An Immunoproteomic Approach to Identifying Cancer-Associated Autoantibody Biomarkers

Muneerah Smith

University of Cape Town

Thesis presented for the degree of Doctor of Philosophy

In the Department of Integrative Biomedical Sciences

University of Cape Town

June 2018

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the American Chemical Society convention for citation and referencing. Each contribution to, and in quotation, this thesis from the work(s) of other people have been attributed, and has been cited and referenced.
3. This thesis is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
5. I acknowledge that copying some else's assignment or essay, or part of it, is wring and declare that this is my own work.

Signature.....

Acknowledgements

My sincere gratitude and thanks to my supervisor, Prof. Jonathan Blackburn, for sharing his invaluable knowledge and guidance, whilst also allowing me to explore my own research ideas. I also thank him for being an excellent mirror and motivator for mental health in a sometimes overwhelming field.

Thank you to each patient and healthy volunteer within our cohorts for their selfless contributions toward improving our understanding - and allowing the exploration - of cancer biomarkers. My thanks to our collaborators within UCT and at the ONJCRI, who have allowed an extended capacity for cancer biomarker exploration. Thanks to each of my funders, including the National Research Foundation (NRF), the Harry Crossley Foundation, Marion Beatrice Waddel and the University of Cape Town, for generously funding me throughout my degree.

A special thanks to all members from the JMB lab, past and present, for their humour, kindness and knowledge; you have inspired growth in all aspects of my life. Kitten.

To my ever-nurturing Mother and Father – thank you for your unwavering faith in me. I thank you both for teaching me to pursue what I love.

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
Ab	Antibody
AGC	Automatic gain control
ANA	Anti-nuclear antibodies
ANOVA	Analysis of variance
AP-MS	Affinity purification-mass spectrometry
APCs	Antigen presenting cells
ASR	Age-specific rates
ATP	Adenosine triphosphate
AUC	Area under the receiver operator characteristic curve
BCCP	Biotin carboxyl carrier protein
BLAST	Basic local alignment search tool
BMI	Body mass index
BSA	Bovine serum albumin
BPG	Biotin-PEG_2-4-glyoxalbenzoic acid
°C	Degree Celsius
CBC	Complete blood count
CEA	Carcinoembryonic antigen
CID	Collision-induced diffraction

CPGR	Centre of Proteomic and Genomic Research
CRAN	Comprehensive R Archive Network
CRC	Colorectal cancer
CT100plus	Cancer-testis antigen microarray
CTA	Cancer-testis antigen
CTLA-4	Cytotoxic T-lymphocyte-associated antigen 4
CV	Coefficient of variance
Cy5	Cyanine 5
Cy3	Cyanine 3
Da	Daltons
DCs	Dendritic cells
DLBCL	Diffuse large B-cell lymphoma
DMP	Dimethyl pimelimidate dihydrochloride
ECD	Electron capture dissociation
E.coli	*Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-to-mesenchymal transition
ESI	Electrospray ionization
FAP	Familial adenomatous polyps
Fold-Δ	Fold-change
FP	Free protein
FPW	Free protein in wash

FTIC	Fourier transform ion cyclotron
FWER	Family wise error rate
GC	Germinal centre
HC	Healthy control
HCs	Healthy controls
HCD	High-energy collision dissociation
HNPCC	Hereditary nonpolyposis colorectal cancer
HPLC	High-performance liquid chromatography
iBAQ	Intensity-based absolute quantification
ICL	Insect cell lysate
IFN-γ	Interferon-γ
Igs	Immunoglobulins
irAEs	Immune-related adverse events
iTRAQ	Isobaric tag for relative and absolute quantitation
kDa	Kilodaltons
MAD	Median absolute deviation
MALDI	Matrix assisted laser desorption ionization
m/z	Mass-to-charge ratio
MC	Mast cells
ms	millisecond
MS	Microsatellite
MSI	Microsatellite instability

MSI-H	High microsatellite instability
MSI-L	Low microsatellite instability
mg/ml	Milligrams per millilitre
µg/ml	Micrograms per millilitre
µg	Micrograms
µl	Microliter
ms	Milliseconds
ng/µl	Nanograms per microliter
nm	Nanometres
NCBI	National Centre for Biotechnology Information
NK	Natural killer cells
NSCLC	Non-small cell lung cancer
N/A	Not available
N	Number of patients
ONJCRI	Olivia Newton-John Cancer Research Institute
PAD	Protein arginine deiminase
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline supplemented with Tween®-20
PCA	Principle component analysis
PD-1	Programmed death 1
PD-L1	Programmed death ligand 1
PD-L2	Programmed death ligand 2

PDB	Protein data bank
PEG	Polyethylene glycol
PEP	Posterior error of probability
pg/ml	Picogram per millilitre
PMT	Photomultiplier tube
P. pastoris	*Pichia pastoris*
PSA	Prostate specific antigen
PSC	Post source decay
PTM	Post-translational modification
PTMs	Post-translational modifications
RA	Rheumatoid arthritis
RB	Retinoblastoma-associated
RCC	Renal cell carcinoma
RFU	Relative fluorescence unit
RFUs	Relative fluorescence units
ROC	Receiver operator characteristic
RT	Room temperature
S. cerevisiae	*Saccharomyces cerevisiae*
SCC	Squamous cell carcinoma
SD	Standard deviations
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling by amino acids in cell culture

SLE	Systemic Lupus Erythematosus
SSA	Sjogren's syndrome A
SSB	Sjogren's syndrome B
STRING	Search tool for the retrieval of interacting genes/proteins
SEREX	Serological analysis of recombinant cDNA expression libraries
SERPA	Serological proteome analysis
SD	Standard deviation
SDs	Standard deviations
SILAC	Stable isotope labelling by amino acids in cell culture
TAA	Tumour-associated antigen
TAAs	Tumour-associated antigens
TMT	Tandem mass tags
TNF	Tumour necrosis factor
TNM	Tumour, node, metastasis
TOF	Time-of-flight
TP	Timepoint
TPs	Timepoints
T_{regs}	Regulatory T-cells
TSA	Tumour-specific antigen
TSAs	Tumour-specific antigens
TSP-1	Thrombospondin-1
UV	Ultraviolet

U/µl	Unit per microliter
VEGF-A	Vascular endothelial growth factor-A
WHO	World Health Organisation

Table of Contents

List of Equations

List of Figures

List of Tables

Abstract

Cancer is a heterogenous disease capable of forming and spreading in most tissues of the human body. Cancer screening and diagnosis can be performed through medical procedures, which are highly invasive, requiring an intensive infrastructure. It is therefore important to create cost-effective, non-invasive cancer diagnostic tools that also gives an indication of disease prognosis. With this in mind, the Blackburn lab previously created a cancer-testis antigen microarray (CT100plus) functionalised with tumour-associated and tumour-specific antigens, capable of detecting plasma- or serum-derived autoantibodies in the picogram per millilitre (pg/ml) range.

In this thesis, a newly established statistical pipeline was used to analyse colorectal cancer (CRC) patient-derived CT100plus data. Using the pipeline, the 10 antigens with the highest receiver operator characteristic (ROC)-derived area under the ROC curve (AUC)-values were identified as potential autoantibody-based biomarkers. The top 10 antigen biomarker candidates include CEACAM 1, COL6A1, GRWD1, MAGEA1, MAGEA5, MAGEA10, NY-CO-1, SGY-1, SPANXB1 and THEG. Using these biomarker candidates, distinct clusters of healthy controls (HCs) and CRC patients were observed using both unsupervised hierarchical clustering and principle component analysis (PCA) analysis. Combinatorial ROC analysis indicates that CEACAM1 and GRWD1 as the top autoantigen combination for CRC diagnosis, together producing sensitivity-, and specificity-, and AUC-values of 1.00, 0.77 and 0.94, respectively. Furthermore, other top autoantigens, including COL6A1, THEG and CEACAM7, a homologue of CEACAM1, were also identified in this thesis by affinity purification-mass spectrometry (AP-MS) for patients from the same cohort, providing supporting evidence that these antigens are associated with CRC.

The CT100plus microarray content was enzymatically modified to include citrullinated proteins, with the subsequent assessment of CRC patient autoantibody response. Significantly (p-value ≤ 0.05; adjusted p-value ≤ 0.05) higher signal intensities were detected in CRC patients *versus* HCs for citrullinated CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1, whereas no significant (adjusted p-value > 0.05) difference in autoantibody signal was detected for these autoantigens on the non-citrullinated microarray for the same patient cohort. ROC analyses of these antigens resulted in

an AUC-, sensitivity- and specificity-values of 0.91, 0.87 and 0.89, respectively. Together, this thesis shows for the first time that cancer patients elicit an autoantibody response to citrullinated proteins, resulting in potential novel CRC biomarkers.

A novel AP-MS assay was developed to detect autoantibody responses to autologous native CRC tissue proteins. Using the optimised methodology, proteins or homologues of proteins that were significantly (> cut-off value) detected on the CT100plus microarray for the same 5 patients were re-identified by the orthogonal AP-MS method. Using the methodology, PAD2, an enzyme that catalyses the conversion of arginine to citrulline was also identified. In addition, citrullinated antigens associated with cancer were identified, including homologues of CDK7 and MAGEB supporting the conclusion that citrullinated homologues of these proteins induce an autoantibody response in CRC patients.

Finally, serum and/or plasma samples of a cohort melanoma patients were analysed using the CT100plus microarray, and autoantibody signals were compared to those of healthy control (HC) samples. Using the established statistical pipeline, the 10 antigens with the highest ROC-derived AUC-values were identified as potential biomarkers. The top 10 biomarker autoantigen candidates for melanoma included CEACAM 1, DPPA2, FGFR2, ITGB1, MAGEA10, NANOG, PIM1, SPANXB1, THEG and XAGE1B. Using these biomarker candidates, distinct clusters of HCs and melanoma patients were identified in both unsupervised hierarchical clustering and PCA analysis. Combinatorial ROC analysis indicates that CEACAM1 and FGFR2 were identified as the top antigens for melanoma diagnosis, together producing sensitivity-, and specificity-, and AUC-values of 0.96, 0.94 and 0.93, respectively.

In conclusion, a statistical pipeline was established for microarray data, enabling the identification of potential antigens associated with cancer diagnosis, and the potential to determine disease prognosis. Using the established pipeline, cancer antigens associated with CRC and melanoma were identified. In addition, an AP-MS assay was developed for the identification of known and novel cancer antigen that can be added to the customisable CT100plus microarray.

Chapter 1

Introduction

1.1. Hallmarks of cancer

Cancer is a heterogenous disease capable of forming and spreading almost anywhere in the body. The disease is complex, where healthy cells encounter multiple insults before progression to a state of malignancy. Several characteristics mark a cell as cancerous; these characteristics have been termed the six hallmarks of cancer by Hanahan & Weinberg (Hanahan and Weinberg, 2000) (Figure 1.1).

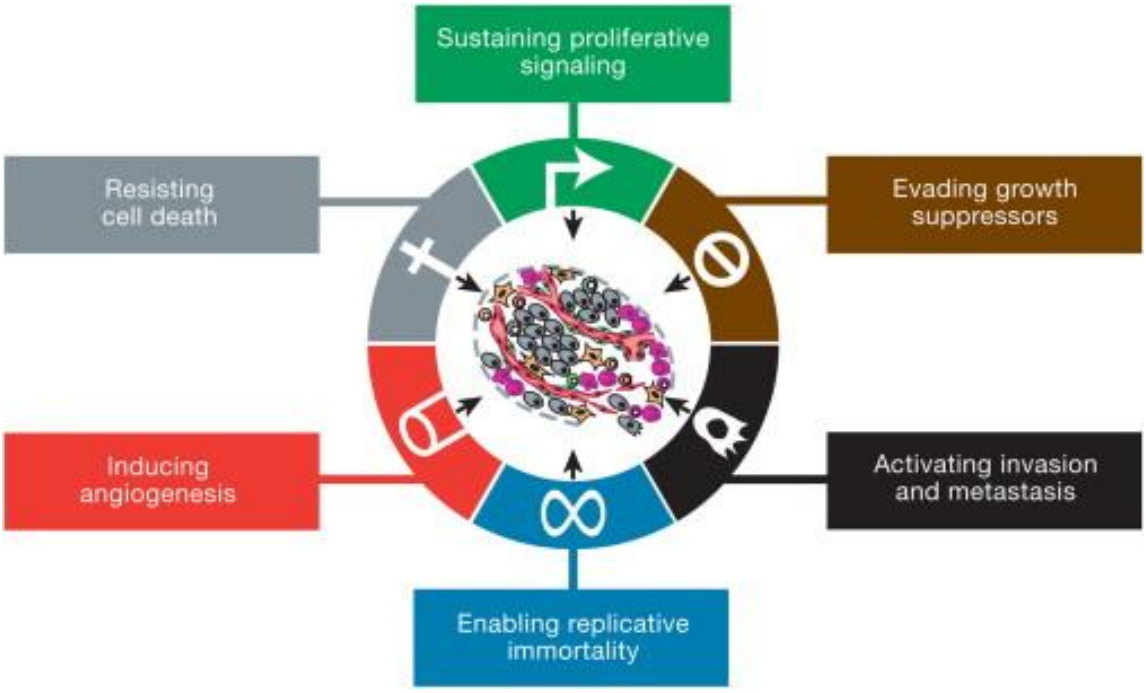


Figure 1.1. The six hallmarks of cancer. The six hallmarks of cancer provide a logical framework for the altered biological properties acquired by cells during tumour development. The six hallmarks include: resistance to cell death, sustained proliferative signalling, evasion of growth suppressors, evasion of metastasis, enabled replicative immortality, and the induction of angiogenesis [Adapted from (Hanahan and Weinberg, 2000)].

The hallmarks, which provide a logical outline for understanding the disruption of tissue homeostasis leading to cancer development, can be summarized as follows:

1. Sustained proliferative signalling: Healthy cells control the production and release of growth-promoting signals that instruct cells to undergo cell division. In cancer, these mitogenic signals are disrupted to sustain proliferative signalling either through autocrine proliferative signalling; growth factor-supply from normal cells (Bhowmick et al., 2004; Cheng et al., 2008); disrupting negative-feedback mechanisms that attenuate proliferative signalling (O'Reilly et al., 2006; Yuan and Cantley, 2008; Jiang and Liu, 2009; Sudarsanam and Johnson, 2010); upregulating receptor expression or augmenting protein structure to enhance growth factor binding; and/or growth factor-independent proliferation through pathways downstream to the receptor (Yuan and Cantley, 2008; Jiang and Liu, 2009; Davies and Samuels, 2010).

2. Evasion of growth suppressors: Complementary to sustained proliferative signalling, cancer cells additionally acquire the ability to evade growth suppressors. Tumour suppressors are inactivated in many cancers, including the prototypical tumour suppressors, retinoblastoma-associated (RB) protein and p53, which act as central nodes with complementary cellular mechanisms that govern cellular proliferation.

3. Activation of invasion and metastasis: Once the cancer has metastasised, it is almost impossible to treat. Cancer invasion and metastasis is a multistep process termed the invasion-metastasis cascade, depicted as local invasion, cancer cell intravasation into nearby blood and lymphatic vessels, transit of cancer cells through lymphatic and hematogenous systems, cancer cell escape from the lumina, followed by the formation of small cancer cell nodules, and finally growth of metastatic lesions (Fidler, 2003; Talmadge and Fidler, 2010). The invasion-metastasis cascade results from changes in cancer cell morphology and their attachment to other cells, caused by the loss of E-cadherin, a key cell-to-cell adhesion molecule (Cavallaro and Christofori, 2004; Berx and van Roy, 2009).

4. Resistance to cell death: In healthy cells, apoptosis serves as a natural barrier to cancer development by balancing pro- and anti-apoptotic proteins (Evan, 1998; Lowe et al., 2004; Adams and Cory, 2007). Tumour cells develop strategies that circumvent apoptosis through the loss of

the p53 tumour suppressor gene; increased expression of anti-apoptotic proteins, Bcl-2 and Bcl-X_L; and the downregulation of pro-apoptotic proteins, Bax, Bim and Puma.

5. Enabled replicative immortality: Healthy cells pass through a limited number of successive cell growth and division cycles before ending in senescence or cell death. Cancer cells circumvent senescence and cell death, enabling unlimited proliferation, and a great wealth of evidence suggests it is due to telomere protection of chromosome ends (Shay and Wright, 2000; Blasco, 2005). The length of telomeric DNA at the ends of DNA dictates the number of successive cell generations its progeny can pass through. Telomerase, the specialised DNA polymerase that adds telomeric repeats at DNA ends, is almost absent in non-immortal cells, but present at significantly higher levels in the majority (~90%) of immortalised cells, including cancer.

6. Angiogenesis: Like normal tissues, tumours require nutrients and oxygen, and the removal of metabolic waste and carbon dioxide. Angiogenesis normally only occurs during embryogenesis, wound healing and the female reproductive cycling. However, for the acquisition of nutrients and the removal of waste, angiogenesis is constitutively activated during tumourigenesis and sustained after neoplastic growth (Hanahan and Folkman, 1996). Evidence indicates that this angiogenic switch is controlled by countervailing factors that are inhibitory, e.g. thrombospondin-1 (TSP-1), or stimulatory, e.g. vascular endothelial growth factor-A (VEGF-A) (Bergers and Benjamin, 2003; Baeriswyl and Christofori, 2009).

Following on from their first paper on the Hallmarks of Cancer, Hanahan & Weinberg published additional *enabling characteristics* and *emerging hallmarks* of cancer (Figure 1.2) (Hanahan and Weinberg, 2011). These *emerging hallmarks* of cancer can be summarized as follows:

7. Reprogramming cellular energy: Under aerobic conditions, healthy cells undergo aerobic respiration for energy production, although cancer cells undergo anaerobic respiration, resulting in an ~18-fold decrease in adenosine triphosphate (ATP) production. Although opting for anaerobic metabolism appears to be counterintuitive, it is hypothesized that cancers use the glycolytic intermediates in the production of nucleosides and amino acids that are required to assemble new cancer cells.

8. Evading immune destruction: It is thought that cells and tissues are under constant immune surveillance that recognizes and eliminates nascent tumours. Thus, it appears that developed cancers evade detection and destruction by the immune system. These findings are supported by *in vivo* studies where tumours developed more frequently in immunodeficient mice (Kim et al., 2007; Teng et al., 2008). Clinical epidemiology also shows evidence of anti-tumour immune responses in some forms of human cancer (Nelson, 2008; Bindea et al., 2010; Ferrone and Dranoff, 2010)

These *enabling characteristics* constitute the development of genomic instability in cancer cells and the inflammatory state of premalignant and malignant tissues, both promoting tumour progression. Taken together, it is clear that numerous complementary insults that affect DNA and protein expression leads to cancer development. Understandable, these molecules make for excellent targets in principle for improved cancer chemo- and immuno-therapy, and cancer diagnostics and screening.

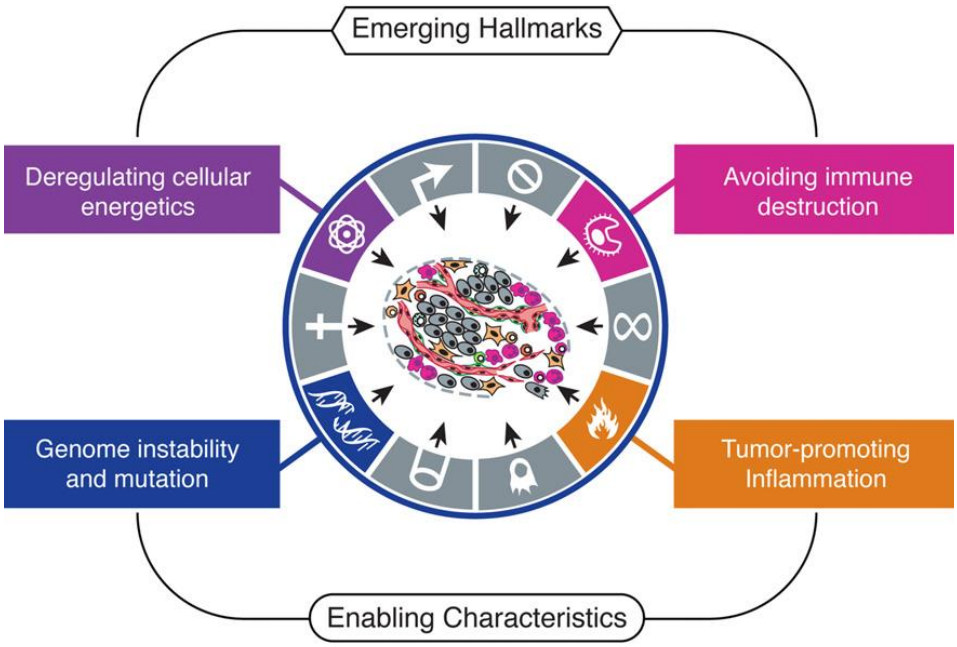


Figure 1.2. Emerging hallmarks and enabling characteristics. Besides the 6 core hallmarks of cancer previously described, two additional hallmarks emerged, including the deregulation of cellular energetics and cancer evasion of immunological destruction. Additionally, two consequential enabling characteristics of neoplasia facilitate the acquisition of core and emerging hallmarks, including genomic instability and inflammation by innate immune cells.[Adapted from (Hanahan and Weinberg, 2011)]

1.2. Global cancer epidemiology

In 2015, the World Health Organisation (WHO) reported that of the 56.4 million deaths worldwide, more than half were due to the top 10 diseases: ischaemic heart disease (8.8 million), stroke (6.2 million deaths), chronic obstructive pulmonary disease (3 million deaths), lung cancer (including trachea and bronchus cancers) (1.7 million deaths), diabetes mellitus (1.6 million), dementia (1.5 million), diarrhoeal disease (1.4 million), tuberculosis (1.4 million), and road kill injury (1.3 million) (http://www.who.int) (Figure 1.3). Lung, trachea and bronchus cancers increased from 1.26 million to 1.69 million deaths from 2000 to 2015. The world population size increased from ~6.1 billion in 2000 to ~7.2 billion in 2015, indicating that the rate of lung, trachea and bronchus cancers increased.

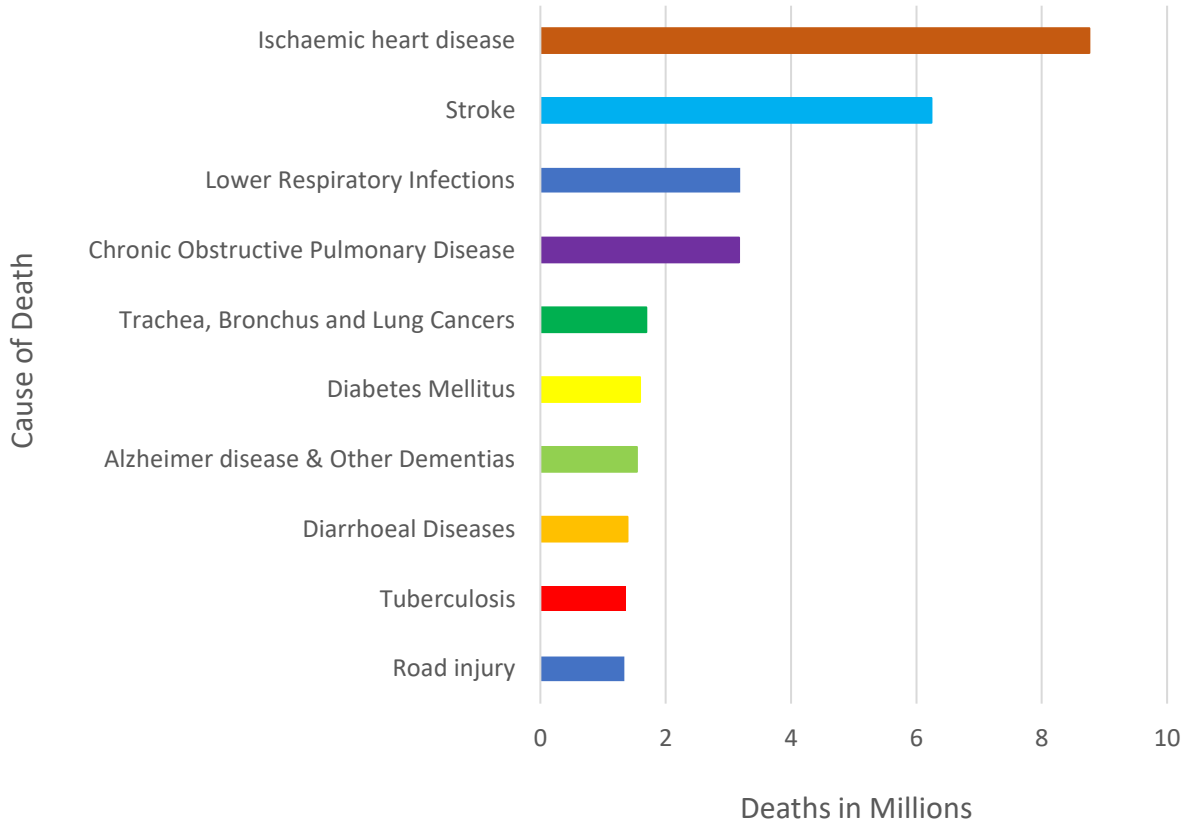


Figure 1.3. Top 10 causes of death worldwide in 2015. The worldwide leading causes of death are summarized as deaths per million. Cancers of the trachea, bronchus and lung increased from 1.26 million deaths in 2000 to 1.69 million deaths in 2015, as the fifth leading cause of death worldwide (adapted from: http://www.who.int).

Although trachea, bronchus and lung cancers are the fifth leading cause of death worldwide, all cancer-types are cumulatively responsible for most deaths worldwide: Based on GLOBOCAN 2008, ~12.7 million cancer cases and ~7.6 million cancer deaths occurred in 2008; of these, 56% of the cases and 64% of the deaths occurred in developing countries (Jemal et al., 2011). According to the GLOBOCAN (2012), these numbers increased to ~14.1 million new cancers, and ~8.2 million deaths in 2012 worldwide (http://globocan.iarc.fr).

The GLOBOCAN summarized the global cancer incidence and mortality rates, depicted in global incidence (Figure 1.4) and mortality (Figure 1.5) maps, based on age-specific rates (ASR) across all cancer types, excluding non-melanoma skin cancers. Here, it is evident that areas presenting the highest incidence numbers and highest mortality numbers do not overlap. These disparities are probably due to differential risk factors, screening practices, and/or the availability of cancer treatment in developing countries (Jemal et al., 2011).

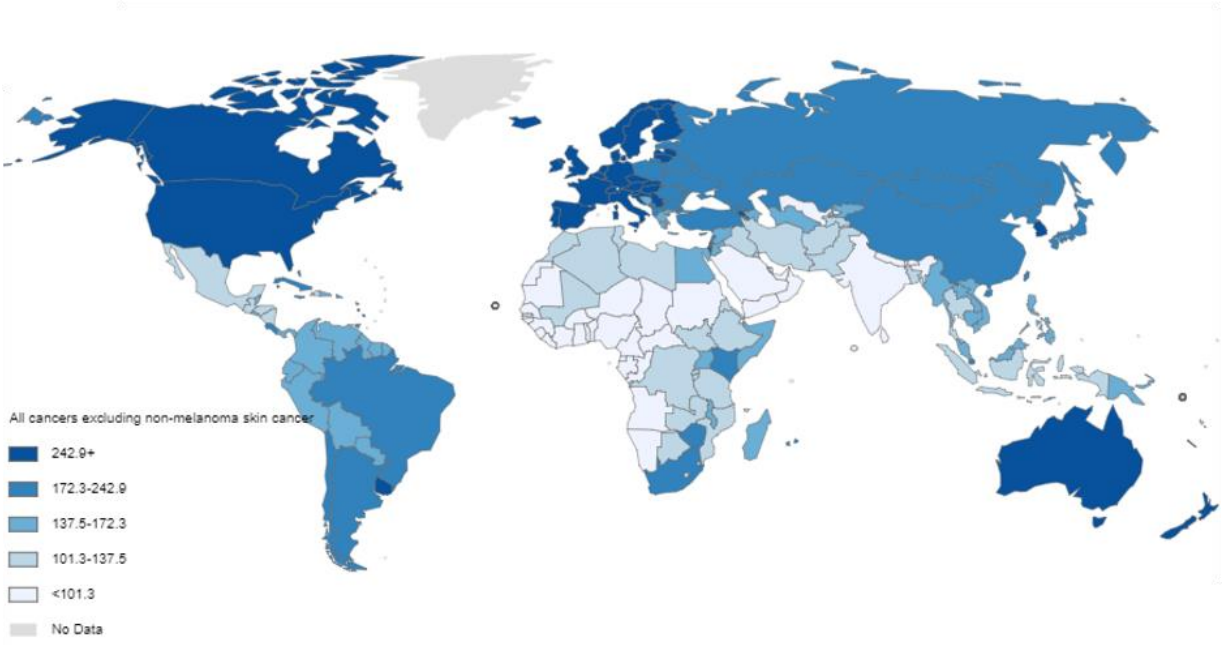


Figure 1.4. Global cancer incidence rates. The map represents the age-specific rates (ASR) of global incidence for all cancer-types, except for non-melanoma skin cancers. The blue gradient scale correlates with cancer incidence, while grey indicates that no data is available. [Adapted from: http://globocan.iarc.fr]

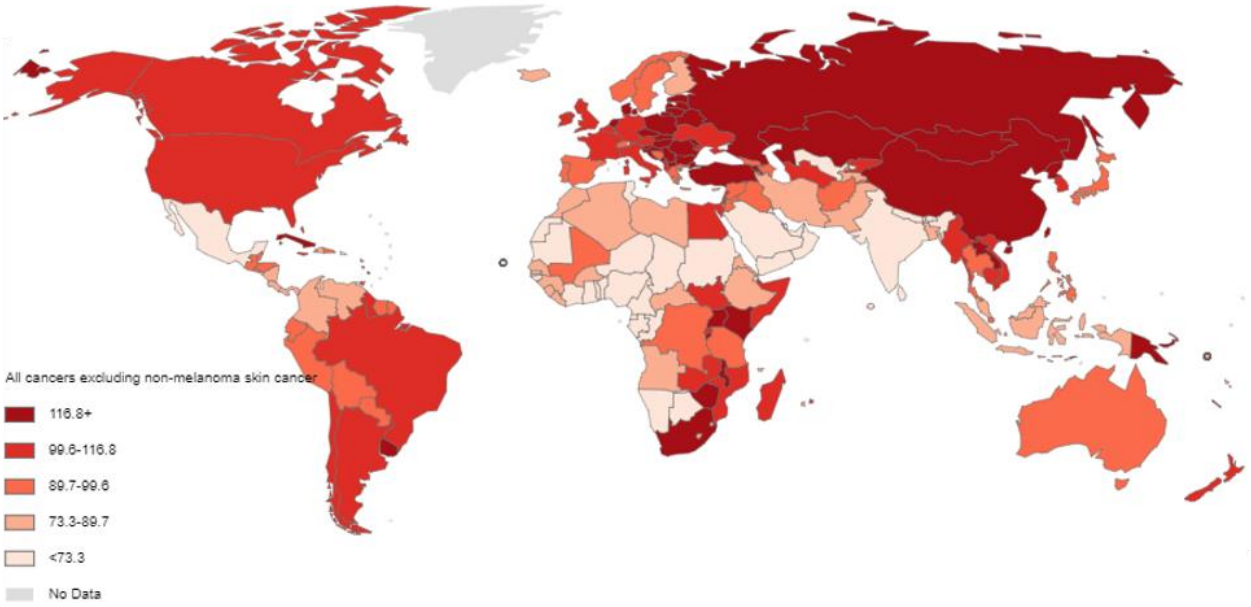


Figure 1.5. Global cancer mortality rates. The map represents the age-specific rates (ASR) of global mortality for all cancer-types, except for non-melanoma skin cancers. The blue gradient scale correlates with cancer incidence, while grey indicates that no data is available. [Adapted from: http://globocan.iarc.fr].

Although the cancer incidence rates in developing countries are half of those reported in developed countries, the overall mortality rates are similar, regardless of patient gender. The most prevalent cancers worldwide include breast, prostate and colorectal cancer (CRC). The mortality and incidence rates for the top 20 cancers, including both genders, are summarised in Figure 1.6.

For females, breast cancer is the most frequently diagnosed (23% of total cases) cancer, and the leading cause of cancer death (14% of total deaths) in developed and developing countries; whereas lung cancer is most frequently diagnosed (17%) cancer in men, accounting for 23% of total cancer deaths (Jemal et al., 2011). Furthermore, cancer incidence rates are almost 25% higher in men, at 205 per 100 000 males *versus* 165 per 100 000 females.

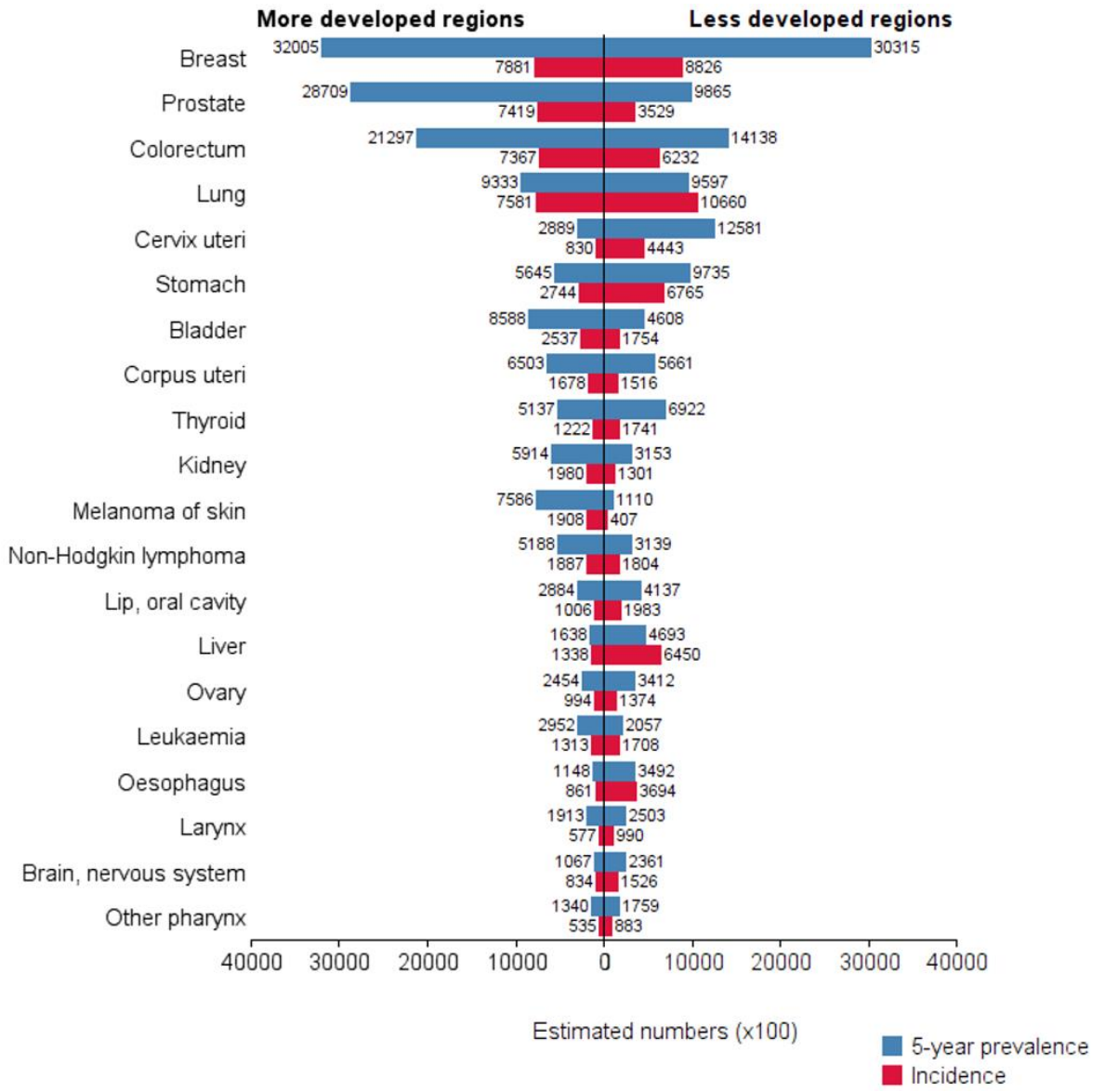


Figure 1.6. Global cancer prevalence in developed and developing regions. The 5-year cancer prevalence (blue bars) and cancer incidence (red bars) is summarised, with estimations numbers (×100), for developed and developing countries for both males and females. [Adapted from: http://globocan.iarc.fr].

Together, the incidence and mortality rates highlight the need to reduce cancer burden worldwide. Prevention modalities include cancer screening and increased awareness of risk factors, such as smoking, alcohol consumption, diet, obesity, poor exercise. Treatment modalities include the use and development of chemotherapy, radiotherapy, surgery and cancer vaccines.

1.3. Cancer immunology

The innate and adaptive immune responses are raised against cancer. The innate immune response is rapid, and the first line of defence. It consists of soluble factors (e.g. complement proteins) and cellular components, including granulocytes (including basophils, eosinophils and neutrophils), mast cells (MC), macrophages, dendritic cells (DCs) and natural killer cells (NK). The adaptive immune response is slower, although it is antigen-specific and provides immunological memory. The components of the adaptive immune system consist of T-lymphocytes, which include T-helper cells ($CD4^+$), cytotoxic T-cells ($CD8^+$), regulatory T-cells (T_{regs}). Other immune cells of the adaptive immune system are B-lymphocytes (B-cells), which are responsible for antibody (Ab) secretion, natural killer T-cells and γδ T-cells which are both cytotoxic lymphocytes at the interface of innate and adaptive immunity (Figure 1.7) (Dranoff, 2004).

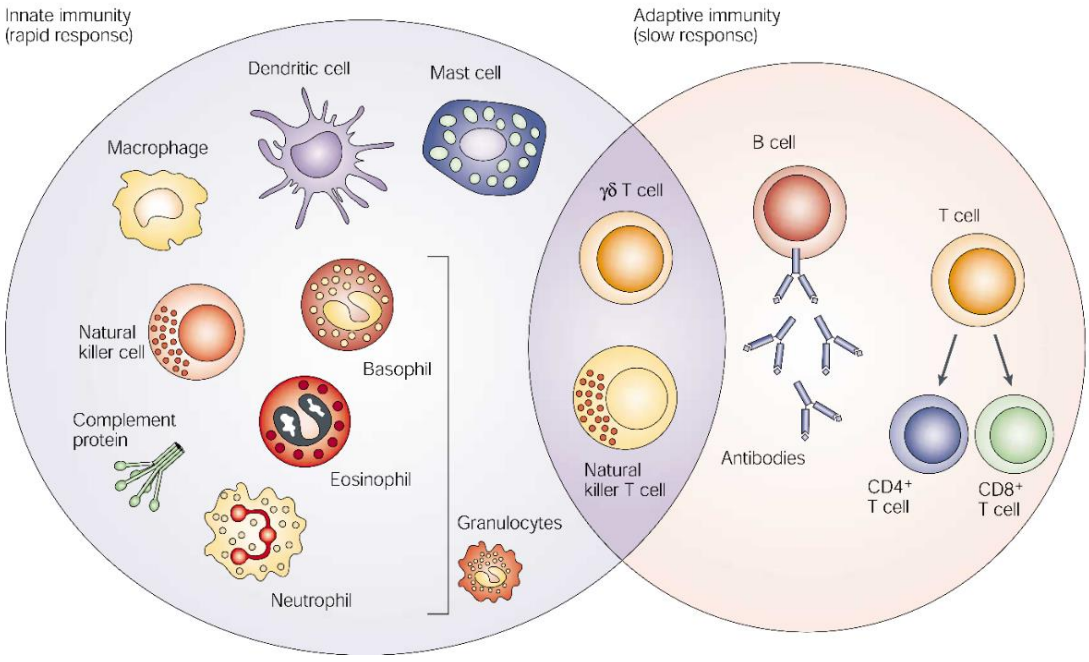


Figure 1.7. Components of the innate and adaptive immune system. The innate immune response is rapid, and includes soluble components (e.g. complement proteins), granulocytes (e.g. basophils, eosinophils and neutrophils), mast cells (MC), dendritic cells (DCs), macrophages, natural killer cells (NK). The adaptive immune, which is slower, but more specific and provides immune memory retention. It includes T lymphocytes (e.g. $CD4^+$ T-cell, $CD8^+$ T-cells, regulatory T-cells) and B lymphocytes. Natural killer T-cells and γδ T-cells are cytotoxic lymphocytes at the interface of innate and adaptive immunity [Adapted from (Dranoff 2004)].

In the anti-cancer immune response, antigen presenting cells (APCs) lyse and present tumour-associated antigens (TAAs) on major histocompatibility complex (MHC) molecules at the surface of T-cells. MHC I molecules occur on all nucleated cells and platelets, but not on red blood cells. The MHC I complex is composed only of β2 subunits, and are recognised by $CD8^+$ T-cell receptors trigging a cascade of events that lead to $CD8^+$ T-cell proliferation, cytokine production and target cell lysis (Johnsen et al., 1999; Seliger et al., 2000). For MHC I production, a large mutlicatalytic protease complex called the proteasome degrades endogenous proteins, e.g. TAAs. The peptides are transported to the lumen of the endoplasmic reticulum (ER) by transporter-associated with antigen processing (TAP) proteins. Accessory proteins facilitate the folding and assembly of complete MHC I molecules which interacts with the immunogenic peptide in the ER, which is subsequently transported from the ER through the Golgi apparatus for antigen presentation on the cell surface. MHC II molecules have β1 and β2 subunits, and are crucial for the activation of $CD4^+$ T-cells. MHC II molecules can be expressed on all cell-types, although they are usually expressed on professional APCs including macrophages, B-cells and DCs. MHC II molecules are also expressed on tumour cells, including colorectal and breast carcinomas. TAAs are degraded to form peptides by degradative processes in multivesicular bodies (MVBs) in tumour cells. In the ER, three MHC II αβ heterodimers bind to trimeric proteins called invariant chains, forming a nonameric complex. The complex is transported through the Golgi apparatus or cell surface to MVBs, where proteases degrade the invariant chain. Concurrently, non-classical MHC II HLA-DM and -DO dimerize in the ER and re-locate to the MVBs. Free HLA-DM interacts with HLA-DR and mediate peptide exchange following the inward budding of the limiting membrane. Peptide-loaded MHC II molecules then egress to the cell surface for antigen presentation (Thibodeau et al., 2012) to interact with $CD4^+$ T-cells.

Multiple mechanisms have been proposed for the immune evasion of tumours; one such mechanism is through downregulated expression of MHC molecules in and on cancer cells as a result of somatic mutations. MHC I is almost replete in multiple cancers, reportedly as low as 13% in CRC (Cabrera et al., 1998), 4% in breast cancer (Cabrera et al., 1996), 4% in cervical carcinomas (Koopman et al., 2000), and 30% in laryngeal cancers (Cabrera et al., 2000). Multiple studies propose that the major force contributing to the appearance of MHC negative tumour clones is

due to T-cell immunoselection, whereby T-cells recognise tumour antigens displayed on MHC positive tumour cells. The remaining MHC-defective tumour cells evade immune surveillance, and acquire a growth advantage that allow the tumour to expand (Kaklamanis and Hill, 1992; Lehmann et al., 1995; Jäger et al., 1997). The loss of MHC can occur during HLA synthesis, assembly, transport, or expression on the cell surface (Garrido et al., 1997; Ruiz-Cabello and Garrido, 1998). A molecular mechanism analysis provided evidence that MHC I-deficient metastatic nodes in immunocompetent mice showed co-ordinated suppression of multiple components of the MHC I antigen presenting machinery. Interestingly, treatment with interferon-γ (IFN-γ) transcriptionally induces the re-expression of downregulated antigen processing machinery, and consequently enhances MHC I surface expression (Garcia-Lora et al., 2003); thus the inclusion of IFN-γ with immunotherapy treatment may improve patient outcomes.

The antigen-specific B-cells encounter, recognize and bind activated $CD4^+$ T-cell cells (Goodnow et al., 2005); this process is maximized by lymphocyte circulation into secondary lymphoid organs and peripheral tissues. The interaction between T-cells and B-cells activates cytokine and chemokine production, leading to B-cell proliferation. A portion of the B-cells serve as memory cells, whereas the remainder act as effector cells that differentiate into antibody-producing plasma cells responsible for the production and release of antigen-specific antibodies (Wardemann and Nussenzweig, 2007). The B-cell response to a specific antigen is thus dependent on T-cells and occurs in three phases: extrafollicular, follicular and germinal centre (GC) reaction (Qi, 2016). During the first phase, B cell growth is driven by T- helper cells and early waves of plasma cells are produced from the T-cell – B-cell border of the follicle. In the second phase, T-cells and B-cells interact inside the follicle, culminating in GC formation followed by cycles of B-cell somatic mutation, clonal proliferation and selection that result in affinity maturation in the third phase (Qi, 2016).

Peripheral tolerance mechanisms usually ensure that self-reactive T- and B-cells (i.e. displaying T- or B-cell receptors for self-antigens) are suppressed. However, in certain circumstances, peripheral tolerance can be broken, resulting in proliferation of autoantigen-specific T- and B-cells. Simplistically, peripheral tolerance can be broken for a number of reasons, for example if

the self-antigen is significantly over-expressed in a tissue or if neoantigens are somehow presented to the host immune system. Such neoantigens can include mutated peptide epitopes, aberrantly spliced or aberrantly post-translationally-modified epitopes, or new discontinuous epitopes resulting from misfolding of the antigen. One consequence of this loss of peripheral tolerance can be the production of self-antigen-specific autoantibodies.

The interaction between autoantibodies and TAAs leads to the destruction of transformed cancer cells containing the B-cell primed antigen by labelling them for faster macrophage recognition and phagocytosis. Autoantibody-induced tumour killing subsequently occurs as a result of direct interaction between autoantibodies and TAAs, which blocks tumour cell surface receptors associated with tumour cell proliferation and survival. The autoantibodies can also induce tumour-associated antigen (TAA) uptake via DC Fc gamma receptors, leading to antigen cross-presentation and strong $CD4^+$ and $CD8^+$ T-cell responses, complement-dependent cytotoxicity and NK cell-mediated antibody-dependent cellular cytotoxicity (Carter, 2001).

Antibodies are secreted heterodimeric proteins comprising of light and heavy chains produced through recombination of V(D)J segments in developing B-lymphocytes. At any given time, there are thought to be of the order of 10^7-10^8 different Ab sequences present in human serum. In response to the presence of foreign antigens or pathogens, somatic hypermutation processes drive the affinity maturation of specific Ab sequences, resulting in the production of high affinity, antigen-specific antibodies. Also, known as immunoglobulins (Igs), affinity matured antibodies are produced by plasma cells and secreted into the blood stream where they scavenge their cognate antigen for destruction. Antibodies thus play a crucial adaptive role in mammalian defence mechanisms against harmful components that can cause disease. There are 5 classes of antibodies; IgG, IgM, IgE, IgD and IgA, which differ in their structures and immune functions. IgG is the major Ab in the blood and has the longest serum half-life of all Ig isotypes (Schroeder and Cavacini, 2010). It contributes directly to an immune response including neutralisation of toxins and viruses. IgA is also involved in the direct neutralization of toxins, viruses and bacteria; however, it concentrates particularly in mucosal surfaces. IgM, a pentameric Ig, is the largest Ab and it is associated with a primary immune response and is frequently used to diagnose acute exposure to an immunogen or pathogen (Schroeder and Cavacini, 2010). IgD and IgE are found

in trace amounts in the blood with short half-lives. IgD remains membrane-bound and is involved in regulation of cell activation while IgE is associated with hypersensitivity and allergic reactions (Schroeder and Cavacini, 2010). Autoantibodies are antibodies produced by the immune system that target one or more of the host proteins, resulting in autoimmune diseases. Classical autoantibodies are typically IgMs and include: anti-nuclear antibodies (ANA), which bind to nuclear membrane, nucleoplasm, nucleoli and nuclear organelles of cells. Rheumatoid Factor (RF), which binds with relatively low affinity to the Fc region of IgGs and which is found in the serum of Rheumatoid Arthritis (RA) patients (Ma et al., 2017); Anti-double-stranded DNA antibodies, anti-Sm antibodies, antiphospholipid antibodies, anti-Ro, anti-ribonucleoprotein and anti-La Antibodies which are all frequently found in Systemic Lupus Erythematosus (SLE) patients (Ma et al., 2017); & Anti-Sjogren's syndrome A (SSA) and -B (SSB) antibodies, which are found in many patients with Sjogren's Syndrome (Ma et al., 2017).

In cancer, multiple aberrations result in autoantibody production against TAAs, including tolerance defects and inflammation, changes in TAA expression levels, altered protein structure, and cellular death mechanisms (Figure 1.8), which are summarised below (Zaenker et al., 2016):

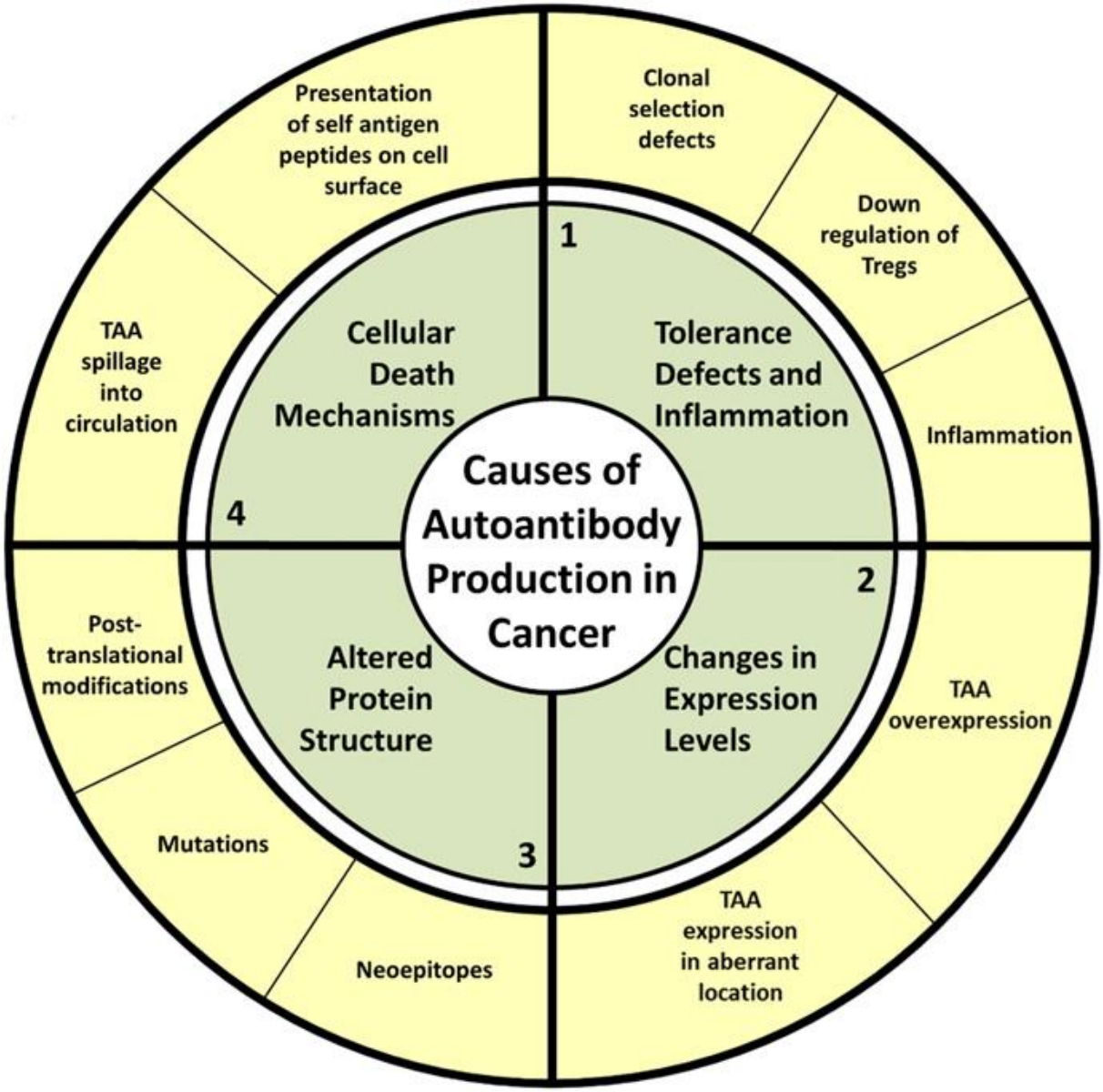


Figure 1.8. The proposed causes of antibody production in cancer [Adapted from (Zaenker et al., 2016)**]**. Abbreviations include: TAA = tumour-associated antigen.

1. Tolerance defects and inflammation: Roughly half of the lymphocytes in lymphocytic organs are directed against autoantigens. Upon circulation, these self-reactive lymphocytes enter a series of checkpoints that control central tolerance. Thus, only lymphocytes that are non-self-reactive will mature. However, self-tolerance is complex and subject to error. For example, maintaining clonal anergy requires constant receptor interaction and signalling, which can be

reversed by antigen dissociation, leading to autoantibody production. Tolerance defects also stem from the downregulation of T_{regs}. Studies indicate delayed tumour growth due to decreased T_{regs} levels, which correlate with increased levels of effector T helper cells, germinal centre B-cells, and high titres of autoantibodies (Kim et al., 2010). Lastly, lasting chronic inflammatory responses are associated with cancer. Inflammation assists with increased vasculature permeability allowing immune cell accumulation at the tumour site (Alvarez Arias et al., 2014). The inflammatory response is suggested to facilitate the release and exposure of intracellular antigens to the immune system, resulting in autoantibody production in cancer patients (Carl et al., 2005).

2. Changes in protein expression levels: Autoimmune responses to TAAs are largely due to antigen overexpression during tumourigenesis. Studies with hepatocellular carcinoma showed an association with increased CENPF levels with anti-CENPF autoantibody production (Ménard et al., 2004). Breast cancer patients with increased Her2/Neu and MUC1 produced anti-HER2/Neu and anti-MUC1 autoantibodies, whereas autoantibodies were not detected in patients with low Her2/Neu and MUC1 expression levels, respectively. Autoantibody production may result from protein expression in an aberrant location. Cancer-testis antigen (CTA) expression is normally confined to immunoprivileged sites, e.g. testis, ovaries, or trophoblast of the placenta (Scanlan et al., 2002), and expression in many cancer types results in autoantibody production (Simpson et al., 2005). Similarly, 21% of hepatocellular carcinoma patients produce autoantibodies against oncofoetal antigen IMP2, a protein normally expressed only during prenatal development and shortly after birth (Zhang et al., 1999).

3. Altered protein structure: Neoepitopes may be created by somatic mutations that alter protein structure, or due to the exposure of epitopes normally located within enclosed protein regions (Liu et al., 2012). Genetic instability, a hallmark of cancer (See section 1.1), also results in the expression of neoantigens that induce an autoimmune response (Joseph et al., 2014). Autoantibody responses have also been raised against mutated cancer proteins. For example, p53 is frequently mutated in various cancers, resulting in autoantibody production. However, only 20-40% of patients produce autoantibodies against mutated p53, indicating that not all protein aberrations induce autoantibody responses (Soussi, 2000) or that not all cancers have

mutated p53. Although protein products of missense mutations result in autoantibody production, protein products of stop, splice and frameshift mutations have been reported to not cause autoantibody production in lung cancer (Reuschenbach et al., 2009). In contrast, frame shift mutations do induce immunological responses in CRC patients (Schwitalle et al., 2008). The mRNA translation from alternative open reading frames can also generate proteins capable of triggering an autoimmune response. Examples of such proteins are CTAG1B (Wang et al., 1998) and opioid growth factor receptor (OGFr) (Mollick et al., 2003). Lastly, aberrant post-translational modifications (PTMs) on cancer proteins, e.g. glycosylation, methylation, phosphorylation, sumoylation, adenosine diphosphate-ribosylation, ubiquitination, acetylation and citrullination, are also capable of inducing an autoantibody response.

4. Cellular death mechanisms:

The exact mechanism leading to an autoimmune response against cancer is still to be elucidated. Several studies provide evidence that post-apoptosis, insufficient clearance of accumulated nuclear and cytoplasmic proteins plays an important role. Also, cell lysis following tumour cell necrosis and autophagy results in the spillage of autologous proteins into blood, triggering an immune response that leads to autoantibody production.

Tumour cell lysis may also occur through cytokine secretion, including tumour necrosis factor (TNF) and IFN-γ by T-cells. Also, TAAs present on the surface of tumours cells may induce an immune response that promotes autoantibody production. Furthermore, apoptotic cancer cells present altered cleavage products and post-translationally modified self-antigens on the surface of blebs, which also promotes autoimmunity (Joseph et al., 2014).

1.4. Tumour antigens and cancer-testis antigens

Tumour antigens are antigenic substances produced in tumour cells that trigger an autoimmune response in the host. Tumour antigens can be defined as TAAs, which are antigens similar to proteins found in normal cells, but are modified or overexpressed; or TSA, which are antigens found only in or on tumour cells, and not on normal cells.

Tumours were first observed to be immunogenic during the 1940s and 1950s, when rodents showed immune-mediated rejection of syngeneic transplanted tumours (Gross, 1943). Subsequently, the identities of many tumour antigens, in mice and humans revealed that they can be oncovirus proteins such as E6 and E7 (Stauss et al., 1992); mutated proteins such as RAS and p53 (Peace et al., 1991; Skipper and Stauss, 1993); fusion proteins such as BCR-Able (Chen et al., 1992); over-expressed proteins such as HER2/neu (Schechter et al., 1984); differentiation proteins such as CD20 (Anderson et al., 1984); or contain aberrant PTMs. A list defining cancer antigen-types, definitions and examples are described in Table 1.1.

Table 1.1. List of cancer antigens-types, definitions and examples. Tumour antigens can be defined as TAAs, which are antigens similar to proteins found in normal cells, but are modified or overexpressed; or TSA, which are antigens found only in or on tumour cells, and not on normal cells. The table below lists cancer-antigen-types, definitions and examples. Abbreviations: TAAs = tumour-associated antigens; TSAs = tumour-specific antigens.

Cancer antigen-type	Definition	Examples	Reference
Cancer-testis antigens	Cancer proteins with restricted expression to the testis or placenta, but also expressed in cancer.	NY-ESO-1 MAGE SSX	(Jäger et al., 1999) (Daudi et al., 2014) (Smith and McNeel, 2010)
Differentiation antigens	Cancer proteins expressed during at least one stage of cellular differentiation resulting in tumour development	MART-1 HER-2/neu IL22RA1	(Colella et al., 2000; Yee et al., 2000) (Slamon et al., 2001) (Qi and Ding, 2018)
Virus-associated antigens	Viral stain-associated proteins that promote cancer initiation and/or progression.	EBV HPV ATLA	(Liu et al., 2018) (Schäfer et al., 2017) (Hinuma et al., 1982)
Mutational antigens	Cancer proteins that result from mutations in otherwise “normal” proteins	P53 BRAF EGF	(Peace et al., 1991; Deniger et al., 2018) (Tomei et al., 2015) (Lynch et al., 2004)
Over-expressed antigens	Cancer proteins that are overexpressed in cancers, usually conferring tumorigenic effect and an immune-suppressor.	VEGF/VEGFR AR HER2/Neu	(Li et al., 2016; Zhu et al., 2017) (Bupp and Jorgensen, 2018) (Schechter et al., 1984)

CTA, also known as germ-line antigens, were added to the list of TAAs in 1991 when Van Der Bruggen *et al*. (van der Bruggen et al., 1991) showed expression of melanoma antigen family A, 1 (MAGE-1) antigen in the MZ2-MEL melanoma cell-line using autologous typing. Subsequently, cloning of the gene revealed that the gene encoding MAGE-1, *MAGEA1*, is a member of the 12 gene family whose expression is normally restricted to the testis, but can be detected in a wide range of tumour-types (De Plaen et al., 1994). Further investigations lead to the discovery of other tumour antigens that were encoded by testis-restricted genes, including *BAGE* (Boël et al., 1995) and *GAGE* (De Backer et al., 1999) genes. Other antigens isolated also include NY-ESO-1 (Türeci et al., 1996), HOM-MEL-40 (Türeci et al., 1998) and SYNSP1 (Chen et al., 1997). Given the reactivation and subsequent immunogenicity of testis proteins in tumours, such antigens were termed CTAs. Numerous methods have been developed and employed for the discovery of CTAs, with nearly 225 genes identified, however, many of these were identified by gene expression data alone and have unknown antigenic status. The comprehensive CTA database, including the list of proteins with gene names, and mRNA, protein expressed and immune response summary, can be found at http://www.cta.lncc.br/.

The roles of CTAs in the gamete are unclear, although gene expression and knockout studies have shown that these proteins have diverse functions. Studies on X-lined CTAs, e.g. *MAGEA1* and *NY-ESO-1*, have shown their expression in the initial stages of spermatogenesis (Jungbluth et al., 2000b). Expression of *GAGE* genes has been detected in $OCT4^{+}$ primordial germ cells, suggesting that the protein may be important in germ stem cell function (Gjerstorff et al., 2007). Studies on mice lacking single CTAs frequently showed reduced or even ablated fertility. Furthermore, SPATA19 and COX6B2 have been shown to be essential in sperm metabolism (Goldberg et al., 2010), AKAP3, CABYR, SPA17 and ROPN1 are important in sperm movement and SPO11, SYCE1, SYCP1, HORMADI and TEX15 are important for meiosis in sperm cells (Brown et al., 2003; Eddy et al., 2003; Chiriva-Internati et al., 2008; Fiedler et al., 2008). The diverse function of CTAs in their native setting suggests that they may be important in a number of tumourigenic features.

The functions of CTAs in cancers have previously been studied, appearing to promote tumourigenicity. A CTA called hCGH was detected in a variety of cancers including bladder, breast, renal, ovarian, prostate, pancreatic and CRCs (Stenman et al., 2004). hCGH, normally

produced in the placenta, is required for progesterone secretion by the corpus leuteum in order to promote the formation of blood vessels and capillaries to sustain the developing foetus, thus, hCGH may contribute to the process of angiogenesis in tumours (Cole, 2012). The CTA, ATAD2, acts as a cofactor for Myc-mediated transcription, where its depletion reduces tumour cell proliferation by inhibiting S-phase entry (Ciró et al., 2009), suggesting that ATAD2 may promote tumourigenicity. Furthermore, CTAs including PRAME (Epping et al., 2005), CCDC62 (Chen et al., 2009) and MAGE (Doyle et al., 2010), have been implicated in tumourigenic signalling, whereas CEP55 (Chang et al., 2012), TEX14 (Mondal et al., 2012), CASC5, TTK (Cheeseman et al., 2008) and NUF2 (DeLuca et al., 2002) are associated with tumour cell division. The tumourigenic properties of CTAs make them potential targets for cancer immunotherapy and useful as targets for cancer diagnosis.

1.5. Autoantibodies against TAA for cancer diagnosis

To counteract high cancer mortality rates, diagnostic tools need to be developed for early stage detection of cancers before they reach incurable metastatic stages. Detection of cancer is currently typically restricted to morphological and histochemical diagnosis of primary tumour samples. More recently, autoantibodies against autologous TAAs have been investigated as potential cancer biomarkers. Using autoantibodies for cancer diagnosis is advantageous as (**1**) serum extraction is minimally invasive; (**2**) increased autoantibody levels are detected at early cancer stages (Zayakin et al., 2013); (**3**) autoantibody production precedes clinical signs of tumour progression by months to years (Caron et al., 2007); (**4**) autoantibodies are stable serological proteins (Anderson and LaBaer, 2005); and (**5**) autoantibodies persist for long periods in blood due to limited proteolysis and clearance from circulation (Pedersen and Wandall, 2011).

No single autoantibody has been used alone as a cancer biomarker due to low sensitivity and specificity of single markers. However, biomarker panels of multiple TAAs can in principle result in high sensitivity and specificity, and are therefore sought after to produce biomarker panels to detect specific cancer-types at early-stages. According to the WHO, breast, prostate and CRC are the most prevalent cancer-types globally (Figure 1.6). As such, autoantibodies that may serve as biomarkers for these cancer-types will be discussed:

For breast cancer, CA 15-3, CA 27-29, and carcinoembryonic antigen (CEA) are clinically used as biomarkers, although they are more informative for advanced disease states. The first autoantibody biomarkers for breast cancer include HER2 (Disis et al., 1997), p53 (Crawford et al., 1982), MUC1 (von Mensdorff-Pouilly et al., 1996), and NY-ESO-1 (Stockert et al., 1998). A panel of autoantigens, including HER2, p53, p16 and c-Myc produced sensitivity- and specificity-values of 44% and 97.6%, respectively. Using serological analysis of recombinant cDNA expression libraries (SEREX), SERAC1, RELT, SOCS and ASB-9 achieved sensitivity- and specificity-values of 77% and 82.8%, respectively (Zhong et al., 2008). Serological proteome analysis (SERPA) was used to assess the autoantibody response of advanced breast cancer patients, where HSP60 autoantibodies were detected in 47.5% of patients, but only in 4.7% of healthy control (HC) sera (Hamrita et al., 2008).

For prostate cancer, prostate specific antigen (PSA) serum levels are used for diagnosis, although its specificity is less than 50%, leading to high false-positive results (Zeliadt et al., 2011). Furthermore, the PSA biomarker cannot reliably distinguish between benign and malignant prostate cancer (Stamey et al., 1987). Furthermore, PSA levels can increase due to non-cancerous conditions e.g. urinary infection (Pan and McCahy, 2012). Using protein microarrays on prostate cancer and HC sera, increased autoantibody levels against BRD2, elf-4G1, RPL22, RPL13a and XP_373908 were detected in prostate cancer patients, producing sensitivity- and specificity-values of 81.6% and 88.2%, respectively (Wang et al., 2005). Another microarray study aimed to identify autoantigens that distinguish prostate cancer from benign prostate hyperplasia in patients with high PSA levels. The potential biomarkers include TLN1, TARDBP, LEDGF, CALD1, and PARK7. Alone, PSA produced sensitivity- and specificity-values of 12.2 and 80%, respectively, whereas the collective panel produced sensitivity- and specificity-values of 95% and 80%, respectively (O'Rourke et al., 2012). Using the cancer-testis antigen microarray (CT100plus) containing CTAs and other TAAs, revealed that prostate cancer patients had higher autoantibody titres against GAGE1, ROPN1, SPANZA1 and PRKCZ compared to patients with benign prostate hyperplasia or disease control (Adeola et al., 2016).

For CRC, CEA is the only clinically used serological biomarker, however, it has low sensitivity and specificity (Liu et al., 2009). Autoantibodies against Fas reportedly show high specificity for early-

stage CRC. In addition, higher anti-Fas titres were detected in patients with colorectal adenomas versus colorectal adenocarcinomas, resulting in sensitivity- and specificity-values of 17% and 100% for CRC (Reipert et al., 2005), suggesting that Fas may not be useful to detect CRC, but rather to confirm the absence of disease. An enzyme-linked immunosorbent assays (ELISA)-based assay was used to detect MUC5AC between CRC patients, disease controls and healthy volunteers, producing sensitivity- and specificity-values of 73% and 54%, respectively (Kocer et al., 2006). Autoantibodies against HSP60 (He et al., 2007), p53 (Cioffi et al., 2004; Yoshizawa et al., 2007; Belousov et al., 2008) and Calnuc (Chen et al., 2007) have also been detected in CRC patients, although autoantibodies against these autoantigens are also detected in patients with other cancer-types, too, indicating that these biomarkers are not specific to CRC.

Taken together, it is clear that emerging autoantibody biomarkers have applicability in the clinical setting, although the current issues regarding specificity and sensitivity need to be addressed. The apparent solution involves the use of multiple biomarkers for cancer detection, that, according to the literature, improves both the sensitivity and specificity of the test. It is also apparent that biomarkers specific to a cancer-type are needed to prevent diagnostic ambiguity. Although ELISAs provide effective means of identifying novel antigens, these assays are limiting with regards to the number of biomarkers that can be tested, and often involves tedious procedures. Thus, alternate robust and rapid assaying methods are required, for example proteomics which tests hundreds to thousands of proteins as potential biomarkers for cancer.

1.6. Methods for cancer autoantigen and autoantibody identification

Cancer patients produce autoantibodies against proteins that are mutated, misfolded, improperly glycosylated, overexpressed, truncated or aberrantly localised within tumour cells. Research on CTAs and TAAs, as well as their related autoantibodies, has provided an abundance of targets for therapy, prognostication as well as biomarkers of disease and response to therapy. Techniques used for the identification of cancer associated autoantibody biomarkers are discussed below:

Western blotting is a widely used technique whereby proteins are separated according to their molecular weight by electrophoresis, and then transferred to a membrane; a primary Ab specific

to the protein of interest is then used to detect the presence and relative abundance of the target protein. Conventional western blotting allows the detection of specific proteins at the level of single isotypes. However, it is associated with poor reproducibility, lack of accurate quantitation, and a lengthy time to results, whilst non-specific cross-reactivity of mono- and poly-clonal primary and secondary antibodies on the blots is common. In the context of biomarker discovery, western blotting is used as a validation method rather than a primary method of identifying biomarkers as it is time-consuming, laborious, requires a large sample volume, and allows the assessment of only a few samples at a time.

ELISAs, unlike western blotting, are adaptable to high throughput as it is performed in a 96-well microtiter plates, whereby plate handling and detection systems can be automated. Furthermore, ELISAs can be used to determine the exact amount of protein in a sample, making it readily quantitative. The signals can be produced by chromogenic reaction or chemiluminescence, which is quantified by spectrophotometry. Considering its rapidity and high-throughput capability, as well as the generation of quantitative information, ELISA is an excellent tool for screening, verification and validation of a small number of biomarkers. However, standard ELISAs often have relatively low sensitivity levels and detection usually depends on enzymatic amplification of signal at the end of the assay. In addition, ELISA can give false-positive outputs due to cross-reactivity of the detecting antibodies and other proteins in the sample. Since sensitivity and specificity are prerequisites of any biomarker discovery platform, traditional ELISA may not be the ideal choice when it comes to identifying biologically relevant and meaningful disease biomarkers.

Several high throughput techniques have been used to identify novel CTAs and TAAs, which include SEREX, phage display, SERPA and protein micro-array technology, which are discussed below:

SEREX was developed by Sahin *et al*. in 1995, and utilizes human cDNA derived from patients to profile autoantibody repertoires from the same patient in a process called autologous typing. The approach involves the extraction of mRNA from cancer tissue or tumour cell and the subsequent construction of a tumour-derived cDNA expression library into a bacteriophage for infection of

E.coli. Recombinant proteins are expressed during the lytic phase and blotted onto a nitrocellulose membrane. Sero-reactive proteins can then be identified by sequencing the phage cDNA from positive plaques (Sahin et al., 1995) (Figure 1.9). The approach is highly sensitive, given the use of DNA detection, compared to direct protein identification-based methods that are often limited by absolute protein abundance. Furthermore, the method detects potential antigens from patient-derived tissues and identifies several TAAs in one experiment. The first cancer testis antigen, NY-ESO-1, was identified by SEREX by analysing TAAs that elicit high titre IgG Ab in sera from patients with different types of cancer (Yuan et al., 2016). SEREX has been used to identify several TAAs that generate a humoral response in cancers such as those from kidney, lung, breast and colon (Martin et al., 2011). However, a crucial limitation of SEREX is that it lacks the ability to differentiate or detect PTMs that are likely to play a significant role in autoimmune diseases (Ganesan et al., 2016) and cancers (Martin et al., 2011). Furthermore, identification of TAAs is limited to those that are expressed by the patient tumour or cell line in which the cDNA library was derived, which means that more than one cDNA library may be required to identify comprehensive set of TAAs for difference cancers (Martin et al., 2011).

The phage display approach was developed by Smith, 1985. For the method, a phage display library is constructed from cancer tissue, and biopanning, an affinity selection technique which selects for peptides that bind to a given target, is used to enrich for target peptides. Several cycles of biopanning are performed to enrich for peptides that are specific to cancer sera. Reactive phages are cloned to monoclonality and propagated prior to printing on a microarray. The microarray is incubated with fluorescently-labelled antibodies against the autoantigen of interest. The microarray is also incubated with a fluorescently-labelled Ab specific for the phage capsid protein, which is used for antigen signal normalization (Smith, 1985a) (Figure 1.10). Phage display is useful for the direct detection of TAAs from tumour tissues or a patient-derived tumour cell-line, and simultaneous screening of a large number of antigens against the sera of cancer patients relative to the sera of HCs. However, like SEREX, phage display cannot detect alternate tumour-associated PTMs (Zaenker and Ziman, 2013).

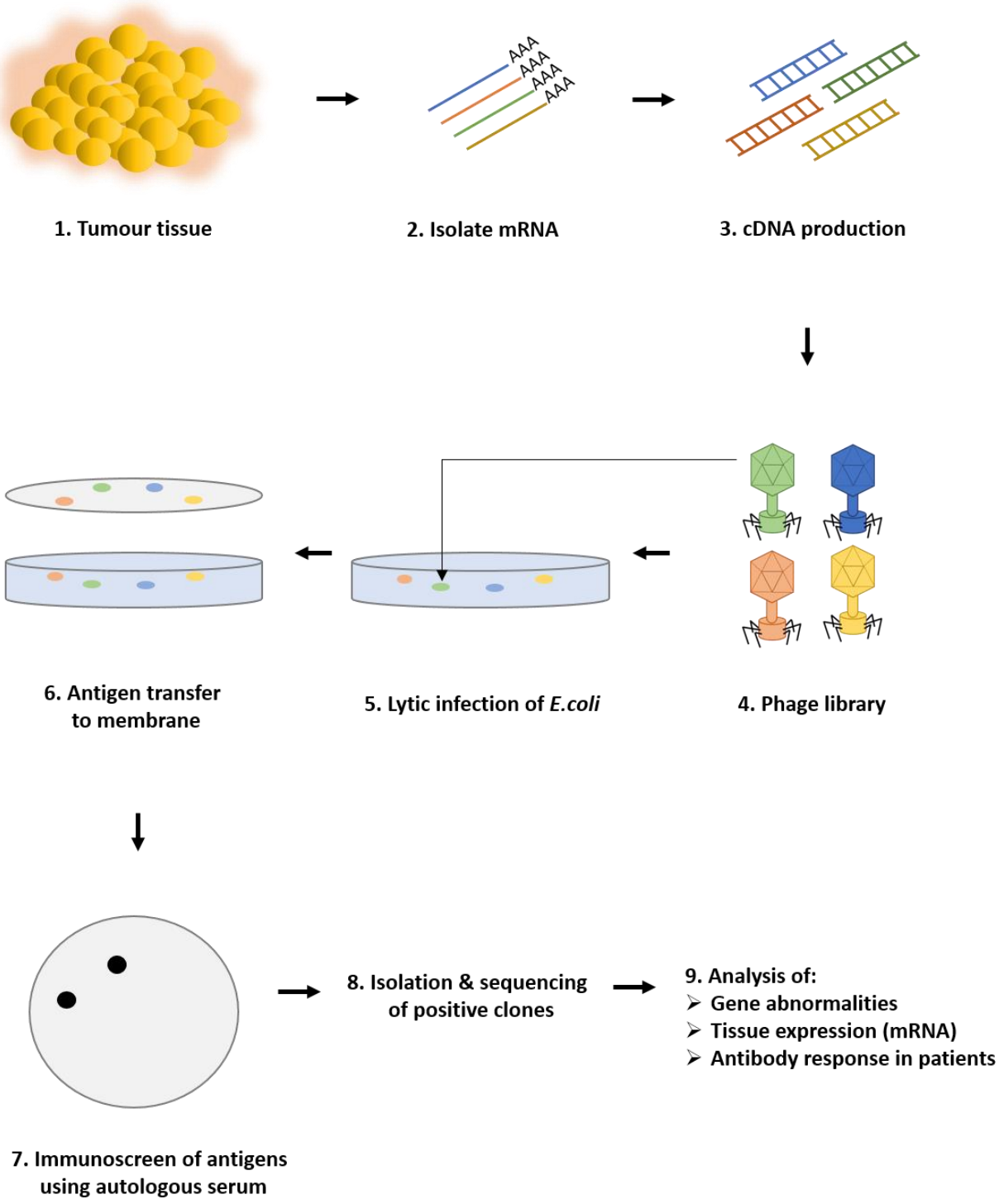


Figure 1.9. The strategy followed in serological analysis of recombinant cDNA expression libraries (SEREX). [Adapted from (Gunawardana & Diamandis 2007)].

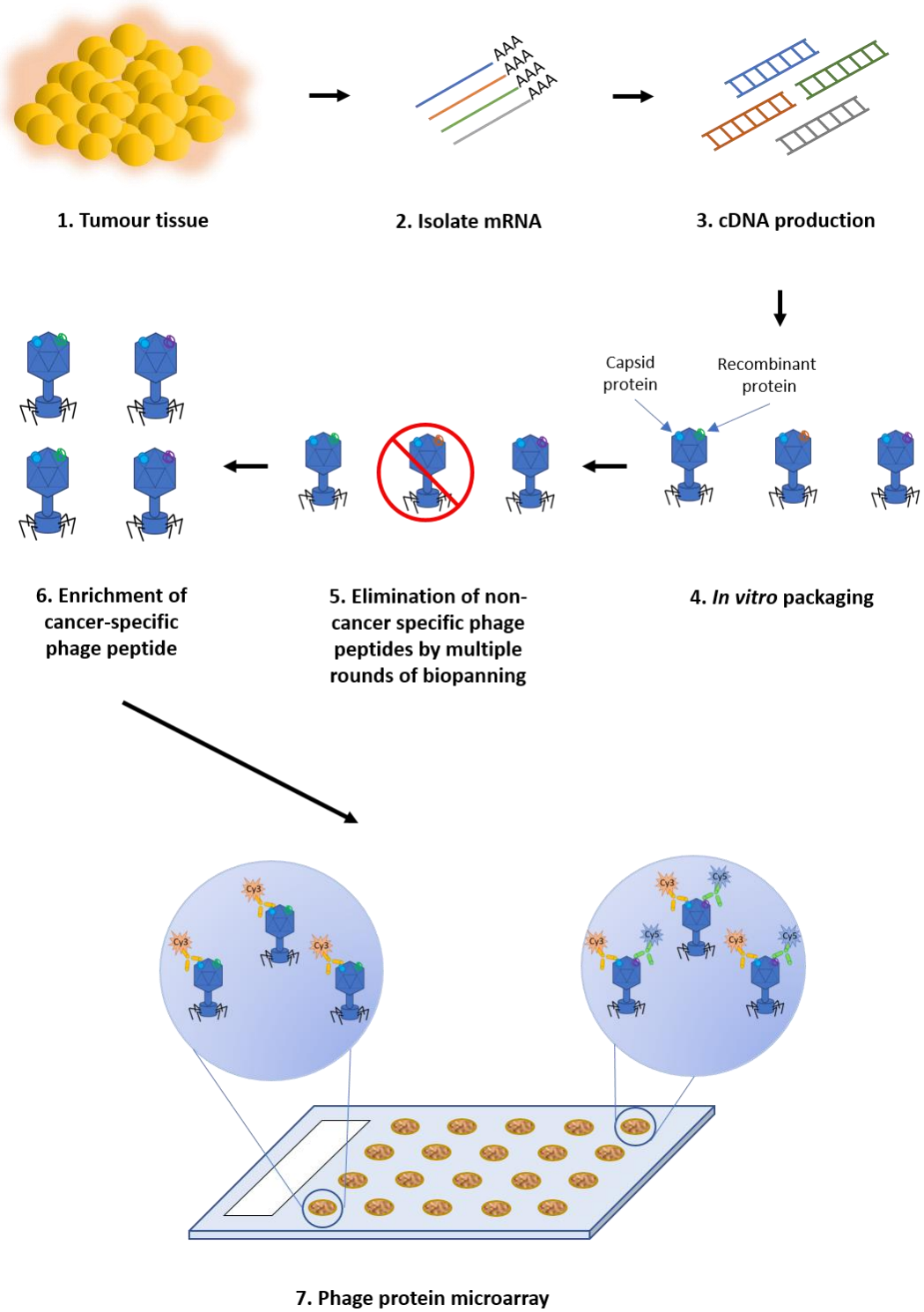


Figure 1.10. Overview of the phage display protein microarray production strategy. [Adapted from (Gunawardana & Diamandis 2007)].

SERPA was developed by Klade *et al*. in 2001, using a combination of two-dimensional (2D) gel electrophoresis and serological analysis. The method involves separating proteins using three replicate 2-dimensional electrophoresis (2-DE) gels under identical conditions. Two gels are blotted to a nitrocellulose membrane, and probed with cancer or normal sera, whereas the third gel is stained with Coomassie blue. Unique immunoreactive spots present in patient and not in the healthy sample are identified and excised from the third gel, and subjected to trypsin digestion followed by protein/peptide identification using mass spectrometry, and subsequently validated using an ELISA (Figure 1.11) (Klade et al., 2001). The advantages of SERPA includes the detection of autoantigens from *in vivo* material, the identification of tumour-specific PTMs, and a parallel analysis of cancer patient and HC plasma. However, the inherent gel-to-gel variability and relatively low resolving power of individual gels impacts on the accuracy of spot picking and imposes a limitation, which is especially true for low-abundance proteins. Several modifications have been suggested to address such limitations, including the triangulation approach involving rigorous quality control steps to accurately identify the protein as well as fluorescence-based 2-dimensional immunoproteomic approaches (Ganesan et al., 2016), but these still so not address the fundamental issues of resolving power and limit of detection for SERPA.

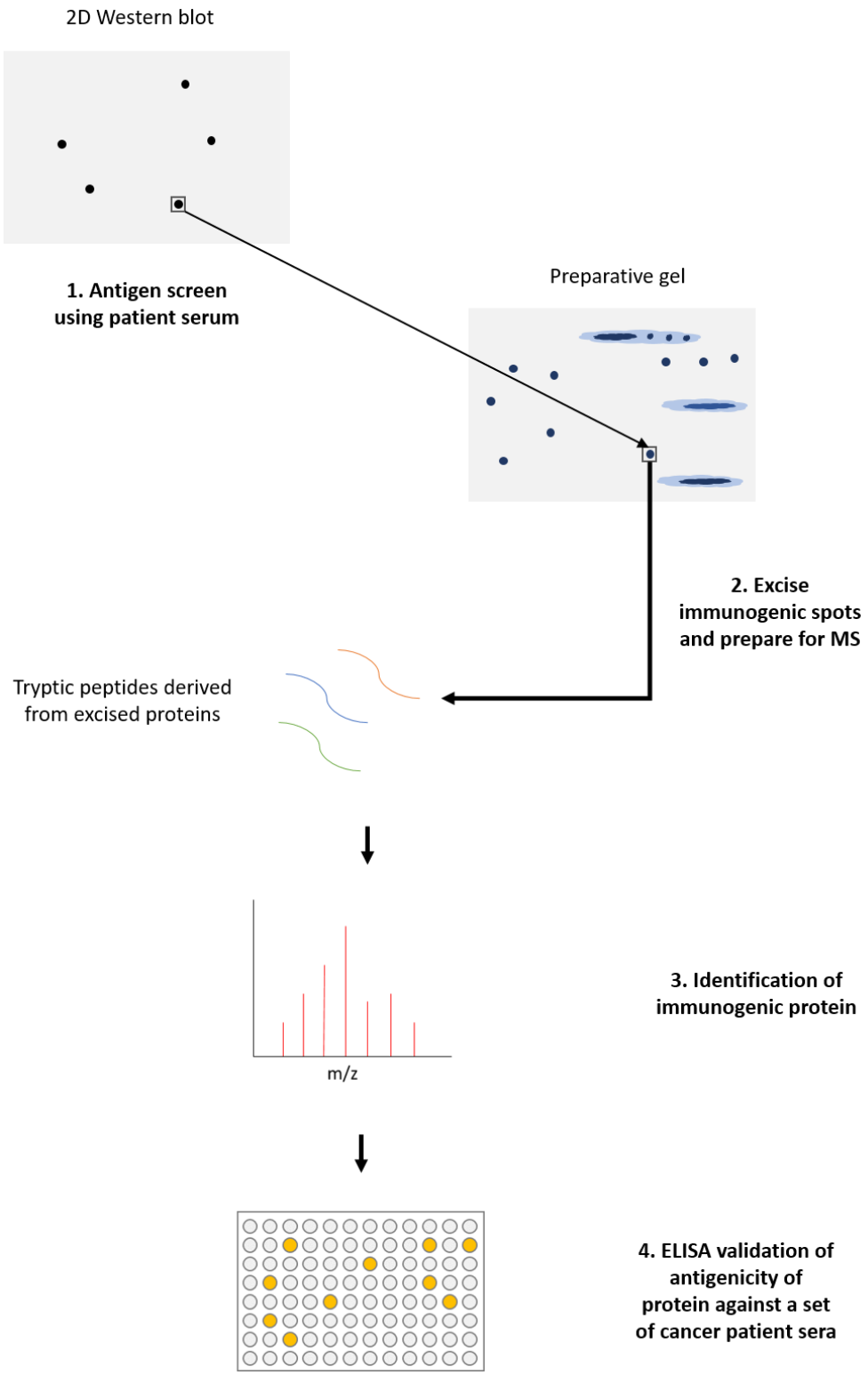


Figure 1.11. Serological proteome analysis (SERPA). Abbreviations include: 2D = two-dimensional; MS = mass spectrometry; ELISA = enzyme-linked immunosorbent assay [Adapted from (Gunawardana & Diamandis 2007)].

Protein microarrays are easily customisable, miniaturised platforms used to simultaneously characterise the biomolecular interactions of hundreds of proteins that are spotted in defined locations on a solid support. Microarrays allow the screening of a wide variety of analytes, such as antibodies, proteins, DNA, RNA, small molecules, lipids, enzymes, as well as peptides, allowing the quantification of analytes that bind to arrayed proteins. Typically, three types of protein microarrays are commonly used, including analytical, function and reverse-phase microarrays. Analytical protein arrays, or Ab arrays, are ideal for quantification of different proteins in a biological sample, monitoring protein expression levels and protein profiling in what amounts to miniaturised, highly multiplexed ELISAs. Functional protein arrays can be used for autoantibody and immune response profiling, biomolecular interaction profiling and the identification of enzyme substrates, amongst others (Duarte and Blackburn, 2017), whilst reverse phase protein arrays are suited for detection of altered proteins or other targets that contribute to disease. Moreover, protein microarrays can be applied in diagnostic and therapeutic research, through new biomarker discovery for disease staging and monitoring, potential drug-target evaluation and for identification of new drug targets.

Microarray slide surface chemistries vary: aldehyde and epoxy-derivatized glass surfaces are used for random attachment through amines, whereas nitrocellulose, hydrogel or metal surfaces are used for the attachment of affinity-purified proteins. An ideal surface chemistry should resist non-specific adsorption, provide sufficient stability for the 3D structures and allow the immobilisation of native proteins and provide highly specific linking chemistry that eliminates the need for protein pre-purification (Duarte and Blackburn, 2017).

The *E.coli* expression system rapidly produces proteins at relatively low costs, require simple and rapid culture conditions and are highly scalable. However, inefficient disulphide bond formation, insolubility, aggregation and poor folding of proteins have been reported using this method, as well as the minimal capability of forming post-translational modifications (Broadway, 2012). Yeast protein expression systems, e.g. *Saccharomyces cerevisiae* (*S. cerevisiae*) and *Pichia pastoris* (*P. pastoris*), are common alternatives to prokaryotic expression systems and fold efficiently and provide numerous PTMs (Broadway, 2012). However, the major disadvantage of yeast expression systems is that it does not mimic protein glycosylation patterns from

mammalian cells, with proteins tending to be hyperglycosylated. Moreover, yeast lysis conditions are harsh and induce many endogenous proteases, meaning that the extracted recombinant proteins can be significantly proteolyzed. Baculoviruses belong to a diverse group of large double-stranded DNA virus that infect many different species of insects as their natural host. Baculovirus stains are highly species-specific. The system shows good expression levels especially for intracellular proteins and typically produces soluble and functionally active recombinant proteins containing proper folding, disulphide bond formation and oligomerization, in addition to performing mammalian-like post-translation modifications. In particular, the glycosylation of proteins is more like mammalian cells, which indicates that the protein produced using this method is both structurally and functionally similar to its native counterpart (Broadway, 2012). Mammalian cells, e.g. HeLa, human embryonic kidney-derived (HEK293) epithelial cells, Chinese hamster ovary cells (CHOs) and African green monkey kidney cells (COS), are preferred by some researchers as they produce more "humanised" proteins. Despite producing eukaryotic recombinant proteins, mammalian expression systems require more demanding culture conditions compared to other systems (Broadway, 2012), and so are significantly more challenging for high throughput expression purposes.

Previously, Hanash *et al*. (2003) demonstrated the feasibility of manufacturing a natural protein microarray using chromatographic techniques and microarray printing technology. Briefly, cell lysate from a cancer cell-line is resolved using liquid-based isoelectric focusing into 20 fractions. Each fraction is further divided into 92 fractions using reverse-phase liquid chromatography. The fractionated proteins are lyophilized and re-suspended in a suitable buffer for printing on a nitrocellulose-based array. The protein array is screened using sera from cancer patients and healthy individuals as controls. Protein are subsequently identified using mass spectrometry (Figure 1.12) (Nam et al., 2003).

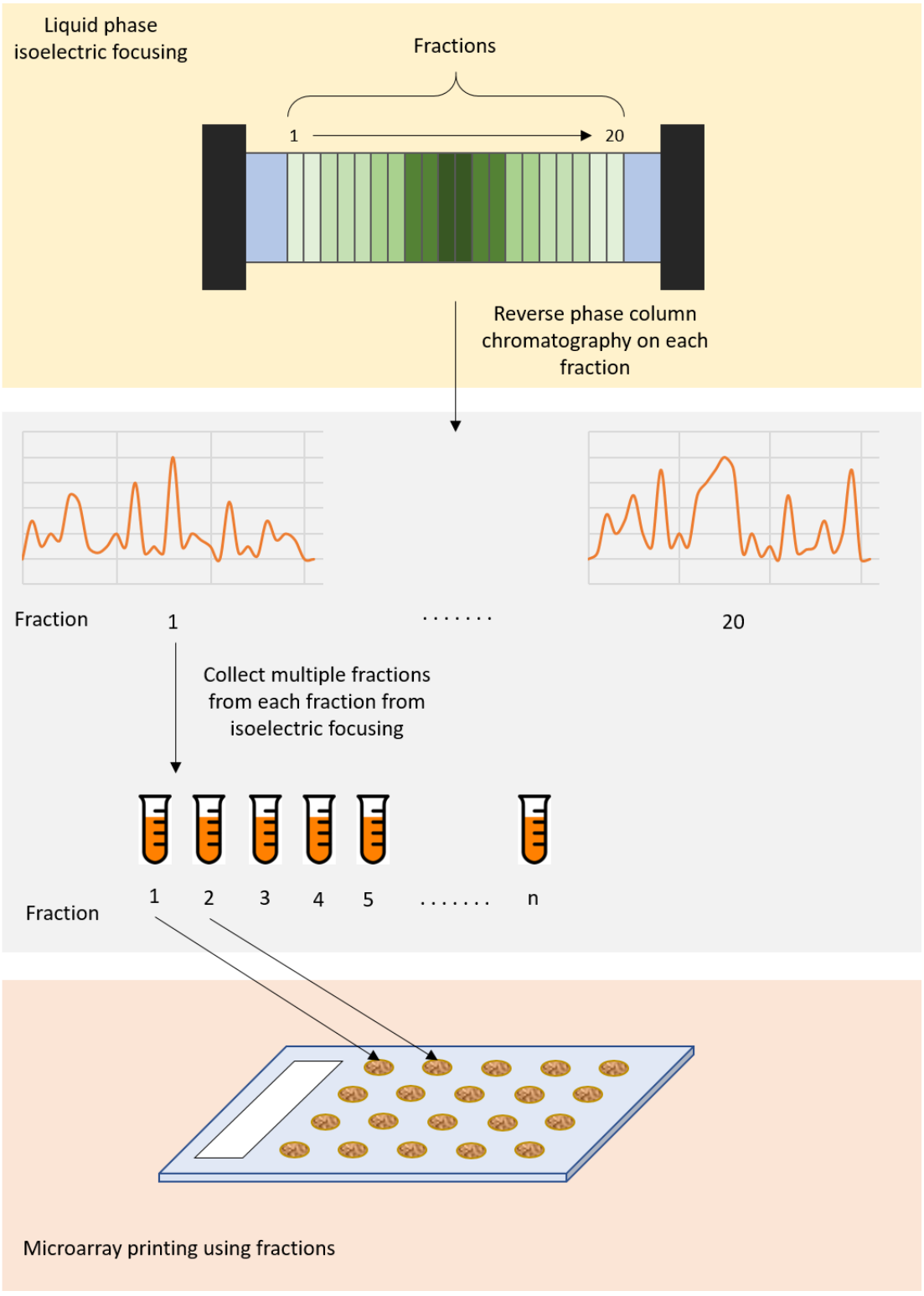


Figure 1.12. Overview of the natural protein microarray production strategy. [Adapted from (Gunawardana & Diamandis 2007)].

SEREX, phage display, SERPA and cell fractionation methods have provided advanced and effective means of identifying numerous novel TAAs that are applicable in the clinical setting, for cancer diagnosis and in some cases, cancer treatment. The most significant drawbacks include the use of non-native TAAs that are expressed in *E.coli* or proteins that have been denatured (e.g. in SERPA), resulting in missed TAAs due to the loss of conformational epitopes. Thus, to identify TAAs that can be used to reproducibly detect cancer, autoantibody and autoantigen detection methods should be improved to identify near-native proteins with relevant PTMs. An approach proposed in this thesis involves the use of affinity purification-mass spectrometry (AP-MS), whereby the plasma-derived IgG molecules are purified from CRC patients using Protein G beads, which is subsequently used to capture native cancer antigens from autologous patient lysate. Using the relevant controls, potential antigens enriched from cancer-specific AP-MS samples can be identified for use as cancer biomarkers (See Chapter 4).

1.7. Citrullination and cancer

Aberrant PTMs on cancer proteins can activate an autoimmune response in cancer patients. Examples of reported immune-activating PTMs include glycosylation, methylation, phosphorylation, sumoylation, adenosine diphosphate-ribosylation, ubiquitination, acetylation and citrullination. Although citrullination is not as extensively investigated as other PTMs e.g. glycosylation, aberrant citrullination of proteins has been reported for cancer patients (Chang et al., 2009; Baka et al., 2011; Assmann et al., 2014).

Protein citrullination is a calcium-dependent hydrolytic reaction that converts arginine to the citrulline amino acid, resulting in the loss of ammonia (Figure 1.13) (Rogers and Simmonds, 1958). The reaction is catalysed by the PAD family of proteins, consisting of 5 isozymes (PAD1-4 and 6), each with unique and overlapping target substrates, but with tissue-specificity (Table 1.2) (Knuckley et al., 2010; Darrah et al., 2012). Under physiological conditions, PADs are not normally active unless stimulated by calcium, resulting in the citrullinated of structural proteins e.g. vimentin, filaggrin and keratin (Inagaki et al., 1989; Senshu et al., 1995); and proteins involved in gene transcription e.g. histones H1, H2A, H3 and H4 (Wang, 2004; Saiki et al., 2009; Christophorou et al., 2014).

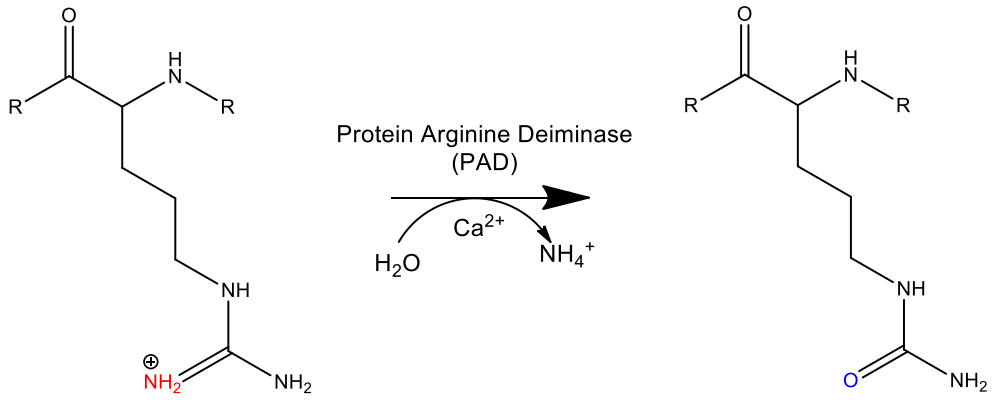


Figure 1.13. PAD-catalysed conversion of peptidyl arginine to peptidyl citrulline. PAD isozymes catalyse the calcium-dependent hydrolytic conversion of peptidyl arginine to peptidyl citrulline, with the loss of ammonia.

Table 1.2. Substrates and locations of PAD isozymes in humans. Abbreviations include: PAD = protein arginine deiminase. [Adapted from (Witalison et al., 2015)]

Isozyme	Known Substrates	Location
PAD1	Keratin and filaggrin	Epidermis, hair follicles and the uterus
PAD2	Myelin basic protein, vimentin, actin and histones	Central nervous system, spleen, skeletal muscle and leukocytes, colon and tumours
PAD3	Filagrin, trichohyalin, apoptosis-inducing factor and vimentin	Epidermis and hair follicles
PAD4	Histones, ING4, p300, p21, nucleophosmin and nuclear lamin C	Macrophages, neutrophils, mammary glands and tumours
PAD6	Unknown	Ovaries, eggs and embryo

Neeli *et al*. were the first to demonstrate the connection between histone acetylation and neutrophil extracellular "trap" (NET) release to confer protection against pathogens (Neeli et al., 2008), although it was later found that deiminated histones were antigens of human autoantibodies, aiding in pathogen removal (Dwivedi et al., 2012). Citrullinated proteins, and resulting autoantibody responses, have also been associated with autoimmune diseases, including rheumatoid arthritis (Schellekens et al., 1998), Alzheimer's disease (Ishigami et al., 2005) and cancer (Chang and Fang, 2010).

PAD2 and PAD4 expression has been reported to be dysregulated in multiple cancer-types. For example, PAD2 levels are reportedly elevated in breast cancer cell-lines (Zhang et al., 2012), whereas decreased PAD2 expression is reported for CRC and is associated with poor disease prognosis (Funayama et al., 2017). Increased PAD4 levels were reported for cancerous tissues, but not in benign tumours or non-cancerous inflammatory disease (Chang et al., 2009). Citrullination of cancer proteins culminates in an autoantibody response in lung cancer (Baka et al., 2011) and diffuse large B-cell lymphoma patients (DLBCL) (Assmann et al., 2014).

1.8. Cancer immunotherapy

Cancer immunotherapy is a biological therapy used to stimulate or suppress the immune system to aid in cancer eradication. The various arms of the immune system have been implicated in cancer detection and eradication.

Early *In vivo* mice experiments showed that T-cell populations cultured with leukaemia cells eradicate leukaemia when introduced into mice models (Cheever et al., 1981a, 1981b). Since then, clinical trials have provided evidence of a powerful anti-leukaemia response in patients treated with allogeneic hematopoietic cells (Wu et al., 2005; Wu and Ritz, 2006). In addition, metastatic melanoma patients displayed metastatic regression after treatment with $CD8^+$ T-cells specific toward melanoma proteins MART-1 (Yee et al., 2002; Morgan et al., 2006), Melan-A, and gp100 (Yee et al., 2002). Cancer-specific monoclonal antibodies can also produce potent anti-tumour activity in cancer patients.

Numerous studies have shown clinical efficacy of monoclonal antibodies against HER2-positive breast cancer (Piccart-Gebhart et al., 2005; Romond et al., 2005), B-cell lymphomas (Cheson and Leonard, 2008), as well as head and neck (Kim et al., 2004), non-small-cell lung (Giaccone, 2005) and colorectal (Mendelsohn, 1997) cancers. Compelling evidence confirms immune reactivity toward tumour antigens and other molecules, although patient responses are marginal, possibly due to late disease presentation (patients receive immunotherapy after other treatments have failed), and/or tumour immune evasion. Of particular interest to this study is tumour immune evasion through checkpoint inhibitors mediated by cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and programmed death 1(PD-1), discussed below (see Section 1.9).

1.9. Immune checkpoint inhibitors: CTLA-4 and PD-1

The discovery of immune checkpoint proteins such as CTLA-4 and PD-1, is regarded as a big step forward in cancer immunotherapy. The roles of these checkpoint inhibitors in immune deactivation towards tumours since their inhibition can reactivate the immune system against cancer. The roles of CTLA-4 and PD-1 in immune deactivation and as cancer vaccine targets are further discussed below:

Besides TCR and MHC interaction (See Section 1.3), T-cell activation requires co-stimulatory signals. One essential interaction occurs between CD28 on T-cells and B7-1 or B7-2 molecules on APC, which aids T-cell proliferation and survival, the production of growth cytokines (e.g. IL-2), and increased energy metabolism (Buchbinder and Desai, 2016). CTLA-4 is a CD28 homologue also found in or on T-cells, but has a higher affinity toward B7 molecules than CD28 (Chambers et al., 2001; Collins et al., 2002). The interaction between CTLA-4 and B7-1 or B7-2 molecules prevents T-cell proliferation and survival, the production of growth cytokines, and increased energy metabolism. On naïve T-cells, CTLA-4 is normally located in the cytosolic side of the cell membrane (Linsley et al., 1996). However, T-cell activation and stimulatory signals result in increased CTLA-4 levels on the T-cell surface (Krummel and Allison, 1996). The ratio of CD28 to B7 versus CTLA-4 to B7 binding that determines T-cell activation or no immune response (also known as anergy) (Krummel and Allison, 1995) (Figure 1.14). The exact molecular mechanism of CTLA-4 remains controversial: molecular studies indicate that CTLA-4 may function by recruiting phosphatases to TCRs resulting in signal attenuation (Lee et al., 1998), and/or by modulating cell motility and signalling through PI3 kinase (Knieke et al., 2012). Other *in vivo* studies suggest that CTLA-4 captures and removes B7 from the membranes of APC, preventing CD28 and B7 interaction required for T-cell activation (Qureshi et al., 2011).

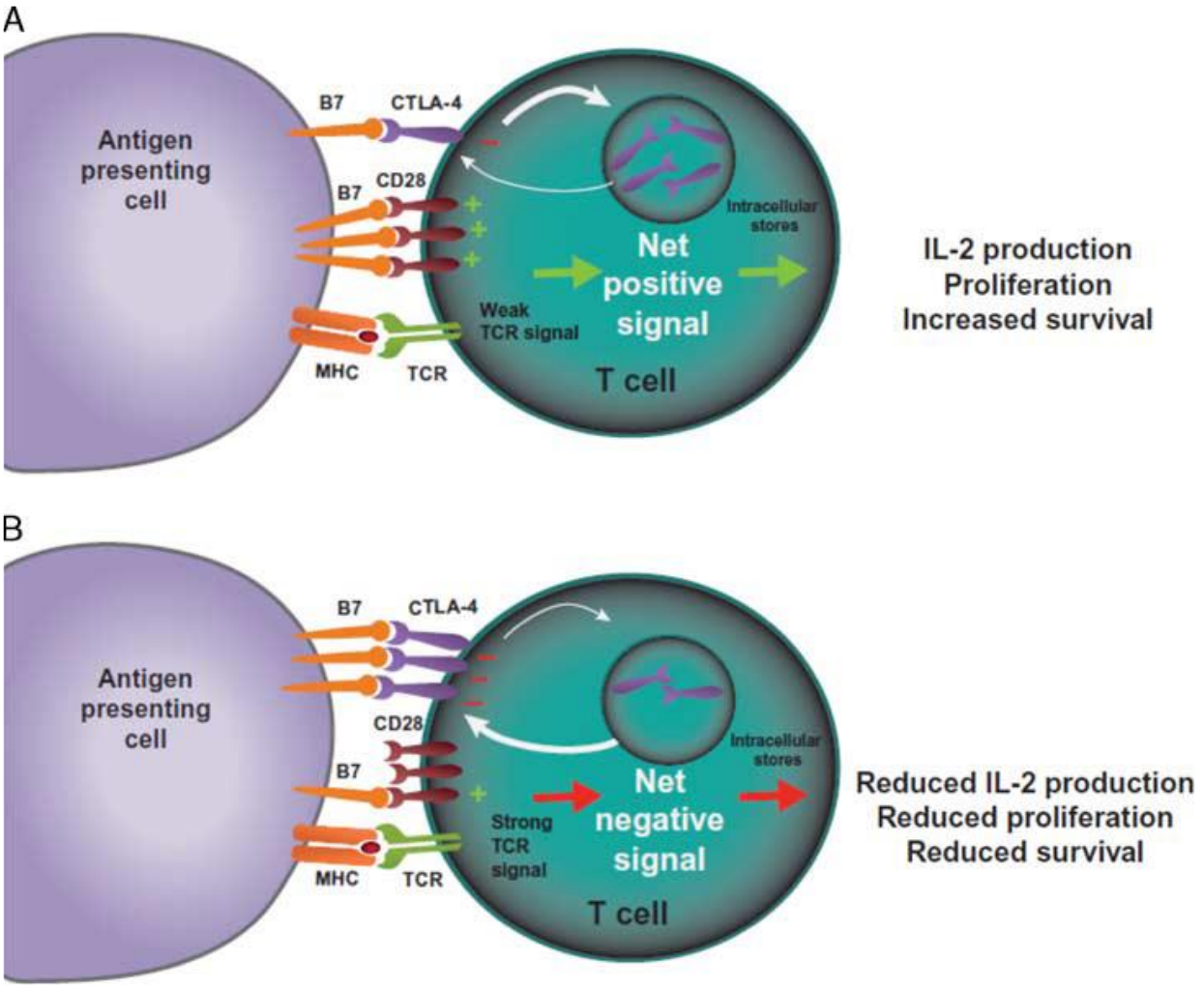


Figure 1.14. CTLA-4-mediated inhibition of T-cell. T-cells become activated upon interaction between TCRs and peptide antigens presented on MHC molecules, in concert with CD28 and B7-mediated co-stimulation. (**A**). A weak TCR stimulus in which CD28 and B7 binding results in IL-2 production, proliferation and increased survival. (**B**) Overstimulating TCRs result in CTLA-4 expression and transport to the cell surface. CTLA-4 has a high affinity for B7, and the interaction results in decreased IL-2 production, T-cell proliferation and survival. [Adapted from (Buchbinder & Desai 2016)]

PD-1 regulates T-cell activation by binding to its ligands, programmed death ligand 1 (PD-L1) and programmed death ligand 2 (PD-L2). Upon binding, the PD-1/PD-L interaction generates signals that prevent phosphorylation of key TCR signalling intermediates, reducing T-cell activation (Bennett et al., 2003; Parry et al., 2005). Like CTLA-4, PD-1 binding to its ligands also inhibits T-cell proliferation and survival, but also inhibits IFN-γ, TNF-α and IL-2 production (Keir et al., 2008)

(Figure 1.15). PD-1 is normally expressed on T-cells after high levels of stimulation, leading to suboptimal control of chronic infections and cancer (Wherry, 2011).

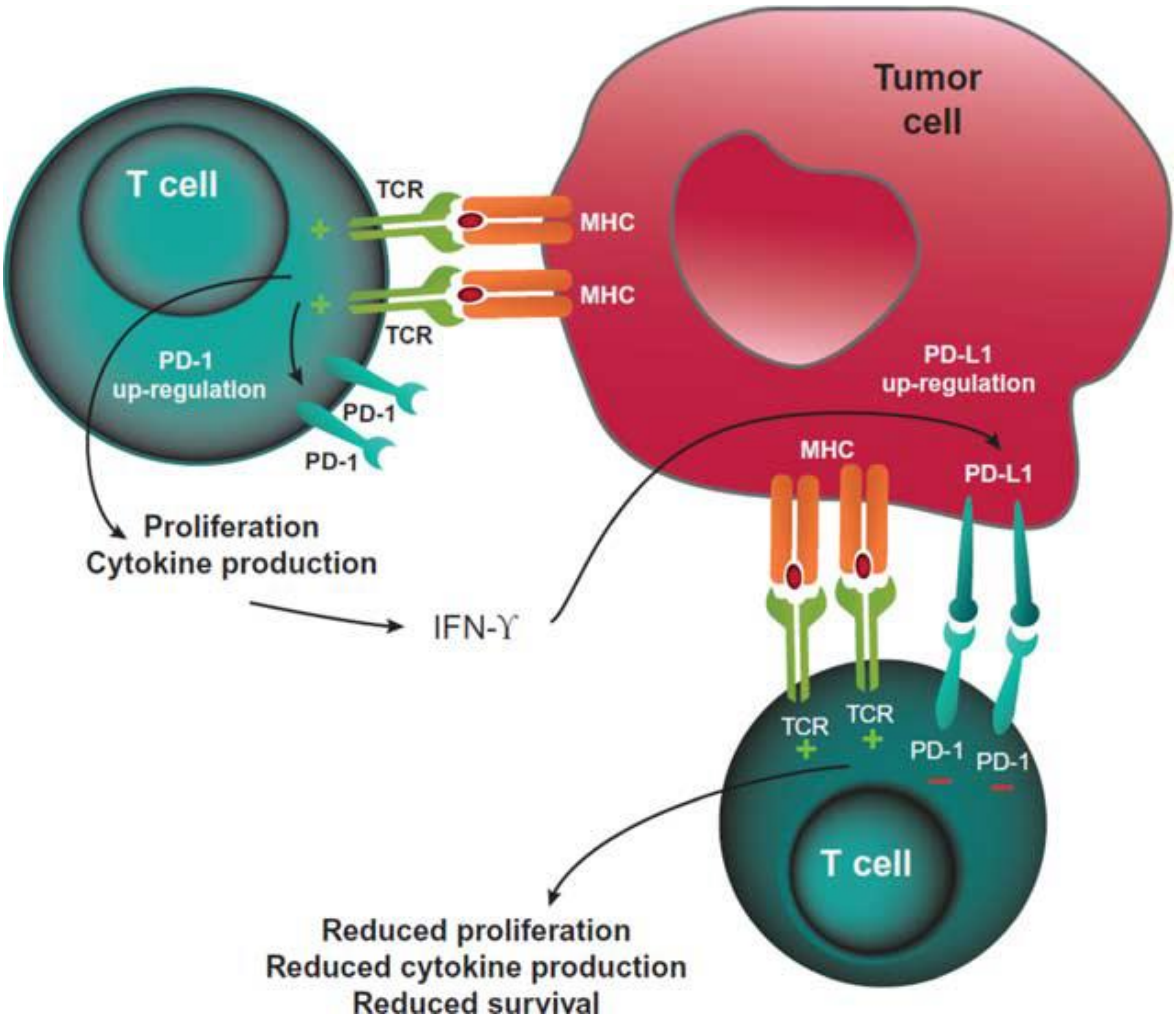


Figure 1.15. PD-1-mediated inhibition of T-cells. Prolonged TCR stimulation causes the upregulation of PD-1 expression. Tumours cells expressed PD-L1 and PD-L2 (not shown) in response to inflammatory cytokines and/oncogenic signalling pathways. This interaction inhibits TCR-mediated T-cell activation, leading to decreased cytokine production, as well as T-cell proliferation and survival. [Adapted from (Buchbinder & Desai 2016)]. Abbreviations include: PD-1 = programmed death 1; PD-L1 = programmed death ligand 1; PD-L2 = programmed death ligand 2; TCR = T-cell receptor; IFNγ = interferon γ.

Both CTLA-4 and PD-1 are also constitutively expressed on T_{reg} cells, resulting in the dampening of immune responses. While CTLA-4 expression is restricted to T-cells, PD-1 is more broadly expressed on activated T-cells, B-cells and myeloid cells. Also, whilst B7 is restricted to APCs, PD-

1 ligands are expressed by tumour cells and tumour-infiltrating lymphocytes. Lastly, CTLA-4 appears to function in the priming phase of T-cell activation in the lymph nodes and spleen, whereas PD-1 functions during the effector phase in peripheral tissues.

Preclinical studies with CLTA-4 and PD-1 blockade led to decreased tumour burden, thus, immune checkpoint inhibitors were further developed for cancer patient treatment. Monoclonal antibodies that block CLTA-4 and PD-1 have been developed for melanoma, non-small lung-, kidney-, prostate- and head and neck-cancers. Agents targeting PD-L1 are also being developed.

Ipilimumab was the first anti-CTLA-4 blockade to prolong survival in patients with advanced melanoma (Hodi et al., 2010; Robert et al., 2011). Long term survival analysis indicated a 3-year survival of 22%, 26% and 20% in all patients with sufficient follow-up, in treatment of naïve patients, and in previously treated patients, respectively (Schadendorf et al., 2015). PD-1 blockade has improved survival for metastatic melanoma, non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC) patients (Herbst et al., 2014; Topalian et al., 2014; Garon et al., 2015; Gettinger et al., 2015; McDermott et al., 2015). In one trial, advanced melanoma patients treated with pembrolizumab (anti-PD-1 antibody) and ipilimumab (anti-CTLA-4 antibody) showed a superior response in pembrolizumab-treated versus ipilimumab-treated patients, with response rates of 34% and 12%, respectively. In conjunction with this data, patient survival rates were also high in pembrolizumab versus ipilimumab treatment, at 74% and 58%, respectively (Farolfi et al., 2012). Nivolumab is a human IgG4 anti-PD-1 monoclonal Ab that selectively blocks the interaction between PD-1 and PD ligands, restoring T-cell immune activity directed against cancer. Nivolumab was the first checkpoint inhibitor approved for advanced NSCLC following chemotherapy, resulting in increased response rates, survival and progression-free survival when compared to intravenous docetaxel alone. Furthermore, nivolumab was better tolerated than docetaxel, as nivolumab-induced immune adverse events were more manageable (Keating, 2015). Stage III/IV melanoma patients achieved a partial tumour response, with a median progression free survival of 172 days, with only 18% experiencing grade 2 or 4 adverse events (Deeks, 2014).

Combination check-point inhibitor treatments, targeting both CTLA-4 and PD-1, have also shown strong promise, with clinical trial data in untreated melanoma patients reporting objective response rates up to 72% (amongst patients with PD-L1-positive tumours) and with median progression-free survival for 11.5 months for ipilimumab plus nivolumab, compared to 2.9 months with ipimilumiab alone and 6.9 months with nivolumab alone (Larkin et al., 2015). However, high grade immune-related adverse events (irAEs) occurred in 55% of those in the combination treatment group (Larkin et al., 2015) and similarly high rates if irAEs have been reported elsewhere for anti-CTLA-4 and anti-PD-1 treatments (Michot et al., 2016).

Biomarkers that could predict treatment outcomes would be hugely beneficially in providing optimal patient treatment, whilst limiting potential adverse treatment-side-effects. Exhausted T-cells are defined as T-cell dysfunction that arises during chronic infections and cancer, defined poor effector function, sustained expression of inhibitory receptors and a transcriptional state distinct from functional effector or memory T-cells (Wherry, 2011). Upregulation of PD-1 on exhausted-cells and of PD-L ligands on tumour cells, and/or tumour infiltrating cells, may serve as biomarkers that predict patient response to PD-1 or PD-1 ligand vaccines. Preliminary results of multiple tumour-types indicate that patients with PD-L1-expressing tumours or infiltrating immune cells typically have higher response rates to anti-PD-1 or anti-PD-L1 versus patients with low or no PD-L1 expression (Herbst et al., 2014; Taube et al., 2014), however, these results remain controversial as patients with NSCLC did not show a correlation between PD-L1 levels and treatment outcomes (Borghaei et al., 2015; Brahmer et al., 2015; Herbst et al., 2016; Passiglia et al., 2016). It has been more challenging to identify predictive biomarkers for ipilimumab-treatment, due to low levels CTLA-4 expression, and wide range expression of B7 ligands. Retrospective studies have identified markers associated with treatment response, including absolute lymphocyte count, increased levels of T-cell activation marker ICOS, and T-cell response to NY-ESO-1 (Callahan et al., 2013), although neither biomarker has since been validated. In addition, a study concluded that neoantigen presentation on tumours correlates with overall survival in CTLA-4-treated melanoma patients (Snyder et al., 2014). Together, it is evident that reliable biomarkers for treatment response remain to be discovered in order to enable beneficial

management of patient treatment, thereby improving treatment response rate and reducing immune related adverse events.

Hypothesis

The underlying hypothesis for the studies described in this thesis is that measurable differences in autoantibody repertoires can be detected in CRC and melanoma patients compared to HCs, which alone or as part of a biomarker panel, reproducibly detects cancer. In addition, we hypothesise that the autoantibody response against aberrant cancer proteins can also be used to assess disease prognosis and response to therapy.

Aims and objectives

The overall aim of this study is to identify TAAs associated with CRC and melanoma, that can reproducibly detect cancer and predict disease prognosis. The specific objectives therefore are:

i. To develop a robust statistical pipeline that enables the assess the biomarker potential of TAAs identified using microarray technology.
ii. To assess the autoantibody response of CRC patients on the CT100plus microarray, and use the established statistical pipeline to identify top biomarker candidates linked to disease diagnosis and prognosis.
iii. To assess the autoantibody response of CRC patients toward citrullinated antigens using the CT100plus microarray, and use the established statistical pipeline to identify top biomarker candidates linked to disease diagnosis and prognosis.
iv. To develop and test an AP-MS assay for the detection of TAAs associated with CRC, and compare the results to responses observed on the CT100plus microarray.
v. To assess the autoantibody response of melanoma patients to checkpoint inhibitor treatment, and use the established statistical pipeline to identify top biomarker candidates of response to therapy.

Chapter 2

Analysis of CRC patient plasma on the CT100plus microarray

Abstract

Cancer patients elicit an autoantibody response toward TAA. The aberrant TAAs arise due to altered tissue-specific expression patterns, mutations, PTMs, *inter alia*, and induce autoimmune and autoantibody responses in the host. The resulting autoantibody response in cancer patients can in principle be used for cancer diagnosis and predicting disease prognosis. In this study, we assessed the autoantibody response of 57 CRC patients and 14 HCs on a custom CT100plus microarray. Significantly (p-value ≤ 0.05, adjusted p-value ≤ 0.05) higher autoantibody signals were detected in CRC patients *versus* HCs toward CEACAM 1, COL6A1, GRWD1, MAGEA1, MAGEA5, MAGEA10, NY-CO-1, SGY-1, SPANXB1 and THEG. Furthermore, receiver operator characteristic (ROC) analyses indicate area under the ROC curve (AUC)-, sensitivity- and specificity-ranges of 0.81 – 0.86, 0.70 – 0.88 and 0.64 – 0.86, respectively. Combined ROC results indicate that CEACAM1 and GRWD1 were the best autoantigen combination, producing sensitivity-, and specificity-, and AUC-values of 1.00, 0.77 and 0.94, respectively. Multivariate PCA and unsupervised hierarchical clustering analyses resulted in the identification of distinct clusters of CRC patients and HCs. Interestingly, COL6A1, THEG and CEACAM7, a homologue of CEACAM1, were also later identified by AP-MS (See Chapter 4), providing supporting evidence for biomarker identification. Patient clinical information was available for the CRC patients, although no significant association was found for patient clinicopathological features and antigenic signal. Together, we provide strong evidence using orthogonal methods that CRC patients elicit an autoantibody response to COL6A1, THEG and CEACAM1, which have potential use in CRC cancer diagnosis and treatment.

Introduction

CRC refers to cancers of the colon and rectum. CRC is a major cause of morbidity and mortality throughout the world, being the third most common cancer worldwide and the fourth most common cause of cancer-related deaths. The likelihood of CRC increases after the age of 40, with more than 90 % of CRC cases presented by patients aged 50 years or older. World-wide, CRC represents 10.1 % of all cancer incidences in women and 9.4 % in men. Although most CRC cases occur in developed countries e.g. the United States, Australia, New Zealand and Western Europe, an increase in the incidence number has been observed in developing countries. CRC is generally considered a cancer-type associated with the environmental risk factors e.g. cultural, social and lifestyle factors. Approximately 5-10% of CRC are hereditary, with the most common inherited conditions being hereditary nonpolyposis colorectal cancer (HNPCC) (accounts for ~ 2-6 % of CRC cases) and familial adenomatous polyps (FAP) (accounts for ~ 1 % of CRC cases) (Jackson-Thompson et al., 2006). Genes for hereditary CRC have been identified, with HNPCC being associated with genes involved in DNA repair (*MLH1* and *MSH2*) (Papadopoulos et al., 1994), and FAP being caused by mutations in the tumour suppressor gene *APC* (Wilmink, 1997). CRC typically starts as a non-cancerous polyp, which develops in the lining of the colon or the rectum. Adenomous polyps give rise to adenocarcinomas, which makes up about 96 % of CRC cases (Levine and Ahnen, 2006). Once the cancer forms in the large intestine, it can grow through the intestinal lining into the wall of the colon or rectum. Cancers which have invaded the wall are also able to penetrate the blood or lymph vessels, leading to cancer metastasis (Siegel et al., 2014).

CRC diagnosis typically starts with a positive faecal occult blood test, or observed blood in the faeces, followed by a colonoscopy. The patient medical history is also collected, including family history of CRC, and then a physical exam. Blood tests are often performed, including a complete blood count (CBC), as CRC patients tend to have anaemia due to tumour bleeding. Liver enzymes are also assessed as CRC can spread to the liver (www.cancer.org). Lastly, the tumour marker, CEA, is used for CRC diagnosis, although it reportedly has low sensitivity and specificity (Liu et al., 2009). An autoantibody response is activated toward cancer-related aberrant proteins, including

TSA and TAA. Autoantibodies against Fas show high specificity for early-stage CRC. In addition, higher anti-Fas titres were detected in patients with colorectal adenomas versus colorectal adenocarcinomas. The resulting sensitivity- and specificity-values of 17% and 100% for CRC (Reipert et al., 2005) suggest that Fas may not be useful to detect CRC, but rather to confirm absence of disease. An ELISA-based assay was used to detect differential abundance of MUC5AC between CRC patients, disease controls and healthy volunteers, producing sensitivity- and specificity-values of 73% and 54%, respectively (Kocer et al., 2006). Autoantibodies against HSP60 (He et al., 2007), p53 (Cioffi et al., 2004; Yoshizawa et al., 2007; Belousov et al., 2008) and Calnuc (Chen et al., 2007) have also been detected in CRC patients, although autoantibodies against these autoantigens are also detected in patients with other cancer-types, indicating that these biomarkers are not specific to CRC.

The Blackburn lab created a customised cancer antigen microarray, called the CT100plus microarray. The microarray is fabricated with biotinylated native cancer-testis antigens and tumour associated antigens, printed in triplicate within 8 subarrays. Each subarray contains 5, 10 and 15 ng/µl Cy5-labelled biotinylated bovine serum albumin (BSA) which is used for downstream normalization. Other controls include anti-human IgG antibody, used to confirm the presence of human serum/plasma and human IgG, used to confirm the addition of the detection Ab (fluorescently labelled anti-human IgG). The negative controls include lysis buffer only, insect cell lysate (ICL) and the vector containing the biotin carboxyl carrier protein (BCCP)-tag only (Figure 7.4).

The CT100plus microarray is built on a glass slide coated with a hydrogel layer for the covalent immobilisation of proteins. The hydrogel layer is a permeable polymer coating that helps preserve the three-dimensional (3D) structure of proteins, thus maintaining protein conformation and functionality. In-house slide derivatisation resulting in a uniform streptavidin layer covalently bound to the hydrogel, which greatly diminishes non-specific binding. Biotinylated cancer proteins are produced in insect cells to maintain the eukaryotic integrity of proteins. The whole ICL is spotted on the microarray and washed, resulting in the purified cancer protein remaining within the spot. The sandwich layout of the CT100plus microarray is depicted in Figure 2.1. Previously, the CT100plus microarray has been used to assay the serum samples of

melanoma (Duarte et al, unpublished data) and prostate cancer (Adeola et al., 2016) patients, and in this thesis was used to assess the autoantibody profiles of CRC (Chapter 2 and 3) and stage IV melanoma (Chapter 5) patients.

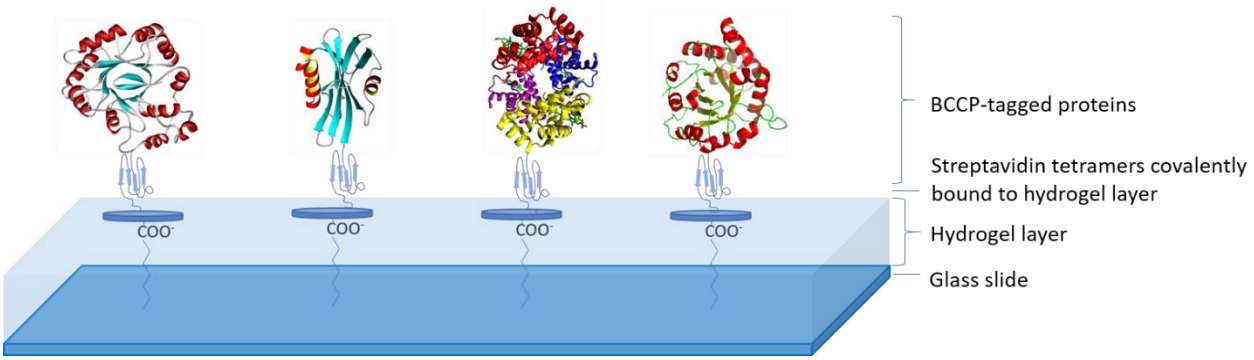


Figure 2.1. Depiction of the slide surface chemistry of the CT100plis microarray. Individually purified BCCP-tagged proteins are immobilized onto custom hydrogel-coated glass slide surfaces such that they retain their folded structure and function in an aqueous environment and behave as though they are in free solution.

The Blackburn group previously collected plasma (with buffycoat) samples, CRC and normal mucosa tissues, and clinical information from 63 CRC patients undergoing colonic resection surgery at Groote Schuur Hospital (GSH) (HREC 269/2011). The clinical information was gathered within a database format by Dr. Katie Lennard and Dr. Ryan Goosen. Genomic and bacteria-association studies on the patient-matched CRC tissue samples were conducted by Dr. Katie Lennard. The 63 plasma samples were assayed and analysed by Dr. Jessica Duarte on a custom CT100plus microarray using a 1000-relative fluorescence units (RFUs) cut-off. In this Chapter, the CT100plus-derived CRC data is re-analysed by comparison to HC data to identify the top 10 antigens for CRC diagnosis. The analysis pipeline includes ROC analysis, hierarchical clustering analysis, principle component analysis (PCA), and Kruskal-Wallis and Benjamini-Hochberg analysis (Figure 2.2). In addition to the clinical information provided for each patient, the microsatellite (MS) status of patients, identified by Dr. Katie Lennard, were also included in the analysis. Lastly, the obtained results were compared to AP-MS results (see Chapter 4) to search for potential CRC biomarkers identified using immunoproteomics as an orthogonal approach.

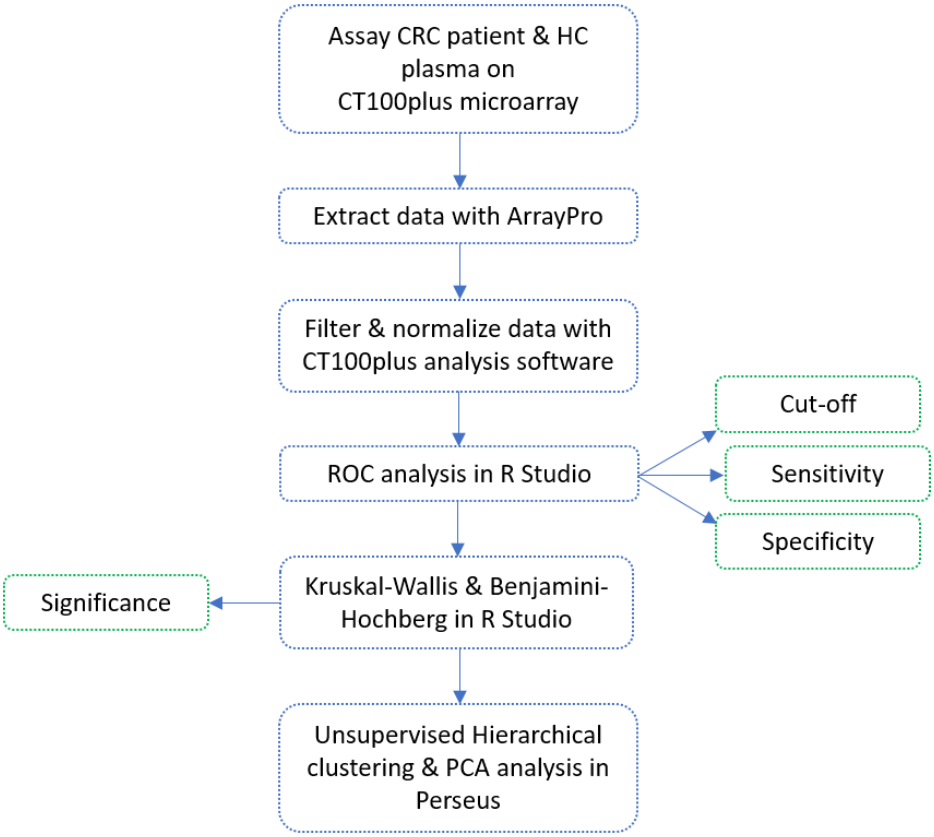


Figure 2.2. Data analysis pipeline. Sixty-three plasma samples were previously assayed (Duarte *et al*, unpublished data) on the CT100plus microarray, and potential CRC biomarkers were identified using a 1000-RFU cut-off. In this Chapter, the CT100plus-derived CRC data were re-analysed by comparison to HC data to identify the top 10 antigens for CRC diagnosis. The analysis pipeline includes ROC analysis, hierarchical clustering analysis, PCA, Kruskal-Wallis and Benjamini-Hochberg analysis, described in this Figure. Abbreviations include: CRC = colorectal cancer; HC = healthy control; RFU = relative fluorescence units; ROC = receiver operator characteristic; PCA = principle component analysis.

Results

2.1. CT100plus microarray quality controls

A total of 63 plasma samples, derived from individual CRC patients (i.e. single timepoint (TP) per patient), were assayed on the CT100plus microarray. Three controls assays were performed, using Cyanine 3 (Cy3)-labelled anti-c-Myc Ab (Figure 2.3), pooled negative control plasma (Figure 2.4), or pooled positive control plasma (Figure 2.5). Each of the 123 recombinant proteins on the CT100plus microarray (Figure 7.4) contains a c-Myc tag that can be used to confirm the successfully immobilisation of recombinant proteins on the slide surface with an anti-c-Myc antibody. It is evident from Figure 2.3 that antigens do not have the same signal intensity across the microarray. This is because each of the 123 recombinant proteins is expressed at different levels within insect cells and are not pre-normalised prior to array fabrication. Although the signal for each antigen is not visibly discernible on the image, protein expression was confirmed by Western blot analysis (Beeton-Kempen et al., 2014). The antigens are in a 3D conformation on the microarray, and in some cases, the c-Myc tag may be buried with in the protein structure, preventing interaction between the c-Myc tag and the anti-c-Myc antibody; thus, signal would be reduced for these antigens on the CT100plus microarray.

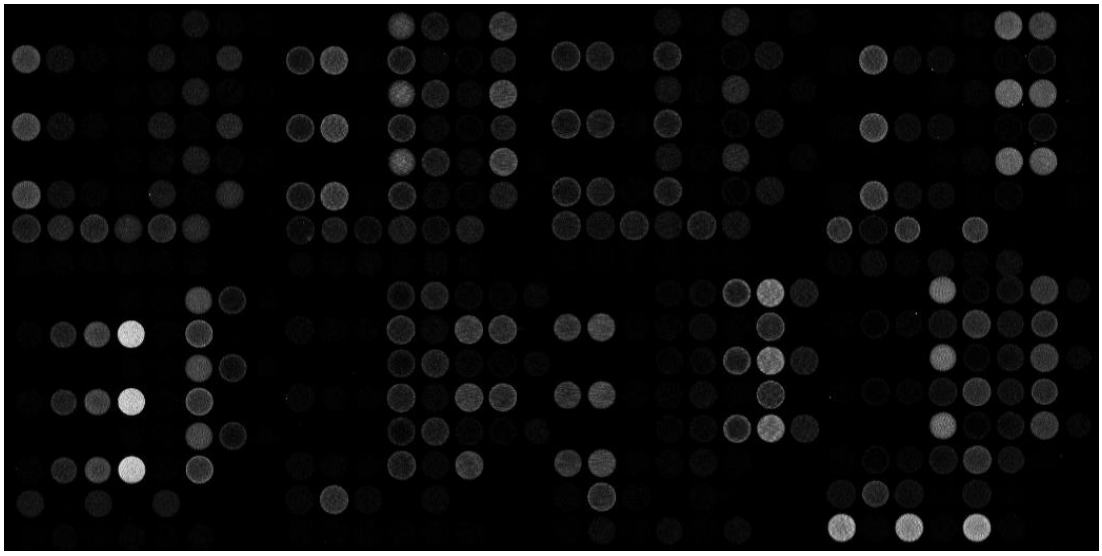


Figure 2.3. CT100plus microarray c-Myc control. Each of the 123 recombinant proteins on the CT100plus microarray contains a c-Myc tag. The microarray was treated with Cy3-labelled anti-c-Myc antibody, which was used to confirm the successfully immobilisation of recombinant proteins on the slide surface with an anti-c-Myc antibody, as depicted in the microarray image.

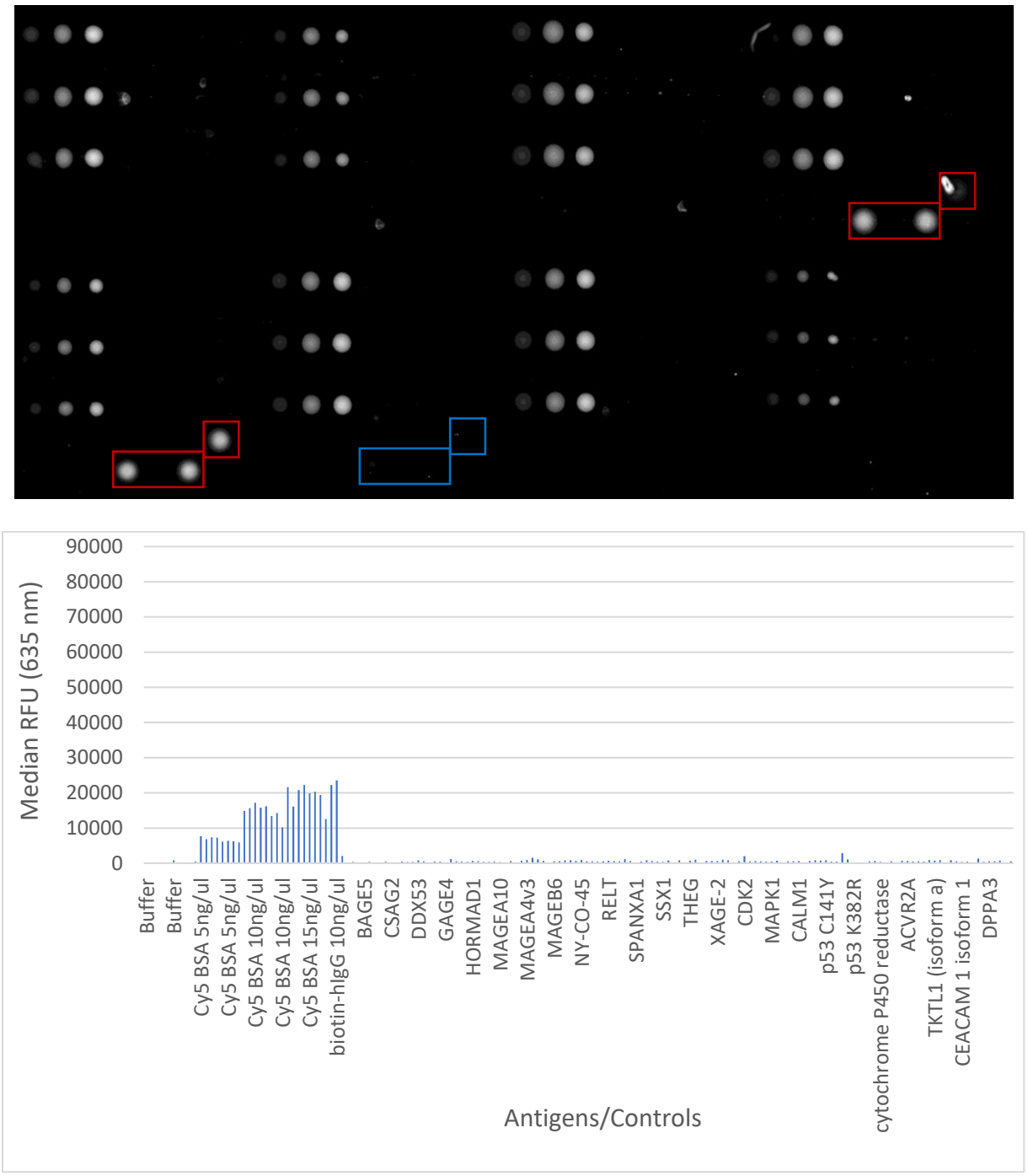


Figure 2.4. CT100plus microarray negative control pooled plasma sample. As a negative control, the CT100plus microarray was incubated with pooled plasma of HCs. Low or no signal is detected for the buffer ICL, and BCCP negative control spots, indicating low levels of non-specific binding. High signal was detected for the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots, anti-human IgG was detected (boxed in red), and human IgG was detected (boxed in blue). Low or no signal is detected for recombinant proteins. Abbreviations include: ICL = insect cell lysate; BCCP = biotin carboxyl carrier protein.

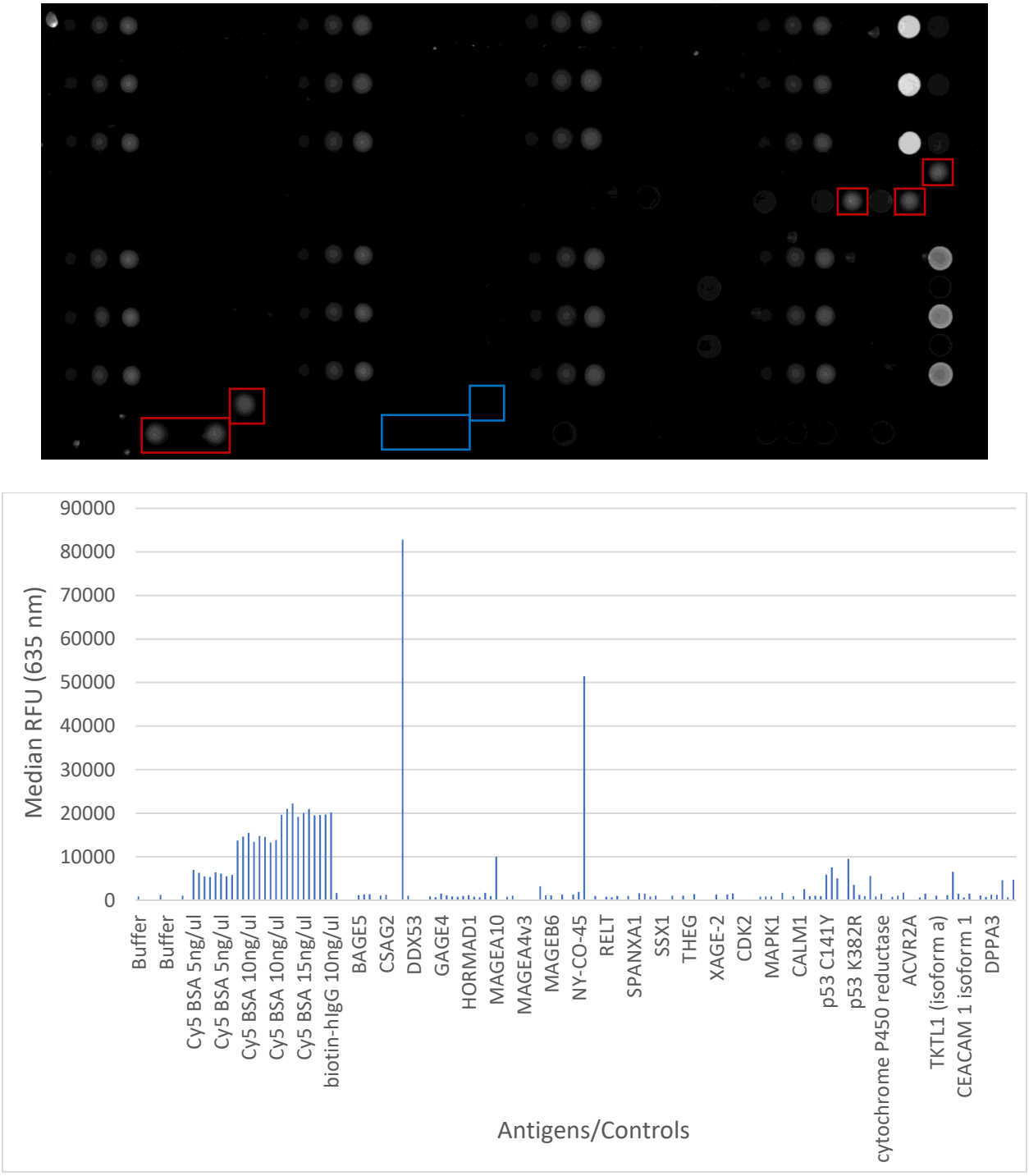


Figure 2.5. CT100plus microarray positive control pooled plasma sample. For the positive control sample, the CT100plus microarray was incubated with pooled plasma of cancer patients. Low or no signal is detected for the buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. High signal was detected for the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots, anti-human IgG was detected (boxed in red), and human IgG was detected (boxed in blue). Distinct and varying autoantibody signal intensity was detected for recombinant proteins. Abbreviations include: HCs = healthy controls; ICL = insect cell lysate; BCCP = biotin carboxyl carrier protein.

Both the positive and negative control assays display low or no signal from buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. Furthermore, the 5, 10 and 15 nanograms per microliter (ng/µl) Cy5-biotin-BSA positive control spots were detected, and used for subsequent downstream data normalisation steps; anti-human IgG was detected, confirming the addition of patient plasma; and human IgG was detected, confirming the addition of detection antibody. Low or no antigen signal was detected for the pooled negative control (Figure 2.4), whereas high signals were detected for antigens within the pooled positive control (Figure 2.5), indicating a high-quality platform that can be used for the specific detection of autoantibodies against CT100plus cancer antigens.

2.2. ROC and hierarchical clustering analysis of CRC patients on CT100plus microarray

The blood plasma samples of CRC patients were previously processed on the CT100plus microarray (Duarte *et al*., unpublished data) as per the methodology described in Chapter 7. The raw data files were extrapolated using the ArrayPro software, and data filtering and normalisation was performed with the CT100plus software with the following criteria: whole array coefficient of variation (CV) ≤ 25%; probe replicate CV ≤ 20%; noise-threshold ≥ 2 standard deviations (SDs) above background; maximum threshold = 65355 nm; whole array filtering control = 15 ng/µl Cy5-biotin-BSA; 57 patient data files passed the filtering and normalisation quality control procedure. The median signal intensity for each antigen was obtained and used for subsequent ROC and hierarchical clustering analyses.

A ROC analysis was performed, which is a sensitivity and specificity report of a tested parameter, which in this study refers to antigens. The ROC analysis produces an AUC-value, a measure of the discriminating efficacy of antigen for patients from HCs; a sensitivity-value, indicating the fraction of patients with high cancer antigen intensity; and a specificity-value, indicating the fraction of HCs with no or significantly lower signal intensity for antigens (Metz, 1978; Vining and Gladish, 1992; Zweig and Campbell, 1993). The higher the AUC-, sensitivity- and specificity-values, the better an antigen distinguishes patients from HCs. The ROC curve is widely accepted for selecting an optimal cut-off point. The cut-off point is used to determine the presence or absence of a disease, where values above the cut-off are disease-positive, and values below the cut-off are

considered disease-negative (Akobeng, 2007). An *a priori* ROC sample size was performed using MedCalc software (version 17.2), with a hypothesised AUC-value of 0.80, a null hypothesis value of 0.5, and a sample ratio of 0.25. The results produced a α-value (significance) of 0.05 and β-value (1-Power) of 0.01 for the 57 CRC patients and 14 HCs, indicating that our sample size was sufficient for the ROC analysis.

The ROC analysis was performed in R Studio (1.0.136) using the ROCR package (Sing et al., 2005) to produce AUC-, sensitivity- and specificity-values for all 123 CT100plus antigens. A top 10 protein list was generated from proteins with the highest AUC-values, which included CEACAM 1 (isoform 1), COL6A1, GRWD1, MAGEA1, MAGEA10, MAGEA5, NY-CO-45, SGY-1, SPANXB1 and THEG (Table 2.1). These 10 proteins were selected for further ROC analysis in CombiROC (Mazzara et al., 2017) to determine the combination of proteins that produce the highest AUC-, sensitivity-, and specificity-values. Here, the CombiROC analysis results indicate that a combination of CEACAM1 and GRDW1 produced the highest AUC-, sensitivity- and specificity-values of 0.94, 1.0 and 0.77, respectively. Figure 2.6 represents the combined ROC curve for CEACAM1 (isoform 1) and GRWD1.

Table 2.1. Sensitivity-, specificity-, AUC-, and cut-off-values for top 10 antigens for CRC patients on the CT100plus microarray. Abbreviations include: AUC = area under the receiver operator characteristic curve.

Antigen	Sensitivity	Specificity	AUC	Cut-off
CEACAM 1 (isoform 1)	0.88	0.86	0.86	408
COL6A1	0.70	0.79	0.82	513
GRWD1	0.82	0.64	0.82	471
MAGEA1	0.82	0.71	0.81	452
MAGEA10	0.72	0.86	0.83	467
MAGEA5	0.88	0.86	0.88	481
NY-CO-45	0.72	0.86	0.83	558
SGY-1	0.82	0.71	0.84	542
SPANXB1	0.82	0.79	0.84	441
THEG	0.81	0.86	0.86	474

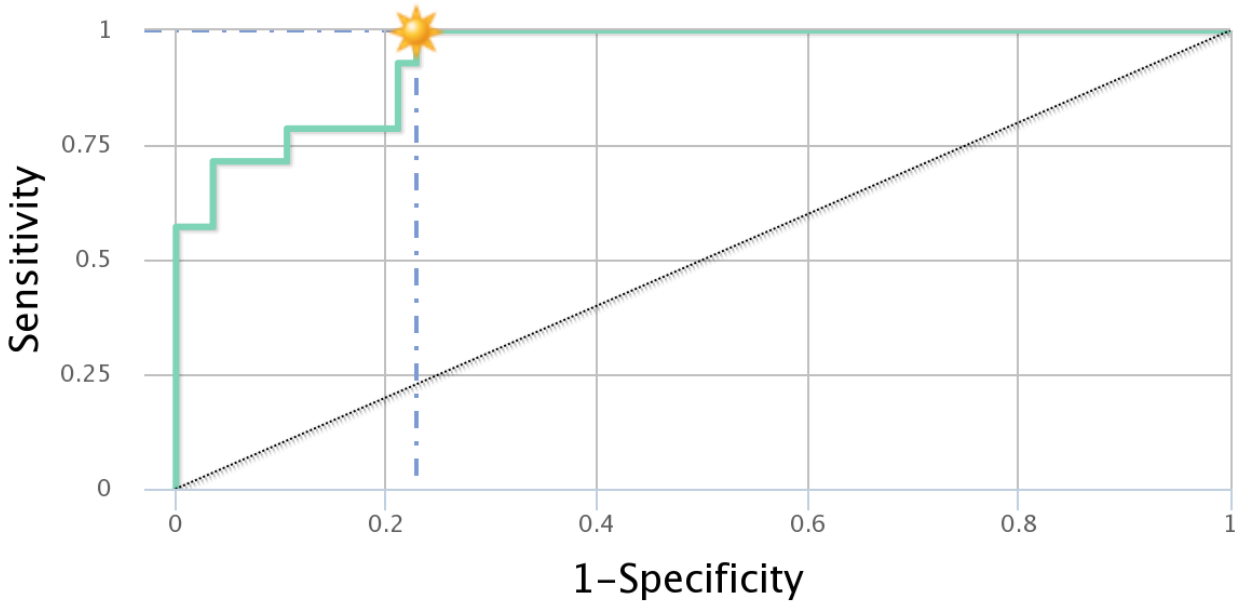


Figure 2.6. ROC curve for CombiROC-derived top 2 antigens for CRC. The ROC analysis was performed in CombiROC (http://combiroc.eu/) for the top 10 proteins; the ROC curve represents the best combination, which included CEACAM1 and GRWD1. The resulting sensitivity-, specificity- and AUC-values of 1.0, 0.77 and 0.94, respectively. The yellow star indicates the optimal cut-off point. Abbreviations include: ROC = receiver operator characteristic; CRC = colorectal cancer; AUC = area under the ROC curve.

To visualize the differential signal intensities for the top 10 proteins, a bar graph was constructed using the median RFU (± median absolute deviation (MAD)) for all patients and HCs. The signal intensity for each of the top 10 antigens was higher in CRC patients when compared to HCs: HCs produce a median signal intensity ranging of 0 - 208 RFU for the top 10 antigens; whereas CRC patients produce median signal intensities ranging from 548 – 771 RFU (Figure 2.7).

Further statistical tests were performed in R Studio (1.0.136), and the Kruskal-Wallis test analysis was performed with the PMCMR package (Maintainer and Pohlert, 2016): The Shapiro-Wilk (Shapiro and Wilk, 1965) results indicate a non-normal data distribution (p-value < 0.05), as such, the Kruskal-Wallis (Kruskal and Wallis, 1952) and Benjamini-Hochberg (Benjamini and Hochberg, 1995) tests were performed, indicating a significant increase (p-value < 0.05; adjusted p-value < 0.05) in signal intensity for CRC patients versus HCs on the CT100plus microarray (Table 2.2).

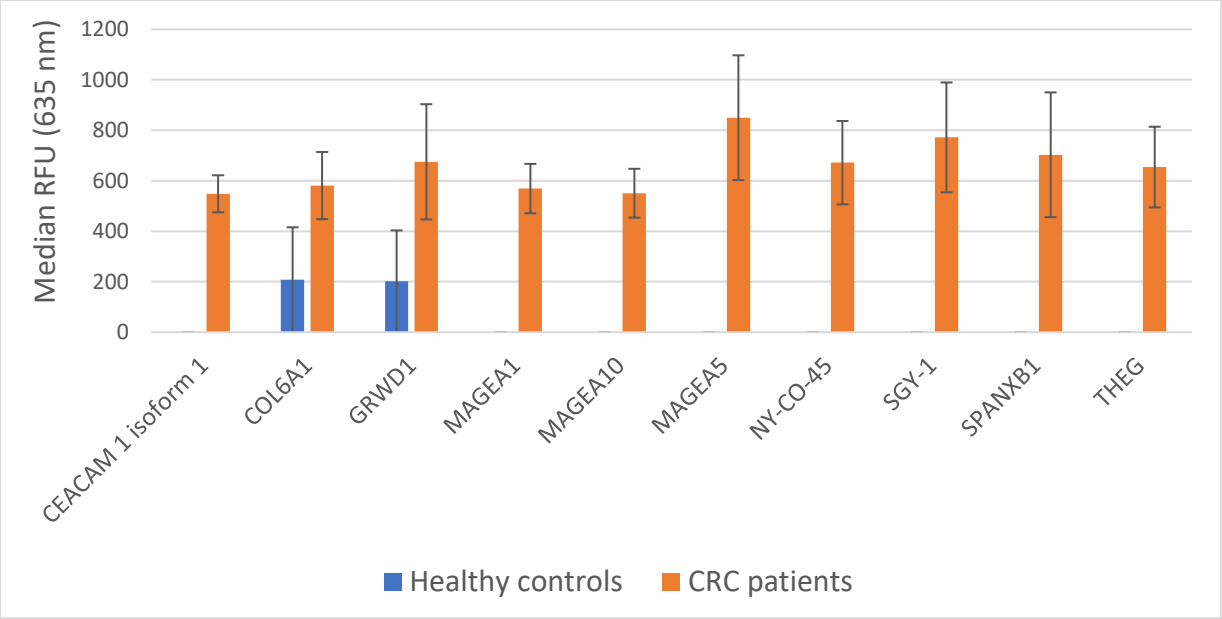


Figure 2.7. Signal intensities of top 10 antigens for CRC patients and HCs. CRC patient (*N* = 57) and HC (*N* = 14) samples were processed on the CT100plus microarray. The top 10 antigens were identified; the graph displays the median RFU-values for the top 10 antigens across all patients or HCs, and the error bars represents the MAD. Abbreviations include: CRC = colorectal cancer; HC = healthy control; *N* = number of patients; RFU = relative fluorescence units; MAD = median absolute deviation.

Table 2.2. Shapiro-Wilk, Kruskal-Wallis and Benjamini-Hochberg tests values for the top 10 antigens for CRC on the CT100plus microarray.

Antigen	Shapiro-Wilk (*p*-value)	Kruskal-Wallis (*p*-value)	Benjamini-Hochberg (adjusted *p*-value)
CEACAM 1 (isoform 1)	5.10×10^{-6}	2.67×10^{-5}	1.13×10^{-3}
COL6A1	8.05×10^{-4}	2.49×10^{-4}	3.40×10^{-3}
GRWD1	1.19×10^{-6}	2.49×10^{-4}	3.40×10^{-3}
MAGEA1	8.48×10^{-12}	3.49×10^{-4}	3.40×10^{-3}
MAGEA10	6.15×10^{-10}	1.62×10^{-4}	2.85×10^{-3}
MAGEA5	8.15×10^{-15}	8.39×10^{-6}	1.03×10^{-3}
NY-CO-45	1.36×10^{-4}	1.46×10^{-4}	2.85×10^{-3}
SGY-1	2.37×10^{-4}	8.05×10^{-5}	1.98×10^{-3}
SPANXB1	5.54×10^{-5}	6.21×10^{-5}	1.90×10^{-3}
THEG	8.30×10^{-6}	2.78×10^{-5}	1.13×10^{-3}

To further assess the biomarker potential of the top 10 proteins, a PCA and unsupervised hierarchical clustering, were performed in Perseus (version 1.5.4.1). The signal intensity values were log_2-transformed prior to the analysis.

The one-dimensional (1D) and two-dimensional (2D) PCA plots display two segregating clusters, each representing CRC patients (green) or HCs (orange), although overlap is evident between CRC patients and HCs; **CRC005** and **CRC057** cluster close with HCs, whereas **HC1**, **HC5** and **HC12** cluster with CRC patients (Figure 2.8).

Unsupervised hierarchical clustering analysis also produced segregated clusters of CRC patients and HCs. In agreement with the PCA analysis, **CRC005** and **CRC057** clustered with HCs, suggesting low autoantibody titres against the top 10 proteins for these patients; whereas **HC1** clustered with **CRC032**, **CRC046** and **CRC060**, suggesting higher autoantibody titres against the top 10 proteins for **HC1** (Figure 2.9).

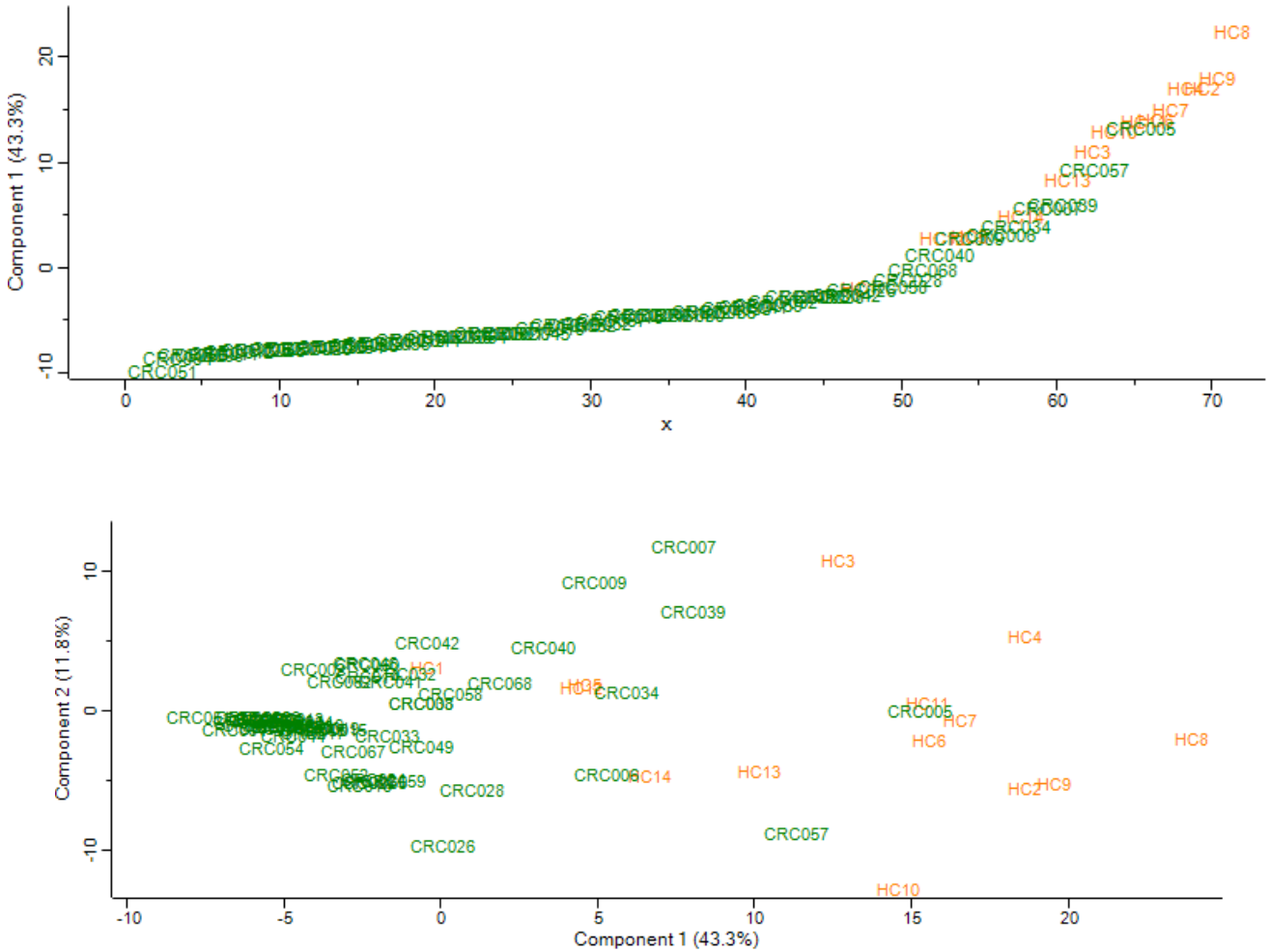


Figure 2.8. **PCA between CRC patients and HCs.** CRC patient (*N* = 57) and HC (*N* = 14) plasma were processed on the citrullinated CT100plus microarray, and the 10 with the highest AUC-values were log_2-transformed and selected for 1D and 2D PCA multivariate testing using the Perseus (version 1.5.4.1). The 1D and 2D PCA plots display CRC patient (green) and HC (orange) clusters, with overlap too. Abbreviations include: PCA = principle component analysis; CRC = colorectal cancer; HC = healthy control; AUC = area under the receiver operator characteristic (ROC) curve; 1D = one-dimensional; 2D = two-dimensional.

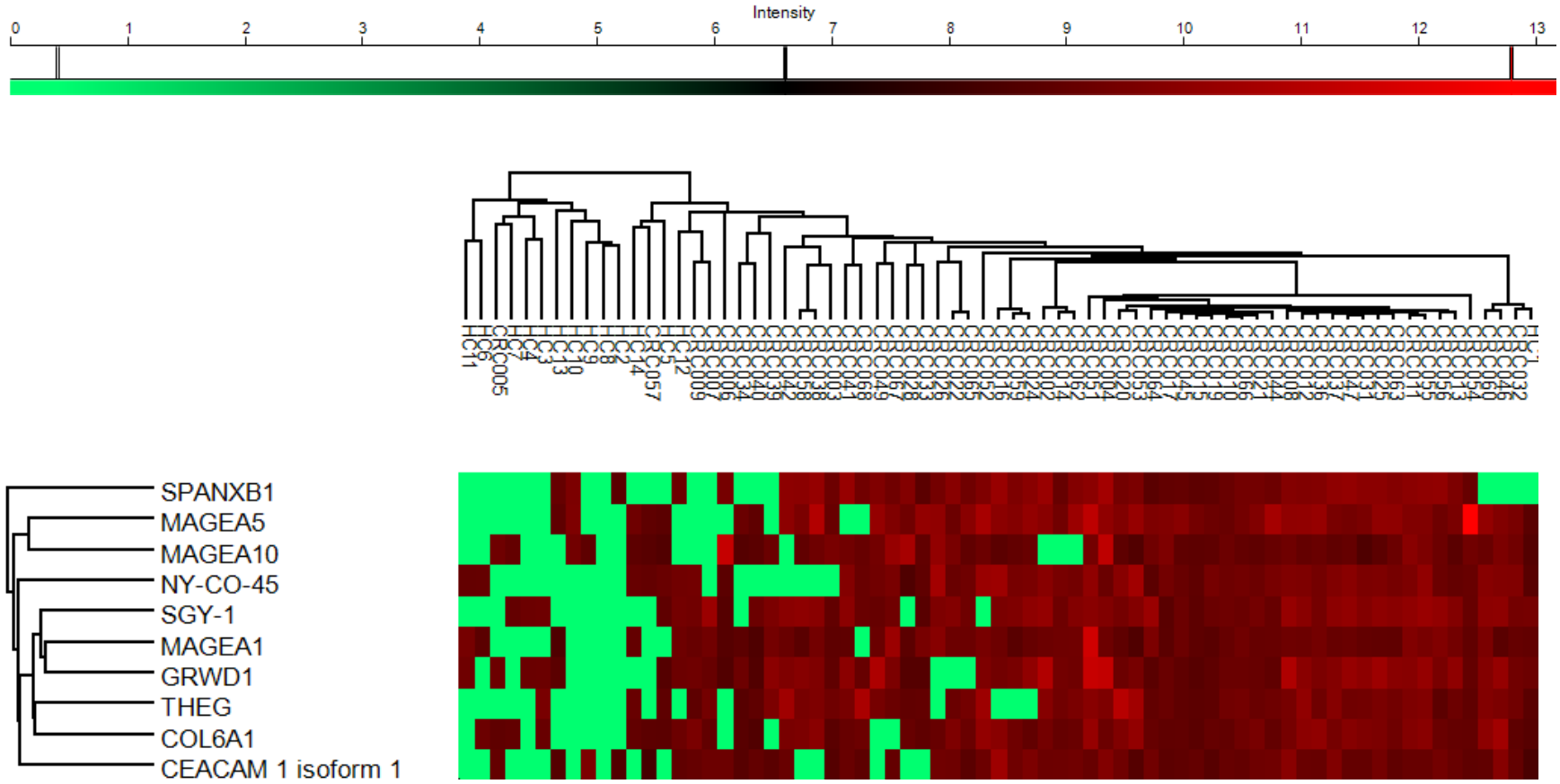


Figure 2.9. Unsupervised clustering of CRC patients and HCs. CRC patient (*N* = 57) and HC (*N* = 14) plasma were processed on the CT100plus microarray, and the 10 most significantly upregulated antigens in CRC patients were $\log_2$-transformed and selected for unsupervised hierarchical clustering using the Perseus software (version 1.5.4.1). Here, the HCs and CRC patients form two major clusters; except for **CRC057** and **CRC005**, who cluster with HCs, and **HC1** that clustered with CRC patients, **CRC0032**, **CRC046** and **CRC060**. Abbreviations include: CRC = colorectal cancer; HC = healthy control; *N* = number of patients.

2.3. CT100plus antigens associated with CRC patient clinicopathological features

We were provided with the clinicopathological information for 61 CRC patients whose plasma samples were assayed on the CT100plus microarray, summarised in Table 2.3.

Table 2.3. Clinicopathological features of CRC patients. The clinicopathological data of CRC patients tested on the CT100plus microarray are summarised. Abbreviation include: *N* = number of patients; SD = standard deviation; N/A = not available; TNM = tumour, node, metastasis; MSI = microsatellite instability; MSI-H = high microsatellite instability; MSI-L = low microsatellite instability; HNPCC = hereditary non-polyposis colorectal cancer; BMI = body mass index.

Characteristics	**Number of patients (*N* = 61)**
Mean age (SD, range)	57 (15.8, 23-84)
Gender (male/female) (percent)	30/30 (49/49%)
Ethnicity	
Black	8 (13%)
Indian	2 (3%)
Mixed	38 (62%)
White	8 (13%)
N/A	5 (8%)
TNM stage (percent)	
I	6 (9%)
II	20 (33%)
III	25 (41%)
IV	4 (7%)
N/A	6 (10%)
MSI status	
MSI-H	13 (21%)
MSI-L	3 (5%)
MSS	22 (36%)
N/A	23 (38%)
Sporadic	55 (90%)
HNPCC	6 (10%)
Family history	7 (11%)
Anatomical location (percent)	
Proximal colon	13 (21%)
Distal colon	45 (74%)
N/A	3 (5%)
Inflammation (percent)	8/61 (12%)
Radiotherapy	24/61 (39%)
Mean BMI (SD, range)	26.8 (4.7, 18.15 – 40.48)
Hypertensive	8/61 (13%)
Smoking	12/61 (20%)
Alcohol	4/61 (7%)

The clinical features include patient age, gender, ethnicity, family history, body mass index (BMI), hypersensitivity, smoking and alcohol intake. The pathology information includes cancer stage at disease presentation; MS stability or instability status; whether the disease was sporadic or classified as HNPCC; family history of CRC; anatomical presentation of the cancer specimen; whether inflammation was detected; and whether the patient received radiotherapy – these features were used for subsequent statistical analysis to identify associations between clinicopathological features and autoantibody detection.

The baseline-corrected antigen signal values were obtained prior to statistical analysis; here the baseline-corrected values were obtained by subtracting the ROC-derived cut-off-values from CT100plus-derived median-values for each antigen. The ROC-derived cut-off-values used for baseline subtraction is shown in Figure 2.10 for each antigen, resulting in a median-value of 473 (MAD = 46; Range = 325 - 661).

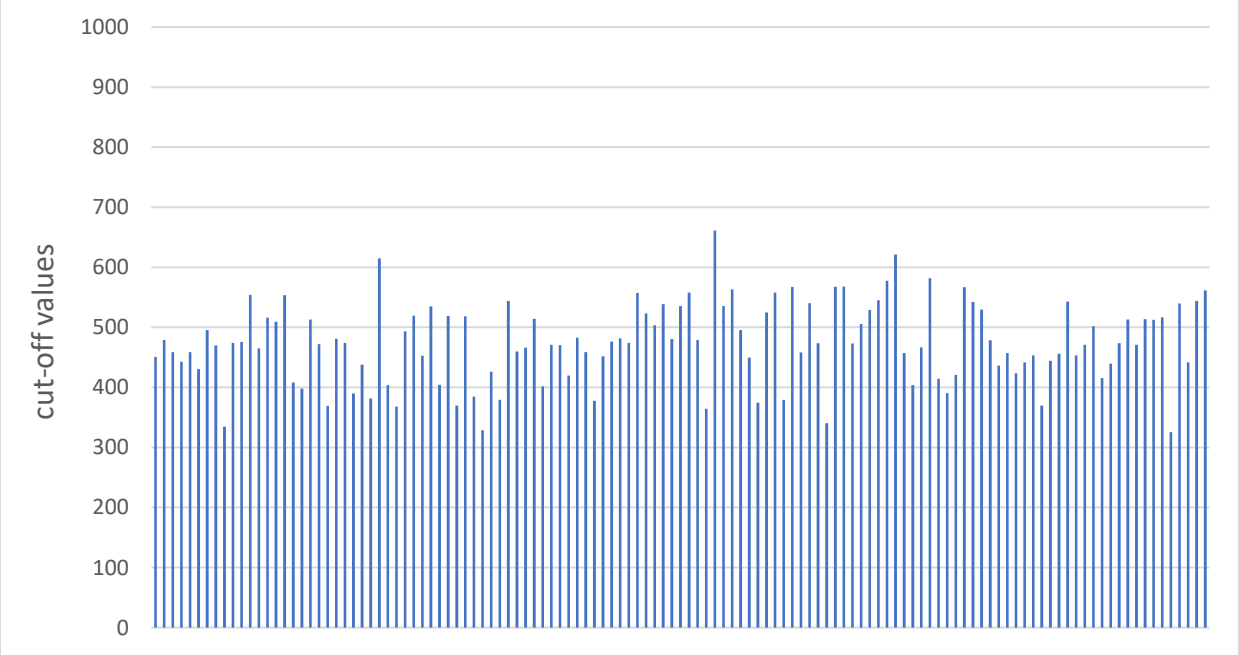


Figure 2.10. ROC-derived cut-off-values for CRC patients. ROC analyses between CRC patients and HCs produced cut-off-values for each antigen; each bar represents the cut-off-value per antigen. These values were used for baseline-correction for each antigen prior to clinicopathological statistical analysis. Abbreviations include: ROC = receiver operator characteristic curve; CRC = colorectal cancer; HCs = healthy controls.

Clinicopathological features and autoantibody signal associations were determined using the Mann-Whitney *U* test (Mann and Whitney, 1947) when comparing two groups (e.g. Recurrence), and the Kruskal-Wallis test (Kruskal and Wallis, 1952) when comparing three or more groups (e.g. Cancer stage). Statistical significance was assessed using the Benjamini-Hochberg test (Benjamini and Hochberg, 1995), where an adjusted *p*-value ≤ 0.05 was deemed significant. All tests were performed in R Studio (1.0.136), and the Kruskal-Wallis test analysis was performed with the PMCMR package in R (Maintainer and Pohlert, 2016) (Figure 2.11).

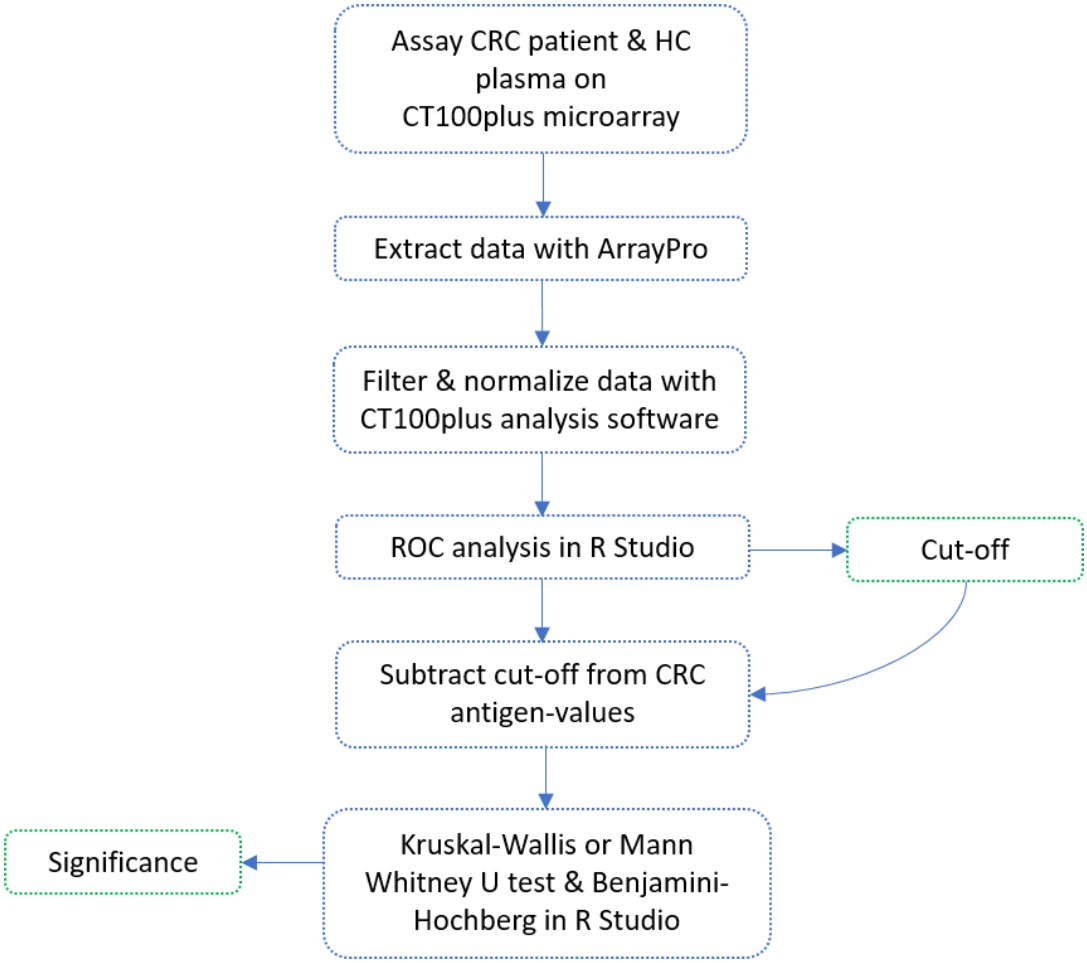


Figure 2.11. Clinicopathological data analysis pipeline. The CT100plus-derived CRC data were analysed by subtracting ROC-derived cut-off values, and subsequently performing Kruskal-Wallis or Mann-Whitney *U* tests, as well as Benjamini-Hochberg tests. Abbreviations include: CRC = colorectal cancer; ROC = receiver operator characteristic.

Although *p*-values ≤ 0.05 were obtained for associations between clinicopathological features (including gender, hypertensive, MS status, recurrence and stage) and autoantibody signal, the Benjamini-Hochberg correction test indicated that these changes were non-significant. The results for *p*-values ≤ 0.01 and adjusted *p*-values are summarised in Table 2.4.

Table 2.4. Clinicopathological features and autoantibody signal with *p*-values ≤ 0.01. Mann-Whitney *U* and Kruskal-Wallis test indicate a significant difference in antigenic signal for gender, hypertensive state, MS status, recurrence and cancer stage, although it is not statistically significant after the Benjamini-Hochberg test is applied. Abbreviation include: TAA = tumour-associated antigen; MS = microsatellite.

Clinical feature	TAA	*p*-value	Adjusted *p*-values
Gender	ACVR2B	0.007	0.80
Hypertensive	NY-ESO-1	0.009	0.58
MS status	SSX1	0.008	0.59
	MART-1	0.01	0.59
Recurrence	BAGE2	0.001	0.17
Stage	TSSK6	0.001	0.13
	MAGEA3	0.007	0.46

Discussion and conclusion

Previously, a group of 63 CRC patient plasma samples were assayed on the CT100plus microarray (Duarte *et al*., unpublished data). Only 57 (*N* = 57/63; 91%) samples passed quality control tests, and were thus used for further analysis. For the analyses, an arbitrary 1000 RFU cut-off was applied, derived from the average signal obtained for the pooled negative control sample – all antigenic signal above this threshold was considered significant. The most frequently significant antigens included SILV (*N*=19/62. 31%), p53 Q136x (*N*=13/62, 21%), CDK4 (*N*=12/62, 19%), MAGEA5 (*N*=12/62, 19%), p53 S6A (*N*=12/62, 19%), PRKCZ (*N*=12/62, 19%), 5T4/TPBG (*N*=12/62, 19%), CTAG2/LAGE-1b/LAGE1L (*N*=11/62, 18%), MAPK1 (*N*=11/62, 18%), p53 K382R (*N*=11/62, 18%), CREB1 (*N*=10/62, 16%), p53 C141Y (*N*=10/62, 16%), p53 T18A (*N*=10/62, 16%), ROPN1A (*N*=10/62, 16%). The aforementioned proteins were identified as potential CRC biomarkers relating to diagnosis or therapy.

The drawback of employing the 1000-RFU cut-off relates to the variation in antigen threshold due to variable antigen expression, evident in c-Myc control microarray in Figure 2.3. The effect is compounded when considering the variation in protein physicochemical properties between antigens. This idea is supported by results for the pooled negative control plasma, which for all CT100plus antigens produced a median signal intensity of 519 (MAD = 129; Range: 0 – 2811). Thus, the results may be more accurate and informative if the analysis process accounted for the variable antigen intensities on the CT100plus microarray. To account for threshold variation, the CT100plus microarray data from CRC patients and HCs were compared, and the top candidates were selected by employing receiver operating characteristic (ROC) analysis.

ROC curve analysis is used to identify antigens with a significantly higher expression in cancer patients versus HCs. The ROC curve displays the true positive rate (sensitivity) as a function of the false-positive rate (1 - specificity) for different cut-off points. Thus, each point on the ROC curve represents a sensitivity/specificity pair corresponding to a decision threshold. A perfect discriminatory score produces a ROC curve that passes through the upper left corner (sensitivity = 1; specificity = 1); thus the closer the ROC curve is to the upper left corner, the better the overall accuracy of the test (Zweig and Campbell, 1993). The ROC analysis produces an AUC-value, a

measure of the discriminating efficacy of antigen for patients from HCs; a sensitivity-value, indicating the fraction of patients with high cancer antigen intensity; and a specificity-value, indicating the fraction of HCs with no, or significantly lower, signal intensity for antigens (Metz, 1978; Vining and Gladish, 1992; Zweig and Campbell, 1993).

ROC analysis was employed to analyse CT100plus microarray data from 57 CRC patients and 14 HCs samples, where after further analysis into biomarker potential was performed using the 10 proteins with the highest AUC-values. It is important to note that the HC samples were not age- nor gender-matched to the CRC patients. Furthermore, although none of the HCs were previously diagnosed with cancer, tests were not performed to confirm that the HCs were CRC-free. However, since our intention here was to acquire preliminary results for a potentially larger study, we decided to use the plasma samples available to assess autoantibody profiles between CRC patient and HCs.

The top 10 proteins after ROC analysis were CEACAM 1, COL6A1, GRWD1, MAGEA1, MAGEA10, MAGEA5, NY-CO-45, SGY-1, SPANXB1 and THEG, of which only MAGEA5 was significant in a minimum of 10 patients when using the 1000 RFU cut-off. Individually, these proteins produced AUC-, sensitivity- and specificity-values ranging from 0.81 – 0.86, 0.70 – 0.88 and 0.64 – 0.86, respectively (Table 2.1). These 10 proteins were selected for further ROC analysis in CombiROC (Mazzara et al., 2017) to determine the combination of proteins that produce the highest AUC-, sensitivity-, and specificity-values. Here, the CombiROC analysis results indicate that a combination of CEACAM1 and GRDW1 produced the highest AUC-, sensitivity- and specificity-values of 0.94, 1.0 and 0.77, respectively (Figure 2.6). Literature-reported CRC biomarkers include CEA, IL-8, Apo AI, C9, and CA-19-9 producing AUC-values ranging from 0.62 - 0.90 (Murakoshi et al., 2011; Lee, 2013; Jin et al., 2014; Zhang et al., 2015). Together, these results indicate that the top 10 proteins identified with the CT100plus microarray individually produced AUC-, sensitivity- and specificity-values in the same range as the best reported CRC biomarkers, whereas combinatorial ROC results outperformed individual and literature-reported potential biomarkers. Furthermore, Kruskal-Wallis test with Benjamini-Hochberg *post hoc* analysis indicated a significant difference in antigen signals between CRC patients and HCs, producing *p*-values and adjusted *p*-values less than 0.05 (Table 2.2), indicating a significant difference in antigen signal

intensity between CRC patients and health controls. Furthermore, the PCA and hierarchical clustering results display two segregating clusters, each representing CRC patients (green) or HCs (orange), although overlap is evident on the 1D and 2D PCA plots. Together, these results indicate that the top 10 proteins may be potential CRC biomarkers, although alternate methods e.g. validating with another CRC cohort or using tumour tissues for histological confirmation, is required for biomarker validation

In Chapter 4, the development and results of an AP-MS assay is described. The AP-MS assay was developed to identify known and novel CRC biomarkers, and amongst others corroborates the results described here from the CT100plus microarray. For the AP-MS assay, CRC patient samples including **CRC002**, **CRC004**, **CRC017**, **CRC021** and **CRC031** were assayed using the AP-MS method described in Chapter 7. Of the top 10 proteins, THEG and COL1A6 were identified by AP-MS. Although CEACAM1, COL1A6, GWRD1 and MAGEA1, MAGEA5, and MAGEA10 were not identified by AP-MS, homologues of these proteins were identified, including CEACAM7, COL1A2, WRD8 and MAGED2 (Table 2.5). These outcomes imply the possibility that the CT100plus microarray detects autoantibodies toward homologues of proteins printed on the microarray, although this should be further investigated.

Since COL6A1 and THEG were identified on the CT100plus microarray as well as by AP-MS, these proteins were further investigated in literature for their biomarker potential. Additionally, since CEACAM1 and CEACAM7 share a high degree of homology (Table 2.5), CEACAM7 was also further investigated for its biomarker potential:

Table 2.5. Top 10 CRC proteins on CT100plus microarray matched to AP-MS protein list. CRC patient and HC plasma were tested on the CT100plus microarray. Data filtering and normalisation was performed using the CT100plus application, and the median intensity-values were used for ROC analysis. Antigens with the highest AUC-values were selected and compared to proteins identified for CRC patients by AP-MS (Chapter 4). Protein and homologue identification is summarised with query coverage (%), identity (%), and E-value. Abbreviation include: CRC = colorectal cancer; AP-MS = affinity purification-mass spectrometry; HC = healthy control; N/A = not applicable.

Protein	AP-MS result	Homologue	Query coverage (%)	Identity (%)	E-value
CEACAM1	No	CEACAM7	75	70	8^{-116}
COL6A1	Yes	N/A	100	100	0
GRWD1	No	RWD82	65	28	3^{-5}
MAGEA1	No	MAGED2	67	39	1^{-49}
MAGEA10	No	MAGED2	53	40	2^{-56}
MAGEA5	No	MAGED2	59	39	0.06
NY-CO-45	No	No	N/A	N/A	N/A
SGY-1	No	No	N/A	N/A	N/A
SPANXB1	No	No	N/A	N/A	N/A
THEG	Yes	N/A	100	100	0

Collagen alpha-1(VI) chain, also known as COL6A1 (http://www.uniprot.org/uniprot/P12109), is a collagen protein that has a cell-binding function. *COL6A1* gene expression is reportedly increased in CRC cell-lines after 5-aza-2'-deocycitidine treatment (Khamas et al., 2012). It is significantly deregulation in CRC tissue samples, determined by label-free mass spectrometry (Yang et al., 2012; Sethi et al., 2015) and confirmed by Western blot analysis, and thus proposed as to be potential biomarker for CRC (Yang et al., 2012). Although COL6A1 is not extensively reported for CRC, it is reportedly upregulated and correlates with a poor prognosis in lung (Chiu et al., 2011), renal (Wan et al., 2015) and prostate (Zhu et al., 2015) cancer. COL6A1 was previously investigated amongst a panel of proteins by ELISA to detect autoantibodies in breast cancer patients versus HCs, although it was not found to be significantly dysregulated (Piura and Piura, 2011). According to our knowledge, we show for the first time an autoantibody response toward COL6A1 in CRC patients using two orthogonal immunoproteomic assays, which could serve as a potential biomarker for diagnosis or disease therapy.

Testicular haploid expressed gene, also known as THEG (http://www.uniprot.org/uniprot/Q9P2T0), is expressed only in spermatid cells and cancer, although its function is unknown. It is a relatively newly proposed CTA, and thus also not extensively reported. However, in a study of 22 cell-lines, high levels of THEG mRNA was detected in the HCT15 colon cancer cell-line (Chen et al., 2005). According to our knowledge, an autoimmune response toward THEG has not previously been reported, thus we show for the first time an autoantibody response toward THEG using two orthogonal immunoproteomics assay for CRC patients, which could also serve as a potential biomarker for diagnosis or disease therapy.

Carcinoembryonic antigen (CEA), also known as CEACAM (http://www.uniprot.org/uniprot/Q13984), is a widely-used tumour biomarker. Its main application is for gastrointestinal cancers, especially in CRC. In 1965, CEA was identified in foetal colon and colon adenocarcinoma, but absent from healthy adult colon (Gold and Freedman, 1965a, 1965b). In 1969, increased serum concentrations of CEA were reported for 35 of 36 patients with CRC, whereas low levels were detected in healthy volunteers, pregnant women, patients with non-gastrointestinal cancer, or benign gastrointestinal disease (Thomson et al., 1969). Further investigations revealed that CEA levels increase with increasing disease stage (Wanebo et al., 1978), although serum levels decrease with liver metastasis likely due to CEA uptake by Kupffer cells, which modify CEA by removing its sialic acid residues and the asialo CEA is endocytosed and degraded by liver parenchymal cells (Thomson et al., 1969; Begent, 1984). CEA concentrations are also higher in well-differentiated CRC neoplasms (Rieger and Wahren, 1975; Goslin et al., 1981; Bhatnagar et al., 1999), malignancies on the right side of the colon (Wanebo et al., 1978; Slater et al., 1979), male smokers (Wilson et al., 1999), and aneuploid CRC versus diploid CRC (Rognum, 1986). CEACAM7 is a family member of CEA that is expressed in normal colorectal epithelia, although it is downregulated in adenomas and carcinomas (Thompson et al. 1997; Schö et al. 2000; Birkenkamp-Demtroder et al. 2005). CEACAM7 decreased expression in rectal cancers is linked to disease recurrence (Messick et al., 2010), whereas elevated serum levels were detected for CRC patients with liver metastasis (Kleivi et al., 2007). According to our knowledge, an autoantibody response has not been reported against CEACAM1 or CEACAM7 in CRC patients. Here, using the CT100plus microarray, we again show for

the first time using two orthogonal immunoproteomics methodologies an autoantibody response in CRC patients.

We did not find a significant (adjusted *p*-value < 0.05) association between patient clinicopathological information and autoantibody signal intensity. Although *p*-values ≤ 0.05 were obtained for associations between clinicopathological features (including gender, hypertensive, MS status, recurrence and stage) and autoantibody signal, the Benjamini-Hochberg correction test indicated that these changes were non-significant. Unlike the melanoma cohort (see Chapter 5), these patient samples were not collected as part of a clinical cohort. Thus, when considering the diversity of the cohort (Table 2.3), and only assessing 123 TAA, this result is plausible. It is therefore important that future CT100plus assays include carefully selected patient samples that form part of a controlled clinical cohort, which may result in statistically significant associations between clinicopathological features and autoantibody signal intensity.

In conclusion, when comparing CRC patients and HCs, we identified top antigens supported using several statistical tests that are either used clinical biomarkers, e.g. CEACAM1, or have been associated with CRC cancer, e.g. COL6A1 and THEG. Although CEACAM1 is a known clinical biomarker for CRC, autoantibody responses have not been reported against COL1A6 or THEG and CEACAM in CRC. It is important to note though that alternate methods e.g. validating with another CRC cohort or using tumour tissues for histological confirmation, are required for downstream biomarker validation.

Chapter 3

Detecting CRC patient autoantibody responses to citrullinated CT100plus microarray proteins

Abstract

Cancer patients elicit an autoantibody response toward TAAs which can result from altered tissue specific expression patterns, mutations, PTMs, *et alia*. Protein citrullination is a post-translational modification (PTM) resulting from the PAD-induced conversion of arginine to citrulline. Citrullinated proteins are known to induce an autoantibody response in patients with rheumatoid arthritis, Alzheimer's disease and cancer. The expression of isozymes PAD2 and PAD4 are dysregulated in many cancer-types, including CRC, resulting in alternate PAD substrates. In this study, we assessed the autoantibody response of 59 CRC patients and 11 HCs toward citrullinated and non-citrullinated antigens on the CT100plus microarray. Only 1 microgram per millilitre (µg/ml) PAD4 was required to add the citrulline modification on CT100plus microarray antigens. CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1 produced a medium (2000 < RFU ≤ 10 000) or high signal (relative fluorescent unit (RFU) > 10 000) on the citrullinated microarray, but low signal (RFU ≤ 2000) on the non-citrullinated microarray. Significantly (p-value ≤ 0.05; adjusted p-value ≤ 0.05) higher signal intensities were detected in CRC patients *versus* HCs for citrullinated CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1, whereas no significant (p-value > 0.05) difference in autoantibody signal was detected on the non-citrullinated microarray for the same patient cohort. Furthermore, ROC analyses of these antigens resulted in an AUC-, sensitivity- and specificity-ranges of 0.91, 0.87 and 0.89, respectively. PCA results indicate distinct groups of CRC patients and HC clusters. Together, here we show for the first time that CRC patients elicit an autoantibody response to citrullinated proteins, resulting in potential novel CRC biomarkers.

Introduction

Both the humoral and T-cell arm of the immune system are reportedly activated in response to a TAAs and TSAs. In this Chapter, we investigated the autoantibody response to citrullinated CT100plus proteins:

Protein citrullination is a calcium-dependent hydrolytic reaction that converts arginine to the citrulline amino acid (Figure 3.1) (Rogers and Simmonds, 1958). The reaction is catalysed by the Protein Arginine Deimidase (PAD) family of proteins (Knuckley et al., 2010; Darrah et al., 2012). Under physiological conditions, PADs are normally inactive until stimulated by calcium, resulting in the citrullination of structural proteins (Inagaki et al., 1989; Senshu et al., 1995), and proteins involved in gene transcription (Wang, 2004; Saiki et al., 2009; Christophorou et al., 2014).

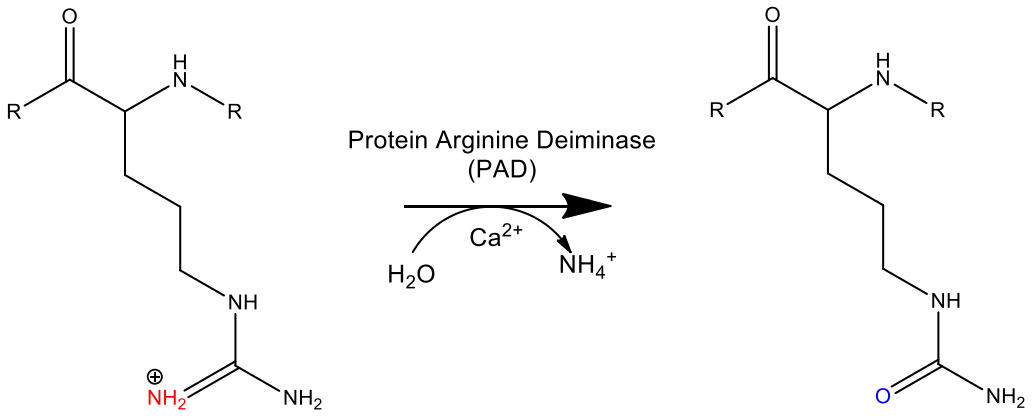


Figure 3.1. PAD-catalysed conversion of peptidyl arginine to peptidyl citrulline. PAD isozymes catalyse the calcium-dependent hydrolytic conversion of peptidyl arginine to peptidyl citrulline, with the expulsion of ammonia. Here, the charged amine group (indicated in red) is converted to neutral oxygen (indicated in blue).

Citrullinated proteins induce an autoantibody response in patients with rheumatoid arthritis (Schellekens et al., 1998), Alzheimer's disease (Ishigami et al., 2005) and cancer (Chang and Fang, 2010). Increased PAD4 expression and activity in cancer-cells have previously been reported (Denman, 2005; Chang et al., 2009; Doyle et al., 2010) for colorectal, lung, ovarian and uterine adenocarcinomas, as well as hepatocellular, oesophageal, renal, uterine and bladder carcinomas

(Wang et al., 2010). More recently, growing evidence implicates the role of PAD2 in cancer progression, where changes in PAD2 expression correlates cancer progression, depending on the cancer-type. For example, *in vitro* studies showed that PAD2 overexpressed in human squamous cell carcinoma (SCC) cell-line resulted in elevated markers for inflammation and epithelial-to-mesenchymal transition (EMT). *In vivo* studies supported this finding as transgenic mice with overexpressed PAD2 developed neoplastic skin lesions, demonstrating increased expression of markers of invasion, inflammation and EMT (McElwee et al., 2014). In contrast, studies of breast cancer demonstrated a correlation between PAD2 depletion or inhibition with cancer migration (Horibata et al., 2017). Also, *in vitro* studies concluded that while the *PADI2* (isoform 1) transcript is expressed in normal colon tissue, the expression is significantly lower in colon cancer cells. Moreover, overexpression of *PADI2* in colon cancer cells suppressed cell proliferation, suggesting that *PADI2* may have a tumour suppressor role in colon cancer. Despite decreased PAD2 expression in colon cancer, citrullinated proteins have been detected by immunoblot analysis (Funayama et al., 2017).

PAD2 substrates in normal tissues include myelin basic protein (Lamensa and Moscarello, 1993), vimentin (Hsu et al., 2014), actin and histones (Zhang et al., 2012). Citrullinated proteins have been detected in various cancer cell-lines including colon cancer cells. Cancer-associated citrullinated proteins include ENO1, HSP60, KRT8, TUBB, which are associated with tumour growth, invasion, and a poor prognosis (Jiang et al., 2013). Altogether, these results provide evidence that PAD2 may have alternate substrates in cancer, seemingly associated with cancer progression.

In this study, we aim to identify CRC patient autoantibodies against citrullinated CT100plus antigens. Here, we compared autoantibody responses between citrullinated and non-citrullinated CT100plus antigens to identify autoantibodies against citrullinated antigens. Furthermore, we compare these autoantibody responses between CRC patients and HCs to determine antigen discriminatory efficiency. Lastly, we identified citrullinated antigens also detected in our AP-MS study (Chapter 4), which we compare to results obtained on the CT100plus microarray.

Results

3.1. Citrullinating antigens on the CT100plus microarray

To determine whether CRC patients produce an autoantibody response to citrullinated CT100plus antigens, CRC patient plasma was incubated on a citrullinated version of the microarray. For this, the microarray was incubated with PAD4 enzyme, to convert a proportion of surface-exposed arginine residues to citrulline, and subsequently incubated with CRC patient plasma, followed by fluorescently-labelled anti-human IgG to detect patient autoantibodies. The optimal PAD4 concentration was determined before assessing the CRC patient autoantibody response to citrullinated antigens. For this assay, the CT100plus microarray was treated with 0, 0.25, 0.5, 1.0, 2.0, 4.0 and 8.0 µg/ml PAD4 for 2 hours, whereafter citrulline groups were detected with fluorescently-labelled rabbit anti-citrulline IgG. Raw GenePix files were used to obtain the median of all median RFU-values for the foreground (antigen spot) and background (streptavidin) signals, and a line graph constructed to assess the dose-dependent effect of PAD4-induced citrullination on the CT100plus microarray (Figure 3.2).

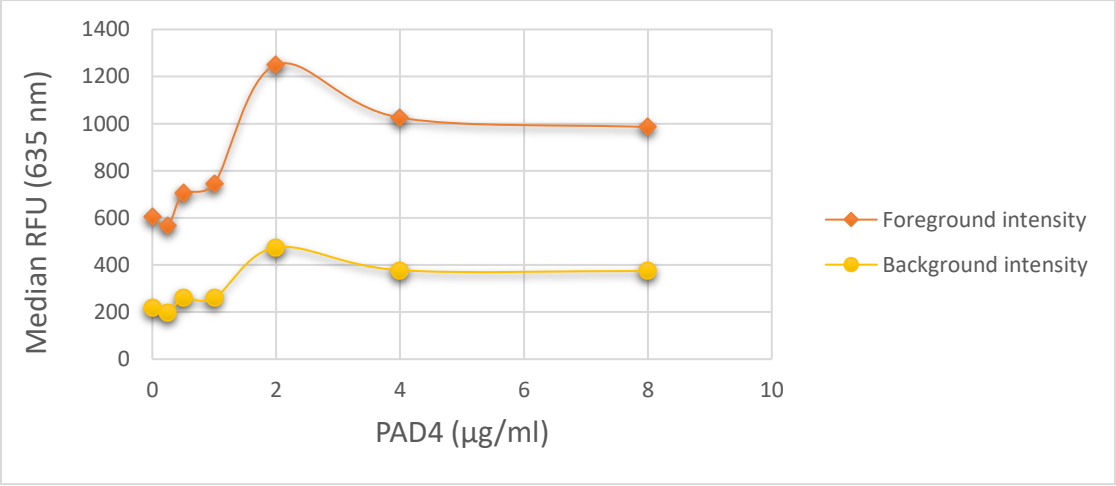


Figure 3.2. Detecting citrullinated antigens on the CT100plus microarray. The CT100plus microarray was incubated with 0.25, 0.5, 1.0, 2.0, 4.0 and 8.0 µg/ml PAD4 enzyme to determine the optimal concentration for protein citrullination. Citrullinated antigens were detected using anti-citrulline antibody, and subsequent detected with a fluorescent-labelled secondary Ab (635 nm). The median of the median net intensity of all antigens was obtained. Here, we detect a higher foreground signal versus background signal. Abbreviations include: PAD4 = protein arginine deiminase isoform 4; Ab = antibody.

A higher foreground (antigen spots) intensity *versus* background (streptavidin) intensity was detected, suggesting a higher frequency of citrulline groups present on CT100plus antigen spots *versus* streptavidin. Together, the results indicate that PAD4 citrullinated antigens on the CT100plus microarray. It is important to determine the optimal PAD4 concentration required for significant increase in citrullination on the CT100plus microarray. For this, the CT100plus software-derived median net intensities (median foreground minus median background) were obtained for all antigens across the microarray, and a Mann-Whitney *U* test (Mann and Whitney, 1947) was performed to assess statistical significance, and a line graph was constructed to determine the linear range of detection. Figure 3.3 displays the entire range of PAD4 concentrations assayed, whereas Figure 3.4 displays the signal intensity within the linear range of detection (0.25 – 2 µg/ml PAD4). A significant (*p*-value = 0.005; Mann-Whitney *U* test) increase in citrulline detection was observed for 1 µg/ml PAD4 *versus* the untreated microarray, which falls within the linear range detection, and was thus used for subsequent CT100plus citrullination assays.

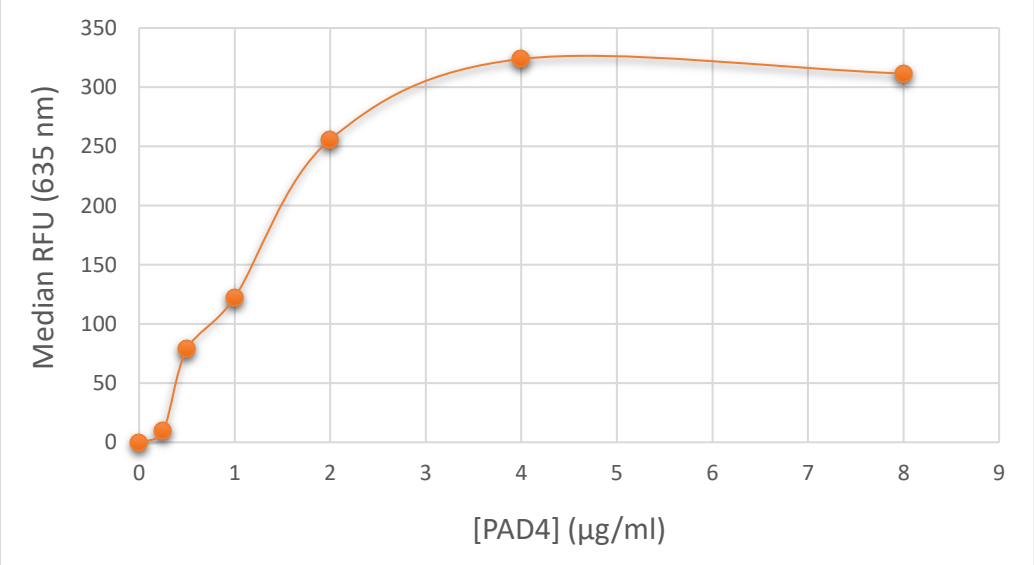


Figure 3.3. Detecting Citrullination after CT100plus software processing. To determine the optimal PAD4 concentration to citrullinate CT100plus antigens, the microarray was treated with 0.25, 0.5, 1.0, 2.0, 4.0 and 8.0 µg/ml PAD4 enzyme. Citrullinated antigens were detected using anti-citrulline antibody, and subsequently detection with a fluorescent-labelled Ab at 635 nm. The CT100plus software was used to filter and normalise raw data. The processed data shows a dose-dependent increase in signal. Abbreviations include: PAD = protein arginine deiminase; Ab = antibody.

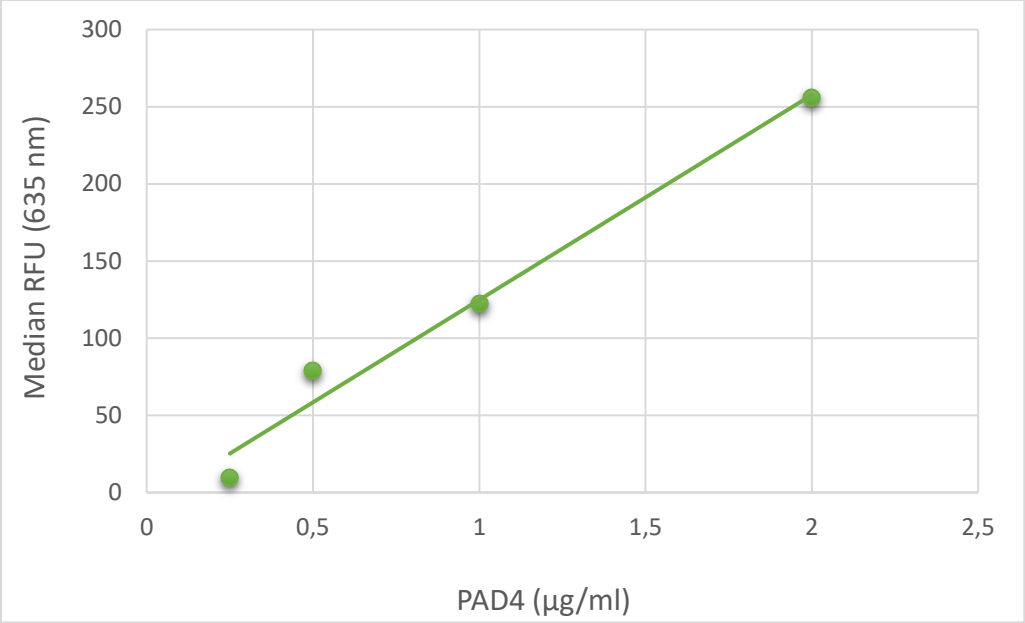


Figure 3.4. Linear range when detecting PAD4-induced citrullination. The CT100plus microarray was treated with 0.25, 0.5, 1, 2, 4 and 8 µg/ml PAD4. The Mann-Whitney *U* test result showed a significant (*p*-value = 0.005) increase in citrullination when using 1 µg/ml PAD4, which falls within the linear range of detection. Thus, 1 µg/ml PAD4 was subsequently used to citrullinated CT100plus microarrays before analysing CRC patient plasma. Abbreviations include: PAD = protein arginine deiminase.

3.2. CT100plus microarray quality controls

A significant increase in citrullination was detected when treating with 1 µg/ml PAD4, as such, it was used for all subsequent assays. Two control assays were performed, including a c-Myc (Figure 3.5) and pooled negative control plasma (Figure 3.6). Each of the 123 recombinant proteins on the CT100plus microarray (Figure 7.4) contains a c-Myc tag, used to confirm the successfully immobilisation of recombinant proteins on the slide surface with an anti-c-Myc antibody. It is evident from Figure 3.5 that antigens do not have the same signal intensity across the microarray. This is because each of the 123 recombinant proteins are expressed at different levels within insect cells. Although the signal for each antigen is not visibly discernible on the image, protein expression was confirmed by Western blot analysis (Beeton-Kempen et al., 2014). The antigens are in a 3D conformation on the microarray and in some cases the c-Myc tag may be occluded within the protein structure, preventing interaction between the c-Myc tag and the anti-c-Myc antibody; thus, signal would not be detected for these antigens on the CT100plus microarray.

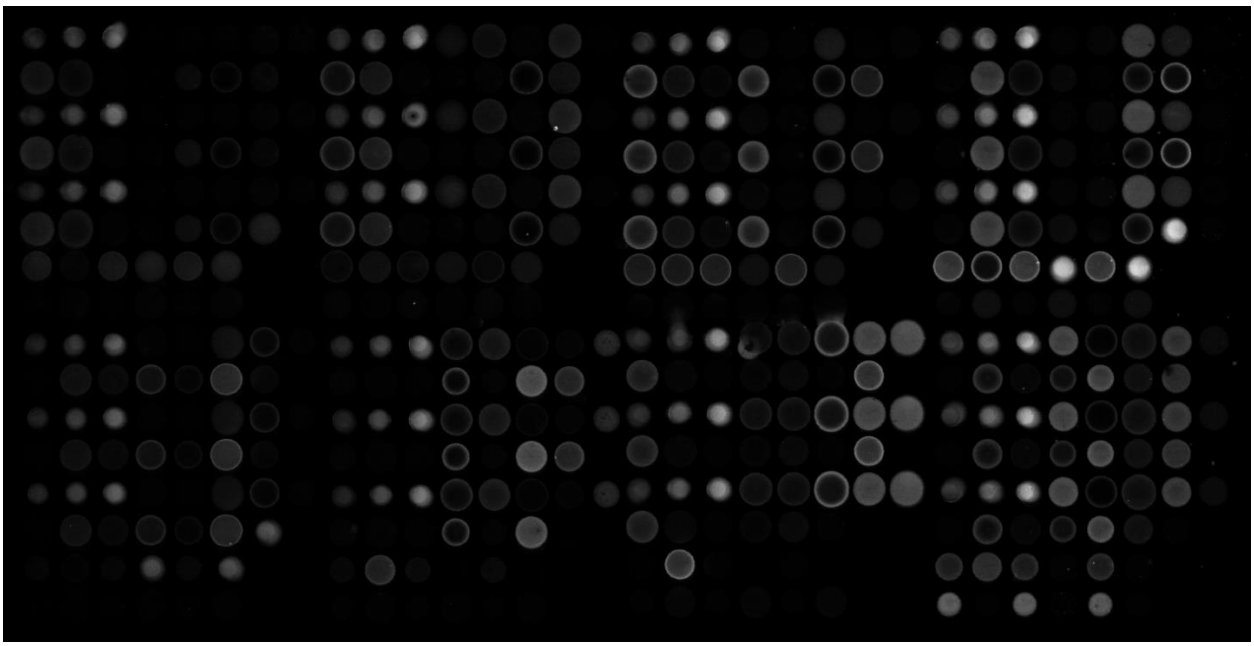


Figure 3.5. CT100plus microarray c-Myc control. Each of the 123 recombinant proteins on the CT100plus microarray contains a c-Myc tag. The microarray was treated with Cy3-labelled anti-c-Myc antibody, which was used to confirm the successfully immobilisation of recombinant proteins on the slide surface with an anti-c-Myc antibody, as depicted in the microarray image.

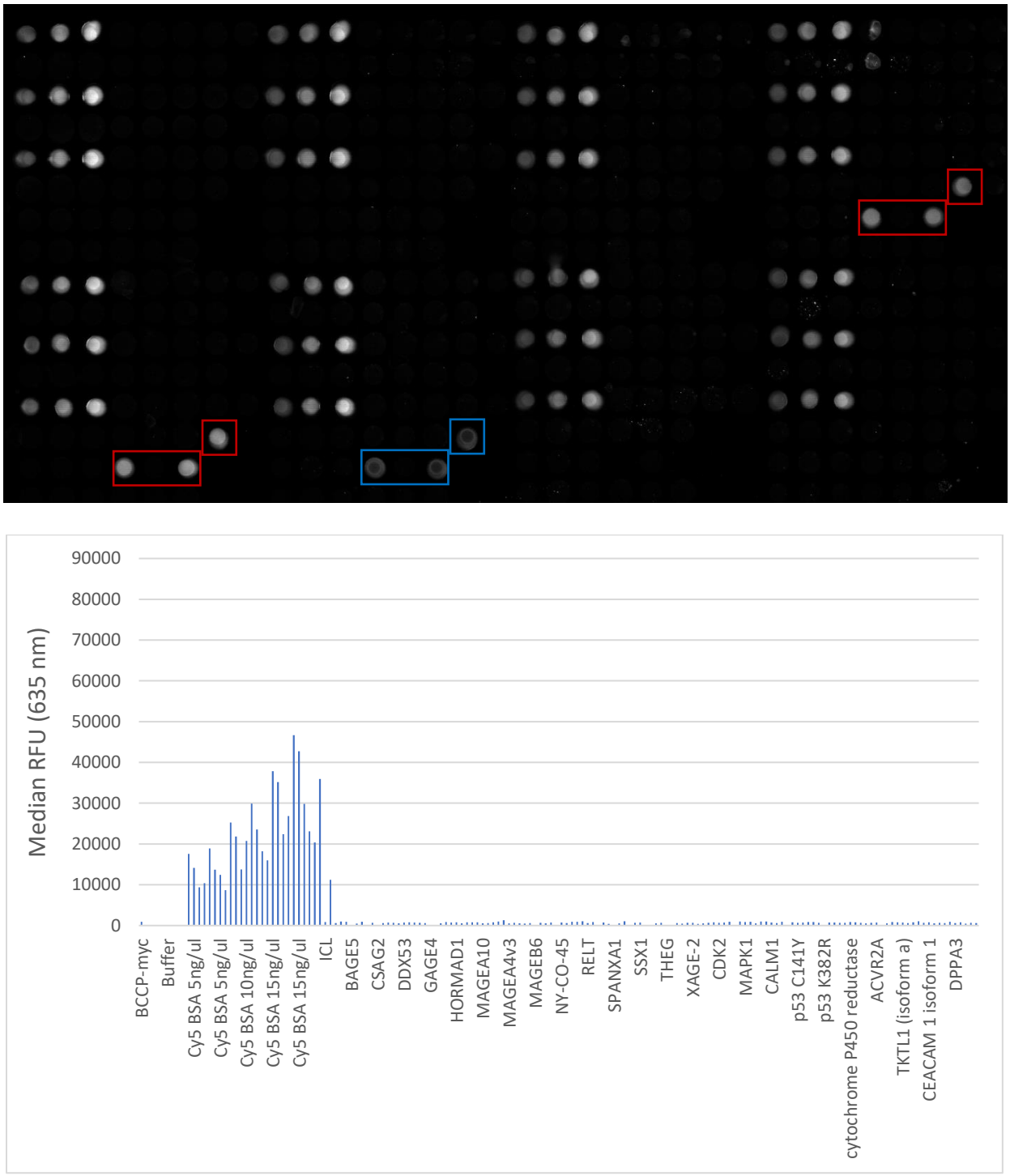


Figure 3.6. Citrullinated CT100plus microarray negative control pooled plasma sample. As a negative control, the CT100plus microarray was incubated with pooled plasma of HCs. Low or no signal is detected for the buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. High signal was detected for the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots, anti-human IgG was detected (boxed in red), and human IgG was detected (boxed in blue). Low or no signal is detected for recombinant proteins. Abbreviations include: HCs = healthy controls; ICL = insect cell lysate; BCCP = biotin carboxyl carrier protein.

The negative control assay displays low or no signal from buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. Furthermore, the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots were detected, and used for subsequent downstream data normalisation steps; anti-human IgG was detected, confirming the addition of patient plasma; and human IgG was detected, confirming the addition of detection antibody.

Recombinant proteins were detected on the CT100plus microarray by the anti-C-Myc Ab (Figure 3.5), with low or no signal is detected for the pooled negative control (Figure 3.6), indicating a high-quality platform that can be used for the specific detection of autoantibodies against CT100plus cancer antigens.

3.3. Detecting CRC patient autoantibodies toward citrullinated cancer antigens

A significant increase in citrulline groups was detected when incubating the CT100plus microarray with 1 µg/ml PAD4. As such, this concentration was used for all subsequent assays to detect autoantibodies against citrullinated antigens in CRC patients (Groote Schuur Hospital, HREC REF Number: 269/2011). Briefly, CT100plus microarrays were pre-treated with 1 µg/ml PAD4, and subsequently incubated with CRC patient plasma (1:800) and then with fluorescently-labelled detection antibody. The microarray images were obtained at 635 nanometres (nm) using the automatic gain control (AGC) setting.

To identify the citrullinated antigens that induce autoantibody response in CRC patients, the antigenic signals for the citrullinated and non-citrullinated CT100plus microarrays were compared. All data files and images for the non-citrullinated microarray were obtained previously (Duarte *et al*., unpublished data). Here, the CT100plus microarrays were also incubated with CRC patient plasma (1:800) and fluorescently-labelled detection Ab only, and microarray images were obtained at 635 nm using the AGC setting. The raw data files from the citrullinated and non-citrullinated CT100plus microarrays were extrapolated using the ArrayPro software, and data filtering and normalisation was performed using the CT100plus software with the following criteria: whole array CV ≤ 25%; probe replicate CV ≤ 20%; noise-threshold ≥ 2 SD above background; maximum threshold = 65355 nm; whole array filtering control = 15 ng/µl Cy5-biotin-BSA. Here, 43 patient data files passed the filtering and normalisation quality control procedure.

The median signal intensity for each antigen was obtained and used for subsequent statistical analyses.

Interestingly, 26/43 (62%) CRC patients produced moderately higher baseline values on the citrullinated microarray compared the non-citrullinated microarray, increasing from a median antigen RFU of 637 (±120 MAD) to 1100 (±402 MAD), respectively. Thus, a direct comparison, e.g. the Mann-Whitney *U* test, between antigens on the citrullinated and non-citrullinated microarrays may lead to false-positive or false-negative results. Comparisons between citrullinated and non-citrullinated microarrays were based on signal intensities ranges, which were classified as those that were low (0 - 2000 RFU), medium (2001 - 10 000 RFU), or high (> 10 000 RFU). Here, 11/43 (26%) CRC patients displayed low RFU signals on both array-types; an example is shown using **CRC042** in Figure 3.7. A total of 21/43 (49%) patients produced medium or high signals for antigens on the non-citrullinated microarray, but only 6/43 (14%) retained the signal toward the same antigens on the citrullinated microarray; an example is shown for **CRC010** in Figure 3.8. Here, signal is detected for DDX53 on both microarray-types, although a high signal is observed on the non-citrullinated microarray whilst a medium signal is observed on the citrullinated microarray; this may be due to the conversion of arginine to citrulline on antigen surfaces, decreasing autoantibody-epitope affinity. A total of 28/43 (65%) patients elicited a medium or high signal intensity to at least one of the citrullinated antigens, but low signal on the non-citrullinated microarray. Here, the antigens that produced medium and/or high signal on the citrullinated microarray, but low signal on the non-citrullinated microarray, included CDK7 and MAGEB6, each detected for 14 (number of patients (*N*) = 14/43, 33%) patients; MAGEB1 in 12 (*N* = 12/43, 28%) patients; MAGEB5 and SYCP1, each detected for 11 (*N* = 11/43, 26%) patients; MAGEA4v2 and SPANXA1, each detected in 6 (*N* = 6/43, 14%) patients; SOX1 in 5 (*N* = 5/43, 12%) patients; MAGEA3 and CDK2, each detected in 4 (*N* = 4/43, 9%) patients; NLRP, p53 S46A and p53 S6A, each detected in 2 (*N* = 2/43, 5%) patients; and CDCC33, CTNNB1, GAGE5, MAP9, p53, PBK, SILV, and TKTL1, each detected in 1 (*N* = 1/43, 2%) patient. An example is shown for **CRC045** in Figure 3.9, which showed increased signal intensity for MAGEB1, MAGEB5 and MABEG6, which was detected at lower levels on the non-citrullinated microarray. Together, these results provide

evidence that CRC patients produce autoantibody signals toward citrullinated antigens on the CT100plus microarray.

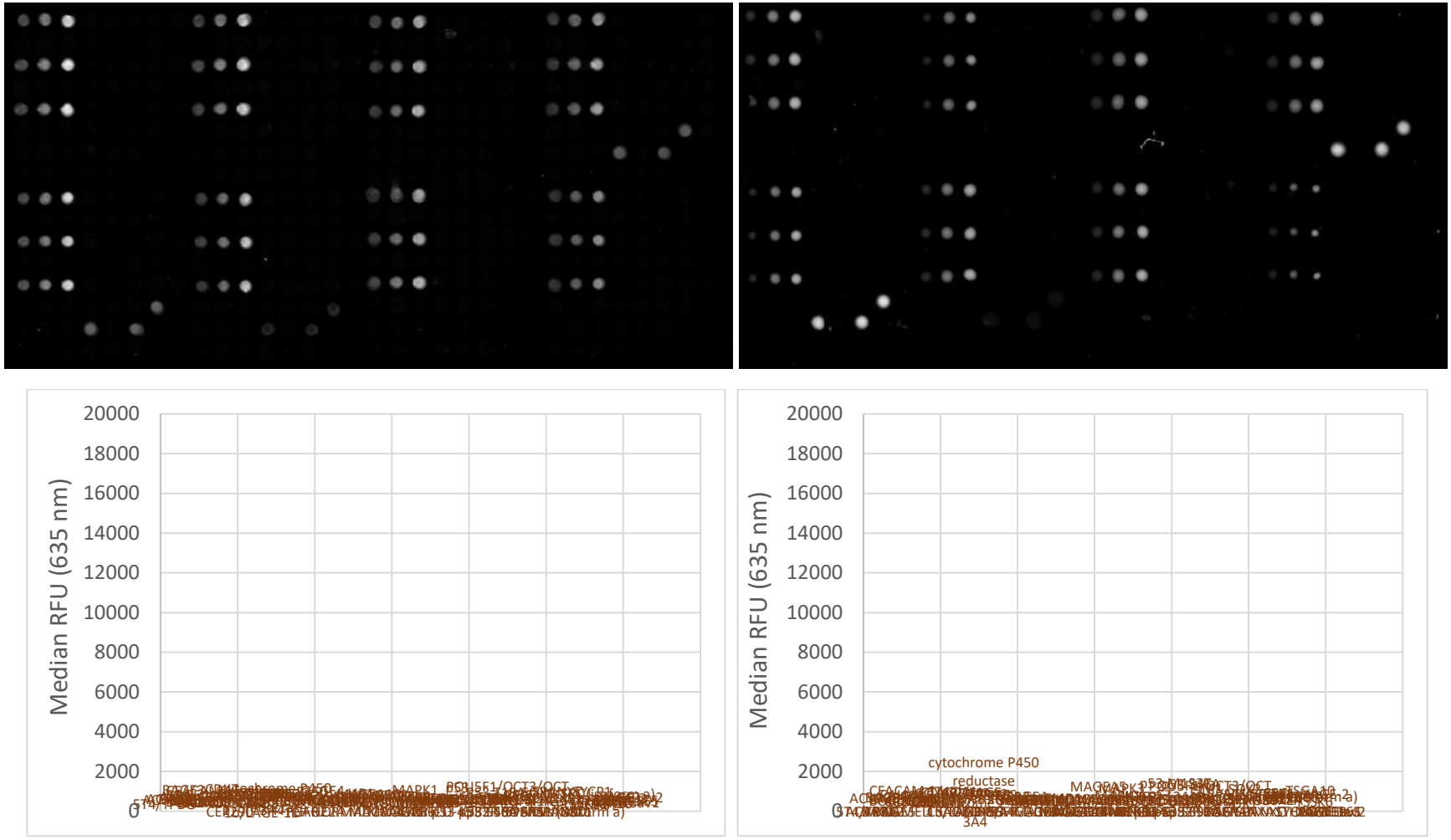


Figure 3.7. CT100plus antigenic signal for CRC042 on citrullinated and non-citrullinated microarrays. Microarray images were scanned Images, and the data extrapolated, filtered and normalised. The CT100plus microarrays treated with (left image and graph) and without (right image and graph) 1 µg/ml PAD4, then subsequently plasma from patient **CRC042**, and detection antibody. Abbreviations include: RFU = relative fluorescence units; PAD4 = protein arginine deiminase isoform 4.

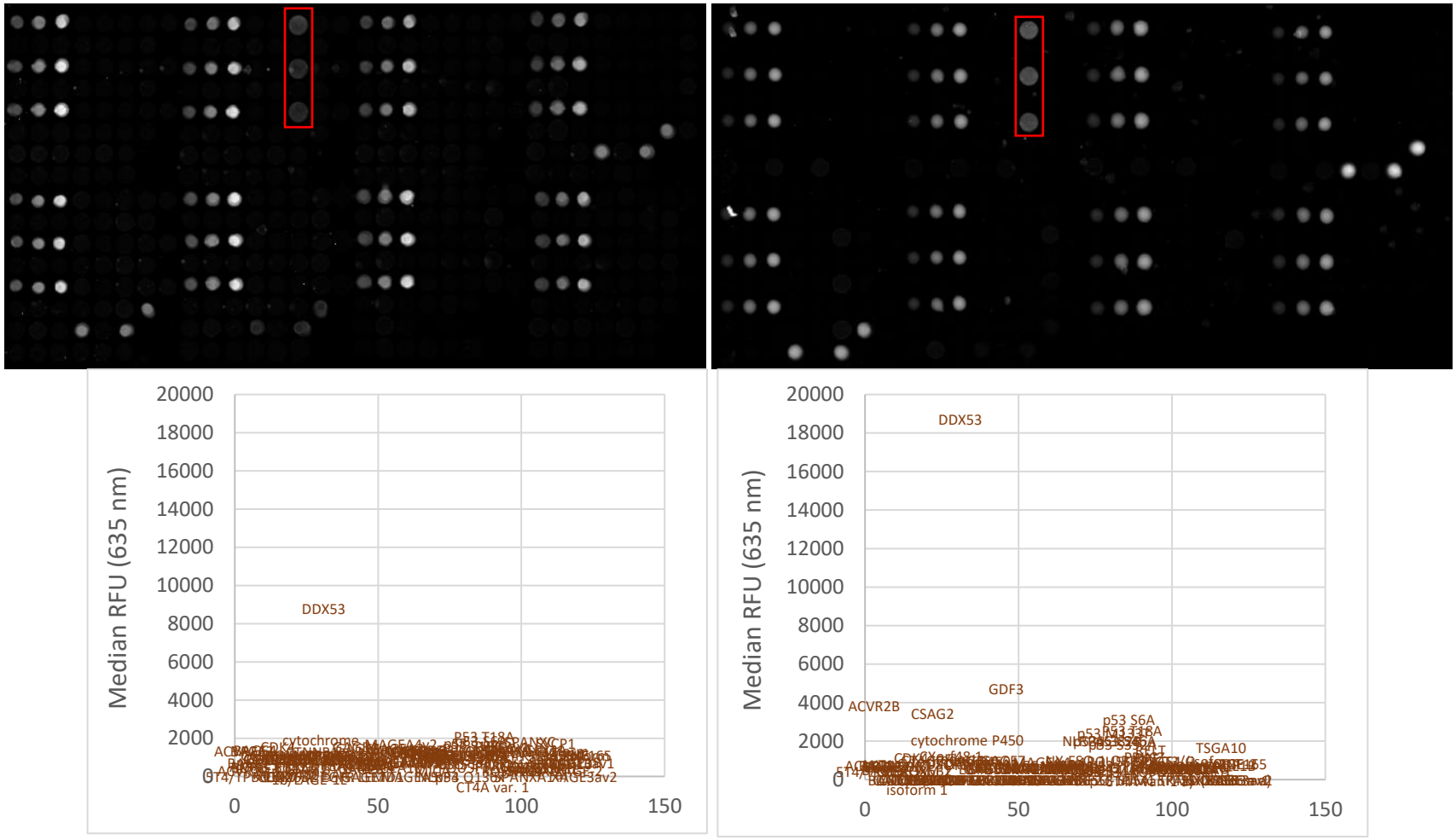


Figure 3.8 CT100plus antigenic signal for CRC010 on citrullinated and non-citrullinated microarrays. Microarray images were scanned Images, and the data extrapolated, filtered and normalised. The CT100plus microarrays treated with (left image and graph) and without (right image and graph) 1 µg/ml PAD4, then subsequently plasma from patient **CRC010**, and detection antibody. Higher signal intensities were detected for DDX53 (boxed in red) on the citrullinated microarrays *versus* non-citrullinated microarray. Abbreviations include: RFU = relative fluorescence units; PAD4 = protein arginine deiminase isoform 4.

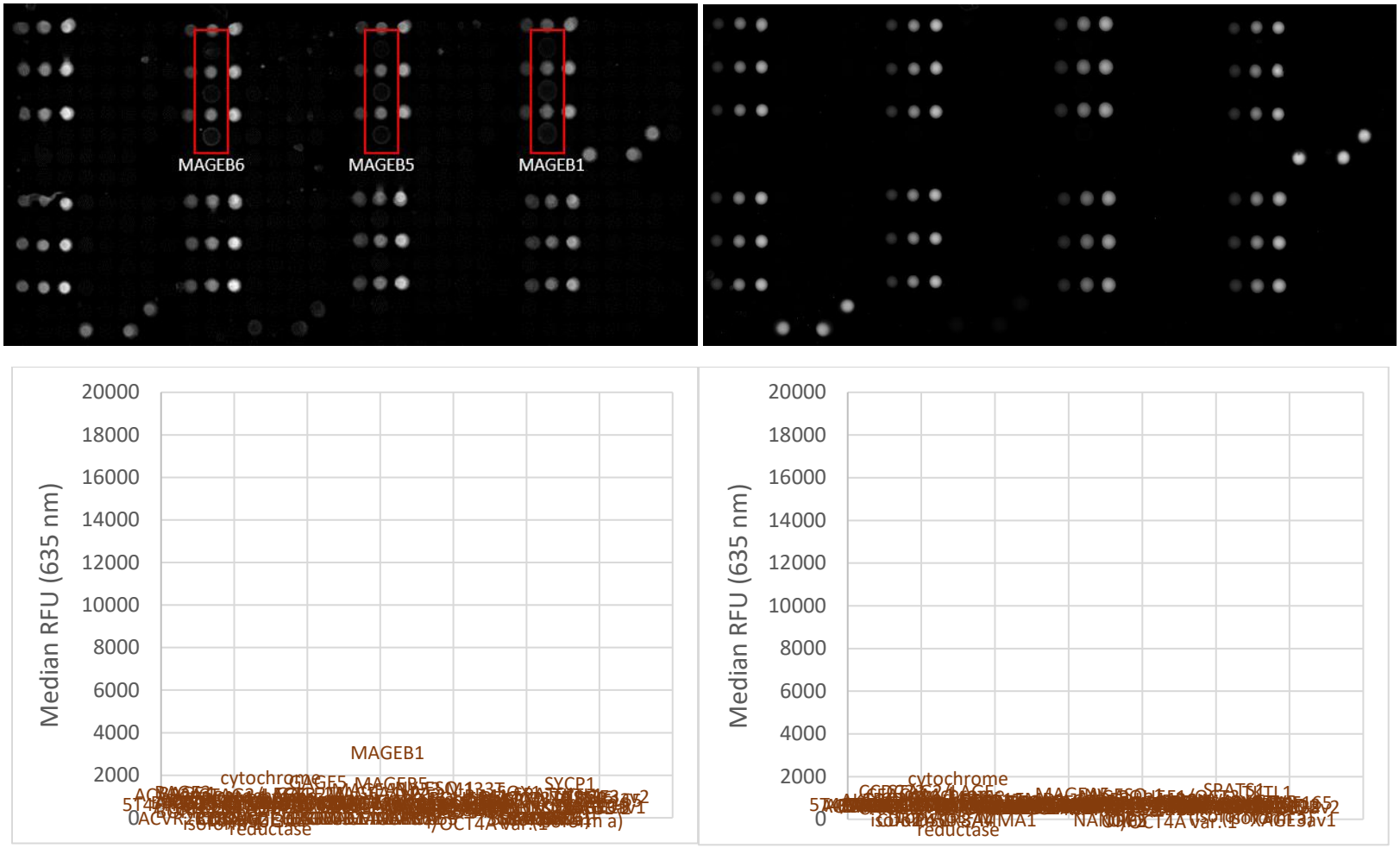


Figure 3.9. CT100plus antigenic signal for CRC045 on citrullinated and non-citrullinated microarrays. Microarray images were scanned Images, and the data extrapolated, filtered and normalised. The CT100plus microarrays treated with (left image and graph) and without (right image and graph) 1 µg/ml PAD4, then subsequently plasma from patient **CRC045**, and detection antibody. Higher signal intensities were detected for MARGEB1, MAGEB5 and MAGEB6 on the citrullinated microarrays *versus* non-citrullinated microarrays. Abbreviations include: RFU = relative fluorescence units; PAD4 = protein arginine deiminase isoform 4.

3.4. CRC and healthy individuals: autoantibody responses toward citrullinated CT100plus antigens

When assessing a biological molecule for its potential as a biomarker, it is critical that the molecule should be absent, or at significantly lower levels, in healthy individuals. We therefore compared the autoantibody response between 59 CRC patients and 11 HCs on the citrullinated CT100plus microarray. It is important to note that the HC samples were not age- nor gender-matched to the CRC patients. Furthermore, although none of the HCs were previously diagnosed with cancer, tests were not performed to confirm that the HCs were CRC-free. However, since our intention here is to acquire preliminary results for a potentially larger study, we decided to use the plasma samples available to assess autoantibody profiles between CRC patient and HCs.

The raw data files from citrullinated and non-citrullinated microarrays were processed through the CT100plus application, and the median antigen intensity obtained. Citrullinated antigens with high or medium intensity, but low intensity on the non-citrullinated microarray, were identified and the top 5 citrullinated antigens, to include CDK7 (*N* = 14/43, 33%), MAGEB6 (*N* = 14/43, 33%), MAGEB1 (*N* = 12/43, 28%), MAGEB5 (*N* = 11/43, 26%), and SYCP1 (*N* = 11/43, 26%), were used for further analysis. For this analysis, ROC analysis was performed, which reports the sensitivity and specificity of tested parameter *i.e.* antigens. An *a priori* ROC sample size was performed using MedCalc software (version 17.2), with a hypothesised AUC-value of 0.80, a null hypothesis value of 0.5, and a sample ratio of 0.19. The results produced a α-value (significance) of 0.01 and β-value (1-Power) of 0.20 for the 59 CRC patients and 11 HCs, indicating that our sample size was sufficient for the ROC analysis.

The ROC analysis was performed in R Studio (1.0.136) using the ROCR package (Sing et al., 2005). The ROC analysis produced sensitivity-, specificity-, AUC-, and cut-off-values for the top 5 antigens (Table 3.1). These 5 proteins were selected for further ROC analysis in CombiROC (Mazzara et al., 2017) to determine the combination of proteins that produce the highest AUC-,sensitivity-, and specificity-values. Here, the CombiROC analysis results indicate that the combination of CDK7 and MAGEB6 produced the highest AUC-, sensitivity- and specificity-values of 0.88, 1.0 and 0.82, respectively. Figure 3.10 represents the combined ROC curve for CDK7 and

MAGEB6. It is evident that MAGEB1, MAGEB5 and MAGEB6 have higher AUC-values than the combined antigens, indicating that these antigens alone outperform CDK7 and MAGEB6 together.

Table 3.1. Sensitivity-, specificity-, AUC-, and cut-off-values for top 5 citrullinated antigens, and a combination thereof. Abbreviation include: AUC = area under the receiver operator characteristic (ROC) curve.

Antigen	Sensitivity	Specificity	AUC	Cut-off
CDK7	0.75	0.82	0.74	955
MAGEB1	0.87	0.91	0.89	791
MAGEB5	0.85	1.00	0.89	790
MAGEB6	0.83	0.91	0.92	848
SYCP1	0.78	0.91	0.87	1190

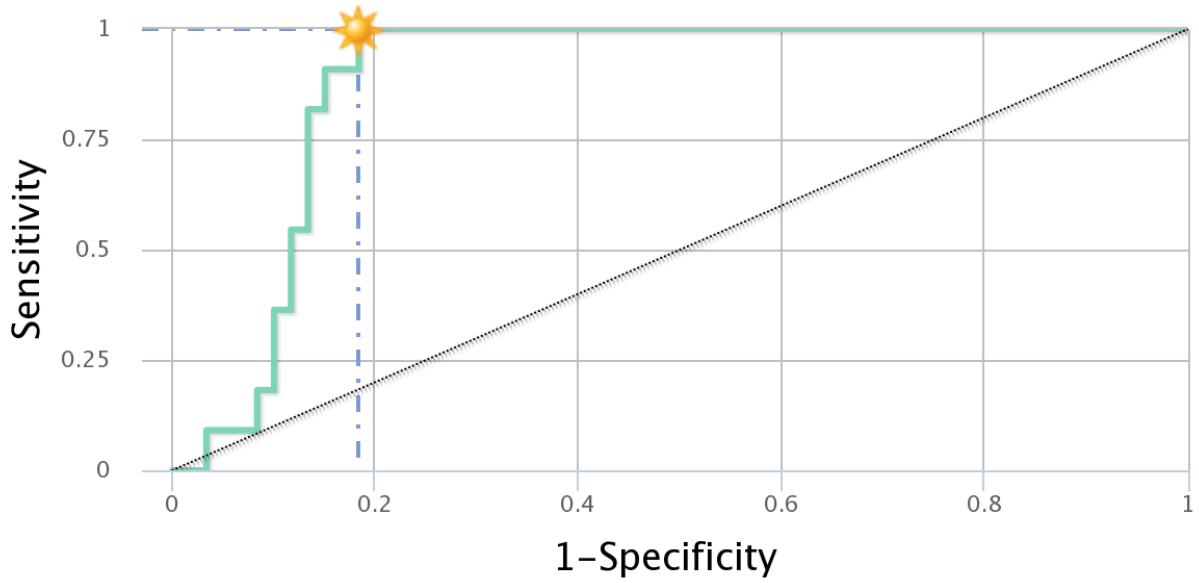


Figure 3.10. ROC curve for CombiROC-derived top 2 antigens for CRC patients. The ROC analysis was performed in CombiROC (http://combiroc.eu/) for the top 5 proteins (CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1). The ROC curve represents the best protein combination, which included CDK7 and MAGEB6. The resulting sensitivity-, specificity- and AUC-values are 1.0, 0.82 and 0.88, respectively. The yellow star indicates the optimal cut-off point. Abbreviations include: ROC = receiver operator characteristic; AUC = area under the ROC curve.

By comparison, it is evident that the signal intensity for the top 5 antigens was higher in CRC patients when compared to healthy individuals (Figure 3.11) on the citrullinated microarray. TO determine whether the difference was statistically significant, the following statistical tests were performed in R Studio (1.0.136): The Shapiro-Wilk (Shapiro and Wilk, 1965) results indicate a non-normal data distribution (p-value < 0.05), as such, the Mann-Whitney *U* (Mann and Whitney, 1947) and Benjamini-Hochberg (Benjamini and Hochberg, 1995) tests were performed, indicating a significant increase (p-value < 0.05; adjusted p-value < 0.05) in signal intensity for CRC patients *versus* HCs on the citrullinated CT100plus microarray (Table 3.2).

It is evident that the signal intensity for the top 5 citrullinated antigens was similar for CRC patients and healthy individuals on the non-citrullinated microarray (Figure 3.12). Again, the Shapiro-Wilk analysis indicated a non-normal data distribution (p-value < 0.05), as such, the Mann-Whitney *U* and Benjamini-Hochberg tests were performed. Here, no significant (p-value > 0.05; adjusted p-value > 0.05) change in signal intensity was detected on the non-citrullinated CT100plus microarray (Table 3.3).

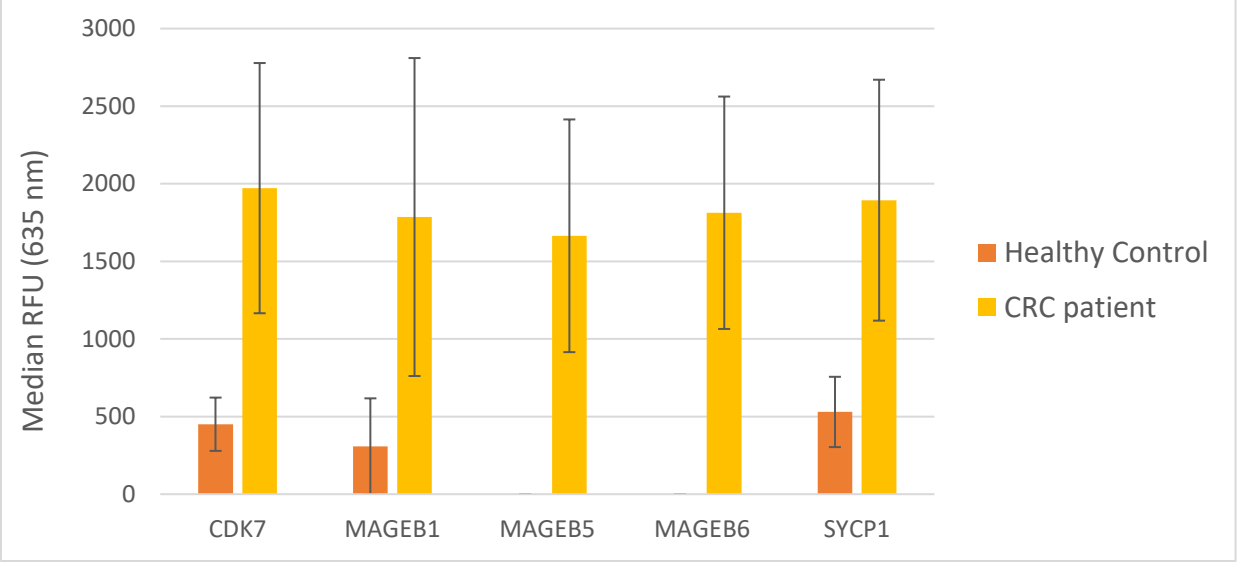


Figure 3.11. Citrullinated microarray: top 5 antigens citrullinated antigens. CRC patient and HC samples were processed on the citrullinated microarray. The top 5 citrullinated antigens were identified; the graph displays the median RFU-values for CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1 across all CRC patients or HCs, and the error bars represents the MAD on the citrullinated microarray. Abbreviations include: CRC = colorectal cancer; HC = healthy control; RFU = relative fluorescence units; nm = nanometre; MAD = median absolute deviation.

Table 3.2. Shapiro-Wilk, Kruskal-Wallis and Benjamini-Hochberg tests values for the top 5 citrullinated antigens on the citrullinated CT100plus microarray.

Antigen	Shapiro Wilk (*p*-value)	Mann-Whitney *U* (*p*-value)	Benjamini-Hochberg (adjusted p-value)
CDK7	1.04×10^{-13}	3.41×10^{-3}	4.51×10^{-3}
MAGEB1	1.20×10^{-6}	2.80×10^{-5}	2.46×10^{-4}
MAGEB5	2.63×10^{-3}	6.73×10^{-6}	2.31×10^{-4}
MAGEB6	2.43×10^{-8}	1.82×10^{-5}	2.31×10^{-4}
SYCP1	3.82×10^{-7}	5.12×10^{-5}	2.88×10^{-4}

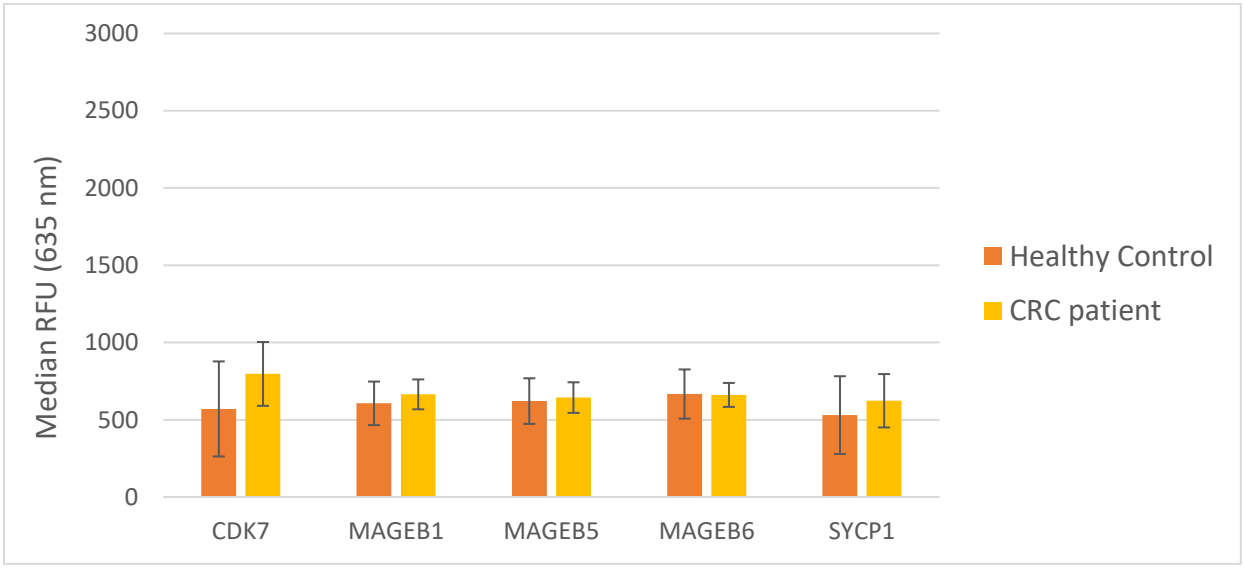


Figure 3.12. Non-citrullinated microarray: top 5 citrullinated antigens. CRC patient and HC samples were processed on the citrullinated microarray. The top 5 citrullinated antigens were identified; the graph displays the median RFU-values for CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1 across all CRC patients or HCs, and the error bars represents the MAD on the non-citrullinated microarray. Abbreviations include: CRC = colorectal cancer; HC = healthy control; RFU = relative fluorescence units, nm = nanometre; MAD = median absolute deviation.

Table 3.3. Shapiro-Wilk, Mann-Whitney *U* and Benjamini-Hochberg tests values for the top 5 citrullinated antigens on the non-citrullinated CT100plus microarray.

Antigen	Shapiro Wilk (*p*-value)	Mann-Whitney *U* (*p*-value)	Benjamini-Hochberg (adjusted *p*-value)
CDK7	2.54×10^{-4}	0.07	0.14
MAGEB1	1.15×10^{-9}	0.14	0.22
MAGEB5	4.83×10^{-5}	0.38	0.50
MAGEB6	5.28×10^{-9}	0.73	0.75
SYCP1	5.55×10^{-15}	0.07	0.14

To further assess the biomarker potential of the top 5 citrullinated antigens, PCA and unsupervised hierarchical clustering were performed in Perseus (version 1.5.4.1). The signal intensity values were $\log_2$ transformed prior to the analysis.

The PCA results display two segregating clusters, each representing CRC patients (green) or HCs (orange), although overlap is evident on the 1D and 2D PCA plots. For both plots, CRC patients and HCs separately, although **CRC064**, **CRC060**, **CRC050**, **CRC039** and **CRC050**, clustering with **HC2**, **HC3**, **HC6**, **H7** and **HC1** (Figure 3.13).

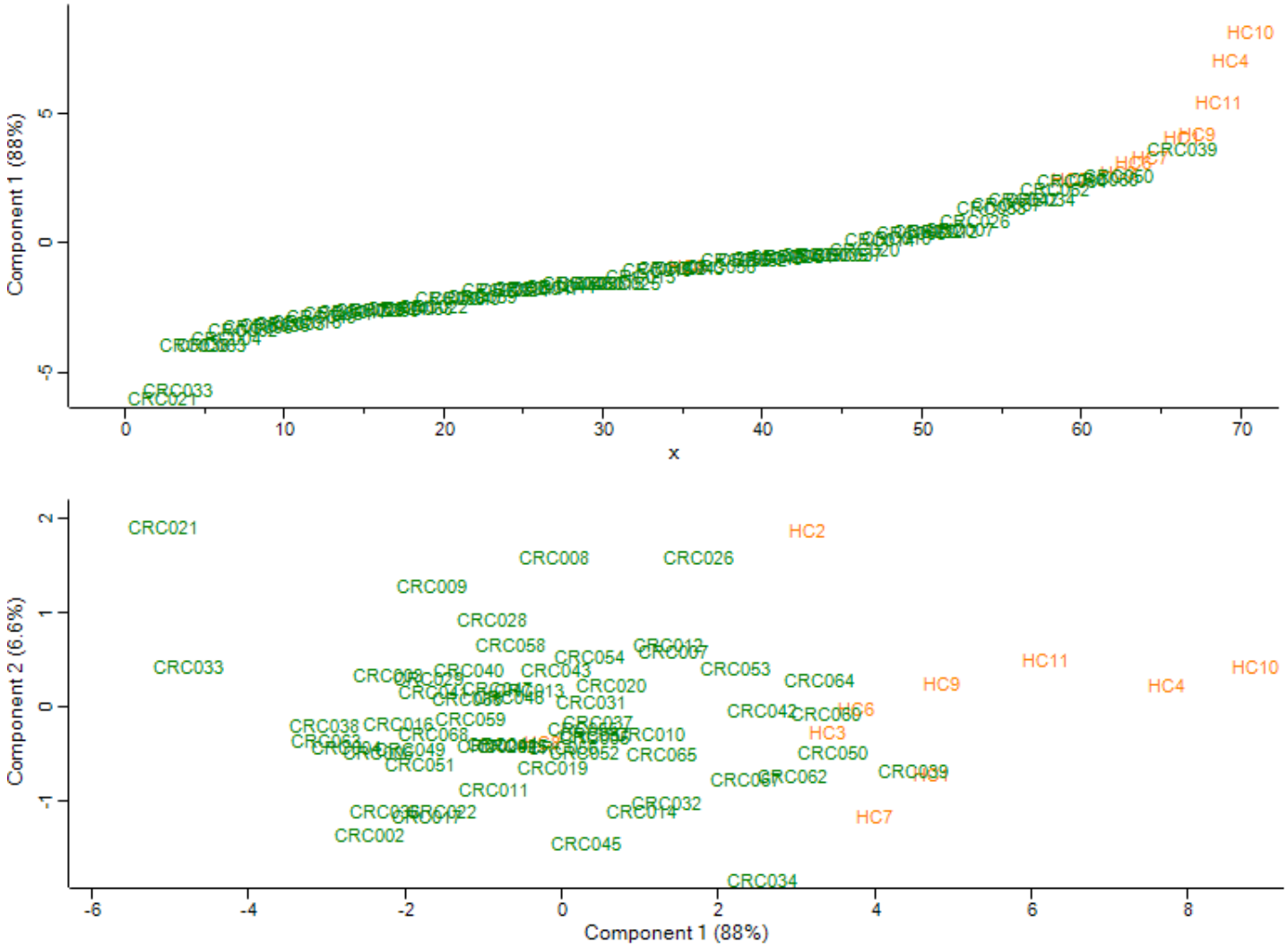


Figure 3.13. PCA between CRC patients and HCs. CRC patient and HC plasma were processed on the citrullinated CT100plus microarray, and the top 5 citrullinated antigens in CRC patients were log_2-transformed and selected for 1D and 2D PCA multivariate testing using the Perseus (version 1.5.4.1). The 1D and 2D PCA plots display distinct CRC patient (green) and HC (orange) clusters. Abbreviations include: PCA = principle component analysis; CRC = colorectal cancer; HC = healthy control; 1D = one-dimensional; 2D = two-dimensional.

3.5. CT100plus antigens associated with CRC patient clinicopathological features

The clinicopathological information of 61 CRC patients whose plasma samples were assayed on the citrullinated CT100plus microarray are summarised in Table 2.3.

The clinical features include patient age, gender, ethnicity, family history, body mass index (BMI), hypersensitivity, smoking and alcohol intake. The pathology information includes cancer stage at disease presentation; MS stability or instability status; whether the disease was sporadic or classified as HNPCC; family history of CRC; anatomical presentation of the cancer specimen; whether inflammation was detected; and whether the patient received radiotherapy. These features were used for subsequent statistical analysis to identify associations between clinicopathological features and autoantibody detection.

The baseline-corrected antigen signal values were obtained prior to statistical analysis. Here, the baseline-corrected values were obtained by subtracting the ROC-derived cut-off-values from CT100plus-derived median-values for each antigen. The ROC-derived cut-off-values used for baseline subtraction is shown in Figure 3.14 for each antigen, resulting in a median-value of 619 (MAD = 95; Range = 500 - 1190).

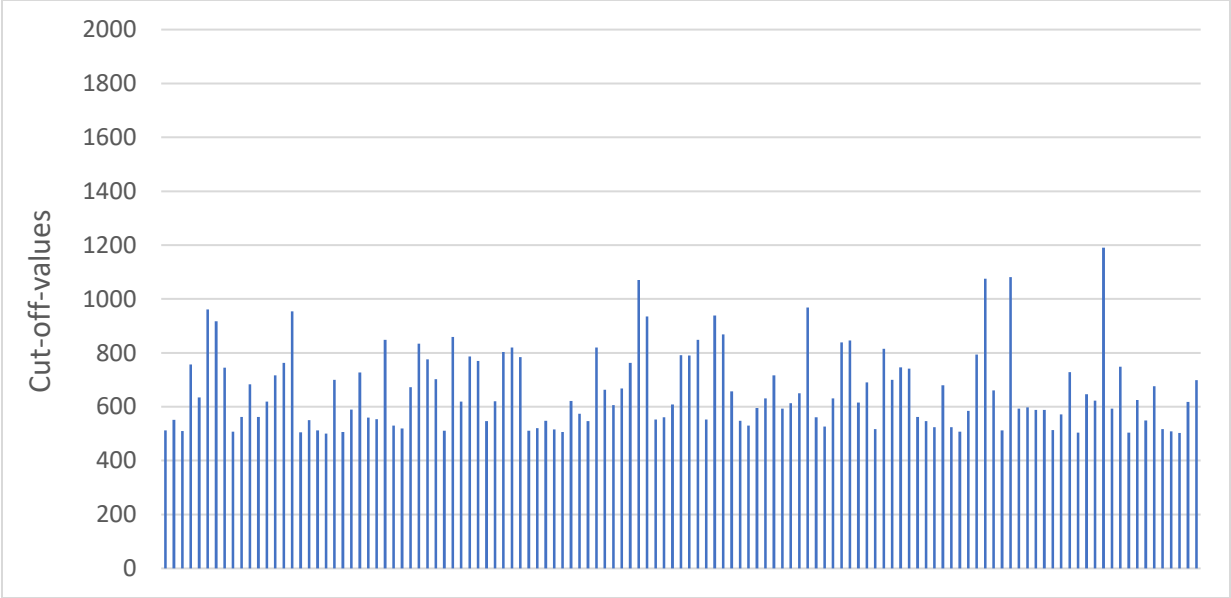


Figure 3.14. ROC-derived cut-off-values for CRC patients. ROC analyses between CRC patients and HCs produced cut-off-values for each antigen; each bar represents the cut-off-value per antigen. These values were used for baseline-correction for each antigen prior to clinicopathological statistical analysis. Abbreviations include: ROC = receiver operator characteristic; CRC = colorectal cancer; HCs = healthy controls.

Clinicopathological features and autoantibody signal associations were determined using the Mann-Whitney *U* test (Mann and Whitney, 1947) when comparing two groups (e.g. Recurrence), and the Kruskal-Wallis test (Kruskal and Wallis, 1952) when comparing three or more groups (e.g. Cancer stage). Statistical significance was assessed using the Benjamini-Hochberg test (Benjamini and Hochberg, 1995), where an adjusted *p*-value ≤ 0.05 was deemed significant. All tests were performed in R Studio (1.0.136), and only the Kruskal-Wallis test analysis was performed with the PMCMR package (Maintainer and Pohlert, 2016).

No significant association between clinicopathological features and autoantibody signal were identified for CRC patients assayed on the citrullinated CT100plus microarray. Although a *p*-value of ≤ 0.05 was obtained for associations between “smoking” and the signal intensity of MICA, the Benjamini-Hochberg correction test indicated that the changes were non-significant. The results for *p*-values ≤ 0.01 and adjusted *p*-values are summarised in Table 3.4.

Table 3.4. Clinicopathological features and autoantibody signal with p-values ≤ 0.01. Mann-Whitney U test indicate a significant difference in antigenic signal for the smoking, although it is not statistically significant after the Benjamini-Hochberg test was applied. Abbreviation include: TAA = tumour-associated antigen.

Clinical feature	TAA	*p*-value	Adjusted *p*-value
Smoking	ACVR2B	0.007	0.80

Discussion and conclusion

Arginine is a positively charged amino acid, which when converted to citrulline results in a monoisotopic mass difference of +0.984016 Daltons (Da), attaining a neutral charge (Rogers and Simmonds, 1958). The shift in charge caused by citrullination affects protein-protein interactions, hydrogen bond formation, protein structure, and in some cases, can cause denaturation (Tarcsa et al., 1996; Knuckley et al., 2010). In fact, it is hypothesized that citrullination-induced changes in protein surface results in protein structure alterations that activate autoimmune responses when the protein is released in to the surrounding interstitial tissues. Interestingly, for our assays, CRC plasma produced a higher median signal intensity on the citrullinated (1100 RFU, MAD = ±402) *versus* non-citrullinated (637 RFU; MAD = ±120) CT100plus microarray, possibly due to increased non-specific binding on citrullinated antigens caused by increased hydrophobicity of the citrulline compared to arginine.

For this study, our aim was to assess the autoantibody response of CRC patients to citrullinated antigens on the CT100plus microarray. The first objective was to determine the appropriate PAD concentration needed to citrullinated antigens on the CT100plus microarray to acquire a significant level of citrullination. Since PAD4 was commercially available, and since it catalyses the same citrullination reaction as PAD2, it was used to citrullinate antigens on the CT100plus microarray. It is expected that surface-exposed arginine amino acids will be converted to citrulline on all proteins on the slide surface, including streptavidin, antigens and positive controls. From our results, it is evident that the median signal intensity from streptavidin is lower than the median signal intensity from antigens (Figure 3.2). There are several reasons why the foreground (antigen spots) intensity is greater than the background (streptavidin) intensity: The foreground signal may result from the cumulative intensity derived from citrullinated streptavidin and antigen present within antigen spot, resulting in a higher foreground *versus* background intensity. Alternatively, the higher foreground signal may be derived from citrullinated antigen alone as steric hindrance may prevent streptavidin citrullination within the antigen spot. If the foreground signal is derived from antigen alone, there are several possible explanations for decreased levels of citrullinated streptavidin *versus* antigens: (**1**) The

streptavidin-biotin interaction changes streptavidin conformation, limiting PAD4-induced citrullination, (**2**) On average, the number of surface arginine residues is higher for the antigens *versus* streptavidin, or (**3**) the arginine residues in streptavidin are flanked by amino acids which limit PAD-induced citrullination. Together, the results indicate the successful conversion of arginine to citrulline on antigens present on the CT100plus microarray. To determine the concentration of PAD4 required, a Mann-Whitney *U* test was performed comparing untreated and the 8 PAD4-treated microarrays, and here we found that 1 µg/ml PAD4 was sufficient for a significant (*p*-value = 0.005) increase in citrullination, which also falls within the linear range when detecting citrullination (Figure 3.4). It is important to incubate the microarray with PAD4 in the linear range concentration, as a low PAD4 concentration may result in under citrullination whereas high PAD4 concentration may result in autocitrullination, which reportedly inactivates the PAD4 enzyme (Andrade et al., 2010). Thus, 1 µg/ml PAD4 was used to citrullinate antigens on the CT100plus microarray before assaying CRC patient plasma.

The second objective was to compare the differential CRC patient autoantibody signals on the citrullinated and non-citrullinated CT100plus microarray. The comparative analysis showed that 11/43 (26%) of patients produced low RFU signal for all antigens on both microarray-types; an example is shown for patient **CRC042** in Figure 3.7, indicating that this patient subgroup does not produce autoantibodies against citrullinated or non-citrullinated CT100plus microarray antigens. Twenty-one of the 43 (49%) patients elicited a medium or high response on the unmodified microarray, but only 6/43 (14%) patients retained the signal toward these antigens on the citrullinated microarray. An example of decreased signal is shown for **CRC010** (Figure 3.8), who produced an autoantibody response toward DDX54 on both array-types, although a high signal is observed on the non-citrullinated microarray whilst a medium signal is observed on the citrullinated microarray. Decreased autoantibody signal on the citrullinated microarray may result from citrullination-induced epitope destruction. Interestingly, 28/43 (65%) patients produced a medium and/or high autoantibody response to antigens on the citrullinated microarray, although baseline intensity levels were detected on the non-citrullinated microarray. An example is shown for **CRC045** (Figure 3.9) who produced baseline signals on the non-

citrullinated microarray, but higher signals intensities for MAGEB1, MAGEB5 and MAGEB6 on the citrullinated microarray.

A critical consideration when selecting a potential biomarker molecule is that the molecule should be detectable in patients, but absent or at significantly lower levels in healthy individuals. Our final objective was therefore to compare the autoantibody responses of 59 CRC patients and 11 healthy individuals toward citrullinated antigens on the CT100plus microarray. The top 5 antigens, to include CDK7, MAGEB6, MAGEB1, MAGEB5, and SYCP1, with medium or high signal on the citrullinated, but low signal on the non-citrullinated microarray were analysed:

The ROC analysis produced sensitivity-values ranging from 0.75 – 0.87; specificity-values ranging from 0.82 - 1.0; and AUC-values ranging from 0.74 – 0.92. A combination analysis was also performed for all antigens, and the best combination was for CDK7 and MAGEB6, which together produced sensitivity-, specificity-, and AUC-values of 1.0, 0.82 and 0.88, respectively (Table 3.1). In the literature, ROC assessments are reported for potential and currently used CRC protein biomarkers, including CEA, IL-8, Apo AI, C9, and CA-19-9. Here, the AUC-values, either alone or in combination, ranged from 0.62 - 0.90 (Murakoshi et al., 2011; Lee, 2013; Jin et al., 2014; Zhang et al., 2015). Thus, our citrullinated antigen panel produced similar AUC-values to the best performing CRC biomarkers tested. To further assess the selected top 5 antigen panel, a multivariate PCA was performed in Perseus (version 1.5.4.1). Distinct groupings of CRC patients and HCs formed on the 1D and 2D PCA plots, with overlap between CRC patients and **HC1**, **HC3, HC6, HC7** and **HC8** (Figure 3.13). Overlap may be expected as a ROC specificity-value of 0.91 was obtained, indicating a relatively high, but not perfect, discrimination score. Together, the Mann-Whitney U *test*, ROC and PCA results provide evidence for the first time that citrullinated antigens discriminate between CRC patients and HCs with high sensitivity and specificity, and may be potential CRC diagnostic biomarkers.

Autoantibody responses to citrullinated and non-citrullinated CDK7, MAGEB6, MAGEB1, MAGEB5, and SYCP1 have not been reported previously. CDK7 and MAGE proteins are reportedly expressed in CRC, which is summarised below:

Cyclin-dependent kinase, or CDK7 (http://www.uniprot.org/uniprot/P50613), is a serine/threonine kinase involved in cell cycle control and RNA polymerase II-mediated RNA transcription. Cyclin-dependent kinases are activated by binding to cyclin and mediate the progression through the cell cycle. A study with multiple cell-lines, including the HCT116 colon cancer cell-line, proposed that CDK7 may promote cancer cell-cycle progression and improve cell survival, as low doses of CDK7 inhibitors caused delays in cell-cycle progression and diminished tumour cell survival. High doses of CDK7 inhibitor caused cell death, which they concluded resulted from the inability of CDK7-induced mRNA synthesis (Kelso et al., 2014). Autoimmune and autoantibody responses toward non-citrullinated and citrullinated CDK7 have not been reported. However, using the solved structure of CDK7 in PyMOL (version 2.0.4), we identified arginine residues on the surface of CDK7 (Figure 3.15) available for conversion to citrulline in CRC cells or on the CT100plus microarray. In conclusion, CRC-related CDK7 activity has only been described for *in vitro* colon cancer cell-lines but here we provide supporting evidence for the citrullinated form of CDK7 in CRC patients.

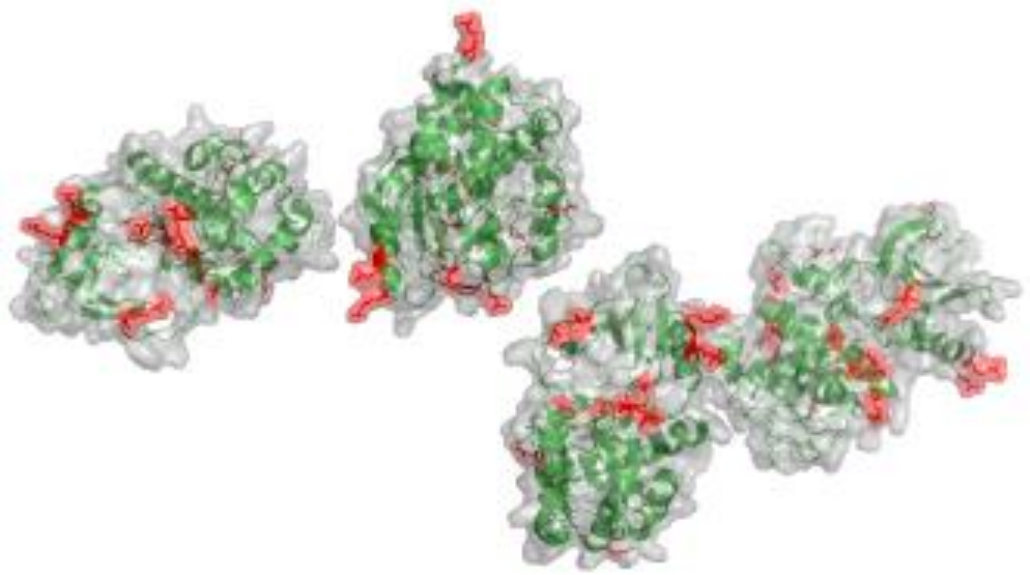


Figure 3.15. Crystal structure of CDK7 with arginine residues highlighted in red. CDK7 was detected at medium and high signal intensities on the citrullinated microarray, but low signal intensities on the non-citrullinated microarray, indicating that the citrullinated version CDK7 induces an autoantibody response in CRC patients. The image represents the crystal structure of CDK7 (PDB: 1UA2). The arginine-residues available for PAD-induced citrullination are indicated in red. Abbreviations include: CRC = colorectal cancer; PAD4 = protein arginine deiminase isoform 4; PDB = protein data bank.

Melanoma antigen gene, also known as MAGE, is a large protein family of highly conserved proteins that share the MAGE homology domain. MAGE family members can be divided in to two categories: Type I, which are considered CTAs in humans, and include MAGE-A, -B, and -C subfamily members; and Type II, which include MAGE-D, -E, -F, -G, -H, -L subfamilies, and are expressed in many tissues throughout the body (Weon and Potts, 2015). Jungbluth *et al*. did not detected the expression of MAGEA-1, -3, -4, -6 and -12 in normal nor cancer tissues of colon and rectum, which they suggest may be due to the poor sensitivity of immunohistochemistry (Jungbluth et al., 2000a). In contrast, a large number of studies identified MAGE mRNA and protein expression in CRC cell-lines and tissues (Mori et al., 1996; Li et al., 2005; Chung et al., 2010; Jeon et al., 2011; Choi and Chang, 2012). Several of these studies observed a significant association between increased MAGE expression in patient cancer tissues and liver metastasis (Mori et al., 1996; Li et al., 2005; Choi and Chang, 2012). Autoimmune and autoantibody responses to cancer antigens have been reported, suggesting that the immune system may be an excellent target for cancer diagnosis, screening and treatment. Consistent with this idea, a study of 20 CRC patients treated with a MAGEA DC vaccine showed a median progression-free survival of 2.4 months, with 2 patients showing progression-free survival for more than 27 months (Toh et al., 2009). MAGEB1, MAGEB5 and MAGEB6 are not reportedly expressed in CRC, but MAGEB1 is reportedly expressed in other gastrointestinal carcinomas, including oesophageal SCCs (Nagashima et al., 2002). *MAGE-D2* transcripts and the MAGED2 protein are reportedly overexpressed in CRC (Hashimoto et al., 2015). Thus, the autoantibody signal detected on the CT100plus microarray may be directed against an epitope shared between MAGEB and MAGED2. Interestingly, MAGED2 was also identified in our AP-MS results (Chapter 4). Although MAGEB and MAGED subfamilies form part of type I and type II MAGEs, respectively, both MAGE-types share a MAGE homology domain (MHD) that spans 170 amino acids, which on average is 46% conserved amongst all human MAGEs. It is possible that an autoimmune response is raised against citrullinated MAGED2, and that the autoantibodies cross react with MAGEB1, MAGEB5 and MAGEB6 on the CT100plus microarray. Unfortunately, the structures of MAGEB1, MAGEB5 and MAGEB6 are not reported, and potential surface epitopes could not be identified. In

conclusion, although CRC-related MAGE activity has previously been reported, for the first time we provide evidence for the detection of citrullinated forms in CRC patients.

Synaptomenal complex protein 1, or SYCP1 (www.uniprot.org/uniprot/Q15431), is a major component of the transverse chromosomes during meiotic prophase. SYCP1 is needed for normal assembly of the central element of synaptomenal complexes, normal centromere pairing during meiosis, normal meiotic chromosome synapsis during oocyte and spermatocyte development, and for normal male and female fertility. SYCP1 is not reportedly expressed in CRC, including autoimmune and autoantibody responses. Although, using the solved structure of SYCP1 in PyMOL (version 2.0.4), we identified arginine residues on the surface of SYCP1 (Figure 3.16) available for conversion to citrulline in CRC cells or on the CT100plus microarray. In conclusion, we provide supporting evidence for the citrullinated form in CRC patients.

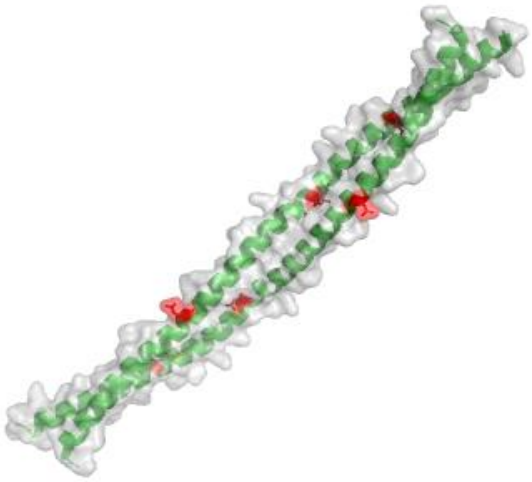


Figure 3.16. Crystal structure of SYCP1 with arginine residues highlighted in red. SYCP1 was detected at medium and high signal intensities on the citrullinated microarray, but low signal intensities on the non-citrullinated microarray, indicating that the citrullinated version SYCP1 induces an autoantibody response in CRC patients. The image represents the crystal structure of SYCP1 (PDB: 4YTO). The arginine-residues available for PAD4-induced citrullination is indicated in red. Abbreviations include: CRC = colorectal cancer; PAD 4 = protein arginine deiminase isoform 4; PDB = protein data bank.

In Chapter 4, the AP-MS assay was used to identify novel and known TAAs and TSAs of CRC. We obtained supporting evidence that a fraction of significantly expressed CT100plus antigens were

also identified by AP-MS (Chapter 2 and 4). PAD2 was also detected in the AP-MS data for 2 patients: **CRC002** showed PAD2 expression in *Cancer* only, whereas **CRC004** showed PAD2 expression in *Cancer* and *Normal*. The MaxQuant search was extended to include the detection of citrullinated peptides; a total of 58 proteins, 87 peptides and 108 citrullination sites were identified. Vimentin and actin were identified, which are normal substrates of PAD2 (Zhang et al., 2012; Hsu et al., 2014). In addition, citrullinated tubulin, heat shock protein, and fibrinogen were also identified which are PAD2 substrates in cancer cell-lines and are associated with tumour growth, invasion, and a poor prognosis (Jiang et al., 2013). In this Chapter, a top 5 antigen panel was selected to include CDK7, MAGEB6, MAGEB1, MAGEB5, and SYCP1 as the most frequently detected antigens on the citrullinated microarray. Although citrullinated peptides of these proteins were not detected by AP-MS, non-citrullinated peptides of the CDK5 and MAGED2 homologues were identified. CDK5 and CDK7 share 46% homology, whereas MAGED2 shares ~ 40% homology with MAGEB1, MAGEB5 and MAGEB6. Together, these results suggest that the detected CT100plus autoantibody signals may be toward homologues of CT100plus antigens, although this should be confirmed through alternate methods. In addition, citrullinated peptides may not have been sufficiently enriched, and were therefore below the limit of detection for the mass spectrometer. Citrulline-enrichment using Biotin-PEG_2-4-glyoxalbenzoic acid (BPG) improved the detection from 119 to 3600 citrullinated peptides (Tutturen et al., 2014), suggesting that citrullinated peptide-enrichment may be important to detect more citrullinated peptides in future studies.

We found a non-significant (adjusted *p*-value > 0.05) association between patient clinicopathological information and autoantibody signal intensity. Although *p*-values ≤ 0.05 were obtained for associations between “smoking” and the signal intensity for MICA, the Benjamini-Hochberg multiple testing correction indicated that these changes were non-significant. Unlike the melanoma cohort, these patient samples were not collected as part of a clinical cohort. Thus, when considering the diversity of the cohort (Table 2.3), and only assessing 123 TAA, this result is plausible. It is therefore important that future CT100plus assays include carefully selected patient samples that form part of a controlled clinical cohort, which may result in statistically significant associations between clinicopathological features and autoantibody signal intensity.

In conclusion, to our knowledge, this is the first time citrullinated cancer proteins have been used to detect autoantibodies in CRC patients, indicating that citrullinated proteins can be used for CRC diagnosis. Thus, it may be critical to consider protein citrullination when developing diagnostic and screening tools for CRC diagnosis, as has been done here for the CT100plus microarray.

Chapter 4

CRC antigen biomarker identification using the IMMUNOME™ microarray and affinity purification-mass spectrometry (AP-MS)

Abstract

TAAs can induce an autoantibody response in its host; these autoantigens have been identified using high-throughput methods including phage display, SEREX and SERPA. These methods identify multiple potential antigens in a single experiment; however, the disadvantages include the use of non-native TAAs that lead to the loss of important epitopes due to expression in *Escherichia coli* (*E.coli*), protein denaturation (e.g. in SERPA), and absent PTMs. We previously processed a cohort of 62 CRC patient plasma samples on the CT100plus microarray which contains 123 TAAs (Chapter 2 and 3). In this Chapter, we processed two CRC plasma samples on the SENGENICS IMMUNOME™ microarray functionalized with 1622 native eukaryotic proteins. The highest autoantibody signals were toward proteins not present on the CT100plus microarray, including NOL4 and TXN2 for patient **CRC038**, and LGALS1 and KRT19 for patient **CRC050**. We also developed an AP-MS assay to detect autoantibody responses to autologous native tissue proteins. For the assay, 5 CRC patient samples were processed, whose plasma were also processed on the CT100plus microarray. For 5/5 (100%) patients we identified proteins via AP-MS that were also detected on the CT100plus microarray. Furthermore, we also identify citrullinated proteins associated with cancer progression. In conclusion, we identify common TAAs for potential diagnostic use using complementary immunoproteomic techniques that contain native proteins.

Introduction

Cancer patients can produce autoantibodies to proteins that are mutated, misfolded, improperly glycosylated, over-expressed, truncated or aberrantly localised within tumour cells. Research on TAAs, as well as their related autoantibodies, has provided an abundance of targets for therapy, prognostication, diagnosis, and response to therapy. Several techniques have been used to identify TAAs, including SEREX, SERPA and protein micro-array technology (Gunawardana and Diamandis, 2007), discussed in Chapter 1.

Using the CT100plus microarray, we identified the autoantigenic targets of autoantibodies that are expressed in CRC patients (Chapter 2 and 3). Since only 123 antigens are assessed on the CT100plus microarray, we sought to employ tools or assays that allow the assessment of a wider array of protein targets. Thus, we assayed CRC patient samples on the SENGENICS IMMUNOME™ microarray, which is functionalized with 1622 proteins and based on the same underlying technology used to construct the CT100plus microarray. Thus, we expect robust and reliable detection of autoantibodies toward autoantigens. To expand the autoantigen identification capabilities even further, AP-MS was used to capture CRC patient IgG and subsequently capture native cancer antigens from autologous CRC patient tissue samples. Since patient IgG was used to capture antigens from whole tissue lysates, the autoantibody response to the entire CRC proteome was assessed. In this chapter, we discuss the results of CRC patient autoantibody response on the SENGENICS IMMUNOME™ microarray, and discuss the development and results of the AP-MS assay.

Results

4.1. Autoantibody detection on the SENGENICS IMMUNOME™ microarray

The SENGENICS IMMUNOME™ microarray contains 1622 proteins, comprised of cancer antigens, transcription factors, kinases and signalling proteins (Suwarnalata et al., 2016). The signal intensities from SENGENICS IMMUNOME™ microarrays treated with detection Ab only (blank) and pooled HCs plus detection Ab were provided by SENGENICS®. The data from the SENGENICS IMMUNOME™ microarrays were extrapolated using stringent processing and filtered steps: whole array CV ≤ 20%; probe replicate CV ≤ 10%; noise-threshold ≥ 2 SD above background; maximum threshold = 65355 nm; whole array filtering control = Cy3-biotin-BSA.

The autoantibody profiles for the two selected patients were compared between the SENGENICS IMMUNOME™ microarray and the CT100plus microarray. A blank SENGENICS IMMUNOME™ microarray slide was assessed to determine signal levels from a microarray incubated with Cy3-conjugated anti-human IgG detection antibody. Here, the median antigen signal intensity was 144 (MAD = 12) RFU, and most antigens produced signal intensities below 500 RFU, except for IGHG1 which produced a signal intensity of 14115 RFU (Figure 4.1), which confirmed the addition of Cy3-conjugated anti-human IgG detection antibody. The SENGENICS IMMUNOME™ microarray was also incubated with pooled HC plasma - the signal intensities displayed in Figure 4.2. The median signal intensity was 844 (± 96) RFU, reflecting low non-specific binding between the functionalised proteins and plasma proteins. Although 95.7% of antigens displayed low signal intensities (200-2000 RFU), 67 antigens produced medium signal intensity (2000-10000 RFU) and 2 antigens produced high signal intensity (>10000 RFU). As expected, a high signal intensity was detected for IGHG1 (15162 RFU), the constant region of the IgG heavy chain, confirming the addition of detection Ab which binds to the constant region of human IgG molecules. However, a high signal intensity was also detected for RBPJ (10954 RFU), which – according to the array manufacturer - always produces a high signal intensity on the microarray, even when assaying plasma from healthy individuals.

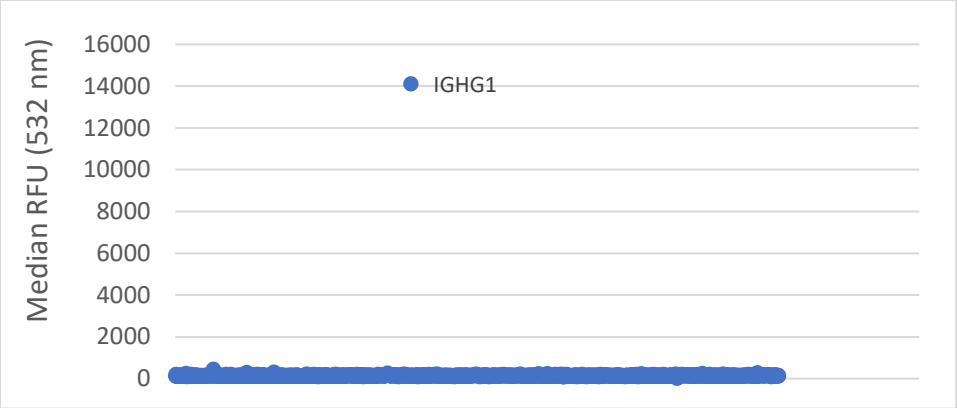


Figure 4.1. Antigen intensities for a Blank SENGENICS IMMUNOME™ microarray. The SENGENICS IMMUNOME™ microarray was incubated with serum albumin buffer (PBS, BSA and Triton™ X-100) and Cy3-conjugated anti-human IgG detection antibody. Fluorescence images were obtained at 532 nm, and numerical data extrapolated. The net intensity for each antigen was obtained, and the data was subsequently filtered using stringent criteria (control CV ≤ 20%; antigen CV ≤ 10%; noise threshold ≥ 2 SD above background). The scatter plot displays the median intensity-values for each antigen. The image represents only a quarter of the microarray, which contains one of the 4 quadruplicate antigens spots. Abbreviations include: PBS = phosphate buffer saline; BSA = bovine serum albumin; Cy3 = cyanine 3; IgG = immunoglobulin G; nm = nanometres; CV = coefficient of variation; SD = standard deviation.

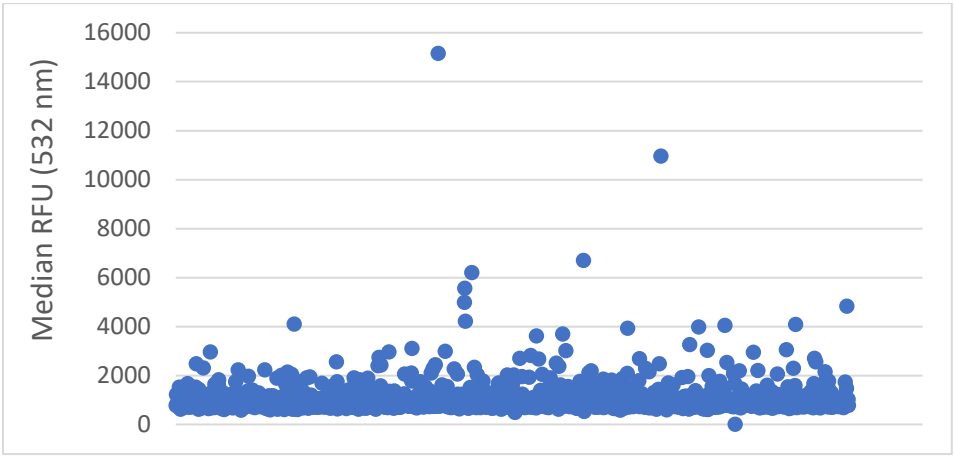


Figure 4.2. Antigen intensities for the SENGENICS IMMUNOME™ microarray treated with pooled HC plasma. The SENGENICS IMMUNOME™ microarray was incubated with pooled HC plasma and Cy3-conjugated anti-human IgG detection Ab to determine baseline signal intensities for various antigens. Fluorescence images were obtained at 532 nm, and numerical data extrapolated. The net intensity for each antigen was obtained, and the data was subsequently filtered using stringent criteria (control CV ≤ 20%; antigen CV ≤ 10%; noise threshold ≥ 2 SD above background). The scatter plot displays the median intensity-values for each antigen. The image represents only a quarter of the microarray, which contains one of the 4 quadruplicate antigens spots. Abbreviations include: HC = healthy control; Cy3 = cyanine 3; IgG = immunoglobulin G; Ab = antibody; nm = nanometres; CV = coefficient of variation; SD= standard deviation.

The plasma samples for **CRC038** and **CRC050** were previously processed on the CT100plus microarray (Duarte *et al*., unpublished data), and showed high autoantibody signals. These patient plasma samples were therefore also assessed on the SENGENICS IMMUNOME™ microarray to detect autoantibody signals toward other cancer antigens. Unlike the CT100plus microarray, the SENGENICS IMMUNOME™ microarrays were processed manually in a slide chamber as the Tecan HS4800 Pro automated microarray hybridization station modules cannot accommodate the physical dimensions of the SENGENICS IMMUNOME™ microarray. The SENGENICS IMMUNOME™ microarray images were obtained at 523 nm using the AUTOPMT setting

We were limited by the number of SENGENICS IMMUNOME™ microarrays available, and so we obtained the immunome profile of 2 patients only: **CRC038** and **CRC050**. When assessing the autoantibody response of **CRC038** (Figure 4.3) and **CRC050** (Figure 4.4) on the CT100plus microarray, the median intensity for all antigens across the microarray was 0 (MAD = 0) and 642 (MAD = 267), respectively. The higher median intensity was likely due to low non-specific binding from **CRC050** plasma, a phenomenon observed previously for patients within this cohort (Chapter 2 and 3) and the melanoma cohort (Chapter 5). When assessing the median autoantibody response of **CRC038** (Figure 4.5) and **CRC050** (Figure 4.6) on the SENGENICS IMMUNOME™ microarray, the median intensity for all antigens across the microarray was 847 (MAD = 135) and 9584 (MAD = 1610), respectively. Taken together, the median intensity was higher for **CRC050** than **CRC038** on both microarray-types, although the difference was more marked on the SENGENICS IMMUNOME™ microarray. These results are supported by the microarray images presented for **CRC038** (Figure 4.5) and **CRC050** (Figure 4.6). The methods for slide preparation are different for each microarray-type: all SENGENICS IMMUNOME™ microarray processing, *i.e.* plasma incubation, detection Ab incubation and washes, were performed manually in a microarray slide holder, whereas all CT100plus microarray processing is confined and automated in the Tecan HS4800 Pro automated microarray hybridization station, resulting in the production of higher quality microarray data.

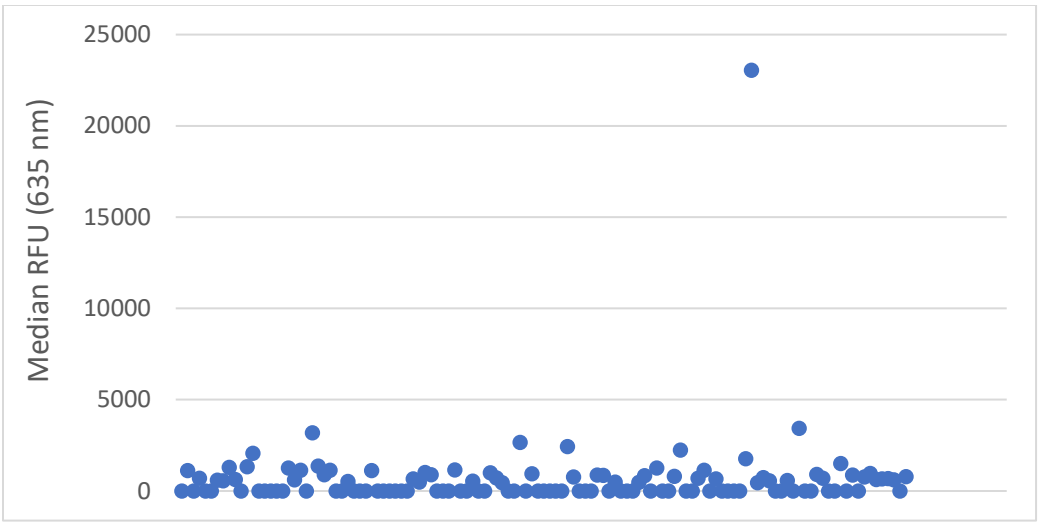


Figure 4.3. Scatterplot representing the autoantibody intensities for CRC038 on CT100plus microarray. The CT100plus microarray was incubated with the blood plasma of **CRC038**, and subsequently incubated with fluorescently-labelled anti-human IgG. Fluorescence images were obtained at 635 nm, and numerical data extrapolated using the CT100plus.jar software: Antigen net intensities were obtained, and the data was processed and filtered: whole array CV ≤ 25%; replicate probe CV ≤ 20%; noise threshold ≥ 2 SD above background). The scatter plot displays the median intensity-values for each antigen. Abbreviations include: IgG = immunoglobulin G; nm = nanometres; CV = coefficient of variation; SD = standard deviation.

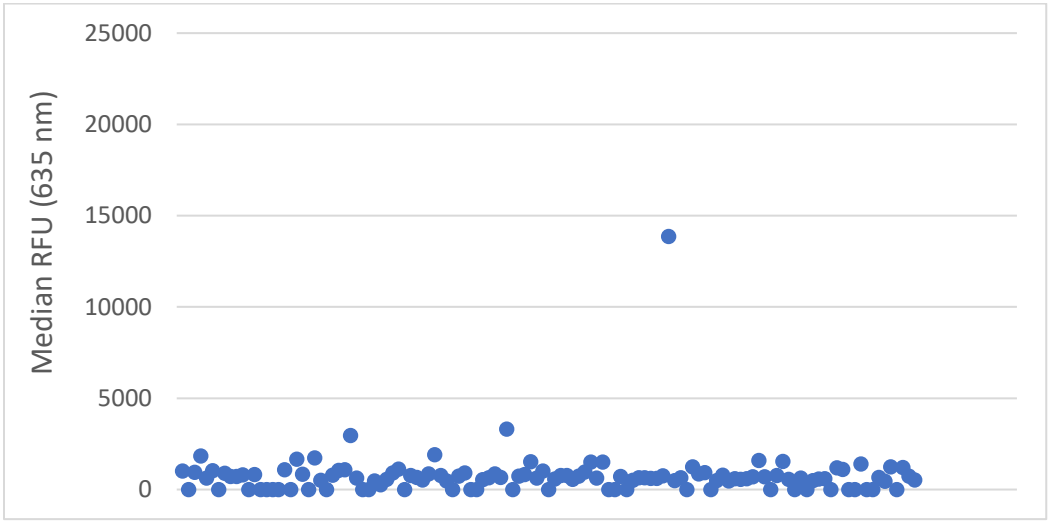


Figure 4.4. Scatterplot representing the autoantibody intensities for CRC050 on CT100plus microarray. The CT100plus microarray was incubated with the blood plasma of **CRC050**, and subsequently incubated with fluorescently-labelled anti-human IgG. Fluorescence images were obtained at 635 nm, and numerical data extrapolated using the CT100plus.jar software: Antigen net intensity were obtained, and the data was processed and filtered: whole array CV ≤ 25%; replicate probe CV ≤ 20%; noise threshold ≥ 2 SD above background). The scatter plot displays the median intensity-values for each antigen. Abbreviations include: IgG = immunoglobulin G; nm = nanometres; CV = coefficient of variation; SD = standard deviation.

When assessing the autoantibody response of **CRC038** on the CT100plus microarray (Figure 4.3), there was high signal toward SILV; a medium signal toward SPATS1 (isoform a), CT47.11, MAGEA4v2, MAPK1, p53 S46A and CRC25A; and low or no signal toward the remaining antigens. However, when assessing the autoantibody response of **CRC038** on the SENGENICS IMMUNOME™ microarray (Figure 4.5), the highest autoantibody response detected was against NOL4 (15477 RFU), TXN2 (10218 RFU), PPP2R4 (8956 RFU), SDCCAG8 (7611 RFU) and TFG (7565 RFU), which are all antigens not present on the CT100plus microarray. When assessing the autoantibody response of **CRC050** on the CT100plus microarray (Figure 4.4), there was high signal toward the mutant form p53 Q136x; a medium signal toward MAGEA11 and cytochrome p450 3A4; and low or no signal toward the remaining antigens. However, when assessing the autoantibody response of **CRC050** on the SENGENICS IMMUNOME™ microarray (Figure 4.6), the highest autoantibody response was detected was for LGALS1 (34263 RFU), KRT19 (22412 RFU), FTH1 (21505 RFU), PLK1 (19694 RFU) and p53 (19253 RFU), which except for p53, are not present on the CT100plus microarray. Together, these results indicate that CRC patients produce autoantibody signals toward proteins not present on the CT100plus microarray, suggesting that the CT100plus microarray content should be expanded to allow for the detection of novel cancer specific antigens. These antigens could potentially be used for cancer diagnosis and prognosis, and even lead to the identification of improved cancer vaccines and T-cell therapies.

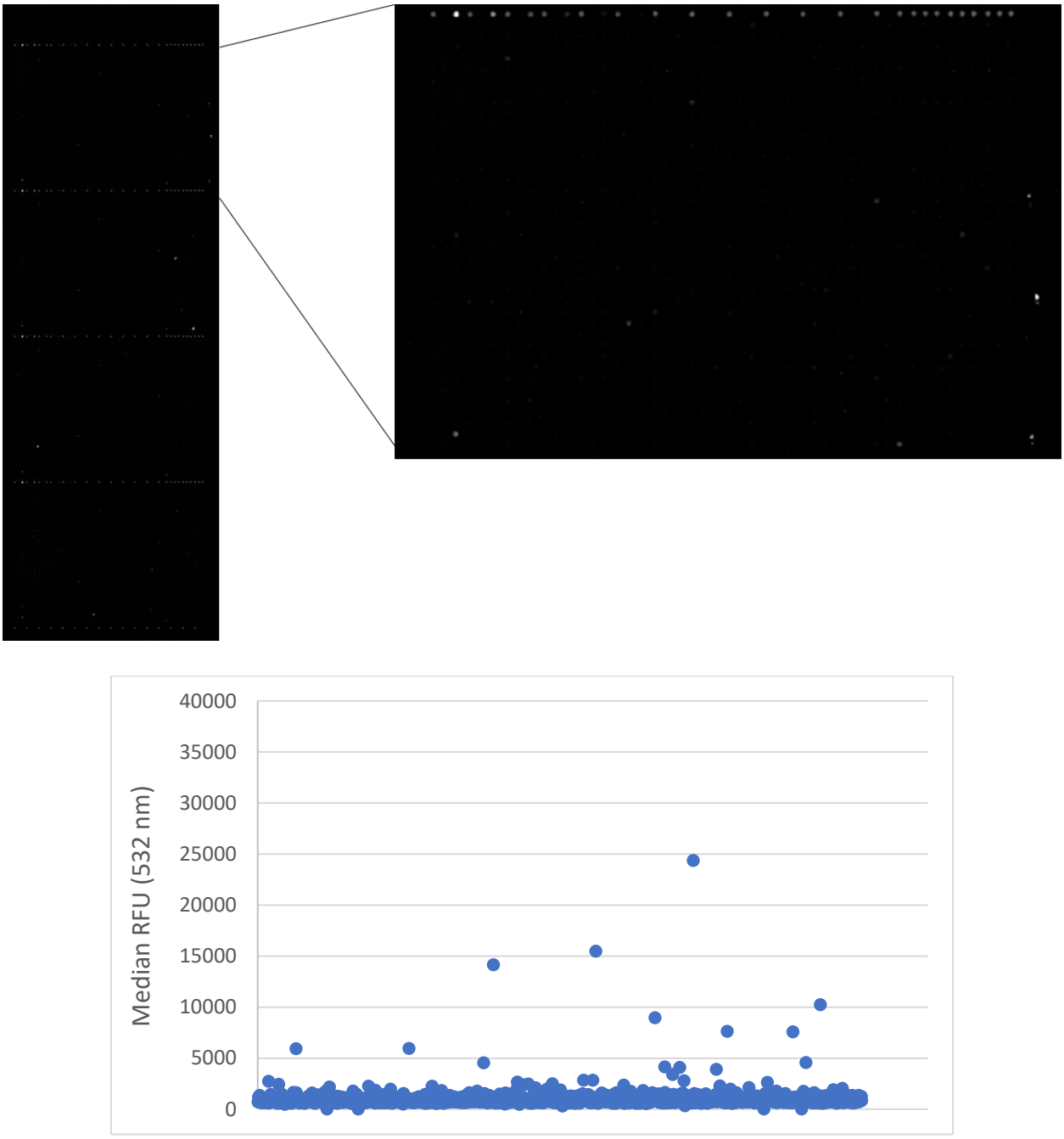


Figure 4.5. Microarray image and scatter plot for CRC038 plasma on the SENGENICS IMMUNOME™ microarray. The SENGENICS IMMUNOME™ microarray was incubated with the blood plasma of **CRC038**, and subsequently incubated with Cy3-conjugated anti-human IgG detection antibody. Fluorescence images were obtained at 532 nm, and numerical data extrapolated. The net intensity for each antigen was obtained, and the data was subsequently filtered using stringent criteria (control CV ≤ 20%; antigen CV ≤ 10%; noise threshold ≥ 2 SD above background). The scatter plot displays the median intensity-values for each antigen. The image represents only a quarter of the microarray, which contains one of the 4 quadruplicate antigens spots. Cy3 = cyanine 3; IgG = immunoglobulin G; nm = nanometres; CV = coefficient of variation; SD = standard deviation.

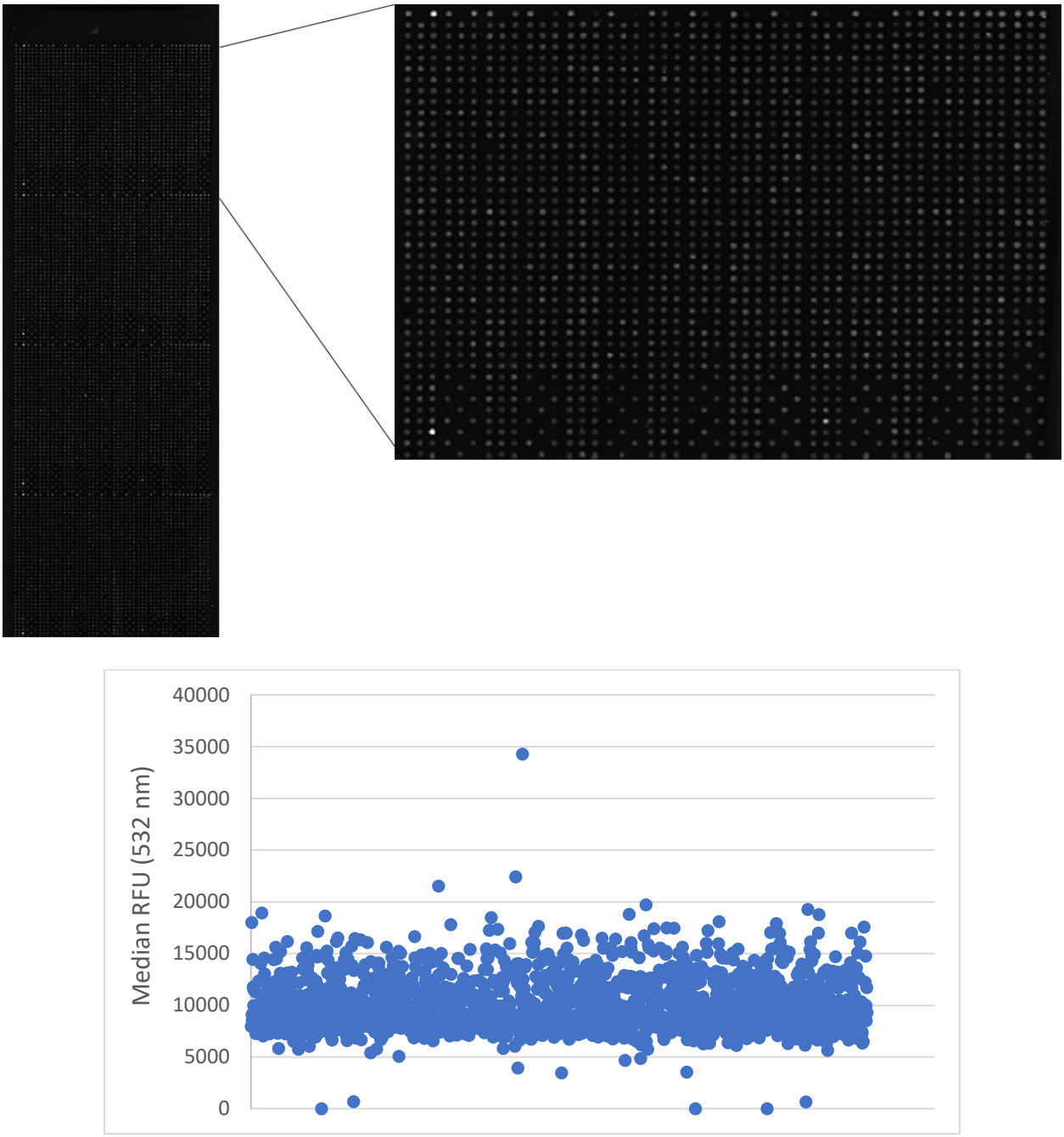


Figure 4.6. Microarray image and scatter plot for CRC050 plasma on the SENGENICS IMMUNOME™ microarray. The SENGENICS IMMUNOME™ microarray was incubated with the blood plasma of **CRC050**, and subsequently incubated with Cy3-conjugated anti-human IgG detection antibody. Fluorescence images were obtained at 532 nm, and numerical data extrapolated. The net intensity for each antigen was obtained, and the data was subsequently filtered using stringent criteria (control CV ≤ 20%; antigen CV ≤ 10%; noise threshold ≥ 2 SD above background). The scatter plot displays the median intensity-values for each antigen. The image represents only a quarter of the microarray, which contains one of the 4 quadruplicate antigens spots. Cy3 = cyanine 3; IgG = immunoglobulin G; nm = nanometres; CV = coefficient of variation; SD = standard deviation.

4.2. Developing and employing an affinity purification-mass spectrometry (AP-MS) assay

An AP-MS assay was developed to identify novel and known cancer antigens associated with CRC, which we hypothesize could serve as potential biomarkers for CRC diagnosis and prognosis. For this assay, IgG from CRC patient plasma was affinity purified using MagReSyn® Protein G magnetic beads, whereas IgA, IgG and IgM were captured using MagReSyn® Protein A magnetic beads. The Protein G beads were used for assay optimizations and AP-MS experiments, whereas Protein A was used in an AP-MS experiment only to compare the results from the two bead-types. To prevent or decrease IgG contamination during the antigen elution step, IgG (and any non-specific proteins from plasma) molecules were cross-linked to Protein A or G using dimethyl pimelimidate dihydrochloride (DMP). The IgG-coated Protein A or G beads were subsequently incubated with autologous native lysate from CRC patients to capture potential antigens, which were eluted under mild conditions and then subjected to mass spectrometric analysis. The assay optimization and results are described below:

The first step in the AP-MS assay was to optimize the correct ratio of CRC patient plasma volume to Protein G magnetic beads. For this assay, 0 microliters (µl), 2.5 µl, 5 µl, 15 µl, 30 µl, 60 µl, 90 µl and 120 µl of **CRC010** plasma was added to Protein G beads, and IgG eluted with 2.5 % acetate. The bound proteins were quantified using the BCA assay (Figure 4.7), and a fraction of the sample was separated on a reducing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel and stained with AQUASTAIN (Figure 4.8).

From Figure 4.7, it is evident that the protein concentration increases linearly from 0 to 15 µl plasma, where after it starts to plateau. This result is supported by the SDS-PAGE gel and densiometric analysis of the Ig light chain (Figure 4.8). Proteins bands are seen at 25 kilodaltons (kDa) and 55 kDa, corresponding to the molecular weight of light and heavy chains, respectively, of IgG (as Protein G reportedly binds only IgG of all Igs) after reducing disulphide bonds - indicating that IgGs were successfully enriched from plasma samples, and that 50 µl Protein G beads was sufficient to capture most of the Ig from 15 µl patient plasma. Protein bands that do not correspond to IgG were also detected, and are evident between 70 and 100 kDa, 130 and 250 kDa, and 35 and 55 kDa. Although these bands appear faint for 2.5 µl plasma, the intensity is

pronounced with an increase in plasma volume, and is likely due to non-specific binding of plasma proteins to the Protein G magnetic beads. Densiometric analysis of the 25 kDa band showed that the protein levels start to plateau at 15 µl plasma. Thus, to extract most Igs, and to minimize non-specific binding, a ratio of 15 µl CRC plasma to 50 µl MagReSyn® Protein A or G magnetic beads was used for subsequent AP-MS assays.

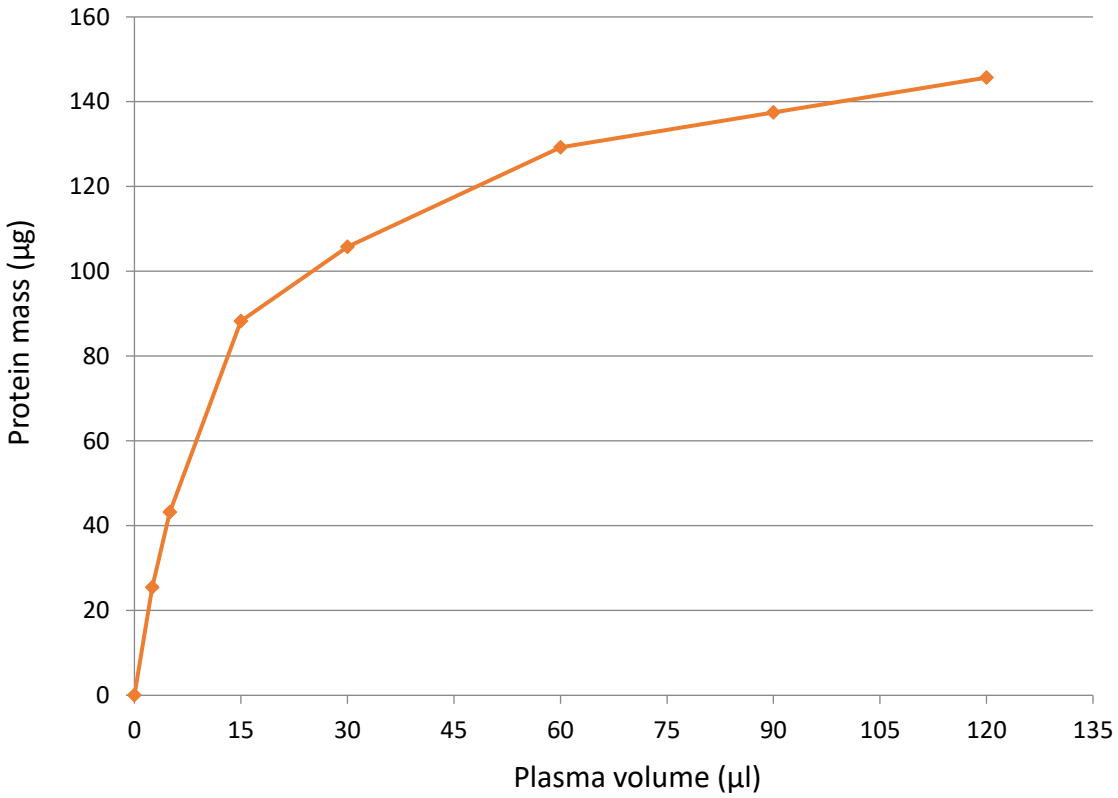


Figure 4.7. CRC patient plasma volume as a function of IgG eluted from Protein G beads. CRC010 patient plasma, at 0 µl, 2.5 µl, 5 µl, 15 µl, 30 µl, 60 µl, 90 µl or 120 µl, was incubated with 50 µl MagReSyn® Protein G magnetic beads to capture IgG. Bound proteins, including IgG and any non-specific proteins, were eluted and the resulting protein mass (µg) is displayed as a function of plasma volume (µl).

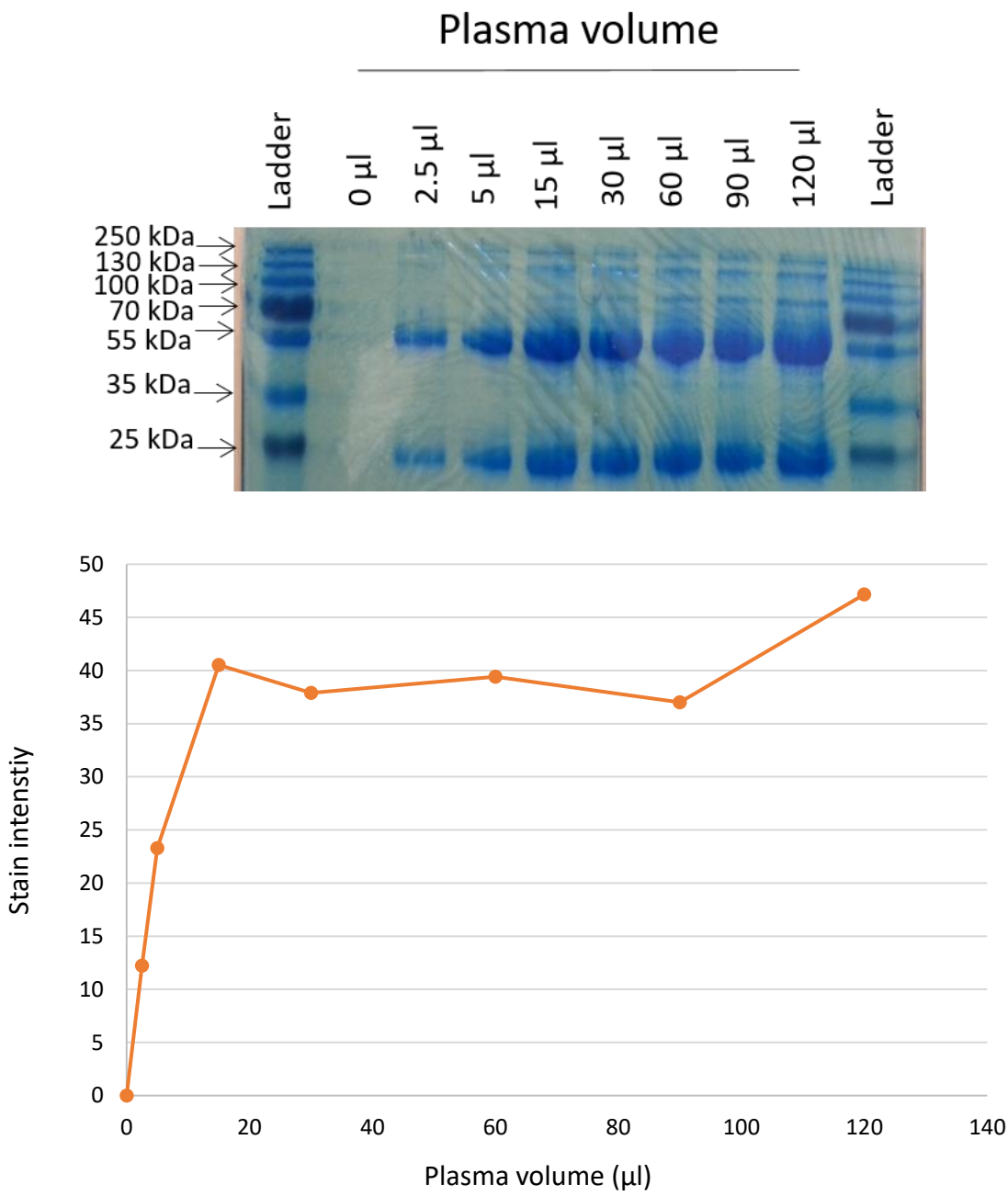


Figure 4.8. Relative quantitation of CRC plasma-derived IgG eluted from MagReSyn® Protein G magnetic beads. CRC010 patient plasma, at 0 µl, 2.5 µl, 5 µl, 15 µl, 30 µl, 60 µl, 90 µl or 120 µl, was incubated with 50 µl MagReSyn® Protein G magnetic beads to capture IgG. Bound proteins were eluted, and separated by SDS-PAGE, and stained with AQUASTAIN. Distinct protein bands formed at 25 kDa and 55 kDa from 2.5 µl plasma, corresponding to the light and heavy chains, respectively, of IgG. Densitometric analysis was performed for the 25 kDa band, and the resulting stain intensity is displayed as a function of plasma volume (µl). Abbreviations include: CRC = colorectal cancer; IgG = immunoglobulin G; SDS-PAGE = sodium dodecyl sulphate polyacrylamide gel electrophoresis.

To determine whether Protein A or G captures more IgG from plasma, and to confirm that the major bands identified at 25 kDa and 55 kDa (Figure 4.8) were indeed Igs, plasma from **CRC038** or **CRC050** was incubated with MagReSyn® Protein A or G magnetic beads, and the eluents from each bead-type were analyzed by Western blot analysis using peroxidase-conjugated anti-human IgG antibody, with subsequent densiometric analysis. Here, **FP** is free protein, and refers to removed unbound plasma protein; **FPW** is free protein in wash, and refers to proteins that were removed during wash steps; **Ab** is antibody, and refers to antibodies (and non-specific binders) removed through acid elution. For **CRC038**, Protein A and G captures 73 % and 78 % of IgG, respectively (Figure 4.9); whereas for **CRC050** Protein A and G captures 58 % and 93 % of IgG**,** respectively (Figure 4.10). This confirms that Protein A and G captured IgG from CRC plasma, and it appears that Protein G captures more IgG that Protein A for the plasma tested.

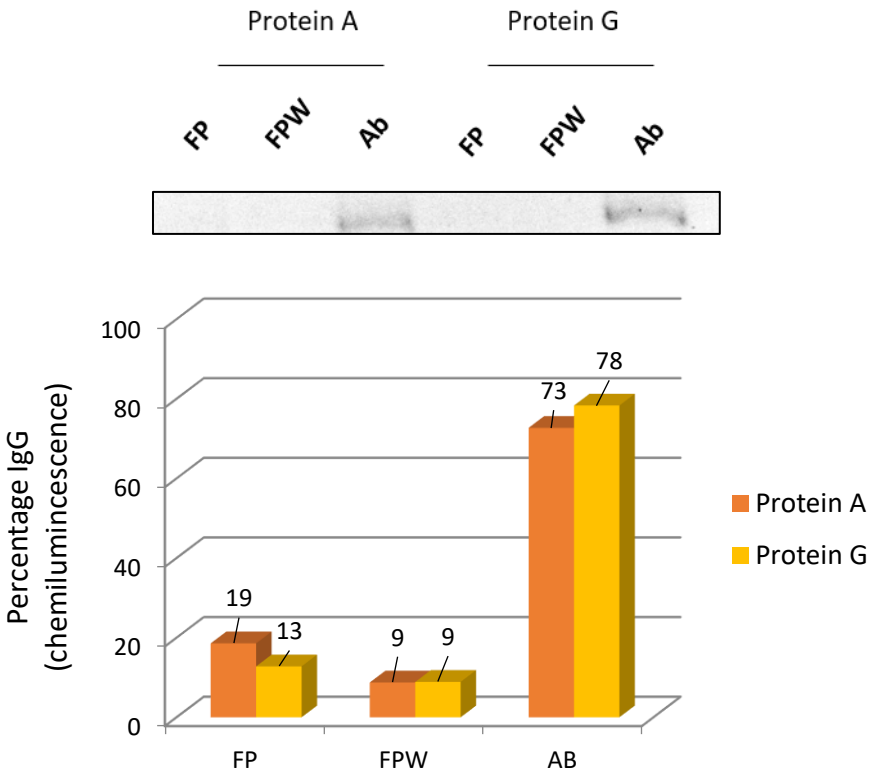


Figure 4.9. Western blot analysis of IP eluents for CRC038. CRC patient plasma, at 0 µl, 2.5 µl, 5 µl, 15 µl, 30 µl, 60 µl, 90 µl or 120 µl, was incubated with 50 µl MagReSyn® Protein G magnetic beads to capture IgG from **CRC038** plasma. Unbound proteins and bound proteins were eluted, and IgG light and heavy chains were detected by Western blot. Densitometric analysis was performed for the 25 kDa band, and the resulting stain intensity is displayed. Abbreviations include: IgG = immunoglobulin G, FP = free protein, FPW = free protein in wash, Ab = antibody.

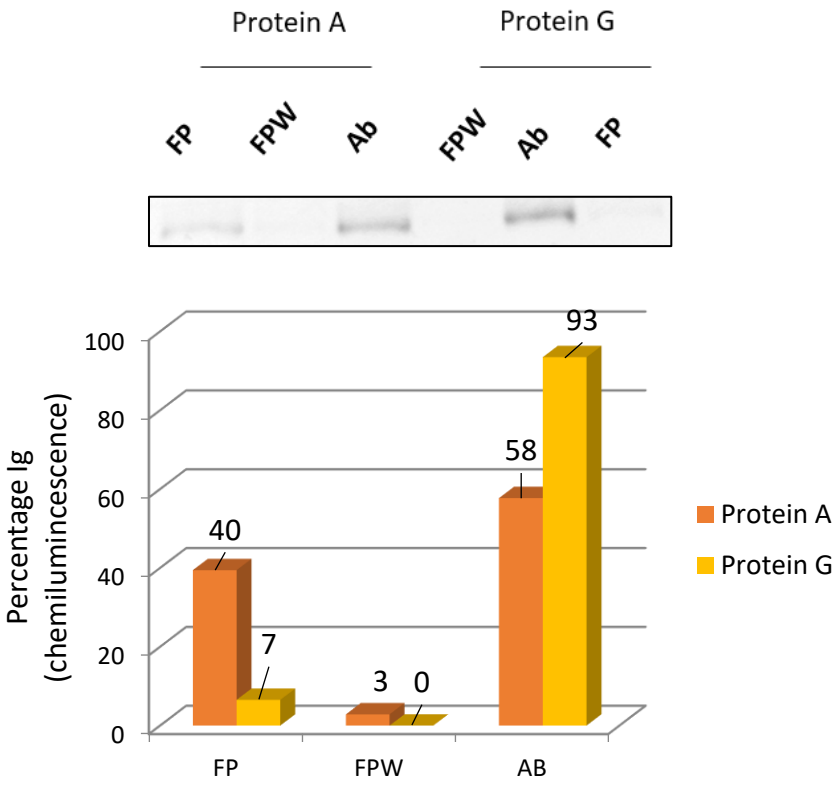


Figure 4.10. Western blot analysis of IP eluents for CRC050. CRC patient plasma, at 0 µl, 2.5 µl, 5 µl, 15 µl, 30 µl, 60 µl, 90 µl or 120 µl, was incubated with 50 µl MagReSyn® Protein G magnetic beads to capture IgG from **CRC050** plasma. Unbound proteins and bound proteins were eluted, and IgG light and heavy chains were detected by Western blot. Densitometric analysis was performed for the 25 kDa band, and the resulting stain intensity is displayed. Abbreviations include: IgG = immunoglobulin G, FP = free protein, FPW = free protein in wash, Ab = antibody.

To minimize IgG contamination during antigen elution, the second aim was to crosslink IgG to Protein G with DMP using an adapted protocol from Holzmann *et al*. (Holzmann et al., 2011). For this assay, 50 µl Protein G beads were incubated with 7.5 µl or 15 µl plasma; the two concentrations were selected to determine whether the IgG load on the beads affected cross-linking. After the wash steps, Protein G-bound proteins were crosslinked with DMP. The beads were subsequently incubated with 2.5% acetate (pH 2.87) to elute proteins that were not cross-linked. As a negative control, Protein G beads were incubated with 15 µl plasma, and after several wash steps, bound proteins were eluted with 2.5% acetate (pH 2.87). The resulting eluents were quantified using the BCA protein quantitation kit and analysed by SDS-PAGE and AQUASTAIN stain. The negative control produced 1519 µg/ml protein, confirming that non-crosslinked

proteins elute under acidic conditions i.e. 2.5% acetate (pH 2.87). When Protein G-bound proteins were cross-linked with DMP, the concentration of eluted protein decreases to 0 (0%) and 17 µg/ml (1%) for 7.5 µl and 15 µl plasma, respectively (Figure 4.11). This result was supported by the SDS-PAGE and AQUASTAIN analysis of eluted proteins: the negative control displayed protein bands at ~25 kDa and ~55 kDa, corresponding to the molecular weight of the light and heavy chain of IgG, respectively. The DMP-treated samples did not appear to have any protein bands seen in the negative control (Figure 4.12). Together, these results indicate that IgG and any other non-specific proteins from plasma are crosslinked to Protein G, decreasing the amounts of non-specific proteins to be analysed on the mass spectrometer.

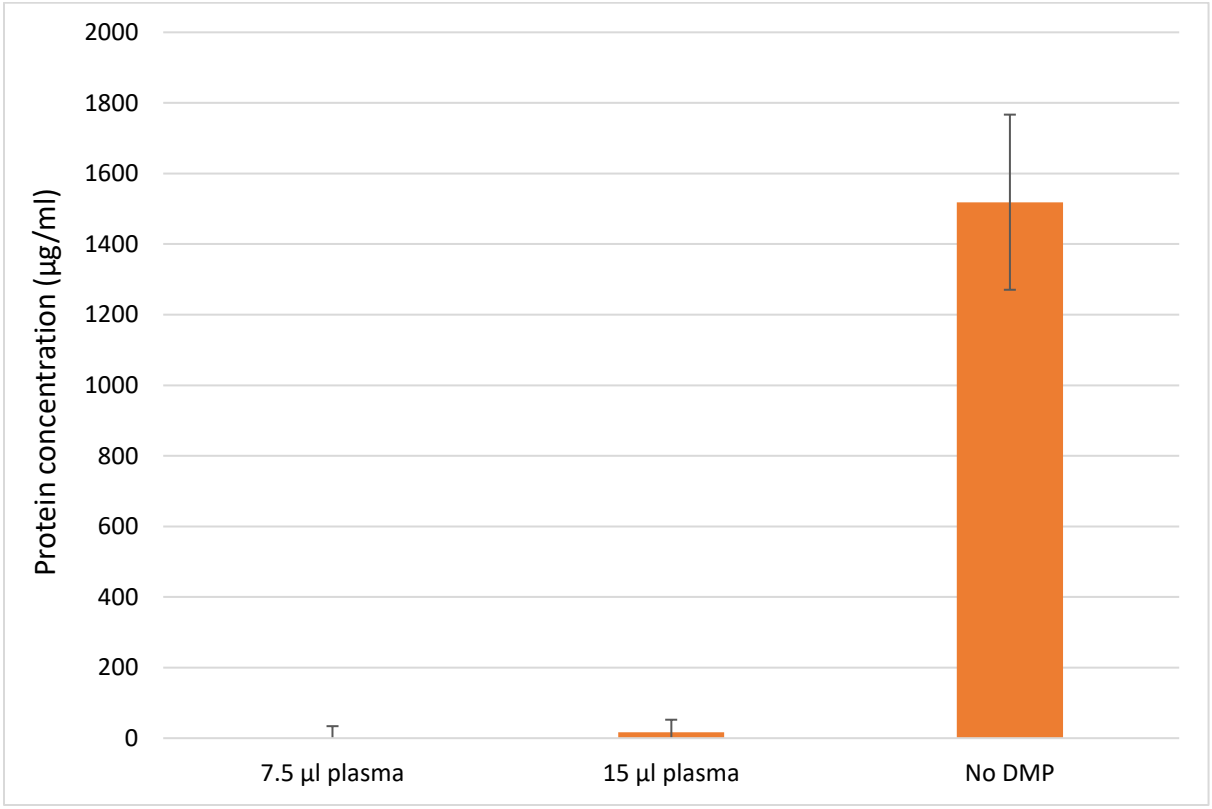


Figure 4.11. Protein quantitation of eluents after crosslinking IgG to MagReSyn® Protein G magnetic beads with DMP. IgG was crosslinked to Protein G beads with DMP to prevent, or decrease, IgG contamination when eluting potential cancer antigens. Here, 7.5 µl or 15 µl plasma was incubated with 50 µl Protein G magnetic beads, and cross-linked using DMP. Beads incubated with 15 µl plasma, with no crosslinking were used as a negative control. All proteins that were not crosslinked were eluted and quantified using the BCA protein quantification kit. Abbreviations include: IgG = immunoglobulin G; DMP = dimethyl pimelimidate dihydrochloride; BCA = bicinchoninic acid.

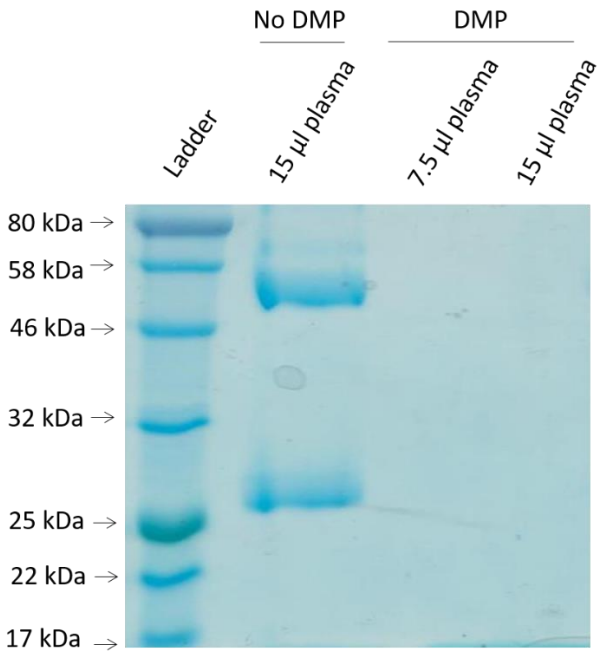


Figure 4.12. SDS-PAGE and AQUASTAIN of eluents after crosslinking IgG to MagReSyn® Protein G magnetic beads with DMP. IgG and Protein G were crosslinked with DMP to prevent, or reduce, IgG contamination when eluting potential cancer antigens. Here, 7.5 µl or 15 µl plasma was incubated with 50 µl Protein G magnetic beads, and cross-linked using DMP. Beads incubated with 15 µl plasma, with no crosslinking was used as a negative control. All proteins that was not cross-linked was eluted and separated on an SDS-PAGE gel and stain with AQUASTAIN. Abbreviations include: SDS-PAGE = sodium dodecyl sulphate polyacrylamide gel electrophoresis; IgG = immunoglobulin G; DMP = dimethyl pimelimidate dihydrochloride.

The next step was to incubate the IgG-bound beads with autologous tissue lysate to capture and enrich for potential cancer antigens. Therefore, the third aim was to produce native protein lysates from CRC tissues. Since we were limited with the amount of fresh frozen tissues available for assay optimization, SHSY5Y neuroblastoma cells were used to test the efficacy of the buffer with the various detergents. The buffer system contained 20 mM 4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES) (pH 7.9), 1 mM $MgCl_2$, 5 mM EDTA, 1 unit per microliter (U/µl) benzonase, 1× protease inhibitor, and a detergent. Detergents tested included 1% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 1% IGEPAL® CA-630, 1% Triton™ X-100, a mixture of detergents (0.25% CHAPS, 0.5% IGEPAL® CA 630, and 0.5% Triton™ X-100), or a negative control without detergent. Three lysis conditions were also assessed for protein yield: *lysis buffer only* refers to lysates generated by incubating with lysis buffer and on

ice with intermittent vortexing only; *lysis buffer & PE* refers to lysates generated by incubating with lysis buffer and homogenizing using a Potter-Elvehjem homogenizer, and then incubated on ice with intermittent vortexing; *lysis buffer & shaking* refers to lysates generated by incubating with lysis buffer and homogenizing with a Potter-Elvehjem homogenizer, and then continuous vortexing at 4 degree Celsius (°C). From Figure 4.13 it is evident for the *lysis buffer only* samples that the buffer containing Triton™ X-100 produced the highest protein yield, followed by "mix", IGEPAL® then CHAPS, with no protein detected in the buffer without detergent. However, for the *lysis buffer & PE* samples, the protein concentration increased for all buffers, excluding CHAPS. Here, again the buffer containing Triton™ X-100 produced the highest concentration of protein, followed by IGEPAL®, "mix", CHAPS, and lowest concentration of protein was detected for the buffer without detergent. The condition producing the highest concentration was *lysis buffer & shaking*, which resulted in similar protein levels for all detergent-types used, except for the buffer without detergent which produced a significantly lower protein concentration. Since the highest protein concentration was produced using Triton™ X-100, the buffer containing Triton™ X-100 was used for subsequent assays. Since *lysis buffer & shaking* produced the highest concentration of protein, this method was used for CRC tissue lysis.

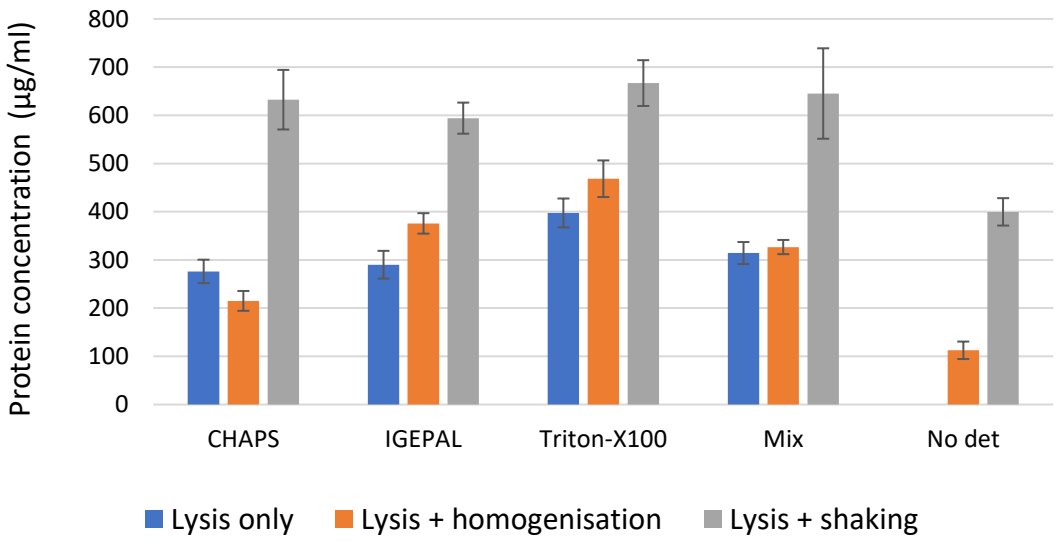


Figure 4.13. Protein quantitation of SHSY5Y cell-line lysates. SH-SY5Y human neuroblastoma cells were lysed with non-denaturing lysis buffers containing different detergents, including 1% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 1% IGEPAL® CA-630, 1% Triton™ X-100, a mixture of detergents (0.25% CHAPS, 0.5% IGEPAL® CA 630, and 0.5% Triton™ X-100), and a negative control with no detergent (No det). Various lysis conditions were also assessed for protein yield: lysis only refers to lysates generated by incubating with lysis buffers and incubated on ice with intermittent vortexing only; lysis + homogenization refers to lysates generated by incubating with lysis buffers and homogenizing using a Potter-Elvehjem homogenizer, and then incubated on ice with intermittent vortexing; lysis + shaking refers to lysates generated by incubating with lysis buffers and homogenizing with a Potter-Elvehjem homogenizer, and then continuous vortexing at 4 °C.

After the Protein A or G beads were incubated with CRC patient plasma and autologous tissue lysate, and the beads were thoroughly washed to remove proteins that bind non-specifically, the antigens and any remaining non-specifically bound proteins were eluted. Here, proteins were eluted using on-bead tryptic digestion, a method adapted from Turriziani *et al*. (Turriziani et al., 2014), to elute tryptic fragments that were further used for mass spectrometry analysis.

Next, we determined the number of proteins identified from an AP-MS experiment using Protein A or Protein G magnetic bead-types. The AP-MS experiment was performed with either Protein A or G magnetic beads, using the blood plasma, and native cancer and tissue lysates, from **CRC004**. Here, *Cancer* refers to antigens and non-specific binders eluted from AP-MS assays using blood plasma and native cancer tissue lysate. The negative controls included *Plasma*, which refers to proteins from AP-MS assay with **CRC004** plasma only; and *Normal*, which refers to proteins

from an AP-MS assay with **CRC004** plasma and paired normal tissue. The raw files were processed through MaxQuant (version 1.5.3.12) using intensity-based absolute quantification (iBAQ) (Schwanhüusser et al., 2011) and match between runs, with a minimum of two unique peptides. The iBAQ quantitation values were median normalised. Cancer and normal tissues were collected for the CRC patients. The normal mucosa was excised from tissues flanking the cancer tissue, and it is therefore possible that normal tissue may have been in the process of transforming to cancer tissue. Thus, for the analysis, we identified cancer-derived proteins that were present in *Cancer* only samples (*i.e. Cancer*-unique), to identify potential cancer antigens unique to cancer tissues, or proteins detected at ≥ 2-fold iBAQ in *Cancer* than *Normal* (*i.e. Cancer*-enriched), to identify potential cancer antigens that may also be present in transformed normal mucosa. For the analysis, *Cancer*-unique and *Cancer*-enriched protein groups identified using Protein A or G bead-types were compared to determine whether common protein groups were enriched using the two bead-types. The data analysis pipeline is outlined in Figure 4.14.

The iBAQ quantitation values were median normalised and is presented as a boxplot in Figure 4.15. Protein A magnetic beads captured 994, 898 and 117 proteins for *Cancer*, *Normal* and *Plasma*, respectively; the number of shared proteins is indicated in Figure 4.16, whereas Protein G magnetic beads captured 945, 812 and 499 proteins for *Cancer*, *Normal* and *Plasma*, respectively; the number of shared proteins is indicated in Figure 4.17. A total of 317 *Cancer*-enriched proteins were identified from Protein A and G, of which70 (22%) proteins were shared between the bead-types, whereas 170 (54%) and 77 (24%) were unique to Protein A or G, respectively (Figure 4.18). A total of 268 *Cancer*-unique proteins were identified from Protein A and G, of which 80 (30%) proteins were common between the bead-types, whereas 75 (28%) and 113 (42%) were unique to Protein A or G, respectively (Figure 4.19). Together, these results indicate that peptides and proteins were identified from plasma and tissue lysates using Protein A or G magnetic beads. Both bead-types identified more proteins for *Cancer* versus *Normal*, with the lowest number of proteins identified for *Plasma*. Furthermore, 70 and 80 proteins were commonly identified as enriched or unique, respectively, to *Cancer* for Protein A and G. It would be ideal to compare the AP-MS results to the CT100plus microarray results, and since an IgG autoantibody response was detected on the CT100plus microarray, and that Protein G binds all

IgG subtypes but Protein A binds weakly to a subgroup of IgG subtypes, we continued to use Protein G magnetic beads for further assays.

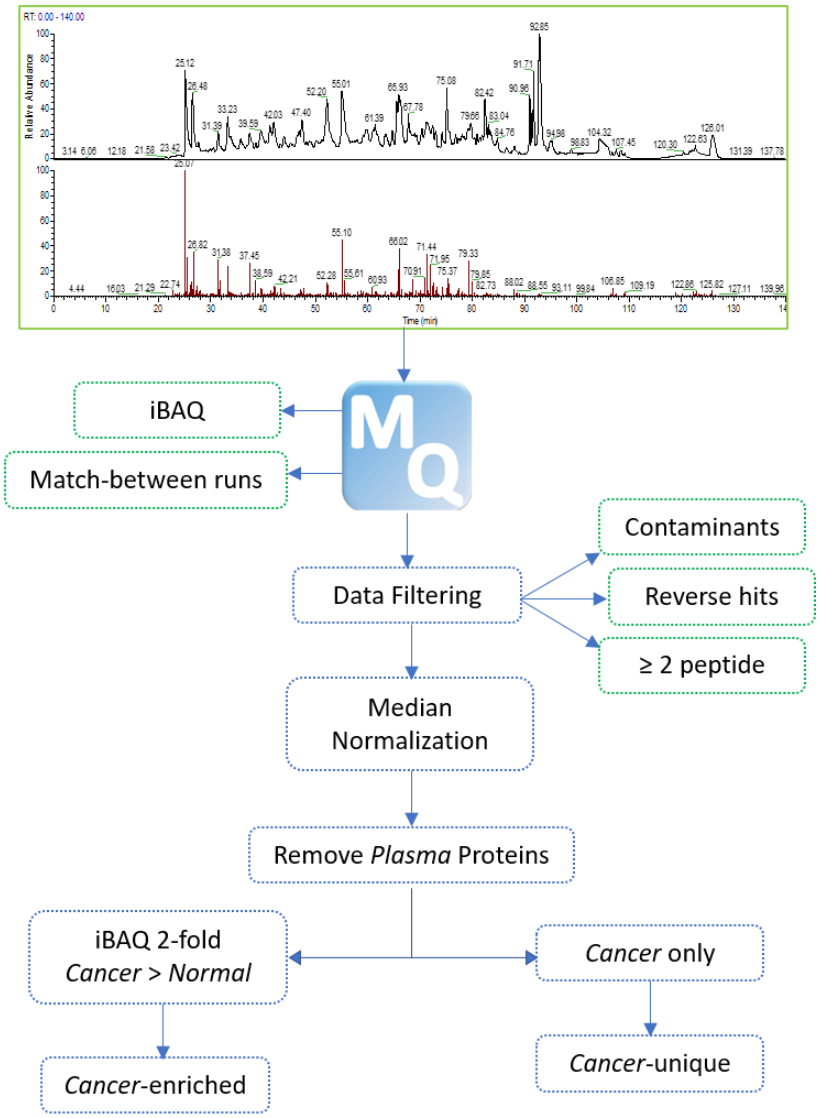


Figure 4.14. AP-MS data analysis pipeline. *Cancer* (AP-MS with plasma and CRC tissue), *Normal* (AP-MS with plasma and normal mucosa) and *Plasma* (AP-MS with plasma only) samples of **CRC004** were processed using Protein A or Protein G beads, and analysed using the Q Exactive mass spectrometer. The raw files were processed through MaxQuant (version 1.5.3.12) using intensity-based absolute quantification (iBAQ) and match between runs. Subsequent data filtering included the removal of reverse hits, contaminants and protein groups with < 2 peptides, and the data was normalised using median iBAQ values. *Plasma* proteins were removed from *Cancer* and *Normal* for Protein A and Protein G samples. *Cancer*-enriched, defined as proteins detected at ≥ 2-fold iBAQ in *Cancer* than *Normal*, and *Cancer*-unique, defined as cancer-derived proteins that were present in *Cancer* only, were subsequently identified. Abbreviations include: MQ = MaxQuant.

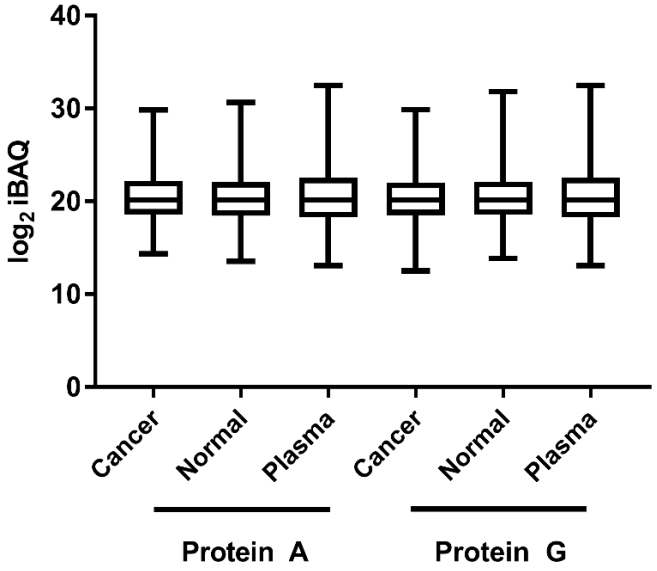


Figure 4.15. Box plot of median-based iBAQ normalisation. *Cancer* (AP-MS with plasma and CRC tissue), *Normal* (AP-MS with plasma and normal mucosa) and *Plasma* (AP-MS with plasma only) samples of **CRC004** were processed using Protein A or Protein G beads, and analysed using the Q Exactive mass spectrometer. The raw files were processed through MaxQuant (version 1.5.3.12) using iBAQ and match between runs. Subsequent data filtering included the removal of reverse hits, contaminants and protein groups with < 2 peptides, and the data was normalised using median iBAQ values. Abbreviations include: AP-MS = affinity purification-mass spectrometry; iBAQ = intensity-based absolute quantification.

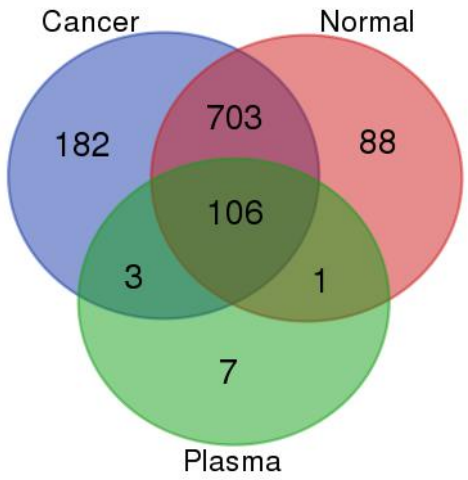


Figure 4.16. Number of proteins identified for AP-MS with MagReSyn® Protein A magnetic beads for CRC004. For the AP-MS assay, Igs from **CRC004** were captured using Protein A magnetic beads, and proteins were subsequently captured from **CRC004** cancer and paired normal native lysate. The captured proteins were eluted, analysed by mass spectrometry. The negative controls included *Plasma*, which refers to proteins from Protein A magnetic beads incubated with **CRC004** plasma; and *Normal*, which refers to proteins from Protein A magnetic beads incubated with **CRC004** plasma and paired normal tissue. Abbreviations include: AP-MS = affinity purification-mass spectrometry; Igs = immunoglobulins.

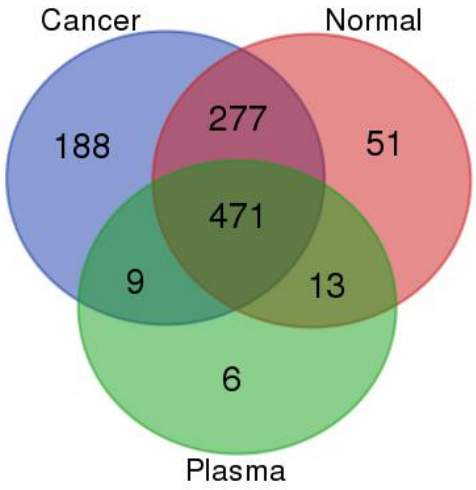


Figure 4.17. Number of proteins identified for AP-MS with MagReSyn® Protein G magnetic beads for CRC004. For the AP-MS assay, Igs from **CRC004** were captured using Protein G magnetic beads, proteins were subsequently captured from **CRC004** cancer and paired normal native lysate. The captured proteins were eluted, analysed by mass spectrometry. The negative controls included *Plasma*, which refers to proteins from Protein G magnetic beads incubated with **CRC004** plasma; and *Normal*, which refers to proteins from Protein G magnetic beads incubated with **CRC004** plasma and paired normal tissue. Abbreviations include: AP-MS = affinity purification-mass spectrometry; Igs = immunoglobulins.

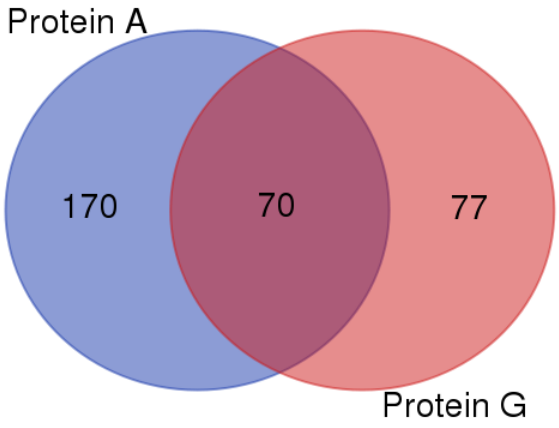


Figure 4.18. *Cancer*-enriched proteins shared proteins between Protein A and G. For the AP-MS assay, Igs from **CRC004** were captured using Protein A or G magnetic beads, and proteins were subsequently captured from **CRC004** cancer and paired normal native lysate. Proteins were identified using MaxQuant software (version 1.5.3.12). Proteins identified in the *Plasma* data were removed from *Cancer* and *Normal* data, and proteins with higher signal in *Cancer* versus *Normal* (i.e. *Cancer*-enriched) were identified, displayed in the Venn diagram. Abbreviations include: AP-MS = affinity purification-mass spectrometry; Igs = immunoglobulins.

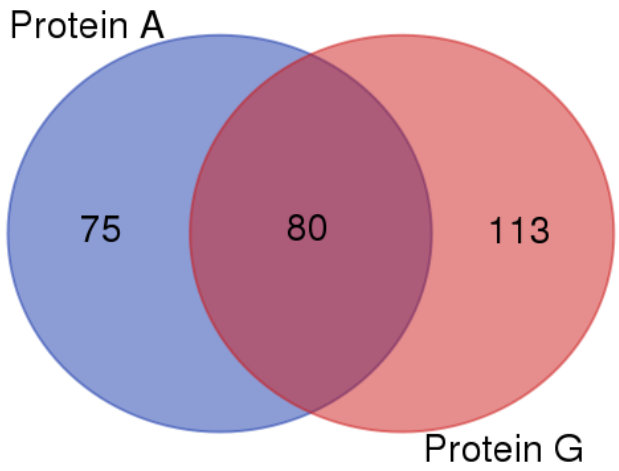


Figure 4.19. *Cancer*-unique shared proteins between Protein A and G. For the AP-MS assay, Igs from **CRC004** were captured using Protein A or G magnetic beads, and proteins were subsequently captured from **CRC004** cancer and paired normal native lysate. Proteins were identified using MaxQuant software (version 1.5.3.12). Proteins identified in the *Plasma* and *Normal* data were removed from *Cancer* data, and proteins specific to *Cancer* (i.e. *Cancer*-unique) were identified, and displayed in the Venn diagram. Abbreviations include: AP-MS = affinity purification-mass spectrometry; Igs = immunoglobulins.

Cancer-enriched and *Cancer*-unique proteins from AP-MS experiments with Protein A and Protein G were analysed using the functional association network software search tool for the retrieval of interacting genes/proteins (STRING) (version 10.5) (https://string-db.org/). The settings included active interaction sources from textmining, experiments, databases and co-expression, and a medium (0.4) minimum interaction score. The edges indicate evidence of interaction, whereas nodes contain query proteins. Although some proteins do not interact physically with other proteins in the network, the possibility of the co-expression or predicted linkage is plausible. The network generated from STRING indicates multiple links between input proteins, and ACLY is a central protein linked to ABCC3, ADH1B, AGPS, CSK, EFTUD1, MCCC2, PCK2, PHGDM, PRDX4, RDH13 and UGDH (Figure 4.20).

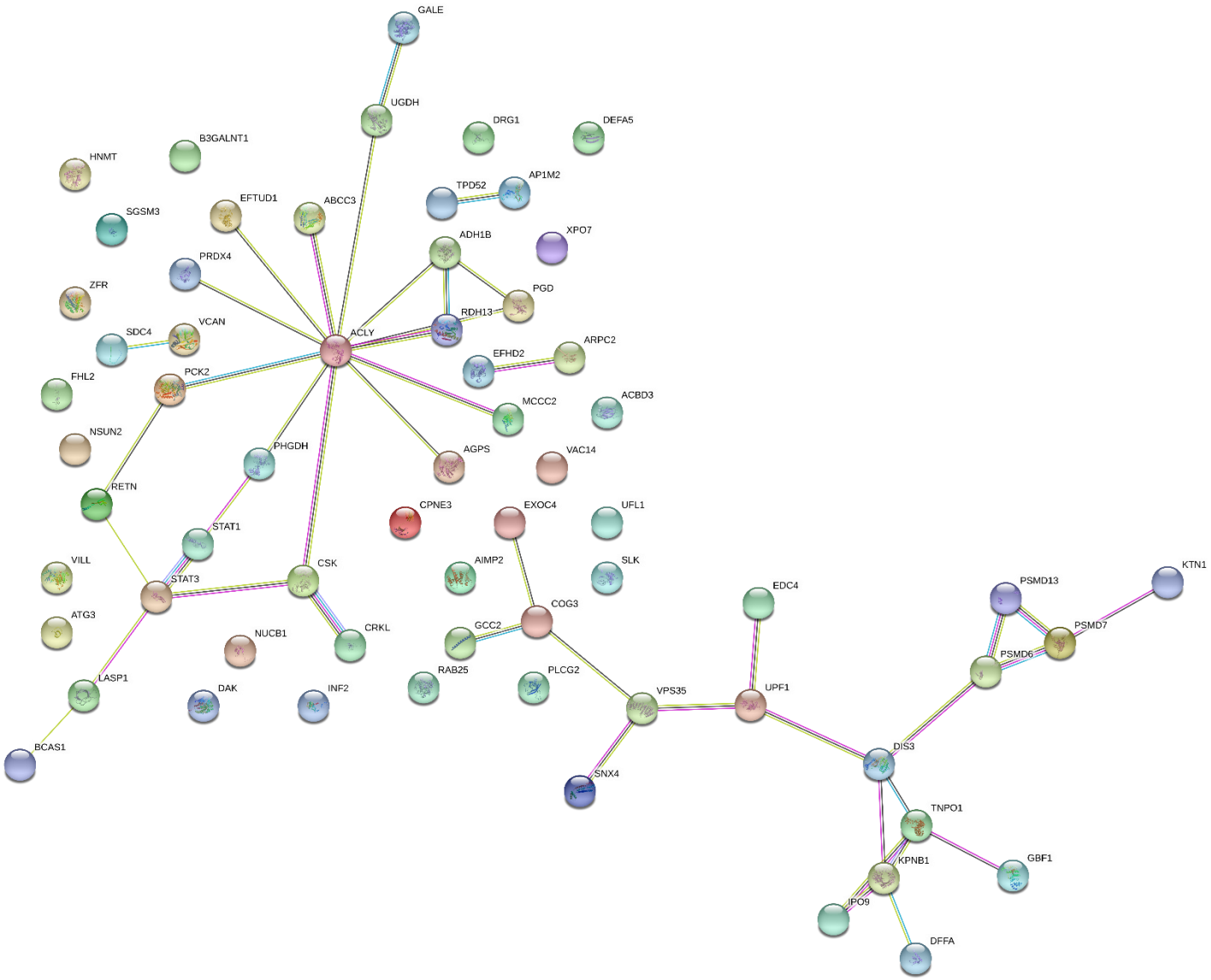


Figure 4.20. String analysis of Cancer proteins common for Protein A and Protein G. *Cancer*-enriched and *Cancer*-unique proteins were identified for AP-MS samples with Protein A and Protein G and analysed using STRING (https://string-db.org).

We next performed the AP-MS assay using Protein G magnetic beads for 4 additional CRC patient samples for whom we had the highest amount of tissue specimen available. The patients included **CRC002**, **CRC017**, **CRC021** and **CRC031**, for which the *Cancer*, *Normal* and *Plasma* protein elutions were analysed by mass spectrometry. Thus, a total of 5 patient samples were analysed by AP-MS. Of the 5 patients tested, **CRC017**, **CRC021** and **CRC031** had polyethylene glycol (PEG) detected in the mass spectrum. PEG is a charged molecule derived from detergents, e.g. Triton™ X-100, Tween® 20 or IGEPAL® CA-630, used in non-denaturing buffers like the buffer used to produce native lysate from CRC tissues. Since it is charged, PEG is detected in samples analysed by mass spectrometry. At high levels, PEG suppresses the signal intensity from peptides-derived ions, resulting in lower levels, or no peptide, detected. Thus, patients whose spectra contained high levels of PEG were not used for quantitative analysis, although, these were used to identify proteins detected on the CT100plus microarray and to assess citrullination.

Potential cancer antigens identified by AP-MS and CT100plus microarray:

We next compared the proteins identified in the AP-MS experiment and CT100plus microarray. For this comparison, CT100plus microarray antigen signal intensities greater than or equal to cut-off-value (Figure 2.10) and present in the AP-MS results, without any data filtering were identified. We also identified homologues of significant CT100plus antigens for **CRC002** and **CRC004**. CT100plus antigens and antigen homologues identified by AP-MS for **CRC002** are summarised in Table 4.1 and Table 4.2, respectively, whereas CT100plus antigens and antigen homologues identified by AP-MS for **CRC004** are summarised in Table 4.3 and Table 4.4. CT100plus antigens identified by AP-MS for **CRC017**, **CRC021** and **CRC031** are summarised in Table 4.5, Table 4.6 and Table 4.7, respectively. Since different CRC patients were assayed on the SENGENICS IMMUNOME™ microarray and the AP-MS assays, comparisons could not be made between identified proteins.

Table 4.1. Proteins matched between CT100plus microarray and AP-MS for CRC002. For **CRC002**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification-mass spectrometry.

Protein	Cancer	Normal	Plasma
CALM1	Yes	Yes	No
COL6A1	Yes	Yes	No
CTNNB1	Yes	Yes	No
ITGB1	Yes	Yes	No
LDHC	Yes	Yes	Yes
MAPK1	Yes	Yes	No
MAPK3	Yes	Yes	No
SPAG9	Yes	Yes	No
THEG	Yes	Yes	No

Table 4.2. Proteins homologues matched between CT100plus microarray and AP-MS for CRC002. For **CRC002**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification mass-spectrometry, N/A = not available.

Protein	Homologue	Cancer	Normal	Plasma	Query cover	E-value	Identity
CEACAM1	**CEACAM7**	Yes	Yes	No	84%	2.00e-120	70%
COX6B2	**COX6B1**	Yes	Yes	No	96%	2.00e-35	54%
MAGEA1	**MAGED2**	Yes	No	No	30%	1.00e-48	41%
MAGEA10	**MAGED2**	Yes	No	No	32%	1.00e-55	40%
MAGEA11	**MAGED2**	Yes	No	No	38%	1.00e-54	41%
MAGEA2	**MAGED2**	Yes	No	No	35%	3.00e-36	38%
MAGEA3	**MAGED2**	Yes	No	No	38%	7.00e-45	40%
MAGEA4v2	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEA4v3	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEA4v4	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEA5	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEB1	**MAGED2**	Yes	No	No	33%	2.00E-55	43%
MAGEB5	**MAGED2**	Yes	No	No	32%	3.00E-46	40%
MAGEB6	**MAGED2**	Yes	No	No	32%	1.00E-52	42%
PRKCZ	**PRKCD**	Yes	No	No	79%	3.00E+113	40%
ROPN1	**ROPN1B**	Yes	Yes	No	100%	4.00E-157	96%
ROPN1A	**ROPN1B**	Yes	Yes	No	100%	2e-120	96%

Table 4.3. Proteins matched between CT100plus microarray and AP-MS for CRC004. For **CRC004**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification-mass spectrometry.

Protein	Cancer	Normal	Plasma
CALM1	Yes	Yes	Yes
COL6A1	Yes	Yes	Yes
CTNNB1	Yes	Yes	Yes
ITGB1	Yes	Yes	Yes
LDHC	Yes	No	No
MAPK1	Yes	Yes	Yes
MAPK3	Yes	Yes	No
SPAG9	Yes	No	No
THEG	Yes	Yes	Yes

Table 4.4. Proteins homologues matched between CT100plus microarray and AP-MS for CRC004. For **CRC004**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification-mass spectrometry, N/A = not available.

Protein	Homologue	Cancer	Normal	Plasma	Query cover	E-value	Identity
CEACAM 1	**CEACAM7**	Yes	Yes	Yes	84%	2.00e-120	70%
COX6B2	**COX6B1**	Yes	Yes	Yes	96%	2.00e-35	54%
MAGEA1	**MAGED2**	Yes	No	No	30%	1.00e-48	41%
MAGEA10	**MAGED2**	Yes	No	No	32%	1.00e-55	40%
MAGEA11	**MAGED2**	Yes	No	No	38%	1.00e-54	41%
MAGEA2	**MAGED2**	Yes	No	No	35%	3.00e-36	38%
MAGEA4v2	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEA4v3	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEA4v4	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEA5	**MAGED2**	Yes	No	No	N/A	N/A	N/A
MAGEB1	**MAGED2**	Yes	No	No	33%	2.00e-55	43%
MAGEB5	**MAGED2**	Yes	No	No	32%	3.00e-46	40%
MAGEB6	**MAGED2**	Yes	No	No	32%	1.00e-52	42%
ROPN1	**ROPNB1**	Yes	Yes	No	100%	4.00e-157	96%
ROPN1A	**ROPNB1**	Yes	Yes	No	100%	2e-120	96%

Table 4.5. Proteins matched between CT100plus microarray and AP-MS for CRC017. For **CRC017**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification-mass spectrometry.

Protein	Cancer	Normal	Plasma
CALM1	Yes	Yes	No
MAPK3	Yes	No	No
THEG	Yes	Yes	Yes

Table 4.6. Proteins matched between CT100plus microarray and AP-MS for CRC021. For **CRC021**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification-mass spectrometry.

Protein	Cancer	Normal	Plasma
CALM1	No	Yes	No
COL6A1	Yes	Yes	Yes
CTNNB1	No	Yes	No
ITGB1	No	Yes	No
LDHC	Yes	Yes	No
MAPK3	No	Yes	No
THEG	No	Yes	No

Table 4.7. Proteins matched between CT100plus microarray and AP-MS for CRC031. For **CRC031**, significant (> cut-off) cancer antigens on CT100plus microarray were searched against same patient AP-MS results. Abbreviations include: AP-MS = affinity purification-mass spectrometry.

Protein	Cancer	Normal	Plasma
COL6A1	Yes	No	No
CTNNB1	Yes	No	No
THEG	Yes	No	Yes

Potential cancer antigens identified by AP-MS:

The next step was to identify autoantigens from the AP-MS data that were not present on the CT100plus or SENGENICS IMMUNOME™ microarrays. Here, autoantigens specific to *Cancer*-unique or *Cancer*-enriched proteins were matched to the AAgAtlas human autoantigen database (version 1.0) with "cancer" as the disease-term. The gene symbols in the AAgAtlas human autoantigen database (version 1.0) matched to 2981 UniProt accession numbers, which was used for further analysis. Due to PEG contamination in **CRC017**, **CRC021** and **CRC031**, only **CRC002** and **CRC004** were used for further analysis.

The data analysis pipeline is described in Figure 4.14. Briefly, the raw files were processed through MaxQuant (version 1.5.3.12) using iBAQ (Schwanhüusser et al., 2011) and 'match between runs', with a minimum of two unique peptides. After the filtering was applied, for **CRC002** we identified 1440, 1420 and 303 proteins for *Cancer*, *Normal* and *Plasma*, respectively; whereas for CRC004 we identified 1307, 1152 and 883 proteins for *Cancer*, *Normal* and *Plasma*, respectively. The iBAQ quantitation values were median normalised, presented as a boxplot in Figure 4.21.

Cancer and normal tissues were collected for the CRC patients. The normal mucosa was excised from tissues flanking the cancer tissue, and it is therefore possible that normal tissue may have been in the process of transforming to cancer tissue. Thus, for the analysis, we identified cancer-derived proteins that were present in *Cancer* only samples (*i.e. Cancer*-unique), to identify potential cancer antigens unique to cancer tissues; or proteins detected at ≥ 2-fold iBAQ in *Cancer* than *Normal* (*i.e. Cancer*-enriched), to identify potential cancer antigens that may also be present in transformed normal mucosa.

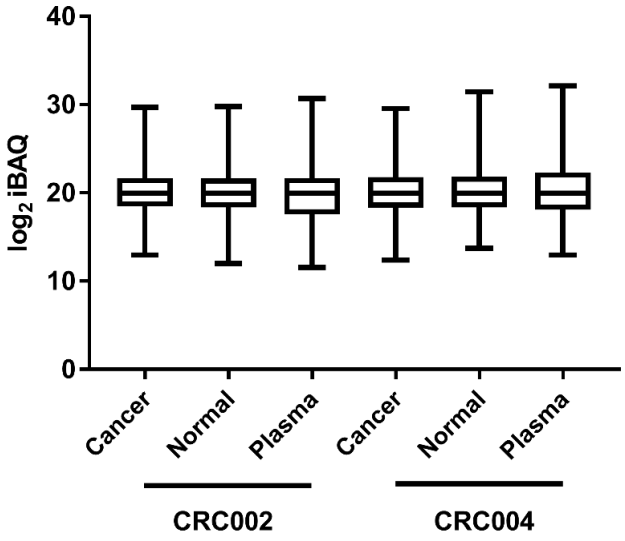


Figure 4.21. Box plot of median-based iBAQ normalisation. *Cancer* (AP-MS with plasma and CRC tissue), *Normal* (AP-MS with plasma and normal mucosa) and *Plasma* (AP-MS with plasma only) samples of **CRC002** and **CRC004** were processed by AP-MS. The raw files were processed through MaxQuant (version 1.5.3.12) using iBAQ and match between runs. Subsequent data filtering included the removal of reverse hits, contaminants and protein groups with < 2 peptides, and the data was normalised using median iBAQ-values. Abbreviations include: CRC = colorectal cancer; iBAQ = intensity-based absolute quantification; AP-MS = affinity purification-mass spectrometry.

Cancer-unique and *Cancer*-enriched proteins were identified for **CRC002** and **CRC004**. A total of 68 and 193 *Cancer*-unique proteins were identified for **CRC002** and **CRC004**, respectively. There were no protein shared between **CRC002**, **CRC004** and AAgAtlas, although 1 and 8 proteins matched to AAgAtlas for **CRC002** and **CRC004**, respectively (Figure 4.22). Furthermore, 17 *Cancer*-unique proteins were shared between **CRC002** and **CRC004** only. *Cancer*-enriched proteins were

also identified for **CRC002** and **CRC004**. A total of 98 and 146 *Cancer*-unique proteins were identified for **CRC002** and **CRC004**, respectively. Here, there was no protein shared between **CRC002**, **CRC004** and AAgAtlas, although 1 and 1 proteins matched to AAgAtlas for **CRC002** and **CRC004**, respectively. Furthermore, 1 *Cancer*-unique protein was shared between **CRC002** and **CRC004** (Figure 4.23).

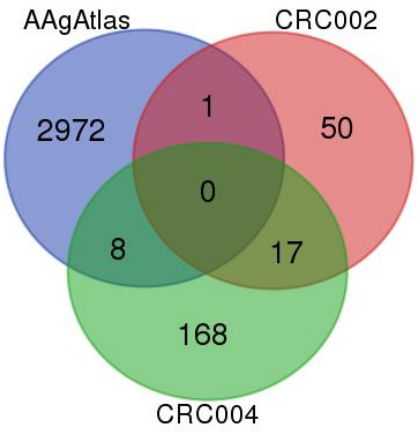


Figure 4.22. *Cancer*-unique proteins matched to the AAgAtlas autoantigen database. *Cancer*-unique proteins, defined as proteins characterised with at least 2 unique peptides specific to the *Cancer* sample, for **CRC002** and **CRC004** were matched it to the AAgAtlas human autoantigen database (version 1.0). The Venn diagram represents the number of proteins that are unique to, or overlap between, *Cancer*-unique and the AAgAtlas proteins.

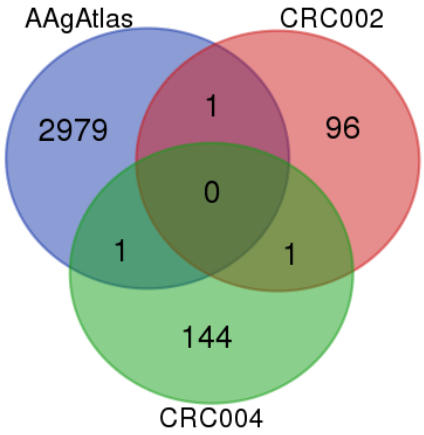


Figure 4.23. *Cancer*-enriched proteins matched to the AAgAtlas autoantigen database. *Cancer*-enriched proteins, defined as proteins with higher expression in *Cancer* versus *Normal* (and *Plasma*-specific proteins removed), for **CRC002** and **CRC004** were matched to the AAgAtlas human autoantigen database (version 1.0). The Venn diagram represents the number of proteins that are unique to, or overlap between, *Cancer*-enriched and the AAgAtlas proteins.

Citrullinated peptides detected in AP-MS assays:

In Chapter 3, the autoantibody response of CRC patients to citrullinated CT100plus microarray proteins was investigated. Here, autoantibody signals were detected against citrullinated versions of the proteins, indicating that a PAD enzyme might be present and active in the CRC tissues. From the AP-MS results, PAD2 was detected for *Cancer* sample of **CRC002**, and the *Cancer* and *Normal* samples of **CRC004**, but it was not detected for the remaining patient samples. To assess PAD activity, the raw spectra for each patient were processed through MaxQuant (1.5.3.12) to include the detection of citrulline residues. Stringent quality control conditions were used: Andromeda score > 40; Δ-score > 8; posterior error of probability (PEP) score < 0.01; and a localisation score of > 0.75. After the quality control measures were applied, a total of 58 proteins, 87 peptides and 108 citrullination sites were identified. Here, vimentin and actin were identified, which are normal substrates of PAD2 (Zhang et al., 2012; Hsu et al., 2014). We also identified citrullinated tubulin, heat shock protein and fibrinogen, which are PAD2 substrates in cancer cell-lines and associated with tumour growth, invasion, and a poor prognosis (Jiang et al., 2013). Although the citrulline modification may be expected for *Cancer* and *Normal* samples, as PAD2 is reportedly expressed in colon tissues, 15 citrullinated proteins were also detected in the *Plasma* sample. The list of proteins identified from citrullinated peptides for the *Cancer*, *Normal* and *Plasma* samples is summarised for each patient in Appendix B (Table B2 – B16). Together, these results indicate that PAD2 was detected, and was active, in normal and cancer colorectal tissues.

Discussion and conclusion

Cancer is a heterogenous disease capable of forming and spreading in most tissues of the human body. Cancer diagnosis usually involves invasive extraction techniques required for histological confirmation. The operations performed are risky and expensive, with a long waiting time in public sector hospitals in South Africa. It is therefore important to create a cost-effective, non-invasive cancer diagnostic tools that also indicates disease prognosis. With these in questions mind, the Blackburn lab created the CT100plus microarray, a tool that detects autoantibody signals from the blood of cancer patients, detectable in the (picogram per millilitre) pg/ml range. The CT100plus microarray is functionalised with native TSA and TAA, which include CTA. Thus, the autoantibody signals detected for patients are specific to aberrant cancer proteins, and should therefore be detected in cancer patients only (Beeton-Kempen et al., 2014). Despite the technological advancements of microarrays, autoantibody signals are not identified in all patients. Furthermore, none of the biomarkers have yet been associated with cancer prognosis. We therefore sought to assess a wider range of proteins functionalized similarly to CT100plus platform, for example, the SENGENICS IMMUNOME™ microarray.

Like the CT100plus microarray, the SENGENICS IMMUNOME™ microarray is functionalized with replicates of 1622 full-length, correctly folded proteins that contain conformational and linear epitopes, as well as carrying eukaryotic PTMs. The processed data of the blank SENGENICS IMMUNOME™ microarray indicates signal intensity below 500 RFU for all proteins, although a high signal intensity for IGHG1, confirming the addition of the detection antibody, Cy3-conjugated anti-human IgG. The slide incubated with pooled HC plasma produced high signal for IGHG1 (15162 RFU) and RBPJ (10954), and although 95.7% of the proteins produced low signal intensity (200-2000 RFU), 67 proteins produced medium signal intensity (2000-10000 RFU). The medium and high signal intensities may be due to genuine Ab signals present at low frequencies in the HC pooled sample; notably, young HC sera show very low binding to either the CT100plus or SENGENICS IMMUNOME™ microarrays (J. Blackburn, *personal communication*), but the average age of the commercial HC pooled sample is not described in the product sheet.

When assaying the CRC patient plasma, we were limited by cost to only 2 SENGENICS IMMUNOME™ microarrays, and assayed the plasma of **CRC038** (Figure 4.3) and **CRC050** (Figure 4.4) only, as these patient samples displayed high autoantibody signal on the CT100plus microarray. Since the 2 microarrays tested produced variable responses, it was a challenge to set a baseline threshold, and so we opted to assess the top 5 antigens with the highest signal intensity for each microarray. The highest signal intensity detected on the SENGENICS IMMUNOME™ microarray for both patients was toward proteins not present on the CT100plus microarray, except for p53 which was detected in **CRC050**. The highest signal intensity for **CRC038** was NOL4 (15477 RFU), TXN2 (10218 RFU), PPP2R4 (8956 RFU), SDCCAG8 (7611 RFU) and TFG (7565 RFU); whereas the highest signal intensity for **CRC050** was LGALS1 (34263 RFU), KRT19 (22412 RFU), FTH1 (21505 RFU), PLK1 (19694 RFU) and p53 (19253 RFU). Although the CT100plus microarray did not detect significant levels (>1000 RFU) of p53 for **CRC050**, high signal was detected against the p53 Q136x mutant form. Together, these results indicate that CRC patients produce autoantibody signals against proteins not present on the CT100plus microarray, implying that the CT100plus antigen content should be expanded, or that the SENGENICS IMMUNOME™ microarray should be used, to identify antigens linked with CRC diagnosis and disease prognosis. Importantly, more patient samples should be analysed to obtain enough power to obtain statistically significant correlations between autoantigens to CRC diagnosis and prognosis.

Although assessing the autoantibody response to 1622 proteins on the SENGENICS IMMUNOME™ microarray is an improvement from the 123 proteins on the CT100plus microarray, here we aimed to develop an assay that assessed the autoantibody response toward the CRC proteome. Several techniques have already been used to identify cancer antigens, including SEREX (Sahin et al., 1995), SERPA (Klade et al., 2001), phage display (Smith, 1985b), and natural protein microarrays (Qiu et al., 2004). Although these techniques have led to the discovery of cancer autoantigens, they are limited as most proteins are not identified in their native state and/or lack the correct PTMs, suggesting that TAAs or TSAs may be missed as conformational epitopes may be lost. We therefore sought to use a modified AP-MS assay, where CRC patient IgGs were captured using Protein G beads. The IgG-functionalized beads were used to capture cancer antigens from autologous native tissue lysates, which were subsequently

eluted and identified using mass spectrometry. Here, assay optimisation was performed, and protein quantitation (Figure 4.7) and SDS-PAGE (Figure 4.8) confirmed that 50 µl Protein G magnetic beads captured most of the IgG present in 15 µl CRC plasma. Western blot analysis with an anti-human IgG detection Ab confirmed that the 25 kDa band identified by SDS-PAGE was indeed IgG, and that Protein G captured more IgG than Protein A for **CRC038** (Figure 4.9) and **CRC050** (Figure 4.10). This result is supported by the reports that Protein G has a higher affinity for all IgGs, when compared to Protein A. Since the IgG response is assessed on the CT100plus and SENGENICS IMMUNOME™ microarrays, Protein G magnetic beads were used for subsequent assays.

In order to decrease the level of plasma protein contamination during the antigen elution step, IgG (and proteins that bind non-specifically) from plasma were crosslinked to Protein G beads. Although protein quantitation (Figure 4.11) and SDS-PAGE (Figure 4.12) analyses indicate that little or no protein was detected after cross-linking, proteins were detected by mass spectrometry. The high number of *Plasma* proteins could be rationalized by the nature of the antigen elution step: after Protein G and IgG were crosslinked, proteins were eluted for 30 minutes in a mild elution buffer containing trypsin. Thus, even though *Plasma* proteins were crosslinked to Protein G beads, *Plasma* peptides may have subsequently eluted due to trypsin digestion. However, DMP crosslinks at primary amines present on lysine residues and peptide N-termini, decreasing the number of lysine residues available for trypsin digestion. Thus, although *Plasma* peptides are detected by mass spectrometry, crosslinking plasma proteins may have decreased non-specific peptide signal by preventing proteins from eluting due to the presence of 2 M urea, and due to the unavailability of crosslinked lysine residues for trypsin cleavage.

After the crosslinking step, the IgG-functionalised Protein G beads were incubated with native CRC tissue lysate. Since CRC tissue was precious and limited, an SHSY-5Y human neuroblastoma cell-line was first used to optimize lysis buffer composition and lysis conditions. The lysis buffer included HEPES (pH 7.4), which is a Good's buffer characterised by its high compatibility with biological analysis, good solubility in water and minor salt effects (Good et al., 1966). Benzonase was added to digest nucleic acids. Protease inhibitors were added to avoid premature protein degradation. Most critical for the isolation of native proteins was the use of detergents – we

tested Triton™ X-100, IGEPAL® CA 630 and CHAPS, which are non-denaturing detergents. Varying lysis conditions were also tested, and from the results, it is evident that the buffer containing Triton™ X-100 with continuous vortexing at 4 °C produced the highest yield of protein from SHSY-5Y cell-line, and it was therefore used to produce native CRC tissue lysates. For the AP-MS assays, 200 micrograms (µg) native CRC lysate was tested which was sufficient to identify proteins by mass spectrometry.

Since IgG bind to both Protein A and G, the AP-MS assay with **CRC004** was performed to determine the specificity of each bead-type. The results showed high levels of presumed non-specific binding for the negative controls, as 117 and 453 protein groups were identified for the *Plasma* sample from Protein A and G, respectively. Since 94.0 % and 98.7 % of *Plasma* proteins from Protein A and G, respectively, were identified in the *Cancer* and/or *Normal* samples. These *Plasma* proteins were therefore removed from *Cancer* and *Normal* samples for downstream analysis. Furthermore, for the *Normal* sample, 898 and 812 protein groups were detected for Protein A and G, respectively. Together, these results indicate that further steps should be taken to decrease non-specific binding in the AP-MS experiments. It is expected that the difference in physicochemical properties of Protein A and G should affect non-specific binding. Despite this, 70 and 80 protein groups are shared between *Cancer*-enriched and *Cancer*-unique, respectively, indicating that the same CRC lysate proteins bound to IgG for each bead-type. To obtain support that cancer antigens might be enriched in the AP-MS assay, proteins that displayed significant signal intensity (> cut-off) on the CT100plus microarray were searched for in the AP-MS protein list.

For all patients, significant proteins (greater than cut-off value) detected on the CT100plus microarray were also detected by AP-MS. For **CRC002**, 9 proteins that displayed significant intensity on the CT100plus microarray were also detected by AP-MS (Table 4.1), whereas the homologues of 17 CT100plus antigens were detected by AP-MS (Table 4.2). For **CRC004**, 9 proteins that displayed significant intensity on the CT100plus microarray were also detected by AP-MS (Table 4.3), whereas the homologues of 15 CT100plus antigens were detected by AP-MS (Table 4.4). It is possible that antigens detected on the CT100plus microarray share an epitope present on the homologues identified by AP-MS. Despite PEG contamination issues, 3, 7 and 3

significant CT100plus antigens were also identified by AP-MS for **CRC017** (Table 4.5), **CRC021** (Table 4.6) and **CRC031** (Table 4.7), respectively. There are several possible reasons why antigens with significant intensity on the CT100plus microarray were not detected in the AP-MS assay: (**1**) Cancer is a heterogenous composition of abnormal cells, and only a ~3 mm^3 section of cancer specimen was assessed in the AP-MS assays. It is therefore possible that tissue sections that contain other antigens of interest were not assessed; (**2**) The peptides for the unidentified significant CT100plus proteins do not ionize well, and were therefore not detected by the mass spectrometer; (**3**) The unidentified significant CT100plus protein levels were too low to be detected by the mass spectrometer. Notwithstanding this, for the first time, we have demonstrated that antigens detected through protein microarray-based autoantibody assays in CRC patients can also be detected by an alternate method i.e. AP-MS.

To further assess the proteins identified by AP-MS, *Cancer*-unique and *Cancer*-enriched proteins were searched against the AAgAtlas human autoantigen database (version 1.0) with "cancer" as the disease-term (Wang et al., 2017). For *Cancer*-unique, there was no protein shared between **CRC002**, **CRC004** and AAgAtlas, although 1 (PARP1) and 8 (RRL12, VWf, PDLI1, RO60, IF2B2, B2MG, HYES and MRP3) proteins matched to AAgAtlas for **CRC002** and **CRC004**, respectively (Figure 4.22). Furthermore, 17 *Cancer*-unique proteins were shared between **CRC002** and **CRC004** only. For *Cancer*-enriched, 0 proteins were shared between **CRC002**, **CRC004** and AAgAtlas, although 1 (SYG) and 1 (HNRNPL) proteins matched to AAgAtlas for **CRC002** and **CRC004**, respectively (Figure 4.23). Furthermore, 1 *Cancer*-enriched protein was shared between **CRC002** and **CRC004** only. These results indicate that potential CRC antigens were identified that are not present on the CT100plus or SENGENICS IMMUNOME™ microarray. It is essential that more patient samples are assayed in order to obtain enough power to produce statistically significant results. Once a selected group of proteins for cancer diagnosis and prognosis are identified by AP-MS, the proteins of interest should be expressed in insect cells and printed on the CT100plus microarray to confirm that an autoantibody response is detected in the CRC patients.

For the AP-MS analysis, citrullination detection was included in the MaxQuant search. Although PAD2 was detected for only **CRC002** and **CRC004**, citrullinated peptides were detected for all

samples tested. Although cancer-associated citrullinated peptides were detected in the AP-MS data, we did not the detect citrullinated peptides of proteins (CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1) associated with an autoantibody response on the citrullinated CT100plus microarray. However, homologues of CDK7, MAGEB1, MAGEB5, MAGEB6 were identified including CDK5, MAGED2, suggesting that we may have in fact detected autoantibody signals toward citrullinated homologues on the citrullinated CT100plus microarray. There are several alternative explanations for not detecting CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1 by AP-MS: (**1**) if the citrullinated epitope interacts with an autoantibody, steric hindrance could prevent tryptic cleavage of the citrullinated peptide during the on-bead elution step; (**2**) citrullinated peptides do not ionize well and may not be detected by the mass spectrometer; or (**3**) The citrullinated peptide may not have been sufficiently enriched, and was therefore below the limit of detection for the mass spectrometer. Previous studies reported that citrulline-enrichment, e.g. using Biotin-PEG_2-4-glyoxalbenzoic acid (BPG), improved the detection from 119 to 3600 citrullinated peptides (Tutturen et al., 2014), suggesting that citrullinated peptide-enrichment may be important to detect more citrullinated peptides in future studies.

In conclusion, we have detected autoantibody responses in CRC patients using both the CT100plus and SENGENICS IMMUNOME™ microarrays. Furthermore, for all for 5 patients studied, cancer antigens detected on the CT100plus microarray were also detected by AP-MS. Several AP-MS-identified proteins that are not present on the CT100plus microarray matched to the AAgAtlas database. Lastly, citrullinated cancer proteins were detected by AP-MS, and whilst the citrullinated CT100plus antigens were not identified by AP-MS, homologues of these proteins were identified. These results are promising, but it will be important to assess at least 15 more patients to increase statistical power.

Chapter 5

A temporal CT100plus autoimmune assessment of stage IV melanoma patients treated with checkpoint inhibitors

Abstract

Melanoma is curable if detected early, whereas advanced melanoma is associated with high mortality rates. Immunotherapy with checkpoint inhibitors pembrolizumab, nivolumab and ipilimumab has improved survival in advanced stage melanoma patients. Biomarkers that predict treatment outcomes are hugely beneficial in providing optimal patient treatment, whilst limiting potential adverse treatment-side-effects. In this study, we assessed the autoantibody response of 52 melanoma patients before and after treatment with checkpoint inhibitors. Significantly (p-value ≤ 0.05, adjusted p-value ≤ 0.05) higher autoantibody signals were detected in melanoma patients *versus* HCs toward CEACAM 1, DPPA2, FGFR2, ITGB1, MAGEA10, NANOG, PIM1, SPANXB1, THEG and XAGE1B, which produced AUC-, sensitivity- and specificity-ranges of 0.79-0.87, 0.77 – 1.00 and 0.64 – 0.79, respectively. The combined ROC result indicates that CEACAM1 and FGFR2 were the best antigen combination, producing AUC-, sensitivity-, and specificity-values of 0.93, 0.96 and 0.94, respectively. Multivariate PCA and supervised hierarchical clustering analyses resulted in distinct groups of melanoma patients and HC clusters. Patient treatment information was available for the melanoma patients, although the longitudinal assessment indicates that no significant association was found for patient treatment-type and antigenic signal. In conclusion, although no association was found between immunotherapy treatment, we identified potential autoantibody biomarkers for the detection of advanced stage melanoma.

Introduction

Melanoma is a type of skin cancer that develops from melanocytes - a cell-type of the epidermis that produces the UV-protective melanin pigment. Although melanomas account for ~1 % of all cancer cases, it is responsible for most skin-cancer-related deaths, particularly after having metastasized. In fact, worldwide statistics revealed approximately 133 000 new cases of melanoma per year (Boyle and Levin, 2014), with 3.1 million active disease cases (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016), and 59 800 melanoma-related deaths in 2015 alone (GBD 2013 Mortality and Causes of Death Collaborators 2015), with the highest rates reported per year from Australia and New Zealand (WHO, 2012; Boyle and Levin, 2014). In 2014, the Australian melanoma mortality rate was reportedly 5.5 deaths per 100 000 persons (8.1 for males; 3.3 for females), estimating to have increased to 6.3 deaths per 100 000 persons (9.5 for males; 3.6 for females) in 2017 (Whiteman et al., 2008). This increased incidence was associated with increased recreational exposure to UV radiation and/or early stage detection of lesions. Although the incidence rates have increased, the overall trend shows that melanoma incidence is beginning to stabilize. Furthermore, the 5-year survival rate from 1988 to 2013 increased from 86% to 90%, likely associated with protection from ultraviolet (UV) light, early disease detection, improved chemo- and immuno-therapy, and surgery techniques (Clegg, 2002; Azoury and Lange, 2014).

In this study, we use the CT100plus microarray to investigate the autoantibody response of advanced stage melanoma patients from the Olivia Newton-John Cancer Research Institute (ONJCRI) in Australia. For the Australian population, the age-standardized incidence and mortality rate increases with age for men and women, with a preponderance toward men (www.aihw.gov.au/acim-books). Melanoma typically forms on the legs of women and the trunks of men, differences unlikely due to UV exposure as these regions of the body are not necessarily exposed to sun light. Furthermore, mortality studies show that women tend to have thinner tumours and higher survival rates, suggesting the potential influence of X-linked gene expression or hormonal factors (de Vries et al., 2008).

Despite progress in the melanoma field, 13 941 new cases (8392 males; 5549 females), and 1839 deaths (1280 males; 559 females) are predicted in 2017 for Australia alone (www.aihw.gov.au/acim-books). These estimates suggest that further melanoma research into early detection, recurrence and therapeutic response is needed, especially research that assesses the immune response because immunotherapy has significantly increased the 5-year relative survival rate.

Preclinical studies with CLTA-4 and PD-1 blockade led to decreased tumour burden, thus immune checkpoint inhibitors have been further developed for cancer patient treatment. Monoclonal antibodies that block CLTA-4 and PD-1 have been developed for melanoma, lung, kidney, prostate and head and neck cancers. Agents targeting PD-L1 are also under development. Ipilimumab, was the first anti-CTLA-4 treatment to prolong survival in patients with advanced melanoma (Hodi et al., 2010; Robert et al., 2011). Long-term survival analysis indicated a 3-year survival of 22%, 26% and 20% in all patients with sufficient follow-up, in treatment naïve patients, and in previously treated patients, respectively (Schadendorf et al., 2015). PD-1 blockade has improved survival for metastatic melanoma, NSCLC and RCC patients (Herbst et al., 2014; Topalian et al., 2014; Garon et al., 2015; Gettinger et al., 2015; McDermott et al., 2015). In one trial, advanced melanoma patients treated with pembrolizumab (anti-PD-1 antibody) and ipilimumab (anti-CTLA-4 antibody) showed a superior response in pembrolizumab-treated versus ipilimumab-treated patients, with response rates of 34% and 12%, respectively. In conjunction with these data, patient survival rates were also high in pembrolizumuab versus ipilimumab treatment, at 74% and 58%, respectively (Farolfi et al., 2012). Nivolumab is a human IgG4 anti-PD-1 monoclonal Ab that selectively blocks the interaction between PD-1 and PD ligands, restoring T-cell immune activity directed against cancer. Nivolumab was the first checkpoint inhibitor approved for advanced NSCLC following chemotherapy, resulting in increased response rates, survival and progression-free survival when compared to intravenous docetaxel. Nivolumab was better tolerated than docetaxel, as nivolumab-induced immune adverse events were more manageable (Keating, 2015). Stage III/IV melanoma patients achieved a partial tumour response, with a median progression free survival of 172 days, with only 18% experiencing grade 2 or 4 adverse events (Deeks, 2014).

Biomarkers that predict treatment outcomes would be hugely beneficially in providing optimal patient treatment, whilst limiting potential adverse treatment-side-effects. Upregulation of PD-1 on exhausted T-cells and of PD-L ligands on tumour cells, and/or tumour infiltrating cells, may serve as biomarkers that predict patient response to PD-1 or PD-1 ligand vaccines. Preliminary results of multiple tumour-types indicated that patients with PD-L1-expressing tumours or infiltrating immune cells typically have higher response rates to anti-PD-1 or anti-PD-L1 versus patients with low or no PD-L1 expression (Herbst et al., 2014; Taube et al., 2014). However, PD-L1 expression on tumours does not correlate with treatment outcomes for all cancer-types, as PD-L1 levels in squamous NSCLC did not correlate with patient response to immunotherapy (Borghaei et al., 2015; Brahmer et al., 2015; Herbst et al., 2016; Passiglia et al., 2016). It has been more challenging to identify predictive biomarkers for ipilimumab-treatment, due to low levels of CTLA-4 expression, and the wide range expression of B7 ligands. Retrospective studies have identified markers associated with treatment response, including absolute lymphocyte count, increased levels of T-cell activation ICOS, and T-cell response to NY-ESO1 (Callahan et al., 2013), although none of these biomarkers have since been validated. In addition, another study concluded that neoantigen presentation on tumours correlates with overall survival in CTLA-4-treated melanoma patients (Snyder et al., 2014). Together, it is evident that reliable biomarkers for treatment response are required which would be beneficial in managing patient treatment, thereby improving treatment response rates and reducing immune related adverse events.

The Blackburn group gained access to plasma and serum samples of 52 advanced stage melanoma patients before and at various timepoints (TPs) after treatment, amounting to 304 plasma and/or sera samples, from the ONJCRI. The samples were collected from clinical cohorts (HREC/14/Austin/425 - Biomarkers in Cancer) where patients were treated with checkpoint inhibitors; ipilumumab, nivolumab, pembrolizumab, combinations thereof or placebos. Here, melanoma patients and HCs were compared using the CT100plus microarray platform to identify advanced state melanoma-associated TAA; and the treatment information was used to correlate autoantibody signatures with treatment undergone.

Results

5.1. CT100plus microarray quality controls

A total of 304 samples, corresponding to 52 stage IV melanoma patients (Table 5.3 – 5.10), were assayed on the CT100plus microarray. Three controls were performed, including a c-Myc (Figure 5.1), pooled negative control plasma (Figure 5.2), or pooled positive control plasma (Figure 5.3). Each of the 123 recombinant proteins on the CT100plus microarray (Figure 7.4) contains a c-Myc tag that is used to confirm the successful immobilisation of recombinant proteins on the slide surface. It is evident from Figure 5.1 that antigens do not all have the same intensity across the microarray. This is because each of the 123 recombinant proteins is expressed at different levels within insect cells. Although the signal for every antigen is not visibly discernible on the image, protein expression and biotinylation was confirmed by Western blot analysis (Beeton-Kempen et al., 2014).

Both the positive and negative control assays display low or no signal from buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. Furthermore, the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots were detected, which was used in subsequent data normalisation steps; anti-human IgG was detected, confirming the addition of patient plasma; and human IgG was detected, confirming the addition of detection antibody. Low or no signal is detected for the pooled negative control (Figure 5.2), whereas high signal is detected for the pooled positive control (Figure 5.3), indicating a high-quality platform that can be used for the specific detection of autoantibodies against CT100plus cancer antigens.

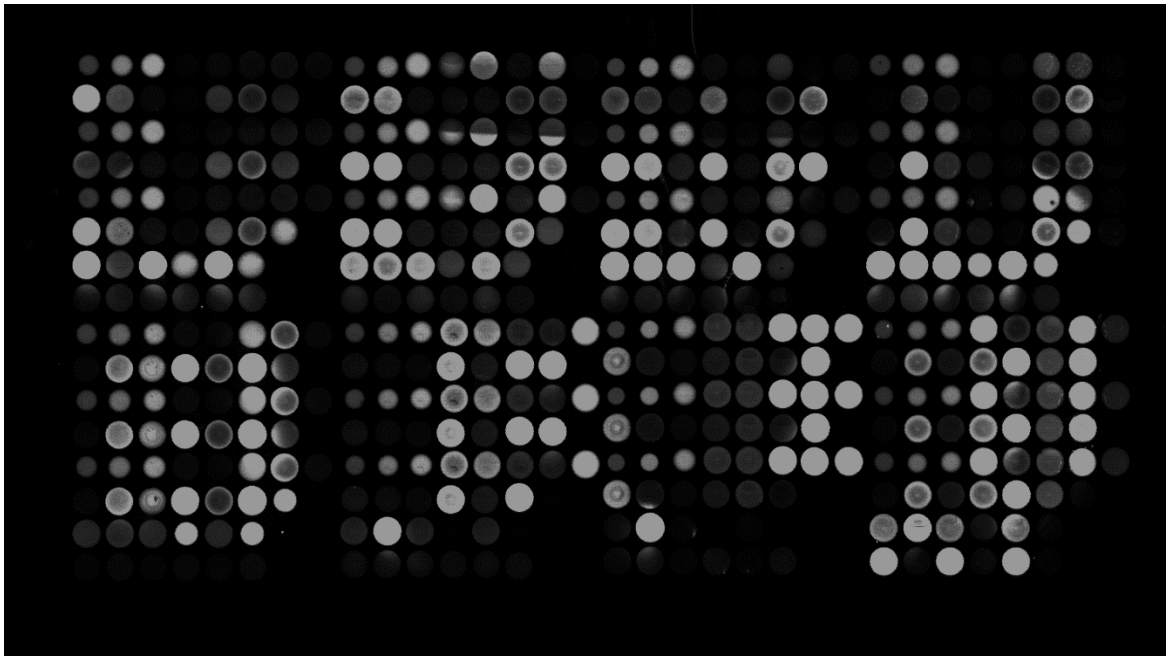
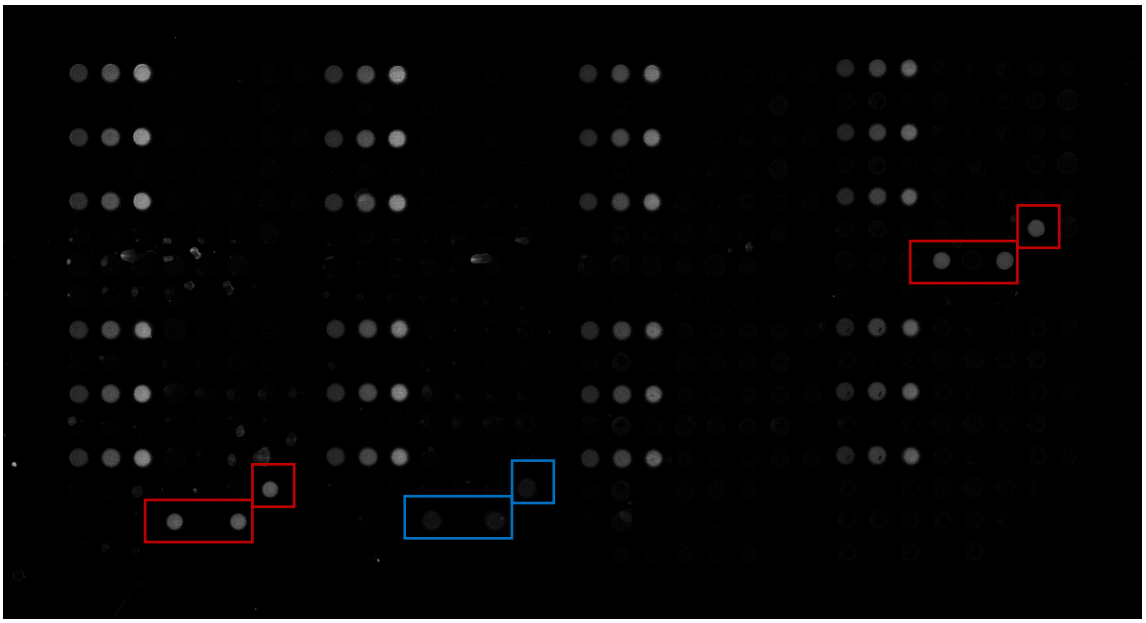


Figure 5.1. CT100plus microarray c-Myc control. Each of the 123 recombinant proteins on the CT100plus microarray contains a c-Myc tag. The microarray was treated with Cy3-labelled anti-c-Myc antibody, which was used to confirm the successfully immobilisation of recombinant proteins on the slide surface with an anti-c-Myc antibody, as depicted in the microarray image. Abbreviations include: Cy3 = cyanine 3.



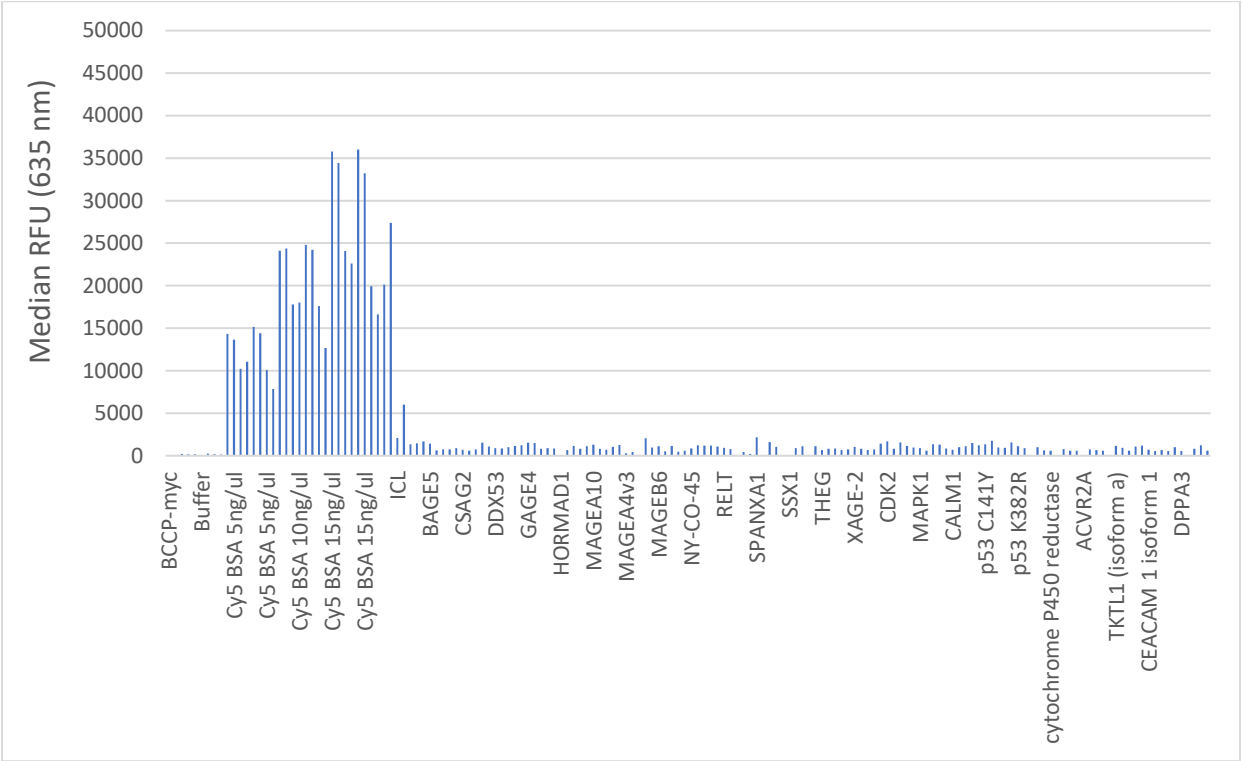


Figure 5.2. CT100plus microarray negative control pooled plasma sample. As a negative control, the CT100plus microarray was incubated with pooled plasma of HCs. Low or no signal is detected for the buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. High signal was detected for the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots, anti-human IgG was detected (boxed in red), and human IgG was detected (boxed in blue). Low or no signal is detected for recombinant proteins. Abbreviations include: HCs = healthy controls; ICL = insect cell lysate; BCCP = biotin carboxyl carrier protein; IgG = immunoglobulin G; Cy5 = cyanine 5; BSA = bovine serum albumin.

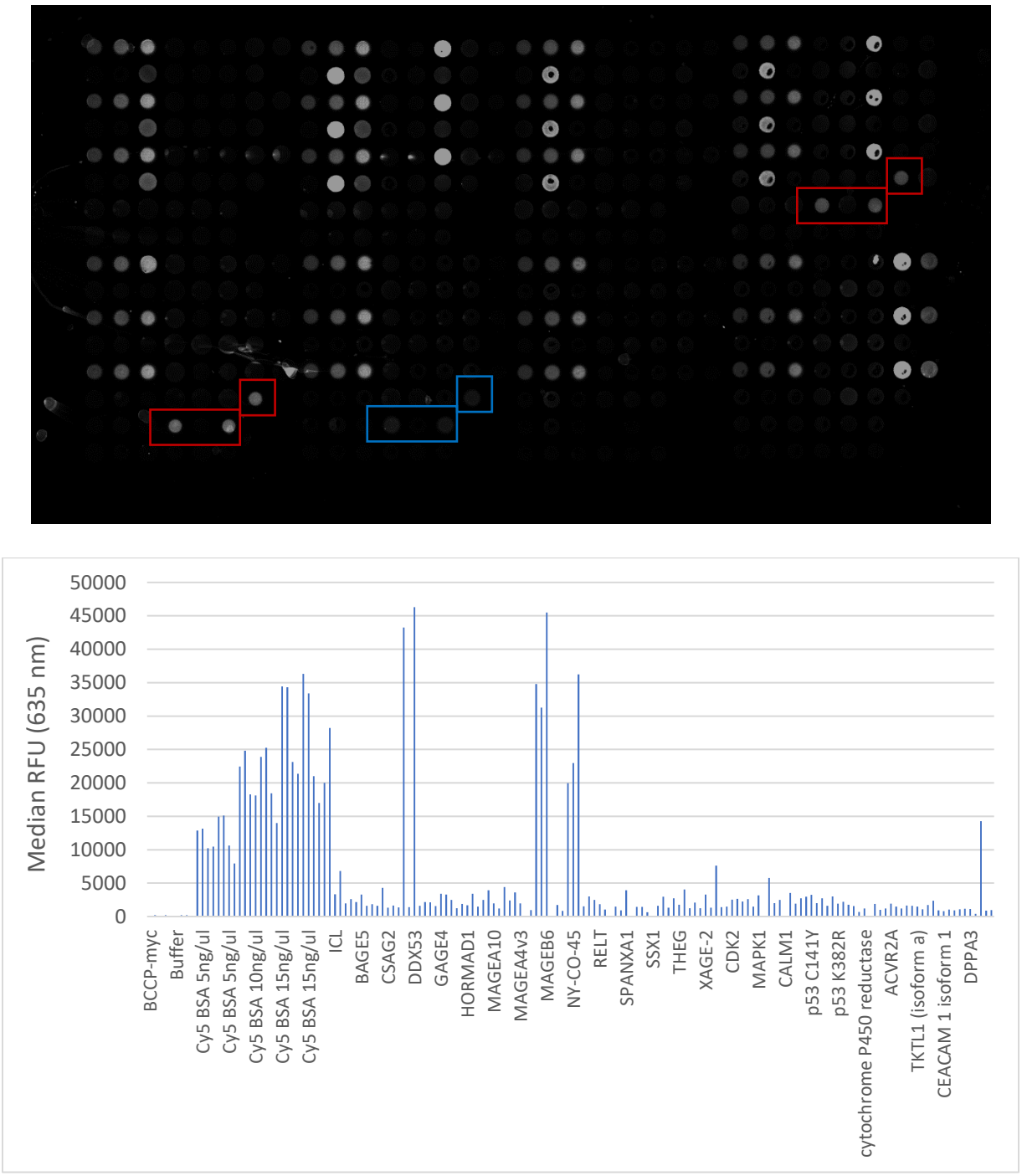


Figure 5.3 CT100plus microarray positive control pooled plasma sample. For the positive control sample, the CT100plus microarray was incubated with pooled plasma of cancer patients. Low or no signal is detected for the buffer, ICL, and BCCP negative control spots, indicating low levels of non-specific binding. High signal was detected for the 5, 10 and 15 ng/µl Cy5-biotin-BSA positive control spots, anti-human IgG was detected (boxed in red), and human IgG was detected (boxed in blue). Distinct and varying autoantibody signal intensity was detected for recombinant proteins. Abbreviations include: HC = healthy controls; ICL = insect cell lysate; BCCP = biotin carboxyl carrier protein; IgG = immunoglobulin G; Cy5 = cyanine 5; BSA = bovine serum albumin.

5.2. Batch-to-batch normalisation

A total of 307 samples were processed on the CT100plus microarray, consisting of 304 advanced melanoma patient samples, and the 3 controls described in Section 5.1. Due to the large cohort size, the samples were assayed in two separate batches using two separate microplate sources. Furthermore, to improve microarray image and data quality, images that failed visual and CT100plus software assessments were re-assayed in two separate batches. Thus, the resulting data are derived from four sample batches, which will further be referred to as batch **B1**, **B2**, **B3** or **B4**.

A batch-effect was observed for all samples processed in separate batches, and patient **14** is here used as an example. For patient **14**, sera samples were processed for nine TPs (**14A**-**I**) in batch **B1**, although samples **14B** and **14C** were re-assayed in batch **B3** to further improve the quality of the microarray images. To assess the batch-effect, the signal intensity of cytochrome p450 3A4 was plotted for patient **14** across various TPs either from the same batch (blue line) or separate batches (orange line) in Figure 5.4. Here, it is evident that assaying samples in separate batches (orange line) results in a greater signal intensity variation between time points, compared to the same samples processed together (blue line). To assess the batch-effect for all proteins on the CT100plus microarray, a volcano plot was constructed in RStudio (version 1.0.136) using the ggplot2 package for proteins between all batches. Blue dots indicate a significant difference (adjusted p-value ≤ 0.05, fold-$\Delta \geq 0.25$) in signal intensity between batches, whereas the red dots indicate a non-significant (adjusted p-value > 0.05, fold-$\Delta < 0.25$) difference in signal intensities. Here, it is evident that many of the proteins are significantly different between batches, and furthermore, significant differences were observed between **B1** and **B2**, **B3** as well as **B4** (Figure 5.5), a rare and unusual result for CT100plus microarray data, especially since low baseline signal intensities were observed for the majority of the antigens on most microarray images.

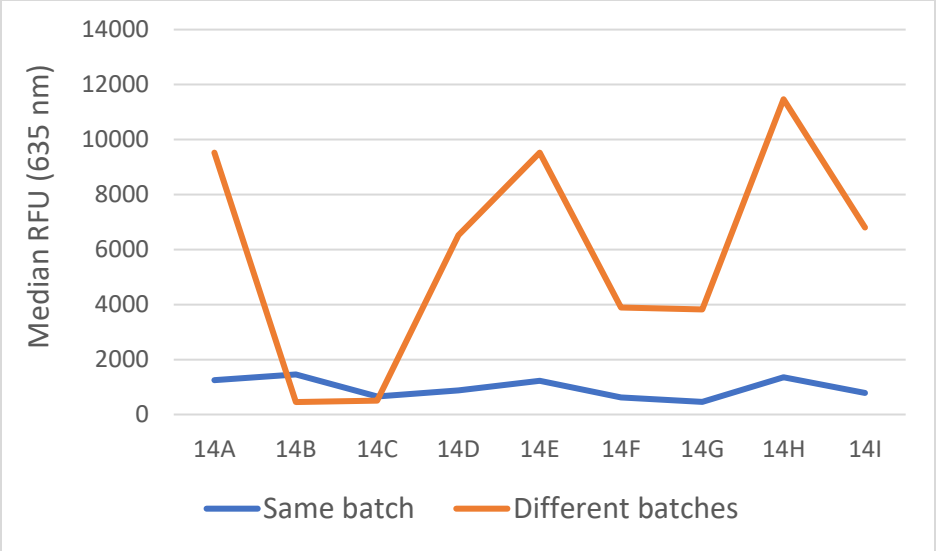


Figure 5.4. Temporal assessment of cytochrome P450 3A4 for patient 14 pre-normalisation. The samples of patient **14** were assayed in the same and separate batches, and here use to assess batch-to-batch effects with cytochrome p450 3A4 as an example. It is evident that assaying samples in separate batches (orange line) results in a greater variation in signal intensity compared to processing the samples within the same batch (blue line). Abbreviations include: RFU = relative fluorescence units.

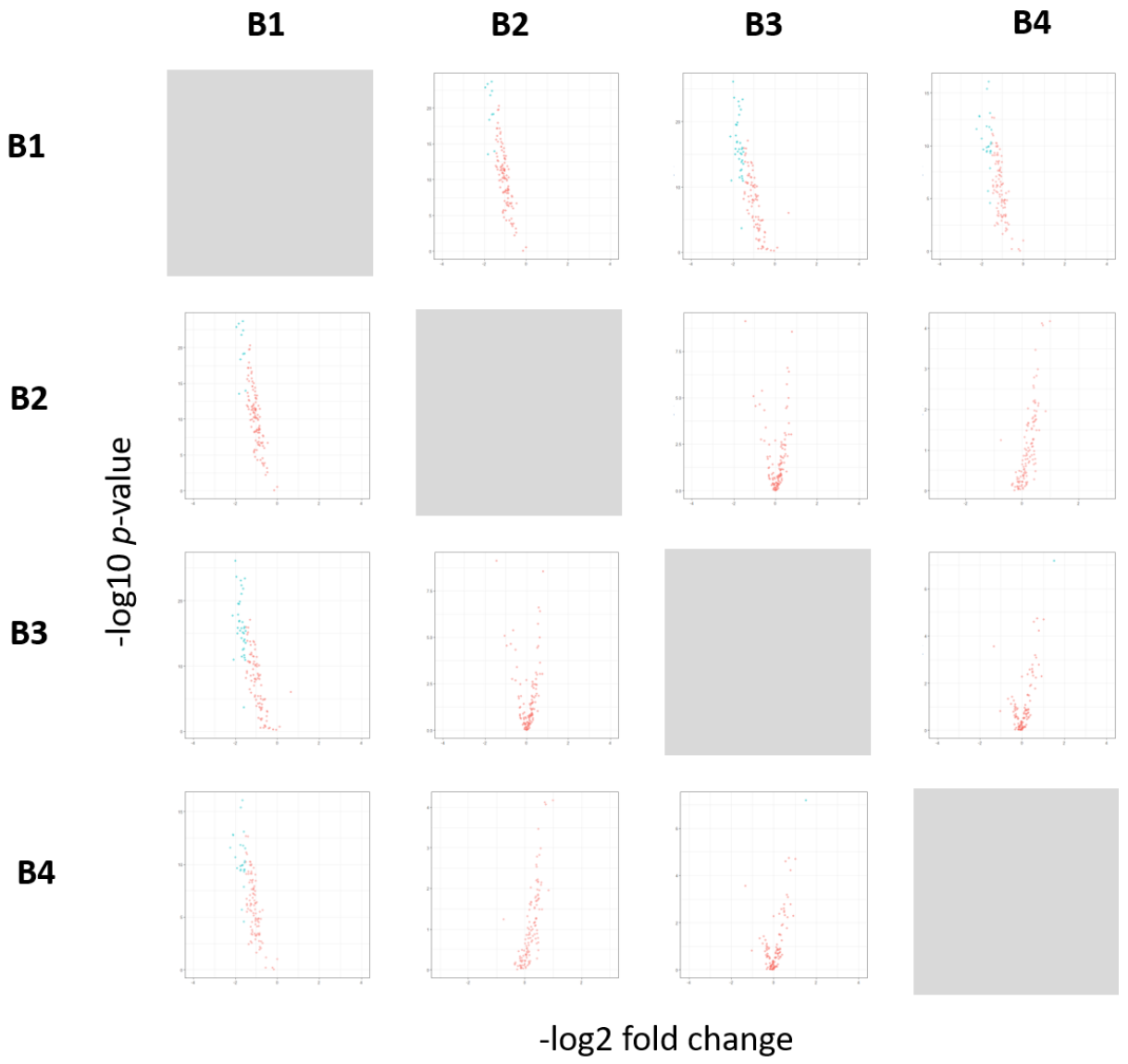


Figure 5.5. Volcano plot for the CT100 antigens pre-normalisation. A volcano plot was constructed in RStudio (version 1.0.136) to determine significant fold-changes (fold-Δ) in signal intensities between sample batches. Blue dots indicate a significant difference (adjusted *p*-value ≤ 0.05, fold-Δ ≥ 0.25) in signal intensity between batches, whereas the red dots indicate a non-significant (adjusted *p*-value > 0.05, fold-Δ < 0.25) difference in signal intensities. Here, it is evident that all antigen signals are dysregulated between batches, and furthermore, significant differences were detected between **B1** and **B2**, **B3** as well as **B4**, despite similar intensities viewed on the CT100plus images.

To account for batch-to-batch variation, we used an adapted normalisation method based on the total intensity-based module (Quackenbush, 2001). The total intensity-based module assumes that post-normalisation, all microarrays have the same total intensity-value for positive control

spots, given by $\sum_{i=1}^{N_{\text{spots}}} X_i$ .Microarray *k* is normalised using the normalisation factor a_k, given by Equation 5.1, where $\sum_{i=1}^{N_{\text{spots}}} X_{ik}$ is the total intensity of all positive control spots on array *k* prior to normalisation (Wang, 2013).

Equation 5.1. Normalisation factor for array-to-array normalisation.

$$a_k = \frac{\sum_{i=1}^{N_{\text{spots}}} X_i}{\sum_{i=1}^{N_{\text{spots}}} X_{ik}}$$

For batch-to-batch normalisation, the equation has been adapted to assume that post-normalisation the median of the median pixel intensity of all antigens across on a microarray is the same across all batches. Batch i is normalised using the normalisation factor a_i given by Equation 5.2.

Equation 5.2. Normalisation factor for batch-to-batch normalisation.

$$\alpha_i = \frac{\mu(X)}{\mu(X_i)}$$

α_i	:	Normalisation factor for array i
μ	:	Median
X	:	Set of median foreground pixel intensities for all microarrays
i	:	Microarrays from 1,2,3 . . . n

After batch-to-batch normalisation, it is evident that the signal intensity for patient **14** across all TPs are now similar (grey line) to the signal intensities for samples processed within the same batch (blue line) *versus* non-normalised data (orange line) (Figure 5.6). As expected, the volcano plot indicates no difference in the median antigen signal between batches (Figure 5.7).

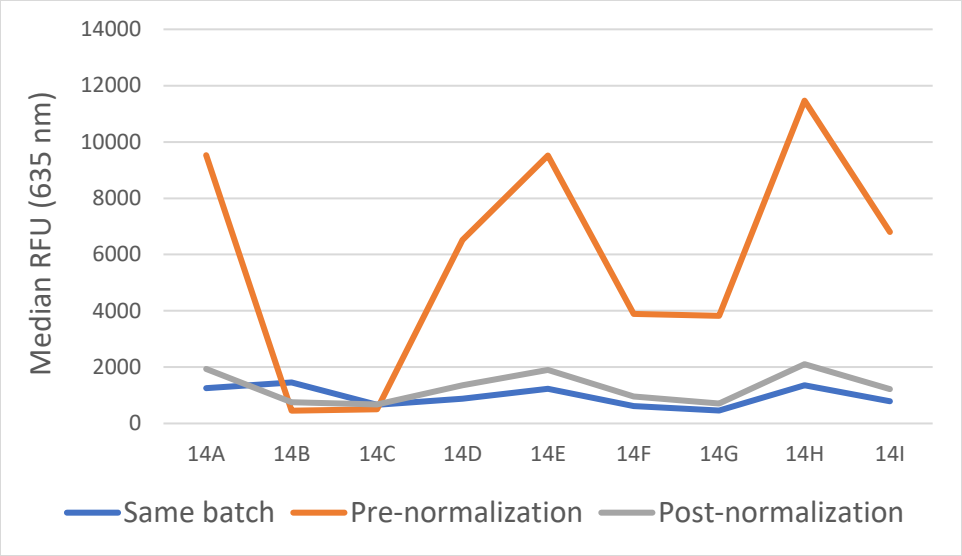


Figure 5.6. Temporal assessment of cytochrome P450 3A4 for patient 14 post-normalisation. The samples of patient **14** were assayed in the same and separate batches, although assays performed in separate batches result highly variable signal intensity that does not correlate with visual intensity. To account for batch-to-batch effects seen pre-normalisation (orange line), sample batches were normalised using Equation 5.2. The results indicate that signal intensities post-normalisation (grey line) are now likened to signal intensities of samples processed together (blue line). Abbreviations include: RFU = relative fluorescence units

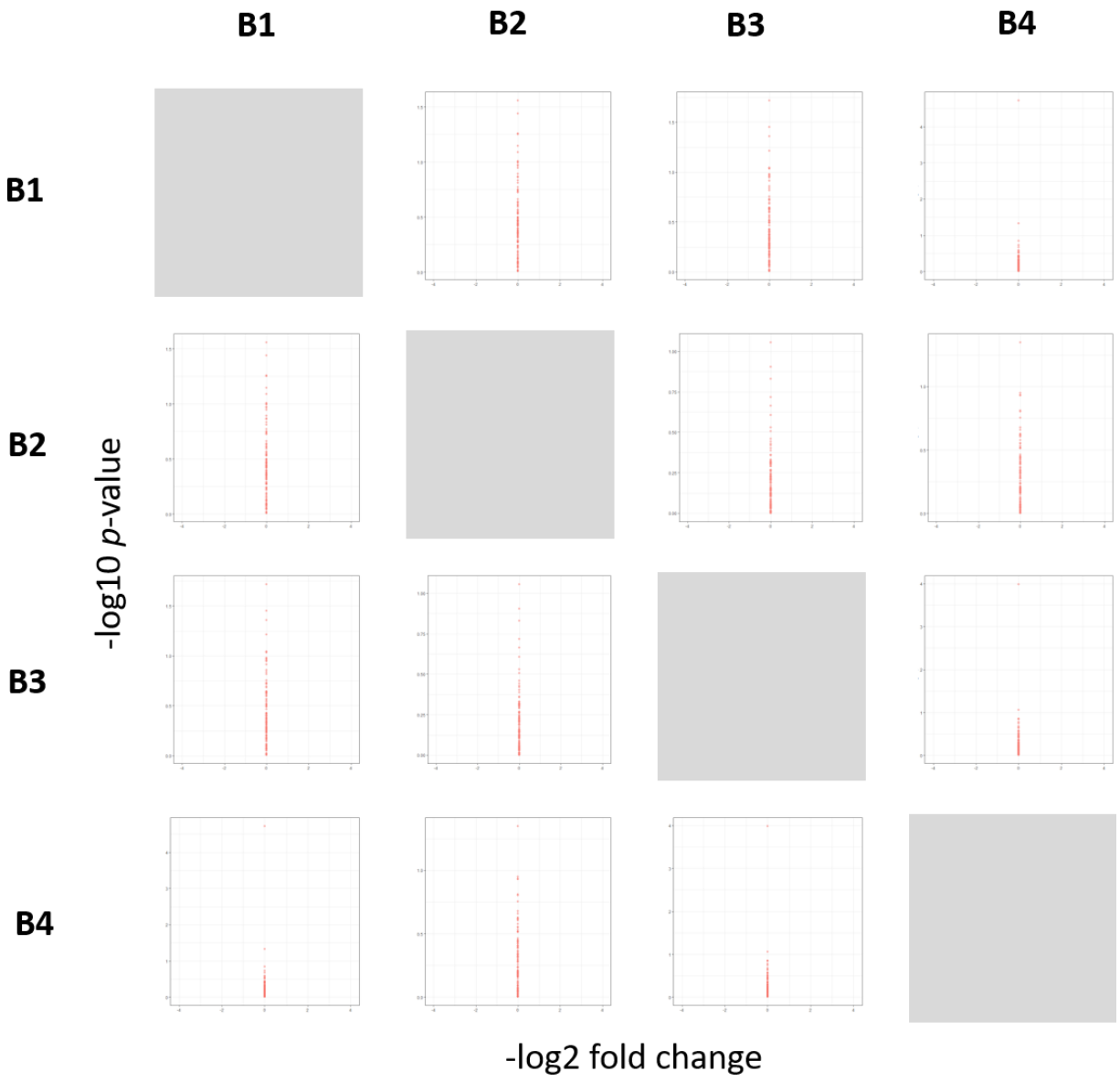


Figure 5.7. Volcano plot for the CT100 antigens for batches post-normalisation. A volcano plot was constructed in RStudio (version 1.0.136) to determine significant fold-changes (fold-Δ) in signal intensities between sample batches. Blue dots indicate a significant difference (adjusted *p*-value ≤ 0.05, fold-Δ ≥ 0.25) in signal intensity between batches, whereas the red dots indicate a non-significant (adjusted *p*-value > 0.05, fold-Δ < 0.25) difference in signal intensities. Now, antigen signals are no longer dysregulated between batches.

5.3. Assessing the autoantibody response between healthy controls and melanoma patients

The blood plasma and/sera of 52 patients across multiple TPs, amounting to 304 samples, were assayed on the CT100plus microarray with the methods described in Chapter 7. For each patient, blood samples were collected before (TP = 0) and after (TP = 1, 2, 3, …*n*) immunotherapy treatment; in most cases, multiple plasma and/or sera samples were collected post-treatment. In this section, the statistical results of the analysis between HCs (**HC1** to **HC14**) and melanoma patients (**1A** to **52A**) before immunotherapy are described. The CT100plus microarray images were used to extrapolate raw data files using the ArrayPro software, after which data filtering was performed using the CT100plus software with the following settings: whole array CV ≤ 25%; probe replicate CV ≤ 20%; noise-threshold ≥ 2 SD above background; maximum threshold = 65355 nm; whole array filtering control = 15 ng/µl Cyanine 5 (Cy5)-biotin-BSA; 57 patient data files passed the filtering and normalisation quality control procedure. The median signal intensity for each antigen was obtained, and used for subsequent statistical analyses.

A ROC analysis was performed, which reports the sensitivity and specificity for a tested parameter, and in this study, refers to CT100plus antigens. An *a priori* ROC sample size was performed using MedCalc software (version 17.2), with a hypothesised AUC-value of 0.80, a null hypothesis value of 0.5, and a sample ratio of 0.17. The results produced a α-value (significance) of 0.01 and β-value (1-Power) of 0.10 for the 52 CRC patients and 14 HCs, indicating that our sample size was sufficient for the ROC analysis.

The ROC analysis was performed in R Studio (1.0.136) using the ROCR package (Sing et al., 2005). A top10-list was generated from proteins with the highest AUC-values. The top 10 protein list included CEACAM 1 isoform 1, DPPA2, FGFR2, ITGB1, MAGEA10, NANOG, PIM1, SPANXB1, THEG and XAGE1B, which were selected for further ROC analysis to produce a sensitivity and specificity report, displayed in Table 5.1. Combinatory ROC analysis was also performed for the top 10 antigens using CombiROC (Mazzara et al., 2017), indicating that a combination of CEACAM1 and FGFR2 was the best biomarker combination, producing a sensitivity-, specificity-, and AUC-value

of 0.93, 0.89, and 0.92, respectively, indicating that this combination outperforms individual proteins. The ROC curve in Figure 5.8 represents a combination for CEACAM1 and FGFR2.

Table 5.1. Sensitivity-, specificity-, AUC-, and cut-off-values for top 10 antigens for melanoma patients on the CT100plus microarray. Abbreviations include: AUC = area under the receiver operator characteristic curve.

Antigen	Sensitivity	Specificity	AUC	Cut-off
CEACAM 1 (iso1)	0.81	0.71	0.87	587
DPPA2	1.00	0.71	0.85	444
FGFR2	0.94	0.79	0.83	336
ITGB1	0.96	0.71	0.87	307
MAGEA10	0.77	0.71	0.80	760
NANOG	0.90	0.64	0.82	326
PIM1	0.94	0.79	0.79	353
SPANXB1	0.90	0.71	0.83	307
THEG	0.96	0.71	0.81	406
XAGE1B	0.92	0.79	0.82	451

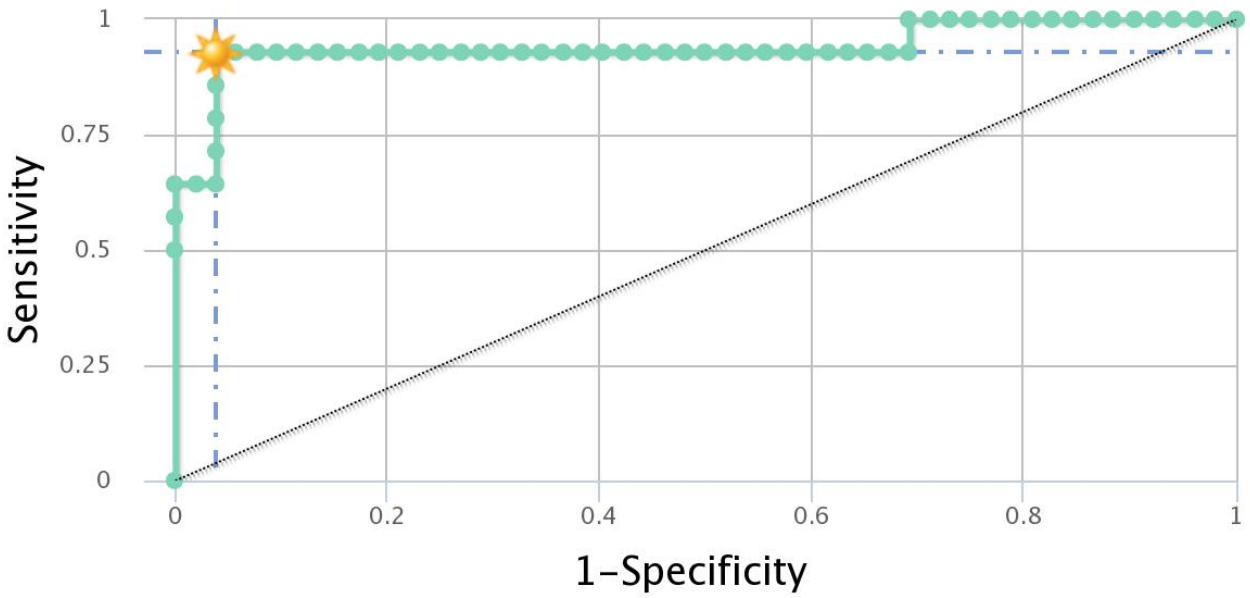


Figure 5.8. ROC curve for melanoma patients and HCs. The ROC curve was constructed using CombiROC. A combination of the CEACAM1 and FRGR2 produce the highest AUC is displayed, with the resulting sensitivity-, specificity- and AUC-values of 0.93, 0.96 and 0.94, respectively. The yellow star indicates the optimal cut-off point. Abbreviations include: ROC = receiver operating characteristic; AUC = area under the ROC curve.

It is evident that the signal intensity for each of the top 10 antigens was higher in melanoma patients when compared to healthy individuals (Figure 5.9). Further statistical tests were performed in R Studio (1.0.136): The Shapiro-Wilk (Shapiro and Wilk, 1965) results indicate a non-normal data distribution (p-value ≤ 0.05), as such, the Mann-Whitney U (Mann and Whitney, 1947) and Benjamini-Hochberg (Benjamini and Hochberg, 1995) tests were performed, indicating a significant increase (p-value ≤ 0.05; adjusted p-value < 0.05) in signal intensity for melanoma patients *versus* HCs on the CT100plus microarray (Table 5.2).

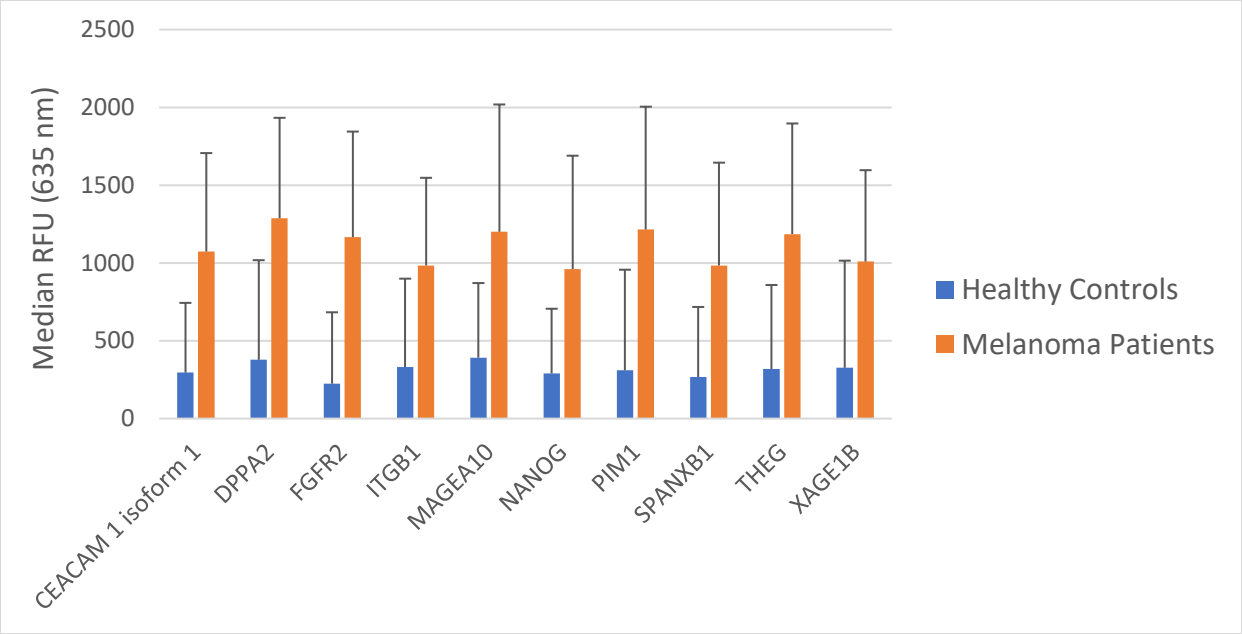


Figure 5.9. Pre-immunotherapy-treated melanoma patient and HC raw data files were batch processed using the CT100plus software, and a ROC test performed to identify the top 10 upregulated antigens in cancer patients. The graph displays the median RFU-values for the top 10 potential antigen biomarkers on the CT100plus microarray, and the error bars represents the MAD. Abbreviations include: HC = healthy control; ROC = receiver operator characteristic; RFU = relative fluorescence units; MAD = median absolute deviation.

Table 5.2. Shapiro-Wilk, Mann-Whitney U and Benjamini-Hochberg test values for the top 10 antigens for melanoma patients on the CT100plus microarray.

Antigen	Shapiro Wilk (*p*-value)	Mann-Whitney *U* (*p*-value)	Benjamini-Hochberg (adjusted *p*-value)
CEACAM 1 (iso1)	2.14×10^{-4}	5.43×10^{-5}	3.34×10^{-3}
DPPA2	3.30×10^{-3}	1.96×10^{-4}	5.19×10^{-3}
FGFR2	3.91×10^{-4}	2.02×10^{-5}	2.47×10^{-3}
ITGB1	4.26×10^{-3}	6.40×10^{-4}	8.75×10^{-3}
MAGEA10	3.12×10^{-8}	2.53×10^{-4}	5.19×10^{-3}
NANOG	2.20×10^{-6}	7.66×10^{-4}	9.42×10^{-3}
PIM1	1.24×10^{-4}	1.50×10^{-4}	5.19×10^{-3}
SPANXB1	2.67×10^{-4}	4.09×10^{-4}	7.19×10^{-3}
THEG	5.59×10^{-5}	2.18×10^{-4}	5.19×10^{-3}
XAGE1B	7.18×10^{-4}	5.50×10^{-4}	8.45×10^{-3}

To further assess the biomarker potential of the top 10 proteins, unsupervised tests including PCA multivariate testing and unsupervised hierarchical clustering were performed in Perseus (version 1.5.4.1):

Unsupervised hierarchical clustering analysis produced two major segregating clusters of melanoma patients and HCs. Most of the HCs clustered together, except for HCs **HC10** and **HC12** who cluster separate from the major HC cluster. Furthermore, **HC1** and **HC9** are separate from other HCs, clustering with patients **17A** and **19A** (Figure 5.10).

In agreement with the hierarchical clustering analysis, the 1D and 2D PCA plots display two major segregating clusters, each representing CRC patients (red) or HCs (blue), although **HC10** cluster closer to CRC patients, as is evident from the 1D and 2D PCA plots (Figure 5.11).

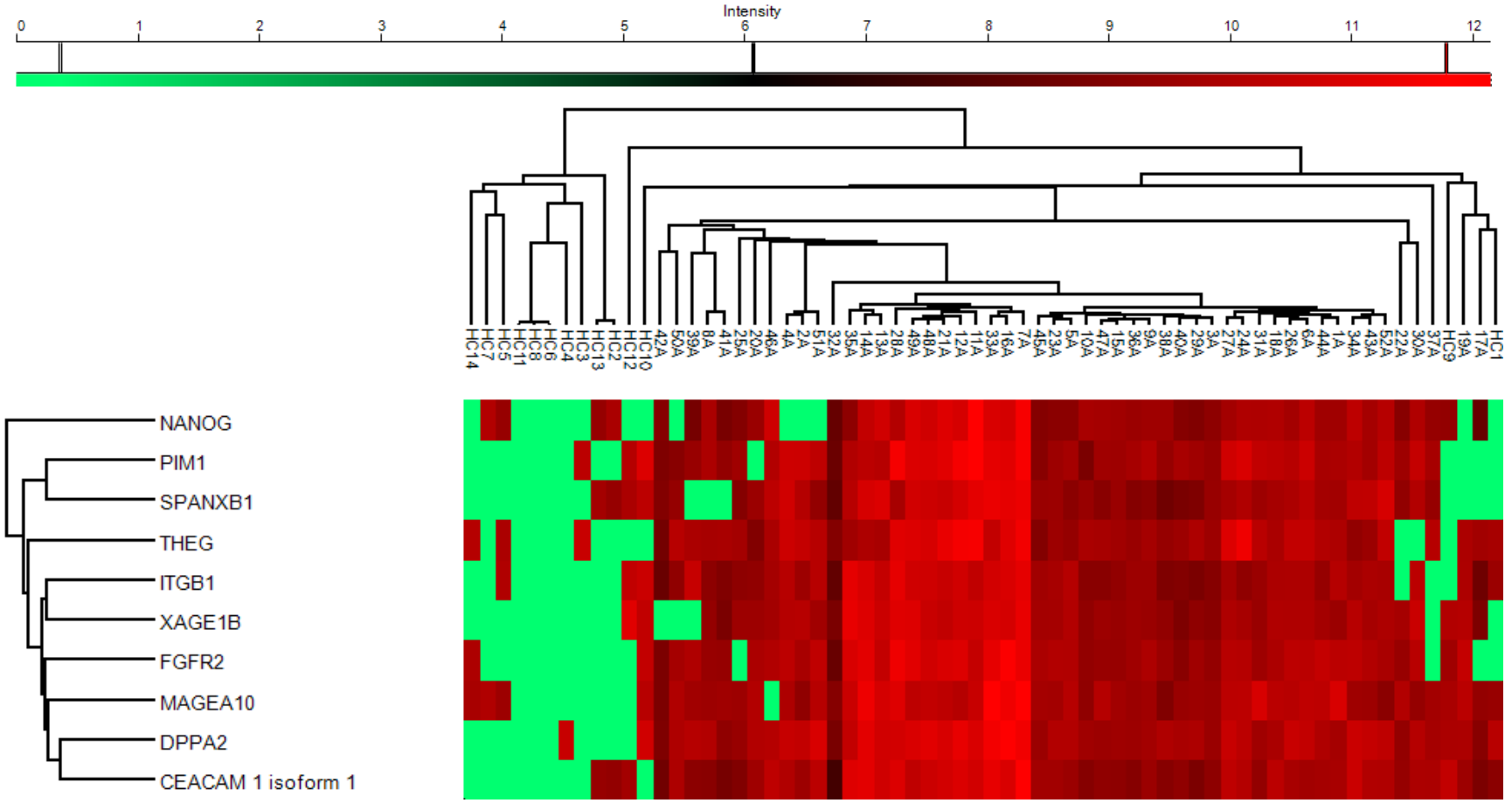


Figure 5.10. Unsupervised clustering of pre-immunotherapy melanoma patients and HCs. Pre-immunotherapy melanoma patient (*N* = 52) and HC (*N* = 14) plasma and/or sera were processed on the CT100plus microarray, and the 10 most significantly upregulated antigens in melanoma patients were $\log_2$-transformed and selected for unsupervised hierarchical clustering using the Perseus software (version 1.5.4.1). Here, the HC and melanoma patients form two major clusters. Most of the HCs clustered together, except for HCs **HC10** and **HC12** who cluster separate from the major HC cluster. Furthermore, **HC1** and **HC9** are separate from other HCs, clustering with patients **17A** and **19A**. Abbreviations include: HCs = healthy controls; *N* = number of patients.

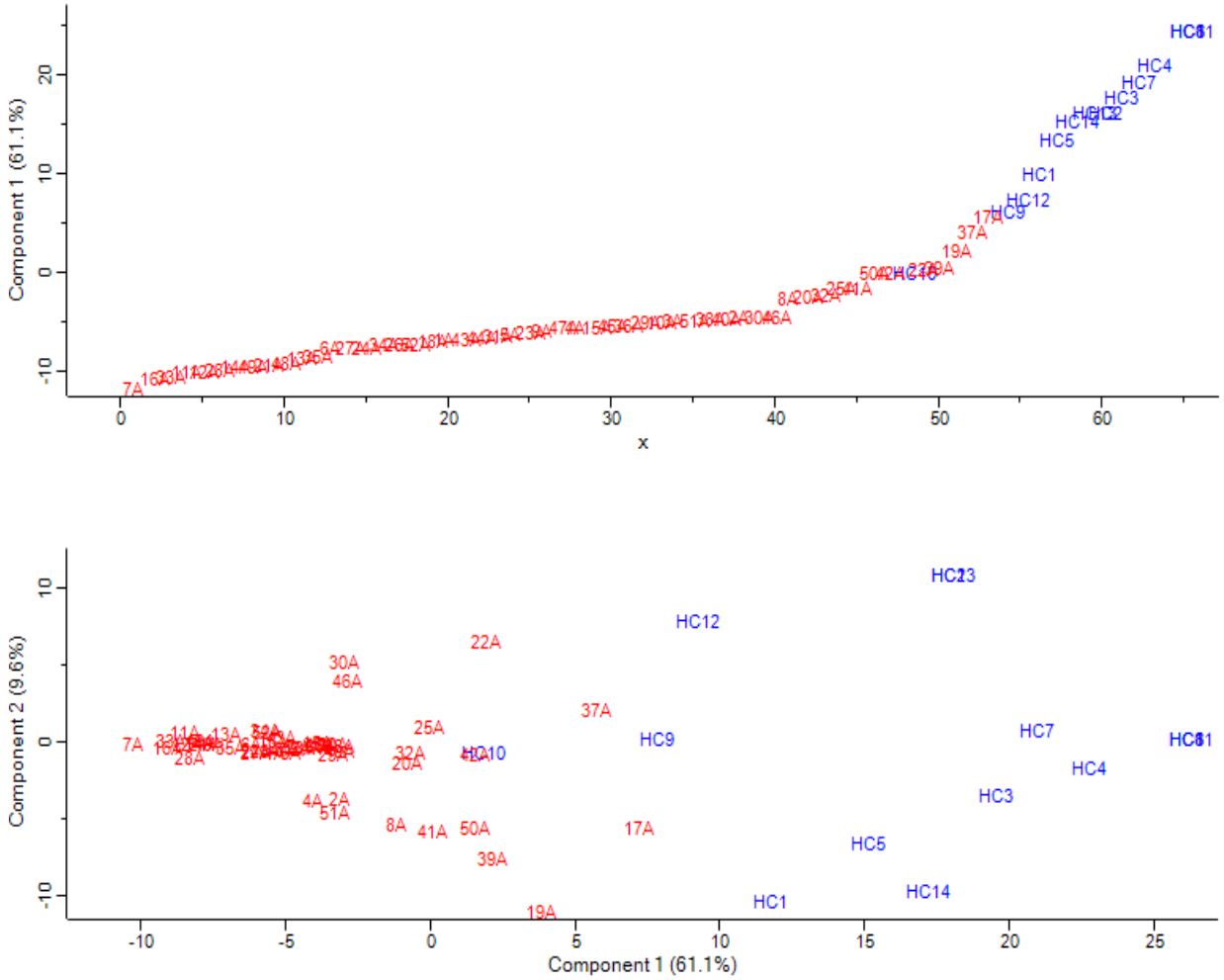


Figure 5.11. PCA between pre-immunotherapy treated melanoma patients and HCs. Pre-immunotherapy treated melanoma patient (*N* = 52) and HC (*N* = 14) plasma and/or sera samples were processed on the CT100plus microarray, and the 10 proteins with the highest AUC-values were log_2-transformed and selected for 1D and 2D PCA multivariate testing using the Perseus (version 1.5.4.1). The 1D and 2D PCA plots display melanoma patient (red) and HC (blue) clusters, with overlap between melanoma patient **42A** and **HC10**. Abbreviations include: PCA = principle component analysis; HC = healthy control; *N* = number of patients; AUC =area under the receiver operator characteristic curve; 1D = one-dimensional; 2D two-dimensional.

5.4. Autoimmune responses toward checkpoint inhibitor treatment

The serum and/or plasma of 52 advanced-stage melanoma patients treated with checkpoint inhibitors, including ipilimumab, pembrolizumab, nivolumab, or combinations thereof, was assayed on the CT100plus microarray. Microarray images were captured and the numerical data extrapolated using the ArrayPro software. The microarray images were re-assayed if images showed evidence of spot running, missing spots (due to pin sticking), washing artefacts, dust particles and/or faded arrays. The numerical data was filtered and normalised using the CT100plus software bioinformatic pipeline for pre-processing and quality control. The data filtering and normalisation criteria was as follows: whole array CV ≤ 25%; probe replicate CV ≤ 20%; noise-threshold ≥ 2 SD above background; maximum threshold = 65355 nm; whole array filtering control = 15 ng/µl Cy5-biotin-BSA. Batch-to-batch normalization was performed as described in Section 5.2. Re-assaying was repeated only twice after failing visual and computation filtering criteria. Here, 99% (*N* = 300/304) of the samples passed the computation quality control procedure; thus, samples **2B**, **4B**, **13D** and **13E** were excluded from all further analysis.

The patient therapy targets the immunomodulatory checkpoint inhibitors: PD-1 and CTLA-4. A subgroup of the patients received BRAF inhibitor treatment prior to immunotherapy, although the patient identities are unknown. Also, the treatment regimen varied, depending on patient response. Thus, the cohort consists of those patients who received one treatment-type only (e.g. Ipilimumab only), those that received differing consecutive treatments (e.g. Ipilimumab then Pembrolizumab), and those that received co-treatments (e.g. Ipilimumab and Pembrolizumab). For each patient, a serum and/plasma sample was collected before receiving immunotherapy, with single or multiple sera and/or plasma samples collected after treatment, amounting to 304 samples.

Due to the on-going nature of the study and the time of writing this thesis, clinical patient information was limited to immunotherapy undergone, which was used for subsequent data analysis. The treatment regimen of patients **26** and **33** were not disclosed, and these patients were consequently excluded from treatment-type analysis. Patients and their corresponding treatment-day and -type are summarised in Table 5.3 to Table 5.10.

Table 5.3. Summary of melanoma patients 1 to 7 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days trial initiation	Immunotherapy treatment
1	a – 0	Ipilimumab
	b – 21	Ipilimumab
	c – 46	Ipilimumab
	d – 67	Ipilimumab
2	a – 0	Ipilimumab
	b – 20	Ipilimumab
	c – 42	Ipilimumab
	d – 63	Ipilimumab
3	a – 0	Ipilimumab
	b – 23	Ipilimumab
	c – 63	Ipilimumab
	d – 189	No treatment
	e – 252	No treatment
4	a – 0	Ipilimumab and Pembrolizumab
	b – 42	Ipilimumab and Pembrolizumab
	c – 86	Pembrolizumab
	d – 126	Pembrolizumab
5	a – 0	Ipilimumab
	b – 20	Ipilimumab
	c – 49	Ipilimumab
6	a – 3 days after BCG	Ipilimumab
	b – 21	Ipilimumab
	c – 75	Ipilimumab
7	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 64	Ipilimumab
	d – 134	No treatment
	e – 188	No treatment
	f – 286	No treatment
	g – 496	No treatment
	h – 509	No treatment
	i – No data	Ipilimumab
	j – No data	Ipilimumab
	k – No data	Ipilimumab

Table 5.4. Summary of melanoma patients 8 to 11 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
8	a – 0	Ipilimumab
	b – 42	Ipilimumab
	c – 63	Ipilimumab
	d – 0	Nivolumab
	e – 32	Nivolumab
	f – 74	Nivolumab
	g – No data	Nivolumab
9	a – 0	Nivolumab
	b – 28	Nivolumab
	c – 56	Nivolumab
	d – 84	Nivolumab
	e – 112	Nivolumab
	f – 140	Nivolumab
	g – 182	Nivolumab
	h – 210	Nivolumab
	i – 413	Nivolumab
10	a – 0	Ipilimumab
	b – 43	Ipilimumab
	c – 64	Ipilimumab
	d – no data	No treatment
	e – 253	No treatment
	f – 0	Pembrolizumab
	g – 22	Pembrolizumab
	h – 43	Pembrolizumab
	i – 64	Pembrolizumab
	j – 858	Pembrolizumab
	k – 148	Pembrolizumab
	l – 169	Pembrolizumab
	m – 190	Pembrolizumab
	n – 232	Pembrolizumab
	o – 253	Pembrolizumab
	p – 462	No treatment
	q - 491	No treatment
11	a – 0	Ipilimumab
	b – 43	Ipilimumab
	c – 0	Nivolumab
	d – 29	Nivolumab
	e – data	Nivolumab

Table 5.5. Summary of melanoma patients 12 to 16 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
12	a – 0	Pembrolizumab
	b – 22	Pembrolizumab
	c – 43	Pembrolizumab
	d – 64	Pembrolizumab
	e – 85	Pembrolizumab
	f – 169	Pembrolizumab
	g – 253	Pembrolizumab
	h – 337	Pembrolizumab
	i – 421	Pembrolizumab
	j – 505	Pembrolizumab
	k – 589	Pembrolizumab
	l – 673	Pembrolizumab
13	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 43	Ipilimumab
	d – 64	Ipilimumab
	e – 85	Ipilimumab
	f – 169	No treatment
	g – 253	No treatment
14	a – 0	Ipilimumab
	b – 8	Ipilimumab
	c – 21	Ipilimumab
	d – 78	Ipilimumab
	e – No data	Nivolumab
	f – No data	Nivolumab
	g – No data	Nivolumab
	h – No data	Nivolumab
	i – No data	Nivolumab
15	a – 0	Pembrolizumab
	b – 21	Pembrolizumab
	c – 42	Pembrolizumab
	d – 63	Pembrolizumab
	e – 84	Pembrolizumab
	f – 168	Pembrolizumab
16	a – 0	Pembrolizumab
	b – 29	Pembrolizumab
	c – 56	Pembrolizumab
	d – 85	Pembrolizumab
	e – 169	Pembrolizumab
	f – 0	Ipilimumab
	g – 20	Ipilimumab
	h – 66	Ipilimumab

Table 5.6. Summary of melanoma patients 17 to 24 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
17	a – 0	BCG
	b – 56	Ipilimumab
	c – 82	No treatment
	d – 120	No treatment
18	a – 0	Ipilimumab
	b – 43	Ipilimumab
	c – 64	Ipilimumab
	d – 0	Nivolumab
	e – 72	Nivolumab
19	a – 0	Pembrolizumab
	b – 28	Pembrolizumab
	c – 56	Pembrolizumab
	d – 84	Pembrolizumab
	e – 168	Pembrolizumab
20	a – 0	Ipilimumab
	b – 29	Ipilimumab
	c – 71	Ipilimumab
	d – 106	No treatment
	e –190	No treatment
	f – 358	No treatment
	g – 442	No treatment
	h – 533	No treatment
	i – 617	No treatment
	j – 701	No treatment
	k – No data	No treatment
21	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 64	Ipilimumab
22	a – 0	Pembrolizumab
	b – 21	Pembrolizumab
23	a – 0	Ipilimumab
	b – 21	Pembrolizumab
24	a – 0	Pembrolizumab
	b – 21	Pembrolizumab
	c – 42	Pembrolizumab
	d – 62	Pembrolizumab
	e –191	No treatment
	f – 225	No treatment
	g – 253	No treatment
	h – 281	No treatment
	i – 323	No treatment

Table 5.7. Summary of melanoma patients 25 to 36 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
25	a – 0	Ipilimumab
	b – 44	Ipilimumab
27	a - 0	Ipilimumab
	b – 29 (1 week after second treatment)	Ipilimumab
	c – 43	Ipilimumab
28	a – 0	Pembrolizumab
	b – 28	Pembrolizumab
	c – 56	Pembrolizumab
	d – 153	Pembrolizumab
	e –237	Pembrolizumab
29	a – 0	Pembrolizumab
	b – 29	Pembrolizumab
	c – 55	Pembrolizumab
	d – 83	Pembrolizumab
30	a – 0	Ipilimumab
	b – 64	Ipilimumab
	c – 85	No treatment
	d – 175	No treatment
31	a – 22 days before day 0	No treatment
	b – 24	Ipilimumab
	c – 41	Ipilimumab
	d – 62	Ipilimumab
32	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 43	Ipilimumab
	d – 0	Nivolumab
	e –29	Nivolumab
	f - 71	Nivolumab
34	a – 0	Nivolumab
	b – 28	Nivolumab
	c – No data	Nivolumab
35	a – 0	Pembrolizumab
	b – 64	Pembrolizumab
	c – 85	Pembrolizumab
	d – 169	Pembrolizumab
36	a – 0	Ipilimumab and Nivolumab
	b – 44	Ipilimumab and Nivolumab
	c – 78	No treatment
	d – 218	No treatment
	e –264	No treatment

Table 5.8. Summary of melanoma patients 37 to 45 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
37	a – 0	Nivolumab
	b – 28	Nivolumab
	c – 70	Nivolumab
	d – 154	Nivolumab
	e –No data	Nivolumab
38	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 43	Ipilimumab
	d – 73	No treatment
	e – 157	No treatment
39	a – 0	Ipilimumab
	b – 57	Ipilimumab
	c – 99 (no treatment)	No treatment
	d – No data	No treatment
	e – No data	No treatment
40	a – 0	Ipilimumab
	b – 43	Ipilimumab
	c – 64	Ipilimumab
	d – 86	No treatment
	e – 126	No treatment
	f – 0	Nivolumab
	g – 72	Nivolumab
41	a – 0	Ipilimumab
	b – 43	Ipilimumab
42	a – 0	Ipilimumab
	b – 21	Ipilimumab
43	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 43	Ipilimumab
	d – 64	Ipilimumab
	e – 85	No treatment
	f – 99	No treatment
44	a – 0	Ipilimumab
	b – 20	Ipilimumab
	c – 44	Ipilimumab
45	a – 0	Nivolumab
	b – 29	Nivolumab
	c – 71	Nivolumab
	d – 23	Nivolumab

Table 5.9. Summary of melanoma patients 46 to 51 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
46	a – 0	Pembrolizumab
	b – 57	Pembrolizumab
	c – 85	Pembrolizumab
	d – 169	Pembrolizumab
	e –253	Pembrolizumab
	f – 337	Pembrolizumab
	g – 421	Pembrolizumab
	h – 505	Pembrolizumab
	i – 589	Pembrolizumab
	j – 673	Pembrolizumab
47	a – 0	Pembrolizumab
	b – 106	Pembrolizumab
	c – No data	No treatment
48	a – 0	Ipilimumab and Nivolumab
	b – 22	Ipilimumab and Nivolumab
	c – 59	Ipilimumab and Nivolumab
	d – 135	Nivolumab
	e –149	Nivolumab
	f – 177	No treatment
49	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 44	Ipilimumab
	d – 64	Ipilimumab
	e – 211	No treatment
	f – 295	No treatment
	g – 407	No treatment
	h – 483	No treatment
50	a – 0	Ipilimumab
	b – 21	Ipilimumab
	c – 79	Ipilimumab
	d – 98	Ipilimumab
51	a – 0	Pembrolizumab
	b – 42	Pembrolizumab
	c – 63	Pembrolizumab
	d – 145	Pembrolizumab

Table 5.10. Summary of melanoma patient 52 before and after treatment with checkpoint inhibitors at various TPs. Abbreviations include: TP = timepoints.

Patient	Timepoint - days post-treatment	Immunotherapy treatment
52	a – 0	Ipilimumab
	b – 22	Ipilimumab
	c – 43	Ipilimumab
	d – 71	Ipilimumab
	e – 92	No treatment
	f – 106	No treatment
	g – 190	No treatment
	h – 236	No treatment
	i – 259	No treatment
	j – 315	No treatment
	k – 0	Pembrolizumab
	l – 22	Pembrolizumab
	m – 43	Pembrolizumab
	n – 64	Pembrolizumab
	o – 125	Pembrolizumab
	p – 146	Pembrolizumab
	q – 187	Pembrolizumab

The patients were assessed based on the treatment-type undergone. For the analysis, the treatment-induced effects were assessed on the CT100plus microarray by comparing plasma/serum autoantibody levels before treatment (TP0) and the first sample collected thereafter (TP1). The Shapiro-Wilk normality test was used to assess data distribution (p-value ≤ 0.05), which indicated a non-normal distribution. Thus, non-parametric tests were performed when comparing TP0 and TP1, which included Mann-Whitney *U* test (Mann and Whitney, 1947) and Benjamini-Hochberg (Benjamini and Hochberg, 1995) *post hoc* correction test.

Longitudinal assessments of autoantibody titres were used to determine the long-term changes in autoantibody in response to immunotherapy. Longitudinal assessments of autoantibody titres were performed using the Kruskal-Wallis test (Kruskal and Wallis, 1952) for multiple samples and Benjamini-Hochberg *post hoc* correction test. The Mann-Whitney *U*, Kruskal-Wallis and Benjamini-Hochberg tests were performed in R (version 3.3.0) using the PMCMR package (Maintainer and Pohlert, 2016)

All treatment-types

The combinatorial immunomodulatory effects of all immunotherapies were assessed by comparing TP0 and TP1 for all patients on the CT100plus microarray. The Mann-Whitney *U* test results indicate a significant difference in signal intensity which became non-significant after the Benjamini-Hochberg *post hoc* test for ROPN1 (*p*-value = 0.03, adjusted *p*-value = 0.72), AKT1 (*p*-value = 0.03, adjusted *p*-value = 0.72), and DPPA (*p*-value = 0.05, adjusted *p*-value = 0.72). This finding is supported by ROC analysis between TP0 and TP1, performed in RStudio using the ROCR package. The ROC analysis was performed for antigens that produced the lowest p-values, including ROPN1, which produced AUC-, sensitivity-, and specificity-values of 0.38, 0.50 and 0.38, respectively; AKT1, which produced AUC-, sensitivity-, and specificity-values of 0.38, 0.24 and 0.73; and DPPA2, which produced AUC-, sensitivity-, and specificity-values of 0.39, 0.34 and 0.62. The ROC curves for ROPN1, AKT1 and DPPA2 are shown in Figure 5.12.

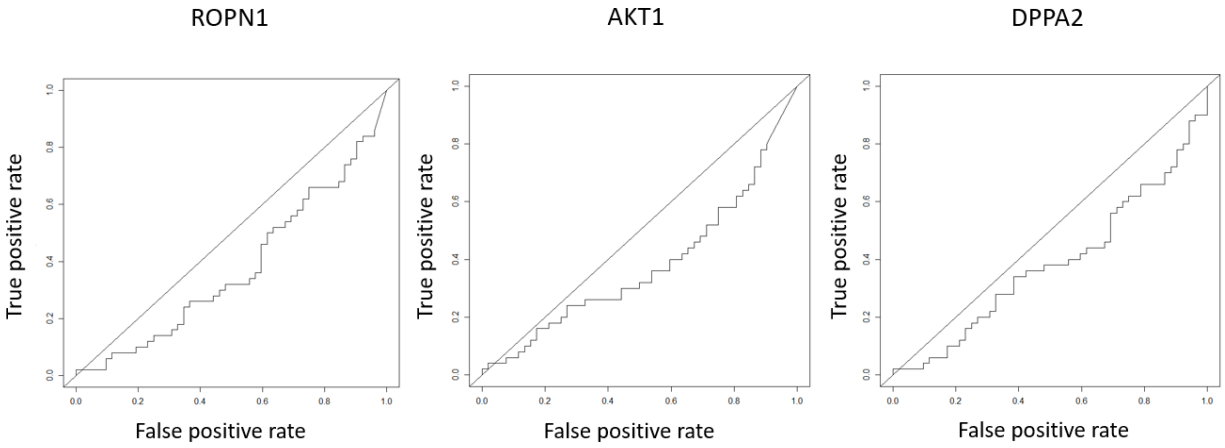


Figure 5.12. ROC curves for ROPN1, AKT1 and DPPA2. Melanoma patient serum/plasma before (TP0) and after (TP0) immunotherapy were assayed on the CT100plus microarray to determine changes of autoantibody levels against CT100plus antigens. A non-significant change (adjusted *p*-value > 0.05, Benjamini-Hochberg) in autoantibody levels were detected. This result was supported by ROC analysis for the 3 proteins with the lowest *p*-values, to include ROPN1, AKT1 and DPPA2, which produced AUC-values of 0.38, 0.38 and 0.39, respectively. ROC = receiver operator characteristic curve; AUC = area under the ROC curve.

Although a non-significant change in autoantibody titres was detected at various TPs, a subset of patients produced high autoantibody signal intensities (> 10 000 RFU) toward autoantigens. Only proteins detected in more than one patient are reported: For TP0, seven (*N* = 7/52; 14%) patients produced a high (> 10 000 RFU) signal intensity to at least one antigen, including CTAG2 (*N* = 6/52, 12%), NY-ESO-1 (*N* = 6/52, 12%), and PRKCZ (*N* = 2/52, 4%). For TP1, eleven (*N* = 11/50, 22%) patients produced a high (> 10 000 RFU) signal intensity to at least one antigen including CTAG2 (*N* = 6/50, 12%), NY-ESO-1 (*N* = 6/50, 12%), MAGEB6 (*N* = 2/50, 4%), NXF2 (*N* = 2/50, 4%), NY-CO-45 (*N* = 2/50, 4%), ACVR2B (*N* = 2/50, 4%), and CTAG2/LAGE-1b/LAGE-1L (*N* = 2/50, 4%).

Taken together, it is evident that high (> 10 000 RFU) signal was detected for CTAG2 and NY-ESO-1 regardless of treatment, although PRKCZ was detected in 2 patients before treatment, whereas signals for MAGEB6, NXF2, NY-CO-45, ACVR2B and CTAG2/LAGE-1b/LAGE-1L were each detected in 2 patients after treatment.

Treatment with Nivolumab

Nivolumab is an anti-PD-1 cancer immunotherapeutic. Of the 52 patients assessed, 4 patients were treated with nivolumab alone. Six other patients received ipilimumab treatment prior to nivolumab treatment, and were thus excluded from the analysis due to the possible immune modulating effects of ipilimumab.

Nivolumab treatment-induced effects were assessed on the CT100plus microarray by comparing plasma/serum autoantibody levels before treatment (TP0) and the first sample collected thereafter (TP1). Here, we observed non-significant (*p*-value > 0.05) changes in autoantibody levels after nivolumab treatment. Together, these results suggest that anti-PD-1-treatment does not affect autoantibody titres toward proteins on the CT100plus microarray. Longitudinal assessments of autoantibody titres performed also indicate no significant (*p*-value > 0.05; adjusted *p*-value > 0.05) difference in autoantibody titres between the different TPs for patients treated with nivolumab.

As expected, most patients produced low signal (200 – 2000 RFU) toward antigens across various TPs, although patient **9** produced a high (> 10 000 RFU) signal intensity against DDX53.

Treatment with Pembrolizumab

Pembrolizumab is another anti-PD-1 cancer immunotherapeutic. Of the 52 patients assessed, 12 patients were treated with pembrolizumab alone. Three other patients received ipilimumab treatment prior to pembrolizumab treatment, and were thus excluded from the analysis due to the possible immune modulating effects of ipilimumab. The treatment-induced effects were assessed on the CT100plus microarray by comparing plasma/sera autoantibody titres before treatment (TP0) and the first sample collected thereafter (TP1). Here, we observed non-significant (p-value > 0.05) changes in autoantibody titres after pembrolizumab treatment. Together, these results again suggest that anti-PD1-treatment does not affect autoantibody titres toward proteins on the CT100plus microarray. Longitudinal assessments of autoantibody titres also indicate no significant (p-value > 0.05; adjusted p-value > 0.05) difference in autoantibody titres between the different TPs for patients treated with pembrolizumab.

As expected, most patients produced low signal toward antigens across various TPs, whereas a subset of patients produced high signal (> 10 000 RFU) at various TPs toward specific antigens. Six (N = 6/12; 50%) patients produced a high signal intensity to at least one antigen, for at least one TP, although the antigens were not associated with pre-treatment or post-treatment, or treatment duration. The proteins detected with a high signal included CTAG2 (N = 3/12, 25%), NY-ESO-1 (N = 3/12, 25%), SPANXA1 (N = 3/12, 25%), CTAG2/LAGE-1b/LAGE-1L (N = 2/12, 17%), DDX53 (N = 1/12, 8%), MAGEA1 (N = 1/12, 8%), MAGEA4v4 (N = 1/12, 8%), MAGEB1 (N = 1/12, 8%), MAGEB5 (N = 1/12, 8%), MAGEB6 (N = 1/12, 8%).

Treatment with Ipilimumab

Ipilimumab is an anti-CTLA-4 cancer immunotherapeutic. Of the 52 patients assessed, 32 patients were treated with either ipilimumab alone, or prior to treatment with the anti-PD-1 immunotherapeutics, nivolumab or pembrolizumab.

The treatment-induced effects were assessed on the CT100plus microarray by comparing plasma/sera autoantibody titres before treatment (TP0) and the first sample collected thereafter (TP1). A significant (p-value = 0.03) difference in GRWD1 autoantibody titres after treatment was detected, although this was to be a non-significant (adjusted p-value = 0.77) result after the *post*

hoc correction test. A ROC analysis was performed for GRWD1, which produced AUC -value of 0.38. The ROC curves for GRWD1 is shown in Figure 5.13. Together, these results indicate that anti-CTLA-4-treatment does not affect autoantibody titres toward proteins on the CT100plalso us microarray. Longitudinal assessments of autoantibody titres for anti-CTLA-4-treated patients indicate no significant (*p*-value > 0.05; adjusted *p*-value > 0.05) difference in autoantibody titres between the different TPs for patients treated with Ipilimumab.

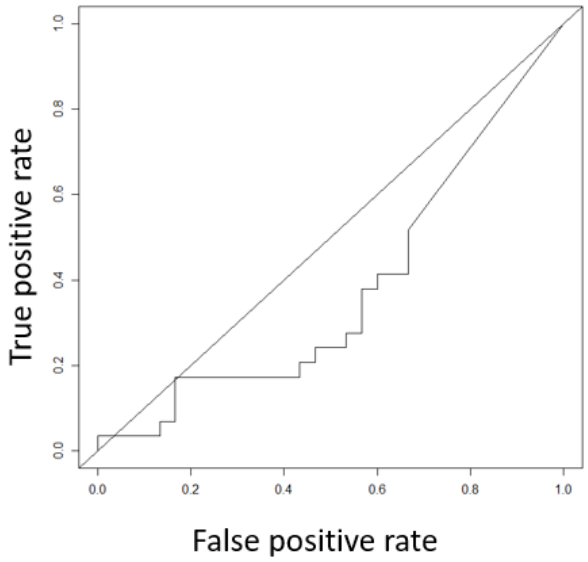


Figure 5.13. ROC curves for GRWD1. The serum/plasma of ipilimumab-treated melanoma patients before (TP0) and after (TP0) immunotherapy were assayed on the CT100plus microarray to determine changes of autoantibody levels against CT100plus antigens. A non-significant change (adjusted *p*-value > 0.05, Benjamini-Hochberg) in autoantibody levels were detected. This result was supported by ROC analysis for the proteins with the lowest *p*-value, i.e. GRWD1, which produced an AUC-value of 0.38.

Although a non-significant change in autoantibody titres was detected for patients at various TPs, some patients produced high (> 10 000 RFU) autoantibody signal intensities toward individual autoantigens. As expected, most patients produced low signal toward antigens across various TPs, whereas a subset of patients produced high signal at various TPs toward specific antigens. Eight (*N* = 8/32; 25%) patients produced a high (> 10 000 RFU) signal intensity to at least one antigen, for at least one TP, although the antigens were not associated with pre-treatment or

post-treatment, or treatment duration. The top 10 proteins detected with a high signal included CTAG2 (*N* = 4/32, 13%), NY-ESO-1 (*N* = 4/32, 13%), CTAG2/LAGE-1b/LAGE1L (*N* = 4/32, 13%), MAGEB6 (*N* = 3/32, 9%), NXF2 (*N* = 3/32, 9%), NY-CO-45 (*N* = 3/32, 9%), PRKCZ (*N* = 3/32, 9%), MAGEB1 (*N* = 2/32, 6%), MAGEB5 (*N* = 2/32, 6%), SPANXA1 (*N* = 2/32, 6%).

Consecutive immunotherapeutic treatment with ipilimumab and then nivolumab

Six (N = 6/52, 12%) patients, including **8**, **11**, **14**, **18**, **32** and **40**, received Ipilimumab treatment followed by nivolumab treatment. To assess the autoantibody response to changed immunotherapeutic treatment, all samples before and after treatment change was compared. The Mann-Whitney *U* test results indicate a significant difference in signal intensity which became non-significant after the Benjamini-Hochberg *post hoc* test for p53 C141Y (*p*-value = 0.03, adjusted *p*-value = 0.40), MAGEB6 (*p*-value = 0.04, adjusted *p*-value = 0.4), DPPA3 (*p*-value = 0.04, adjusted *p*-value = 0.4), SPAG9 (*p*-value = 0.04, adjusted *p*-value = 0.4), and CEP290 (*p*-value = 0.04, adjusted *p*-value = 0.4).

Although a non-significant change in autoantibody titres was detected at various TPs, a subset of patients produced high autoantibody signal intensities (> 10 000 RFU) toward autoantigens. When patients were treated with Ipilimumab only, two (*N* = 2/6, 33%) patients produced a high (> 10 000 RFU) signal intensity to at least one antigen. The top 4 proteins detected with a high signal included CTAG2 (*N* = 1/6, 17%), NY-ESO-1 (*N* = 1/6, 17%), ACVR2B (*N* = 1/6, 17%), and CTAG2/LAGE-1b/LAGE-1L (*N* = 1/6, 17%).

After patients received nivolumab treatment, two (*N* = 2/6, 33%) patients produced a high (> 10 000 RFU) signal intensity to at least one antigen. The top 10 proteins detected with a high signal included CTAG2 (*N* = 4/6, 67%), NY-ESO-1 (*N* = 4/6, 67%), NY-CO-45 (*N* = 4/6, 67%), NXF2 (*N* = 3/6, 50%), p53 S15A (*N* = 2/6, 33%), CTAG2/LAGE-1b/LAGE-1L (*N* = 2/6, 33%), MAGEB6 (*N* = 1/6, 18%), p53 S6A (*N* = 1/6, 18%), p53 T18A (*N* = 1/6, 18%), and p53 S392A (*N* = 1/6, 18%).

Taken together, it is evident that high (> 10 000 RFU) signal was detected for CTAG2 and NY-ESO-1 regardless of treatment undergone. Furthermore, PRKCZ was detected in only 2 patients before treatment, whereas MAGEB6, NXF2, NY-CO-45, ACVR2B and CTAG2/LAGE-1b/LAGE-1L were each detected in 2 patients after treatment. Only 1 patient received pembrolizumab treatment

followed by Ipilimumab treatment, whereas 2 patients received Ipilimumab treatment followed by pembrolizumab treatment, which were omitted from further analysis due to poor statistical power.

Three patients received co-treatment of ipilimumab with nivolumab or ipilimumab with pembrolizumab, although again due to the small samples size, statistical analyses could not be performed.

Discussion and conclusion

In total 52 melanoma patients, amounting to 304 sera and/or plasma samples, were assessed on the CT100plus microarray. Quality control tests were performed to ensure that conditions were optimal to produce high quality microarrays.

Due to the large cohort size, the samples were assayed in two separate batches using two separate microplate sources. Furthermore, to improve microarray image and data quality, images that failed visual and CT100plus software assessments were re-assayed in two separate batches. Thus, the resulting data are derived from four sample batches, each referred to as batch **B1**, **B2**, **B3** or **B4**. From the volcano plot, it was evident that all proteins were dysregulated between batches (Figure 5.5). After batch-to-batch normalisation, the volcano plots indicate no significant (adjusted *p*-value > 0.05, fold-Δ > 0.25) difference between the batches (Figure 5.7). After the data was normalised, the remaining statistical tests were performed.

ROC analysis was employed to analyse CT100plus microarray data from 301 CRC patients and 14 HCs samples, where further analysis into biomarker potential was performed using 10 proteins with the highest AUC-values. It is important to note that the HC samples were neither age- nor gender-matched to the CRC patients. Furthermore, although neither of the HCs were previously diagnosed with cancer, tests were not performed to confirm that the HCs were CRC-free. However, since our intention here is to acquire preliminary results for a potentially larger study, we decided to use the plasma samples available to assess autoantibody profiles between CRC patient and HCs.

The top 10 proteins include CEACAM 1, DPPA2, FGFR2, ITGB1, MAGEA10, NANOG, PIM1, SPANXB1, THEG and XAGE1B; these proteins produced AUC-, sensitivity- and specificity-values ranging from 0.79 – 0.87, 0.77 – 1.00 and 0.64 – 0.79, respectively (Table 5.1). A combinatorial ROC analysis of the top 10 proteins produced AUC-, sensitivity- and specificity-values of 0.92, 0.93 and 0.89, respectively. Melanoma Inhibitory Activity (MIA) serological marker is reportedly increased in the serum of patients with metastatic uveal melanoma when compared to progression-free patients; levels of MIA were previously assessed in a cohort of 503 uveal melanoma patients and analysed by ROC analysis, producing an AUC level of 0.84 (Klingenstein

et al., 2011). Together, these results indicate that the top 10 proteins identified with the CT100plus microarray produced AUC-, sensitivity- and specificity-values in the same range as the best reported biomarker, whereas combinatorial ROC results outperformed individual and literature-reported potential biomarkers. Furthermore, Kruskal-Wallis and Benjamini-Hochberg tests indicated a significant difference in antigen signals between melanoma patients and HCs, producing *p*-values and adjusted *p*-values less than 0.05 (Table 5.2). The hierarchical clustering and PCA results display two segregating clusters, each representing melanoma patients or HCs, although overlap is evident between **HC1**, **HC9** and **HC10** and melanoma patients. Together, these results indicate that the top 10 proteins may be potential CRC biomarkers, although alternate methods e.g. validating with another melanoma cohort or using tumour tissues for histological confirmation, is required for biomarker validation.

The 3 proteins with the highest AUC-values, including CEACAM1, FGFR2 and PIM1, were further investigated:

CEA, also known as CEACAM (http://www.uniprot.org/uniprot/Q13984), is a widely-used tumour biomarker in melanoma and CRC. In malignant melanoma, surface expression of CEACAM1 correlates with tumour progression and poor survival (Markel et al., 2010; Sivan et al., 2012). Secreted CEACAM1 later emerged as a potential biomarker of malignant melanoma; an increase in serum levels is associated with disease progression and lower serum levels positively correlate with treatment outcome (Sivan et al., 2012; Ortenberg et al., 2015). CEACAM is also reportedly involved in the melanomas ability to evade immune attacks (Sivan et al., 2012). Although autoantibody responses to CEACAM1 are not reported for malignant melanoma, its increased expression on the cell surface of melanoma and release into patient blood could induce an autoantibody response detectable on the CT100plus microarray.

Fibroblast growth factor receptor 2, also known as FGFR2 (http://www.uniprot.org/uniprot/P21802), is a tyrosine-protein kinase that acts as a cell surface receptor for fibroblast growth factors, and plays an essential role in the regulation of cell proliferation, differentiation, migration, apoptosis, and in the regulation of embryonic development. FGFR2 expression with loss-of-function mutations is reported in melanocytes

(Diez de Medina et al., 1997; Zhang et al., 2001; Naimi et al., 2002; Gartside et al., 2009; Amann et al., 2010). A study of 47 melanoma cell-lines identified mutations in FGFR2, but not for FGFR1, for 12% of the tested cell-lines, and in 12% metastatic melanomas (Gartside et al., 2009). Another study concluded that FGFR2 genetic variation was identified in melanoma and non-melanocytic skin cancer, although there was no association between the genetic variations and risk of disease (Nan et al., 2009). The CT100plus microarray contains a non-mutated form of the FGFR2 protein; we would therefore not expect to detect autoantibodies directed at mutated epitopes of the protein. Instead, an autoantibody response toward FGFR2 overexpression may be detected. Although FGFR2 overexpression is not reported for melanoma, neoantigen-specific $CD4^+$ T cells may drive affinity maturation of B-cells that have taken up and processed the mutated FGFR antigen, with the resulting autoantibody response being directed toward a wild-type epitope.

Provirus integration site for moloney murine leukaemia virus, also known as PIM1 (http://www.uniprot.org/uniprot/P11309), is a proto-oncogene with serine/threonine kinase activity involved in cell survival and cell proliferation, thus providing a selective advantage in tumourigenesis. PIM1 is expressed in melanoma tumours pre- and post-treatment. *In vivo* studies concluded that PIM contributes to melanoma cell proliferation and tumour growth, where PIM inhibition with SGI-1776 reduced melanoma proliferation and survival in preclinical models of melanoma, indicating that PIM inhibitors are promising additions to targeted therapies available to melanoma patients (Shannan et al., 2016). PIM1 has also been shown to be a direct target of miR-542-3p, a microRNA which is a tumour suppressor in various cancers, including melanoma. miR-542-3p exogenous expression inhibited cell migration, invasion and epithelial-mesenchymal transition (EMT) *in vitro* and lung metastasis *in vivo*. Overexpression of PIM1 partially rescued miR-542-3p-mediated suppression of cell migration, invasion and EMT (Rang et al., 2016). Although PIM1 RNA has been reported in skin, protein expression is not detected, thus the presence of PIM1 in melanoma likely results in the autoantibody response that was detected by the CT100plus microarray.

Both PD-1 and CTLA-4 prevent autoimmunity, and their effects are reportedly executed at different stages of the immune response. Blockade of CTLA-4 impacts T-cell activation in the lymph nodes when CTLA-4 expressing T_{regs} remove B7 from the surface of APCs, limiting their

ability to stimulate tumour-specific T-cells (Qureshi et al., 2011). The anti-CTLA-4 can also take effect on exhausted CTLA4-4-expressing T-cells and T_{regs} accumulated within the tumour microenvironment (Curran et al., 2010; Salvi et al., 2012). Blocking with anti-PD-1 is thought to mainly take place at the effector stage of the immune response as the anti-PD-1 blocks the interaction between PD-1 on T-cell and PD-L1 on cancer cells. Since B-cells, the only cell type that produces Igs, also express PD-1, treatment with anti-PD-1 may directly increase Ab production in response to TAA.

The CT100plus microarray was used to determine the immunomodulatory effects of all immunotherapy treatment. Here, the autoantibody levels of all patients, regardless of treatment-type, were compared for TP0 and TP1. The Mann-Whitney *U* test indicates a non-significant (*p*-value > 0.05) change in autoantibody titre post-treatment. For all treatment-types in the melanoma cohort, an increased T-cell and B-cell response is expected in response to immunotherapy. However, our results indicate that none of the 123 proteins on the CT100plus microarray correlates with treatment type when assessing outcomes using the Kruskal-Wallis test. A power analysis for the Kruskal-Wallis test was performed in R with the pwr package. For the power analysis, an effect size of 0.8, a significance level of 0.05, and a power of 0.95 was used. The results indicate that a subset of patient within each treatment-type did not have enough timepoints to reach the required statistical power (Table 5.11). It would therefore be important to capture more timepoints for these patients in this on-going trial.

Table 5.11. Power analysis for treatment-types for melanoma patient cohort. A power analysis was performed for the melanoma cohort to determine the number of patients required to reach statistical power for each treatment group. Here, the effect size was set at 0.8, the significance level was set at 0.05, and the power was set at 0.95. Abbreviations include: *N* = number of patients; TP = timepoints.

Treatment	*N*	Number of TP required	Patients with sufficient TP
Nivolumab	4	7.77	1
Pembrolizumab	12	4.17	7
Ipilimumab	32	2.56	27
Ipilumab-Nivolumab	6	6.16	4

Although a non-significant change in autoantibody titres was detected at various TPs, a subset of patients produced high autoantibody signal intensities toward autoantigens. Interestingly, seven (*N* = 7/52; 14%) patients had a high (> 10 000 RFU) signal intensity at T0 to at least one antigen, including CTAG2 (*N* = 6/52, 12%), NY-ESO-1 (*N* = 6/52, 12%), and PRKCZ (*N* = 2/52, 4%). Furthermore, at T1, eleven (*N* = 11/50, 22%) patients produced a high (> 10 000 RFU) signal intensity to CTAG2 (*N* = 6/50, 12%), NY-ESO-1 (*N* = 6/50, 12%), MAGEB6 (*N* = 2/50, 4%), NXF2 (*N* = 2/50, 4%), NY-CO-45 (*N* = 2/50, 4%), ACVR2B (*N* = 2/50, 4%), or CTAG2/LAGE-1b/LAGE-1L (*N* = 2/50, 4%). These results indicate that a subset of the patient cohort have a change in autoantibody signature which may result from an autoantibody response towards TAAs released from tumours after immunotherapy. It would therefore be interesting to follow up with these patients to determine whether the identification of autoantibody signatures relates to patient treatment outcome and recurrence.

High signals were detected for CTAG2 and NY-ESO-1 before and after treatment, regardless of treatment-type. However, due to the low detection frequency, which is supported by statistical tests outcomes, it is highly unlikely that these proteins could reproducibly detect melanoma or be used for disease prognosis. Although CTAG is not reportedly expressed in melanoma, its homologue NY-ESO-1 is reportedly expressed at high levels in metastatic (28 - 32 %) *versus* primary (0 - 16 %) malignant melanoma tissues samples. NY-ESO1 is also found to be associated with cancer state, primary tumour thickness, CD3+ T-cell tumour infiltration, although no association was found with patient survival, gender, age, tumour site, ulceration or lymph node sentinel status (Velazquez et al., 2007; Aung et al., 2014; Giavina-Bianchi et al., 2015). Our cohort consists of only metastatic melanoma patients, six (*N* = 6/52, 12%) of which produced high levels of NY-ESO-1. Here, we did not find a correlation between NY-ESO-1 levels and patient treatment regimen. However, these results may correlate with other patient clinical features that will be obtained in future when the clinical trial un-blinds. NY-ESO1 and CTAG2 are also frequently detected together, which is also likely due to a common epitope targeted by melanoma autoantibodies, as these proteins are spliced variants (Table 5.12).

Table 5.12. Shared homology between frequently detected melanoma antigens. Melanoma sera and/or plasma were assayed in the CT100plus microarray. Frequently detected antigens with high signal intensity (> 10 000 RFU) were identified, and the protein that shares high homology is summarised with query coverage (%), identity (%), and E-values. Abbreviations include: RFU = relative fluorescence units.

Protein	Protein	Query coverage (%)	Identity (%)	E-value
NY-ESO-1	CTAG2	78	77	1^{-38}

In conclusion, using the CT100plus microarray, we identified a list of 10 proteins that distinguish HCs from metastatic melanoma patients. Furthermore, we observed high and distinct signals for a subgroup of patients, although we have not yet found an association with patient treatment regimen. For future investigations, we will be looking in to possible correlations between other patient clinical features, and looking in to the autoantibody profiles of early stage melanoma patients for the identification of diagnostic biomarkers for melanoma.

Chapter 6

Summary, Conclusions and Future Prospectives

6.1. Summary and conclusions

Cancer is a heterogenous disease capable of forming and spreading in most tissues of the human body. Cancer screening and diagnosis can be performed through medical procedures, which are typically highly invasive and require an intensive infrastructure e.g. for CRC; protein biomarkers, which have low specificity resulting in a high false positive rate, e.g. prostate-specific antigen (PSA) for prostate cancer; and medical imaging, which exposes patients to harmful radiation and may not be effective in some cases e.g. mammographies are unsuccessful for women with dense breasts. It is therefore important to create cost-effective, non-invasive cancer diagnostic tools that also give an indication of disease prognosis. With these concepts in mind, the Blackburn lab created the CT100plus microarray, a tool that detects autoantibody signals from the blood of cancer patients, detectable in the pg/ml range. The CT100plus microarray is functionalised with native TSA and TAA, which include CTA. Thus, the autoantibody signals detected for patients are specific to aberrant cancer proteins, and should therefore be detected in cancer patients only (Beeton-Kempen et al., 2014). Despite the technological advancements of microarrays, autoantibody signals are only identified in a subset of the patients, although the reason for this is not entirely clear.

The data from the CT100plus microarray is filtered and normalised through a robust in-house developed pipeline built into the CT100plus software. However, the statistical analysis thereafter has been limited to using an arbitrary 1000 RFU cut-off, and only antigenic signal intensities above this threshold were considered significant. In our study, we re-analysed the CRC patient-derived CT100plus data using a newly established statistical pipeline (Figure 2.1), including ROC analyses, Mann-Whitney *U* test, unsupervised hierarchical clustering and PCA analysis. Using this pipeline, we identified CEACAM1 and GRWD1 as our top antigens for CRC diagnosis, together producing sensitivity-, and specificity-, and AUC-values of 1.00, 0.77 and 0.94, respectively, and

distinct clusters of HCs and CRC patients for both unsupervised hierarchical clustering and PCA analysis. CEACAM1 is a known serological biomarker for CRC, and its presence in blood could induce an autoimmune response. Furthermore, COL6A1, THEG and CEACAM7, a homologue of CEACAM1, were also identified by AP-MS (See Chapter 2), providing additional supporting evidence for biomarker identification. Unfortunately, we did not observe antigen association with clinicopathological features. Taken together, we have generated a statistical pipeline that distinguishes CRC patients from HCs, which was applied to subsequent cohorts (Chapter 3 and 5).

PAD2 is expressed in CRC, and citrullinated cancer-associated proteins have been identified for CRC, which we have also identified by AP-MS (See Chapter 4). The CT100plus microarray content was modified to include citrullinated proteins, with the subsequent assessment of CRC patient autoantibody response. Significantly (p-value ≤ 0.05; adjusted p-value ≤ 0.05) higher signal intensities were detected in CRC patients *versus* HCs for citrullinated CDK7, MAGEB1, MAGEB5, MAGEB6 and SYCP1, whereas no significant (p-value > 0.05) difference in autoantibody signal was detected on the non-citrullinated microarray for the same patient cohort. Furthermore, ROC analyses of these antigens resulted in AUC-, sensitivity- and specificity-ranges of 0.91, 0.87 and 0.89, respectively Unsupervised hierarchical clustering analyses did not produce distinct clusters of patients and HCs, although the PCA results indicate distinct groups of CRC patient and HC clusters. Together, here we show for the first time that cancer patients elicit an autoantibody response to citrullinated proteins, resulting in potential novel CRC biomarkers.

Two CRC plasma samples were also processed on the SENGENICS IMMUNOME™ microarray functionalized with 1622 native eukaryotic proteins. The highest autoantibody signals were toward proteins not present on the CT100plus microarray. Furthermore, we also developed an AP-MS assay to detect autoantibody responses to autologous native tissue proteins. Although PEG was detected in the spectra, we also detected the proteins or homologues of proteins identified on the CT100plus microarray for the same 5 patients. A high degree of non-specific binding is detected, although, we account for this by using negative controls (*i.e. Plasma* and *Normal*). However, it is important that more patient samples are assessed in future to identify CRC proteins associated with disease diagnosis and prognosis.

Finally, we assayed the serum and/or plasma samples of melanoma patients, and compared autoantibody signals to that of HC samples. CEACAM1 and FGFR2 were identified as the top antigens for melanoma diagnosis, together producing sensitivity-, and specificity-, and AUC-values of 0.96, 0.94 and 0.93, respectively, and distinct clusters of HCs and melanoma patients for both unsupervised hierarchical clustering and PCA analysis. CEACAM1 is a known serological biomarker for melanoma also, and its presence in blood could induce an autoimmune response. Each of the 52 patients received treatment with an immunotherapeutic (nivolimumab, ipilimumab or pembrolizumab) alone, consecutive treatments with two immunotherapeutics, or simultaneous treatment with two immunotherapeutics. Here, we did not observe a significant change in autoantibody levels following immunotherapeutic treatments. Alternatively, response to treatment may be assessed on microarrays with more proteins e.g. on the SENGENICS IMMUNOME™ microarray, or if the autologous tissue samples are available, AP-MS could be used to identify known and novel cancer antigens associated with the response of advanced melanoma patients to therapy.

6.2. Future prospectives

Using two independent immunoproteomics approaches, we show evidence of autoantibody signals against TAAs and TSAs. For future work, obtaining additional HC serum and/plasma samples are required to improve statistical power. In the same regard, it is imperative that a larger cohort be tested for the AP-MS experiments.

It would be useful to establish a user-friendly software application available to analyse CT100plus microarray – or any other microarray - derived data, which can be performed using R or Java. The CT100plus microarray could be developed further to include additional TAAs or TSAs as the microarray layout allows for triplicate printing of 170 proteins.

Finally, it is evident that out robust, highly reproducible, sensitive and specific microarray has potential in a clinical setting. Furthermore, the tool can be used in a research capacity to identify novel diagnostic, disease progression, prognostic, treatment resistant and predictive biomarkers, which could aid in the detection and management of cancer. Our in-house developed microarray platform represents a sensitive, high-throughput and readily customizable means to detect and

quantify macromolecule interactions. Thus, the platform can be modified to assess the autoantibody response in other autoimmune disease e.g. rheumatoid arthritis or Tuberculosis, providing a non-invasive means of disease diagnosis and prognosis.

Chapter 7

Materials and Methods

7.1. Coating NEXTERION® H slides with streptavidin

7.1.1. Preparing streptavidin solution

A vial of 100 mg lyophilised streptavidin (ProSpec) was removed from -20°C storage, equilibrated at room temperature (RT) for 1 hour, and reconstituted in 10 ml slide coating buffer (see Appendix A for preparation) to a final concentration of 1 milligram per millilitre (mg/ml) streptavidin. The solution was stored at -20°C until ready to use.

7.1.2. Coating NEXTERION® H Slides

The enclosed NEXTERION® H slide (Schott) package, and the 1 mg/ml streptavidin solution, was removed from -20°C storage and equilibrated for 1 hour at RT. A 6-chamber slide processing dish was cleaned with triple-distilled Millipore water, and air-dried before adding 5 ml 1 mg/ml streptavidin solution to each chamber. Six NEXTERION® H slides were removed from their packaging, the barcodes were recorded, and each slide transferred to a chamber containing the 1 mg/ml streptavidin solution. The chamber slide processing dish was placed on the orbital shaker (ORBITAL SHAKER SO3, Stuart Scientific) at 100 rotations per minute (RPM) for 1 hour at RT. After coating, the slides were immersed in slide wash buffer (see Appendix A for preparation) for 5 minutes whilst shaking at RT - the wash step was repeated twice. The slides were immersed in triple-distilled Millipore water (high-performance liquid chromatography (HPLC) grade) for 5 minutes at RT, and dried by centrifugation (Megafuge 1.0R, Heraeus) at 272 ×g for 2 minutes at 22°C. Once dry, the slides were re-packaged in a slide holder, sealed in a zip seal bag, and stored at -20°C. This processed was repeated until the required number of slides were derivatized. The remaining streptavidin solution (NB: this solution can be used to coat up to 100 slides) was

combined and transferred to a 50 ml tube (Greiner), sealed with parafilm (Bemis), and stored at -20°C. Containers with uncoated slides were labelled, placed in a zip seal bag and stored at -20°C.

7.1.3. Slide coating quality control

For quality control purposes, a derivatized slide from each batch of NEXTERION® H slides was immersed in Cy5-biotin-BSA to assess the variability in coating across the slide surface. For the quality control assessment, the streptavidin coated slides were removed from the -20°C storage, and equilibrated at RT for 1 hour. A 12 µl aliquot of 0.5 mg/ml Cy5-biotin-BSA and 10 µl aliquot of 1M DTT were removed from -20°C storage, and thawed on ice until ready to use. Fresh quality control buffer (See Appendix A for preparation) was prepared, and 5 µl 1M DTT and 10 µl 0.5 mg/ml Cy5-biotin-BSA were added to the quality control buffer, mixed, and transferred a chamber in the 6-chamber slide processing dish.

The equilibrated slide was immersed in the quality control buffer (See Appendix A for preparation), and incubated on a shaker at 100 RPM for 1 hour at RT, protected from light. The slide was washed in 1× PBST (1× phosphate buffered saline (PBS) and 0.1% Tween®-20) for 5 minutes on a shaker at 100 RPM at RT - the wash step was repeated twice. After removing the wash solution, the slide was finally rinsed in Millipore water (HPLC) for 5 minutes on a shaker at 100 RPM at RT. The slide was dried by centrifugation at 272 ×g for 2 minutes at 22°C.

The entire streptavidin-coated area of the NEXTERION® H slide, as well as the four microarray printing areas, were scanned using the GenePix 4000A fluorescence microarray scanner (Axon Instruments, USA) using the automatic photon multiplier tube (AUTOPMT) setting a 10 µm resolution. The GenePix Pro software (Axon Instruments, USA) was used to obtain the mean and the standard deviation (SD) of the net RFU. The resulting CV was calculated, where a CV of ≤ 5% across the microarray printing area was considered acceptable.

7.2. Printing microarrays

7.2.1. Preparing CT100plus microarray cancer antigen lysates and controls

Stocks of cancer antigen lysates, BCCP lysate, ICL, human IgG (Rocklands), anti-human IgG (SIGMA®), and the 0.5 mg/ml Cy5-biotin BSA stock solution were generated and prepared by Dr. Aubrey Shoko at the Centre of Proteomic and Genomic Research (CPGR).

The working concentrations of the cancer antigen lysates, BCCP lysate, ICL, human IgG and anti-human IgG were prepared by diluting 1:1 in 40% sucrose (in triple-distilled Millipore water) and vortexing (Vortex Genie 2, Scientific Industries). The stock and working solutions were subsequently stored at -80°C until it was needed to prepare the 384-well Genetix plates (Genetix, Separations). Before preparing the working concentrations of the Cy5-biotin-BSA controls, the concentration of the protein as well as the molar ratio of Cy5:BSA was measured, and recorded. The working concentrations of the Cy5-biotin-BSA controls were prepared by adding 16, 32 and 48 µl 0.5 mg/ml Cy5-biotin-BSA to 1.584, 1.568 and 1.552 ml CT100plus control buffer (See Appendix A for preparation) to generate 5, 10 and 15 ng/µl Cy5-biotin-BSA, respectively. The Cy5-biotin-BSA working solutions were stored on ice and protected from light. The "buffer" control on the microarray refers to CT100plus control buffer (See Appendix A for preparation). The 0.5 mg/ml Cy5-biotin-BSA aliquoted stock solutions were stored at -20°C, and the working solutions were stored at -80°C.

7.2.2. Transferring cancer antigens and controls to 384-well Genetix microplates

After removal from -80°C storage, the cancer antigens and controls were thawed on ice. A 384-well Genetix microplate was placed on ice, 20 µl of each antigen and control were transferred to their respective wells; the contents and layout for plate are indicated in Figure 7.1 and Table 7.1, respectively, whereas the layout for plate are indicated in Figure 7.2 and Table 7.2, respectively.

Table 7.1. The 384-well Genetix microplate 1 containing 60 cancer antigens (labelled with gene name) and the Cy5-biotin-BSA, anti-human IgG Ab and human IgG controls, as well as the control buffer.

Well	Component	Well	Component	Well	Component
1A-D & 2A-D/8A-D & 9A-D	5 ng/µl Cy5-biotin-BSA	4E / 11E	GAGE1	5K / 12K	MAGEB6
1E-H & 2E-H/8E-H & 9E-H	10 ng/µl Cy5-biotin-BSA	4F / 11F	GAGE2A	5L / 12L	MART-1/MLANA
1I-L & 2I-L/8I-L & 9I-L	15 ng/µl Cy5-biotin-BSA	4G / 11G	GAGE4	5M / 12A	MICA
3B / 10B	BAGE2	4H / 11H	GAGE5	5N / 12N	NLRP4
3C / 10C	BAGE3	4I / 11I	GAGE6	5O / 12O	NXF2
3D / 10D	BAGE4	4J / 11J	GAGE7	5P / 12P	NY-CO-45
3E / 10E	BAGE5	4K / 11K	GRWD1	6A / 13A	NY-ESO-1
3F / 10F	CCDC33	4L / 11L	HORMAD1	6B / 13B	OIP5
3G / 10G	CEP290	4M / 11M	LDHC	6C / 13C	p53
3H / 10H	COL6A1	4N / 11N	LEMD1	6D / 13D	PBK
3I / 10I	COX6B2	4O / 11O	LIPI	6E / 13E	RELT
3J / 10J	CSAG2	4P / 11P	MAGEA1	6F / 13F	ROPN1
3K / 10K	CT47.11	5A / 12A	MAGEA10	6G / 13G	SGY-1/
3L / 10L	CT62	5B / 12B	MAGEA11	6H / 13H	SILV
3M / 10M	CTAG2/NY-ESO-2	5C / 12C	MAGEA2	6I / 13I	SPAG9
3N / 10N	CXorf48.1	5D / 12D	MAGEA3	6J / 13J	SPANXA1
3O / 10O	DDX53	5E / 12E	MAGEA4 v2	6K / 13K	SPANXB1
3P / 10P	MMA1	5F / 12F	MAGEA4 v3	6L / 13L	SPANXC
4A / 11A	FTHL17	5G / 12G	MAGEA4 v4	6M / 13M	SPANXD
4B / 11B	Control buffer	5H / 12H	MAGEA5	6N / 13N	SPO11
4C / 11C	Anti-human IgG antibody	5I / 12I	MAGEB1	6O / 13O	SSX1
3A/10A & 4D / 11D	Human IgG	5J / 12J	MAGEB5	6P / 13P	SSX2a

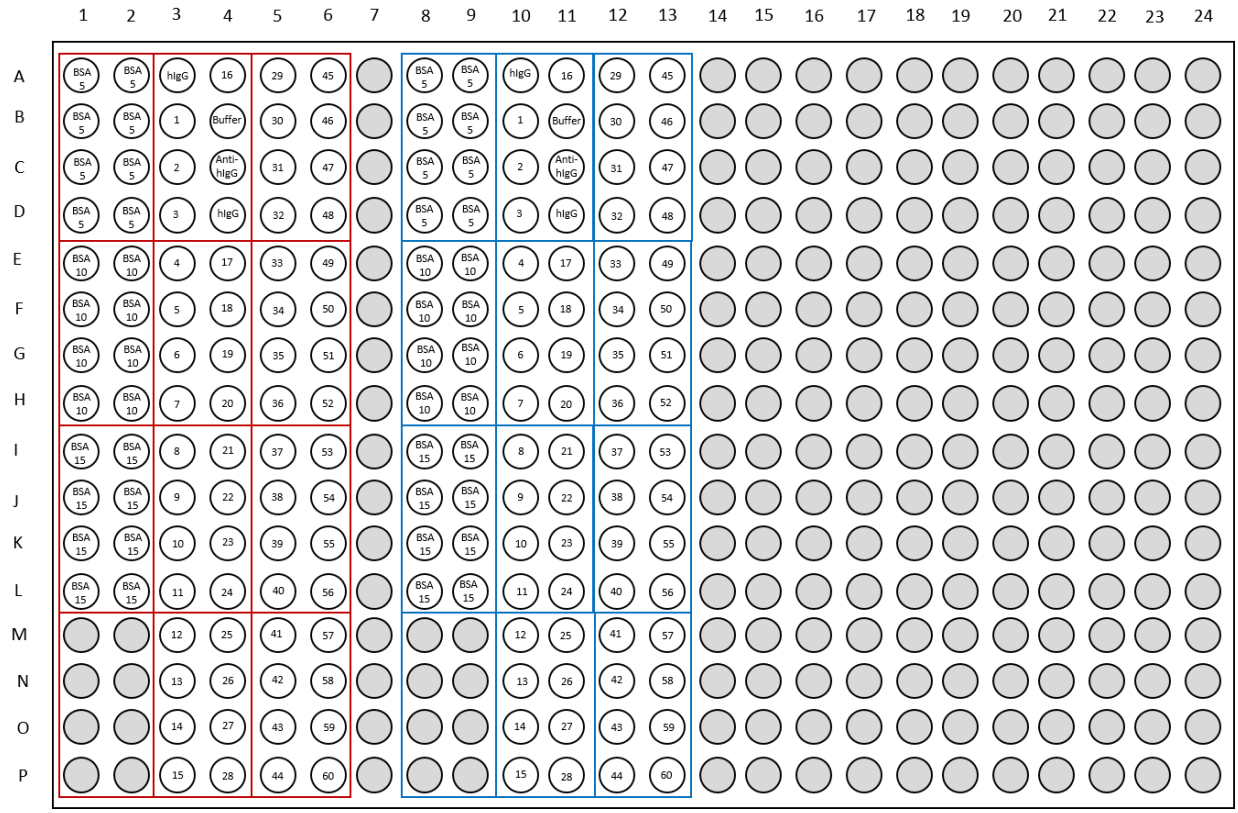


Figure 7.1. 384-well Genetix microplate 1. The figure depicts the layout and contents of a 384-well Genetix microplate 1 compatible with the Qarrayer microarray printer. The microplate was prepared by adding cancer antigens labelled 1-60 (Table 7.1), as well as the controls: Cy5-biotin-BSA, human IgG, anti-human IgG Ab and control buffer to the wells indicated. Grey-shaded wells were empty. Abbreviations include: Cy5 = cyanine 5; BSA = bovine serum albumin; IgG = immunoglobulin G; Ab = antibody.

Table 7.2. The 384-well Genetix microplate 2 containing 63 cancer antigens (labelled with gene name) and the BCCP-c-Myc, ICL controls, as well as the control buffer. Abbreviations include: BCCP = Biotin carboxyl carrier protein; ICL = insect cell lysate.

Well	Component	Well	Component	Well	Component
1A / 8A	SSX4	2G / 9G	SRC	3I / 10I	ACVR2B
1B / 8B	SYCE1	2H / 9H	CALM1	3J / 10J	ITGB1
1C / 8C	SYCP1	2I / 9I	CDC25A	3K/ 10K	MAP9
1D / 8D	THEG	2J / 9J	CREB1	3L / 10L	PIM1
1E / 8E	TPTE	2K/ 9K	CTNNB1	3M / 10M	TKTL1 Isoform a
1F / 8F	TSGA10	2L / 9L	p53 S6A	3N / 10N	SPATS1 isoform 1
1G / 8G	TSSK6	2M / 9M	p53 C141Y	3O / 10O	DPPA2
1H / 8H	TYR	2N / 9N	p53 S15A	3P/ 10P	SOX1
1I / 8I	XAGE-2	2O / 9O	p53 T18A	4E / 11E	ROPN1A
1J / 8J	XAGE3a v1	2P/ 9P	p53 Q136X	4F / 11F	CEACAM1 isoform 1
1K / 8K	XAGE3a v2	3A / 10A	p53 S46A	4G / 11G	POU5F1
1L / 8L	ZNF165	3B / 10B	p53 K382R	4H / 11H	NANOG
1M / 8M	AKT1	3C / 10C	p53 S392A	4I / 11I	BORIS B0
1N / 8N	CDK2	3D / 10D	p53 M133T	4J / 11J	DPPA4
1O / 8O	CDK4	4A / 11A	p53 L344P	4K / 11K	DPPA3
1P / 8P	CDK7	4B / 11B	CYP3A4	4L / 11L	GDF3
2A / 9A	FES	4C / 11C	CYPR	4M / 11M	CTAG2/LAGE-1b/LAGE-1L
2B / 9B	FGFR2	4D / 11D	EGFR	4N / 11N	CAMEL
2C / 9C	MAPK1	3E / 10E	5T4/TPBG	4O / 11O	NY-ESO-1 ORF2
2D / 9D	MAPK3	3F/ 10F	XAGE1B	4P / 11P	BCCP
2E/ 9E	PRKCZ	3G / 10G	SOX2	5A / 12A	Insect cell lysate
2F / 9F	RAF	3H / 10H	ACVR2A	5B-D & 6A-D/12B-D & 13A-D	Control buffer

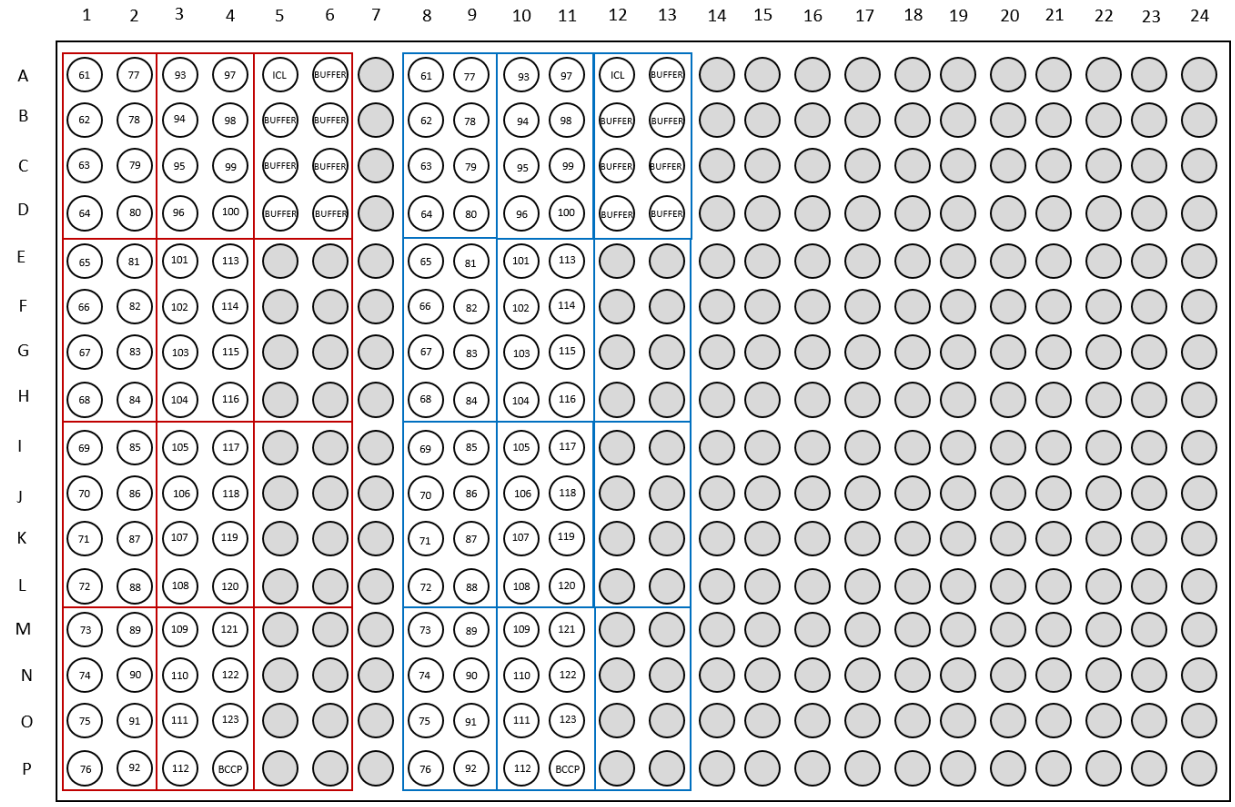


Figure 7.2. 384-well Genetix microplate 2. The figure depicts the layout and contents of a 384-well Genetix microplate compatible with the Qarrayer microarray printer. The microplate was prepared by adding cancer antigens labelled 61-123 (Table 7.2), as well as the controls: BCCP-c-Myc and control buffer (Buffer) to the wells indicated. Grey-shaded wells were empty.

To prevent photobleaching, the Cy5-biotin-BSA controls were transferred last to the microplate, while protected from light. Once the transfer was complete, both 384-well Genetix microplates were either immediately used to print the CT100plus microarray or the microplate was covered with aluminium-foil sealing films (EXCEL Scientific), protected in tin foil (SAPPHIRE), and stored at -80°C. The remaining cancer antigen and control working solutions were stored at -80°C until required. It is important to note that the Cy5-biotin-BSA controls were stored in tubes protected from light to prevent photobleaching.

7.2.3. Printing CT100plus microarrays

7.2.3.1. Preparing derivatised NEXTERION® H slides and the Genetix 384-well microplates

Streptavidin-coated NEXTERION® H slides were removed from -20°C storage, and equilibrated for 1 hour at RT. If streaking and/or particles were seen, the slides were rinsed in triple-distilled Millipore water (HPLC grade) for 2 minutes at RT, while shaking at 100 RPM and protected from light. Slides ware dried by centrifugation at 272 ×g for 2 minutes at 22°C. The two 384-well Genetix microplates containing the cancer antigens and controls were removed from -80°C storage, and thawed on ice.

7.2.3.2. Preparing Qarray2 robotic arrayer

The Qarray2 robotic arrayer (Genetix, Berkshire, UK), equipped with 8 X 300 µm flat-tipped solid pins (Genetix aQU 300 µm), was used to print the CT100plus microarray. The print settings are summarised in Table 7.3.

Before starting the printing procedure, the following settings for the Qarray2 robotic arrayer were checked: The humidifier and water wash container were topped with triple-distilled Millipore water (HPLC grade), and the ethanol wash container was topped with 80% ethanol (in triple-distilled Millipore water). Three water and ethanol washes were activated to confirm that the hydraulic systems were functional. The printing pins were sonicated for 5 minutes at RT in triple-distilled Millipore water (HPLC grade), sprayed with 80% ethanol (in triple-distilled Millipore water), and dried with lint-free tissue (Munktell).

Table 7.3. Qarrayer parameter for environmental conditions, print settings and pin wash settings are described. The same settings were used for every CT100plus print run. Abbreviations include: ms = milliseconds.

Parameter	Setting
Environmental conditions	
Temperature	~19 °C
Humidity	~ 50%
Print settings	
Inking time	500 ms
Microarray pattern	8 X 8 spots/grid
Spacing	562 µm
Max. stamps per ink	1
Stamps per spot	1
Stamp time	0 ms
Print depth	150 µm
Pin wash settings	
Water wash	60 sec.
Ethanol wash	10 sec.
Drying	10 sec.
Pause	1 sec.

Datum pointing was performed to calibrate the position of the printing pins relative to the slide. A print depth of 150 µm (Table 7.3) was found to be the optimal depth to ensure homogenized printing across the slide bed. After completing the datum pointing check, the remaining pins were inserted into the printer head. Pin set one printed the first set of microarrays, and pin set 2 printed the second set of microarrays on the slide (Figure 7.3). Thereafter, the required number

of equilibrated slides were loaded in the arrayer, the vacuum was applied to prevent slides from moving during the printing procedure, and the array printing area was purged to remove potential interfering particles or contaminants. Thereafter, the humidity was set to ~ 50% and the Qarray2 light was switched off. Once the above-mentioned parameters were checked, the printing procedure was started.

Replica CT100plus microarrays were printed in a four-plex format on streptavidin-coated NEXTERION® H slides. Each CT100plus microarray is subdivided into 8 subarrays, each containing cancer antigens, positive and/or negative controls printed in triplicate. The human immunoglobulin G "hIgG" positive control was printed in subarrays 4 and 5, and the "anti-human immunoglobulin G (anti-humanIgG)" was printed in triplicate in subarray 6. Subarrays 1-3 and 5-8 contained the "buffer" negative control printed in triplicate, subarray 4 contained the ICL negative control printed in triplicate, and subarray 5 contained the "BCCP" negative control printed in triplicate. The layout of the entire slide, CT100plus microarray and subarrays is shown in Figure 7.4.

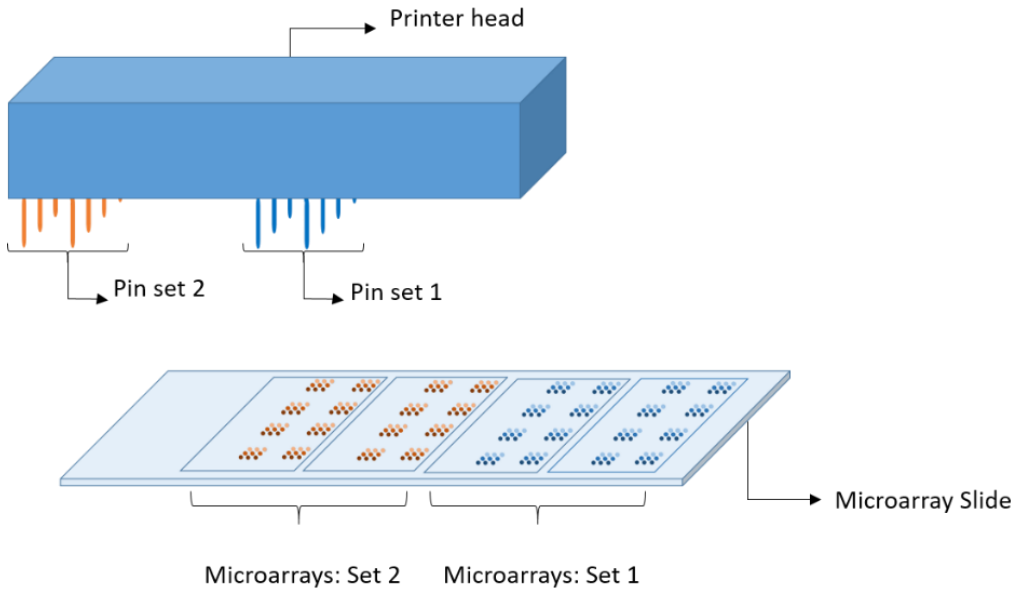


Figure 7.3. Pin positioning for CT100plus printing. The Qarray2 robotic arrayer printer head and pin positioning use to print CT100plus microarrays in a four-plex format. Pin set 1 prints microarray set 1; pin set 2 prints microarray set 2 on a streptavidin-coated NEXTERION® H slide. The 5, 10 and 15 ng/µl Cy5-biotin-BSA controls were used for downstream pin-to-pin and array-to-array normalisation.

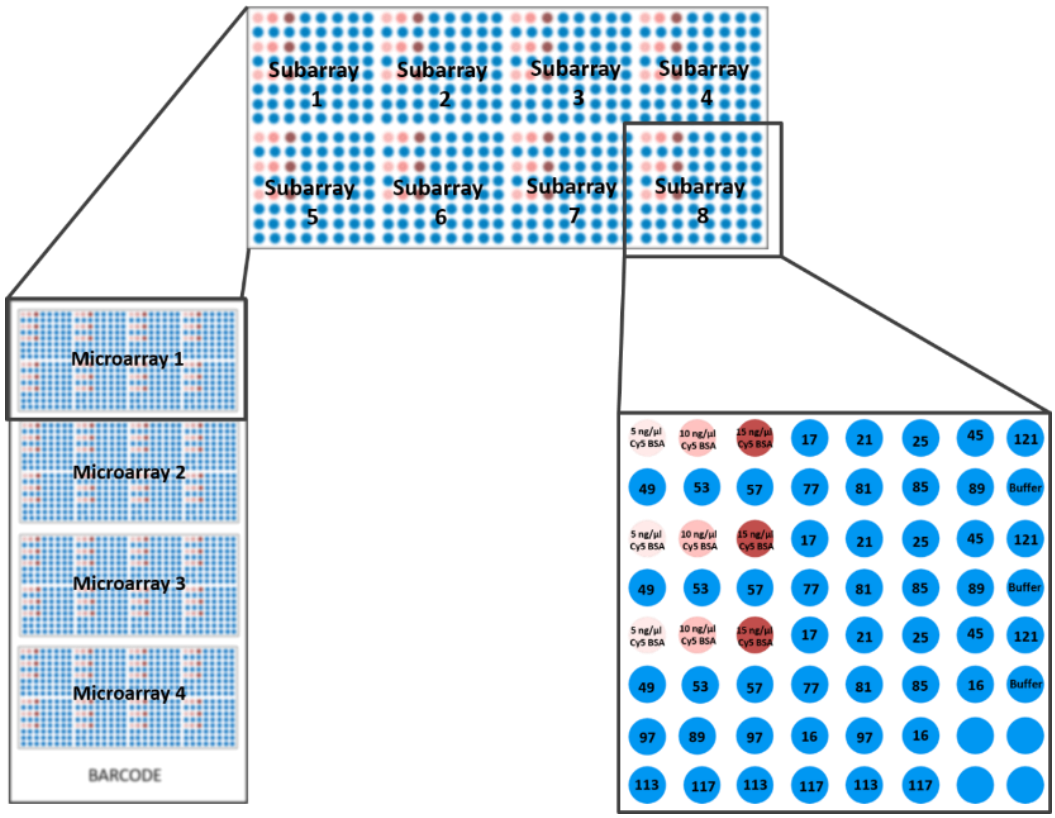


Figure 7.4 Layout of the CT100plus microarray. CT100plus microarrays were printed in a 4-plex format on streptavidin-coated NEXTERION H slides. Each microarray is subdivided into 8 subarrays, each containing 123 cancer antigens, as well as positive (Cy5-biotin-BSA, human IgG, anti-human IgG) and/or negative controls (buffer, BCCP-c-Myc, ICL) printed in triplicate. Abbreviations include: Cy5 = cyanine 5; BSA = bovine serum albumin; IgG = immunoglobulin G; BCCP = biotin carboxyl carrier protein, ICL = insect cell lysate.

After the printing procedure was complete, the humidity and vacuum was switched off. Slides were removed from their holders and visually inspected for missing spots, spot merging and/or drying and dust particles. The slides were re-packaged in slide holders and stored at 4°C, protected from light, until the day of assaying.

7.3. CT100plus hybridisation assay

7.3.1. Preparing PAD4 (only Chapter 3)

The PAD4 enzyme (SIGMA®) was removed from -80°C storage and, thawed on ice. Each microarray was treated with 1 µg/ml PAD4 enzyme in PAD4 buffer (See Appendix A for preparation).

7.3.2. Preparing sera /plasma samples and detection antibody

Patient and HC sera and/or plasma were removed from -80°C storage and thawed on ice. The serum and/or plasma was diluted 1:800 in 1× PBST (1× PBS and 0.1% Tween®20), mixed by vortexing, and stored on ice until the CT100plus assay began. The Alexa fluor 647-coupled goat anti-human IgG (H+L) (Invitrogen) was used as a detection Ab at a dilution of 20 µg/ml (in PBST).

7.3.3. Preparing CT100plus microarray slides

NB: The CT100plus microarray slides were prepared at RT on a shaker (100 RPM) and protected from light.

The CT100plus microarray slides were removed from 4°C storage, and immediately immersed in blocking buffer (see Appendix A for preparation) for 1 hour. The slides were subsequently washed in 1× PBST (1× PBS and 0.1% Tween®-20) for 5 minutes – the wash was repeated twice, and finally rinsed in triple-distilled Millipore water (HPLC grade) for 5 minutes. The slides were dried by centrifuging at 272 ×g for 2 minutes at 22°C. Blocked slides were temporarily stored at RT and protected from light until the Tecan Hybridization station (Tecan, Austria) was ready to perform the CT100plus assay.

7.3.4. Tecan hybridisation station

7.3.4.1. Rinsing procedure

Before starting the rinsing routine, the Tecan Hybridization station PC, as well as the extension and liquid units were switched on. The nitrogen gas supply was set at 270 kPa, and the valve was opened to release the nitrogen pressure.

A routine rinsing procedure was performed before and after each assay. For the rinsing procedure, the Tecan Hybridization wash/rinse bottle 6 was topped with fresh 2.25 L triple-distilled Millipore water (HPLC grade), and the washing/rinsing tubes were placed in bottle 6 of the hybridization station. Sterilized slide holders were placed in each slide holder. The hybridization chambers were placed in their allocated modules, module lid carefully closed, and each injected port plug tightened. The rinsing procedure was activated, with a final drying step.

Once the rinsing procedure was completed, the module lids were opened and the slide and slide holders were inspected for water leaks. Once the rinsing procedure was completed, the CT100plus assay could start.

7.3.4.2. Tecan Hybridisation: CT100plus assay

Before starting the assay, the CT100plus assay hybridization settings (Table 7.4) were loaded in the HS Pro Control Manager (Version 4.00).

Table 7.4. Tecan Hybridization protocol settings for the citrullinated CT100plus microarray. Abbreviations include: PAD4 = protein arginine deiminase isoform 4; PBST = phosphate buffered saline supplemented with tween-20; tH_2O = triple-distilled water; min. = minutes.

Parameter	Setting
Incubation: PAD4	2 hours (37°C)
Wash time (PBST) / Soak time	3 × 1 min. / 3 × 0.5 min.
Wash time (tH_2O) / Soak time	2 × 1 min. / 2 × 0.5 min.
Incubation: Plasma	1 hour (23°C)
Wash time (PBST) / Soak time	3 × 1 min. / 3 × 0.5 min.
Wash time (PBST) / Soak time	2 × 1 min. / 2 × 0.5 min.
Incubation: Detection antibody	30 min. (23°C)
Wash time (PBST) / Soak time	4 × 1 min. / 4 × 0.5 min.
Wash time (tH_2O)	1 × 0.5 min.
Slide drying time/ machine temperature	1.5 min. / 30°C

The blocked CT100plus microarray slides were placed in the appropriate module positions, and the module lid was closed and covered with a black lid to protect the microarrays from light. The

Tecan Hybridization bottles 1 and 2 were topped with 2 L PBST (1× PBS, 0.1% Tween®-20), and bottle 1 was topped with 1 L triple-distilled Millipore water (HPLC grade), and tubes 1, 2 and 3 were placed in their complementary bottles.

Upon completion of the CT100plus assay, the slides were removed from the modules and immediately scanned. The routine Tecan Hybridization station rinse protocol was activated (see Section 1.3.3.1.), where after each chamber was manual cleaned by rinsing with 30% ethanol (in triple-distilled water) and triple-distilled Millipore water (HPLC grade), and finally dried with lint-free tissue.

7.4. CT100plus microarray scanning and visual assessment

The microarray images were scanned using the parameters summarised in Table 7.5. Unique alpha-numeric description was ascribed to each microarray, and the images were subsequently used for data extrapolation using the ArrayPro data extraction and analysis software (Media Cybernetics, Rockville MD).

Table 7.5. GenePix 4000B scanner settings for the CT100plus microarray.

Parameter	Description
Scanner	GenePix 4000B
Software	GenePix Pro
Gain setting	AUTOPMT
Saturation threshold	0 %
Resolution	10 µm
Image type	*.TIFF

To ensure that the numerical outputs from the ArrayPro and CT100+ programs represented biologically meaningful data, each CT100plus microarray *.TIFF image was visually assessed for spot homogeneity, spot merging, spot bleeding, dust particles, speckling, and variation in

background. If either of the aforementioned artefacts affected at least 2 of the triplicate spots, a sample re-run was performed until a high-quality image was produced.

7.5. ArrayPro data extrapolation and CT100+.jar settings

The ArrayPro software was used to extrapolate numerical information from the CT100plus microarray images. The bioinformatic pipeline for CT100plus quality control measures was developed in the Blackburn laboratory. The initial prototype was prepared by Mr. Jean-Michel Serufuri, and was later upgraded by Dr. Jessica Duarte, and set in a Java script by Dr. Ryan Goosen. Altogether, a robust and user-friendly pipeline was developed, and the following quality control measures were applied to the data set for this cohort:

For the CRC cohort, the default noise- (2 SD above background) and saturation- thresholds (65535 RFU) were applied. The 15 ng/µl Cy5-biotin-BSA control was used to assess variance across microarrays, with a CV threshold set at 25%. The median net RFU was obtained, with the CV-threshold for antigens set at 20%. Table 7.6 provides descriptions for each filtering and normalisation step in the CT100plus application.

Table 7.6. Description for filtering and normalisation steps in the CT100plus application. Abbreviations include: BG = background; SDs = standard deviations; CV = coefficient of variation; Cy5 = cyanine 5; BSA = bovine serum albumin.

Parameter	Description
1. Neighbourhood background	The median-value of the 8 neighbouring BG-values flanking each antigen.
2. Net intensity	Raw intensity minus neighbourhood background.
3. Acceptable noise threshold	Antigen net intensity greater than 2 SDs from background.
4. CV threshold: Antigens	User-defined CV-threshold for mean net intensity of antigens.
5. CV threshold: Cy5-biotin-BSA	User-defined CV-threshold for mean net intensity of Cy5-biotin-BSA.
6. Pin-pin normalisation	Subarrays normalised with Cy5-biotin-BSA controls using Equation 1.
7. Array-to-array normalisation	Arrays normalised with Cy5-biotin-BSA controls using Equation 1.

7.6. SENGENICS IMMUNOME™ microarray assay

7.6.1. Blocking SENGENICS IMMUNOME™ microarray slides

NB: The CT100plus microarray slides were prepared at RT on a shaker (100 RPM) and protected from light.

The SENGENICS IMMUNOME™ microarrays were removed from their storage solution, and washed thrice with 1× PBST (PBS and 0.1% Tween®-20) for 5 minutes – the wash step was repeated twice, and the slides were finally rinsed in triple-distilled Millipore water for 5 minutes. The slides were then immediately immersed in blocking solution (see Appendix A for preparation) for 1 hour (hr). The slides were washed in 1× PBST (PBS and 0.1% Tween®-20) for 5 minutes – the wash step was repeated twice, and the slides were finally rinsed in triple-distilled Millipore water for 5 minutes, and dried by centrifuging at 272 ×g for 2 minutes at 22°C.

7.6.2. Preparing CRC plasma and detection antibody

Patient plasmas were removed from -80°C storage, and thawed on ice. The serum and/or plasma was diluted 1:800 in 1× PBST (PBS and 0.1% Tween®-20), mixed by vortexing, and stored on ice until the CT100plus assay began. The Alexa fluor 647-coupled goat anti-human IgG (H+L) (Invitrogen) was used as a detection Ab at a dilution of 20 µg/ml (in PBST).

7.6.3. Assaying CRC plasma

Blocked SENGENICS IMMUNOME™ microarrays were incubated with **CRC038** or **CRC050** plasma for 1 hr. The slides were subsequently washed in PBST (PBS and 0.1% Tween®-20) for 5 minutes – the wash step was repeated twice, and the slides were finally rinsed in triple-distilled Millipore water for 5 minutes. Each slide was incubated with 20 µg/ml Alexa fluor 647-coupled goat anti-human IgG (H+L) (Invitrogen) for 30 minutes. The slides were washed again in PBST (PBS and 0.1% Tween®-20) for 5 minutes – the wash step was repeated twice, and the slides were finally rinsed in triple-distilled Millipore water for 5 minutes. The slides were dried by centrifuging at 240 ×g for 2 minutes at 22°C.

7.6.4. SENGENICS IMMUNOME™ microarray scanning and visual assessment

The microarray images were scanned using the parameters summarised in Table 7.7. The images were subsequently used for data extrapolation.

Table 7.7. GenePix 4000B scanner settings for the SENGENICS IMMUNOME™ microarray.

Parameter	Description
Scanner	GenePix 4000B
Software	GenePix Pro
Gain setting	900
Saturation threshold	0 %
Resolution	10 µm
Image type	*.TIFF

7.7. Optimising the AP-MS assay

7.7.1. Capturing IgG from CRC patient plasma

Protein G is a 22 kDa protein from Streptococcus group C and G that has a high affinity toward all subclasses of IgG e.g. IgG1, IgG2, IgG3 and IgG4. For our assays, Protein G-coated magnetic beads were used to capture IgG molecules from CRC patient plasma, which we verified through protein quantitation, SDS-PAGE with AQUASTAIN and Western blot analyses. The protocols that were used are described below:

7.7.2. Equilibrating MagReSyn® Protein G magnetic beads

Suspended MagReSyn® Protein G magnetic beads (ReSyn Biosciences) were thoroughly mixed, and 50 µl was transferred to a 1.5 ml tube. The shipping solution was removed, and the beads were washed with 300 µl binding buffer (See Appendix A for preparation) for 5 minutes – the wash step was repeated twice. The beads were then ready to capture IgG from CRC patient plasma.

7.7.3. IgG capture from CRC patient plasma

CRC patient plasma was removed from -80°C storage, and thawed on ice. For each patient, 15 µl CRC plasma was re-suspended in 135 µl binding buffer (See Appendix A for preparation), and

mixed by aspirating. The diluted plasma sample was added to the washed Protein G magnetic beads (see section 7.7.2), and incubated for 1 hr. Thereafter, unbound proteins were removed, and stored for further analysis. The beads were washed with 300 µl binding buffer for 5 minutes – this wash step was repeated twice. All bound proteins were eluted with 50 µl 2.5% acetic acid (SIGMA®) for 5 minutes – the elution step was repeated twice, and the eluents combined. The bound protein eluents were neutralized with 7 µl 5 M NaOH, and stored for protein quantitation (see Section 7.7.6), SDS-PAGE with AQUASTATION and Western Blot analysis (see Section 7.7.7).

7.7.4. Crosslinking Protein G and IgG

Dimethyl pimelimidate (DMP) is a crosslinker that contains an amine reactive imidoester group at each end of a 7-atom spacer arm. It is reactive toward primary amines present on arginine and N-termini of peptides and proteins, resulting in the formation of an imidine bond. For the AP-MS assay, DMP was used to crosslink Protein G-bound plasma proteins to reduce plasma protein contamination during further MS analysis. Prior to crosslinking, 50 µl Protein G beads were incubated with either 7.5 or 15 µl patient plasma, and then crosslinked with DMP before low pH elution. As a negative control, 50 µl Protein G beads were incubated with 15 µl patient plasma without crosslinking followed by low pH elution. The protocol used is described below:

Protein G beads functionalized with plasma IgG (and non-specific proteins) were incubated with crosslinking buffer, containing 20 mM DMP in 0.1 M sodium borate (pH 9.2), for 45 minutes at RT. The crosslinking buffer was removed, and the beads were washed with 300 µl 1× PBS for 5 minutes at RT. The crosslinking and wash steps were repeated twice, and the beads were finally washed in 500 µl 0.2 M Tris-HCl (pH 8.0) for 5 minutes at RT. Unbound proteins were eluted with 50 µl 2.5% acetic acid (SIGMA®) for 5 minutes – the elution step was repeated twice, and the eluents for each sample-type combined. The eluents were neutralized with 7 µl 5 M NaOH, and stored for protein quantitation (see Section 7.7.6) and SDS-PAGE with AQUASTAIN (see Section 7.7.7).

7.7.5. Generating native protein lysate

For the AP-MS assay, native lysate was generated through the use of the non-denaturing lysis buffer described before (Alhamdani et al., 2010). The lysis buffer included HEPES (pH 7.4), which

is a Good's buffer characterised by its high compatibility with biological analysis, good solubility in water and minor salt effects (Good et al., 1966). Benzonase was added to block interference from nucleic acids. To avoid protein degradation, protease inhibitors (Roche) were added. Most critical for the isolation of native proteins was the detergents used – the list of tested detergents include Triton™ X-100, IGEPAL® CA-630 and CHAPS, or combinations thereof. Since CRC tissue was precious and limited, an SHSY-5Y human neuroblastoma cell line was used to optimize lysis buffer composition and lysis conditions. The protocol that was used is described below:

7.7.5.1. Cell culture

SHSY-5Y neuroblastoma cells were grown to 80% confluency in 12 sterile 10 cm dishes (Greiner). The DMEM media (Gibco®), supplemented with 10% fetal calf serum (Gibco®) and 1% pen/strep (LONZA), was removed, and the cells were washed with 1× PBS (Gibco®). The cells were lifted with trypsin (LONZA), transferred to a 15 ml tube and centrifuged at 4200 ×g for 5 minutes. The supernatant was discarded, and the pellet was further processed, as displayed in Figure 7.5.

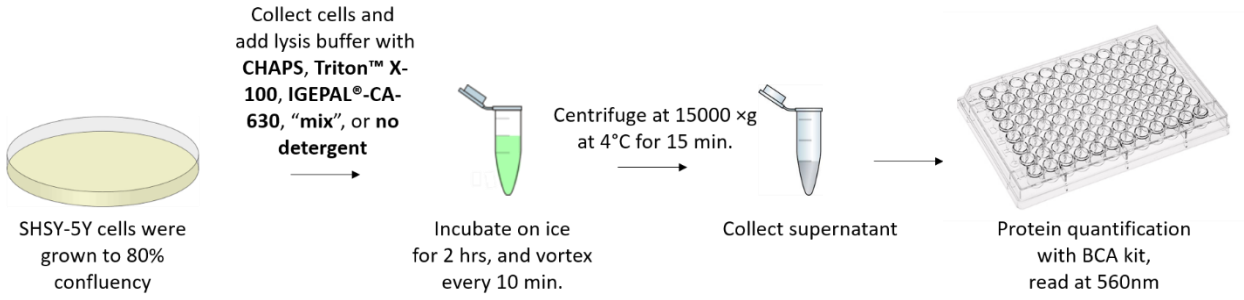


Figure 7.5. SHSY-5Y lysate generation. SHSY-5Y neuroblastoma cells were used to optimise lysis buffer detergent composition for CRC tissues in the AP-MS assay. The non-denaturing lysis buffer was supplemented with either CHAPS, Triton™ X-100, IGEPAL® CA-630, "mix", or no detergent. The supernatants were collected, and protein quantification was performed using the BCA protein quantitation kit (Pierce™). Here the lysis conditions are described for *lysis buffer only* (generated with 350 µl lysis buffer on ice for 2 hours, with intermittent vortexing), although the same procedure was used for *lysis buffer & PE* (generated with 350 µl lysis buffer and homogenizing using a Potter-Elvehjem homogenizer, and incubation on ice for 2 hours, with intermittent vortexing) and *lysis buffer & shaking* (generated with 350 µl lysis buffer and homogenizing with a Potter-Elvehjem homogenizer, and continuous vortexing 4°C). Abbreviations include: CRC = colorectal cancer; AP-MS = affinity purification-mass spectrometry; BCA = bicinchoninic acid; PE = Potter-Elvehjem; min. = minute.

The pelleted SHSY-5Y neuroblastoma cells were lysed in 350 µl non-denaturing lysis buffer (see Appendix A for preparation), each containing various specific non-denaturing detergent, including either 1% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) (SIGMA®), 1% Triton™ X-100 (SIGMA®), 1% IGEPAL® CA-630 (SIGMA®), or a mixture detergents (0.25% CHAPS, 0.5% IGEPAL® CA 630, and 0.5% Triton™ X-100). As a negative control, the SHSY-5Y neuroblastoma cell pellet was incubated with lysis buffer without detergent.

Three lysis conditions were also assessed for protein yield: **1**. *Lysis buffer only* refers to lysates generated by incubating with 350 µl lysis buffer on ice for 2 hours, with vortexing every 10 minutes. **2**. *lysis buffer & PE* refers to lysates generated by incubating cells with 350 µl lysis buffer and homogenizing using a Potter-Elvehjem homogenizer, and then incubated on ice for 2 hours, with vortexing every 10 minutes. **3**. *Lysis buffer & shaking* refers to lysates generated by incubating cells with lysis buffer and homogenizing with a Potter-Elvehjem homogenizer, and then continuous vortexing for 2 hrs at 4°C. Thereafter, the lysates were centrifuged at 15 000 ×g for 15 min at 4°C. The supernatants were collected, and the protein concentration determined using the BCA protein quantitation kit (See Section 7.7.6).

7.7.6. BCA protein quantitation

The bicinchoninic acid (BCA) colorimetric assay was used to quantify proteins. The assay is based on the reduction of Cu^{2+} to Cu^{1+} by proteins in an alkaline medium. The accumulated cuprous (Cu^{1+}) cation is quantified colorimetrically with BCA at 562 nm. For our assays, we employed the BCA protein quantitation kit (Pierce™). The protocol that was used is described below:

For the BCA protein quantitation, 10 µl sample or 10 µl BSA standard (2, 1.5, 1, 0.75, 0.5, 0.25, 0.125, 0.025 mg/ml) were transferred in triplicate to a 96-well plate (Greiner bio-one). Thereafter, 200 µl working reagent, prepared by mixing BCA solution and copper solution at a ratio of 1:50, was added to each well containing sample or standard. The plate was incubated for 30 minutes at 37°C, protected from light. The plate was read at 560 nm, and the data extracted and analysed.

7.7.7. Detecting proteins by SDS-PAGE and Western blot analysis

SDS-PAGE is a method used to separate denatured proteins. Using a molecular weight marker, the molecular weight of protein bands can also be estimated. For our studies, SDS-PAGE was used to detect denatured and reduced IgG molecules using AQUASTAIN or Western blot analysis. The protocols that were used are described below:

After proteins were eluted from Protein G magnetic beads, and neutralized, 16 µl protein was mixed with 4 µl Laemmli buffer (see Appendix A for preparation) to a final volume of 20 µl, and heated at 95 °C for 5 minutes. The lysate was centrifuged at 15 000 ×g for 30 s. The samples were subsequently loaded onto a polyacrylamide gel with a 4% stacking gel (see Appendix A for preparation) and 10% separating gel (see Appendix A for preparation), and protein separation performed in 1× running buffer (see Appendix A for preparation). The initial voltage was set to 100 volts (V) until the dye front reached the separating gel. The voltage was increased to 200 V until the dye front was 0.5 cm from the bottom of the gel. The protein bands were detected by incubating with AQUASTAIN for 15 minutes, followed by image capture.

IgG protein bands were also detected using Western blot analysis. After SDS-PAGE was completed, proteins were transferred onto a nitrocellulose membrane (BioTrace™ NT, PALL Corporation) at 100 V for 1 hr in ice-cold transfer buffer (See Appendix A for preparation). Thereafter, the nitrocellulose membrane was blocked in 5% fat-free milk powder diluted in 1× TBST (See Appendix A for preparation) at RT for 1 hr. The membrane was incubated with goat anti-human IgG peroxidase Ab (SIGMA®) (1:10000) at 4°C with gentle shaking overnight. Once labelled with antibody, the membrane was washed with 1× TBST (See Appendix A for preparation) for 10 minutes – the wash was repeated twice. The membrane was treated with chemiluminescence substrate (Advansta) for 2.5 minutes, and the relative chemiluminescence was detected using the G:BOX (Syngene). Relative quantitation for AQUASTAIN and Western blot was determined using densitometry performed with Image J (version 1.5.1).

7.8. AP-MS assays with CRC plasma and tissue lysates

Once the various steps of the AP-MS assay were optimised, the protocol was used to assay the blood plasma and tissues of 5 CRC patients: **CRC002**, **CRC004**, **CRC017**, **CRC021** and **CRC031**. The protocol for the optimized assay is described below:

7.8.1. CRC tissue lysis

Normal colorectal mucosa and CRC tissues were removed from -80°C storage, and thawed on ice. A ~3 mm^3 tissue section was removed and minced with a sterile blade (HI-CARE $_{INT}$). The minced tissue was immediately re-suspended in non-denaturing lysis buffer with Triton® X-100 (See Appendix A for preparation), and homogenized with a Potter-Elvehjem homogenizer with ~30 strokes. The suspension was transferred to a 1.5 ml tube (Greiner) and vortexed at 4°C for 2 hrs. Thereafter, the lysates were centrifuged at 15 000 ×g at 4°C for 15 minutes. Lysate pellets and supernatants were stored separately at -80°C. Before the AP-MS assays, the supernatant was diluted to 0.4 mg/ml in binding buffer (See Appendix A for preparation).

7.8.2. Equilibrating MagReSyn® Protein G magnetic beads

The protocol describes the AP-MS assay conditions for one patient, although this method was applied to all patient samples tested. It is important to note that the washing and incubation steps were performed at RT whilst vortexing, unless stated otherwise.

Suspended MagReSyn® Protein G magnetic beads (ReSyn Biosciences) were thoroughly mixed, and 150 µl were transferred to a 1.5 ml tube. The shipping solution was removed, and the beads were washed with 900 µl binding buffer (See Appendix A for preparation) for 5 minutes – the wash step was repeated twice. The beads were then ready to capture IgG from CRC patient plasma.

7.8.3. Capturing IgG from CRC patient plasma

CRC patient plasma was removed from -80°C storage and thawed on ice. To capture IgG, 45 µl CRC plasma was re-suspended in 405 µl binding buffer (See Appendix A for preparation) and mixed by aspirating. The diluted plasma sample was added to the washed Protein G magnetic

beads (see section 7.7.2) and incubated for 1 hr. Thereafter, unbound proteins were removed, and the beads were washed with 900 µl binding buffer for 5 minutes – this wash step was repeated twice. The beads were then ready for crosslinking.

7.8.4. Crosslinking IgG to MagReSyn® Protein G magnetic beads

Functionalized MagReSyn® Protein G magnetic beads were crosslinked with 100 µl 20 mM DMP in 0.1 M sodium borate (pH 9.2) for 45 minutes. The crosslinking buffer was removed, and the beads were washed with 900 µl binding buffer (See Appendix A for preparation) for 5 minutes – the wash step was repeated twice. This process was repeated twice, with a final wash in 1200 µl 0.2 M Tris-HCl (pH 8.0) for 5 minutes. The beads were then ready for incubation with native colorectal lysates.

7.8.5. Adding patient lysate tissues to IgG-functionalised MagReSyn® Protein G magnetic beads

Once the beads were crosslinked, they were separated into 3 tubes each containing ⅓ of the functionalized beads. Each tube was incubated with either 500 µl 0.4 mg/ml native CRC tissue lysate, 500 µl 0.4 mg/ml native normal colon or rectal tissue lysate, or with 500 µl binding buffer only overnight at 4°C. Thereafter, unbound proteins were removed, and the beads were washed with 300 µl binding buffer (See Appendix A for preparation) for 5 minutes – the wash step was repeated four times. The beads were then washed with 300 µl 1× PBS for 5 minutes – the wash step was repeated four times. The beads were then ready for the elution step.

7.8.6. Eluting proteins from functionalised MagReSyn® Protein G magnetic beads

After the beads were functionalized, they were incubated with **elution buffer 1** (see Appendix A for preparation) for 30 minutes. The beads were washed twice with **elution buffer 2** (see Appendix A for preparation), and the eluents mixed. The protein samples were digested at 27 °C for 18 hrs, and then incubated with 5.5 mM IAA (AMRESCO®) for 30 minutes protected from light. Trypsinization was stopped with 0.15% formic acid (SIGMA®), and the samples were ready for desalting.

7.8.7. Detergent removal

The HiPPR™ Detergent Removal Resin Column Kit (Thermo Scientific) was used to remove Triton X-100 according to the manufacturers protocol:

Empty spin columns were placed in an empty 2 ml collection tube (Greiner). Thereafter, 200 µl evenly suspended resin was placed in the spin column and centrifuged at 1500 ×g for 1 minute to remove the storage buffer. The slurry was equilibrated with 200 µl PBS (pH 7.5), then centrifuged at 1500 ×g for 1 minute to remove the equilibration buffer – this step was repeated twice. After inserting the bottom plug, peptides were applied directly to the compact resin bed, and incubated for 10 minutes at RT. Lastly, the cap was slightly loosened, and the bottom plug removed, then centrifuged at 1500 ×g for 1 minute to elute detergent-free peptides.

7.8.8. Desalting AP-MS peptides

Digested peptides were desalted using reverse phase extraction, which separates molecules based on their hydrophobicity. C-18 was used as the hydrophobic stationary phase to remove salts and impurities with an aqueous phase, and hydrophobic molecules eluted for further analysis. While solid phase extraction kits are commercially available, we made use of in-house prepared stage tips. The protocol is described below:

As described before (Rappsilber et al., 2007), stage tips were prepared by inserting two C-18 (Empore™) discs into an irradiated 200 µl tip up to the hilt, which was subsequently inserted into the pierced lid of an irradiated 2 ml tube. The C-18 was equilibrated with 100 µl **solvent B** (see Appendix A for preparation), then centrifuged for 2 minutes at 7000 ×g – this equilibration step was repeated twice. The C-18 was washed with 100 µl **solvent A** (see Appendix A for preparation), then centrifuged for 2 minutes at 7000 ×g – this wash step was repeated once. The washed C-18 stage tip was incubated with 10 µg digested peptide, then washed with 50 µl **solvent A** (see Appendix A for preparation) and centrifuged for 2 minutes at 7000 ×g – this wash step was repeated twice. The desalted peptides were subsequently eluted in to a glass vial (MACHEREY-NAGEL) with 50 µl **solvent C** (see Appendix A for preparation), then centrifuged for 2 minutes at 7000 ×g – this elution step was repeated twice. The desalted peptides were lyophilized in a SpeedVac (Thermo Fisher Scientific, USA) at 35 °C for 45 minutes. The peptides were re-

suspended in 30 µl **solvent A** (see Appendix A for preparation) and stored at -20 °C before analysing on the mass spectrometer.

7.8.9. Mass spectrometer-based sample processing

7.8.9.1. High-performance liquid chromatography (HPLC)

High-performance liquid chromatography (HPLC) is an analytical chemistry technique used to simplify, identify and quantify components within a mixture. During the process, fine chromatographic materials are compacted to form the stationary phase to improve mobile phase (i.e. the solvent) and stationary phase surface interaction, allowing the use of smaller quantities of analyte. Each analyte in the sample interacts slightly differently with the stationary phase, resulting in component-specific elution rates. In HPLC, analytes can be separated by ion exchange, size exclusion, reverse-phase chromatography. For proteomics-based mass spectrometry, peptides, which typically have varying degrees of hydrophobicity, are separated using a reverse-phase column, which in our study consisted of a compacted C18 resin. To effectively simplify the peptide mixture prior to mass spectrometric analysis, an elution gradient of increasing acetonitrile concentration is used to separate peptides based on their hydrophobicity. Molecule separation is visualized by a detector, where the time taken for compound to elute from the column to the detector is known as the retention time. The protocol is described below:

The vials of 30 µl re-suspended peptides were placed in labelled autosampler glass vials (MACHEREY-NAGEL), and loaded in the autosampler tray of the Dionex UltiMate® RSnano LC system (Thermo Fisher, San Jose, CA, USA). The nanoflow ultra-high-performance liquid chromatography (UHPLC) was performed using a C-18 reverse phase precolumn trap with Luna beads (100 µm × 2 cm; 5 µM, 100Å) (Separations, South Africa) connected to an analytical column (75 µm × 40 cm; 5 µM, 100Å) (Separations, South Africa) packed with C18 Aeris beads (5 µm diameter, 100 Å pore size; Phenomenex 04A-5452) connected to an Ultimate 3500 RS nano UPLC system (Dionex). The gradient chromatography was performed at 40 °C with a flow rate of 400 nL/min and eluted with a 5-35% gradient of water-acetonitrile for 70 minutes using a binary mobile system with **buffer A** and **buffer B** (See Appendix A for preparation). The gradient for

peptide elution was 5% of **buffer B** (See Appendix A for preparation) for 10 minutes, which increased to 35% of buffer B at 70 minutes. After each run, the analytical column was briefly washed in-line with 80% buffer B before equilibration at 2% buffer B.

7.8.9.2. Mass spectrometry

Mass spectrometry is an analytical method used to ionize analytes, which are sorted according to their mass-to-charge (*m/z*) ratios. This analysis is performed by a mass spectrometer, which is a highly sensitive and sophisticated weighing device. Analyte ionization forms an integral part of the technique, as mass spectrometry involves the acceleration of ionized analytes through a magnetic or electrical field, whereby the analytes acceleration is measured in relation to its *m/z* in a vacuum, producing MS1. In mass spectrometry, analytes are ionized in a gaseous phase mainly by two methods: matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), although higher sensitivity is achieved using nanospray ESI (Jorabchi and Smith, 2009). Post-ionization, charged analytes are detected by mass analysers, which include the Orbitrap, Fourier transform ion cyclotron (FTIC), ion trap or time-of-flight (TOF). Proteomics-based MS identification relies on matching sequence tags to theoretical databases, *de novo* interpretation of peptide sequences, or assigning correlation scores for experimental-to-theoretical spectral matching (McHugh and Arthur, 2008; Frank, 2009; Wang and Wilson, 2013).

Tandem MS, also known as MS^2 or MS/MS, involves the detection of analytes and its fragment products by two mass analysers in tandem. Firstly, MS1 are produced as mentioned before, and ions of a particular *m/z* are selected to produce fragment ions. The fragment ions can be unimolecular or bimolecular (Cooks and Mueller, 2013). Unimolecular fragmentation, also known as post source decay (PSC), occurs when single metastable ions fragment suddenly, producing in *y/b* ions. Bimolecular fragmentation, also known as collision-induced diffraction (CID), occurs when precursor ions are bombarded with inert noble gas in a collision cell, also resulting in *y/b* ions. Electron capture dissociation (ECD) and electron transfer dissociation are other form of bimolecular fragmentation, although they typically generate *z/c* fragments (Boersema et al., 2009b). Tandem MS determines protein identity based on peptides and their fragment ions (i.e. bottom-up) (Washburn et al., 2001; Wolters et al., 2001), although intact protein analysis (i.e.

top-town) are reported too (Kelleher, 2004; Zhou et al., 2012). In proteomics, absolute quantitation is achieved using stable isotope labelling by amino acids in cell culture (SILAC) for *in vivo* labelling (Ong et al., 2002; Ong and Mann, 2006), while isobaric tag for relative and absolute quantitation (iTRAQ) (Ross et al., 2004), dimethyl labelling (Boersema et al., 2009a), and tandem mass tags (TMT) (Thompson et al., 2003) are *in vitro* labelling approaches. Label-free identification involves peak intensity measurements and MS1 spectral peak integration using the "top three" approach and iBAQ. The protocol we used is described below:

The QExactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher, San Jose, CA, USA) was used in-line with the UltiMate® 3500 RSnano LC system. Mass spectra were collected in a data-dependent manner with automatic switching between MS and MS/MS scans using a top 10 method. The peptides were ionised by electron spray ionisation, and mass spectra were acquired with a resolution of 70 000 at a target-value of 3×10^6, or a maximum integration time of 250 milliseconds (ms). The scan range was restricted between 300 and 1750 *m/z*. Peptide fragmentation was performed by high-energy collision dissociation (HCD) and normalised collision energy was set at 25. Intensity threshold for ion selection was fixed at 1.7×10^4 with a charge exclusion of z = 1 and z > 5. MS^2 spectra were acquired at a resolution of 17 500, at a target-value of 2×10^5 ions or a maximum integration time of 120 ms and the isolation window set at 4.0 *m/z*.

7.9. Data analysis

7.9.1. CT100plus analysis program

The CT100plus program was developed to provide a customised normalisation and qualitative sample clustering using the parameters described in the following sections:

7.9.1.1. Spot homogeneity

Homogeneous signals are expected for all pixels within a spot. Deviations from this may be due to the "doughnut effect", caused by inadequate pin height; dust particles, from poor storage conditions; high humidity (> 50%), resulting in condensation and uneven intensity. Temperature fluctuations and low humidity (< 35%) can also result in spot evaporation. The CV, which is a ratio

of median intensity and SD of all pixel intensities with in a spot, is used to evaluate variations between replicates – the CV should not exceed 25%.

7.9.1.2. Spot-to-spot variation

In principle, it is expected that signals from triplicate spots should be similar, and that spots across the microarray should be uniform. Deviations from this may be due to "spot bleeding", whereby a spot spread into the background area; "spot merging", whereby spots positioned closely run into each other. "Spot bleeding" and "spot merging" may result from inadequate spotting buffer composition or high humidity (> 50%). Inadequate washing steps may also result in the formation of washing artefacts or speckles, resulting in negative spots or additional smaller spots, resulting in spot-to-spot variation.

7.9.1.3. Signal-to-noise ratio

The net spot intensity, which results from foreground intensity minus background intensity, is expected to be higher than the neighbourhood or local background. However, washing artefacts, speckles and dust particles may result in a background higher than the foreground. The CT100plus program used a signal-to-noise ratio of ≥ 2 as acceptable, where "signal" refers to foreground intensity and the "noise" refers to 2 SD above the background intensity.

7.9.1.4. Pixel saturation

Accurate relative quantitation relies on signal intensity detection in the linear range of the fluorescence scanner. The GenePix scanner has limit of detection of 65535 RFU. If saturated spots are detected, the slides should be rescanned at a lower photomultiplier tube (PMT) gain setting until saturation is no longer detectible across the microarray.

7.9.1.5. Background subtraction and correction

The net intensity is derived by subtracting background from foreground intensity. Inaccurate local background values arise from speckling, dust particles, *et alia*. Alternatively, the neighbourhood background is derived from the background values of neighbouring spots, and accounts for artefact-effects that may arise from using local background. Thus, the CT100plus

software calculates net intensities by subtracting neighbourhood background from the median local foreground pixel intensity.

7.9.1.6. Data normalisation

Normalisation is an efficient and robust method of correcting technical and systematic bias in array-to-array and pin-to-pin discrepancies. For the CT100plus microarray employs total intensity based module described by Quackenbush, 2001 (Quackenbush, 2001), which assumes that all spots have similar intensities, meaning that intensity sum for all positive controls is constant. The scaling method factors in flagged spots – if a positive control spot is flagged, the corresponding spot is flagged across all arrays to minimise discrepancies in the number of positive control spots used across the microarray.

7.9.2. MaxQuant analysis

Proteomics experiments produce large amounts of complex data. Thus, computational analyses, algorithms and workflows are required for systems-level identification of biologically significant differences. The MaxQuant software uses a C++ based algorithm for the analysis of high throughput mass spectrometry proteomic data. MaxQuant allows the proteome-wide identification and quantification of proteins from spectra-based peptide masses generated from precursor scans (Cox and Mann, 2008). MaxQuant is integrated with the Andromeda search engine which does a score-based identification of fragment ions, peptides and protein groups. Proteins that share peptides are grouped together and called protein groups. The Andromeda scoring system is based on the binomial distribution probability (Cox et al., 2011). A binomial experiment requires repeated independent trials which can result in just two possible outcomes (i.e. success or failure), and the probability of success is the same for every trial. Peptide scoring functions are assigned in a binomial distribution based probabilistic algorithm by comparing spectral peak characteristics in experimental to theoretical target databases.

MaxQuant version 1.5.3.12 was used for the analysis. Raw Xcalibur files generated from the mass spectrometer were loaded onto MaxQuant, and an experimental template was generated. The minor modifications (Table 7.8) were used with default experimental parameters (Table 7.9). The number of threads were set to the number of Xcalibur files plus 1.

Table 7.8. Minor modifications to MaxQuant settings.

Parameter	Settings
Threads	Raw files plus 1
iBAQ	Ticked
Upload FASTA file	Yes
Variable modifications	Deimination (R)

Table 7.9. Default MaxQuant settings.

Parameter	Setting
Variable modifications	Oxidation (M) and Acetyl (Protein N-term)
Multiplicity	2
Enzyme	Trypsin (P)
First search PPM	20
Main Search PPM	6
Separate variable modifications for first search	Unticked
Maximum number of modifications per peptide	5
Maximum labelled AAs	3
Maximum missed cleavage	2
Maximum charge	7
Individual peptide mass tolerance	Ticked
Type	Standard
MS/MS tolerance for FTMS	20
MS/MS tolerance for ITMS	0.5
MS/MS tolerance for TOF	0.1
Unknown	0.5
Fixed modifications	Carbamidomethyl
Peptide FDR	0.01
Site FDR	0.01
Maximum peptide PEP	1
Minimum peptides	1
Minimum razor + unique peptide	1
Minimum unique peptide	0
Protein FDR	0.01
Applied site FRD separately	Ticked
Minimum peptide length	6
Minimum score	0
Filter labelled amino acid	Ticked
Second peptide	Ticked
Discard unmodified counterpart peptides	Ticked
Minimum ratio count	2
Site quantification mode	Use least modified peptide
Use of occupancies	Normalised ratio
Protein quantification use only unmodified peptides	Ticked
Requantify	Ticked
Keep low scoring version of identified peptides	No
Match between runs	Unticked
iBAQ	Unticked
Label free quantification	Unticked
Thread	1

7.9.3. RStudio

R is a free, open source, C++ based programming language that is useful for large-scale statistical and graphical computing of high throughput omics data. RStudio (www.rstudio.com/) is an R interface available as a desktop application. Various codes and scripts are deposited as packages in the Comprehensive R Archive Network (CRAN) repository (https://cran.r-project.org/). The RStudio software (version 1.0.136) was used to analyse and visualise our data.

For the analyses, required packages were downloaded from CRAN and installed. The working directory was set to the location of the experimental dataset, which was subsequently loaded on to RStudio. Pre-set scripts were loaded, and the datasets were processed to produce the required statistical outputs or graphs.

7.9.4. Perseus

Perseus is a user-friendly based software for downstream computational analysis of proteomics data. Various statistical tools are integrated into Perseus, including data normalisation, data imputation, *t*-test, analysis of variance (ANOVA), clustering, correlation, enrichment, volcano plots, motif identification and scatter plots.

Perseus (version 1.5.4.1) was used for our analysis. The raw data was loaded in to Perseus, $\log_2$ transformed, the categorical row annotated, missing values replaced by zero, and the relevant statistical tests performed.

7.9.5. Data pre-processing and normalisation

Raw data requires pre-processing before downstream analyses, as noisy data and missing values biases statistical inference and may mask the actual biological meaning of the data. Thus, the filtering steps are applied to make the data more amenable to statistical analysis. The filtering criteria for microarray or mass spectrometry data are defined in the results section of Chapters 2, 3, 4 and 5.

It is also important determine whether the data follows a normal Gaussian distribution to determine whether parametric or non-parametric statistical tests should be applied. Data distribution can be tested using Shapiro-Wilk (Shapiro and Wilk, 1965), Anderson-Darling

(Anderson and Darling, 1952) or Kolmogorov-Smirnov (Feller, 1948) tests. In this study, we used Shapiro-Wilk to test data distribution.

7.9.6. Mann-Whitney *U*, Kruskal-Wallis test and Benjamini-Hochberg correction

The Mann-Whitney *U* test, also known Wilcoxon rank-sum test, is a non-parametric test of the null hypothesis that it is equally likely that a random selected value from one sample is less than or greater than a randomly selected value from a second sample (Mann and Whitney, 1947). Unlike the *t*-test, the Mann-Whitney *U* test does not require the assumption of a normal distribution. The test was used to identify statistical significant differences between HCs and cancer patients, to identify CT100plus antigens associated with clinicopathological features, and to construct volcano plots.

The Kruskal-Wallis test (Kruskal and Wallis, 1952) is like the Mann-Whitney *U* test, although it is used to assess statistical significant differences between multiple groups *i.e.* more than two groups. The test was used to identify CT100plus antigens associated with clinicopathological features.

Type I errors, which are false-positive results, can be reduced by applying multiple testing correction. Multiple correction can be applied adjusting the false discovery rate (FDR) as in the Benjamini-Hochberg test (Benjamini and Hochberg, 1995) or by Bonferroni's correction which maintains family wise error rate (FWER).

7.9.7. Hierarchical clustering

Hierarchical clustering analysis is an analysis method that seeks to build a hierarchy of clusters. Two general clustering strategies are known i.e. agglomerative, which is a *bottom up* approach where each observation starts as its own cluster and pairs of clusters merge as one moves up the hierarchy, and divisive, which is a top down approach where all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

Hierarchical clustering is used to identify group differences, by using a training set (supervised) or by identifying random clustering patters without a training set (unsupervised). In this study, hierarchical clustering was performed in Perseus using $\log_2$ transformed microarray data using

Euclidian distance, average linkage and pre-processing with K-means for the column and row trees.

7.9.8. Principle Component Analysis (PCA)

PCA is multivariate test that searches for correlations between variables that are orthogonally transformed to linear uncorrelated sets of variables known as principle components. The first principle component has the largest variance, and subsequent principle components follow with decreasing variance. The analysis typically produces 1D, 2D or 3D plots for clusters identification. I this study, PCA was performed using Perseus on $\log_2$ transformed microarray data.

7.9.9. Venn diagram

A Venn diagram, also known as a primary diagram, shows all possible logical relations between a finite collection of datasets. The tool is particularly useful for identifying unique and shared between study groups. A free online software was used to produce Venn diagrams in this study, which can be used at http://bioinformatics.psb.ugent.be/webtools/Venn/.

7.9.10. Sample size calculation: area under the ROC curve

The sample size calculator is used to compare the area under the ROC curve with a null hypothesis value. The sample size calculator considers the required significance level and power of the test. The inputs required include (**1**) type I error, which is the probability of making a Type I error (α-level, two sided), *i.e.* the probability of rejecting the null hypothesis when it is in fact true; (**2**) type II error, which is the probability of making a Type II error (β-level), i.e. the probability of accepting the null hypothesis when it is in fact false; (**3**) area under the ROC curve, which is the hypothesised area under the ROC curve; (**4**) a null hypothesis, which is a null-hypothesis AUC-value that is 0.5, indicating that the null hypothesis states no difference between the two groups tested; and (**5**) ratio of sample sizes in negative and positive groups, which is the desired ratio of negative and positive cases. The outputs show the required sample size for different Type I and Type II Error levels.

In this study, the ROC sample size was performed using MedCalc software (version 17.2), with a hypothesised area under the ROC curve of 0.80, a null hypothesis value of 0.5. The remaining

cohort-specific parameters (*e.g.* sample size ratio, α-value and β-value) are described within the relevant Chapters.

7.9.11. ROC analysis

The ROC analysis was used in this study, which is a sensitivity and specificity report of a tested parameter, which in this study refers to antigens.

In a binary classification, the outcomes are labelled as either positive or negative, resulting in four possible outcomes from a binary classifier. If the outcome from a prediction and the actual value are both positive, then it is called a *true positive*; however, if the actual value is negative, then it is said to be a *false-positive*. Conversely, a *true negative* occurs when both the prediction outcome and the actual value are negative, and a *false negative* is when the prediction outcome is negative while the actual value is positive.

To draw a ROC curve, only the *true positive* rate and the *false positive* rate are needed. The *true positive* rate (also known as the sensitivity) defines how many correct positive results occur amounts all of the positive samples in the tests. The *false positive* rate (equal to 1 - specificity) defines how many incorrect positive results occur among all negative samples in the test. The ROC space is defined by FPR and TPR as x- and y-axes, depicting the relative trade-offs between true-positives and false-positive. Thus, each prediction result of the matrix represents one point in the ROC space.

In our study, the ROC analysis produces an AUC-value, a measure of the discriminating efficacy of antigen for patients from HCs; a sensitivity-value, indicating the fraction of patients with high cancer antigen intensity; and a specificity-value, indicating the fraction of HCs with no or significantly lower signal intensity for antigens (Metz, 1978; Vining and Gladish, 1992; Zweig and Campbell, 1993). The higher the AUC-, sensitivity- and specificity- values, the better an antigen distinguishes patients from HCs. The ROC curve is widely accepted for selecting an optimal cut-off point. The cut-off point is used to determine the presence or absence of a disease, where values above the cut-off are disease-positive, and values below the cut-off are considered disease-negative (Akobeng, 2007).

In this study, we performed the ROC analysis in RStudio using the ROCR package and CombiROC (http://www.combiroc.eu/).

7.10. Databases

7.10.1. NCBI protein BLAST

Basic local alignment search tool (BLAST) is an algorithm used to compare primary biological sequences of information, like proteins or nucleotides. The BLAST search allows the comparison between a query sequence with a database of sequences. In the study, the National Centre for Biotechnology Information (NCBI) BLAST tool (https://blast.ncbi.nlm.nih.gov/) was used which produces a query cover-value, E-value, identify-value which provides scores for the comparison between the query sequence and the database sequence.

7.10.2. Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)

STRING (http://string-db.org) is a web-based and biological database source used to interrogate protein-protein interactions (Szklarczyk et al., 2011, 2015). The data sources in STRING are derived from curated literature, computational predictions and experimental data, with a resulting physical interaction network assigned weighted confidence scores. The majority of STRING associations are from genome-based prediction algorithms or by pre-computed interlog transfer. The probabilistic confidence score is derived from separately benchmarked associated groups against the manually curated functional scheme of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. For the analysis, gene or protein lists are entered, and the organism species is specified. The output is a network which includes the prediction of co-expression, neighbourhood evidence, occurrence in similar biological pathway and gene fusion events (Szklarczyk et al., 2011).

7.10.3. AAgAtalas v1.0

The human autoantigen database (AAgAtlast database 1.0) provides a user-friendly interface to browse, retrieve and download human autoantigens as well as their associated diseases that is freely accessible at http://biokb.ncpsb.org/aagatlas/. To develop the, AAgAtlast database 1.0, the authors extracted 45830 autoantigen-related abstracts and 94313 sentences from PubMed using

“autoantigen” or “autoantibody” or their lexical variants as keywords. These abstracts were reduced to 25520 abstract and 3984 autoantigen candidates using bio-entity recogniser based on protein ontology. In total, 1126 genes were identified as autoantigens related to 1071 diseases which was used to construct the AAgAtlast database 1.0. The database is a valuable resource for understanding human autoantigens and their research functions (Wang et al., 2017).

Appendix A: Preparation of stock and working solutions and buffers

A.1. Slide coating buffer (150 mM NaPO4, pH 8.5)

150 mM NaH_2PO_4, pH 4.0 (solution 1)
- Dissolve 9 g NaH_2PO_4 in 500 ml triple-distilled Millipore water

150 mM Na_2HPO_4, pH 9.0. (solution 2)
- Dissolve 26.7 g Na_2HPO_4 in 800 ml triple-distilled Millipore water.

150 mM NaPO4, pH 8.5, 0.001% Tween®-20
- Titrate 150 mM Na_2HPO_4 (solution 1) with 150 mM NaH_2PO_4 (solution 2) until a pH 8.5 is reached.
- Add 100 µl Tween®-20 to 100 ml 150 mM $NaPO_4$, pH 8.5.

A.2. Slide washing buffer (50 mM KPO_4, pH 8.0)

1 M KPO_4
- Dissolve 13.94 g K_2HPO_4 and 2.72 g KH_2PO_4 to a final volume of 100 ml triple-distilled Millipore water.
- Adjust pH with KOH or H_3PO_4 if needed.

50 mM KPO_4, pH 8.0
- Dilute 50 ml of 1 M KPO4 in 800 ml triple-distilled Millipore water.
- Adjust to pH 80. With KOH.
- Add triple-distilled Millipore water to a final volume of 1000 ml.

A.3. Quality control buffer (150 mM NaPO4, pH 8.5)

150 mM NaH_2PO_4, pH 4.0 (solution 1)
- Dissolve 9 g NaH_2PO_4 in 500 ml triple-distilled Millipore water

150 mM Na_2HPO_4, pH 9.0. (solution 2)
- Dissolve 26.7 g Na_2HPO_4 in 800 ml triple-distilled Millipore water.

150 mM NaPO4, pH 8.5, 0.001% Tween®-20
- Titrate 150 mM Na_2HPO_4 (solution 1) with 150 mM NaH_2PO_4 (solution 2) until a pH 8.5 is reached.
- Add 100 µl Tween®-20 to 100 ml 150 mM $NaPO_4$, pH 8.5.

A.4. CT100plus control Buffer

- Mix the reagents to the specified final concentration (Table A1), and add triple-distilled Millipore water.

Table A1. Table of basic reagents in CT100plus control buffer.

Reagent	Stock solution	Final concentration	Manufacturer
2.5 M HEPES, pH 7.5	2.5 M	12.5 mM	SIGMA®
KCl	1 M	25 mM	EMSURE
Glycerol	100%	10 %	univAR
Triton X-100	10 %	0.05 %	SIGMA®
DTT	1 M	1 mM	SIGMA®
Sucrose	80 %	20% (40% for Cy5-biotin-BSA)	

A.5. Blocking buffer

- Mix the reagents to the specified final concentration (Table A2), and add triple-distilled Millipore water.

Table A2. Table of basic reagents in CT100plus blocking buffer.

Reagent	Stock solution	Final concentration	Manufacturer
2.5 M HEPES, pH 7.5	2.5 M	25 mM	SIGMA®
KCl	1 M	50 mM	EMSURE
Glycerol	100%	20 %	univAR
Triton X-100	10 %	0.1 %	SIGMA®
Biotin	50 mM	50 µ	SIGMA®
DTT	1 M	1 mM	SIGMA®

- Add DTT just before use.

A.6. PAD4 buffer (prepare fresh before use**)**

- Mix the reagents to the specified final concentration (Table A3), and add triple-distilled Millipore water.

Table A3. Table of basic reagents in PAD4 buffer.

Reagent	Stock solution	Final concentration	Manufacturer
2.5 M HEPES, pH 7.5	2.5 M	50mM	SIGMA®
$CaCl_2$	500 mM	5 mM	SIGMA®
NaCl	1000 mM	150 mM	Merck
Tween®-20	100 %	0.2 %	SIGMA®

A.7. 1× PBST
- 5 PBS tablets (SIGMA®) in triple distilled Millipore water up to 1 L.
- All 1 ml 100 % Tween® 20 (uniLAB).
Store at RT.

A.8. Binding buffer
Prepare fresh before use:
- 1× PBS tablet (SIGMA®) in triple distilled Millipore water up to 200 ml.
- All 50 µl 100 % Triton™ X-100 (SIGMA®).
Store at RT.

A.9. Non-denaturing lysis buffer (Prepare fresh before use**)**

- Mix the reagents to the specified final concentration (Table A4), and add triple-distilled Millipore water.

Table A4. Table of basic reagents in non-denaturing lysis buffer.

Reagent	Stock solution	Final concentration	Manufacturer
2.5 M HEPES, pH 7.5	2.5 M	12.5 mM	SIGMA®
KCl	1 M	25 mM	EMSURE, MERCK
Protease inhibitor	10 ×	1 ×	ROCHE
Benzonase	250 U	1U/µl	SIGMA®

- Add detergents to a final concentration of 1% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) (SIGMA®), 1% Triton™ X-100 (SIGMA®), 1% IGEPAL® CA-630 (SIGMA®), a mixture detergents (0.25% CHAPS, 0.5% IGEPAL® CA 630, and 0.5% Triton™ X-100), or no detergent.

A.10. 5× Laemmli buffer

- Dissolve 1.75 g Tris (SIGMA®), 30 ml glycerol (KIMIX) and 5 g SDS (SIGMA®) to a final volume of 50 ml triple-distilled Millipore water, pH 6.0.
- Just before use, add 5 µl 25% bromophenol blue and 150 mM DTT
- Dilute to 1× laemmli buffer in triple-distilled Millipore water.

A.11. 4% Stacking gel

Prepare immediately before use:
- 0.625 ml 1 M Tris (SIGMA®) at pH 6.8.
- 0.65 ml 40% acrylamide/bisacrylamide (SIGMA®)
- 50 µl 10 % SDS (SIGMA®)
- 3.6 ml triple-distilled Millipore water.
- 25 µl 20 % ammonium persulfate (APS) (SIGMA®)
- 5 µl 100 % *N,N,N',N'*-Tertramethylethylenediamine (TEMED) (SIGMA®)

A.12. 10% Separating gel

Prepare immediately before use:
- 3.75 ml 1 M Tris (SIGMA®) at pH 8.8.
- 2.5 ml 40% acrylamide/bisacrylamide (SIGMA®).
- 100 µl 10 % SDS (SIGMA®).
- 3.615 ml triple-distilled Millipore water.
- 25 µl 20 % APS (SIGMA®)
- 10 µl 100 % (TEMED) (SIGMA®)

A.13. 10× Running buffer

- 3.03 g Tris (SIGMA®)
- 14.5 g glycine (SIGMA®)
- 35 mM SDS (SIGMA®)
- Dissolve in 1 litre (L) triple-distilled Millipore water.

Store at RT.

A.14. 10× Transfer buffer

- 3.03 g Tris (SIGMA®)
- 14.5 g glycine (SIGMA®)
- Dissolve in 1 litre (L) triple-distilled Millipore water.

Store at RT.

A.15. 1× TBST
- 50 ml 1 M tris (SIGMA®)
- 20 ml 5 M NaCl (Merck)
- 250 µl 100% Tween® 20
- Dissolve in 100 ml triple-distilled Millipore water.

Store at RT.

A.16. Elution buffer 1

Prepare immediately before use
- 2M urea (SIGMA®)
- 50 mM Tris-HCl (pH 7.5)
- 5 µg/µl trypsin (Promega)

A.17. Elution buffer 2

Prepare immediately before use
- 2M urea (SIGMA®)
- 50 mM Tris-HCl (pH 7.5)
- 1 mM DTT (SIGMA®)

A.18. Solvent A
- 2% acetonitrile (ACN) (SIGMA®) and 0.1% formic acid (FA) (SIGMA®) in triple-distilled Millipore water.

A.19. Solvent B
- 80% ACN (SIGMA®) and 0.1% FA (SIGMA®) in triple-distilled Millipore water.

A.20. Solvent C
- 60% ACN (SIGMA®) and 0.1% FA (SIGMA®) in triple-distilled Millipore water.

A.21. Buffer A
- 0.1% ACN (SIGMA®) and 0.1% FA (SIGMA®) in HPLC grade water.

A.22. Buffer B
- 70% ACN (SIGMA®) and 0.1% FA (SIGMA®) in HPLC grade water.

Appendix B: Supplementary Figures and Tables

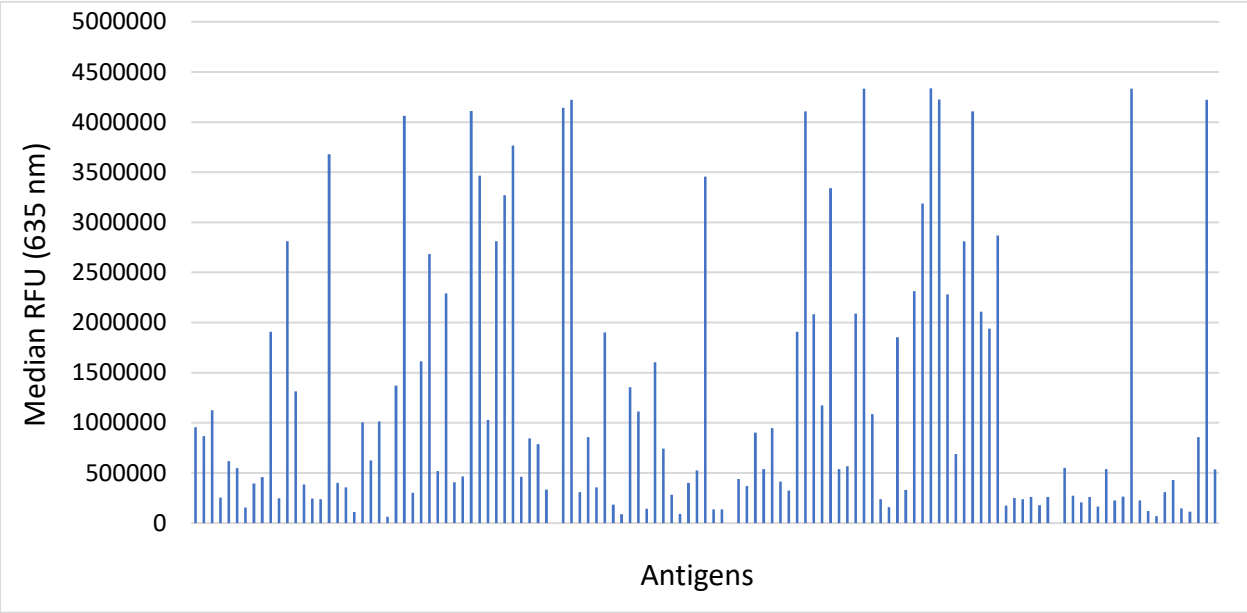
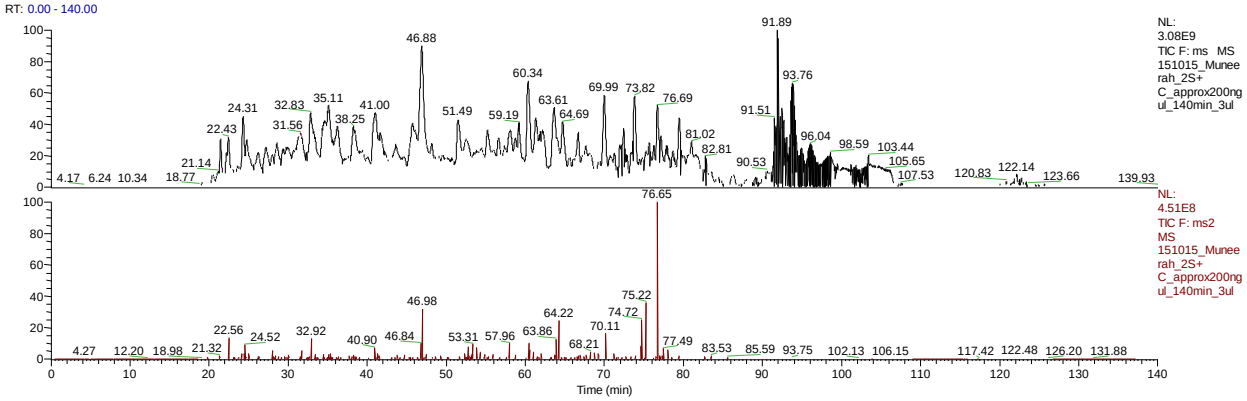


Figure B1. Signal intensity for antigens on anti-c-Myc-treated CT100plus microarray. The CT100plus microarray antigens each have a c-Myc tag which is used as a quality control measure to confirm the presence of the CT100plus antigens. For the assay, the CT100plus microarray is treated with Cy3-labelled anti-c-Myc antibody, detected at 532 nm.

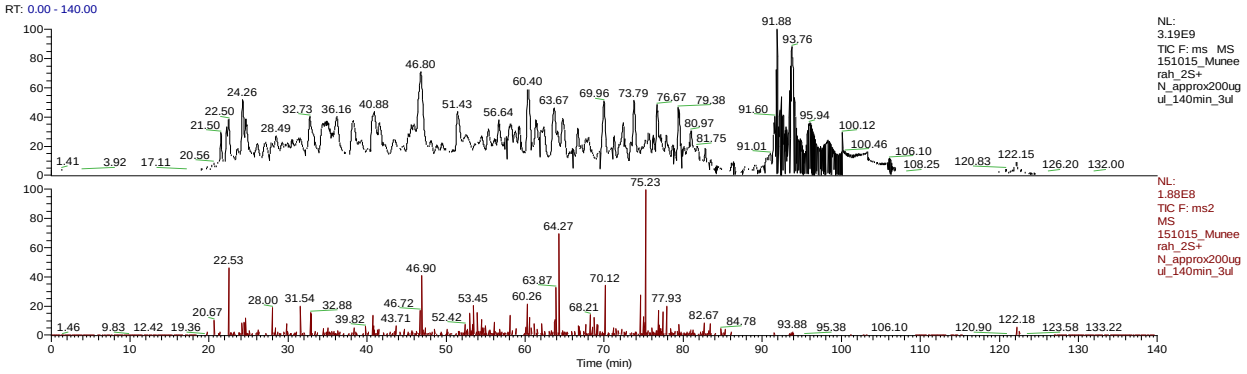
Table B1. Potential CRC biomarkers identified by CT100plus microarray and AP-MS. CRC patients and HC autoantibody signals against CT100plus antigens were assayed and the top 10 antigens with the highest AUC-values were identified. These proteins were searched against proteins identified by AP-MS (See Chapter 4), and COL6A1 and THEG were identified. *Cancer* refers to the AP-MS experiment with patient CRC patient plasma and tissue; *Normal* refers to the AP-MS experiment with patient CRC patient plasma and normal colon tissue; and *Plasma* refers to the AP-MS experiment with CRC patient plasma only. Abbreviations include: CRC = colorectal cancer; HC = healthy control; AUC = area under the receiver operator characteristic curve; AP-MS = affinity purification-mass spectrometry.

Patient	**CRC002**			**CRC004**			**CRC017**			**CRC021**			**CRC031**		
Sample	**Cancer**	**Normal**	**Plasma**	**Cancer**	**Normal**	**Plasma**	**Cancer**	**Normal**	**Plasma**	**Cancer**	**Normal**	**Plasma**	**Cancer**	**Normal**	**Plasma**
COL6A1	Yes	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No
THEG	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes

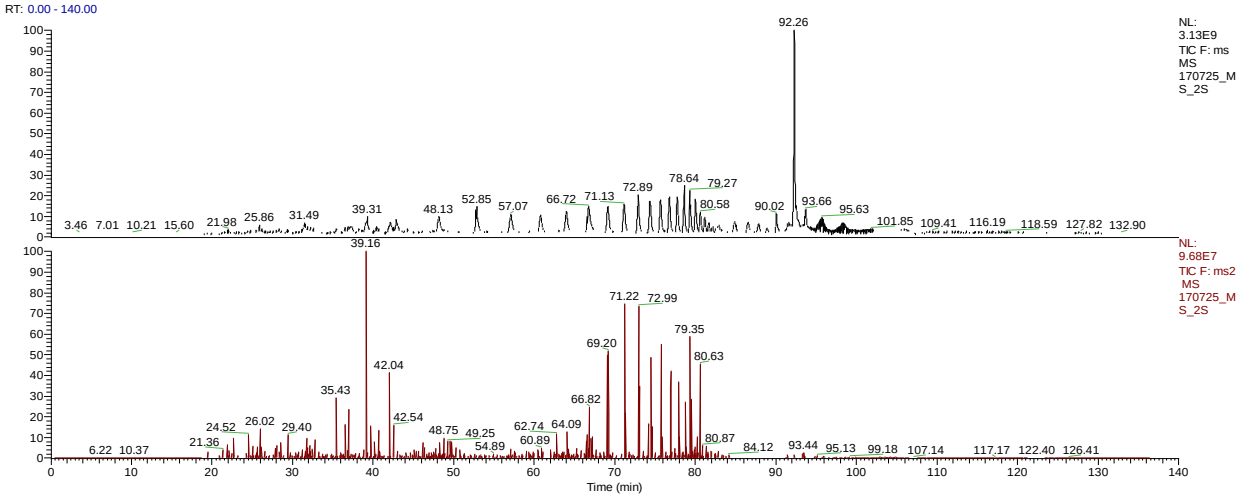
CRC002, *Cancer*



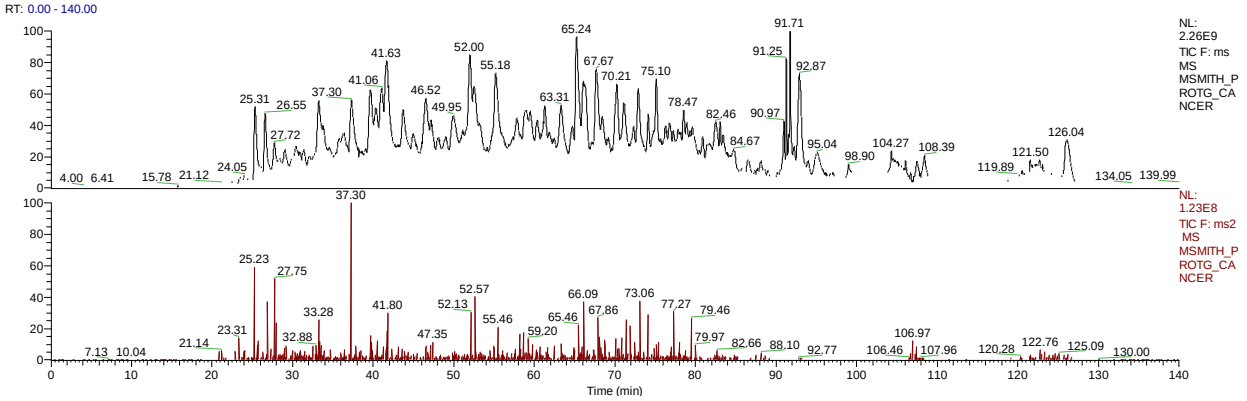
CRC002, *Normal*



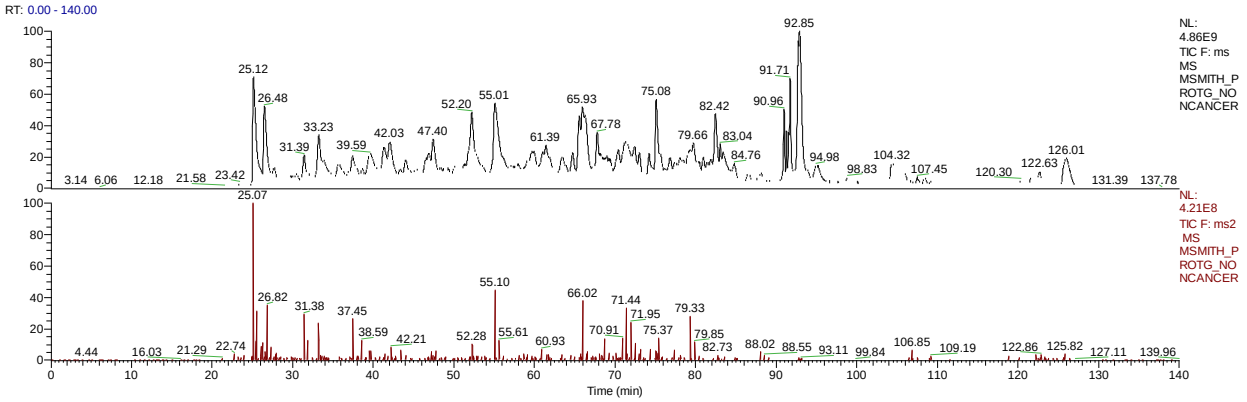
CRC002, *Plasma*



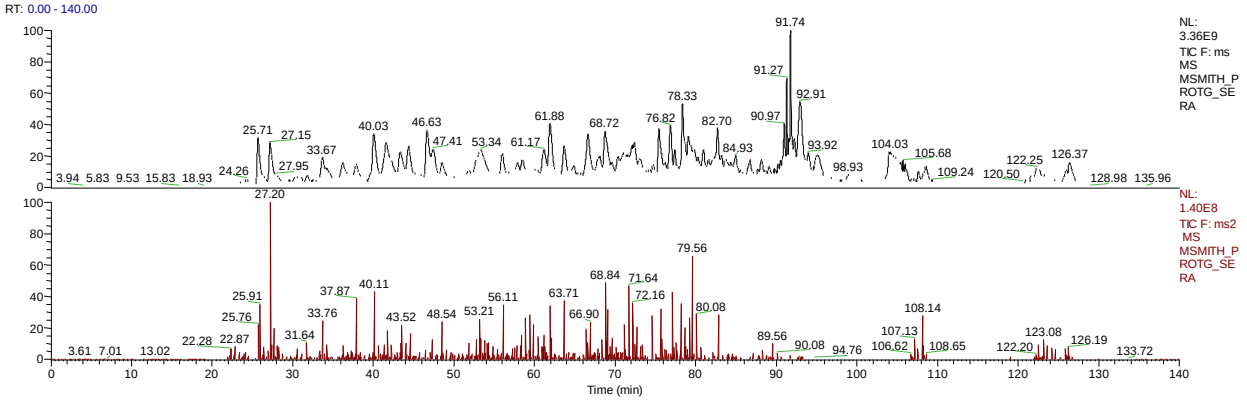
CRC004, *Cancer*



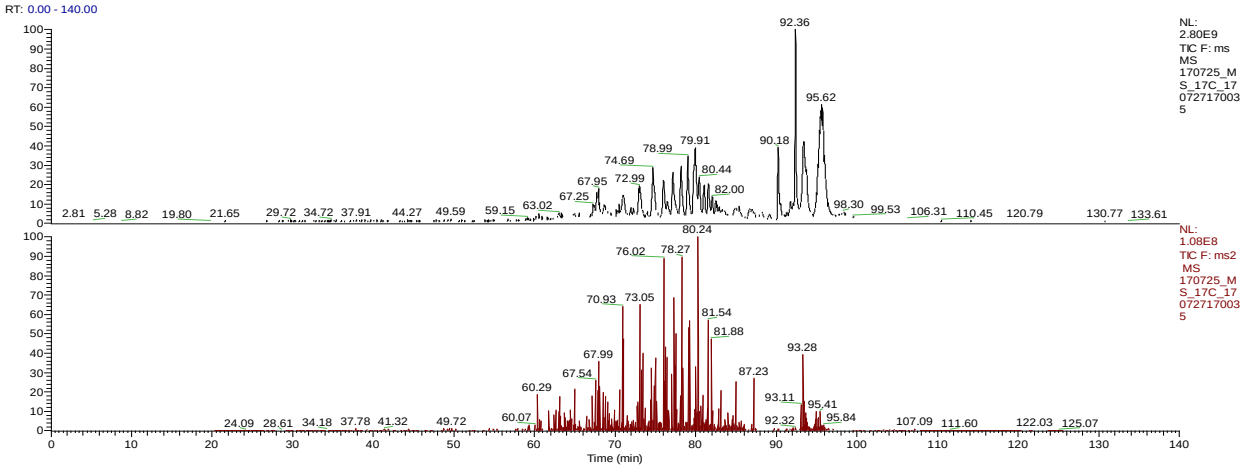
CRC004, *Normal*



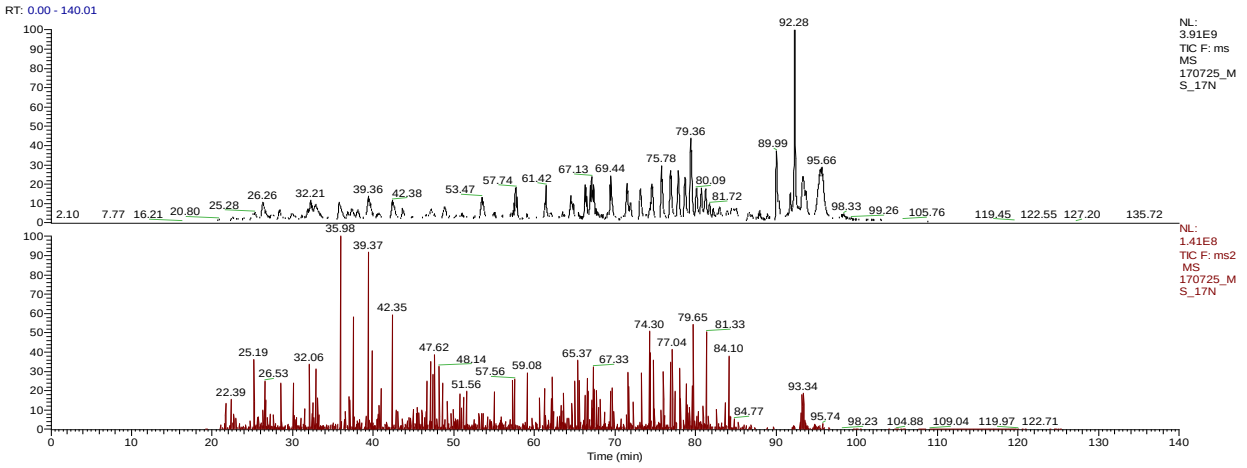
CRC004, *Plasma*



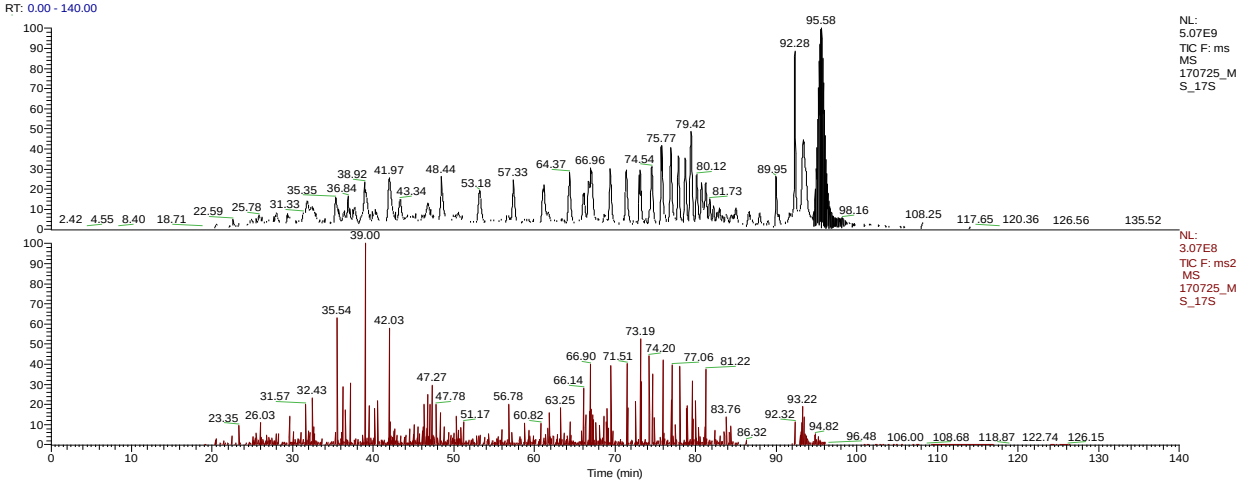
CRC0017, *Cancer*



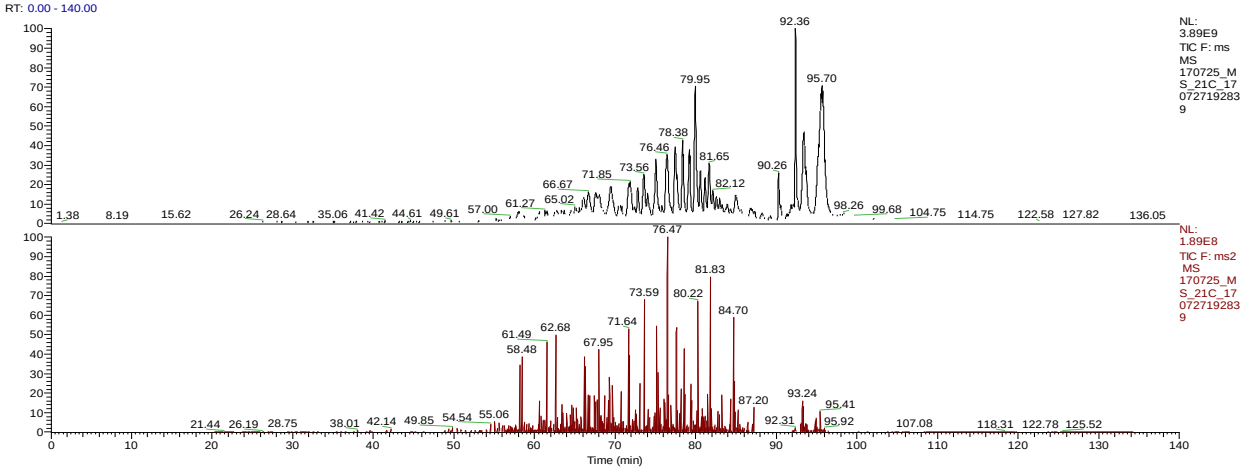
CRC0017, *Normal*



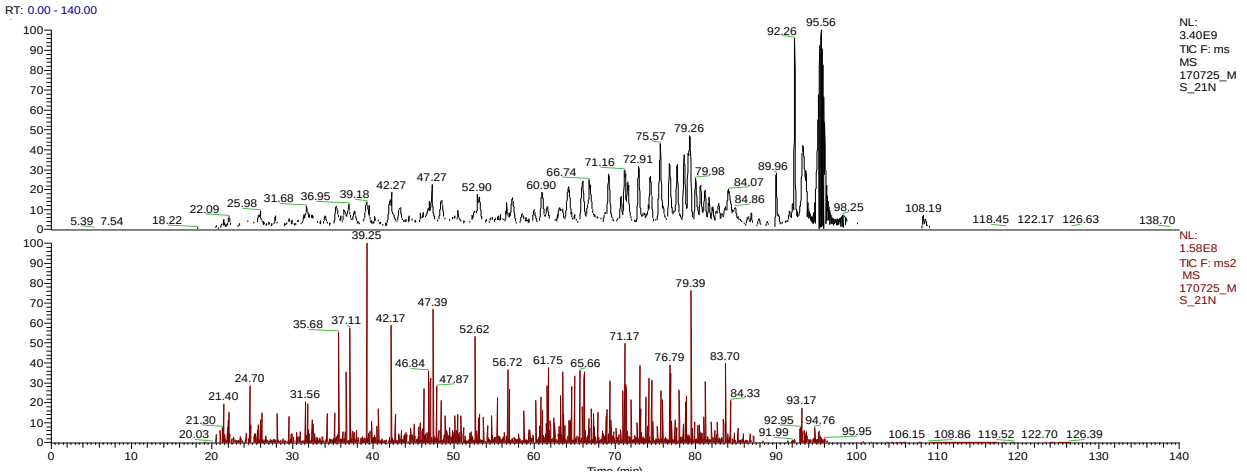
CRC0017, *Plasma*



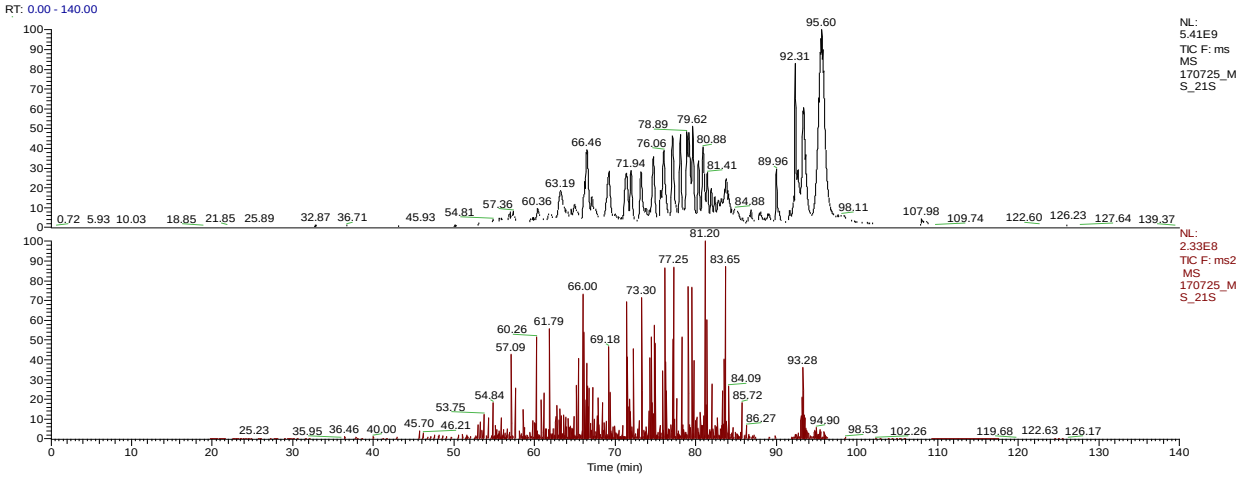
CRC0021, *Cancer*



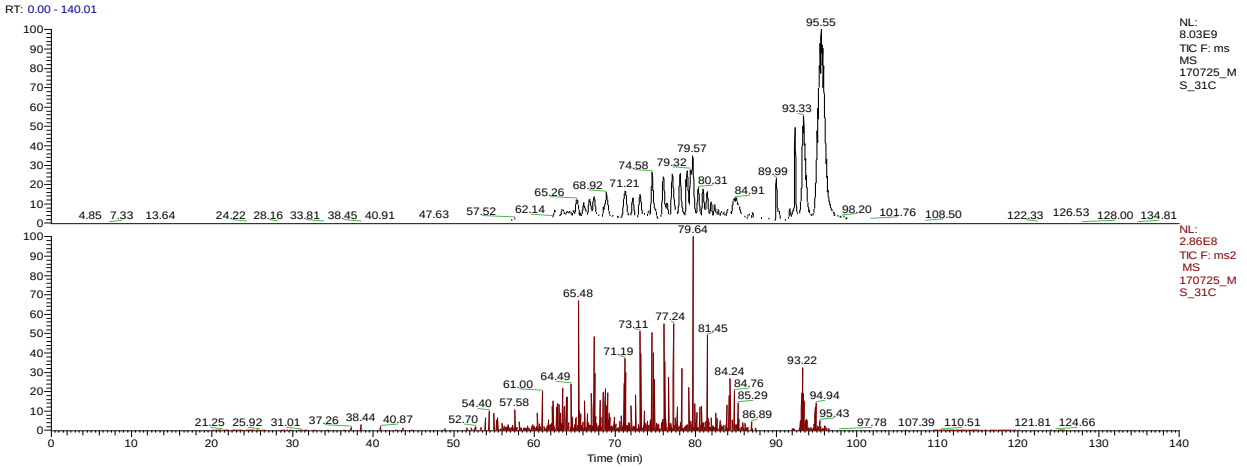
CRC0021, *Normal*



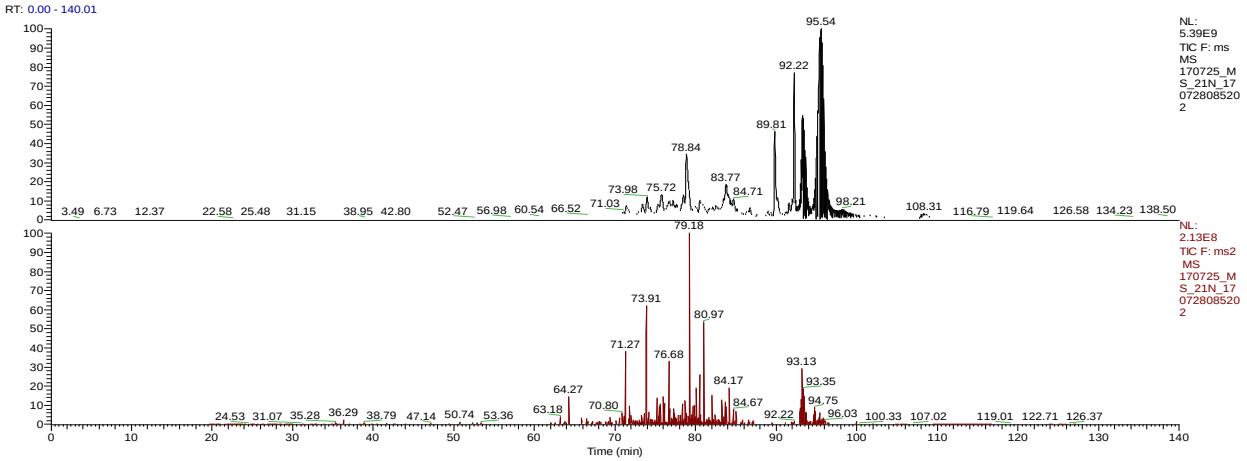
CRC0021, *Plasma*



CRC0031, *Cancer*



CRC0031, *Normal*



CRC0031, *Plasma*

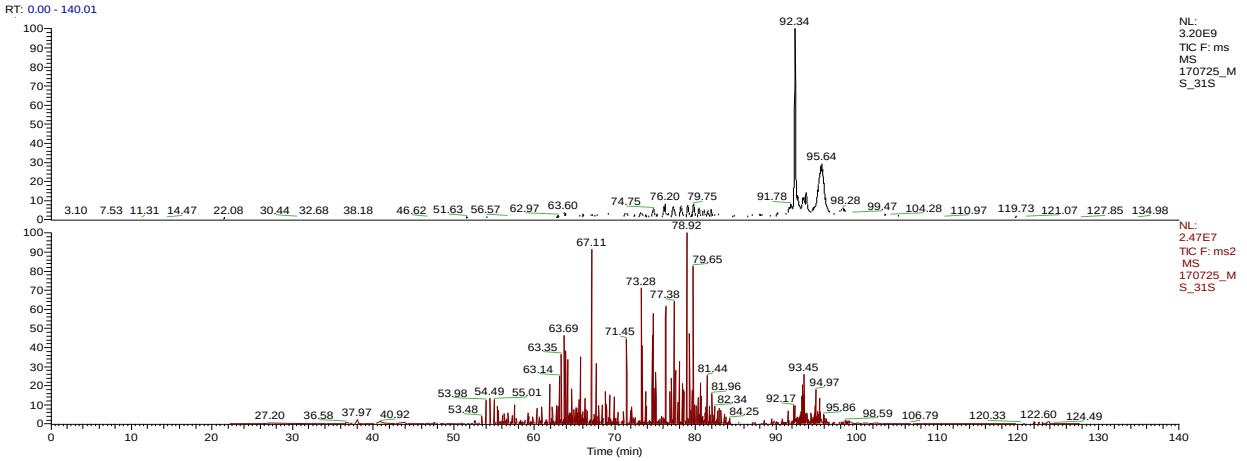


Table B2. Citrullination sites for *Cancer* sample of CRC002

Protein Gene Symbol	Peptide (modification)
A0A075B6N8	SSAVR(1)LR(0.999)SSVPGVR
A0A087WUK2	SCDTPPPCPR(1)
A0A087WVV2	GSQQYR(1)ALTVPELTQQMFDAK
A0A0X1KG71	GSQQYR(1)ALTVPELTQQMFDAK
ACTA	DLYANNVLSGGTTMYPGIADR(1)MQK LDINTLTGEER(1)VPVVNK
ACTB	PR(1)HQGVMVGMGQK AGFAGDDAPR(1)AVFPSIVGRPR R(1)NLGSINTELQDVQR NTR(1)TNEKVELQELNDR
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
CAP1	LQIWDTAGQER(1)FR(0.999)TITSSYYR(0.001)
CHD7	R(1)SR(1)NTDEMVELR
COG5	LEAVSHTSDMHR(1)GYADSPSK
EF1A3	DGNASGTTLLEALDCILPPTR(0.024)PTDKPLR(0.976)LPLQDVYK
EFTU	FAVR(1)DMR(1)QTVAVGVIK
EHD4	RGPAEESSSWR(0.986)DSSR(0.014)
EIF3C	YEEIDNAPEER(0.972)AR(0.028)
EPIPL	SYR(1)TDISMSDFENSR
FIBA	SR(1)TSVQTEDDQLIAGQSAR
FLNA	ADIQTER(1)AYQK
G3P	TASGSSVTSLDGTR(0.964)SR(0.036) THLENNR(1)FGGSGSQVDSAR LWR(1)DGR(1)GALQNIIPASTGAAK
GFPT1	SSSHSR(1)AGQSAAGAAPGGGVDTR
H0YFA4	LWR(1)DGR(1)GALQNIIPASTGAAK
HNRPK	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK
HNRPU	SR(1)ANEEYGLR
IF4A3	GPSR(1)ASSVTTFTGEPNTCPR
IMA3	IREEYPDR(1)IMNTFSVVPSPK
IQGA1	SLQER(1)NQEQEKVER
J3QR09	LTR(1)GEADR(0.991)DTYR(0.009)
K2C8	LLEGEESR(1)LESGMQNMSIHTK
LAMB4	HVPR(1)AVFVDLEPTVIDEVR
NDRG1	AQALR(1)DNSTMGYMMAK
PDCD4	AQALR(1)DNSTMGYMAAK
POTEE	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK
Q5JP53	ISLPLPNFSSLNLR(1)ETNLDSLPLVDTHSK TSYSYR(1)QSSATSSFGGLGGGSVR
RAB1A	LQIWDTAGQER(1)FR(0.999)TITSSYYR(0.001) PR(1)HQGVMVGMGQK
RS14	AVFPSIVGR(0.828)PR(0.172)
SAMH1	TESASVQGR(1)NTDVAQSPEAPKQEAPAK
TBA1B	AVLVDLEPGTMDSVR(0.997)SGPFGQIFR(0.003)PDNFVFGQSGAGNNWAK AVFVDLEPTVIDEVR(0.984)TGTYR(0.016) SR(1)NTDEMVELR
TBB3	EVVTSEDGSDCPEAMDLGTLSGIGTLDGFR(0.77)HR(0.23)
TBB4A	GSCNLSR(1)VDSTTCLFPVEEK
TBB4B	ADNEKLDNQR(1)LK
TCPD	MR(1)PEIAGLK
VIME	PENVAPR(0.999)SGATAGAAGGR(0.001) TTEYQLSTLEER(1)DIK AQYEDIANR(0.98)SR(0.02) LHFFMPGFAPLTSR(1)GSQQYR(1)ALTVPELTQQVFDAK TVETR(1)DGQVINETSQHHDDLE
ZO2	FAVR(1)DMR(1)QTVAVGVIK

Table B3. Citrullination sites for *Normal* sample of CRC002.

Protein Gene Symbol	Peptide (modification)
A0A075B6N8	SSAVR(1)LR(0.999)SSVPGVR(0.001)
A0A087WVV2	GSQQYR(1)ALTVPELTQQMFDAK
A0A0X1KG71	GSQQYR(1)ALTVPELTQQMFDAK
ACTA	DLYANNVLSGGTTMYPGIADR(1)MQK LDINTLTGEER(1)VPVVNK
ACTB	LHFFMPGFAPLTSR(0.81)GSQQYR(0.19)ALTVPELTQQMFDAK PR(1)HQGVMVGMGQK AGFAGDDAPR(1)AVFPSIVGRPR NTR(1)TNEKVELQELNDR
CAP1	LQIWDTAGQER(1)FR(0.999)TITSSYYR(0.001)
CHD7	R(1)SR(1)NTDEMVELR
COG5	LEAVSHTSDMHR(1)GYADSPSK
CTNA2	DLYANTVLSGGTTMYPGIADR(1)MQK
EF1A3	ATTATMATSGSAR(1)K
EFTU	FAVR(1)DMR(1)QTVAVGVIK
EIF3C	YEEIDNAPEER(0.972)AR(0.028)
EPIPL	SYR(1)TDISMSDFENSR
FLNA	ADIQTER(1)AYQK
G3P	TASGSSVTSLDGTR(0.964)SR(0.036)
GFPT1	SSSHSR(1)AGQSAAGAAPGGGVDTR
HNRPK	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK
HNRPU	SR(1)ANEEYGLR
HS90B	LHFFMPGFAPLTSR(1)GSQQYR(1)ALTVPELTQQVFDAK
IMA3	IREEYPDR(1)IMNTFSVVPSPK
IQGA1	SLQER(1)NQEQEKVER
J3QR09	LTR(1)GEADR(0.991)DTYR(0.009)
K1C20	ISNSR(1)HTVNYGSDLTGGGDLFVGNEK
K2C8	LLEGEESR(1)LESGMQNMSIHTK
NDRG1	AQALR(1)DNSTMGYMMAK
PDCD4	AQALR(1)DNSTMGYMAAK
POTEE	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK
Q5JP53	ISLPLPNFSSLNLR(1)ETNLDSLPLVDTHSK TSYSYR(1)QSSATSSFGGLGGGSVR
RAB1A	LQIWDTAGQER(1)FR(0.999)TITSSYYR(0.001) PR(1)HQGVMVGMGQK
RS14	AVFPSIVGR(0.828)PR(0.172)
S4R435	MQR(1)ADSEQPSKRPR
TBA1B	AVLVDLEPGTMDSVR(0.997)SGPFGQIFR(0.003)PDNFVFGQSGAGNNWAK AVFVDLEPTVIDEVR(0.984)TGTYR(0.016) SR(1)NTDEMVELR
TBB4A	GSCNLSR(1)VDSTTCLFPVEEK
VIME	PENVAPR(0.999)SGATAGAAGGR(0.001) LHFFMPGFAPLTSR(1)GSQQYR(1)ALTVPELTQQVFDAK AQYEDIANR(0.98)SR(0.02) TTEYQLSTLEER(1)DIK FADLSEAANR(1)NNDALR TVETR(1)DGQVINETSQHHDDLE
ZO2	FAVR(1)DMR(1)QTVAVGVIK

Table B4. Citrullination sites for *Plasma* sample of CRC002.

Protein	Peptide (modification)
ACTA	LDINTLTGEER(1)VPVVNK
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
H0YAW0	ETDTASESSYQLSR(1)HK

Table B5. Citrullination sites for *Cancer* sample of CRC004.

Protein	Peptide (modification)
A0A075B6N8	SSAVR(1)LR(0.999)SSVPGVR(0.001)
A0A087WUK2	SCDTPPPCPR(1)
ACTA	LDINTLTGEER(1)VPVVNK
ACTB	NTR(1)TNEKVELQELNDR AGFAGDDAPR(1)AVFPSIVGRPR R(1)NLGSINTELQDVQR
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
CTNA2	DLYANTVLSGGTTMYPGIADR(1)MQK
EFTU	FAVR(1)DMR(1)QTVAVGVIK
EIF3C	YEEIDNAPEER(0.972)AR(0.028)
FLNA	ADIQTER(1)AYQK
GFPT1	SSSHSR(1)AGQSAAGAAPGGGVDTR
HNRPK	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK LTGMAFR(1)VPTANVSVVDLTCR
HNRPU	SR(1)ANEEYGLR
IMA3	IREEYPDR(1)IMNTFSVVPSPK
IQGA1	SLQER(1)NQEQEKVER
J3QR09	LTR(1)GEADR(0.991)DTYR(0.009)
K1C19	MELER(1)PGGNEITR
K1C20	ISNSR(1)HTVNYGSDLTGGGDLFVGNEK
M0QZC5	SYELPDGQVITIGNER(0.965)FR(0.035)
NDRG1	AQALR(1)DNSTMGYMMAK
PDCD4	AQALR(1)DNSTMGYMAAK
PDXD1	LTAIIEEAEEAPGAR(1)PQLQDAWR(0.767)GPR(0.233) LTAIIEEAEEAPGAR(1)PQLQDAWR(0.77)GPR(0.23)
POTEE	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK
Q5JP53	TSYSYR(1)QSSATSSFGGLGGGSVR
RS14	AVFPSIVGR(0.828)PR(0.172)
S4R435	MQR(1)ADSEQPSKRPR
SAMH1	TESASVQGR(1)NTDVAQSPEAPKQEAPAK
TBA1B	AVFVDLEPTVIDEVR(0.984)TGTYR(0.016) SR(1)NTDEMVELR
TBB3	EVVTSEDGSDCPEAMDLGTLSGIGTLDGFR(0.77)HR(0.23)
TBB4A	GSCNLSR(1)VDSTTCLFPVEEK
TBB4B	ADNEKLDNQR(1)LK
VIME	FADLSEAANR(1)NNDALR LHFFMPGFAPLTSR(1)GSQQYR(1)ALTVPELTQQVFDAK TVETR(1)DGQVINETSQHHDDLE
ZO2	FAVR(1)DMR(1)QTVAVGVIK

Table B6. Citrullination sites for *Normal* sample of CRC004.

Protein	Peptide (modification)
A0A075B6N8	SSAVR(1)LR(0.999)SSVPGVR(0.001)
ACTA	LDINTLTGEER(1)VPVVNK
ACTB	NTR(1)TNEKVELQELNDR AGFAGDDAPR(1)AVFPSIVGRPR
EF1A3	DGNASGTTLLEALDCILPPTR(0.024)PTDKPLR(0.976)LPLQDVYK
EFTU	FAVR(1)DMR(1)QTVAVGVIK
FLNA	ADIQTER(1)AYQK
HNRPK	AGFAGDDAPR(1)AVFPSIVGR(1)PR(1)HQGVMVGMGQK
HNRPU	SR(1)ANEEYGLR
IMA3	IREEYPDR(1)IMNTFSVVPSPK
IQGA1	SLQER(1)NQEQEKVER
K1C20	ISNSR(1)HTVNYGSDLTGGGDLFVGNEK
PDCD4	AQALR(1)DNSTMGYMAAK
PDXD1	LTAIIEEAEEAPGAR(1)PQLQDAWR(0.767)GPR(0.233) LTAIIEEAEEAPGAR(1)PQLQDAWR(0.77)GPR(0.23)
PSMD2	DGNASGTTLLEALDCILPPTR(0.5)PTDKPLR(0.5)LPLQDVYK
Q5JP53	TSYSYR(1)QSSATSSFGGLGGGSVR
TBA1B	SR(1)NTDEMVELR
VIME	TVETR(1)DGQVINETSQHHDDLE FADLSEAANR(1)NNDALR LHFFMPGFAPLTSR(1)GSQQYR(1)ALTVPELTQQVFDAK

Table B7. Citrullination sites for *Plasma* sample of CRC004.

Protein	Peptide (modification)
A0A075B6N8	SSAVR(1)LR(0.999)SSVPGVR(0.001)
ACTA	LDINTLTGEER(1)VPVVNK
EHD4	RGPAEESSSWR(0.986)DSSR(0.014)
HNRPU	SR(1)ANEEYGLR
IMA3	IREEYPDR(1)IMNTFSVVPSPK
K1C20	ISNSR(1)HTVNYGSDLTGGGDLFVGNEK
PDCD4	AQALR(1)DNSTMGYMAAK
VIME	TVETR(1)DGQVINETSQHHDDLE FADLSEAANR(1)NNDALR AQYEDIANR(0.98)SR(0.02)

Table B8. Citrullination sites for *Cancer* sample of CRC017.

Protein	Peptide (modification)
A0A075B6N8	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	IQDLNLSR(0.825)QAKADQLR(0.175)

Table B9. Citrullination sites for *Normal* sample of CRC017.

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	LDINTLTGEER(1)VPVVNK
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
H0YAW0	ETDTASESSYQLSR(1)HK
IMA3	IREEYPDR(1)IMNTFSVVPSPK
K2C8	LLEGEESR(1)LESGMQNMSIHTK
PDCD4	AQALR(1)DNSTMGYMAAK
VIME	TVETR(1)DGQVINETSQHHDDLE

Table B10. Citrullination sites for *Plasma* sample of CRC017.

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	LDINTLTGEER(1)VPVVNK
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)

Table B11. Citrullination sites for *Cancer* sample of CRC021.

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	LDINTLTGEER(1)VPVVNK
H0YAW0	ETDTASESSYQLSR(1)HK
S4R435	MRYLQQKVTR(1)

Table B12. Citrullination sites for *Normal* sample of CRC021.

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	LDINTLTGEER(1)VPVVNK
ACTB	NTR(1)TNEKVELQELNDR
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
EHD4	RGPAEESSSWR(0.986)DSSR(0.014)
H0YAW0	ETDTASESSYQLSR(1)HK
K2C8	LLEGEESR(1)LESGMQNMSIHTK
S4R435	MRYLQQKVTR(1)
VIME	TVETR(1)DGQVINETSQHHDDLE

Table B13. Citrullination sites for *Plasma* sample of CRC021:

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	LDINTLTGEER(1)VPVVNK
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
H0YAW0	ETDTASESSYQLSR(1)HK
S4R435	MRYLQQKVTR(1)

Table B14. Citrullination sites for *Cancer* sample of CRC031

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTA	LDINTLTGEER(1)VPVVNK
EF1A3	SR(1)WQQGNIFSCSVMHEALHNR
H0YAW0	ETDTASESSYQLSR(1)HK
IMA3	IREEYPDR(1)IMNTFSVVPSPK

Table B15. Citrullination sites for *Normal* sample of CRC031

Protein	Peptide (modification)
A0A0U1RQH7	QVLESR(0.995)LQR(0.005)PLPEDLAR
ACTBL	IQDLNLSR(0.825)QAKADQLR(0.175)
EF1A3	SR(1)WQQGNIFSCSVMHEALHNR
H0YAW0	ETDTASESSYQLSR(1)HK

Table B16. Citrullination sites for *Plasma* sample of CRC031

Protein	Peptide (modification)
ACTA	LDINTLTGEER(1)VPVVNK
H0YAW0	ETDTASESSYQLSR(1)HK

References

Adams, J. M.; Cory, S. The Bcl-2 Apoptotic Switch in Cancer Development and Therapy. *Oncogene* **2007**, *26* (9), 1324–1337.

Adeola, H. A.; Smith, M.; Kaestner, L.; Blackburn, J. M.; Zerbini, L. F. Novel Potential Serological Prostate Cancer Biomarkers Using CT100+ Cancer Antigen Microarray Platform in a Multi-Cultural South African Cohort. *Oncotarget* **2016**, *7* (12), 13945–13964.

Akobeng, A. K. Understanding Diagnostic Tests 3: Receiver Operating Characteristic Curves. *Acta Paediatr.* **2007**, *96* (5), 644–647.

Alhamdani, M. S. S.; Schröder, C.; Werner, J.; Giese, N.; Bauer, A.; Hoheisel, J. D. Single-Step Procedure for the Isolation of Proteins at near-Native Conditions from Mammalian Tissue for Proteomic Analysis on Antibody Microarrays. *J. Proteome Res.* **2010**, *9* (2), 963–971.

Alvarez Arias, D. a; Kim, H.-J.; Zhou, P.; Holderried, T. a W.; Wang, X.; Dranoff, G.; Cantor, H. Disruption of CD8+ Treg Activity Results in Expansion of T Follicular Helper Cells and Enhanced Antitumor Immunity. *Cancer Immunol. Res.* **2014**, *2* (3), 207–216.

Amann, T.; Bataille, F.; Spruss, T.; Dettmer, K.; Wild, P.; Liedtke, C.; Mühlbauer, M.; Kiefer, P.; Oefner, P. J.; Trautwein, C.; et al. Reduced Expression of Fibroblast Growth Factor Receptor 2IIIb in Hepatocellular Carcinoma Induces a More Aggressive Growth. *Am. J. Pathol.* **2010**, *176* (3), 1433–1442.

Anderson, K. C.; Bates, M. P.; Slaughenhoupt, B. L.; Pinkus, G. S.; Schlossman, S. F.; Nadler, L. M. Expression of Human B Cell-Associated Antigens on Leukemias and Lymphomas: A Model of Human B Cell Differentiation. *Blood* **1984**, *63* (6), 1424.

Anderson, K. S.; LaBaer, J. The Sentinel within: Exploiting the Immune System for Cancer Biomarkers. *J. Proteome Res.* **2005**, *4* (4), 1123–1133.

Anderson, T. W.; Darling, D. A. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *Ann. Math. Stat.* **1952**, *23* (2), 193–212.

Andrade, F.; Darrah, E.; Gucek, M.; Cole, R. N.; Rosen, A.; Zhu, X. Autocitrullination of Human Peptidyl Arginine Deiminase Type 4 Regulates Protein Citrullination during Cell Activation. *Arthritis Rheum.* **2010**, *62* (6), 1630–1640.

Assmann, G.; Shihadeh, K.; Poeschel, V.; Murawski, N.; Conigliarou, J.; Ong, M. F.; Pfreundschuh, M. Prevalence of Anti-Citrullinated Protein Antibodies (ACPA) in Patients with Diffuse Large B-Cell Lymphoma (DLBCL): A Case-Control Study. *PLoS One* **2014**, *9* (2), e88177.

Aung, P. P.; Liu, Y.-C.; Ballester, L. Y.; Robbins, P. F.; Rosenberg, S. A.; Lee, C.-C. R. Expression of New York Esophageal Squamous Cell Carcinoma-1 in Primary and Metastatic Melanoma. *Hum. Pathol.* **2014**, *45* (2), 259–267.

Azoury, S. C.; Lange, J. R. Epidemiology, Risk Factors, Prevention, and Early Detection of Melanoma. *Surg. Clin. North Am.* **2014**, *94* (5), 945–962.

De Backer, O.; Arden, K. C.; Boretti, M.; Vantomme, V.; De Smet, C.; Czekay, S.; Viars, C. S.; De Plaen, E.; Brasseur, F.; Chomez, P.; et al. Characterization of the GAGE Genes That Are Expressed in Various Human Cancers and in Normal Testis. *Cancer Res.* **1999**, *59* (13), 3157–3165.

Baeriswyl, V.; Christofori, G. The Angiogenic Switch in Carcinogenesis. *Semin. Cancer Biol.* **2009**, *19* (5), 329–337.

Baka, Z.; Barta, P.; Losonczy, G.; Krenács, T.; Pápay, J.; Szarka, E.; Sármay, G.; Babos, F.; Magyar, A.; Géher, P.; et al. Specific Expression of PAD4 and Citrullinated Proteins in Lung Cancer Is Not Associated with Anti-CCP Antibody Production. *Int. Immunol.* **2011**, *23* (6), 405–414.

Beeton-Kempen, N.; Duarte, J.; Shoko, A.; Serufuri, J. M.; John, T.; Cebon, J.; Blackburn, J. Development of a Novel, Quantitative Protein Microarray Platform for the Multiplexed Serological Analysis of Autoantibodies to Cancer-Testis Antigens. *Int. J. Cancer* **2014**, *135* (8), 1842–1851.

Begent, R. H. J. The Value of Carcinoembryonic Antigen Measurement in Clinical Practice. *Ann. Clin. Biochem. An Int. J. Biochem. Lab. Med.* **1984**, *21* (4), 231–238.

Belousov, P. V; Kuprash, D. V; Sazykin, a Y.; Khlgatian, S. V; Penkov, D. N.; Shebzukhov, Y. V; Nedospasov, S. a. Cancer-Associated Antigens and Antigen Arrays in Serological Diagnostics of Malignant Tumors. *Biochem. Biokhimiia* **2008**, *73* (5), 562–572.

Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **1995**, *57* (1), 289–300.

Bennett, F.; Luxenberg, D.; Ling, V.; Wang, I.-M.; Marquette, K.; Lowe, D.; Khan, N.; Veldman, G.; Jacobs, K. A.; Valge-Archer, V. E.; et al. Program Death-1 Engagement Upon TCR Activation Has Distinct Effects

on Costimulation and Cytokine-Driven Proliferation: Attenuation of ICOS, IL-4, and IL-21, But Not CD28, IL-7, and IL-15 Responses. *J. Immunol.* **2003**, *170* (2), 711–718.

Bergers, G.; Benjamin, L. E. Angiogenesis: Tumorigenesis and the Angiogenic Switch. *Nat. Rev. Cancer* **2003**, *3* (6), 401–410.

Berx, G.; van Roy, F. Involvement of Members of the Cadherin Superfamily in Cancer. *Cold Spring Harb. Perspect. Biol.* **2009**, *1* (6).

Bhatnagar, J.; Tewari, H. B.; Bhatnagar, M.; Austin, G. E. Comparison of Carcinoembryonic Antigen in Tissue and Serum with Grade and Stage of Colon Cancer. *Anticancer Res.* **1999**, *19* (3 B), 2181–2187.

Bhowmick, N. A.; Neilson, E. G.; Moses, H. L. Stromal Fibroblasts in Cancer Initiation and Progression. *Nature* **2004**, *432* (7015), 332–337.

Bindea, G.; Mlecnik, B.; Fridman, W.-H.; Pagès, F.; Galon, J. Natural Immunity to Cancer in Humans. *Curr. Opin. Immunol.* **2010**, *22* (2), 215–222.

Birkenkamp-Demtroder, K.; Olesen, S. H.; Sørensen, F. B.; Laurberg, S.; Laiho, P.; Aaltonen, L. A.; Orntoft, T. F. Differential Gene Expression in Colon Cancer of the Caecum versus the Sigmoid and Rectosigmoid. *Gut* **2005**, *54* (3), 374–384.

Blasco, M. A. Telomeres and Human Disease: Ageing, Cancer and Beyond. *Nat. Rev. Genet.* **2005**, *6* (8), 611–622.

Boël, P.; Wildmann, C.; Sensi, M. L.; Brasseur, R.; Renauld, J. C.; Coulie, P.; Boon, T.; van der Bruggen, P. BAGE: A New Gene Encoding an Antigen Recognized on Human Melanomas by Cytolytic T Lymphocytes. *Immunity* **1995**, *2* (2), 167–175.

Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat. Protoc.* **2009a**, *4* (4), 484–494.

Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Analysis by Mass Spectrometry. *J. Mass Spectrom.* **2009b**, *44* (6), 861–878.

Borghaei, H.; Paz-Ares, L.; Horn, L.; Spigel, D. R.; Steins, M.; Ready, N. E.; Chow, L. Q.; Vokes, E. E.; Felip, E.; Holgado, E.; et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2015**, *373* (17), 1627–1639.

Boyle, P.; Levin, B. World Cancer Report 2014. *World Cancer Rep. 2014* **2014**, *i*, 630.

Brahmer, J.; Reckamp, K. L.; Baas, P.; Crinò, L.; Eberhardt, W. E. E.; Poddubskaya, E.; Antonia, S.; Pluzanski, A.; Vokes, E. E.; Holgado, E.; et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2015**, *373* (2), 123–135.

Broadway, N. Recombinant Protein Expression: Vector-Host Systems. *Mater. Methods* **2012**, *2*.

Brown, P. R.; Miki, K.; Harper, D. B.; Eddy, E. M. A-Kinase Anchoring Protein 4 Binding Proteins in the Fibrous Sheath of the Sperm Flagellum. *Biol. Reprod.* **2003**, *68* (6), 2241–2248.

van der Bruggen, P.; Traversari, C.; Chomez, P.; Lurquin, C.; De Plaen, E.; Van den Eynde, B.; Knuth, A.; Boon, T. A Gene Encoding an Antigen Recognized by Cytolytic T Lymphocytes on a Human Melanoma. *Science (80-.).* **1991**, *254* (5038), 1643–1647.

Buchbinder, E. I.; Desai, A. CTLA-4 and PD-1 Pathways. *Am. J. Clin. Oncol.* **2016**, *39* (1), 98–106.

Bupp, M. R. G.; Jorgensen, T. N. Androgen-Induced Immunosuppression. *Front. Immunol.* **2018**, *9* (APR), 1.

Cabrera, T.; Angustias Fernandez, M.; Sierra, A.; Garrido, A.; Herruzo, A.; Escobedo, A.; Fabra, A.; Garrido, F. High Frequency of Altered HLA Class I Phenotypes in Invasive Breast Carcinomas. *Hum. Immunol.* **1996**, *50* (2), 127–134.

Cabrera, T.; Collado, A.; Fernandez, M. A.; Ferron, A.; Sancho, J.; Ruiz-Cabello, F.; Garrido, F. High Frequency of Altered HLA Class I Phenotypes in Invasive Colorectal Carcinomas. *Tissue Antigens* **1998**, *52* (2), 114–123.

Cabrera, T.; Salinero, J.; Fernandez, M. A.; Garrido, A.; Esquivias, J.; Garrido, F. High Frequency of Altered HLA Class I Phenotypes in Laryngeal Carcinomas. *Hum. Immunol.* **2000**, *61* (5), 499–506.

Callahan, M. K.; Postow, M. A.; Wolchok, J. D. Immunomodulatory Therapy for Melanoma: Ipilimumab and Beyond. *Clin. Dermatol.* **2013**, *31* (2), 191–199.

Carl, P. L.; Temple, B. R. S.; Cohen, P. L. Most Nuclear Systemic Autoantigens Are Extremely Disordered Proteins: Implications for the Etiology of Systemic Autoimmunity. *Arthritis Res. Ther.* **2005**, *7* (6), R1360-74.

Caron, M.; Choquet-Kastylevsky, G.; Joubert-Caron, R. Cancer Immunomics Using Autoantibody

Signatures for Biomarker Discovery. *Mol. Cell. Proteomics* **2007**, *6* (7), 1115–1122.

Carter, P. Improving the Efficacy of Antibody-Based Cancer Therapies. *Nat. Rev. Cancer* **2001**, *1* (2), 118–129.

Cavallaro, U.; Christofori, G. Cell Adhesion and Signalling by Cadherins and Ig-CAMs in Cancer. *Nat. Rev. Cancer* **2004**, *4* (2), 118–132.

Chambers, C. A.; Kuhns, M. S.; Jackson, G.; Allison, J. P. CTLA -4-Mediated Inhibition in Regulation of T Cell Respones: Mechanisms and Manipulation in Tumor Immunotherapy. *Annu Rev Immunol* **2001**, *19*, 565–594.

Chang, X.; Fang, K. PADI4 and Tumourigenesis. *Cancer Cell International*. BioMed Central March 12, 2010, p 7.

Chang, X.; Han, J.; Pang, L.; Zhao, Y.; Yang, Y.; Shen, Z. Increased PADI4 Expression in Blood and Tissues of Patients with Malignant Tumors. *BMC Cancer* **2009**, *9* (1), 40.

Chang, Y. C.; Wu, C. H.; Yen, T. C.; Ouyang, P. Centrosomal Protein 55 (Cep55) Stability Is Negatively Regulated by P53 Protein through Polo-like Kinase 1 (Plk1). *J. Biol. Chem.* **2012**, *287* (6), 4376–4385.

Cheeseman, I. M.; Hori, T.; Fukagawa, T.; Desai, A. KNL1 and the CENP-H/I/K Complex Coordinately Direct Kinetochore Assembly in Vertebrates. *Mol Biol Cell* **2008**, *19* (2), 587–594.

Cheever, M. A.; Greenberg, P. D.; Fefer, A. Adoptive Therapy of Established Syngeneic Leukemia by Cells Primarily Sensitized in Vitro. *Cancer Res.* **1981a**, *41* (7), 2658–2663.

Cheever, M. A.; Greenberg, P. D.; Fefer, A. Specific Adoptive Therapy of Established Leukemia with Syngeneic Lymphocytes Sequentially Immunized in Vivo and in Vitro and Nonspecifically Expanded by Culture with Interleukin 2. *J. Immunol.* **1981b**, *126* (4), 1318–1322.

Chen, M.; Ni, J.; Chang, H.-C.; Lin, C.-Y.; Muyan, M.; Yeh, S. CCDC62/ERAP75 Functions as a Coactivator to Enhance Estrogen Receptor Beta-Mediated Transactivation and Target Gene Expression in Prostate Cancer Cells. *Carcinogenesis* **2009**, *30* (5), 841–850.

Chen, W.; Peace, D.; Rovira, D.; You, S.; Cheever, M. T-Cell Immunity to the Joining Region of P210BCR-ABL Protein. *Proc Natl Acad Sci U S A* **1992**, *89* (4), 1468–72.

Chen, Y.-T.; Scanlan, M. J.; Sahin, U.; Tureci, O.; Gure, A. O.; Tsang, S.; Williamson, B.; Stockert, E.;

Pfreundschuh, M.; Old, L. J. A Testicular Antigen Aberrantly Expressed in Human Cancers Detected by Autologous Antibody Screening. *Proc. Natl. Acad. Sci.* **1997**, *94* (5), 1914–1918.

Chen, Y.-T.; Scanlan, M. J.; Venditti, C. A.; Chua, R.; Theiler, G.; Stevenson, B. J.; Iseli, C.; Gure, A. O.; Vasicek, T.; Strausberg, R. L.; et al. Identification of Cancer/Testis-Antigen Genes by Massively Parallel Signature Sequencing. *Proc. Natl. Acad. Sci.* **2005**, *102* (22), 7940–7945.

Chen, Y.; Lin, P.; Qiu, S.; Peng, X.-X.; Looi, K.; Farquhar, M. G.; Zhang, J.-Y. Autoantibodies to Ca2+ Binding Protein Calnuc Is a Potential Marker in Colon Cancer Detection. *Int. J. Oncol.* **2007**, *30* (5), 1137–1144.

Cheng, N.; Chytil, A.; Shyr, Y.; Joly, A.; Moses, H. L. Transforming Growth Factor-Beta Signaling-Deficient Fibroblasts Enhance Hepatocyte Growth Factor Signaling in Mammary Carcinoma Cells to Promote Scattering and Invasion. *Mol. Cancer Res.* **2008**, *6* (10), 1521–1533.

Cheson, B. D.; Leonard, J. P. Monoclonal Antibody Therapy for B-Cell Non-Hodgkin's Lymphoma. *N. Engl. J. Med.* **2008**, *359* (6), 613–626.

Chiriva-Internati, M.; Cobos, E.; Da Silva, D. M.; Kast, W. M. Sperm Fibrous Sheath Proteins: A Potential New Class of Target Antigens for Use in Human Therapeutic Cancer Vaccines. *Cancer Immun.* **2008**, *8*.

Chiu, K.-H.; Chang, Y.-H.; Wu, Y.-S.; Lee, S.-H.; Liao, P.-C. Quantitative Secretome Analysis Reveals That COL6A1 Is a Metastasis-Associated Protein Using Stacking Gel-Aided Purification Combined with ITRAQ Labeling. *J. Proteome Res.* **2011**, *10* (3), 1110–1125.

Choi, J.; Chang, H. The Expression of MAGE and SSX, and Correlation of COX2, VEGF, and Survivin in Colorectal Cancer. *Anticancer Res.* **2012**, *32* (2), 559–564.

Christophorou, M. A.; Castelo-Branco, G.; Halley-Stott, R. P.; Oliveira, C. S.; Loos, R.; Radzisheuskaya, A.; Mowen, K. A.; Bertone, P.; Silva, J. C. R.; Zernicka-Goetz, M.; et al. Citrullination Regulates Pluripotency and Histone H1 Binding to Chromatin. *Nature* **2014**, *507* (7490), 104–108.

Chung, F.-Y.; Cheng, T.-L.; Chang, H.-J.; Chiu, H.-H.; Huang, M.-Y.; Chang, M.-S.; Chen, C.-C.; Yang, M.-J.; Wang, J.-Y.; Lin, S.-R. Differential Gene Expression Profile of MAGE Family in Taiwanese Patients with Colorectal Cancer. *J. Surg. Oncol.* **2010**, *102* (2), 148–153.

Cioffi, M.; Riegler, G.; Vietri, M. T.; Pilla, P.; Caserta, L.; Carratu, R.; Sica, V.; Molinari, A. M. Serum P53 Antibodies in Patients Affected with Ulcerative Colitis. *Inflamm.Bowel.Dis.* **2004**, *10* (5), 606–611.

Ciró, M.; Prosperini, E.; Quarto, M.; Grazini, U.; Walfridsson, J.; McBlane, F.; Nucifero, P.; Pacchiana, G.; Capra, M.; Christensen, J.; et al. ATAD2 Is a Novel Cofactor for MYC, Overexpressed and Amplified in Aggressive Tumors. *Cancer Res.* **2009**, *69* (21), 8491–8498.

Clegg, L. X. Impact of Reporting Delay and Reporting Error on Cancer Incidence Rates and Trends. *J. Natl. Cancer Inst.* **2002**, *94* (20), 1537–1545.

Cole, L. A. HCG Variants, the Growth Factors Which Drive Human Malignancies. *Am. J. Cancer Res.* **2012**, *2* (1), 22–35.

Colella, T. A.; Bullock, T. N. J.; Russell, L. B.; Mullins, D. W.; Overwijk, W. W.; Luckey, C. J.; Pierce, R. A.; Restifo, N. P.; Engelhard, V. H. Self-Tolerance to the Murine Homologue of a Tyrosinase-Derived Melanoma Antigen Implications for Tumor Immunotherapy. *J. Exp. Med.* **2000**.

Collins, A. V.; Brodie, D. W.; Gilbert, R. J. C.; Iaboni, A.; Manso-Sancho, R.; Walse, B.; Stuart, D. I.; Van Der Merwe, P. A.; Davis, S. J. The Interaction Properties of Costimulatory Molecules Revisited. *Immunity* **2002**, *17* (2), 201–210.

Cooks, R. G.; Mueller, T. Through a Glass Darkly: Glimpses into the Future of Mass Spectrometry. *Mass Spectrom. (Tokyo, Japan)* **2013**, *2* (Spec Iss), S0001.

Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.

Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.

Crawford, L. V.; Pim, D. C.; Bulbrook, R. D. Detection of Antibodies against the Cellular Protein P53 in Sera from Patients with Breast Cancer. *Int. J. Cancer* **1982**, *30* (4), 403–408.

Curran, M. A.; Montalvo, W.; Yagita, H.; Allison, J. P. PD-1 and CTLA-4 Combination Blockade Expands Infiltrating T Cells and Reduces Regulatory T and Myeloid Cells within B16 Melanoma Tumors. *Proc. Natl. Acad. Sci.* **2010**, *107* (9), 4275–4280.

Darrah, E.; Rosen, a.; Giles, J. T.; Andrade, F. Peptidylarginine Deiminase 2, 3 and 4 Have Distinct Specificities against Cellular Substrates: Novel Insights into Autoantigen Selection in Rheumatoid Arthritis. *Ann. Rheum. Dis.* **2012**, *71* (1), 92–98.

Daudi, S.; Eng, K. H.; Mhawech-Fauceglia, P.; Morrison, C.; Miliotto, A.; Beck, A.; Matsuzaki, J.; Tsuji, T.;

Groman, A.; Gnjatic, S.; et al. Expression and Immune Responses to MAGE Antigens Predict Survival in Epithelial Ovarian Cancer. *PLoS One* **2014**, *9* (8), e104099.

Davies, M. A.; Samuels, Y. Analysis of the Genome to Personalize Therapy for Melanoma. *Oncogene* **2010**, *29* (41), 5545–5555.

Deeks, E. D. Nivolumab: A Review of Its Use in Patients with Malignant Melanoma. *Drugs* **2014**, *74* (11), 1233–1239.

DeLuca, J. G.; Moree, B.; Hickey, J. M.; Kilmartin, J. V.; Salmon, E. D. HNuf2 Inhibition Blocks Stable Kinetochore-Microtubule Attachment and Induces Mitotic Cell Death in HeLa Cells. *J. Cell Biol.* **2002**, *159* (4), 549–555.

Deniger, D. C.; Pasetto, A.; Robbins, P. F.; Gartner, J. J.; Prickett, T. D.; Paria, B. C.; Malekzadeh, P.; Jia, L.; Yossef, R.; Langhan, M. M.; et al. T-Cell Responses to *TP53* "Hotspot" Mutations and Unique Neoantigens Expressed by Human Ovarian Cancers. *Clin. Cancer Res.* **2018**.

Denman, R. B. PAD: The Smoking Gun behind Arginine Methylation Signaling? *BioEssays* **2005**, *27* (3), 242–246.

Diez de Medina, S. G.; Chopin, D.; Marjou, A. El; Delouvée, A.; LaRochelle, W. J.; Hoznek, A.; Abbou, C.; Aaronson, S. A.; Thiery, J. P.; Radvanyi, F. Decreased Expression of Keratinocyte Growth Factor Receptor in a Subset of Human Transitional Cell Bladder Carcinomas. *Oncogene* **1997**, *14* (3), 323–330.

Disis, M. L.; Pupa, S. M.; Gralow, J. R.; Dittadi, R.; Menard, S.; Cheever, M. A. High-Titer HER-2/Neu Protein-Specific Antibody Can Be Detected in Patients with Early-Stage Breast Cancer. *J. Clin. Oncol.* **1997**, *15* (11), 3363–3367.

Doyle, J. M.; Gao, J.; Wang, J.; Yang, M.; Potts, P. R. MAGE-RING Protein Complexes Comprise a Family of E3 Ubiquitin Ligases. *Mol. Cell* **2010**, *39* (6), 963–974.

Dranoff, G. Cytokines in Cancer Pathogenesis and Cancer Therapy. *Nat. Rev. Cancer* **2004**, *4* (1), 11–22.

Duarte, J. G.; Blackburn, J. M. Advances in the Development of Human Protein Microarrays. *Expert Rev. Proteomics* **2017**, *14* (7), 627–641.

Dwivedi, N.; Upadhyay, J.; Neeli, I.; Khan, S.; Pattanaik, D.; Myers, L.; Kirou, K. A.; Hellmich, B.; Knuckley, B.; Thompson, P. R.; et al. Felty's Syndrome Autoantibodies Bind to Deiminated Histones and Neutrophil Extracellular Chromatin Traps. *Arthritis Rheum.* **2012**, *64* (4), 982–992.

Eddy, E. M.; Toshimori, K.; O'Brien, D. A. Fibrous Sheath of Mammalian Spermatozoa. *Microsc. Res. Tech.* **2003**, *61* (1), 103–115.

Epping, M. T.; Wang, L.; Edel, M. J.; Carlée, L.; Hernandez, M.; Bernards, R. The Human Tumor Antigen PRAME Is a Dominant Repressor of Retinoic Acid Receptor Signaling. *Cell* **2005**, *122* (6), 835–847.

Evan, G. A Matter of Life and Cell Death. *Science (80-.).* **1998**, *281* (5381), 1317–1322.

Farolfi, A.; Ridolfi, L.; Guidoboni, M.; Nicoletti, S. V. L.; Piciucchi, S.; Valmorri, L.; Costantini, M.; Scarpi, E.; Amadori, D.; Ridolfi, R. Ipilimumab in Advanced Melanoma: Reports of Long-Lasting Responses. *Melanoma Res.* **2012**, *22* (3), 263–270.

Feller, W. On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *Ann. Math. Stat.* **1948**, *19* (2), 177–189.

Ferrone, C.; Dranoff, G. Dual Roles for Immunity in Gastrointestinal Cancers. *J. Clin. Oncol.* **2010**, *28* (26), 4045–4051.

Fidler, I. J. Timeline: The Pathogenesis of Cancer Metastasis: The "seed and Soil" Hypothesis Revisited. *Nat. Rev. Cancer* **2003**, *3* (6), 453–458.

Fiedler, S. E.; Bajpai, M.; Carr, D. W. Identification and Characterization of RHOA-Interacting Proteins in Bovine Spermatozoa. *Biol. Reprod.* **2008**, *78* (1), 184–192.

Frank, A. M. A Ranking-Based Scoring Function for Peptide-Spectrum Matches. *J. Proteome Res.* **2009**, *8* (5), 2241–2252.

Funayama, R.; Taniguchi, H.; Mizuma, M.; Fujishima, F.; Kobayashi, M.; Ohnuma, S.; Unno, M.; Nakayama, K.; Nakayama, C. K. Protein-Arginine Deiminase 2 Suppresses Proliferation of Colon Cancer Cells through Protein Citrullination. *Cancer Sci* **2017**, *108*, 713–718.

Ganesan, V.; Ascherman, D. P.; Minden, J. S. Immunoproteomics Technologies in the Discovery of Autoantigens in Autoimmune Diseases. *Biomol. Concepts* **2016**, *7* (2), 133–143.

Garcia-Lora, A.; Algarra, I.; Garrido, F. MHC Class I Antigens, Immune Surveillance, and Tumor Immune Escape. *J. Cell. Physiol.* **2003**, *195* (3), 346–355.

Garon, E. B.; Rizvi, N. A.; Hui, R.; Leighl, N.; Balmanoukian, A. S.; Eder, J. P.; Patnaik, A.; Aggarwal, C.; Gubens, M.; Horn, L.; et al. Pembrolizumab for the Treatment of Non-Small-Cell Lung Cancer. *N. Engl. J.*

Med. **2015**, *372* (21), 2018–2028.

Garrido, F.; Ruiz-Cabello, F.; Cabrera, T.; Pérez-Villar, J. J.; López-Botet, M.; Duggan-Keen, M.; Stern, P. L. Implications for Immunosurveillance of Altered HLA Class I Phenotypes in Human Tumours. *Immunol. Today* **1997**, *18* (2), 89–95.

Gartside, M. G.; Chen, H.; Ibrahimi, O. A.; Byron, S. A.; Curtis, A. V; Wellens, C. L.; Bengston, A.; Yudt, L. M.; Eliseenkova, A. V; Ma, J.; et al. Loss-of-Function Fibroblast Growth Factor Receptor-2 Mutations in Melanoma. *Mol. Cancer Res.* **2009**, *7* (1), 41–54.

GBD 2013 Mortality and Causes of Death Collaborators. Global, Regional, and National Age-Sex Specific All-Cause and Cause-Specific Mortality for 240 Causes of Death, 1990-2013: A Systematic Analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)* **2015**, *385* (9963), 117–171.

GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 310 Diseases and Injuries, 1990-2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet* **2016**, *388* (10053), 1545–1602.

Gettinger, S. N.; Horn, L.; Gandhi, L.; Spigel, D. R.; Antonia, S. J.; Rizvi, N. A.; Powderly, J. D.; Heist, R. S.; Carvajal, R. D.; Jackman, D. M.; et al. Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer. TL - 33. *J. Clin. Oncol.* **2015**, *33 VN-r* (18), 2004–2012.

Giaccone, G. Epidermal Growth Factor Receptor Inhibitors in the Treatment of Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2005**, *23* (14), 3235–3242.

Giavina-Bianchi, M.; Giavina-Bianchi, P.; Sotto, M. N.; Muzikansky, A.; Kalil, J.; Festa-Neto, C.; Duncan, L. M. Increased NY-ESO-1 Expression and Reduced Infiltrating CD3+ T Cells in Cutaneous Melanoma. *J. Immunol. Res.* **2015**, *2015*, 1–8.

Gjerstorff, M. F.; Kock, K.; Nielsen, O.; Ditzel, H. J. MAGE-A1, GAGE and NY-ESO-1 Cancer/Testis Antigen Expression during Human Gonadal Development. *Hum. Reprod.* **2007**, *22* (4), 953–960.

Gold, P.; Freedman, S. O. Demonstration of Tumor-Specific Antigens in Human Colonic Carcinomata By Immunological Tolerance and Absorption Techniques. *J. Exp. Med.* **1965a**, *121*, 439–462.

Gold, P.; Freedman, S. O. Specific Carcinoembryonic Antigens of the Human Digestive System. *J. Exp. Med.* **1965b**, *122* (3), 467–481.

Goldberg, E.; Eddy, E. M.; Duan, C.; Odet, F. LDHC: The Ultimate Testis-Specific Gene. *J. Androl.* **2010**, *31* (1), 86–94.

Good, N.; Winget, G.; Winter, W. Hydrogen Ion Buffers for Biological Research. *Biochemistry* **1966**, *310* (2), 300–310.

Goodnow, C. C.; Sprent, J.; de St Groth, B. F.; Vinuesa, C. G. Cellular and Genetic Mechanisms of Self Tolerance and Autoimmunity. *Nature* **2005**, *435* (7042), 590–597.

Goslin, R.; O'Brien, M. J.; Steele, G.; Mayer, R.; Wilson, R.; Corson, J. M.; Zamcheck, N. Correlation of Plasma CEA and CEA Tissue Staining in Poorly Differentiated Colorectal Cancer. *Am. J. Med.* **1981**, *71* (2), 246–253.

Gross, L. Intradermal Immunization of C3H Mice against a Sarcoma That Originated in an Animal of the Same Line. *Cancer Res.* **1943**, *3* (5), 326–333.

Gunawardana, C. G.; Diamandis, E. P. High Throughput Proteomic Strategies for Identifying Tumour-Associated Antigens. *Cancer Lett.* **2007**, *249* (1), 110–119.

Hamrita, B.; Chahed, K.; Kabbage, M.; Guillier, C. L.; Trimeche, M.; Chaïeb, A.; Chouchane, L. Identification of Tumor Antigens That Elicit a Humoral Immune Response in Breast Cancer Patients' Sera by Serological Proteome Analysis (SERPA). *Clin. Chim. Acta* **2008**, *393* (2), 95–102.

Hanahan, D.; Folkman, J. Patterns and Emerging Mechanisms of the Angiogenic Switch during Tumorigenesis. *Cell* **1996**, *86* (3), 353–364.

Hanahan, D.; Weinberg, R. A. The Hallmarks of Cancer. *Cell* **2000**, *100* (1), 57–70.

Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The next Generation. *Cell* **2011**, *144* (5), 646–674.

Hashimoto, R.; Kanda, M.; Takami, H.; Shimizu, D.; Oya, H.; Hibino, S.; Okamura, Y.; Yamada, S.; Fujii, T.; Nakayama, G.; et al. Aberrant Expression of Melanoma-Associated Antigen-D2 Serves as a Prognostic Indicator of Hepatocellular Carcinoma Outcome Following Curative Hepatectomy. *Oncol. Lett.* **2015**, *9* (3), 1201–1206.

He, Y.; Wu, Y.; Mou, Z.; Li, W.; Zou, L.; Fu, T.; Zhang, A.; Xiang, D.; Xiao, H.; Wang, X. Proteomics-Based Identification of HSP60 as a Tumor-Associated Antigen in Colorectal Cancer. *Proteomics - Clin. Appl.* **2007**, *1* (3), 336–342.

Herbst, R. S.; Soria, J.-C.; Kowanetz, M.; Fine, G. D.; Hamid, O.; Gordon, M. S.; Sosman, J. A.; McDermott, D. F.; Powderly, J. D.; Gettinger, S. N.; et al. Predictive Correlates of Response to the Anti-PD-L1 Antibody MPDL3280A in Cancer Patients. *Nature* **2014**, *515* (7528), 563–567.

Herbst, R. S.; Baas, P.; Kim, D. W.; Felip, E.; Pérez-Gracia, J. L.; Han, J. Y.; Molina, J.; Kim, J. H.; Arvis, C. D.; Ahn, M. J.; et al. Pembrolizumab versus Docetaxel for Previously Treated, PD-L1-Positive, Advanced Non-Small-Cell Lung Cancer (KEYNOTE-010): A Randomised Controlled Trial. *Lancet* **2016**, *387* (10027), 1540–1550.

Hinuma, Y.; Komoda, H.; Chosa, T.; Kondo, T.; Kohakura, M.; Takenaka, T.; Kikuchi, M.; Ichimaru, M.; Yunoki, K.; Sato, I.; et al. Antibodies to Adult T-Cell Leukemia-Virus-Associated Antigen (Atla) in Sera from Patients with Atl and Controls in Japan: A Nation-Wide Sero-Epidemiologic Study. *Int. J. Cancer* **1982**, *29* (6), 631–635.

Hodi, F. S.; O'Day, S. J.; McDermott, D. F.; Weber, R. W.; Sosman, J. A.; Haanen, J. B.; Gonzalez, R.; Robert, C.; Schadendorf, D.; Hassel, J. C.; et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* **2010**, *363* (8), 711–723.

Holzmann, J.; Fuchs, J.; Pichler, P.; Peters, J. M.; Mechtler, K. Lesson from the Stoichiometry Determination of the Cohesin Complex: A Short Protease Mediated Elution Increases the Recovery from Cross-Linked Antibody-Conjugated Beads. *J. Proteome Res.* **2011**, *10* (2), 780–789.

Horibata, S.; Rogers, K. E.; Sadegh, D.; Anguish, L. J.; McElwee, J. L.; Shah, P.; Thompson, P. R.; Coonrod, S. A. Role of Peptidylarginine Deiminase 2 (PAD2) in Mammary Carcinoma Cell Migration. *BMC Cancer* **2017**, *17* (1), 378.

Hsu, P.-C.; Liao, Y.-F.; Lin, C.-L.; Lin, W.-H.; Liu, G.-Y.; Hung, H.-C. Vimentin Is Involved in Peptidylarginine Deiminase 2-Induced Apoptosis of Activated Jurkat Cells. *Mol. Cells* **2014**, *37* (5), 426–434.

Inagaki, M.; Takahara, H.; Nishi, Y.; Sugawara, K.; Sato, C. Ca2+-Dependent Deimination-Induced Disassembly of Intermediate Filaments Involves Specific Modification of the Amino-Terminal Head Domain. *J. Biol. Chem.* **1989**, *264* (30), 18119–18127.

Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; et al. Abnormal Accumulation of Citrullinated Proteins Catalyzed by Peptidylarginine Deiminase in Hippocampal Extracts from Patients with Alzheimer's Disease. *J. Neurosci. Res.* **2005**, *80* (1), 120–128.

Jackson-Thompson, J.; Ahmed, F.; German, R. R.; Lai, S.-M.; Friedman, C. Descriptive Epidemiology of Colorectal Cancer in the United States, 1998–2001. *Cancer* **2006**, *107* (S5), 1103–1111.

Jäger, E.; Ringhoffer, M.; Altmannsberger, M.; Arand, M.; Karbach, J.; Jäger, D.; Oesch, F.; Knuth, A. Immunoselection in Vivo: Independent Loss of MHC Class I and Melanocyte Differentiation Antigen Expression in Metastatic Melanoma. *Int. J. Cancer* **1997**, *71* (2), 142–147.

Jäger, E.; Srockert, E.; Zidianakis, Z.; Cnen, Y. T.; Karbach, J.; Jäger, D.; Arand, M.; Ritter, G.; Old, L. J.; Knuth, A. Humoral Immune Responses of Cancer Patients against "Cancer-Testis" Antigen NY-ESO-1: Correlation with Clinical Events. *Int. J. Cancer* **1999**, *84* (5), 506–510.

Jemal, A.; Bray, F.; Center, M. M.; Ferlay, J.; Ward, E.; Forman, D. Global Cancer Statistics. *CA Cancer J Clin* **2011**, *61*, 69–90.

Jeon, C. H.; Kim, D. D.; Lee, H. Il; Oh, H. K.; Chae, H. D. Melanoma-Associated Antigen (MAGE) Expression in the Normal Mucosa around Colorectal Cancer after Curative Resection: Presence of Undetectable Free Cancer Cells? *Int. J. Biol. Markers* **2011**, *26* (2), 88–93.

Jiang, B. H.; Liu, L. Z. Chapter 2 PI3K/PTEN Signaling in Angiogenesis and Tumorigenesis. *Adv. Cancer Res.* **2009**, *102*, 19–65.

Jiang, Z.; Cui, Y.; Wang, L.; Zhao, Y.; Yan, S.; Chang, X. Investigating Citrullinated Proteins in Tumour Cell Lines. *World J. Surg. Oncol.* **2013**, *11*.

Jin, W.-J.; Xu, J.-M.; Xu, W.-L.; Gu, D.-H.; Li, P.-W. Diagnostic Value of Interleukin-8 in Colorectal Cancer: A Case-Control Study and Meta-Analysis. *World J. Gastroenterol.* **2014**, *20* (43), 16334.

Johnsen, A. K.; Templeton, D. J.; Sy, M.; Harding, C. V. Deficiency of Transporter for Antigen Presentation (TAP) in Tumor Cells Allows Evasion of Immune Surveillance and Increases Tumorigenesis. *J. Immunol.* **1999**, *163* (8), 4224–4231.

Jorabchi, K.; Smith, L. M. Single Droplet Separations and Surface Partition Coefficient Measurements Using Laser Ablation Mass Spectrometry. *Anal. Chem.* **2009**, *81* (23), 9682–9688.

Joseph, C. G.; Darrah, E.; Shah, A. A.; Skora, A. D.; Casciola-Rosen, L. A.; Wigley, F. M.; Boin, F.; Fava, A.; Thoburn, C.; Kinde, I.; et al. Association of the Autoimmune Disease Scleroderma with an Immunologic Response to Cancer. *Science (80-.).* **2014**, *343* (6167), 152–157.

Jungbluth, A. A.; Busam, K. J.; Kolb, D.; Iversen, K.; Coplan, K.; Chen, Y. T.; Spagnoli, G. C.; Old, L. J.

Expression of MAGE-Antigens in Normal Tissues and Cancer. *Int. J. Cancer* **2000a**, *85* (4), 460–465.

Jungbluth, A. A.; Stockert, E.; Chen, Y. T.; Kolb, D.; Iversen, K.; Coplan, K.; Williamson, B.; Altorki, N.; Busam, K. J.; Old, L. J. Monoclonal Antibody MA454 Reveals a Heterogeneous Expression Pattern of MAGE-1 Antigen in Formalin-Fixed Paraffin Embedded Lung Tumours. *Br. J. Cancer* **2000b**, *83* (4), 493–497.

Kaklamanis, L.; Hill, A. MHC Loss in Colorectal Tumours: Evidence for Immunoselection? *Cancer Surv.* **1992**, *13*, 155–171.

Keating, G. M. Nivolumab: A Review in Advanced Squamous Non-Small Cell Lung Cancer. *Drugs* **2015**, *75* (16), 1925–1934.

Keir, M. E.; Butte, M. J.; Freeman, G. J.; Sharpe, A. H. PD-1 and Its Ligands in Tolerance and Immunity. *Annu. Rev. Immunol.* **2008**, *26* (1), 677–704.

Kelleher, N. L. Peer Reviewed: Top-Down Proteomics. *Anal. Chem.* **2004**, *76* (11), 196 A-203 A.

Kelso, T. W. R.; Baumgart, K.; Eickhoff, J.; Albert, T.; Antrecht, C.; Lemcke, S.; Klebl, B.; Meisterernst, M. Cyclin-Dependent Kinase 7 Controls MRNA Synthesis by Affecting Stability of Preinitiation Complexes, Leading to Altered Gene Expression, Cell Cycle Progression, and Survival of Tumor Cells. *Mol. Cell. Biol.* **2014**, *34* (19), 3675–3688.

Khamas, A.; Ishikawa, T.; Mogushi, K.; Iida, S.; Ishiguro, M.; Tanaka, H.; Uetake, H.; Sugihara, K. Genome-Wide Screening for Methylation-Silenced Genes in Colorectal Cancer. *Int. J. Oncol.* **2012**, *41* (2), 490–496.

Kim, E. S.; Vokes, E. E.; Kies, M. S. Cetuximab in Cancers of the Lung and Head & Neck. *Semin. Oncol.* **2004**, *31*, 61–67.

Kim, H.-J.; Verbinnen, B.; Tang, X.; Lu, L.; Cantor, H. Inhibition of Follicular T-Helper Cells by CD8+ Regulatory T Cells Is Essential for Self Tolerance. *Nature* **2010**, *467* (7313), 328–332.

Kim, R.; Emi, M.; Tanabe, K. Cancer Immunoediting from Immune Surveillance to Immune Escape. *Immunology* **2007**, *121* (1), 1–14.

Klade, C. S.; Voss, T.; Krystek, E.; Ahorn, H.; Zatloukal, K.; Pummer, K.; Adolf, G. R. Identification of Tumor Antigens in Renal Cell Carcinoma by Serological Proteome Analysis. *Proteomics* **2001**, *1*, 890–898.

Kleivi, K.; Lind, G. E.; Diep, C. B.; Meling, G. I.; Brandal, L. T.; Nesland, J. M.; Myklebost, O.; Rognum, T. O.; Giercksky, K. E.; Skotheim, R. I.; et al. Gene Expression Profiles of Primary Colorectal Carcinomas, Liver Metastases, and Carcinomatoses. *Mol. Cancer* **2007**, *6* (2).

Klingenstein, A.; Haritoglou, I.; Schaumberger, M. M.; Nentwich, M. M.; Hein, R.; Schaller, U. C. Receiver Operating Characteristic Analysis. *Melanoma Res.* **2011**, *21* (4), 352–356.

Knieke, K.; Lingel, H.; Chamaon, K.; Brunner-Weinzierl, M. C. Migration of Th1 Lymphocytes Is Regulated by CD152 (CTLA-4)-Mediated Signaling via PI3 Kinase-Dependent Akt Activation. *PLoS One* **2012**, *7* (3).

Knuckley, B.; Causey, C. P.; Jones, J. E.; Bhatia, M.; Dreyton, C. J.; Osborne, T. C.; Takahara, H.; Thompson, P. R. Substrate Specificity and Kinetic Studies of PADs 1, 3, and 4 Identify Potent and Selective Inhibitors of Protein Arginine Deiminase 3. *Biochemistry* **2010**, *49* (23), 4852–4863.

Kocer, B.; McKolanis, J.; Soran, A. Humoral Immune Response to MUC5AC in Patients with Colorectal Polyps and Colorectal Carcinoma. *BMC Gastroenterol.* **2006**, *6*, 4.

Koopman, L. A.; Corver, W. E.; van der Slik, A. R.; Giphart, M. J.; Fleuren, G. J. Multiple Genetic Alterations Cause Frequent and Heterogeneous Human Histocompatibility Leukocyte Antigen Class I Loss in Cervical Cancer. *J Exp Med* **2000**, *191* (6), 961–976.

Krummel, M. F.; Allison, J. P. CD28 and CTLA-4 Have Opposing Effects on the Response of T Ceils to Stimulation. *J. Exp. Med.* **1995**, *182* (August), 459–465.

Krummel, M. F.; Allison, J. P. CTLA-4 Engagement Inhibits IL-2 Accumulation and Cell Cycle Progression upon Activation of Resting T Cells. *J. Exp. Med.* **1996**, *183* (6), 2533–2540.

Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47* (260), 583–621.

Lamensa, J. W. E.; Moscarello, M. A. Deimination of Human Myelin Basic Protein by a Peptidylarginine Deiminase from Bovine Brain. *J. Neurochem.* **1993**, *61* (3), 987–996.

Larkin, J.; Chiarion-Sileni, V.; Gonzalez, R.; Grob, J. J.; Cowey, C. L.; Lao, C. D.; Schadendorf, D.; Dummer, R.; Smylie, M.; Rutkowski, P.; et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N. Engl. J. Med.* **2015**, *373* (1), 23–34.

Lee, J. H. Clinical Usefulness of Serum CYFRA 21-1 in Patients with Colorectal Cancer. *Nucl. Med. Mol. Imaging (2010).* **2013**, *47* (3), 181–187.

Lee, K.-M.; Chuang, E.; Griffin, M.; Khattri, R.; Hong, D. K.; Zhang, W.; Straus, D.; Samelson, L. E.; Thompson, C. B.; Bluestone, J. A. Molecular Basis of T Cell Inactivation by CTLA-4. *Science (80-.).* **1998**, *282* (5397), 2263–2266.

Lehmann, F.; Marchand, M.; Hainaut, P.; Pouillart, P.; Sastre, X.; Ikeda, H.; Boon, T.; Coulie, P. G. Differences in the Antigens Recognized by Cytolytic T Cells on Two Successive Metastases of a Melanoma Patient Are Consistent with Immune Selection. *Eur. J. Immunol.* **1995**, *25* (2), 340–347.

Levine, J. S.; Ahnen, D. J. Clinical Practice. Adenomatous Polyps of the Colon. *N. Engl. J. Med.* **2006**, *355* (1), 2551–2557.

Li, M.; Yuan, Y.-H.; Han, Y.; Liu, Y.-X.; Yan, L.; Wang, Y.; Gu, J. Colorectal Cancer Tissue and Adjacent Normal Tissue Expression Profile of Cancer-Testis Genes in 121 Human Expression Profile of Cancer-Testis Genes in 121 Human Colorectal Cancer Tissue and Adjacent Normal Tissue. *Clin Cancer Res* **2005**, *11* (September 2003), 1809–1814.

Li, Y.-L.; Zhao, H.; Ren, X.-B.; Li, Y.-L.; Zhao, H.; Ren, X.-B. Relationship of VEGF/VEGFR with Immune and Cancer Cells: Staggering or Forward? *Cancer Biol. Med.* **2016**, *13* (2), 206–214.

Linsley, P. S.; Bradshaw, J.; Greene, J.; Peach, R.; Bennett, K. L.; Mittler, R. S. Intracellular Trafficking of CTLA-4 and Focal Localization towards Sites of TCR Engagement. *Immunity* **1996**, *4* (6), 535–543.

Liu, T.; Wang, X.; Karsdal, M. A.; Leeming, D. J.; Genovese, F. Molecular Serum Markers of Liver Fibrosis. *Biomark. Insights* **2012**, *7*, 105–117.

Liu, W.; Wang, P.; Li, Z.; Xu, W.; Dai, L.; Wang, K.; Zhang, J. Evaluation of Tumour-Associated Antigen (TAA) Miniarray in Immunodiagnosis of Colon Cancer. *Scand. J. Immunol.* **2009**, *69* (1), 57–63.

Liu, Z.; Poiret, T.; Meng, Q.; Rao, M.; von Landenberg, A.; Schoutrop, E.; Valentini, D.; Dodoo, E.; Peredo-Harvey, I.; Maeurer, M. Epstein-Barr Virus- and Cytomegalovirus-Specific Immune Response in Patients with Brain Cancer. *J. Transl. Med.* **2018**, *16* (1), 182.

Lowe, S. W.; Cepero, E.; Evan, G. Intrinsic Tumour Suppression. *Nature* **2004**, *432* (7015), 307–315.

Lynch, T. J.; Bell, D. W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R. A.; Brannigan, B. W.; Harris, P. L.; Haserlat, S. M.; Supko, J. G.; Haluska, F. G.; et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2004**, *350* (21), 2129–2139.

Ma, W. T.; Chang, C.; Gershwin, M. E.; Lian, Z. X. Development of Autoantibodies Precedes Clinical Manifestations of Autoimmune Diseases: A Comprehensive Review. *J. Autoimmun.* **2017**, *83*, 95–112.

Maintainer, T. P.; Pohlert, T. Package "PMCMR" Title Calculate Pairwise Multiple Comparisons of Mean Rank Sums. **2016**.

Mann, H. B.; Whitney, D. R. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Ann. Math. Stat.* **1947**, *18* (1), 50–60.

Markel, G.; Ortenberg, R.; Seidman, R.; Sapoznik, S.; Koren-Morag, N.; Besser, M. J.; Bar, J.; Shapira, R.; Kubi, A.; Nardini, G.; et al. Systemic Dysregulation of CEACAM1 in Melanoma Patients. *Cancer Immunol. Immunother.* **2010**, *59* (2), 215–230.

Martin, K.; Ricciardelli, C.; Hoffmann, P.; Oehler, M. K. Exploring the Immunoproteome for Ovarian Cancer Biomarker Discovery. *Int. J. Mol. Sci.* **2011**, *12* (1), 410–428.

Mazzara, S.; Rossi, R. L.; Grifantini, R.; Donizetti, S.; Abrignani, S.; Bombaci, M. CombiROC: An Interactive Web Tool for Selecting Accurate Marker Combinations of Omics Data. *Sci. Rep.* **2017**, *7*, 45477.

McDermott, D. F.; Drake, C. G.; Sznol, M.; Choueiri, T. K.; Powderly, J. D.; Smith, D. C.; Brahmer, J. R.; Carvajal, R. D.; Hammers, H. J.; Puzanov, I.; et al. Survival, Durable Response, and Long-Term Safety in Patients With Previously Treated Advanced Renal Cell Carcinoma Receiving Nivolumab. *J. Clin. Oncol.* **2015**, *33* (18), 2013–2020.

McElwee, J. L.; Mohanan, S.; Horibata, S.; Sams, K. L.; Anguish, L. J.; McLean, D.; Cvitaš, I.; Wakshlag, J. J.; Coonrod, S. A. PAD2 Overexpression in Transgenic Mice Promotes Spontaneous Skin Neoplasia. *Cancer Res.* **2014**, *74* (21).

McHugh, L.; Arthur, J. W. Computational Methods for Protein Identification from Mass Spectrometry Data. *PLoS Comput. Biol.* **2008**, *4* (2).

Ménard, S.; Casalini, P.; Campiglio, M.; Pupa, S. M.; Tagliabue, E. Role of HER2/Neu in Tumor Progression and Therapy. *Cell. Mol. Life Sci.* **2004**, *61* (23), 2965–2978.

Mendelsohn, J. Epidermal Growth Factor Receptor Inhibition by a Monoclonal Antibody as Anticancer Therapy. *Clin. Cancer Res.* **1997**, *3* (12 Pt 2), 2703–2707.

von Mensdorff-Pouilly, S.; Gourevitch, M. M.; Kenemans, P.; Verstraeten, A. A.; Litvinov, S. V; van Kamp, G. J.; Meijer, S.; Vermorken, J.; Hilgers, J. Humoral Immune Response to Polymorphic Epithelial Mucin

(MUC-1) in Patients with Benign and Malignant Breast Tumours. *Eur J Cancer* **1996**, *32A* (8), 1325–1331.

Messick, C. A.; Sanchez, J.; DeJulius, K. L.; Hammel, J.; Ishwaran, H.; Kalady, M. F. CEACAM-7: A Predictive Marker for Rectal Cancer Recurrence. *Surgery* **2010**, *147* (5), 713–719.

Metz, C. E. Basic Principles of ROC Analysis. *Semin. Nucl. Med.* **1978**, *8* (4), 283–298.

Michot, J. M.; Bigenwald, C.; Champiat, S.; Collins, M.; Carbonnel, F.; Postel-Vinay, S.; Berdelou, A.; Varga, A.; Bahleda, R.; Hollebecque, A.; et al. Immune-Related Adverse Events with Immune Checkpoint Blockade: A Comprehensive Review. *Eur. J. Cancer* **2016**, *54*, 139–148.

Mollick, J. A.; Hodi, F. S.; Soiffer, R. J.; Nadler, L. M.; Dranoff, G. MUC1-like Tandem Repeat Proteins Are Broadly Immunogenic in Cancer Patients. *Cancer Immun.* **2003**, *3*, 1–17.

Mondal, G.; Ohashi, A.; Yang, L.; Rowley, M.; Couch, F. J. Tex14, a Plk1-Regulated Protein, Is Required for Kinetochore-Microtubule Attachment and Regulation of the Spindle Assembly Checkpoint. *Mol. Cell* **2012**, *45* (5), 680–695.

Morgan, R. A.; Dudley, M. E.; Wunderlich, J. R.; Hughes, M. S.; Yang, J. C.; Sherry, R. M.; Royal, R. E.; Topalian, S. L.; Kammula, U. S.; Restifo, N. P.; et al. Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes. *Science (80-.).* **2006**, *314* (5796), 126–129.

Mori, M.; Inoue, H.; Mimori, K.; Shibuta, K.; Baba, K.; Nakashima, H.; Haraguchi, M.; Tsuji, K.; Ueo, H.; Barnard, G. F.; et al. Expression of MAGE Genes in Human Colorectal Carcinoma. *Ann. Surg.* **1996**, *224* (2), 183–188.

Murakoshi, Y.; Honda, K.; Sasazuki, S.; Ono, M.; Negishi, A.; Matsubara, J.; Sakuma, T.; Kuwabara, H.; Nakamori, S.; Sata, N.; et al. Plasma Biomarker Discovery and Validation for Colorectal Cancer by Quantitative Shotgun Mass Spectrometry and Protein Microarray. *Cancer Sci.* **2011**, *102* (3), 630–638.

Nagashima, H.; Sadanaga, N.; Mashino, K.; Yamashita, K.; Inoue, H.; Mori, M.; Sugimachi, K. Expression of Hqk Encoding a KH RNA Binding Protein Is Altered in Human Glioma. *Japanese J. Cancer Res.* **2002**, *93* (2), 167–177.

Naimi, B.; Latil, A.; Fournier, G.; Mangin, P.; Cussenot, O.; Berthon, P. Down-Regulation of (IIIb) and (IIIc) Isoforms of Fibroblast Growth Factor Receptor 2 (FGFR2) Is Associated with Malignant Progression in Human Prostate. *Prostate* **2002**, *52* (3), 245–252.

Nam, M. J.; Madoz-Gurpide, J.; Wang, H.; Lescure, P.; Schmalbach, C. E.; Zhao, R.; Misek, D. E.; Kuick, R.;

Brenner, D. E.; Hanash, S. M. Molecular Profiling of the Immune Response in Colon Cancer Using Protein Microarrays: Occurrence of Autoantibodies to Ubiquitin C-Terminal Hydrolase L3. *Proteomics* **2003**, *3* (11), 2108–2115.

Nan, H.; Qureshi, A. A.; Hunter, D. J.; Han, J. Genetic Variants in FGFR2 and FGFR4 Genes and Skin Cancer Risk in the Nurses' Health Study. *BMC Cancer* **2009**, *9*, 172.

Neeli, I.; Khan, S. N.; Radic, M. Histone Deimination As a Response to Inflammatory Stimuli in Neutrophils. *J. Immunol.* **2008**, *180* (3), 1895–1902.

Nelson, B. H. The Impact of T-Cell Immunity on Ovarian Cancer Outcomes. *Immunol. Rev.* **2008**, *222* (1), 101–116.

O'Reilly, K. E.; Rojo, F.; She, Q.-B.; Solit, D.; Mills, G. B.; Smith, D.; Lane, H.; Hofmann, F.; Hicklin, D. J.; Ludwig, D. L.; et al. MTOR Inhibition Induces Upstream Receptor Tyrosine Kinase Signaling and Activates Akt. *Cancer Res.* **2006**, *66* (3), 1500–1508.

O'Rourke, D. J.; DiJohnson, D. A.; Caiazzo, R. J.; Nelson, J. C.; Ure, D.; O'Leary, M. P.; Richie, J. P.; Liu, B. C. S. Autoantibody Signatures as Biomarkers to Distinguish Prostate Cancer from Benign Prostatic Hyperplasia in Patients with Increased Serum Prostate Specific Antigen. *Clin. Chim. Acta* **2012**, *413* (5–6), 561–567.

Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–386.

Ong, S. E.; Mann, M. A Practical Recipe for Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *Nat. Protoc.* **2006**, *1* (6), 2650–2660.

Ortenberg, R.; Sapoznik, S.; Zippel, D.; Shapira-Frommer, R.; Itzhaki, O.; Kubi, A.; Zikich, D.; Besser, M. J.; Schachter, J.; Markel, G. Serum CEACAM1 Elevation Correlates with Melanoma Progression and Failure to Respond to Adoptive Cell Transfer Immunotherapy. *J. Immunol. Res.* **2015**, *2015*, 1–8.

Pan, D.; McCahy, P. Patient Knowledge about Prostate-Specific Antigen (PSA) and Prostate Cancer in Australia. *BJU Int.* **2012**, *109* (SUPPL. 3), 52–56.

Papadopoulos, N.; Nicolaides, N. C.; Wei, Y. F.; Ruben, S. M.; Carter, K. C.; Rosen, C. A.; Haseltine, W. A.; Fleischmann, R. D.; Fraser, C. M.; Adams, M. D. Mutation of a MutL Homolog in Hereditary Colon Cancer.

Science **1994**, *263* (5153), 1625–1629.

Parry, R. V.; Chemnitz, J. M.; Frauwirth, K. A.; Lanfranco, A. R.; Braunstein, I.; Kobayashi, S. V.; Linsley, P. S.; Thompson, C. B.; Riley, J. L. CTLA-4 and PD-1 Receptors Inhibit T-Cell Activation by Distinct Mechanisms. *Mol. Cell. Biol.* **2005**, *25* (21), 9543–9553.

Passiglia, F.; Bronte, G.; Bazan, V.; Natoli, C.; Rizzo, S.; Galvano, A.; Listì, A.; Cicero, G.; Rolfo, C.; Santini, D.; et al. PD-L1 Expression as Predictive Biomarker in Patients with NSCLC: A Pooled Analysis. *Oncotarget* **2016**, *7* (15), 19738–19747.

Peace, D. J.; Chen, W.; Nelson, H.; Cheever, M. A. T Cell Recognition of Transforming Proteins Encoded by Mutated Ras Proto-Oncogenes. *J. Immunol. (Baltimore, Md. 1950)* **1991**, *146* (6), 2059–2065.

Pedersen, J. W.; Wandall, H. H. Autoantibodies as Biomarkers in Cancer. *Lab. Med.* **2011**, *42* (10), 623–628.

Piccart-Gebhart, M. J.; Procter, M.; Leyland-Jones, B.; Goldhirsch, A.; Untch, M.; Smith, I.; Gianni, L.; Baselga, J.; Bell, R.; Jackisch, C.; et al. Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *N. Engl. J. Med.* **2005**, *353* (16), 1659–1672.

Piura, E.; Piura, B. Autoantibodies to Tailor-Made Panels of Tumor-Associated Antigens in Breast Carcinoma. *J. Oncol.* **2011**, *7*.

De Plaen, E.; Traversari, C.; Gaforio, J. J.; Szikora, J. P.; De Smet, C.; Brasseur, F.; van der Bruggen, P.; Lethé, B.; Lurquin, C.; Chomez, P.; et al. Structure, Chromosomal Localization, and Expression of 12 Genes of the MAGE Family. *Immunogenetics* **1994**, *40* (5), 360–369.

Qi, H. T Follicular Helper Cells in Space-Time. *Nat. Rev. Immunol.* **2016**, *16* (10), 612–625.

Qi, L.; Ding, Y. Screening of Differentiation-Specific Molecular Biomarkers for Colon Cancer. *Cell. Physiol. Biochem.* **2018**, *46* (6), 2543–2550.

Qiu, J.; Madoz-Gurpide, J.; Misek, D. E.; Kuick, R.; Brenner, D. E.; Michailidis, G.; Haab, B. B.; Omenn, G. S.; Hanash, S. Development of Natural Protein Microarrays for Diagnosing Cancer Based on an Antibody Response to Tumor Antigens. *J. Proteome Res.* **2004**, *3* (2), 261–267.

Quackenbush, J. Computational Analysis of Microarray Data. *Nat. Rev. Genet.* **2001**, *2* (6), 418–427.

Qureshi, O. S.; Zheng, Y.; Nakamura, K.; Attridge, K.; Manzotti, C.; Schmidt, E. M.; Baker, J.; Jeffery, L. E.;

Kaur, S.; Briggs, Z.; et al. Trans-Endocytosis of CD80 and CD86: A Molecular Basis for the Cell-Extrinsic Function of CTLA-4. *Science* **2011**, *332* (6029), 600–603.

Rang, Z.; Yang, G.; Wang, Y. W.; Cui, F. MIR-542-3p Suppresses Invasion and Metastasis by Targeting the Proto-Oncogene Serine/Threonine Protein Kinase, PIM1, in Melanoma. *Biochem. Biophys. Res. Commun.* **2016**, *474* (2), 315–320.

Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips. *Nat. Protoc.* **2007**, *2* (8), 1896–1906.

Reipert, B. M.; Tanneberger, S.; Pannetta, A.; Bedosti, M.; Poell, M.; Zimmermann, K.; Stellamor, M. T. Increase in Autoantibodies against Fas (CD95) during Carcinogenesis in the Human Colon: A Hope for the Immunoprevention of Cancer? *Cancer Immunol. Immunother.* **2005**, *54* (10), 1038–1042.

Reuschenbach, M.; Von, M.; Doeberitz, K.; Wentzensen, N. A Systematic Review of Humoral Immune Responses against Tumor Antigens. *Cancer Immunol Immunother* **2009**, *58*, 1535–1544.

Rieger, A.; Wahren, B. CEA Levels at Recurrence and Metastases; Importance for Detecting Secondary Disease. *Scand J Gastroenterol* **1975**, *10* (8), 869–874.

Robert, C.; Thomas, L.; Bondarenko, I.; O'Day, S.; Weber, J.; Garbe, C.; Lebbe, C.; Baurain, J.-F.; Testori, A.; Grob, J.-J.; et al. Ipilimumab plus Dacarbazine for Previously Untreated Metastatic Melanoma. *N. Engl. J. Med.* **2011**, *364* (26), 2517–2526.

Rogers, G. E.; Simmonds, D. H. Content of Citrulline and Other Amino-Acids in a Protein of Hair Follicles. *Nature* **1958**, *182* (4629), 186–187.

Rognum, T. O. A New Approach in Carcinoembryonic Antigen-Guided Follow-up of Large-Bowel Carcinoma Patients. *Scand. J. Gastroenterol.* **1986**, *21* (6), 641–649.

Romond, E. H.; Perez, E. A.; Bryant, J.; Suman, V. J.; Geyer, C. E.; Davidson, N. E.; Tan-Chiu, E.; Martino, S.; Paik, S.; Kaufman, P. A.; et al. Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer. *N. Engl. J. Med.* **2005**, *353* (16), 1673–1684.

Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; et al. Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169.

Ruiz-Cabello, F.; Garrido, F. HLA and Cancer: From Research to Clinical Impact. *Immunol. Today* **1998**, *19*

(12), 539–542.

Sahin, U.; Tureci, O.; Schmitfr, H.; Cochlovius, B.; Johannes, T.; Schmits, R.; Stenner, F.; Luo, G.; Schobert, I.; Pfreundschuht, M. Human Neoplasms Elicit Multiple Specific Immune Responses in the Autologous Host. *Immunology* **1995**, *92*, 11810–11813.

Saiki, M.; Watase, M.; Matsubayashi, H.; Hidaka, Y. Recognition of the N-Terminal Histone H2A and H3 Peptides by Peptidylarginine Deiminase IV. *Protein Pept. Lett.* **2009**, *16* (9), 1012–1016.

Salvi, S.; Fontana, V.; Boccardo, S.; Merlo, D. F.; Margallo, E.; Laurent, S.; Morabito, A.; Rijavec, E.; Dal Bello, M. G.; Mora, M.; et al. Evaluation of CTLA-4 Expression and Relevance as a Novel Prognostic Factor in Patients with Non-Small Cell Lung Cancer. *Cancer Immunol. Immunother.* **2012**, *61* (9), 1463–1472.

Scanlan, M. J.; Gure, A. O.; Jungbluth, A. A.; Old, L. J.; Chen, Y.-T. Cancer/Testis Antigens: An Expanding Family of Targets for Cancer. *Immunol. Rev.* **2002**, *188*, 22–32.

Schadendorf, D.; Hodi, F. S.; Robert, C.; Weber, J. S.; Margolin, K.; Hamid, O.; Patt, D.; Chen, T. T.; Berman, D. M.; Wolchok, J. D. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin Oncol* **2015**, *33* (17), 1889–1894.

Schäfer, G.; Graham, L. M.; Lang, D. M.; Blumenthal, M. J.; Bergant Marušič, M.; Katz, A. A. Vimentin Modulates Infectious Internalization of Human Papillomavirus 16 Pseudovirions. *J. Virol.* **2017**, *91* (16), e00307-17.

Schechter, A. L.; Stern, D. F.; Vaidyanathan, L.; Decker, S. J.; Drebin, J. A.; Greene, M. I.; Weinberg, R. A. The Neu Oncogene: An Erb-B-Related Gene Encoding a 185,000-Mr Tumour Antigen. *Nature* **1984**, *312* (5994), 513–516.

Schellekens, G. A.; De Jong, B. A. W.; Van Den Hoogen, F. H. J.; Van De Putte, L. B. A.; Van Venrooij, W. J. Citrulline Is an Essential Constituent of Antigenic Determinants Recognized by Rheumatoid Arthritis-Specific Autoantibodies. *J. Clin. Invest.* **1998**, *101* (1), 273–281.

Schö, S.; Zimmermann, W.; Schwarzkopf, G.; Grunert, F.; Rogaczewski, B.; Thompson, J. Carcinoembryonic Antigen Family Members CEACAM6 and CEACAM7 Are Differentially Expressed in Normal Tissues and Oppositely Deregulated in Hyperplastic Colorectal Polyps and Early Adenomas. *Am J Pathol* **2000**, *156*, 595–605.

Schroeder, H. W.; Cavacini, L. Structure and Function of Immunoglobulins. *J. Allergy Clin. Immunol.* **2010**,

125 (2), S41--S52.

Schwanhüusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global Quantification of Mammalian Gene Expression Control. *Nature* **2011**, *473* (7347), 337–342.

Schwitalle, Y.; Kloor, M.; Eiermann, S.; Linnebacher, M.; Kienle, P.; Knaebel, H. P.; Tariverdian, M.; Benner, A.; von Knebel Doeberitz, M. Immune Response Against Frameshift-Induced Neopeptides in HNPCC Patients and Healthy HNPCC Mutation Carriers. *Gastroenterology* **2008**, *134* (4), 988–997.

Seliger, B.; Maeurer, M. J.; Ferrone, S. Antigen-Processing Machinery Breakdown and Tumor Growth. *Immunol. Today* **2000**, *21* (9), 455–464.

Senshu, T.; Akiyama, K.; Kan, S.; Asaga, H.; Ishigami, A.; Manabe, M. Detection of Deiminated Proteins in Rat Skin: Probing with a Monospecific Antibody after Modification of Citrulline Residues. *J. Invest. Dermatol.* **1995**, *105* (2), 163–169.

Sethi, M. K.; Thaysen-Andersen, M.; Kim, H.; Park, C. K.; Baker, M. S.; Packer, N. H.; Paik, Y. K.; Hancock, W. S.; Fanayan, S. Quantitative Proteomic Analysis of Paired Colorectal Cancer and Non-Tumorigenic Tissues Reveals Signature Proteins and Perturbed Pathways Involved in CRC Progression and Metastasis. *J. Proteomics* **2015**, *126*, 54–67.

Shannan, B.; Watters, A.; Chen, Q.; Mollin, S.; Dörr, M.; Meggers, E.; Xu, X.; Gimotty, P. A.; Perego, M.; Li, L.; et al. PIM Kinases as Therapeutic Targets against Advanced Melanoma. *Oncotarget* **2016**, *7* (34), 54897–54912.

Shapiro, S. S.; Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52* (3/4), 591.

Shay, J. W.; Wright, W. E. Hayflick, His Limit, and Cellular Ageing. *Nat. Rev. Mol. Cell Biol.* **2000**, *1* (October), 72–76.

Siegel, R.; Desantis, C.; Jemal, A. Colorectal Cancer Statistics, 2014. *CA Cancer J. Clin.* **2014**, *64* (1), 104–117.

Simpson, A. J. G.; Caballero, O. L.; Jungbluth, A.; Chen, Y.-T.; Old, L. J. Cancer/Testis Antigens, Gametogenesis and Cancer. *Nat. Rev. Cancer* **2005**, *5* (8), 615–625.

Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: Visualizing Classifier Performance in R. *Bioinformatics* **2005**, *21* (20), 3940–3941.

Sivan, S.; Suzan, F.; Rona, O.; Tamar, H.; Vivian, B.; Tamar, P.; Jacob, S.; Gal, M.; Michal, L. Serum CEACAM1 Correlates with Disease Progression and Survival in Malignant Melanoma Patients. *Clin. Dev. Immunol.* **2012**, *2012*, 290536.

Skipper, J.; Stauss, H. J. Identification of Two Cytotoxic T Lymphocyte-Recognized Epitopes in the Ras Protein. *J. Exp. Med.* **1993**, *177* (5), 1493–1498.

Slamon, D. J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; et al. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *N. Engl. J. Med.* **2001**.

Slater, G.; Papatestas, A. E.; Aufses, A. H. Preoperative Carcinoembryonic Antigen Levels in Colorectal Carcinoma. *Arch. Surg.* **1979**, *114* (1), 52–53.

Smith, G. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (80-.).* **1985a**, *228* (4705), 1315–1317.

Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985b**, *228* (4705), 1315–1317.

Smith, H. A.; McNeel, D. G. The SSX Family of Cancer-Testis Antigens as Target Proteins for Tumor Therapy. *Clin. Dev. Immunol.* **2010**, *2010*.

Snyder, A.; Makarov, V.; Merghoub, T.; Yuan, J.; Zaretsky, J. M.; Desrichard, A.; Walsh, L. a; Postow, M. a; Wong, P.; Ho, T. S.; et al. Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* **2014**, 2189–2199.

Soussi, T. P53 Antibodies in the Sera of Patients with Various Types of Cancer: A Review. *Cancer Res.* **2000**, *60* (7), 1777–1788.

Stamey, T. A.; Yang, N.; Hay, A. R.; McNeal, J. E.; Freiha, F. S.; Redwine, E. Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate. *N. Engl. J. Med.* **1987**, *317* (15), 909–916.

Stauss, H. J.; Davies, H.; Sadovnikova, E.; Chain, B.; Horowitz, N.; Sinclair, C. Induction of Cytotoxic T Lymphocytes with Peptides in Vitro: Identification of Candidate T-Cell Epitopes in Human Papilloma Virus. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (17), 7871–7875.

Stenman, U.-H.; Alfthan, H.; Hotakainen, K. Human Chorionic Gonadotropin in Cancer. *Clin. Biochem.* **2004**, *37* (7), 549–561.

Stockert, E.; Jäger, E.; Chen, Y.-T.; Scanlan, M. J.; Gout, I.; Karbach, J.; Arand, M.; Knuth, A.; Old, L. J. A Survey of the Humoral Immune Response of Cancer Patients to a Panel of Human Tumor Antigens. *J. Exp. Med.* **1998**, *187* (8), 1349–1354.

Sudarsanam, S.; Johnson, D. E. Functional Consequences of MTOR Inhibition. *Curr. Opin. Drug Discov. Devel.* **2010**, *13* (1), 31–40.

Suwarnalata, G.; Tan, A. H.; Isa, H.; Gudimella, R.; Anwar, A.; Loke, M. F.; Mahadeva, S.; Lim, S. Y.; Vadivelu, J. Augmentation of Autoantibodies by Helicobacter Pylori in Parkinson's Disease Patients May Be Linked to Greater Severity. *PLoS One* **2016**, *11* (4).

Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; et al. The STRING Database in 2011: Functional Interaction Networks of Proteins, Globally Integrated and Scored. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 1).

Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.; et al. STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Res.* **2015**, *43* (D1), D447–D452.

Talmadge, J. E.; Fidler, I. J. AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective. *Cancer Research*. 2010, pp 5649–5669.

Tarcsa, E.; Marekov, L. N.; Mei, G.; Melino, G.; Lee, S. C.; Steinert, P. M. Protein Unfolding by Peptidylarginine Deiminase: Substrate Specificity and Structural Relationships of the Natural Substrates Trichohyalin and Filaggrin. *J. Biol. Chem.* **1996**, *271* (48), 30709–30716.

Taube, J. M.; Klein, A.; Brahmer, J. R.; Xu, H.; Pan, X.; Kim, J. H.; Chen, L.; Pardoll, D. M.; Topalian, S. L.; Anders, R. A. Association of PD-1, PD-1 Ligands, and Other Features of the Tumor Immune Microenvironment with Response to Anti-PD-1 Therapy. *Clin. Cancer Res.* **2014**, *20* (19), 5064–5074.

Teng, M. W. L.; Swann, J. B.; Koebel, C. M.; Schreiber, R. D.; Smyth, M. J. Immune-Mediated Dormancy: An Equilibrium with Cancer. *J. Leukoc. Biol.* **2008**, *84* (4), 988–993.

Thibodeau, J.; Bourgeois-Daigneault, M.-C.; Lapointe, R. Targeting the MHC Class II Antigen Presentation Pathway in Cancer Immunotherapy. *Oncoimmunology* **2012**, *1* (6), 908–916.

Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by

MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904.

Thompson, J.; Seitz, M.; Chastre, E.; Ditter, M.; Aldrian, C.; Gespach, C.; Zimmermann, W. Down-Regulation of Carcinoembryonic Antigen Family Member 2 Expression Is an Early Event in Colorectal Tumorigenesis. *Cancer Res.* **1997**, *57* (9), 1776–1784.

Thomson, D. M.; Krupey, J.; Freedman, S. O.; Gold, P. The Radioimmunoassay of Circulating Carcinoembryonic Antigen of the Human Digestive System. *Proc. Natl. Acad. Sci. U. S. A.* **1969**, *64* (1), 161–167.

Toh, H. C.; Wang, W. W.; Chia, W. K.; Kvistborg, P.; Sun, L.; Teo, K.; Phoon, Y. P.; Soe, Y.; Tan, S. H.; Hee, S. W.; et al. Clinical Benefit of Allogeneic Melanoma Cell Lysate-Pulsed Autologous Dendritic Cell Vaccine in MAGE-Positive Colorectal Cancer Patients. *Clin. Cancer Res.* **2009**, *15* (24), 7726–7736.

Tomei, S.; Bedognetti, D.; De Giorgi, V.; Sommariva, M.; Civini, S.; Reinboth, J.; Al Hashmi, M.; Ascierto, M. L.; Liu, Q.; Ayotte, B. D.; et al. The Immune-Related Role of BRAF in Melanoma. *Mol. Oncol.* **2015**, *9* (1), 93–104.

Topalian, S. L.; Sznol, M.; McDermott, D. F.; Kluger, H. M.; Carvajal, R. D.; Sharfman, W. H.; Brahmer, J. R.; Lawrence, D. P.; Atkins, M. B.; Powderly, J. D.; et al. Survival, Durable Tumor Remission, and Long-Term Safety in Patients with Advanced Melanoma Receiving Nivolumab. *J. Clin. Oncol.* **2014**, *32* (10), 1020–1030.

Türeci, O.; Sahin, U.; Zwick, C.; Koslowski, M.; Seitz, G.; Pfreundschuh, M. Identification of a Meiosis-Specific Protein as a Member of the Class of Cancer/Testis Antigens. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (9), 5211–5216.

Türeci, Ö.; Sahin, U.; Schobert, I.; Koslowski, M.; Schmitt, H.; Schild, H. J.; Stenner, F.; Seitz, G.; Rammensee, H. G.; Pfreundschuh, M. The SSX-2 Gene, Which Is Involved in the t(X;18) Translocation of Synovial Sarcomas, Codes for the Human Tumor Antigen HOM-MEL-40. *Cancer Res.* **1996**, *56* (20), 4766–4772.

Turriziani, B.; Garcia-Munoz, A.; Pilkington, R.; Raso, C.; Kolch, W.; von Kriegsheim, A. On-Beads Digestion in Conjunction with Data-Dependent Mass Spectrometry: A Shortcut to Quantitative and Dynamic Interaction Proteomics. *Biology (Basel).* **2014**, *3* (2), 320–332.

Tutturen, A. E. V; Fleckenstein, B.; De Souza, G. A. Assessing the Citrullinome in Rheumatoid Arthritis

Synovial Fluid with and without Enrichment of Citrullinated Peptides. *J. Proteome Res.* **2014**, *13* (6), 2867–2873.

Velazquez, E. F.; Jungbluth, A. A.; Yancovitz, M.; Gnjatic, S.; Adams, S.; O'Neill, D.; Zavilevich, K.; Albukh, T.; Christos, P.; Mazumdar, M.; et al. Expression of the Cancer/Testis Antigen NY-ESO-1 in Primary and Metastatic Malignant Melanoma (MM)--Correlation with Prognostic Factors. *Cancer Immun.* **2007**, *7*, 11.

Vining, D. J.; Gladish, G. W. Receiver Operating Characteristic Curves: A Basic Understanding. *Radiographics* **1992**, *12* (6), 1147–1154.

de Vries, E.; Nijsten, T. E. C.; Visser, O.; Bastiaannet, E.; van Hattem, S.; Janssen-Heijnen, M. L.; Coebergh, J. W. W. Superior Survival of Females among 10 538 Dutch Melanoma Patients Is Independent of Breslow Thickness, Histologic Type and Tumor Site. *Ann. Oncol.* **2008**, *19* (3), 583–589.

Wan, F.; Wang, H.; Shen, Y.; Zhang, H.; Shi, G.; Zhu, Y.; Dai, B.; Ye, D. Upregulation of COL6A1 Is Predictive of Poor Prognosis in Clear Cell Renal Cell Carcinoma Patients. *Oncotarget* **2015**, *6* (29), 27378–27387.

Wanebo, H. J.; Rao, B.; Pinsky, C. M.; Hoffman, R. G.; Stearns, M.; Schwartz, M. K.; Oettgen, H. F. Preoperative Carcinoembryonic Antigen Level as a Prognostic Indicator in Colorectal Cancer. *N. Engl. J. Med.* **1978**, *299* (9), 448–451.

Wang, D.; Yang, L.; Zhang, P.; LaBaer, J.; Hermjakob, H.; Li, D.; Yu, X. AAgAtlas 1.0: A Human Autoantigen Database. *Nucleic Acids Res.* **2017**, *45* (D1), D769–D776.

Wang, L.; Chang, X.; Yuan, G.; Zhao, Y.; Wang, P. Expression of Peptidylarginine Deiminase Type 4 in Ovarian Tumors. *Int. J. Biol. Sci.* **2010**, *6* (5), 454–464.

Wang, P.; Wilson, S. R. Mass Spectrometry-Based Protein Identification by Integrating de Novo Sequencing with Database Searching. *BMC Bioinformatics* **2013**, *14*, S24.

Wang, R. F.; Johnston, S. L.; Zeng, G.; Topalian, S. L.; Schwartzentruber, D. J.; Rosenberg, S. A.; Suzanne, L.; Schwartzentruber, D. J.; Steven, A.; Topalian, S. L.; et al. A Breast and Melanoma-Shared Tumor Antigen: T Cell Responses to Antigenic Peptides Translated from Different Open Reading Frames. *J. Immunol.* **1998**, *161* (7), 3598–3606.

Wang, X. *Bioinformatics of Human Proteomics (Google EBook)*; 2013.

Wang, X.; Yu, J.; Sreekumar, A.; Varambally, S.; Shen, R.; Giacherio, D.; Mehra, R.; Montie, J. E.; Pienta, K.

J.; Sanda, M. G.; et al. Autoantibody Signatures in Prostate Cancer. *N. Engl. J. Med.* **2005**, *353* (12), 1224–1235.

Wang, Y. Human PAD4 Regulates Histone Arginine Methylation Levels via Demethylimination. *Science (80-.).* **2004**, *306* (5694), 279–283.

Wardemann, H.; Nussenzweig, M. C. B-Cell Self-Tolerance in Humans. *Adv. Immunol.* **2007**, *95*, 83–110.

Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.

Weon, J. L.; Potts, P. R. The MAGE Protein Family and Cancer. *Curr. Opin. Cell Biol.* **2015**, *37*, 1–8.

Wherry, E. J. T Cell Exhaustion. *Nat. Immunol.* **2011**, *131* (6), 492–499.

Whiteman, D. C.; Bray, C. A.; Siskind, V.; Green, A. C.; Hole, D. J.; Mackie, R. M. Changes in the Incidence of Cutaneous Melanoma in the West of Scotland and Queensland, Australia: Hope for Health Promotion? *Eur. J. Cancer Prev.* **2008**, *17*, 243–250.

WHO. Globocan 2012 - Home. *Globocan 2012*. 2012.

Wilmink, A. B. M. Overview of the Epidemiology of Colorectal Cancer. *Dis. Colon Rectum* **1997**, *40* (4), 483–493.

Wilson, A. P. M.; Van Dalen, A.; Sibley, P. E. C.; Kasper, L. A.; Durham, A. P.; El Shami, A. S. Multicentre Tumour Marker Reference Range Study. *Anticancer Res.* **1999**, *19* (4 A), 2749–2752.

Witalison, E. E.; Thompson, P. R.; Hofseth, L. J. Protein Arginine Deiminases and Associated Citrullination: Physiological Functions and Diseases Associated with Dysregulation. *Curr Drug Targets* **2015**, *16* (7), 700–710.

Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, *73* (23), 5683–5690.

Wu, C. J.; Ritz, J. Induction of Tumor Immunity Following Allogeneic Stem Cell Transplantation. *Adv Immunol* **2006**, *90*, 133–173.

Wu, C. J.; Biernacki, M.; Kutok, J. L.; Rogers, S.; Chen, L.; Yang, X.-F.; Soiffer, R. J.; Ritz, J. Graft-versus-Leukemia Target Antigens in Chronic Myelogenous Leukemia Are Expressed on Myeloid Progenitor Cells. *Clin. Cancer Res.* **2005**, *11* (12), 4504–4511.

Yang, H.-Y.; Kwon, J.; Park, H.-R.; Kwon, S.-O.; Park, Y.-K.; Kim, H.-S.; Chung, Y.-J.; Chang, Y.-J.; Choi, H.-I.; Chung, K.-J.; et al. Comparative Proteomic Analysis for the Insoluble Fractions of Colorectal Cancer Patients. *J. Proteomics* **2012**, *75* (12), 3639–3653.

Yee, C.; Thompson, J.; Roche, P.; Byrd, D.; Lee, P.; Piepkorn, M.; Kenyon, K.; Davis, M.; Riddell, S.; Greenberg, P. Melanocyte Destruction after Antigen-Specific Immunotherapy of Melanoma: Direct Evidence of T Cell-Mediated Vitiligo. *J. Exp. Med* **2000**, *192* (11), 252.

Yee, C.; Thompson, J. A.; Byrd, D.; Riddell, S. R.; Roche, P.; Celis, E.; Greenberg, P. D. Adoptive T-Cell Therapy Using Antigen-Specific CD8+ T-Cell Clones for the Treatment of Patients with Metastatic Melanoma: In Vivo Persistence, Migration, and Antitumor Effect of Transferred T-Cells. *Proc Natl Acad Sci U S A* **2002**, *99* (25), 16168–16173.

Yoshizawa, S.; Matsuoka, K.; Inoue, N.; Takaishi, H.; Ogata, H.; Iwao, Y.; Mukai, M.; Fujita, T.; Kawakami, Y.; Hibi, T. Clinical Significance of Serum P53 Antibodies in Patients with Ulcerative Colitis and Its Carcinogenesis. *Inflamm.Bowel.Dis.* **2007**, *13* (7), 865–873.

Yuan, J.; Hegde, P. S.; Clynes, R.; Foukas, P. G.; Harari, A.; Kleen, T. O.; Kvistborg, P.; Maccalli, C.; Maecker, H. T.; Page, D. B.; et al. Novel Technologies and Emerging Biomarkers for Personalized Cancer Immunotherapy. *J. Immunother. cancer* **2016**, *4*.

Yuan, T. L.; Cantley, L. C. PI3K Pathway Alterations in Cancer: Variations on a Theme. *Oncogene* **2008**, *27* (41), 5497–5510.

Zaenker, P.; Ziman, M. R. Serologic Autoantibodies as Diagnostic Cancer Biomarkers — A Review. **2013**, 1–22.

Zaenker, P.; Gray, E. S.; Ziman, M. R. Autoantibody Production in Cancer-The Humoral Immune Response toward Autologous Antigens in Cancer Patients. *Autoimmun. Rev.* **2016**, *15* (5), 477–483.

Zayakin, P.; Ancans, G.; Siliaņa, K.; Meistere, I.; Kalniņa, Z.; Andrejeva, D.; Endzeliņš, E.; Ivanova, L.; Pismennaja, A.; Ruskule, A.; et al. Tumor-Associated Autoantibody Signature for the Early Detection of Gastric Cancer. *Int. J. Cancer* **2013**, *132* (1), 137–147.

Zeliadt, S. B.; Hoffman, R. M.; Etzioni, R.; Gore, J. L.; Kessler, L. G.; Lin, D. W. Influence of Publication of US and European Prostate Cancer Screening Trials on PSA Testing Practices. *J. Natl. Cancer Inst.* **2011**, *103* (6), 520–523.

Zhang, J. Y.; Chan, E. K.; Peng, X. X.; Tan, E. M. A Novel Cytoplasmic Protein with RNA-Binding Motifs Is an Autoantigen in Human Hepatocellular Carcinoma. *J. Exp. Med.* **1999**, *189* (7), 1101–1110.

Zhang, S. Y.; Lin, M.; Zhang, H. B. Diagnostic Value of Carcinoembryonic Antigen and Carcinoma Antigen 19-9 for Colorectal Carcinoma. *Int. J. Clin. Exp. Pathol.* **2015**, *8* (8), 9404–9409.

Zhang, X.; Bolt, M.; Guertin, M. J.; Chen, W.; Zhang, S.; Cherrington, B. D.; Slade, D. J.; Dreyton, C. J.; Subramanian, V.; Bicker, K. L.; et al. Peptidylarginine Deiminase 2-Catalyzed Histone H3 Arginine 26 Citrullination Facilitates Estrogen Receptor Target Gene Activation. *Proc. Natl. Acad. Sci.* **2012**, *109* (33), 13331–13336.

Zhang, Y.; Wang, H.; Toratani, S.; Sato, J. D.; Kan, M.; McKeehan, W. L.; Okamoto, T. Growth Inhibition by Keratinocyte Growth Factor Receptor of Human Salivary Adenocarcinoma Cells through Induction of Differentiation and Apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (20), 11336–11340.

Zhong, L.; Ge, K.; Zu, J.; Zhao, L.; Shen, W.; Wang, J.; Zhang, X.; Gao, X.; Hu, W.; Yen, Y.; et al. Autoantibodies as Potential Biomarkers for Breast Cancer. *Breast Cancer Res.* **2008**, *10* (3), R40.

Zhou, H.; Ning, Z.; E. Starr, A.; Abu-Farha, M.; Figeys, D. Advancements in Top-down Proteomics. *Anal. Chem.* **2012**, *84* (2), 720–734.

Zhu, P.; Hu, C.; Hui, K.; Jiang, X. The Role and Significance of VEGFR2+regulatory T Cells in Tumor Immunity. *Onco. Targets. Ther.* **2017**, *10*, 4315–4319.

Zhu, Y.-P.; Wan, F.-N.; Shen, Y.-J.; Wang, H.-K.; Zhang, G.-M.; Ye, D.-W. Reactive Stroma Component COL6A1 Is Upregulated in Castration-Resistant Prostate Cancer and Promotes Tumor Growth. *Oncotarget* **2015**, *6* (16), 14488–14496.

Zweig, M. H.; Campbell, G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.* **1993**, *394* (4), 561–577.