The development of autoantibody diagnostics in renal cell carcinoma

Maria T. Tsoumpeli

Student ID: 4317158

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, May 2021.

Supervisors: Prof Kevin Gough, Dr Ben Maddison, Dr Nigel Mongan, Dr Cinzia Allegrucci, Prof Richard Emes.

Abstract

Renal Cell Carcinoma (RCC) is one of the most commonly diagnosed cancers worldwide and it lacks an early diagnostic biomarker. Here, the autoantibody responses against tumour-associated antigens were mapped to identify mimotopes that could be used in serological assays. An epitope mapping technique was applied that coupled the diagnostic potential of phage display peptide libraries, with the analytical depth of Next Generation Sequencing, a process called Next Generation Phage Display (NGPD). Firstly, a synthetic 16mer peptide phage library was constructed (diversity 5 x 10^9). This was used in the optimisation of an epitope mapping approach using known mAbs in solution/sera. The method consistently identified panels of mAb specific peptides, confirmed by phage-peptide ELISA. In addition, the NGS frequencies of the enriched peptides could be summed (an immunosignature) and used to distinguish between spiked and non-spiked sera samples. This immunosignature method was at least as sensitive and specific as ELISAs using phage-peptide mimotopes. The optimised NGPD approach was applied to sera collected from a mouse model for early stage RCC (TRACK model). In total, 14 (training set) and 29 (testing set) samples from TRACK or Wild Type (WT) mice were assessed. Enriched peptides identified by NGS (n=42) were screened against TRACK and WT sera by ELISA, and particularly one of them showed 72% sensitivity and 85.7% specificity, when sera from 12 month or older mice were used. The immunosignature approach was also applied and was 76.5% sensitive and 100% specific, when applied to sera from 12 month or older

TRACK mice. The strategy was then applied to sera derived from cc-RCC (n=100) or healthy (n=50) patients using same phage-peptide sub-library enriched against RCC–specific autoantibodies in TRACK mice. An immunosignature assay demonstrated 85% sensitivity and 90% specificity in the training cohort used to develop the assay, and 30% sensitivity and 100% specificity in a test cohort. Taken all the data together, mapping host's autoimmune response against RCC by NGPD could be a novel and effective strategy to developing serological based diagnostics to cc-RCC and other diseases that may trigger an autoantibody response.

Acknowledgements

First of all, I am deeply grateful to both of my supervisors Prof Kevin Gough and Dr Ben Maddison for their continuous support throughout my PhD. I am incredibly privileged to have had their advice and knowledge available when needed and it has been a true honour working with them.

Special thanks to Dr Jon Owen and his tremendous support and patience with me, especially during NGS analysis. I would also like to thank Dr Nigel Mongan for setting up the mouse model collaboration, Weill Cornell Medical College for providing me with TRACK sera, and Leeds multiNHR biobank for providing me with RCC and healthy samples. Furthermore, thanks to RSK-ADAS and University of Nottingham for my funding.

I would also like to thank my lab mates Juan, Brittany and Morena for making this experience enjoyable. Moreover, my partner Ashley will always have my sincere gratitude for his overwhelming support and patience throughout this journey.

Last but not least, I would like to thank K. Without her, none of this would have been possible.

‘’You can, you should and if you are brave enough, you will’’- Stephen King

Table of Contents

List of tables

List of figures

Abbreviations

NH_4Cl	Ammonium chloride
A	Alanine
AA	Amino acids
AAbs	Autoantibodies
Abs	Antibodies
APCs	Antigen Presenting Cells
APEx	Anchored Periplasmic Expression
BCR	B cell receptor
Bp	Base pairs
BR	Broad range
BSA	Bovine Serum Albumin
C	Cysteine
CFU	Colony Forming Unit
CO_2	Carbon dioxide
CT	Computed Tomography
D	Aspartic acid
DEPC	Diethyl pyrocarbonate
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
E	Glutamic acid
E. coli	Escherichia Coli
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
F	Phenylalanine
Fab	Fragment antigen-binding
FACS	Fluorescence activated cell sorting
Fc domain	Fragment crystallisable region domain
G	Glycine
H	Histidine
HCL	Hydrochloric acid
HIF-a	Hypoxia Induced factor α
HPLC	High-performance liquid chromatography
I	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Lysine
KH_2PO_4	Monopotassium phosphate
L	Leucine
LOD	Limit of Detection
M	Methionine
mAb	Monoclonal antibody
$MgCl_2$	Magnesium chloride

MHC	Major histocompatibility complex
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
N	Asparagine
$Na_2HPO_4.7H_2O$	Disodium hydrogen phosphate heptahydrate
NaCl	Sodium Chloride
Nap	Non-animal protein
NGPD	Next Generation Phage Display
NGS	Next Generation Sequencing
NMS	Normal Mouse Serum
OD	Optical density
P	Proline
p	Opal stop codon
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PTMs	Post Translational Modifications
Q	Glutamine
q	Amber stop codon
QC	Quality control
qPCR	Quantitative PCR
R	Arginine
RCC	Renal Cell Carcinoma
RNA	Ribonucleic acid
RT	Room temperature
S	Serine
scFv	Single chain variable region fragment
SDM	Site Directed Mutagenesis
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
ssDNA	Single-stranded DNA
T	Threonine
TAAs	Tumour Associated Antigens
TBS	Tris-Buffered Saline
TMB	3,3',5,5'-Tetramethylbenzidine
TRACK	Transgenic model of cancer of the kidney
TRIS	Trisaminomethane
V	Valine
VHL	Von Hippel-Lindau tumour suppressor
W	Tryptophan
WB	Western Blot
WHO	World Health Organisation
WT	Wild Type
Y	Tyrosine

1. Introduction

1.1. Cancer

Cancers are a group of diseases that involve mainly the abnormal growth and differentiation of cells and their spread across tissues after evading immunosurveillance (American Cancer Society, 2018; Postel et al., 2018). All people would develop cancer if they lived long enough, and as well as age, lifestyle can play a dramatic role in the development of a malignancy (Lichtenstein, 2017).

Worldwide, cancer is one of the leading causes of death. A recent study estimated the total global cancer burden to be 18 million new cases and almost 10 million cancer related deaths per year (Bray et al., 2018). 1 out of 8 men and 1 out of 10 women will develop this disease in their lifetime with men having a 50% higher mortality rate than women. Lung cancer is the most common disease (18.4% of diagnosed patients), followed by breast cancer (11.6%). However, there is high variability in the reported incidences globally which is likely due to not all countries having an established, standardised, and accurate reporting system. Cancer ranks as either the first or second-leading cause of death in almost half of the countries, whereas in 12% of them it ranks as the third or fourth-leading cause (Bray et al., 2018). In the USA, cancer is the second leading cause of death, only surpassed by heart disease. Around 2 million new cancer cases and almost 600,000 new deaths related to cancer are reported annually in this country, which accounts for 20% of the total deaths. The peak of cancer related deaths in the USA was 30 years ago, and since then

there has been a relative decline of almost 30% (Siegel et al., 2020). However, worldwide total cancer cases have been predicted to continue rising over the next 20 years (Seretis et al., 2019). Indeed, the total number of diagnosed cancer cases has been steadily increasing due to various reasons, such as the global increase of the human population as well as the westernisation of lifestyle in developing countries. Of note is that Europe accounts for less than 10% of the global population, yet its cancer rate and morbidity accounts for 23% of the total estimated cases. There are 36 cancer types according to the latest broad cancer classification (Bray et al., 2018). The top twenty most common cancer types account for 86% of the total diagnosed cases and 85% of the cancer related deaths (Figure 1.1.1) (Bray et al., 2018).

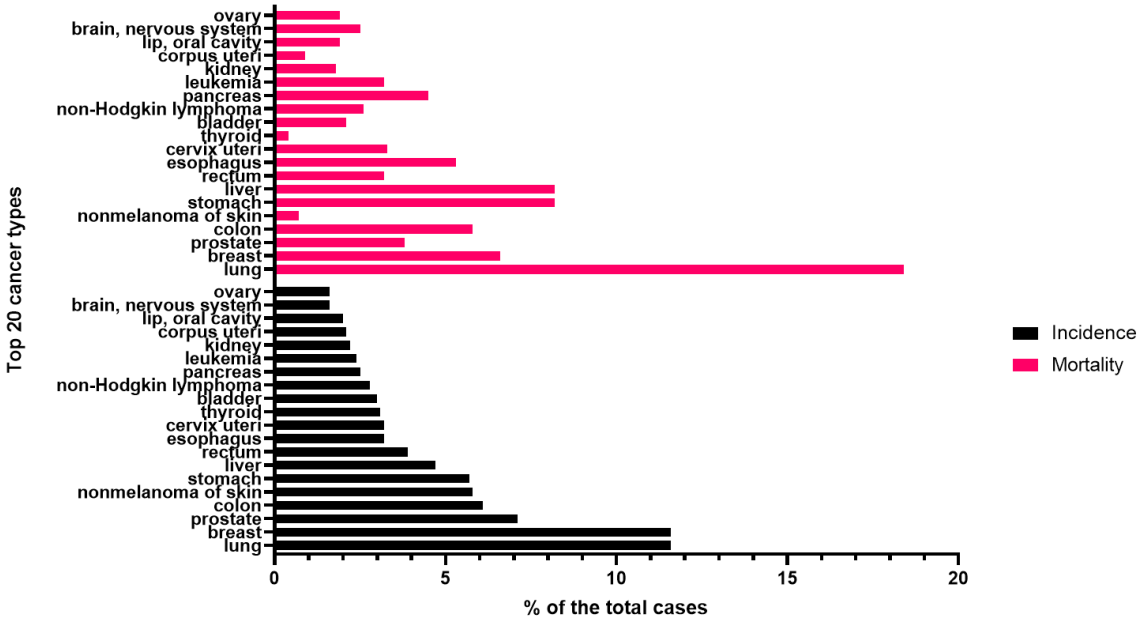


Figure 1.1.1 Incidence and mortality rate of the 20 most common cancer types in 196 countries.
The most common cancer type is lung, followed by breast cancer. Lung cancer mortality is the biggest, followed by deaths from liver and stomach cancer, adapted from (Bray et al., 2018).

Numerous studies suggest that somatic mutations are part of normal evolution, with some of them inevitably leading to cancer progression (Gerstung et al., 2020). The idea of the tumour microenvironment was first introduced after the realisation that a tumour is not a homogenous carcinogenic mass of cells. Tumours, like normal cells, require oxygen and nutrients as well as ways to get rid of their metabolic waste such as CO_2. Therefore, enhanced angiogenesis is vital for tumour progression and once this pathway is activated, it is never turned off. Tumour vascularisation is characterised by large, abnormal, newly formed vessels with unstable blood flow, increased levels of endothelial cell proliferation factors, and decreased levels of apoptotic factors. Therefore, it has been included in one of the hallmarks of cancer (Hanahan and Weinberg, 2011). As well as an enhanced and sustained blood supply, cancer cells are also thought to largely avoid the immune system. One theory for how this happens is cancer immunoediting which suggests that cancer can at first be eliminated successfully by the immune system (the elimination phase), followed by the equilibrium phase, during which malignancy is controlled but active, and finally the escape phase, in which the immune response fails and cancer cells escape the defence mechanisms (Kobold et al., 2010; Vesely et al., 2011).

1.1.1. Renal Cell Carcinoma (RCC)

1.1.1.1. RCC statistics - Epidemiology

Cancer of the urinary system is the 5th most diagnosed cancer in the USA and accounts for 9% of the total diagnosed cases. Kidney cancer accounts for 50%

of these cases, and almost 40% of its mortality rate. Kidney cancer cases continue to increase every year, and its survival rate has generally increased, except for cases related to uterine cancer (Siegel et al., 2020). Worldwide, kidney cancer accounts for 2% of cancer mortality and incidence, with almost 400,000 new cases and 175,000 deaths per year (Bray et al., 2018). Epidemiologically, the highest incidence has been described in Native and African Americans, and less in Asian Americans. Men have an almost 50% higher chance of developing kidney cancer than women. The mean age for developing this type of cancer is 64 years old, consistent with studies showing that people older than 65 years old have an increased chance of developing this type of cancer. Specifically, Renal Cell Carcinoma (RCC) accounts for almost 90% of the kidney cancer cases (Hsieh et al., 2018; Saad et al., 2019). RCC is within the ten most common types of cancer worldwide: (Hsieh et al., 2018). In the last 30 years, a slow regression in the mortality rate of RCC has been reported, especially in western countries. The incidence rate plateaued since 2008, compared to it increasing by 2.4% every year before that since 1992. Increased use of CT scan or has led to improved diagnosis (Chow et al., 2010; Kumar et al., 2014; Saad et al., 2019; Liu et al., 2018) and almost 50% of the newly diagnosed patients are at stage I. However, half of the patients will relapse after nephrectomy, and almost 16% of those diagnosed have already developed metastasis. Metastatic patients have only a 12-20% chance of 2-year survival (Huang and Hsieh, 2020; Botta et al., 2017). Extremely rare cases of patients presenting metastatic RCC (mRCC) but without the existence of detectable mass in the kidneys have also been reported (Kumar et al., 2014).

1.1.1.2. Classification

RCC is a heterologous disease with numerous different types categorised based on their histological, genetic, molecular, and epigenetic profiles. There are over 10 types of RCC including: clear cell (ccRCC), papillary (with further classification of 2 further subtypes named type I-basophilic and type II-eosinophilic), chromophobe, collecting duct (Bellini's), unclassified RCC, medullary hereditary leiomyomatosis RCC, hypermethylated CpG island methylator phenotype-associated (CIMP) RCC and metabolically divergent chRCC, renal oncocytome, mucinous tubular and spindle type, tubulocystic carcinoma, mucinous tubular and spindle cell carcinoma, Xp11.2 translocation carcinoma, renal medullary carcinoma and carcinoma associated with neuroblastoma. CcRCC has the poorest survival of all and accounts for 75% of cases. Its name is derived from the higher concentration of glycogen and lipids compared to normal cells, making these cells look ''clear'' (Moch et al., 2016; Linehan and Ricketts, 2019; Oosterwijk, 2011; Penticuff and Kyprianou, 2015; Brodaczewska et al., 2016). RCC derives from the renal tubular epithelial cells and the parenchymal cell as an adenocarcinoma, and the cells that are the most compromised are the ones of the tubules close to the nephrons, causing structural instability. RCC is a very heterogeneous type of cancer caused by the intra and inter-tumour complexity. It is a solid tumour, and it is usually present in one side of the nephron. The growth rate of renal tumours is 0.28 cm/year (Penticuff and Kyprianou, 2015; Hsieh et al., 2018; Fu et al., 2011). RCC is a very challenging disease to diagnose and it is usually unaffected by radiation

therapy (Huang and Hsieh, 2020). Currently, non-metastatic RCC is treated surgically by whole or partial nephrectomy (the latter can be done laparoscopically which is preferred for smaller tumours because it is a less invasive approach), ablation (by cryotherapy or cryotherapy radiofrequency), active surveillance or chemotherapy. Metastatic RCC is less responsive to chemotherapy and its common metastatic sites are lungs, lymph nodes, bones and liver (Pastore et al., 2015; Hsieh et al., 2018; Saad et al., 2019; Kumar et al., 2014). There are four stages of RCC, ranging from I to IV, depending on its spread to lymph nodes (Table 1.1.1.).

Table 1.1.1 Description of the four different stages of RCC and an estimated 5-year survival rate (Hsieh et al., 2018).

Stage	Tumour specification	5-year survival rate
I	<7 cm, limited to kidney	95%
II	>7cm, limited to kidney	88%
III	>7 cm, lymph node affected	59%
IV	>7cm, adrenal gland affected, metastasis	20%

Specifically, there is an additional grading system to classify clear cell carcinoma, as it has been revised by the WHO, and it classifies this specific type of cancer in four different stages. Basically, grade refers to the degree of differentiation of the cancer with grade IV having the worst prognosis. It should not get confused with the aforementioned stage classification which refers to how advanced the cancer is (Dagher et al., 2017).

Diagnosis still happens incidentally and almost exclusively by imaging techniques. Currently, MRI (Magnetic Resonance Imaging) and CT (Computed Tomography) scan are the only available options for diagnosis followed by

histopathological analysis in order to determine the stage of the RCC (Richard et al., 2013). Symptoms include abdominal pain and haematuria, but most people are asymptomatic until diagnosis. Additionally, other blood tests are complementing the diagnosis or the prognosis of patients such as blood cell count (red blood cells, white blood cells and platelets) and calcium and lactate dehydrogenase levels (Hsieh et al., 2018; Rajandram et al., 2019).

1.1.1.3. RCC lifestyle related risk factors

Several cross-sectional studies suggest the common life-style related risk factors for developing RCC are obesity, high blood pressure and smoking. It has also been suggested that moderate alcohol consumption might have a protective effect (Capitanio and Montorsi, 2016; Saad et al., 2019). There is a growing body of literature that recognises the importance of these risk factors. Over half of the diagnosed cases were reported to be associated with at least one of them (Hsieh et al., 2018; Williams, 2014). Measurements related to obesity such as body mass index (BMI), waist to hip ratio and adipocity are all related to increased oxidative stress. The worldwide increase in obesity cases might have contributed to the increased incidences of RCC (Hsieh et al., 2018; Chow et al., 2010; Rajandram et al., 2019; Saad et al., 2019). A large meta-analysis of 9 million participants showed a clear correlation between body mass index (BMI) and risk of kidney cancer in both men and women, in a dose response matter. This correlation could be due to several reasons such as resistance to insulin, high level secretion of specific hormones (like estrogen) or adipokines (like TNF-alpha). Specifically, high estrogen and IGF (insulin like

growth factor) levels have been associated with carcinogenesis. Overweight participants (BMI 25-29.9) had a 35% more chance of developing kidney cancer, and the risk was doubled in obese participants (BMI> 30) (Liu et al., 2018). Furthermore, the incidence rate is increased by almost 12% by an increase in one unit of BMI ($kg.m^2$) (Williams, 2014).

Existing research also recognises the critical role played by tobacco usage as an established RCC risk due to the toxicity caused on the renal tubules and the increased DNA damage (Ridge et al., 2014). Smoking has also been associated with an increased risk of metastasis (Saad et al., 2019). On average, there is a 35% higher chance for a smoker to develop cancer than a non-smoker. Other contributing effects from tobacco smoking could be hypoxia, the presence of the carcinogens n-nitrosamine and benzo[α]pyrene diolepoxide, and the shortening of telomere length (Chow et al., 2010).

Hypertension (increased blood pressure) has been positively associated with RCC (amongst other cancers). Its association has been described in both women and men, in a dose-dependent manner with approximately 5-7% elevated risk, although whether it is a direct outcome or a cause is yet to be established (Seretis et al., 2019).

A variety of studies suggests that there is a protective effect of alcohol for RCC risk. Notably, different type of alcoholic drinks had different effects on RCC risk; for example, consuming wine was reported more protective for women, whereas beer and spirits were more protective for men. Possible reasons for the protective effect of alcohol consumption could be the presence of specific

antioxidants that reduce the oxidative stress, or the diuretic effect allowing less time with the renal epithelial cells to be in contact with carcinogenic substances (Xu et al., 2015; Hsieh et al., 2018; Capitanio and Montorsi, 2016).

Another potential risk factor is reduced physical activity; incidence of RCC has been reported to be 61% lower for people that undertake regular physical activity (Williams, 2014). Moreover, meat intake might elevate RCC risk, and specifically mutagens derived from cooking red meat in high temperatures. However, this could be also explained by other confirmed associated risks like obesity or smoking that was common between the participants of this study (Daniel et al., 2011). Aspirin usage was positively correlated with the increase of risk whereas the opposite has been reported for statins, that are normally prescribed for reducing the cholesterol levels (Hsieh et al., 2018). Occupation exposure has also been reported to be implicated in the progression of RCC, for example exposure to asbestos, rubber, paints or trichloroethylene (Kabaria et al., 2016; Chow et al., 2010). Malnutrition has been negatively associated with the risk of RCC development. Specifically, large intake of micronutrients like carotenoids and vitamin A and C have a protective effect. In fact, the role of the concentration of micronutrient vitamin C in a cohort of 10 studies was investigated and its daily intake was associated with lower RCC risk, possibly due to its antioxidant function (Bock et al., 2018; Jia et al., 2015). Healthy diet with fruit and vegetables has a protective effect with an almost 30% reduction in incidence rate (Chow et al., 2010). In addition to life-related risk factors, other pre-existing medical conditions have been associated with RCC. These

include chronic kidney disease, acquired kidney cystic disease, diabetes mellitus, long term dialysis, kidney transplantation, Cowden syndrome (multiple hamartoma syndrome), hyperparathyroidism, jaw tumour syndrome, and a family history of RCC (Hsieh et al., 2018; Pastore et al., 2015).

1.1.1.4. RCC genetic risk factors

As in many other cancers, there are also genetic factors for RCC. The conversion of methylated cytosine to thymine for instance has been described for RCC, with most of these mutations happening at around the age of 60 (Gerstung et al., 2020). Von Hippel-Lindau tumour suppressor (VHL) protein plays a dramatic role in RCC progression. It has also been associated with other types of cancer (for example haemangioblastomas, pheochromocytomas and pancreatic neuroendocrine tumours) but there is a 70% chance of developing ccRCC by the age of 60 in carriers of mutations of this gene (Mantovani et al., 2008; Fu et al., 2011). The mechanism is common in the different cancers (as illustrated in Fig 1.1.2.). VHL has 2 isoforms, both phosphorylated, one of 30 kDa and the other of 19 kDa (Penticuff and Kyprianou, 2015). It is a member of an E3 ubiquitin-protein ligase complex. The gene for VHL is located on chromosome 3 and the encoded protein is 213 amino acid long (Patel et al., 2006). In normoxia (when the oxygen level is normal) VHL is phosphorylated and it ubiquitinates the hydroxylated hypoxia inducible factor (HIF-α) transcription factor thus leading to HIF-α degradation. In hypoxia (when the oxygen level is reduced), and when VHL is mutated or truncated (deletion is more likely than promoter hypermethylation or rearrangement), HIF-α

escapes degradation by the ubiquitination complex. HIF-α is no longer hydroxylated in the hypoxic tumour microenvironment so it binds to HIF-responsive element. This leads to the upregulation of HIF-α protein, as well as the recruitment of other transcription factors. This promotes the overexpression of a range of proteins, including VEGF (Vascular endothelial growth factor) and EPO (erythropoietin) (Patel et al., 2006; Cairns, 2012; Richard et al., 2013; Fu et al., 2011; Bielecka et al., 2014), leading to angiogenesis and cancer development. Moreover, hypermethylation of the promoter of the VHL gene (gene expression is downregulated) is the second most frequently observed event (37% of RCC cases) (Linehan and Ricketts, 2019).

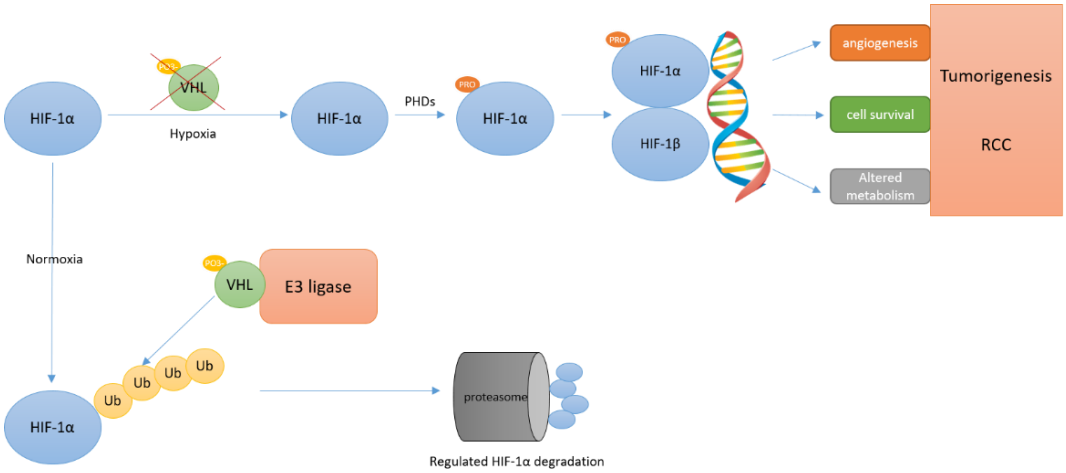


Figure 1.1.2 Illustration of the Hypoxia induced pathway involving HIF-1α.
In normoxia phosphorylated VHL ubiquitinates the transcription factor HIF-1α leading to its degradation by the proteasome. In contrast, in hypoxia or in the absence of functional VHL, HIF-1α and β lead to the upregulation of the transcription of other genes that promote angiogenesis and cancer formation, adapted from (Bielecka et al., 2014).

The risk factors for the remaining 50% of diagnosed RCC cases needs to be investigated. For instance, the abnormal level of DNA methylation in white

blood cells, and especially leukocytes has been suggested to be an event correlated with the risk of RCC development (Liao et al., 2011). Numerous other genes are mutated in several types of RCC (Table 1.1.2), with PBRM1 (Protein polybromo-1) being the 2nd most commonly mutated gene (Capitanio and Montorsi, 2016; Linehan and Rathmell, 2012; Oosterwijk, 2011; Brodaczewska et al., 2016; Voss et al., 2018). The loss of the 3p25 chromosome arm has been described as an important genetic event in the development of RCC, since PBRM1, SETD2, BAP1 and VHL genes are all closely clustered in this region. Presence of the other mutated genes varies from patient to patient and even within the tumour mass in a phenomenon called intratumoral genetic heterogeneity. (Hsieh et al., 2018; Oosterwijk, 2011). Some additional single nucleotide polymorphisms (SNPs) in various chromosome loci that have been reported in association with RCC are the following: 1q24.1, 2p21, 2q22.3, 8q24.21, 11q13.3, 12p11.23 and 12q24.31, VHL_rs779805 (Henrion et al., 2015; Oosterwijk, 2011; Liu et al., 2020; Hsieh et al., 2018; van de Pol et al., 2020). Impairment of the TERT gene promoter has been reported to be associated with worse survival rate of RCC, too (Casuscelli et al., 2019).

In some cases, RCC patients did not have any observed genetic mutations, but they presented similar unusual methylation patterns. Sequencing of these patterns are likely to facilitate monitoring the transcription levels of other proteins (Junker et al., 2013).

Table 1.1.2. List of reported genes that have been implicated to RCC and their corresponding protein name (Chow et al., 2010; Capitanio and Montorsi, 2016; Hsieh et al., 2018; Linehan and Ricketts, 2019).

Protein name	Gene name
Annexin a4	ANXA4
Apolipoprotein C1	APOE/C1
AT-rich interactive domain-containing protein 1A	ARID1A
Putative Polycomb group protein	ASXL1
BRCA1 associated protein-1/ubiquitin carboxy-terminal hydrolase	BAP1
Zinc finger protein basonuclin-1	BNC1
Carbonic anhydrase 9	CAIX
Caveolin	CAV1
Neprilysin	CD10
Integrin beta chain-2	CD18
p16	CDKKN2A
cytokeratin-8	CK8
Core component of multiple cullin-RING-based ECS	CUL2
Fibrillin-2 precursor	FBN2
Follicullin fumarate hydratase	FH
Glutathione S-transferase Mu 1	GSTM1
Glutathione S-transferase P	GSTP1
GSTT1 - Glutathione S-transferase theta-1"	GSTT1
Glutathione S-transferase	GST-α
Lysin-specific demethylase 5C	KDM5C
Lysyl oxidase	LOX
Tyrosin-protein kinase Met/hepatocyte growth receptor	MET
Microphthalmia-associated transcription factor	MITF
Mixed-lineage leukemia protein 3	MLL3
Mammalian target of rapamycin	MTOR
Protein polybromo-1 (PB1)/BRG1-associated factor 180	PBRM1
Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	PIK3CA
Phosphatase and tensin homolog	PTEN
Succinate Dehydrogenase Complex Iron Sulfur Subunit B	SDHB
Succinate dehydrogenase complex subunit C	SDHC
Succinate dehydrogenase complex subunit D	SDHD
SET domain containing 2	SETD2
Secreted frizzled-related protein 1	SFRP1
Transcription activator BRG1	SMARCA4
Transcription elongation factor B polypeptide 1	TCEB1
Transcription factor E3	TFE3
Transcription factor EB	TFEB
p53	TP53
Tuberous sclerosis 1/Hemartin	TSC1
Tuberous sclerosis complex 2	TSC2

Vitamin D3 receptor	VDR
Vimentin	VIM

1.1.1.5. Management/Therapy/Research

Interferon and IL-2 administration were the only available treatments for RCC until 2005 but no more than 10% of the treated patients responded to them (Huang and Hsieh, 2020). The role of VEGF in cancer progression was demonstrated 25 years ago, establishing it immediately as an appealing therapeutic target. Since then, several anti-VEGF antibodies have been developed and used for RCC treatment with a patient response up to 40%. The past 15 years have seen increasingly rapid advances in the field of approved therapeutic agents, some of them are the following (by chronological order): sorafetinib, sunitinib, temsirolimus, everolimus, bevacizumab, pazolanib, axitinib, nivolumab, cabozantinib, lenvatinib, ipilimumab, pembrolizumab, and avelumab (targeting a wide range of protein targets) (Huang and Hsieh, 2020; Hsieh et al., 2018; Saad et al., 2019). Some combinations of the aforementioned reagents (depending on the genetic predisposition of the patients) are also applied (Nabi et al., 2018). An example of the efficacy of the treatments is the administration of interferon to metastatic patients that has increased the survival rate by on average 6 months and its combination with several other agents such as IL-1 and other VEGF-inhibitors has been shown a further improvement of up to 4 months. A family of monoclonal antibodies called checkpoint inhibitors have also been used to treat RCC and they enhance the T-cell response against the tumour cells. Examples of the overall survival

with checkpoint inhibitors are 24 months with nivolumab, 20 months for everolimus, and 29 months for atezolizumab in combination with avelumab. In one of the trials, the response rate for the administration of pembrolizuab and IL-2 has been increased by 30%, and the three year survival rate to 46% (Botta et al., 2017). T cell vaccination is another treatment for RCC that is under development. This approach aims to activate the immune system's response against the cancer cells using extracted RNA from RCC tumours. Phase II results were very optimistic, with an increase of overall survival up to 2.5 years (Penticuff and Kyprianou, 2015; Botta et al., 2017). Patients that have received different types of immunotherapy had complete eradication of RCC, even though they were in the late stages of the disease. This indicates that RCC, in some cases, might be sensitive to the immune system of the host, although larger data set is required (Moch et al., 2014). Interestingly, RCC T cell infiltration rate was the highest amongst 19 cancers (Hsieh et al., 2018).

For research purposes, it is challenging to recapitulate RCC *in vitro* because it is a highly vascularised tumour *in vivo*. *In vitro* research in RCC is being facilitated by the use of cell lines derived from patients after nephrectomy. These serve as a first line platform for testing the efficacy and toxicity of the therapeutic reagents. Some ccRCC cell lines are: 786-O, UM_RC-2, SKRC-44, SNU-333, SKRC-45, and for metastatic ccRCC: Caki-1 (Brodaczewska et al., 2016). Additionally, it has been very challenging to develop a mouse model for ccRCC since a double knockout of the main driver gene of the disease (VHL) is not sufficient; homologous deletions of VHL and BAP1 have also been

attempted but led to early death of the mice, and unrelated tumours developed when heterologous deletions of VHL were present (Hsieh et al., 2018). There is a promising RCC model, with an active HIF1α topologically expressed in the kidney proximal tubules named TRACK (transgenic model of cancer of the kidney). This was achieved by mutation of HIF in three different positions (P402A, P564A, and N803A). The histopathological analysis of the mice revealed similar characteristics as the ones observed in human ccRCC after 3 months with notable carcinoma *in situ* in mice older than a year (Fu et al., 2011).

To summarise, RCC has been characterised as the cancer with the most deaths among genitourinary cancers. Patients have higher survival chances if they have been diagnosed early and if the tumour is removed surgically, and there is no biomarker available for stratification, prognosis, or diagnosis. Thus, there is a need for the development of a diagnostic tool for RCC that will allow diagnosis at an early stage. Due to its high heterogeneity of RCC, there is also a need for biomarker(s) that can predict the stage of the disease and to personalise the therapy for each individual patient (Cairns, 2012; Craven et al., 2013).

1.1.2. Autoantibodies in cancer

Immunity against foreign pathogens/antigens has been traditionally divided into innate and adaptive immunity. All the cells involved in the immune response are derived from a common precursor hematopoietic cell. Innate immunity is the first line of the immune response, rapid (normally within

hours) and against a wide range of antigens. This is followed by adaptive immunity that has higher affinity and specificity and forms memory so that upon second infection from the same antigen, there is an immediate and specific response against it (Murphy et al., 2008; Riera Romo et al., 2016; Netea et al., 2019). Synergy occurs between the major cell subtypes for the adaptive immune response (B and T cells) (Chaplin, 2010; Netea et al., 2019). B cells respond to foreign antigens by producing high affinity, specific antibodies. The early antibody response is made up of mostly IgM, although later in the response B cells can class-switch to produce other isotypes such as IgG, which makes up the vast majority of the immunoglobulins (antibodies) present in sera (Table 1.1.3). Antibodies are proteins that can be divided into Fab fragments, which contain two identical heavy and two identical lights chains the variable regions of which bind specifically to a target antigen, and an Fc region, which is responsible for effector functions. Naïve B cells patrol the secondary lymphoid organs until they recognise an antigen through their B cell receptor (BCR). They then become activated and can differentiate into plasmablasts and plasma cells which secrete large amounts of antibody that makes up the early protective antibody response, or they can enter germinal centre reactions where they interact with T-helper cells and undergo affinity maturation and differentiate into plasma cells producing high affinity antibody or memory B cells. Memory B cells are crucial in the case of a secondary infection as they are able to rapidly differentiate into plasma cells which secrete high affinity and high specificity antibodies (Chaplin, 2010; Nutt et al., 2015; Akkaya et al., 2020). Checkpoints normally ensure that B cells expressing

a BCR that is specific to a self-antigen are eliminated by clonal deletion, clonal anergy, or by receptor editing, and this concept of self-tolerance is crucial to prevent pathology due to an autoimmune response. In autoimmune diseases, the immune system fails to distinguish between foreign and self-antigens and these cases are often characterised by the production of self-reactive autoantibodies (AAbs) (Pedersen and Wandall, 2011; Alberts et al., 2002). Cancers can also elicit autoantibodies to 'self-antigens' and the B cell response against tumour associated antigens (TAAs) is similar to that against pathogens or foreign antigens. TAAs may evade tolerance and elicit a B cell response despite being self-antigens for many possible reasons, for example, they may be mutated versions of physiological antigens, or they may have abnormal post-translational modifications causing the formation of neo-epitopes (epitopes that would not normally be present) (Kobayashi et al., 2020). Three prevailing theories currently exist for the existence of immune responses in cancer: a) the immune response promotes disease progression; b) the immune response suppresses disease progression; or c) it is an unrelated phenomenon (Kobold et al., 2010). ''Which came first: cancer or inflammation?'' has yet to be answered. In some cases, autoimmunity pre-exists cancer and in other cases cancer affects the immune system and results in autoimmunity. There is a well-established link between these two conditions and trying to diagnose one, could potentially lead to diagnosis of the other (Mantovani et al., 2008). Specifically for cancer autoimmunity, it has been reported that different subclasses of antibody are responsible for the autoimmune response

depending on the stage of the cancer progression (Kobayashi et al., 2020) (Table 1.1.3) (Chaplin, 2010).

Table 1.1.3. The 5 subtypes of immunoglobulins that are present in human sera, their molecular weight and their concentration as well as a brief description of their physiological role and their role in autoimmunity in cancer (Murphy et al., 2008; Chaplin, 2010).

Immunoglobulin subtype	IgG (monomer)	IgA (monomer/dimer)	IgM (pentamer)	IgD (monomer)	IgE (monomer/dimer)
Molecular Weight (kDa)	150	160-400	950	175	190
Sera concentration (mg/ml)	17.5	2.5	2	0.03	0.003
Role (physiological)	Neutralisation of bacteria/viruses, complement activation	Alternative complement activation, mucous pathway, inflammation	Neutralisation of bacteria/viruses, complement activation, endocytosis	B cell /mast cell activation	Allergy. Binding to mast cells/ basophils
Role (autoimmunity in cancer)	Predominantly present during the whole tumour progression	Later stages, multiple antigens	Early stages, nuclear proteins	-	-

The first report of innate immune response in cancer was described by Baldwin in mice, reporting the variability of the individuals when challenged with tumour grafts (Baldwin, 1966) and describes the quantification of the autoimmune response against TAAs by using immunofluorescence (Baldwin, 1971). There are a lot of potential causes of autoantibody production in cancer (Figure 1.1.3). The rate of DNA mutations is increased in cancer, and this could cause the formation of different epitopes (neo-epitopes) caused by a conformational change in protein structure by a mutation, truncation of the gene or a translation of an alternative open reading frame. Post-translational modifications (PTMs) also play a significant role in generating an immune response. The most common reported PTMs in cancer are glycosylation and

methylation, probably caused by mutations of enzymes that facilitate these modifications. Furthermore, immune responses against self-antigens can be initiated even against antigens that have not been mutated. A factor contributing to this process could be the upregulated apoptosis and necrosis of cells, releasing a plethora of intracellular components that are not normally present, such as nuclear antigens, thus immune response is elevated against them. Another reason for immunogenicity could be the presence of TAAs in tissue where it is not normally expressed. Interestingly, autoantibodies could also be against whole vesicles, like tumour derived exosomes. The existence of AAbs may or may-not be beneficial, but they could be an indicator of the disease long before the development of symptoms. Indeed, AAbs can be produced in the early stages of cancer, and they have been observed in patients with breast, lung, gastrointestinal, ovarian, and prostate cancer. AAbs could therefore potentially be ideal biomarkers since their level is relatively higher than the protein they are specific to (Zaenker et al., 2016; Kobayashi et al., 2020; Yadav et al., 2019).

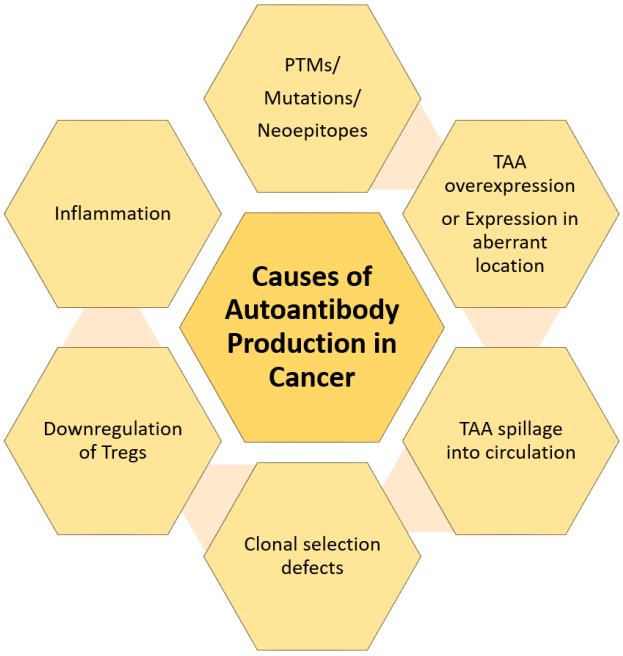


Figure 1.1.3 A summary of factors that contribute to autoimmune response in cancer, adapted from (Zaenker et al., 2016).

1.1.3. Cancer biomarkers

Biological markers (biomarkers) are any of a wide range of components (proteins, metabolites, DNA and epigenetic alterations, RNAs, hormones, peptides, antibodies) that can be objectively measured and facilitate in diagnosis. They can represent markers of prognosis, disease stratification, disease progression, or response to therapy. A ''good'' biomarker should have high specificity and sensitivity, facilitate in early screening, contribute to pharmacogenomics (choosing the right treatment based on someone's genome profile), differentiate between stages or severity, be cost effective and ideally non-invasive. They might be present in body fluids (blood, saliva) or tissue (biopsy required) (Costa-Pinheiro et al., 2015; Etzioni et al., 2003; Füzéry et al., 2013; Mordente et al., 2015; Henry and Hayes, 2012). Sensitivity is the ability of a method or a diagnostic test to detect the patients of the specific

disease whereas its specificity, it is determined as the ability to not identify patients with unrelated disease (Füzéry et al., 2013). There is an urgent need to increase the current available repertoire of tumour biomarkers in order to increase the quality of life of patients, personalise the available therapies/treatments, increase the chances of total eradication and decrease the economic burden of cancer (Mordente et al., 2015; Kirwan et al., 2015). Most importantly, early detection in cancer patients is crucial; it relies on the simple fact that tumours should be treated before cancer metastasizes because at that stage the cancer has a better chance to be eliminated (Etzioni et al., 2003; Wu and Qu, 2015). Early detection in cancer is currently limited by the diagnostic tools that are available therefore there is a global need for the development of biomarkers (Desmetz et al., 2009). Even for the most commonly used biomarkers for screening such as PSA (prostate specific antigen) and CEA (carcinoembryonic antigen), only a relatively small fraction of patients has elevated levels of these proteins. Therefore, it is apparent that a combination of biomarkers is required in order to accurately detect the condition. Biomarkers such as mutations in p53 or hypermethylated promoter regions of cancer-associated genes are one of the first that have been described but unfortunately, they were not specific for one type of cancer. Moreover, the majority of the FDA approved biomarkers lack a single biomarker recommendation approach and the majority of the available biomarkers are not for early detection or prevention (Sidransky, 2002; Kirwan et al., 2015). There are few cancer biomarkers that have been approved by the FDA and broadly used for diagnosis (Table 1.1.4). For instance, AFP is a protein

that is normally expressed during pregnancy, but it is also an indicator of liver and pancreatic tumours amongst others. However, its specificity is low as it is present in increased concentration in a variety of tumours, and its sensitivity for hepatocellular carcinoma varies from 40-60%. Furthermore, PSA is a protein normally expressed in the prostate and its elevated concentration is indicative of prostate tumour but its low specificity (70%) can lead to overdiagnosis and increased use of imaging techniques (Kirwan et al., 2015). Even though there is a growing body of literature reporting numerous biomarkers, most of them never get approval from the FDA. This failure to secure approval is due to various reasons such as lack of reproducibility, bias introduced by the type of the diagnostic assay (e.g. ELISAs) or introduced by the lack of a diverse ethnicity background of population included in the studies (e.g. Great response against White-European, low response against Asian-American) (McShane et al., 2005; Ludwig and Weinstein, 2005). The discovery of a single tumour biomarker is very challenging, mostly because cancer is a heterogeneous disease and differs from patient to patient and even intra-tumour. On the other hand, a panel of biomarkers is more likely to provide a specific and sensitive assay. Circulating protein biomarkers (mainly measuring the level of the proteins in body fluids) have been described. However, proteins have a short half-life and most importantly, their concentration is relatively low and cannot easily be amplified *in vitro* like nucleic acids. Free circulating nucleic acids have been an attractive target as biomarkers, but their clinical use has been limited due to their lack of specificity for a single cancer type (Ludwig and Weinstein, 2005; Wu and Qu, 2015; Kirwan et al., 2015).

Specifically, circulating tumour DNA (ctDNA) have been successfully detected in lung cancer patients with sensitivity and specificity that reached up to 100% for stage III and IV patients, but significant lower (50%) for stage I. Elevated levels of ctDNA has also been described in non-cancer conditions, such as pregnancy, inflammation, and diabetes. Therefore, its incorporation into clinic has been delayed. Moreover, non-coding RNAs such as miRNAs could also serve as biomarkers, although their low concentration, sequence homology and abnormal structures have made them diagnostically challenging. Furthermore, the detection of whole circulating tumour cells (CTCs) has recently been approved by the FDA, although its efficacy varies between studies (Wu and Qu, 2015; Postel et al., 2018). Additionally, combinations of biomarkers have shown some efficacy. For instance, a combination of detection of IL-7 and CA125 (an FDA approved cancer biomarker) has improved the accuracy of ovarian cancer diagnosis to 69% (Eftimie and Hassanein, 2018). Generally, a biomarker with ~80% sensitivity and ~70% specificity is considered ''good'', and some of the already widely used biomarkers fall outside this margin (Schiffman et al., 2015).

Table 1.1.4 List of FDA approved biomarkers used for screening, diagnosis, stratification and prognosis for various cancer types (Ludwig and Weinstein, 2005; Füzéry et al., 2013; Kirwan et al., 2015; Mordente et al., 2015; Schiffman et al., 2015; Postel et al., 2018; Vieira and Schmitt, 2018; Eggener et al., 2019; National Cancer Institute, 2019).

Biomarker	Cancer type	Description
PAP test	Cervical	Screening – Detection of abnormal cells (immunohistochemistry)
Carcino-embryonic antigen (CEA)	Multiple types, mainly colorectal, gastric	Prognostic & Monitoring – Detection of antigen
Alpha-fetoprotein (AFP)	Liver, germ cells, testicular, pancreas, brain	Monitoring – Detection of antigen
Human chorionic gonadotropin-β (β-hGC)	Testicular	Monitoring – Detection of antigen
Prostate-Specific Antigen (PSA) – free or in a complex	Prostate	Screening/Diagnostic & Monitoring – Detection of antigen
Nuclear Mitotic Apparatus protein (Nmp22)	Bladder	Screening/Diagnostic & Monitoring – Detection of antigen
Bladder tumour antigen (BTA)	Bladder	Monitoring – Detection of antigen
Thyroglobulin	Thyroid	Monitoring – Detection of antigen
CA27-29	Breast	Monitoring – Detection of antigen
CA15-3	Breast	Monitoring – Detection of oligosaccharide of antigen
CA-125	Ovarian	Monitoring – Detection of antigen
HER-2/neu	Breast, ovarian, bladder, pancreatic, stomach	Stratification for therapy – Detection of antigen and/or gene (immunohistochemistry)
Progesterone receptor	Breast	Prognostic & Stratification for therapy – Detection of antigen (immunohistochemistry)
Epidermal growth factor receptor (EGFR)	Colorectal	Prognostic – Detection of antigen (immunohistochemistry)
Estrogen receptor	Breast	Prognostic & Stratification for therapy – Detection of antigen (immunohistochemistry)

CA19-9	Pancreatic, bile duct, gastric, ovarian	Monitoring – Detection of antigen
c-kit	Gastrointestinal stromal, acute myeloid leukaemia	Monitoring – Detection of antigen (immunohistochemistry)
p63	Prostate	Differential diagnosis – Detection of antigen (immunohistochemistry)
Circulating Tumour Cells	Breast, prostate, colorectal	Differential diagnosis & prognosis – Detection of EpCAM antigen (immunofluorescence)
AFP-L3%	Hepatocellular	Risk assessment – Detection of antigen (HPLC)
Fibrin (DR-70)	Colorectal, Bladder	Monitoring – Detection of antigen
HE4	Ovarian	Monitoring – Detection of antigen
OVA1 (multiple proteins)	Ovarian	Prediction – Detection of 5 antigens
BRCA1 and BRCA2 gene mutations	Breast	Stratification of therapy - Detection of gene mutation
Cytokeratin fragment 21-1	Lung	Monitoring – Detection of antigen
Cytokeratins	Breast	Prognosis – Detection of antigen (immunohistochemistry)
PD-L1	Non-small cell lung, liver, stomach lymphomas	Stratification of therapy – Detection of antigen
Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor (PAI-1)	Breast	Stratification of therapy – Detection of antigens
FoundationOne® CDx (F1CDx) genomic test	All solid tumours	Stratification of therapy – Detection of 324 gene alterations
21-Gene signature (Oncotype DX®)	Breast	Stratification of therapy – Detection of 324 gene alterations
46-Gene signature (Prolaris®)	Prostate	Stratification of therapy – Detection of 46 genes (RNA expression)
70-Gene signature (Mammaprint®)	Breast	Prognosis – Detection of 70 genes
Epi ProColon®	Colorectal cancer	Diagnosis – Detection of methylated DNA
Not approved by the FDA but commonly used in clinical practise		

ALK gene	Lung	Prognostic – Detection of gene mutation
Antibody gene rearrangement	B-cell lymphoma	Diagnostic – Detection of gene mutation
Bladder Tumour Antigen (BTA)	Bladder	Monitoring – Detection of antigen
Chromosome 3, 7, 17, 21	Bladder	Prognostic – Detection of chromosome deletion/mutation
EGFR gene mutations	Non-small cell lung cancer	Prognostic – Detection of gene mutation
FGFR2/3 gene mutations	Bladder	Stratification of therapy – Detection of gene mutation
FLT3 gene mutations	Acute myeloid leukaemia	Stratification of therapy – Detection of gene mutation
IDH1/2 gene mutations	Acute myeloid leukaemia	Stratification of therapy – Detection of gene mutation
KRAS gene mutations	Colorectal, non-small cell lung caner	Stratification of therapy – Detection of gene mutation
Lactate dehydrogenase	Germ cell, lymphoma, leukaemia, melanoma, neuroblastoma	Prognostic & Stratification for therapy – Detection of antigen
Prostatic Acid Phosphatase	Prostate	Differential diagnosis – Detection of antigen
ROS1 gene rearrangement	Non-small cell lung	Differential diagnosis – Detection of gene alteration
Soluble Mesothelin-Related Peptides	Mesothelioma	Monitoring – Detection of antigen
Somatostatin Receptor	Neuroendocrine tumours	Stratification of therapy – Detection of antigen (immunohistochemistry)
Urine catecholamines (VMA, HVA)	Neuroblastoma	Aid in diagnosis – Detection of antigen
17-Gene signature (Oncotype DX GPS test®)	Prostate	Management – Detection of 17 gene alterations

1.1.3.1. RCC biomarkers

Unfortunately, there are no FDA approved RCC biomarkers. Some studies have identified potential biomarkers, mainly focused in prognosis but none of them

for early detection-screening (Figure 1.1.4) (Pastore et al., 2015; Hsieh et al., 2018).

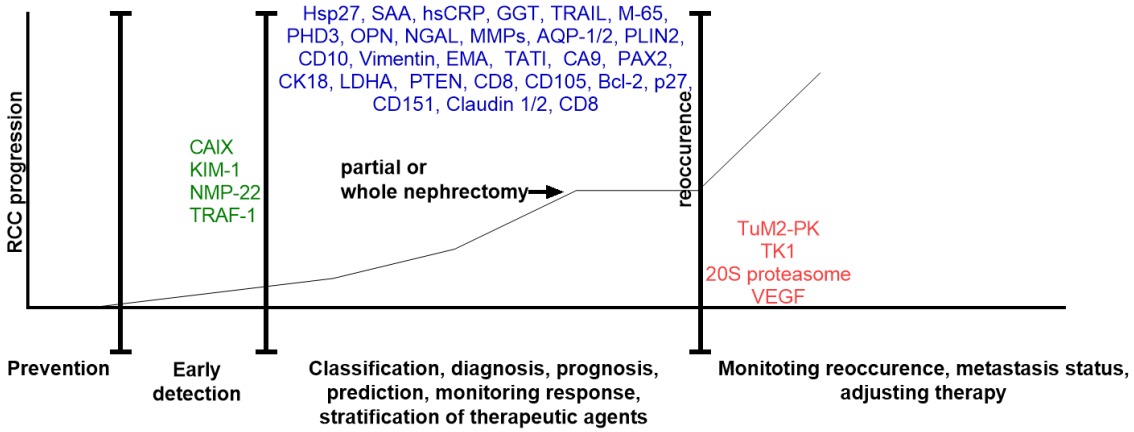


Figure 1.1.4 Illustration of the potential usage of RCC specific biomarkers during different tumour progression stages.
These biomarkers are mainly for diagnosis/confirmation of diagnosis (green text), for monitoring the disease or to personalise the therapy (red text) or are complimentary to histopathological tests and CT scans (blue text), adapted from (Moch et al., 2014; White et al., 2014; Pastore et al., 2015; Golovastova et al., 2017; Klatte et al., 2018).

MRI scans might have a predictive value of the response to inhibitors (everolimus and temsirolimus) in metastatic RCC in mouse models, with the human *in vivo* value yet to be determined (Yiyu, 2019). Profiling of the human leukocyte antigens suggested different patterns present in the RCC when compared to control patients, and it might be indicative of prognosis or even diagnosis (Yilmaz et al., 2010). Predictive markers such the circulation of certain combinations of adipokines could have potential as biomarkers. These include leptin, adiponectin, omentin, apelin, visfantin, resistin (Rajandram et al., 2019). Increased CAIX level within sera was related to metastic RCC, too (Oosterwijk, 2011). VHL mutation or loss cannot serve as a prognostic marker since its incidence is universal and so prominent in all the RCC cases (Voss et

al., 2018). Tumour Necrosis Factor Receptor-Associated Factor-1 (TRAF-1) even though it has been described to be upregulated in cases of RCC, however, the number of samples within the training cohort were small in this study (n=15) (Rajandram et al., 2014). Serum amyloid A (SAA) correlated well with the development of RCC, and its higher concentration was able to distinguish between the different stages with 78% sensitivity and 82% specificity when compared with healthy controls, but the cohort of patients was relative low (n=45), and it was only indicative of the later stages of RCC (Fischer et al., 2012). Additionally, antibodies against a key enzyme named prolyl hydroxylase 3 (PHD3) for the RCC progression known for its involvement in the hypoxia pathway have been detected to be elevated in RCC patients with 86% sensitivity and 58% specificity, but again the tested sample numbers were low (n=17) and there was no comparison with other types of cancer (Tanaka et al., 2011). Furthermore, the concentration of Bcl2, a known apoptotic marker, has been elevated in RCC patients with 50% specificity/sensitivity, but it has also been reported in other types of cancers (Golovastova et al., 2017).

In summary, most of the aforementioned biomarkers lack reproducibility or require further validation in larger cohorts, including non RCC cancer cases. For instance, even though increased levels of NMP-22 have been reported in RCC cases, the same protein is a FDA approved biomarker for urothelial cancer (Pastore et al., 2015).

Non protein markers could be biomarkers, too. For instance, four long non-coding RNAs (lncRNAs) recently reported as prognostic biomarkers, with very

significant association with the survival rate ($p < 0.001$) for ccRCC (Haoran and Tao Ye, Xiaoqi Yang, Peng Lv, Xiaoliang Wu, Hui Zhou, Jin Zeng, Kun Tang, 2020). Additionally, epigenetic signature in RCC may provide alternative biomarkers. Promoter hypermethylation of APC, RARβ2 and RASSF1A genes contributed towards the confirmation of RCC diagnosis with 94% sensitivity and specificity, but the number of RCC patients' samples was low (n=26) and these markers were not assessed for other types of cancer or types or RCC. Moreover, different combinations of micro-RNAs have been reported to successfully detect cc-RCC with 99% sensitivity and 100% specificity (miR-141, miR-200b) but their identification was in RCC tissues for confirming a diagnosis of a kidney mass, rather than using them in screening. Additionally, poor prognosis of RCC has been associated with several histone modifications (H3K4me2, HEK18Ac, H3K9me2, H3K4me1, and H3K4me3). However, the application of epigenetic markers in clinical practice is not commonly used, but might complement diagnosis, due to reproducibility issues and lack of discrimination between different types of RCC (Costa-Pinheiro et al., 2015).

Despite the promise of the aforementioned biomarkers, there is still a need for clinic validation in larger patient cohorts. Therefore, a need for the development of an early detection or screening test for renal cell carcinoma with high specificity and sensitivity remains, since the economic burden of a false positive cancer diagnosis is more than 1000 USD, especially in cases of unnecessary follow-up treatment (Lafata et al., 2004). In addition, the psychological burden of false positive diagnoses might last up to 3 years

(Brodersen and Siersma, 2013). Collectively, around 1.6 billion USD are estimated to be spent in some countries for metastatic RCC. Furthermore, an increase in the total RCC incidents until 2030 has been predicted, therefore there is an emerging need for treatment and earlier diagnosis of this type of cancer (Wong et al., 2017).

1.1.3.2. Autoantibody biomarkers

Autoantibodies against tumour associated antigens could be utilised as early detection biomarkers as studies have suggested that they are present up to 5 years before patients develop any symptoms (Doseeva et al., 2015). In fact, the driver mutations causing the tumour progression might be present even up to decades before the diagnosis (Gerstung et al., 2020). Autoantibodies against TAAs are essentially amplified by the immune system (Kobayashi et al., 2020) and so the produced AAbs against TAAs are in relatively high abundance, even if the expression of TAAs themselves is low. Thus, there is an advantage in utilising AAbs instead of TAAs as biomarkers. Recent papers suggest that it is more likely a panel of AAbs could be a diagnostic tool, rather than the use of single antigens (Etzioni et al., 2003). This is because the level of the AAbs in the serum of an individual may be low for one particular antigen, but a combination of them could be sufficient for diagnosis across patient cohorts (Reuschenbach et al., 2009). Epitope mapping the circulating autoantibodies in cancer patients might also lead to the identification of new cancer antigens, as well as elucidating new pathways related to the disease. Such an approach may even be able to stratify patients based on similar patterns/levels of the

immune response (Moritz et al., 2019). For example, p53 has been labelled the "guardian of the genome" as it signals for the destruction of cells with DNA damage, and is also often the most frequently mutated gene; its mutation frequency reaches up to 94% in some cancer types (but only 2% in RCC) (Ushigome et al., 2018). Consequentially, this can lead to an immune response against mutated p53, even though as a transcription factor it is an intracellular component and autoantibodies against it are not normally present (Kobold et al., 2010). Therefore, it is possible that AAbs will be raised against this very commonly mutated protein, but p53 by itself would be insufficient for a specific diagnostic as its mutations are present in the majority of cancer types. As an example of the need for panels of AAbs, in colorectal cancer, when measuring AAbs against 6 antigens the prediction rate was 56%, but by measuring AAbs against 17 tumour specific antigens the rate was increased to 73% (Ushigome et al., 2018). The same approach has been reported for lung cancer. Autoantibodies against three TAAs were detected with 70% sensitivity and 76% specificity in lung cancer patients when compared with benign lung disease in a large cohort of 322 lung cancer patients and 322 healthy controls within the training set and 124/124 within the validation set (Jiang et al., 2019). Another example of lung cancer detection utilising the immune response against TAAs is the combination of 4 antigens (Cyclin B1, survivin or p53, HCC1) with 65% sensitivity and 100% specificity (Li et al., 2017). One of the latest and more promising immunoassay examples is "PAULA's" test (Protein Assays Utilizing Lung cancer Analytes). This test is unique as it combines antigens and autoantibodies in the same test for the first time. Additionally, elevated

autoantibodies against p16, p53, and c-myc were detected in breast cancer patients with 44% sensitivity and 97.6% specificity (Yadav et al., 2019). One of the latest examples of the usage of the autoantibodies as a diagnostic tool for breast cancer uses 5 antigens to detect AAbs, p53, cyclinB1, p16, p62, and 14-3-3ξ. The resulting ELISA had specificity and sensitivity on average of 93% and 54%, respectively (Qiu et al., 2020). Therefore, a panel of autoantibodies against TAAs could provide a strong diagnostic tool for early diagnosis in cancers, including RCC.

1.1.3.3. Methods for autoantibody screening

Various methods have been used in attempts to analyse the human antibody repertoire in order to discover new antigens responsible for diseases like cancer or autoimmunity. Some of the key methods are serological analysis of expression cDNA libraries (SEREX), serological proteome analysis (SERPA) and protein and peptide arrays.

SEREX involves the construction of a library of cDNA from a patients' tumour mRNA encoding the TAAs that would be topologically highly expressed, and the library is cloned to phage vectors. *E. coli* are transformed with these vectors and protein expression produces bacteria forming lytic plaques (containing the expressed antigens). These are then transferred into a nitrocellulose membrane which then is incubated with the same patient's serum that potentially contains tumour specific autoantibodies. Phage vector of positive bacterial clones were sequenced in order to identify the sequence of the corresponding tumour antigens. The method allowed the analysis of

multiple TAAs at the same time although there were some limitations. PTMs are not considered, and the response can be patient specific and therefore not applicable universally (Zaenker and Ziman, 2013; Kobayashi et al., 2020). SERPA is an advanced version of SEREX, with the difference that protein fractions from tumours are used directly without preparing a cDNA library. Electrophoresis of a nitrocellular membrane containing extracted TAAs results in 2D protein separation based on molecular weight and isoelectric points. The membrane is then probed with patient's sera in a classical Western Blot. The latest version of this technique utilises mass spectrometry to identify TAAs making it more sensitive than SEREX. A limitation is that only linear epitopes, due to protein denaturation, are recognised but it is possible to distinguish between different PTMs (Zaenker and Ziman, 2013; Kobayashi et al., 2020), Again, the identified AAbs can be patient-specific and therefore have very limited diagnostic potential.

With protein microarrays, proteins are either directly purified from cancer serum (by liquid chromatography) or commercially synthetic proteins (taking into account the whole human peptidome). These are spotted onto arrays either with beads or on glass sides and then positive hits (that bind patient sera) are identified by fluorescent signal or conventional western blot. This is a quite an expensive technique requiring excellent protein purification; its common limitations are the absence of conformational epitopes and PTMs but an advantage is that high numbers of proteins can be tested simultaneously (Zaenker and Ziman, 2013; Kobayashi et al., 2020).

Arrays of random peptides can be probed with patient's sera (containing autoantibodies) followed by detection by a fluorescent conjugated antibody. Every disease has a different positive pattern and that has been labelled an immunosignature (Sykes et al., 2013).

1.1.3.4. Immunosignature

An immunosignature is the pattern of positive signal when a random peptide microarray is probed with sera containing antibodies (Stafford et al., 2014). Random peptides were first used for the epitope mapping of known antibodies using approximately 5,000 random peptides (Reineke et al., 2002). This term though was first introduced in 2011, in an attempt to utilise a random 10,000 peptide microarray platform for the global characterisation of the immune response against a pathogen, a vaccine or for non-infectious disease diagnostics. The goal was to reuse the same microarray for the characterisation of a wide range of diseases, due to the fact that each will have a distinct pattern of recognition with the host's antibody response. These short (20 amino acid) random peptides are mimicking the epitopes of proteins or possibly other disease related macromolecules. This approach is appealing as the immunosignature characterisation is not depended on the pathogen (Legutki et al., 2010). These peptides do not need to be the exact epitopes; antibodies can recognise mimotopes that form similar shapes to the epitope. By sorting the top 500 peptides by signal intensity, a different signature for every monoclonal or polyclonal antibody was determined, with 20% of them being unique to each antibody that was tested (Halperin et al., 2011). The authors

suggested that the constant monitoring of healthy individuals' immunosignature over a period of time could be indicative of an early onset of a disease (cancer, autoimmunity, infection etc.) but could also be beneficial for the improvement of the health care system in remote areas, or even for epidemiological usage as a quicker response to a pandemic caused by a zoonotic pathogen (Stafford and Johnston, 2011). The affinity of the antibody interaction with these random peptides was determined to be at the micromolar range (Stafford et al., 2012). This revolutionary method has already been applied for distinguishing different types of cancer (Hughes et al., 2012), for mapping the Alzheimer's disease immune response (Restrepo et al., 2011), discovery of new antimicrobial agents (Johnston et al., 2017) and for rapid diagnosis of stroke (O'Connell et al., 2019). It has been also reported as a way to assess the quality of antibodies (either mono or polyclonal) (Stafford et al., 2020). A thorough study, with a large dataset of samples form 5 different cancers in comparison with healthy controls, revealed that immunosignatures could be an ideal biomarker test with up to 98% accuracy and a clear demonstrable ability for disease classification (Stafford et al., 2014). A more recent version of the immunosignature approach is described that uses more than 300,000 random peptides per assay (Legutki et al., 2014). This updated version was used for successfully mapping new immunosignatures from 7 different diseases, all caused by different pathogens (Richer et al., 2015). The immunosignatures of the healthy individuals is highly variable between people, but this unique signature alters very quickly in the case of the progression of an infection or autoimmunity to produce a recognisable and specific pattern

(Stafford et al., 2012). A healthy person's immunosignature will, broadly speaking, remain the same; slight changes depending on the time of day have been reported. Interestingly, significant differences were observed depending on which kind of body fluid when comparing antibodies from serum and saliva (Stafford et al., 2016).

1.2. Ligand display technologies

The display of libraries of ligands all with potentially unique binding properties (antibody fragments or peptides) have a myriad of applications and can be achieved with a wide range of methods (bacteriophage, yeast, cell etc. display). The following section will be primarily focused on phage display.

1.2.1. M13 biology and architecture

M13 is a filamentous bacteriophage that is shaped like a cylinder. Its length is around 1 µm long and its genome is single stranded DNA, around 6500 bp in total. Its genome contains 11 genes, 5 of which encode coat proteins named pVIII, pVI, pIII, pVII and pIX with 5 copies for each of them in the phage particle except pVIII. The major coat protein pVIII is the most represented with 2700 copies and encapsulates the bacteriophage DNA (Løset and Sandlie, 2012; Carmen and Jermutus, 2002) (Figure 1.2.1). The M13 genome has overlapping regions of genes, packaging signals as well as ribosome binding sites, making it difficult for genome engineering (Smeal et al., 2017a). M13 bacteriophage only infect male bacteria via interaction with the F pili that is bound by pIII. There are 3 distinct domains of pIII. N-terminal domain 1 interacts with a complex named TolQRA that is located in the periplasm of the *E. coli*, and N-terminal

domain 2 helps this interaction by bringing these two proteins into close proximity. The C-terminal domain is required for the release of the phage after its assembly. Once inside the bacterium cytoplasm, the M13 genome is quickly converted to double stranded, followed by mRNA transcription and translation of its genes. When there is enough pV protein, it binds to the ssDNA to prevent its degradation. Although the bacteriophage life cycle is well studied, there are few key processes that are still not fully understood, such as how the assembly of bacteriophage proteins is attached to the cell membrane and the specific mechanism of exporting the whole virion. One bacterial cell can produce approximately a thousand bacteriophages. All coat proteins have been reported to be used in phage display, with the pIII and pVIII being by far the most frequently used. An antibody or peptide fusion is displayed at the C or N-terminal of the pIII or pVIII coat protein (Carmen and Jermutus, 2002; Smeal et al., 2017a).

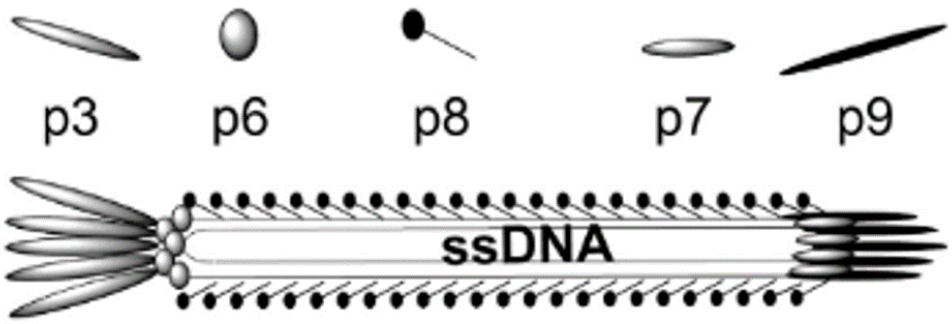


Figure 1.2.1 Filamentous phage architecture (Smeal et al., 2017b)

1.2.2. Phage display

Phage display is a molecular technology that was first described in 1985 and won a Nobel Prize in Chemistry in 2018 (Krištof Bozovicar and Tomaž Bratkovič,

2020). George P. Smith fused a peptide with the pIII coat protein displayed by a filamentous phage. This concept, has now been broadly used for many different applications including antibody discovery for 'difficult' antigens, for example toxic or self-antigens (Smith, 1985). The fundamental principle of this technology is that there is a direct connection between the phenotype (the binding properties of the ligand fused with the coat protein) and the genotype (the DNA sequence coding for the corresponding ligand). The identification of the desired phenotype allows the identification of its genotype and therefore the binding ligand. Different phage display systems are based on either a phage vector (where the library of ligand genes is cloned into the phage genome fused to one of the coat proteins) or phagemid vector system (where a plasmid contains the library of ligand genes fused to the coat protein (Figure 1.2.2). A complete set of phage genes are provided by a helper phage that are packaging deficient, providing larger transformation efficiency and thus larger diversity can be achieved (Clackson and Lowman, 2004; Russel et al., 2004; Lowman, 2013; Ledsgaard et al., 2018).

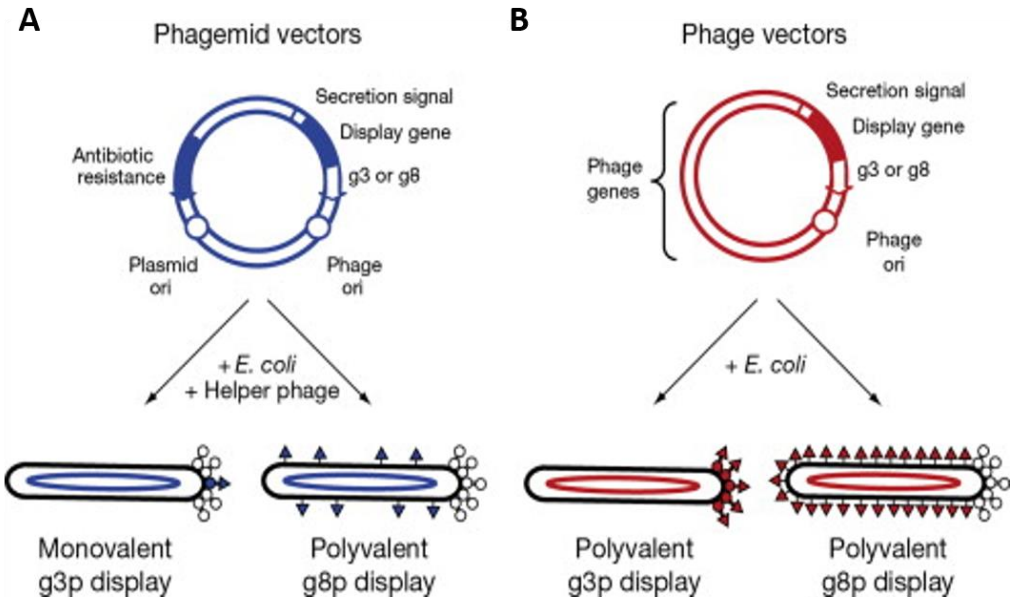


Figure 1.2.2 The difference between the usage of phagemid and phage vectors in phage display.
A. Phagemid vectors contain g3 or g8 (gene encoding either pIII or pVIII proteins, respectively), followed by a secretion signal. Helper phage rescue provides the rest of the phage proteins that are needed leading to monovalent pIII display or polyvalent pVIII display (but less than 2000 copies). **B.** Phage vectors contain all the phage genes and display gene is followed by a secretion signal. The addition of helper phage is not needed; this leads to polyvalent pIII display (5 copies) and pVIII (>2000 copies) (Lowman, 2013).

When using a phagemid system, helper phage is needed to provide the rest of the essential genes for assembly and replication that the phagemid lacks. Helper phage also have a wild type copy of the coat protein that is used for the display and so both wild type and ligand fusion are packaged into the surface of phage particles. In terms of the DNA packaging into the phage particles, the helper phage genome's intergenic region is mutated to make its packaging less efficient. In this way only one phagemid is packaged inside each phage particle, linking the genotype with the phenotype. The advantage of using phagemid vectors is that they display fewer ligands per phage particle, and this helps for the display of larger molecules such as large peptides or Ab fragments that can

interfere with the biological function of the phage coat protein unless a wild type version is displayed alongside the fusion. Phagemid antibody libraries usually display Ab fragments such as single chain variable region fragments (scFvs), fragment antigen-binding (Fabs), camelid heavy chain-only antibody fragments (known as VHHs) or shark derived antibody fragments (IgNar) (Dooley et al., 2003; Nissim et al., 1994). Libraries can be either naïve (constructed from B cells that have not been exposed to any antigen), immunised (constructed from B cells from individuals that have been immunised with a specific antigen) or synthetic (constructed *in vitro* using randomised primers and PCR, creating artificial potential antigen binding sites). Peptide libraries can be displayed on either the pIII and pVIII coat proteins. pVIII display provides polyvalence due to the increased amount of gene copies but only the insertion of short peptides can be tolerated; pIII display provides monovalency, but access to the displayed protein can be limited (Russel et al., 2004).

The three main steps of phage display methods are library construction, biopanning and screening (Figure 1.2.3). Genes encoding antibody fragments or randomised peptides are cloned into the phagemid vector to create a library of peptide-coat protein fusions. Once the library is constructed, phage particles are produced with the addition of helper phage that provides the rest of the essential genes. Phage displaying peptides or antibody fragments are panned against potential targets that are usually immobilised on a solid surface. Phage displaying specific Abs or peptides will remain bound to the target after

intensive washing when the majority of target non-specific phage will be washed away. Selected phage will be eluted, usually with a pH shift. This procedure, called panning, is usually repeated for several iterative rounds. After each round, eluted phage would be used to infect bacteria and bacteria grown on plates before used to produce a sub-library of enriched phage. After the final round of panning, some individual bacterial clones (usually a few hundred) would be screened with various binding assays, usually ELISA (enzyme-linked immunosorbent assay). This is followed by Sanger sequencing to identify unique positive binding clones. Phage display involves ''finding the needle in the haystack'', rare clones that are propagated and enriched through the panning rounds (Carmen and Jermutus, 2002; Wu et al., 2016; Leow et al., 2017; Hammers and Stanley, 2014; Hay and Lithgow, 2019).

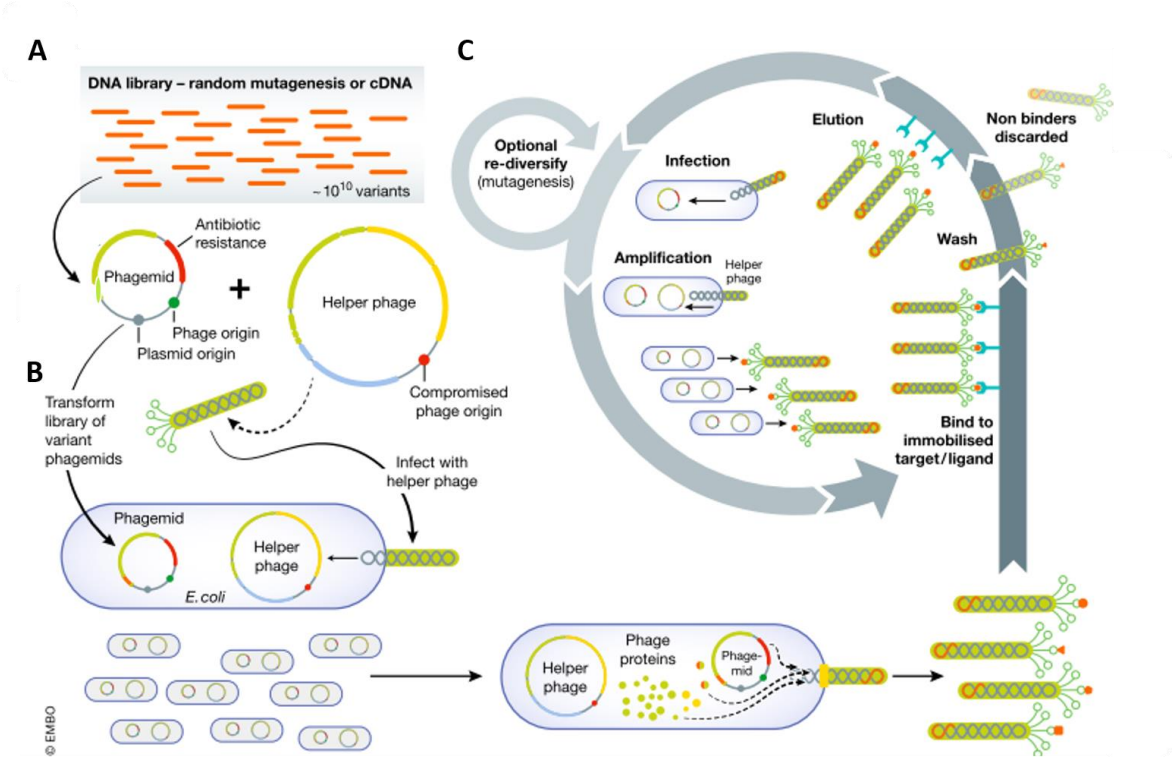


Figure 1.2.3. Schematic representation of phage display using a phagemid vector.
A. Library construction: Genes encoding peptides or antibody fragments derived either from cDNA synthesis or from *in vitro* synthesis are cloned into the phagemid vector **B.** Library production: DNA library is transformed into bacteria and phage particles are produced with the addition of helper phage. **C.** Biopanning: Virions are panned against immobilised or non-targets. Target non-specific bacteriophages are washed away and the specific ones are eluted and reintroduced to bacteria for amplification for iterative biopanning rounds, adapted from (Hay and Lithgow, 2019).

There are numerous different ways for library construction (Table 1.2.1). The two main approaches are either by focused mutagenesis (e.g. site saturation mutagenesis by using randomised codons NNK) or by random mutagenesis (by using error-prone PCR) (Krištof Bozovicar and Tomaž Bratkovič, 2020).

Table 1.2.1 Different ways for antibody and/or peptide library construction either by focused or random mutagenesis (Krištof Bozovicar and Tomaž Bratkoviˇ, 2020).

Focused mutagenesis	Random mutagenesis
Oligonucleotide-directed mutagenesis	Mutagenesis plasmids
Kunkel mutagenesis	Error-prone PCR
Overlap extension PCR	Sequence saturation mutagenesis
Gibson assembly	Isothermal rolling circle amplification under error-prone conditions
Alanine-scanning mutagenesis	DNA shuffling
Cassette mutagenesis	Mutagenic organized recombination process by homologous in vivo grouping
Site-saturation mutagenesis (randomised codons: NNN, NNK, NNS and NNB N= A/C/G/T, K= G/T, S= C/G and B= C/G/T)	In vivo assembly
"Tailored" randomization	
22c-trick with usage of degenerate codons	

1.2.3. Epitope mapping using phage display

When considering epitope mapping of antibodies with phage peptide libraries, the advantage of having a randomised library is the increased probability of these peptides to form mimotopes (a neo-epitope that will resemble the original epitope) in comparison with libraries that are derived from a known protein, which is important when considering non-linear epitopes or those due to PTMs (Ibsen and Daugherty, 2017). The word mimotope was first used to describe a peptide mimicking a discontinuous epitope of a virus, even though the tertiary structure was unknown (Geysen et al., 1986). This is especially relevant for epitope mapping since epitopes (also known as antigenic determinants) can be either linear (amino acid sequence is continuous) or conformational (amino acid sequence is interrupted but close in space). These

two possibilities could be covered in a large diversity peptide library since the estimated length of a linearized epitope is 6 residues (Larman et al., 2011).

Epitope mapping has facilitated the discovery of disease specific epitopes/mimotopes, by panning peptide libraries against serum of patients that suffer from viral infections or autoimmune diseases (Wu et al., 2016). Interestingly, phage display has already been reported to be used in the diagnosis of cancer. Phage libraries constructed from ovarian cancer antigens (total RNA was isolated) were panned against cancer patient's serum and several peptides were discovered representing known proteins (Chatterjee et al., 2006).

Serum contains the circulating antibody repertoire and if it is analysed thoroughly, it can be informative for the past and current status of the immune response. However, serum is very difficult to analyse since it contains a myriad of antibodies that could potentially have more than one binding site. Moreover, multiple antibodies could be generated that bind the same antigen, but to a different epitope. The estimated diversity of B cells that have already been activated against a specific antigen is about 10^6 (Ryvkin et al., 2012). Random peptide libraries have been utilised in the past to explore the whole immune repertoire, using either pIII or pVIII display systems. These synthetic peptide libraries could be panned against polyclonal antibodies and affinity selected against them. There are more than 2000 copies of the pVIII coat protein per phage, and this creates polyvalent display of peptides (~200 copies per virion) which improves the potential of binding due to the avidity effect.

Griffiths et al. suggests that the bigger the phage library is, the higher the chances are of finding an antibody/peptide interaction (Griffiths et al., 1994).

1.2.4. Complimentary/alternative methods to phage display

Another method used for epitope mapping instead of phage display is the magnetic activated cell sorting (MACS); bacteria displaying peptides or Ab fragments are incubated with the biotinylated antigen of interest. Positive clones are then sorted through an anti-biotin magnetic column (Salema and Fernández, 2017). Fluorescent activated cell sorting (FACS) is another alternative, too. Virions are very small particles that cannot be detected with FACS. Whilst bacterial or yeast display are methods compatible with FACS the diversity of libraries that can be displayed on bacteria and FACS sorted is limited (~1 million). It has been reported that even one round of phage biopanning, followed by FACS sorting using the sub library of phage ligands recloned to be displayed on bacteria or yeast, increases the chances of finding high affinity clones. This couples the very high diversity phage display library in round 1 of selection with the very high selectivity of sorting in round 2 once the diversity has been reduced to a level compatible with bacterial/yeast surface display. It has also been demonstrated that repeated FACS ''rounds'' amplifies the positive clones further (Mazor et al., 2008, 2010). Periplasmic Expression with Cytometric Screening (PECS) has also been reported to be used instead of phage display. A bacterial system that expresses libraries into the periplasmic space of *E. coli* followed by fluorescent probe labelling after target probing. This method is amenable for less diverse libraries (~10^6 clones).

Unbound fluorescent conjugate is defused from the outer membrane of the bacterial cell. There is no need for subcloning a library designed for phage display for use in PECS due to the fact that pIII fusions are anchored onto the inner membrane by the C-terminal before their incorporation onto phage (Chen et al., 2001). Anchored Periplasmic Expression (APEx) could potentially be another phage display alternative. Antibody fragments are fused with either *E. coli* lipoprotein NlpA on the N-terminal or with pIII minor coat phage M13 protein on the C-terminus so they are anchored to the inner membrane of Gram negative bacteria. Cells are permeabilised chemically with the addition of lysozyme and the chelator EDTA followed by antigen probing and conventional FACS staining. This method has only been reported with antibody fragment libraries and not with random peptide libraries (Harvey et al., 2004). A drawback of APEx is that bacterial cells without their outer membrane (spheroplasts) are no longer viable so in order to recover the positive clones, extra manipulation is needed by PCR amplification of the ligand genes or recovery of the expression vector (Salema and Fernández, 2017).

1.3. Next Generation Sequencing (NGS)

Recently, next generation sequencing (NGS) platforms have been made available for mass sequencing analysis and accelerate the number of reads and made sequencing faster, easier and at much less cost than the available technologies at the time. One of the sequencing technologies, is Ion Torrent (licenced by Life Technologies). Preparation of NGS libraries is similar across the different technologies, incorporating adapters flanking the DNA of interest,

allowing bridge amplification on a solid phase (van Dijk et al., 2014). The incorporation of barcoded primers is employed so that in one sequencing sample, many experiments can be analysed to further reduce the costs (Matochko et al., 2012).

1.3.1. Next generation phage display

The output of a phage panning was traditionally screened by immunoassays and/or direct Sanger Sequencing of 10^2 or 10^3 random clones. With the incorporation of NGS, the number of clones that can be interrogated is increased by 10^4 fold (Rouet et al., 2018). This is a significant advantage combining these two techniques. The first reported use of NGS with phage display was reported by E. Dias-Neto *et al.* (Dias-Neto et al., 2009). Usually, the diversity of the library is reduced to approximately 10^6 after the first round of panning, making it feasible to be analysed by NGS since the enrichment even after one round of selection can be sufficient to enrich for binders above background phage. The insight that could be achieved with NGS is unprecedented since the enriched clones could be quantitatively measured, their copy number, and a specific amino acid ''trend'' (consensus sequence) among positive clones could also be identified. The reads that are usually obtained by Ion Torrent Sequencing are up to 400 bp, enough to cover the variable heavy or light region of an antibody or an entire peptide. In naïve libraries, the sequencing depth of NGS could be informative on the actual diversity of the library and any bias it may contain. This data on the naïve library can serve as quality control, since overrepresented clones, clones with

frame shift or containing stop codons could be assessed before the beginning of biopanning to eliminate them from further analysis after panning (Glanville et al., 2015). By comparison, conventional sequencing of clones screened in binding assays will be incomplete due to the small number of clones analysed compared to the diversity present, especially at round one (Zhang et al., 2011). A common problem in phage display is that some clones preferentially proliferate in bacteria or bind to the solid phase or blocker used in panning. The incorporation of NGS in phage display can overcome these problems by sequencing all binders and ranking them based on their frequency or enrichment against the target compared to a control (Christiansen et al., 2015). Therefore, NGS can facilitate a more thorough approach to screen for potential target specific clones than conventional methods.

1.4. Hypothesis and Aims

The hypothesis being tested in this study is that RCC generates a range of autoantibodies and that Next Generation Phage Display can reveal large panels of mimotopes recognised specifically by these autoantibodies. Enriched peptides mimicking the epitopes of the TAAs will then be used to develop a sero-diagnostics for this particular type of cancer.

Aims:

- Construction of a synthetic randomised 16mer peptide library displayed on the pVIII coat phage protein in order to represent a large TAA repertoire.

- Optimisation of the epitope mapping technique using known monoclonal antibodies and its application in sera spiked with known mAbs (mimicking the polyclonal response).
- Optimisation of screening serum-based assays and the development of alternative assays to conventional using the sum of frequencies of peptide mimotopes (immunosignature panels).
- Application of the optimised epitope mapping technique to sera from TRACK (transgenic model of cancer of the kidney) mice, kindly provided by Weill Cornell Medical College in order to identify individual peptides that can distinguish between wild-type and TRACK sera.
- Application of the optimized NGDP in sera derived from healthy and RCC patients, kindly provided by Leeds biobank NIHR in order to identify either individual human RCC specific peptides or elucidate the RCC immunosignature.

2. Material and Methods

2.1. Material

- Glucose Solution 25%: 25 g of glucose (Fisher Scientific) was dissolved in 100 ml of autoclaved water and then filter sterilized with a 0.22 µm sterile filter unit (Merck Millipore).
- 2YT Media: 16 g of tryptone (VWR chemicals), 10 g of yeast extract (Merck Millipore) and 5 g of NaCl (Fisher Scientific) was dissolved in 1 L of water. It was then sterilized by autoclaving.
- 2YT Agar: 15 g of agar (VWR chemicals), 16 g of tryptone (VWR chemicals), 10 g of yeast extract (Merck Millipore) and 5 g of NaCl (Fisher Scientific) and dissolving it in 1 L of water. It was then sterilized by autoclaving.
- Washing buffers:
    - 1xPBS (Phosphate buffered saline) was made by dissolving 1 tablet of pre-made PBS (VWR chemicals) in 500 ml DEPC water (Gibco). Solution was then sterilized by autoclaving.
    - PBS-500 mM DTT was prepared by adding 38.625 g of Dithiothreitol (DTT) (Fisher Scientific) to 0.5 L of autoclaved 1x PBS solution.
    - 0.1% PBST was prepared by adding 1 ml of Tween-20 (MP Biomedicals) to 1 L of autoclaved 1×PBS solution.
    - 2% PBST was prepared by adding 20 ml of Tween-20 to 1 L of autoclaved 1×PBS solution.

- 2% PBST + 1M NaCl (Fisher Scientific) was prepared by adding 58.44 g of NaCl (Fisher Scientific) to 1 L of 2% PBST solution.

- Blocking buffers:
  - 3% PBSM was prepared by adding 3 g instant dried skimmed milk (Co-op) to 100 ml of autoclaved 1×PBS solution.
  - 18% PBSM was prepared by adding 1.8 g instant dried skimmed milk (Co-op) to 10 ml of autoclaved 1×PBS solution.
- Antibiotics: Ampicillin (Applichem Lifescience) and kanamycin antibiotic (Alfa-Aesar) stock solutions were made with autoclaved water to final concentrations of 150 mg/ml and 50 mg/ml, respectively. They were filtered using a 0.22 µm sterile filter unit (Merck Millipore), aliquoted and stored at -20°C.
- Tris base buffer 1M was prepared by dissolving 121 g of Tris base (Fisher Scientific) in 1 L of DEPC water (pH = 7.5). It was then sterilized by autoclaving.
- EDTA buffer 0.5M was made by adding 18.61 g of EDTA (Ethylenediaminetetraacetic acid; Fisher Scientific) to 100 mL of DEPC water. It was then sterilized by autoclaving.
- NaCl (Fisher Scientific) 1M: NaCl (Fisher Scientific) buffer was prepared by adding 58.44 g of NaCl (Fisher Scientific) to 1L of autoclaved DEPC water.
- TES buffer: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 100 mM NaCl (Fisher Scientific). TES buffer was prepared by adding 50 µl of Tris-base 1M, 10 µl of EDTA 0.5M and 500 µl of NaCl (Fisher Scientific) 1M to

4.440 ml of autoclaved DEPC H20 (final concentration: 10 mM Tris-HCl pH 7.5), 1 mM EDTA, and 100 mM NaCl.

- Elution reagents:
    - Triethylamine 100 mM: Solution was prepared by adding 700 ml of 7.18M stock solution (Fisher Scientific) to 50 ml of DEPC water.
    - Glycine 0.2M (pH 2.2): Solution was prepared by adding 15.01 g glycine (Fisher Scientific) in 1 L autoclaved water.
    - 50 mM DTT: Solution was prepared by adding 7.7 mg of DTT to 1 mL of autoclaved DEPC water.
    - Papain (Sigma-Aldrich) solution was prepared at final concentration 10 mg/ml and by adding 0.04M EDTA and 0.04M L-cysteine (Merck Millipore).
- TAE buffer (Tris-acetate-EDTA), 50x: 57.1 ml glacial acid, 100 ml of 0.5 M EDTA and 242 g of Tris base was added in a final volume of 1 L of water.
- TBS buffer (Tris buffered saline) 10x: 60.6 g Tris base and 87.6 g NaCl (Fisher Scientific) was added to a final volume of 1 L of DEPC water. 0.1% TBST was prepared by adding 1 ml of Tween-20 (MP Biomedicals) to 1 L of autoclaved 1×TBS solution. 1%-3% TBSTM was prepared by adding 1 or 3 g instant dried skimmed milk (Co-op) to 100 ml of autoclaved 0.1% TBST solution.

- M9 salts solution was prepared by adding 32 g of $Na_2HPO_4.7H_2O$ (Fisher Scientific), 7.5 g of KH_2PO_4 (Sigma-Aldrich), 2.5 g of NH_4Cl (Fisher Scientific) and 1.25 g NaCl (Fisher Scientific) to 500 ml of DEPC water.
- $MgCl_2$ 20% solution was prepared by adding 10 g to 50 ml of DEPC water; solution was sterilized by autoclaving.
- Thiamine Hydrochloride was prepared by adding 100mg of thiamine HCl to 10 ml of DEPC water.
- TG1 minimal agar was prepared in order to select only F' pilus (+) TG1 cells (necessary for bacteriophage infection) by adding 1.5 g agar in 75 ml of DEPC water; solution was sterilized by autoclaving. 25ml of M9 salts (5x), 1.6 ml of 25% glucose, 100 µl of $MgCl_2$ and 50 µl of Thiamine Hydrochloride were added to the melted agar and the solution poured into petri dishes.
- Urea buffer (8M) was prepared by adding 48 g of urea (Fisher Scientific) in 100 ml of autoclaved DEPC water.
- Polyethylene glycol (PEG) buffer was prepared by adding 200 g of PEG8000 (Fisher Scientific) and 146.1 g of NaCl (Fisher Scientific) in 1 L of water. It was then sterilized by autoclaving.
- Coomassie Blue was prepared by adding 2 g Coomasie blue R-250 (ThermoFisher Scientific) in a solution containing 45% (v/v) Methanol, 45% (v/v) DEPC water and 10% (v/v) glacial acetic acid.
- Destaining buffer was prepared in a solution with 50% (v/v) DEPC water, 40% (v/v) Methanol and 10% (v/v) glacial acetic acid.
- Immunoassay substrates:

- Insoluble: SIGMA-FAST™ BCIP/NBT tablet
- Soluble: SIGMA-FAST™ p-Nitrophenyl phosphate Tablets (Sigma-Aldrich)
- 3,3',5,5'-Tetramethylbenzidine (TMB, ThermoFisher Scientific)

- IPTG (isopropyl-b-d-thiogalactopyranoside) (VWR chemicals)
- NBF fixation buffer (ThermoFisher)
- Ready-Lyse™ Lysozyme Solution (Lucigen)
- UltraPure™ DNase/RNase-Free Distilled Water 500mL (Invitrogen)
- TG1 Electroporation-Competent Cells (Agilent Technologies)
- SYBR Green qPCR Master Mix; PowerUp SYBR Green Master Mix; Sample Type: DNA; 2 x 1mL; 200 Reactions (ABI Applied Biosystems)
- Super G blocking buffer (Grace Bio-Labs)
- NAP blocker (Non animal protein) (G-Biosciences)
- Ampure (Beckman coulter)
- Antibodies that have been used in immunoassays:
  - Secondary; Goat anti-Mouse IgG (H+L)-Alexa Fluor 488 conjugate (ThermoFisher)
  - Secondary; Rabbit anti-fd Bacteriophage unconjugated antibody (Sigma-Aldrich)
  - Secondary; Goat anti-Rabbit IgG (H+L)-Alkaline phosphatase conjugate (ThermoFisher)
  - Secondary; Goat Anti mouse IgG-Alkaline phosphatase conjugate (ThermoFisher)

- Secondary; Goat anti mouse-IgG-HRP conjugate (ThermoFisher)
- Secondary; Goat anti-Mouse IgG (H+L)-DyLight® 650 conjugate (Invitrogen).
- Secondary; Rabbit anti-human IgG unconjugated (Sigma-Aldrich)
- Secondary; Rabbit anti-M13 antibody- Horseradish Peroxidase conjugate (Sigma-Aldrich)

2.2. Methods

2.2.1. Gel electrophoresis

1 g (for >1 kb DNA fragments) or 3 g of agarose (for <1 kb DNA fragments) (Fisher Scientific) was added to 100 ml of 1x TAE buffer. The solution was microwaved for approximately 2 min until it was clear. 3 µl of Nancy (Sigma-Aldrich) was added to the solution and it was poured into the gel tray with well combs already present. 6x loading dye (NEB) was added to the samples and then they were loaded into the wells, alongside a 1 kb molecular weight marker (NEB). Gels were running for 45 min at 80 V.

2.2.2. Qubit

DNA concentration was determined by Qubit™ dsDNA BR Assay Kit (ThermoFisher Scientific) using a Qubit 2.0 Fluorometer 2.0 (Invitrogen, UK). Qubit working solution was prepared by diluting the BR Reagent 1 in 200 in the BR buffer. 10 µl of the standards (0 ng and 100 ng) were added to 190 µl of working solution. 2 µl of the samples were added to 198 µl of working solution.

2.2.3. Site directed mutagenesis – Cloning of desired peptide sequences

A full description of the pc89 vector map that was used and its features are shown (Fig 2.2.1.)(Felici et al., 1991). Briefly, the M13 origin of replication (ori) is used for single stranded DNA replication (virus DNA) and the plasmid ori for its replication (double stranded DNA). Flanking sequences of the inserts (in this case 16mer peptides) are stated as flanking peptides 5' (left, AEGEF) and flanking peptide 3' (right, DPAKA) motifs, respectively, and the gene encoding pVIII is marked as red. Lac operon components (responsible for RNA transcription) are present before the signal peptide (responsible for protein translocation). The online reverse translation tool that was used was http://www.bioinformatics.org/sms2/rev_trans.html and the option for the most likely used codons in *E. coli* was chosen for the design of primers for clone rescue. Plasmid template was amplified using primers 1-2 for SDM and primers 3-158 for cloning of desired peptide sequences (Table 2.2.1.). PCR reaction was prepared by mixing 25.5 µl DEPC water, 10 µl Q5 buffer, 10 µl GC enhancer, 1 µl dNTPs, 1 µl of forward primer (10 nM/ml), 1 µl of reverse primer (10 nM/ml), 0.5 µl of Q5 enzyme and 1 µl of 10 ng template. Reaction was incubated in a thermocycler at 95°C for 5 min, [95°C for 30s, 55°C for 50s, 72°C for 5 min] x 30, 72°C for 10 min. 10 µl of PCR products were loaded into a 1% (w/v) agarose gel and after electrophoresis separation of bands the band of the correct size was excised and cleaned up using a Nucleospin kit (Macherey Nagel); protocol was followed according to manufacturer instructions. The concentration of DNA preparation was measured with a NanoDrop™ ND-8000

spectrophotometer (ThermoFisher) at a wavelength of 260 nm and converted to DNA concentration. Digestion with the restriction enzyme DpnI (New England Biolabs) followed in order to remove methylated DNA, hence to eliminate the DNA template. DNA was digested according to NEB recommendations for 2h at 37°C and then 20 min at 80°C for inactivation of the enzyme. DNA was purified using a Nucleospin kit (Macherey Nagel) according to manufacturer's instructions. For non-phosphorylated primers, an extra step of phosphorylation was conducted. 1 µg of DNA was phosphorylated with AP-T4 ligase according to NEB recommendations for 30 min at 37°C and then 20 min at 80°C for inactivation of the enzyme.

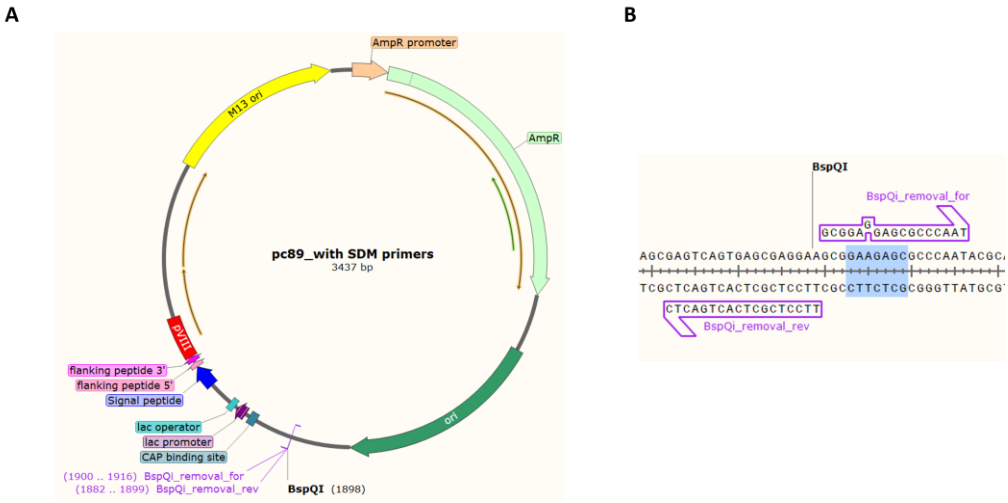


Figure 2.2.1 The pc89 vector map and SDM strategy.
A. Important features are highlighted by different colours in the pc89 map. AmpR is the gene encoding ampicillin resistance. Left (AEGEF) and right (DPAKA) motifs are the flanking regions of the library insert site that contains a TAG stop codon, present within the pVIII gene and after the signal peptide sequence. M13 origin of replication (ori) and plasmid ori are present, indicated. The size of the pc89 vector is stated (3437 bp). Image was obtained with SnapGene viewer. **B.** Inverse PCR primer design strategy for site directed mutagenesis of codon GAA to GAG in pc89 vector to remove the indicated BspQI cleavage site.

Table 2.2.1 Primer sequences used for SDM (1-2) and clone rescue (5-158).

Primer ID	DNA sequence	Usage
1	GCGGAGGAGCGCCCAAT	SDM
2	TTCCTCGCTCACTGACTC	SDM
3	CGCCTGCCGAACGCGCCGGTGACCGATCCCGCAAAAGCG	Clone rescue
4	GCGGATAGCAGCAGCATTATGGTGGATCCCGCAAAAGCG	Clone rescue
5	AACTATGAAAAACTGACCCTGGCGGATCCCGCAAAAGCG	Clone rescue
6	CATAACCCGACCCGCACCAAAACCGATCCCGCAAAAGCG	Clone rescue
7	CAGAACGAACTGCGCAACAGCGCGGATCCCGCAAAAGCG	Clone rescue
8	GCGATGGGCATTTATGAAGCGCCGGATCCCGCAAAAGCG	Clone rescue
9	GTGCAGGAACAGACCACCAACAGCGATCCCGCAAAAGCG	Clone rescue
10	GCGCAGATCGCGGGTGCTGCTCACGAATTCACCCTCAGCAGC	Clone rescue
11	GTTGTTCAGGCTGTTCATGGTGCTGAATTCACCCTCAGCAGC	Clone rescue
12	CATCATCATCAGGCGCATGCCCGGGAATTCACCCTCAGCAGC	Clone rescue
13	TTTCAGATCGTTCGGGCTCAGATAGAATTCACCCTCAGCAGC	Clone rescue
14	CACCAGGTTGCGGGTGCTGCTGCTGAATTCACCCTCAGCAGC	Clone rescue
15	ATCCAGCACAATGCGCGCGCCCGGGAATTCACCCTCAGCAGC	Clone rescue
16	AATCTGCATGCCCGGGCGGCTGGTGAATTCACCCTCAGCAGC	Clone rescue
17	GCGCAGCCGACCATTGAACTGACCGATCCCGCAAAAGCG	Clone rescue
18	CCGCCGCCGAACCCGAGCAACAGCGATCCCGCAAAAGCG	Clone rescue
19	CGCCTGGCGATGAAACCGTATGTGGATCCCGCAAAAGCG	Clone rescue
20	GAAGCGCCGGTGCATCCGACCACCGATCCCGCAAAAGCG	Clone rescue
21	CCGAGCGCGATTAGCGGCCCGGTGGATCCCGCAAAAGCG	Clone rescue
22	TGGCGCACCTATCAGGAAAAAGCGGATCCCGCAAAAGCG	Clone rescue
23	GAAGGCCATCAGCCGGCGCATCGCGATCCCGCAAAAGCG	Clone rescue
24	CTGAGCACCAACTTTCCGATTGGCGATCCCGCAAAAGCG	Clone rescue
25	AGCAACGGCAGCACCAGCCATGCGGATCCCGCAAAAGCG	Clone rescue
26	ATTCAGATGACCGCGCGCCCGCAGGATCCCGCAAAAGCG	Clone rescue
27	CAGATCGCGCGCCGCGCCCCACGGGAATTCACCCTCAGCAGC	Clone rescue
28	GCTCAGGGTCGGTTTATGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
29	CGGTTCGCCGCGGCGCGGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
30	ATAGTTCTGGCTGCGCGGCAGCATGAATTCACCCTCAGCAGC	Clone rescue
31	ATGCCAGCCGCTGCGTTTGCCAATGAATTCACCCTCAGCAGC	Clone rescue
32	GCCGCTGCTCGGGTTCTGCTGTTTGAATTCACCCTCAGCAGC	Clone rescue
33	CTGCAGGGTCTGGCTGCGCGGCAGGAATTCACCCTCAGCAGC	Clone rescue
34	TTCCAGGTTCGGATACACCTGGCGGAATTCACCCTCAGCAGC	Clone rescue
35	CACCTGGCTGCGCGGATGCAGCAGGAATTCACCCTCAGCAGC	Clone rescue
36	CCAGCCATGCTGCATGCTCAGCTGGAATTCACCCTCAGCAGC	Clone rescue
37	CGCCCGAGCATTCCGCCGATTGGCGATCCCGCAAAAGCG	Clone rescue
38	ATGTTTGAAGCGCTGAAAAGCAAAGATCCCGCAAAAGCG	Clone rescue
39	GTGTATGATCAGACCCGCAGCCATGATCCCGCAAAAGCG	Clone rescue
40	CTGGAATATGCGGCGCGCTATACCGATCCCGCAAAAGCG	Clone rescue
41	GATAGCTATCATACCCAGATTACCGATCCCGCAAAAGCG	Clone rescue
42	GCGGCGCGCAAAGTGGCGTGCGCGGATCCCGCAAAAGCG	Clone rescue
43	AGCGGCCTGAAACCGTATCTGTGCGATCCCGCAAAAGCG	Clone rescue
44	AGCGGCCTGGATCCGGCGGTGGTGGATCCCGCAAAAGCG	Clone rescue
45	CAGCAGGCGACCGGCCTGCGCGCGGATCCCGCAAAAGCG	Clone rescue
46	CGCGCGGGCATTGTGACCTGCCAGGATCCCGCAAAAGCG	Clone rescue
47	ATGTTTGAAGCGCTGAAAAGCAAAGATCCCGCAAAAGCG	Clone rescue
48	ATGTGCCCGACCGGCCTGAGCAAAGATCCCGCAAAAGCG	Clone rescue

49	ACCCCGCCGACCACCGTGCCGGCGGATCCCGCAAAAGCG	Clone rescue
50	TGCAGCTTTCTGCAGACCCATAGCGATCCCGCAAAAGCG	Clone rescue
51	TATGCGGGCATTACCGATAGCCCGGATCCCGCAAAAGCG	Clone rescue
52	GCGAGCGATAACCCGATGCGCAACGATCCCGCAAAAGCG	Clone rescue
53	ACCGGCCATGATACCGATTGGAACGATCCCGCAAAAGCG	Clone rescue
54	ACCCCGCTGGAATTTCAGCAGCGCGATCCCGCAAAAGCG	Clone rescue
55	ACCAACCCGACCTGGACCCGCTATGATCCCGCAAAAGCG	Clone rescue
56	AGCACCCCGACCGTGCATGAAACCGATCCCGCAAAAGCG	Clone rescue
57	CTGCTGCCGCGCAGCCCGCAGACCGATCCCGCAAAAGCG	Clone rescue
58	AAACCGAGCAAACGCGGCATTCGCGATCCCGCAAAAGCG	Clone rescue
59	AAACTGCTGAAAGCGGCGCATAAAGATCCCGCAAAAGCG	Clone rescue
60	CCGCGCCGCCAGCTGGTGCGCAAAGATCCCGCAAAAGCG	Clone rescue
61	CGCCGCGGCGAACAGCATGAACGCGATCCCGCAAAAGCG	Clone rescue
62	TGGGAAGATCTGCAGCTGAGCGGCGATCCCGCAAAAGCG	Clone rescue
63	AGCAGCCGCCCGCCGCTGATGAGCGATCCCGCAAAAGCG	Clone rescue
64	CGCCTGCCGACCCAGCATAGCGATGATCCCGCAAAAGCG	Clone rescue
65	GCGAAAAGCACCAAACGCCAGTATGATCCCGCAAAAGCG	Clone rescue
66	CGCCGCGGCAGCCAGAAAATGAGCGATCCCGCAAAAGCG	Clone rescue
67	GCGCGCGATGGCAGCCTGTTTAAAGATCCCGCAAAAGCG	Clone rescue
68	GCGACCAGCACCCCGCCGGCGAACGATCCCGCAAAAGCG	Clone rescue
69	AACAAAACCCATGAAGCGGTGACCGATCCCGCAAAAGCG	Clone rescue
70	GAAGTGTTTGCGAACCTGAGCACCGATCCCGCAAAAGCG	Clone rescue
71	AGCACCAGCACCCTGCAGCGCACCGATCCCGCAAAAGCG	Clone rescue
72	TGCCGCGCGAAAAGCAGCGATTGCGATCCCGCAAAAGCG	Clone rescue
73	CTGGAACGCAACAAACGCCGCACCGATCCCGCAAAAGCG	Clone rescue
74	GTGGCGCCGACCCAGCATGCGCTGGATCCCGCAAAAGCG	Clone rescue
75	GATCCGGGCCTGGTGCCGTATTGCGATCCCGCAAAAGCG	Clone rescue
76	GGCGATCCGCGCGCGAACGTGACCGATCCCGCAAAAGCG	Clone rescue
77	GCGACCCGCGTGAGCAAAGTGCATGATCCCGCAAAAGCG	Clone rescue
78	CAGCGCCCGCGCAGCACCAAAAGCGATCCCGCAAAAGCG	Clone rescue
79	CGGGCCCGCTTTCAGCTGGTTCAGGAATTCACCCTCAGCAGC	Clone rescue
80	CTGCACGCGCGCTTTAAACACGTTGAATTCACCCTCAGCAGC	Clone rescue
81	GCTATAGCGATGCAGCAGCACGTTGAATTCACCCTCAGCAGC	Clone rescue
82	GCGATATTCCAGCATGGTCTGGCGGAATTCACCCTCAGCAGC	Clone rescue
83	GCTTTTCACCGGCAGCATGCCGCTGAATTCACCCTCAGCAGC	Clone rescue
84	GCAGGTGCGCATATGCAGCATGCTGAATTCACCCTCAGCAGC	Clone rescue
85	CGCCATCGCGGTGGTGCGATGGCAGAATTCACCCTCAGCAGC	Clone rescue
86	GCTAATCTGCGGCGCCAGCTGATGGAATTCACCCTCAGCAGC	Clone rescue
87	CACGCGATACGCCAGCATGTTCAGGAATTCACCCTCAGCAGC	Clone rescue
88	GCACGCCGCGCTGCTATACGGGTTGAATTCACCCTCAGCAGC	Clone rescue
89	CTGCACGCGCGCTTTAAACACGTTGAATTCACCCTCAGCAGC	Clone rescue
90	GCCGCTGCTATAGCAGCTCGCCGGGAATTCACCCTCAGCAGC	Clone rescue
91	TTTGCTATACGGGCGCATGCCGCTGAATTCACCCTCAGCAGC	Clone rescue
92	CATGCCGCTCAGCGCGCACGGGCTGAATTCACCCTCAGCAGC	Clone rescue
93	ATGGCGGCGCATCTGCAGCACGGTGAATTCACCCTCAGCAGC	Clone rescue
94	GGTATGCAGCCATTCGGTCGGCACGAATTCACCCTCAGCAGC	Clone rescue
95	ATGCGGCGCCGCCTGAAACATCGCGAATTCACCCTCAGCAGC	Clone rescue
96	CGGGCTGCGCAGCGGCGGGCTAAAGAATTCACCCTCAGCAGC	Clone rescue
97	GCGCTGATCGCGATGTTTGGTGTTGAATTCACCCTCAGCAGC	Clone rescue
98	TTTGCGCAGCGGCGGGCCGCCCGGGAATTCACCCTCAGCAGC	Clone rescue
99	CGGGCACGGATCGCGGCGCGGGCGGAATTCACCCTCAGCAGC	Clone rescue
100	TTTCAGGGTCTGCGGAATGCTGGTGAATTCACCCTCAGCAGC	Clone rescue
101	CGCGCGTTCCAGCAGCGCCAGCGCGAATTCACCCTCAGCAGC	Clone rescue
102	CAGGCCCTGGCGGCGGCGATAGCCGAATTCACCCTCAGCAGC	Clone rescue

103	ATGGCCATCGCCCGGCGGGCGATGGAATTCACCCTCAGCAGC	Clone rescue
104	CATATCGCCCGGATGGCGCGCCAGGAATTCACCCTCAGCAGC	Clone rescue
105	TTCAATGGTCGCAATCGCGCGCACGAATTCACCCTCAGCAGC	Clone rescue
106	GGTCAGCGGTTTCTGCGCGCTCACGAATTCACCCTCAGCAGC	Clone rescue
107	ATGCATCAGGCTCGCGCGCTGATAGAATTCACCCTCAGCAGC	Clone rescue
108	ATCGCACGCCTGGGTCGCAAACGGGAATTCACCCTCAGCAGC	Clone rescue
109	TTTATCGCGCATGGTGCTTTTCGGGAATTCACCCTCAGCAGC	Clone rescue
110	CAGGGTGCGCACATAGCGCGGCGGGAATTCACCCTCAGCAGC	Clone rescue
111	CGGGGTGGTGGTCATCGCGCGCGGGAATTCACCCTCAGCAGC	Clone rescue
112	TTCCAGCGGGTTCGCATGGGTCGGGAATTCACCCTCAGCAGC	Clone rescue
113	GTTCGGCTGCAGCACAATGGTCTGGAATTCACCCTCAGCAGC	Clone rescue
114	ATACAGAAACAGATGGCTGGTCTGGAATTCACCCTCAGCAGC	Clone rescue
115	TTTAATCGCGCCTTTCTGTTTGCGGAATTCACCCTCAGCAGC	Clone rescue
116	CGGCTGGCGCGGCATCGGCTGGCTGAATTCACCCTCAGCAGC	Clone rescue
117	GGTCTGCGCCCACGCATCGCTGCTGAATTCACCCTCAGCAGC	Clone rescue
118	CACCAGGCGGCTCAGCCAGCAGGTGAATTCACCCTCAGCAGC	Clone rescue
119	CAGCAGGGTCTGGCTGCTCGGGGTGAATTCACCCTCAGCAGC	Clone rescue
120	CCACACGCGGTTGGTCCACGGGGTGAATTCACCCTCAGCAGC	Clone rescue
121	CCGTATCCGAGCCCGGCGTATGTGGATCCCGCAAAAGCGGCC	Clone rescue
122	CTGCCGAGCCTGGAACTGCTGAAAGATCCCGCAAAAGCGGCC	Clone rescue
123	AACGCGCTGATTTATGATACCGCGGATCCCGCAAAAGCGGCC	Clone rescue
124	ACCACCCCGCTGCCGAACACCCCGGATCCCGCAAAAGCGGCC	Clone rescue
125	GAACATCCGCTGATGTATAGCGCGGATCCCGCAAAAGCGGCC	Clone rescue
126	GCGCCGAGCACCCCGAGCCAGCAGGATCCCGCAAAAGCGGCC	Clone rescue
127	GCGCATCTGACCCATACCGGCCGCGATCCCGCAAAAGCGGCC	Clone rescue
128	GCGAGCCGCACCAAAGATACCGTGGATCCCGCAAAAGCGGCC	Clone rescue
129	CCGAACAGCATTGTGCGCGTGCCGGATCCCGCAAAAGCGGCC	Clone rescue
130	AGCGTGGTGCGCGATATTGCGAACGATCCCGCAAAAGCGGCC	Clone rescue
131	ATTGCGCTGAGCACCCTGTTTGATGATCCCGCAAAAGCGGCC	Clone rescue
132	CTGCCGAACAGCACCAGCGTGGATGATCCCGCAAAAGCGGCC	Clone rescue
133	CTGGCGCCGCTGCATACCGAAGTGGATCCCGCAAAAGCGGCC	Clone rescue
134	GTGCAGCGCGCGGAAGCGAACCGCGATCCCGCAAAAGCGGCC	Clone rescue
135	CCGGTGCGCATTAACTATCTGGATGATCCCGCAAAAGCGGCC	Clone rescue
136	TATCCGAGCCTGGTGCCGTATAGCGATCCCGCAAAAGCGGCC	Clone rescue
137	AGCCGCGTGGATAAAGCGTTTTGCGATCCCGCAAAAGCGGCC	Clone rescue
138	CCGCAGGATACCGTGACCCCGCATGATCCCGCAAAAGCGGCC	Clone rescue
139	GTGGGCGATAAATATAGCGCGAACGATCCCGCAAAAGCGGCC	Clone rescue
140	GCTCAGCGGCGGGGTCGGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
141	GCTCGGCGGATAGCTCGGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
142	GGTGGTCGGCGGGCTCGGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
143	CATCGCCGGATGCAGCGGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
144	GCTAATCGGCATTTTCGGCCACGGGAATTCACCCTCAGCAGC	Clone rescue
145	CAGGGTCGGATAATGATGCGGATAGAATTCACCCTCAGCAGC	Clone rescue
146	ATCGCGGCCCGCCAGGCCCGGATAGAATTCACCCTCAGCAGC	Clone rescue
147	GCCCGCCAGGCTCGGATACGGGGTGAATTCACCCTCAGCAGC	Clone rescue
148	CATCTGGTTCGGATACGGCAGGCTGAATTCACCCTCAGCAGC	Clone rescue
149	GGTCAGCGCCAGGTTCGGATACGGGAATTCACCCTCAGCAGC	Clone rescue
150	GGTATGCAGCTGATGCGGATACGGGAATTCACCCTCAGCAGC	Clone rescue
151	GCCCGGCGCCAGCGCCGGATACGGGAATTCACCCTCAGCAGC	Clone rescue
152	CACCGGATACGGCACCGGTTTCGGGAATTCACCCTCAGCAGC	Clone rescue
153	GGTCGCCAGGTTCGGATAATGCGGGAATTCACCCTCAGCAGC	Clone rescue
154	GGTCGCCATCTGTTTGTTAAACGGGAATTCACCCTCAGCAGC	Clone rescue
155	CGGCACGGTGCTCGGCATCGGATGGAATTCACCCTCAGCAGC	Clone rescue
156	CTGCGCCAGGTTCGGATACGGAAAGAATTCACCCTCAGCAGC	Clone rescue

157	GTTGCGCGCCAGGTTCGGATACGCGAATTCACCCTCAGCAGC	Clone rescue
158	CAGGTTCGGATAGCGGCTGCCGTTGAATTCACCCTCAGCAGC	Clone rescue

2.2.4. Ligation/Transformation

The concentration of DNA preparation was determined by Qubit and then 200 ng of DNA was mixed with 4 µl of ligase buffer and 0.4 µl of T4 ligase (NEB) in a total 40 µl reaction. DNA was ligated at 4°C for 6h, 16° for 6h, 4°C for 6h and followed by 10 min at 65°C for inactivation of the enzyme. All DNA was dialysed for desalting of buffers using V-series membranes (Filter disks, Millipore) and 4 µl of ligated DNA were transformed in TG1 electrocompetent cells (Agilent). Cells were pulsed using a BioRad GeneP Pulsar Xcell Electroporator at 1.8 kV and then recovered for 1 h at 37°C with vigorous shaking in 1 ml of SOC Outgrowth Medium (New England Biolabs). 100 µl of the transformation volume was then plated on 2YT agar plates with 150 ug/ml ampicillin and 1% glucose and incubated overnight at 37°C static.

2.2.5. Library construction

Phagemid was amplified to introduce diversity using degenerate primers (Table 2.2.2.) The schematics of the molecular strategy that was followed is depicted in Figure 2.2.1. The NNK approach was used in order to introduce this diversity which is stated as MNN in the reverse complement sequence. 40 PCR reactions consisted of (each) 10 µl Q5 buffer (NEB), 10 µl GC enhancer, 1 µl dNTPs (1mM), 1 µl of forward primer (10 nM/ml), 1 µl of reverse primer (10 nM/ml), 0.5 µl of Q5 enzyme and 1 µl of 10 ng template in a total volume of 50 µl. The PCR programme was 95°C for 3 min, then 30 cycles of 95°C for 30s,

57.5°C for 30s and 72°C for 2 min and 20 s, followed by an incubation at 72°C for 5 mins. PCR products were extracted from a 1% (w/v) agarose gel and purified using a Nucleospin kit (Macherey Nagel); protocol was followed according to the manufacturer instructions and with an elution step using 20 µl of DEPC water. The concentration of DNA was determined by Qubit. Digestion with the restriction enzymes DpnI and SapI (isoschizomer of BspQI that works on the same buffer as DpnI) (NEB) was carried out for 10 reactions. In each reaction, 1 ug of DNA was digested according to NEB recommendations for 2h at 37°C and then 20 min at 80°C for inactivation of the enzymes. DNA was purified using a Nucleospin kit (Macherey Nagel) according to the manufacturer's instructions. The concentration of DNA was measured with Qubit. 1-2 µg of DNA were ligated with 2 µl of T4 DNA ligase in 200 µl total reaction for 6h at 4°C, 18h at 16°C, 6h at RT and 20 min at 80°C for enzyme heat inactivation. DNA was purified using a Nucleospin kit (Macherey Nagel) according to the manufacturer's instructions followed by the measurement of the DNA concentration with Qubit. 500 ng of DNA were transformed in 50 µl of TG1 electrocompetent cells (Agilent). Cells were pulsed using a BioRad Gene Pulser Xcell Electroporator at 1.8 kV and then recovered for 1 h at 37°C with vigorous shaking in 1 ml of SOC Outgrowth Medium (New England Biolabs). 100 µl of the transformation volume was then plated on 9 cm 2YT agar plates with 150 ug/ml ampicillin and 1% (w/v) glucose, a serial dilution of 10 µl of the transformation volume was also carried out to estimate library size (now called pc89_BspQI$^-$) and all plates incubated overnight at 37°C static; the rest of the volume of the transformation was plated on a bioassay dish with 2YT agar, 150

ug/ml ampicillin and 1% (w/v) glucose. Bacterial cells were scraped the next day and stored in -80°C after the addition of 30-50% (v/v) glycerol. Colonies from the library titration plates were individually picked in 5 ml 2YT media with 150 ug/ml ampicillin and 1% glucose and shaken at 37°C for 16h. Plasmid was extracted following the manufacturer instructions for Qiagen Miniprep kit and sent for Sanger sequencing (Source BioScience).

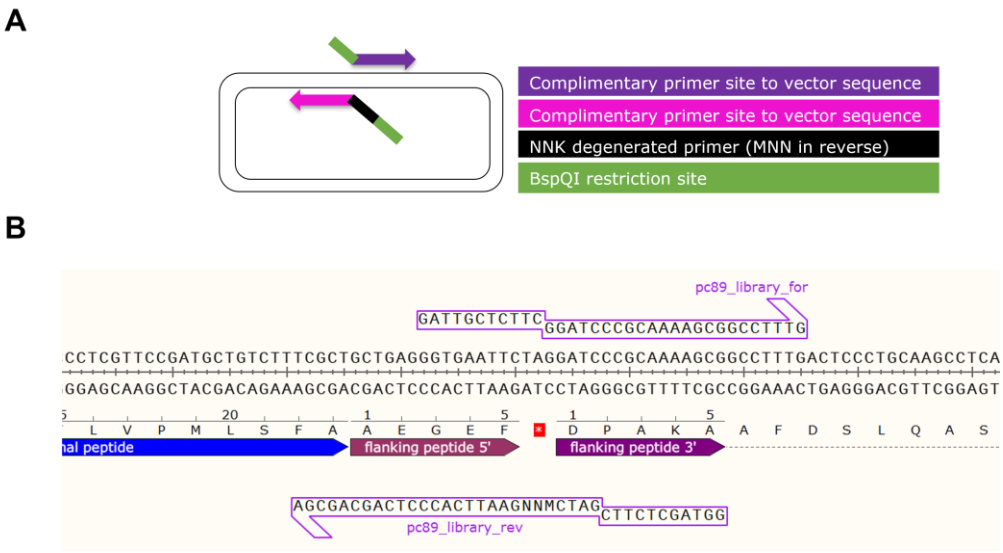


Figure 2.2.2 NNK approach for peptide library.
A. Schematic overview of the NNK randomisation strategy that was adopted in peptide library. **B.** Primers' annealing position for the pVIII display library (produced in SnapGene) N.B MNN is only represented once instead of 16x for presentation purposes.

Table 2.2.2 Primer sequences used for library construction

Primer ID	DNA sequence	Usage
161	GATTGCTCTTCGGATCCCGCAAAAGCGGCCTTTG	pVIII library
162	GGTAGCTCTTCGATC(MNNX16)GAATTCACCCTCAGCAGCGA	pVIII library

2.2.6. Colony lifting

The protocol was adapted from the Merck Millipore website (https://www.merckmillipore.com/GB/en/search/novatope?search=&Trackin

gSearchType=SB++Search+Box&SearchContextPageletUUID=&SearchTerm=novatope, accessed Dec 2017). Transformants were spread at low density (<3000 colonies) on 2YT agar plates containing 150 ug/ml ampicillin and 0.1 mM IPTG (isopropyl-b-d-thiogalactopyranoside) (VWR chemicals) and incubated O/N at 37°C. The next day, nitrocellulose filters were placed on the top of agar plates and were removed after 1 minute; they were placed under the fume hood, in proximity to chloroform for 15 min. Whatman 3MM paper containing denaturation buffer (20 mM Tris-HCl pH 7.9, 6 M urea, 0.5 M NaCl (Fisher Scientific) in excess was placed on the bottom of the filters for 15 min. Filters were then blocked with 3% TBSTM for 1h at RT, followed by washing with TBST for 15 min (x2). Filters were then incubated with primary antibody in 1/2000 dilution in TBST + 3% (w/v) milk for 1h at RT with agitation, followed by washing with TBST for 15 min (x3). Filters were then incubated with anti-mouse AP conjugated antibody (Invitrogen) in 1/2000 dilution in 3% TBSTM for 1h at RT with agitation, followed by washing with TBST for 15 min (x3). SIGMA-FAST™ BCIP/NBT solution was added to the filters for 10 min and they were then washed with DEPC autoclaved water. Positive colonies were identified by their purple colour, and they were further recovered by removing the corresponding region in the agar plates and growing them in 5 ml of 2YT media containing 150 ug/ml ampicillin for 2h at 37°C at 180 rpm. This mix of clones was further spread on a new 2YT agar plate containing 150 ug/ml ampicillin and tested again with the colony lifting technique until single positive colonies were able to be identified and sequenced.

2.2.7. Dotblot

3 µl of PEG precipitated phage was spotted on nitrocellulose membrane, and the membrane was blocked with 1% TBSTM for 1h. Membrane was then incubated with different dilutions of primary antibody (depending on which mAb was used) in 1% TBSTM for 1h at RT agitating, followed by washing with 1% TBSTM for 10 min (x4). Membrane was then incubated with anti-mouse HRP conjugated antibody (Invitrogen) in 1/2000 dilution in 1% TBSTM for 1h at RT with agitation, followed by washing with 1% TBST for 10 min (x4). 1 ml of TMB substrate (ThermoFisher Scientific) was added for 1 min and washed with DEPC autoclaved water. Membrane was read at 420 nm at Biorad Image viewer.

2.2.8. SDS-PAGE analysis of purified IgG from sera

Normal mouse sera (Sigma-Aldrich), spiked or non-spiked with monoclonal antibodies was incubated with protein G beads (Pierce™ ThermoFisher Scientific) in 500 µl of PBS (1x) rotating for 3 h. Beads were washed with 0.1% PBST (x2) and PBS (x2) and then eluted with glycine 0.2 M pH 2.2, followed by neutralisation using Tris buffer. 4x LDS-751 (ThermoFisher) was added to samples in a final volume of 15 µl and denaturated in 95°C for 5 min. Samples were then loaded onto a NuPAGE polyacrylamide gel Bis-Tris 12% (ThermoFisher) prewashed with 1x MOPS solution (G-Bioscience), and run at 200V for 35 min. The gel was then stained for 5 min with Coomassie Blue and destained overnight with destaining buffer.

2.2.9. Anchored Periplasmic Expression (APEx)

The protocol was adapted from Harvey *et al.*(Harvey et al., 2004). In general, a single colony of bacteria containing phagemid vectors or TG1 bacteria only was inoculated in 2YT medium with 150 ug/ml ampicillin and 1% (w/v) glucose and shaken at 37°C until OD_{600} was 0.4-0.6. The bacterial pellet was collected by centrifugation at 5000 g for 15 min and resuspended in the same volume of 2YT medium with 150 ug/ml ampicillin and 1 mM isopropyl-b-d-thiogalactopyranoside (IPTG) (VWR). Cells were shaken at 30°C at 100 rpm for 4h or overnight. Pellet was collected by centrifugation at 5000g for 15 min and resuspended in same volume of PBS and fixed with 10% (v/v) formalin for 20 min at 4°C. Cells were washed with 1x or 6x PBS (2 or 3 times) were collected and resuspended in 25 µl of TES buffer with various amounts of lysozyme (Lucigen) and incubated at RT for 10 min. Cells were washed with 1x or 6x PBS (2 or 3 times) and blocked with various reagents, washed as described before, and then incubated with various dilutions of primary antibody (SAF84 or SAF70) for 45 min at RT. Cells were washed with 1x or 6x PBS (2 or 3 times) and then resuspended in 1 ml of PBS, followed by incubation with one drop of the fluorescent conjugate antibody for 45 min at RT (Alexa 488 anti-mouse-Invitrogen). The pellet was washed with 1x or 6x PBS (2 or 3 times) and was analysed on a Beckman Coulter Astrios EQ Cell Sorter.

2.2.10. NGS preparation

An overview of the PCR strategy for PCR1 and the subsequent PCR2 is described in Fig 4A and B, respectively. The PCR1 primers that were used were

pVIII library specific; they also contained two different linker sequences that contain homologous sequences to primers used in the following amplification round (PCR2). PCR2 forward primer had a complimentary sequence with linker 1, an adapter sequence and a unique DNA barcode (95 barcoded primers were available at the initial stages and 107 at the latest stages) whereas PCR2 reverse primer was universal (P1 primer), being complimentary to linker 2 but also containing a different adapter sequence. Adapter sequences were compatible with the Ion Torrent NGS platform that was being used. DNA from individual output phage sublibraries was purified using a Mini Prep Kit (Qiagen) and were amplified with primers 165 and 166 (Table 2.2.3.). Specifically, 1 µl of the template (10ng) was mixed with 25.5 µl DEPC water, 10 µl Q5 buffer, 10 µl GC enhancer, 1 µl dNTPs, 1 µl of forward primer (10 nM/ml), 1 µl of reverse primer (10 nM/ml) and 0.5 µl of Q5 enzyme. Reaction was then incubated at 95°C for 3 min, [95°C for 30s, 60°C for 30s, 68°C for 30s] x30 and last 72°C for 5 mins. Gel electrophoresis of 10 µl of the PCR1 reaction on a 3% (w/v) agarose gel was carried out to confirm amplification (275 bp). DNA was purified using the Nucleospin kit (Macherey Nagel) according to the manufacturer's instructions. 10 ng of purified DNA was amplified as described before, with a universal reverse primer 167 (P1) for all DNA samples and with forward barcoded primers 168-272 (different for every sample, Table 2.2.3.) with the PCR2 conditions (95°C for 3 min, [95°C for 30s, 63°C for 30s, 68°C for 30s] x12 and last 72°C for 5 mins). Amplification was confirmed by gel electrophoresis (335 bp). The concentration of DNA was measured with a Qubit dsDNA HS Assay Kit (ThermoFisher Scientific) using Qubit 2.0 Fluorometer 2.0

(Invitrogen). All barcoded amplicons were mixed to equal amounts of DNA. The pooled sample was resolved on a 3% (w/v) Metaphor agarose gel alongside a MassRuler™ Low Range DNA ladder and ran for 6h at 50V. The desired DNA band size (335 bp) was extracted and purified using the Nucleospin kit following the manufacturer's instructions. The sample was further purified using the Agencourt AMPure XP Bead Clean-up kit (Beckman coulter) according to the manufacturer's instructions. Samples were sent for sequencing using the commercial service offered by the University of Pennsylvania using the Ion Proton platform with an SS 540 Chip.

Table 2.2.3 Primer sequences for NGS preparation

Primer ID	DNA sequence	Barcode ID
163	GTAATCCTTGTGGTATCGGATGCTGTCTTTCGCTGC	R1
164	CTAGAACATTTCACTTACGGTTTTCCCAGTCACG	R1
165	CCTCTCTATGGGCAGTCGGTGATCTAGAACATTTCACTTAC	R2
166	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGTAATCCTTGTGGTATCG	1
167	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGTAATCCTTGTGGTATCG	2
168	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGTAATCCTTGTGGTATCG	3
169	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGTAATCCTTGTGGTATCG	4
170	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGTAATCCTTGTGGTATCG	5
171	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGTAATCCTTGTGGTATCG	6
172	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGTAATCCTTGTGGTATCG	7
173	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGTAATCCTTGTGGTATCG	8
174	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGTAATCCTTGTGGTATCG	9
175	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGTAATCCTTGTGGTATCG	10
176	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGTAATCCTTGTGGTATCG	11
177	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGTAATCCTTGTGGTATCG	12
178	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGTAATCCTTGTGGTATCG	13
179	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCGTAATCCTTGTGGTATCG	14
180	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCGTAATCCTTGTGGTATCG	15
181	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACGTAATCCTTGTGGTATCG	16
182	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTATTCGTCGTAATCCTTGTGGTATCG	17
183	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCAATTGCGTAATCCTTGTGGTATCG	18
184	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAGTCGGACGTAATCCTTGTGGTATCG	19
185	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGATCCATCGTAATCCTTGTGGTATCG	20
186	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCAATTACGTAATCCTTGTGGTATCG	21
187	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGAGACGCGTAATCCTTGTGGTATCG	22
188	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCCACGAACGTAATCCTTGTGGTATCG	23

189	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCTCATTCGTAATCCTTGTGGTATCG	24
190	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATACGTAATCCTTGTGGTATCG	25
191	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACAACCTCGTAATCCTTGTGGTATCG	26
192	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCGTAATCCTTGTGGTATCG	27
193	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGGAATCGTAATCCTTGTGGTATCG	28
194	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGACCACTCGTAATCCTTGTGGTATCG	29
195	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGGTTATCGTAATCCTTGTGGTATCG	30
196	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCAAGCTGCGTAATCCTTGTGGTATCG	31
197	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTTACACACGTAATCCTTGTGGTATCG	32
198	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTCATTGAACGTAATCCTTGTGGTATCG	33
199	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATCGTTCGTAATCCTTGTGGTATCG	34
200	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGCCATTGTCGTAATCCTTGTGGTATCG	35
201	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGGAATCGTCGTAATCCTTGTGGTATCG	36
202	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGAGAATGTCGTAATCCTTGTGGTATCG	37
203	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGGACGGACGTAATCCTTGTGGTATCG	38
204	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAACAATCGGCGTAATCCTTGTGGTATCG	39
205	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACATAATCGTAATCCTTGTGGTATCG	40
206	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCACTTCGCGTAATCCTTGTGGTATCG	41
207	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACGAATCGTAATCCTTGTGGTATCG	42
208	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGACACCGCGTAATCCTTGTGGTATCG	43
209	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGGCCAGCGTAATCCTTGTGGTATCG	44
210	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGCTTCCTCGTAATCCTTGTGGTATCG	45
211	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGTCCGAACGTAATCCTTGTGGTATCG	46
212	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGCAACCACGTAATCCTTGTGGTATCG	47
213	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTAAGAGACGTAATCCTTGTGGTATCG	48
214	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAACATAACGTAATCCTTGTGGTATCG	49
215	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAATGGCGTAATCCTTGTGGTATCG	50
216	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTCGTAATCCTTGTGGTATCG	51
217	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCATGGAACGTAATCCTTGTGGTATCG	52
218	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGGCAATCCTCGTAATCCTTGTGGTATCG	53
219	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGGAGAATCGCGTAATCCTTGTGGTATCG	54
220	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCACCTCCTCGTAATCCTTGTGGTATCG	55
221	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCATTAATTCGTAATCCTTGTGGTATCG	56
222	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGCAACGGCGTAATCCTTGTGGTATCG	57
223	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAGAACACGTAATCCTTGTGGTATCG	58
224	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTTGATGTTCGTAATCCTTGTGGTATCG	59
225	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGCTCTTCGTAATCCTTGTGGTATCG	60
226	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTCGGATCGTAATCCTTGTGGTATCG	61
227	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTGCTTCACGTAATCCTTGTGGTATCG	62
228	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTTAGAGTTCGTAATCCTTGTGGTATCG	63
229	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGAGTTCCGACGTAATCCTTGTGGTATCG	64
230	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGGCACATCGTAATCCTTGTGGTATCG	65
231	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCAATCATCGTAATCCTTGTGGTATCG	66
232	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTACCAGTCGTAATCCTTGTGGTATCG	67
233	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAAGAAGTTCGTAATCCTTGTGGTATCG	68
234	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCAATTGGCGTAATCCTTGTGGTATCG	69
235	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTACTGGTCGTAATCCTTGTGGTATCG	70
236	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGGCTCCGACGTAATCCTTGTGGTATCG	71
237	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGGCCACACGTAATCCTTGTGGTATCG	72

238	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGCCTGTCGTAATCCTTGTGGTATCG	73
239	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATCGGTTCGTAATCCTTGTGGTATCG	74
240	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGGAATACGTAATCCTTGTGGTATCG	75
241	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGAACCTCGTAATCCTTGTGGTATCG	76
242	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGCGATTCGTAATCCTTGTGGTATCG	77
243	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCCAATTCTCGTAATCCTTGTGGTATCG	78
244	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGGTTGTCGTAATCCTTGTGGTATCG	79
245	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGAAGGCAGGCGTAATCCTTGTGGTATCG	80
246	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGCCATTCGCGTAATCCTTGTGGTATCG	81
247	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCATCTCGTAATCCTTGTGGTATCG	82
248	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGACATTCGTAATCCTTGTGGTATCG	83
249	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTCCATAACGTAATCCTTGTGGTATCG	84
250	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCAGCCTCAACGTAATCCTTGTGGTATCG	85
251	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGGTTATTCGTAATCCTTGTGGTATCG	86
252	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCTGGACGTAATCCTTGTGGTATCG	87
253	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGAACACTTCGTAATCCTTGTGGTATCG	88
254	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGAATCTCGTAATCCTTGTGGTATCG	89
255	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAACCACGGCGTAATCCTTGTGGTATCG	90
256	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGGATGCGTAATCCTTGTGGTATCG	91
257	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGAACCGCGTAATCCTTGTGGTATCG	92
258	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGTCCAATCGTAATCCTTGTGGTATCG	93
259	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCGACAAGCGTAATCCTTGTGGTATCG	94
260	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAGATCGTAATCCTTGTGGTATCG	95
261	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAAGCGGTCGTAATCCTTGTGGTATCG	96
262	CCATCTCATCCCTGCGTGTCTCCGACTCAG GGTGGAATACC GTAATCCTTGTGGTATCG	97
263	CCATCTCATCCCTGCGTGTCTCCGACTCAG AATCCTTAGGC GTAATCCTTGTGGTATCG	98
264	CCATCTCATCCCTGCGTGTCTCCGACTCAG CAAGTTCATAC GTAATCCTTGTGGTATCG	99
265	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTCAAGGCCGC GTAATCCTTGTGGTATCG	100
266	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCCGATAGAGC GTAATCCTTGTGGTATCG	101
267	CCATCTCATCCCTGCGTGTCTCCGACTCAG TACACGCTCCC GTAATCCTTGTGGTATCG	102
268	CCATCTCATCCCTGCGTGTCTCCGACTCAG GGAGTAGATTC GTAATCCTTGTGGTATCG	103
269	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACGCTTGGACC GTAATCCTTGTGGTATCG	104
270	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGTATACGGAC GTAATCCTTGTGGTATCG	105
271	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCTCAGTACAC GTAATCCTTGTGGTATCG	106

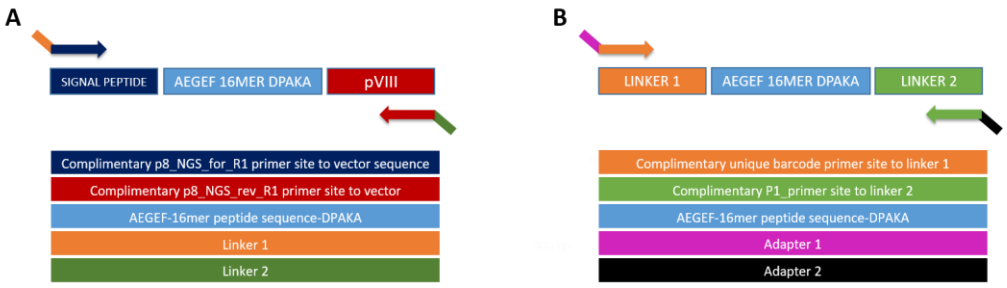


Figure 2.2.3 NGS PCR overview.
A. Schematic overview of the PCR1 **B.** Schematic overview of the subsequent PCR2 amplification of the amplicon from PCR1.

2.2.11. NGS analysis

Next generation sequencing results were received in fastq format. The following scripts were carried out in Ubuntu, a software operating system provided by Linux. Any input file name will be stated in green font and any output file name in red. The following script (called pipeline1.pl) was responsible for various different functions: a) Conversion from fastq format to fasta file format (seqtk seq -a xxx.fastq >xxx.fasta).

b) Demultiplexing of the data according to the barcodes (barcode identifying sequences had been already supplied, example of the command for the barcode CTAAGGTAAC: cat xxx.fasta | gep –no -goup-separator -E -B 1 '^.{0,5}CTAAGGTAAC' | sed 's/^.{0,5}CTAAGGTAAC//' > split_barcode_BC01.fasta).

c) Translation to all possible 3 forward reading frames (example being: perl translate.pl -i split_barcode_BC01.fasta -s opq -f 1 > Sept2019_BC01.frame1.fasta, perl translate.pl -i split_barcode_BC01.fasta -s opq -f 2 > Sept2019_BC01.frame2.fasta, perl translate.pl -i split_barcode_BC01.fasta -s opq -f 3 >Sept2019_BC01.frame3.fasta),

d) Pool of all the possible translation outcomes in one fasta file format (example being: cat Sept2019_BC01.frame1.fasta Sept2019_BC01.frame2.fasta Sept2019_BC01.frame3.fasta > Sept2019_BC01.frames),

e) Identification of the flanking motifs (AEGEF being the left and DPAKA being the right motif, respectively, example of command being: perl iterate_motifs.pl Sept2019_BC01.frames > Sept2019_BC01.frames.LR.fasta).

Mainly, there were two different approaches that were then followed for the further analysis of these data: a) based on Z score and b) based on frequency analysis. Z score analysis takes into consideration the total amount of sequences that are present in each barcode, as well as the frequency of every peptide sequence in positives and negatives samples; the exact equation is described in (Zhang et al., 2011).

$$z = \frac{p1 + p2}{\sqrt{\underset{n1}{p1(1-p1)} \underset{n2}{p2(1-p2)}}}$$

Z scores were calculated in order to compare the frequency of each peptide in a given barcode compared to within a negative pool. For instance, barcodes were pooled for the creating of a negative pool that all the rest of the barcodes will be compared to (as example command being: cat Sept2019_BC07.frames.LR.fasta Sept2019_BC08.frames.LR.fasta Sept2019_BC09.frames.LR.fasta Sept2019_BC10.frames.LR.fasta Sept2019_BC11.frames.LR.fasta Sept2019_BC12.frames.LR.fasta > WT2months_pool.fa), followed by a comparison of positive samples (e.g. BC01) with the negative pool (example command being: perl compare.collapsed.sequences_2a.Zscore.pl Sept2019_BC01.frames.LR.fasta WT2months_pool.fa > 01vsWT2_pool.table). Then, by setting up a Z score cut off at either 2, 4, 5, or 8, peptides were identified as being specific for the

positive sample when it had a Z score above that cut off. This was done for each positive replicate sample and for each peptide sequences were ranked by Z score (example commend being: awk ' NR>1 && $4>=5 {print $1;}' 01vs68-85.table> 01_Z5vs68-85_seq). This was followed by pooling all the peptides that fit in the criteria and sorting them by the example command: cat 01-67_Z5vs68-85_sequences | sort | uniq -c | sort -nr | sed 's/^[\t]*//' | awk '$1>=3 {print $2;}' > 01-67_Z5_vs68-85_seen_3_or_more. This output file contained all the unique sequences that have a Z score more than 5 and they have been seen in 3 or more of the replicate positive samples.

Furthermore, a simpler relative frequency approach was followed taking into consideration peptide sequences that are present in the top 50 or 200 enriched sequences in a positive sample but not present in the top 50 or 200 top enriched sequences in a negative pool. This was achieved by the application of the following script containing firstly a command line, in which the sorting of the sequences to the top 200 was achieved for each barcode (e.g. cat round1_BC01.rank.table | head -n 200 | awk '{print $1}' > BC01_ranked_top200), followed by condensing every replicate sequence in the negative pool to a single occurrence (e.g. sort -u WT_12months_50-67.top200_sequences > WT_12months_50 67.top200_sequences_unique). Then the sequences for each positive sample that were not present in the negative pool were identified (e.g. gep -v -f WT_12months_50-67.top200_sequences_unique TRACK_12months_01-18.top200_sequences > TRACK_12months_01-18_top_200_not_seen_top200_negatives_50-67),

followed by finding the number of times they had been seen across the positive replicates' barcodes (e.g. sort TRACK_2months_19-49.top200_not_seen_top200_negatives_68-85 | uniq -c | sort -nr | sed 's/^[\t]*//' > TRACK_2months_19-49.top200_not_seen_top200_negatives_68-85_seen_in_the_positives) and finally creating a separate file containing a list of the peptide sequence when a cut off was set, this was set as a sequence being seen in at least in the 30% of the positive replicates (unless it is stated otherwise) e.g. awk ' $1>=6 {print $2;}' TRACK_12months_01-18_top_200_not_seen_top200_negatives_50-67_seen_in_the_positives > TRACK_12months_01-18_seen_in_3outof10samples. The final output file was peptidelist.txt_percentage_report.txt containing the percentages of the frequencies of each selected peptide and was dependant on the total amount of sequences per barcode using the following code command (LR.list being the list of all the barcode files that have been identified after the identification of the flanking motifs): perl countpep1.2.pl 57_R2_peptides.txt LR.list > 57_R2_peptides.txt_percentage_report.txt.

Finally, all the frequencies or Z scores of the individual peptides of interest could be searched for in the table files (for Z scores) and fasta files (for frequency) (e.g. grep –w xxx *.table > Peptide_1.txt).

2.2.12. ELISA

- Phage ELISA using an HRP detection system

Normally, 1/2000 of anti-prion monoclonal antibodies (used as control antibodies in the majority of the assays as large quantities were available) were

coated onto Nunc Immunoplate F96 MaxiSorp (ThermoFisher) overnight at 4°C. The plates were then washed 3 times with 1xPBS and blocked with 400 µl of 3% PBSM (Marvel) for 1h at RT. After washing 3 times with PBS, 100 µl of supernatant phage (preblocked with 3% PBSM for 1h at RT) was added to the wells and incubated for 1h at RT, followed by 3 times with 0.1% PBST and 3 times with PBS washing. 100 µl of anti-M13 antibody-HRP (Sigma-Aldrich) pre-diluted 1/2000 in 3% PBSM was then added to the wells and incubated for 1 hour at RT. This was followed by washing 3 times with PBS and 3 times with 0.1% PBST. 100 µl of 3,3',5,5'-Tetramethylbenzidine (TMB) Liquid Substrate (ThermoFisher) was added to the wells. After colour development, plates were read at 450 nm on the Multiskan Ascent plate reader (Thermo Electron Corp).

- Phage ELISA using an AP detection system

This was carried out exactly the same as for the ELISA using the HRP detection system with the exception that the detection system was 100 µl of pre-diluted 1/2000 in 3% PBSM anti-Fd antibody (Sigma-Aldrich) followed by 100 µl of pre-diluted 1/2000 in 3% PBSM anti-rabbit-AP antibody (Sigma-Aldrich). 100 µl of SIGMA-FAST™ p-Nitrophenyl phosphate buffer was added to the wells. Plates were read at 405 nm on the Multiskan Ascent plate reader (Thermo Electron Corp).

- Detection of synthetic peptides in ELISAs

1 ug/ml of biotinylated peptide was coated onto Pierce™ Streptavidin Coated High Capacity Plates or 10 ug/ml of amidated peptide was coated onto Nunc Immunoplate F96 MaxiSorp (ThermoFisher Scientific) overnight at 4°C. The

plates were then washed 3 times with PBS (1x) and blocked with 400 µl of 3% PBSM (Marvel) for 1h at RT. After washing 3 times with PBS, 100 µl of sera or sera spiked with monoclonal anti-prion antibodies or PBS solution with monoclonal anti-prion antibodies (in various dilutions) were added to the wells and incubated for 1h at RT, followed by 3 times 0.1% PBST and 3 times with PBS washing. 100 µl of pre-diluted 1/2000 in 3% PBSM anti-mouse-AP antibody (Sigma-Aldrich) were then added to the wells and incubated for 1 hour at RT. This was followed by washing 3 times with PBS and 3 times with 0.1% PBST. 100 µl of SIGMA-FAST™ p-Nitrophenyl phosphate buffer was then added to the wells at RT. Plates were read at 405 nm on the Multiskan Ascent plate reader (Thermo Electron Corp) after 2h and O/N incubation with substrate.

- Detection of phage-peptide in ELISA

100 µl of supernatant bacteriophage or 10 µl of PEG precipitated bacteriophage were coated onto Nunc Immunoplate F96 MaxiSorp (ThermoFisher Scientific) at RT for 1h or O/N at 4°C. The plates were then washed 3 times with PBS (1x) and blocked with 400 µl of 3% PBSM (Marvel) for 1h at RT. After washing 3 times with PBS, 0.5 µl of mouse sera (preblocked with 3% PBSM for 1h) was added to the wells and incubated for 1h at RT, followed by 3 times with 0.1% PBST and 3 times with PBS washing. 100 µl of pre-diluted 1/2000 3% PBSM anti-mouse-AP antibody (Sigma-Aldrich) were then added to the wells and incubated for 1 hour at RT. This was followed by washing 3 times with PBS and 3 times with 0.1% PBST. 100 µl of SIGMA-FAST™ p-Nitrophenyl

phosphate buffer was added to the wells at RT. Plates were read at 405 nm on the Multiskan Ascent plate reader (Thermo Electron Corp) after 2h and O/N.

2.2.13. Phage production

- Production of phage in flasks

One colony from TG1 bacteria containing phagemid vector was inoculated in 500 ml of 2YT medium with 150 ug/ml ampicillin and 1% glucose and shaken at 37°C until OD_{600} is 0.4-0.6. Helper phage (M13K07 for bacteria containing pVIII vector, exphage for bacteria containing pIII vector) was added with 10x multiplicity of infection and then bacteria incubated static at 37°C for 30 min. Bacteria were pelleted (5000g for 20 min) and resuspended in 1L of 2YT with 150 ug/ml ampicillin and 150 ug/ml kanamycin and shaken at 30°C for 16h. 125 ml of PEG solution was added to the supernatant and placed on ice for 1h. Cultures were centrifuged at 8000 g for 20 min, and phage pellet was resuspended in 12 ml of PBS (1x). Serial dilutions of phage was conducted to determine phage titre. 20 µl of each dilution was added to 180 µl of OD_{600} 0.4-0.6 TG1 cells and incubated static at 37°C for 45 min. 100 µl of the cultures were plated on 2YT agar plates containing 150 ug/ml ampicillin and 1% (w/v) glucose and incubated static at 37°C for 16h. Colonies were counted and titration was calculated the next day.

- Production of phage in 96 well plates

One colony from TG1 bacteria containing phagemid vector was inoculated into a single well containing 600 µl of 2YT medium with 150 ug/ml ampicillin and

1% (w/v) glucose and shaken at 37°C until OD_{600} is 0.4-0.6 in an Axygen® 96 deep well plate. Helper phage (M13K07 for bacteria containing pVIII vector, exphage for bacteria containing pIII vector) was added with 10x multiplicity of infection and then bacteria were incubated static at 37°C for 30 min. Bacteria were pelleted (3000g for 15 min) and resuspended in 600 µl of 2YT with 150 ug/ml ampicillin and 150 ug/ml kanamycin and shaken at 30°C for 16h. Supernatant containing phage was obtained after centrifugation at 3500 rpm for 10 min.

2.2.14. Quantitative PCR of phagemid

Phage ssDNA was extracted following the manufacturer instructions for E.Z.N.A.® M13 DNA mini kit (Omega Bio-tek). 1 µl of DNA was added to 9.5 µl of UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen), 12.5 µl of PowerUp™ SYBR™ Green Master Mix (Fisher Scientific), 1 µl of forward and 1µl of reverse primer (psD3: CTGCAGGCTTACTGTTACTG and ATGAAACCATCGATAGCAGC, pc89: ATTACGCCAGCTGGCGAAA). Reaction was incubated at 95° for 2 min, [95° for 15s, 59° for 30s, 72° for 30s]x40 and 72° for 5 min; melting curve was applied from 65° to 95°.

2.2.15. Biopanning

- Epitope mapping using Protein G agarose beads

Anti-prion antibodies (SAF84, SAF70, SAF15 and SAF32) were coupled with prewashed Pierce™ Protein G Agarose beads (1/2,000 in 500 µl of final volume of PBS) rotating for 1h at RT or O/N at 4°. Beads were then washed 3 times

with PBS and blocked with 300 µl of 3% MPBS (Marvel) for 1h at RT. 1 ml of PEG precipitated bacteriophages (derived from the pc89_BspQI$^-$ library) was preblocked with 3% MPBS and uncoupled protein G beads for 1h. Beads were washed 3 times with PBS and phage supernatant was collected (2000g for 3 min) and it was added to the beads and incubated for 3h rotating at RT, followed by 5 times with 1% PBST washing and 5 times with PBS washing (unless it is stated otherwise). 100 µl of 0.2 M Glycine HCl pH 2.6 or 100 µl of 100 mM triethylamine were added to the pelleted beads and incubated at room temperature for 10 minutes. Eluted bacteriophages were neutralised by adding 100 µl 1 M Tris buffer pH 7.4. Alternatively, bacteriophages were eluted with 200 µl of 50 mM DTT. 100 µl of the eluted bacteriophage were inoculated into 10 ml of TG1 bacteria (OD_{600} 0.4-0.6) and then bacteria incubated static at 37°C for 30 min. 10 µl of the bacterial culture were plated on small 2YT agar plates with 150 ug/ml ampicillin and 1% glucose. Bacteria were then pelleted, resuspended in 10 ml of 2YT with 150 ug/ml ampicillin and 1% (w/v) glucose and shaken at 37°C for 16 h for the production of glycerol stocks.

- Epitope mapping using precoated Protein G 96 well plate

96 well Pierce™ Protein G Coated Plates were washed three times with PBS (1x) and coated with anti-prion antibodies (SAF84, SAF70, SAF15 and SAF32) for 1h at RT or O/N at 4°C. Plates were then washed 3 times with PBS and blocked with 300 µl of 3% MPBS (Marvel) for 1h at RT. 100 µl of PEG precipitated bacteriophages (derived from pc89_BspQI$^-$ library) was preblocked with 3% MPBS and uncoupled protein G beads for 1h at RT. Phage

supernatant was collected (2000g for 3 min) and it was added per well (after washing 3x with 1xPBS) and incubated for 3 h with agitation at RT, followed by 10 times with 1% PBS Tween washing and 10 times with PBS washing (unless it is stated otherwise). 100 µl of 0.2 M Glycine HCl pH 2.6 or 100 µl of 100 mM triethylamine were added to the wells and incubated at RT for 10 minutes. Eluted bacteriophages were neutralised by adding 100 µl 1M Tris buffer pH 7.4. Alternatively, bacteriophages were eluted with 200 µl of 50 mM DTT. 100 µl of the eluted bacteriophages were kept separately and inoculated into 10 ml of TG1 bacteria (OD_{600} 0.4-0.6) and then bacteria incubated static at 37°C for 30 min. 10 µl of the bacterial culture were plated on small 2YT agar plates with 150 ug/ml ampicillin and 1% (w/v) glucose. Bacteria were then pelleted, resuspended in 10 ml of 2YT with 150 ug/ml ampicillin and 1% (w/v) glucose and shaken at 37°C for 16 h for the production of glycerol stocks.

- Panning against IgG from sera (human or mouse)

For a subtraction step, 96 well Pierce™ Protein G Coated Plates (ThermoFisher) were washed three times with PBS (1x) and was coated with 0.5 µl of normal sera (mouse pool from Sigma-Aldrich, human pool from Leeds biobank) per well for 3 h. Plates were then washed 3 times with PBS and blocked with 300 µl of 3% MPBS (Marvel) for 3 h at RT. PEG precipitated bacteriophages (derived from pc89_BspQI^{-} library) were preblocked with 3% MPBS and uncoupled protein G beads for 1 h. Phage supernatant was collected (2000 g for 3 min) and 100 µl were added per well (after washing wells 3x with 1xPBS) and incubated for 20 min agitating at RT; unbound phages were collected and the

process was repeated 3 times in total. Depleted bacteriophages were stored at 4°C O/N while new 96 well Pierce™ Protein G Coated Plates were washed three times with PBS (1x) and were coated with 0.5 mouse or human sera per well. Plates were washed three times with PBS (1x) and blocked with 300 µl of 3% PBS with milk (Marvel) for 1h at RT. After washing the plates 3 times with PBS, 100 µl of depleted bacteriophages were added per well and incubated for 3 h agitation at RT, followed by five times washing with 2% PBST +1 M NaCl (Fisher Scientific) PBST and NaCl (Fisher Scientific) and five times washing with 1xPBS. Bacteriophages were eluted with 200 µl of 50 mM DTT. 100 µl of the elution were inoculated to 10 ml of TG1 bacteria (OD_{600} 0.4-0.6) and then bacteria incubated static at 37°C for 30 min. 10 µl of the bacterial culture were plated on small 2YT agar plates with 150 ug/ml ampicillin and 1% (w/v) glucose. Bacteria were then pelleted (3000 g for 15 min), resuspended in 10 ml of 2YT with 150 ug/ml ampicillin and 1% (w/v) glucose and shaken at 37°C for 16 h for the production of glycerol stocks.

2.2.16. Detecting antibody binding to immobilised phage-peptides on a microarray

NOVA Nitrocellulose Film Slides were fitted with Proplate slide modules (Grace Bio-Labs) and washed 3 times with PBS (1x). Slides were blocked with various blocking reagents (the list of which is stated in the results section) for 3h at RT or O/N at 4°C. Preblocked normal mouse sera (NMS) spiked with various monoclonal antibodies or the same sera with monoclonal antibodies but purified using protein G beads, were added to the wells for 1 or 3 h. Slides were then washed with various washing buffers (the list of which is stated at the

results section). 100 µl of anti-mouse antibody DyLight® 650 (Invitrogen) pre-diluted 1/2000 in various blocking reagents was then added to the wells and incubated for 1 h. This was followed by washing 3 times with various washing buffers and 3 times with PBS. Slides were then dried and were read in a GenePix 4000A (Molecular Devices, USA) with the PMT settings at 635 and 532 nm, or 800 and 400 nm respectively.

2.2.17. Bradford assay

Standard concentrations of BSA (Biotium) (1.4, 1.0, 0.5, 0.25 and 0 mg/ml) were prepared by adding BSA powder to PBS. 5 µl of sample or standards were added to 250 µl of Bradford Reagent (Bio-Rad) on Nunc MaxiSorp™ ELISA plate. Plates were incubated at RT for 1 min and read at 600 nm on the Multiskan Ascent plate reader (Thermo Electron Corp).

2.2.18. Production of TG1 competent cells

One TG1 colony (Agilent Technologies) from a minimal agar plate was inoculated into 10 ml of 2YT media for 16 h at 37°C, shaking at 220 rpm. 1 ml of the overnight culture was added to 500 ml of prewarmed 2YT media in a 1L flask. Bacteria were grown shaking at 220 rpm at 37°C till bacteria reached OD_{600} 0.4-0.6. Cultures were placed on ice for 15 min followed by centrifugation at 5000 g for 10 min (x3). Pellets were resuspended in 10 ml of 10% (v/v) glycerol (Fisher Scientific) in DEPC autoclaved water and the culture was centrifuged at 5000 rpm for 10 min. Pellets were resuspended in 10 ml of 10% glycerol in DEPC autoclaved water and placed in dry ice for 15 min. Aliquots of the cultures were stored at -80°C.

2.2.19. Helper phage production

One TG1 colony (Agilent Technologies) from a minimal agar plate was inoculated into 10 ml of 2YT media and grown for 16 h at 37°C, shaking at 220 rpm. All of the overnight culture was added to 500 ml of prewarmed 2YT media in a 1 L flask, shaking at 220 rpm at 37°C till bacteria reach OD_{600} 0.4-0.6. Helper phage stock (M13KO7: CFU: 10^{13}/ml) was added to the culture at MOI of 10. Cultures were incubated static at 37°C for 45 min, centrifuged at 5000g for 10 min and the pellet resuspended in 1L of 2YT media containing 150 ug/ml kanamycin, shaking at 220 rpm at 30°C for 16 h. Cultures were centrifuged at 5000 rpm for 20 min; supernatant was collected; 250 ml of PEG buffer was added to the supernatant and the sample placed on ice for 1 h. Cultures were centrifuged at 8000 g for 20 min, and the phage pellet was resuspended in 12 ml of PBS (1x). Serial dilutions of phage were prepared to determine the phage titre. 20 µl of each dilution was added to 180 µl of OD_{600} 0.4-0.6 TG1 cells and incubated static at 37°C for 45 min. 100 µl of the cultures were plated on 2YT agar plates containing 150 ug/ml kanamycin and incubated static at 37°C for 16h. Phage titration was calculated the next day by colony counting.

3. Phage-peptide library construction

3.1. Introduction

Phage display is a molecular technology that was first described in 1985 (Smith, 1985). Antibody fragments (such as VHH or scFvs) or peptides are usually displayed on the surface of filamentous bacteriophages as fusions with one of the phage coat proteins (usually either pIII or pVIII). This powerful technique links the phenotype with the phenotype, allowing selection of target specific ligands (Chang et al., 2020; Georgieva and Konthur, 2011). Synthetic peptide libraries are widely used for the epitope mapping of monoclonal antibodies or of polyclonal response in autoimmunity and cancer. In fact, pVIII display is generally preferred for peptide display (Rahbarnia et al., 2017; Loh et al., 2018). NNK randomisation is one of the directed mutagenesis approaches for the construction of peptide library, where K =G/T, S = C/G, and B = C/G/T. This approach codes for all 20 amino acids but for only one stop codon (amber) (Krištof Bozovicar and Tomaž Bratkoviˇ, 2020). Amber stop codons can be eliminated in phage display by the usage of amber suppressor bacterial strains (Solemani Zadeh et al., 2019). A high diverse peptide library is thus needed in order to be utilised in future epitope mapping experiments. Quality control of the library with NGS is vital, as potential bias should be identified before any biopanning rounds (Glanville et al., 2015).

Aims

- Construction of a randomised 16mer peptide library with directed mutagenesis (NNK approach).
- Quality control of this library by deep sequencing, accessing its diversity and identify any potential bias.

3.2. Phagemid vector preparation

Site directed mutagenesis of the phagemid vector pc89 was required (Felici et al., 1991). The pc89 vector contained a BspQI (Isoschizomer of SapI) restriction enzyme site that would be used for downstream library cloning. Therefore, site directed mutagenesis (SDM) was conducted mutating GAA to GAG (silent mutation, triplete is encoding for the same amino acid, glutamate). Inverse PCR was conducted as described in method 2.2.3. Gel electrophoresis was performed in order to confirm PCR amplification (Figure 3.2.1A). Purified PCR gel product (~4000 bp) were digested with the DpnI methylation sensitive enzyme, allowing only the newly synthesized vector to be ligated and transformed into TG1 bacteria. Sanger sequencing confirmed that clone was carrying the desired mutation.

Testing the vector functionality after SDM was then carried out. A gene coding for the peptide epitope of monoclonal antibody SAF84 (YYRPVDQYN) was digested with the restriction enzymes XbaI and NheI in order to clone it as a control to the newly pc89_BspQI$^-$ vector that was also digested with the same enzymes. Gel electrophoresis of the digested DNA was conducted followed by gel purification of the desired products (pc89_BspQI$^-$ linearized vector <3500 bp and SAF84 epitope insert 400 bp) (Figure 3.2.1B). The SAF84 epitope insert

was then subcloned in to the pc89_BspQI$^-$ vector, and the SAF84_pc89_BspQI$^-$ clone was confirmed by Sanger Sequencing.

After small-scale phage production, supernatant containing bacteriophages were tested in ELISA against SAF84 and PBS coated wells in order to test the functionality of the new mutated vector. ELISA signal was observed only to SAF84 antibody coated wells when bacteriophages displaying the SAF84 epitope were used. No signal was observed to PBS coated wells (Figure 3.2.1C). In summary, pc89_BspQI$^-$ vector retained its functionality (binding capacity and infectivity) after the aforementioned mutation.

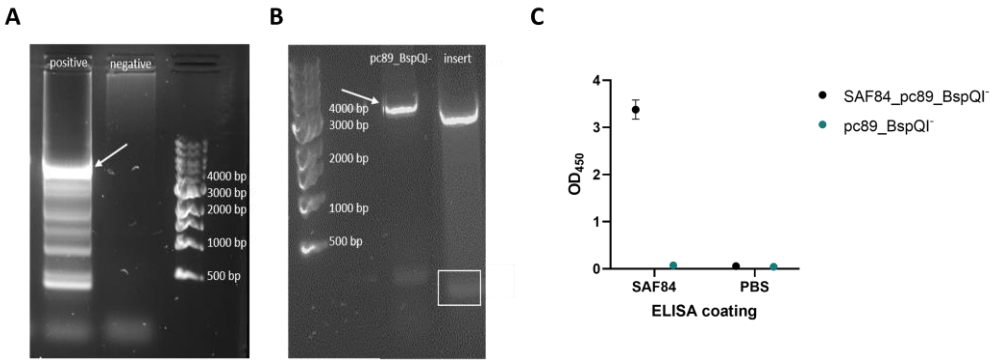


Figure 3.2.1 pc89 SDM and functionality ELISA.
A. Image after agarose gel electrophoresis confirming the PCR amplification at the expected size of pc89_BspQI$^-$ vector (white arrow, 3400bp) (positive). Sample containing no template was included as a control for possible unspecific amplification (negative). **B.** Image of agarose gel electrophoresis performed after restriction enzyme digestion of pc89_BspQI$^-$ vector and pc89 vector with XbaI and NheI restrictions enzymes. Linearized pc89_BspQI$^-$ vector (white arrow, 3400 bp) and SAF84 epitope insert from pc89 vector (400bp, marked with white box) were gel purified. **C.** Phage displaying the SAF84 epitope clone bound to the SAF84 mAb with a mean absorption above 3. No absorption was measured to PBS coated wells.

3.3. Library construction

The library of 16mer randomised peptide sequences in the pc89_BspQI$^-$ vector, was constructed (now named pc89_BspQI$^-$ library). The schematics of the molecular strategy that was followed is depicted in Method 2.2.5. Inverse PCR was used to amplify the phagemid template with introduced diversity and gel electrophoresis confirmed the PCR amplification (x 40 replicates) (Figure 3.3.1.). DpnI and BspQI (SapI isoschizomer) restriction enzyme digestion was performed to the purified PCR amplicons. Ligation and transformation into TG1 bacterial cells followed and the estimated pVIII library size was 5 x 10^9 when titrated. Few clones from pc89_BspQI$^-$ were Sanger sequenced (Table 3.3.1.). The vast majority of these clones were confirmed as 16mers; in fact, only 10% of the clones were religated vector and another 10% with the BspQI site still present.

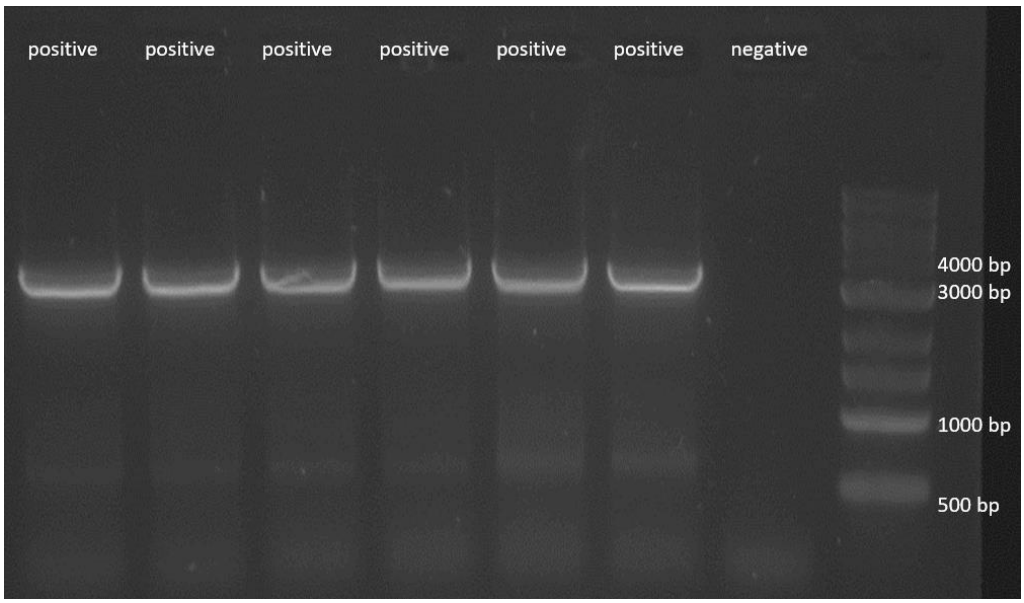


Figure 3.3.1 Representative image after agarose gel electrophoresis confirming amplification after inverse PCR.
Amplicons were observed at the desired size of ~3400 bp. Sample containing no template was included as a control (negative) to ensure no unspecific amplification took place (NEB 1kb ladder marker was used).

Table 3.3.1 Sequences of handpicked library clones

Amino Acid sequences[1]	
NSKVHLTMPRFLPSVP	*NSNACKAMCPPKPYW
*LTRSPWVHL*PHTLR	AFPYRAPCHSSYQQTH
TIGLTKPSTPRMTHM*	YVKPARKPHARSASSK
ALQSHNRNQVAHTPTM	TMNP*KPHIP*KKPLQ
HDMVHGFGSVKESQFL	SRLTSGRNAFLTS*LK
PDNLATQNPLYSRKPA	PPPRSALMAHQYSTRM
MWTRLRIPK*HNHANE	EPSARQRNWCSYAG*P

[1]Amino acid translated sequences of the 16 randomised peptide regions for clone picks that were sanger sequenced.
**represents the presence of an amber stop codon*

3.4. Deep sequencing of the naïve pVIII peptide library

The randomised pc89_BspQI^{-} library was sent for Next Generation Sequencing in order to investigate any potential cloning bias. The region containing the peptide genes was amplified in two sequential steps to prepare it for Ion

Torrent sequencing, and PCR amplicons were observed at the desired size 275 bp and 335 bp for each step, respectively. A negative sample containing water instead of the DNA template was also included in order to determine any unspecific amplification (Figure 3.4.1.).

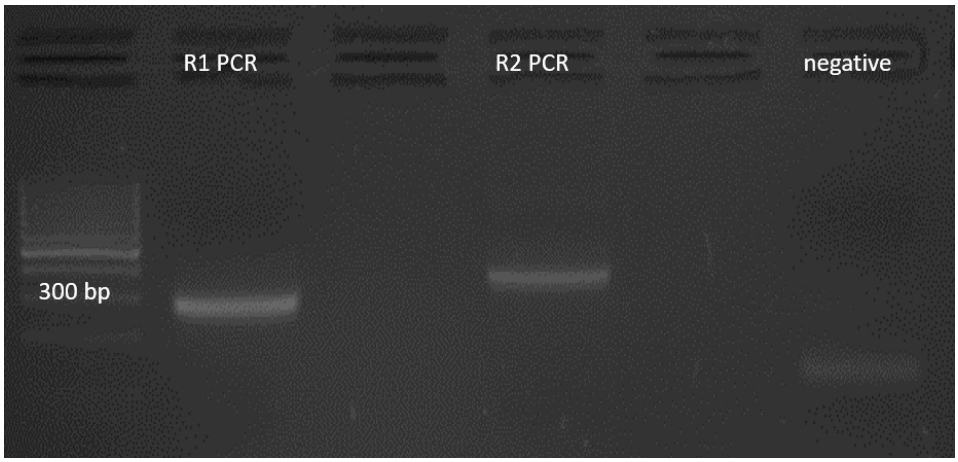


Figure 3.4.1 Representative image after agarose gel electrophoresis confirming PCR amplification for the NGS analysis of the phagemid library.
Amplicons were observed at the desired size at 275 bp for R1 PCR and 335 bp for R2 PCR. Sample containing water instead of a PCR R1 product as a template was performed in order to determine any unspecific amplification (negative).

Approximately 10^6 sequences were obtained containing the required left and right motifs. The number of amino acids between these flanking regions represent the length of the peptides that are displayed by the bacteriophages. The vast majority of the peptides being displayed were 16mers (89.03%), with only 2% being 17mer (Figure 3.4.2A). Furthermore, when all the translated AA sequences were ranked based on their frequency, 73.15% of them were unique; the rest of them containing either amber stop codons or a motif bias. The amino acid sequence of this motif bias is mostly due to DCSS that was present in 2.36% of the whole sequenced population (Figure 3.4.2B). This DCSS

motif was included in the forward primer sequence that should have been removed after cloning, suggesting that the purity of this primer synthesis might have caused the DCSS overrepresentation. Utilising the NKK randomisation strategy, 32 codons (coding for 20 AA and the amber stop codon) could possibly be generated. This theoretical AA distribution for every single AA position of the displayed peptide was compared with the actual obtained AA distribution and it was plotted as fold difference. No difference greater than 2x fold was observed; opal (TGA) and ochre (TAA) stop codon frequency was 0%, as expected (Figure 3.4.3.). In other words, the data reveals always less than 1.5-fold difference over- or under-representation of any amino acids in any given position. Furthermore, the frequency of single copies present in the naïve library was 83% whereas sequences with more than 3 copies represent only 5% (Figure 3.4.4.). Finally, the poisson probability was calculated and the probability of a sequence to be seen more than 3 times in this dataset was very low ($6x10^{-9}$).

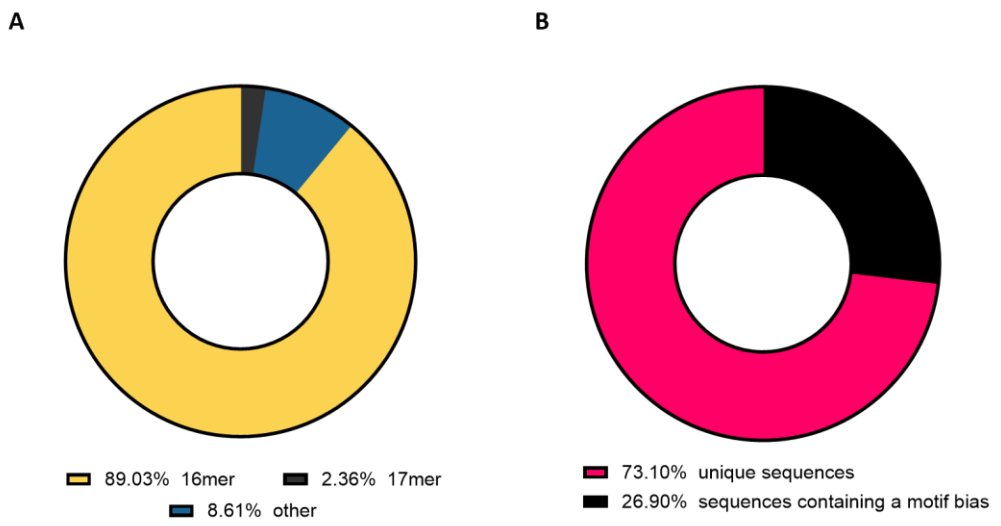


Figure 3.4.2 An overview of the peptide length and single copy number distribution of the pc89_BspQI$^-$ naïve library.
A. The length of the majority of the displayed peptides confirmed to be 16mer (89%). **B.** The vast majority of the peptide sequences were unique (73%); the rest contained either a motif bias (DCSS), the original pc89 template or other minor enrichment of peptide sequences.

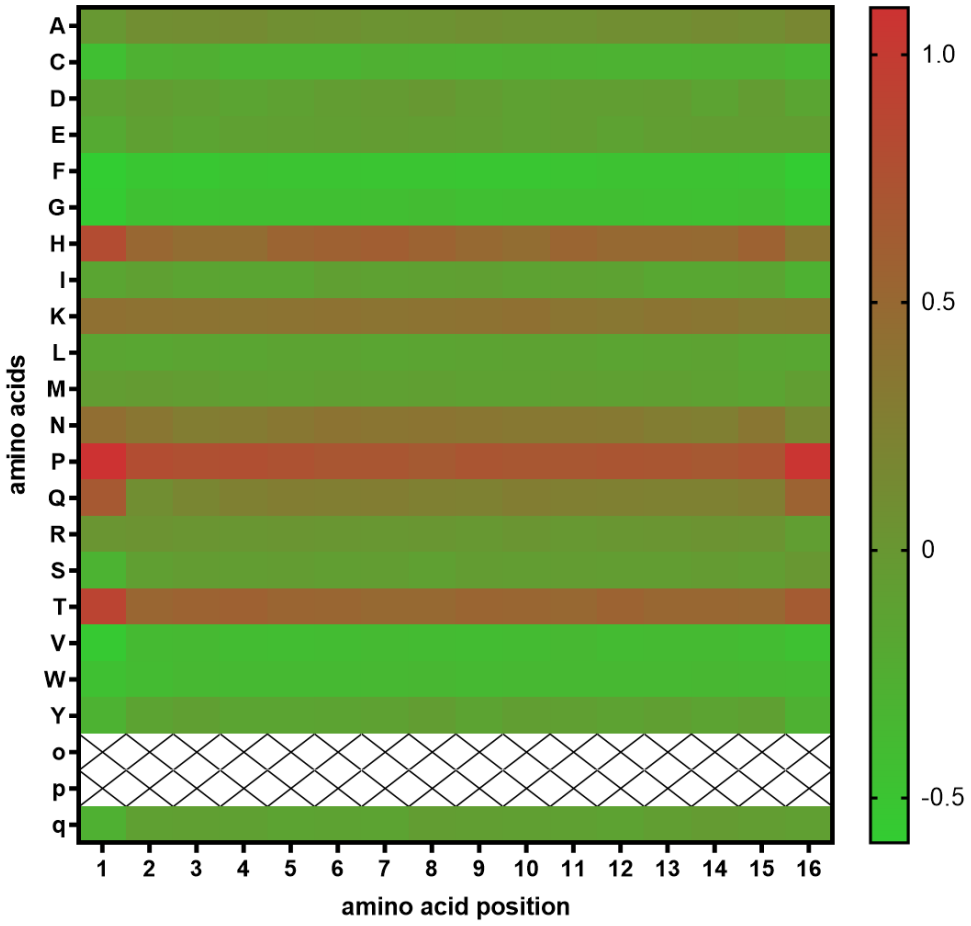


Figure 3.4.3 Heat map of the fold difference between the theoretical and naïve distribution of each amino acid at each position.
Abbreviations of the amino acids were depicted on the y axis of the heat map (o = ochre, p =opal, and q= amber stop codons). Position of every residue of the 16mer peptides is described on the x axis. Fold difference of the theoretical versus the obtained library diversity was observed to be smaller than 2x for every amino acid position.

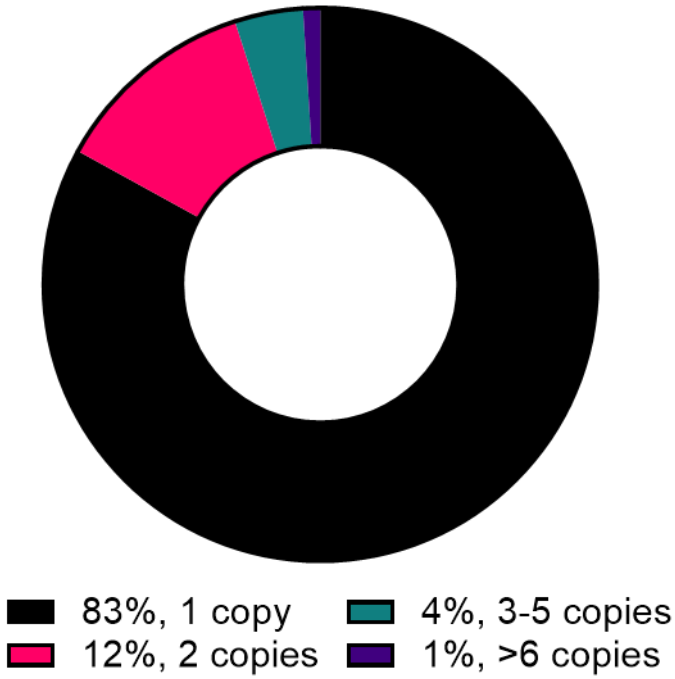


Figure 3.4.4 An overview of the insert sequence diversity.
The vast majority of the in frame clones were singletons (one copy). Sequences present more than 3 times represent 5% of the sequenced clones.

3.5. Discussion

The first phase of this study aimed to produce peptide bacteriophage library that could be used for the epitope mapping of mAbs and then polyclonal immune responses. A randomised 16mer peptide library was constructed into pc89_BspQI$^-$ pVIII vector and its size was estimated to be 5 x 10^9 CFU/ml. This library construction strategy is a very simple approach, using a single step PCR method, and its diversity is comparable with similar approach with diversity up to 10^{10} (Kong et al., 2020; Ryvkin et al., 2018).

Deep sequencing of phage libraries gives unprecedented insight of their quality, diversity and size. With almost a million sequences, the analysis of the naïve pc89_BspQI$^-$ 16mer peptide library was extremely useful for identifying

bias that will allow the exclusion of clones that may be preferentially propagated in future panning experiments. Future libraries would benefit from this initial deep sequencing analysis, since the information that is gained with the laborious technique of Sanger sequencing of only a very limited number of handpicked clones is insufficient to identify clones that are overrepresented in a naïve library (Ibsen and Daugherty, 2017). The diversity of the library was confirmed, after deep sequencing of roughly 1 million sequences, around 70% of those were unique. In accordance with the present results, other reported peptide libraries have similar percentages of singletons from 94-98% (Gough et al., 1999; Kong et al., 2020). Moreover, 3 copies of the same sequences represent 5% of the sequences library, which is comparable with similar percentages (8%) that have been recently reported (Kong et al., 2020). Furthermore, previous studies have calculated the diversity as described in Rebollo *et al*., by applying the percentage of unique sequences in the amount of all the bacterial clones. Taking this into consideration, the diversity of the pc89_BspQI^{-} library was estimated at 3.7 x 10^{9} CFU/ml (Rentero Rebollo et al., 2014).

The chosen molecular approach was a focused randomisation, known as site-saturation mutagenesis using degenerated NNK codons. K stands for the equal mixture of the G and T bases, and N stands for the equal mixture of all the bases, meaning that the probability of the presence of a stop codon is 3.13% for each residue with this approach. This method was optimal because, compared with other site saturation schemes like NNB or NNS, when used in

an E. *coli* system it has the highest expected coverage as well as the highest peptide diversity. In fact, in a comparison to similar approaches, NNK was amongst the best approaches, because only one codon for the amino acid of cysteine is included (Sieber et al., 2015). Another advantage of this approach is the fact that the resulting 32 DNA triplets are coding for all 20 natural amino acids and only one of them is coding for the amber stop codon (TAG). This issue can be largely overcome with the usage of amber suppressor bacterial strains that can translate it to the amino acid glutamine (Galán et al., 2016). Moreover, the presence of stop codons in some of the clones is not leading to non-displaying phage as pVIII display is dependent on the number of gene copies of pVIII (around 2700) leading to the display of hundreds of copies of the peptide in each phage particle. The presence of a stop codon in an amber suppressor strain may lead to the reduction in display levels but as hundreds of copies are present, each phage will still display the encoding peptides.

Deep sequencing of the pc89_BspQI^{-} library revealed that around 2% of the length of the displayed peptides were 17mer instead of 16mer. This is not surprising since the degenerate primer was purified with high performance liquid chromatography (HPLC) and one of its known disadvantages is the coelution of similar size molecules, especially for primers exceeding 50bp length (Pinto et al., 2018). Less than 3% of the sequences obtained carried a motif bias, the translated amino acid sequence of which was DCSS. After closer inspection of the primer design strategy, this exact sequence is present at the 5’ end of the forward primer. These extra base pairs were firstly included in

the primer sequence for the BspQI enzyme digestion. This sequence should have been completely eliminated after the digestion step, however with a combination of incomplete cleavage and blunt end ligation, DCSS sequence was inserted into the library, although its frequency was very low. Knowledge of this bias allows the exclusion of every potential enriched clone containing this motif in future panning experiments. A motif bias (GSSSI) based on incomplete digestion introduced from the reverse primer was not observed, indicating that forward primer secondary structure was more complex than the reverse one. These findings further support the necessity of naïve library's deep sequencing before its usage in future biopanning experiments. Taken together, these data confirm that this 16mer peptide library is highly diverse and has limited peptide bias. The next chapter moves on to discuss the results of using this library on the epitope mapping of mAbs.

4. Optimisation of epitope mapping using phage and/or bacterial display.

4.1. Introduction

Random peptides libraries displayed in bacteriophages have been used extensively for the epitope mapping of monoclonal antibodies. Determination of the exact cognate epitopes of antibodies is crucial to vaccine development, serum based diagnostics and elucidation of the immune response (Potocnakova et al., 2016). Comprehensive epitope mapping of autoimmune responses to cancer is required to identify potential diagnostic epitopes. In the current study, epitope mapping was carried out using a process called Next Generation Phage Display (NGPD) that couples the vast diversity of ligands within phage display libraries, with the analytical depth of next generation sequencing (NGS) (Naqid et al., 2016). Furthermore, Anchored Periplasmic Expression (APEx) is a bacterial display method in which the membrane of Gram-negative bacteria containing phagemids is stripped, periplasmic region is exposed, and antibody or protein fragments are ‘’anchored’’ in the inner bacterial membrane as fusions with the N-terminus of the pIII coat phage protein. Positive bacterial clones could be directly isolated with Fluorescence-activated cell sorting (FACS) (Harvey et al., 2004). This bacterial display could potentially compliment phage display as its incorporation with flow cytometry, minimises the screening effort. In this chapter, known monoclonal antibodies were spiked into buffer and/or sera over a range of titres to determine the limit of detection.

Aims:

- Optimise APEx and combine it with an initial round of phage display panning in order to increase the selection of positive phage-peptide clones.
- Optimise NGPD steps when applied to epitope mapping of monoclonal antibodies (the epitopes of which are known) in order to maximise the sensitivity and specificity of the method, including minimising the biopanning iterative rounds of the NGPD method to retain diversity of peptide binders.
- Determine the limit of detection of NGPD when monoclonal antibodies (the epitopes of which are known) are spiked into normal mouse sera.

4.2. Optimisation of Anchored Periplasmic Expression (APEx)

The following experiments were mainly conducted with bacterial clones containing the pSD3 phagemid vector, used for the pIII display, as APEx is based on the anchored ability of the phage pIII protein with the inner bacterial membrane (Figure 4.2.1A). The following optimisation APEx steps aimed to determine its LOD that could potentially compliment traditional phage display panning by cloning phage output populations to the bacterial display format. Firstly, the epitopes of SAF84 and SAF70 mAbs cloned into psD3 phagemid vector were used. TG1 bacterial cells containing no phagemid were also used as a control and all cells were fixed and permeabilised. In every experiment, cells were incubated with the corresponding antigen, an irrelevant antigen control as well as with the fluorescent conjugated secondary mouse antibody

(Alexa 488) alone as a further background control (Method 2.2.9). Gating the cell population was based on TG1 only fixed sample on the scatter plot. Compensation was conducted taking into consideration any fluorescence that was observed when any sample was incubated with irrelevant primary antibody. The population of cells included in the P2 box was considered positive. Firstly, the optimal dilution of the primary antibody (SAF84) was determined as well as the sensitivity of the assay by performing 2x washes and adding 250 units of lysozyme. A bacterial clone containing the phagemid vector encoding the epitope for the SAF84 antibody (SAF84 epitope clone) was used undiluted or was spiked in various dilutions into randomised peptide pIII bacterial clones, in order to assess the sensitivity of the method (Table 4.2.1). The optimal concentration for the primary antibody was determined to be $2x10^{-3}$ to 10^{-4}, since when it was diluted further no positive cell population was observed (Figure 4.2.1B). When SAF84 epitope clone was spiked in pIII library, positive signal was only observed when the positive clone was present at 50% (Fig 4.2.1C).

Table 4.2.1 List of conditions of APEx to determine the optimal primary antibody concentration and the limit of detection.

Condition	SAF84 epitope clone dilution in pIII library	SAF84 antibody dilution
1	undiluted SAF84 epitope clone	2×10^{-3}
2		10^{-4}
3		5×10^{-5}
4		10^{-6}
5		2×10^{-6}
6	1 in 2	2×10^{-3}
7	10^{-2}	2×10^{-3}
8	10^{-4}	2×10^{-3}
9	5×10^{-5}	2×10^{-3}
10	10^{-6}	2×10^{-3}

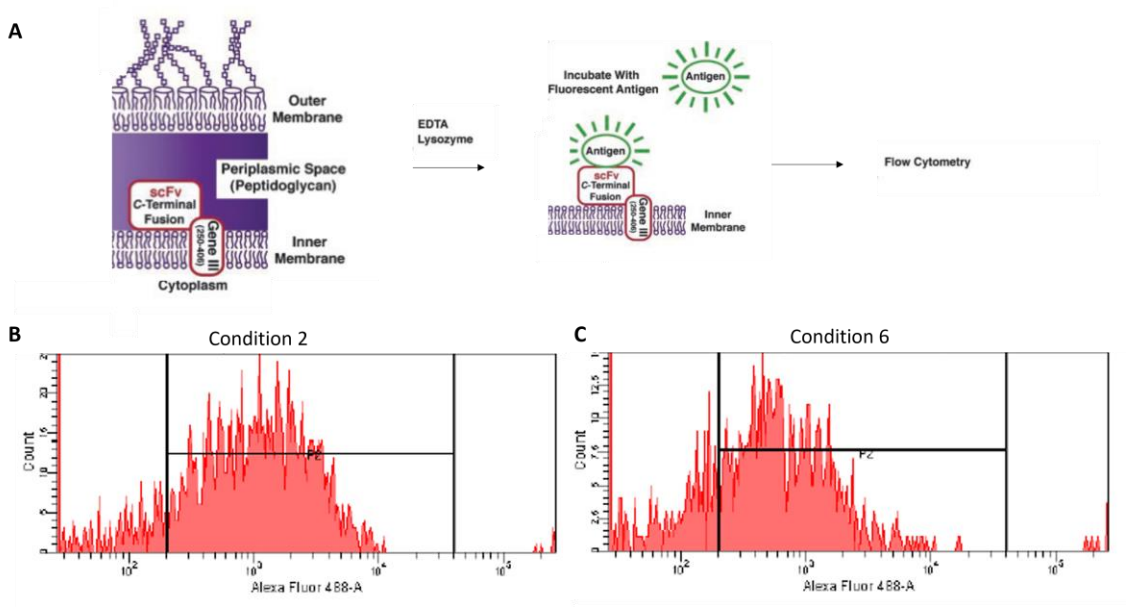


Figure 4.2.1 APEx overview and an initial determination of the sensitivity of APEX as well as the optimal primary antibody dilution (SAF84).
A. APEx overview: Fused scFv (or in this case, SAF84 epitope) with phage pIII protein was anchored in the inner bacterial membrane. Outer bacterial membrane was removed by the addition of EDTA and lysozyme, and anchored complex was exposed. The now called spheroplasts were incubated with fluorescent antigen, followed by flow sorting of the positive clones, adapted from (Harvey et al., 2004). **B and C.** Different primary antibody dilutions; condition 2 (10^{-4}) was determined to be the lowest one in which positive signal was observed (P2 box). Determination of the limit of detection of the method when known positive clone is

present in various dilutions in a pIII library. Positive population was observed only when SAF84 epitope clone was present at 50% of the clones in the pIII library (P2 box); no positive cell population was observed in higher SAF84 epitope clone dilutions (data now shown).

Further optimisation was carried out in order to increase the limit of detection of the assay conducting three instead of two washes between the incubation steps and using an increased amount of lysozyme (3,000 units instead of 250 units) (Table 4.2.2). Lysozyme is the main enzyme used for stripping the outer membrane of Gram-negative bacteria therefore an increased usage of it could potentially increase the amount of positively selected clones. Unfortunately, the limit of detection was not improved but was confirmed to be the same as previously observed (50% of SAF84 clone was spiked in pIII library and 10^{-4} was the optimal primary antibody dilution) (Figure 4.2.2).

Table 4.2.2 List of conditions of APEx to determine the optimum primary antibody dilution and limit of detection with increased washing (from two to three) and increased lysozyme treatment (from 250 to 3000 units).

Condition	SAF84 epitope clone dilution in pIII library	SAF84 antibody dilution
1	SAF84 epitope clone was used undiluted	10^{-4}
2		5×10^{-4}
3		10^{-5}
4		2.5×10^{-5}
5		5×10^{-5}
6	1 in 2	10^{-4}
7	10^{-2}	10^{-4}
8	10^{-4}	10^{-4}
9	5×10^{-5}	10^{-4}
10	10^{-6}	10^{-4}
11	1 in 2	10^{-5}
12	10^{-2}	10^{-5}
13	10^{-4}	10^{-5}
14	5×10^{-5}	10^{-5}
15	10^{-6}	10^{-5}

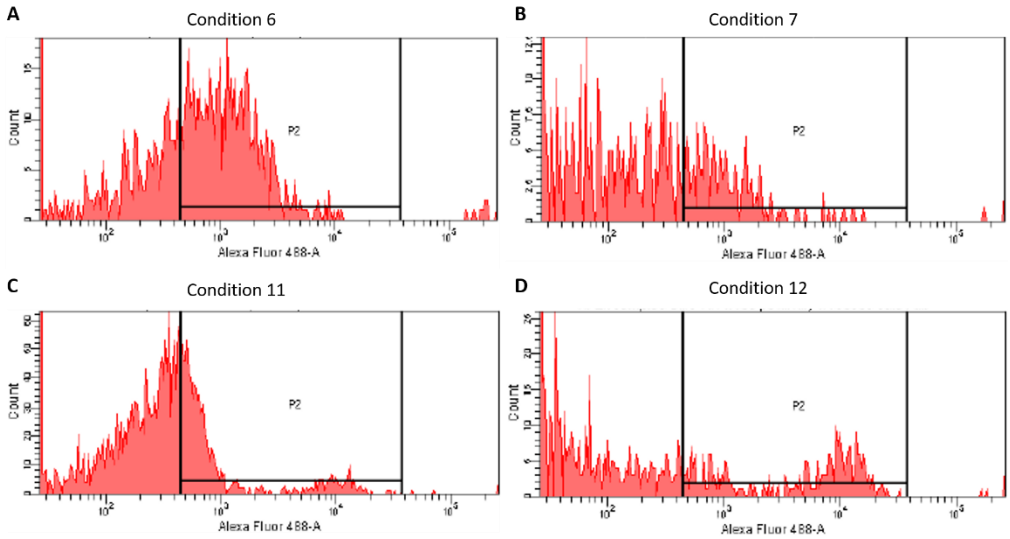


Figure 4.2.2 Assessment of the limit of detection for APEx with increased washing steps and increased lysozyme treatment.
SAF84 epitope clone was present in the pIII library at 50% (A and C) and 10% (B and D); primary antibody was diluted 10^{-4} (A and B) or 10^{-5} (C and D) respectively. Limit of detection was confirmed to be when the known epitope was present at 50% in the pIII library and when the primary antibody was not diluted more than 10^{-4}.

The addition of a blocking reagent (1% PBSM) before the primary antibody incubation could potentially improve the sensitivity by blocking the unspecific interaction between bacterial cells and antibodies. Therefore, marvel skimmed milk was used as it is usually the blocking reagent in phage display. The potential contribution of an increased amount of lysozyme was further assessed in combination with the blocking reagent since increased washing conditions did not seem to have any effect (Table 4.2.3.). Surprisingly, permeabilised cells were immediately dissolved after the addition of the blocking reagent; this had an effect on the observed positive population which was not improved with increasing amounts of lysozyme (Fig 4.2.3.).

Table 4.2.3 List of conditions of APEx to test addition of a blocking agent and increased lysozyme treatment on the sensitivity of the method.

Condition	SAF84 epitope clone dilution in pIII library	SAF84 antibody dilution	Lysozyme (units)
1	SAF84 epitope clone was used undiluted	10^{-4}	250
2		10^{-4}	250
3		10^{-4}	250
4		10^{-4}	3,000
5		10^{-4}	3,000
6		10^{-4}	3,000
7		10^{-4}	60,000
8		10^{-4}	60,000
9		10^{-4}	60,000

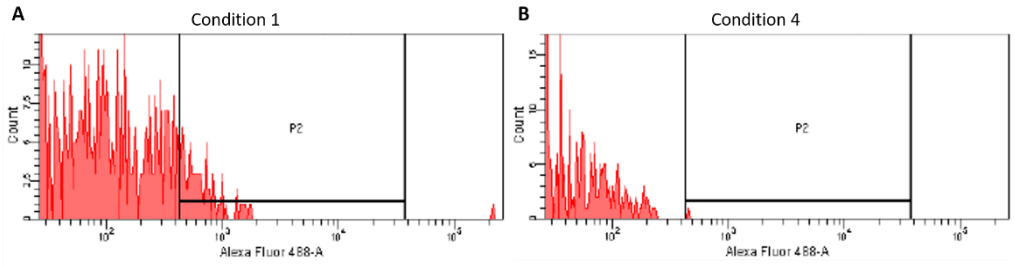


Figure 4.2.3 Assessment of limit of detection of APEx with increased usage of lysozyme and the addition of blocking reagent (1% PBSM).
A. A significant reduction in the observed positive cell population was observed with the usage of the blocking reagent. **B.** Positive cell population was not observed with increased amount of lysozyme and blocking agent.

A verification of the previous findings was necessary in order to conclude on the best combination of conditions such as the optimal lysozyme dilution, the impact of a different blocking reagent (BSA) and the optimal dilution of the secondary fluorescent antibody (Table 4.2.4.). The highest percentage of positive cells was observed with the addition of undiluted secondary antibody (Fig 4.2.4.A. and B). Usage of 12x more of the recommended lysozyme concentration showed an increase of the positive cell population but when sample was blocked with different blocking reagent (BSA), bacterial particles

were still physically distorted (no pellet was visible after centrifugation) and a reduction on the positive cell population was observed (Fig 4.2.4.C and D). The presence of blocking reagents (either Marvel skimmed milk or BSA) was determined to not be beneficial therefore this step was not included in the following experiments.

Table 4.2.4 List of conditions of APEx to assess the effects of blocking cells, increased lysozyme treatment and optimisation of secondary antibody concentration when SAF84 mAb concentration was not changed (1 µl).

Condition	Lysozyme (units)	Fluorescent antibody dilution	Blocked with 1% BSA in PBS
1	250	1 in 4	NO
2	250	1x	NO
3	3,000	1 in 4	NO
4	3,000	1x	NO
5	60,000	1 in 4	NO
6	60,000	1x	NO
7	3,000	1x	YES

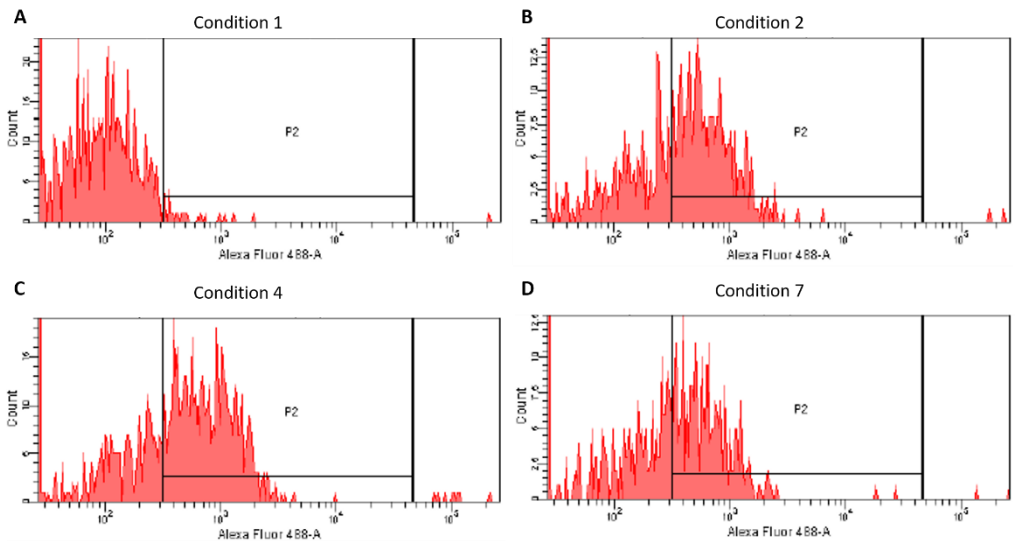


Figure 4.2.4 Assessment of limit of APEx detection with the addition of blocking reagent (1% PBSM) and increased usage of lysozyme as well as the optimal secondary fluorescent conjugate antibody dilution.
Optimal dilution of the fluorescent antibody was determined to be 1x, when directly compared in condition 1 and 2 (Table 4.2.4., A and B). Experimental parameters in condition 4 (C) and 7 (D) were identical, except from the addition of a blocking reagent in the latter, which reduced the positive cell population (P2).

The optimal lysozyme dilution was determined to be 3000 units/µl (12x more than the recommended concentration) and fluorescent conjugated secondary antibody should be used undiluted (one drop per sample, as recommended). The determination of the limit of detection for APEx was repeated. SAF84 epitope clone was spiked in pIII library and it was incubated with various SAF84 antibody dilutions (Table 4.2.5). No positive cell population was detected when SAF84 epitope clone was spiked in higher dilution than 1/100 in pIII library (Fig 4.2.5). Taken all these together, these modifications did not seem to dramatically change the positively identified population or assay's LOD.

Table 4.2.5 List of conditions of APEx to determine the limit of detection of the method.

Condition	SAF84 epitope clone dilution in pIII library	SAF84 antibody dilution
1	SAF84 epitope clone was used undiluted	10^{-4}
2	10^{-1}	10^{-4}
3	10^{-2}	10^{-4}
4	10^{-3}	10^{-4}
5	10^{-4}	10^{-4}

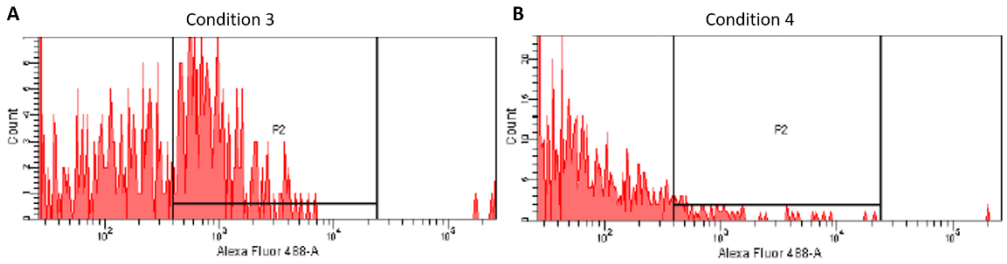


Figure 4.2.5 Assessment of limit of detection for APEx with optimised lysozyme and secondary antibody dilution when known SAF84 epitope clone was spiked into a pIII library.
Positive cell population was observed only when SAF84 epitope clone was spiked in 1/100 in pIII library (A) but not in 1/1000 (B).

SAF84 epitope clone was spiked in various dilutions into the pIII library and the optimal conditions were determined (300 unit of lysozyme, undiluted secondary antibody and not blocking reagent added) (Table 4.2.6.). The positive cell population was identified only when SAF84 epitope clone was used undiluted (Figure 4.2.6A) and not when it was spiked in different dilutions into the pIII library (Figure 4.2.6B and C). More specifically, around 3% of the gated population was positive in condition 1, demonstrating the poor limit of detection of the assay, even when undiluted SAF84 epitope clone was used. Nevertheless, positive clones were successfully recovered by PCR (as described in Method 2.2.4). The expected size of the PCR amplicons was 439 bp and it

was successfully identified (Figure 4.2.6D) from sorted bacterial cells from condition 1 (undiluted SAF84 epitope clone). This result shows that the ideal approach for recovery of positive clones after FACS could be either directly using 10^3 clones or DNA miniprepped from 10^5 sorted clones.

Table 4.2.6 List of APEx experimental conditions once the optimised conditions were identified.

Condition	SAF84 epitope clone dilution in pIII library	SAF84 antibody dilution
1	SAF84 epitope clone was used undiluted	10^{-4}
2	10^{-1}	10^{-4}
3	10^{-2}	10^{-5}

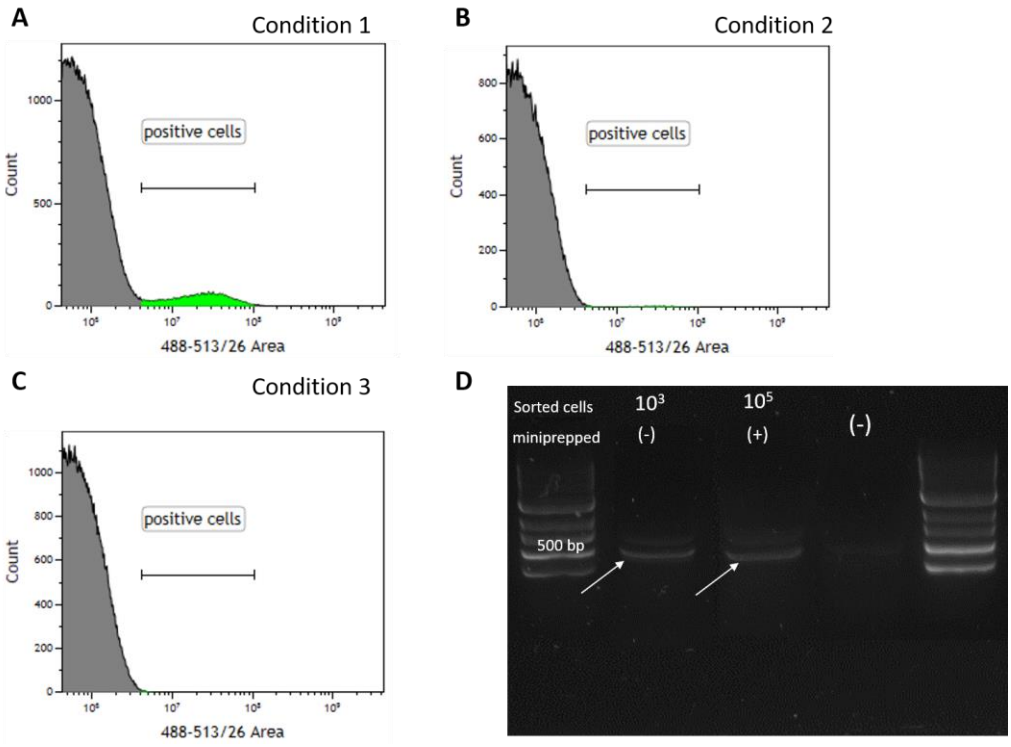


Figure 4.2.6 Assessment of different FACS sorting condition when isolating positive bacterial particles and their successful recovery by PCR. A. Positive cell population was only observed when SAF84 epitope clone was used undiluted and not when it was spiked in pIII library **(B and C). D.** PCR amplification of the peptide sequence derived from the positively sorted cells was carried out. Amplicon at the expected size (as indicated with white arrow) was observed for cells sorted at under both conditions (10^3 or 10^5 cells, and either used directly in PCR or miniprepped first).

The potential use of a different blocking reagent (normal chicken sera) was investigated in order to increase the LOD of APEx (Table 4.2.7). SAF84 epitope clone was preblocked with normal chicken sera, prior to its interaction with primary antibody. The percentage of the positive clones was increased (when compared with samples blocked with milk) (Fig 4.2.7.). However, when bacteria were infected by output phage that was previously panned for one or two rounds against SAF84 antibody (its dilution was 1/500,000 in normal mouse sera), positive cell population was not identified after either one round (Fig 4.2.7.C) or two rounds (Fig 4.2.7.D) of biopanning respectively.

Table 4.2.7 List of conditions for APEx to test blocking with chicken sera

Condition	Bacteria used	Blocking reagent	SAF84 antibody dilution
1	SAF84 epitope clone was used undiluted	-	10^{-4}
2		1% PBSM	10^{-4}
3		10^{-3} chicken sera	10^{-4}
4	R1 output phage	-	10^{-4}
5	R1 output phage	-	10^{-5}
6	R2 output phage	-	10^{-4}
7	R2 output phage	-	10^{-5}

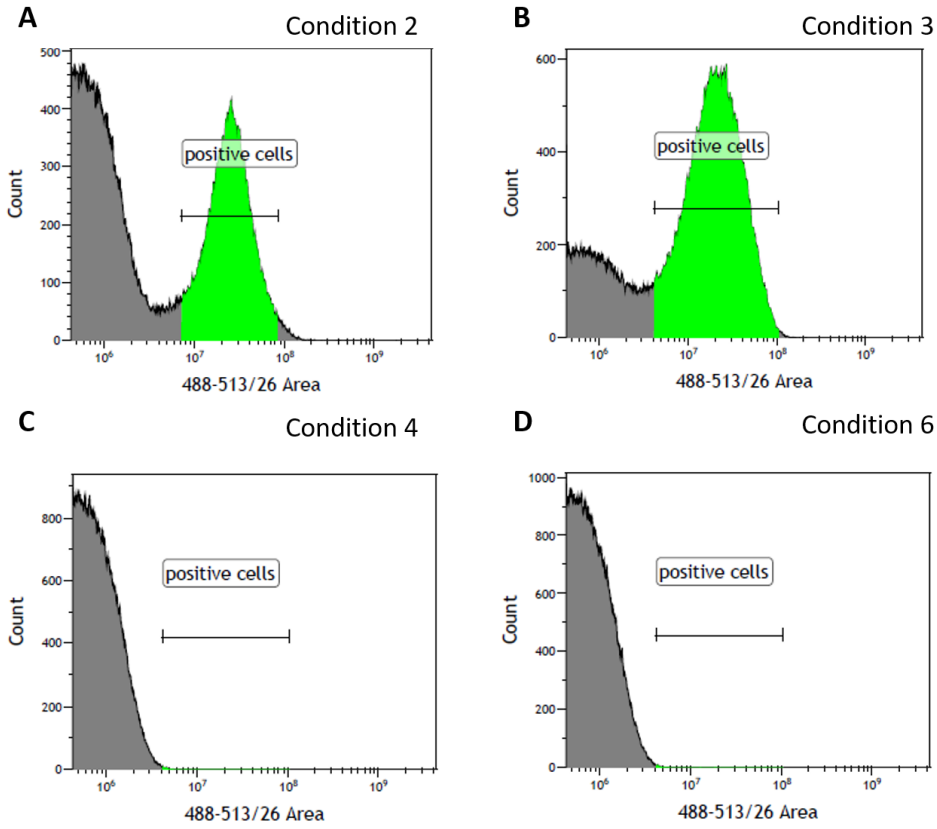


Figure 4.2.7 Assessment of limit of APEx detection with chicken sera as blocking reagent the determination of the LOD in biopanning experiments.

A. Bacterial clones (undiluted SAF84 epitope) were blocked with milk (A) or normal chicken sera (B). The positive population size doubled in the latter condition. A positive cell population was not observed for bacteria infected with output phage previously biopanned for 1 round against SAF84 antibody (5 x 10^{-5} dilution) (C) or for output phage previously biopanned for 2 rounds against SAF84 antibody (5 x 10^{-5} dilution) (D).

Overall, these experiments indicate that the sensitivity of the APEx methodology is very low, is not reproducible and positive cells were identified only when a known epitope was present at 1/100 dilution. Even though the recovery of the control positive clones was successful after cell sorting, positive clones after a biopanning experiment with a monoclonal antibody spiked into normal mouse sera at 5 x 10^{-5} were not able to be identified. Therefore, APEx will not compliment any further phage display experiments.

4.3. Epitope mapping of SAF84 mAb spiked into normal mouse sera with NGPD.

It was investigated whether peptide sequence enrichment can be observed in mouse sera containing a known antibody, when both antibody and its cognate epitope are present in low abundance. This was used as a model for autoantibodies in patient's sera, specifically for a single epitope (Reuschenbach et al., 2009). This biopanning experiment was designed to determine the limit of detection of the Next Generation Phage Display technology. Five replicates of eight different conditions (Table 4.3.1) were included in round 4, including decreasing amount of SAF84 and SAF70 mAbs (Figure 4.3.1). Naïve pc89_BspQI^{-} library was spiked with monoclonal bacteriophages displaying the known epitope for both SAF84 and SAF70 antibody, respectively (dilution factor was calculated to represent one copy per $5x10^{9}$, which is the estimated library's diversity). Four rounds of biopanning were conducted on protein G beads, washed with 0.1% PBST and eluted with triethylamine using spiked NMS with SAF84 and SAF70 mAbs; post selection subtraction step was included in some conditions (pre subtraction steps was always included). The subtraction step was an incubation step where phages were bound to SAF70 spiked mouse sera to ensure depletion of SAF70 binders. R1 and R2 replicates were kept separately; R2 replicates were pooled together to create the R2 input sub-libraries; these sublibraries were used for both positive (SAF84 and SAF70 spiked NMS) and negative (SAF70) selections. R3 replicates were pooled together to create the R3 input sub-libraries; these sublibraries were used for both positive (SAF84 and SAF70 spiked NMS) and

negative (SAF70) selections. Phage DNA from all R4 samples (n=40) were extracted, amplified with NGS specific primers and sent for Ion Torrent Sequencing. Sequences were analyses as described in Method 2.2.11. Enriched sequences within positive replicate (e.g., sample_1 from condition_1, SAF84 and SAF70 spiked sera) were compared with enriched sequences in its corresponding negative replicate, meaning the ones that the same input phage sub-library was used (e.g. sample_21 from condition_2, SAF70 spike sera); any peptide sequences that were present in 2 or more replicates and with a Z score larger than 4 were pooled. Five translated amino acid sequences, (TSRPGMQIVQEQTTNS, STMNSLNNADSSSIMV, PGMRLMMMNYEKLTLA, YLSPNDLKHNPTRTKT, PGARIVLDAMGIYEAP) were the most commonly enriched within at least 3 out of 5 positive samples. These were successfully cloned into the phagemid pc89 vector by inverse PCR, described in method 2.2.3. (Figure 4.3.2.), and their insertion was confirmed by Sanger Sequencing. They were then tested for their SAF84 specificity by phage ELISA but no binding was measured (data not shown).

Table 4.3.1 Conditions and samples ID for biopanning experiment to determine the limit of detection for NGPD epitope mapping.

Condition	Samples ID (Round 4)	Subtraction step	Dilution of mAbs in NMS
1	1-5	NO	SAF84/SAF70 $5x10^{-5}$
2	21-25	NO	SAF70 $5x10^{-5}$
3	6-10	NO	SAF84/SAF70 $5x10^{-6}$
4	26-30	NO	SAF70 $5x10^{-6}$
5	11-15	YES	SAF84/SAF70 $5x10^{-5}$
6	31-35	YES	SAF70 $5x10^{-5}$
7	16-20	YES	SAF84/SAF70 $5x10^{-6}$
8	36-40	YES	SAF70 $5x10^{-6}$

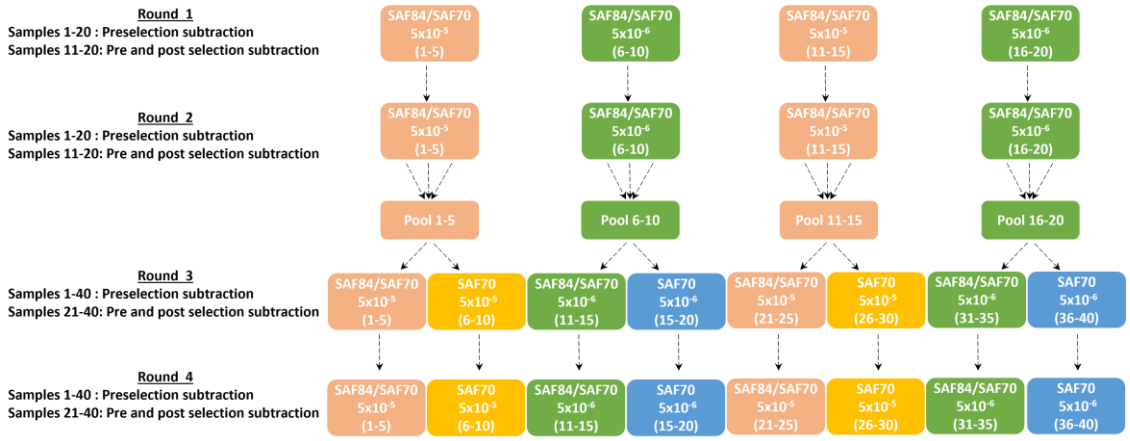


Figure 4.3.1 Overview of the biopanning experiment strategy.
Five replicates were included in all the biopanning rounds using four (R1 and R2) or eight different conditions (R3 and R4) with decreasing amount of spiked mAbs and/or pre and post selection subtraction (Table 4.3.1.). All R4 samples were deep sequenced.

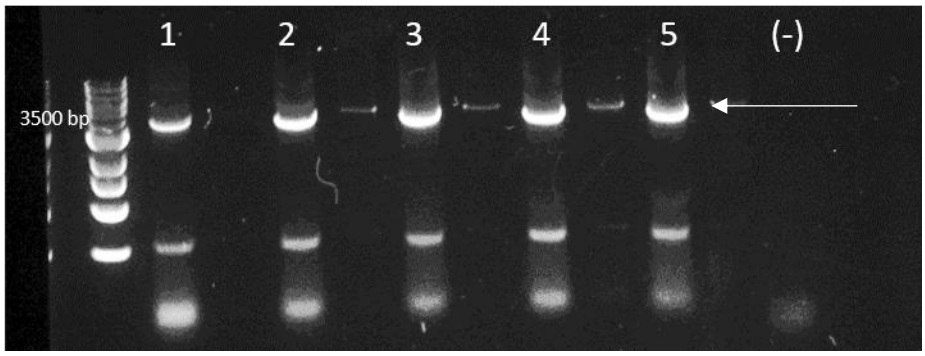


Figure 4.3.2 Amplification of the sequence of enriched peptides after NGS selection from the biopanning experiment.
Inverse PCR resulted in amplicon at the expected size (~3500 bp, white arrow) for all 5 targeted sequences; sample containing no template (-) was included as a control for possible unspecific amplification (lanes in between were left empty. Unspecific amplification was observed in low molecular weight.

Taken together, NGS data from this experiment did not allow the identification of peptides that were specific for SAF84. The number of enriched peptide sequences were seen in multiple conditions were low, no specific motif was

identified and when rescued, the clones (bacteriophages displaying peptides) were not recognised by SAF84. It is possible that enrichment against the mAbs (SAF84 and SAF70) took place during the earlier rounds of the biopanning and were outcompeted by 'parasitic phage' by round 4. To investigate this, 72 monoclonal bacteriophages were randomly picked from conditions 1, 3, 5 and 7 from R4, as well as polyclonal bacteriophages derived from all R4 conditions and they were tested by phage ELISA; no SAF84 binders were observed (Figure 4.3.3.). Controls were always included, and they were bacteriophages containing phagemid encoding the known epitopes for SAF84 or SAF70 respectively, each of them acting as a positive or negative control for the other. Additionally, 283 monoclonal bacteriophages from R1 (conditions 1 and 3) and 283 from R2 (conditions 1 and 3) were tested with phage ELISA; again, no binders were observed (Figure 4.3.4.). An additional assay called colony lifting was then followed in order to confirm our NGS and ELISA findings and also to analyse much higher numbers of phagemid clones (Method 2.2.12). Around 2000 colonies were tested with this method per condition (A: R1-condition 1, B: R1-condition 3, C: R2-condition 1, D: R2-condition 3). Phagemids from R1 and R2 conditions 1 and 3 were analysed. Two positive clones from R1 condition 1 (pool from 1-5 replicate samples) were identified (Figure 4.3.5), recovered and Sanger sequenced. One of the positive clones had a stronger intensity of colour on the colony lift assay, therefore they were named strong and weak binder clones, respectively. A SAF84 epitope motif was identified for the strong binder. The known SAF84 epitope is YY**R**PVD**QY**N, the weak binder amino acid sequence was MRNGERWTPPHTRTNT, and the strong binder amino

acid sequence was HHSETME**R**ARF**QY**AAS (bold AA are the common AA between the epitope and the sequenced clone). These clones were tested in a phage ELISA to confirm their SAF84 specificity; absorption above background was observed only for the strong binder clone (Fig 4.3.6). This represents a frequency of ~1 in 1000 for positive clones and confirms the enrichment for binders when SAF84 antibody was diluted $5x10^{-5}$ in sera, and that this enrichment can be seen in earlier rounds. Naïve library clones were also tested with the aforementioned colony lifting technique against SAF84 and SAF70 mAbs and positive clones were not identified (data not shown).

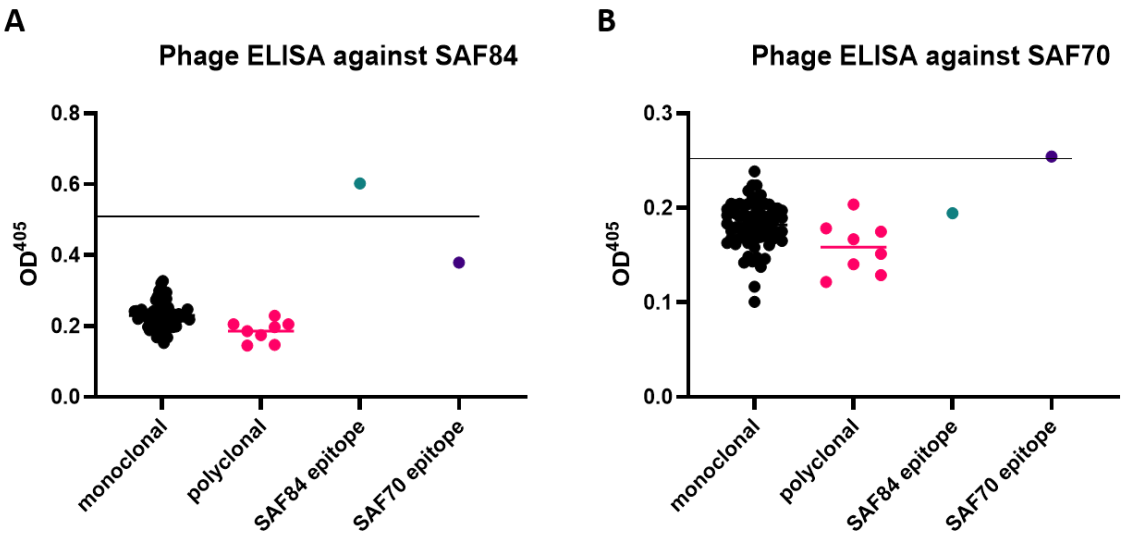


Figure 4.3.3 Specificity of polyclonal and 72 monoclonal bacteriophages derived from the 4th biopanning round were assessed with phage ELISA. **A.** Only positive control (SAF84 epitope) had positive signal when polyclonal and monoclonal bacteriophages were tested against mAb SAF8; line depicts 2x the background signal. **B.** Only positive control (SAF70 positive) had positive signal when polyclonal and monoclonal bacteriophages were tested against mAb SAF70; line depicts 2x the background signal.

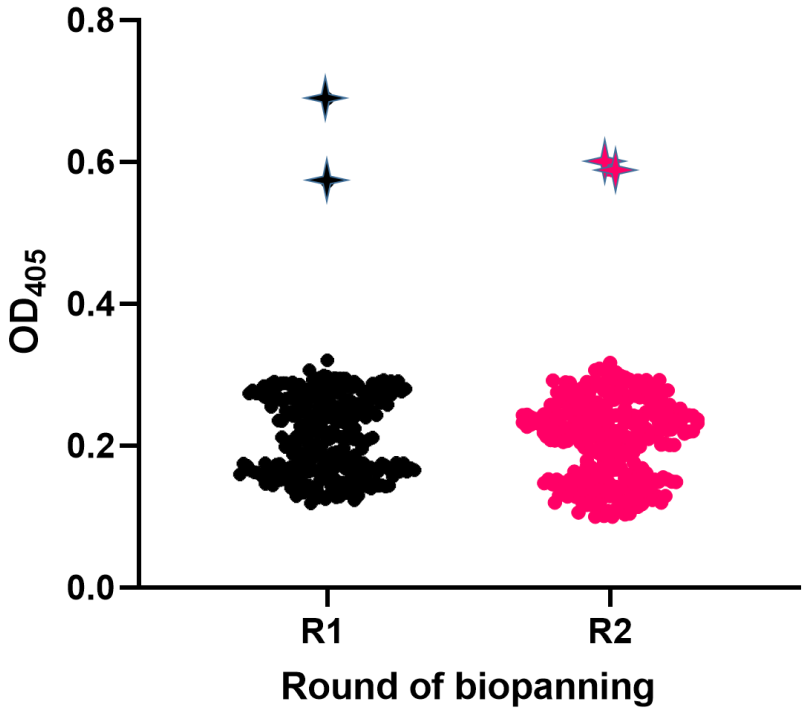


Figure 4.3.4 Specificity of 283 monoclonal bacteriophages derived from round 1 or 2 was assessed with phage ELISA.
X axis represents the round of biopanning these bacteriophages derived from, and y represents axis the OD_{405}. Only positive controls had positive signal on the SAF84 antigen coated plate (marked as stars).

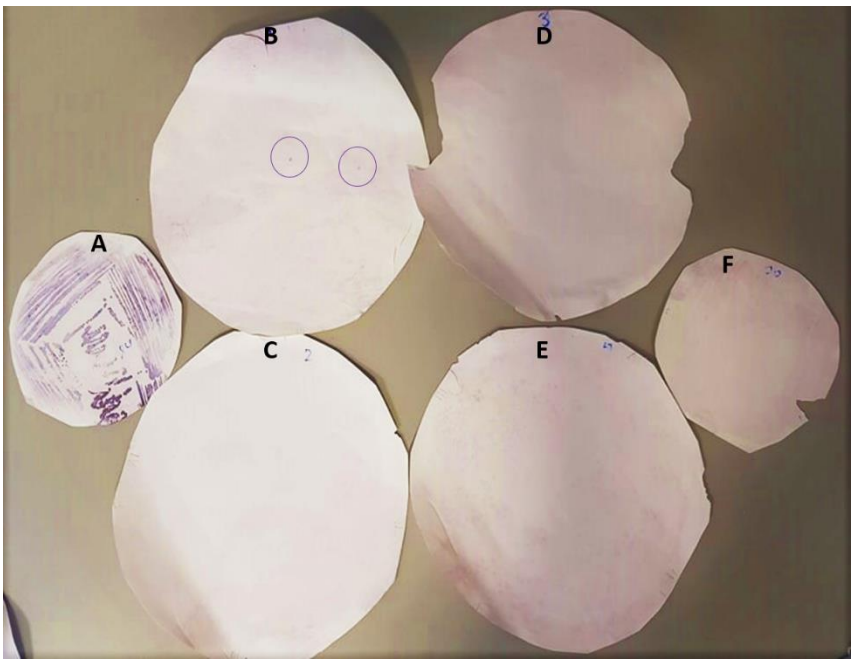


Figure 4.3.5 Colony lifting membranes after the addition of the substrate. Images of the colony lifting membranes after the colour development, including positive (A) and negative (F). A: R1-condition 1, B: R1-condition 3, C: R2-condition 1, D: R2-condition 3. Identified positive dots are circled.

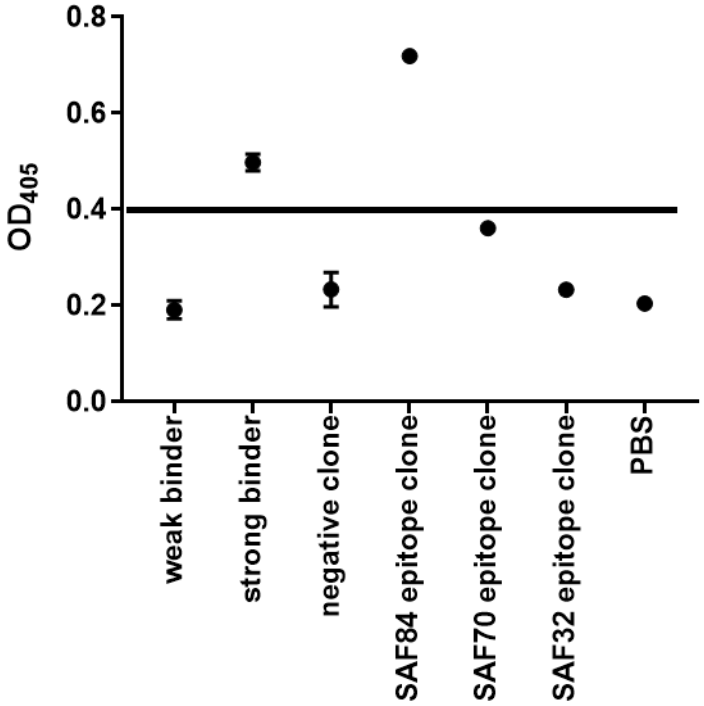


Figure 4.3.6 Previously identified clones from colony lifting were screened for their SAF84 specificity as bacteriophages displaying the corresponding peptides.
Weak and strong binders previously identified from the colony lifting assay were tested with three replicates by phage ELISA against immobilised SAF84 antibody, a negative clone was also tested with three replicates, and positive monoclonal SAF84 as well as two negative monoclonal bacteriophages (SAF70 and SAF32) were included. Black line is the 2x STD from the negative signal (when only PBS was added to the SAF84 coated well).

In summary, epitope mapping of polyclonal sera (normal mouse sera spiked with SAF84 and SAF70 antibodies) required further optimisation as specific enrichment and identification of positive clones was not observed in the NGS data from the fourth round of panning. This optimisation should use earlier rounds of panning as colony lifting of thousands of clones from the first round of panning identified at least one positive binder, indicating specific enrichment.

4.4. Epitope mapping of monoclonal antibodies in a buffer system

Optimisation of the biopanning technique for epitope mapping was essential in order to explore its limit of detection. It was previously observed that antibody enrichment can occur in the first round of panning to an approximate 1/1000 occurrence of positive clones. Here, four different anti-prion antibodies (named SAF84, SAF70, SAF15 and SAF32) were used because their epitopes were known from previous experiments. SAF84 epitope: YRPVDQY, SAF70 epitope: RYPNQVY and SAF15 epitope: QPHGGGWGQ. N.B. SAF32 antibody is not included at this data set due to the fact it showed no interaction with the protein G beads or plates. The conditions that were tested for the optimisation of the epitope mapping were mainly focused on the following areas: a) the decreasing diversity of the pc89_BspQI^{-} library b) the target immobilisation either on protein G beads (in solution) or on protein G plates (solid surface) c) washing conditions (less or more stringent with the addition of Tween, low concentrations of dithiothreitol [DTT] and/or high concentration of salts) and finally d) elution conditions: treatment with triethylamine, glycine, papain or DTT, or direct infection of TG1 cells (Table 4.4.1). Positive and negative controls were always included in the ELISA experiments; monoclonal bacteriophages displaying the known epitope of the SAF84 or SAF70 or SAF15 antibodies, each of them acting as a positive or negative control for the other. Furthermore, cut off of the ELISAs was always determined as 2x the mean of duplicate negative samples' OD_{405}. Epitope mapping of these mAbs were done separately for each

of them in a standard dilution (1/1000) and monoclonal bacteriophages were screened after a single biopanning round.

Table 4.4.1 Optimisation of conditions for the epitope mapping of three monoclonal antibodies in PBS.
Epitope mapping was done separately for each mAb at a standard 1/1000 concentration and positive clones were screened after a single biopanning round.

Condition	Immobilisation of the mAb	Library diversity	Washing	Elution
1	Protein G beads	10^9	0.1% PBST	Triethylamine
2	Protein G beads	10^9	0.1% PBST	Glycine
3	Protein G beads	10^9	0.1% PBST	TG1
4	Protein G beads	10^7	0.1% PBST	Glycine
5	Protein G beads	10^5	0.1% PBST	Glycine
6	Protein G beads	10^9	0.1% PBST	Papain
7	Protein G beads	10^7	0.1% PBST	Papain
8	Protein G beads	10^5	0.1% PBST	Papain
9	Protein G beads	10^9	0.1% PBST	DTT
10	Protein G beads	10^7	0.1% PBST	DTT
11	Protein G beads	10^5	0.1% PBST	DTT
12	Protein G plates	10^7	0.1% PBST	DTT
13	Protein G plates	10^7	500 uM DTT	DTT
14	Protein G plates	10^7	2% PBST	DTT
15	Protein G plates	10^7	2% PBST +1 M NaCl	DTT
16	Protein G plates	10^9	0.1% PBST	Triethylamine
17	Protein G plates	10^9	0.1% PBST	glycine
18	Protein G plates	10^9	0.1% PBST	DTT
19	Protein G plates	10^9	2% PBST +1 M NaCl	Triethylamine
20	Protein G plates	10^9	2% PBST +1 M NaCl	glycine
21	Protein G plates	10^9	2% PBST +1 M NaCl	DTT
22	Protein G beads	10^9	2% PBST +1 M NaCl	Triethylamine
23	Protein G beads	10^9	2% PBST +1 M NaCl	glycine
24	Protein G beads	10^9	2% PBST +1 M NaCl	DTT

Firstly, comparison of the optimal elution conditions when epitope mapping SAF84 antibody revealed that condition 2 (elution with glycine) was the most effective, with 79% of clones being positive by phage ELISA; condition 1 (elution with triethylamine) and condition 3 (direct infection of TG1 bacteria)

had 65% and 64% positive bacteriophages clones identified by phage ELISA, respectively (Figure 4.4.1). In order to confirm these findings, SAF84 was again epitope mapped with condition 2. Monoclonal bacteriophages (n=44) were tested for their specificity against SAF84 and SAF70 mAbs with phage ELISA. The percentage of the SAF84 specific bacteriophages was 73%; none of these bacteriophages showed any non-specific binding to the control antibody (SAF70) (Figure 4.4.2A). DNA from six possible SAF84 epitopes clones were extracted and Sanger sequenced. Four of these amino acids translated sequences contained amino acid motifs resembling the SAF84 known epitope (YRPVDQY) (Figure 4.4.2B).

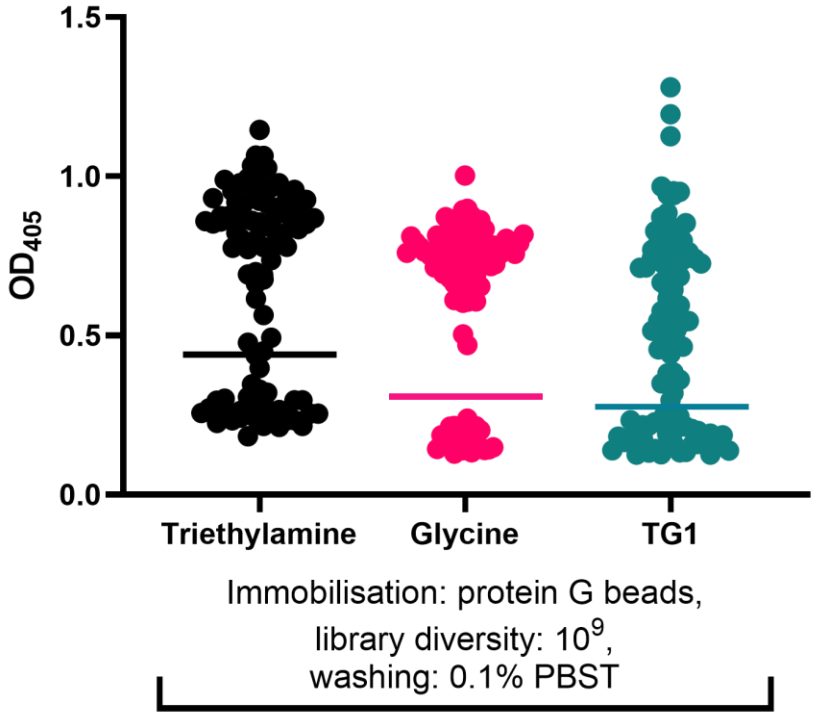
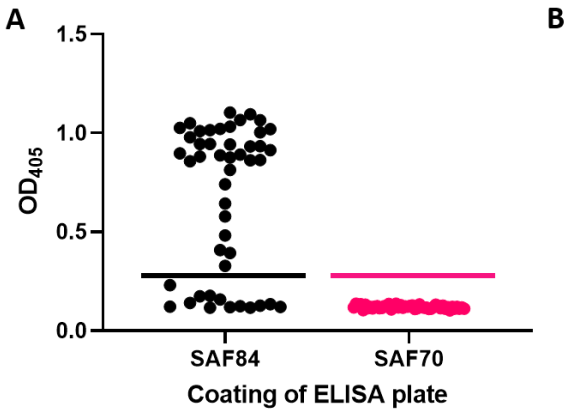


Figure 4.4.1 Optimisation of SAF84 epitope mapping in PBS. Comparison of different elution reagents when epitope mapping SAF84 antibody.
Cut off of the ELISA is represented by the horizontal lines and was determined as 2x the OD405 of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 91 monoclonal bacteriophages were screened per condition; the highest percentage of positive clones was observed in condition 2, glycine elution (79%) whereas results were comparable in condition 1, triethylamine elution (65%) and condition 3, elution with TG1 (64%), respectively (Table 4.4.1).



B

Clone ID	AA sequence
1	T*RPNYQYLRSEIKNV
2	RPMSHYPTPGPLSMHR
3	RERSHWSDPPLKTYAP
4	RPLANYTAETLHTPTM
5	PASERPVTQYPRLVGV
6	RDANNWRQPSGTMSKV

Figure 4.4.2 Epitope mapping of SAF84 with glycine elution; monoclonal bacteriophages were tested against SAF84 and SAF70 for assessment of their specificity.
A. Positive ELISA signal was observed in 72.7% of the hand-picked clones against SAF84 (n=44); none showed any non-specific binding to an irrelevant antibody (SAF70). **B.** DNA from 6 positive clones was extracted and Sanger sequenced; 4 out of 6 amino acid translated sequences contained a loose motif of the known epitope of SAF84 (YRPVDQY; red text).

It was also tested whether decreasing the library diversity had a positive impact on the percentage of positive clones that could be identified by epitope mapping. Therefore, pc89_BspQI^{-} library was diluted 2- and 4- log so the estimated final diversity was 10^7 and 10^5, respectively. Moreover, the optimal antibody concentration of the screening assay (ELISA) was yet to be determined; thus dilutions of coated antibody at 1/2,000 (standard condition), 1/16,000, 1/32,000 and 1/64,000 were tested. Epitope mapping of a different antibody, SAF70, was carried out using protein G beads, glycine as an elution reagent and reducing library diversity (condition 2 using 10^9, 4 using 10^7 and 5 using 10^5 library diversity, Table 4.4.1.). Monoclonal bacteriophages (n=92) were produced from each condition and were screened by phage ELISA. Condition 2 resulted in 23.9% positive clones, whereas decreased library

diversity in condition 4 and 5 resulted in 18.4% and 4.3% of positives, respectively (Figure 4.4.3.). Clones (n=24) with a wide range of ELISA signals were miniprepped and sent for sequencing, revealing that approximately half of them had a motif of the SAF70 epitope whereas 30% of them contained a PW motif in the first and second amino acid position, which could potentially represent a distinct mimotope of the SAF70 epitope (Table 4.4.2).

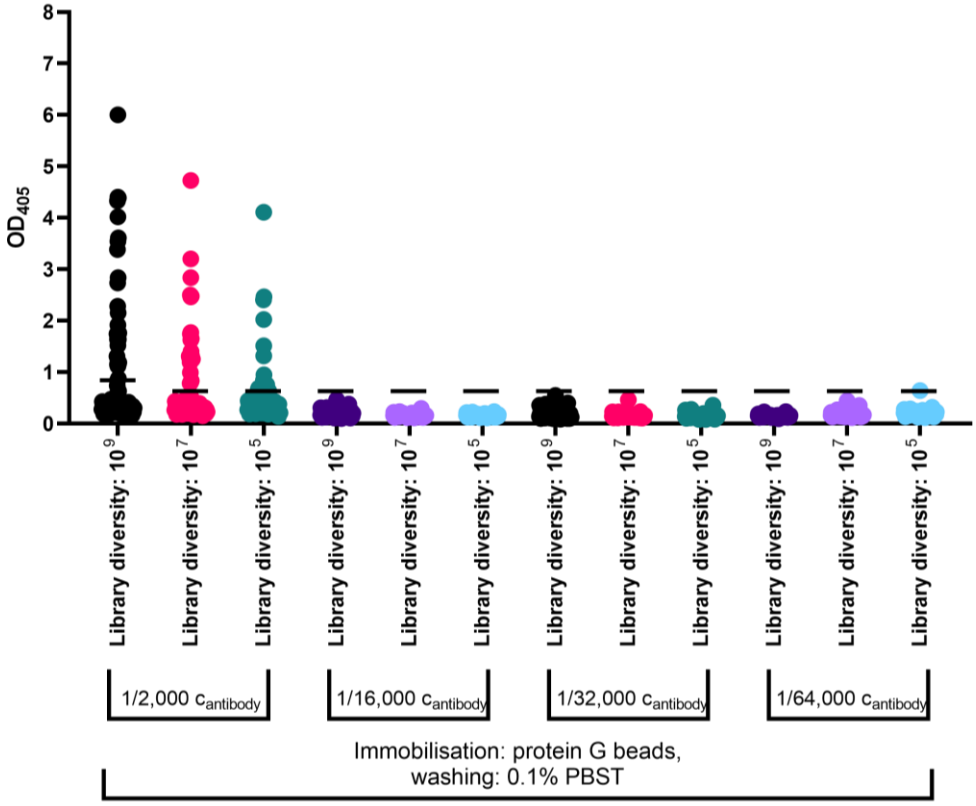


Figure 4.4.3 Epitope mapping of mAb SAF70 with decreased library diversity and decreasing amount of coated antibody.
Comparison of different conditions (2, 4 and 5) when epitope mapping SAF70 antibody with decreased library diversity (condition 2: 10^9, condition 4: 10^7 and condition 5:10^5 library diversity). Cut off of the ELISA is represented by the horizontal lines and it has been determined as 2x the OD_{405} of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 92 monoclonal bacteriophages per condition were screened against the standard coated 1/2,000 SAF70 antibody dilution as well as decreasing amounts (1/16,000, 1/32,000 and 1/64,000). The highest percentage of positive clones was observed in condition 2 (23.9%) whereas the percentage of positive bacteriophages

was lower in conditions 4 (18.4%) and condition 5 (4.3%) (Table 4.4.1). When the same bacteriophages were screened with a decreased amount of coated SAF70 antibody, none retained their positive ELISA signal.

Table 4.4.2 Amino acid translated sequences of 24 positive clones derived from different epitope mapping conditions with a wide range of phage ELISA signals.

Clone ID	Condition	AA Sequence[1]
1	condition 2	QHH*WNLFSQTPYLKH
2	condition 2	**PW**PQTPRTSTPIPNQT
3	condition 2	RDEQSRARKLWKLTYA
4	condition 2	SLNFPRLMERYPTQYP
5	condition 2	**PW**PGSAANSALTGHSR
6	condition 2	VRHLAIPGQSKLNKPR
7	condition 2	QSTLMMPYPHMATYNS
8	condition 2	VTSTYYPHQSARSYPS
9	condition 2	QTKRVLPRTDNQRTTL
10	condition 2	**PW**APSNLLVNS*NATE
11	condition 2	PSHSDYHTGWTVEARI
12	condition 2	YQPPNLNPIICSASLR
13	condition 2	YSWNLTYPNLALLASS
14	condition 4	**PW**PVRHIPTLLEVDTP
15	condition 4	**PW**GHATFHDTYPHTHQ
16	condition 4	RLSHQTGKEK*AGS*G
17	condition 4	**PW**QTARSILNPASTMA
18	condition 4	SKATFTRSTAQSQPTP
19	condition 4	SHGLWTVASSNNRRGI
20	condition 4	YPWQSLMQAGQKALDN
21	condition 4	**PW**RRPT*ERSGPMPVS
22	condition 5	RSHLKPAKAAFDSLQA
23	condition 5	TK*ITPLTLPRVHTPM
24	condition 5	**PW**GPHIQAGCTQSLQQ

[1]50% of these sequences contained a motif of the known SAF70 epitope (RYPNQVY, shown in red text), whereas 30% contained a PW motif in positions 1 and 2 (newly discovered mimotope.

Further optimisation was carried out to assess alternative elution methods. Papain has been reportedly used in order to reduce background in immunocomplexes (Fab fragments with peptides), followed by screening with mass spectrometry (Aibara et al., 2018). Furthermore, DTT has been previously

described as a reducing agent that can stop antibody binding (Liu et al., 2010). Therefore, these reagents were utilised for elution methods where instead of only bacteriophages being eluted, the whole phage-peptide-Ab complex is eluted instead. These elution conditions were assessed for the epitope mapping of SAF70 alongside decreasing library diversity (conditions 6-8, decreasing library diversity and papain elution, conditions 9-11, decreasing library diversity with DTT elution, Table 4.4.1.). Monoclonal bacteriophages (n=92) from each condition were screened by phage ELISA. Elution by DTT was more effective compared with elution by papain; specifically, the percentage of positives for conditions 6, 7 and 8 were 8.7%, 14.1%, and 18.5% whereas the observed percentage of positives for conditions 9, 10 and 11 were 32.6%, 35.9%, and 25%, respectively (Figure 4.4.4.). The same monoclonal bacteriophages were screened with decreased coated SAF70 antibody dilution (1/32,000) in the phage ELISA and there was an overall decrease in clones that were positive (3.3%, 3.3%, 2.2%, 10.8%, 10.8% and 7.6%, for conditions 6-11, respectively). Clones with a wide range of ELISA signals from all the different conditions were miniprepped and sent for sequencing, revealing that 31.5% of them had a motif of the already known SAF70 epitope and 52.6% containing a PW motif in the first and second amino acid position (Table 4.4.3).

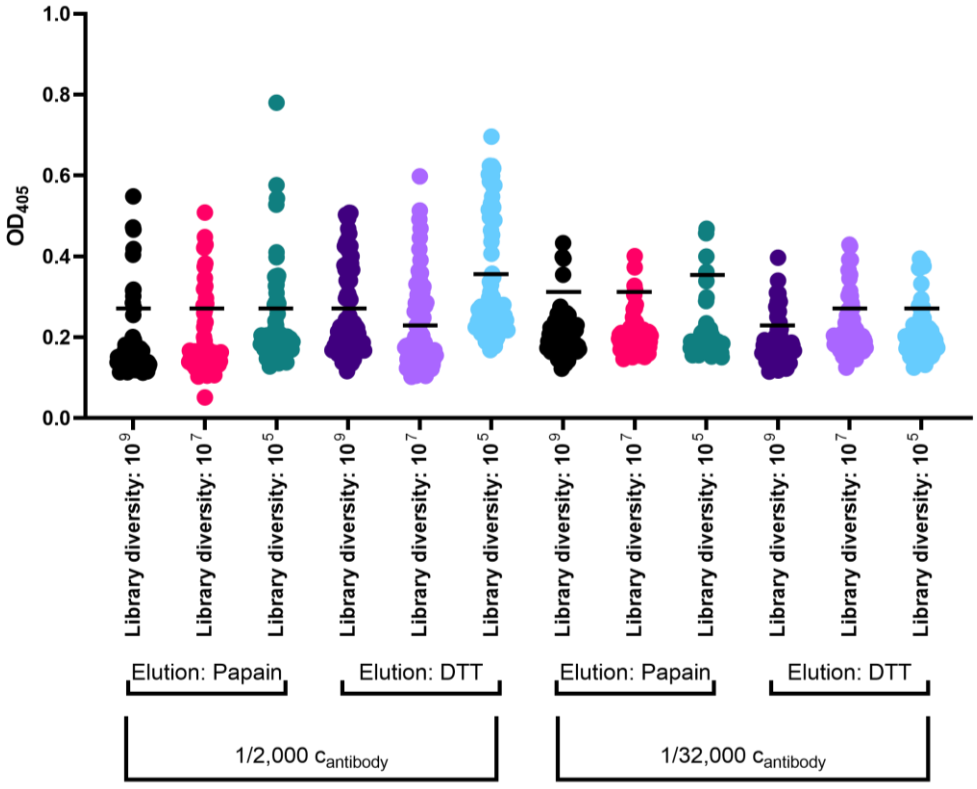


Figure 4.4.4 Epitope mapping of mAb SAF70 with decreased library diversity, different elution methods (condition 6-8: papain, condition 7-11: DTT) and decreasing amount of antibody coated in the screening ELISA.

Comparison of different library diversity and elution conditions when epitope mapping SAF70 antibody. Cut off of the ELISA is represented by the horizontal lines and it has been determined as 2x the OD_{405} of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 92 monoclonal bacteriophages per condition were screened against the standard coated 1/2,000 SAF70 antibody dilution as well as decreasing amounts of it (1/32,000). The highest percentage of positive clones (32.6%, 35.9%, and 25%) was observed when bacteriophages were eluted with DTT (conditions 9, 10, 11 respectively) whereas decreased numbers of positive clones (8.7%, 14.1% and 18.5%) were observed with papain as the elution method (conditions 6, 7, 8) (Table 4.4.1). Additionally, the same bacteriophages were screened with decreased amount of coated SAF70 antibody, 3.3%, 3.3%, 2.2%, 10.8%, 10.8% and 7.6% of them retained their positive ELISA signal for conditions 6-11, respectively.

Table 4.4.3 Amino acid translated sequences of 19 positive clones derived from different epitope mapping conditions with a wide range of positive ELISA signals.

Clone ID	Condition	AA Sequence[1]
1	condition 6	**PW**PASFLASPHNTPEY
2	condition 6	TTPYPNLTPSAMHSAT
3	condition 6	PLTVRGKNYIPFLHLD
4	condition 6	YPVMHLTCH
5	condition 7	KWPVPPRTLNTPPVVT
6	condition 7	PYPVMRMMLPNRGTPQ
7	condition 7	**PW**PKTNPM*TRLHKAY
8	condition 7	**PW**KEVPAREHTMLASE
9	condition 8	**PW**PHPSAPRNNSYQLQ
10	condition 8	**PW**GPPIQAGCTQSLQQ
11	condition 8	**PW**GPPIQAGCTQSLQQ
12	condition 8	ITNYYKRPPSLDGTCR
13	condition 9	RQVYPNLELSTNFPIG
14	condition 9	**PW**LLGNSARMPMQPLT
15	condition 9	LGARDMRTEQPTGSHS
16	condition 10	**PW**PDHRRLDATPTYDQ
17	condition 10	**PW**SSLRSVFVSPVPRA
18	condition 10	**PW**SVSADEARPALPKP
19	condition 11	THPNPLYIRYETPEHP

[1]31.5% of these sequences contained a motif of the known SAF70 epitope (RYPNQVY, red text), whereas 52.6% of them contained a PW motif in positions 1 and 2.

More extensive and more stringent washing conditions could also potentially positively increase the percentage of the identified clones. Therefore, four different conditions were tested using four different washing reagents, with DTT elution and 10^7 library diversity (condition 12: 0.1% PBST, condition 13: PBS with 500 um DTT, condition 14: 2% PBST and condition 15: 2% PBST + 1M NaCl, Table 4.4.1.). In addition, SAF70 antibody was immobilised on protein G plates, instead of using soluble protein G beads, in order to assess this method

of this immobilisation as a higher throughput format. 92 monoclonal bacteriophages were produced from each condition and were screened by phage ELISA; dilutions of coated SAF70 antibody were 1/2000 and 1/32,000. Washing reagent 2% PBST + 1M NaCl seemed to be the most effective; specifically, the percentage of positives for this conditions 15 were 32.6% whereas the observed percentage of positives for conditions 12, 13 and 14 were 8.7%, 14.1%, and 18.5%, respectively (Figure 4.4.5.). The same monoclonal bacteriophages were screened with decreased coated SAF70 antibody dilution by phage ELISA in order to possibly identify high affinity clones that still interact with low antibody concentration and there was an overall decreased observed positive ELISA signal (0%, 2.2%, 0%, and 3.3%, respectively). Clones with a wide range of ELISA signals from all of the different conditions were miniprepped and sent for sequencing, revealing that 12% of these clones had a motif of the already known SAF70 epitope, whereas 62.5% of them containing a PW motif in the first and second amino acid position (Table 4.4.3).

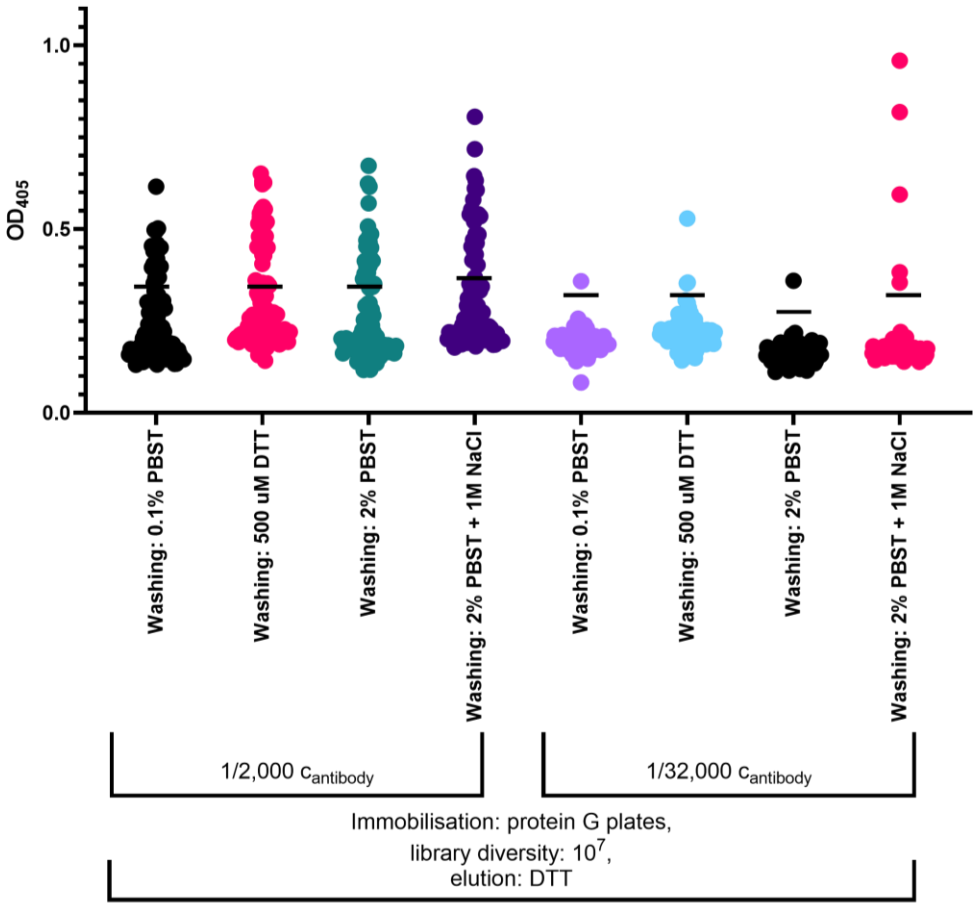


Figure 4.4.5 Epitope mapping of mAb SAF70 with 10^7 library diversity, different washing conditions (condition 12: 0.1% PBST, condition 13: PBS with 500 um DTT, condition 14: 2% PBST and condition 15: 2% PBST + 1M NaCl) and decreasing amount of antibody coated in the screening ELISA to identify potentially higher affinity binders that interact with low levels of Ab.

Comparison of washing conditions when epitope mapping SAF70 antibody. Cut off of the ELISA is represented by the horizontal lines and it has been determined as 2x the OD_{405} of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 92 monoclonal bacteriophages per condition were screened against the standard coated 1/2,000 SAF70 antibody dilution as well as a decreased amount (1/16,000). The highest percentage of positive clones (32.6%) was observed when protein G plates were washed with 2% PBST + 1M NaCl (condition 15), whereas decreased numbers of positive clones (8.7%, 14.1% and 18.5%) was observed with different washing reagents (conditions 12, 13, 14, respectively). Additionally, the same bacteriophages were screened with decreased amount of coated SAF70 antibody (1/32,000) - 0%, 2.2%, 0% and 3.3% of them retained their positive ELISA signal for conditions 12-15, respectively.

Table 4.4.4 Amino acid translated sequences of 16 positive clones derived from different epitope mapping conditions with a wide range of ELISA signals.

Clone ID	Condition	AA Sequence[1]
1	condition 12	**PW**SVRPSLAAREDMAP
2	condition 12	**PW**PAQLRNRTPTSNEV
3	condition 12	YILHYSRPAGYVEYLS
4	condition 12	QRLVRLHMFPSSSLVP
5	condition 13	**PW**PGMPAPSERPLGST
6	condition 13	**PW**ILHPILSWHANSPC
7	condition 13	YPNQHVELDRSSQPVP
8	condition 13	**PW**MKYTEQSSMINASI
9	condition 14	**PW**PCELMRMPSGATRD
10	condition 14	TQDTFYPSQARACIVP
11	condition 14	SRPKEYLLAGMRRSHI
12	condition 14	**PW**PPNTLQLNIGPPAP
13	condition 15	**PW**NGPGQRSRTLHQSF
14	condition 15	NWPLEPSSSVKNTSMY
15	condition 15	**PW**GYTHRQCPVTGPTP
16	condition 15	**PW**PSTFLSRHLCQNCY

[1]12% of these sequences contained a motif of the known SAF70 epitope (RYPNQVY, red text), whereas 62.5% of them contained a PW motif in positions 1 and 2.

The optimal epitope mapping conditions derived for SAF70 were retested using a distinct antibody, SAF15. This antibody was immobilised on protein G plates and then unspecific bacteriophages were washed away with three different wash reagents (condition 12: 0.1% PBST, condition 13: PBS with 500 um DTT and condition 15: 2% PBST + 1M NaCl, Table 4.4.1.). 92 monoclonal bacteriophages were produced from each condition and were screened by phage ELISA; dilution of coated SAF15 antibody was 1/2,000. The percentage of the positive clones were comparable in all the three tested conditions; 2.2%, 3.3% and 2.2% in condition 12, 15 and 13, respectively. Clones with a wide

range of ELISA signals from conditions 12 and 15 were miniprepped and sent for sequencing, revealing that 67% of these clones had a motif of the already known SAF15 epitope (Table 4.4.5).

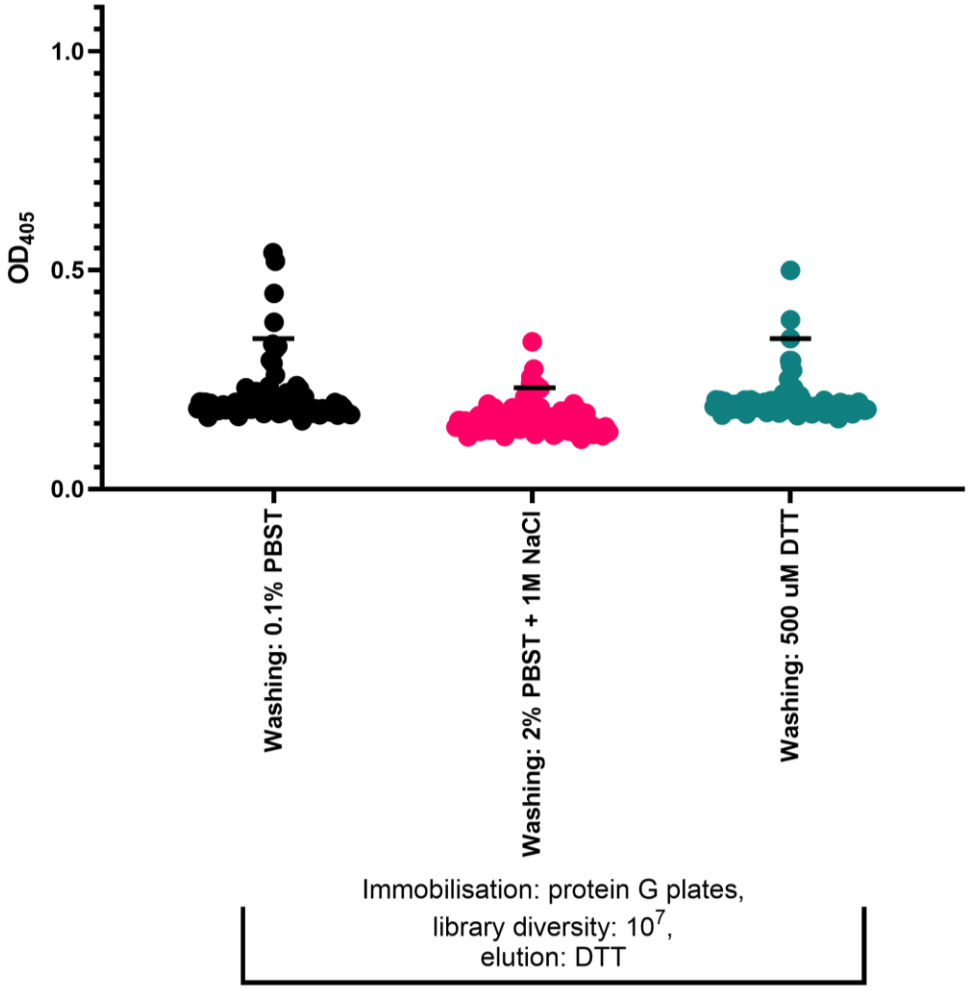


Figure 4.4.6 Epitope mapping of mAb SAF15 with 10^7 library diversity and different washing conditions (condition 12: 0.1% PBST, condition 13: PBS with 500 um DTT and condition 15: 2% PBST + 1M NaCl).
Comparison of washing conditions when epitope mapping SAF15 antibody. Cut off of the ELISA is represented by the horizontal lines and it has been determined as 2x the OD_{405} of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 92 monoclonal bacteriophages per condition were screened against the standard coated 1/2,000 SAF15 antibody dilution. The percentage of positive clones (2.2%, 3.3% and 2.2%) across all conditions (12, 15 and 13, respectively) was comparable (Table 4.4.1.).

Table 4.4.5 Amino acid translated sequences of 6 positive clones derived from 2 different epitope mapping conditions with a wide range of affinities, as determined by phage ELISA.

Clone ID	Condition	AA Sequence[1]
1	condition 12	YESTQPNRSGWNPLSA
2	condition 12	WLVTTKGWHPPSASSS
3	condition 12	RSAQDNYRAGWGDPML
4	condition 15	SPYQPSRHGWYNLNVH
5	condition 15	APLDPRLPRAQFRSSS
6	condition 15	MMPRAQFNATVPNNYY

[1]67% of these sequences contained a loose motif of the known SAF15 epitope (QPHGGGWGQ) (highlighted with red colour).

In order to confirm our findings, a direct comparison of 12 of the most promising conditions were used for the epitope mapping of SAF70. The pc89_BspQI^{-} library was biopanned with its original diversity (since 2-fold dilution of the library had comparable results with undiluted phage libray in previous experiments), SAF70 antibody was immobilised on protein G plates (conditions 16-21) or bound to soluble protein G beads (condition 1, 2, 9, 22, 23 and 24) and different elution (Triethylamine: conditions 1, 22, 16 and 19, Glycine: conditions 2, 23, 17 and 20, DTT: conditions 9, 24, 18 and 21) as well as washing conditions (0.1% PBST: conditions 1, 2, 9, 16, 17, 18 and 2% PBST + 1M NaCl: conditions 19-24) were assessed (Table 4.4.1.). 44 bacteriophages were produced from each condition and were screened by phage ELISA; dilution of coated SAF70 antibody was 1/2,000. The highest percentage of positive clones was determined to be with condition 23 (immobilisation: beads, elution with glycine, 45.8%), followed by condition 21 (immobilisation: plates, elution with DTT, 29.5%) and condition 20 (immobilisation: plates,

elution with glycine, 27.1%) (Figure 4.4.7). Notably, the washing reagent for these conditions was the same (2% PBST + 1M NaCl). The most promising of the conditions were applied for the epitope mapping of SAF15 antibody in order to assess different conditions with this additional antibody (Table 4.4.1). The highest percentage of positive clones was identified with condition 21 (immobilisation: plates, elution with DTT, 9.1%); no positive signal was observed with other conditions (Figure 4.4.8). The amount of positive clones of condition 21 were determined with a picture, which was taken the day after the ELISA experiment, since the readings of condition 21 did not depict the true colour development of the positive wells (OD was manually calculated).

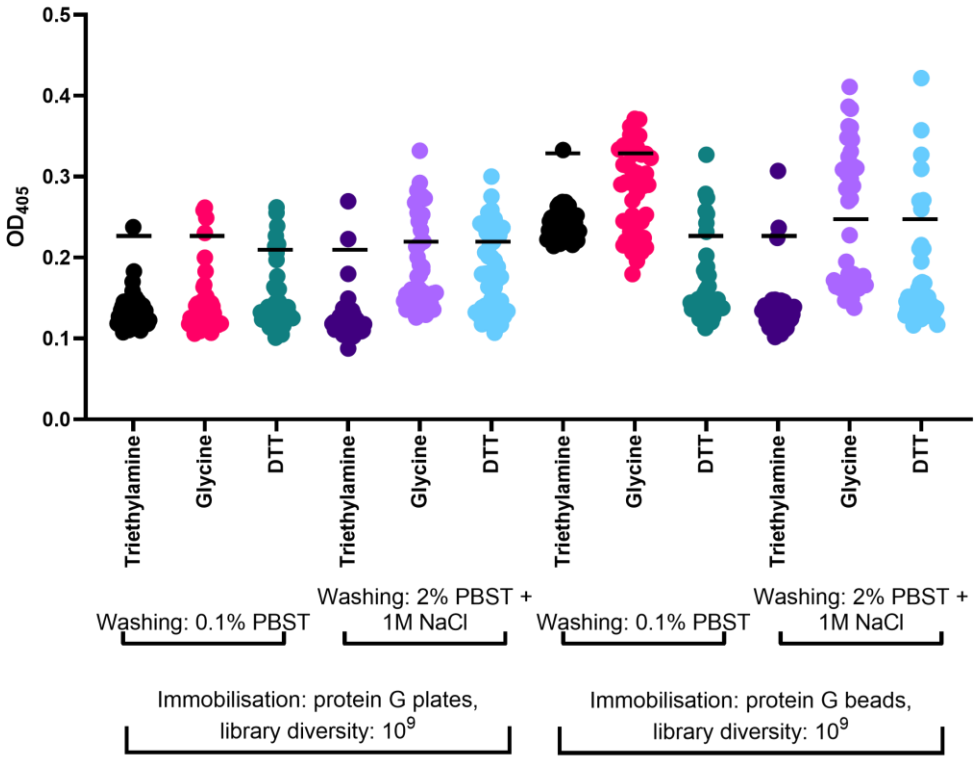


Figure 4.4.7 Epitope mapping of mAb SAF70 with 10^9 library diversity, different washing conditions (0.1% PBST or 2% PBST + 1M NaCl), different immobilisation of SAF70 antibody (Protein G beads or Protein G plates) and different elution methods (triethylamine, glycine, DTT).
Comparison of 12 different conditions when epitope mapping SAF70 antibody (Table 4.4.1). Cut off of the ELISA is represented by the horizontal lines and it has been determined as 2x the OD_{405} of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 44 monoclonal bacteriophages per condition were screened against the standard coated 1/2,000 SAF70 antibody dilution. The highest percentage of positive clones (45.8%) was observed when protein G beads were washed with 2% PBST + 1M NaCl and positively selected clones were eluted with glycine (condition 23), followed by 29.5% of positives when protein G plates were washed with 2% PBST + 1M NaCl and positively selected clones were eluted with DTT (condition 21) and 27.1% of positive clones in condition 20 (protein G plates were washed with % PBST + 1M NaCl and positively selected clones were eluted with glycine). Decreased comparable amount of positive clones (2.1%, 4.5%, 12.5%, 0%, 2.1%, 18.2%, 14.6%, 2.3%, 9.1%) were observed with different conditions (16, 17, 18, 19, 1, 2, 9, 22, 24, respectively).

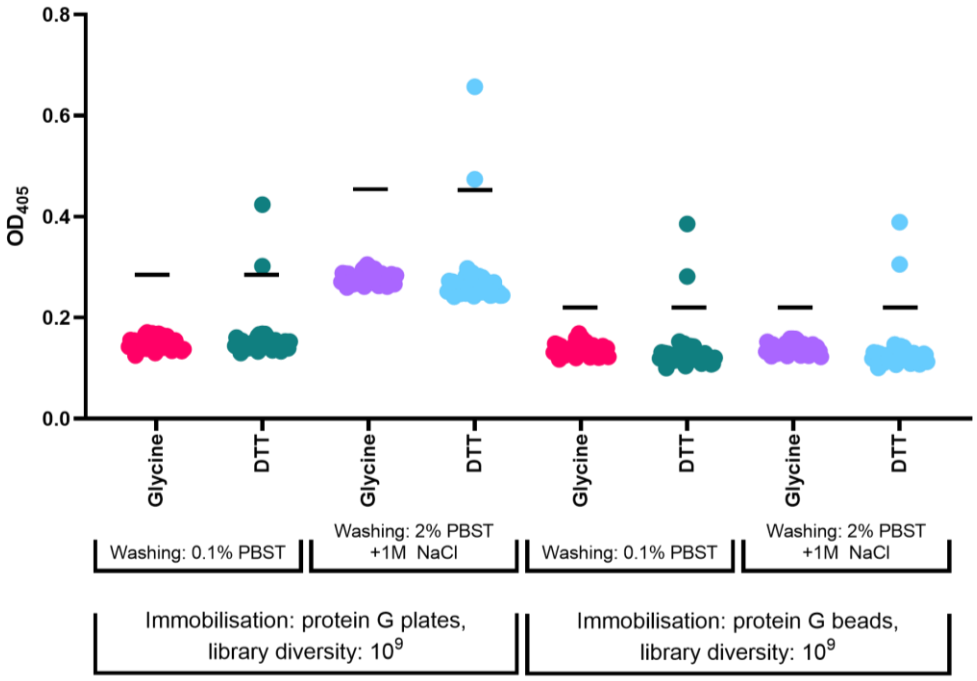


Figure 4.4.8 Epitope mapping of mAb SAF15 with 10^9 library diversity, different washing conditions (0.1% PBST or 2% PBST + 1M NaCl), different immobilisation of SAF15 antibody (Protein G beads or Protein G plates) and different elution methods (glycine or DTT).
Comparison of 8 different conditions when epitope mapping SAF15 antibody (Table 4.4.1). Cut off of the ELISA is represented by the horizontal lines and it has been determined as 2x the OD405 of duplicate control wells containing bacteriophages displaying irrelevant peptide to the antibody epitope. 44 monoclonal bacteriophages per condition were screened against the standard coated 1/2,000 SAF15 antibody dilution. The highest percentage of positive clones (9.1%) was observed when protein G plates were washed with 2% PBST + 1M NaCl and positively selected clones were eluted with DTT (condition 21). No positive signal was observed for the rest of the conditions (17, 18, 20, 2, 9, 23, and 24). OD for positively identified clones were manually calculated since the plate reading was failed.

Taken together, the highest percentages of positive clones are obtained across distinct antibodies when using panning condition 21: library diversity $5x10^9$; immobilisation on protein G coated plates; washing condition: 2% PBST + 1 M NaCl (5 times), followed by PBS (5 times) and elution with DTT. Therefore, this

condition was used in future experiments as the optimal method for epitope mapping monoclonal antibodies.

4.5. Epitope mapping of mAbs SAF70 and SAF15 spiked into normal mouse sera with optimised NGPD conditions.

Following the optimisation of conditions for epitope mapping of monoclonal antibodies in buffer, these conditions were applied to the epitope mapping of monoclonal antibodies spiked into mouse sera at high dilutions, mimicking the low titration of autoantibodies in cancer sera. Additionally, a subtraction method with NMS was introduced in order to deplete bacteriophage-peptides that represent epitopes to the background murine antibody repertoire. Briefly, protein G plates were coated with 0.5 µl of NMS, blocked with 3% PBSM, and preblocked (with 18% PBSM and protein G beads) bacteriophages from pc89_BspQI^{-} library were introduced to the NMS coated wells. Sequentially, depletion was carried out 3 times and unbound bacteriophages were then collected and introduced in pre-blocked protein G plates that had been coated with antibody spiked NMS. Unbound bacteriophages were washed and eluted as described above using with optimised condition 21, library diversity 5x10^{9}; immobilisation on protein G coated plates; washing condition: 2% PBST + 1 M NaCl (5 times), followed by PBS (5 times) and elution with DTT. Normal mouse sera was spiked with two monoclonal antibodies (SAF70 and SAF15), the epitopes of which are already known, in two different dilutions. Dilution of SAF15 was less since it was apparent from previous experiments that its binding capacity to protein G was not as strong as for SAF70 antibody, as the amount of positively identified clones were also lower than the ones from

SAF70 or SAF84 epitope mapping (data not shown). Two different blocking reagents were also tested (either PBS or NMS). Samples without spiked NMS or with PBS only coated wells were included as negative controls. Triplicates from eight different conditions and six replicates from standard epitope mapping conditions of SAF70 and SAF15 in PBS were included in the biopanning (Table 4.5.1.). Output phage titration from all the different conditions was comparable (around 10^5). Colony lifting was performed in the first 8 conditions in order to assess the success of the epitope mapping (Method 2.2.6). Around 2000 colonies (bacteria containing phagemid encoding for peptide) were tested with this method for each condition. Membranes of conditions 1-4 were probed with either SAF70 or SAF15 antibody; membranes of conditions 5-8 were probed with both SAF70 and SAF15 antibodies together. The percentage of the positive clones was calculated for all the conditions, assuming 2000 colonies per membrane (Figure 4.5.1.). The highest percentage of SAF70 positive clones was observed in condition 1, 3 and 4 whereas the percentage of SAF15 positive clones was lower than the ones of SAF70, but higher than the control groups.

Table 4.5.1 Description of the different conditions that were carried out in this biopanning experiment.

Condition	Diluent	Antibody dilution	Blocking reagent	Samples ID
1	NMS	SAF70 1/250, SAF15 1/50	3% PBSM	1-3
2		SAF70 1/250, SAF15 1/50	3% PBSM + 1/1000 NMS	4-6
3		SAF70 1/500, SAF15 1/100	3% PBSM	7-9
4		SAF70 1/500, SAF15 1/100	3% PBSM + 1/1000 NMS	10-12
5		-	3% PBSM	13-15
6		-	3% PBSM + 1/1000 NMS	16-18
7	PBS	-	3% PBSM	19-21
8		-	3% PBSM + 1/1000 NMS	22-24
9		1/1,000 SAF70	3% PBSM	25-30
10		1/1,000 SAF15	3% PBSM	31-36
11		-	3% PBSM	37-42

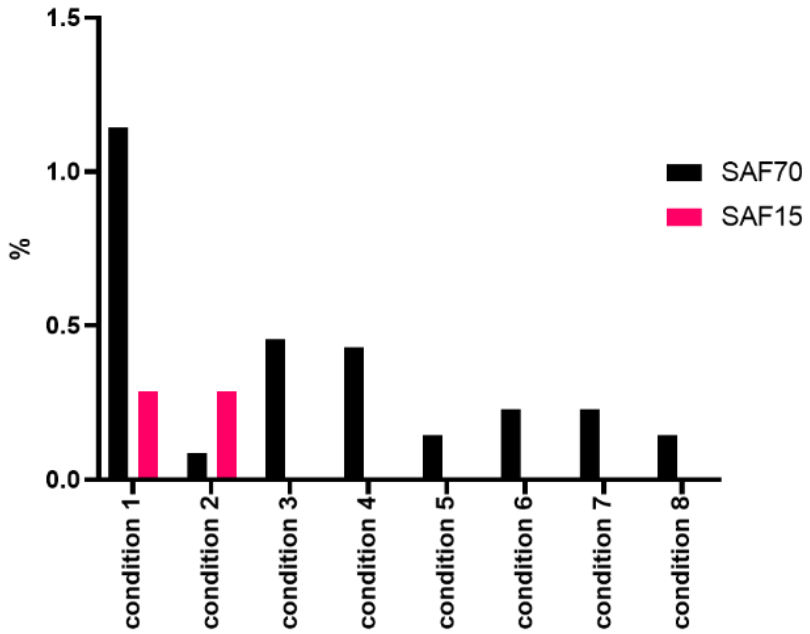


Figure 4.5.1 Assessment of the SAF70 and SAF15 epitope mapping by a colony lifting.
Percentage of the positively identified clones when membranes when probed with SAF70 and/or SAF15. The largest proportion of positive clones was identified with condition 1; SAF15 specific clones were also identified, only in conditions 1 and 2, in the lowest SAF15 dilution (Table 4.5.1.)

DNA was extracted from all the individual samples and amplified in R1 (Figure 4.5.2A) and R2 (Figure 4.5.2B) PCR. Barcoded DNA was pooled together in equal quantities, and sent for Ion Torrent Sequencing.

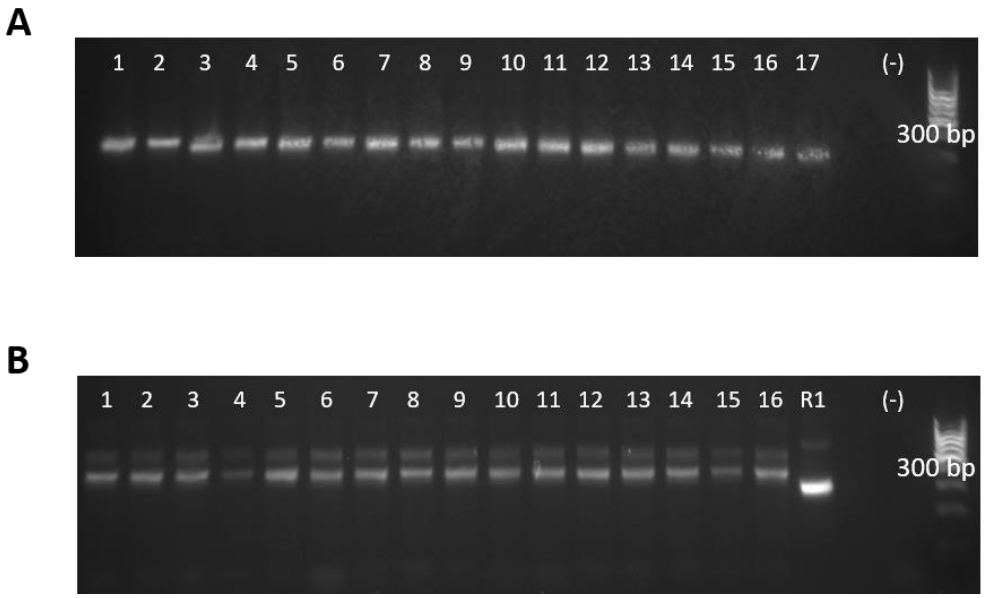


Figure 4.5.2 PCR amplification of sublibraries for NGS preparation (representative image after agarose gel electrophoresis).
A. Amplicons from samples 1-17 were observed at the expected size at 275 bp for R1 PCR, negative control was included (contained water instead of DNA template). **B.** Amplicons from samples 1-16 were observed at the expected size at 335 bp for R2 PCR. Sample containing water instead of a PCR R1 product as a template was performed in order to determine any unspecific amplification (negative). R1 amplicon was also included to the gel electrophoresis in order to confirm the insertion of the barcodes (R2 amplicons are larger by 60 bp).

Sequences were demultiplexed and translated, followed by a selection of those that were between the flanking regions of AEGEF and DPAKA (Method 2.2.11.). Enriched sequences of either 3 or 6 replicates were pooled together (11 different sets) (Figure 4.5.3.). Analysis was designed to look for specific peptides that were common between related datasets and not seen in the control datasets. Firstly, condition 1 and 2 were assessed (NMS was spiked with

SAF70: 1/250 and SAF15: 1/50). These conditions were comparable since only the blocking reagents that were used, were different. Sequences from subset 1 and subset 2 (training subsets) were ranked and the top 50 sequences were compared revealing 46 common peptides (Figure 4.5.4A.). The frequency of these 46 sequences was then calculated as a percentage in positive subset 3 (test subset) as well as in negative subsets in which wells were coated with NMS only (set 7, 8 and 9) or PBS (set 10, 11 and 12). All frequencies were normalised depending on the amount of total sequences per barcode. The sum of the percentages of these 46 sequences was calculated for all the aforementioned data subsets (training and testing) (Figure 4.5.4B). These peptide sequences seem to be enriched when mAbs SAF70 and SAF15 were spiked into NMS, and were not present in NMS or PBS only controls (subsets 7-12). The percentage frequency of the 46 commonly enriched peptides were also plotted as individual values for the positive and negative subsets (Figure 4.5.5.). The data shows that the majority of the peptides' frequencies are close to zero in all the negative subsets. Peptides 15, 9, 10 and 14 had the highest frequency in the positive subsets, indicating that they are binding specifically to the spiked mAbs.

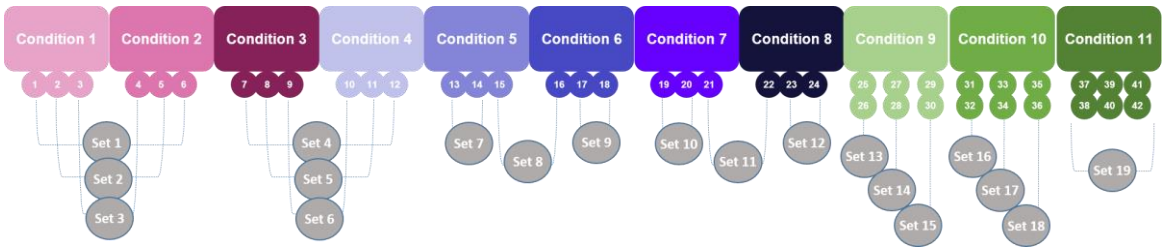


Figure 4.5.3 Overview of this optimised epitope mapping biopanning experiment and an overview of the pool of sequences per condition (pool of replicates).
Two different antibodies (SAF70 and SAF15) were spiked in Normal Mouse Sera (NMS) at 2 dilutions (conditions 1, 2,3 and 4); samples such as NMS only coated wells (conditions 5,6), or PBS only (conditions 7,8) coated wells were included as controls. Standard epitope mapping of SAF70 (condition 9) and SAF15 (condition 10) in PBS were included along with PBS coated well controls (condition 11; Table 4.5.1.). Nineteen different subsets of pooled sequencing data (as indicated) were assessed and compared with their equivalent negative control subsets.

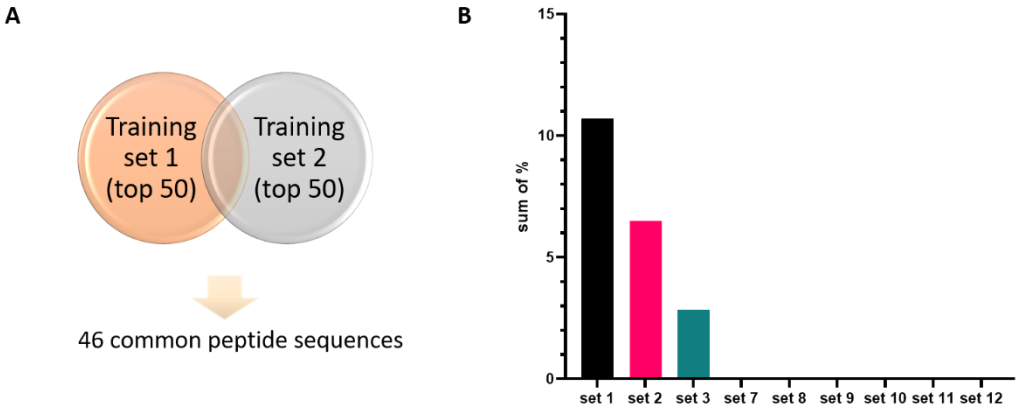


Figure 4.5.4 Comparison of the 50 most enriched sequences between two sets and the sum of frequency of the selected common enriched peptide sequences (n=46) (NMS was spiked with 1/250 SAF70 and 1/50 SAF15).
A. The fifty most enriched peptide sequences from set 1 and set 2 were compared and 46 of them were common between these training sets. **B.** The sum of the frequency of these 46 enriched peptide sequences was calculated for all the sets (training 1, 2 and testing 3-12). These sequences seem to be preferentially enriched in sets 1-3, in which NMS was spiked with SAF70 and SAF15 mAbs, indicating their specificity.

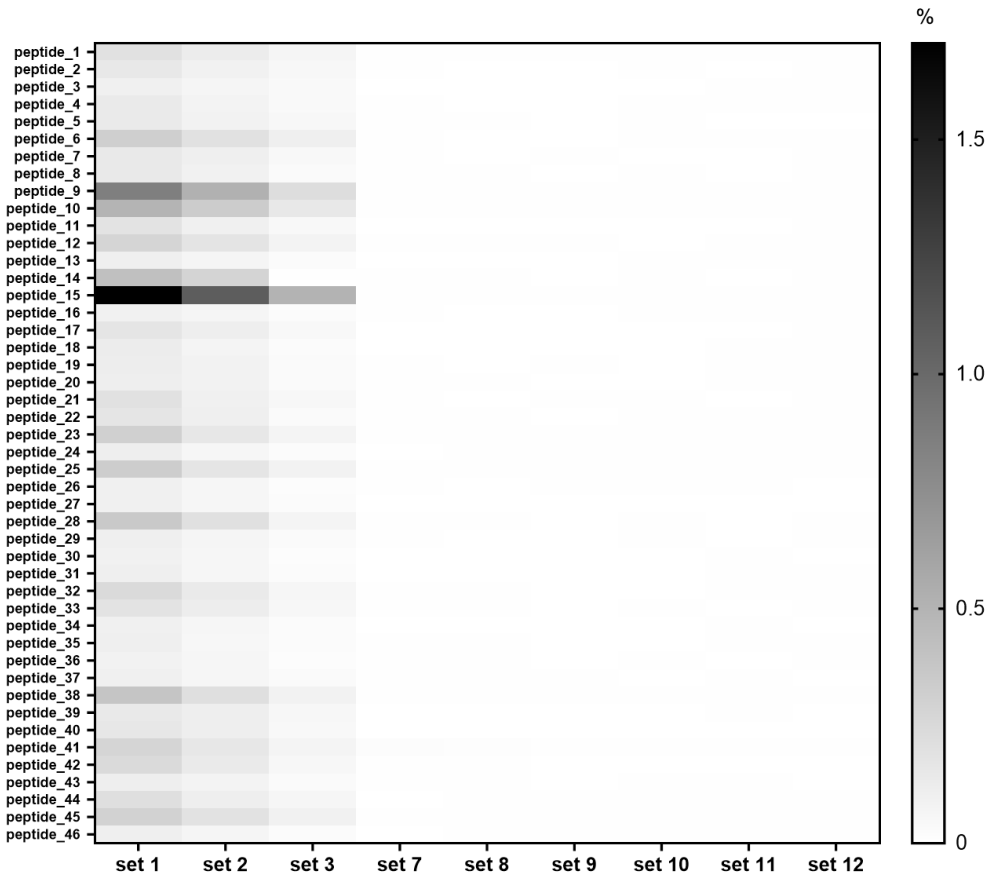


Figure 4.5.5 Assessment of individual frequencies of selected enriched peptide sequences (common in training sets 1 and 2, n=46); they were then calculated and plotted in all the sets (training and testing).
Normalised percentage of the 46 enriched sequences in all the training (1, 2) and testing tests (3-12). The majority of the sequences seem to be preferentially enriched in sets 1-3 (condition 1, 2 Table 4.5.1.). Peptide 15 is up to 1.5% of all sequences in the positive.

Similarly, condition 3 and 4 were assessed (NMS was spiked with SAF70: 1/500 and SAF15: 1/100). These conditions were comparable since only the blocking reagents that were used, were different. Sequences from subset 4 and subset 5 (training subsets) were ranked and the top 50 sequences were compared revealing 22 common peptides (Figure 4.5.6A.). The frequency of these 22 sequences was then calculated as a percentage in positive subset 6 (test

subset) as well as in negative subsets in which wells were coated with NMS only (set 7, 8 and 9) or PBS (set 10, 11 and 12). All frequencies were normalised depending on the amount of total sequences per barcode. The sum of the percentages of these sequences was calculated for all the aforementioned data subsets (training and testing) (Figure 4.5.6B). These peptides were enriched when mAbs SAF70 and SAF15 were spiked in NMS, and were in lower abundance when panning against NMS or PBS only (sets 7-12). The percentages of the 22 common peptide enriched sequences were also plotted as individual values for the positive and negative sets (Figure 4.5.7.). Heatmap shows that the majority of the peptide sequences frequencies are enriched in the positive sets, in up to 0.1% frequency. Specifically, peptides 18, 19 and 20 seem to have comparable frequency between negative and positive sets whereas peptides 3, 2, 15 and 15 seem to be preferentially enriched in the positive sets, indicating that they might be specific for the spiked mAbs.

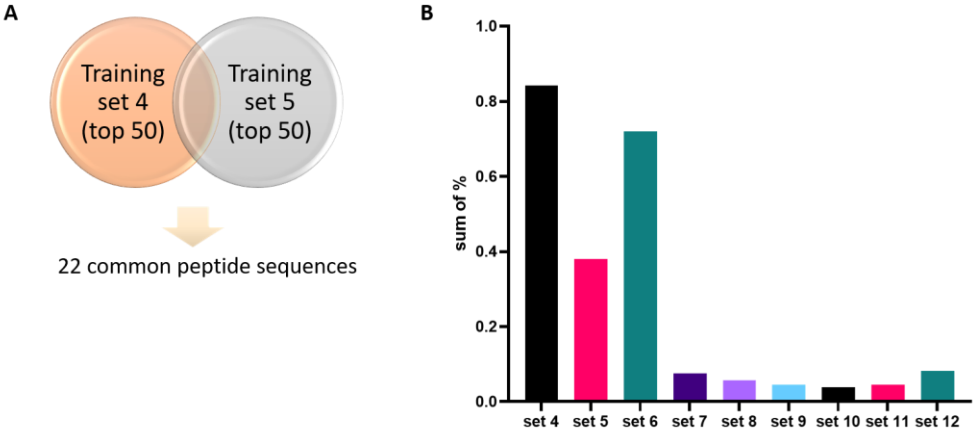


Figure 4.5.6. Comparison of the 50 most enriched sequences between two sets and the sum of frequency of the selected common enriched peptide sequences (n=22) (NMS was spiked with 1/500 SAF70 and 1/100 SAF15).

A. The fifty most enriched peptide sequences from set 1 and set 2 were pooled, and 22 of them were common between these training sets. **B.** The sum of the frequency of these 22 enriched peptide sequences was calculated for all the sets (training 4, 5 and testing 6-12). These sequences seem to be preferentially enriched in sets 4-6, in which NMS was spiked with SAF70 and SAF15 mAbs (Table 4.5.1.), indicating their specificity.

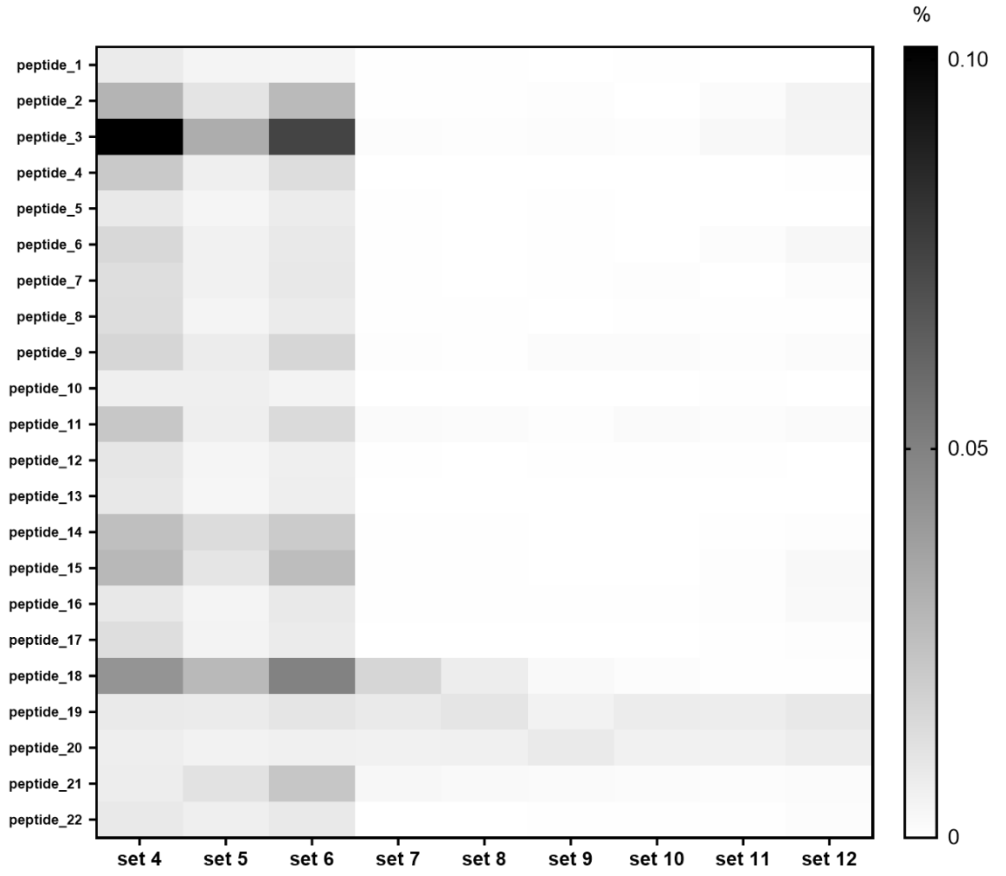


Figure 4.5.7 Assessment of individual frequencies of selected enriched peptide sequences (common in training sets 4 and 5, n=22); they were then calculated and plotted in all the sets (training and testing).
Normalised percentage of the 22 enriched sequences in all the training (4, 5) and testing tests (6-12) was plotted as individual values. The majority of the sequences seem to be preferentially enriched in sets 4-6 (condition 3, 4 Table 4.5.1.) in comparison with the negative sets (7-12) in which there were not any spiked mAbs. Especially, peptide 3 seem to have the highest frequency (0.1%) in the positive sets, followed by less enriched peptides, the percentage of which in positive sets was still higher rather than the ones in the negative sets (set 7-12). Peptide sequences of 18, 19 and 20 seem to be present in negative sets, but still in lower abundance.

Standard epitope mapping condition was also assessed (dilution of mAb SAF70 in PBS was 1/1000). Sequences from subset 13 and subset 14 were ranked, and the top 50 sequences were compared revealing 39 common peptide sequences (Figure 4.5.8A.). All frequencies were normalised depending on the amount of total sequences per barcode. The percentages of the 22 common

peptide enriched sequences were also plotted as individual values for the positive and negative sets (Figure 4.5.8B.). Heatmap shows that the majority of the peptide sequences frequencies are enriched in the positive sets, in up to 0.1% frequency. Specifically, peptides 17, 8 and 9 seem to have comparable frequency between negative and positive sets, indicating that they might be specific for mAb SAF70.

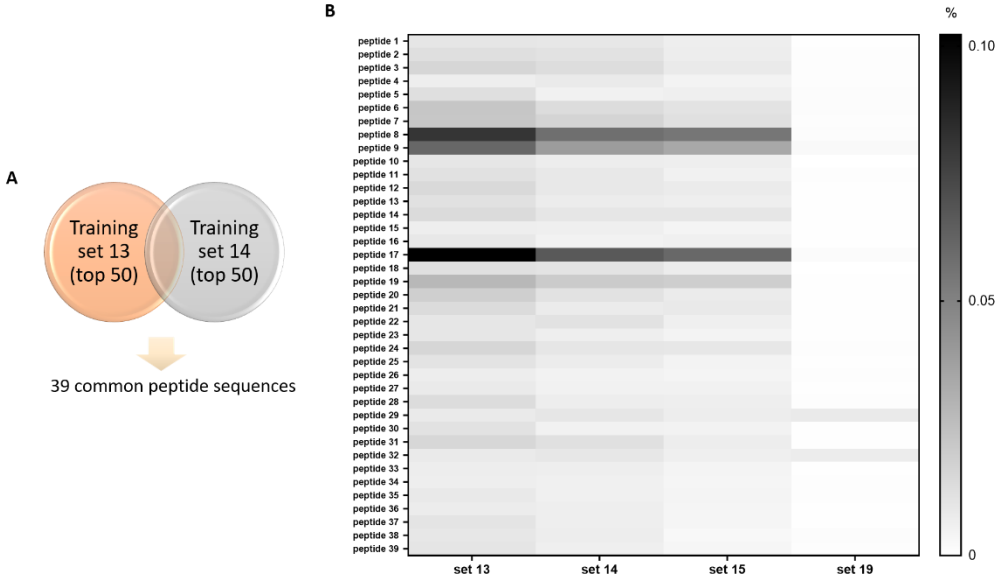


Figure 4.5.8 Comparison of the 50 most enriched sequences between training sets 13 and 14 and assessment of the frequency of 39 common enriched peptide sequences within training (set 13 and 14), positive testing (set 15) and negative set (set 19).
A. The fifty most enriched peptide sequences from training sets 13 and 14 were compared, revealing 39 common enriched peptide sequences. **B.** Normalised percentage of the 39 enriched sequences in all the training (13, 14) and testing tests (15 and 19) was plotted as individual values. The majority of the peptides seem to be preferentially present with condition when SAF70 was present; especially, peptides 17, 8 and 9 had the highest percentage (up to 0.1%) in the positive sets in comparison with the negative set 19 (PBS only).

Finally, standard epitope mapping condition was also assessed in SAF15 (dilution of mAb SAF15 in PBS was 1/1000). Sequences from subset 16 and

subset 17 were ranked, and the top 50 sequences were compared revealing 20 common peptide sequences (Figure 4.5.9A.). The frequency of these 20 sequences was then calculated as a percentage in positive subset 18 (test subset) as well as in negative subset 19 (PBS only). All frequencies were normalised depending on the amount of total sequences per barcode. The percentages of the 20 common peptide enriched sequences were also plotted as individual values for the positive and negative sets (Figure 4.5.9B). These peptides were enriched when mAb SAF15 was spiked in NMS, and were in lower abundance in PBS only (set 19). Heatmap shows that the majority of the peptide sequences frequencies are enriched in the positive sets, in up to 0.01% frequency. Specifically, peptides 7, 4 and 5 seem to have comparable frequency between negative and positive sets, indicating that they might be specific for mAb SAF15.

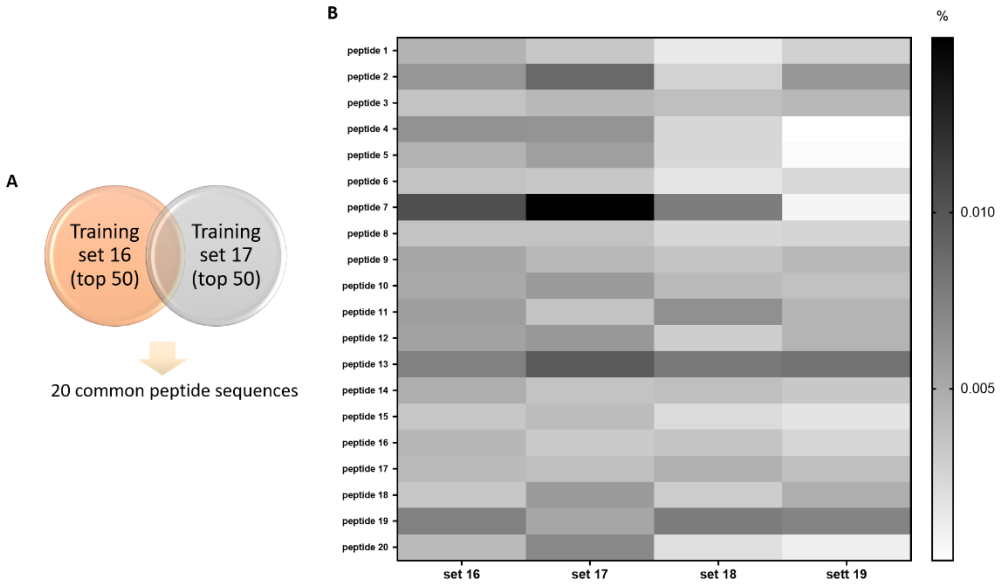


Figure 4.5.9 Frequency of 20 common enriched peptide sequences in sets 16 and 17 (standard epitope mapping of SAF15: dilution 1/1,000) and their frequency in all of the sets.
A. Fifty most enriched peptide sequences from set 16 and set 17 were pooled, and revealed that 20 of them were common between these training sets. **B.** Normalised percentage of the 20 enriched sequences in all the training (16, 17) and testing tests (18 and 19). The majority of the sequences seem to be non-specifically enriched in both positive and negative sets, and their total frequency seems to be significant low. Notably, only peptides 4, 5 and 7 had the highest percentage (up to 0.01%) in the positive sets in comparison with the negative set 19 (PBS only).

Seventeen of the identified peptides were in common in conditions 1, 2, 3 and 4 and 6 of these peptides were common between conditions 1-4 and also 9 or 10 (Figure 4.5.10.). To confirm that peptides identified by NGPD analysis are binding specifically to SAF70 or SAF15, 10 peptide sequences were selected depended on their ranking in all the different conditions (Table 4.5.2.). These were amplified with specific primers (Table 2.2.6.) and cloned into pc89_BspQI^{-} vector (Figure 4.5.11A.). Six transformants for bacteriophages displaying each of these peptides were tested for their SAF70 or SAF15 specificity with phage ELISA. Peptides 1, 2, 3 and 8 were

confirmed as binding to SAF70 and no signal was observed in SAF15 coated wells. Interestingly, peptide 8 was previously identified in the SAF70 epitope mapping optimisation experiments, validating further the efficacy of the technique. Peptides 1, 2 and 3 had the amino acids PW in positions 1 and 2, a motif that was identified previously in SAF70 binding peptides. A single transformant for peptide 10 gave a positive signal against SAF15 in comparison with SAF70 coated wells and contained a GW sequence, which was previously identified as a possible SAF15 binding motif (Figure 4.5.11B.). All clones were sequence confirmed afterwards, revealing that in some of them there were deletions, explaining that e.g., in the case of peptide 10, only one of the replicates had a positive SAF15 signal. Taken all these together, optimised epitope mapping conditions were successfully confirmed to be able to map antibodies in lower abundance in normal polyclonal sera with LOD of spiked mAbs up to 1/500.

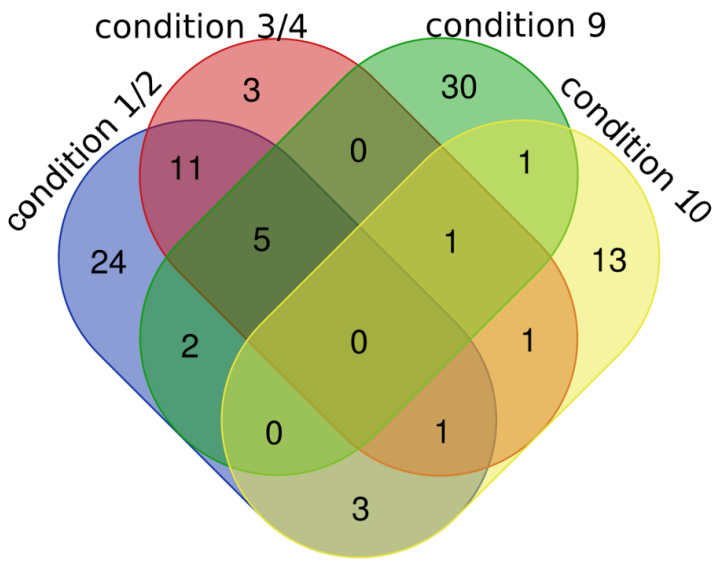


Figure 4.5.10 Venn diagram of the 95 unique enriched peptides sequences that were identified by stringent NGS analysis.
7 enriched sequences from condition 9 (SAF70 1/1000) were also common between condition 1/2 and/or 3/4 (Table 4.5.1.) indicating that these ones could be specific for SAF70 antibody. Another 5 enriched sequences from condition 10 (SAF15 1/1000) were common with condition 1/2 and/or 3/4 (Table 4.5.1.) indicating that these ones could be specific for SAF15 antibody.

Table 4.5.2 List of the selected for cloning and expression enriched peptides.

Clone ID	AA sequence	Ranking	Condition
peptide 1	PWGAARDLAQPTIELT	1st	Condition 9
peptide 2	PWHKPTLSPPPNPSNS	2nd	Condition 9
peptide 3	PWPRRGEPRLAMKPYV	3rd	Condition 9
peptide 4	MLPRSQNYEAPVHPTT	1st	Condition 1, 2, 3, 4 & 10
peptide 5	IGKRSGWHPSAISGPV	2nd	Condition 1, 2
peptide 6	KQQNPSSGWRTYQEKA	3rd	Condition 1, 2
peptide 7	LPRSQTLQEGHQPAHR	4th	Condition 1, 2
peptide 8	RQVYPNLELSTNFPIG	2nd	Condition 3, 4
peptide 9	LLHPRSQVSNGSTSHA	3rd	Condition 3, 4
peptide 10	QLSMQHGWIQMTARPQ	4th	Condition 3, 4

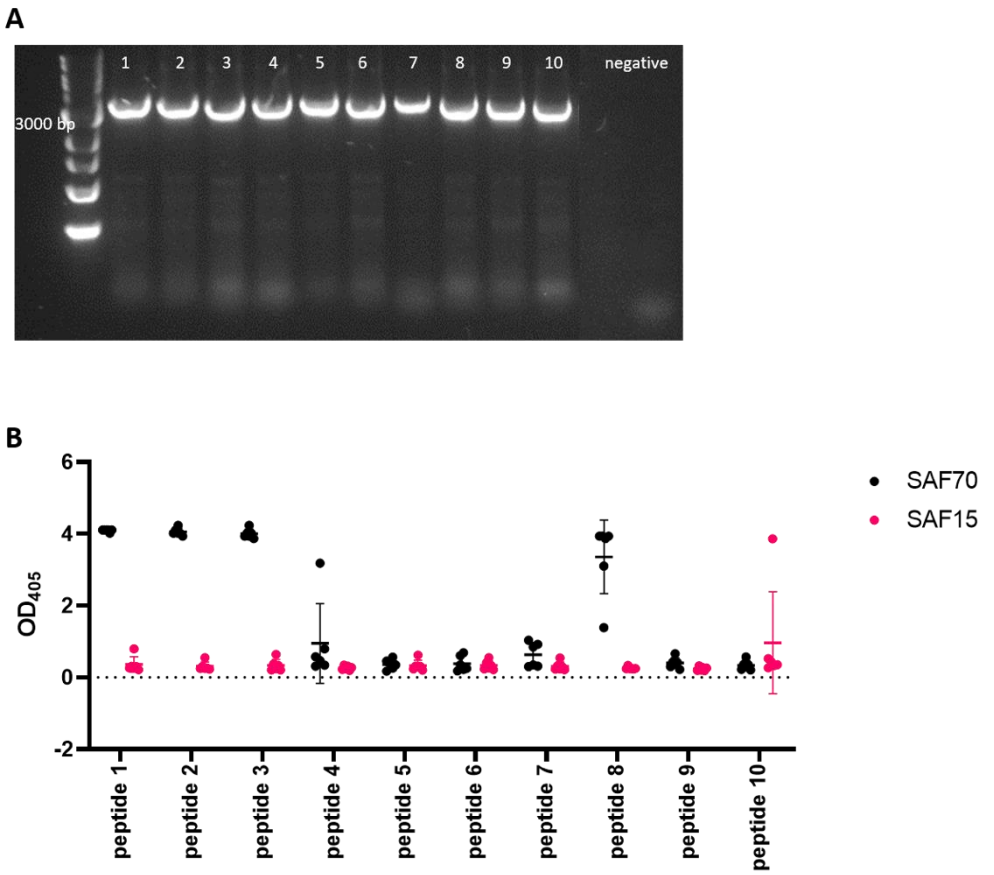


Figure 4.5.11 Assessment of rescued clones SAF70 and SAF15 specificity. **A.** Image after agarose gel electrophoresis confirming amplification after inverse PCR at the expected size (~3500 bp); sample containing no template (negative) was included as a control for possible unspecific amplification. **B.** Six transformants for each bacteriophage displaying one of the selected peptides 1-10 were tested for their SAF70 or SAF15 specificity with phage ELISA. Peptides 1, 2, 3 and 8 showed positive signal in SAF70 coated wells and no signal was observed in SAF15 coated wells. Peptide 10 showed a single transformant that was positive for SAF15 coated wells and no signal was observed in SAF70 coated wells.

4.6. Discussion

The aim of this study was to optimise the epitope mapping method utilising monoclonal antibodies in either a buffer system or spiked in sera to mimic the autoantibody titration in cancer sera. Firstly, Anchored Periplasmic Expression (APEx) was used to map a polyclonal sera. This method could potentially couple

the diversity of phage display peptide libraries with the screening power of FACS. APEx is a bacterial display method in which antibody or protein fragments are “anchored” in the inner membrane of Gram-negative bacteria by fusion with the N terminus of the pIII. Screening of the potential positive clones can be achieved by flow cytometry, allowing approximately 1 million clones to be assessed for their specific binding at once (Harvey et al., 2004). Recovery of the DNA is required after flow cytometry sorting to allow positive populations to be identified (by sequencing) and then cloned for characterisation or further rounds of selection (Ramesh et al., 2015). Here, DNA was successfully recovered from the spheroplasts, since after the removal of their outer membrane the bacteria cannot be re-cultured. Unfortunately, the sensitivity of the APEx assay, even after its thorough optimisation, was never greater than 1/100 and the mAb dilution never exceeded 10^{-3} (known epitope clone dilution in a bacterial library) which is inadequate for its applications to epitope mapping rare antibodies in polyclonal sera. Importantly, no positive cell population was observed when clones derived from two iterative rounds of biopanning. This outcome was not comparable with the successful epitope mapping of anti-VP2 polyclonal Ab (Guo et al., 2014) that suggested that APEx could be utilised for epitope mapping of polyclonal antibodies to some extent. However, when utilising a 10^9 bacterial library, seven rounds of APEx screening were needed in order to identify positive clones against Fc domain (Lee et al., 2017). This suggests that APEx could be laborious even when it is successful. Therefore, APEx was not used in further experiments.

Evidence that phage displaying random peptides could be utilised for epitope mapping, even when the epitopes are not linear has been described more than 20 years ago (Luzzago et al., 1993). The limit of detection of NGPD was yet to be determined, therefore normal mouse sera was spiked with monoclonal antibodies (the epitopes of which were known), in order to mimic the physiological concentration of AAbs in mouse or human sera. Four iterative rounds of biopanning were conducted using pc89_BspQI^{-} library, and deep sequencing of the R4 output phage population revealed no motifs resembling the epitope of the spiked mAbs, nor any mAb binding specificity of the five peptides that were enriched in the positive sets. NGS analysis was not conducted at earlier rounds of panning but the application of colony lift screening after a single round of panning revealed that an enrichment for mAb binding peptides was observed in earlier biopanning rounds and a single peptide (out of ~2000 screened) was confirmed as a binder to one of the mAbs. The colony lifting technique has been reported to be useful for screening clones in the first biopanning rounds in order to reduce potential bias introduced by growth advantage (Rodi and Makowski, 1999). Whilst colony lifts are a very efficient way to screen for epitope/mimotopes, the method relies on probing the membrane with a pure antibody and the method is not suitable to epitope map polyclonal sera and it could be very technically challenging to recover the positive clones. The failure of the NGS analysis to reveal binders could be due to the very high dilution of the mAbs (up to 10^{-6}), as well as the multiple rounds of biopanning. If only a small fraction of the clones is enriched against the specific targets, further iterative biopanning

rounds may result in their loss, as some could have a disadvantage at the propagation stage and be outcompeted by 'parasitic phage'. Phage display is always limited to the fact that during the amplification steps (growth in bacteria), some clones will have growth and amplification advantages that are unrelated to their binding activity, and these will be carried over to the next round of biopanning (Halperin et al., 2011). Library diversity might be reduced even after single bacteria propagation (Derda et al., 2011). These issues could be addressed by reducing the number of selection rounds, using more specific elution methods, or the incorporation of NGS screening (Ryvkin et al., 2018). The latter can be used to exclude any enriched sequences that can be observed in both positive and negative samples. Coupling phage display with NGS offers great advantages due to the fact, that rare enriched sequences can be identified, eliminating the laborious step of colony picking. A useful piece of information that can be extracted by deep sequencing of phage output when epitope mapping, is consensus motifs that could predict where the peptide-Ab interaction is. Peptides with similar sequences are very likely to share the same epitope. By ranking the enriched peptides by abundance or sequence homology (same patterns), it has been shown that it is enough to find high-affinity peptides, even after only one round of biopanning (Rentero Rebollo et al., 2014). NGS offers an insight to the enrichment of non-specific to target binders (such as binders to the blocking reagent, plastic, non-paratopic regions, protein A) but additional subtraction steps could also facilitate the stringent selection of target-specific binders (Christiansen et al., 2015; Huang et al., 2012; Matochko et al., 2014). When considering the limitations of phage

display, there is a possibility of lower representation for rare codons that are not widely present in E*. coli*. Clones that are toxic for the cells, even if their affinity is high, could fail to propagate efficiently as the host cells are under stress producing the ligand fusion protein (Rodi and Makowski, 1999). On the other hand, clones that propagate easier and have a less toxic effect, will be over-represented in a phage library. This bias could be eliminated if phage display is complemented with NGS in which even rare clones could be sequenced and analysed (Spiliotopoulos et al., 2015).

Therefore, extensive optimisation of the epitope mapping methodology of monoclonal antibodies was carried out to refine the method. This used monoclonal antibodies diluted in buffer (Figure 4.6.1A). The key result from this optimisation was the conditions that allowed the identification of the highest percentage of positive clones after only one round of panning. Different elution methods were first addressed, as a typical elution is achieved by pH shift, addition of mild acid or alkali or with the addition of trypsin or detergent.; However, the stronger the antibody-epitope interaction is, the harder the best binders are to elute (Solemani Zadeh et al., 2019; Huang et al., 2012). Another condition that was assessed was the immobilisation of the target mAbs. Immobilisation of targets is commonly on a solid surface either directly on coated plates e.g. with streptavidin or using precoated agarose beads (Szardenings, 2003; Huang et al., 2012), thus our approach was the binding of Abs into protein G either on beads or plates. The optimised combination of conditions was consistently the most effective for the different

antibodies tested (Figure 4.6.1B.). Specifically, these conditions are a library diversity of $5x10^9$, antibody immobilisation on protein G coated plates, plate washing 5 times with 2% PBST + 1 M NaCl followed by 5 times with PBS, and elution of specific bacteriophages with DTT.

Traditionally, epitope mapping is achieved after multiple rounds of biopanning (Ulrich Reineke* and Mike Schutkowski, 2009; Krištof Bozovicar and Tomaž Bratkoviˇ, 2020), thus our approach is superior since mimotopes of all three antibodies were able to be detected after only one round (and up to 60% positively identified clones). These results reflect those of Halperin *et al*. that epitope mapping of five different mAbs by using a random peptide microarray, revealed 37-75% positively identified peptides sequences for each mAb (Halperin et al., 2011). With this model system, motifs resembling the known epitopes of the mAbs were identified from Sanger Sequencing of selected positive clones. In the case of SAF70, an unknown mimotope was also described that contained PW amino acids in the first 2 positions of the 16mer peptide sequences. The positive clones were shown to be specific and did not bind to other mAbs. Similar short motif occurrence have been described when output phage population was deep sequenced after biopanning against a toxin; the identification of short motifs by MEME (e.g. QxQ) was a common approach throughout the different rounds of biopanning. In fact, this occurrence was compared with the same motif's occurrence within the naive library and a steady increase was shown (Braun et al., 2020).

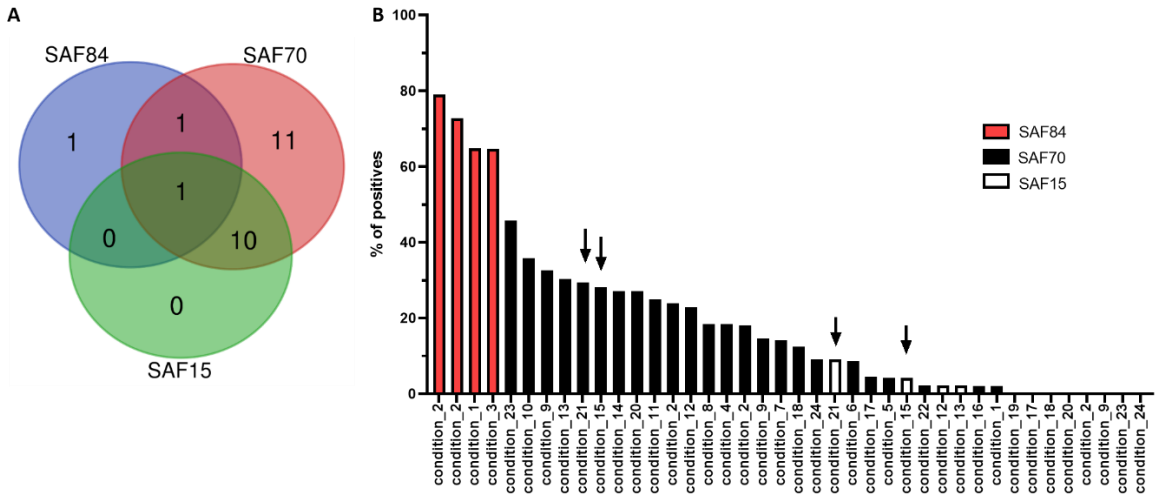


Figure 4.6.1 Summary of the epitope mapping conditions that were tested for three model mAbs and the percentage of positive clones identified by each condition.
A. Venn diagram depicting the number of the conditions that were tested when epitope mapping SAF15, SAF70 and SAF84 antibodies. **B.** Summary of all the different conditions for epitope mapping SAF84, SAF70 and SAF15 antibodies, and the percentage of the identified specific positive clones. Conditions marked with arrows were reproducible with one of the highest percentage of positive clones in both of SAF70 and SAF15a mAbs, therefore condition 21 was chosen to be the optimal condition for epitope mapping monoclonal antibodies.

The previous NGS data indicates that fewer rounds might be required to identify mimotopes. Therefore, optimised epitope mapping conditions were then applied in a single one round of biopanning where SAF70 and SAF15 were spiked into polyclonal sera (SAF70 at 1/500 and SAF15 at 1/250). Various conditions were tested in triplicate, and standard epitope mapping conditions in PBS (1/1,000 SAF70 or SAF15 in PBS) were also included. Enriched sequences were compared between pools of replicate samples (testing sets), in which the dilution of the mAbs in NMS was the same. This identifes peptides that were enriched across two experiments. The frequencies of these peptide sequences was then analysed as a sum of the frequencies of all peptides and as the

frequency of indiviual peptides. This was carried out for positive samples (containing the targeted mAbs) as well as negative samples (that contained no mAbs). In all of the cases, the sum of the frequencies of these peptides was enriched only in the positive samples, indicating their specificity. When plotting the individual values for every peptide, in the vast majority of the cases, these peptide sequences seemed to be highly enriched in the positive samples compared with the negative samples. In some cases, enriched peptides were up to 0.1% of the total seqeunces. The data indicated that one round of biopanning was sufficient for enrichment. Additionally, the conventional epitope mapping of SAF70 and SAF15 was assessed. Once more, the majority of the enriched AA sequences were preferentially enriched in the positive set of SAF70s. However, SAF15 enriched peptide sequences were less frequent, and only the minority of them were enriched specifically in the positive samples which indicates that in this case it was near to its LOD. Some of these peptide sequences were cloned into phagemid vector, and they were tested as phage displayed peptides for their specificity against the aforementioned antibodies. Indeed, four of them were successfully confirmed as SAF70 binders, and one as a SAF15 binder. As has been addressed before, when epitope mapping, the lack of exact sequence similarity with the true epitope is not unusual (Halperin et al., 2011). Similar NGS outcomes after biopanning have been previously reported, with up to 75% successfully identified mimotope sequences (Ibsen and Daugherty, 2017). Taking all these together, these developed optimised conditions allowed the epitope mapping of mAbs and this NGS analysis approach could be used to identify specific

binders to antibodies after a single round of panning even when they were diluted into sera at up to 1/500.

5. Optimisation of screening immunoassays

5.1. Introduction

Normal immunoglobulins are present in the blood circulation and their total concentration is around 1000 mg/dl; a very small fraction of those could be cancer specific autoantibodies. In some cases, complex of antibody and its corresponding tumour associated antigen might have been already formed, further decreasing the amount of free detectable autoantibodies (Kobayashi et al., 2020). Therefore, their utilisation in serum-based immunoassays (such as conventional ELISAs or microarrays) could be very challenging. There is a plethora of methods in which autoantibodies can be detected, with the most popular being ELISA, protein microarray and Western blot. Whole protein, peptides or bacteriophages displaying protein fragments have also been widely used in detection methods (Chen et al., 2014). The establishment of the limit of detection in any of these assays is crucial since the reliability of an assay is directly associated with its sensitivity and specificity. In this chapter, the limit of detection of the following immunoassays was determined, specifically:

Aims:

- Establishing the limit of detection of serum-based immunoassays (ELISA and qPCR) using normal mouse sera spiked with monoclonal antibodies.
- Investigating the limit of detection of phage microarray, in which bacteriophages displaying peptides mimicking the epitopes of

antibodies were used, instead of individually spotted synthetic peptides.

- Exploring the limit of detection of the novel approach of the immunosignature using soluble phage arrays, a computational method in which next generation sequencing data of the output bacteriophage population after biopanning are utilised to identify and measure diagnostic immunosignatures (Figure 5.1.1.).

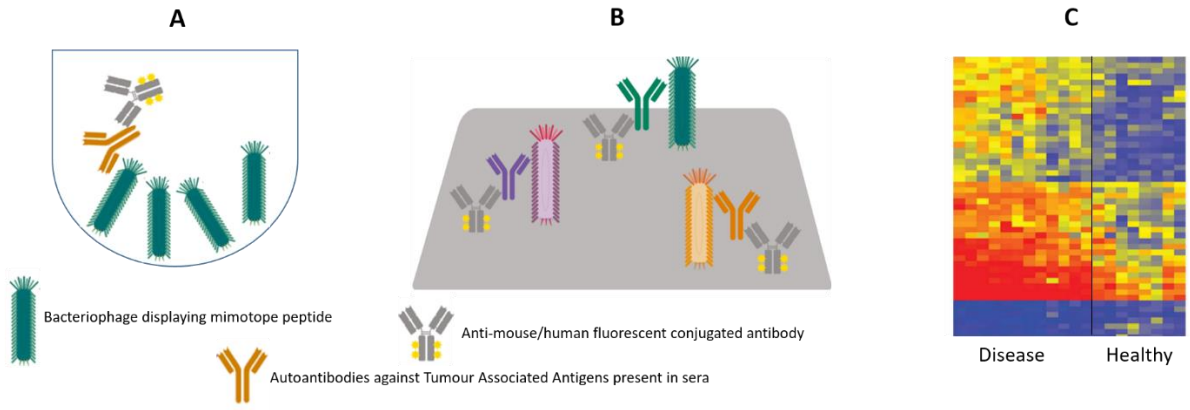


Figure 5.1.1 Three different serum-based immunoassays used for the epitope mapping of the autoimmune response against Tumour Associated Antigens (TAAs).
A. Phage ELISA: serum autoantibodies are captured by bacteriophages displaying peptides that mimic the epitopes of TAAs coated in ELISA wells; the complex is detected by an anti-phage antibody and a specific fluorescent conjugated secondary antibody. **B.** Phage Microarray: serum autoantibodies are captured by bacteriophages displaying peptides that mimic the epitopes of TAAs coated onto nitrocellulose membrane (microarray format, high throughput); the complex is detected by an anti-mouse antibody and a specific fluorescent conjugated antibody. **C.** Cancer Immunosignature: Next Generation Phage Display is applied to phage-peptides being captured by immobilised IgG from cancer and healthy sera. Enriched peptide sequences are identified and ranked by bioinformatics tools, and cancer specific immunosignature patterns are identified.

5.2. Limit of detection of conventional ELISA assays

The golden standard format for diagnostic serological assays is ELISA (Enzyme Linked ImmunoSorbent Assay). Its indirect version (where antigen is immobilised on a solid surface and a positive signal is detected by the addition of sera followed by a conjugated secondary antibody) is one of the most widely used in research and diagnostics. Here, its limit of detection was investigated using normal mouse sera (NMS), spiked with monoclonal antibodies (SAF84, SAF70 and SAF15) in different concentrations. The epitopes of these mAbs had been previously identified and they were used either as biotinylated synthetic peptides, amidated (free C terminus) peptides or as bacteriophages displaying the corresponding epitopes. Notably, biotin was present at the N terminus of the peptides, followed by the amino acid region AEGEF (N-terminal residues in our phage display system), the corresponding 9mer mAb epitope (SAF84: YYRPVDQYN, SAF70: NRYPNQVYY and SAF15: YEDRYYREN), followed by the amino acid region DPAKA (again, residues present in our phage display system). Peptides displayed by bacteriophages were fused with the pVIII coat protein. Twenty different conditions were assessed (Table 5.2.1.) in which different dilutions of normal mouse sera was spiked with the aforementioned mAbs. Streptavidin plates were coated with biotinylated peptides, whereas conventional Nunc MaxiSorp plates were coated with either amidated or bacteriophage-peptides representing the mAbs epitopes. Normal mouse sera that did not contain any mAbs (negative control) was used at a corresponding dilution in order to establish a background cut off (positive signal was

determined as 2-fold difference compared to the negative control). Notably, it was difficult to see any dose dependent effect on phage assay, due to pVIII display avidity effect (>2,000 copies per virion). In the case of SAF84 mAb, the most sensitive assay for its detection was the phage-peptide assay since a positive signal was observed in almost all the spiked dilutions and when only 0.1 µl of sera was used (Figure 5.2.1.C). On the contrary, the amidated-peptide assay was the least sensitive, with a positive signal only observed at the highest concentration of spiked mAb and sera used (Figure 5.2.1.A). Notably, the limit of detection of the biotinylated assay was determined to be when 10^{-1} SAF84 was spiked and 0.1 µl sera was used (condition 11, Figure 5.2.1.C). No positive signal was observed when 0.01 µl of sera was used cause the background signal of the phage assay was always higher than the other ELISA formats, so conditions 16 and 18 that were observed above the 2-fold cut off were considered false positive (Figure 5.2.1.D).

Table 5.2.1 Conditions that were tested in order to assess the ELISA limit of detection for biotinylated, amidated or phage-peptides (peptides displayed by bacteriophages as pVIII fusions).

Condition	mAb dilution[1]	NMS dilution[2]
1	10^{-1}	10^{-1}
2	10^{-2}	10^{-1}
3	10^{-3}	10^{-1}
4	10^{-4}	10^{-1}
5	-	10^{-1}
6	10^{-1}	10^{-2}
7	10^{-2}	10^{-2}
8	10^{-3}	10^{-2}
9	10^{-4}	10^{-2}
10	-	10^{-2}
11	10^{-1}	10^{-3}
12	10^{-2}	10^{-3}
13	10^{-3}	10^{-3}
14	10^{-4}	10^{-3}
15	-	10^{-3}
16	10^{-1}	10^{-4}
17	10^{-2}	10^{-4}
18	10^{-3}	10^{-4}
19	10^{-4}	10^{-4}
20	-	10^{-4}

[1]Normal mouse sera (NMS) was spiked with various amounts of mAb (SAF84, SAF70 or SAF15).

[2]NMS was then diluted within the assay. Non spiked NMS (-) was also used in the same dilutions in order to determine a cut off value.

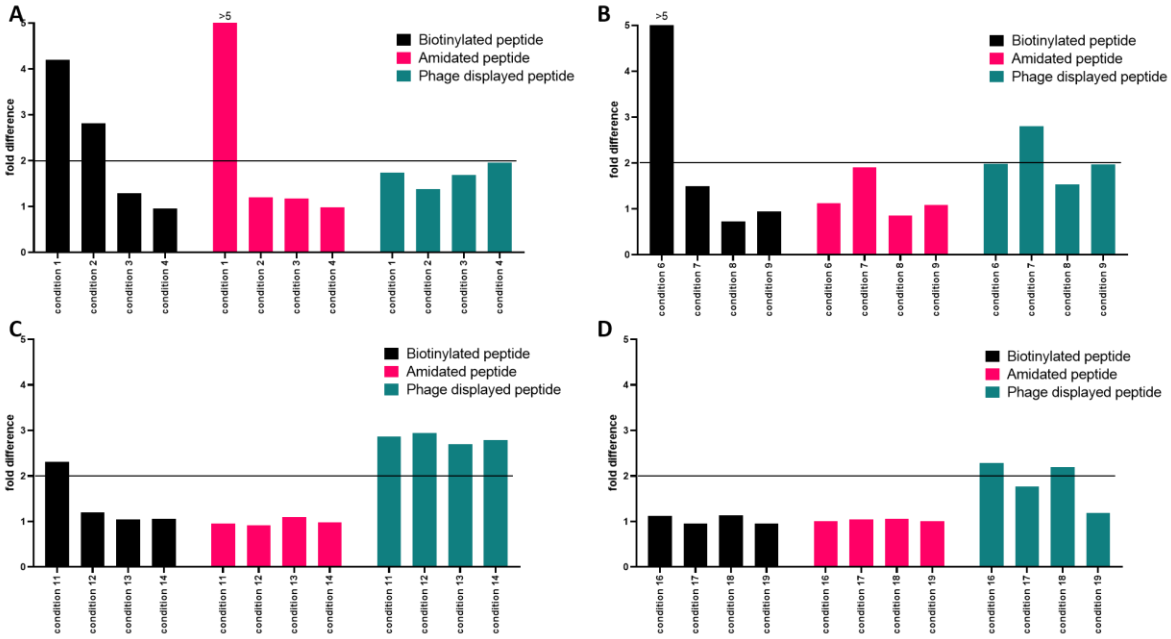


Figure 5.2.1 Comparison of different ELISA methods for detecting spiked sera with either biotinylated, amidated or phage-peptides; mAb SAF84 and its epitope are shown. Cut off value was determined as 2-fold difference from the negative control (same dilution of non-spiked NMS was used, black line).
A. Monoclonal SAF84 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 1-4, respectively) and 10 µl of NMS was used in the ELISA assay (final volume 100 µl). The use of biotinylated peptides seemed to increase the sensitivity of the assay since positive signal was detected in 10^{-2} mAb dilution. Fold difference when amidated peptide was used was >5 (actual value = 11.4). **B.** Monoclonal SAF84 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 6-9, respectively) and 1 µl of NMS was used in the ELISA assay (final volume 100 µl). Biotinylated assay seems to be the most sensitive (detectable positive signal was observed in 10^{-1} mAb dilution). Its fold difference was >5 (actual value = 8.4). **C.** Monoclonal SAF84 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 11-14, respectively) and 0.1 µl of NMS was used in the ELISA assay (final volume 100 µl). Phage-peptide seemed to be the most sensitive (detectable positive signal was observed in all the mAb dilutions). **D.** Monoclonal SAF84 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 16-19, respectively) and 0.01 µl of NMS was used in the ELISA assay (final volume 100 µl). No signal above the cut-off was observed. Notably, background signal of the phage ELISA was higher than any other ELISA formats, therefore conditions 16-19 were not considered false positive for the latter condition.

Similarly, the limit of detection of ELISA was further investigated using the aforementioned conditions with the mAb SAF70. The most sensitive assay

again seemed to be the phage assay since positive signal was observed in almost all the spiked dilutions and when only 0.1 µl of sera was used (Figure 5.2.2.C). On the contrary, amidated assay was the least sensitive one because positive signal was only observed at the highest amount of spiked mAb and sera (fold difference was 6.6 but showed as >5 for presentation purposes, Figure 5.2.2.A). The limit of detection of the biotinylated assay was determined when 10^{-1} SAF70 was spiked and 0.1 µl sera was used (condition 11, Figure 5.2.2.C), similar to what was observed when SAF84 was used. No positive signal was observed when 0.01 µl sera was used, again due to the high background observed in phage ELISAs conditions 16-18 were considered false positives (Figure 5.2.2.D).

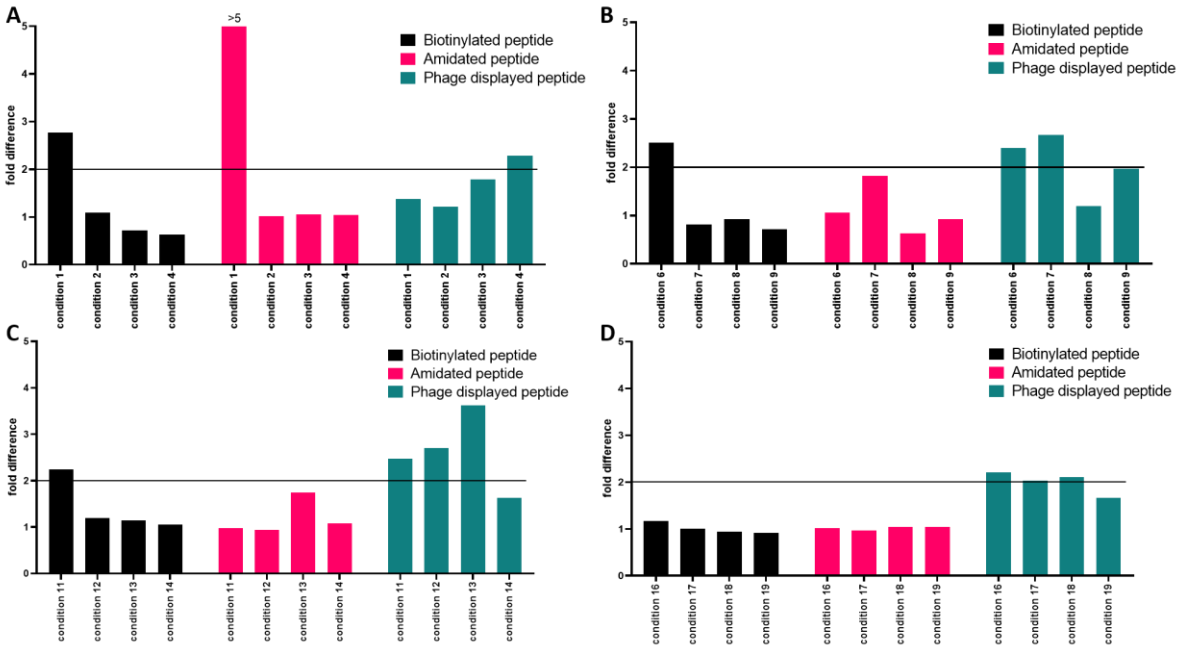


Figure 5.2.2 Comparison of different ELISA methods for detecting spiked sera with either biotinylated, amidated or phage-peptides; mAb SAF70 and its epitope are shown. Cut off was determined as 2x fold difference from the negative control (same amount of non-spiked NMS was used, black line).

A. Monoclonal SAF70 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 1-4, respectively) and 10 µl of NMS was used in the ELISA assay (final volume 100 µl). The most sensitive assay seemed to be when amidated peptide (SAF70 epitope) was used with fold difference 6.6 but showed as >5, for presentation purposes. **B.** Monoclonal SAF70 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 6-9, respectively) and 1 µl of NMS was used in the ELISA assay (final volume 100 µl). Phage assay seems to be the most sensitive (detectable positive signal was observed in 10^{-2} mAb dilution, condition 7). **C.** Monoclonal SAF70 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 11-14, respectively) and 0.1 µl of NMS was used in the ELISA assay (final volume 100 µl). Bacteriophage assay seems to be the most sensitive (detectable positive signal was observed in almost all the mAb dilutions). **D.** Monoclonal SAF70 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 16-19, respectively) and 0.01 µl of NMS was used in the ELISA assay (final volume 100 µl). No signal above the cut off was observed, and conditions 16-19 were considered as false positives due to the high background signal.

Finally, an additional mAb (SAF15) was used with the aforementioned conditions in order to confirm the limit of detection of the different kinds of ELISA. The most sensitive assay again seemed to be the phage-peptide assay

since positive signal was observed when only 0.1 or 0.01 µl of sera was used, in almost all the spiked dilutions (Figure 5.2.2.C and D); notable in the case of condition 9 (mAb dilution 10^{-4}) fold difference was 5.3 (>5, for presentation purposes). Amidated assay was the least sensitive one because positive signal was only observed at the highest amount of spiked mAb and sera (fold difference was 14.1 but showed as >5 for presentation purposes, Figure 5.2.3.A). No signal was observed in the case of biotinylated assay which was not the case for the other antigen presentations.

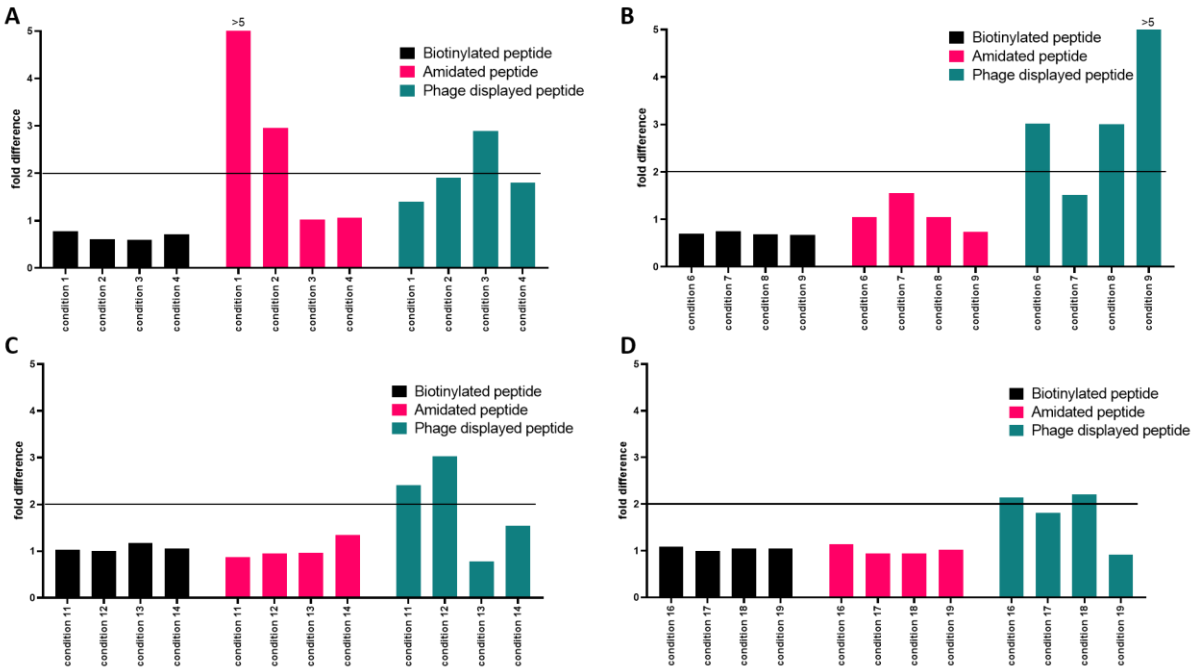


Figure 5.2.3 Comparison of different ELISA methods when biotinylated, amidated or displayed by bacteriophage peptides were used; mAb SAF15 and its epitope were used in this case. Cut off was determined as 2-fold difference from the negative control (same amount of non-spiked NMS was used, black line).

A. Monoclonal SAF15 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 1-4, respectively) and 10 µl of NMS was used in the ELISA assay (final volume 100 µl). The most sensitive assay seemed to be when amidated peptide (SAF15 epitope) was used although positive signal was observed in phage assay when the dilution of SAF15 was 10^{-2} (condition 3). **B.** Monoclonal SAF15 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 6-9, respectively) and 1 µl of NMS was used in the ELISA assay (final volume 100 µl). Phage assay seemed to be the most sensitive (detectable positive signal was observed in all mAb dilutions, fold difference of condition 9 was 14.1, but showed as >5 for presentation). **C.** Monoclonal SAF15 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 11-14, respectively) and 0.1 µl of NMS was used in the ELISA assay (final volume 100 µl). Bacteriophage assay seemed to be the most sensitive (detectable positive signal was observed in condition 12 (NMS was spiked with 10^{-2} SAF15). **D.** Monoclonal SAF15 antibody was diluted 10^{-1}-10^{-4} in NMS (condition 16-19, respectively) and 0.01 µl of NMS was used in the ELISA assay (final volume 100 µl). No signal above the cut off was observed, and conditions 16 and 18 were considered false positive due to the high background present in phage assays.

Taken together, these results suggest that bacteriophages displaying the corresponding epitope as peptide-coat pVIII protein fusion was consistently

the most sensitive assay, as signal was detected above the cut-off value when mAb dilution was from 10^{-2}-10^{-4} and when 0.1 µl of sera was used (NMS dilution 10^{-3}). Moreover, the optimal sera concentration for the phage-peptide assay seemed to be when 1 or 0.1 µl of sera was used, suggesting that higher amounts of sera containing more Abs interferes negatively with assay's sensitivity.

5.3. Limit of detection of a phage qPCR assay

Quantitative PCR (qPCR) could potentially be used as a screening assay. Enriched peptides that would have been previously identified by NGPD, would be expressed as pVIII fusions, and displayed on bacteriophages. These phage-peptide complexes could be tested for their specificity in a soluble assay, when probed with purified autoantibodies from cancer sera. Initially, normal mouse sera (NMS) was spiked with mAb SAF84, in various concentrations in order to assess the sensitivity of phage-qPCR. Briefly, bacteriophages displaying either the known SAF84 epitope (named SAF84 epitope clone) or random peptides derived from the pc89_BspQI^{-} library (negative control), were interacted with antibodies purified from spiked NMS with the same mAb in various concentrations (Table 5.3.1.). Bound bacteriophages were eluted with triethylamine and their ssDNA was extracted and amplified with a SAF84 epitope specific primer (Table 2.2.4). Firstly, control samples were assessed with conventional PCR using 2 different sets of primers (SAF84 epitope specific and ampicillin gene specific, Table 2.2.4.) in order to confirm amplification, to assess the ssDNA purification and its quality. Control samples with no DNA

template, ssDNA purified directly from pc89_BspQI$^-$ library clones, ssDNA from SAF84 epitope clones and dsDNA from pc89_BspQI$^-$ library clones were included. DNA amplicons with the SAF84 specific primer set were estimated at 200 bp, whereas amplicons with ampicillin specific primers were estimated at 280 bp. Even though a slight contamination was observed in the control using no ssDNA template-amplified with ampicillin specific primers, no amplicons were observed within the negative controls whereas the positive controls produced amplicons at the expected size (Figure 5.3.1.). Following the successful amplification with conventional PCR, ssDNA was extracted from decreasing mAb or SAF84 epitope clone concentration conditions (Table 5.3.1) and was amplified with qPCR. Unfortunately, background was very high and it was impossible to determine the limit of detection of the assay when known epitope (SAF84 clone) was present in spiked NMS (data not shown).

Table 5.3.1 Description of all conditions for the determination of limit of detection of the initial phage qPCR assay.

Sample ID	SAF84 mAb dilution[1]	SAF84 epitope clone in pc89_BspQI$^-$ library[2]
1	-	undiluted
2	10^{-1}	10^{-1}
3	10^{-2}	10^{-2}
4	10^{-3}	10^{-3}
5	10^{-4}	10^{-4}
6	10^{-5}	10^{-5}
7	-	undiluted
8	10^{-1}	10^{-1}
9	10^{-2}	10^{-2}
10	10^{-3}	10^{-3}
11	10^{-4}	10^{-4}
12	10^{-5}	10^{-5}
13	-	undiluted
14	10^{-1}	10^{-1}
15	10^{-2}	10^{-2}
16	10^{-3}	10^{-3}
17	10^{-4}	10^{-4}
18	10^{-5}	10^{-5}
19	-	undiluted

20	10^{-1}	10^{-1}
21	10^{-2}	10^{-2}
22	10^{-3}	10^{-3}
23	10^{-4}	10^{-4}
24	10^{-5}	10^{-5}
25	-	undiluted
26	10^{-1}	10^{-1}
27	10^{-2}	10^{-2}
28	10^{-3}	10^{-3}
29	10^{-4}	10^{-4}
30	10^{-5}	10^{-5}
31	-	undiluted
32	10^{-1}	10^{-1}
33	10^{-2}	10^{-2}
34	10^{-3}	10^{-3}
35	10^{-4}	10^{-4}
36	10^{-5}	10^{-5}
[1]*SAF84 mAb dilution was decreased when spiked into normal mouse sera*		
[2]*SAF84 epitope clone was spiked in pc89_BspQI⁻ library*		

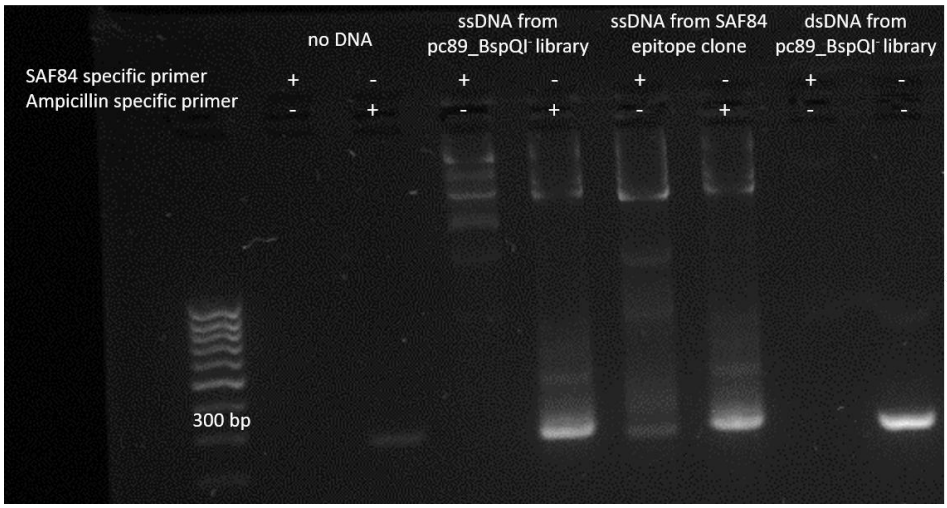


Figure 5.3.1 PCR amplification of SAF84 epitope gene fragment or an ampicillin resistant gene fragment from phagemid DNA (ssDNA) or plasmid DNA (dsDNA).

Amplicons were observed at the expected size (200 or 280 bp, respectively). Samples containing no DNA template were included as negative controls to ensure no unspecific amplification. Sample containing ssDNA derived from pc89_BspQI⁻ library was negative when SAF84 specific primer was used and an amplicon was observed when an ampicillin gene specific primer was used, as expected. Amplification of ssDNA from SAF84 epitope clone with SAF84 epitope specific primer was successfully observed at the expected size (280bp), and a 200 bp amplicon was observed when the ampicillin gene specific primer was used. Similarly, dsDNA derived from pc89_BspQI⁻ library was negative when SAF84 specific

primer was used and an amplicon was observed when an ampicillin gene specific primer was used, as expected. 1kb ladder (NEB) was used.

Further optimisation of phage qPCR was carried out in order to reduce the background thus specific bacteriophages were eluted after four instead of three washes. Additional control was included, such as probed sample with an irrelevant mAb (SAF70) instead (Table 5.3.2.). Background was lower, and unspecific amplification was observed after the 27th cycle (Figure 5.3.2.). Therefore, the limit of detection of the phage-qPCR assay was determined based on sample 28 in which the lowest detectable concentration of mAb SAF84 and its epitope clone was observed (SAF84 antibody dilution 10^{-3}, SAF84 epitope clone spiked 10^{-4} in the naïve library). This was comparable with the aforementioned limit of detection of the conventional phage-ELISA.

Table 5.3.2 Description of all conditions for the determination of limit of detection for the second phage qPCR assay.

Sample ID	mAb dilution[1]	SAF84 epitope in pc89_BspQI$^-$ library[2]
1	-	-
2	SAF70 10^{-1}	-
3	SAF84 10^{-2}	-
4	SAF84 10^{-3}	-
5	SAF84 10^{-4}	-
6	SAF84 10^{-5}	-
7	-	10^{-1}
8	SAF70 10^{-1}	10^{-1}
9	SAF84 10^{-2}	10^{-1}
10	SAF84 10^{-3}	10^{-1}
11	SAF84 10^{-4}	10^{-1}
12	SAF84 10^{-5}	10^{-1}
13	-	10^{-2}
14	SAF70 10^{-1}	10^{-2}
15	SAF84 10^{-2}	10^{-2}
16	SAF84 10^{-3}	10^{-2}
17	SAF84 10^{-4}	10^{-2}
18	SAF84 10^{-5}	10^{-2}
19	-	10^{-3}

20	SAF70 10^{-1}	10^{-3}
21	SAF84 10^{-2}	10^{-3}
22	SAF84 10^{-3}	10^{-3}
23	SAF84 10^{-4}	10^{-3}
24	SAF84 10^{-5}	10^{-3}
25	-	10^{-4}
26	SAF70 10^{-1}	10^{-4}
27	SAF84 10^{-2}	10^{-4}
28	SAF84 10^{-3}	10^{-4}
29	SAF84 10^{-4}	10^{-4}
30	SAF84 10^{-5}	10^{-4}
31	-	10^{-5}
32	SAF70 10^{-1}	10^{-5}
33	SAF84 10^{-2}	10^{-5}
34	SAF84 10^{-3}	10^{-5}
35	SAF84 10^{-4}	10^{-5}
36	SAF84 10^{-5}	10^{-5}
[1]*SAF84 mAb dilution was decreased when spiked into normal mouse sera*		
[2]*SAF84 epitope clone was spiked in pc89_BspQI⁻ library*		

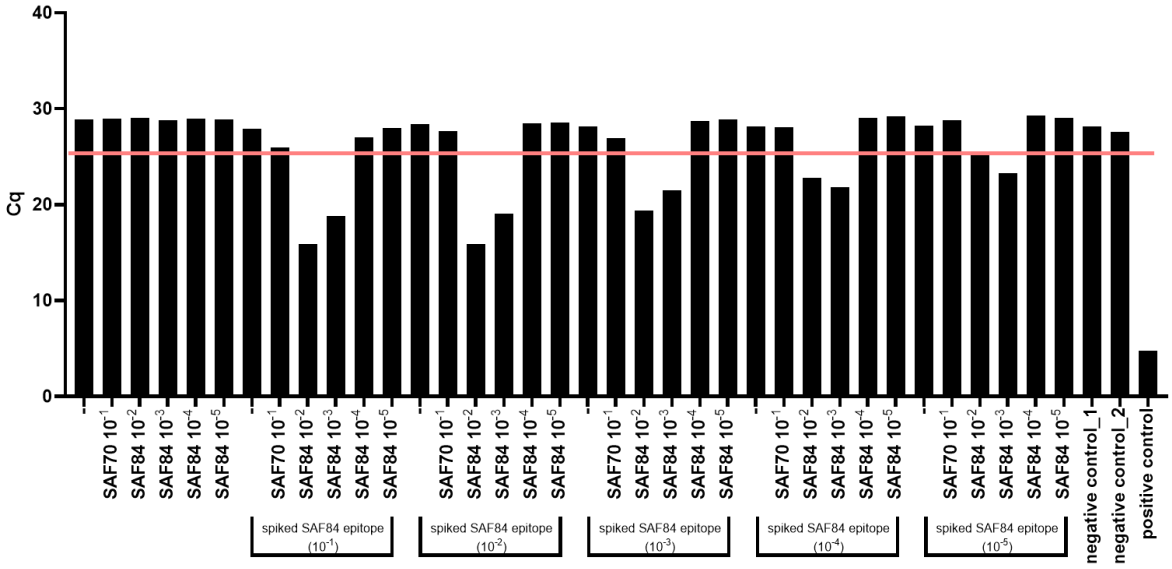


Figure 5.3.2 Assessment of the limit of detection of phage-qPCR experiment with increased washing conditions and decreased amount of spiked SAF84 epitope (10^{-1}-10^{-5}).
ssDNA was amplified with SAF84 epitope specific primers. Unspecific amplification was observed after the 26th cycle so this Cq cut off value of 25.5 was established. Limit of detection was determined when the dilution of the spiked SAF84 epitope was 10^{-4} and the spiked SAF84 mAb dilution 10^{-3}.

5.4. Limit of detection of phage dotblots and phage microarray

Protein/peptide microarrays are widely being used in diagnostics and research due to their high throughput advantages. Bacteriophages displaying peptides instead of the synthetic peptides were used in this instance, in order to introduce an inexpensive screening assay utilising phage-peptides previously identified by Next Generation Phage Display. These peptides had been previously identified by the conventional epitope mapping of known monoclonal antibodies (SAF84, SAF70 and SAF15), and were confirmed to mimic the already known epitopes of these mAbs. The aim of the following preliminary experiment was to confirm the specificity of the selected bacteriophage clones (initially involving only SAF70 specific ones) displaying the 16mer mimotope peptides and also their stability over prolonged storage at 4°C as spotted on membrane or at -20°C as glycerol aliquots. Initially, four bacteriophages displaying peptides mimicking the SAF70 epitope were used (SAF70_mimotopes_1-4, Table 5.4.1.). These clones had a wide range of relative affinities as assessed by ELISA (strong = equal or higher than the positive control signal, medium = half of the positive control signal and weak = mean+ 2xSTD of the negative control signal) (Table 5.4.1.). Each monoclonal bacteriophage sample was titrated to an estimated 10^9 bacteriophages per µl. This was followed by serial dilutions in PBS (Table 5.4.1) and samples were then spotted onto nitrocellulose membrane (Method 2.2.6). Two of them already contained a loose motif resembling the SAF70 epitope (YPN) and two of them contained PW in the first two amino acid positions, a sequence described in

Chapter 4.2.4. Positive control (monoclonal bacteriophage displaying the known SAF70 epitope, referred to as SAF70 epitope clone) and negative controls (monoclonal bacteriophage displaying the known SAF84 epitope, referred to as SAF84 epitope clone, and a PBS only sample) were included. Nitrocellulose membranes were spotted in Day 1 and they were immediately stored at 4°C. These membranes (referred to as ''stored'') were directly compared with freshly spotted membranes in which the same monoclonal bacteriophages were used after their storage at -20°C (referred to as ''new''). Comparison of these different storage conditions took place at days 7, 30, 90, 180, 270, and 360 after initial storage. All clones retained their SAF70 specificity when nitrocellulose membranes were probed with SAF70 antibody (Figure 5.4.1.). A slight reduction in the signals of the positive samples in some dilutions was observed after 12 months in newly spotted membranes (Figure 5.4.1.B); no positive signal was obtained when SAF70_mimotope_4 was spotted, confirming its low SAF70 specific ELISA signal. Negative controls (SAF84 epitope clone and PBS only) did not produce any signal, as expected all throughout this stability assay. In summary, bacteriophages seemed to be highly stable during storage on membranes at 4°C for at least 360 days.

Table 5.4.1 Description of bacteriophage clones that were used for the initial dotblot storage assessment.
Amino acid sequences and their approximate affinities (as observed by ELISAs when epitope mapping SAF70 mAb) were described.

Specificity[1]	AA sequence[2]	Affinity[3]
SAF70_mimotope_1	RQVYPNLELSTNFPIG	Strong
SAF70_mimotope_2	TTPYPNLTPSAMHSAT	Strong
SAF70_mimotope_3	PWGPPIQAGCTQSLQQ	Medium
SAF70_mimotope_4	PWSVSADEARPALPKP	Weak
SAF84 epitope clone	YYRPVDQYN	(-)
SA70 epitope clone	NRYPNQVYY	(+)
PBS	-	(-)
[1]Specificity as this was assessed when epitope mapping SAF70 mAb		
[2]Amino acid sequence of all the bacteriophage clones as this was assessed by Sanger Sequencing of the purified dsDNA		
[3]Approximation of the affinity as this was measured by conventional ELISA		

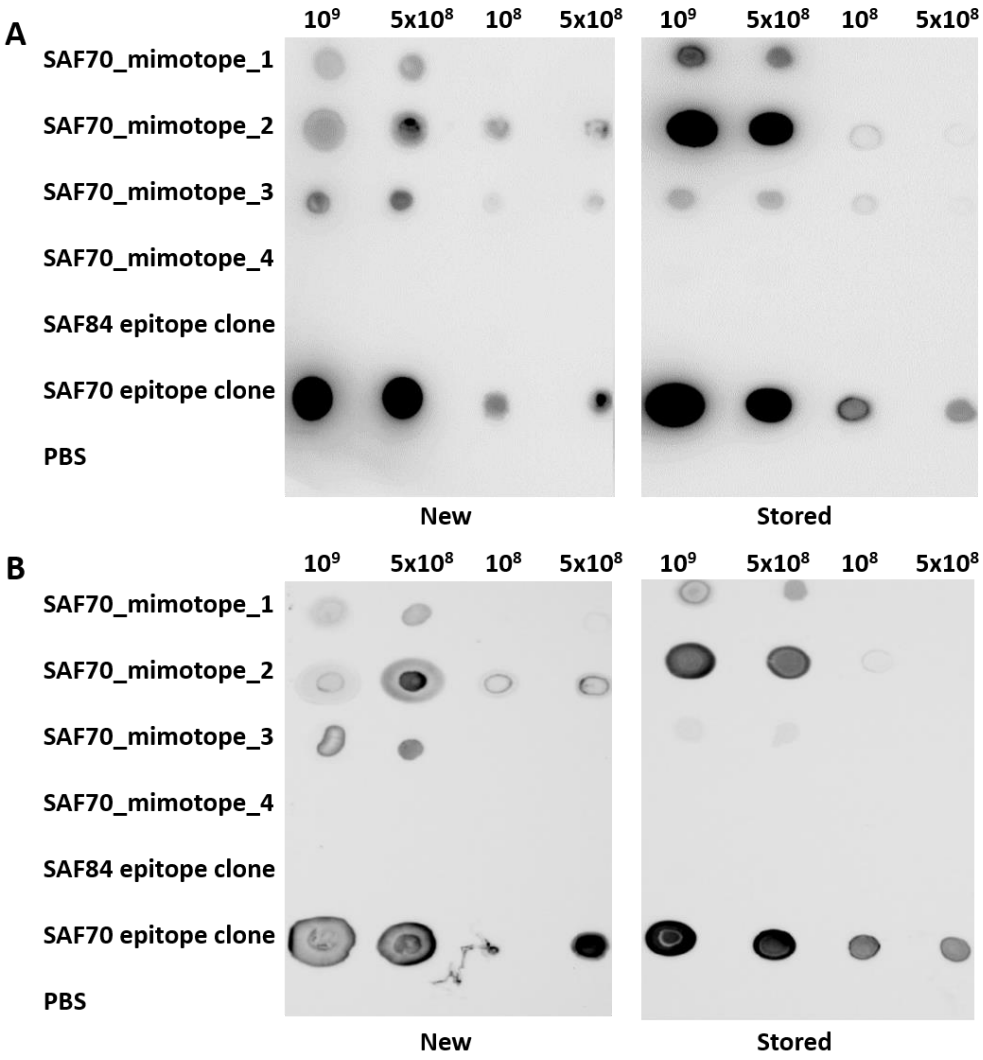


Figure 5.4.1 Stability assay over 2 different time points throughout 12 months in order to assess three storage conditions on the phage dotblots.

Bacteriophages displaying SAF70_mimotopes_1-4, SAF84 epitope clone, SAF70 epitope clone and PBS only were spotted in four different dilutions ($5x10^8$-10^9) on nitrocellulose membranes and either stored at 4°C (labelled as stored) or were freshly prepared from bacteriophages stored at -20°C (labelled as new). Comparison of new with stored nitrocellulose membrane after 1 month (A), and 12 months (B). Bacteriophage-peptides retained their binding specificity when membranes were probed with SAF70 antibody. No positive signal was detected in negative samples (SAF84 epitope clone and PBS). A slight reduction of single was observed in some dilutions; no positive signal was observed for SAF70_epitope_4, confirming its low signal as observed by previous ELISA experiments.

Twelve bacteriophages displaying mimotopes of three different mAbs (SAF15, SAF70 and SAF84) had been previously identified by conventional epitope mapping of the aforementioned antibodies and were selected for the development of a novel microarray assay, in which bacteriophages displaying peptides were used instead of synthetic peptides (Table 5.4.2.). Firstly, their specificity was confirmed by conventional phage ELISA (Figure 5.4.2.). All monoclonal bacteriophages were probed with all the aforementioned mAbs; PBS coated wells were included as negative control; all in duplicates. Clones displaying the known mAbs epitopes were included, acting as negative controls for each other. Cut off of the assay was determined as 2x above negative signal of PBS only coated wells (black line, Figure 5.4.2.). Specificity of SAF84 mimotope clones (1-3) was confirmed. Signal of the SAF70 mimotope clones was lower but above the cut off value for 4 out of the 5 clones analysed. Positive signal of SAF15 specific clones was also confirmed; no unspecific background signal was detected in PBS coated wells. Bacteriophages displaying mAb mimotopes did not cross react with other mAbs (e.g. SAF84_mimotope_1 was SAF84 specific and no positive ELISA signal was observed in SAF15, SAF70 or PBS coated wells) (Figure 5.4.2.). Following this confirmation, their specificity was further confirmed in dotblots. The same clones were spotted on nitrocellulose membranes and they were then probed with either SAF15, SAF70 or SAF84 mAb (Figure 5.4.3.A, B and C, respectively). Positive signal was confirmed for SAF15 specific clones (Figure 5.4.3.A). On the other hand, only strong SAF70 specific clones had positive signal when the

membrane was probed with mAb SAF70, and their overall signal was lower than the others (Figure 5.4.3.B). This relative binding strength when comparing the mAbs was similar to that seen in the ELISA. SAF84 specific clones 1 and 2 (Figure 5.4.3.C) were confirmed, whereas no positive signal was observed for clone 3 with relatively weak binding (Figure 5.4.3.C). When dotblot was probed with spiked normal mouse sera with SAF84 and SAF15 mAbs, unspecific background signal was observed across all samples (data not shown). This unspecific signal was not improved when membrane was pre-blocked with normal chicken sera (Figure 5.4.3.D.). This was indicative of a cross reactivity that could occur in the following microarray experiments.

Table 5.4.2 Description of the selected bacteriophage clones that were used for the development of the phage microarray assay.

Specificity[1]	AA sequence[2]	Affinity[3]
SAF84_mimotope_1	PASERPVTQYPRLVGV	Strong
SAF84_mimotope_2	SPRPSYQYQPSSTE*L	Medium
SAF84_mimotope_3	NRPETSYATRISTHHH	Weak
SAF70_mimotope_1	RQVYPNLELSTNFPIG	Strong
SAF70_mimotope_2	TTPYPNLTPSAMHSAT	Strong
SAF70_mimotope_3	PYPVMRMMLPNRGTPQ	Strong
SAF70_mimotope_4	PWGPPIQAGCTQSLQQ	Medium
SAF70_mimotope_5	PWSVSADEARPALPKP	Weak
SAF15_mimotope_1	YESTQPNRSGWNPLSA	Strong
SAF15_mimotope_2	SPYQPSRHGWYNLNVH	Strong
SAF15_mimotope_3	APLDPRLPRAQFRSSS	Strong
SAF15_mimotope_4	WLVTTKGWHPPSASSS	Medium
SAF84 epitope clone	YYRPVDQYN	Strong
SAF70 epitope clone	NRYPNQVYY	Strong
SAF15 epitope clone	QPHGGGWGQ	Strong
PBS	-	-
M13K07	-	-
[1]Specificity as this was assessed when epitope mapping SAF70 mAb		
[2]Amino acid sequence of all the bacteriophage clones as this was assessed by Sanger Sequencing of the purified dsDNA		
[3]Approximation of the affinity as this was measured by conventional ELISA		

Amino acid sequences of these clones, their specificities as well as their approximate affinities (as described by ELISAs when epitope mapping these mAbs) were described.

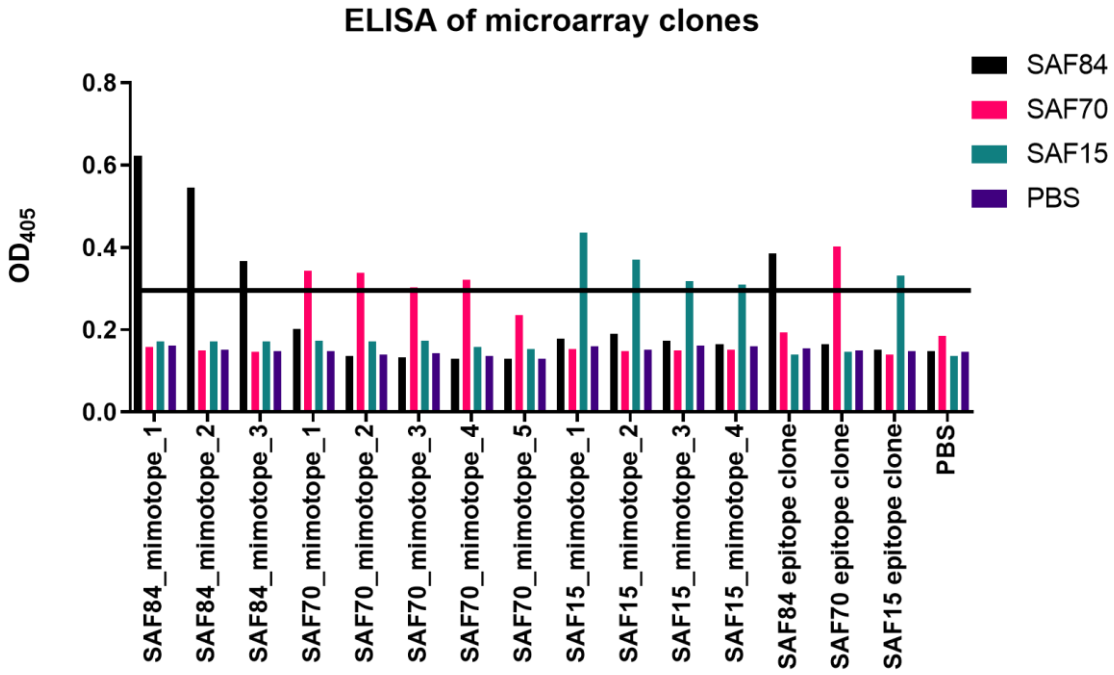


Figure 5.4.2 Confirmation of the mAb specificity of selected bacteriophage clones that were used for the development of a phage microarray assay.

All clones were tested against 3 different mAbs and against PBS coated wells (as indicated). Black line represents the cut off of the ELISA assay (2xSTD above negative signal, PBS only wells). Bacteriophages displaying the known mAbs epitope were included as positive controls. For SAF84 binding, positive signal was observed in SAF84 mimotope clones 1-3 (Table 5.4.2.) and SAF84 epitope clone, as expected. For SAF70 binding, all mimotope clones except SAF70_mimotope_5 were detected (Table 5.4.2.), the SAF70 epitope clone was also detected, as expected. For SAF15 binding, positive signal was observed in all SAF15 mimotope clones (1-4) (Table 5.4.2.) and SAF15 epitope clone when screened against mAb SAF15.

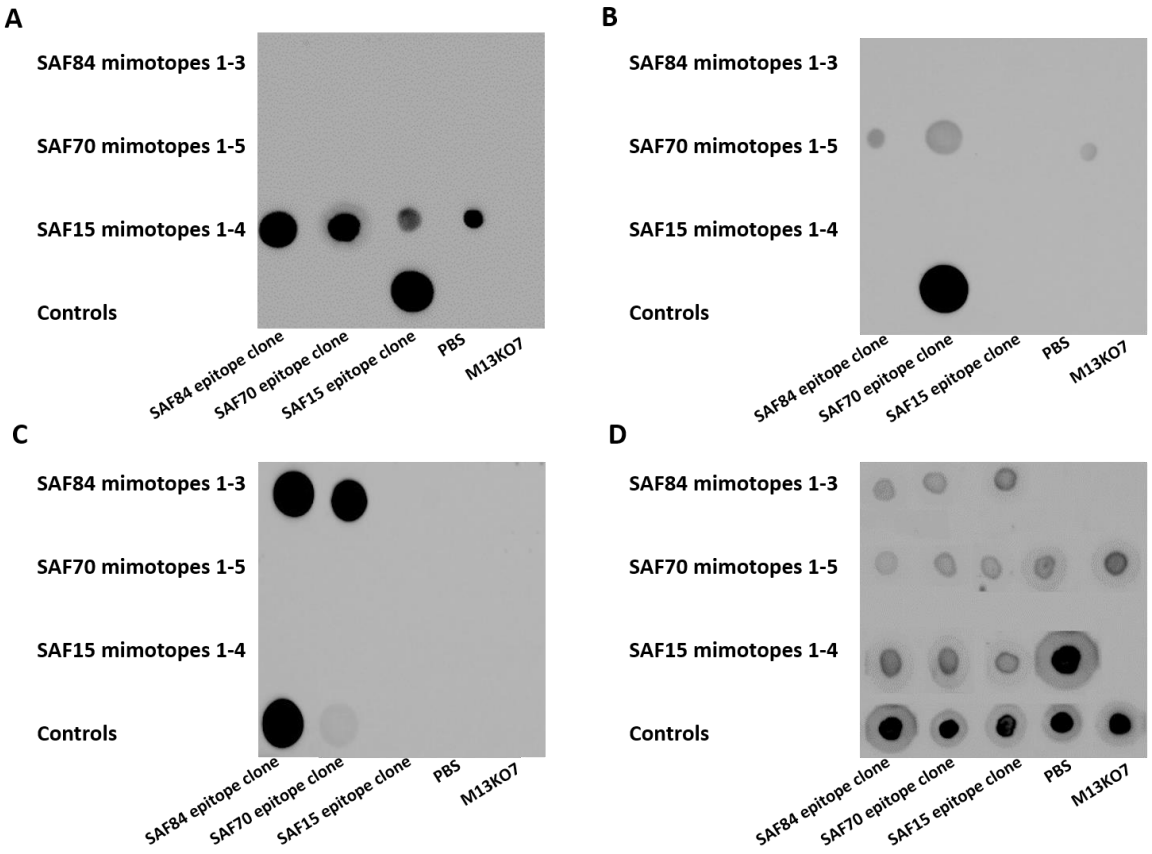


Figure 5.4.3 Detection of monoclonal bacteriophages displaying mimotopes bound by mAbs SAF84, SAF70 and SAF15 in dotblots.
Bacteriophages were spotted onto nitrocellulose membranes, and then were probed with mAbs diluted in PBS (A-C for MAbs SAF15, SAF70 and SAF84, respectively), or preblocked with normal chicken sera (NCS) and then probed with normal mouse sera spiked with all 3 MAbs (D). Helper phage M13K07 and PBS only samples were included as negative controls; monoclonal bacteriophages displaying the known epitopes for the mAbs were included as positive controls (epitope clones) and each of them acting as further negative control for each other.

All the aforementioned monoclonal bacteriophage clones were spotted onto glass slides coated with nitrocellulose membrane in a microarray format (Arrayjet®). Each glass slide had 16 patches of all the aforementioned samples in quadruplicates, therefore 16 conditions could be tested per slide. All slides were scanned at 800/400 fluorescent channel, and the background signal of the four replicates was averaged and subtracted from the mean F635 signal of

the quadruplicates of all samples. Firstly, the optimal blocking conditions were investigated since high unspecific background was observed in previously describe dotblots when they were probed with spiked NMS. 1 or 0.1 µl of NMS was added in each patch (final volume was 100 µl) and then were blocked with a range of blocking reagents (Table 5.4.3.). Collectively, the lowest background signal was observed using NAP blocking reagent diluted 50% in 1% PBST (Figure 5.4.4.). Therefore, this blocking reagent was used for all the following experiments.

Table 5.4.3 Eight different blocking reagents were used for the initial microarray experiment in order to determine the optimal one and then patches were probed with either 1 or 0.1 µl of NMS.

Condition	NMS dilution	mAbs	Blocking reagent
1	10^{-2}	-	0.1% TBST + 1% Marvel skimmed milk
2	10^{-2}	-	1% TBST + 3% Marvel skimmed milk
3	10^{-2}	-	3% BSA in TBS
4	10^{-2}	-	3% BSA, in 1% TBST
5	10^{-2}	-	Superblock ™ buffer
6	10^{-2}	-	¼ NAP in PBS
7	10^{-2}	-	½ NAP in 1% PBST
8	10^{-2}	-	1% PBST + 3% Marvel skimmed milk
9	10^{-3}	-	0.1% TBST + 1% Marvel skimmed milk
10	10^{-3}	-	1% TBST + 3% Marvel skimmed milk
11	10^{-3}	-	3% BSA in TBS
12	10^{-3}	-	3% BSA, in 1% TBST
13	10^{-3}	-	Superblock™ buffer
14	10^{-3}	-	¼ NAP in PBS
15	10^{-3}	-	½ NAP in 1% PBST
16	10^{-3}	-	1% PBST + 3% Marvel skimmed milk

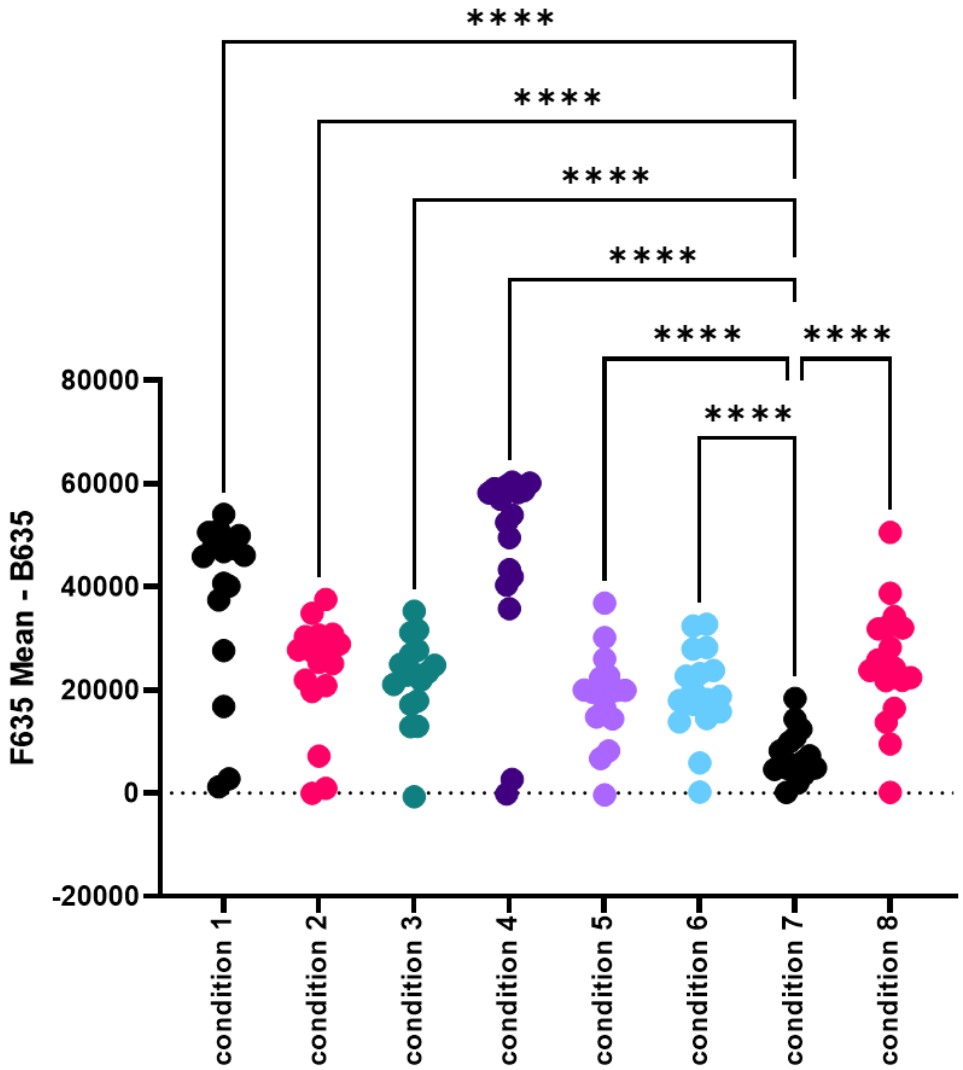


Figure 5.4.4 Optimisation of blocking conditions for phage-peptide microarrays.
Microarray patches were blocked with eight different blocking reagents and then probed with 1 µl of NMS (condition 1-8) or 0.1 µl (condition 9-16) of NMS. Condition 7 (NAP blocking reagent was added 50% (v/v) in 1% (v/v) PBST) had the lowest collective background signal, when compared with other blocking reagents.

Once the optimal blocking conditions were established, a confirmation of the specificity of the spotted bacteriophage clones was necessary. Therefore, patches were probed with different dilutions of monoclonal Abs individually (condition 3-8) or in combinations (9-16); first two patches (condition 1 and 2)

were probed with only the secondary conjugated antibody (Table 5.4.4). Average signal of the four replicate samples (after subtraction of the background signal) was expressed as a fold difference when compared with the mean signal of negative samples (containing either PBS or M13KO7 helper phage, 8 replicates in total). The cut off of the assay was determined as 2-fold difference above the negative samples (black line). The majority of the bacteriophages retained their specificity when patches were probed with mAbs or with combination of them. For instance, positive signal of 3 out of 5 SAF70 mimotope clones (1, 2, 4) was detected in condition 5 (dilution of SAF70 mAb was 10^{-4}, Table 5.4.4.); in fact SAF70 epitope clone's fold difference was almost 170, but showed as 22-fold difference for presentation purposes (Figure 5.4.5.). Furthermore, SAF84 and SAF15 mimotope clones were positive only when patches were probed with both of SAF84 and SAF15 mAbs (except SAF84 mimotope_3) (Figure 5.4.6.).

Table 5.4.4 Microarray patches were probed with SAF84, SAF70 and SAF15 mAbs in different concentrations and/or in combination; dilutions were all in PBS.

Condition	NMS dilution	mAbs	mAbs dilution
1	-	-	-
2	-	-	-
3	-	SAF84	10^{-4}
4	-	SAF84	10^{-5}
5	-	SAF70	10^{-4}
6	-	SAF70	10^{-5}
7	-	SAF15	10^{-4}
8	-	SAF15	10^{-5}
9	-	SAF84 & SAF70	10^{-4}
10	-	SAF84 & SAF70	10^{-5}
11	-	SAF70 & SAF15	10^{-4}
12	-	SAF70 & SAF15	10^{-5}
13	-	SAF84 & SAF15	10^{-4}
14	-	SAF84 & SAF15	5×10^{-4}
15	-	SAF84 & SAF15	10^{-5}
16	-	SAF84 & SAF15	5×10^{-5}

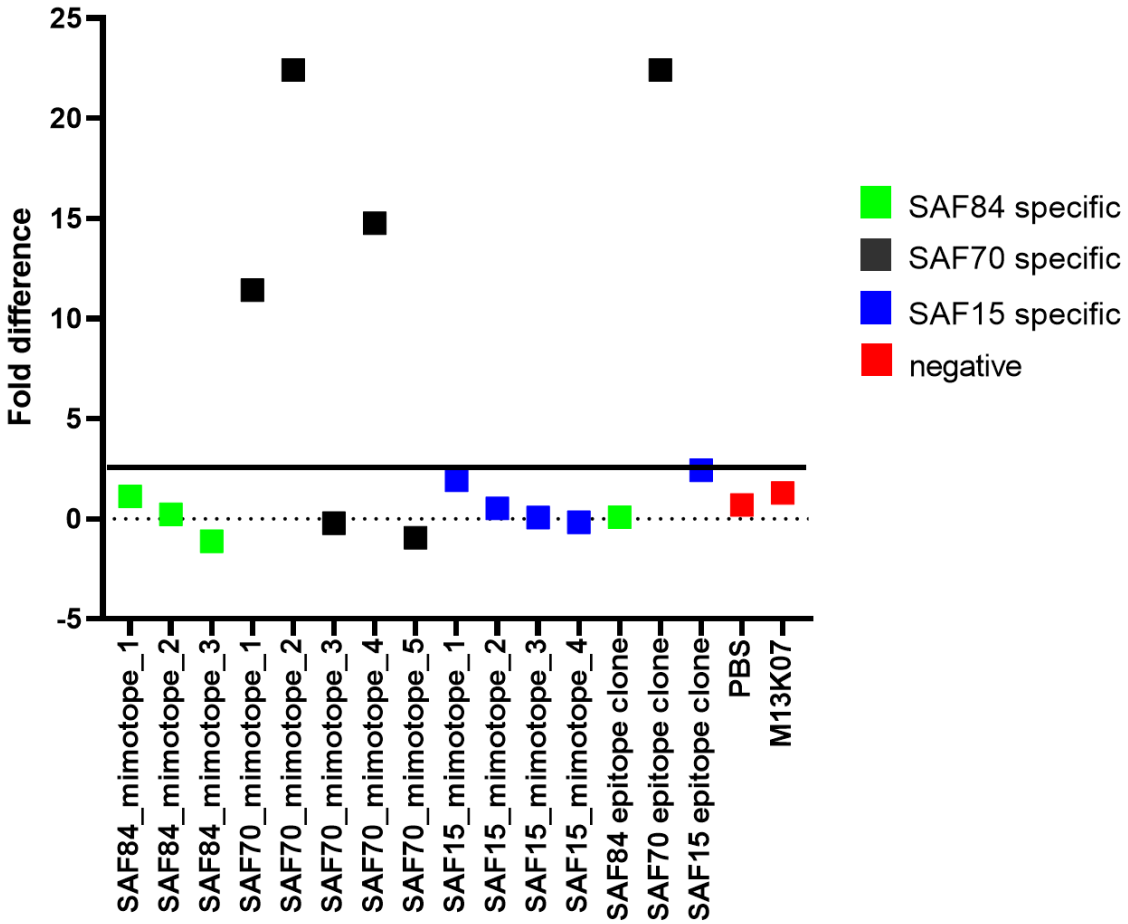


Figure 5.4.5 Testing of a phage-peptide microarray when probed with mAb SAF70.

A microarray patch was probed with SAF70 (condition 5, Table 5.4.4) and three out of five SAF70 mimotope clones retained their SAF70 specificity in this microarray format (clones 1, 2, and 4) in presence of 10^{-4} diluted SAF70 in PBS; fold difference for the positive clone (bacteriophage displaying the known SAF70 epitope, SAF70 epitope clone) was almost 170, but shown as 22 for presentation purposes. Cut off of the assay was determined as 2x the average of negative signal (PBS and M13K07, black line). No unspecific signal from SAF84 or SAF15 specific clones was detected.

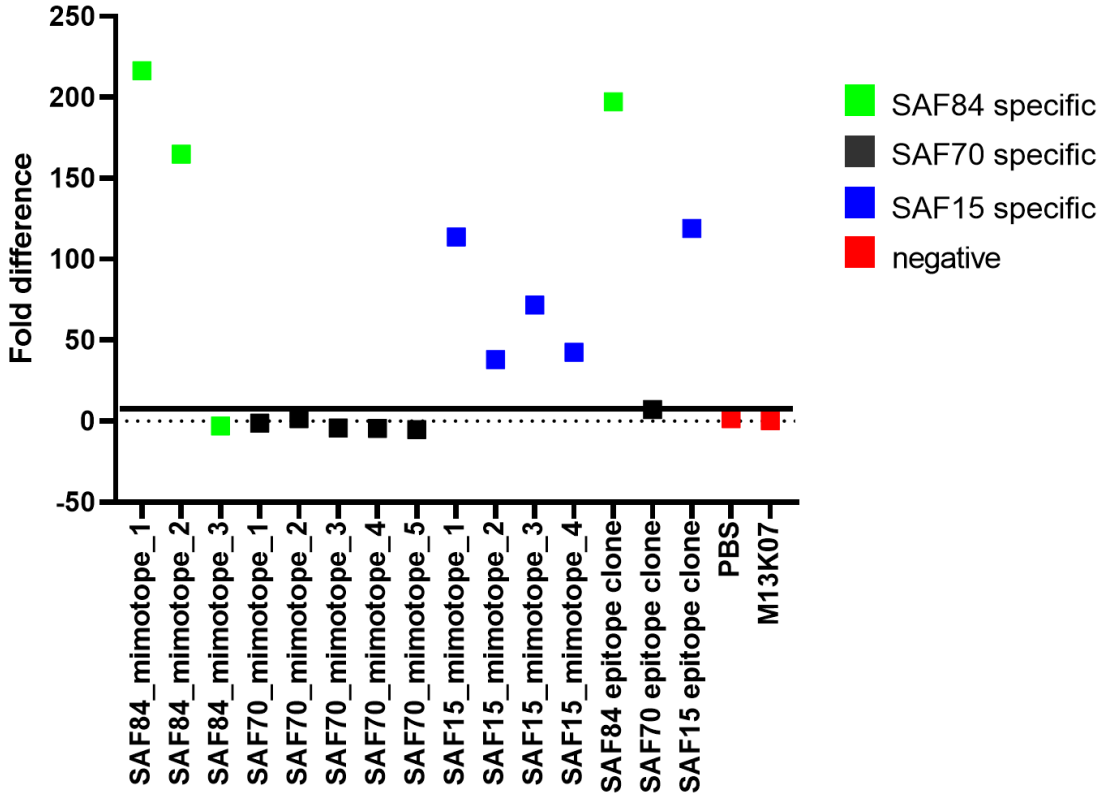


Figure 5.4.6 Testing of a phage-peptide microarray when probed with mAb SAF84 and SAF15.
A microarray patch was probed with SAF84 and SAF15 (condition 13, Table 5.4.4) and almost all SAF84 and SAF15 mimotope clones (except SAF84 mimotope_3) retained their specificity. Cut off of the assay was determined as 2x the average of negative signal (PBS and M13K07, black line). No unspecific signal from SAF70 mimotope clones was detected, except SAF70 epitope clone.

Following the encouraging results of the retained mAb specificity in dilutions up to 10^{-5} in PBS, the limit of detection of the microarray assay was investigated in order to explore its sensitivity and specificity in sera. The aforementioned mAbs were spiked in various concentrations in NMS in combinations (most frequently SAF84 and SAF15). A range of conditions were tested (18 slides x 16 patches = 288 different conditions). The variations tested included: antibody dilution spiked in NMS, amount of spiked sera,

duration/temperature/agitation of primary or secondary incubations, depletion of NMS with helper phage, number of washing steps, different washing reagents or use of purified antibody samples (purified on protein G beads) from spiked NMS. Unfortunately, unspecific background was very high in almost all the combinations of conditions and positive clones could not be identified above the cut off. Additionally, reproducibility between slides or even from patches within the same slides could not be achieved either. For example, various conditions were tested in an attempt to reduce the unspecific background (Table 5.4.5.) in which either spiked NMS was depleted with M13KO7 helper phage or mAbs were purified from spiked NMS. Specifically, SAF84 and SAF15 were spiked in 10^{-3} dilution in NMS, followed by its depletion with helper phage and then 1 or 0.1 µl were used per patch (condition 3 and 4, respectively, Table 5.4.5.). Fold difference was calculated as the ratio of the different signal between condition 3 and 1 or condition 4 and 2 (same amount of non-spiked NMS was used as the control in each case). Cut off was determined as above 2-fold difference; no positive signal was detected (Figure 5.4.7. and 8).

Table 5.4.5 Microarray patches were probed with normal mouse that was either depleted with helper phage (conditions 1-8) or with mAbs that were previously purified from spiked normal mouse sera with a protein G beads purification kit (condition 9-16).

Condition		NMS dilution	mAbs	mAb dilution
1	NMS was depleted with helper phage	10^{-2}	-	-
2		10^{-3}	-	-
3		10^{-2}	SAF84 & SAF15	10^{-3}
4		10^{-3}	SAF84 & SAF15	10^{-3}
5		10^{-2}	SAF84 & SAF15	10^{-4}
6		10^{-3}	SAF84 & SAF15	10^{-4}
7		10^{-2}	SAF84 & SAF15	10^{-5}
8		10^{-3}	SAF84 & SAF15	10^{-5}
9	Purified Abs with protein G beads	10 ul	SAF84 & SAF15	10^{-2}
10		1 ul	SAF84 & SAF15	10^{-2}
11		10 ul	SAF84 & SAF15	10^{-3}
12		1 ul	SAF84 & SAF15	10^{-3}
13		10 ul	SAF84 & SAF15	10^{-4}
14		1 ul	SAF84 & SAF15	10^{-4}
15		10 ul	SAF84 & SAF15	10^{-5}
16		1 ul	SAF84 & SAF15	10^{-5}

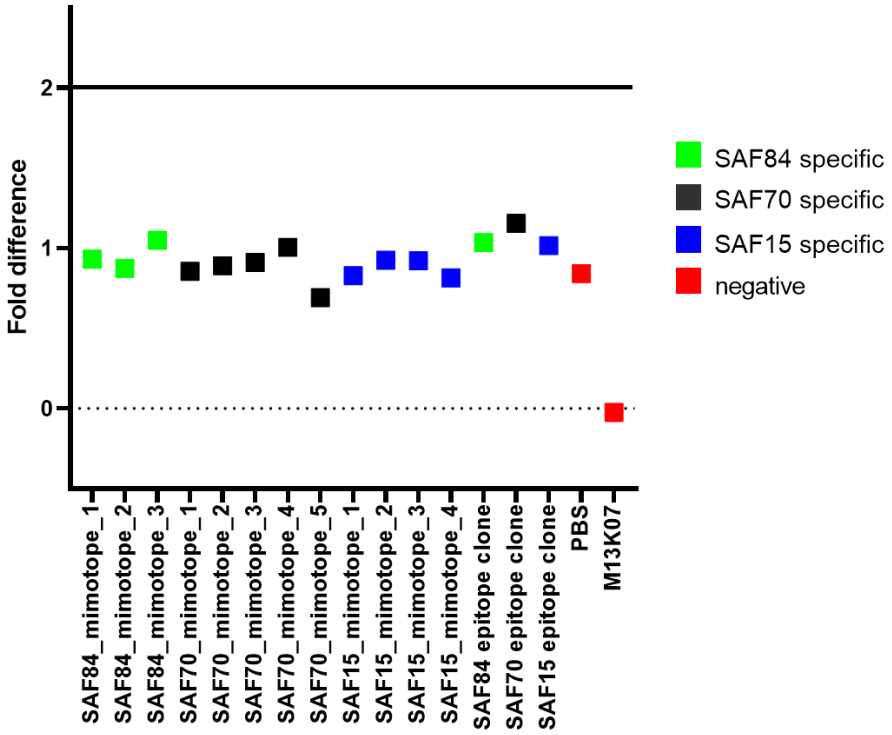


Figure 5.4.7 Testing of a phage-peptide microarray when probed with 1 µl of normal mouse sera (spiked with SAF84 and SAF15 mAbs) depleted by binding to M13K07 helper phage.

A microarray patch was probed with NMS spiked with SAF84 and SAF15 (condition 3, Table 5.4.5.) that was pre-blocked with M13K07 helper phage. Cut off of the assay was determined 2x fold difference (black line).

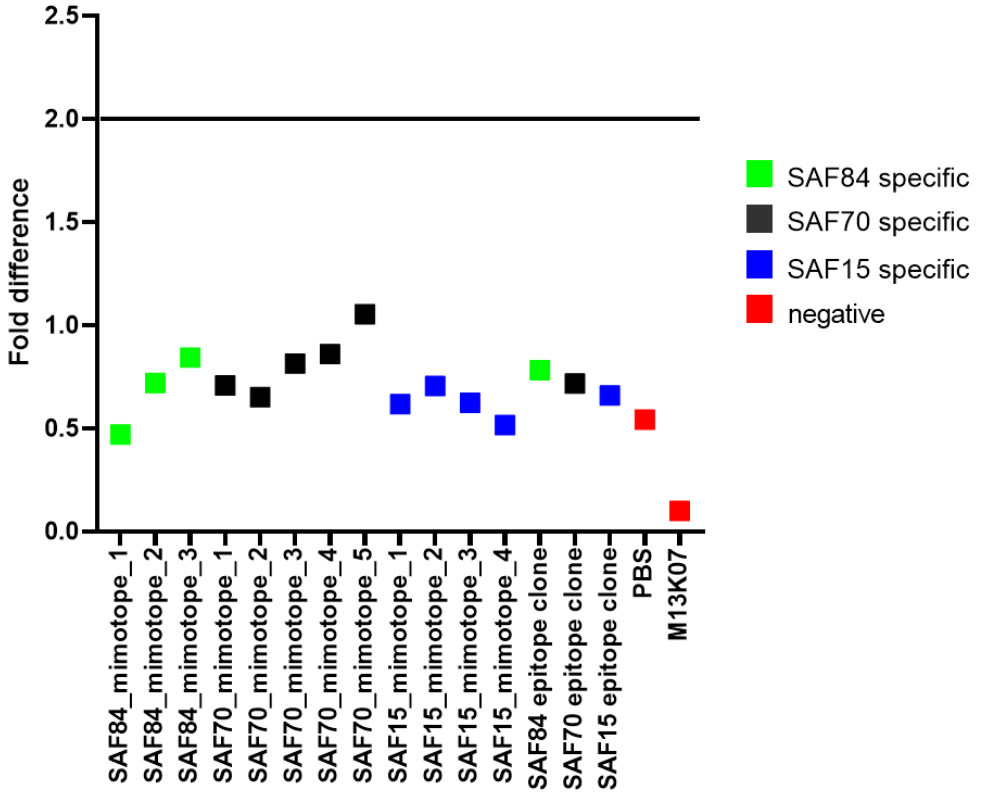


Figure 5.4.8 Testing of a phage-peptide microarray when probed with 0.1 µl of normal mouse sera (spiked with SAF84 and SAF15 mAbs) and was depleted with M13K07 helper phage.
A microarray patch was probed with NMS spiked with SAF84 and SAF15 (condition 4, Table 5.4.5.) that was pre-blocked with M13K07 helper phage. Cut off of the assay was determined 1x fold difference (anything above the negative signal, black line). Unspecific signal from SAF70 mimotope_5 was the only one detected above the cut-off value.

As a further example, as the background signal was extremely high, and positive clones were not able to be distinguished patches were probed with antibodies derived from protein G beads purification of spiked or no spiked normal mouse sera in different mAbs dilutions (Table 5.4.6.). Patches were probed with 50 µl of the purified mAbs (produced from between 1 and 30 µl of sera) for all conditions tested. Firstly, purification of antibodies was confirmed by SDS-PAGE (Figure 5.4.9.A.). Antibody

purification did not improve the specificity or the sensitivity of the microarray assay though. For instance, purified mAbs from spiked NMS were used instead of crude NMS was used to probe a microarray patch (condition 2, Table 5.4.6.). Unspecific signal was observed from SAF70 mimotope_2 as well as from sample containing only PBS (Figure 5.4.9.B.)

Table 5.4.6. Normal mouse sera was spiked with mAbs SAF84 and SAF15 in various dilutions, and Abs were then purified with protein G beads. Patches were then probed with 50 µl of the elution.

Table 5.4.6. Microarray patches were probed with mAbs that were previously purified from spiked normal mouse sera with a protein G beads purification kit using different volume of beads.

Condition	mAbs	mAbs dilution	Sera used	Beads used
1	-	-	1 ul	10 ul
2	SAF84 & SAF15	10^{-3}	1 ul	10 ul
3	SAF84 & SAF15	10^{-4}	1 ul	10 ul
4	SAF84 & SAF15	10^{-5}	1 ul	10 ul
5	-	-	5 ul	10 ul
6	SAF84 & SAF15	10^{-3}	5 ul	10 ul
7	SAF84 & SAF15	10^{-4}	5 ul	10 ul
8	SAF84 & SAF15	10^{-5}	5 ul	10 ul
9	-	-	10 ul	20 ul
10	SAF84 & SAF15	10^{-3}	10 ul	20 ul
11	SAF84 & SAF15	10^{-4}	10 ul	20 ul
12	SAF84 & SAF15	10^{-5}	10 ul	20 ul
13	-	-	30 ul	60 ul
14	SAF84 & SAF15	10^{-3}	30 ul	60 ul
15	SAF84 & SAF15	10^{-4}	30 ul	60 ul
16	SAF84 & SAF15	10^{-5}	30 ul	60 ul

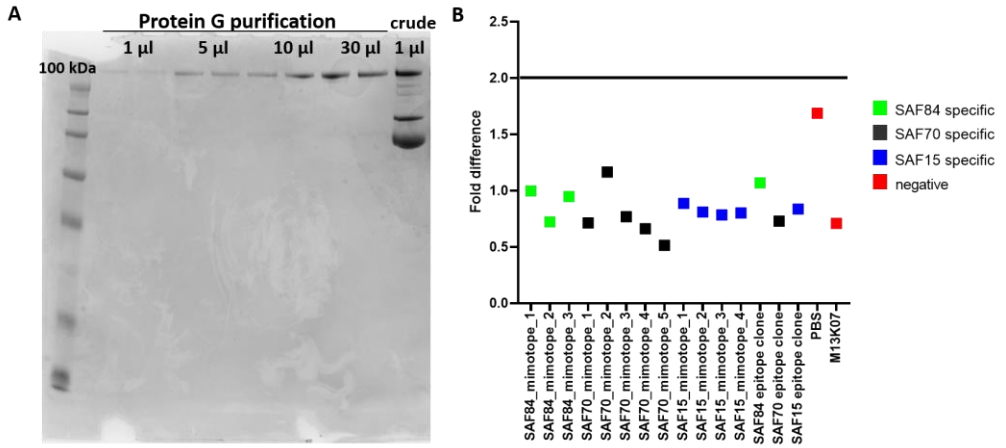


Figure 5.4.9 Assessment of antibody protein G purification followed by a phage-peptide microarray assay when probed with the same purified Abs from spiked with SAF84 and SAF15 mAbs NMS.
A. Purification of antibodies with protein G beads was confirmed with SDS-PAGE for representative samples derived from a range of NMS volumes (1-30 µl). The size of the heavy chains was 100 kDa, as expected. The increasing amount of Abs correlated with the increasing sera volume that was used for the purification. Crude sample (NMS without purification, 1 µl) was included. **B.** A microarray patch was probed with purified Abs derived from NMS spiked with SAF84 and SAF15 mAbs (condition 2, Table 5.4.6.) Cut off of the assay was determined 2x fold difference; no positive signal was detected.

The reproducibility of the assay was needed to be validated within the same slide when patches 2 and 10 were probed with the same condition (NMS was spiked with 10^{-2} SAF84 and SAF15 mAbs, 10 µl of it was purified with protein G beads and 50 µl of the elution was used for each patch, condition 2, Table 5.4.7.) whereas patches 6 and 14 were probed with the same condition but with higher volumes of purified antibody (NMS was spiked with 10^{-2} SAF84 and SAF15 mAbs, 10 µl of it was purified with protein G beads and 100 µl of the elution was used for each patch, condition 6, Table 5.4.7.). Notably, even though these four patches were probed with identical or similar conditions (only the amount of elution used was different), overall signals were extremely

variable and statistically significant between patches probed with the same NMS conditions (Figure 5.4.10).

Table 5.4.7 Normal mouse sera was spiked with mAbs SAF84 and SAF15 in various dilutions, and Abs were then purified with protein G beads.

Condition	mAbs added	mAb dilution	Sera (ul)	Beads (ul)	Elution (ul)
1	-	-	10	20	50
2	SAF84 & SAF15	10^{-2}	10	20	50
3	SAF84 & SAF15	10^{-3}	10	20	50
4	SAF84 & SAF15	10^{-4}	10	20	50
5	-	-	10	20	100
6	SAF84 & SAF15	10^{-2}	10	20	100
7	SAF84 & SAF15	10^{-3}	10	20	100
8	SAF84 & SAF15	10^{-4}	10	20	100

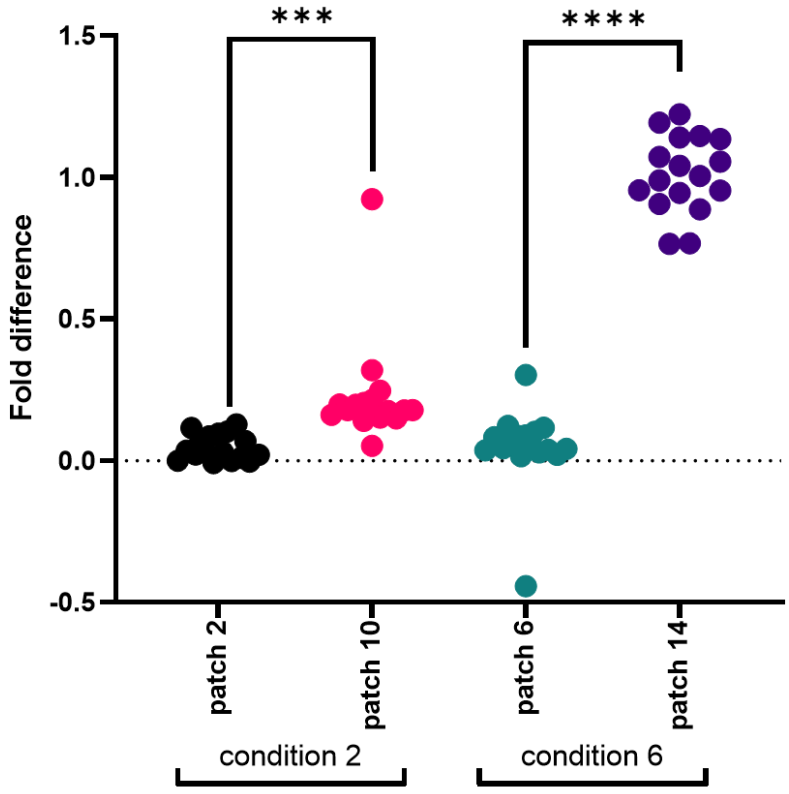


Figure 5.4.10 Assessment of the reproducibility of the phage-peptide microarray assay when probed with purified antibodies from spiked NMS with SAF84 and SAF15 mAbs.
Two patches per condition were probed with NMS spiked with 10^{-2} SAF84 and SAF15 mAbs (condition 2 and 6, Table 5.4.7.). There was significant difference between the obtained signals from different patches when probed with the same spiked NMS conditions.

To summarise, bacteriophages displaying mimotopes of known mAbs were tested with a wide range of conditions in order to reduce the background and increase the specificity and sensitivity of the phage-peptide microarray assay. Unfortunately, when patches were probed with sera or antibodies purified from sera, there was no evidence that this assay can be reproducible and its limit of detection could not be determined because in the vast majority of the experiments, bacteriophages displaying SAF70 mimotopes or negative control

samples (PBS only) were positive when slides were probed with NMS spiked with SAF84 and SAF15 mAbs in various concentrations.

5.5. The development of a Next Generation Phage Display Immunosignature assay

An alternative screening method for cancer diagnosis could be the use of the immunosignature, which is the positive signal that is being obtained when a random peptide array is being probed with sera containing (auto)antibodies and can contribute towards the diagnosis of any disease in which autoantibodies are present (Stafford et al., 2014). Similar to this, this tested approach is that instead of utilising a random peptide array, a soluble assay in which previously identified by Next Generation Phage Display enriched peptides could be applied instead. Firstly, the limit of detection of this novel approach was investigated. Two iterative rounds of biopanning were conducted. Initially, pc89_BspQI^{-} 16mer library was panned against normal mouse sera that was spiked with the monoclonal antibodies SAF70 and SAF15 at 10^{-2} (condition 1), 10^{-3} (condition 2) and 10^{-4} (condition 3). During the first biopanning round, 5 replicates were included per condition (Figure 5.5.1.). The resulting phage sub-libraries from these replicated samples were then pooled together, creating three different input bacteriophage sub-libraries. Each of these sub-libraries were panned against 5 replicates of NMS spiked with three different SAF70 and SAF15 mAbs dilutions: spike at 10^{-2} (condition 1), 10^{-3} (condition 2) and 10^{-4} (condition 3). Fifteen replicates of condition 4 (normal mouse sera containing no known mAbs) were also included as negative

controls (Figure 5.5.1). All 90 samples were individually amplified by PCR and analysed by NGS.

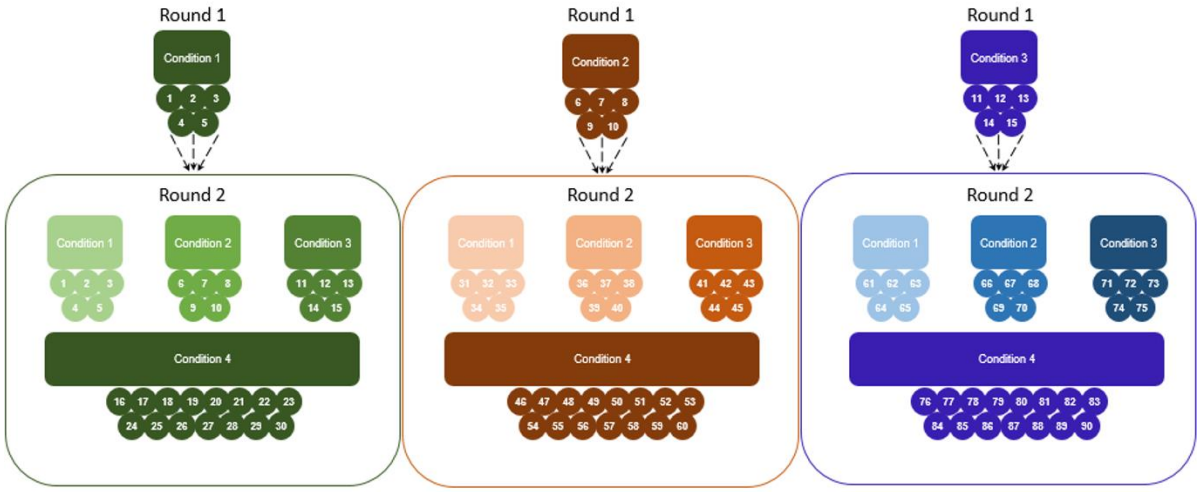


Figure 5.5.1 Schematic overview of the immunosignature biopanning strategy that was followed to test its limit of detection.
NMS was spiked with SAF70 and SAF15 mAbs in 10^{-2} (condition 1), 10^{-3} (condition 2) and 10^{-4} (condition 3). The pc89_BspQI^{-} 16mer library was interacted with 5 replicates from each condition (Round 1). Eluted samples were pooled together from each condition creating three phage sublibraries (colour coded). Each sublibrary was used for the second round of biopanning against 5 replicate samples from condition 1, 5 replicate samples from condition 2, 5 replicate samples from condition 3 and 15 replicate non-spiked NMS samples (condition 4).

Initially, data from replicate samples were pooled together and 12 different datasets were created (Table 5.5.1.). The percentage of the previously identified amino acid motifs that resemble the true epitopes of the SAF70 and SAF15 mAbs (PWP for SAF70 and GW for SAF15, in any position) was calculated within all the datasets and plotted as fold difference in comparison with their corresponding negative samples (e.g. percentage of PWP in a round 2 condition 1 dataset vs percentage of PWP in the corresponding round 2 condition 4 dataset). For presentation, every calculated fold difference above 20, was equalised to 20. The cut off of the assay was 2-fold (black line, Figure

5.5.2.). The occurrence of the SAF70 motif (PWP) was present in almost all the sets, with its fold difference ranging from 5 to 80 (Figure 5.5.2, black bars). On the contrary, the percentage of the putative SAF15 epitope motif (GW) in all the data sets was lower in comparison with the SAF70 epitope motif and was below the cut-off (Figure 5.5.2, pink bars). Taken together, the SAF70 epitope motif was predominant in some cases (up to 80% of the total sequences in some cases) but that was not the case for the SAF15 epitope motif. In fact, the lowest dilution of SAF70 mAb within NMS in which the percentage of the SAF70 motif was detectable above the cut-off value was 10^{-3} for the first biopanning round and 10^{-4} for the second round. This was the most sensitive yet described limit of detection of a method when compared with conventional ELISA or phage qPCR that were prior attempted using the same mAbs spiked in NMS.

Table 5.5.1 Description of the conditions of the Next Generation Phage Display Immunosignature panning, and the data sets derived from different input (R1) and immunosignature (R2) panning conditions.

Data set	Samples ID	Input phage	Immunosignature panning
set 1	1-5	Condition 1	Condition 1
set 2	6-10	Condition 1	Condition 2
set 3	11-15	Condition 1	Condition 3
set 4	16-30	Condition 1	Condition 4
set 5	31-35	Condition 2	Condition 1
set 6	36-40	Condition 2	Condition 2
set 7	41-45	Condition 2	Condition 3
set 8	46-60	Condition 2	Condition 4
set 9	61-65	Condition 3	Condition 1
set 10	66-70	Condition 3	Condition 2
set 11	71-75	Condition 3	Condition 3
set 12	76-90	Condition 3	Condition 4
**condition 1: 10^{-2} SAF70 and SAF15 in NMS* **condition 2: 10^{-3} SAF70 and SAF15 in NMS* **condition 3: 10^{-4} SAF70 and SAF15 in NMS* **condition 4: NMS only*			

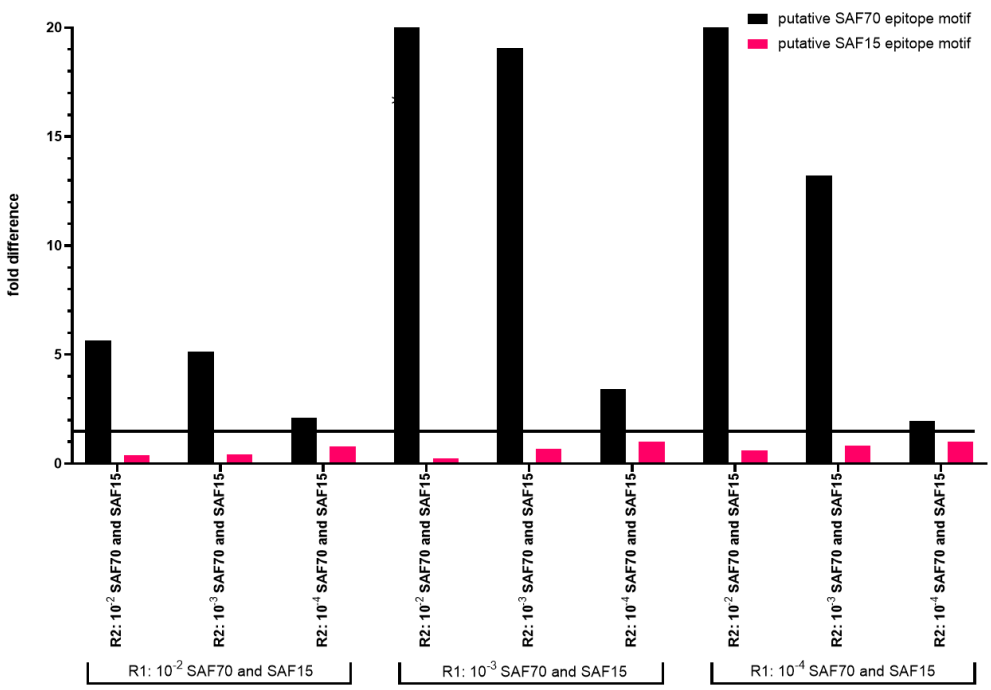


Figure 5.5.2 Detection of known epitope motifs of SAF70 and SAF15 mAbs in an immunosignature assay using NGPD.
The SAF70 (PWP, black bars) and SAF15 epitope motif percentage (GW, pink bars) was calculated in all the data sets (Table 5.2.2), and expressed as fold difference with their corresponding negative control (samples with the same input phage, no SAF70/SAF15 mAb in R2). For presentation the highest fold difference shown is 20 (actual fold difference was 41 and 89, respectively). The cut off of the assay was 2-fold difference (black line); PWP motif fold difference was above the baseline in almost all the sets whereas GW motif occurrence was significantly lower and similar in all the groups.

The aim of the following NGS analysis was to apply more stringent criteria to select potential diagnostic peptides that can form a mAb specific-immunosignature pattern which can be used as an alternative screening method. Firstly, all samples in which the first biopanning round condition was the same (condition 2, samples 31-60, Table 5.5.1.) were set as the training cohort. This specific condition would be indicative of the specificity and the sensitivity of this approach since the concentration of the spiked SAF70 and SAF15 mAbs in the NMS was between the ranges of all the tested

concentrations (10^{-3}). Therefore, the 200 most enriched sequences were identified within samples 31-45 (same R1 condition, R2 condition 1-3, spiked mAbs concentration varied from 10^{-2}-10^{-4}) and 46-60 (same R1 condition, R2 condition 4, NMS only) were pooled and condensed. Peptide sequences that were absent from the 200 most enriched sequences in condition 4 (R2, NMS only) and were present within at least 4 out of the 15 replicates (samples 31-45) were selected (n=157, Table 5.5.2.). The sum of these peptides frequency was averaged, both within training and testing cohorts; the latter consisted of samples 1-30, same R1 condition (SAF70 and SAF15 10^{-2}) and samples 46-90, same R1 condition (SAF70 and SAF15 10^{-4}). The cut-off of the assay was determined as 10x the mean of frequencies of these immunosignature specific peptides within the control group of the training cohort (condition 4, R1: 10^{-3} SAF70 and SAF15, R2: NMS only). According to these criteria, the lowest mAb concentration in which the sum of these peptides frequency was above the cut-off value, and therefore detectable, was 10^{-3} in both conditions 1 and 2 (Figure 5.5.3.). The sensitivity of this assay when using these specific peptides was determined as 66.7% and its specificity as 100% (as none of the NMS only sum of frequencies exceeded the cut-off value).

Table 5.5.2 Enriched amino acid peptide sequences that were selected with stringent NGPD criteria (present within ≥ 3 replicates when SAF70 and SAF15 spiked dilution in NMS was 10-3 and absent from NMS only) (n=157).

Immunosignature specific peptides (n=157)			
YPHMAASQSTTMQSTP	PWPHYIDGLVEKNTGV	PWPLSDPKSTTTAAYA	PWPSPENPMAMLPTLH
PWPTLPHMALSYSMPE	PWPHTPSEVLKFTEPP	PWPIPSGNAPNLAVRP	PWPSEVAQSMGAQGRP
PWPSAPSHAVPTDTLL	PWPELPQNMHTSLHTP	PWPHPPTDPHTGPHLA	PWPSDTSLPTHSVHVE
PWPTPKGPHPTPMPPA	PWPATPPHLPPSMLIP	PWPDPSGQPKWPEQAA	PWPSDLSRTPALRAES
PWPMVPPPEFRSRNVP	PWPAPTLLYEEPLKSY	PWPAPRTADTKNVVDT	PWPSAMPPPTNPGVRT
PWPAKPSPIHTCRGTV	PWPAPLHPSRADVPPA	PWPANLSITSQGSIRP	PWPRTRDVAPNLPAPW
PWPMPTSTSVPEHRTA	PWPAIPSHLIKHTATA	PWQVPSMPAPLQVGPA	PWPRMASANYATSRPN
PWPVPELALKPATMPH	PWPAAAPSPSSRHLNP	PWPVPASTLELPANSP	PWPPVTPAPGLLTSSP
PWPTAAPADLNPPDAT	PWPAAAGFIHQMTPSH	PWPVLPSPNHAKSPND	PWPPSPPNTPLHIST
PWPRYAPKWCGATPDH	PWPVMVPFERVAPTPP	PWPVAGSPSVPPSELT	PWPPMGPSTQSARLTS
PWPSPVHYVCLPCSEP	PWPVAVMTNSPENYTP	PWPTAHPKEHSGSTTS	PWPMTAPKPNVMTSQT
PWPSFPDNNVITVNPE	PWPTVVRDSAVIIQSS	PWPSIPHHLTLSTQNV	PWPLVIHEHKHTLEIS
PWPRLPLGQQDSSSSA	PWPTPTPDSGKSTASS	PWPSHITEVPFVLAAH	PWPIVAPRSTCAVIPP
PWPNKPPPYTSPATNA	PWPTPRGSINMPTAPR	PWPSAPYT	PWPHRSVDTFMHAPTP
PWPARPTSLQGSSLIT	PWPTPPTYNANQEHRV	PWPRPGDTPRLSPITA	PWPHQEQMHSETTAAI
PWPVRLESGTLLPLSQ	PWPSPPKPLQNPSKPL	PWPRIPKFDGVQPGLH	PWPDVRAQPGMASLTS
PWPVPTWGPTIEPTHK	PWPSLPTSTPLPLLHP	PWPRETPDLPAKELYQ	PWPDADPVSRALDTPH
PWPVPPSKTPAYTLLR	PWPNPVPEGVHSSQTD	PWPQRPQSTSQLRELP	PWPARPLPNPTRAEWS
PWPVPNKNRPEHTHAG	PWPAAPSSS	PWPQPHCPTMAPINRT	NSTNSYFPWPSTPNHT
PWPVNPPTASDYENWH	PWPTRQESHTYGAYSS	PWPPRPPSAGSTVDTT	PWPTVPELTTFSLQFD
PWPTTPPYELSTLREP	PWPTPSLYTSTWSKTS	PWPNPTTPCESRGNYQ	PWPTSPMLATNIPIFS
PWPTPVPISQSQLRTI	PWPTLPTTNGDKRETV	PWPMQLPSKHMPLKTP	PWPTPVGTTPPKPLPS
PWPTPTPPISPLLKTT	PWPSTYPSATPTAGGN	PWPMPPKHAYYTGTAL	PWPTPGFPSPIVPHGL
PWPTPTGPYRPTSLTH	PWPSPTPLPSWATTDI	PWPMHDPHQATTLHTA	PWPTDPNLNQGAAPMC
PWPTPRPELPTNRLLT	PWPRNPRDLENVPMFL	PWPHPSSLGPNSRVTP	PWPSTDGQTRALSNLH
PWPTPADTLLPTPVKP	PWPQITPGPNNQATLV	PWPHPRPPRATQPILA	PWPSPSANTHGYQLQH
PWPTLPLPSETSPRHL	PWPMYNERPTRTPQSV	PWPARPQSTPSNITPL	PWPSPIAPPEQRSHTL
PWPTLHPTHISPVQEH	PWPEDKTRTHILNAPL	PWPAPPNRYTKQIEIA	PWPSLRNETTTRRNLS
PWPTFQHDPTSAVPNI	PWPARIPDPESYSMPS	PWPANPHSMSPKPPAH	PWPRHAPSTITTPVLS
PWPSTPRKLPDVSPRI	PWPAFPGNTTSQQSLTP	PWQLPPQVAVNSMPLP	PWPPHPVTLSTYMTEL
PWPSRSAMPPLDNHSA	PWPTYAPYLDLTPRPE	PWPTVTLAPAMEPHLL	PWPNPTQPSFTNQPAL
PWPSQPIETRETLIAH	PWPTSPPTHPHKVVYL	PWPTPSTLHSTQMMPH	PWPLPLEALSSKTAER
PWPSPTSQERHVSSPL	PWPTPTRTVNHCYAQR	PWPTPGTPLNIRMSTS	PWPHHAANAPDHTTLA
PWPSPPTQCTRLLTTS	PWPTPPHALGNCKAND	PWPTHLSPANTNDTLP	PWPEFLAETPQRMLGA
PWPSPKNPSAGLMLPL	PWPTMPCTKGTQQKLP	PWPTHGTYTRDSPLPP	PWPDTIFYLNAPDRTQ
PWPSPHHTTLPLPVNT	PWPTKTEAYMQLPATH	PWPTHAESPNPSPTVH	PWPDPPREKIRPVPDY
PWPSLPLPVPARPANV	PWPSPPPTPRLQPLGT	PWPTESHVRINDLPLS	PWPAPESNMLSTQPQS
PWPSIPPHSPSRAADA	PWPPMHPAKMATDIVK	PWPSTYKPHDVYTAQL	
PWPMPPSPRLPTLHNS	PWPPAPLPEPLAFTRL	PWPSSSEKPILTPPLT	
PWPKPPATKSLKVLGE	PWPNPPLSAVTAGHTL	PWPSPRLSPMNTPCIT	

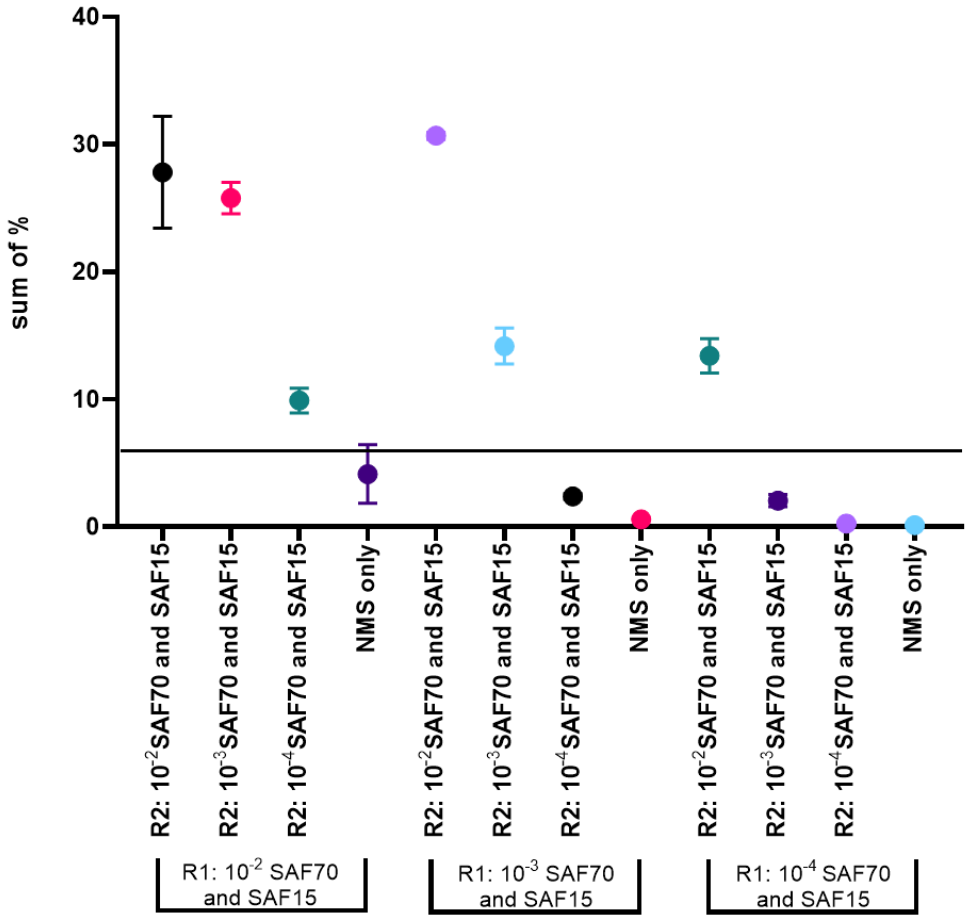


Figure 5.5.3 Assessment of the immunosignature specific peptides that were selected with stringent NGPD criteria (n=157).
Peptides that were enriched within samples with same R1 panning conditions (10-3 SAF70 and SAF15, training cohort) and absent from the corresponding negative training group (NMS only) were selected (n=157). The frequency of these peptides was averaged per replicates and summed in order to create an immunosignature within both the training and testing cohorts. The best cut-off value was determined as 10x (but also 2x, 3x and 5x were assessed, too) the sum of frequencies with the negative group of the training cohort (NMS only). The specificity of the immunosignature was 100% and its sensitivity 66.7%, taking into consideration that NMS only samples were considered negative samples. The lowest concentration in which mAbs were spiked and that was detectable was 10-3 in R1 and R2.

The predominance of the PW SAF70 motif was a common occurrence, especially within samples with the highest SAF70 concentration (10^{-2}, Figure 5.5.4.) Therefore, sequences containing this motif in the first two amino acid

positions that could mask additional mimotopes, were removed from all the data sets, in order to further challenge the screening potential of this assay and potentially identify additional motifs or peptides that mimic the epitopes of these antibodies. Once the same stringent selection criteria were applied, 62 peptides were enriched within the same training cohort and absent from its negative set (NMS only, R1: 10^{-3} SAF70 and SAF15) (Table 5.5.3.). Notably, the motif YP (present within the known SAF70 epitope) was included in 87% of these peptides and YPN was included in 37% of them; on the other hand, GW SAF15 motif was still not detectable. The sum of their frequencies was averaged and summed, as mentioned before, resulting in the same sensitivity and specificity of the immunosignature assay (Figure 5.5.5.). This approach further validates the method, even when the most stringent criteria were applied.

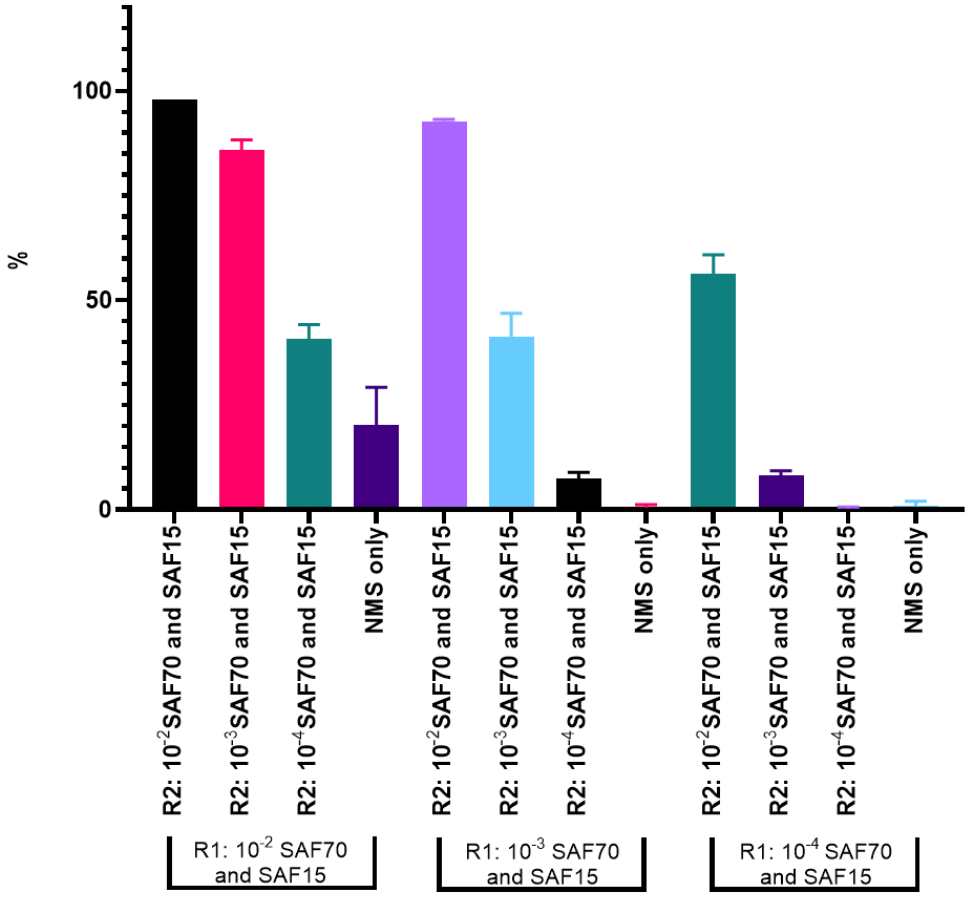


Figure 5.5.4 Estimation of the presence of the putative PW SAF70 motif within the first two amino acid positions of enriched sequences.
Percentage of the 100 most enriched sequences that contained a SAF70 epitope motif (PW) at the first and second position of the amino acids within all NGS data sets.

Table 5.5.3 Enriched amino acid peptide sequences that were selected with stringent NGPD criteria (present within ≥ 3 replicates when SAF70 and SAF15 spiked dilution in NMS was 10^{-3} and absent from NMS only) when sequences with the predominant PW SAF70 motif were removed from the NGS datasets (n=62).

Immunosignature specific peptides (n=62)			
YTLYPVQASASPPNET	PYPNLYPSPASPVTPS	KAYPHLASNPLPNSEW	PYPSLAPAPHPSVWMS
YPVMHGRPAKLTVEAP	PYPNLALTSVVRDIAN	IWYPNLAPTSLNNYVA	PYHLQARPSNESSYPL
YPSQAPSDHSFPSTPS	PYPLQAADMHVNTAAS	IDAIYPNMFHSHLPNL	PTEPLPYPSLFMNELA
YPNLFMTTTPAHAEKL	PYPHVAPNQATTTTTK	HPMPSTVPYPSLVPYS	NYPNQIPHTPISIRTP
YPHQTGLTTATSASPQ	PYPHQLHTIALSTLFD	FTAYPNMAGAATGVPH	NPSAHYKELYPHQHLL
YPHMAASQSTTMQSTP	PYPALAAQSLHLSETL	FPYPNLAQSRVDKAFC	EMPAYPNLANVSPTIH
YPHHYPTLAPSTPSQQ	PVPSQPIETRETLIAH	AYPNLARNPQDTVTPH	TAYPNLALTKAPDVIR
YPGLAGRDAHLTHTGR	PVAYSAWqYoYADCAA	ASMYPLLATHMDKPVS	PYPSQALYSDGSNRSS
TPYPSLAGASRTKDTV	PKPVPYPVLAPLHTEV	YPVMAPYNTEQPVTLT	LYQPYYPNLHLPVTRP
THYPNMHLGPPILVEV	PHYPNLATVQRAEANR	YPNMTLSGPVTTTRLS	LVRPSLYPQLAPHCPD
STPIYLPEYPNQYTPA	PFNKQMATPVRINYLD	YPNMITGQPLNNSPSY	HRPSTKNLVAQYALHQ
SLPYPNQMPNSIVRVP	PEPYPHMVYYPTSLGT	YPHLSRDAQTKELFPS	GRSTKSLIPPAPSTPS
SAVPRYPVLAPQSAPL	NLYPHMAASRNAPTTA	SPPYPYQAQMQAPQRS	ENCNITIQCKTNPVPL
SAGHYPNMHISPSRVL	NHHAYPNMAMYAWRTK	SDTPVPYPILARYEPL	DHPTLTGYPNLATLTP
QYYPILAGFPCSPKKA	NGSRYPNLVGDKYSAN	RRYPNLAAPPPNAVNE	
QTPYPHLASLQREPLP	LPSYPMLAVDSADAHT	QLMPHLAGRAMVDTAS	

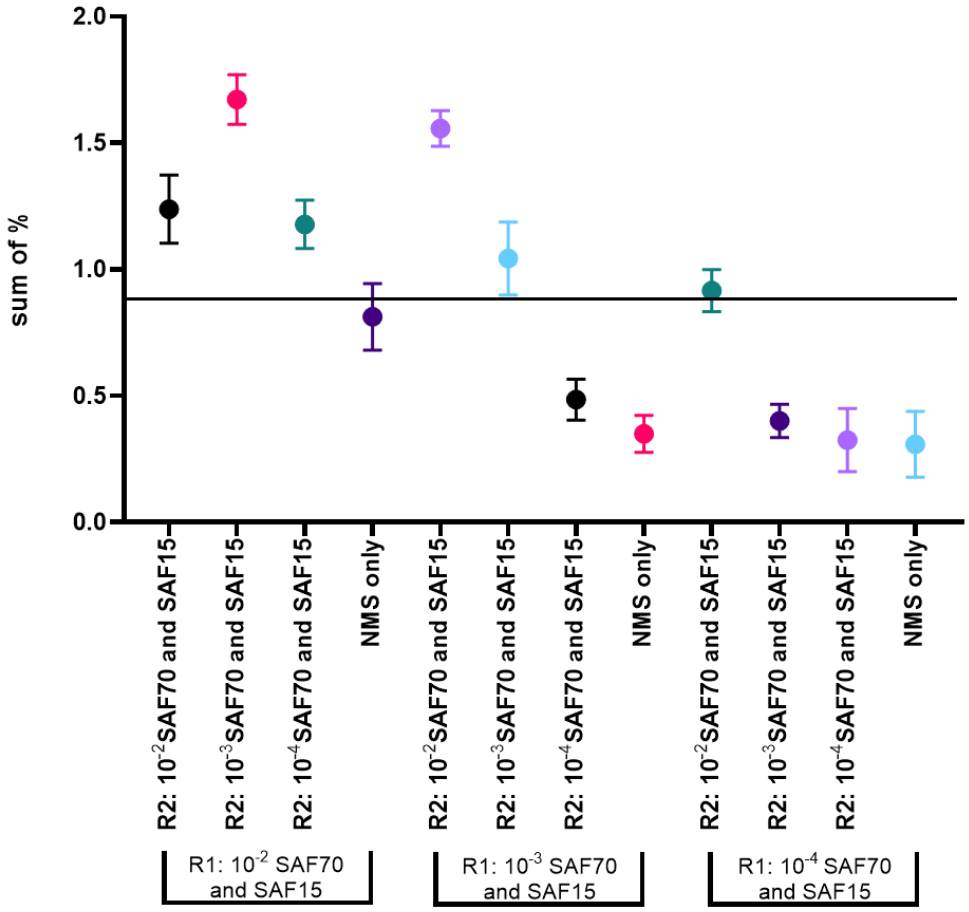


Figure 5.5.5 Assessment of the immunosignature specific peptides that were selected with stringent NGPD criteria, after the removal of SAF70 PW motif (n=62).

Peptides that were enriched within samples with same R1 panning conditions (10^{-3} SAF70 and SAF15, training cohort) and absent from the corresponding negative training group (NMS only) were selected (n=62). The frequency of these peptides was averaged per replicates and summed in order to create an immunosignature within both the training (R1, 10^{-3}) and testing cohorts (R1, 10^{-2} or 10^{-4}). The best cut-off value was determined as 3x (but 2x, 5x and 10x were assessed, too) the sum of frequencies with the negative group of the training cohort (NMS only). The specificity of the immunosignature was 100% and its sensitivity 66.7%, taking into consideration NMS only samples as negatives. The lowest concentration in which mAbs were spiked and that was detectable was 10^{-3} in R1 and R2.

Few peptides were selected (Table 5.5.4.) deriving from all the aforementioned

NGPD criteria in order to be screened as potential SAF70 or SAF15 mimotopes.

Therefore, they were individually cloned using inverse PCR primers (Figure 5.5.6.). Bacteriophages displaying these mimotopes were tested against mAbs SAF70 and SAF15 with conventional phage ELISA in quadruplicates. The vast majority of these clones were SAF70 specific; no SAF15 signal was observed (Figure 5.5.7). In fact, 73.7% of these clones showed SAF70 specificity and none of them cross reacted with a different mAb (SAF15). These data further supported the selection process, confirming SAF70 specificity and validated the immunosignature strategy that enriched peptides can be readily identified in NGS datasets.

Table 5.5.4 Amino acid sequences and clone ID details of 19 peptides identified by NGPD as being enriched against mAbs.
The first five peptides were the 5 most enriched in the majority of the positive datasets; the rest were enriched even when PW motif was excluded from all the data sets.

Clone ID	AA sequence
1	PWPTPPLSPYPSPAYV
2	PWPSYPPSLPSLELLK
3	PWPSPPTTNALIYDTA
4	PWPLHPAMTTPLPNTP
5	PWPKMPISEHPLMYSA
6	YPHHYPTLAPSTPSQQ
7	YPGLAGRDAHLTHTGR
8	TPYPSLAGASRTKDTV
9	SLPYPNQMPNSIVRVP
10	PYPNLALTSVVRDIAN
11	PYPHQLHTIALSTLFD
12	PYPALAPGLPNSTSVD
13	PVAYSAWqYoYADCAA
14	PKPVPYPVLAPLHTEV
15	PHYPNLATVQRAEANR
16	PFNKQMATPVRINYLD
17	HPMPSTVPYPSLVPYS
18	FPYPNLAQSRVDKAFC
19	AYPNLARNPQDTVTPH

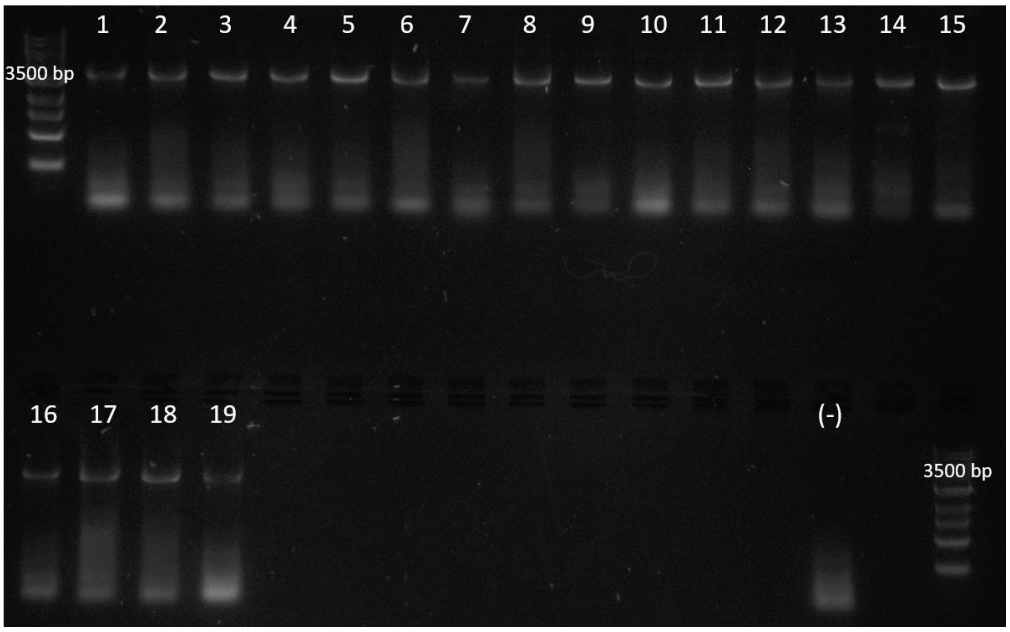


Figure 5.5.6 Confirmation of amplification of rescued clones after inverse PCR (n=19).
Inverse PCR resulted in amplicons at the expected size (~3500 bp) for all 19 targeted sequences; sample containing no template (-) was included as a control for possible unspecific amplification.

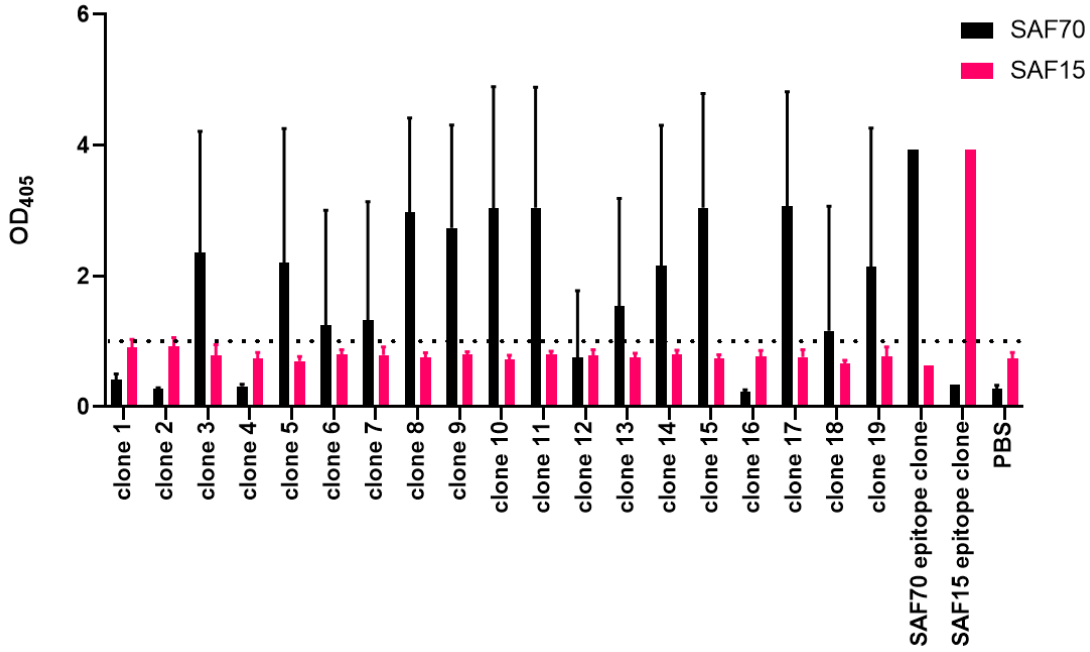


Figure 5.5.7 Confirmation of the specificity of selected bacteriophage clones against SAF70 and SAF15 mAbs.
Monoclonal bacteriophage clones were tested in quadruplicates against 2 different mAbs before the Sanger Sequencing confirmation, thus error bars are high (colour coded). Black line represents the cut off of the ELISA assay (2xSTD above negative signal, PBS only coated wells). Bacteriophages displaying the known mAbs epitope were also included as positive controls. Positive signal was observed in clones 3, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 17, 18 and 19. No positive signal was observed when the same clones were tested against SAF15 mAb.

To sum up, the limit of detection of the immunosignature assay was determined to be up to 10^{-4} initial mAb dilution and up to 10^{-3} after a second round of selection, which was comparable with conventional ELISAs. The selected enriched peptides seem to be able to distinguish the positive from the negative selected samples, suggesting that this assay could be utilised as screening assay for diagnosis.

5.6. Discussion

The standardisation of an immuno-based detection method can be very challenging. The use of reagents that are spiked with dilutions of a known mAb in order to determine the sensitivity and specificity of an assay under development is a common approach. Multiple parameters need to be evaluated, such as the optimal detection system, the dilution factor of the sample and the type of analyte. Especially in diagnostics, the development of a good biomarker candidate relies on its high sensitivity (positive detectable signal even in low abundance) and high specificity (non-occurrence in physiological conditions) (Tighe et al., 2015). ELISAs are the gold standard of serological diagnostics. It is a well-established quantifiable method, however the antigen has to be already identified and immobilised on a solid surface (Legutki et al., 2010). Therefore, the first approach that was tested was the direct comparison of different immobilisation strategies. ELISAs using synthetic peptide, a synthetic peptide immobilised through a biotin tag and phage-peptide were all tested. Phage-peptide detection seemed to be more sensitive than the other approaches, with positive signal detected when the dilution of the spiked mAb was 10^{-3}-10^{-4}, and when spiked NMS was diluted a further 1000x. In the same assay format using amidated synthetic peptide, the LOD was between 10^{-1} and 10^{-2}. This is not surprising as similar results have been reported when the multimer version of a peptide was compared to its monomeric form, and its sensitivity was significantly increased. This is theorised to mimic the polyvalency of the pVIII display (Oyama et al., 2003). In

our case, displayed epitopes were fused with the pVIII coat protein, hence the polyvalency was even greater; the variation on the background of phage ELISAs could be because of the wide range of the level of display on phage particles. However, ELISAs presents its own limitations. Reagents may have cross-reactivity that can lead to false positive results. Furthermore, the hook effect, a phenomenon in which lower signal is observed than the expected one when an excess of analyte is present, is another limitation of this type of immunoassays (Füzéry et al., 2013).

qPCR has been used widely in diagnosis due to its superiority over conventional PCR. qPCR offers a non-binary result and real time quantification of DNA molecules. Determining a cut-off value of a threshold cycle (Ct) is vital for the development of the assay, as it defines the lowest amount of cycles when an amplification can be observed (Caraguel et al., 2011). Taken this into consideration, the pc89 library was spiked with bacteriophages displaying a known mAb epitope (SAF84) and they were then immunoprecipitated in the presence of the same mAb. The limit of detection of this assay was when both the spiking dilutions (mAb and known bacteriophages) were at 10^{-3}, although a high Ct vale had to be established. This sensitivity was similar to the aforementioned ELISA LOD findings.

Alternatively, microarray based detection methods can be used as they have a much higher multiplexing capacity and allow more replicate sample analysis. Moreover, in the case of random peptide arrays they can represent linear protein epitopes or even other immunogenic macromolecules like lipids and

carbohydrates. This assay format has its own caveats because improper protein folding in the case of whole proteins arrays can occur (Legutki et al., 2010; Cekaite et al., 2004). The LOD of the method is also dependent on the immobilisation quality of the peptides/proteins, challenging further the reproducibility of analysis which depends on batch variation (Tighe et al., 2015). Our approach was to use bacteriophages displaying peptides that mimic the known Ab epitopes as the antigens. This approach could potentially lower the cost of the solid microarray assays as purification of the protein/peptide antigens is no longer needed (Cekaite et al., 2004). The limit of detection of a phage-peptide microarray assay was evaluated. Firstly, the stability of the bacteriophage clones that displayed mimotopes of already known mAbs was assessed and their recognition was not affected either by the spotting conditions (in nitrocellulose membrane) nor by the storage over 12 months at 4°C. This has previously been assessed in a study where phage-peptides were spotted onto membrane and breast cancer sera samples were used (Cekaite et al., 2004). However, when microarrays spotted with bacteriophages were probed with spiked normal mouse sera, high background signal was observed in almost all the different conditions despite efforts to resolve these issues with extensive washing, and by testing various blocking conditions and different incubation methods. High background signals have previously been reported with nitrocellulose coated slides, even though this kind of membrane is widely used and they are preferred due to their protein binding characteristics (Tighe et al., 2015). In the presented study, phage auto fluorescence was also considered, but when slides were scanned without the presence of any sera or

detection antibodies, signal was negative (data not shown). Positive signal of microarrays is highly dependent on the surface material. Even reducing the space between spotted peptides by 3-fold, signal can be reduced by 30- or even up to 1,000-fold (Stafford et al., 2012). Moreover, when replicate analysis was carried out in different patches on the same slide, variable results were observed every time. This is supported by the reported low rate of reproducibility of microarrays and their uncertainty for clinical adaptation (Oyama et al., 2003). The data presented with phage-peptide microarrays indicated that a different assay approach may be needed to detect autoantibody biomarkers.

Next, a computational approach in which NGS data on phage-peptide binding was used as an assay readout was investigated, a so-called immunosignature. This approach would be analogous to the outputs from high diversity random-peptide arrays in which random peptides were used for the epitope mapping of mAbs or for profiling the autoantibody response in individuals with cancer or other diseases (Legutki et al., 2010; Hughes et al., 2010; Halperin et al., 2011).

Here, instead of using a large library of randomised peptides immobilised on a solid surface array, two iterative biopanning rounds were conducted in order to identify mAb specific peptide. In the latter biopanning round, the same input phage sublibraries were used when panned against spiked sera containing different dilution of known mAbs (mAb dilutions mimicked possible autoantibodies titrations in cancer sera). This was carried out for 3 different

mAb concentrations in round 1 and 2. Firstly, the identification of known mimotope motifs of SAF70 (PWP) (but not of SAF15) was successful. In fact, SAF70 epitope (PWP) was present in a high percentage within the top 100 enriched sequences, demonstrating the success of epitope mapping after two biopanning rounds. This is in accordance with another group's attempts for epitope map monoclonal and polyclonal antibodies with random peptides (n=10,000) in a microarray format. Specifically, 10 mAbs were epitope mapped and the 500 most enriched peptides for each antibody were very distinct, with only 1.2% overlap between the enriched sequences (Halperin et al., 2011). A plethora of new mimotopes of SAF70 were identified, the vast majority of them containing the aforementioned motif. A handful of newly discovered peptides were tested against SAF70 and SAF15 and 70% of them showed high reactivity against SAF70. Additionally, applying stringent criteria on the NGS analysis, 157 peptides were highly enriched within the samples with spiked NMS. By using the sum of their frequency (the immunosignature of each sample), samples with spiked NMS were able to be discriminated from the non-spiked samples in almost all the mAb dilutions with 67% sensitivity and 100% specificity. In fact, the lowest detectable mAb dilution was 10^{-3}-10^{-4} in both biopanning rounds. Even after the removal of sequences containing the PW motif, 62 peptides were selected as spiked NMS specific; the sensitivity and the specificity of the assay remained at the same level. The data demonstrated the potency of this soluble immunosignature assay to uncover a wide range of specific mimotopes to low concentrations of antibody. Taken all together, deep sequencing in combination with phage display successfully epitope

mapped mAbs that were spiked in NMS (in low concentration). This approach sets a promising alternative method to ELISAs in order to identify peptides or immunosignatures for the identification of autoantibodies to tumour associated antigens.

6. Cancer sera biopanning

6.1. Introduction

Renal Cell Carcinoma (RCC) is amongst the 10 most frequently occurring cancers and its most frequent subtype being clear cell RCC (cc-RCC), making up 75% of cases. The mutation/loss of the VHL gene is an established RCC hallmark, reported approximately in half of the cases which subsequently impairs the hypoxia induced pathway, causing amongst other processes, angiogenesis (Hsieh et al., 2018). Currently, no biomarkers are available for cc-RCC early diagnosis. Tumours derived from RCC are very vascular, therefore this cancer is hard to recapitulate *in vitro*. Animal models are needed in order to investigate the etiopathology of this specific carcinoma. The generation of RCC mouse models has been attempted either with the inactivation of the phosphorylated form of the VHL (Kapitsinou and Haase, 2008) or with a more active form of HIF-α (Fu et al., 2011). Sera from the latter mouse model was available for this study. This mouse model is called TRACK (transgenic model of cancer of the kidney) and it was developed by Weill Cornell Medical College. Briefly, characteristic manifestations of cc-RCC such as the presence of distorted clear cells in the kidney proximal tubule cells with indication of cc-RCC onset from 3-months was present in some cases, renal cysts were confirmed by histochemistry in all mice within 6-7-months and *in situ* carcinoma confirmed within 14-20-months (Fu et al., 2011).

Phage display can be coupled with the analytical depth of next generation sequencing (NGPD). The method was applied to the discovery of enriched

peptides that bind anti-tumour antibodies and mimic the tumour associated antigens (TAA) in order to develop a RCC specific diagnostic assay. In this study, we attempt to identify the best possible immunodiagnostic approach for Renal Cell Carcinoma using initially sera from TRACK mice and then samples from human cc-RCC patients. Two main NGS approaches were applied, either frequency or Z score analysis (Figure 6.1.1.), aiming for 100% sensitivity and specificity; if that was not possible, assay's cut off was adjusted aiming for the highest specificity. All these approaches were applied to all following NGS biopanning data, and the criteria for each analysis were always specified.

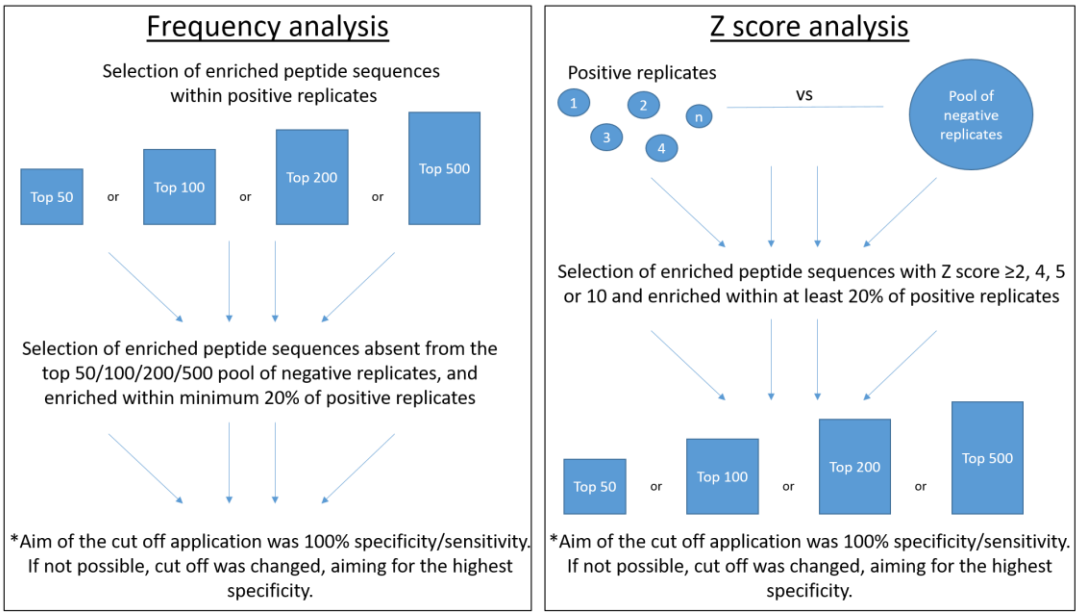


Figure 6.1.1 Overview of the two main approaches of NGS analysis. Frequency analysis: The 50, 100, 200 or 500 enriched peptide sequences were identified within positive replicates. Peptides absent from the top 50, 100, 200 or 500 of the negative pool of replicates and enriched within at least between 20% of positive replicates, were selected. Z score analysis: Positive replicates were compared with a pool of negative replicates. Peptides with Z score ≥ 2, 4, 5, or 10 and enriched within at least 20% of the positive replicates, were selected. Determination of both assays' cut off aimed for 100% sensitivity and specificity but it was adjusted when that was not possible, aiming for the highest sensitivity instead.

Aims:

- Identification of Renal Cell Carcinoma specific peptides that mimic the TAA epitopes recognised by the circulating autoantibodies by applying previously optimised epitope mapping conditions (developed using monoclonal antibodies) in sera derived from the TRACK model.
- Development of a serum-based immunoassay for early RCC detection either by conventional ELISA or a computational NGS based method using the enriched peptide mimotopes in a TRACK model.
- Application of the diagnostic peptides identified with the TRACK model to human samples to develop a diagnostic assay.
- Application of the methods developed with the TRACK model to human RCC sera samples to develop a diagnostic assay.

6.2. First biopanning round against TRACK and WT sera.

Sera from C57BL/6 mouse with the hyperactive HIF-α (TRACK, n=8) or with a wild type HIF-α (WT, n=6) were available for the first part of this study (training set, Table 6.2.1). Bacteriophage (input titration $5x10^{11}$ CFU/ml) of the pc89_BspQI^{-} 16mer peptide library was biopanned against sera from three 12-month old TRACK mice, five 2-month old TRACK mice, three 12-month old WT mice and three 2-month old WT mice, in 6 replicates (Figure 6.2.1.). A subtraction step using a pool of all available WT mice sera was conducted sequentially three times before the biopanning round. Output phage titres of between 2 to 9 x 10^5 CFU/ml were produced. For each sample, plasmid DNA was individually extracted from the bacterial glycerol stocks, and amplified in

two consecutive PCR rounds in order to incorporate Ion Torrent compatible barcode adapters (n=84). Round 1 amplicons (Figure 6.2.2A) and round 2 amplicons (Figure 6.2.2B) were detected at the expected sizes (275 bp and 335 bp respectively). Negative (no template) controls were included in order to ensure the absence of unspecific amplification. These amplicons were quantified by Qubit and a pool of equal DNA contribution was created and gel extracted (Figure 6.2.2C). This DNA pool was further Qubit quantified and a final 50 ng DNA pool was run in a gel electrophoresis in order to confirm the presence of a single band at the correct expected size (Figure 6.2.2D). A 200 ng pool of DNA was sent for Ion Torrent Next Generation Sequencing.

Table 6.2.1 Sera sample ID and details of different ages of TRACK or WT mice.

Sera from either TRACK or WT mice were used in different biopanning rounds and in training or test sample cohorts, as indicated.

ID[1]	Age	R1, R2 & training set	testing set	ID	Age	R1, R2 & training set	testing set
TRACK_1	2-months	y		WT_1	2-months	y	
TRACK_2	2-months	y		WT_2	2-months	y	
TRACK_3	2-months	y		WT_3	2-months	y	
TRACK_4	2-months	y		WT_4	2-months		y
TRACK_5	2-months	y		WT_5	2-months		y
TRACK_6	2-months		y	WT_6	2-months		y
TRACK_7	2-months		y	WT_7	2-months		y
TRACK_8	2-months		y	WT_8	2-months		y
TRACK_9	2-months		y	WT_9	12-months	y	
TRACK_10	12-14-months	y		WT_10	12-months	y	
TRACK_11	12-14-months	y		WT_11	12-months	y	
TRACK_12	12-14-months	y		WT_9	24-months		y
TRACK_1	12-14-months		y	WT_10	24-months		y
TRACK_2	12-14-months		y	WT_11	24-months		y
TRACK_3	12-14-months		y	WT_12	24-months		y
TRACK_4	12-14-months		y	WT_13	8-months		y
TRACK_13	23-months		y	WT_14	8-months		y
TRACK_14	23-months		y				
TRACK_15	23-months		y				
TRACK_16	23-months		y				
TRACK_17	12-14-months		y				
TRACK_18	12-14-months		y				
TRACK_19	12-14-months		y				
TRACK_20	12-14-months		y				
TRACK_21	12-14-months		y				
TRACK_22	12-14-months		y				

[1]Notably, sera from TRACK mice 1-3 were used in two different time points (2 and 12-months) and sera from WT mice 9-11 were used in two different time points (12 and 24-months old).

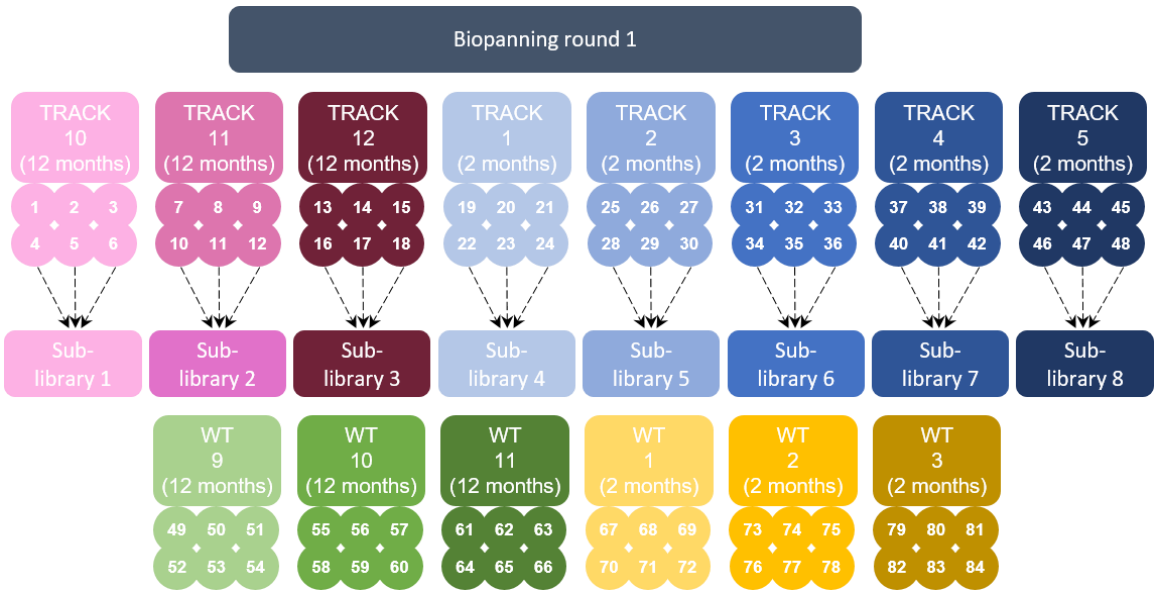


Figure 6.2.1 Overview of the biopanning strategy against TRACK and WT mouse sera of various ages for the first biopanning round.
The first biopanning round was conducted using sera from three different 12-month old TRACK mice, five 2-month old TRACK mice, three 12-month old WT mice and three 12-month old WT mice (Table 6.2.1) in 6 replicates.

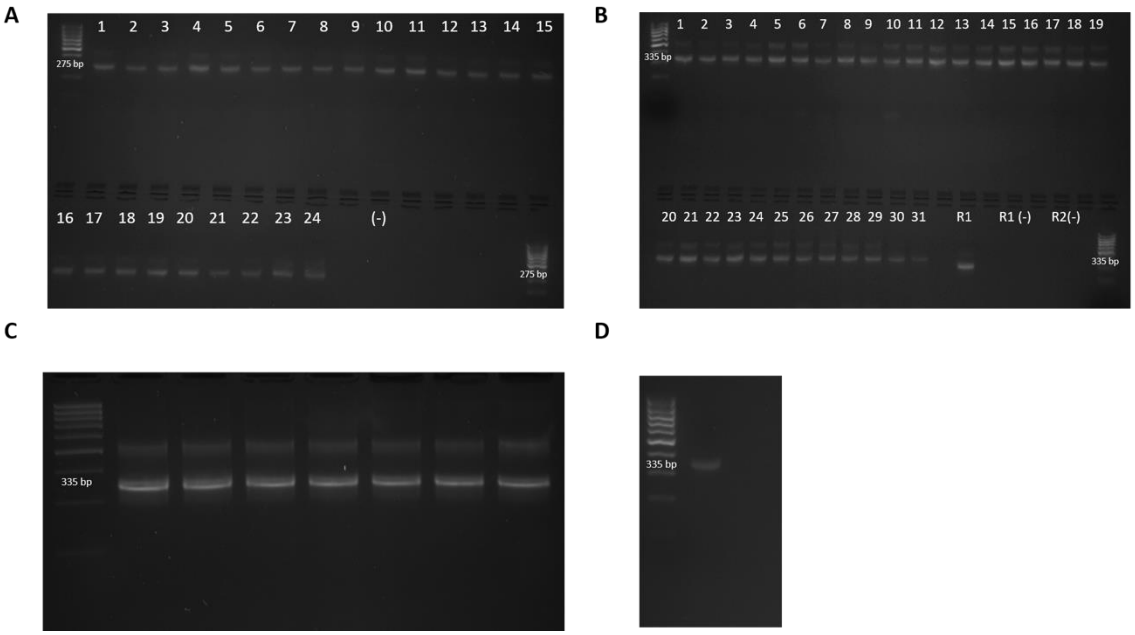


Figure 6.2.2 Production of DNA amplicons for Ion Torrent NGS.
A. Representative gel electrophoresis images after the first PCR round. Amplicons from samples 1-24 from were observed at the expected size (275 bp); negative control (-) was included (contained water instead of DNA template) in order to ensure the absence of unspecific amplification. **B.** Representative gel electrophoresis images after the second PCR round. PCR amplicons from samples 1-31 were observed at the expected size (335 bp). Negative control samples containing either water instead of a PCR R1 product as a template [R2(-)] or containing the R1-no template control [R1(-)] were included in order to determine any unspecific amplification. R1 amplicon was also included to the gel electrophoresis in order to confirm the insertion of the barcodes (R2 amplicons were larger by 60 bp). **C.** DNA concentration of the individual amplicons was assessed and equal amounts of amplified DNA were pooled together. DNA band (335 bp) was gel extracted after gel electrophoresis using Metaphor agarose. **D.** Image from gel electrophoresis after gel extraction and further clean up using Ampure; single band of 50 ng pooled DNA was observed, as expected.

The NGS data was analysed to identify potential diagnostic peptides that mimic TAAs and for their potential to be used in a diagnostic format. Firstly, eight data sets were created containing the 200 most enriched peptide sequences per TRACK sera sample (Figure 6.2.3.). Two further data sets were generated for the pool of the 200 most frequent peptide sequences of 12- and 2-month old WT negative selection (Figure 6.2.3.). Any highly enriched peptides found

within both the TRACK samples and the age matched control groups (e.g. set 9 for 12-month olds or within set 10 for 2-month olds) were excluded (Figure 6.2.4A and B, respectively). Peptides that were recognised by a minimum of 2 positive sera replicates, were selected. Thirteen 12-months of age specific peptides and six 2-months of age specific peptides were identified (Table 6.2.2.). One of these peptide sequences was common between the two age groups. Each peptide's frequency was calculated and their sum was estimated per sera sample using either the 12-month old (Figure 6.2.5A.) or 2-month old TRACK (Figure 6.2.5B.) specific peptides. This was an attempt to design an *in silico* assessment of a panel of peptides that can be used in combination as a diagnostics. A cut-off value was used in order to evaluate this assay (1.5x the frequency sum of the WT age matched control sera). By utilising the sum of percentages the 12-month-specific peptides (13 in total), the sensitivity of the method was estimated at 50%, and its specificity at 100%. Additionally, by utilising the 2-month-specific peptides (6 in total) sum of frequencies, the sensitivity of the method was estimated at 75% and its specificity at 100% (Figure 6.2.5B.). Notably, these peptides seemed to be age specific and did not cross over to detect other samples that were not used in the peptide selection. The sum of the combined TRACK specific peptides frequency (n=18) was also compared with the sum of the same peptides frequency within the WT sera samples. A cut off value was used in order to evaluate this assay was 2-fold difference with the WT age control sera (Figure 6.2.6).

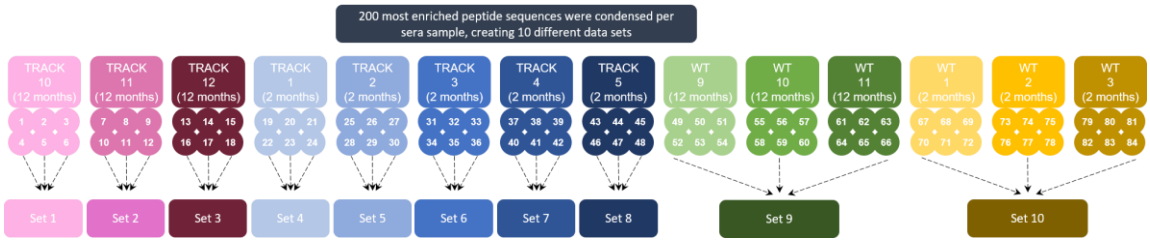


Figure 6.2.3 Overview of the first biopanning round NGS analysis based on the frequency of the enriched sequences.
The 200 most frequent peptide sequences per sera sample were identified from pooled replicate analysis data for each sera sample and were condensed. Sets 9 and 10 were a pool of 18 replicates, creating age-specific negative pools of sequences.

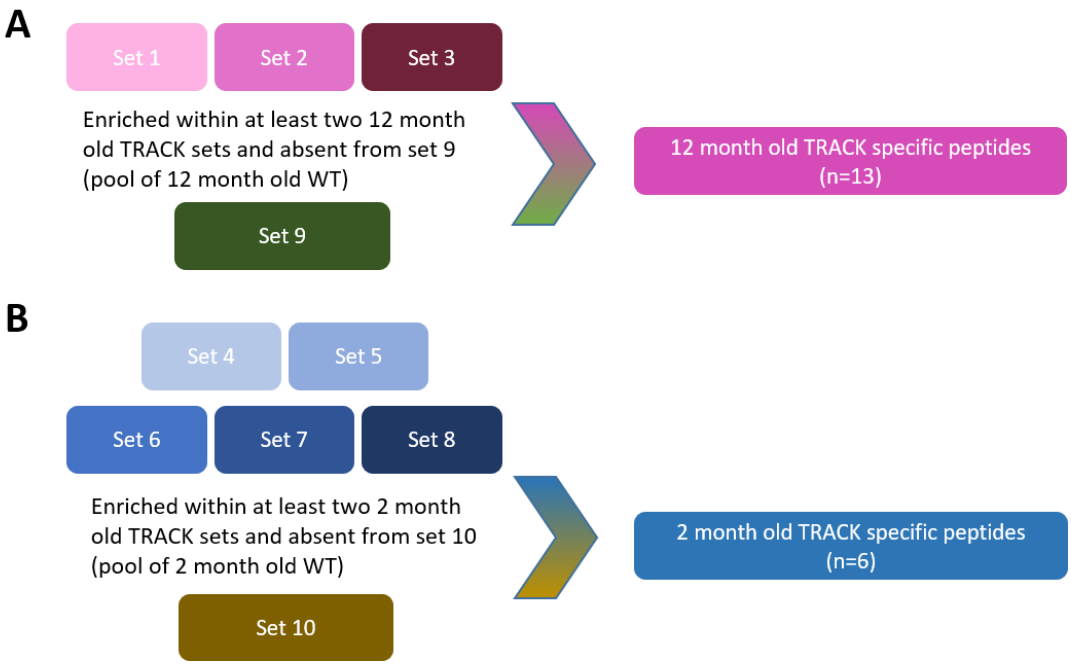


Figure 6.2.4 Overview of the NGS analysis of the first biopanning round in which the 200 most enriched sequences per sera sample were compared with the 200 most enriched sequences per age control sera.
A. Peptides that were enriched in 12-month old TRACK mice samples and absent in set 9 (pool of replicates of 12-months old WT) **B.** Peptides that were enriched in 2-month old TRACK mice samples and absent in set 10 (pool of replicates of 2-months old WT), generated thirteen and six specific peptides, respectively.

Table 6.2.2 List of translated amino acid sequences of the enriched peptides derived from peptide frequency analysis of NGS data from the first biopanning round against TRACK and WT sera.
These peptides were present within at least 2 of the 12 or 2-month old TRACK mice and absent from the 200 most enriched sequences of the WT mice control groups (set 9 and 10, respectively for every age group).

12 month old TRACK specific peptide sequences (n=13)	2 month old TRACK specific peptide sequences (n=6)
TDFRLFHFYYYAAPWN	SHTRMMFQPIRTqSSC
PRAMTTTPNKTHEAVT*	RKQKGAIKLERNKRRT
VqLRKPDLAAPVLQLE	AREQPRVQAGLFPPLI
NNSQPALTSGVRARYT	RSRQMQTLNIPSLVPV
KWMGPKTPRqSqSKLT	**PRAMTTTPNKTHEAVT***
KNVSSSSHHPPPTPTP	LPLPLVPDVPPATLKS
VMPVGTLQEPSLTDTQ	
VHKARSTKRAWDESGL	
PALVHPEPQVQHQQNT	
NqVPqPSAGRTPSIPH	
NEPSCPTKTTqGAEAN	
HSTqLHPTNQSDqSHG	
APPDTHGVATGSNGFA	
***common between the different age groups**	

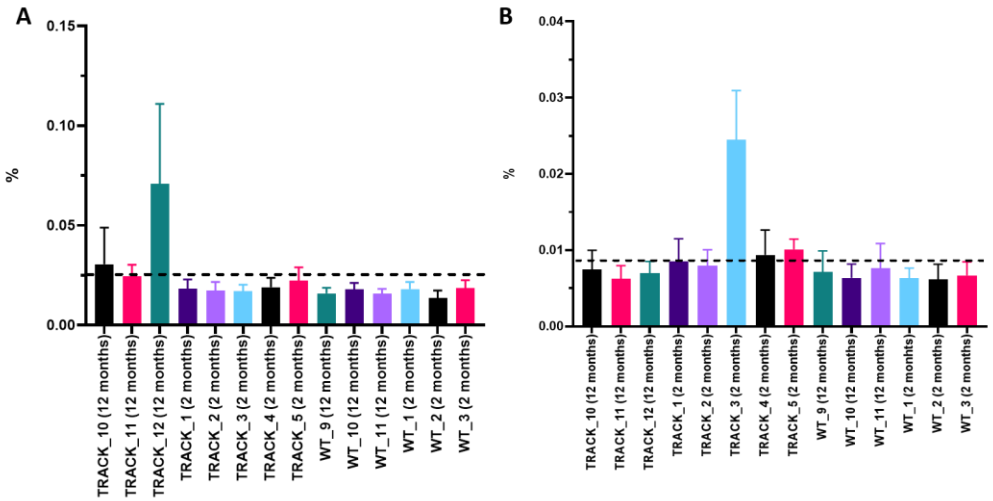


Figure 6.2.5 Assessment of an in silico NGS based diagnostic utilising either the sum of frequencies of the thirteen 12 -month old TRACK-specific peptides or the six 2-month old TRACK-specific peptides.
A. 13 peptides were enriched within at least 2 of 12-month old TRACK mice sets and absent from set 9 (pool of 12-month old WT mice). Their frequency was summed per individual sera sample. A cut off was determined as 1.5x the averaged frequency sum of these 13 peptides within set 9 (negative sample). Positive samples were defined by applying this cut-off value, the 12-month old TRACK sera samples were positive but all 5 2 month TRACK samples were negative, establishing the assay sensitivity and specificity at 50% and 100%, respectively. **B.** 6 peptides were enriched within at least 2 of 2-month old TRACK mice sets and absent from set 10 (pool of 2-month old WT mice). Their frequency was summed per individual sera sample. A cut off was determined as 1.5x the averaged frequency sum of these 6 peptides within set 10 (negative selection). Positive samples were defined by applying this cut-off value, and 4 of 5 2-month old TRACK sera samples were characterised as positives, establishing the assay sensitivity and specificity for all samples at 75% and 100%, respectively.

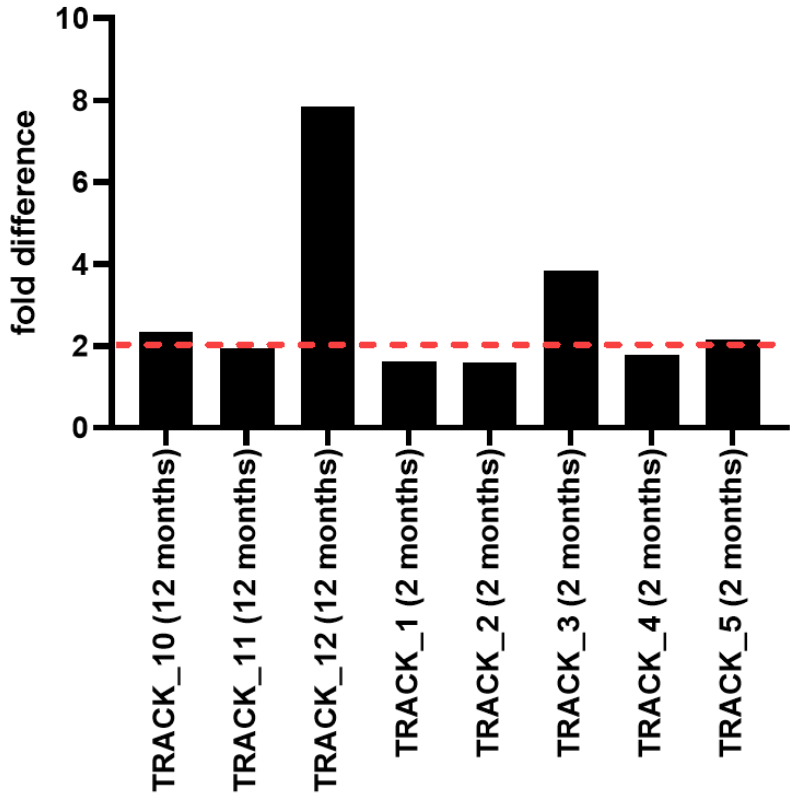


Figure 6.2.6 Assessment of the in silico NGS based diagnostics utilising the sum of frequencies of all TRACK-specific peptides (n=18).
Fold difference between the sum of the 18 selected TRACK specific peptides, as a different way for data visualisation and their sum of frequencies within WT sera. A cut-off value was determined as 2x fold difference and 50% of the samples were considered positive.

Additionally, a two way Z score analysis was also conducted in order to select more peptides that were potential RCC-TAA mimotopes, since it is a more thorough statistical analysis. Two negative pools of sequences were firstly created, by pooling all the sequences of either the 2- or 12-month old WT sera replicates (set 1 and 2, respectively). A Z score for peptide sequences was calculated for data sets from each TRACK sera replicate by comparing the frequency of the peptides within 12-month old TRACK replicates and set 1, or 2-month old TRACK replicates and set 2 (Figure 6.2.7.). Peptides with a Z score ≥5 and present within at least 30% of the TRACK replicates were selected for

each age group. Specifically, fourteen 12-month and seven 2-month old TRACK specific peptides were identified. In total, sixteen unique TRACK specific peptides were selected (Table 6.2.3.).

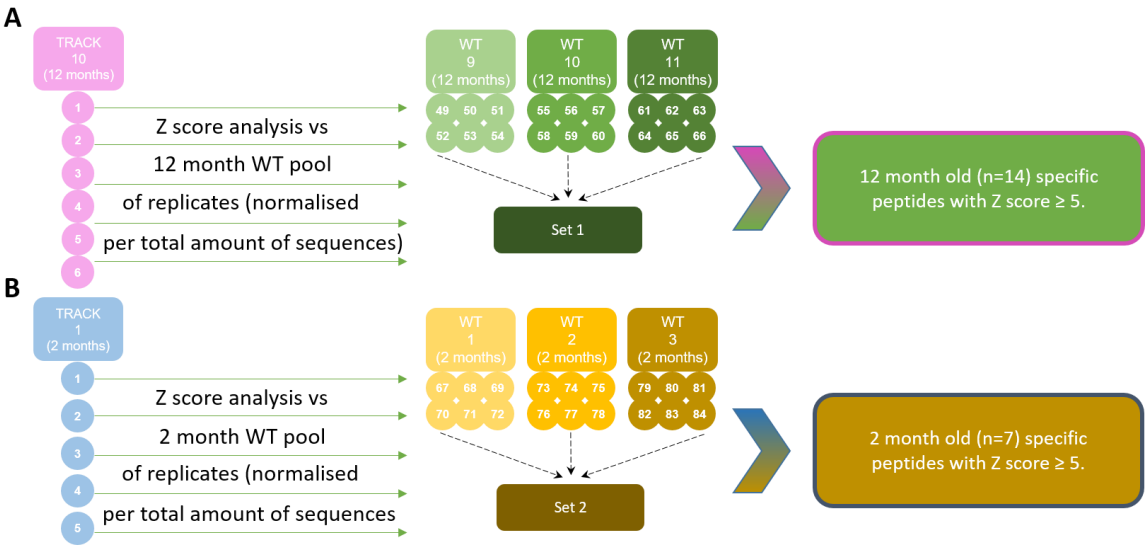


Figure 6.2.7 Flow diagram of the Z score analysis that was conducted utilising NGS data derived from the first biopanning round against TRACK and WT sera (2 and 12-months old) by creating two negative pools of WT replicates based on their age.
A. Z score analysis was conducted between every 12-month old TRACK replicate (e.g. from TRACK 10) and set 1 (pool of sequences of all 12-month old WT mice replicates). 14 peptide sequences with Z score ≥5 and present within 30% of the 12-month TRACK replicates (18 replicates in total) were selected. **B.** Z score analysis was conducted between every 2-month old TRACK replicate (e.g. from TRACK 1) and set 2 (pool of sequences of all 2-month old WT replicates). 7 peptide sequences with Z score ≥5 and present within 30% of the 2-month TRACK replicates (30 replicates in total) were identified.

Table 6.2.3 List of the amino acid sequences of the enriched peptides with Z score ≥5 and enriched within at least 30% of the 12 or 2-month old TRACK replicates.

12-month specific enriched peptide sequences with Z score ≥5 (n=14)	2-month old specific enriched peptide sequences with Z score ≥5 (n=7)
RLNTGWRAALSDIHRQ*	HMMPGDTQPSSASALD
RSRQMQTLNIPSLVPV*	PNWVLENPPPNSWEHG
KNVSSSSHHPPPTPTP*	**RCHTCRDSSNNEPAFT***
HPPFSPELSEWHATQY*	**HPPFSPELSEWHATQY***
RCHTCRDSSNNEPAFT*	**KNVSSSSHHPPPTPTP***
YYWLLHTAWHLSQQNT	**RSRQMQTLNIPSLVPV***
TWFAPLEHTILYRTQA	**RLNTGWRAALSDIHRQ***
SGLLATSGRSPEAYNH	
PAEILSGMRPTNSPDA	
NNSQPALTSGVRARYT	
NNHNSQVLSGIQHRSS	
NASRLNSGMSVQVPR	
HQLAPQISSGLDPAVV	
CHRTTAMASGLKPYLC	
***common between the different age groups**	

The Z score for each diagnostic peptide sequence (n=16, Table 6.2.3.) was calculated for each replicate and their sum was averaged per sera sample. A cut-off value was used in order to assess this assay which was the averaged sum of Z scores of these 16 TRACK specific peptides within the WT pools (Figure 6.2.8.). The eight Z score sum for five out of eight of the TRACK samples were above the cut-off, demonstrating that these peptides ability to be recognised by TRACK sera with 62.5% sensitivity and 83.3% specificity.

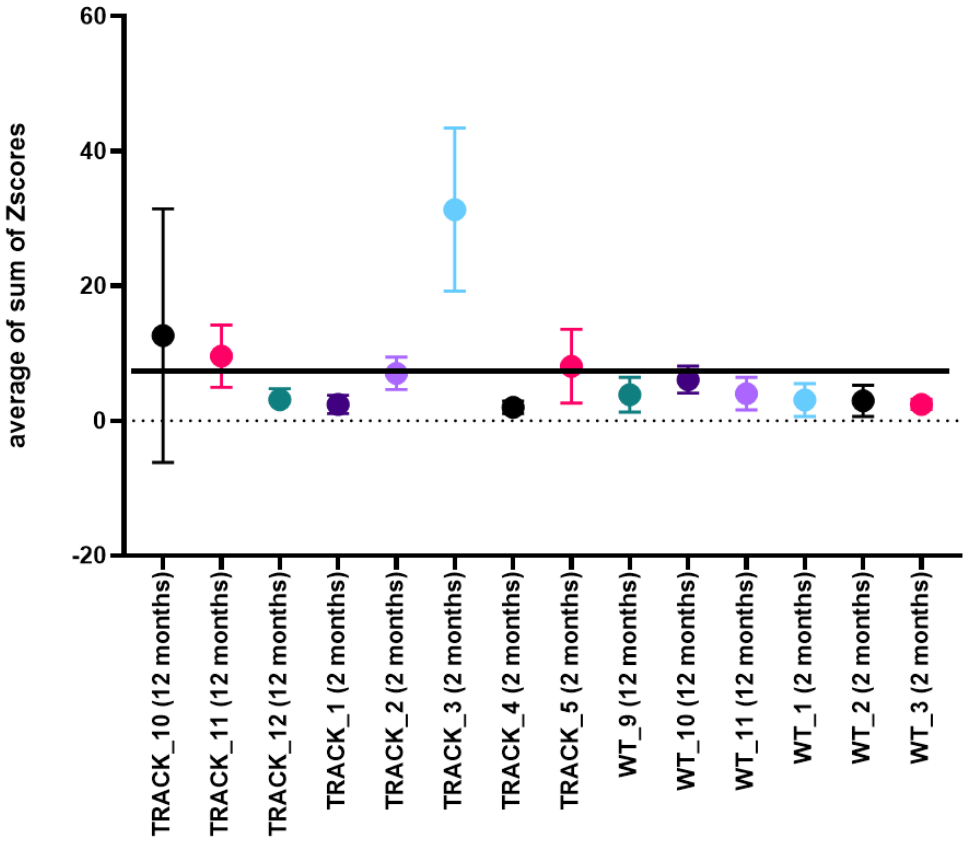


Figure 6.2.8 Evaluation of the averaged sum of Z scores of 16 TRACK specific peptides as a diagnostic in silico approach, derived from the Z score NGS analysis of the first biopanning round using TRACK and WT sera of all ages.
16 unique TRACK specific peptide sequences were identified, summed and averaged per sera sample. A cut-off value was determined as 2x the averaged sum of the Z score within all WT replicates Specificity of this assay was determined as 62.5% and its sensitivity at 83.3%.

6.3. Second biopanning round against TRACK and pools of WT sera.

An additional biopanning round was conducted in order to explore the mimotope repertoire and take advantage of the additional enrichment that can occur after further biopanning rounds. The same sera from TRACK mice at different ages (either 2- or 12-months old) were used; sera from WT mice were pooled depending on their age, creating two sera groups (pool of WT 9-11, 12-months and pool of WT 1-3, 2-months). Eight bacteriophage sub-libraries (1-

8), derived from the first round of output bacteriophage replicates (Figure 6.2.1), were each used for the input phage for the second biopanning round against the sera from the same TRACK mouse (input phage titre $8x10^{10}$ CFU/ml). Three sequential subtraction steps using a pool of all WT mice sera was conducted before the biopanning. To summarise, each sub-library was panned against the same TRACK mice sera as in round 1 as well as against a pool of age matched control WT mice sera (Figure 6.3.1.), all in six replicates. The phage output titres were 10^5 to 10^6 CFU/ml. All replicates were analysed by NGS as described before.

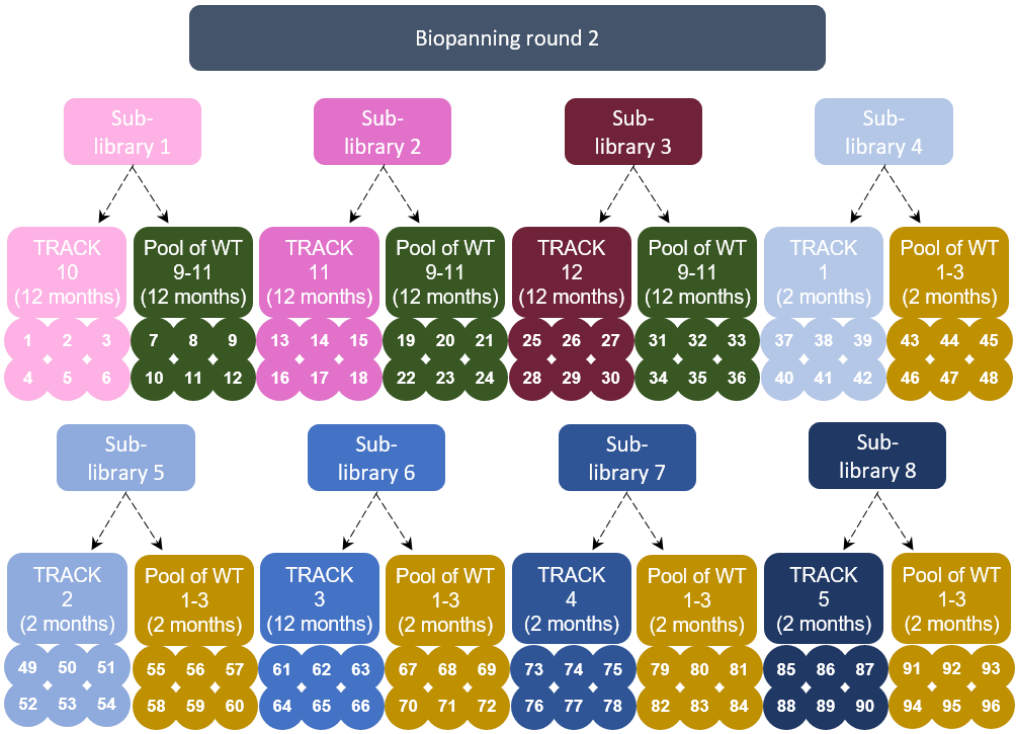


Figure 6.3.1 Overview of the biopanning strategy against TRACK and WT mouse sera of various ages for the second round of biopanning utilising sub-libraries derived from pooling round one output bacteriophages. Output bacteriophage population derived from six round 1 replicates were pooled together creating 8 different sub-libraries which were used for the second round of biopanning. Sub-libraries were also panned against a pool of sera of 12-month old (9-11) or 2-month old WT mice (1-3).

NGS data was analysed to identify potential diagnostic peptides that mimic TAAs and can be used in a downstream diagnostic development assay (Method 2.2.12). Analysis involved ranking peptide candidates using either their frequency or their Z scores. Firstly, 16 different NGS data sets were created that included the 50 most enriched peptide sequences per sera sample (but also 100 and 200 most frequent sequences were tested, too) (Figure 6.3.2.). Sets in which the same input phage sub-library from round 1 was used, were compared. A stringent sorting analysis, in which peptides were absent from their equivalent control sets (e.g. sequences enriched within set 1, but absent from set 2) was applied. 32 unique peptides sequences were identified that were present within at least two TRACK sets (Table 6.3.2.). The sum of their frequency was compared with their frequency within their equivalent WT sets (set 2, 4, 6, 8, 10, 12, 14, and 16) (Figure 6.3.3.). In some cases, the sum of these peptide frequencies exceeded 80-fold difference; all peptide frequencies with more than 2-fold difference were represented with yellow colour whereas the ones with range between 2- and 80-fold differences were represented with black colour (Figure 6.3.3.). A few of these peptides (e.g. 6, 7 and 19) were enriched within different TRACK sera and by more than 10-fold, demonstrating their diagnostic potential.

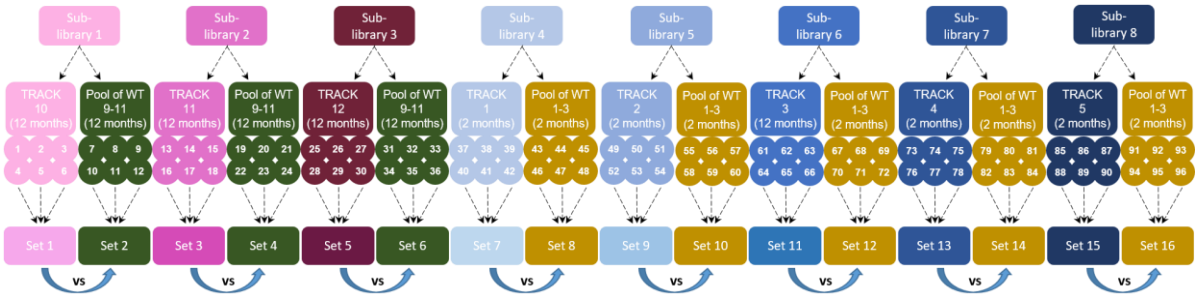


Figure 6.3.2 Overview of the analysis that was conducted using NGS data from the second round of biopanning by creating 16 different sets of the 50 most enriched peptide sequences within different sera samples.
16 different data sets contained the 50 most enriched sequences per sera sample and condensed. Comparison between sets that the same bacteriophage sub-library was used took place. Peptides were ranked in terms of how many times each peptide was enriched against TRACK sera (sets 1, 3, 5, 7, 9, 11, 13 and 15), but absent within WT sera sets (sets 2, 4, 6, 8, 10, 12, 14 and 16).

Table 6.3.1 List of the enriched peptides sequences that were selected based on their frequency within TRACK sera and their absence from the 50 most enriched sequences within WT data sets.
Data sets were derived from the second round of biopanning and 35 peptides were selected that were recognised by a minimum of 2 TRACK sera and absent from their equivalent WT mice control groups.

35 enriched peptide sequences derived from the NGS analysis of second biopanning round based on their frequency	
AMFqAAPHTGHDTDWN	SGMLPVKSDSYHTQIT
TSIPqTLKKPSKRGIR	RKQKGAIKLERNKRRT
SMLHMRTCAARKVACA	qTSHLFLYCRAKSSDC
RQTMLEYRLEYAARYT	QTIVLQPNSTSTLQRT
RPRRDPCPLLPRSPQT	PTHANPLEEVFANLST
PGGPPLRKSTPTVHET	PRAMTTTPNKTHEAVT
NTKHRDQRTNPTWTRY	PPRYVRTLATSTPPAN
FSPPLRSPTPLEFqqR	PKSTMRDKARDGSLFK
YqRASLMHAKSTKRQY	PFATQACDRRGSqKMS
VSAQKPLTRLPTQHSD	PASCYSSGMCPTGLSK
VRAIATIESSRPPLMS	NVFKARVQMFEALKSK
TPWTNRVWqRPRSTKS	NPYSSAACRAGIVTCQ
TPSSqTLLATRVSKVH	LARHPGDMWEDLQLSG
TCWLSRLVGDPRANVT	HRPPGDGHRRGEQHER
SSDAWAQTDPGLVPYC	HQLAPQISSGLDPAVV
SQPMPRQPVAPTqHAL	GYRRRQGLPRRQLVRK
SPCALSGMCSFLQTHS	ALALLERAKLLKAAHK
SGMRPYSKTPPTTVPA	

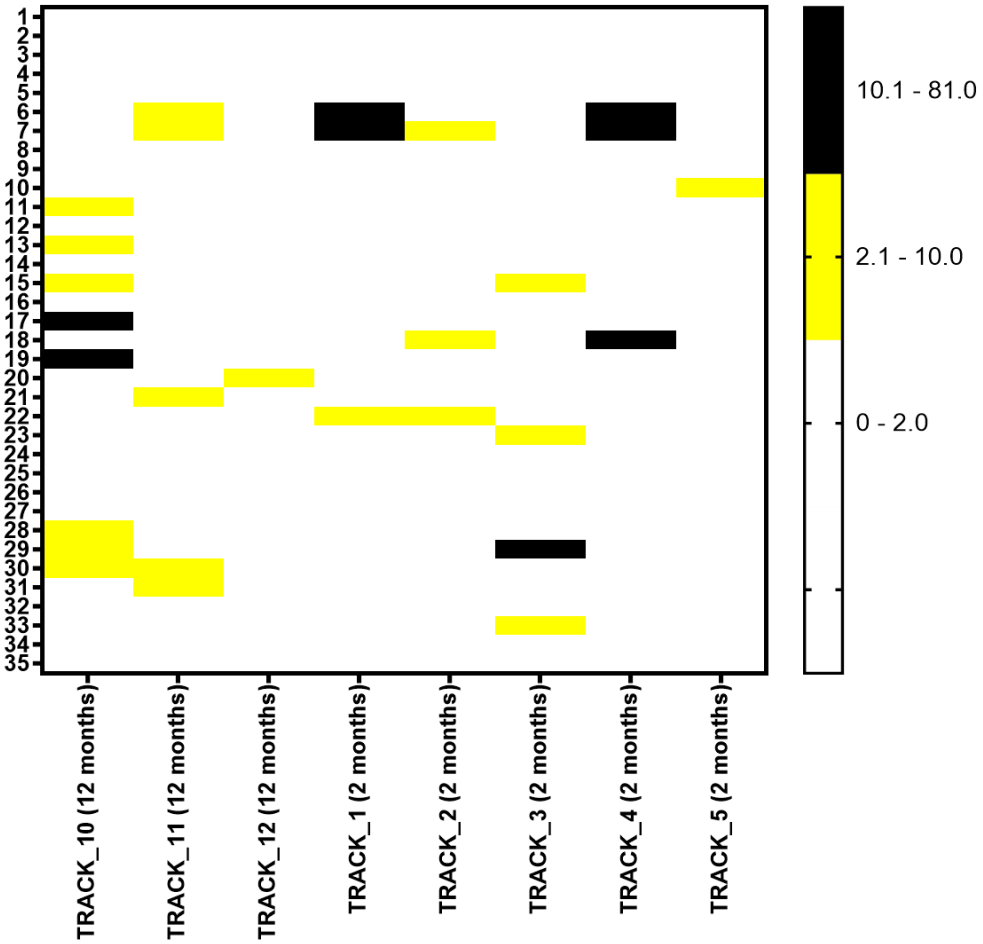


Figure 6.3.3 Comparison of the frequency of 35 TRACK specific peptides within different TRACK sera.
35 peptide sequences were identified within the 50 most enriched sequences within TRACK sets (1, 3, 5, 7, 9, 11, 13 and 15) and absent from the 50 most enriched sequences within their WT negative sets, in which the same input sub-library phage was used (sets 2, 4, 6, 8, 10, 12, 14, 16). These peptides were present within ≥2 TRACK sets. The frequency of all these peptides was summed per replicate and averaged per sera sample. This was then compared with the sum of these peptides frequency within their equivalent negative set (e.g. sum of set 1 vs sum of set 2) and expressed as a ratio.

Additionally, a two way Z score analysis was also conducted in order to expand the repertoire of potential mimotope peptides. A Z score cut off of 4.0 was used to define peptide enrichment between replicate samples in which the same phage-sublibraries were used. Fifteen peptide sequences were identified

with Z score ≥4, and present within at least two TRACK replicates (Table 6.3.3.). A loose motif (SGMxP) was present in some of the peptide sequences. The sum of their frequency within TRACK sera was compared with their frequency within WT sera pool, with the same sub-library. A cut-off value of 2-fold difference was defined in order to validate these peptides potential use as diagnostics. Five positive TRACK samples (fold difference ≥2) were thus identified (Figure 6.3.4.). This data selection was promising towards the development of a common diagnostic, in which peptides were not sera specific, but instead they were TRACK specific. Finally, the frequency of peptides derived from both the analytical methods were combined but no positive TRACK samples were identified (data not shown).

Table 6.3.2 List of the amino acid sequences of the enriched peptides with Z score ≥4 and enriched within at least 2 out of 8 TRACK sera samples. Fifteen peptides were selected with Z score higher than 4 and present within at least 20% of TRACK mice replicates.

15 enriched peptide sequences Z score analysis
SMLHMRTCAARKVACA
LNQLKAGPRPSIPPIG
NVLLHRYSVYDqTRSH
RQTMLEYRLEYAARYT
CHRTTAMASGLKPYLC
HQLAPQISSGLDPAVV
LNMLAYRVQQATGLRA
NPYSSAACRAGIVTCQ
NVFKARVQMFEALKSK
PASCYSSGMCPTGLSK
SGMLPVKSDSYHTQIT
SGMRPYSKTPPTTVPA
SPCALSGMCSFLQTHS
TVLQMRRHYAGITDSP
VPTEWLHTASDNPMRN

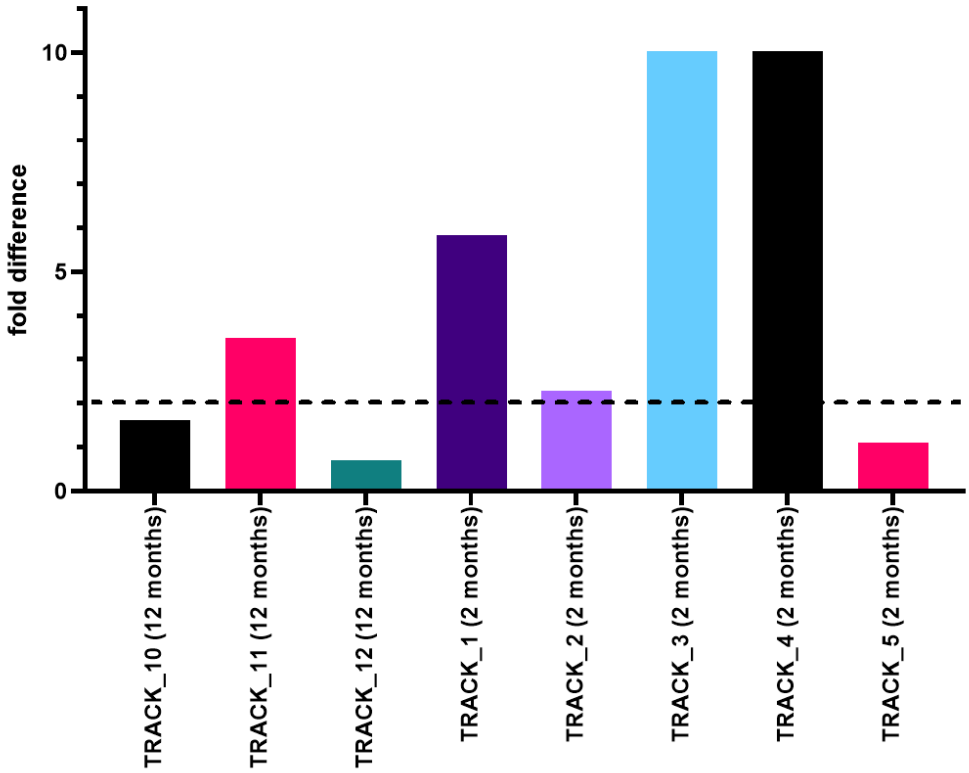


Figure 6.3.4 Comparison of the frequency sum of 16 TRACK specific peptides with Z score ≥4. 16 peptide sequences were identified within at least two TRACK sets and with Z score ≥4.
The frequency of all these peptides was calculated within the different samples, averaged per replicates (six in total) and the sum of all the peptides frequency was calculated as fold difference with the sum of peptide frequency within their equivalent negative set (e.g. sum of set 1 vs sum of set 2).

6.4. ELISA summary utilising the potential diagnostic peptides derived from the second round of TRACK biopanning

The second biopanning round was conducted against TRACK sera (n=8) and a pool of WT sera, depending on their age (2- or 12-month old). Two different approaches for analysis the NGS data were conducted involving ranking peptide sequences using their Z score (≥4) or their actual frequency within different data sets (present within the most 50 enriched sequences against TRACK but absent from the 50 most enriched sequences against WT sera). In

total 41 unique peptide sequences were identified (Table 6.4.1.). These peptides were evaluated for their potential as diagnostic candidates by conventional ELISA, either as phage fusions or synthetic peptides. Firstly, these peptides were expressed as pVIII fusions by using specific inverse PCR primers and cloning the corresponding peptide sequences into phagemid (Method 2.2.6). Inverse PCR was followed by gel electrophoresis, confirming almost all samples were amplified (Figure 6.4.1.). A few of them failed to be recovered (e.g. samples 32 and 49), even after PCR optimisation steps. In total, 83% of these peptides were successfully recovered. Bacteriophages displaying these peptides were tested in an ELISA format against a pool of TRACK and WT sera as an initial screening of these TRACK specific peptides. For each peptide an O/N OD reading was taken and a fold difference was calculated between its TRACK and WT pool sera signal. Four to six transformants for each clone were used to make bacteriophage and screened. This was before their sequence was confirmed by Sanger sequencing, therefore their averaged OD value did not reflect the actual OD of the Sanger confirmed clones. Binding was defined by applying the cut-off value as a 1.5-fold difference between the TRACK and WT sera signals. An unrelated peptide (SAF70 mimotope), expressed as a pVIII fusion was included as a negative control. Sensitivity and specificity determination took into account, that some sera samples were tested twice in different age. Firstly, peptides 11-16 were assessed (Table 6.4.1.) and peptide 15 was considered positive because its fold difference value was above the cut-off (1.5-fold) (Figure 6.4.2.). Additionally, peptides 17-26 were assessed (Table 6.4.1.) and peptide 23 had a positive signal above the cut-off value. One out of

six bacteriophages displaying peptides 20 and 24 were positively recognised by TRACK sera, therefore even though their fold difference was not above the cut off, they were also considered positive (Figure 6.4.3.). Finally, peptides 27-48 were screened (Table 6.4.1.) and none of them was identified as positive (Figure 6.4.4.). Notably, peptide 15 was included in the last two assays as an internal positive control. All of the clones were later Sanger sequencing confirmed, and they correlated with the ELISA results. Monoclonal bacteriophages displaying peptides 15, 20, 23 and 24 were isolated and retested against individual TRACK and WT sera, in duplicates in order to investigate their diagnostic potential. Sera were also tested against monoclonal bacteriophages displaying a SAF70 mimotope peptide, as a negative control. A cut-off value was used in order to define the positive samples which was 2-fold difference of the overnight OD measurement when it was compared with the negative peptide control OD measurement. These assays were repeated 3 times, and each time the results were comparable. The sensitivity of the ELISA assays when bacteriophages displaying peptides 15, 20, 23 and 24 or a mixture of them was 62.5%, 81.3%, 75%, 62.5% and 81.3%, respectively whereas the specificity of the assay was determined at 66.7%, 50%, 66.7%, 75% and 66.7%, respectively (Figure 6.4.5.). The highest sensitivity was marked when peptide 20 and the combination of them were used, whereas peptides with the highest specificity were peptide 24, followed by peptides 15, 23 and the combination of them. One of these peptides (peptide 15) was also produced in a biotinylated format in an attempt to increase the sensitivity and specificity of the assay as using streptavidin plates

(demonstrated in the previous chapter) was a far more sensitive ELISA approach. Therefore, biotinylated peptide 15 was tested against individual TRACK and WT sera; the sensitivity of the assay was decreased at 50% from 62.5%) but its specificity was improved at 83.3% (from 66.7%; Figure 6.4.6.). Furthermore, synthetic peptides 11-26 (Table 6.4.1.) were produced and they were tested against TRACK and WT pools of sera, in duplicate. The cut off for the assay was consistently assessed as 2-fold difference above the WT pool of sera OD. Peptides 17 and 23 (the latter had previously also been identified as being positive for TRACK sera as a phage fusion) were identified as positives (Figure 6.4.7.). Therefore, both of these synthetic peptides were tested against individual TRACK and WT sera in duplicates. The sensitivity and the specificity of the assay were 62.5% and 91.6%, respectively when peptide 17 was used (Figure 6.4.8.) and 18.8% and 100%, respectively when peptide 23 was used (Figure 6.4.9.).

Table 6.4.1 List of enriched peptide sequences derived from either frequency or Z score analysis of the second round's biopanning NGS data. These peptides were present within at least 2 TRACK sets and with Z score ≥4 or they were present within the 50 most enriched sequences of the TRACK sera but absent from their age-matched WT sera.

AA sequence	recovery ID	AA sequence	recovery ID
Z score analysis		**frequency analysis**	
LNQLKAGPRPSIPPIG	peptide_11	AMFqAAPHTGHDTDWN	peptide_27
NVFKARVQMFEALKSK	peptide_12	FSPPLRSPTPLEFqqR	peptide_28
NVLLHRYSVYDqTRSH	peptide_13	NTKHRDQRTNPTWTRY	peptide_29
RQTMLEYRLEYAARYT	peptide_14	PGGPPLRKSTPTVHET	peptide_30
SGMLPVKSDSYHTQIT	peptide_15	RPRRDPCPLLPRSPQT	peptide_31
SMLHMRTCAARKVACA	peptide_16	TSIPqTLKKPSKRGIR	peptide_32
CHRTTAMASGLKPYLC	peptide_17	ALALLERAKLLKAAHK	peptide_33
HQLAPQISSGLDPAVV	peptide_18	GYRRRQGLPRRQLVRK	peptide_34
LNMLAYRVQQATGLRA	peptide_19	HRPPGDGHRRGEQHER	peptide_35
NPYSSAACRAGIVTCQ	peptide_20	LARHPGDMWEDLQLSG	peptide_36
PASCYSSGMCPTGLSK	peptide_22	VRAIATIESSRPPLMS	peptide_37
SGMRPYSKTPPTTVPA	peptide_23	VSAQKPLTRLPTQHSD	peptide_38
SPCALSGMCSFLQTHS	peptide_24	YqRASLMHAKSTKRQY	peptide_39
TVLQMRRHYAGITDSP	peptide_25	PFATQACDRRGSqKMS	peptide_40
VPTEWLHTASDNPMRN	peptide_26	PKSTMRDKARDGSLFK	peptide_41
		PPRYVRTLATSTPPAN	peptide_42
		PRAMTTTPNKTHEAVT	peptide_43
		PTHANPLEEVFANLST	peptide_44
		QTIVLQPNSTSTLQRT	peptide_45
		qTSHLFLYCRAKSSDC	peptide_46
		RKQKGAIKLERNKRRT	peptide_47
		SQPMPRQPVAPTqHAL	peptide_48
		SSDAWAQTDPGLVPYC	peptide_49
		TCWLSRLVGDPRANVT	peptide_50
		TPSSqTLLATRVSKVH	peptide_51
		TPWTNRVWqRPRSTKS	peptide_52

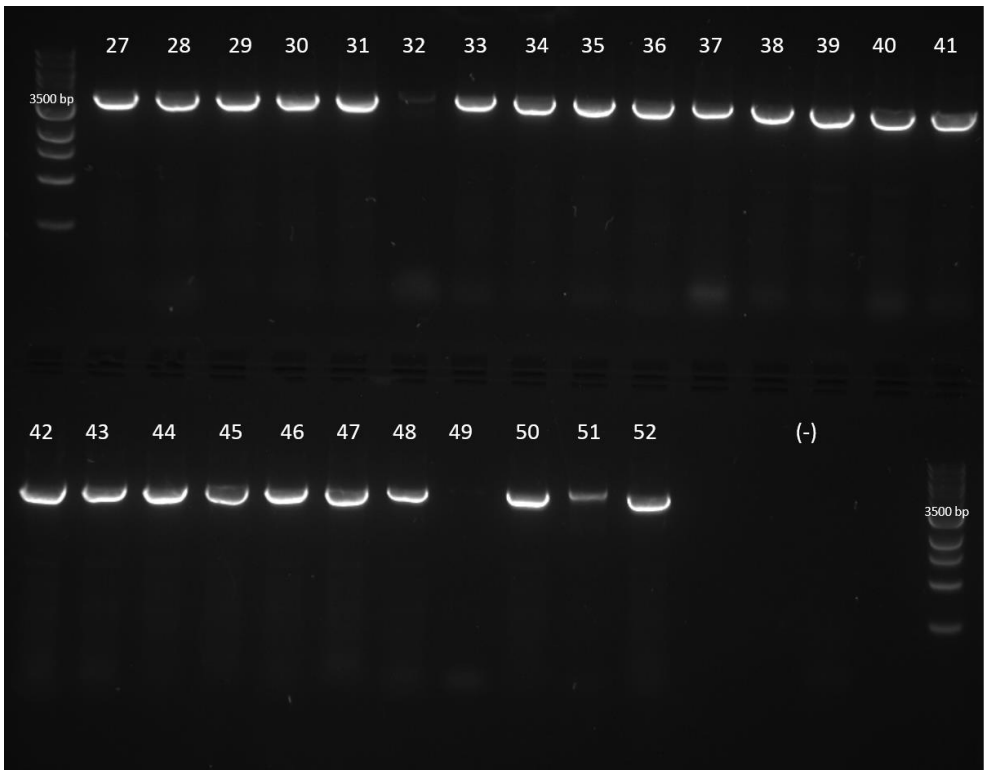


Figure 6.4.1 Amplification of target phage clones using inverse PCR.
PCR amplicons were found at the expected size (~3500 bp) for most of the targeted sequences; sample containing no template (-) was included as a control for possible unspecific amplification. Clones that failed to amplify were repeated with optimised PCR conditions but could not be recovered.

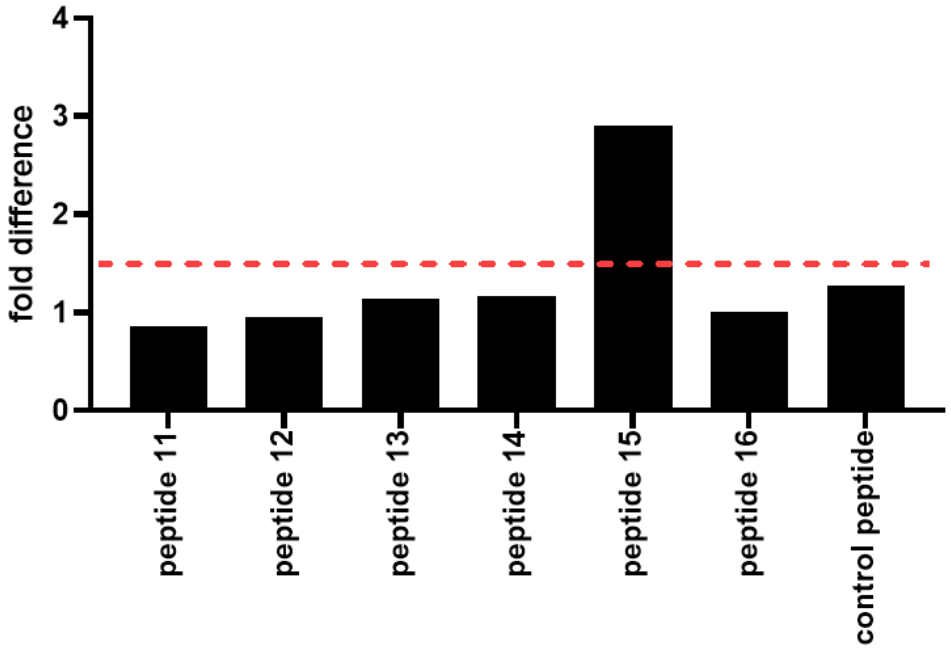


Figure 6.4.2 Assessment of the diagnostic potential of the enriched peptides 11-16 as pVIII fusions in a phage ELISA assay format against a pool of TRACK and WT sera.
Peptides 11-16 (Table 6.4.1.) were tested against pool of TRACK and WT sera, in six replicates. Fold difference of their overnight OD value was calculated between TRACK and WT sera. Notably, peptide insertion was not yet confirmed within all the bacteriophages, therefore some of them could potentially not be displaying the appropriate peptide. Cut off of the assay was determined 1.5- fold difference (dotted line). Peptide 15 seemed to positively identify TRACK over WT sera. Bacteriophage displaying a control peptide (previously identified as a SAF70 mAb mimotope) was included as a negative control.

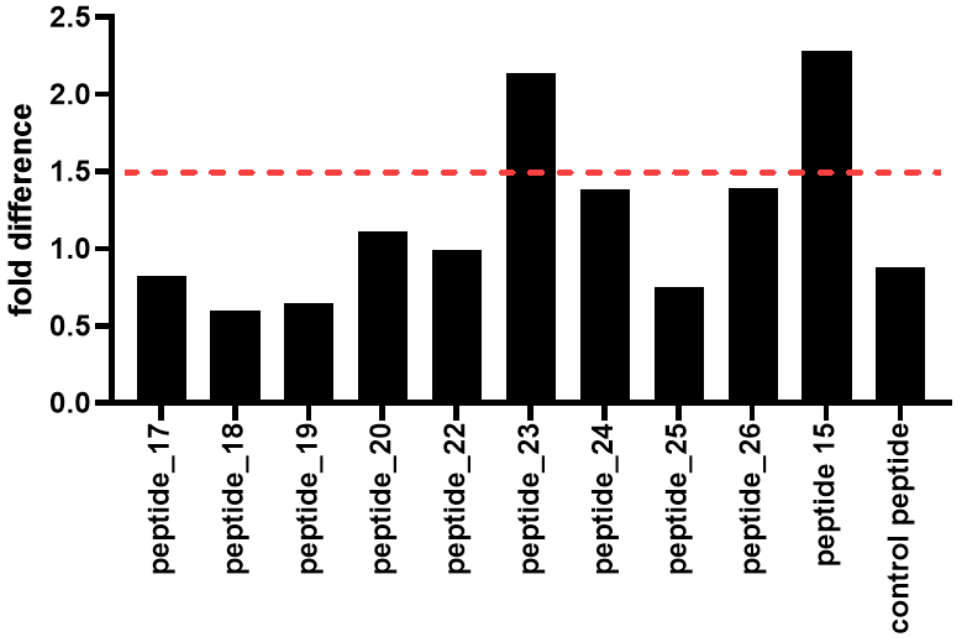


Figure 6.4.3 Assessment of the diagnostic potential of the enriched peptides 17-26 as pVIII fusions in a phage ELISA assay format against a pool of TRACK and WT sera.

Peptides 17-26 (Table 6.4.1.) were tested against pool of TRACK and WT sera, respectively in six replicates. Fold difference of their overnight OD was calculated between TRACK and WT sera. Notably, peptide insertion was not yet confirmed within all the bacteriophages, therefore some of them could potentially not be displaying the appropriate peptide. Cut off of the assay was determined 1.5-fold difference (dotted line). Peptide 23 seemed to positively identify TRACK over WT sera. Notably, individual values of bacteriophages displaying peptide 20 and 23 were also positive for TRACK over WT sera, but was not reflected in this averaged data. Bacteriophage displaying a control peptide (previously identified as a SAF70 mAb mimotope) was included as a negative control and previously identified peptide 15 was included as a positive control.

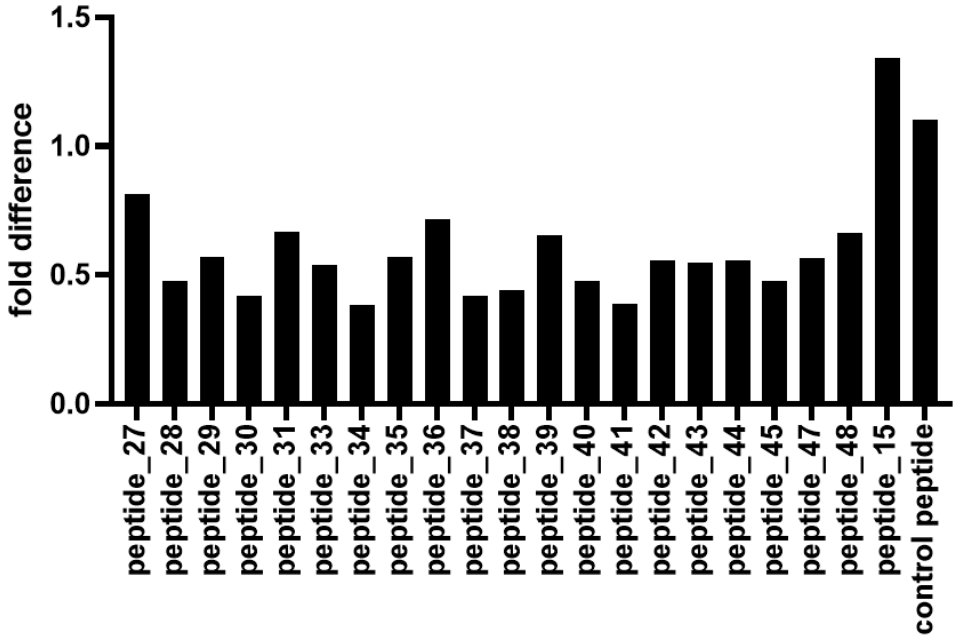


Figure 6.4.4 Assessment of the diagnostic potential of the enriched peptides 27-48 as pVIII fusions in a phage ELISA assay format against a pool of TRACK and WT sera.

Peptides (Table 6.4.1.) were tested against pools of TRACK and WT sera, respectively in four replicates. Fold-difference of their overnight OD was calculated between TRACK and WT sera. Notably, peptide insertion was not yet confirmed within all the bacteriophages, therefore some of them could potentially not be displaying the appropriate peptide. Cut off of the assay was determined 1.5-fold difference. Bacteriophage displaying a control peptide (previously identified as a SAF70 mAb mimotope) was included as a negative control; previously identified peptide 15 was also included as a positive control. None of these peptides were positive for TRACK over WT sera.

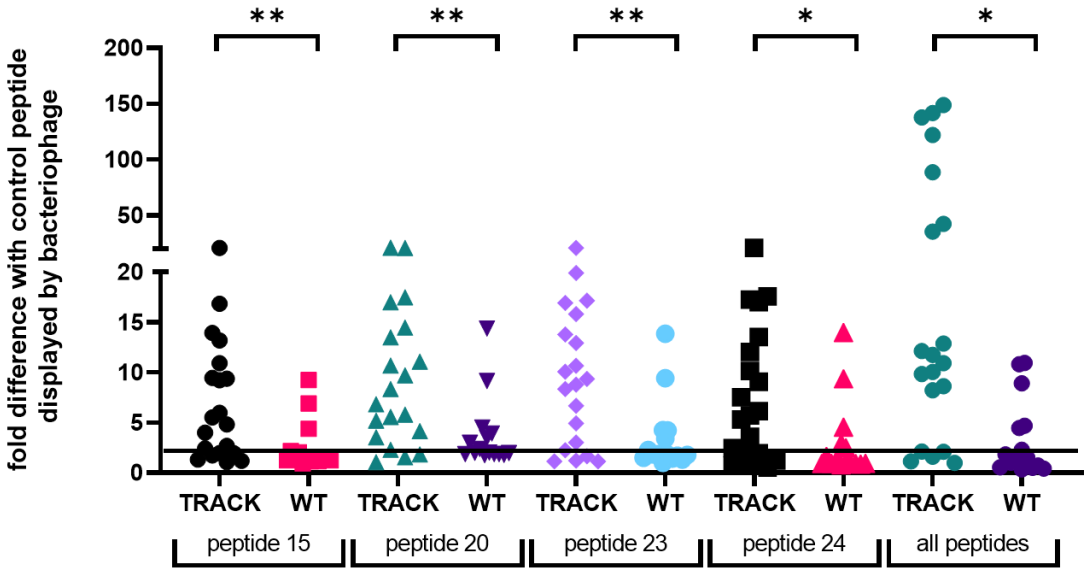


Figure 6.4.5 Assessment of the diagnostic potential of the enriched peptides 15, 20, 23, 24 and their combination as pVIII fusions in a phage ELISA assay format against individual TRACK and WT sera.
These peptides were previously identified to have a positive signal in TRACK over WT pool of sera. Individual TRACK (n= 20) or WT (n=15) of all ages (Table 6.2.1) were tested in duplicates. Overnight sera OD measurements were calculated as fold difference with the OD when negative control phage-peptide was used. Binding was defined by applying the cut-off value as a 2-fold difference. Sensitivity was determined as 62.5%, 81.3%, 75%, 62.5% and 81.3%, whereas specificity was determined as 66.7%, 50%, 66.7%, 75% and 66.7% when peptide 15, 20, 23, 24 or a combination of them were used, respectively.

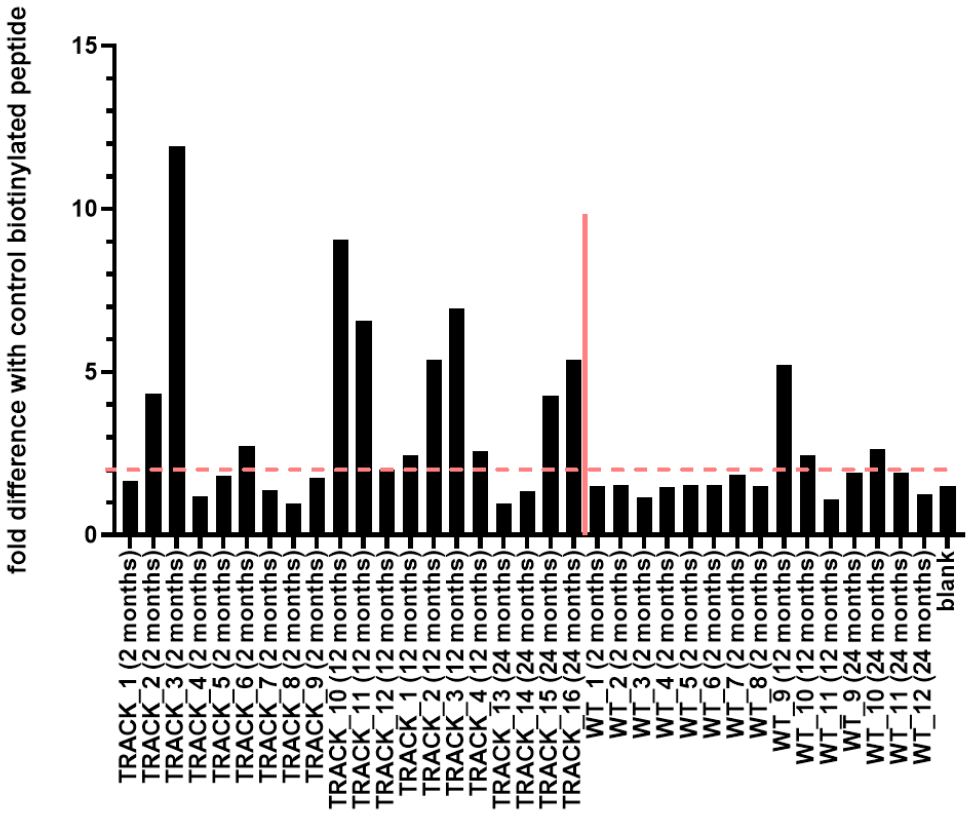


Figure 6.4.6 Assessment of the diagnostic value of biotinylated peptide 15 against individual TRACK and WT sera.

To evaluate the diagnostic potential of this peptide, it was assayed against TRACK and WT sera, in duplicate. Fold-difference of for the same sera between the overnight OD values of the peptides and a control peptide was conducted. Binding was defined by applying the cut-off value as a 2-fold difference. The sensitivity of the assay was determined as 50% whereas its specificity was at 83.3%.

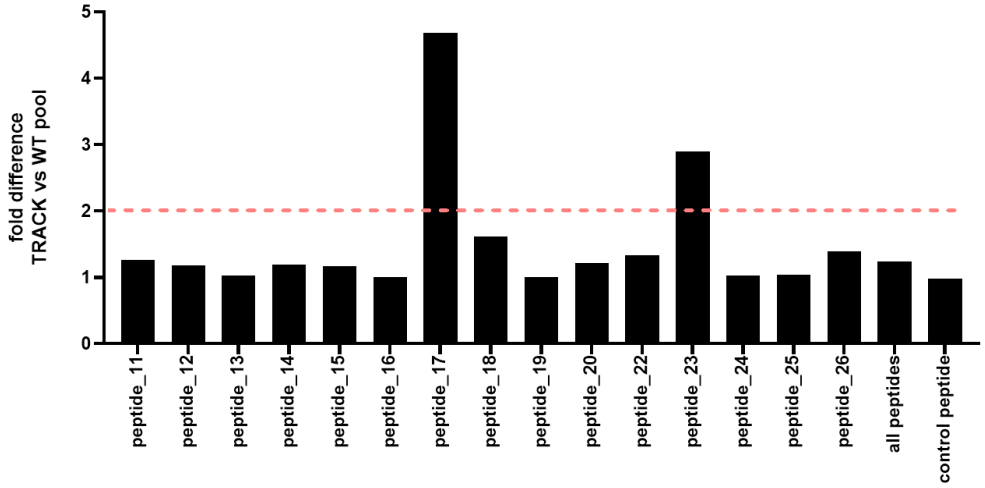


Figure 6.4.7 Assessment of the diagnostic value of synthetic peptides 11-26 (no modification) against pools of TRACK and WT sera.
Peptides 11-26 (Table 6.4.1.) were tested against pools of TRACK and WT sera, in duplicate. Fold difference of signals for the same sera between the overnight OD values of the peptides and a control peptide was conducted. Binding was defined by applying the cut-off value as a 2-fold difference. Peptides 17 and 23 were positively identified in TRACK compared with WT pools of sera.

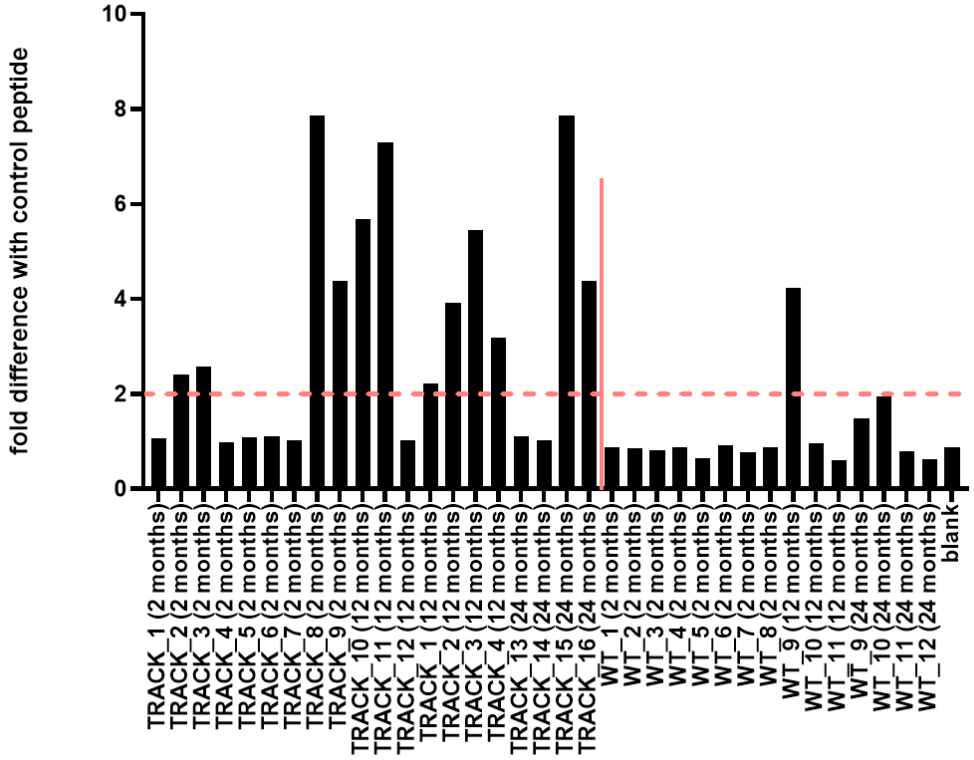


Figure 6.4.8 Assessment of the diagnostic value of synthetic peptide 17 (no modification) against individual TRACK and WT sera.
Fold difference of the same sera between the overnight OD values of the peptides and a control peptide was conducted. Binding was defined by applying the cut-off value as a 2-fold difference. The sensitivity of the assay was determined as 62.5% whereas its specificity was at 92.6%.

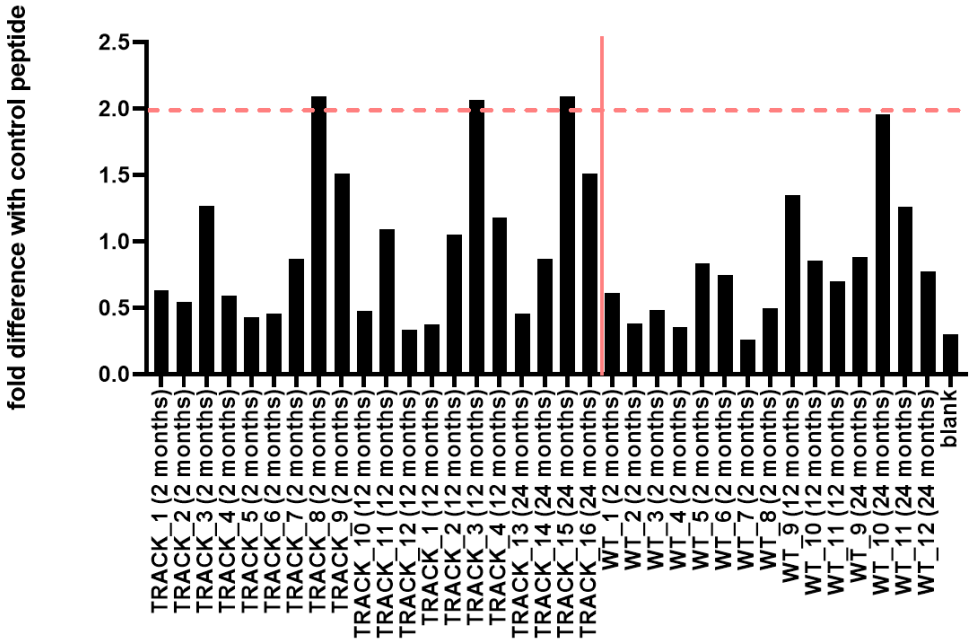


Figure 6.4.9 Assessment of the diagnostic value of synthetic peptide 23 (no modification) against individual TRACK and WT sera.
Fold difference of the signals with same sera between the overnight OD values of the peptides and a control peptide was conducted. Binding was defined by applying the cut-off value as a 2-fold difference. The sensitivity of the assay was determined as 18.8% whereas its specificity was at 100%.

The peptides that were positively identified by TRACK sera had a distinctive consensus motif, as identified by the online motif finder platform MEME (Bailey and Gribskov, 1998). The motif that was common between peptide_15 (SGMLPVKSDSYHTQIT), peptide_17 (CHRTTAMASGLKPYLC), peptide 20 (NPYSSAACRAGIVTCQ), peptide_23 (SGMRPYSKTPPTTVPA) and peptide_24 (SPCALSGMCSFLQTHS) was SGMxPY, and when it was blasted in the online platform of the Immune Epitope Database (IEDB) there was not any protein epitope confirmed (Marcatili et al., 2018). Additionally this motif is not present in the HIF-α protein that was originally mutated for the mouse model TRACK either. The motif could represent a discontinuous epitope, thus the protein could be challenging to identify.

6.5. Alternative round 2 biopanning strategy utilising sublibraries against immunoreactive sera (immunosignature)

An alternative second round biopanning strategy was designed in order to expand the repertoire of renal cell carcinoma mimotopes and potentially utilise them in a NGS-based diagnostic assay. This *in silico* approach, also known as immunosignature, was used to map the autoimmune response against cc-RCC in TRACK mice and in WT mice. TRACK sera that showed the highest immunoreactivity based on their R2 Z score of the selected TRACK-specific peptides from round 1 of biopanning were selected (Figure 6.3.4). A common pool of their output phage populations was created. Therefore, these carefully selected sub-libraries 2, 4 and 7 were pooled and used as a common reagent for this round 2 biopanning, in an attempt to increase the assay's sensitivity and specificity. This common sub-library pool was biopanned against a training and a testing cohort in order to evaluate its diagnostic capacity. The training cohort consisted of sera from TRACK (n=8) and WT (n=6) that were previously used for round 1 and 2 panning. The testing cohort consisted of sera from TRACK and BPS mice, a different version of the TRACK model, (n=18) and WT (n=11). Therefore, the training set was a group of individual sera from three 12-month old TRACK, five 2-month old TRACK, three 12-month old WT and three 2-month old WT mice whereas the testing set was made of four 2-month old TRACK, four 14-month old TRACK, four 23-month old TRACK, three 12-month old TRACK, three 12-month old BPS, five 2-month old WT, two 8-month old WT and four 24-month old WT mice, all tested in duplicates (Figure 6.5.1.). Importantly, during the study it was revealed that

75% of the 24-month old WT mice sera in the testing set had been previously included in the training set when they were 12-months old and 75% of the 14-month old TRACK mice sera in the testing set had been previously included in the training set when they were 2-months old (Table 6.2.1.). Two of those had also been used to create 2/3 of the immunosignature reagent (sub-libraries 4 and 7). This was taken into account for all further statistical analyses, meaning that even though the total number of TRACK sera was 26, the number of individual TRACK mice was 22, whereas the total number of WT sera samples was 17 but the individual WT mice sera was 13 (Table 6.2.1). Firstly, the input bacteriophage titration was estimated at 4x 10^{11} CFU/ml. The sequential subtraction steps using a pool of WT mice sera (from training set) was conducted before the biopanning. Phage output titres were 2 - 5 10^4 CFU/ml (12 samples were randomly titrated). All these replicates were analysed by NGS (n=86) as described in Method 2.2.12. The aim of this study was to apply different bioinformatics strategies to analyse enriched peptide sequences in the training cohort, identify TRACK-specific ones (the RCC immunosignature) and then by applying the same criteria in a testing cohort to produce a diagnostic assay with high specificity and sensitivity. Firstly, all the enriched sequences within the 12-month old WT mice sera replicates were pooled together (set 1). A Z score analysis was conducted between six 12-month old TRACK replicates (Figure 6.5.2A.) and set 1. A cut-off value of Z score ≥5 was used and peptides were selected that were enriched within minimum 30% of the 12-month old TRACK samples replicates (n = 19, Table 6.5.1). The frequency of each peptide was calculated for every replicate and averaged per sera

sample (within training and testing cohorts). The averaged sum of the 12-month old TRACK specific peptides was plotted, to investigate the potential of this immunosignature as diagnostics. The cut off for this assay was determined as 2x the 12-month old WT averaged sum of these peptides frequency in the training set. The sensitivity of the assay was estimated at 59%, and the specificity at 50% (Figure 6.5.2B.).

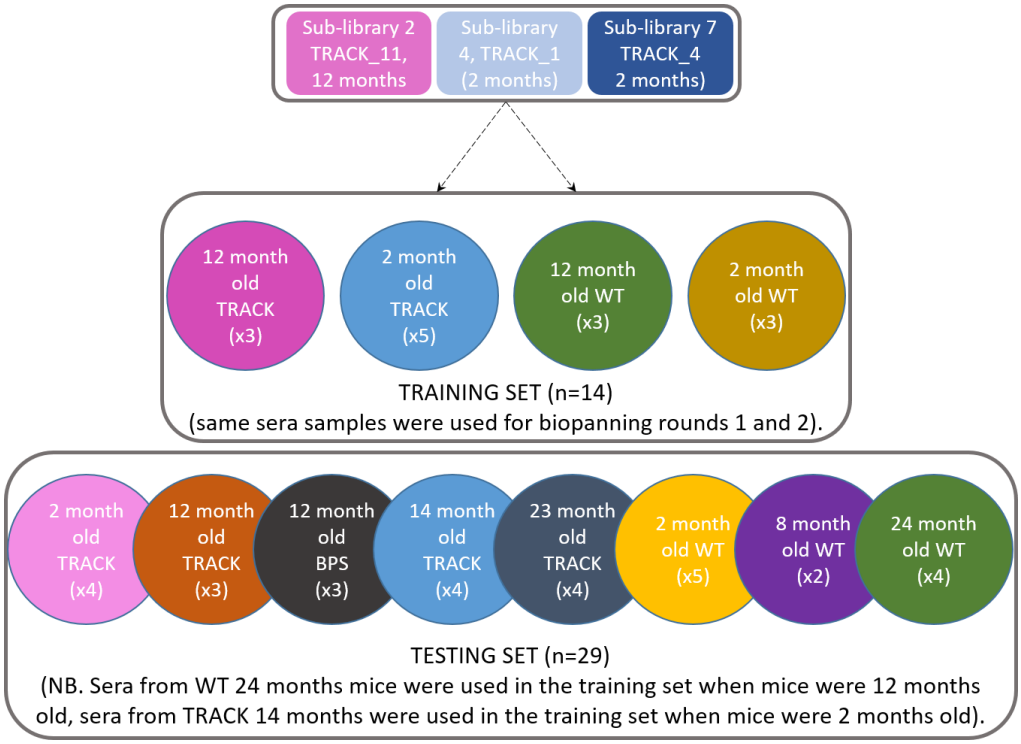


Figure 6.5.1 Overview of the biopanning strategy against TRACK and WT mouse sera of various ages for an immunosignature panning strategy. Three different sub-libraries of output bacteriophage derived from the first biopanning round were used as a common reagent for the immunosignature biopanning. The training set consisted of three 12-month old TRACK mice, five 2-month old TRACK mice, three 12-month old WT mice and three 2-month old WT mice. Additionally, the testing set consisted of four 2-month old TRACK mice, four 14-month old TRACK mice, four 23-month old TRACK mice, three 12-month old TRACK mice, three 12-month old BPS mice, five 2-month old WT mice, two 8-month old WT mice and four 2-month old WT mice. All sera samples were tested in duplicate. Notably, there are sera from 14-month old TRACK mice and 24-month old WT that were used in the training set when mice were 2- and 12--months old, respectively.

Table 6.5.1 List of translated amino acid sequences from the enriched peptides resulted from the Z score NGS analysis of this biopanning round. These peptides were present within at least 30% of the TRACK replicates and with Z score ≥5. Common peptide sequences between the two aged groups were highlighted.

12-month old TRACK specific peptide sequences (n=19)	2-month old TRACK specific peptide sequences (n=34)	
LNQLKAGPRPSIPPIG	QKRSLAMQNSGMVSRT	THRWHCTSRKTATRAD
KHNVLAARVHKYNTDR	QLTPRISSGLAPAVNT	SPCALSGMCSFLQTHS
TVLQMRRHYAGITDSP*	TLPPWAQEYGLDDTPI	SLTAGPRPTAAIQLYH
RQTMLEYRLEYAARYT	SRSMDPNKQTMQQTSV	SGMLPVSTNDWPLTKS
PNNLYNARLTYARERQ	SGMLPVKSDSYHTQIT	RNPHPIATLDPPVKQI
PASCYSSGMCPTGLSK*	qLLSPNKSHCWHGPHP	RHAALVSAARLTVNPS
NVLLWRLKSHFTTEQT	**NPYSSAACRAGIVTCQ***	PSSVRVHTNVEKTDPI
NVLLHRYSVYDqTRSH	NPKVASGIMPPTGSEP	**PASCYSSGMCPTGLSK***
NDPVTMYSEPCRPDLQ	NHIVVRTLDDSYPAYR	**PAEILSGMRPTNSPDA***
LNMLAYRVQQATGLRA	ITPSVRSGMRPASSAT	NKHVSQDPNKPNFVEY
TYRTLDTLPEPKPQAN	YAGPRPQAWRPATNTM	LRLAGPRSISMHAPYI
TVLQLRIMQSRDLEAG	**TVLQMRRHYAGITDSP***	LLTVHPTLDSAPPGHT
PAEILSGMRPTNSPDA*	TTHAAIFSGMASHQKY	LIPTSVRPLKHALTTP
NVMELRHHYYYSRTTV	TQPIISSGISPHTPAN	LAQHRAGMAPMPPELT
NVLHSRYRAWSQRSSE	TQMAGPRPPIKDLLPA	IAEVSAGPYMqSLPPT
NPYSSAACRAGIVTCQ*	TMQRVYTLHKAVRLRR	HQLAPQISSGLDPAVV
NFVPPAWLPRESNELR	TITRLNLEKK	CHRTTAMASGLKPYLC
NASRLNSGMSVQVPR		
MNNVLNARIRHFHALA		
***common between the different age groups**		

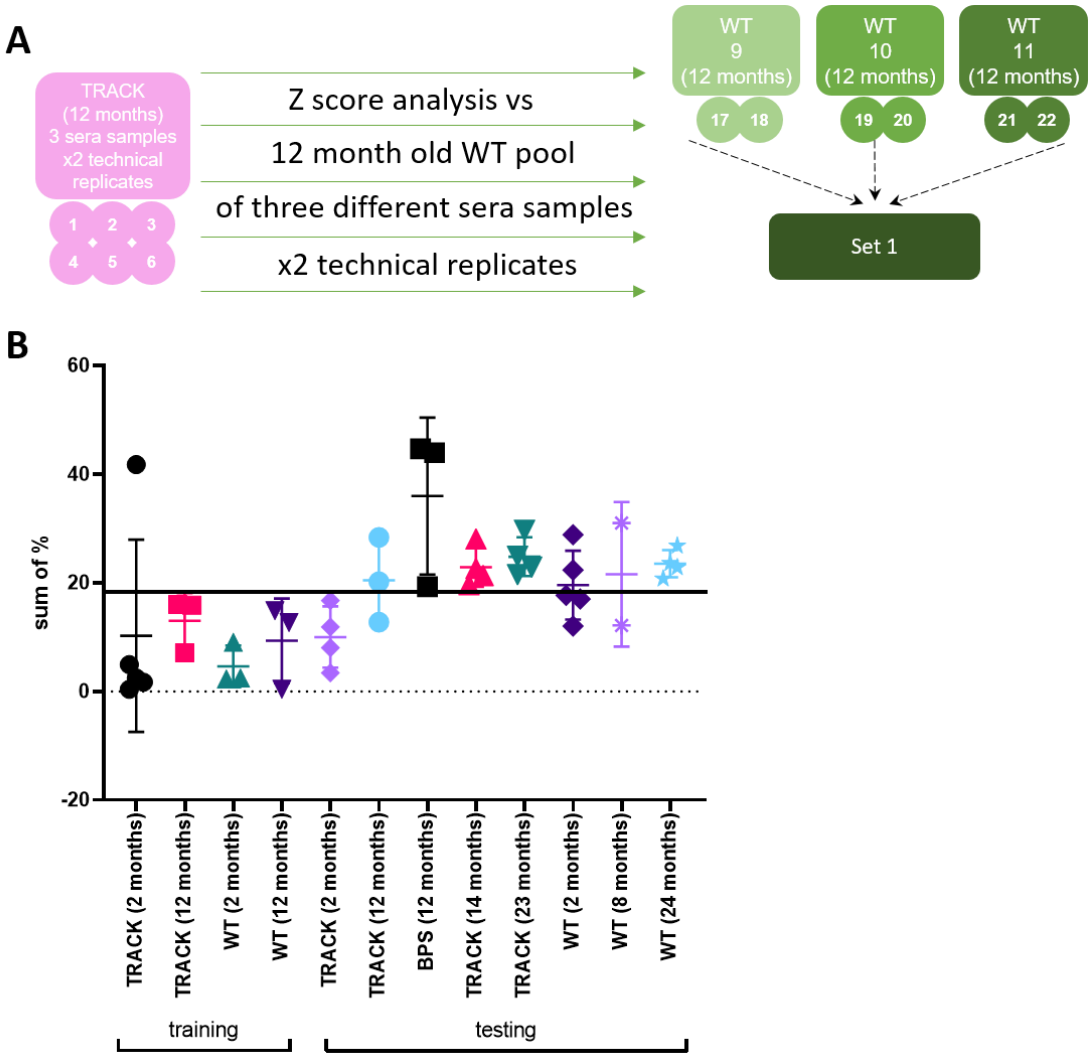


Figure 6.5.2 Assessment of the Z score analysis between 12-month old TRACK replicates and a pool of enriched sequences within 12-month old WT replicates (set 1) and the sum of frequency of the selected peptides within every sera sample.
A. Set12 was a pool of all the enriched sequences of 12-month old WT replicates. A Z score analysis was conducted between six 2-month old TRACK replicates and set 1. Nineteen unique peptide sequences with Z score ≥ 5, and enriched within at least 30% of the TRACK replicates were identified. **B.** The frequency of these 12-month TRACK specific peptides was averaged and summed per every sera sample within training and testing cohorts. The cut-off value was determined as 2x the averaged sum of these 19 peptides frequencies within 12-month old WT replicates (training set). The sensitivity of this assay was determined at 59% and the specificity at 50%.

Similarly, all the enriched sequences within the 2-month old WT mice sera replicates were pooled together (set 2). A Z score analysis was conducted

between ten 2-month old TRACK replicates and set 2 (Figure 6.5.3A.). A cut-off value of Z score ≥5 was defined and peptides that were enriched within minimum 30% of the 2-month old TRACK samples, were selected (n=34, Table 6.5.1). Some sequences were common between the age groups (bold, Table 6.5.1.). The frequency of each peptide was calculated for every replicate and averaged per sera sample (within training and testing cohorts). The averaged sum of the 2-month old TRACK specific peptides was plotted, to investigate the potential of this immunosignature as a diagnostic. The cut off of this assay was determined as 20x (as it generated the best specificity) the 2-month old WT averaged sum of these peptides frequency in the training set. The sensitivity of the assay was estimated at 40.9% and the specificity at 78.6% (Figure 6.5.3B.).

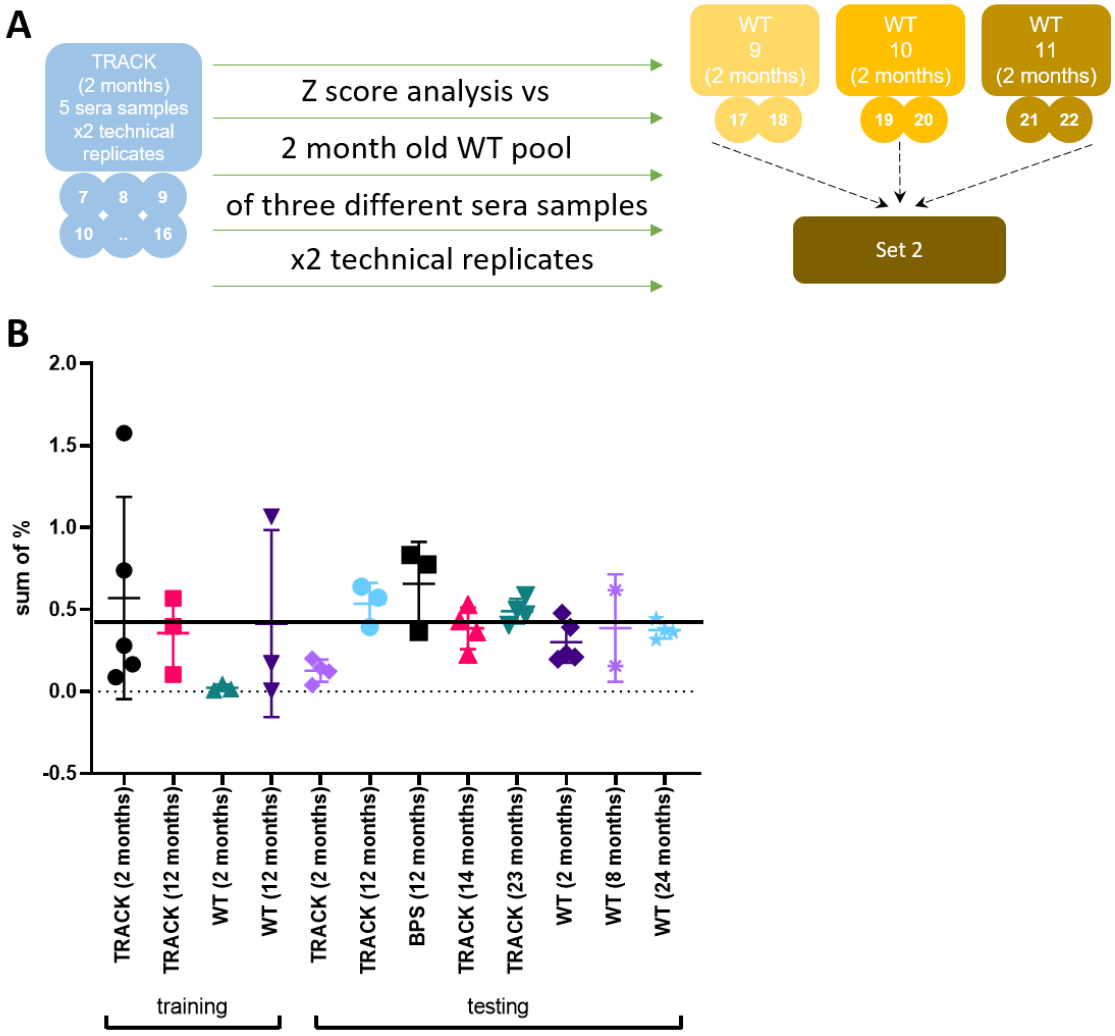


Figure 6.5.3 Assessment of the Z score analysis between 2-month old TRACK replicates and a pool of enriched sequences within 2-month old WT replicates (set 2) and the sum of frequency of the selected peptides within every sera sample.

A. Set 2 was a pool of all the enriched sequences of 2-month old WT replicates. A Z score analysis was conducted between ten 2-month old TRACK replicates and set 2. Thirty four unique peptide sequences with Z score ≥ 5, and enriched within at least 30% of the TRACK replicates were identified. **B.** The frequency of these 2-month TRACK specific peptides was averaged and summed per every sera sample within training and testing cohorts. The cut-off value was determined as 20x the averaged sum of these 34 peptides frequencies within 2-month old WT replicates (training set). The sensitivity of this assay was determined at 40.9% and the specificity at 78.6%.

A frequency analysis was conducted in parallel in an attempt to improve the selection of the TRACK specific peptides (immunosignature). Firstly, two sets

were created containing the 200 most enriched peptide sequences per WT age group (training cohort) (Figure 6.5.5). Set 4 was a pool of the 200 most enriched sequences from 2-month old WT replicates and set 5 was a pool of the 200 enriched sequences from 12-month old WT replicates. Peptides that were present within the 200 most enriched sequences from at least 30% of 2-month old TRACK replicates (training cohort, n=10) and absent from set 4, were selected (Figure 6.5.5A.). Nine peptides met these criteria and their frequency was averaged per sera sample (training and testing cohorts, Table 6.5.3.). Peptides that were present within the 200 most enriched sequences from at least 30% of 12-month old TRACK replicates (training cohort, n=6) and absent from set 5, were selected (Figure 6.5.5B.). Ninety six peptides met these criteria and their frequency was averaged per sera sample (training and testing cohorts). A cut off value was used in order to evaluate this assay which was 10x the frequency sum of the WT age control sera (cut off was determined in order to maximise specificity). By utilising the sum of percentages, the 2-month specific peptides (n=9), the sensitivity of the method was estimated at 36.4%, and the specificity at 85.7% (Figure 6.5.6). Additionally, by utilising the sum of frequencies of the 12-month-specific peptides (n=96), a cut-off value was used in order to evaluate this assay which was 10x the frequency sum of the WT age control sera. The sensitivity of the method was estimated at 45.5% and its specificity at 100% (Figure 6.2.7.). Notably, if the sera from 2-month old mice were excluded, the sensitivity of the assay was estimated at 76.5% and its specificity remained at 100% (taken into consideration that some mice were assessed in multiple time points. Taken all these together, the latter peptide

selection was the most sensitive and specific diagnostic approach, and it may be appropriate to test this immunosignature assay with human sera.

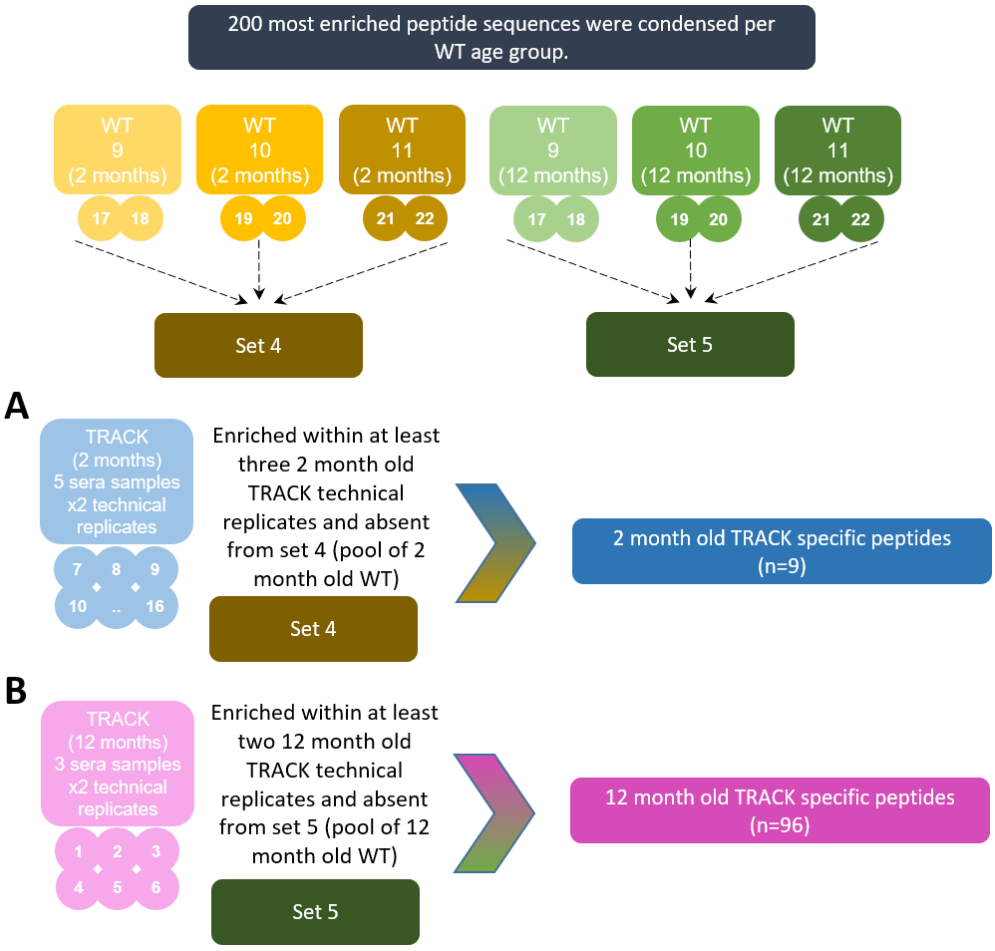


Figure 6.5.4 Overview of the frequency analysis between TRACK replicates of the training set and pools of the 200 most enriched sequences from WT sera.
Peptides enriched within at least 30% of the TRACK replicates and absent from their equivalent negative selection sets (sets 4 and 5) were selected. A. Peptides enriched within at least 30% of the 2-month old TRACK replicates (training cohort, n=10) and absent from set 4 were selected (n=9). B. Peptides enriched within at least 30% of the 12-month old TRACK replicates (training cohort, n=6) and absent from set 5 were selected (n=96).

Table 6.5.2 List of translated amino acid sequences from the enriched peptides resulted from the frequency NGS analysis of the second biopanning round using a defined input phage sub library.
These peptides were present within at least 30% of the 2-month old TRACK replicates and absent from set 4 (pool of 200 most enriched sequences within 2-month old WT replicates.

2-month old specific peptide sequences (frequency analysis, n=9)
TQMAGPRPPIKDLLPA
TMLDLRLAAWHATKPN
QAGPRPPLLPVRNLLT
NMLRHRLARMDMAGPM
LLLAGPRNAMLTHRPP
LFAGPRPRLATTLGVA
LAQHRAGMAPMPPELT
KVGDPLPAAKVFPPMQ
APKPLDPGVLTPPAHN

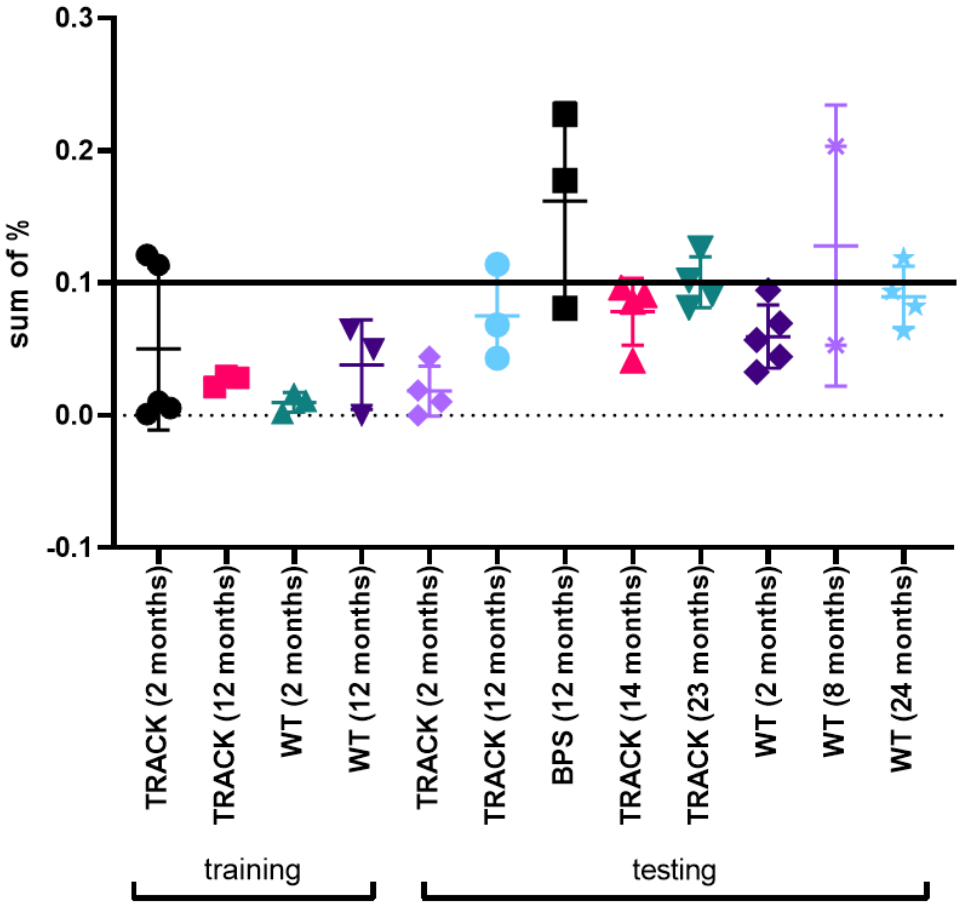


Figure 6.5.5 Assessment of the sum of frequencies of nine 2-month old TRACK specific peptides that were selected with a stringent frequency NGS analysis and their sum was averaged for each sera sample (within training and testing cohorts).

Nine peptides were enriched within at least 30% of 2-month old TRACK replicates and absent from set 4 (pool of enriched sequences of 2-month old WT mice). Their frequency was summed per individual sera sample. A cut off was determined as 10x the averaged frequency sum of these peptides within 2-month old WT mice (negative selection). Positive samples were defined by applying this cut-off value, and 2-month old TRACK sera samples were characterises as positives (n=2), establishing assay's sensitivity and specificity at 36.4% and 85.7%, respectively.

Table 6.5.3 List of translated amino acid sequences from the enriched peptides resulted from the frequency NGS analysis of this immunosignature biopanning round.

These peptides were present within at least 30% of the 12-month old TRACK replicates and absent from set 5 (pool of 200 most enriched sequences within 12-month old WT replicates.

12-month old specific peptide sequences (frequency analysis, n=96)			
TYRTLDTLPEPKPQAN	TPEVKETRLSGDATNY	PNWVLENPPPNSWEHG	KTVVGSLDKGPPDPSP
NDPVTMYSEPCRPDLQ	TIARDTIRTIMLVNNA	PNTPVCSLDTCWNQPH	KKSFNARDDTPVTLAH
YPTKLYGIMGSLDADP	SYAQTLDLPLGQNIRW	PNKAQVTTHVTLDPPQ	HVPLEWMRAPNSGSAS
YPGPAVPTLDMLPPPL	SNGATLDKTEPRIHAP	PMKAPFNASKDPSAKL	HTRGSLDVDDASHSHF
YLHPATSLDEVYNTPR	RYYLHQTHTTLDSPLK	PKRHIVPTVATTLDTT	HTLPALQEGMLLEPWC
WPRPAATLDMPTHAMP	RTLPHGSLDIPMKSAS	PKQTLDLAPMPMKHTE	HRTLDTPAPNPHTDLT
VRTSLDPTPSNQRLTM	RSTTVRPTLDQPATSL	PATTLDHPPRGFNTPE	HPTLDRPPTVHDFYPK
VPTEWLHTASDNPMRN	RRQPKAVPSLDPPLMH	PATTLDGVYLASIELV	HHVHRTLDYSTGFATL
VPAAWLSTPKTDALKH	RQSGTLDPPPSKIPSD	NVLLDRIIAHSVHRGV	HDPVQYYDPSPMTQDA
VHATLDTNDNSGSRPG	RPANTIHVTLDKPLPP	NVKPTLDTVSLRTPDA	HARDASISTILLPISQ
VFYAETLDSYPHREPP	RHTRVVRRTLDEPSAP	NHIVVRTLDDSYPAYR	HAKDIAYSMSPPNATL
TWFAPLEHTILYRTQA	RETYTMPTLDTVKDNH	NGMLSRPQLSLDHNSP	HAKDASPFAVSIYKFA
TTYPTVAHTLDSPTEN	QWMFSLDSPDRSQSKD	NFVPPAWLPRESNELR	GSHVRPTLDQPEQSRD
TTTLAPYGKTLDPAKP	QSTMHTTLDEPHHTNP	NETLTGNIPVHMKCHA	GQTLDLSPQMDTRVVR
TTQNTSRVVHATLDQS	QRLYYQNVATSLDTNY	NASRLNSGMSVQVPR	EVRSAPVHPTLDTVLA
TTPSMPLVHQTLDAPR	QNTAGSLDWVQPTYEP	NARDLNSAQLLQSPDC	DYPTTRPTLDIPTPGV
TTNLKIAASALDSPHI	QGFTNSMTHGSLDPTR	NAMHLGNNAHGPHPSH	DYLRSNSAMATLDPPL
TSVLNSLDSAYMPGST	QAMTTLDTPLPESRRM	NAGALTAHVLTTASTA	DLAPHVFPVQMLMNQG
TSTNTPPHEPVQPQSD	PVYIHRTLDHTVKNTN	NAGAAAPRSTYTALQS	CTLDSCSQAQRLLGEA
TSPADRFLQKVLWDSH	PVPNEWIPPTIKPVLH	MVVNTPPAPTLDTPHA	AWPTLDSPEVFSDKLS
TQGSLDTEHPFPMLAA	PTDWPNTLDNWPSPLP	MHTTLDGEMRTQQHEF	ATHITLDTVQDGPAGT
TPSHWLPEMLPTIPDK	PTARTLDNPPSTSIPT	LMELEALRYLKRVNNS	AKGSLDTTPARQF
TPRTHGSLDLPDASAY	PRDTSAVVVGSLDTNP	LLTVHPTLDSAPPGHT	AIPTLDPHDPIPGRYH
TPIEWLLETTPEVRRW	PPSWSARTLDMPPPLT	LDRWSTPAPSLDLPAP	AETYPYTLDQVQDGIH

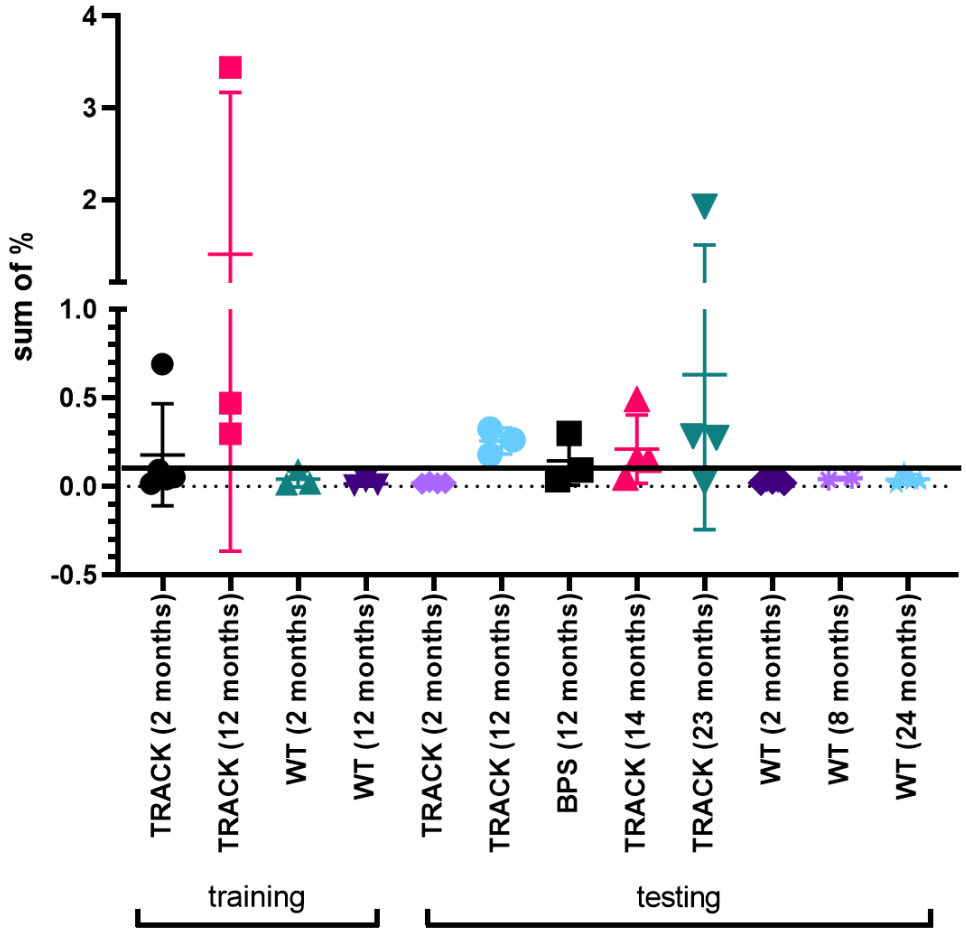


Figure 6.5.6 Assessment of the ninety six 12-month old TRACK specific peptides that were selected with a stringent frequency NGS analysis and their sum was averaged for each sera sample (within training and testing cohorts).

Ninety six peptides were enriched within at least 30% of 12-month old TRACK replicates and absent from set 5 (pool of enriched sequences of 12-month old WT mice). Their frequency was summed per individual sera sample. A cut off was determined as 10x the averaged frequency sum of these peptides within 12-month old WT mice (negative selection). Positive samples were defined by applying this cut-off value, and 12-month old TRACK sera samples were characterised as positives, establishing assay's sensitivity and specificity at 45.5% and 100%, respectively.

6.6. The application of immunosignature assays to human cc-RCC.

In summary, three different biopanning strategies have been described using sera from the TRACK mouse model (transgenic model of cancer of the kidney)

in order to develop a serum based diagnostic assay. The resulting candidate diagnostic peptides were assessed by conventional ELISAs or with an *in silico* diagnostic approach, an NGS based method that used phage-peptides as a soluble array to produce a 'cancer immunosignature'. The best peptide-ELISA was with synthetic peptide 17 and had a sensitivity of 63% (increasing to 72.7% for animals > 12month of age) and a specificity of 92%. The immunosignature approach produced an assay utilising 96 peptides and had 46% sensitivity (increasing to 77% for TRACK animals >12 months of age) and 100% specificity. Driven by the encouraging results from the mouse model, the study set out to investigate the applicability of the immunosignature approach with sera from patients with clear cell RCC (Renal Cell Carcinoma). This type of RCC was chosen, as it is the most common type amongst RCC cases, representing 60-80% (Oosterwijk, 2011). Sera from cc-RCC patients (n=100) and healthy volunteers (n=50) were available from Leeds Multidisciplinary Research Tissue Bank. The majority of the RCC sera samples derived from men (62%). The average year of diagnosis was 60 years and 61% of them diagnosed incidentally (Table 6.6.1.). Patients from all different stages and grades of cc-RCC were included as stage classification is a marker of cancer advancement and grade is a marker of cancer differentiation (Dagher et al., 2017). 17% of the RCC samples had already developed metastasis (88.3% of these metastases being lung cancer) and all of them were from the stage IV cohort. 59% of all RCC patients were current or ex-smokers and 37% of them had hypertension. This was expected as increased blood pressure and smoking are one of the main RCC risk factors (Hsieh et al., 2018). Stage (cancer advancement) and grade

(cancer differentiation) were taken into account when the study was designed in order to have the best possible representation of all different stages, grades, sex and age in both the training and testing cohorts. Sera from 40 RCC patients (mean of age=58) and from 20 volunteers (mean of age=50) were selected to be included in the training set. Half of these 100 RCC patients had developed stages I-II cancer and half of them had developed stages III-IV, with a mixture of grades (Table 6.6.2.).

Table 6.6.1 Summary of the type of stage, grade, sex, age of diagnosis of the clear cell RCC patients (n=100) and sex and age of the healthy volunteers (n=50).

	Total RCC	Stage I	Stage II	Stage III	Stage IV	Healthy
Total samples	100	43	7	33	17	50
Female	38	20	4	11	3	25
Male	62	23	3	22	14	25
Average age (years)	60	58	59	60	64	37
grade 1	4	3	1	0	0	-
grade 2	19	12	3	4	0	-
grade 3	37	19	3	10	5	-
grade 4	30	1	0	19	10	-
Unknown grade/mixture of grades	10	8	0	0	2	-

Table 6.6.2 Summary of the type of stage and grade of the clear cell RCC patients (n=40) that were selected to be included in the training cohort.

Stage	Grade 1	Grade 2	Grade 3	Grade 4	total samples
I	1	5	10	1	17
II	1	1	1	-	3
III	-	1	4	8	13
IV	-	-	2	5	7

Firstly, a biopanning round was conducted using bacteriophage derived from (input titration $1x10^{13}$ CFU/ml) pc89_BspQI^{-} 16mer peptide library and it was

biopanned against sera from the training cohort that was made up of 40 cc-RCC (Table 6.2.2). Three sequential subtraction steps were carried out using a pool of healthy sera (50 healthy sera samples) before the biopanning round. Output phage produced titres between 2 to 5 x 10^4 CFU/ml. All output bacteriophage populations were pooled together and used as a common reagent, referred to as human specific sub-library and it was panned against training (n=60) and testing cohorts (n=90), with phage input titration of 5x10^{11} CFU/ml. Three sequential subtraction step using a pool of healthy sera (n=50) was conducted before the biopanning round. Output phage produced titres 1 - 6 x 10^6 CFU/ml. All samples (no technical replicates were included) were individually PCR prepped as described before, and sent for Ion Torrent Next Generation Sequencing (n=150, samples were split between two different NGS runs, due to the availability of NGS barcode primers). NGS data was analysed to identify potential peptides that are recognised by autoantibodies and mimic TAAs and therefore could be used in a diagnostic assay (Method 2.2.12.). Analysis involved ranking peptide candidates using either their Z scores or their actual frequency (top 200 analysed for each sample). Firstly, all enriched sequences from the healthy training cohort (n=20) were pooled together, creating a negative set. A two-way Z score analysis was conducted comparing enriched sequences within every cc-RCC sera sample from the training cohort (n=40) with the negative set. Peptides with Z score value ≥5 and present within at least 12.5% of the training RCC samples were selected (n=82). The sum of their % was calculated separately per sera sample. Cc-RCC sera samples were not distinguished effectively from healthy samples (Figure 6.6.1.). When

utilising frequency analysis, the 200 most enriched sequences were identified per sera sample in the training set. One negative set was created by pooling the 200 most enriched sequences from all the healthy sera samples (training cohort, n=20) and the most enriched peptides from RCC samples stages I and II (n=20, set 1), stages III and IV (n=20) and all stages (n=40) were listed. Peptides that were present within 12.5% of the positive sets and absent from the negative pool were selected (n=18). A sum of their frequency was calculated per sera samples from training set. The sensitivity and the specificity of this assay was determined at 50% and 80%, respectively. This applied a cut-off value of 1.5x of the summed frequency of these peptides within the healthy sera samples (Figure 6.6.2.). The same strategy taking into account the 500 most enriched sequences in all the samples, was also applied but was not adequate to distinguish between RCC and healthy sera samples (data not shown). In addition, an approach to implement a less stringent peptide selection approach, peptides that were present in less than 20% of the most enriched sequences (applying either the top 200 or 500 peptide sequences for each sample) of the healthy sera, were not excluded. Analogous analysis comparing all RCC samples with the negative samples, failed to identify any positive sera samples above cut-off values (data not shown). Taken all together, a human immunosignature was not as highly sensitive/specific compared to data produced using the mouse model data, and perhaps more biopanning rounds to generate input phage with further enrichment of TAA mimics would be needed to produce a more effective assay.

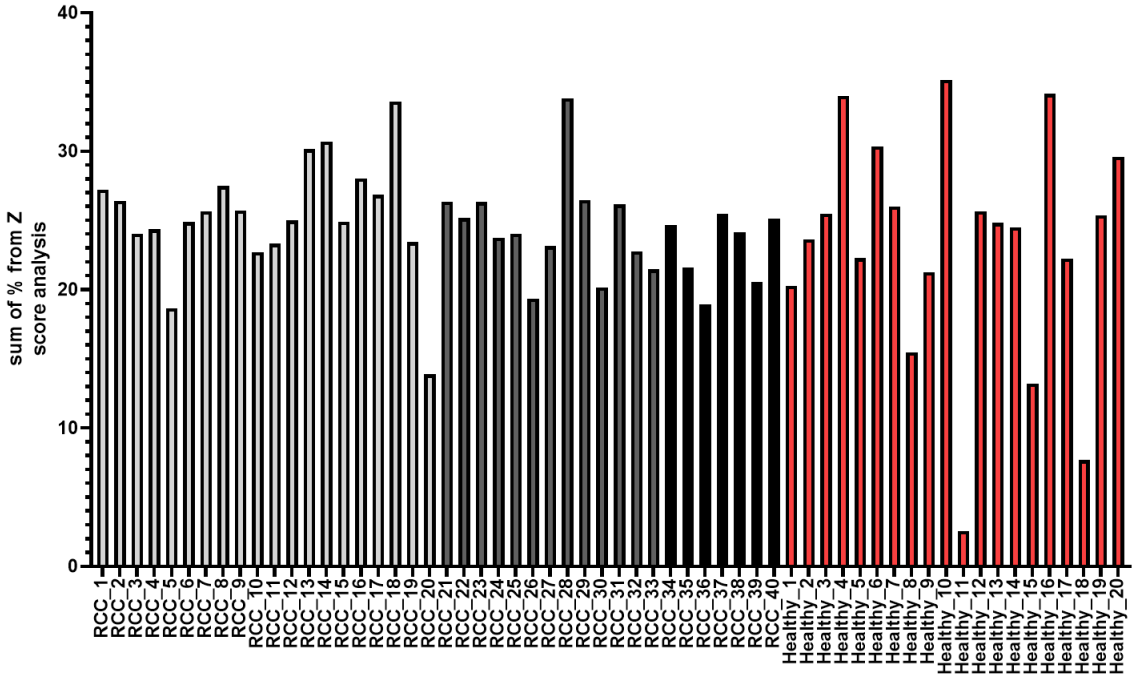


Figure 6.6.1 Assessment of eighty two RCC specific peptides that were selected on the basis of Z score analysis within NGS data (training cohort).

A Z score analysis was conducted between a pool of enriched sequences within the healthy sera (training cohort) and all the sera samples of the training cohort with this set. Peptides with Z score ≥5 and present within at least 12.5% within RCC training samples were selected. Eighty two RCC specific peptides met these criteria and their % were summed per individual sera sample. An appropriate cut off was not established.

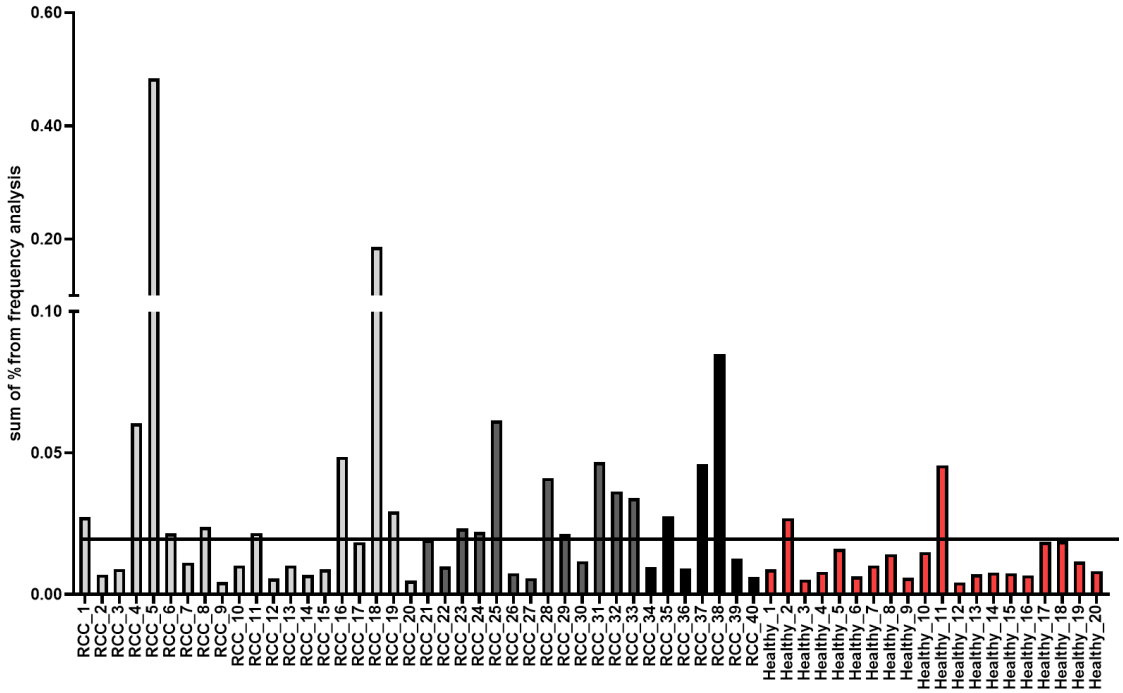


Figure 6.6.2 Assessment of eighteen RCC specific peptides that were selected with a frequency criteria in NGS data (training cohort).
Peptides that were present within at least 12.5% within the 200 most enriched sequences of the RCC samples and absent from the 200 most enriched sequences of the healthy samples were selected. Eighteen RCC specific peptides met these criteria and their frequency was summed per individual sera sample. A cut off was determined as 1.5x the sum of the frequencies of these peptides within healthy sera of training cohort. Positive samples were defined by applying this cut-off value, establishing assay's sensitivity and specificity at 50% and 80%, respectively.

Additionally, one biopanning round was conducted with the mouse immunosignature sub-library (developed in the previous chapter) against the same RCC and healthy sera of the training set, in an attempt to explore the translation of the mouse immunosignature to human biology, as the occurrence of the VHL mutation (impairing the HIF pathway) in cc-RCC is one of the most common ones (Linehan and Ricketts, 2019). Phage sub-library (titration 1x 10^{12} CFU/ml) was panned against the training and the testing cohort of human samples, as before. Three sequential subtraction steps using

a pool of healthy sera (n=50) were conducted before the biopanning round. Output phage titres were between 1 to 6 x 10^5 CFU/ml. All samples (no technical replicates were included) were individually PCR prepped as described before, and sent for Ion Torrent Next Generation Sequencing (n=150, split between two different NGS runs). NGS data was analysed to identify potential diagnostic peptides identified in an original mouse immunosignature panning. Analysis involved ranking peptide candidates using either their Z scores or their frequency (Method 2.2.12.). Firstly, the frequency of the previously selected mouse immunosignature peptides (discussed in Chapter 6.5) was calculated, but unfortunately the majority of the frequencies were 0% (data not shown). Frequency analysis, as described before using either the 200 or 500 most enriched sequences was conducted, but peptides enriched within at least 10% of the RCC samples were not identified (data not shown). Z scores were calculated using pooled data for the healthy sera samples of the training set (n=20) as the negative samples to generate Z scores for each RCC sample. Peptides enriched within at least 12.5% of the RCC samples of the training set (n=40) with Z score ≥5 were selected. 123 peptides met these criteria, and a more stringent sorting method was then applied where peptide sequences that were not 16mers, sequences containing the library's bias motif DCSS within all the data sets, were excluded. 79 peptides were cc-RCC specific and their frequency as well as their individual Z scores were calculated for all the sera samples. Firstly, the sum of these mouse derived cc-RCC selected peptides (n=79) Z scores was calculated for all the sera samples. These samples were included in two different NGS runs which had different sequencing depth

(Figure 6.6.3.). The sum of their Z scores was calculated for all the sera samples, and by applying a cut-off value of 1.5x of the frequency sum within the healthy samples of the training sets, the sensitivity of the assay was 85% and its specificity was 90% for the training cohort (Figure 6.6.4). The same cut-off value was applied for the testing cohort with 30% sensitivity and 100% specificity (Figure 6.6.5). An epitope motif within these peptides was identified by the online MEME platform (Bailey and Gribskov, 1998). Blasting this motif (GSAASGFI) into the IEDB (Marcatili et al., 2018) revealed that this sequence had been described as potential epitope of the immunoglobulin heavy variable genes 3-30 and 3-23.

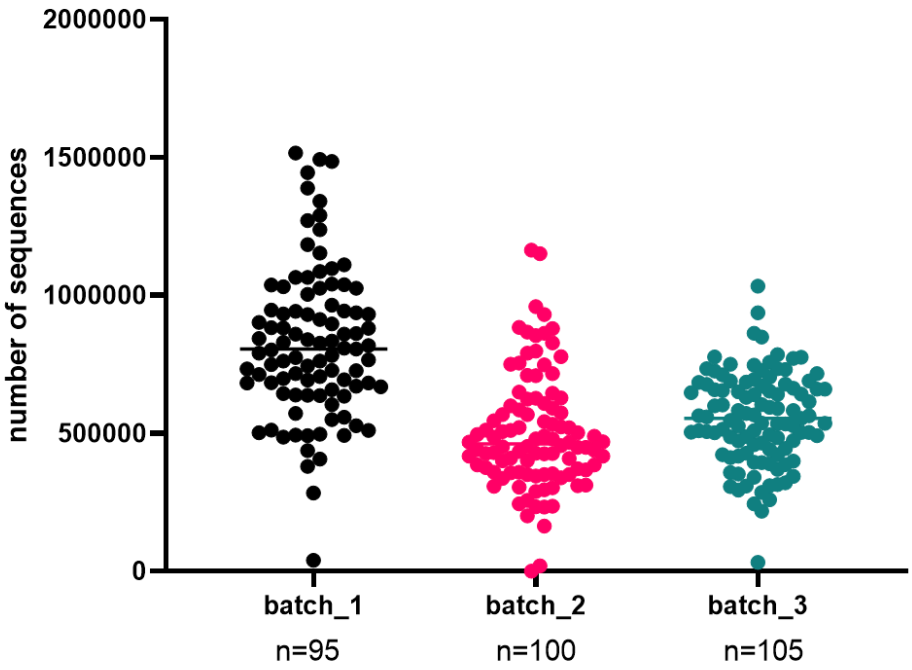


Figure 6.6.3 Sequencing depth between the three different NGS runs that included all human samples analysis, and the number of samples that were used per sequencing run.

The amount of total sequences that were in frame with the flanking regions of the AEGEF and DPAKA were calculated per sample within the three different runs. Samples where the mouse immunosignature sub-library was used in panning were included in batches 1 and 2; samples where the human immunosignature sub-library was used in panning were included in batches 2 and 3. Batch 1 seemed to have disproportionally more sequences included than batch 2, therefore the Z score analysis that was conducted between the two runs, was incompatible.

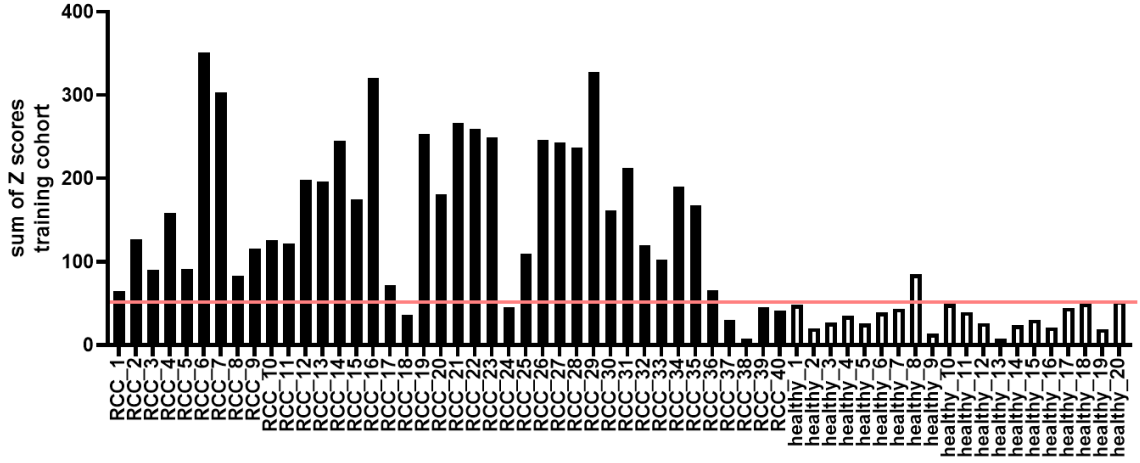


Figure 6.6.4 Assessment of 79 cc-RCC specific peptides that were selected with Z score analysis of NGS data (training cohort).
A Z score analysis was conducted between a pool of enriched sequences within the healthy sera (training cohort) and all the sera samples of the training cohort with this set. Peptides with Z score ≥5 and present within at least 12.5% within RCC samples were selected. 79 met these criteria and their frequency was summed per individual sera sample. A cut off was determined as 1.5x the sum of the frequencies of these peptides within healthy sera of training cohort. Positive samples were defined by applying this cut-off value, establishing assay's sensitivity and specificity at 85% and 90%, respectively.

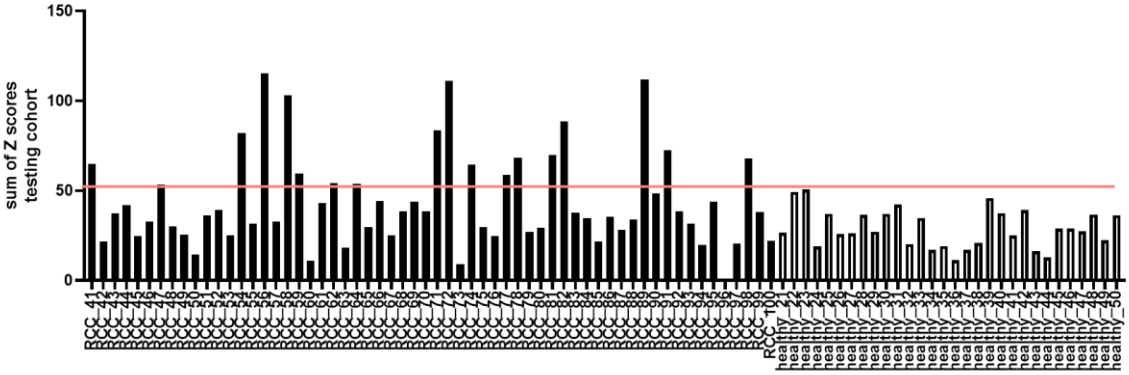


Figure 6.6.5 Assessment of 79 cc-RCC specific peptides that were selected with Z score analysis of NGS data (testing cohort).
A Z score analysis was conducted between a pool of enriched sequences within the healthy sera (training cohort) and all the sera samples of the training cohort with this set. Peptides with Z score ≥5 and present within at least 12.5% within RCC samples were selected. 79 met these criteria and their frequency was summed per individual sera sample. A cut off was determined as 1.5x the sum of the frequencies of these peptides within healthy sera of training cohort. Positive samples were defined by applying this cut-off value, establishing assay's sensitivity and specificity at 30% and 100%, respectively.

Furthermore, TRACK-specific peptides, previously identified by conventional ELISAs in this RCC model in mice were screened against a pool of RCC and healthy samples in order to assess their potential diagnostic value in human sera. An irrelevant peptide was also included as a negative control. The fold-difference between the OD value of RCC pool and healthy pool of sera for every peptide was calculated (Figure 6.6.4.). None of these peptides were above 2-fold difference, thus they were not considered human RCC specific

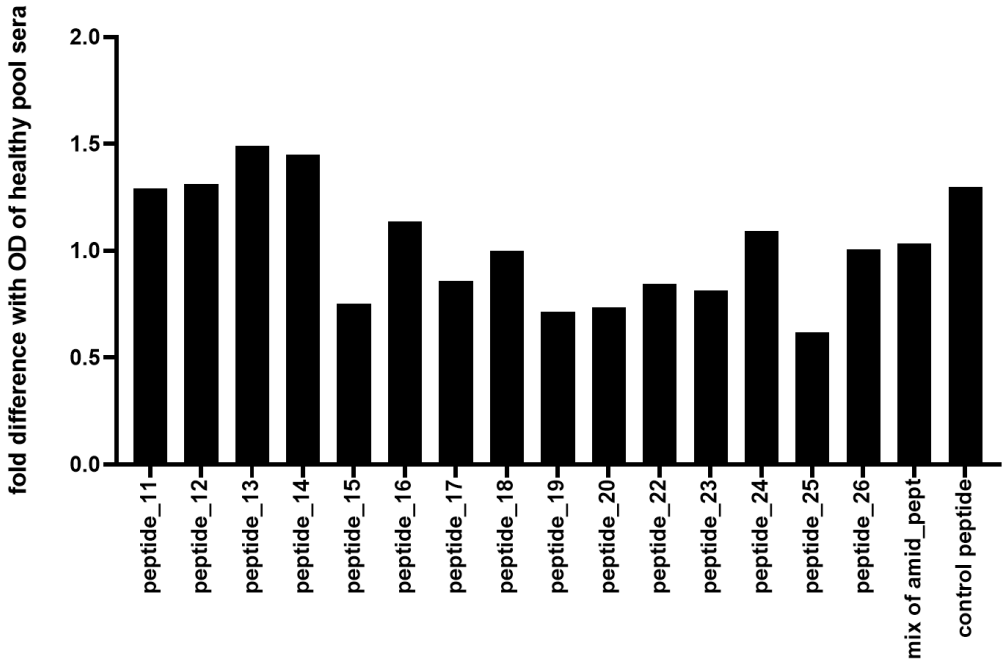


Figure 6.6.6 Assessment of the diagnostic value of TRACK specific peptides (n=15) screened against cc-RCC and healthy pool.
Fifteen TRACK specific peptides (as described in chapter 6.4.) were screened against a pool of sera derived from cc-RCC patients (n=100) and healthy volunteers (n=50). Irrelevant peptide was included as negative control. Fold difference between the OD of the sera pools was not able to identify any RCC specific peptide above the 2-fold cut-off value.

Taken all together, the best available assay to diagnose cc-RCC in human samples was designed by using the mouse TRACK sublibrary of phage-peptides

that was then panned against IgG from human RCC and control samples and peptides selected and analysed by Z score. The assay had a sensitivity of 62.5% and a specificity of 100% on the training cohort, but that was not applicable on the testing cohort across all human samples analysed (100 RCC and 50 healthy controls), suggesting that increased biopanning rounds would possibly increase the selection of enriched mimotope sequences.

6.7. Discussion

Cancer screening in asymptomatic patients is one of the main goals of cancer research (Mordente et al., 2015). Early screening markers would be very useful, especially for Renal Cell Carcinoma, since most of the RCC patients are being diagnosed after metastasis had occurred, resulting in a survival rate of less than 20% (Hsieh et al., 2018). Therefore, to the development of a RCC specific diagnostic assay would be extremely useful. Tumour Associated Antigens, antigens that are responsible for cancer progression and the main drivers of the cancer (auto)immune response, are attractive targets for diagnostics but realistically very difficult to be identified, due to their low abundancy in the circulation (Kobayashi et al., 2020). Therefore, the identification and detection of their epitopes (mimotopes) could be an excellent alternative approach as a priori knowledge of the TAA is not needed. In this chapter, NGPD was applied to screen for peptides that could mimic Tumour Associated Antigens (TAAs) that elicit an autoantibody response in Renal Cell Carcinoma (RCC) in a mouse model (TRACK). This was then followed by the application of the most successful epitope mapping strategies to human

derived samples. These mimotopes could be used either in an ELISA based diagnostic format or in an *in silico* format where NGS data is used as the assay readout. To date, FDA-approved biomarkers are being tested primarily in an immunoassay format (Mordente et al., 2015), as well as other methods like CancerSEEK, looking for methylation patterns using NGS (Cohen et al., 2018), thus an alternative approach could be an *in silico* method, in which peptides mimic the tumour antigen epitope repertoire and can be identified by NGS.

Sera from a RCC mouse model, called TRACK, were available, and an initial biopanning round was conducted using the 16mer peptide pVIII library. The aim was to identify peptides enriched within TRACK sera and absent within age matched WT sera samples. Analysis involved ranking peptide candidates using either Z scores or their actual frequency. A second round of biopanning round using phage-sublibraries derived from the first biopanning round, each panned against the same sera in round 2, in order to increase the enrichment, was conducted. Peptides were ranked based on their Z score or their frequency and 42 were selected. These peptides were screened in ELISAs as phage fusions against TRACK and WT sera pools in order to determine the best peptide that can be recognised by TRACK sera (n=6). When selected peptides were screened as synthetic peptides (including one as a biotinylated peptide), five demonstrated high reactivity to TRACK sera. The best performing ELISA was with synthetic peptide 17 and had a sensitivity of 62.5% sensitivity and 91.6% specificity. Notably, 4 of these TRACK-specific peptides carried the same epitope motif (SGM), but this motif could not be identified by any known

epitope databases. It is not surprising that not all of the enriched peptide sequences that were identified by NGPD, were positively identified by ELISA from TRACK sera, since as it has been discussed in Derda *et al.* that the abundance of the bacteriophage clones does not always correlate with its affinity/reactivity (Derda et al., 2011).

In an alternative approach, it was attempted to create a highly enriched TRACK specific phage sub-library that could be used as a reagent for a diagnostic *in silico* screen. Based on the analysis of the second round of biopanning, three of the TRACK sera showed higher immunoreactivity to their respective input-phage sub-libraries than the other samples, therefore these sub-libraries were pooled and used as the TRACK specific phage sub-library. More TRACK and WT sera that were not previously included were available to screen. From this panning, selection of TRACK-specific peptides was carried out by both Z score analysis and frequency, with the latter producing the most effective assay. The sum of normalised frequencies for 96 TRACK specific peptides were determined for all samples and resulted in an assay able to distinguish RCC with 45.5% and 100% sensitivity and specificity, respectively. Notably, the sensitivity and the specificity were 76.5% and 100% when 2-month old TRACK mice were excluded from the data sets. This is not surprising as the first histological cancer evidence start to develop when TRACK mice were 6 month old (Fu et al., 2011). This development of an immunosignature using NGS analysis as the readout provided an effective assay in this mouse model of RCC. Human and murine immune systems are similar, even though there are some

differences in the immune cell types (Mestas and Hughes, 2004). Following these encouraging results that sera from a mouse model of RCC could be distinguished from healthy mouse sera, a larger study was designed using sera from clear cell-RCC patients (n=100) and from healthy individuals (n=50). Carefully selecting a representative training set which include sera from all of different RCC stages and grades was pivotal. Two different panning strategies were followed. The sub-library from an initial biopanning experiment against sera from the training set using the 16mer library was applied to all samples in a second panning round. NGS analysis of the eluted phage-peptides failed to reveal any enriched peptide sequences that could form an *in silico* immunosignature to diagnose RCC in the training set.

HIF-α is one of the main drivers of the cc-RCC, mainly because of its cancer suppression property and its implication in the main RCC pathway (Hoefflin et al., 2020). Therefore, the mouse-specific immunosignature sub-library derived from the TRACK model, that targeted HIF-α, could be an alternative way to map the auto reactivity in human cc-RCC sera, since there is a high homology between the human and mouse variants (Iyer et al., 1998). Therefore, a bioinformatics strategy was used as a method to define enrichment and identify peptides acting as cancer mimotopes when the immunosignature phage sub-library derived against TRACK samples was used instead of the human derived one. With Z score sorting and analysis, enriched peptide sequences (n=79) were able to distinguish RCC from healthy sera with 85%

sensitivity and 90% specificity in the training cohort and 30% sensitivity and 100% specificity in the testing cohort.

These data are in accordance with a similar approach in which epitope mapping of the autoimmune response was carried out in cases of lupus. NGS analysis allowed the selection of lupus-specific peptides and the detection of lupus with 75% sensitivity and 90% specificity (in a validation cohort). Cut off of their assay was 1.2 fold difference, which is in accordance with similar cut-off values that were established in our case (Wu et al., 2019). As it has been previously mentioned, a traditional way to develop a screening assay for cancer, is to identify TAAs that are implicated with cancer progression. A study identified 55 proteins that might play a role in RCC progression, and a selection of those were able to distinguish RCC from healthy tissues, with 70% accuracy, when only one protein marker was used. Further validation in a testing set was needed (White et al., 2014). In our case, TAAs were not able to be identified by the common motif epitopes present in the enriched sequences.

Epitope mapping strategies of complex sera sample (polyclonal responses) using phage display coupled with next generation sequencing have been previously attempted. For instance, three rounds of biopanning using phage peptide library (diversity 10^9) against sera from patients with peanut allergies led to the discovery of dozens of peptides that mimic the epitopes of the responsible protein (Christiansen et al., 2015). In our case, TRACK specific peptides were identified after less biopanning rounds, using our optimised epitope mapping conditions. By reducing the selection rounds, diversity can be

maintained and bias introduced by amplification minimised. Phage display is always subject to the fact that during the amplification, some clones will be non-specifically amplified (due to growth advantages), and carried over to the next round of biopanning (Halperin et al., 2011). These issues can be addressed by reducing the amount of selection rounds, more specific elution methods as well as the incorporation of NGS analysis to exclude any enriched sequences that can be observed in both positive and negative samples. Cancer sera samples are very complex, therefore the presence of a wide range of peptide sequences is not surprising (n=96).

As an alternative to conventional diagnostic methods in which the discovery of TAAs is followed by a serological based assay using a limited number of epitopes/antigens, the detection of a plethora of peptides (mimicking the TAAs epitopes) would provide a more comprehensive representation of epitopes (Legutki et al., 2014). Similar to this approach, a serum based assay using 10,000 random peptides in a microarray format has been attempted to map the immunosignature of different types of cancer (Stafford et al., 2014). Immunosignatures seemed to be independent of the age and sex within a group of people sharing the same disease. In an analogous method to solid surface arrays, the suggested random peptide phage display libraries can provide a wide range of peptides, but do not involve the expensive cost of manufacturing peptides as well as the instability of the peptides on a solid surface. That being the case, the data presented in the current study

demonstrates that detection of a range of enriched phage-peptides binding to IgG using NGS can provide a novel diagnostic approach.

Prior studies have noted the importance of a combination of TAAs for a successful screening assay. Another alternative strategy to identify TAAs is the construction of bacteriophage library using extracted mRNA from cancer tissues. This has been applied to hepatocellular carcinoma, in which after five iterative biopanning rounds, seven novel TAAs were selected and used in a serological assay, with 82.6% sensitivity and 100% specificity (Liu et al., 2012). The success of this study may be due to the increased biopanning rounds and even though a cancer specific library was used, it may suggest that more biopanning rounds are needed in order to investigate complex cancer sera. Moreover, it indicated the need of a combination of TAAs in order to achieve high sensitivity/specificity, because the individual protein diagnostic value was around 20%. Consistent with these finding, autoimmune responses against cancer is very diverse from patient to patient, therefore it is highly unlikely a single TAA will be enough for early screening. One of Stafford's' approaches was to epitope map with a random peptide array assay (n=10,000), in which patient sera generated a unique immunosignature for Alzheimer's, and the study demonstrated different signatures that can distinguish patients from different disease development stages (Restrepo et al., 2013). Our study involved the biopanning of almost 10^5 more peptides (10^{11} input phage titration with 10^9 diversity) offering a possible explanation the inability to identify human RCC specific peptides due to low abundancy of the potential

specific clones, so possibly a smaller, yet more focused library (RCC- specific immunosignature) would have been an alternative approach. Taking all these together, this new method of generating immunosignatures (as demonstrated in sera from a renal cell carcinoma mouse model) could have a good diagnostic potential. In fact, preliminary data suggest this approach might be applicable to human sera, too.

7. Conclusions and future perspectives

7.1. Conclusions

The main goal of this research was to develop a diagnostic tool for the early detection of Renal Cell Carcinoma. Although this specific cancer type is among the ten most commonly diagnosed cancers worldwide, it is usually asymptomatic and it lacks an early diagnostic biomarker that could detect premetastatic cancer and potentially increase recovery rate; metastatic patients have a 12-20% chance of 2 year survival (Hsieh et al., 2018; Huang and Hsieh, 2020). One possible diagnostic approach could be the mapping of the host's autoimmune response against Tumour Associated Antigens (TAAs) in order to identify peptides that mimic TAA epitopes and could capture the cancer autoantibodies in a serological based assay. Phage display of peptides coupled with the analytical depth of Next Generation Sequencing could be a useful tool to investigate the potential of this diagnostic approach. Firstly, a large synthetic 16mer peptide library was constructed (peptides were displayed as pVIII fusions) and used for initial proof of concept studies for the mapping of monoclonal Abs. This NNK randomisation approach was inexpensive due to the small number of transformations that was required to make a high diversity library. The diversity was estimated to have a diversity of 5×10^9 CFU/ml, which is within the range of previously described naïve synthetic phage libraries (Kong et al., 2020). Deep sequencing of this library offered an unprecedented analytical depth of analysis and potential biases were identified compatible with other reported peptide libraries (Kong et al.,

2020); these biases were then easily removed from datasets in future experiments at the bioinformatics stage. Firstly, the amino acid (AA) distribution was evaluated and never exceeded 1.5-fold difference when compared with the theoretical AA distribution. The vast majority of clones (90%) displayed 16mer peptides, as expected. A small bias was identified (amino acid sequence DCSS) from the presence of undigested primer sequences (2.4%), but once identified, sequences containing this bias motif were easily excluded from all future experiments at the bioinformatic stage. Finally, the vast majority of the sequenced clones were unique (73%) and the probability of the same sequence being present more than 3 times was low (10^{-7}). Confirmation of a highly diverse peptide library with low bias was followed by its application to the epitope mapping of monoclonal antibodies (mAbs) in order to identify the optimal experimental conditions that could be applied to mapping autoantibodies in polyclonal antibody samples. Single biopanning rounds against mAbs with known epitopes were carried out, assessing 24 different biopanning conditions (including a wide range of elution, washing and Ab capture methods). Peptides-pVIII clones were screened in conventional phage ELISA format against the same mAbs, identifying the condition with the highest percentage of positive clones. Consistently, the conditions that were most successful were when mAbs was immobilised on protein G coated plates, washing was performed using 2% PBST + 1 M NaCl solution (5 times), followed by 1x PBS (5 times) and elution was with DTT. These conditions resulted in around half of the tested clones being positive by ELISA. Notably, Sanger Sequencing of the positively identified clones,

confirmed already known epitope motifs (RPxxQY for mAb SAF84, YPN for mAb SAF70 and GW for mAb SAF15); additionally, a new SAF70 epitope motif was identified (PW). Following this, a study was designed to further support the evidence that Next Generation Phage Display (NGPD) can be successfully used for mapping polyclonal antibody responses. NGPD with the optimised epitope mapping conditions was applied to normal mouse sera spiked with known mAbs. Two different findings were described after comprehensive NGS analysis of the output phage sub-libraries. Firstly, selected enriched peptide sequences within the spiked NMS were indeed mAb specific, as confirmed by phage ELISAs. Secondly, by using the sum of the frequencies of stringently selected peptides (the immunosignature), spiked samples were successfully discriminated from non-spiked samples. In fact, the limit of detection of this proposed method was directly compared with conventional screening methods such as ELISAs, qPCR and phage-microarray and it was at least as sensitive, being capable of detecting mAbs diluted 10^{-3}-10^{-4}-fold. This set of data further demonstrated the potency of this method and therefore, it was applied to sera derived from an RCC mouse model (TRACK). Initially, a limited amount of TRACK and WT sera was available (n=14), therefore two iterative biopanning rounds were conducted in order to firstly, identify peptides that mimic the TAA epitopes, and secondly, develop an immunosignature diagnostic approach. Individual selected enriched peptides (n = 42) were assessed for their RCC reactivity either as phage fusions or as synthetic peptides in a serological ELISA format. Some peptides were remarkably reactive against TRACK sera and distinguished TRACK from WT sera with up to

62.5-81.3% sensitivity and 66.7-92.6% specificity. Moreover, the sensitivity/specificity of the NGPD immunosignature application when applied to 12 month or old TRACK were 76.5% and 100%, respectively when applied to a wider set of mouse sera that became available later in the study (training set n=14 and testing set n=29). This evidence further underpins the potency of this serological based method and its implementation for RCC diagnosis. Therefore, sera from cc-RCC patients (n=100) and healthy volunteers (n=50) were used in in two consecutive biopanning rounds, kindly provided by Leeds NIHR biobank. The same biopanning conditions were applied using two approached: 1) the mouse derived immunosignature phage sub-library that was used for the aforementioned TRACK immunosignature biopanning was used as the input phage for the immunosignature assay or 2) A single round of enrichment was performed for the naïve peptide library against 40 human RCC patient sera samples, the resulting sub-library of phage-peptides was then used as the input phage for the immunosignature assay. The sensitivity and specificity of the latter approach were 50 and 80% respectively in the training set, whereas the sensitivity and specificity of the mouse sub-library were 92.5% and 55% in the training set (n=60) but its sensitivity was decreased (35% and 86.7%, respectively) in the testing cohort (n=90). In general, the findings of this study suggest that mapping the host's autoimmune response against RCC by applying Next Generation Phage Display and identifying cancer specific immunosignatures could be a viable screening method for this specific type of cancer.

7.2. Future Perspectives

This approach will prove useful in expanding our understanding of how cancer immunosignatures could be used in early cancer diagnosis. Unfortunately, clinical samples are very physiologically complex, which could have relatively low sensitivity and specificity of the method when applied in human cc-RCC samples, compared to the TRACK murine model. It has been reported that heterogeneity of cancer or even patient's ethnicity commonly limit the usage of AAbs for diagnostics (Yadav et al., 2019). In other words, this NGPD method was successful when applied to TRACK sera, due to the fact that this model's cancer aetiology was based on mutations of one transcriptional factor that impairs the main RCC pathway and thus, the autoimmune response was driven by a single protein and also because all mice were under a specific pathogen free controlled environment. On the other hand, human cc-RCC pathology is a far more complex phenomenon, in which this specific pathway is affected only in up to 50% of the cases. The potential diagnostic value of autoantibody signature needs to be assessed with a plethora of samples from patients ideally from different stages of the progression of cancer, including from early diagnosed patients, which is the bottleneck of all diagnostic tests, asymptomatic and non-clinically detectable patients (Desmetz et al., 2009). A more thorough understanding of the subsets of cancer pathways within selected human samples could allow correlation with sensitivity/specificity of the assays. Another possible way that the assays could have been improved is the implementation of more biopanning rounds to increase the enrichment of

cancer specific peptides in the input phage sub-library. Additionally, only sub-libraries against human sera with high autoimmune response could be pooled together to create a single immunosignature sub-library. This would be analogous the approach taken with the TRACK mice but was not possible in the current study due to time constraints. This could again use an input phage sub-library that contains more enrichment of peptides recognised by the autoantibodies. Moreover, an additional uncontrolled factor is the lack of technical replicates in NGS analysis (but not in TRACK biopanning experiments) that could also minimise any sequencing or experimental variation. Finally, the biopanning method could be further optimised by the use of a fraction of the whole IgG population in the panning. An additional subtraction step could be used using sera from patients with other types of cancer (including those with the same impaired pathways or the most common cancer types), or a positive selection step could be introduced to select IgG that recognise well characterised renal cancer cell lines. Both approached could enrich for anti-cancer autoantibodies before panning.

Taken all together, this research lays the groundwork for the future development of sera-based diagnostic assays. Development of cancer biomarkers is one of the biggest challenges in modern medicine, because biomarkers need to meet a wide range of criteria (such as reproducibility, testing in sera from patients from all over the word and from different environmental and genetic backgrounds) in order to progress into clinic (Sidransky, 2002; Diamandis, 2014). Continuing efforts to develop early

diagnosis in other types of cancers have already emerged. GRAIL's Multi-Cancer Early Detection test can reduce all cancer deaths by up to 26% due to its successful early diagnosis of 485 types of cancer by sequencing cell free DNA (cfDNA) and identifying cancer-specific methylation patterns. Notably, the duration when cancer biomarkers (proteins, AAbs, cfDNA) are detectable is questionable, therefore even in the case of a large screening test, sensitivity may vary (Hubbell et al., 2020). Furthermore, CancerSEEK is a novel combination of genomics and proteomics in which cell free DNA (cfDNA) is sequenced and compared against 61 potential mutations as well as the protein level of eight protein biomarkers is measured. This approach can be informative for the diagnosis of eight different cancer types with sensitivity ranging between 69-98%, and 99% specificity. The data set was large (n=1005), but the sensitivity of CancerSEEK may decrease when applied in large healthy population due to potential other existing ongoing diseases (e.g. autoimmune diseases) (Cohen et al., 2018). CancerSEEK and GRAIL's approaches both aim for high-risk population screening. In other words, both these cancer tests have not yet been implemented to a large population scale and their actual contribution to early cancer diagnosis is yet to be shown. An additional issue that might arise is the 'over-treatment' of non-benign tumours that would not normally be treated, and this could increase the treatment cost and severely effect patient's well-being and mental health (Ahlquist, 2018). Lastly, the incorporation of Artificial Intelligence (AI) seems to be the future of personalised medicine, diagnosis and treatment. An attempt to identify the protein network within a single cell molecule has already been reported, which

can allow the monitoring of emerging pathology, hence identifying potential tumours earlier than symptom development, but the cost and the complexity of this approach is still not applicable to daily basis (Rozenblatt-Rosen et al., 2020). Additionally, AI could facilitate sophisticated peptide or antibody library design that can mimic natural evolution as well in more focused screening methods (Sormanni et al., 2018; Krištof Bozovicar and Tomaž Bratkovi˘, 2020). AI will have a dramatic impact on prognosis and on patient stratification in the near future especially in combination with advanced NGS and bioinformatics technology (Moritz et al., 2019).

The present study has provided a deeper insight into the development of cancer immunosignatures and their possible implementation in early detection. Furthermore, in combination with emerging technologies such as AI and machine learning, large datasets generated by immunosignatures could be further optimised and lead to earlier cancer diagnosis that can reduce psychological and financial burden, and most importantly, increase patients' recovery rate and quality of life.

8. References

Ahlquist, D.A. (2018) Universal cancer screening: revolutionary, rational, and realizable. *Precision Oncology*, 2 (1): 1–5. doi:10.1038/s41698-018-0066-x.

Aibara, N., Kamohara, C., Chauhan, A.K., et al. (2018) Selective, sensitive and comprehensive detection of immune complex antigens by immune complexome analysis with papain-digestion and elution. *Journal of Immunological Methods*, 461 (July): 85–90. doi:10.1016/j.jim.2018.06.021.

Akkaya, M., Kwak, K. and Pierce, S.K. (2020) B cell memory: building two walls of protection against pathogens. *Nature Reviews Immunology*, 20 (4): 229–238. doi:10.1038/s41577-019-0244-2.

Alberts, B., Johnson, A., Raff, M., et al. (2002) *Molecular Biology of the Cell*. doi:10.1093/aob/mcg023.

American Cancer Society (2018) Cancer Facts & Figures 2018. *American Cancer Society*. doi:10.3322/caac.21442.

Bailey, T.L. and Gribskov, M. (1998) Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics*, 14 (1): 48–54. doi:10.1093/bioinformatics/14.1.48.

Baldwin, R.W. (1971) Tumour-associated antigens and tumour-host interactions. *Proceedings of the Royal Society of Medicine*, 64 (10): 1039–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4335921%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1812752.

Baldwin, W. (1966) Tumour specific immunity against spontaneous rat tumours. *international journal of cancer*, 264: 257–264.

Bielecka, Z.F., Czarnecka, A.M. and Szczylik, C. (2014) Genomic analysis as the first step toward personalized treatment in renal cell carcinoma. *Frontiers in Oncology*, 4 JUL (July): 1–15. doi:10.3389/fonc.2014.00194.

Bock, C.H., Ruterbusch, J.J., Holowatyj, A.N., et al. (2018) Renal cell carcinoma risk associated with lower intake of micronutrients. *Cancer Medicine*, 7 (8): 4087–4097. doi:10.1002/cam4.1639.

Botta, G.P., Granowicz, E. and Costantini, C. (2017) Advances on immunotherapy in genitourinary and renal cell carcinoma. *Translational Cancer Research*, 6 (1): 17–29. doi:10.21037/tcr.2017.02.09.

Braun, R., Schönberger, N., Vinke, S., et al. (2020) Application of Next Generation Sequencing (NGS) in Phage Displayed Peptide Selection to Support the Identification of Arsenic-Binding Motifs. *Viruses*, 12 (1360): 1–32.

Bray, F., Ferlay, J., Soerjomataram, I., et al. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36

cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68 (6): 394–424. doi:10.3322/caac.21492.

Brodaczewska, K.K., Szczylik, C., Fiedorowicz, M., et al. (2016) Choosing the right cell line for renal cell cancer research. *Molecular Cancer*, 15 (1): 1–15. doi:10.1186/s12943-016-0565-8.

Brodersen, J. and Siersma, V. (2013) Long-Term Psychosocial Consequences of False-Positive Screening Mammography. *Annals Of Family Medicine*, 11 (2): 106–115. doi:10.1370/afm.1466.INTRODUCTION.

Cairns, P. (2012) Renal cell carcinoma. *Cancer*, 9: 461–473. doi:10.3233/CBM-2011-0176.Renal.

Capitanio, U. and Montorsi, F. (2016) Renal cancer. *The Lancet*, 387 (10021): 894–906. doi:10.1016/S0140-6736(15)00046-X.

Caraguel, C.G.B., Stryhn, H., Gagné, N., et al. (2011) Selection of a cutoff value for real-time polymerase chain reaction results to fit a diagnostic purpose: Analytical and epidemiologic approaches. *Journal of Veterinary Diagnostic Investigation*, 23 (1): 2–15. doi:10.1177/104063871102300102.

Carmen, S. and Jermutus, L. (2002) Concepts in antibody phage display. *Briefings in Functional Genomics and Proteomics*, 1 (2): 189–203. doi:10.1093/bfgp/1.2.189.

Casuscelli, J., Becerra, M.F., Manley, B.J., et al. (2019) Characterization and Impact of TERT Promoter Region Mutations on Clinical Outcome in Renal Cell Carcinoma. *European Urology Focus*, 5 (4): 642–649. doi:10.1016/j.euf.2017.09.008.

Cekaite, L., Haug, O., Myklebost, O., et al. (2004) Analysis of the humoral immune response to immunoselected phage-displayed peptides by a microarray-based method. *Proteomics*, 4 (9): 2572–2582. doi:10.1002/pmic.200300768.

Chang, A., Ting, J., Espada, A., et al. (2020) A novel phage display vector for selection of target-specific peptides. *Protein Engineering, Design, and Selection*, 33: 1–8. doi:10.1093/protein/gzaa023.

Chaplin, D.D. (2010) Overview of the Immune Response. *Journal of Allergy and Clinical Immunology*, 125 (2): s3-23. doi:10.1016/j.jaci.2009.12.980.Overview.

Chatterjee, M., Mohapatra, S., Ionan, A., et al. (2006) Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. *Cancer Research*, 66 (2): 1181–1190. doi:10.1158/0008-5472.CAN-04-2962.

Chen, G., Hayhurst, a, Thomas, J.G., et al. (2001) Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nature biotechnology*, 19 (6): 537–42. doi:10.1038/89281.

Chen, H., Werner, S., Tao, S., et al. (2014) Blood autoantibodies against

tumor-associated antigens as biomarkers in early detection of colorectal cancer. *Cancer Letters*, 346 (2): 178–187. doi:10.1016/j.canlet.2014.01.007.

Chow, W., Dong, L. and Devesa, S. (2010) Epidemiology and risk factors for kidney cancer. *Nature Reviews Urology*, 7 (5): 245–257. doi:10.1038/nrurol.2010.46.Epidemiology.

Christiansen, A., Kringelum, J. V., Hansen, C.S., et al. (2015) High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Scientific Reports*. doi:10.1038/srep12913.

Clackson, T. and Lowman, H.B. (2004) *Phage Display*. doi:10.1007/978-1-4939-7447-4.

Cohen, J.D., Li, L., Wang, Y., et al. (2018) Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*, 359 (6378): 926–930. doi:10.1126/science.aar3247.Detection.

Costa-Pinheiro, P., Montezuma, D., Henrique, R., et al. (2015) Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics*, 7 (6): 1003–1015. doi:10.2217/epi.15.56.

Craven, R.A., Vasudev, N.S. and Banks, R.E. (2013) Proteomics and the search for biomarkers for renal cancer. *Clinical Biochemistry*, 46 (6): 456–465. doi:10.1016/j.clinbiochem.2012.11.029.

Dagher, J., Delahunt, B., Rioux-Leclercq, N., et al. (2017) Clear cell renal cell carcinoma: validation of World Health Organization/International Society of Urological Pathology grading. *Histopathology*, 71 (6): 918–925. doi:10.1111/his.13311.

Daniel, C.R., Schwartz, K.L., Colt, J.S., et al. (2011) Meat-cooking mutagens and risk of renal cell carcinoma. *British Journal of Cancer*, 105 (7): 1096–1104. doi:10.1038/bjc.2011.343.

Derda, R., Tang, S.K.Y., Li, S.C., et al. (2011) Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules*, 16 (2): 1776–1803. doi:10.3390/molecules16021776.

Desmetz, C., Maudelonde, T., Mangé, A., et al. (2009) Identifying autoantibody signatures in cancer: A promising challenge. *Expert Review of Proteomics*, 6 (4): 377–386. doi:10.1586/epr.09.56.

Diamandis, E.P. (2014) Present and future of cancer biomarkers. *Clinical Chemistry and Laboratory Medicine*, 52 (6): 791–794. doi:10.1515/cclm-2014-0317.

Dias-Neto, E., Nunes, D.N., Giordano, R.J., et al. (2009) Next-generation phage display: Integrating and comparing available molecular tools to enable costeffective high-throughput analysis. *PLoS ONE*, 4 (12). doi:10.1371/journal.pone.0008338.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., et al. (2014) Ten years of next-

generation sequencing technology. *Trends in Genetics*, 30 (9): 418–426. doi:10.1016/j.tig.2014.07.001.

Dooley, H., Flajnik, M.F. and Porter, A.J. (2003) Selection and characterization of naturally occurring single-domain (IgNAR) antibody fragments from immunized sharks by phage display. *Molecular Immunology*, 40 (1): 25–33. doi:10.1016/S0161-5890(03)00084-1.

Doseeva, V., Colpitts, T., Gao, G., et al. (2015) Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer. *Journal of Translational Medicine*, 13 (1): 1–15. doi:10.1186/s12967-015-0419-y.

Eftimie, R. and Hassanein, E. (2018) Improving cancer detection through combinations of cancer and immune biomarkers: A modelling approach. *Journal of Translational Medicine*, 16 (1): 1–16. doi:10.1186/s12967-018-1432-8.

Eggener, S., Karsh, L.I., Richardson, T., et al. (2019) A 17-gene Panel for Prediction of Adverse Prostate Cancer Pathologic Features: Prospective Clinical Validation and Utility. *Urology*, 126: 76–82. doi:10.1016/j.urology.2018.11.050.

Etzioni, R., Urban, N., Ramsey, S., et al. (2003) The case for early detection. *Nature Reviews Cancer*, 3 (4): 243–252. doi:10.1038/nrc1041.

Felici, F., Castagnoli, L., Musacchio, A., et al. (1991) Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. *Journal of Molecular Biology*, 222 (2): 301–310. doi:10.1016/0022-2836(91)90213-P.

Fischer, K., Theil, G., Hoda, R., et al. (2012) Serum amyloid A: A biomarker for renal cancer. *Anticancer Research*, 32 (5): 1801–1804.

Fu, L., Wang, G., Shevchuk, M.M., et al. (2011) Generation of a mouse model of Von Hippel-Lindau kidney disease leading to renal cancers by expression of a constitutively active mutant of HIF1α. *Cancer Research*, 71 (21): 6848–6856. doi:10.1158/0008-5472.CAN-11-1745.

Füzéry, A.K., Levin, J., Chan, M.M., et al. (2013) Translation of proteomic biomaarkers into FDA approved cancer diagnostics: Issues and challenges. *Clinical Proteomics*, 10 (1). doi:10.1186/1559-0275-10-13.

Galán, A., Comor, L., Horvatić, A., et al. (2016) Library-based display technologies: Where do we stand? *Molecular BioSystems*, 12 (8): 2342–2358. doi:10.1039/c6mb00219f.

Georgieva, Y. and Konthur, Z. (2011) Design And Screening Of M13 phage display cDNA libraries. *Molecules*, 16 (2): 1667–1681. doi:10.3390/molecules16021667.

Gerstung, M., Jolly, C., Leshchiner, I., et al. (2020) The evolutionary history of

2 ,658 cancers. *Nature*, 578. Available at: https://doi.org/10.1101/161562.

Geysen, H.M., Rodda, S.J. and Mason, T.J. (1986) A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Molecular Immunology*, 23 (7): 709–715. doi:10.1016/0161-5890(86)90081-7.

Glanville, J., D'Angelo, S., Khan, T.A., et al. (2015) Deep sequencing in library selection projects: What insight does it bring? *Current Opinion in Structural Biology*, 33: 146–160. doi:10.1016/j.sbi.2015.09.001.

Golovastova, M.O., Korolev, D.O., Tsoy, L. V., et al. (2017) Biomarkers of Renal Tumors: the Current State and Clinical Perspectives. *Current Urology Reports*, 18 (1). doi:10.1007/s11934-017-0655-1.

Gough, K.C., Cockburn, W. and Whitelam, G.C. (1999) Selection of phage-display peptides that bind to cucumber mosaic virus coat protein. *Journal of Virological Methods*, 79 (2): 169–180. doi:10.1016/S0166-0934(99)00014-2.

Griffiths, a D., Williams, S.C., Hartley, O., et al. (1994) Isolation of high affinity human antibodies directly from large synthetic repertoires. *The EMBO journal*, 13 (14): 3245–60. doi:10.1016/0168-9525(94)90126-0.

Guo, M., Xu, L.M., Zhou, B., et al. (2014) Anchored periplasmic expression (APEx)-based bacterial display for rapid and high-throughput screening of B cell epitopes. *Biotechnology Letters*, 36 (3): 609–616. doi:10.1007/s10529-013-1400-6.

Halperin, R.F., Stafford, P. and Johnston, S.A. (2011) Exploring Antibody Recognition of Sequence Space through Random-Sequence Peptide Microarrays. *Molecular & Cellular Proteomics*, 10 (3): M110.000786. doi:10.1074/mcp.M110.000786.

Hammers, C.M. and Stanley, J.R. (2014) Antibody phage display: Technique and applications. *Journal of Investigative Dermatology*. doi:10.1038/jid.2013.521.

Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: The next generation. *Cell*, 144 (5): 646–674. doi:10.1016/j.cell.2011.02.013.

Haoran, L. and Tao Ye, Xiaoqi Yang, Peng Lv, Xiaoliang Wu, Hui Zhou, Jin Zeng, Kun Tang, Z.Y. (2020) A Panel of Four-lncRNA Signature as a Potential Biomarker for Predicting Survival in Clear Cell Renal Cell Carcinoma. *Journal of Cancer*, 11 (1): 4274–4283. doi:10.1186/s13046-015-0167-0.

Harvey, B.R., Georgiou, G., Hayhurst, A., et al. (2004) Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. *Proceedings of the National Academy of Sciences*, 101 (25): 9193–9198. doi:10.1073/pnas.0400187101.

Hay, I.D. and Lithgow, T. (2019) Filamentous phages: masters of a microbial sharing economy. *EMBO reports*, p. e47427. doi:10.15252/embr.201847427.

Henrion, M.Y.R., Purdue, M.P., Scelo, G., et al. (2015) Common variation at

1q24.1 (ALDH9A1) is a potential risk factor for renal cancer. *PLoS ONE*, 10 (3): 1–11. doi:10.1371/journal.pone.0122589.

Henry, N.L. and Hayes, D.F. (2012) Cancer biomarkers. *Molecular Oncology*, 6 (2): 140–146. doi:10.1016/j.molonc.2012.01.010.

Hoefflin, R., Harlander, S., Schäfer, S., et al. (2020) HIF-1α and HIF-2α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice. *Nature communications*, 11 (1): 4111. doi:10.1038/s41467-020-17873-3.

Hsieh, J.J., Purdue, M.P., Signoretti, S., et al. (2018) Renal cell carcinoma. *nature reviews disease primers*, 3 (17009). doi:10.1038/nrdp.2017.9.Renal.

Huang, J.J. and Hsieh, J.J. (2020) The Therapeutic Landscape of Renal Cell Carcinoma: From the Dark Age to the Golden Age. *Seminars in Nephrology*, 40 (1): 28–41. doi:10.1016/j.semnephrol.2019.12.004.

Huang, J.X., Bishop-Hurley, S.L. and Cooper, M.A. (2012) Development of Anti-Infectives Using Phage Display: Biological Agents against Bacteria, Viruses, and Parasites. *Antimicrobial Agents and Chemotherapy*, 56 (9): 4569–4582. doi:10.1128/aac.00567-12.

Hubbell, E., Clarke, C.A., Aravanis, A.M., et al. (2020) Modeled reductions in late-stage cancer with a multi-cancer early detection test. *Cancer Epidemiology, Biomarkers & Prevention*. doi:10.1158/1055-9965.EPI-20-1134.

Hughes, A.K., Cichacz, Z., Scheck, A., et al. (2012) Immunosignaturing can detect products from molecular markers in brain cancer. *PLoS ONE*, 7 (7): 1–7. doi:10.1371/journal.pone.0040201.

Hughes, R.A., Georgiou, G., Miklos, A.E., et al. (2010) Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nature Biotechnology*, 28 (9): 965–969. doi:10.1038/nbt.1673.

Ibsen, K.N. and Daugherty, P.S. (2017) Prediction of antibody structural epitopes via random peptide library screening and next generation sequencing. *Journal of Immunological Methods*, 451 (June): 28–36. doi:10.1016/j.jim.2017.08.004.

Iyer, N. V., Leung, S.W. and Semenza, G.L. (1998) The human hypoxia-inducible factor 1α gene: HIF1A structure and evolutionary conservation. *Genomics*, 52 (2): 159–165. doi:10.1006/geno.1998.5416.

Jia, L., Jia, Q., Shang, Y., et al. (2015) Vitamin C intake and risk of renal cell carcinoma: A meta-analysis. *Scientific Reports*, 5 (January 1980): 1–6. doi:10.1038/srep17921.

Jiang, D., Wang, Y., Liu, M., et al. (2019) A panel of autoantibodies against tumor-associated antigens in the early immunodiagnosis of lung cancer. *Immunobiology*. doi:10.1016/j.imbio.2019.09.007.

Johnston, S.A., Domenyuk, V., Gupta, N., et al. (2017) A Simple Platform for

the Rapid Development of Antimicrobials. *Scientific Reports*, 7 (1): 1–11. doi:10.1038/s41598-017-17941-7.

Junker, K., Ficarra, V., Kwon, E.D., et al. (2013) Potential role of genetic markers in the management of kidney cancer. *European Urology*, 63 (2): 333–340. doi:10.1016/j.eururo.2012.09.040.

Kabaria, R., Klaassen, Z. and Terris, M.K. (2016) Renal cell carcinoma: Links and risks. *International Journal of Nephrology and Renovascular Disease*, 9: 45–52. doi:10.2147/IJNRD.S75916.

Kapitsinou, P. and Haase, V. (2008) The VHL tumor suppressor and HIF: insights from genetic studies in mice. *Cell Death Differentiation*, 15 (4): 650–659. doi:10.1038/sj.cdd.4402313.The.

Kirwan, A., Utratna, M., O'Dwyer, M.E., et al. (2015) Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *BioMed Research International*, 2015. doi:10.1155/2015/490531.

Klatte, T., Rossi, S.H. and Stewart, G.D. (2018) Prognostic factors and prognostic models for renal cell carcinoma: a literature review. *World Journal of Urology*, 36 (12): 1943–1952. doi:10.1007/s00345-018-2309-4.

Kobayashi, M., Katayama, H., Fahrmann, J.F., et al. (2020) Development of autoantibody signatures for common cancers. *seminars in immunology*, 1044–5323 (January): 101388. doi:10.1016/j.smim.2020.101388.

Kobold, S., Lütkens, T., Cao, Y., et al. (2010) Autoantibodies against tumor-related antigens: Incidence and biologic significance. *Human Immunology*, 71 (7): 643–651. doi:10.1016/j.humimm.2010.03.015.

Kong, X., Carle, V., Cristina, D., et al. (2020) Generation of a Large Peptide Phage Display Library by Self- Ligation of Whole-Plasmid PCR Product. *ACS chemical biology*, 10 (1021). doi:10.1021/acschembio.0c00497.

Krištof Bozovicar and Tomaž Bratkoviˇ (2020) Evolving a Peptide : Library Platforms and Diversification Strategies. *international journal of molecular sciences*, 21 (215).

Kumar, R.M., Aziz, T., Jamshaid, H., et al. (2014) Metastatic renal cell carcinoma without evidence of a primary renal tumour. *Current Oncology*, 21 (3): 521–524. doi:10.3747/co.21.1914.

Lafata, J.E., Simpkins, J., Lamerato, L., et al. (2004) The economic impact of false-positive cancer screens. *Cancer Epidemiology Biomarkers and Prevention*, 13 (12): 2126–2132.

Larman, H.B., Zhao, Z., Laserson, U., et al. (2011) Autoantigen discovery with a synthetic human peptidome. *Nature Biotechnology*, 29 (6): 535–541. doi:10.1038/nbt.1856.

Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., et al. (2018) Basics of antibody phage display technology. *Toxins*, 10 (6). doi:10.3390/toxins10060236.

Lee, C.H., Romain, G., Yan, W., et al. (2017) IgG Fc domains that bind C1q but not effector Fc3 receptors delineate the importance of complement-mediated effector functions. *Nature Immunology*, 18 (8): 889–898. doi:10.1038/ni.3770.

Legutki, J.B., Magee, D.M., Stafford, P., et al. (2010) A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine*, 28 (28): 4529–4537. doi:10.1016/j.vaccine.2010.04.061.

Legutki, J.B., Zhao, Z.G., Greving, M., et al. (2014) Scalable high-density peptide arrays for comprehensive health monitoring. *Nature Communications*, 5: 1–7. doi:10.1038/ncomms5785.

Leow, C., Fischer, K., Leow, C., et al. (2017) Single Domain Antibodies as New Biomarker Detectors. *Diagnostics*, 7 (4): 52. doi:10.3390/diagnostics7040052.

Li, P., Shi, J.X., Xing, M.T., et al. (2017) Evaluation of serum autoantibodies against tumor-associated antigens as biomarkers in lung cancer. *Tumor Biology*, 39 (10): 1–10. doi:10.1177/1010428317711662.

Liao, L.M., Brennan, P., van Bemmel, D.M., et al. (2011) Line-1 methylation levels in leukocyte DNA and risk of renal cell cancer. *PLoS ONE*, 6 (11): 1–8. doi:10.1371/journal.pone.0027361.

Lichtenstein, A. (2017) Cancer: Bad luck or punishment? *Biochemistry (Moscow)*, 82 (1): 75–80. doi:10.1134/S0006297917010084.

Linehan, W.M. and Rathmell, W.K. (2012) Kidney cancer. *Urologic oncology*, 30 (6): 948–51. doi:10.1016/j.urolonc.2012.08.021.

Linehan, W.M. and Ricketts, C.J. (2019) The Cancer Genome Atlas of renal cell carcinoma: findings and clinical implications. *Nature Reviews Urology*. doi:10.1038/s41585-019-0211-5.

Liu, F., Wen, T., Tang, Q., et al. (2020) Impact of Vascular Endothelial Growth Factor Gene Polymorphisms and Their Interactions with Environmental Factors on Susceptibility to Renal Cell Carcinoma. *Nephron*, pp. 1–6. doi:10.1159/000505817.

Liu, H., Chumsae, C., Gaza-Bulseco, G., et al. (2010) Ranking the susceptibility of disulfide bonds in human IgG1 antibodies by reduction, differential alkylation, and LC-MS analysis. *Analytical Chemistry*, 82 (12): 5219–5226. doi:10.1021/ac100575n.

Liu, H., Zhang, J., Wang, S., et al. (2012) Screening of autoantibodies as potential biomarkers for hepatocellular carcinoma by using T7 phase display system. *Cancer Epidemiology*. doi:10.1016/j.canep.2011.04.001.

Liu, X., Sun, Q., Hou, H., et al. (2018) The association between BMI and kidney cancer risk. *Medicine*, 97 (44): e12860. doi:10.1097/md.0000000000012860.

Loh, B., Kuhn, A. and Leptihn, S. (2018) The fascinating biology behind phage display: Filamentous Phage Assembly. *Molecular Microbiology*, pp. 0–2.

doi:10.1111/mmi.14187.

Løset, G.Å. and Sandlie, I. (2012) Next generation phage display by use of pVII and pIX as display scaffolds. *Methods*, 58 (1): 40–46. doi:10.1016/j.ymeth.2012.07.005.

Lowman, H.B. (2013) *Encyclopedia of Biological Chemistry*.

Ludwig, J.A. and Weinstein, J.N. (2005) Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer*, 5 (11): 845–856. doi:10.1038/nrc1739.

Luzzago, A., Felici, F., Tramontano, A., et al. (1993) Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage library of constrained peptides. *Gene*, 128 (1): 51–57. doi:10.1016/0378-1119(93)90152-S.

Mantovani, A., Allavena, P., Sica, A., et al. (2008) Cancer-related inflammation. *Nature*, 454 (7203): 436–444. doi:10.1038/nature07205.

Marcatili, P., Zarebski, L., Jespersen, M.C., et al. (2018) Epitope Specific Antibodies and T Cell Receptors in the Immune Epitope Database. *Frontiers in Immunology*, 9 (November): 1–10. doi:10.3389/fimmu.2018.02688.

Matochko, W.L., Chu, K., Jin, B., et al. (2012) Deep sequencing analysis of phage libraries using Illumina platform. *Methods*, 58 (1): 47–55. doi:10.1016/j.ymeth.2012.07.006.

Matochko, W.L., Cory Li, S., Tang, S.K.Y., et al. (2014) Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Research*, 42 (3): 1784–1798. doi:10.1093/nar/gkt1104.

Mazor, Y., Van Blarcom, T., Carroll, S., et al. (2010) Selection of full-length IgGs by tandem display on filamentous phage particles and Escherichia coli fluorescence-activated cell sorting screening. *FEBS Journal*, 277 (10): 2291–2303. doi:10.1111/j.1742-4658.2010.07645.x.

Mazor, Y., Van Blarcom, T., Iverson, B.L., et al. (2008) E-clonal antibodies: Selection of full-length IgG antibodies using bacterial periplasmic display. *Nature Protocols*, 3 (11): 1766–1777. doi:10.1038/nprot.2008.176.

McShane, L.M., Altman, D.G., Sauerbrei, W., et al. (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). *British Journal of Cancer*. 93 (4) pp. 387–391. doi:10.1038/sj.bjc.6602678.

Mestas, J. and Hughes, C.C.W. (2004) Of Mice and Not Men: Differences between Mouse and Human Immunology. *The Journal of Immunology*, 172 (5): 2731–2738. doi:10.4049/jimmunol.172.5.2731.

Moch, H., Cubilla, A.L., Humphrey, P.A., et al. (2016) The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *European Urology*, 70 (1): 93–105. doi:10.1016/j.eururo.2016.02.029.

Moch, H., Srigley, J., Delahunt, B., et al. (2014) Biomarkers in renal cancer. *Virchows Archiv*, 464 (3): 359–365. doi:10.1007/s00428-014-1546-1.

Mordente, A., Meucci, E., Martorana, G.E., et al. (2015) Cancer Biomarkers Discovery and Validation:State of the Art, Problems and Future Perspectives. *Adv Exp Med Biol.*, 867: 229–44. doi:10.1007/978-94-017-7215-0.

Moritz, C.P., Paul, S., Stoevesandt, O., et al. (2019) Autoantigenomics : Holistic characterization of autoantigen repertoires for a better understanding of autoimmune diseases. *Autoimmunity Reviews*. doi:10.1016/j.autrev.2019.102450.

Murphy, K., Travers, P. and Walport, M. (2008) *Janeway's Immunobiology*. 7th ed.

Nabi, S., Kessler, E.R., Bernard, B., et al. (2018) Renal cell carcinoma: a review of biology and pathophysiology. *F1000Research*, 7 (0): 307. doi:10.12688/f1000research.13179.1.

National Cancer Institute (2019) *Tumor Markers in Common Use*.

Netea, M.G., Schlitzer, A., Placek, K., et al. (2019) Innate and Adaptive Immune Memory: an Evolutionary Continuum in the Host's Response to Pathogens. *Cell Host and Microbe*, 25 (1): 13–26. doi:10.1016/j.chom.2018.12.006.

Nissim, A., Hoogenboom, H.R., Tomlinson, I.M., et al. (1994) Antibody fragments from a "single pot" phage display library as immunochemical reagents. *EMBO Journal*, 13 (3): 692–698. doi:10.1002/j.1460-2075.1994.tb06308.x.

Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M., et al. (2015) The generation of antibody-secreting plasma cells. *Nature Reviews Immunology*, 15 (3): 160–171. doi:10.1038/nri3795.

O'Connell, G.C., Stafford, P., Walsh, K.B., et al. (2019) High-Throughput Profiling of Circulating Antibody Signatures for Stroke Diagnosis Using Small Volumes of Whole Blood. *Neurotherapeutics*, 16 (3): 868–877. doi:10.1007/s13311-019-00720-9.

Oosterwijk, E. (2011) Basic research in Kidney Cancer. *European Urology*, 60 (4): 622–633. doi:10.1016/j.physbeh.2017.03.040.

Oyama, T., Sykes, K.F., Samli, K.N., et al. (2003) Isolation of lung tumor specific peptides from a random peptide library: Generation of diagnostic and cell-targeting reagents. *Cancer Letters*, 202 (2): 219–230. doi:10.1016/j.canlet.2003.08.011.

Pastore, A.L., Palleschi, G., Silvestri, L., et al. (2015) Serum and urine biomarkers for human renal cell carcinoma. *Disease Markers*, 2015. doi:10.1155/2015/251403.

Patel, P.H., Chadalavada, R.S.V., Chaganti, R.S.K., et al. (2006) Targeting von

Hippel-Lindau pathway in renal cell carcinoma. *Clinical Cancer Research*, 12 (24): 7215–7220. doi:10.1158/1078-0432.CCR-06-2254.

Pedersen, J.W. and Wandall, H.H. (2011) Autoantibodies as Biomarkers in Cancer. *Laboratory Medicine*, 42 (10): 623–628. doi:10.1309/LM2T3OU3RZRTHKSN.

Penticuff, J.C. and Kyprianou, N. (2015) Therapeutic challenges in renal cell carcinoma. *American journal of clinical and experimental urology*, 3 (2): 77–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26309897%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4539109.

Pinto, A., Chen, S.X. and Zhang, D.Y. (2018) Simultaneous and stoichiometric purification of hundreds of oligonucleotides. *Nature Communications*, 9 (1): 1–9. doi:10.1038/s41467-018-04870-w.

van de Pol, J.A.A., van den Brandt, P.A., van Engeland, M., et al. (2020) Germline polymorphisms in the Von Hippel-Lindau and Hypoxia-inducible factor 1-alpha genes, gene-environment and gene-gene interactions and renal cell cancer. *Scientific Reports*, 10 (1): 1–9. doi:10.1038/s41598-019-56980-0.

Postel, M., Roosen, A., Laurent-Puig, P., et al. (2018) Droplet-based digital PCR and next generation sequencing for monitoring circulating tumor DNA: a cancer diagnostic perspective. *Expert Review of Molecular Diagnostics*, 18 (1): 7–17. doi:10.1080/14737159.2018.1400384.

Potocnakova, L., Bhide, M. and Pulzova, L.B. (2016) An Introduction to B-Cell Epitope Mapping and in Silico Epitope Prediction. *Journal of Immunology Research*, 2016. doi:10.1155/2016/6760830.

Qiu, C., Wang, P., Wang, B., et al. (2020) Establishment and validation of an immunodiagnostic model for prediction of breast cancer. *OncoImmunology*, 9 (1). doi:10.1080/2162402X.2019.1682382.

Rahbarnia, L., Farajnia, S., Babaei, H., et al. (2017) Evolution of phage display technology: from discovery to application. *Journal of Drug Targeting*, 25 (3): 216–224. doi:10.1080/1061186X.2016.1258570.

Rajandram, R., Perumal, K. and Yap, N.Y. (2019) Prognostic biomarkers in renal cell carcinoma: Is there a relationship with obesity? *Translational Andrology and Urology*, 8 (7): S138–S146. doi:10.21037/tau.2018.11.10.

Rajandram, R., Yap, N.Y., Pailoor, J., et al. (2014) Tumour necrosis factor receptor-associated factor-1 (TRAF-1) expression is increased in renal cell carcinoma patient serum but decreased in cancer tissue compared with normal: Potential biomarker significance. *Pathology*, 46 (6): 518–522. doi:10.1097/PAT.0000000000000145.

Ramesh, B., Frei, C.S., Cirino, P.C., et al. (2015) Functional enrichment by direct plasmid recovery after fluorescence activated cell sorting.

BioTechniques, 59 (3): 157–161. doi:10.2144/000114329.

Reineke, U., Ivascu, C., Schlief, M., et al. (2002) Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *Journal of Immunological Methods*, 267 (1): 37–51. doi:10.1016/S0022-1759(02)00139-4.

Rentero Rebollo, I., Sabisz, M., Baeriswyl, V., et al. (2014) Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic acids research*, 42 (22): e169. doi:10.1093/nar/gku940.

Restrepo, L., Stafford, P. and Johnston, S.A. (2013) Feasibility of an early Alzheimer's disease immunosignature diagnostic test. *Journal of Neuroimmunology*, 254 (1–2): 154–160. doi:10.1016/j.jneuroim.2012.09.014.

Restrepo, L., Stafford, P., Magee, D.M., et al. (2011) Application of immunosignatures to the assessment of Alzheimer's disease. *Annals of Neurology*, 70 (2): 286–295. doi:10.1002/ana.22405.

Reuschenbach, M., Von Knebel Doeberitz, M. and Wentzensen, N. (2009) A systematic review of humoral immune responses against tumor antigens. *Cancer Immunology, Immunotherapy*, 58 (10): 1535–1544. doi:10.1007/s00262-009-0733-4.

Richard, S., Gardie, B., Couvé, S., et al. (2013) Von Hippel-Lindau: How a rare disease illuminates cancer biology. *Seminars in Cancer Biology*, 23 (1): 26–37. doi:10.1016/j.semcancer.2012.05.005.

Richer, J., Johnston, S.A. and Stafford, P. (2015) Epitope Identification from Fixed-complexity Random-sequence Peptide Microarrays. *Molecular & Cellular Proteomics*, 14 (1): 136–147. doi:10.1074/mcp.m114.043513.

Ridge, C.A., Pua, B.B. and Madoff, D.C. (2014) Epidemiology and staging of renal cell carcinoma. *Seminars in Interventional Radiology*, 31 (1): 3–8. doi:10.1055/s-0033-1363837.

Riera Romo, M., Pérez-Martínez, D. and Castillo Ferrer, C. (2016) Innate immunity in vertebrates: An overview. *Immunology*, 148 (2): 125–139. doi:10.1111/imm.12597.

Rodi, D.J. and Makowski, L. (1999) Phage-display technology - Finding a needle in a vast molecular haystack. *Current Opinion in Biotechnology*, 10 (1): 87–93. doi:10.1016/S0958-1669(99)80016-0.

Rouet, R., Jackson, K.J.L., Langley, D.B., et al. (2018) Next-Generation Sequencing of Antibody Display Repertoires. *Frontiers in Immunology*, 9 (February): 1–5. doi:10.3389/fimmu.2018.00118.

Rozenblatt-Rosen, O., Regev, A., Oberdoerffer, P., et al. (2020) The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell*, 181 (2): 236–249. doi:10.1016/j.cell.2020.03.053.

Russel, M., Lowman, H.B. and Clackson, T. (2004) Introduction to phage biology and phage display. *Phage Display: A Practical Approach*, (January 2004): 332. Available at: http://books.google.com/books?hl=en&lr=&id=CRgfEbXfiAUC&pgis=1.

Ryvkin, A., Ashkenazy, H., Smelyanski, L., et al. (2012) Deep panning: Steps towards probing the IgOme. *PLoS ONE*, 7 (8): 1–11. doi:10.1371/journal.pone.0041469.

Ryvkin, A., Ashkenazy, H., Weiss-Ottolenghi, Y., et al. (2018) Phage display peptide libraries: deviations from randomness and correctives. *Nucleic Acids Research*, (May): 1–10. doi:10.1093/nar/gky077.

Saad, A.M., Gad, M.M., Al-Husseini, M.J., et al. (2019) Trends in Renal-Cell Carcinoma Incidence and Mortality in the United States in the Last 2 Decades: A SEER-Based Study. *Clinical Genitourinary Cancer*, 17 (1): 46-57.e5. doi:10.1016/j.clgc.2018.10.002.

Salema, V. and Fernández, L.Á. (2017) Escherichia coli surface display for the selection of nanobodies. *Microbial Biotechnology*, 10 (6): 1468–1484. doi:10.1111/1751-7915.12819.

Schiffman, J.D., Fisher, P.G. and Gibbs, P. (2015) Early Detection of Cancer : Past, Present, and Future. *Cancer Screening amd Surveillance*. doi:10.14694/EdBook_AM.2015.35.57.

Seretis, A., Cividini, S., Markozannes, G., et al. (2019) Association between blood pressure and risk of cancer development: a systematic review and meta-analysis of observational studies. *Scientific Reports*, 9 (1): 1–12. doi:10.1038/s41598-019-45014-4.

Sidransky, D. (2002) Emerging molecular markers of cancer. *Nature Reviews Cancer*, 2 (3): 210–219. doi:10.1038/nrc755.

Sieber, T., Hare, E., Hofmann, H., et al. (2015) Biomathematical description of synthetic peptide libraries. *PLoS ONE*, 10 (6): 1–20. doi:10.1371/journal.pone.0129200.

Siegel, R.L., Miller, K.D. and Jemal, A. (2020) Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70 (1): 7–30. doi:10.3322/caac.21590.

Smeal, S.W., Schmitt, M.A., Pereira, R.R., et al. (2017a) Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. *Virology*, 500: 259–274. doi:10.1016/j.virol.2016.08.017.

Smeal, S.W., Schmitt, M.A., Pereira, R.R., et al. (2017b) Simulation of the M13 life cycle II: Investigation of the control mechanisms of M13 infection and establishment of the carrier state. *Virology*, 500: 275–284. doi:10.1016/j.virol.2016.08.015.

Smith, G. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228 (4705): 1315–

1317. doi:10.1126/science.4001944.

Solemani Zadeh, A., Grässer, A., Dinter, H., et al. (2019) Efficient Construction and Effective Screening of Synthetic Domain Antibody Libraries. *Methods and Protocols*, 2 (1): 17. doi:10.3390/mps2010017.

Sormanni, P., Aprile, F.A. and Vendruscolo, M. (2018) Third generation antibody discovery methods: In silico rational design. *Chemical Society Reviews*, 47 (24): 9137–9157. doi:10.1039/c8cs00523k.

Spiliotopoulos, A., Owen, J.P., Maddison, B.C., et al. (2015) Sensitive recovery of recombinant antibody clones after their in silico identification within NGS datasets. *Journal of Immunological Methods*, 420: 50–55. doi:10.1016/j.jim.2015.03.005.

Stafford, P., Cichacz, Z., Woodbury, N.W., et al. (2014) Immunosignature system for diagnosis of cancer. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1409432111.

Stafford, P., Halperin, R., Legutki, J.B., et al. (2012) Physical Characterization of the "Immunosignaturing Effect." *Molecular & Cellular Proteomics*, 11 (4): M111.011593. doi:10.1074/mcp.m111.011593.

Stafford, P. and Johnston, S. (2011) Microarray technology displays the complexities of the humoral immune response. *Expert Review of Molecular Diagnostics*, 11 (1): 5–8. doi:10.1586/erm.10.113.

Stafford, P., Johnston, S.A., Kantarci, O.H., et al. (2020) Antibody characterization using immunosignatures. *PLoS ONE*, 15 (3): 1–18. doi:10.1371/journal.pone.0229080.

Stafford, P., Wrapp, D. and Johnston, S.A. (2016) General Assessment of Humoral Activity in Healthy Humans. *Molecular & Cellular Proteomics*, 15 (5): 1610–1621. doi:10.1074/mcp.m115.054601.

Sykes, K.F., Legutki, J.B. and Stafford, P. (2013) Immunosignaturing : a critical review. *Cell*, 31 (1): 45–51.

Szardenings, M. (2003) Phage Display of Random Peptide Libraries: Applications, Limits, and Potential. *Journal of Receptors and Signal Transduction*, 23 (4): 307–349. doi:10.1081/RRS-120026973.

Tanaka, T., Kitamura, H., Torigoe, T., et al. (2011) Autoantibody against hypoxia-inducible factor prolyl hydroxylase-3 is a potential serological marker for renal cell carcinoma. *Journal of Cancer Research and Clinical Oncology*, 137 (5): 789–794. doi:10.1007/s00432-010-0940-6.

Tighe, P.J., Ryder, R.R., Todd, I., et al. (2015) ELISA in the multiplex era: Potentials and pitfalls. *Proteomics - Clinical Applications*, 9 (3–4): 406–422. doi:10.1002/prca.201400130.

Ulrich Reineke* and Mike Schutkowski (2009) *epitope mapping protocols, 2nd edition*. (koita afto:Monitoring B Cell Response to Immunoselected Phage-

Displayed Peptides by Microarrays). doi:10.1016/B978-0-12-374407-4.00303-4.

Ushigome, M., Nabeya, Y., Soda, H., et al. (2018) Multi - panel assay of serum autoantibodies in colorectal cancer. *International Journal of Clinical Oncology*, (0123456789). doi:10.1007/s10147-018-1278-3.

Vesely, M.D., Kershaw, M.H., Schreiber, R.D., et al. (2011) Natural Innate and Adaptive Immunity to Cancer. *Annual Review of Immunology*, 29 (1): 235–271. doi:10.1146/annurev-immunol-031210-101324.

Vieira, A.F. and Schmitt, F. (2018) An update on breast cancer multigene prognostic tests-emergent clinical biomarkers. *Frontiers in Medicine*, 5 (SEP): 1–12. doi:10.3389/fmed.2018.00248.

Voss, M.H., Reising, A., Cheng, Y., et al. (2018) Genomically annotated risk model for advanced renal-cell carcinoma: a retrospective cohort study. *The Lancet Oncology*, 19 (12): 1688–1698. doi:10.1016/S1470-2045(18)30648-X.

White, N.M.A., Masui, O., DeSouza, L. V., et al. (2014) Quantitative proteomic analysis reveals potential diagnostic markers and pathways involved in pathogenesis of renal cell carcinoma. *Oncotarget*, 5 (2): 506–518. doi:10.18632/oncotarget.1529.

Williams, P.T. (2014) Reduced Risk of Incident Kidney Cancer from Walking and Running. *Medicine & Science in Sports & Exercise*, 46 (2): 312–317. doi:10.1038/jid.2014.371.

Wong, M.C.S., Goggins, W.B., Yip, B.H.K., et al. (2017) Incidence and mortality of kidney cancer: Temporal patterns and global trends in 39 countries. *Scientific Reports*, 7 (1): 1–10. doi:10.1038/s41598-017-15922-4.

Wu, C.-H., Liu, I.-J., Lu, R.-M., et al. (2016) Advancement and applications of peptide phage display technology in biomedical science. *Journal of Biomedical Science*, 23 (1): 8. doi:10.1186/s12929-016-0223-x.

Wu, F., Lai, D., Ding, H., et al. (2019) Identification of serum biomarkers for Systemic Lupus Erythematosus using a library of phage displayed random peptides and deep sequencing. *molecular cell proteomics*.

Wu, L. and Qu, X. (2015) Cancer biomarker detection: Recent achievements and challenges. *Chemical Society Reviews*, 44 (10): 2963–2997. doi:10.1039/c4cs00370e.

Xu, X., Zhu, Y., Zheng, X., et al. (2015) Does beer, wine or liquor consumption correlate with the risk of renal cell carcinoma? A dose-response meta-analysis of prospective cohort studies. *Oncotarget*, 6 (15): 13347–13358. doi:10.18632/oncotarget.3749.

Yadav, S., Kashaninejad, N., Masud, M.K., et al. (2019) Autoantibodies as diagnostic and prognostic cancer biomarker: Detection techniques and approaches. *Biosensors and Bioelectronics*, 139 (March): 111315.

doi:10.1016/j.bios.2019.111315.

Yilmaz, E., Çekmen, A., Akkuş, E., et al. (2010) The relationship between human leukocyte antigens (HLA) and renal cell carcinoma. *Bosnian Journal of Basic Medical Sciences*, 10 (4): 282–286. doi:10.17305/bjbms.2010.2661.

Yiyu, D. (2019) Hyperpolarized MRI visualizes Warburg effects and predicts treatment response to mTOR inhibitors in patient-derived ccRCC xenograft models. *Cancer Research*, 79 (1): 242–250. doi:10.1158/0008-5472.CAN-18-2231.

Zaenker, P., Gray, E.S. and Ziman, M.R. (2016) Autoantibody Production in Cancer-The Humoral Immune Response toward Autologous Antigens in Cancer Patients. *Autoimmunity Reviews*, 15 (5): 477–483. doi:10.1016/j.autrev.2016.01.017.

Zaenker, P. and Ziman, M.R. (2013) Serologic autoantibodies as diagnostic cancer biomarkers - A review. *Cancer Epidemiology Biomarkers and Prevention*, 22 (12): 2161–2181. doi:10.1158/1055-9965.EPI-13-0621.

Zhang, H., Torkamani, A., Jones, T.M., et al. (2011) Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems. *Proceedings of the National Academy of Sciences*, 108 (33): 13456–13461. doi:10.1073/pnas.1111218108.