. (2020) 'A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed', *ACS Synthetic Biology*, 9(10), pp. 2665–2678. Available at: https://doi.org/10.1021/acssynbio.9b00523.

Gao, H. *et al.* (2012) 'Precise glioma targeting of and penetration by aptamer and peptide dual-functioned nanoparticles', *Biomaterials*, 33(20), pp. 5115–5123. Available at: https://doi.org/10.1016/j.biomaterials.2012.03.058.

Garber, K. (2022) 'The PROTAC gold rush', *Nature Biotechnology*, 40(1), pp. 12–16. Available at: https://doi.org/10.1038/s41587-021-01173-2.

Gelinas, A.D. (2016) 'Embracing proteins: structural themes in aptamer–protein complexes', *Current Opinion in Structural Biology*, p. 11.

Gilboa-Geffen, A. *et al.* (2015) 'Gene knockdown by EpCAM aptamer-siRNA chimeras suppresses epithelial breast cancers and their tumor-initiating cells', *Molecular Cancer Therapeutics*, 14(10), pp. 2279–2291. Available at: https://doi.org/10.1158/1535-7163.MCT-15-0201-T.

Giordano, F.A. *et al.* (2022) 'Radiotherapy and olaptesed pegol (NOX-A12) in partially resected or biopsy-only MGMT-unmethylated glioblastoma: Interim data from the German multicenter phase 1/2 GLORIA trial.', *Journal of Clinical Oncology* [Preprint]. Available at: https://doi.org/10.1200/JCO.2022.40.16_suppl.2050.

Haßel, S.K. and Mayer, G. (2019) 'Aptamers as Therapeutic Agents: Has the Initial Euphoria Subsided?', *Molecular Diagnosis & Therapy*, 23(3), pp. 301–309. Available at: https://doi.org/10.1007/s40291-019-00400-6.

Healy, J.M. *et al.* (2004) 'Pharmacokinetics and Biodistribution of Novel Aptamer Compositions', *Pharmaceutical Research*, 21(12), pp. 2234–2246. Available at: https://doi.org/10.1007/s11095-004-7676-4.

Hentze, M.W. *et al.* (2018) 'A brave new world of RNA-binding proteins', *Nature Reviews Molecular Cell Biology*, 19(5), pp. 327–341. Available at: https://doi.org/10.1038/nrm.2017.130.

Heo, K. *et al.* (2016) 'An aptamer-antibody complex (oligobody) as a novel delivery platform for targeted cancer therapies', *Journal of Controlled Release*, 229, pp. 1–9. Available at: https://doi.org/10.1016/j.jconrel.2016.03.006.

Hermann, T. and Patel, D.J. (2000) 'Adaptive Recognition by Nucleic Acid Aptamers', *Science*, 287(5454), pp. 820–825. Available at: https://doi.org/10.1126/science.287.5454.820.

Hernández-Jiménez, M. *et al.* (2022) 'First-in-human phase I clinical trial of a TLR4-binding DNA aptamer, ApTOLL: Safety and pharmacokinetics in healthy volunteers', *Molecular Therapy - Nucleic Acids*, 28, pp. 124–135. Available at: https://doi.org/10.1016/J.OMTN.2022.03.005.

Hicke, B.J. *et al.* (2001) 'Tenascin-C Aptamers Are Generated Using Tumor Cells and Purified Protein', *Journal of Biological Chemistry*, 276(52), pp. 48644–48654. Available at: https://doi.org/10.1074/jbc.M104651200.

Hoinka, J. *et al.* (2015) 'Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery', *Nucleic Acids Research*, 43(12), pp. 5699–5707. Available at: https://doi.org/10.1093/nar/gkv308.

Hurtado de Mendoza, T. *et al.* (2021) 'Tumor-penetrating therapy for β5 integrin-rich pancreas cancer', *Nature Communications*, 12(1), p. 1541. Available at: https://doi.org/10.1038/s41467-021-21858-1.

Illumina (2013) *16S Metagenomic Sequencing Library Preparation*, *Illumina.com*, pp. 1–28.

Jahangiri, A. *et al.* (2017) 'Cross-activating c-Met/β1 integrin complex drives metastasis and invasive resistance in cancer', *Proceedings of the National Academy of Sciences of the United*

States of America, 114(41), pp. E8685–E8694. Available at: https://doi.org/10.1073/pnas.1701821114.

Javier, D.J. *et al.* (2008) 'Aptamer-Targeted Gold Nanoparticles As Molecular-Specific Contrast Agents for Reflectance Imaging', *Bioconjugate Chemistry*, 19(6), pp. 1309–1312. Available at: https://doi.org/10.1021/bc8001248.

Jiang, W. *et al.* (2017) 'Mirror-image polymerase chain reaction', *Cell Discovery*, 3, pp. 1–7. Available at: https://doi.org/10.1038/celldisc.2017.37.

Jolma, A. *et al.* (2010) 'Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities', pp. 861–873. Available at: https://doi.org/10.1101/gr.100552.109.

Jonasch, E., Walker, C.L. and Rathmell, W.K. (2021) 'Clear cell renal cell carcinoma ontogeny and mechanisms of lethality', *Nature Reviews Nephrology*, 17(4), pp. 245–261. Available at: https://doi.org/10.1038/s41581-020-00359-2.

Juliano, R.L. (2016) 'The delivery of therapeutic oligonucleotides', *Nucleic Acids Research*, 44(14), pp. 6518–6548. Available at: https://doi.org/10.1093/nar/gkw236.

Jung, J.K. *et al.* (2020) 'Cell-free biosensors for rapid detection of water contaminants', *Nature Biotechnology* [Preprint]. Available at: https://doi.org/10.1038/s41587-020-0571-7.

Kawahara, R., Niwa, Y. and Simizu, S. (2018) 'Integrin β1 is an essential factor in vasculogenic mimicry of human cancer cells', *Cancer Science*, 109(8), pp. 2490–2496. Available at: https://doi.org/10.1111/cas.13693.

Keefe, A.D., Pai, S. and Ellington, A. (2010) 'Aptamers as therapeutics', *Nature Reviews Drug Discovery*, 9(7), pp. 537–550. Available at: https://doi.org/10.1038/nrd3141.

Kelly, L. *et al.* (2021) 'A comparative analysis of cell surface targeting aptamers', *Nature Communications*, 12(1). Available at: https://doi.org/10.1038/s41467-021-26463-w.

Khedri, M. *et al.* (2015) 'Cancer immunotherapy via nucleic acid aptamers', *International Immunopharmacology*, 29(2), pp. 926–936. Available at: https://doi.org/10.1016/j.intimp.2015.10.013.

Kim, M.W. *et al.* (2019) 'Anti-EGF receptor aptamer-guided co-delivery of anti-cancer siRNAs and quantum dots for theranostics of triple-negative breast cancer', *Theranostics*, 9(3), pp. 837–852. Available at: https://doi.org/10.7150/thno.30228.

Kramer, S.T. *et al.* (2022) 'FASTAptameR 2.0: A web tool for combinatorial sequence selections', *Molecular Therapy - Nucleic Acids*, 29, pp. 862–870. Available at: https://doi.org/10.1016/j.omtn.2022.08.030.

Krieg, A.M. (1999) 'Mechanisms and applications of immune stimulatory CpG oligodeoxynucleotides', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1489(1), pp. 107–116. Available at: https://doi.org/10.1016/S0167-4781(99)00147-5.

Kulkarni, O. *et al.* (2007) 'Spiegelmer Inhibition of CCL2/MCP-1 Ameliorates Lupus Nephritis in MRL-(Fas)lpr Mice', *Journal of the American Society of Nephrology*, 18(8), pp. 2350–2358. Available at: https://doi.org/10.1681/ASN.2006121348.

Kuwahara, M. and Obika, S. (2013) 'In vitro selection of BNA (LNA) aptamers', *Artificial DNA: PNA & XNA*, 4(2), pp. 39–48. Available at: https://doi.org/10.4161/adna.25786.

Lee, C.H. *et al.* (2015) 'Pharmacokinetics of a Cholesterol-conjugated Aptamer Against the Hepatitis C Virus (HCV) NS5B Protein', *Molecular Therapy. Nucleic Acids*, 4(10), p. e254. Available at: https://doi.org/10.1038/mtna.2015.30.

Lee, Y.-C. *et al.* (2013) 'Targeting Constitutively Activated β1 Integrins Inhibits Prostate Cancer Metastasis', *Molecular Cancer Research*, 11(4), pp. 405–417. Available at: https://doi.org/10.1158/1541-7786.MCR-12-0551.

Lin, Y. *et al.* (1994) 'Modified RNA sequence pools for in vitro selection', *Nucleic Acids Research*, 22(24), pp. 5229–5234. Available at: https://doi.org/10.1093/nar/22.24.5229.

Litke, J.L. and Jaffrey, S.R. (2019) 'Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts', *Nature Biotechnology* [Preprint]. Available at: https://doi.org/10.1038/s41587-019-0090-6.

Liu, F. *et al.* (2014) 'Application of ZnO/graphene and S6 aptamers for sensitive photoelectrochemical detection of SK-BR-3 breast cancer cells based on a disposable indium tin oxide device', *Biosensors and Bioelectronics*, 51, pp. 413–420. Available at: https://doi.org/10.1016/j.bios.2013.07.066.

Lorenz, R. *et al.* (2011) 'ViennaRNA Package 2.0', *Algorithms for Molecular Biology*, 6(1), p. 26. Available at: https://doi.org/10.1186/1748-7188-6-26.

Lyu, C., Khan, I.M. and Wang, Z. (2021) 'Capture-SELEX for aptamer selection: A short review', *Talanta*, 229, p. 122274. Available at: https://doi.org/10.1016/j.talanta.2021.122274.

Maasch, C. *et al.* (2008) 'Physicochemical Stability of NOX-E36, a 40mer L-RNA (Spiegelmer) for Therapeutic Applications', *Nucleic Acids Symposium Series*, 52(1), pp. 61–62. Available at: https://doi.org/10.1093/nass/nrn031.

Mallikaratchy, P.R. *et al.* (2011) 'A multivalent DNA aptamer specific for the B-cell receptor on human lymphoma and leukemia', *Nucleic Acids Research*, 39(6), pp. 2458–2469. Available at: https://doi.org/10.1093/NAR/GKQ996.

Mann, A.P. *et al.* (2011) 'Thioaptamer Conjugated Liposomes for Tumor Vasculature Targeting', *Oncotarget*, 2(4), pp. 298–304. Available at: https://doi.org/10.18632/oncotarget.261.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), p. 10. Available at: https://doi.org/10.14806/ej.17.1.200.

McNamara, J.O. *et al.* (2006) 'Cell type–specific delivery of siRNAs with aptamer-siRNA chimeras', *Nature Biotechnology*, 24(8), pp. 1005–1015. Available at: https://doi.org/10.1038/nbt1223.

Mi, J. *et al.* (2010) 'In vivo selection of tumor-targeting RNA motifs', *Nature Chemical Biology*, 6(1), pp. 22–24. Available at: https://doi.org/10.1038/nchembio.277.

Mia, M.S. *et al.* (2021) 'Integrin β1 Promotes Pancreatic Tumor Growth by Upregulating Kindlin-2 and TGF-β Receptor-2', *International Journal of Molecular Sciences*, 22(19), p. 10599. Available at: https://doi.org/10.3390/ijms221910599.

Miao, Y. *et al.* (2021) 'Bispecific Aptamer Chimeras Enable Targeted Protein Degradation on Cell Membranes', *Angewandte Chemie - International Edition*, 60(20), pp. 11267–11271. Available at: https://doi.org/10.1002/anie.202102170.

Miskin, R.P. *et al.* (2021) 'Integrin α3β1 promotes invasive and metastatic properties of breast cancer cells through induction of the brn-2 transcription factor', *Cancers*, 13(3), pp. 1–17. Available at: https://doi.org/10.3390/cancers13030480.

Moreno-Layseca, P. *et al.* (2019) 'Integrin trafficking in cells and tissues', *Nature Cell Biology*, 21(2), pp. 122–132. Available at: https://doi.org/10.1038/s41556-018-0223-z.

Ni, S. *et al.* (2021) 'Recent Progress in Aptamer Discoveries and Modifications for Therapeutic Applications', *ACS Applied Materials & Interfaces*, 13(8), pp. 9500–9519. Available at: https://doi.org/10.1021/acsami.0c05750.

Ozer, A., Pagano, J.M. and Lis, J.T. (2014) 'New technologies provide quantum changes in the scale, speed, and success of SELEX methods and aptamer characterization', *Molecular Therapy - Nucleic Acids*, 3(June), pp. 1–18. Available at: https://doi.org/10.1038/mtna.2014.34.

Paige, J.S. *et al.* (2012) 'Fluorescence Imaging of Cellular Metabolites with RNA', *Science*, 335(6073), pp. 1194–1194. Available at: https://doi.org/10.1126/science.1218298.

Paul, N.R., Jacquemet, G. and Caswell, P.T. (2015) 'Endocytic Trafficking of Integrins in Cell Migration', *Current Biology*, 25(22), pp. R1092–R1105. Available at: https://doi.org/10.1016/j.cub.2015.09.049.

Pech, A. *et al.* (2017) 'A thermostable D-polymerase for mirror-image PCR', *Nucleic Acids Research*, 45(7), pp. 3997–4005. Available at: https://doi.org/10.1093/nar/gkx079.

Perez-Riverol, Y. *et al.* (2021) 'The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences', *Nucleic Acids Research*, 50. Available at: https://doi.org/10.1093/nar/gkab1038.

Pleiko, K. *et al.* (2021) 'In vivo phage display: identification of organ-specific peptides using deep sequencing and differential profiling across tissues', *Nucleic Acids Research*, pp. 1–11. Available at: https://doi.org/10.1093/nar/gkaa1279.

Porciani, D. *et al.* (2018) 'Modular cell-internalizing aptamer nanostructure enables targeted delivery of large functional RNAs in cancer cell lines', *Nature Communications*, 9(1). Available at: https://doi.org/10.1038/s41467-018-04691-x.

Põšnograjeva, K. *et al.* (2022) 'New Tools for Streamlined In Vivo Homing Peptide Identification', in Ü. Langel (ed.) *Cell Penetrating Peptides: Methods and Protocols*. New

York, NY: Springer US (Methods in Molecular Biology), pp. 385–412. Available at: https://doi.org/10.1007/978-1-0716-1752-6_25.

Qin, W. *et al.* (2021) 'Deciphering molecular interactions by proximity labeling', *Nature Methods*, 18(February). Available at: https://doi.org/10.1038/s41592-020-01010-5.

R Core Team (2016) 'R Development Core Team', *R: A Language and Environment for Statistical Computing*, 55, pp. 275–286. Available at: http://www.R-project.org.

Riopel, M.M. *et al.* (2013) 'β1 integrin–extracellular matrix interactions are essential for maintaining exocrine pancreas architecture and function', *Laboratory Investigation*, 93(1), pp. 31–40. Available at: https://doi.org/10.1038/labinvest.2012.147.

Ritchie, M.E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), pp. e47–e47. Available at: https://doi.org/10.1093/nar/gkv007.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) '{edgeR}: a {Bioconductor} package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26(1), pp. 139–140. Available at: https://doi.org/10.1093/bioinformatics/btp616.

Rogers, K.A. *et al.* (2018) 'Phase 1b study of obinutuzumab, ibrutinib, and venetoclax in relapsed and refractory chronic lymphocytic leukemia', *Blood*, 132(15), pp. 1568–1572. Available at: https://doi.org/10.1182/blood-2018-05-853564.

Rosch, J.C. *et al.* (2020) 'CRISPR-Mediated Isogenic Cell-SELEX Approach for Generating Highly Specific Aptamers Against Native Membrane Proteins', *Cellular and Molecular Bioengineering*, 13(5), pp. 559–574. Available at: https://doi.org/10.1007/s12195-020-00651-y.

RStudio Team (2020) 'RStudio: Integrated Development Environment for R'. Boston, MA: RStudio, PBC. Available at: http://www.rstudio.com/.

Ruckman, J. *et al.* (1998) '2′-Fluoropyrimidine RNA-based Aptamers to the 165-Amino Acid Form of Vascular Endothelial Growth Factor (VEGF165): INHIBITION OF RECEPTOR BINDING AND VEGF-INDUCED VASCULAR PERMEABILITY THROUGH INTERACTIONS REQUIRING THE EXON 7-ENCODED DOMAIN*', *Journal of Biological Chemistry*, 273(32), pp. 20556–20567. Available at: https://doi.org/10.1074/jbc.273.32.20556.

Rusconi, C.P. *et al.* (2004) 'Antidote-mediated control of an anticoagulant aptamer in vivo', *Nature Biotechnology*, 22(11), pp. 1423–1428. Available at: https://doi.org/10.1038/nbt1023.

Sanford, A.A. *et al.* (2021) 'RE-SELEX: restriction enzyme-based evolution of structure-switching aptamer biosensors', *Chemical Science*, 12(35), pp. 11692–11702. Available at: https://doi.org/10.1039/D1SC02715H.

Sarkans, U. *et al.* (2018) 'The BioStudies database-one stop shop for all data supporting a life sciences study', *Nucleic Acids Research*, 46(D1), pp. D1266–D1270. Available at: https://doi.org/10.1093/nar/gkx965.

Sefah, K. *et al.* (2010) 'Development of DNA aptamers using cell-selex', *Nature Protocols*, 5(6), pp. 1169–1185. Available at: https://doi.org/10.1038/nprot.2010.66.

Shangguan, D. *et al.* (2007) 'Aptamers Evolved from Cultured Cancer Cells Reveal Molecular Differences of Cancer Cells in Patient Samples', *Clinical Chemistry*, 53(6), pp. 1153–1155. Available at: https://doi.org/10.1373/clinchem.2006.083246.

Simón-Gracia, L. *et al.* (2021) 'Novel Anthracycline Utorubicin for Cancer Therapy', *Angewandte Chemie International Edition*, pp. 2–12. Available at: https://doi.org/10.1002/anie.202016421.

Slowikowski, K. (2021) 'ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2"'. Available at: https://cran.r-project.org/package=ggrepel.

Steffensmeier, A.C.G. *et al.* (2007) 'Vitreous Injections of Pegaptanib Sodium Triggering Allergic Reactions', *American Journal of Ophthalmology*, 143(3), pp. 512–513. Available at: https://doi.org/10.1016/j.ajo.2006.10.007.

Steurer, M. *et al.* (2019) 'Olaptesed pegol (NOX-A12) with bendamustine and rituximab: a phase IIa study in patients with relapsed/refractory chronic lymphocytic leukemia', *Haematologica*, 104(10), pp. 2053–2060. Available at: https://doi.org/10.3324/haematol.2018.205930.

Strauss, S. *et al.* (2018) 'Modified aptamers enable quantitative sub-10-nm cellular DNA-PAINT imaging', *Nature Methods*, 15(9), pp. 685–688. Available at: https://doi.org/10.1038/s41592-018-0105-0.

Subbaram, S. and Dipersio, C.M. (2011) 'Integrin α3β1 as a breast cancer target', *Expert Opinion on Therapeutic Targets*, 15(10), pp. 1197–1210. Available at: https://doi.org/10.1517/14728222.2011.609557.

Sullivan, R. *et al.* (2019) 'Analyzing Secondary Structure Patterns in DNA Aptamers Identified via CompELS', *Molecules*, 24(8), p. 1572. Available at: https://doi.org/10.3390/molecules24081572.

Sunbul, M. *et al.* (2021) 'Super-resolution RNA imaging using a rhodamine-binding aptamer with fast exchange kinetics', *Nature Biotechnology*, 39(June). Available at: https://doi.org/10.1038/s41587-020-00794-3.

Swayze, E.E. *et al.* (2007) 'Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals', *Nucleic Acids Research*, 35(2), pp. 687–700. Available at: https://doi.org/10.1093/nar/gkl1071.

Takahashi, M., Sakota, E. and Nakamura, Y. (2016) 'The efficient cell-SELEX strategy, Icell-SELEX, using isogenic cell lines for selection and counter-selection to generate RNA aptamers to cell surface proteins', *Biochimie*, 131, pp. 77–84. Available at: https://doi.org/10.1016/j.biochi.2016.09.018.

Tao, K. *et al.* (2008) 'Imagable 4T1 model for the study of late stage breast cancer', *BMC Cancer*, 8(1), p. 228. Available at: https://doi.org/10.1186/1471-2407-8-228.

Tao, W. *et al.* (2016) 'Polydopamine-Based Surface Modification of Novel Nanoparticle-Aptamer Bioconjugates for *In Vivo* Breast Cancer Targeting and Enhanced Therapeutic Effects', *Theranostics*, 6(4), pp. 470–484. Available at: https://doi.org/10.7150/thno.14184.

Thomas, B.J., Porciani, D. and Burke, D.H. (2022) 'Cancer immunomodulation using bispecific aptamers', *Molecular Therapy - Nucleic Acids* [Preprint]. Available at: https://doi.org/10.1016/j.omtn.2022.01.008.

Tobi, A. *et al.* (2020) 'Silver Nanocarriers Targeted with a CendR Peptide Potentiate the Cytotoxic Activity of an Anticancer Drug', *Advanced Therapeutics*, 2000097, p. 2000097. Available at: https://doi.org/10.1002/adtp.202000097.

Trapnell, C. *et al.* (2012) 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nature Protocols*, 7(3), pp. 562–578. Available at: https://doi.org/10.1038/nprot.2012.016.

Tuerk, C. and Gold, L. (1990) 'Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase', *Science*, 249(4968), pp. 505–510. Available at: https://doi.org/10.1126/SCIENCE.2200121.

Vaught, J.D. *et al.* (2010) 'Expanding the Chemistry of DNA for in Vitro Selection', *Journal of the American Chemical Society*, 132(12), pp. 4141–4151. Available at: https://doi.org/10.1021/ja908035g.

Von Hoff, D.D. *et al.* (2013) 'Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine', *New England Journal of Medicine*, 369(18), pp. 1691–1703. Available at: https://doi.org/10.1056/NEJMoa1304369.

Waldman, A.D., Fritz, J.M. and Lenardo, M.J. (2020) 'A guide to cancer immunotherapy: from T cell basic science to clinical practice', *Nature Reviews Immunology*, 20(11), pp. 651–668. Available at: https://doi.org/10.1038/s41577-020-0306-5.

Wan, Q. *et al.* (2022) 'Aptamer Targets Triple-Negative Breast Cancer through Specific Binding to Surface CD49c', *Cancers*, 14(6), p. 1570. Available at: https://doi.org/10.3390/cancers14061570.

Weiss, Z. and DasGupta, S. (2022) 'REVERSE: a user-friendly web server for analyzing next-generation sequencing data from *in vitro* selection/evolution experiments', *Nucleic Acids Research*, 50(W1), pp. W639–W650. Available at: https://doi.org/10.1093/nar/gkac508.

White, R. *et al.* (2001) 'Generation of Species Cross-reactive Aptamers Using "Toggle" SELEX', *Molecular Therapy*, 4(6), pp. 567–573. Available at: https://doi.org/10.1006/mthe.2001.0495.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Available at: https://ggplot2.tidyverse.org.

Wickham, H. and Bryan, J. (2022) 'readxl: Read Excel Files'.

Yang, J. and Bowser, M.T. (2013) 'Capillary Electrophoresis–SELEX Selection of Catalytic DNA Aptamers for a Small-Molecule Porphyrin Target', *Analytical Chemistry*, 85(3), pp. 1525–1530. Available at: https://doi.org/10.1021/ac302721j.

Zadeh, J.N. *et al.* (2011) 'NUPACK: Analysis and design of nucleic acid systems', *Journal of Computational Chemistry*, 32(1), pp. 170–173. Available at: https://doi.org/10.1002/jcc.21596.

Zanjani, L.S. *et al.* (2018) 'Increased expression of CD44 is associated with more aggressive behavior in clear cell renal cell carcinoma', *Biomarkers in Medicine*, 12(1), pp. 45–61. Available at: https://doi.org/10.2217/bmm-2017-0142.

Zeng, Y. *et al.* (2017) 'Salinomycin-loaded lipid-polymer nanoparticles with anti-CD20 aptamers selectively suppress human CD20 + melanoma stem cells', *Nature Publishing Group*, 39(2), pp. 261–274. Available at: https://doi.org/10.1038/aps.2017.166.

Zhang, Y. *et al.* (2022) 'Binding blockade between TLN1 and integrin β1 represses triple-negative breast cancer', *eLife*, 11, pp. 1–21. Available at: https://doi.org/10.7554/eLife.68481.

Zhao, B. *et al.* (2015) 'Designing activatable aptamer probes for simultaneous detection of multiple tumor-related proteins in living cancer cells', *Biosensors and Bioelectronics*, 68, pp. 763–770. Available at: https://doi.org/10.1016/j.bios.2015.02.004.

Zhou, J. and Rossi, J. (2017) 'Aptamers as targeted therapeutics: Current potential and challenges', *Nature Reviews Drug Discovery*, 16(3), pp. 181–202. Available at: https://doi.org/10.1038/nrd.2016.199.

Zuker, M. (2003) 'Mfold web server for nucleic acid folding and hybridization prediction', *Nucleic Acids Research*, 31(13), pp. 3406–3415. Available at: https://doi.org/10.1093/nar/gkg595.