Characterization of antibody landscapes following curative immunotherapy responses and induced immunological memory in a murine model of melanoma

By:

Anna Hoefges

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Cellular and Molecular Pathology)

At the
UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: June 6th, 2023
The dissertation is approved by the following members of the final oral committee:
Paul M. Sondel, Professor, Human Oncology, Pediatrics
Christian M. Capitini, Associate Professor, Pediatrics
Manish Patankar, Associate Professor, Obstetrics and Gynecology
Zachary S. Morris, Assistant Professor, Human Oncology
Zsuzsana Fabry, Professor, Pathology and Laboratory Medicine

Acknowledgements

With the completion of this thesis I am ending my life as a student and am looking forward to the next stage of my career and beyond. I am grateful for the experiences and opportunities I have gained throughout my time as a graduate student and the people I have encountered along the way.

I would like to express my deepest gratitude to Dr. Paul Sondel, my mentor throughout this PhD. Who would have thought that a small research project I did during my masters program in Sweden at Uppsala University, which was my first introduction to cancer immunotherapy, would have led me to pursuing a PhD in Pauls lab. This small research project caught my interest in cancer immunotherapy and moved me to send out emails to laboratories across the US working on cancer immun otherapy. Paul answered my email and it seemed like the logistics were already figured out. I interviewed with him and Sasha and a couple months later I arrived in Madison to conduct my master thesis research work. After completing my Masters in Sweden, I returned for my PhD. Paul has been an incredible mentor, he is one of the kindest people I know and I admire his down-to-earth mentality, his involvement in the research and strong belief that our research can and will make a difference in someones life. I am grateful for the opportunity Paul gave me back then as a master student and again, as a PhD student in his lab, for his patience, time, effort and endless revisions of manuscripts, help with preparations for talks and posters, and scientific, as well as personal advice.

Dr. Sasha Rakhmilevich, thank you for being my first mentor in the lab, teaching me how to handle mice and all your puns throughout the years. I am grateful I got to know you and had you as a mentor. I love your sharp thinking and experimental design skills, how to make sure that the original question we set out to answer will actually be answered with the experiment and how to not get carried away with other interesting finds along the way.

Dr. Amy Erbe-Gurel, thank you for your friendship, mentorship, positive can-do attitude and endless ideas of experiments to do and questions to investigate. Thank you for always being there for me when I needed someone, for loving my children and spending time with Emma while having an incredibly busy schedule yourself, for always being willing to discuss my project and recent experimental findings.

Another special thanks goes out to Dr. Jackie Hank, thank you for being a friend, always a great colleague and paying attention to detail while never loosing track of the big picture, for discussions, help with designing experiments and for being a role model, on a scientific and personal level.

I would like to thank all the students that I have worked with over the years. To Mackenzie Heck, my first "own" undergraduate student who worked with me for 3 years before joining the lab after her graduation for a year as a full-time research intern. To Nick Mathers, who joined as a Shapiro summer student and became an instrumental part of the SDTG-analysis and is still invlovled in the project. To Drew Melby and Angie Xu, who helped me get a first understanding of this large dataset we generated and how to analyze, what to look for and how to organize our preliminary results while working with bioinformatics to develop the analysis method we needed.

To Arika Feils, who has been in the lab for the same time as me, first as an undergraduate student, and now as a research specialist. Thank you for all your help with organization, lab work, meetings, your friendship and help with Emma!

Thank you Dr. Alexa Heaton for being such a kind and amazing scientist and for taking the time to proof-read my thesis.

A special thanks goes out to former lab members Dr. Claire Baniel, Dr. Julie Voeller, Dr. Peter Carlson, Dr. Alex Pieper, and Dr. Lauren Zebertavage. I value your friendships and having had the chance to work alongside of these amazing people who are set out to make this world a better place! Lauren, I want to

thank you for all your career advice and help especially through this last year of my PhD, teaching me how to do western blots and always taking time to listen and be there for me in all areas of life.

To all present and former Sondel lab members, thank you so much for being a part of this journey and all the fantastic and engaging discussions throughout the years about the ongoing science in the lab.

Of course there are many people outside of the Sondel lab that I have to thank as well.

The two most important here are Dr. Irene Ong and Dr. Sean McIlwain. Sean and Irene, without your help we would have not been able to do this work. You were instrumental in generating the analysis method to analyze our large dataset. I thank you for your patience with us as we went through many iterations of the analysis method until we found the one. Sean, thank you for all the other bioinformatic work you have put into this project besides the generation of the algorithm. I truly appreciate you! Irene, I look forward to joining your lab later this summer and working with you and Sean, and getting to know your other lab members.

A special thanks goes out to Dr. Amber Bates and Dr. Ravi Patel, former members of the Morris lab that have become good friends and helpful mentors throughout their time at UW Madison and beyond. Ravi introduced lab happy hours and was instrumental in fostering a fun and team-oriented work environment and reminding everyone to work hard, but also enjoy life outside of the lab.

Another thank you goes out to my committee members, Dr. Zsuzsa Fabry, Dr. Manish Patankar, Dr. Christian Capitini and Dr. Zach Morris. I want to thank you for your service, advice and suggestions concerning my projects and overall career. I am grateful to have you as my mentors.

My CMP cohort, Rebecca, Lindsey, Steven, Caleb and Phil, thank you for being the best cohort I could have asked for. From day one I felt like we had a great connection and were always focused on helping each other. A special thanks goes out to Lindsey, Rebecca and Steven for your friendship even past graduate school.

Besides people who have influenced me here at UW, I also have to express my deepest gratitude towards my Parents, Iris & Carsten, for always supporting me in whatever I had decided would be the next step, letting me explore the world and giving me the most amazing childhood and parents I could ever ask for. Since becoming a parent myself, I do have a deeper understanding of what that meant. I truly appreciate you! I love watching you with Emma and William and seeing the joy we bring to your life. I am so grateful for all your support and presence especially during these last couple weeks of my PhD journey.

To Jesse, Emma and William. Jesse, thank you for sharing your life and your family with me. Thank you for being my supporter and my calm. Emma and William, you are my motivation to strive for better. You make me a better person, help me to see where I need to improve and what I want to pass on to you. You surely changed my life, and while it did not make my progress with this PhD easier, I wouldn't have it any other way. You gave my life meaning I didn't know it needed, seeing you grow every day and being your mom gives me the greatest joy.

To John, Jodi, Amy and the Rotar family, for including me in your family, helping out with the kids whenever needed and being the best auntie and grandparents always only a phone call away for Emma and William. We are truly lucky to have you!

To my mom friend group, for always being there for me and supporting me through life as a mother, but also as a graduate student and providing different perspectives helping me to see the full picture. You definitely made a difference in my parenting and life and I am eternally grateful to Meriter MBH for bringing us together and having you as a big part of my life.

Abstract

Cancer immunotherapy has revolutionized cancer treatment and has helped thousands of patients (Couzin-Frankel, 2013; Patel & Minn, 2018). However, most patients are still not showing positive responses to current cancer immunotherapy treatment regimens (Patel & Minn, 2018). Using radiation therapy (RT) and intratumoral injections of immunocytokine (IC), our lab has developed a local *in situ* vaccine regimen capable of curing mice bearing B78 melanoma tumors with protective immune memory (Morris et al., 2016). Our *in situ* vaccine (RT+IC) cures 70% of treated mice bearing a single large B78 tumor (which expresses GD2) and creates strong immunologic memory to reject a second challenge of B78 melanoma. We have also demonstrated that our *in situ* vaccine causes epitope spread; 75% of cured mice reject a challenge with B16 melanoma cells (which do not express the GD2 antigen), and we observed strong antibody-binding to B16 cells using serum from cured as compared to naïve mice (Baniel, Heinze, et al., 2020; Morris et al., 2016).

Although we observed epitope spread, the exact antigen targets of these endogenous antibodies were unknown. Knowledge of these additional targets could help to identify biomarkers of positive responses, as well as identify possible new therapeutic targets. In this thesis we utilized a peptide array approach to probe every mouse protein (broken into 16-mer peptides in a stepwise overlapping fashion) to identify antibody targets, using serum from cured mice vs. their matched naïve sample to identify immunodominant tumor antigens on cold murine tumors.

We were able to develop a robust analysis method (HERON) to identify a number of proteins recognized by multiple mice that we are further investigating (explained in detail in Chapter 2). We furthermore saw that within the top epitopes (measured by signal strength and detection in multiple mice) peptides containing a specific four amino acid long sequence were highly overrepresented. The presence of this sequence seemed to be associated with binding, although not being the sole source recognized for

antibody binding. Multiple analysis approaches have led us to know more about this motif while yet not being able to elucidate it fully (described in chapter 3).

Further investigation is needed into top proteins co-recognized by multiple immune mice as well as the four amino acid motif and possible implications and uses as biomarkers or treatment alternatives. We are pursuing experimental as well as bioinformatic approaches to enhance our knowledge of the identified antibody targets.

The methods generated will provide useful tools to identify the immunodominant tumor-specific antigens which may be translatable to mechanisms of resistance (and how to overcome such resistance) in human cancers and find potential uses as biomarkers or new treatment targets.

Table of Contents

Listing of Figures

List of Abbreviations

aa	Amino acid
ADCC	Antibody-Dependent Cell-Mediated Cytotoxicity
APC	Antigen presenting cell
ATP	Adenosine-5'-Triphosphate
BCC	Basal cell carcinoma
CD4	CD4+ T cell
CD8	CD8+ T cell
CHOP	Childrens Hospital of Philadelphia
CTL	Cytotoxic T Lymphocyte
CTLA4	Cytotoxic T-Lymphocyte Associated protein 4
DAMPS	Damage-Associated Molecular PatternS
DC	Dendritic cells
EBRT	External Beam Radiation Therapy
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme Linked Immunosorbent Assay
FDA	U.S. Food and Drug Administration
FDR	False Discovery Rate
Fmoc	Fluorenylmethoxycarbonyl
GD2	Disialognaglioside 2
GM-CSF	Granulocyte-macrophage colony-stimulating factor
Gy	Gray: SI unit of radiation defined as absorption of one joule of radiation energy per kilogram of matter
HERON	Hierarchical antibody binding Epitopes and pROteins from liNear peptides
HMGB1	High-Mobility Group protein B1
IC	Immunocytokine
ICC	Intraclass Correlation Coefficient
ICD	Immunogenic Cell Death
IFN-γ	Interferon-gamma
IgG	Immunoglobulin G
IL-2	Interleukin 2
ISV	In Situ Vaccine
KRAS	Kirsten rat sarcoma virus
mAb	Monoclonal Antibody
MDSC	Myeloid-Derived Suppressor Cell
MHC	Major histocompatibility complex
NK	Natural Killer
NKG2D	Natural Killer group 2 member D

NKT	Natural Killer T
NLRP3	NOD-Like Receptor family, Pyrin domain containing -3n protein
nm	nanometer
NPPOC	photosensitive 2-(2-nitrophenyl) propyl oxy- carbonyl
NSCLC	Non-small cell lung cancer
OD	Optical Density
PD1	Programmed cell death receptor 1, CD279
PDL1	Programmed cell Death receptor ligand 1, B7-H1, CD274
ROS	Reactive Oxygen Species
RT	Radiation Therapy, radiotherapy
SCC	Squamous cell carcinoma
scRNAseq	single cell RNA sequencing
SD	Standard Deviation
STK11	Serine/Threonine Kinase 11
TBS	Tris-Buffered Saline
TFA	Tri-Fluoroacetic Acid
TLR4	Toll-Like Receptor 4
TLS	Tertiary Lymphoid Structure
TME	Tumor microenvironment
Treg	T regulatory cell
TTDS	TrioxaTriDecan-Succinamic acid
TUP	Target Unrelated Peptide
VEGFR	Vascular Endothelial Growth Factor Receptor
γδ T cel	T cell expressing TCRγ and TCRδ subunits
µm	micrometer

Chapter 1: Introduction & Background

Melanoma

Melanoma is a cancer of the skin; in more detail, a cancer that originated in melanocytes. It primarily arises in the skin but can also arise from melanocytes in the CNS, gastrointestinal mucosa, genitourinary mucosa, and the uvea (Tas et al., 2011). Most skin cancers originate in keratinocytes and develop to basal cell carcinoma (BCC) or squamous cell carcinoma (SCC), with BCC accounting for over 4 million and SCC for over 1 million new diagnoses per year in the US alone (Siegel et al., 2022). Melanoma accounted for only 1-2% of all skin cancer cases with over 99000 cases in the US in 2022 (Siegel et al., 2022). It is the 5th most common cancer in males and females for 2022 estimates. Even though it is only responsible for 1-2% of skin cancer diagnoses a year, it is responsible for almost all deaths due to skin cancer with nearly 10000 deaths in the US alone per year. Risk factors for melanoma include fair skin (with Caucasians being 20x more likely to develop melanoma than African Americans), exposure to UV radiation (via sun exposure or tanning beds) and a history of blistering sunburns in childhood as well as having more than 50 small nevi or having dysplastic or atypical nevi (Gandini et al., 2005).

While melanoma in stage I and II (localized) has a 5-year survival rate of 99%, this steadily decreases to 30% for stage IV (distant stage, metastatic melanoma). The primary line of treatment is surgical resection. For patients with advanced disease, a systemic therapy will be used. Until 2011 this was generally chemotherapy alone. More recently, breakthrough discoveries, using immunotherapy have translated into successful (at times curative) therapeutic approaches for some patients with unresectable stage III or metastatic melanoma (stage IV). Mainly, immunotherapy has shown great success in increasing survival over the last decade from less than 15% to now 30% for metastatic melanoma (Albittar et al., 2020). The first immunotherapy approved for metastatic melanoma was

high-dose IL-2. An overall response rate of 16% was achieved in a pooled analysis of 270 patients that received high-dose IL-2 treatment between 1985 and 1993 where patients that did show a complete response at 30 months continued to show no progression (Atkins et al., 1999). While IL-2 showed some success, most patients were not responding to high-dose IL-2; thus additional therapies needed to be developed. One way to augment the high-dose IL-2 therapy was via adoptive cell therapy where tumor-derived T cells are expanded *in vitro* and given back to the patient. This treatment showed an overall response rate of 51% in a small trial with 35 patients, but only 9% showed a complete response (Dudley et al., 2005). Another approach was the use of immune checkpoint inhibitors, like anti-CTLA4 and anti-PD1. Ipilimumab, a fully human IgG1 monoclonal antibody inhibiting CTLA4, was FDA-approved in 2011 for unresectable metastatic melanoma. Ipilimumab led to a significantly longer overall survival (Hodi et al., 2010; Robert et al., 2011). Other successful checkpoint inhibitors in the treatment of metastatic melanoma are Nivolumab, a fully human IgG4 anti-PD1 antibody, and Pembrolizumab, a fully humanized IgG4 anti-PD-1 antibody with some success (Robert, Long, et al., 2015; Robert, Schachter, et al., 2015). Various combinations of these 3 and other similar agents are showing better outcomes than treatment with single agent checkpoint inhibitors.

Another approach is the tumor vaccine approach where tumor associated antigens or mutation-derived antigens are used to target the tumor specifically and induce an adaptive immune response with memory to further increase overall and long-term survival. One such approach was taken by Schwartzentruber et al where a gp100 peptide vaccine was given in conjunction with high-dose IL-2 (Schwartzentruber et al., 2011). The response rate was increased significantly but the overall survival, while improved, was not statistically significant. To potentially improve on these results, checkpoint blockade is currently being tested in different combination treatments with chemotherapy, radiotherapy, vaccines, cytokines, and other checkpoint inhibitors. Other

combinatorial approaches are being tested (Albertini et al., 2016; Hsueh & Morton, 2003; Sahin et al., 2020).

Cancer Immunotherapy

The 4 pillars of cancer treatment

Prior to the rise of immunotherapy, the mainstays of cancer treatment were surgery, radiation, and chemotherapy. While all of these are still widely used and the current standard of care for many cancers, immunotherapy offers hope for less toxic treatments than chemotherapy and radiation alone as well as a higher chance for long term survival without recurrence of the tumor, for certain cancer types. Surgery, chemotherapy, and radiation rely on effectively removing or killing all tumor cells. However, after treatment is completed, the presence of any residual live tumor, means that the non-eradicated tumor is free to regrow. Immunotherapy offers a way to harness the immune system's power of immune recognition and destruction, to potentially enable destruction of all tumor cells, which would thereby prevent regrowth of residual viable tumor cells. Furthermore, chemotherapy is not tumor-specific and will harm other tissues and cells usually leaving patients with quite severe side effects.

Cancer immunotherapy has seen an explosion of interest in the 21st century thanks to important research and discoveries made in the 20th century concerning the immune system and immune cell functions and types. These discoveries helped provide greater understanding of the interactionbetween the immune system and cancerous cells. Our immune system is constantly surveying, recognizing, and killing cells that have cancerous or an otherwise harmful potential. If a cancer develops nonetheless, these cells have already managed to evade the immune system's recognition as potentially harmful. The purpose of immunotherapy is to enable the immune system to recognize and eliminate these cancer cells that had previously evaded immune recognition and

destruction. The process leading to cancer cell escape and tumor formation is described in the "three E" hypothesis.

The "three Es"

The three Es concept is an important hypothesis by Dunn, Old, and Schreiber published in 2004 as "the three Es of cancer immunoediting" consisting of elimination, equilibrium, and escape (Dunn et al., 2004) which predicts biological interactions and immune editing processes between the immune system and evolving cancerous cells.

The elimination phase consists of the classical view of cancer immunosurveillance. In this phase, a cancerous cell is seen by the immune system and eliminated. If this process is executed to completion, the other two phases of equilibrium and escape will not be reached, and the host stays cancer-free. To achieve this, an integrated response requiring both the innate and adaptive immune system is necessary (Janeway, 1989). The first change that the immune system notices is a change in the local stroma surrounding the cancerous growth in the form of one or multiple of the hallmarks of cancer (Hanahan, 2022; Hanahan & Weinberg, 2000). Stromal changes cause the release of proinflammatory molecules and chemokines (sometimes produced by the cancerous cells themselves) which in turn attract innate immune cells like natural killer (NK) cells, macrophages, natural killer T (NKT) cells and γδ T cells (Matzinger, 1994; Wrenshall et al., 1999). These cells recognize certain surface molecules on the cancerous cells and produce interferon gamma (IFN-γ) and proceed to kill these cells via direct immune cell killing mechanisms or indirectly through increasing levels of IFN-γ which then activates a number of IFN-γ-dependent processes like cell cycle inhibition (Bromberg et al., 1996), apoptosis (Kumar et al., 1997), angiostasis (Coughlin et al., 1998; Qin & Blankenstein, 2000), and induction of macrophage tumoricidal activity (Hibbs et al., 1977; Pace et al., 1983) leading to killing of a proportion of cancerous cells/the developing tumor. Debris

from cancerous cells that underwent immunogenic cell death can now be taken up by activated dendritic cells (DCs). The activated and antigen-bearing DCs then proceed to travel to lymph nodes (Sallusto et al., 2000) where they present tumor antigens to naïve CD4 and CD8 T cells and prime and activate these T cells against the presented tumor antigens. These tumor specific CD4 and CD8 T cells can home to the tumor site to then actively kill cancerous cells that express the antigens they were primed with.

This process of elimination is most likely constantly ongoing in the body and a continuous process which is repeated every time antigenically distinct neoplastic or cancerous cells arise. One reason for the ageing population to have a higher cancer prevalence is the ageing immune system which shows decline in its function and therefore also in its cancer surveillance capabilities. Another age associated cause of cancer may involve greater exposure to environmental toxins or DNA damage, leading to greater chances for somatic mutations that can lead to neoplastic transformation of previously healthy cells.

If the immune response in the elimination phase was unable to eliminate all neoplastic/cancerous cells, the equilibrium phase arises.

The equilibrium phase is comprised of a dynamic equilibrium between the host immune system and cancerous cells; this results in a tumor that made it through the elimination phase without elimination. During this time the host's immune system is successfully able to control tumor progression, but unable to fully eradicate it. During this phase, many of the original tumor cells are destroyed, but different variant subpopulations of the original neoplastic clone arise until some can evade detection and eradication by the immune system. This equilibrium phase can be short, but can often have the longest duration of the three phases. It has been shown that there can be a 20-year gap between exposure to a carcinogen and clinical presentation of a tumor (Loeb et al., 2003).

In the last phase, referred to as the Escape phase, tumor cells that escape from the growth inhibition of the equilibrium phase are selected for their ability to continue growing. These "escapees" can grow within the host, despite an intact immune system, without detection. This progressive tumor growth then leads to formation of a large tumor mass that is symptomatic and clinically detectable. Tumors either directly or indirectly obstruct the anti-tumor immune response. This can be through immunosuppressive cytokines, mechanisms involving T regulatory (Treg) cells, altered gene and protein expression on the tumor itself, such as loss of major histocompatibility complex (MHC) molecules (Marincola et al., 2000) and natural killer group 2 member D (NKG2D) (Groh et al., 2002) components, defects in signaling pathways, or anti-apoptotic signal expression (Catlett-Falcone et al., 1999; Takeda et al., 2002), or via development of an insensitivity to IFNγ (Kaplan et al., 1998).

The cytokine IL-2 as an immunotherapeutic agent

IL-2 was first identified and described as T cell growth factor in 1976. It is a proinflammatory cytokine mainly produced by CD4 T cells but also secreted by CD8 T cells, NK cells and dendritic cells (Leonard, 2001; Paliard et al., 1988; Rosenberg, 2014; Yui et al., 2004) and consists of a small 15.5kDa four alpha-helical bundle (Jiang et al., 2016). It plays a major role in the initial immune response as well as the development of a memory response (Boyman & Sprent, 2012). IL-2 has been used clinically in metastatic melanoma and metastatic renal cell carcinoma due to its potential to activate a systemic anti-tumor immune response resulting in immune mediated tumor shrinkage at multiple sites of metastatic disease (Atkins et al., 1999; Klapper et al., 2008; Rosenberg, 2014). It has, in rare cases, been shown to maintain a durable anti-tumor immune response for decades (Rosenberg, 2014).

IL-2 is used for its stimulating capabilities to increase an existing anti-tumor immune response and overcome a threshold for therapeutic efficacy. It is also used to proliferate T cells ex vivo and keep them alive during infusions back into patients (Rosenberg, 2014). However, response rates to IL-2 monotherapy in metastatic melanoma and metastatic renal cell carcinoma are limited (~16%) (Atkins et al., 1999; Davar et al., 2017; Klapper et al., 2008). One major issue of IL-2 is its very short *in vivo* half-life of under 1 hour in humans poses one possible reason for its low *in vivo* efficacy. Ways to circumvent this issue are by injecting IL-2 directly into the tumor which ensures a high amount of IL-2 delivered to the target site (Vaage, 1987) or the use of an immunocytokine (IC) which significantly increases the half-life, and the localization to the tumor, of the IL-2 molecule. There is currently more ongoing research on how to increase the short half-life of IL-2 that could also augment its efficacy while potentially decreasing some of its side effects.

Intratumoral injections allowed for lower doses of IL-2 to be given to the patients and overall, while still enabling higher doses to reach the target site. This was tested in a phase II clinical trial with patients with stage III or IV melanoma where IL-2 was injected directly into the skin metastases. This treatment resulted in complete responses in 78% of treated lesions while side effects were kept minimal (Weide et al., 2010).

The use of a tumor-targeted immunocytokine (IC) is another option of targeted delivery of IL-2 as an immunocytokine offers an increase of the IL-2 half-life as well as potency and efficacy and reduction of side effects (Neri & Sondel, 2016). Immunocytokines are monoclonal antibodies (mAbs) which generally target a tumor antigen and a cytokine or other small molecule attached somewhere at the constant region of the antibody. Due to the specificity of the mAb for the tumor, ICs localize to the tumor microenvironment (TME) and deliver the cytokine or other small molecule directly to the TME. This direct delivery then increases the activation and recruitment of T cells and NK cells in the TME (Neri & Sondel, 2016; Yang et al., 2012). In addition to the delivery of the cytokine, ICs also

have the capability of inducing ADCC or phagocytosis of the tumor cells due to NK cells and macrophage recognition of the tumor-bound antibody and engagement with the Fc portion of the antibody via their Fc receptors. The efficacy of the immunocytokine was also enhanced by local intratumoral injection vs. systemic delivery (Baniel, Sumiec, et al., 2020; Yang et al., 2012). This anti-tumor effect was further enhanced by the addition of radiation therapy (RT) and showed strong synergy (Morris et al., 2016).

In situ cancer vaccines

The first ever used cancer vaccine was Coley's toxin which consisted of killed *Streptococcus pyogenes* and *Serratia marcescens*. Coley successfully used this toxin as a localized treatment of cancer in the late 19th century (Coley, 1893). This toxin, or vaccine, was a powerful inflammatory which primarily functioned via activation of the innate immunity. It lacked capabilities to activate adaptive immunity to specific targets and therefore caused a local immune reaction but failed to initiate antigen-specific memory (Guo et al., 2013). Another kind of cancer vaccine is prophylactic vaccination. One example of this type of vaccine is the HepB vaccine targeting Hepatitis B and is about 70% effective in the prevention of development of hepatocellular carcinoma (Chang & Chen, 2015). Another vaccine in this class is the HPV (human papilloma virus) vaccine which has been largely successful in preventing HPV-linked cancer development of head and neck, cervical, penile, vulvar and anal cancers (De Vincenzo et al., 2014). These vaccines function by activating lifelong adaptive immune responses against specific virus peptides which allows early clearance of a virus infection (De Vincenzo et al., 2014; Yang et al., 2017). However, these mostly work in preventing establishment of disease, not clearing an already existing viral infection or an already existing cancer (Yang et al., 2017).

Generating a cancer vaccine against an already existing cancer faces the same difficulties and has been shown to be largely ineffective even if a good target is identified and present, either as a neoantigen (mutated versions of proteins that are only present on tumor cells), overexpression of certain proteins, expression of embryonal proteins or other proteins otherwise unique to the tumor (Albittar et al., 2020; Durgeau et al., 2018).

However, recent advances in cancer immunotherapy have shown promising results using an *in situ* vaccine rather than a vaccine against one specific cancer antigen. For an *in situ* vaccination approach, the cancer vaccine is generated *in vivo* rather than against a specific previously defined antigen. Generally, the *in situ* vaccine consists of adaptive immunity stimulating components delivered directly to a site of cancer; this then uses the tumor-associated antigens on that cancer that are being released via immunogenic tumor cell death (Hammerich et al., 2015) which can, for example, be caused by radiation. These tumor-associated antigens are then taken up by antigen presenting cells (APCs) and presented to T and B cells. One way to enhance this process is by giving other immune modulators such as IL-2, a monoclonal antibody or immune checkpoint engagers like anti-CTLA4, anti-PD1, or anti-PDL1. Once a potent adaptive, T cell-mediated anti-tumor response is generated, it should then be possible to result in systemic anti-tumor immunity as fully activated effector T cells can circulate to all sites in vivo, no longer require costimulatory signals to kill their target cells and are also less susceptible to inhibitory signals (Gudmundsdottir et al., 1999; Hammerich et al., 2015; London et al., 2000; Suresh et al., 2001).

GD2: a Ganglioside

Gangliosides are carbohydrate-containing sphingolipids (glycosphingolipids) that are composed of a ceramide (generally sphingosine, a long-chain amino alcohol, attached by an amide group to a fatty acid core with varying chain lengths from C18 to C20) bound to a sialic acid (an acidic carbohydrate

with a nine-carbon backbone) linked to one or more monosaccharide units (Fleurence et al., 2017; Schnaar et al., 2009; Varki et al., 2022). Naming of gangliosides generally starts with a G for ganglioside, followed by the letter M, D, T, Q, P, H or S that refer to mono-, di-, tri-, quatra-, penta-, hexa- and septa-sialo-gangliosides, based on the number of sialic acid residues, and ending with the numbers 1, 2, or 3 indicating the order of ganglioside migration on thin layer chromatography (Liu et al., 2018; Svennerholm, 1963). Therefore, GD2, is a disialoganglioside (di-sialo-ganglioside) 2, containing two sialic acid residues and is linked to two monosaccharide units. A chemical structure and IUPAC name for GD2 is available here: https://pubchem.ncbi.nlm.nih.gov/compound/53481124#section=3D-Status.

Gangliosides are synthesized intracellularly starting with the formation of a ceramide core, followed by generation of the monosaccharide units and then translocation to the plasma membrane where they are attached via the ceramide moiety (Berois & Osinaga, 2014; Yu et al., 2011). Gangliosides interact with membrane proteins and other membrane lipids to regulate signaling molecule responsiveness and as mediators and modulators of signal transduction (Lopez & Schnaar, 2009). The monosaccharide units of the ganglioside reach the extracellular space where they facilitate cell-cell recognition and adhesion and display antigenic properties (Battula et al., 2012; Krengel & Bousquet, 2014; Lopez & Schnaar, 2009). Many of the gangliosides are expressed across a wide variety of normal human tissues (like GM3, GM2, GM1 and GD1) (Krengel & Bousquet, 2014; Ledeen & Wu, 2018; Ledeen & Yu, 1982; Yu et al., 2011). This wide expression makes most subtypes of Gangliosides unsuitable as targets for immunotherapy. Functions of these gangliosides vary, but mostly involve cell recognition and regulation of membrane-bound signaling proteins like epidermal growth factor receptor (EGFR) and vascular endothelial growth factor receptor (VEGFR) (Brodeur, 2003; Krengel & Bousquet, 2014).

However, GD2 is somewhat different from most other gangliosides, in that it is expressed on very few normal tissues, mainly limited to the central nervous system, peripheral nerve fibers, dermal melanocytes, lymphocytes, and mesenchymal stem cells (Cavdarli et al., 2019; Hersey & Jamal, 1989; Navid et al., 2010; Yoshida et al., 2001). While GD2 seems to play a role in cell signaling, the exact function of GD2 in normal cell physiology is poorly understood (Dobrenkov & Cheung, 2014). Its function in cancer is better understood. GD2 contributes to elevated tumor cell proliferation, migration, motility, adhesion, and invasion as well as assists with resistance to apoptosis (Chung et al., 2009; Esaki et al., 2018; Liu et al., 2014; Shibuya et al., 2012; Yoshida et al., 2002).

Due to the limited areas of expression of GD2 in normal tissue and the strong overexpression in neuroectodermal tumors like neuroblastoma, melanoma, small cell lung cancer, Ewing sarcoma, osteosarcoma, soft tissue sarcoma, glioma, retinoblastoma, and some tumors not of neuroectodermal origin like some breast and bladder cancers, and its cell surface expression, it is a useful target for antibody-based cancer immunotherapy.

Anti-GD2 immunotherapy

Antibody-based anti-GD2 immunotherapy has been shown to be successful in neuroblastoma, a childhood cancer of neuroectodermal origin and with a very high percentage of GD2-positive tumor cells (Manenq et al., 2020; Mujoo et al., 1987; Sariola et al., 1991), when given as a monoclonal antibody in combination with IL-2 and GM-CSF (Yu et al., 2010; Yu et al., 2021) or without GM-CSF and with or without IL-2 (Ladenstein et al., 2018). It showed a dramatic increase in survival and became the standard of care treatment for neuroblastoma. Consequently it is currently under investigation for use on other solid neuroectodermal tumors expressing high levels of GD2 (Anderson et al., 2022; Nazha et al., 2020). It has also been shown that anti-GD2 mAb-based

therapies are associated with a higher long-term survival and disease-free rate (Cheung et al., 1987; Voeller & Sondel, 2019; Yu et al., 2010).

The proposed mechanisms of action for anti-GD2 monoclonal antibodies (mAbs) against GD2-expressing tumor cells is 1) antibody-dependent cell-mediated cytotoxicity (ADCC) via NK cells and granulocytes and phagocytosis by macrophages, 2) complement-dependent cytotoxicity which results in tumor cell lysis, or 3) direct induction of cell death due to specific binding of anti-GD2 mAbs to GD2 and antibody crosslinking (Anderson et al., 2022; Cavdarli et al., 2019; Perez Horta et al., 2016).

The primary mechanism of action is generally ADCC when using an anti-tumor monoclonal antibody. For ADCC, the antibody bound to the tumor engages NK cells and granulocytes via surface Fcγ receptors which triggers the release of granzyme B and perforin which causes pore generation in the target cell membrane which then causes Fc-dependent phagocytosis and lysis of the tumor cells (Gómez Román et al., 2014; Hanton & Pastoret, 1984; Osinska et al., 2014; Siebert et al., 2014). It has been shown that high levels of baseline ADCC correlate with overall survival in cancer patients treated with mAb against their cancer (Lo Nigro et al., 2019).

To increase the ADCC capability, Yu et al added the cytokine IL-2 to the treatment regimen.

This was further observed in mouse model studies utilizing intratumoral injections of anti-GD2 antibody or immunocytokine in combination with IL-2 and/or radiation which showed a strong synergistic effect (Morris et al., 2016; Yang et al., 2012).

Radiation as an immune modulator

Traditionally, radiation has been used as a potent anti-cancer treatment by inducing localized cell death via targeted administration of beams of radiation (usually X-rays or γ-rays). To this day, about half of all cancer patients will undergo radiation therapy (RT) at some point during their treatment

course, either as treatment with curative intent or as a palliative treatment. Radiation offers a possibility of treatment especially for patients with inoperable tumors or tumors where a complete resection was not possible, but is limited by the sensitivity of normal tissue to radiation and its associated toxicities (Barnett et al., 2009; Durante & Loeffler, 2010).

One way to limit toxicity in healthy tissues and enhance the efficacy in tumor tissue is fractionated radiation where the same amount of radiation is administered in many small doses rather than one large dose. A current standard dosing regimen for patients consists of daily fractions of 1.8-2 Gray (Gy) over a course of 6-7 weeks. These small doses result in a cumulative dose of 50-70 Gy of radiation administered to the tumor (Crocenzi et al., 2016; Durante & Loeffler, 2010), while trying to minimize delivery of radiation to nearby normal tissues.

Recent developments in radiation therapy are moving towards hypo-fractionated regimes where radiation is administered in higher doses (usually 8-30Gy) per session of radiation and the sessions are more spread out and have become standard of care especially for NSCLC (Ko et al., 2018; Nedzi, 2008; von Reibnitz et al., 2018; Whelan et al., 2010). One reason to use hypo- rather than hyper-fractionated radiation is that lymphocytes are very sensitive to radiation-induced cell death. With constant low doses of radiation, the immune response to the tumor might be ablated before having a chance of becoming effective (Lee et al., 2009; Trott, 1982). A study in mice showed that one high dose of radiation followed with a series of low doses of radiation was less effective in inducing tumor clearance than a single high dose alone. This correlated further with a decrease in CD8+ T cells which are especially radiosensitive, and with an increase in more radioresistant myeloid-derived suppressor cells (MDSCs) (Filatenkov et al., 2015). Other studies have also shown that the effect of radiotherapy is largely connected to a functioning immune system as radiation efficacy is attenuated in immune incompetent mouse models (Chakravarty et al., 1999; Meng et al., 2005;

North, 1986; Stone et al., 1979) and an accumulation of immune cells can often be seen around regressing tumors following irradiation (Stewart, 1933).

Mechanism of action of radiation-induced tumor regression

Radiation currently used for therapy usually refers to ionizing radiation, which means it can excite electrons in atoms to higher energy states. This causes some electrons to be ejected from the outer shell and will induce nonspecific damage to proteins and DNA. This is then followed by damage repair responses and the potential for cell death. Roughly half of the damage is caused by water radicals [reactive oxygen species (ROS)] and the other half is induced by direct ionization of non-water molecules like protein and DNA (Breimer, 1988; Sevilla et al., 2016). Alterations to the DNA in some cells can cause lasting damage to the irradiated tissue as the DNA changes (breakages, translocations, point mutations and deletions/insertions) can cause direct cell death and can also be passed on through cell divisions to surviving clones of daughter cells, making it the primary cause of radiation induced therapeutic anti-tumor responses (Thompson, 2012).

One major way radiation engages the immune system is through immunogenic cell death (ICD). ICD is defined as "a form of regulated cell death that is sufficient to activate an adaptive immune response in immunocompetent syngeneic hosts" (Galluzzi et al., 2018). The ability of regulated cell death to drive adaptive immunity depends on antigenicity and adjuvanticity. Antigenicity implies that the host has T cell clones that can recognize antigens selectively expressed on infected or malignant cells (Han & Lotze, 2020; Palucka & Coussens, 2016). These might be neoepitopes which are highly immunogenic or immunogenic epitopes through gaps in central tolerance or incomplete peripheral tolerance (Goldszmid et al., 2014). Adjuvanticity refers to the spatiotemporally coordinated release or exposure of danger signals [such as DAMPS (damage-associated molecular patterns) and cytokines] necessary for recruitment, maturation and phagocytotic activity of antigen

presenting cells (APCs) (Bloy et al., 2017; Garg et al., 2015). These APCs then migrate to lymph nodes and prime a cytotoxic T lymphocyte (CTL) dependent immune response against the MHC-presented engulfed antigenic material of the cancer cell. In addition, interaction with naïve B cells takes place, where B cells start developing antibodies against the presented antigens from the tumor cells through their interaction with antigen presenting cells.

There are three specific molecular signals leading to a perceived "dangerous" cell death by the immune system which aids adjuvanticity. One of those is the translocation of calreticulin to the surface of dying tumor cells which then facilitates uptake of tumor cell material by dendritic cells via scavenger receptors (Obeid et al., 2007). Another is the release of high-mobility group protein B1 (HMGB1) which then binds TLR4 (toll-like receptor 4) which are both directly induced by radiation (Apetoh et al., 2007; Obeid et al., 2007). The third signal is the active release of ATP of cells that are committed to undergo apoptotic cell death. This then activates the NLRP3 (NOD-like receptor family, pyrin domain containing-3n protein) inflammasome (NLRP3-dependent caspase-1 activation complex) on dendritic cells which results in antigen presentation by this dendritic cell to T cells and B cells (Ghiringhelli et al., 2009). ATP-release is generally driven by autophagy, which is also promoted by ionizing radiation (Rieber & Rieber, 2008; Rodriguez-Rocha et al., 2011). This illustrates that ionizing radiation is well capable of inducing immunogenic cell death and can be used as an *in situ* cancer vaccine (Demaria & Formenti, 2012).

Synergy between radiation and immunocytokine in a B78 melanoma mouse model

Morris and colleagues demonstrated in 2016 that 12 Gy RT synergizes with intratumoral injections of an immunocytokine linking anti-GD2 mAb to IL-2 when treating GD2-positive B78 murine melanoma tumors in C57BL6 mice (Morris et al., 2016). They went on to show that the combination of radiation [given on day 1 via external beam radiation therapy (EBRT)] and the immunocytokine

given intratumorally on days 6 to 10 once daily, activated a potent T cell mediated *in situ* vaccine. This treatment was able to cure 70% of tumor-bearing mice of their moderate size (100-200 mm^3) tumor. 90% of these cured mice were able to demonstrate immunologic memory against the B78 melanoma as shown by successful rejection of a rechallenge with the B78 melanoma tumor around one to three months following primary cure. The authors furthermore were able to show immunologic memory to a related, but phenotypically distinct tumor cell line, B16, which does not express GD2. This immunologic memory to a related tumor lacking GD2 expression shows antigen spread and recognition of other tumor markers present on both cancer cell lines by the mouse's immune system.

Furthermore, the authors showed that the day 6-10 timing of the immunotherapy following radiation was important. Early (days 1-5) or late (days 11-15) administration of the immunotherapy resulted in the loss of the synergistic effect seen when it was given on days 6-10. The timing used here (days 6-10) correlated with the largest expression changes observed in B78 melanoma after radiation with 12 Gy EBRT (Werner et al., 2017). Furthermore, it was shown that this treatment utilized, at least in part, anti-GD2 antibody mediated recognition to generate this anti-tumor response as mice lacking Fcγ receptors were not able to control tumor growth. Furthermore, Morris et al. found that this treatment response resulted in a memory T cell response as mice depleted of T cells were not able to reject a rechallenge with the same tumor type while mice without T cell depletion rejected 90% of rechallenged B78 tumors.

This treatment regimen of combining 12 Gy EBRT with intratumoral administration of an immunocytokine targeting GD2 coupled to IL-2 was used as the treatment received by all animals further studied in this thesis.

Antibody generation and B cells in the context of a tumor

Usually much fewer B cells and plasma cells are found in tumor infiltrates compared to T cells (Bindea et al., 2013; Chevrier et al., 2017; Jackson et al., 1996; Schoorl et al., 1976), however, several studies have shown that tumor infiltrating B cells can be an important prognostic factor in cancer (Castino et al., 2016; Erdag et al., 2012; Germain et al., 2014; Kroeger et al., 2016; Ladanyi et al., 2011; Lund & Randall, 2010). Just a few plasma cells in the tumor microenvironment can produce a large amount of antibody and cytokines (Dang et al., 2014) which can influence the TME and immune response to the tumor via ADCC and phagocytosis (Gilbert et al., 2011; Kurai et al., 2007), complement activation and enhancing antigen presentation via dendritic cells (Carmi et al., 2015). In addition, it has also been shown that B cells can act as antigen presenting cells (APCs) to CD4 and CD8 T cells and shape antigen-specific immune responses in the tumor microenvironment (Bruno et al., 2017; Rivera et al., 2001; Rossetti et al., 2018). While many anti-cancer immune responses, especially when associated with long-term response and memory, are based on T cells, more recent work has established a crucial interaction between T and B cells either in tertiary lymphoid structures or in clusters of tumor-infiltrating lymphocytes (Sharonov et al., 2020). It has also been hypothesized that B cells contribute to the formation of tumor associated tertiary lymphoid structures in which the maturation and isotype switching of tumor-specific B cells as well as T cells can take place (Pitzalis et al., 2014; Sautes-Fridman et al., 2019; Zhu et al., 2015). When B cells were eliminated, an anti-GITR antibody was not capable of activating an antitumor T cell response in a colorectal and breast carcinoma mouse model (Zhou et al., 2010) and a CD73 inhibitor was not able to activate T helper cells in melanoma (Forte et al., 2012), further highlighting the need for B cells even in a T cell mediated response. Intratumoral B cell to T cell interactions can be explained by the B cell capability to serve as an APC (Rubtsov et al., 2015) and can help maintain additional T cell

expansion intratumorally after initial activation in the lymph nodes via dendritic cells (Bruno et al., 2017; Rossetti et al., 2018).

B cells located in tertiary lymphoid structures have been shown to undergo clonal expansion, somatic hypermutation, isotype switching and tumor-specific antibody production; which further suggests an active participation in the anti-tumor immune response (Bolotin et al., 2017; Cipponi et al., 2012; Coronella et al., 2002; Germain et al., 2014; Kroeger et al., 2016; Mose et al., 2016; Nzula et al., 2003; Sautes-Fridman et al., 2019). High levels of immunoglobulin mRNA in the tumor (measured by RNAseq) have been shown to be associated with increased survival in melanoma (Bolotin et al., 2017; Mose et al., 2016) and lung adenocarcinoma (Isaeva et al., 2019), further hinting at the importance of B cells and linking B cell responses to cancer driver mutations like KRAS and STK11 to B cell responses (Isaeva et al., 2019).

Tumor-specific antibodies

Tumor-specific antibodies are generally found in the tumor microenvironment and in serum and are usually directed against an array of tumor-expressed and self-antigens which can include overexpressed or differentially expressed self-antigens, modified proteins as well as normal intracellular molecules and, importantly, neoantigens created by mutated proteins (Gnjatic et al., 2010; Reuschenbach et al., 2009; Sahin et al., 1995; Stockert et al., 1998). Serum antibodies can be useful and potent prognostic markers as they are easily accessible and have been validated in later stages of disease (Fremd et al., 2016; Garaud et al., 2018; Gnjatic et al., 2010; Hamanaka et al., 2003; Hirasawa et al., 2000; Reuschenbach et al., 2009). Additionally, measurable amounts of antibodies against known tumor antigens can be detected in the serum of up to 50% of patients with breast cancer (Coronella-Wood & Hersh, 2003; Lu et al., 2008). As established above, it has been shown that anti-tumor antibodies can be produced directly in TLSs in the tumor

microenvironment (Bolotin et al., 2017; Cipponi et al., 2012; Kroeger et al., 2016), but they can also be produced by plasma cells in classical niches like the bone marrow or spleen as systemic B cell responses in the form of tumor-specific B cells, plasma blasts and plasma cells have been found in peripheral blood (DeFalco et al., 2018; Gilbert et al., 2011). DeFalco et al. reported that levels of plasma blasts in the blood of cancer patients was higher and antibodies cloned from some of these blasts were able to bind tumor tissue from other patients with the same kind of tumor.

Technologies to identify antibody binding sites

Identification of antibody targets is important in many settings. For example, knowing the exact target of an antibody is important to gain basic insight into antibody specificity and sensitivity as well as to facilitating the identification and design of antigens that can then be used for the generation of therapeutic antibodies and vaccines as well as for the generation of reagents for proteomics research (Irving et al., 2001; Nelson et al., 2010; Uhlen et al., 2010). Many different methods can be utilized to determine the epitopes of antibodies. Some examples are mass spectrometry (Baerga-Ortiz et al., 2002), bacterial display (Rockberg et al., 2008), phage display (Christiansen et al., 2015; Larman et al., 2011; Petersen et al., 1995) as well as peptide arrays (Bruschi et al., 2022; Buus et al., 2012; Forsstrom et al., 2014; Geysen et al., 1984). Each one of these techniques has its own specific benefits and downfalls. Phage display is based on diverse peptide particles that are fused to phage surface proteins. A selection process called bio-panning is generally performed to reduce the number of targets (Pande et al., 2010). One issue with phage and bacterial display is the large selection of target unrelated peptides (TUPs) due to either providing proliferative advantage to the phage or to binding the constant part of the screening platform (Menendez & Scott, 2005; Vodnik et al., 2011). Besides these issues, phage display generally is a laborious process involving many iterations of screening and functional testing of individual phage clones (Pande et al., 2010). The big advantage of mass spectrometry to identify antibody epitopes is that it can identify

discontinuous epitopes. However, the process is very labor intensive and limited in resolution. Since epitopes are identified either via epitope excision or limited proteolysis and comparing the bound vs. unbound protein-antibody complex, mapping is limited to epitope regions with cleavage sites. The identified peptide sequences are generally 30 to 60 residues long while a typical antibody binding epitope usually is much shorter (Baerga-Ortiz et al., 2002). Peptide arrays now offer a larger scale, faster approach to antibody mapping of multiple samples throughout the whole proteome (Bruschi et al., 2022; Forsstrom et al., 2014).

Peptide array technology

Peptide arrays have been developed for the past ~30 years with commercial products being available since the early 2000s. Most arrays typically consist of hundreds to thousands and now moving into millions of distinct peptide sequences and have been used for determining substrate specificities of enzymes, antibody profiling and identifying binding epitopes as well as ligands that mediate cell adhesion and studying ligand-receptor interactions.

Arrays started out in the early 1990s with oligonucleotide arrays developed by Patrick Brown, Stephen Fodor and Edwin Southern (Fodor et al., 1991; Maskos & Southern, 1992; Schena et al., 1995) and were able to profile roughly 1000 genes by 1996 (Schena et al., 1996). Oligonucleotide arrays are nowadays standard tools in many laboratories and often include several millions of oligonucleotides (Davies et al., 2012; Telenti et al., 2016). Thanks to the fast advances in technical development of the arrays, it is now possible to use robust and inexpensive arrays for clinical applications to help understand human disease and viruses (Abegglen et al., 2015; Jabara et al., 2016).

However, the development of peptide arrays has been slower. For scientific discovery, it is vital to also utilize peptide, protein, and carbohydrate arrays as there often is a disconnect between gene

expression, mRNA, and protein activity as proteins are regulated by many additional factors, including post-translational modifications, alternative splicing, allosteric ligands, colocalization and degradation. Another issue with screening peptides and proteins is the increase in numbers of different proteins vs. a rather limited number of genes. Looking at the human genome, there are roughly 20,000 genes (International Human Genome Sequencing, 2004), however, this already increases to ~100,000 transcriptomes (Pan et al., 2008) and further expands to over 1 million proteoforms (Harper & Bennett, 2016; Jensen, 2004) requiring a much more computationally heavy and involved analysis.

Other issues with protein arrays are the generation of proteins, stable immobilization of the protein in a specific orientation, and maintaining protein activity while immobilized while also preventing denaturation. One way around most of these issues is using a peptide array. Short peptides are relatively easy to synthesize, stable and compatible with different immobilization chemistries like biotinylation. One major downfall of peptides however is that they cannot represent the tertiary structure of a protein and therefore cannot mimic for example antibody binding to conformational epitopes. The first such array was performed in 1984 where Geysen et al. successfully mapped a linear epitope on the coat protein of hand foot and mouth disease virus by using a tiled approach of overlapping peptides (Geysen et al., 1984). Since then, Maier et al. were able to adapt the technology to facilitate a high-throughput epitope-mapping screen of a total of 2304 overlapping peptides comprising the vitamin D receptor (Maier et al., 2010). In the next few years peptide array technology rapidly expanded. Just 2 years later, in 2012, Buus et al. published that they successfully used an *in situ* synthesis on microarrays to generate 70,000 peptides of varying length (between 4 and 20 aa long) for epitope mapping of antibodies (Buus et al., 2012). Then in 2014, Forsström et al. were able to utilize a peptide array with 2.1 million overlapping peptides covering the whole human proteome in 12-mer peptides overlapping by six amino acids (Forsstrom et al., 2014) and has since

expanded further to over 7 million peptides (Bruschi et al., 2022). Peptide array technology is now utilized in many ways, these include: characterizing binding epitopes of monoclonal antibodies, quantitatively evaluating protein kinase activity (Houseman et al., 2002; Jarboe et al., 2012), defining autoantibody signatures in different diseases (Sahlstrom et al., 2020; Yan et al., 2019; Zandian et al., 2017) and assessing CD8 T cell epitope recognition via MHC class 1 binding peptides (Haj et al., 2020).

Peptide arrays to characterize antibody epitopes

Antibody epitopes are generally categorized as conformational (or discontinuous), or linear (continuous). Conformational epitopes require an intact protein structure for the antibody to bind and are therefore difficult to characterize concerning exact binding sites and binding sequence. However, linear epitopes can be assessed using peptide arrays. To evaluate antibody epitope sequences using peptide arrays, specific peptides need to be generated and placed in a defined location. Generally, this used to be done using SPOT™ technology which can synthesize a couple thousand peptides in parallel (Frank, 2002; Halperin et al., 2011; Winkler et al., 2011). For specific epitope sequence identification peptides are often used as a mapping tool by either utilizing overlapping peptide sequences or by N or C-terminal truncations, amino-acid scan, or random single, double, or triple substitutions (Buus et al., 2012).

The ultra-high-density peptide microarray is a relatively new way of using peptide arrays (developed within the last 10 years as highlighted above). It utilizes generation of peptides in a combined maskless photolithographic and solid phase peptide synthesis strategy employing a digital mirror device to project light of a specific wavelength (365nm) onto photosensitive 2-(2-nitrophenyl) propyl oxy-carbonyl (NPPOC)-photo-protected amino groups on a glass surface. The NPPOC-photoprotection as well as the standard fluorenylmethoxycarbonyl (Fmoc)-protected amino acids

allow for individually predefined peptides to be generated in each synthesis field (Li et al., 2005; Wang et al., 2022). Side chain protection groups were removed using trifluoroacetic acid (TFA) treatment which leaves peptides attached to the glass slide at the C-terminus without the need for a biotin-streptavidin connection or other linker. This technology can generate on a single slide (a "chip") over 7 million distinct microscale "lawns", each presenting many identical copies of a single peptide, thereby allowing for whole proteome scanning of antibody binding of linear epitopes (Bruschi et al., 2022). Once the peptides are generated on the slide, the array is incubated with the serum or purified antibody sample. Primary sample binding of the antibody in the serum or purified antibody sample is detected utilizing an AlexaFluor ® 647 conjugated secondary antibody against the Immunoglublin of the primary host species. For example, if the sample being evaluated involves serum from an immune mouse, the secondary antibody is an AlexaFluor ® 647 conjugated goat anti-mouse IgG antibody. The fluorescent signal of the secondary antibody is detected via scanning at 635nm at a 2µm resolution using an MS200 microarray scanner and fluorescent intensity values are extracted for each peptide that can then be further analyzed to determine antibody binding to specific peptides. A schematic of the principles of *in situ* peptide array synthesis and antibody binding analysis can be seen in **Figure 1.1** (Forsstrom et al., 2014). Using a tiled approach removes the necessity of exact replicate samples as consecutive peptides will serve as surrogate replicates and help in the identification of exact binding sequences.

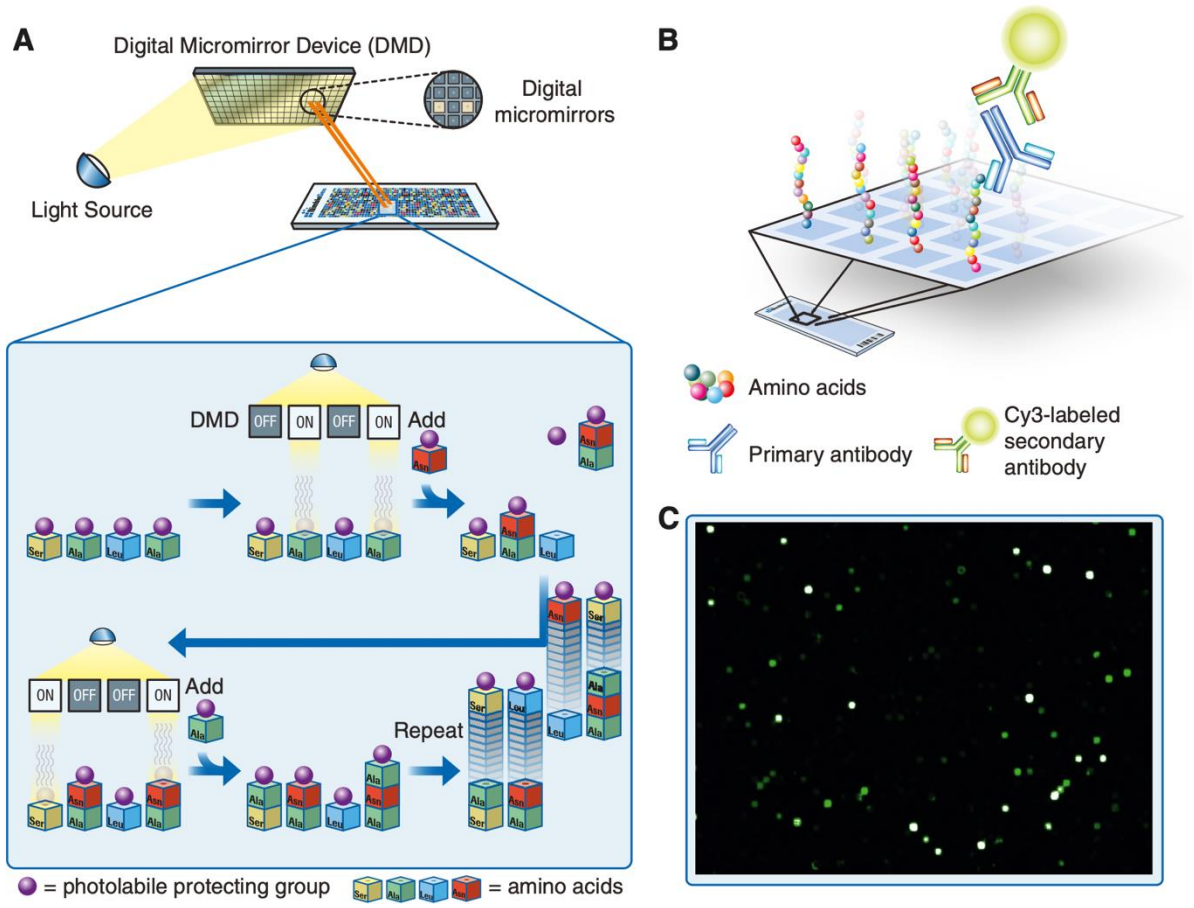


Figure-1.1: The principle of in situ peptide array synthesis and subsequent antibody binding analysis.

A: digital micromirrors individually activate square features on the array by reflecting light on the photo-labile protecting groups of the previously incorporated amino acids. Repeated cycles of selective activation, addition of amino acids, and removal of excess amino acids enables parallel synthesis of peptides with unique sequences. **B**: schematic picture of incubation of the peptide array with the primary antibody and fluorophore-labeled secondary antibody. **C**: a scan image of a part of a planar ultra-dense peptide array in which the bright spots correspond to peptide features bound by antibodies. (Forsstrom et al., 2014)

Thesis structure and hypotheses

Cancer immunotherapy has the potential to engage a patient's own immune system to systemically attack a patient's cancer. Not just at the local site of disease, but also at distant sites and potentially prevent recurrence through immune memory. However, current response rates of patients to immunotherapy remain low. Ways to help the immune system selectively identify and destroy the tumor are still needed. One way to assist with this is identifying new possible tumor targets that differ from normal tissue either via mutations or differential expression patterns. This thesis aims to show a way to help identify new tumor targets that one's own immune system might be able to recognize, but an additional immunostimulatory boost may be needed to facilitate a sufficiently effective immune response.

Hypotheses: This thesis focuses on work I performed as a Ph.D. student examining sera from immunocompetent naïve mice and serum from immunocompetent mice that were implanted with the syngeneic B78 melanoma, cured via treatment with radiotherapy followed by intratumoral hu14.18-IL-2 immunocytokine , and then proven to be immune by their rejection of a rechallenge of the B78 tumor. These immune serum samples, from multiple tumor-immune mice, were shown to be able to bind to the syngeneic murine melanoma, and were then tested against the entire C57BL/6 proteome using the high-density peptide array. Subsequent validation methods were then used, and analyses were performed to evaluate the scope of the antibody repertoire demonstrated, and to evaluate the nature of antigens recognized. We hypothesized that:

A. Mice that reject a rechallenge with the same, or a related, tumor type have developed antibodies that recognize the cancer cells, and that some immunodominant antigens are recognized by antibodies in immune sera from multiple mice cured from the same tumor

B. A whole-proteome peptide array will be a useful tool in identification of antibody targets on cancer cells

C. Antibody targets co-recognized by the antibodies of multiple immune mice may potentially be useful as future therapeutic targets, and possibly as a way to monitor patient's antibody responses as a biomarker of anti-tumor immunity.

Chapter 2 addresses a new method of high-density peptide array tehnology to analyze the whole mouse proteome of normal proteins for antibody targets elicited by a potent anti-tumor *in situ* vaccine immunotherapy regimen in a C57BL6 mouse model of a low immunogenic B78 melanoma. The reproducibility of the methodology, the very large number of peptide targets recognized by the sera of immune mice, and the validation of this recognition using two separate ELISA like assays are presented. The work in chapter 2 is complete; and the manuscript is available as a preprint in BioRxiv (Hoefges et al., 2023), and has been submitted to an immunology journal for formal peer review for subsequent publication.

Chapter 3 explores a phenomenon we have identified within the results of chapter 2 where half of the mice we analyzed using the whole proteome screen exhibited antibodies against an immunodominant motif consisting of four amino acids (SDTG). Most of the work is complete, but some components are still in progress. After the completion of the few remaining ongoing experiments, this manuscript will be deposited in BioRxiv, and submitted for peer review for publication in the summer of 2023.

Chapter 4 summarizes the findings from chapter 2 and 3 and highlights conclusions we can already draw from this work. It discusses ongoing work still underway, but not far enough along to include in this thesis. It then presents future directions on how to further pursue the antibody targets we have already identified using this technology in these mice cured of B78 melanoma. It then briefly asks

additional questions that could now be pursued in further analyzing the immune targets recognized by sera of mice cured of B78 melanoma, or of mice with separate tumors we (and others) are curing using effective immunotherapy. Finally, we present our current ideas on how these methodologic approaches might be used to better understand the antibody targets of sera from human patients cured of their cancers via immunotherapy, and how this information might be used in the future to potentially benefit human cancer patients.

Chapter 2: Antibody landscape of C57BL/6 mice cured of B78 melanoma via immunotherapy

Preface

This chapter includes the extensive work that was done to generate and analyze the peptide array datasets containing proteome-wide data in the form of 16-mer peptides overlapping by 12 or 14aa. The work leading up to the use of a whole proteome peptide array contained the generation of a melanoma and cancer-specific list of proteins to investigate via a targeted peptide array approach with roughly 700 proteins, the determination of the right serum concentration to use on such peptide arrays and the planning and generation of mouse serum samples to be used on these arrays. The serum sample selection process involved following tumor growth and rechallenge for extended periods of time, rechallenging with B78 melanoma as well as the parental B16 melanoma cell line, serum collection at pre-specified time points and flow cytometric analysis of tumor cell binding of the mouse serum. After generation and run of the screening chip (~700 melanoma and general cancer-specific proteins) with different serum dilutions, an anlysis approach needed to be developed. To develop the analysis method, we partnered with a bioinformatics lab with expertise in large datasets to develop a method answering exactly our question. It took some refinement and time to get to the final result and the analysis method developed is highlighted in Appendix A.

The work in this chapter has been submitted for publication to an immunology journal and is currently available as a preprint in BioRxiv (Hoefges et al., 2023).

Abstract

Antibodies can play an important role in innate and adaptive immune responses against cancer, and in preventing infectious disease. Flow cytometry analysis of sera of immune mice that were

previously cured of their melanoma through a combined immunotherapy regimen with long-term memory showed strong antibody-binding against melanoma tumor cell lines. Using a high-density whole-proteome peptide array, we assessed potential protein-targets for antibodies found in immune sera. Sera from 6 of these cured mice were analyzed with this high-density, whole-proteome peptide array to determine specific antibody-binding sites and their linear peptide sequence. We identified thousands of peptides that were targeted by 2 or more of these 6 mice and exhibited strong antibody binding only by immune, not naive sera. Confirmatory studies were done to validate these results using 2 separate ELISA-based systems. To the best of our knowledge, this is the first study of the "immunome" of protein-based epitopes that are recognized by immune sera from mice cured of cancer via immunotherapy.

Introduction

Cancer immunotherapy has revolutionized cancer treatment and has helped thousands of patients (Couzin-Frankel, 2013; Patel & Minn, 2018). However, most patients are still not showing positive responses to current cancer immunotherapy treatment regimens (Chiriva-Internati & Bot, 2015; Patel & Minn, 2018). Using radiation therapy (RT) and intratumoral injections of immunocytokine (IC), we have developed a local *in-situ* vaccine (ISV, RT+IC) regimen capable of curing immunocompetent C57BL/6 mice bearing syngeneic B78 melanoma tumors and resulting in protective immune memory (Morris et al., 2016). Even though B78 is considered a functionally "cold" tumor due to its lack of response to checkpoint inhibitors (Gentles et al., 2015; Morris et al., 2018), our RT+IC regimen can cure many of them. With our *in-situ* vaccine, RT acts to increase the immunogenicity of the tumor by modifying its phenotype and releasing immune stimulatory cytokines. IC is an engineered fusion protein consisting of a tumor-specific monoclonal antibody targeting disialoganglioside (GD2) linked to IL-2. GD2 is a molecule expressed on the surface of most neuroectodermal tumors and some nerve fibers. We also demonstrated that our *in-situ* vaccine

causes epitope spread; 75% of cured mice reject a challenge with B16 melanoma cells (Morris et al., 2016; Yang et al., 2012). B16 melanoma cells do not express the GD2 antigen and are the parental cell line to B78 (Haraguchi et al., 1994; Silagi, 1969; Silagi et al., 1972). We observed strong antibody-binding to B16 cells using serum from cured as compared to naïve mice (Baniel, Heinze, et al., 2020). These antibodies might enable MHC-independent, CD8-T cell independent anti-tumor adaptive immune responses via macrophage-mediated antibody-dependent direct tumor cell killing (Jagodinsky et al., 2022). However, the exact antigen targets of these endogenous antibodies are unknown.

Identifying epitopes on tumor cells that are recognized by antibodies may help identify the immunodominant antigens of cold human tumors, which may help in overcoming immune resistance in these cancers (Sasaki et al., 2020; Shen et al., 2013; Tarp et al., 2007). With the RT+IC regimen, although we are targeting GD2, the memory response does not require GD2 (Morris et al., 2016). Knowledge of these additional antigenic targets may help to identify biomarkers of positive responses and identify potential new therapeutic targets.

In this paper, we utilized a high-density peptide array approach to probe every protein of the mouse proteome, broken down into 16-mer peptides in a 2 or 4 amino acid (aa) tiling approach, to identify antibody targets, using serum from cured mice vs. their matched naïve sample. This high-density peptide array technology has been used for several productive applications recently (Engmark et al., 2016; Haj et al., 2020; Lo et al., 2020; Lyamichev et al., 2017; Mishra et al., 2021; Shen et al., 2019). Using this approach, we identified many tumor antigens expressed by cold murine tumors in individual mice as well as some tumor antigens that are recognized by multiple mice.

Methods

Mice and *in vivo* tumor treatment:

The treatment model used here was previously described in detail (Baniel, Heinze, et al., 2020; Morris et al., 2016; Morris et al., 2018). In brief, B78-D14 (B78) tumor bearing mice were treated when tumors reached ~ 100 mm^3 with a combination of 12 Gy local radiotherapy (RT), followed 5 days later with 5 daily intratumoral (IT) injections of the hu14.18-IL-2 immunocytokine (IC). Mice that were cured were rechallenged after 90 days with an additional injection of the B78 tumor. Mice that rejected the rechallenge were considered immune (**Figure 2.1A**). At indicated timepoints (**Figure 2.1A**), blood via mandibular bleed was collected into BD serum collection tubes and serum was harvested. For select animals a terminal bleed was obtained via cardiac puncture immediately following euthanasia via CO_2 to obtain larger volumes of serum from immune mice. Experiments were performed under an animal protocol approved by the Institutional Animal Care and Use Committee. A list of all serum samples from individual naïve and immune mice, used to generate the data presented in this report is included as **Supplemental Table 2.1**.

Tumor cells:

B78-D14 ["B78", obtained from Ralph Reisfeld (Scripps Research Institute) in 2002] melanoma is a poorly immunogenic cell line derived from B78-H1 cells, which were originally derived from B16 melanoma (Becker et al., 1996; Haraguchi et al., 1994; Silagi, 1969). B78-D14 cells lack melanin, but were transfected with functional GD2/GD3 synthase to express the disialoganglioside GD2 (Becker et al., 1996; Haraguchi et al., 1994), which is overexpressed on the surface of many human tumors including melanoma (Nazha et al., 2020). B16-F10 melanoma was obtained from American Type Culture Collection (ATCC) in 2005. The murine pancreatic ductal adenocarcinoma cell line Panc02 was purchased from ATCC. Panc02, B78 and B16 cells were grown *in vitro* in RPMI-1640 (Mediatech)

supplemented with 10% FBS, 2mMol L-glutamine, 100U/ml penicillin, and 100μg/ml streptomycin. Mycoplasma testing via PCR was routinely performed.

Flow cytometry:

$0.5x10^6$ cells of B16, Panc02 or B78 were used per tube and incubated with 1μl of serum for 45minutes. After incubation, cells were washed with 3ml flow buffer (PBS with 2% FBS) at 300xg and stained with goat anti-mouse IgG-APC (BioLegend, clone Poly4053, catalog # 405308) and rat anti-mouse IgM-PE (ThermoFisher, clone eB121, catalog # 12-5890-82) polyclonal antibodies. Cells were washed again at 300xg for 5min with 3ml flow buffer and resuspended in 50-100μl flow buffer. A drop of DAPI (BioLegend, catalog # 422801) was added to each tube before data was acquired on a ThermoFisher Attune flow cytometer. Data analysis was performed using the software FlowJo version 10.

High-density peptide array:

Design of mouse whole proteome peptide microarray:

The mouse whole proteome peptide microarray was designed based on the protein set downloaded from UniProt in December of 2018 for C57BL/6 mice (The UniProt, 2017). The library was generated *in silico* for synthesis on high-density peptide microarrays (Nimble Therapeutics, Madison WI). The library consisted of overlapping 16-mers representing the entire mouse proteome tiled at every second amino acid for reviewed proteins and every 4 amino acids for most unreviewed proteins. All redundant (non-unique) peptides were only printed once but later computationally mapped back to all UniProt IDs containing this peptide. The individual peptides in the library were randomly assigned to positions on the microarray to minimize the impact of spatial biases.

Peptide array sample binding:

Mouse serum samples were diluted 1:100 in binding buffer (0.01M Tris-Cl, pH 7.4, 1% alkali-soluble casein, 0.05% Tween-20). Diluted sample aliquots were bound to arrays overnight for 16–20 hours at 4°C. After binding, the arrays were washed 3x in wash buffer (1x TBS, 0.05% Tween-20), 10 minutes per wash. Sample binding was detected via goat-anti-mouse IgG Alexa Fluor 647 conjugated polyclonal antibody (Jackson ImmunoResearch, 115-605-071). The secondary antibody was diluted in secondary binding buffer (1x TBS, 1% alkali-soluble casein, 0.05% Tween- 20) and incubated with arrays for 3 hours at room temperature, then washed 3x in wash buffer (10 minutes per wash) and 30 seconds in reagent-grade water. Then the array was washed 2x for 1 minute in 1x TBS and washed once for 30 seconds in reagent-grade water. Fluorescent signal of the secondary antibody was detected by scanning at 635 nm at 2µm resolution and 25% gain, using a micro-array scanner. Data were reported as arbitrary fluorescence units.

Peptide array data processing:

The datasets generated and analyzed for this study can be found on Zenodo under the following DOI: 10.5281/zenodo.7871566.

For each serum sample, the fluorescence intensity data from a single chip, for each unique peptide, was assayed and processed once; then results from identical peptides redundant to multiple proteins (i.e., were present in more than one protein represented) were restored to each protein. Raw fluorescence intensity signals from primary antibodies binding to peptides on the array, and secondary antibodies with a fluorescent tag binding to primary antibodies were reported. The amount of fluorescence signal is influenced by both the titer and affinity of primary antibodies binding to each peptide sequence.

Data analysis workflow/pipeline of whole proteome data:

Detailed bioinformatic/biostatistical data modeling, algorithms, analyses, and graphic presentation methodologies are beyond the scope of this manuscript focusing on the biology and immunology of what is detected using the sera of these immune mice. These issues, and their justification/rationale are presented in detail in a separate manuscript (McIlwain et al., 2023).

JPT peptide array:

Samples were sent to JPT (JPT innovative Peptide Solutions, Berlin, Germany) and a custom designed PepStar Multiwell Peptide Microarray was performed following manufacturers protocol using a manufacturing process based on SPOT synthesis as described previously (Nahtman et al., 2007; Zerweck et al., 2016). Peptides were chosen based on different criteria from the high-density peptide array results, as described in the results section. We included 376 16-mer peptides with a range of signal from the high-density peptide array data and tested those on the same serum samples as well as additional serum samples from immune and naïve mice at a dilution of 1:100. Raw data obtained by JPT for these analyses were sent to us for further analysis and processing. Data were reported as arbitrary fluorescence units.

Peptide ELISA:

For the peptide ELISA, 16 separate JPT BioTides™ Biotinylated Peptides were purchased containing a TTDS-linker and biotinylation at the N-terminus. The peptides were generated using the same SPOT synthesis as the larger peptide array (Nahtman et al., 2007). Peptides were synthesized from C- to N-terminus ensuring that only full-length peptides will have a biotin at the N-terminus. Coating of streptavidin plates was performed per manufacturers instruction with a 250-fold dilution of lyophilized BioTide peptides. ELISA was performed according to JPTs peptide ELISA protocol with the adaptation to a 384 well plate instead of the standard 96 well plate to conserve on serum samples.

Neutravidin coated 384 well plates by ThermoScientific (#15400) were used. Stop solution was added after a 30-minute TMB incubation. Plates were read at regular intervals during TMB substrate incubation (reads at 655 nm) and right after addition of stop solution (reads at 450 nm). Optical density values were used to analyze results.

Choosing of peptides for JPT and ELISA

Peptides for JPT analysis were chosen before the second dataset of whole proteome data using the high-density peptide array was generated and analyzed. 376 peptides were chosen based on different signal strength and reactivity to sample types. In more detail, peptides were chosen based on high signal (>500) in at least one immune sample and low to no signal (<20) in naïve samples. Some peptides were included because they shared, or partially shared, amino acid sequences. Others were chosen because they exhibited no antibody binding in any tested sample or because they had binding in every single sample. We also chose some peptides that exhibited low to medium signal.

For ELISA validation we chose a total of 16 peptides, 2 peptides without any reactivity in any tested sample, 5 peptides based on good correlation between JPT and whole-proteome results and 12 peptides (3 of which were also included in the JPT to whole-proteome correlation category) that showed significant binding in at least 3 immune samples in the moderate (or restrictive) category. We also confirmed that the expected binding sequence within each of these 16 peptides did not have the same, or very similar, sequence to those of any of the other peptides in this group of 16 peptides, to help ensure that each peptide would be identifying relatively distinct antibodies.

Statistical analysis:

Peptide array processing

Data from 13 total unique serum samples were tested in the high-density microarray: 5 from naïve mice, 6 immune samples were obtained from mice following their RT+ IC induced cure from their initial B78 tumor, and then 8-12 days following their rechallenge with another injection of B78 tumor; and two samples (replicates) were obtained from separate mice after a 2nd rechallenge injection of B78 tumor (**Figure 2.1A**). These 13 serum samples were assayed for antibody binding to 6,090,593 unique sequence probes mapped to a total of 8,459,970 unique probe IDs (due to redundancies in tiling across protein sequences and using a mixed tiling of either 2aa or 4aa across each protein), or a total of 53,640 individual proteins. Using spatially corrected processed data from Nimble Therapeutics, the data were log2 transformed, quantile normalized, and further processed using a sliding average mean window across the protein location of +/-8aa.

HERON (Hierarchical antibody binding Epitopes and pROteins from liNear peptides) (McIlwain et al., 2023) was developed and used to determine thresholds for calling antibody binding at the probe, epitope (consecutive probes), and protein level for each sample using meta-analyses methods to summarize binding across subjects in the post-rechallenge condition. Briefly, 1) a global p-value was calculated using a z-test for each probe signal using all sample and probe values, and 2) a differential p-value was calculated between the average of the naïve samples and each individual post-rechallenge (Tumor-free) sample. The global p-value and differential p-value for each post-rechallenge sample were then combined using the Wilkinson's max meta p-value method (Wilkinson, 1951). After correcting for false discoveries using the Benjamini-Hochberg (BH) method (Benjamini & Hochberg, 1995), the individual probes for each post-rechallenge sample are considered bound by antibodies if their false discovery rates (FDR) are below a threshold. Epitope

regions were identified by applying the skater algorithm (AssunÇão et al., 2006) to identify groups of antibody-bound probes (spatially and across subjects), and epitope meta p-values were calculated using the Wilkinson's max method on the 2nd highest probe p-value. Protein p-values were calculated using Wilkinson's min (or Tippett's) method (Tippett, 1931). After correcting the epitope and protein p-values using the BH algorithm, the epitope and protein sample calls were made using an FDR cutoff. To avoid prioritization of peptides that may be due to spurious noise, singleton probe and epitope calls without calls of neighboring probes or if the singleton call was not present in repeat immune samples it was removed. The number of samples that were bound by antibodies for each probe, epitope, and protein were tabulated as K of N statistics (K = # of samples with antibody binding; N = total # of samples).

ICC score

Statistical analysis was conducted using R (v. 4.1.1; R Core Team 2021) and the packages 'lme4' (v. 1.1.27.1; (Bates et al., 2015) and 'specr' (v. 0.2.1; (Masur, 2020) for computing intraclass correlation coefficients (ICC). To analyze the agreement in the high-density whole-proteome, JPT, and ELISA instrument readings among selected peptides, log-transformed readings/intensity was modeled and compared using linear mixed-effects models, in which individual samples and the instruments were modeled, respectively, as random effects, while tumor stage and peptide, when applicable, were modeled as fixed effects. Intra-mouse correlation and intra-instrument correlation were accommodated via random intercepts. The ICC was computed from the instrument random effect to estimate their share of variance in the log-transformed readings. An ICC of 0-0.5 is considered poor reliability; 0.5-0.75 is considered moderate reliability, 0.75-0.9 are considered good reliability; 0.9-1 are considered excellent reliability.

Linear regression/r^2 score

Simple linear regression was performed using GraphPad Prism (Version 9.5.0, 2022) and r^2 values were reported. These values were used to describe how predictive the high-density peptide data for the same sample is of the JPT peptide data. The closer the r^2 value is to 1, the more predictive the high-density peptide array value is of the JPT value.

Test of proportions

A test of proportions was used to compare the portion of positive reactivity between different peptide groups at a threshold of 2 or higher for OD readings. A p-value of 0.05 was considered statistically significant. The proportion of reactivity in the randomly selected peptides (1 of 200 with OD $\geq$2) were found to be significantly less reactive than the HERON validation set (3+ FDR 0.05, 48 of 240 with OD$\geq$2), respectively.

Hypergeometric testing

The ELISA data replicates were first averaged together. A threshold of $\geq$2 O.D. (optical density) units was used to call positive antibody binding for each ELISA data point. For each peptide, the fraction of peptides with antibody binding was calculated for the immune samples in the original and validated ELISA set and for the naïve samples in the validated set. A peptide was considered validated if 25% or more of the respective samples were found to be positive.

To calculate the likelihood of getting n antibody-bound (positive) peptides out of a random sample of 14 draws, a hypergeometric distribution was used to calculate the probability of getting at least *x* positive hits out of 14 random draws, where the total pool of peptides tested by high-density array

has K positive peptides[1]. As the true fraction of positive peptides (K / 6090593) within the pool of ~6 million possible peptides to test from the high-density array is not known, we simulated the calculated p-value from the hypergeometric using different proposed fractions of the total positive peptides within the whole set of ~6 million peptides tested by Nimble.

Results

Mice elicit a tumor-specific adaptive humoral response to RT+ IC treatment

B78 melanoma bearing mice treated with RT + IC + anti-CTLA4 generated an antibody response to surface proteins on the B78 (or B16) tumor cells that was measurable at day 22 post tumor implantation (Baniel, Heinze, et al., 2020). To further investigate these antibody responses and to ensure that the RT+IC treatment alone (without added anti-CTLA4) can elicit a similar antibody response, we collected serum at multiple times before, during and after successful RT+IC treatment of B78-bearing mice (**Figure 2.1A**). Serum was collected from mice at the following timepoints: before tumor cells were implanted (Naïve); once tumors reached treatment size but prior to treatment (Pre-treatment); within a week of mice completing the RT+IC regimen (Post-treatment); when the tumors were regressing but still present, weeks later after mice were deemed tumor-free and prior to a rechallenge (Tumor-free); and 8-12 days after subcutaneous rechallenge with injection of B78 cells, ~90 days after treatment and >30 days after the mice were tumor free (Immune). At this point, a strong memory response was demonstrated based on the rejection of the

[1] To simplify our explanation, we use the binomial distribution, which assumes the probability is the same for every trial (assumes replacement): if there are 10000 positive peptides within the pool of total possible peptides (6000000), the probability of getting k=8 positive peptides out of n=14 selected peptides (if all peptides are equally likely) is p=((6000000 – 10000)/6000000)^k. The probability of 14-k failures is (1-p)$^{(14-k)}$. However, there are $\binom{14}{k}$ different ways of distributing k=8 successes in a sequence of n=14 trials, so $P(X = 8) = \frac{14!}{8!6!}p^8(1-p)^6$. Finally, we calculate the sum of the probabilities using 8 to 14 peptides to obtain the cumulative distribution function P(X >= 8).

rechallenged B78 tumors. These mice were monitored for an additional 5 weeks to ensure complete tumor clearance of the re-engrafted B78 tumor, proving that these mice were immune.

Using flow cytometry, we tested the serum from each of these timepoints for IgG antibody binding to B78 cells. We observed the presence of endogenous anti-tumor antibodies against B78 in tumor-bearing mice starting at the pre-treatment timepoint, with increasing levels of antibody detected by flow cytometry at all subsequent timepoints (**Figure 2.1B**). To determine the specificity of these anti-tumor antibodies, we also incubated these serum samples with B16 melanoma cells, the parental line to B78 that is GD2 negative as well as a separate syngeneic pancreatic adenocarcinoma cell line, Panc02. Serum antibodies showed recognition of B16 to a very similar degree as to B78 and a lower recognition of Panc02 (**Figure 2.1B**). Recognition of Panc02 cells by these serum samples might reflect some shared surface antigens between B78 and Panc02 cancer cell lines. The data presented in Figure 1 are the summed flow cytometry results for 3 of the 6 mice studied subsequently in the high-density peptide array, described below; individual mice showed slight variations in the strength of the responses to these 3 tumor lines at different timepoints (**Supplemental Figure 2.1**).

Whole proteome peptide array results are reliable and repeatable at high signal levels

To investigate what these antibodies are recognizing on the tumor cells, we used a whole proteome peptide array to profile antibody recognition comparing serum from the naïve vs. the immune timepoints (as shown in **Figure 2.1A**). First, we plotted the signal for each of the 6 mice against all 8.46 x10^6 individual peptides, referred to as probes, as a boxplot for each individual sample (**Figure 2.2A**). The overall appearance of immune and naïve samples is very similar, with most probes giving a signal near the baseline, and a small fraction of probes giving signals 100-1000 fold higher than baseline. Even so, more of the probes (detailed numerically in the next paragraph) have even stronger signals in the immune sera, such that the mean of all immune samples is greater than the mean of all naïve samples (**Figure 2.2A**). Our hypothesis was that specific peptides would show

significantly higher binding in immune samples compared to naïve samples. Overall, since we are measuring antibody responses to all native peptides within the mouse proteome, we did expect to see antibody binding to some of these peptides in naïve as well as immune samples as previously seen by Hulett *et al* in 2018 (Hulett et al., 2018).

Prior to identifying high binding peptides recognized by individual or multiple mice, we evaluated how reproducible the signal strength is using this high-density peptide array system. Serum samples from an individual immune mouse (mouse B2), taken after rejection of rechallenge (the immune timepoint in **Figure 2.1A**) was divided and separate aliquots were analyzed in the same array assay, on independent "chips", each quantifying the binding signal against all 8.46×10^6 16-mer peptides. The paired values for each of these peptides, in the 2 parallel samples are plotted on the X and Y axes in **Figure 2.2B**. We first looked at all peptides with significantly higher binding than the mean overall signal, defined as a signal that is larger than three standard deviations (SDs) above the mean (inclusive) (left panel of **Figure 2.2B, red**). Although difficult to appreciate due to the number of data points overlying each other, there are 8,424,675 black data points (corresponding to both of the values for that peptide being $\leq$3SD from the mean value), and only 35,295 non-black data points in the Inclusive group, indicating that at least one of the data points was >3SD (**Figure 2.2B, red)**. A similar analysis was performed for 2 separate aliquots of immune serum from the same blood sample, but from a separate immune mouse (mouse PD1), that was performed on 2 separate identical chips against all peptides, but the analyses were run on separate days, ~ one year apart (**Supplemental Figure 2.2A**). As for **Figure 2.2B**, there are 8,423,302 black data points, and only 36,668 non-black data points (Suppl. **Figure 2.2A)**. However, when looking at all probes that fit these criteria and plotting replicate sample results against each other (**Figure 2.2B** and Suppl. **Figure 2.2A**), we noted that at 3SDs, we found a number of probes where one of the values was >3SD from the mean but where the replicate sample gave a result that was $\leq$3SD from the mean. In **Figure 2.2B**,

these are the probes shown off the diagonal in light-red or dark-red, while all the probes shown along the diagonal in red correspond to *those where both values* from the two separate "runs" of the same serum sample gave concordant values >3SD from the mean. In **Figure 2.2B**, the number of probes in red, (i.e., seen by both replicates, and designated in the legend box as "InclusiveXY"), corresponds to 65% of the non-black probes with 35% of the non-black probes comprised of the lighter and darker red probes (**Figure 2.2B**). Because one of the values for these lighter and darker red probes was not >3SD from the mean, these values were not consistent or reproducible, therefore less reliable to call as antibody binding hits.

To enhance reliability and reproducibility of results, we increased the signal strength criteria to ≥6SDs above the mean (middle panel of **Figure 2.2B**, moderate, in blue) which included the top ~0.1% of peptides compared to the top 0.4% of peptides at 3SDs in the inclusive category. This moderate category showed a larger concordance of recognition between the two replicates, with 80% of the non-black probes being identified at this level by both samples, shown in blue (6653 probes); only 20% of the probes showed discordant signals (one sample >6SD with the paired sample being ≤6SD) in the lighter (591 probes) and darker blue (1027 probes). The best reproducibility between the 2 paired samples on a peptide level was achieved with the restrictive category which was set at 10 SDs (right panel of **Figure 2.2B**, restrictive, green), which includes only 0.02% of all peptides. At 10SDs, over 85% of probes that are not black (at least one of the 2 values >10SD) were in the green category, with both probes >10SD (1422 probes) while only 15% of probes showed discordant signals (one sample >10SD with the paired sample being ≤10SD), in the lighter (95 probes) or darker (149 probes) green. The specific numbers of probes in each category, for each of the paired serum samples for **Figures 2.2B** and **Supplemental Figure 2.2A** are provided in **Supplemental Table 2.2**.

We developed the HERON algorithm to identify consecutive overlapping, reproducible probes with high signal, and categorized the shared aa sequences represented by those highly recognized probes as epitopes based on specified thresholds (**Figure 2.3A**). The mean signal of an epitope was calculated based on the mean signal of all peptides that comprise the epitope. We again used the categories of inclusive, moderate, and restrictive (based on the single probe calls, but now based on standard deviations as well as false discovery rates (FDRs) to assure that all epitopes with significant signal were counted. Reliability and reproducibility for each of these categories increased significantly by looking at epitopes rather than probes alone and helped eliminate many of the non-reproducible binding events seen only on X or Y axes, but not both (**Figure 2.2C** and **Supplemental Figure 2.2B, Supplemental Table 2.2**). From samples from the same run, the percent of epitopes in the inclusive group of epitopes co-recognized by both samples improved from 65% (for peptides in **Figure 2.2B, Supplemental Table 2.2**) to 72% (for epitopes in **Figure 2.2C, Supplemental Table 2.2**); in the moderate category from 80% to 94% and in the inclusive category from 85% to 98.5% (**Figure 2.2C, Supplemental Table 2.2**). When looking at data from separate runs of the same sample, it improved from 37% on the inclusive probe level to 43% on the epitope level. For the moderate category, it increased from 49% to 63%, and in the restrictive category, it increased from 53% to 73% (**Supplemental Figure 2.2B, Supplemental Table 2.2**).

We further assessed reactivity of the sera to which protein these peptides and epitopes could correspond to based on specific criteria. Proteins recognized by immune sera were defined as proteins containing epitopes recognized by the immune sera, using the same criteria for defining inclusive, moderate and restrictive categories of proteins based on the epitope signal for the strongest epitope signal in that protein being either >3, 6 or 10 x SD significantly greater than the mean at FDR adjusted levels of 0.2, 0.05, and 0.01 respectively, clustering across probe hits using skater (AssunÇão et al., 2006) to find epitopes, and filtering regions that are likely spurious signals

reflecting individual probes with a positive signal, but without a signal for the flanking peptides, and rescoring using the Wilkinson's 2nd highest max and the Wilkinson's min/Tippets (Dewey, 2022; Tippett, 1931; Wilkinson, 1951) for the epitope and protein p-values respectively and using Benjamini Hochberg to calculate adjusted p-values on the probe, epitope, and protein level before making calls at the corresponding FDR levels (Benjamini & Hochberg, 1995)(as shown in **Figure 2.4A**). At the protein level, we again were able to see an increase in reliability of called proteins based on signal strength (**Figure 2.2D**). The stronger a protein was recognized (based on signal strength within each epitope in the protein), the higher the percentage of co-recognition by the 2 replicate serum assays (plotted on the X and Y axes) were found: the inclusive category showing 78%, moderate category showing 95% and restrictive category showing 99% of proteins co-recognized by both replicate samples (**Figure 2.2D, Supplemental Table 2.2)**. When comparing data from the same sample from different runs, the percent of proteins co-recognized by both replicate samples also increased in comparison to probes and epitopes, with 50% in inclusive, 66% in moderate and 75% in restrictive (**Supplemental Figure 2.2C, Supplemental Table 2.2**). Overall, we see increasing levels of reproducibility (namely co-recognition of the same peptides, epitopes or proteins by the 2 replicate samples evaluated on separate chips on the same day, or on different days), when going from the inclusive to the moderate to the exclusive category. Furthermore, we see increasing levels of reproducibility within each of these 3 categories, when going from peptide to epitope to protein.

Some epitopes are identified by multiple mice

Using consecutive peptides that show high fluorescence signals in immune sera but not in naïve sera enabled us to identify binding epitopes as well as which part of the peptides contained the binding sequence. **Figure 2.3A** shows at the top an exemplary 16 aa sequence of the protein Titin (UniProt ID A2ASS6), ranging from aa position 8901 to position 8917. Under it are 8 more consecutive 16-mer

peptides, each shifted 2 positions to the right from the one above it, thus overlapping with it by 14 amino acids. For this section of Titin, no binding was observed to any of these 9 peptides by the 2 naïve sera tested (A3 & B2). However, strong binding was seen, reflected in high signal numbers, by sera from 2 immune mice, PD1 and B2. The center 5 peptides all show strong binding by both of the immune sera, indicating that the shared 8 aa sequence of these 5 peptides, SSDSGEYI, reflects the antibody binding sequence. The shared 8 aa sequence, recognized in these 5 overlapping peptides, is referred to as an epitope. Overall, using data from the high-density peptide array, we were able to identify an average of 6400 epitopes in the inclusive category, 2200 epitopes in the moderate category and just under 500 epitopes in the restrictive category by the immune serum samples from each of the 6 immune mice studied (**Figure 2.3B**). Of the identified epitopes, many were recognized by only one mouse, while some epitopes were recognized by sera from 2 or more mice, with one epitope being recognized by sera from all 6 mice in the inclusive category (**Figure 2.3C**). However, with increasing signal strength requirements, that same epitope was seen by fewer than 6 mice when using the moderate or restrictive category. In the highest binding (restrictive) category, twelve epitopes are each recognized by sera from 4 mice (**Figure 2.3C**), while 2450 of 2644 epitopes are recognized by only a single mouse (different epitopes for different mice). For the moderate category, 11491 of 12327 epitopes are recognized by only a single mouse. In the inclusive category, 46493 of 51664 total epitopes are recognized by only a single mouse. These findings are in line with previous studies looking at protein arrays where the abundant and heterogeneous nature of plasma and serum auto-antibodies, regardless of disease status, was discussed (Ayoglu et al., 2013; Nagele et al., 2013).

Previous reports stated that an average length for a linear B cell epitope is around 5 to 12 amino acids (Buus et al., 2012; Engmark et al., 2016; Kringelum et al., 2013). Consistent with this, over 90% of epitopes within the restrictive category are between 7 and 16 amino acids long (**Figure 2.3D).**

Note that a very small fraction of epitopes is identified with a length of 1-2 aa. These small epitopes may be an artifact of the computer algorithm, and most likely suggest that at least two separate antibodies in an individual mouse's serum are binding to overlapping epitopes in this 1-2 aa region, such that we are actually measuring the overlap of the 2 longer epitopes. However, with the data that we have, it is impossible to determine the start and end for each individual overlapping epitope within the region. In general, we found that epitope length varies slightly across binding strength categories, an increase in shorter epitopes is visible in the lower categories (**Supplemental Figure 2.3A & B**).

A greater fraction of proteins than epitopes are bound by sera from multiple mice

A greater fraction of recognized proteins was bound by sera from multiple mice than were found when evaluating epitopes. This difference in proteins vs. epitopes recognized by multiple mice reflects the different requirements for the determination of recognition of a protein vs. an epitope. For an epitope to be recognized by sera from 2 separate mice, the 2 serum samples need to recognize the same epitope. In contrast, for a protein to be recognized by sera from 2 separate mice, each of the 2 serum samples need to recognize that protein, but not necessarily at the same place on the protein; in other words, if the 2 serum samples recognize distinct epitopes, even at opposite ends of the protein, then these 2 serum samples still recognize that individual protein, as shown schematically in **Figure 2.4A**. For each of the 6 mice tested, an average of 5089 recognized proteins were in the inclusive category as compared to 1963 in the moderate and 447 in the restrictive category (**Figure 2.4B**). However, using sera from multiple mice, 4323 proteins were recognized by at least 3 mice within the inclusive category, but only 136 proteins were found to be recognized by sera from all 6 mice. In the restrictive category, 74 proteins were recognized by sera from at least 3 of 6 mice, and 469 proteins from the moderate category were recognized by at least 3 of 6 mice (**Figure 2.4C**).

To broaden the criteria for recognition by sera from multiple mice, we focused on all proteins that were detected in the restrictive category by at least one mouse (2188 total unique UniProt IDs) and looked at these UniProt IDs to see if they were detected by sera from any of the other 5 mice using the restrictive to moderate (purple) criteria (**Figure 2.4D**) to see how many of these mice would recognize these same proteins when the signal strength requirement was loosened. We were able to detect 2 proteins that were now recognized by all 6 mice in this restrictive to moderate category. Overall, 33% of proteins seen by at least one mouse using the restrictive category were seen by 2 or more mice, and 11.4% were seen by 3 or more mice. A similar analysis is also shown for proteins recognized by at least one mouse in the restrictive category, and by other mice using the inclusive (orange) criteria (**Figure 2.4D**). This showed 66% of proteins seen strongly by at least one mouse are recognized by two or more immune mice, while 40% are recognized by 3 or more mice. These analyses indicate that there are several proteins recognized by more than one mouse, while the strength of the recognition signal of the peptide array system can vary from mouse to mouse.

Separate peptide ELISA techniques validate whole proteome peptide array data

After we established HERON, the method used above for the detailed analyses of peptide array data of the proteome recognized by immune sera from mice (and detailed further in a separate companion bioinformatic manuscript, (McIlwain et al., 2023), we wanted to validate our findings with a separate, independent, antibody detection system for 16-mer aa probes that uses a different technology. For this we used a JPT multi-well peptide array to test 189 16-mer peptides per slide, allowing for testing of a larger number of serum samples. We chose peptides to use in this JPT system based on the results obtained using HERON analysis applied to data from the analyses of the entire proteome, summarized above. We chose to include a small number of peptides that showed no binding in any (naïve and immune) serum samples and chose a larger number of peptides that showed significant level of binding by one or more of the immune serum samples using the data

from the proteome analyses from the 6 immune mice tested. The full panel of peptides selected, and the level of their reactivity with naïve and immune serum samples are presented in **Supplemental Table 2.3**. We used some of the same serum samples that we previously tested on the whole proteome high-density array (**Figure 2.5A**) to test these 189 peptides on the JPT array (**Figure 2.5B**)**.** We show the mean reactivity for these same peptides and these same sera using the whole proteome data and the JPT system data **(Figure 2.5A&B).** Both naïve samples show no binding in either the high signal peptide or no signal peptide groups on the whole proteome peptide array as well as the JPT multi-well peptide array (**Figures 2.5A&B**). Immune serum samples showed very similar trends, with higher mean signals seen for the high signal peptides than for the no signal peptides. The A3 and A4 immune samples have a low mean signal for the high signal peptides in the whole proteome array as well as JPT. Overall, these results show that Nimble peptide array data can be qualitatively reproduced using an independent JPT multi-well peptide array. Note that the peptides in the Nimble system are biotinylated at the opposite end from that for the JPT peptides, and thereby fixed to the plate at opposite ends; this makes the peptide available to the sera in reverse orientation, thereby partially accounting for non-identical recognition patterns for the same peptides in these two systems.

The assessment of responses to some of the individual single peptides tested in both systems, demonstrates a qualitative relationship between the magnitude of responses by individual immune mouse serum samples, when tested on the same peptide in the Nimble and JPT systems (**Figure 2.5C**). We examined the same peptides recognized by the same serum samples as shown in **Figure 2.5C** utilizing a separate peptide ELISA system (**Figure 2.5D).** The ELISA data showed these same serum samples show a qualitatively similar pattern to that seen using the Nimble data for these same peptides. A summary of Nimble to JPT to ELISA comparison for 11 peptides using 4 immune and 2 naïve samples is shown in **Figure 2.5E**. The overall intraclass correlation coefficient (ICC) for

instrument (Nimble, JPT, ELISA), considering peptide, tumor stage, and intra-mouse correlation, was 0.86. At the peptide-level, accounting for naïve vs. immune and intra-mouse correlation, these comparisons show 4 peptides with excellent ICCs (> 0.90: Gria4, Scn4b, Srsy and Vsig2) and 7 with good ICCs (0.75-0.90). None of the tested peptides received a moderate (0.5-0.75) or poor (< 0.5) ICC, thereby demonstrating that these 3 ways of measuring antibody responses are not important sources of variation in the measurement of antibodies to these peptides. Overall, these three assay systems showed similar patterns of response for the peptides we chose to evaluate.

Single peptides follow a similar trend in reactivity as seen with surface staining via flow cytometry

We chose 3 peptides to test all timepoints of serum collection shown in **Figure 2.1** using samples from 2 mice. These peptides were chosen based on high signals for these 3 peptides using most immune samples tested, as well as low signals with most naïve samples tested. **Figure 2.6A** shows how the level of antibody from mouse B2 towards the specific peptide increases with each subsequent serum sample (as detected by peptide ELISA) until reaching a plateau and then remains at that peak level while **Figure 2.6B** shows overall lower levels of antibody (as detected by peptide ELISA) for mouse A3 towards the selected peptides. Furthermore, we were able to observe stable antibody concentrations from post-treatment to tumor-free timepoints followed by an increase in antibody in our immune timepoints. In contrast to the flow data reactivity to B16 cells (**Figure 2.1B**), we were not able to see that the post-treatment timepoint exhibited the highest antibody binding for these three specific peptides.

Validation cohort shows binding to most of the 14 peptides selected for binding in immune samples

Lastly, we hypothesized that peptides recognized by immune (but not naïve) sera by 50% or more of our initial cohort of 6 mice would also be seen by sera from additional similarly treated immune mice (as in **Figure 2.1A**) that were not ever previously tested on any peptide array or ELISA. To test this hypothesis, we selected 14 well-recognized peptides with binding by sera from at least 3 of 6 immune mice in the moderate category (**Figure 2.2B**) as tested in the whole proteome peptide array and selected 2 peptides without any significant binding in any serum samples from the Nimble system. Of the 6.09×10^6 unique peptides tested in the Nimble system for these 6 immune mice, only 316 peptides (0.005%) showed recognition at the moderate level for at least 3 of 6 immune mice. **Figure 2.7A** shows, by heat map, the original whole proteome data for these 14 recognized and 2 non-recognized peptides for 5 of the original 6 mice. We then tested these same 16 peptides via ELISA on the same naïve and immune samples as we had run on the whole-proteome peptide array (in **Figure 2.7A**) and obtained qualitatively comparable results (**Figure 2.7B**). We were able to test the same 5 mice but didn't have enough serum left for all peptides with mice A4 or PD1. In this ELISA 12 of the 14 previously selected peptides show significant recognition by at least 1 immune mouse, and 8 of the 14 peptides are recognized by at least 2 of these 5 mice. We then proceeded to run these same 14 reactive peptides (and 2 non-recognized peptides) on new naïve and immune samples (previously untested by array or ELISA) collected from mice who had the same B78 tumor and received the same RT + IC therapy as our initially treated mice. ELISA results of the 20 new immune and 14 new matched naïve serum samples are shown in **Figure 2.7C**. Overall, we were able to show that ~13 of the 20 new immune mice (65%) have antibodies against at least 1 of the 14 reactive peptides with 10 mice showing reactivity to multiple peptides. All new samples exhibited no antibody binding against the 16-mer peptides from Gria4 or P53, just like none of the original

samples did in **Figures 2.7A&B**. However, 6 peptides showed binding to at least 1 naïve sample with 4 of the 14 naïve samples showing some antibody reactivity against at least 1 peptide (**Figure 2.7C**).

To contrast the large amount of antibody binding observed within this targeted selection of peptides we developed and utilized HERON to choose peptides recognized by at least 3 of the 6 mice included on the whole proteome dataset, we used a random number generator to pick 10 peptides out of the whole proteome array dataset of 6,090,593 unique peptide sequences. The log-transformed fluorescence intensity values associated with these 10 random peptides and the negative control Gria4 peptide, used previously from the whole proteome peptide array, are shown in **Figure 2.8A**. All 10 of these random peptides showed virtually no reactivity with any of the sera from the 6 immune mice tested, except for one peptide that showed low, but detectible, reactivity with the B2 immune sample on the original whole proteome dataset (Whrn). This one somewhat positive reaction out of the 60 possible combinations of 10 random peptides with 6 serum samples in Figure 8A corresponds to 1.7% positive. We used these 10 random peptides to probe the immune serum samples from the same 20 validation set mice utilized in **Figure 2.7C** for antibody binding to any of these randomly selected peptides (**Figure 2.8B**). We observed moderate antibody binding by one of the 20 validation set immune samples (V16) to one of the 10 tested random peptides (Podnl1). No other validation serum samples showed detectible binding to any of these 10 peptides. Thus, of 200 possible combinations of the 20 serum samples with the 10 randomly selected peptides in Figure 8B, only one (0.5%) was positive. Contrasting the relatively absent reactivity of these 20 new validation immune samples to these randomly selected peptides, we now show the relatively strong reactivity of these same 20 validation immune serum samples to the 12 peptides from Figure 7C selected utilizing HERON analyses of the original Nimble data, that were recognized by at least 3 of the original 6 immune mice. These data are shown in **Figure 2.8C**, using a selection of data also shown in **Figure 2.7C**. Unlike the 1 positive reaction out of 200 possible combinations shown in **Figure 2.8B**,

these same 20 immune serum samples now show 20% positive reactions with these 12 HERON selected peptides (48 reactions with an OD reading of >=2 out of 240 possible combinations = 20%). This substantial reactivity of these 20 validation sera to these 12 HERON selected peptides is significantly greater ($p < 0.001$) from the 0.5% reactivity in the randomly selected peptides.

We acknowledge that our sample size of 10 randomly selected peptides in **Figure 2.8B** is a small fraction of the over 6 million peptides present on the array. To approach and analyze the ability of the HERON method to identify peptides from the initial Nimble data with the original 6 mice that will show greater than chance reactivity with a new set of immune serum samples, using a calculation that includes a larger number of randomly selected peptides, we employed a model utilizing the hypergeometric distribution (Supplemental Figure 4). When calculating the probability of having a quarter of the previously untested mice recognize a specific peptide with a high ELISA threshold (minimum O.D. signal of 2) if the peptides would have been chosen at random using a hypergeometric test (P(X >= 8 given 14 draws out of a pool of ~6 million), the chance of having this occur with a separate set of mice is almost zero (**Supplemental Figure 2.4).** For example, if we assume that 1% (60,906) of the peptides from the set of the ~6 million unique peptides are reactive, the probability of randomly choosing 14 peptides from the pool of possible peptides and finding that at least 8 of the peptides that are responsive to 25% or more of the mice in the new validation set is 1.9×10^{-15}. If we are sampling with replacement from the pool of samples, i.e. using a binomial distribution rather than hypergeometric, the probability is still 2.85×10^{-13}. The similar result between the hypergeometric and binomial approach is due to the low likelihood of randomly choosing the same peptide twice amongst the large pool of possible peptides. These analyses indicate that peptides recognized at the moderate level using the Nimble array data for immune sera from 50% of multiple mice are highly likely to be recognized by separate, similarly immunized mice, in a validation set, using ELISA data as a validation system.

Since only 0.45% of peptides tested are recognized at the moderate level by at least one mouse (27639 peptides), and only 0.005% of peptides tested are recognized at the moderate level by 3 or more mice (316 peptides), the fact that 13 of 20 (65%) of independent immune mice from the validation set are recognizing at least one of the 14 peptides (selected from the 0.005% of peptides recognized in the Nimble system by 3 or more mice) by the ELISA system indicates that these peptides co-recognized by multiple mice in the Nimble system are identifying peptides likely to be recognized by independent (validation set) immune mice.

Discussion and conclusions

The aim of this study was to establish a method to utilize a high-density overlapping stacked array of peptides representing the entire C57BL/6 proteome in order to identify the "immunome" of epitopes recognized by antibody induced in mice that received curative immunotherapy associated with complete and durable eradication of B78 melanoma tumors with induction of tumor-specific immune memory. In this work, we demonstrated the utility of high-density peptide microarrays for profiling the antibody repertoire in immune serum samples by using a proteome-scale peptide microarray representing all proteins in the mouse proteome. This enabled a fine-mapping of all regions of linear epitopes recognized by circulating antibodies induced during the growth and subsequent complete rejection of a syngeneic murine melanoma. Although these whole-proteome peptide microarrays contained peptides representing the proteome, this study is not a complete analysis of the antibody detected "immunome". The length of the 16-mer peptides is a limitation, as conformational (or discontinuous) epitopes may remain undetected. Nevertheless, these proteome-scale peptide microarrays, along with the development and use of HERON analytic methods, provided an in-depth snapshot of the information stored in the antibody repertoire of mice immune to B78 melanoma after successful RT+IC immunotherapy.

We were able to achieve improved reliability and reproducibility when considering epitopes rather than single peptide probes (**Figures 2.2B vs. 2.2C**). A couple of factors contribute to this; first, there are many more probes than epitopes in the proteome, giving a larger number of possible mismatches. Second, an individual epitope can be a component of several overlapping probes; our HERON algorithm for detecting epitopes recognized by separate assessments of serum samples, requires a degree of similar recognition of the related epitope containing probes by the 2 samples, but does not require complete identity of probe recognition and signal. This enables higher reproducibility of epitopes recognized with high signals than peptides recognized with high signals when replicate chips are evaluated for separate aliquots of the same immune serum sample (**Figures 2.2B vs. 2.2C**). Somewhat similarly, when evaluating proteins that are recognized, since a single protein might be recognized by different mice at different regions, the number of proteins recognized by 4, 5 or 6 of the 6 immune mice (at inclusive, moderate and restrictive recognition levels) is substantially higher than the number of epitopes mutually recognized by 4, 5 or 6 of the 6 immune mice (comparing **Figures 2.3C vs. 2.4C**).

While we were able to use signal strength as a predictor for peptide binding reliability, it cannot be used as a measure of antibody affinity (Buus *et al.*, 2012). Signal strength in peptide arrays is determined by many factors, including quality of synthesized peptide, variations in peptide solvation, presence, or absence of high-affinity antibodies as well as presence or absence of multiple lower-affinity antibodies towards the peptide. As seen in **Figure 2.3B**, the number of recognized epitopes is similar across all 6 immune mice, while the epitopes recognized by individual mice show a large heterogeneity between mice. This heterogeneity in epitopes recognized is demonstrated by the very large number of epitopes recognized by at least one mouse, compared to the substantially smaller number of epitopes with mutual recognition by any 2 of the 6 immune mice (**Figure 2.3C**) and only a much smaller fraction of epitopes mutually recognized by 3 or more (50%) of the 6

immune mice. A large heterogeneity in antibody repertoire between individuals has been shown before in humans (Ayoglu *et al.*, 2013; Nagele *et al.*, 2013) and was expected due to the stochastic nature of V-D-J recombination leading to the specific binding characteristics of an individual antibody generated by a clonally expanded mature B cell.

Interestingly, when validating just a small cohort of 12 peptides, representing ~2.86% of the peptides examined out of the total of 420 peptides, which were each recognized by at least 3 of our original 6 mice based on the Nimble system data, we were able to show reactivity to at least one of these peptides in 65% of our validation cohort of 20 separate immune mice (**Figure 2.7D**). While we did not achieve the same rate of recognition for each individual peptide, having at least one peptide recognized by some of these additional 20 mice supports the biological relevance of these proteins being antibody targets by multiple mice in our system. This biological importance is further supported by the testing of random peptides with immune serum samples from 20 additional mice (**Figure 2.8**) where 10 randomly selected peptides showed only one of the 20 mice recognized just one of the 10 peptides barely above the threshold of an OD value of 2 (mean value of 2.28), corresponding to 0.05% positive reactions of 200). In contrast when these same 20 validation immune serum samples were used to recognize the 12 HERON-selected peptides that showed reactivity with at least 3 of the original 6 mice in the Nimble data, 48 out of the 240 possible combinations had an OD reading of 2 or higher (20%, $p < 0.001$). This validation indicates that the HERON method of selecting peptides from the Nimble data successfully identifies peptides that are being recognized reproducibly in validation assays at a rate far greater than would be seen merely by chance. More importantly, because the antibody repertoire is determined by stochastic gene rearrangements of V-D-J immunoglobulin gene components, the antibody repertoires of distinct genetically identical mice, should have substantial differences. Thus the ability of the HERON method to identify peptides based on their recognition by an initial set of mice using the Nimble

data and these same peptides are subsequently strongly recognized using an independent ELISA assay, on a separate set of previously untested validation immune serum samples, indicates that the peptides (and epitopes) identified by the HERON-method have immunologic importance for other mice from the same strain immunized to the same B78 tumor using the same immunotherapy regimen.

There are several limitations to the current assay configuration to evaluate peptide binding by serum antibodies using a high-density peptide array technology. The assay is set up to provide end-point binding of a complex mixture of antibodies at a single serum dilution. It is difficult to estimate the absolute binding affinity of each antibody clone in the complex sera from such a mixture model. By using different dilutions of known concentrations of well characterized mAbs known to recognize specific epitopes or peptides on the Nimble array, a "standard curve" could be created enabling one to interpolate signals seen with immune sera to the standard curve with the mAb dilutions, allowing calculation of a "mAb concentration equivalent". We have not pursued this and do not think this has been done yet by others using this technology. By evaluating serial dilutions of multiple different mAbs, at the same concentrations, on this high-density proteome array, one might be able to investigate some general patterns to allow quantitative assessments of binding to elements of the proteome with this technology. Once this knowledge has been acquired, this peptide array might be an optimal way to characterize the binding and specificity/cross-reactivity for new mAbs being developed.

It is also possible that this array misses the antibody target of some clinically important antibody responses. These include antibody reactivity against conformationally determined epitopes that are not generated in relatively small 16-mer peptides. Second, as the peptides used in this array are strictly 16-mer aa sequences, no glycosylation is applied to these peptides. Thus, circulating antibodies that recognize differentially glycosylated peptides would not be detected. Third, some

clinically important antibody targets have no peptide component. A major example is the GD2 disialoganglioside, proven to be a clinically important target on neuroblastoma; this glycolipid has no peptide component and is recognized by the Dinutuximab mAb (Yu et al., 2010), and the hu14.18-IL-2 immunocytokine used to cure mice of B78 melanoma in this study (Baniel, Heinze, et al., 2020; Morris et al., 2016), and also recognized by circulating antibody in patients immunized with a GD2-containing vaccine (Cheung et al., 2021). This peptide array would not be able to identify antibodies that might have been turned on to such non-peptide antigens, even if they were strongly induced in the process of these mice rejecting, and developing, a memory immune response to these B78 tumors. However, the array would be able to detect antibody binding to peptide mimotopes of such antigens as they are cross-reactive with the non-peptide antigens (Bolesta et al., 2005; Horwacik et al., 2015; Wondimu et al., 2008). Finally, some of the antigenic targets on tumors that have been recognized by adaptive immunity are mutation-driven neo-antigens, with aa sequences different from that controlled by the inherited germline genome. As each individual tumor will have its own unique set of neoantigens, the detection of antibodies to these neoantigens using this high-density peptide array technology would require independently created proteome arrays to be established for each individual tumor being evaluated. This would seem currently impractical.

As such, while other epitope discovery methods are superior in probing more limited numbers of targets to define discontinuous or conformational or glycosylated, or non-peptide, or mutated epitopes and immunodominant responses, this technology appears useful in identifying immunoreactive regions within the entire proteome, not previously considered as potential epitopes, on a large scale.

Beyond the possible utility of identifying biomarkers for effective immune responses induced by cancer immunotherapy, we are hopeful that this technology can be used in profiling antibody

responses to many other diseases. This tool was able to detect known and previously unknown protein targets of antibody responses throughout the mouse proteome. This approach could potentially be applied to other cancers to advance diagnostic and cancer vaccine development.

This initial description of the results of our analyses to probe the detection of linear epitopes recognized by sera of mice cured of their B78 tumors, relies on the novel bioinformatic approaches developed to analyze these large data sets, reported separately (McIlwain et al., 2023). This report presents: 1) the immunologic methods used to obtain data and validate it using additional JPT and ELISA systems; 2) the spectrum of peptides, epitopes and proteins recognized; and 3) initial description of what fraction of targets recognized by at least one immune mouse are also recognized by some other mice, despite the stochastic nature of each mouse's individual B-cell repertoires. Important additional analyses are still underway and are beyond the scope of this initial report. These include characterizing which antigens are recognized by these immune sera and determining their relationship to the B78 melanoma tumor that responded to the immunotherapy in the process of turning on these adaptive antibody responses. These ongoing studies also include identifying which of the proteins recognized by these immune antisera are expressed or over-expressed by the tumor cells themselves, and if expressed by the tumor cells, what is their cellular location (membrane, cytoplasmic or nuclear). Furthermore, even though these antibodies were not seen in naïve mice, and were thus induced by bearing the tumor, and responding to the immunotherapy (as demonstrated in **Figure 2.5D**), given the very large number of proteins recognized by these immune sera (~10,000 proteins recognized by at least 1 immune mouse, as shown in **Figure 2.4C**), it seems very unlikely that all of these are selectively expressed by the tumor cells and not normal tissues. As such these antibodies induced in these mice by implanting and successfully treating these tumors may also reflect antibodies that can recognize proteins from normal tissues and may thereby be considered "auto-antibodies". Such auto-antibodies may be the

mechanism behind auto-immune, paraneoplastic, syndromes seen frequently in patients with cancer (Ong et al., 2022; Villagran-Garcia et al., 2023). Finally, even though multiple distinct proteins are recognized by these immune sera, might some of these antibodies be recognizing shared or similar amino acid sequences on these distinct proteins, reflecting possible immune cross-reactivity of similar antibodies to seemingly distinct proteins? These issues are now being pursued and will be presented in a subsequent separate report (Hoefges et al., 2022).

In summary, this work shows that peptide array technology can be used to detect the linear antibody-recognized "immunome" of sera from mice immune to B78 tumors through RT+IC treatment. While we saw a large heterogeneity between individual mice, some proteins were strongly recognized by sera from multiple immune mice and may potentially be of importance in achieving immunity to the cancer, or as a biomarker of a potent adaptive response to the cancer. This same type of workflow could be applied to other types of cancer or diseases as well as to the analyses of patients that have received effective immunotherapy associated with a clear immune mediated anti-tumor response to their cancer to evaluate the equivalent antibody-recognized human tumor "immunome". Some work in this cancer realm, and in analyses of auto-immunity and anti-viral immunity has been reported and is underway (Heffron et al., 2021; Mergaert et al., 2022; Potluri et al., 2020; Zheng et al., 2021).

Acknowledgements and clarification of collaborative efforts from others for the work presented in this chapter

Over the course of my PhD, and specifically this project that started towards the end of 2017, a multitude of people in the Sondel lab, as well as collaborating labs, have contributed to this work. I want to specifically thank Irene Ong and Sean McIlwain from the Ong lab for their expertise in bioinformatics, programming and writing analysis algorithms, their ideas for what we should try next

and how else to look at it and always providing different perspectives. I want to specifically thank Paul Sondel and Amy Erbe for their help with developing this project and assisting me in driving it forward. Besides Paul's help in driving this project, he always knew how to motivate me and keep the enthusiasm and excitement for this project going. A special thanks goes out to McKenzie Heck, who started in the lab as my undergraduate student in 2016, and who helped with mouse serum collection, treatment of mice and record keeping for mouse measurements. I also would like to acknowledge Drew Melby, Angie Xu and Nick Mathers who helped me with analysis of the data as well as cross-checking current analysis methods and evaluating how well the current algorithm captured what we were looking for based on random and specific samples and comparing raw data to analyzed data. All three were exceptional students that I had the privilege of mentoring and working with.

I want to acknowledge the specific contributions to the manuscript associated with this chapter. I designed the experiments, collected the data, designed, and performed the analysis, interpreted the data, generated the figures, and drafted the manuscript for this study. Sean McIlwain designed and performed the analysis, generated figures, and assisted in writing and editing the manuscript. Nicholas Mathers, Angie Xu and Drew Melby assisted in data collection and analysis and editing the manuscript. Kaitlin Tetreault, Trang Le and Kyungmann Kim helped in the analysis of experiments and editing the manuscript. Richard Pinapati, Brad Garcia and Jigar Patel assisted in design and execution of experiments and editing of the manuscript. Mackenzie Heck, Arika Feils and Noah Tsarovsky assisted in collecting the data and editing the manuscript. Jackie Hank and Zach Morris assisted in designing the experiments and editing the manuscript. Amy Erbe assisted in designing the experiments and interpreting the data as well as editing the manuscript. Paul Sondel and Irene Ong assisted in designing the experiments and analysis, interpreting the data, and editing the manuscript.

The manuscript has been been submitted for peer review and publication and is now in the review process, and has also been deposited and is available in BioRxiv and can be found online as:

Hoefges, A., McIlwain, S. J., Erbe, A. K., Mathers, N., Xu, A., Melby, D., Tetreault, K., Le, T., Kim, K., Pinapati, R. S., Garcia, B., Patel, J., Heck, M., Feils, A. S., Tsarovsky, N., Hank, J. A., Morris, Z. S., Ong, I. M., & Sondel, P. M. (2023). Antibody landscape of C57BL/6 mice cured of B78 melanoma via immunotherapy. *bioRxiv*. https://doi.org/10.1101/2023.02.24.529012

Figures

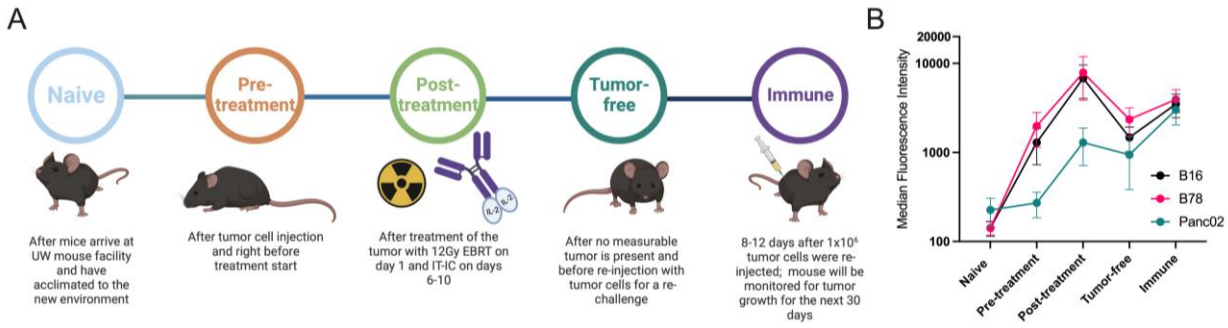


Figure 2.1: Mice develop antibodies against melanoma tumors throughout treatment

A: Timeline of blood serum collection. C57BL/6 purchased from vendors are allowed to acclimate 1-2 weeks prior to Naïve sample collection and B78 tumor implantation. After measurable tumors have established, Pre-treatment samples are collected prior to initiation of radio-immunotherapy [12 Gy external beam radiotherapy (EBRT) and intratumoral hu14.18-IL-2 immunocytokine (IT-IC)]. Following completion of therapy, Post-treatment samples are collected. Tumor-free samples are collected from animals that have no palpable tumors ~30 days following treatment initiation. ~90 days post treatment initiation, these "cured" animals are rechallenged with tumor cells and Immune samples are collected the following week. Schematic created using BioRender.

B: Flow cytometric analysis of serum antibody binding to tumor cells. Murine blood serum was incubated with murine tumor cells prior to staining with fluorescently tagged anti-mouse IgG antibodies and flow cytometric analysis. Median fluorescence intensity values corresponding to the timepoints described in A are shown. Serum samples were tested against B16 melanoma (black), B78 melanoma (pink) and Panc02 pancreatic adenocarcinoma (green) murine tumor cell lines. Error bars show standard error of the mean, n=3 mice for each datapoint.

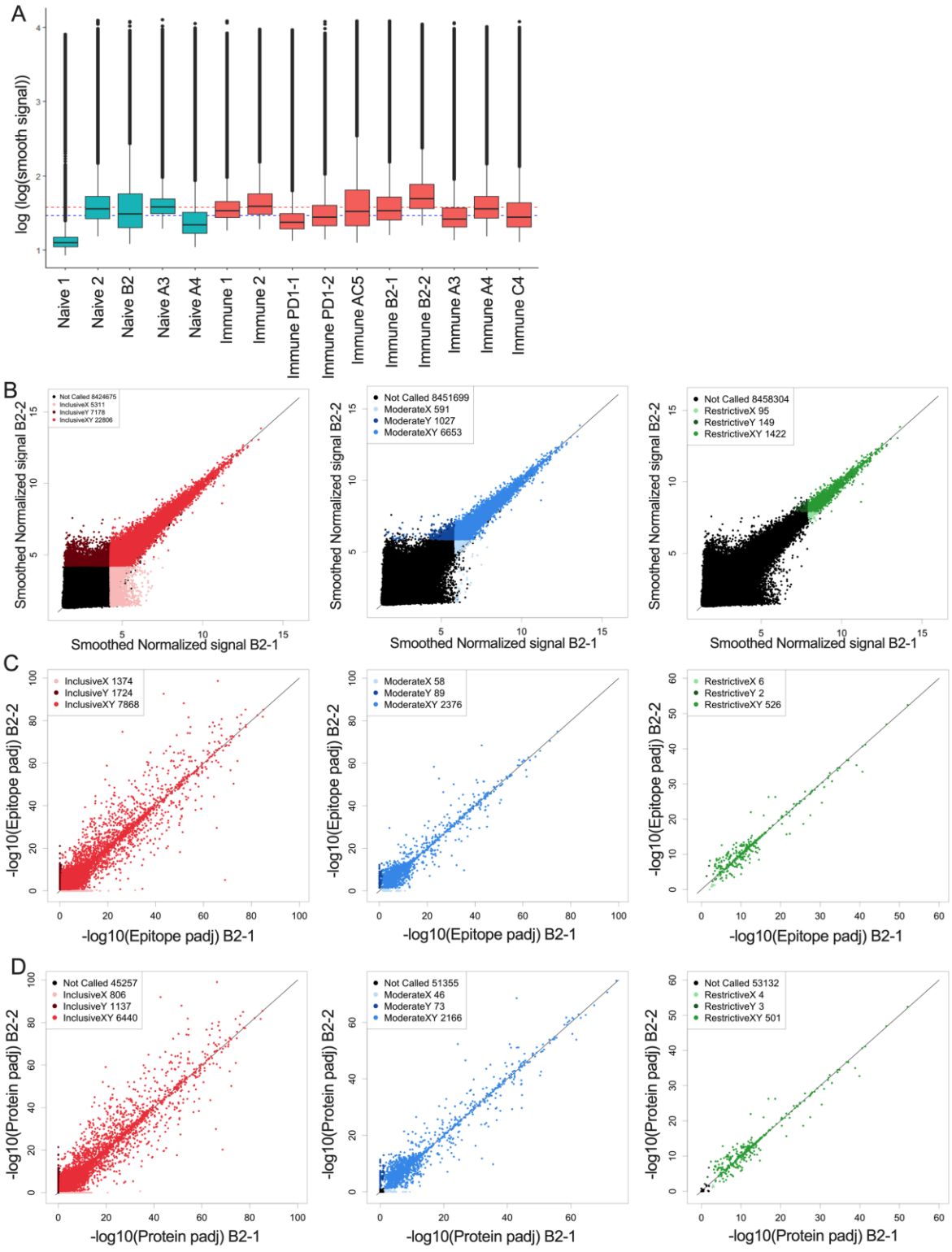


Figure 2.2: Overview of array data and reproducibility and reliability of probe, epitope, and protein calls.

Figure 2.2: Overview of array data and reproducibility and reliability of probe, epitope, and protein calls.

Serum samples from mice described in Figure 1A were run on the Nimble Therapeutics mouse whole proteome peptide microarray.

A: Summary boxplot of signal intensity for sera from each of 6 mice, for which Immune sera were tested (designated: PD1, AC5, B2, A3, A4 and C4). For 3 of these mice, (B2, A3 and A4) naive sera results are also shown. Naive-1, and Naïve-2 are pools of sera from 6 separate naive mice (4 individual mouse samples per pool), and immune-1 and immune-2 are pools of Immune sera from 6 mice (4 individual mouse samples per pool). PD1-1 and PD1-2 are 2 replicate serum samples of cryopreserved Immune sera run independently, one year apart, on separate whole proteome microarray chips; B2-1 and B2-2 similarly are Immune sera from 2 replicate cryopreserved serum samples, run independently on separate whole proteome microarray chips within a day of each other. Data are presented as log-log transformed smoothed fluorescent intensities for all peptides in the array. Median values for all Immune (red) and Naïve samples (blue) shown are represented with dotted horizontal lines.

B: Correlation of fluorescence intensity values from two separate whole-proteome microarray chips run one day apart (B2-1 and B2-2) on the same Immune-serum from one representative mouse. Each dot represents the log transformed processed raw array data for an individual called peptide. These peptides were then separated into 3 categories based on their signal strength: restrictive (highest signal, >10xSD above the mean), moderate (>6xSD above the mean), and inclusive (>3x SD above the mean) based on the statistical significance for each value above the mean signal strength for all peptides. Lighter colored dots represent peptide only called in the B2-1 assay, darker colored dots represent peptides only called in the B2-2 assay, for its respective category (Red: inclusive,

Blue: moderate and green: restrictive). For each graph, the black dots are those peptides that were not called for that graph (at the indicated signal strength level) either in the B2-1 or the B2-2 assay.

C: Scatter plots of epitope-level data based on the peptide data shown in **Figure 2.2B**, again segmented into restrictive, moderate, and inclusive rankings. Epitopes were identified based on overlapping consecutive recognized peptides and values plotted based on the -log10 p-values. Lighter colored dots represent peptides only called for sample B2-1, darker colored dots represent peptides only called for sample B2-2, for its respective category. **D:** Scatter plots of predicted protein-level data based on the peptide and epitope data shown in **Figure 2.2B&C**, segmented into restrictive, moderate, and inclusive rankings. Proteins were identified by combining epitope data and generating a protein p-value and values plotted based on the -log10 p-values. Lighter colored dots represent peptide only called in replicate sample B2-1, darker colored dots represent peptides only called in replicate sample B2-2, for its respective category. For **Figures 2.2 B-D** the numbers in the legend box within each figure indicate the number of dots in each category.

A

SEQ_ID	Probe Sequence	Position	Naïve B2	Naïve A3	PD1 Immune	B2 Immune
A2ASS6	NNVATLVFTQVDSSDS	8901	1.91	2.62	2.31	3.42
A2ASS6	VATLVFTQVDSSDSGE	8903	3.98	2.04	4.21	4.01
A2ASS6	TLVFTQVDSSDSGEYI	8905	1.34	2.42	1882.34	4913.14
A2ASS6	VFTQVDSSDSGEYICR	8907	7.31	2.76	5142.83	8717.05
A2ASS6	TQVDSSDSGEYICRAE	8909	1.96	2.86	2649.60	6446.21
A2ASS6	VDSSDSGEYICRAENS	8911	6.35	8.10	4656.16	6316.83
A2ASS6	SSDSGEYICRAENSVG	8913	1.80	2.41	5615.02	7869.04
A2ASS6	DSGEYICRAENSVGEV	8915	2.64	3.99	3.77	5.26
A2ASS6	GEYICRAENSVGEVSS	8917	7.79	2.47	5.32	2.95

Epitope sequence

antibody binding to peptides

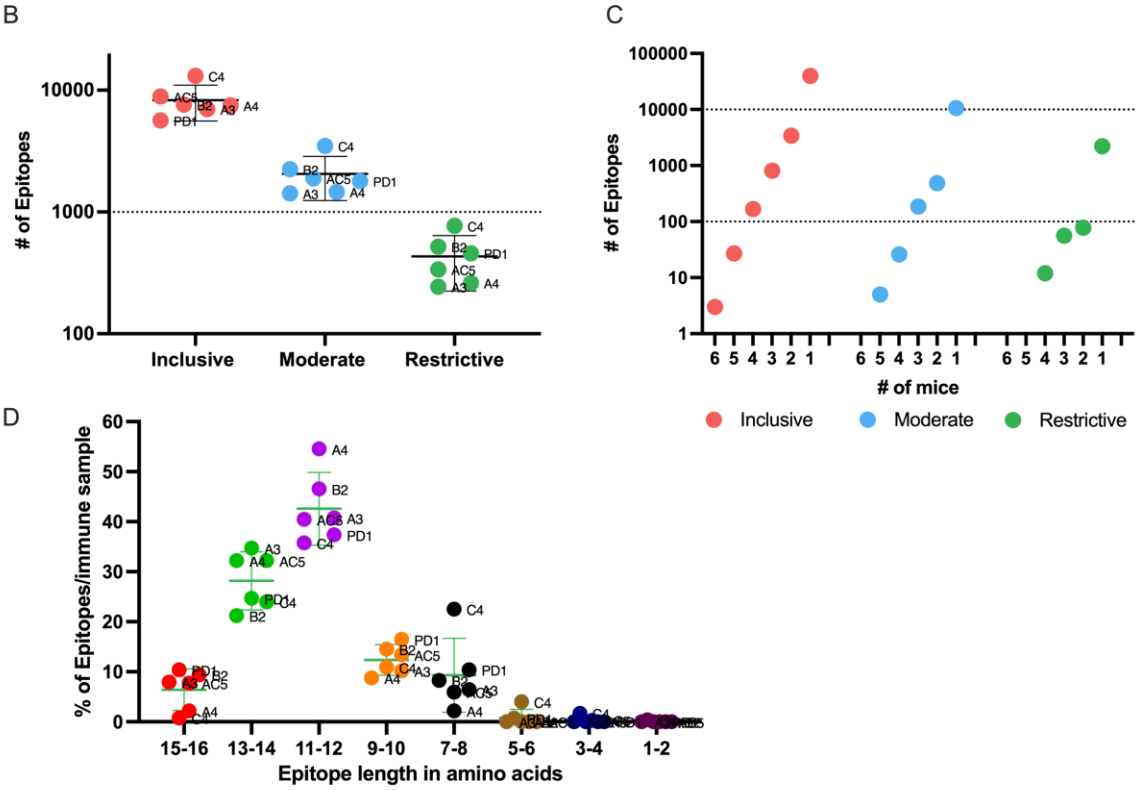


Figure 2.3: Number of epitopes identified and categorized from mouse whole proteome peptide microarray for all Immune samples

A: Example of raw data highlighting a predicted epitope, defined as a clustered and overlapping antibody binding region in the peptide microarray. A section of the titin protein is shown, with 9 stacked 16-mer peptides, each shifted by 2 aa positions, starting at aa position 8901- 8917.

Fluorescence intensity results are shown for each of these 9 16-mer peptides for separate serum samples from 2 naïve mice (naïve B2 and naïve A3) and 2 immune mice (PD1 and B2). Five of the consecutive 16-mers show strong binding by the 2 immune sera, while the other 4 16-mers show very weak binding by all 4 sera shown. The 5 well recognized 16mers each share the 8 sequential aa shown in the green box, indicating a recognized epitope.

B: Number of inclusive, moderate, and restrictive epitopes identified in the Immune samples with significantly higher antibody binding in Immune serum than in Naïve serum samples. Each dot represents the number of epitopes in that category, for each of the 6 separate mice tested. The individual mouse identifications are indicated next to each dot.

C: Number of unique epitopes each recognized by any individual immune mouse, or co-recognized by 2, 3, 4, 5 or 6 Immune mice (of 6 total mice), segmented by cutoff category of inclusive, moderate, and restrictive of the epitopes. Within each category, the single dot plotted above the individual numbers plotted on the X axis indicate the number of epitopes recognized by exactly that number of mice.

D: Categorization of epitopes by peptide length, based on the clustering as in Fig. 3A, using data from the restrictive category. Above each pair of numbers (i.e.: 1-2, 3-4, etc.) on the X axis are 6 colored dots each indicating the number of epitopes of that aa length recognized by each of the 6 mice tested.

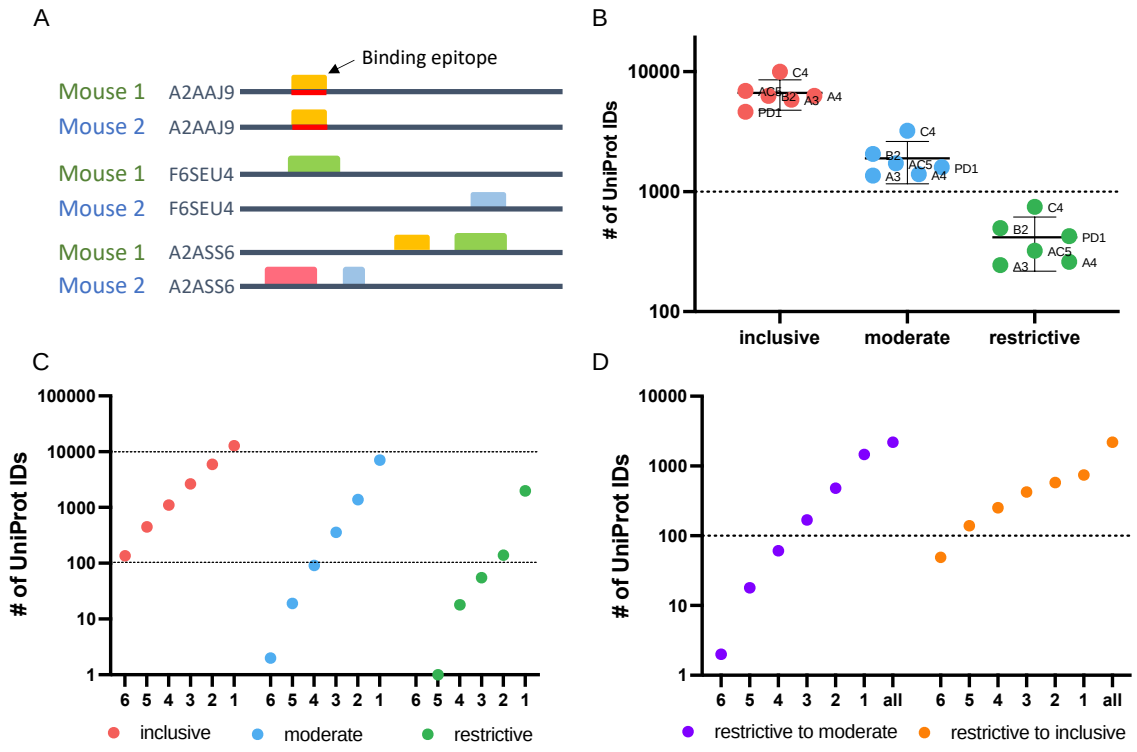


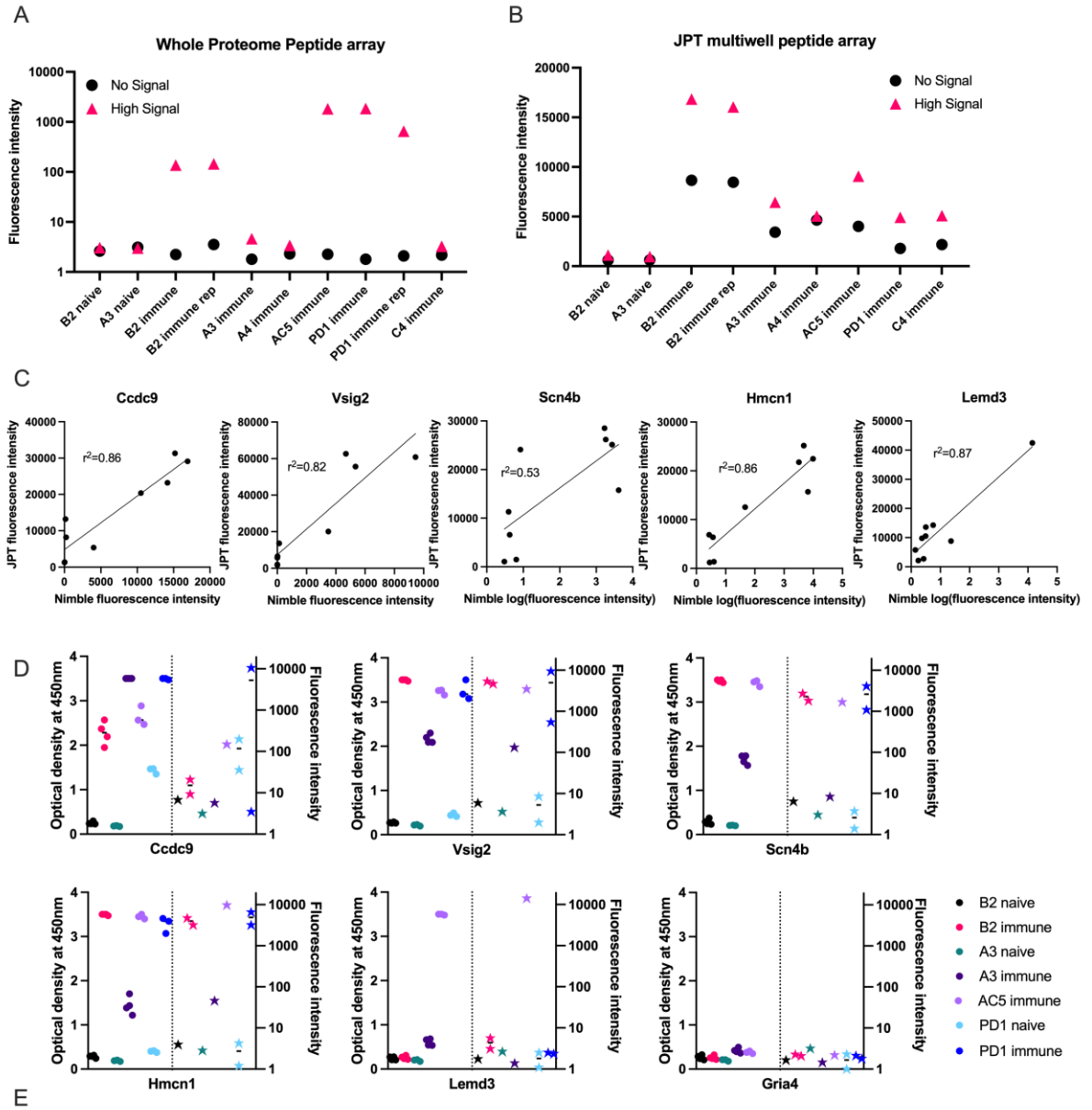
Figure 2.4: Protein level analysis of Nimble Therapeutics mouse whole proteome peptide microarray data and identified epitopes

A: Schematic showing examples of conditions that can lead to identification of one protein recognized by sera from 2 separate mice, including conditions where the same epitope within the protein is not recognized by antibodies from both mice, even though the protein is recognized by sera from both mice.

B: Number of unique UniProt IDs recognized per Immune sample, segmented by the inclusive, moderate, and restrictive categorization of the called protein. Mouse IDs are labeled on each dot to demonstrate similar overall distribution within each category.

C: Number of inclusive, moderate, and restrictive unique proteins recognized by at least 1, 2, 3, 4, 5 or 6 of the 6 mice tested.

D: Number of proteins identified by the given number of mice on the X axis, where at least one mouse recognized the protein within the restrictive category and the other mice identified that same protein at least in the moderate category (purple dots) or at least in the inclusive category (orange dots) in the other samples.



E

	IMMUNE												NAÏVE						ICC
	A3			AC5			B2			PD1			A3			B2			
	Nimble	ELISA	JPT	Nimble	ELISA	JPT	Nimble	ELISA	JPT	Nimble	ELISA	JPT	Nimble	ELISA	JPT	Nimble	ELISA	JPT	
Ccdc9																			0.85
Gria4																			0.99
Hmcn1																			0.89
Junb																			0.82
Lemd3																			0.82
Lzts3																			0.89
Scn4b																			0.91
Srsy																			0.92
Ube2v1																			0.89
Uck2																			0.85
Vsig2																			0.92

Figure 2.5: Comparisons of data from Nimble and JPT systems, for the same peptides and serum samples.

A & B: Comparison of results using the same 10 serum samples tested in both JPT and whole proteome (Nimble) systems, for 11 peptides selected from the whole proteome data to show no significant signal with any serum samples (naïve or immune) vs. 272 peptides showing a high signal with at least one immune serum sample (A&B).

A: Median fluorescence intensity values (from the whole proteome system) for peptides with a high signal (>1000 fluorescence units, >10SD over the mean) in at least 1 immune serum sample that were also tested on the JPT peptide array (pink triangle, high signal, 272 peptides) and on 12 peptides with a signal below 10 fluorescence units in all samples (identified based on the immune and naïve samples from the first whole proteome chipset) (black circle, no signal, 11 peptides) are displayed for 10 serum samples tested in the whole proteome system.

B: Median values for the same peptides as shown in A are shown for the same samples (minus the repeat PD1 sample which was only run once on JPT) run on JPT multi-well peptide array.

C: JPT (Y-axis) vs. Nimble (whole proteome, X-axis) fluorescence signal data-comparison plots for 5 exemplary peptides.

D: Whole proteome data to ELISA comparison plots for 6 representative peptides (5 of which are the same as displayed in **C**) on 7 separate serum samples. For each peptide shown, the left Y axis shows ELISA data as optical density readings, and the right Y axis shows original whole proteome peptide array fluorescence intensity data for the same serum samples on the same peptide. The 7 individual serum samples are displayed in each graph with the same color, ELISA data are shown as circles, whole proteome data are shown as stars. The vertical dotted line separates ELISA from Nimble data.

Multiple datapoints [dots (ELISA) or stars (Nimble, whole proteome) for one sample] show replicates.

E: Heat map of 11 peptides from 11 different proteins with results from 4 immune serum samples and 2 naïve samples across 3 different peptide binding assays. Results from Nimble whole proteome peptide array as well as JPT peptide array and peptide ELISA were performed on the same serum samples and peptides; results are visualized via heatmap. Eight of the peptides shown were selected based on significant binding by at least 50% of immune serum samples in the whole proteome system. ICC: Intraclass correlation coefficient, is the reliability measure of the instrument for that specific peptide accounting for time of treatment and mouse. ICC scores of 0-0.5 show poor reliability,0.5 - 0.75: Moderate reliability, 0.75 -0.9: Good reliability and 0.9-1 excellent reliability.

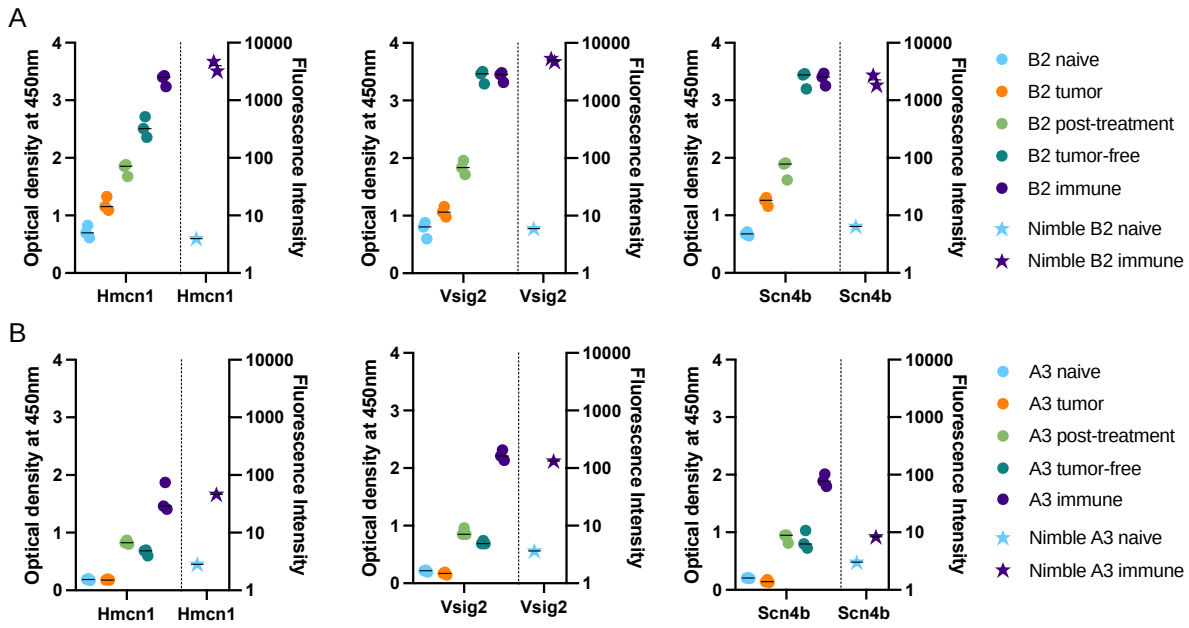


Figure 2.6: Time-course analysis and validation of Nimble peptide array results via peptide ELISA.

A&B: Peptide ELISA of three exemplary peptides (16-mer peptides belonging to Hmcn1, Vsig2 and Scn4b) on all serum collection timepoints shown in **Figure 2.1** on the indicated separate serum samples from 2 immune mice [B2, top row (**A**) and A3, bottom row (**B**)] are shown as optical density on the left Y axis. Right Y axis displays the corresponding fluorescence intensity from the Nimble Peptide array system for the indicated naïve and immune timepoints. Three separate replicate data points are shown for each serum specimen for each peptide in the ELISA (left Y axis), and 2 replicate data points are shown (at times these overlap) for each serum sample on each peptide for the immune Nimble (whole proteome) data (right Y-axis).

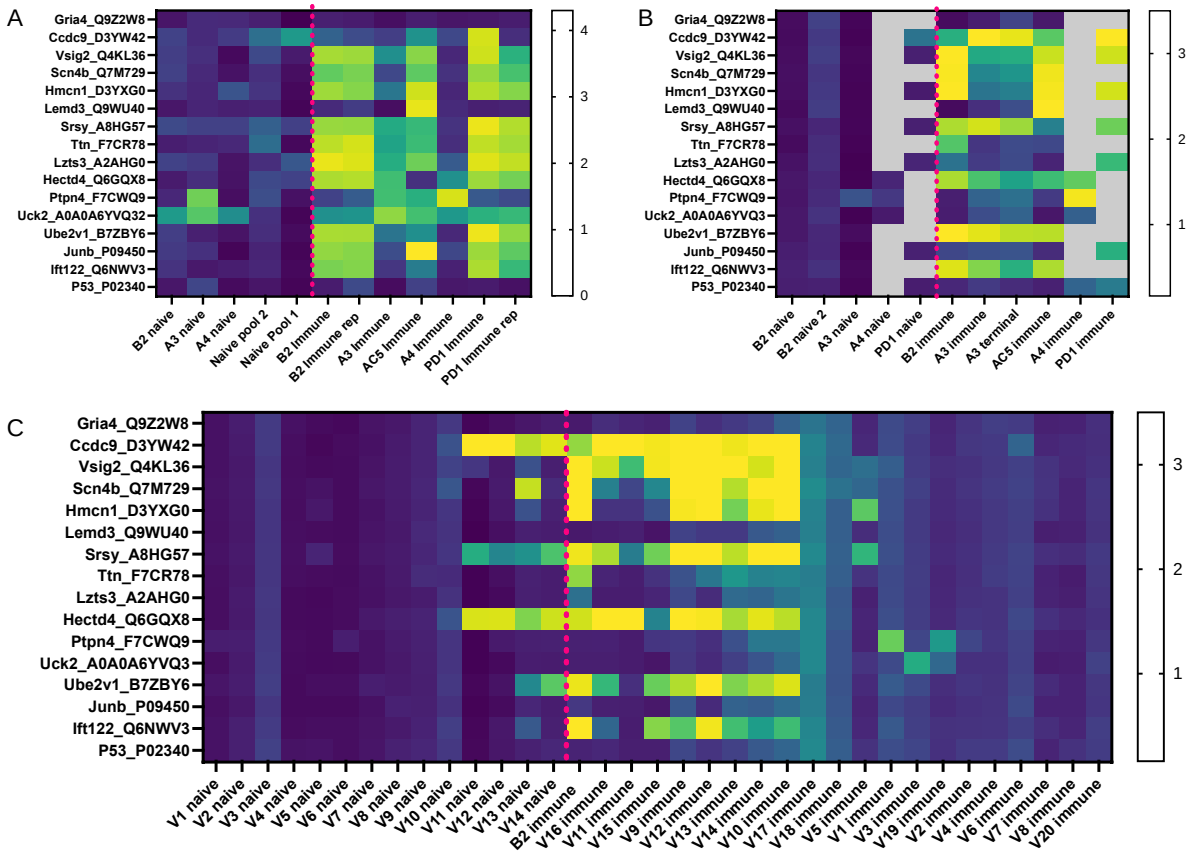


Figure 2.7: Peptides identified by the whole proteome array are also seen by ELISA testing for the same 6 immune mice, and for a separate validation set of 20 separate immune mice.

A: Heatmap of 16 chosen peptides from whole proteome peptide array displaying whole proteome peptide array sample results for 12 serum samples including 2 replicate samples (B2 immune and PD1 immune). Data are shown as log transformed fluorescence intensity. Dotted pink line separates naïve serum samples on left from immune serum samples on right. Fourteen of these 16 Peptides were chosen, based on whole proteome data, demonstrating significant binding by serum samples from at least 3 of the 6 immune samples. Two of these 16 peptides shown were selected because they exhibited no binding by any of the immune or naïve serum samples tested in the whole proteome system (Gria4 & P53, at the top and bottom of the list shown). Data shown are log10 of the fluorescence units of the peptide array signal.

B: Heatmap of ELISA results using the same peptides and serum samples as in **Figure 2.7A**. Grey areas indicate peptides not tested for the 4 indicated serum samples. Data shown are Optical Density (O.D.) values read at 450 nm length on a scale from 0 to 3.5.

C: ELISA data for the same peptides as in A & B but using immune mouse serum samples never tested before from 20 separate mice that have received the same treatment to cure their B78 cancer (together with matched naïve serum samples for 14 of these 20 immune mice). Also included here is a repeat immune serum sample from one of the 6 immune mice used in the original whole proteome samples as an internal control (B2 immune, also shown in whole proteome data in **Figure 2.7A**, and ELISA data for original whole proteome samples in **Figure 2.7B**). Data shown are optical density values read at 450nm length on a scale from 0 to 3.5.

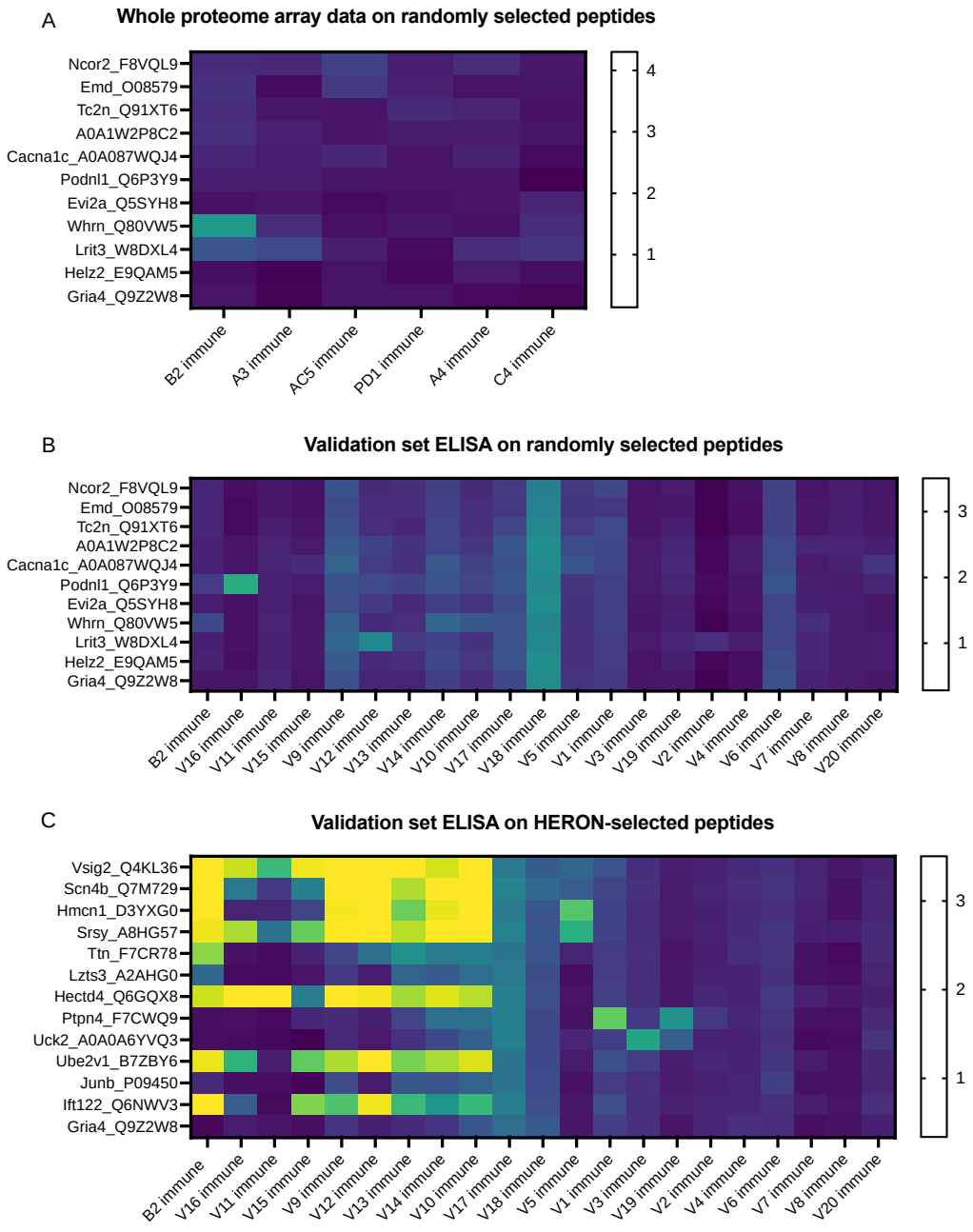


Figure 2.8: Peptides selected at random from whole proteome array

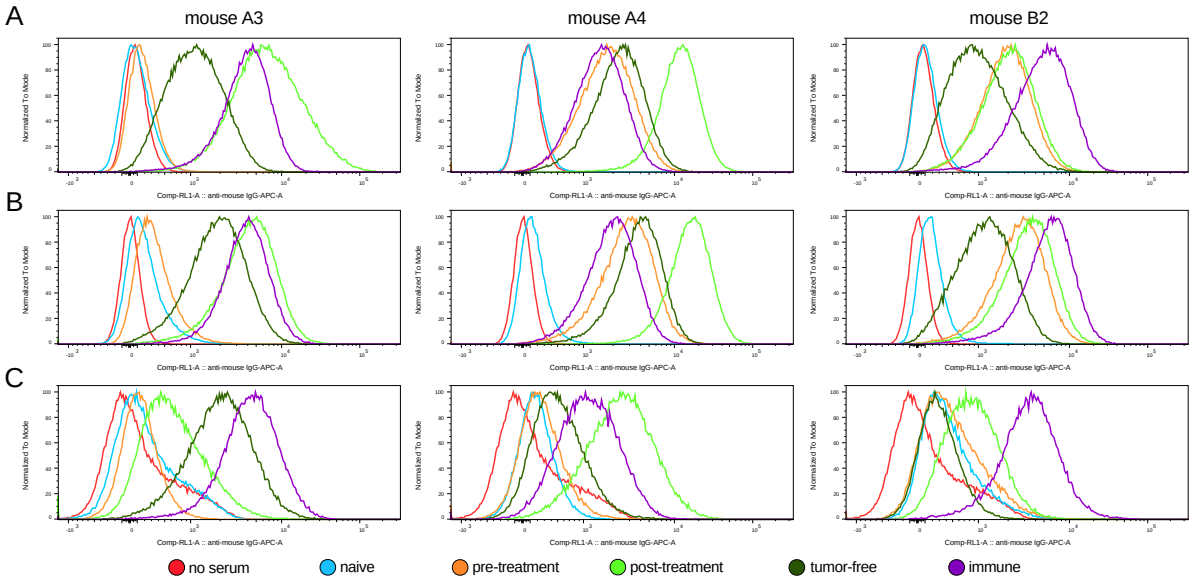
Peptides selected at random from whole proteome array are similarly recognized by ELISA testing for a separate validation set of 20 separate immune mice and show much lower antibody recognition in comparison to HERON-identified peptides present in 50% or more of the original cohort.

A: Heatmap of 11 peptides from the whole proteome dataset displaying whole proteome peptide array sample results for 6 immune serum samples (the same 6 immune serum samples used for figures 2 & 3). Ten of these 11 Peptides were chosen at random utilizing a random number generator out of all probed peptides from the whole proteome array. One peptide is included as a negative control peptide that was intentionally selected as a negative control; we have never observed antibody binding to it in any of our original or validation tested samples (Gria4, at the bottom of the list). Data shown are log10 of the fluorescence units of the peptide array signal.

B: ELISA data for the same peptides as in A but using immune mouse serum samples not tested on the whole proteome array from 20 separate immune mice that have received the same treatment to cure their B78 cancer. These 20 new immune serum samples are identical to the 20 new immune serum samples shown in Figure 7C. Also included here is a repeat immune serum sample from one of the 6 immune mice used in the original whole proteome samples as an internal control (B2 immune, also shown in whole proteome data in **Figure 2.8A**). Data shown are optical density values read at 450nm length on a scale from 0 to 3.5.

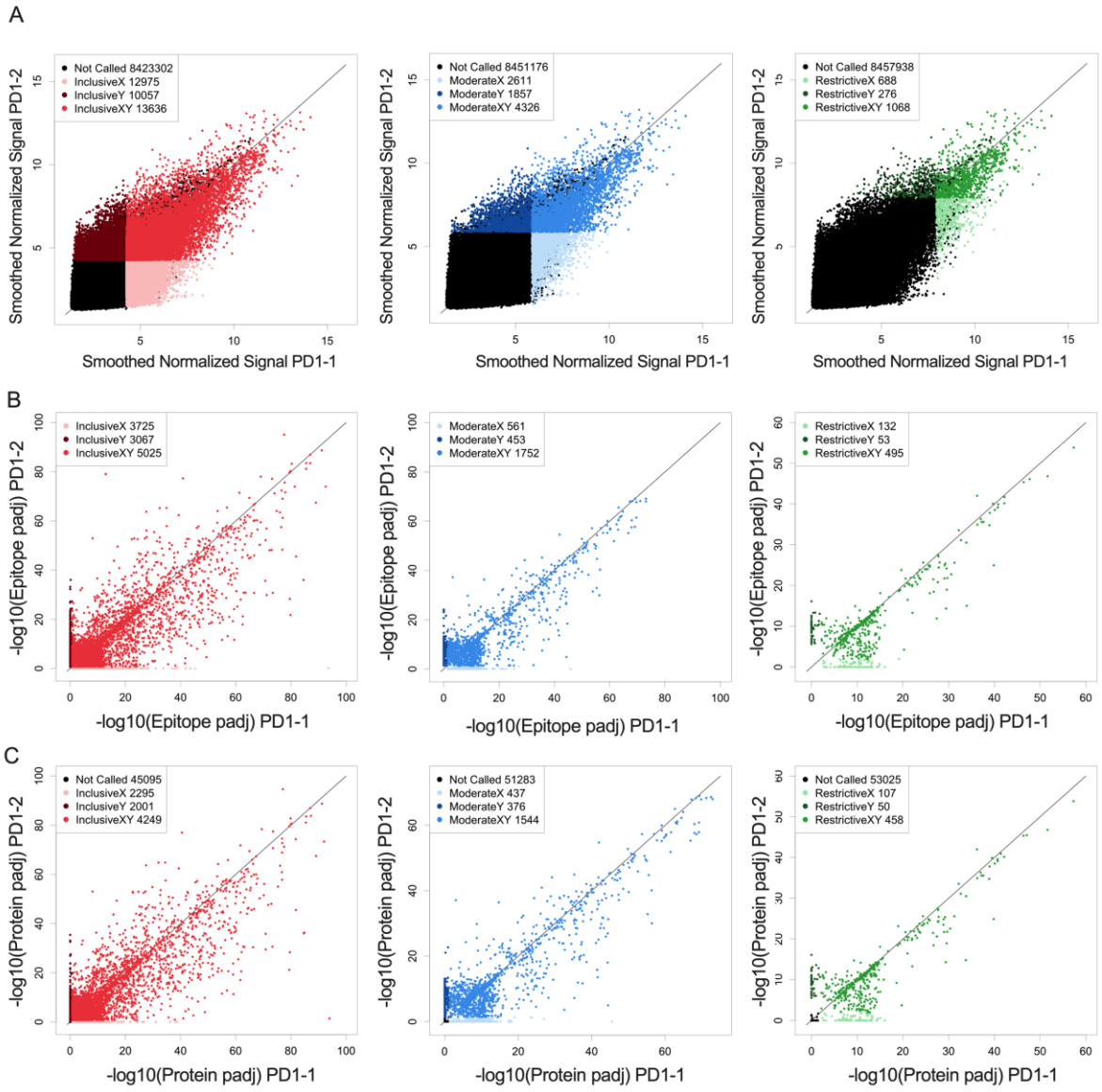
C: ELISA data for 13 of the 16 peptides highlighted in Figure 7 on the same immune serum samples as in **Figure 2.7C**. The peptides included here reflect 12 peptides chosen for strong antibody reactivity in 3 or more of the original 6 mice tested on the whole proteome array. As in Fig. 8A and B, the Gria4 peptide is included as a negative control peptide (at the bottom of the list). Also included here is a repeat immune serum sample from one of the 6 immune mice used in the original whole proteome samples as an internal control (B2 immune, also shown in the whole proteome data in **Figure 2.7A**, and ELISA data for original whole proteome samples in **Figure 2.7B**). Data shown are optical density values read at 450nm length on a scale from 0 to 3.5. Scales used for the heatmaps in **Figure 2.8 A-C** are consistent with the scales used in **Figure 2.7 A-C**.

Supplemental Figures and Tables



Supplemental Figure 2.1: Histograms of IgG binding to tumor cells for three individual mice (A3, A4 and B2) on 3 different tumor cell lines.

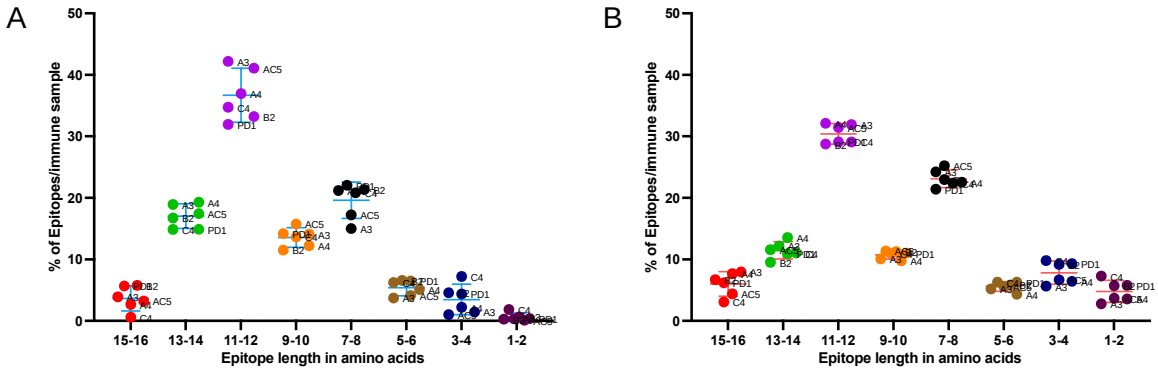
Timepoints correspond to the sample collection timeline in **Figure 2.1A**. **A:** Binding of serum antibodies to B16 tumor cells as measured via flow cytometry. **B:** Binding of serum antibodies to B78 tumor cells as measured via flow cytometry. **C:** Binding of serum antibodies to Panc02 tumor cells as measured via flow cytometry. Data are shown as fluorescence intensities detected in the red channel measuring fluorescence signal for APC. The samples for each individual mouse are normalized to mode to enable comparison between the different time points for each mouse.



Supplemental Figure 2.2: Reproducibility and reliability of probe, epitope, and protein calls

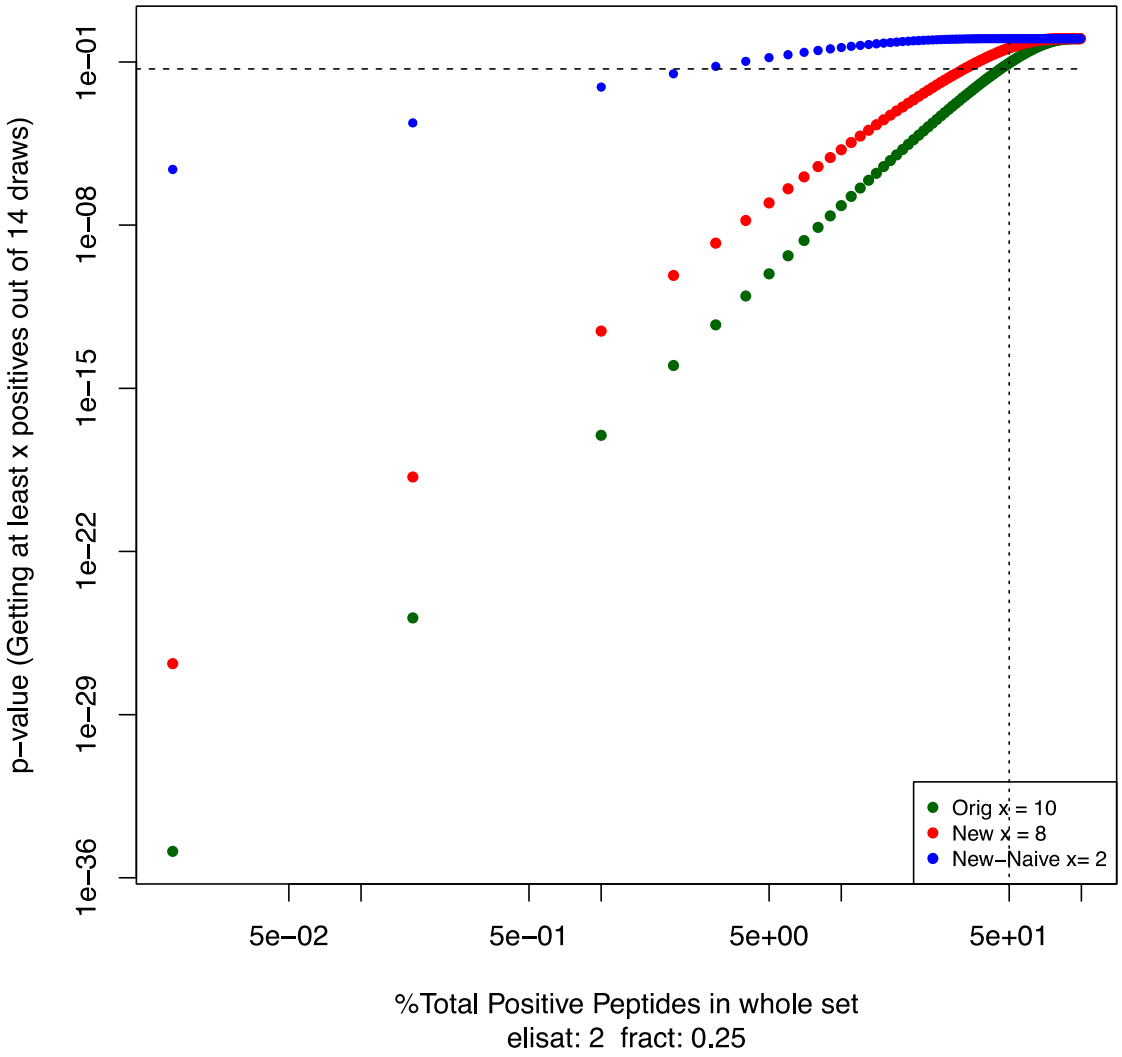
A-C: Reliability of Peptide array: Called peptides were separated in 3 categories [left panel inclusive; middle panel moderate; right panel restrictive] based on signal strength of the bound peptides. **A:** Cryopreserved serum samples from the same Immune blood sample from mouse PD1, were tested independently (sample PD1-1 and PD1-2) on identical whole proteome chips, in assays that were performed 1 year apart. These showed high correlation between signal strength and repeatability of results after data pre-processing on a probe level. Shown are the log transformed processed raw

peptide array data generated in arbitrary fluorescence units. In the inclusive category (signal >3xSD greater than the mean) 58% of probes recognized in sample PD1-2 are also recognized in sample PD1-1. In the moderate category (signal >6xSD greater than the mean) 70% of probes recognized in sample PD1-2 are also recognized in sample PD1-1 and in the restrictive category (Signal > 10xSD greater than the mean) 80% of probes recognized in sample PD1-2 are also recognized in sample PD1-1. Shown are the log transformed values of background corrected smoothed raw peptide array data generated in arbitrary fluorescence units. **B:** Scatter plot of the same serum sample from **A** looking at epitope level data instead of individual peptides. Plotted are the -log10 values of the epitope p values for each epitope recognized in sample PD1-1 and sample PD1-2. Graphs are split up into each category, B-left, showing inclusive category with 43% of epitopes co-recognized by both samples, B-middle showing moderate category with 63% of epitopes co-recognized by both samples and B-right showing restrictive category with 73%. of epitopes co-recognized by both samples, **C:** Scatter plot of the same serum sample from A & B looking at protein level data instead of individual peptides or epitopes. Plotted are the -log10 values of the protein p values for each protein recognized in sample PD1-1 and sample PD1-2. Graphs are split up into each category, C-left showing inclusive category with 50% of proteins co-recognized by both samples, C-middle showing moderate category with 66% of proteins co-recognized by both samples and C-right showing restrictive category with 75% of proteins co-recognized by both samples.



Supplemental Figure 2.3: Epitope length divided by intensity of binding signal

A & B: Epitopes were split in epitope length spanning between 1 and 8 consecutive 16-mer probes tiled at 2 amino acids. Over 80% of epitopes required amino acid sequences of 7-16 amino acids for binding in the moderate (**A**) and inclusive category(**B**). Only linear epitopes were probed, but small conformational epitopes are possible within a 16-mer peptide. The X axis shows the length of the detected epitope in 2aa steps as most peptides were tiled at 2 amino acids. The Y-axis shows the percentage of all epitopes per sample, for each of the 6 mice tested, with the corresponding epitope length. The designation for each of the individual 6 mice is shown in small letters next to each of the 6 dots appearing in each of the columns.



Supplemental Figure 2.4: Plot of the hypergeometric p-value of getting at least x or more positives out of a sample of 14 randomly chosen peptides versus the % of proposed total peptides in the whole set of ~6.5 million unique peptides. Green - original ELISA results on immune mice, Red - validation ELISA results on immune mice, Blue - validation ELISA results on Naive mice. Both the red and green line show very significant p-values for up to almost 50% of peptides of the 6.5million unique peptides giving a positive signal. However, looking at the Naïve samples a significant p-Value is only reached by a maximum of 5% of positive peptides in this simulation using values derived from the Nimble and ELISA data generated.

Supplemental Table 2.1: List of serum samples used across Nimble Peptide arrays, JPT multi-well peptide array and peptide ELISA with data displayed in this manuscript. Identifiers refer to individual mouse IDs, used for naïve as well as immune (noted in column labeled status). Samples marked as (Pool) on Nimble whole proteome array 1 were only used for **Figure 2.2A**, Naïve 1, Naïve 2 and Immune 1, Immune 2. These were array chips run with pooled samples where it was not possible to distinguish which serum sample caused each specific antibody binding. Samples were used at a 1:200 concentration for each specific sample, 1:50 concentration for overall serum concentration in the corresponding array chip.

Identifier	Status	Nimble whole proteome array 1	Nimble whole proteome array 2	JPT multiwell peptide array	ELISA
AC5	naïve	X			
	immune	X		X	X
PD1	naïve	X			X
	immune	X	X	X	X
B2	naïve		X	X	X
	immune		X	X	X
A3	naïve		X	X	X
	immune		X	X	X
C4	immune		X	X	
A4	naïve		X		X
	immune		X	X	X
naïve pool 1	naive	X			
naïve pool 2	naive	X			
immune pool 1	immune	X			
immune pool 2	immune	X			
Validation cohort for ELISA only:					
V1	naïve				X
	immune				X
V2	naïve				X
	immune				X
V3	naïve				X
	immune				X
V4	naïve				X
	immune				X
V5	naïve				X
	immune				X
V6	naïve				X
	immune				X
V7	naïve				X
	immune				X
V8	naïve				X
	immune				X
V9	naïve				X
	immune				X
V10	naïve				X
	immune				X
V11	naïve				X
	immune				X
V12	naïve				X
	immune				X
V13	naïve				X
	immune				X
V14	naïve				X
	immune				X
V15	immune				X
V16	immune				X
V17	immune				X
V18	immune				X
V19	immune				X
V20	immune				X

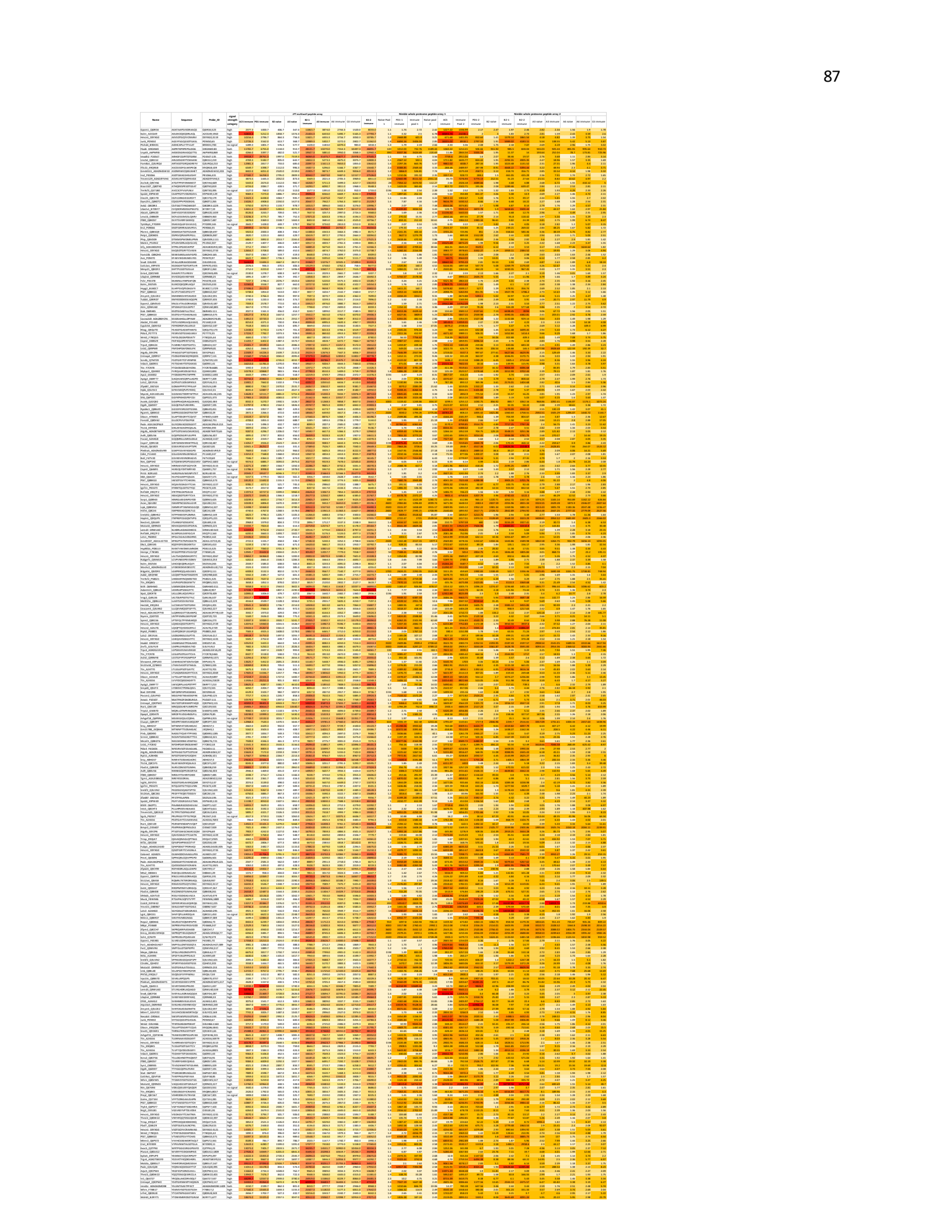
Supplemental Table 2.2: Co-recognized probe, epitope and protein calculations between repeat samples.

Values corresponding to **Figure 2.2** and supplemental **Figure 2.2** and percentage calculations based on the respective called probes, epitopes, or proteins for serum samples from the same mice at the same timepoints (using the same cut-offs as shown in **Figure 2.2** and **Supplemental Figure 2.2**) either run within a day of each other (for mouse B2 in **Figure 2.2**) or a year apart (for mouse PD1, in **Supplemental Figure 2.2**). In addition: calculations comparing serum from 2 different mice tested in the same run (B2 vs. repeat of PD1) show that different mice demonstrate co-recognition of a small fraction of the samples seen by those same individual mice. Calculations for % were done as following: overall=X only +Y only + X&Y; % overall=(overall/all probes)x100; % X&Y=(X&Y/all probes)x100; % of called (X)=(X&Y/(X only + X&Y))/100; % of called (Y)=(X&Y/(Y only + X&Y))x100; % of called X&Y=(X&Y/overall)x100.

	Sample	X (.1) only	Y (.2) only	X&Y	overall	all probes	not called	% overall	% X&Y	% of called (X)	% of called (Y)	% of called X&Y
Probe	B2 Z3	5311	7178	22806	35295	8459970	8424675	0.417	0.270	81.11	76.06	64.62
	PD1 Z3	12975	10057	13636	36668	8459970	8423302	0.433	0.161	51.24	57.55	37.19
	B2 Z6	591	1027	6653	8271	8459970	8451699	0.098	0.079	91.84	86.63	80.44
	PD1 Z6	2610	1857	4326	8793	8459970	8451177	0.104	0.051	62.37	69.97	49.20
	B2 Z10	95	149	1422	1666	8459970	8458304	0.020	0.017	93.74	90.52	85.35
	PD1 Z10	688	276	1068	2032	8459970	8457938	0.024	0.013	60.82	79.46	52.56
	B2vsPD1 (same run, different mice)											
	Z3	25536	21112	2581	49229	8459970	8410741	0.582	0.031	9.18	10.89	5.24
	Z6	6408	5347	836	12591	8459970	8447379	0.149	0.010	11.54	13.52	6.64
	Z10	1355	1128	216	2699	8459970	8457271	0.032	0.003	13.75	16.07	8.00
Epitope	B2 Z3	1374	1724	7868	10966					85.13	82.03	71.75
	PD1 Z3	3725	3067	5025	11817					57.43	62.10	42.52
	B2 Z6	58	89	2376	2523					97.62	96.39	94.17
	PD1 Z6	561	453	1752	2766					75.75	79.46	63.34
	B2 Z10	6	2	526	534					98.87	99.62	98.50
	PD1 Z10	132	53	495	680					78.95	90.33	72.79
	B2vsPD1 (same run, different mice)											
	Z3	8036	6886	1206	16128					13.05	14.90	7.48
	Z6	2434	1844	361	4639					12.92	16.37	7.78
	Z10	427	443	105	975					19.74	19.16	10.77
Protein	B2 Z3	806	1137	6440	8383	53640	45257	15.628	12.006	88.88	84.99	76.82
	PD1 Z3	2295	2001	4249	8545	53640	45095	15.930	7.921	64.93	67.98	49.72
	B2 Z6	45	73	2166	2284	53639	51355	4.258	4.038	97.96	96.74	94.83
	PD1 Z6	437	375	1544	2356	53639	51283	4.392	2.879	77.94	80.46	65.53
	B2 Z10	4	3	501	508	53640	53132	0.947	0.934	99.21	99.40	98.62
	PD1 Z10	107	50	458	615	53640	53025	1.147	0.854	81.06	90.16	74.47
	B2vsPD1 (same run, different mice)											
	Z3	5058	4062	2188	11308	53640	42332	21.081	4.079	30.20	35.01	19.35
	Z6	1767	1475	445	3687	53640	49953	6.874	0.830	20.12	23.18	12.07
	Z10	404	407	101	912	53640	52728	1.700	0.188	20.00	19.88	11.07

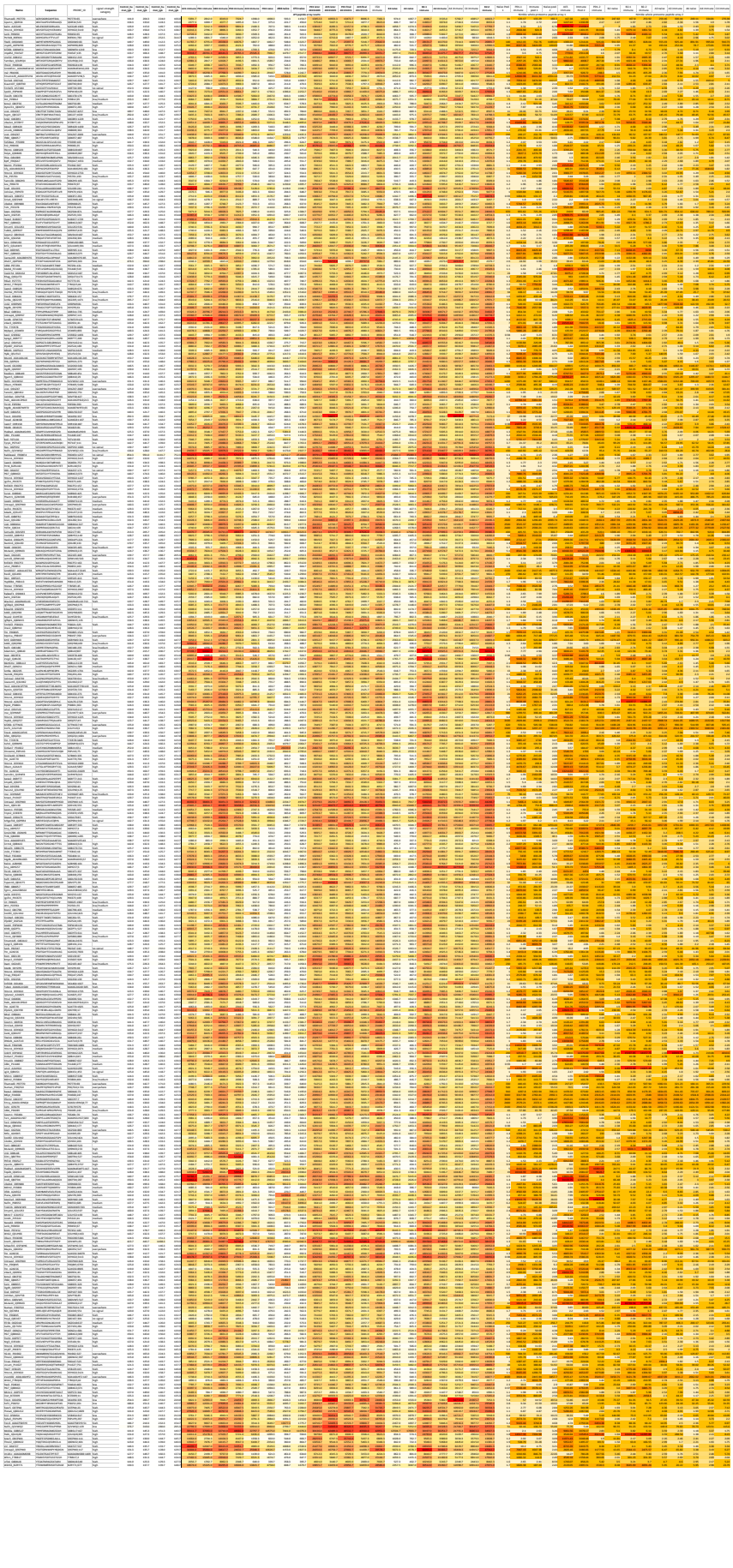
Supplemental Table 2.3: Peptides used in Figure 5 with Nimble and JPT data

284 peptides tested on JPT, and Nimble platform used in **Figure 2.5** in no or high signal category. Data provided for all samples from Nimble and JPT shown in **Figure 2.5**. Categories were chosen based on Nimble whole proteome array 1 and then tested on JPT and Nimble whole proteome array 2. All data from JPT multi-well peptide array runs and matched Nimble peptides on all samples can be found in Supplemental Data 1.



Supplemental Data 2.1:

376 peptides tested on JPT and Nimble platform. Data provided for all samples from Nimble and JPT runs performed. Categories were chosen based on Nimble whole proteome array 1 and then tested on JPT and Nimble whole proteome array 2.



Chapter 3: Prevalent binding motif in C57BL6 mice cured of B78 melanoma via immunotherapy

Preface

This chapter highlights an interesting finding we came across when analyzing epitopes that were recognized by multiple mice as highlighted in chapter 2. We found that half of the analyzed mice exhibited antibiodies against an immunodominant motif consisting of four amino acids (SDTG). While most of the work is complete, some key questions are left unanswered. We have a few ongoing experiments to address these and will update the results as we are able to obtain them.

One key limiting factor for this chapter was availability of mouse serum containing antibodies directed against SDTG-containing peptides to use for various experiments. We were able to show that additional mice injected with the same tumor cell line and treated the same way were able to generate an antibody response against some SDTG-containing peptides. These additional samples will be helpful for some of the planned experiments that will require larger volumes of serum containing antibodies against SDTG-containing sequences. We have tried many ways but have yet to identify the exact cause of SDTG recognition as well as what else besides SDTG causes antibody recognition of these peptides.

We are planning to submit the manuscript later this summer to a peer reviewed journal and make it available on BioRxiv at the same time. It is tentatively titled and authored by:

Hoefges A, Mathers N, McIlwain SJ, Tetreault K, Hampton A, Feils A, Tsarovsky N, Pinapati R, Garcia B, Patel J, Morris ZS, Erbe AK, Ong IM, Sondel PM. Prevalent binding motif in C57BL6 mice cured of B78 melanoma via immunotherapy. In preparation for submission; 2023.

Abstract

Background: Using a high-density peptide array, we assessed potential protein-targets for antibodies detected in mice cured of melanoma through a combined immunotherapy regimen. Our goal was to identify linear peptide sequences recognized by anti-tumor antibodies produced in mice cured of melanoma following immunotherapy.

Methods: Mice bearing B78 melanoma tumors were treated with a combination immunotherapy (local radiation therapy + intratumoral anti-GD2 mAb linked to IL2) capable of inducing an "*in situ* vaccine" effect (ISV), enabling mice to be cured of their tumors with long-term immune memory. Naïve (prior to tumor injection) and immune (post-rechallenge/after cure) sera were collected from these mice. Using flow cytometry, immune sera showed strong antibody-binding against B16 (parental cell line of B78) and B78 melanomas. These sera were then used on a whole-proteome peptide-array with 16-mer linear peptides overlapping by 12 or 14 aa to determine specific antibody-binding sites, and data were analyzed using HERON (**H**ierarchical antibody binding **E**pitopes and p**RO**teins from li**N**ear peptides), a dynamic programming method that scans adjacent peptides to determine whether a peptide is bound by antibodies. Epitopes were selected if peptides were bound using immune sera but not bound, or bound significantly less, with the sera from naïve mice.

Results: We identified many binding epitopes only present in immune mice. Among the epitopes found using our moderate binding category (namely the epitopes showing strong antibody binding), we noticed a repeating motif consisting of 4 amino acids (aa), SDTG, that was a component of over 60% of epitopes that are well recognized by antibodies induced in at least 50% of the cured mice. Even though the epitope analysis indicated that this SDTG motif was a component of the peptide actually recognized on these 16-mers by these immune antibodies, this SDTG 4 amino-acid motif is not the only reason for antibody binding to these epitopes, as ~ 1/3 of the 16-mer peptides

including this SDTG motif do not show detectible binding. The specific aa before and the specific 2 aas following the SDTG motif seem to be important for antibody binding. Using an independent cohort of mice, we were able to show binding of selected peptides containing the identified SDTG motif by serum from these additional immune mice.

Conclusions: This motif might be an important piece in the immune response of some mice to the immunotherapeutically induced cure of syngeneic B78 melanoma. We are further investigating what causes binding vs. no binding to the motif, and if the antibodies against it originated as a response to one specific immunogenic protein vs. an immune response to multiple immunogenic proteins containing the same motif. The presence of antibodies against this motif might be a useful biomarker to predict response to our ISV regimen and might have the potential to be used for other immunotherapy treatments.

Introduction

Recent cancer immunotherapy has shown substantial clinical progress especially in melanoma. Even so, melanoma still causes about 10000 cancer-related deaths in the US alone. Current patients with late-stage melanoma still have a low 5-year survival rate (~30%) which reflects an improvement from the 5-year survival rate for comparable patients prior to the introduction of modern immunotherapy (~15%) (Albittar et al., 2020; Siegel et al., 2022). Current therapeutic measures include surgical resection and immunotherapy utilizing checkpoint blockade like anti-CTLA4 and anti-PD1 (Albittar et al., 2020).

However, a 30% 5-year survival rate is not satisfactory, and many patients still do not show responses to current cancer immunotherapy treatment regimens (Chiriva-Internati & Bot, 2015; Patel & Minn, 2018).

Utilizing an *in situ* vaccination (ISV) regimen previously developed in our lab, consisting of local administration of radiation and an immunocytokine (IC), we are able to cure C57BL/6 mice with B78 melanoma tumors with lasting protective immune memory (Morris et al., 2016). Within this treatment, the low-dose radiation acts on the tumor to increase immunogenicity by modifying the tumor phenotype and releasing immune stimulatory cytokines. The IC is an engineered fusion protein that contains an anti-GD2 monoclonal antibody linked to IL-2. GD2 is a surface antigen on many cancers of neuroectodermal origin, including melanoma and neuroblastoma (Navid et al., 2010). Previous research showed epitope spread in mice cured of B78 via the ISV treatment, as demonstrated by rejection of the parental B16 melanoma cell line lacking GD2 expression (Haraguchi et al., 1994; Morris et al., 2016; Silagi, 1969; Silagi et al., 1972; Yang et al., 2012). Serum samples from these cured mice contained antibodies able to bind to the tumor cells, which was not seen in serum samples from naïve mice (Baniel, Heinze, et al., 2020; Hoefges et al., 2023).

To identify possible new antigenic targets that can be used in a therapeutic approach or as biomarkers, we utilized a whole proteome peptide array to determine the peptide targets of these antibodies found in the serum samples from these cured mice. In 2 recent manuscripts we detailed the immunologic methodology involved for detecting binding of these immune antibodies to thousands of peptides in the whole proteome array (Hoefges et al., 2023) and presented a bioinformatic analysis approach to determine which peptides, epitopes and proteins are well recognized by these immune serum samples (McIlwain et al., 2023).

This paper evaluates the group of peptides found to be most strongly recognized by sera from six immune mice evaluated using this whole proteome peptide array. When we focused on those 100 distinct epitopes that were most strongly recognized by at least 3 of these 6 mice, we found a dominant peptide motif of the 4 aa sequence SDTG, that was contained in roughly half of these 100

distinct epitopes. This manuscript presents the data identifying this immunodominant epitope and explores its possible meaning and implications.

Methods

Mice and in vivo tumor treatment:

The treatment model used here has been previously described in detail (Baniel, Heinze, et al., 2020; Morris et al., 2016; Morris et al., 2018). In brief, B78-D14 (B78) tumor bearing mice were treated when tumors reached ~ 100 mm^3 with a combination of 12 Gy local radiotherapy (RT), followed 5 days later with 5 daily intratumoral injections of the hu14.18-IL2 immunocytokine (IC). For mice that were cured, after 90 days they were rechallenged with an additional injection of the B78 tumor, and mice that rejected the rechallenge were considered immune. At specific timepoints, as previously described in detail in (Hoefges et al., 2023), serum was collected via mandibular bleed using BD serum collection tubes. For select animals, a terminal bleed was obtained via cardiac puncture under general anesthesia, prior to euthanasia, to obtain larger volumes of serum from immune mice. Experiments were performed under an animal protocol approved by the Institutional Animal Care and Use Committee.

Tumor cells:

B78-D14 ["B78", obtained from Ralph Reisfeld (Scripps Research Institute) in 2002] melanoma is a poorly immunogenic cell line derived from B78-H1 cells (obtained from Dr. Matthew Krummel at UCSF) which were originally derived from B16 melanoma (Becker et al., 1996; Binnewies et al., 2019; Broz et al., 2014; Haraguchi et al., 1994; Silagi, 1969). B78-D14 cells lack melanin, but were transfected with functional GD2/GD3 synthase to express the disialoganglioside GD2 (Becker et al., 1996; Haraguchi et al., 1994), which is overexpressed on the surface of many human tumors including melanoma (Nazha et al., 2020). B78-D14 and B78-H1 cells were grown in vitro in RPMI-

1640 (Mediatech) supplemented with 10% FBS, 2mMol L-glutamine, 100U/ml penicillin and 100µg/ml streptomycin. Mycoplasma testing via PCR was routinely performed.

Nimble peptide array:

The mouse whole proteome peptide microarray was designed and generated as previously described (Hoefges et al., 2023) and based on the protein set downloaded from UniProt in December of 2018 for C57BL/6 mice (The UniProt, 2017). Fluorescent intensity unit scoring and analyses for each peptide were analyzed as described (Hoefges et al 2023, McIlwain et al 2023).

Peptide ELISA:

For peptide ELISA, 29 separate JPT BioTides™ Biotinylated Peptides were purchased containing a TTDS-linker and biotinylation at the N-terminus. The peptides were generated using SPOT synthesis (Nahtman et al., 2007). Peptides were synthesized from C- to N-terminus ensuring that only full-length peptides will have a biotin at the N-terminus. Coating of streptavidin plates was performed per manufacturers instruction with a 250-fold dilution of lyophilized BioTides. ELISA was performed according to JPTs peptide ELISA protocol with the adaptation to a 384 well plate instead of the standard 96 well plate to conserve on serum samples. Neutravidin coated 384 well plates by ThermoScientific (#15400) were used. TMB incubation was set to 30min before adding of the stop solution and plates were read at regular intervals during TMB substrate incubation (reads at 655nm) and 1min and 3min after addition of stop solution (reads at 450nm). Optical density values were used to analyze results.

scRNAseq

Mice bearing B78-D14 tumors were treated with no treatment or external beam RT (12 Gy, Day 1) + IT-IC (50µg/dose, Day 6-Day 9). On Day 9, 1 hr post-intratumoral injection of IC, mice were

euthanized via CO_2 and tumors were excised. Duplicate tumors per treatment group were finely chopped and suspended in 2.35 ml of RPMI 1640 + Enzyme D (100µL), R (50µL) and A (12.5µL) (Miltenyi), transferred to a C Tube (130-093-237, Miltenyi) and placed on a gentleMACS Octo Dissociator with Heaters. On the gentleMACs dissociator, program m_imp Tumor_02 was used, after which the tubes were then inverted and placed in a shaking incubator at 37°C x 40min, and then returned to the gentleMACS dissociator and run on program m_imp Tumor_03. Dissociated tumors were filtered through a 40µm cell strainer (352340, Corning), washed with 15ml of RPMI 1640, and centrifuged at 300xg for 10 min. After removal of the supernatant, live cells were purified from the tumor cell pellet using Miltenyi's Dead Cell Removal Kit (130-090-101) according to the manufacturer's protocol. Cell viability was determined and confirmed to be >90%. The Chromium Next GEM Single Cell 5' Reagent Kits v2 (Dual Index) protocol from 10X Genomics was followed for scRNAseq processing. Quantity and quality of the RNA were assessed using High Sensitivity D5000 ScreenTape on an Agilent 4200 TapeStation System (UW Biotech Center). Around 10,000 cells per sample captured for library preparation, and sequenced on an Illumina NovaSeq6000 (Novogene, Sacramento, CA).

Statistical analysis:

Peptide array processing

Data from 13 total samples (5 naïve, 6 immune and 2 replicate samples of 2 individual immune samples) were assayed for antibody binding from 6,090,593 unique sequence probes mapped to a total of 8,459,970 probes (due to redundancies in tiling across protein sequences and using a mixed tiling of either 2aa or 4aa across each protein), or a total of 53,640 individual proteins. Using spatially corrected processed data from Nimble Therapeutics, the data were log2 transformed,

quantile normalized, and further processed using a sliding average mean window across the protein location of +/-8 aa.

HERON (McIlwain et. al 2023) was used to determine thresholds for calling antibody binding at the probe, epitope (consecutive probes), and protein level for each sample using meta-analyses methods to summarize binding across subjects in the post-rechallenge condition. Briefly, 1) a global p-value was calculated using a z-test for each probe signal using all sample and probe values, and 2) a differential p-value was calculated between the average of the naïve samples and each individual immune (tumor-free) sample. The global p-value and differential p-value for each immune sample were then combined using the Wilkinson's max meta p-value method (Wilkinson, 1951). After correcting for false discoveries using the Benjamini-Hochberg (BH) method (Benjamini & Hochberg, 1995), the individual probes for each post-rechallenge sample are considered bound by antibodies if their false discovery rates (FDR) are below a threshold. Epitope regions were identified by applying the skater algorithm (AssunÇão et al., 2006) to identify groups of antibody-bound probes (spatially and across subjects), and epitope meta p-values were calculated using the Wilkinson's max method on the 2nd highest probe p-value. Protein p-values were calculated using Wilkinsons's min (or Tippet's) method (Tippett, 1931). After correcting the epitope and protein p-values using the BH algorithm, the epitope and protein sample calls were made using an FDR cutoff. The number of samples that were bound by antibodies for each probe, epitope, and protein were tabulated as K of N statistics (K = # of samples with antibody binding; N = total # of samples).

MAFFT alignment tool

MAFFT (Multiple Alignment using Fast Fourier Transform) (Katoh et al., 2019) is a high-speed multiple sequence alignment program for amino acid or nucleotide sequences. It was used here to align epitope sequences for all epitopes identified in a certain category to visualize alignment of

sequences within these epitopes. Input of the top 100 16-mers sorted by average immune fluorescent score from the moderate category recognized in many individual immune mouse samples showed strong alignment for these top peptide candidates. Peptide candidates were previously filtered to represent only one protein isoform for each gene.

WebLogo creation

WebLogo is a web-based application to make the generation of sequence logos easy and painless. Sequence logos are a graphical representation of an amino acid or nucleic acid multiple sequence alignment developed by Tom Schneider and Mike Stephens. The logo consists of a stack of symbols for each position in the sequence. Level of sequence conservation at a certain position is displayed by the overall height of the stack while a specific amino acids conservation and frequency is indicated by the height and size of the specific letter. In general, a sequence logo provides a richer and more precise visual description of a binding site than would a consensus sequence. WebLogo was created by Gavin E. Crooks, Gary Hon, John-Marc Chandonia and Steven E. Brenner, Computational Genomics Research Group, Department of Plant and Microbial Biology, University of California, Berkeley and was described in detail by Crooks et al (Crooks et al., 2004) and is based on Schneider and Stephens sequence logo alignment (Schneider & Stephens, 1990).

Observed vs. expected motif frequency in the proteome

Observed vs. expected frequencies were calculated following Shen et al's probabilistic model for the analysis of frequencies of amino acid pairs within protein sequences (Shen et al., 2006). We utilized the protein sequences present on the whole proteome peptide array and calculated frequencies for each amino acid as well as the count and frequency for each 4-mer peptide. We then followed the calculations described in Shen et al. If the log2 of the observed divided by expected frequency is

larger than zero, the sequence occurs more often than expected, if it is smaller than zero, it occurs less often than expected.

Expected frequency for SDTG=P(S) x P(D) x P(T) x P(G) = 0.0856 x 0.0482 x 0.0542 x 0.0635 = 0.00001419; SDSG= 0.00002240

Observed frequency of SDTG: 310 SDTG-mers found in the dataset = 0.00001382.

Observed frequency for SDSG: 806 SDSG-mers found in dataset = 0.00003594.

Log2(Observed/Expected) for SDTG: -0.03795

Log2(Observed/Expected) for SDSG: 0.68209

Random forest model

All unique sequences bearing the xSDTGxx peptides were found using a regex search.

In total there are 848 sequences. The following features were extracted:

- motif.pos1 - the amino acid at the 1st position before the SDTG
- motif.pos6 - the amino acid at the 1st position after the SDTG
- motif.pos7 - the amino acid at the 2nd position after the SDTG
- start - the starting position of the SDTG within the peptide
- sequence_length - the length of the peptide sequence

Using the alakazam R package, the following features were collected over the whole peptide:

- gravy - the hydrophobicity of the amino acid sequences
- bulk - the average bulkiness of the amino acid sequences
- polar - the average polarity of the amino acid sequence
- aliphatic - the average aliphaticity of the amino acid sequence
- charge - the overall charge of the amino acid sequence
- region_ACIDIC - the fraction of amino acids that are acidic
- region_BASIC - the fraction of amino acids that are basic
- region_PHOS - the fraction of amino acids containing phosphorylation sites (STY)
- region_AROMATIC - the fraction of amino acids containing aromatics

We also estimate features for the gravy, bulk, polarity, aliphaticity, and charge of the motif.pos1, motif.pos6, motif.pos7, and the motif.pos67 amino acid sequences.

We regressed the average post-rechallenge signal using random forests from the random forest package. After investigating the mean squared error and mean absolute error using 10-fold cross-validation, the also investigated the variable importance to estimate the largest contributors to the random forest model. We find that charge, polarity, and the bulkiness of the amino acids at the 1st and 2nd position after the SDTG have the greatest effect on the regression modeling.

2-way ANOVA

An ordinary 2-way ANOVA was carried out using GraphPad Prism version 9.5.0 using data displayed in **Figure 3.6A**. Multiple comparison results used Tukey's multiple comparisons test.

DNA sequencing analysis pipeline

DNA sequences from B16 and B78 cell lines were kindly provided by the Maris Laboratory at Childrens Hospital of Philadelphia (CHOP) and compared to the ensemble GRCm39.107 mouse genome. Paired-end reads were aligned using bwa mem, converted to sorted bams utilizing samtools (https://samtools.github.io/bcftools/howtos/csq-calling.html), gatk was used to analyze for single nucleotide variations, small indels and frameshift mutations (https://gatk.broadinstitute.org/hc/en-us). Duplicates were marked using gatk as well as to add or replace read groups. Gatk Haplotype caller was used for variant finder and bcftools csq for consequence calling on protein sequences. Vcf2prot (https://github.com/ikmb/vcf2prot) was utilized to find new proteins.

scRNAseq

Raw reads were aligned to the mm10 reference genome together with UMI (unique molecular identifier) counting using the Cell Ranger pipeline (V3) from 10X Genomics. Data was filtered using DoubletFinder (McGinnis et al., 2019) to remove potential doublets. Further filtering was done to include only the cells with low mitochondria contents (<=10%) and more than 200 genes covered by

the mapping. To integrate the scRNAseq, we used a fuzzy clustering-based integration method (Harmony method) (Korsunsky et al., 2019) to account for potential technical variance across samples.

Downstream analysis for all cells and for only tumor cells were based on Seurat single-cell analysis package (Stuart et al., 2019) including: principal component analysis with standard deviation saturation elbow plot to select the optimal number of principal components, graph-based clustering using FindCluster with different resolutions from 0.1 to 2 to justify the number of clusters based on representative markers overlaid in the hierarchical tree across different resolutions (Clustree R package), differential expression analysis using MAST (Finak et al., 2015) implemented in Seurat with the cutoff average log2FC 0.5, and at least 30% of cells expressed the markers. Data was further analyzed using GraphPad Prism version 9.5.0.

Results

Most of the highly recognized epitopes share a common motif

We have shown previously that B78 melanoma bearing mice successfully treated and cured with ISV could generate many different antibodies against endogenous proteins on the cell surface of B78 melanoma and in the proteome of C57BL/6 mice (Hoefges et al., 2023). We used a high-density whole proteome peptide array system to focus on identifying antigenic epitopes recognized with strong binding by antibodies in immune serum samples (after the mice had been successfully cured of their tumor and rejected a rechallenge with the same or a related tumor type) but no binding by antibodies in naïve mouse serum samples (before tumors were implanted). An example of how the raw data from this high-density array system can be analyzed to identify highly recognized epitopes is presented in **Figure 3.1A**. As we reported, we used the HERON system, evaluating the strength/intensity of binding data to define 3 analysis categories of peptide and epitope binding,

based on how many standard deviations above the mean was used to define the threshold for each category. We utilized a very strict (> 10 S.D. over the mean), a more moderate (> 6 S.D. over the mean) and an inclusive (> 3 S.D. over the mean) threshold. When focusing on the moderate threshold, we saw around 2000 distinct epitopes on average recognized by each of 6 individual immune mice probed on a whole-proteome peptide array (**Figure 3.1B**).

To identify which of these epitopes might be more important and need further investigation, we looked at epitopes that were recognized by multiple immune mice. **Figure 3.1C** shows the number of these moderate category epitopes that were co-recognized by at least 3 of our 6 whole proteome profiled immune mice (highlighted by green dots in **Figure 3.1C**). We then focused on those epitopes with the highest average immune signals. We furthermore filtered out different splice variants of proteins if the identified epitope was in a conserved area (i.e., the peptides were the same for this area between different splice variants). One epitope entry per specific sequence was used to ensure each was a unique epitope, and made certain that the binding component (main aa sequence recognized) of the epitope was present and then selected the top 100 of these epitopes. We then collected the sequence for the whole epitope and entered it into a sequence alignment program (MAFFT, Multiple Alignment using Fast Fourier Transform) to check for common aa sequences and possible overlap between proteins and epitopes that otherwise was not apparent to us (**Figure 3.1D**). Interestingly, we saw that over 50% of our top 100 sequences aligned to a specific motif four amino acids in length. It was either an SDTG (Serine, Aspartic acid, Threonine, and Glycine) or an SDSG, where the T (Threonine) was replaced by another S (Serine). SDTG was present 45 times while SDSG was present only 13 times. This high representation of SDTG within these 100 sequences was further illustrated in **Figure 3.1E**.

Motif presence in the proteome

To investigate the meaning of this motif we checked for overrepresentation of SDTG and SDSG in the proteome. Utilizing Shen et al's developed algorithm to determine if a specific amino acid sequence's observed frequency is more or less than expected, we found that there are 310 SDTG-mers belonging to unique UniProt IDs in the whole proteome peptide array dataset, which leads to an observed frequency of 0.00001382 vs. an expected frequency of 0.00001419 (Shen et al., 2006). The Log2(Observed frequency/expected frequency) gives information if the sequence occurs less or more often than expected. If the result is greater than 0, the sequence occurs more than expected, if it is less than zero, it occurs less than expected. Log2(0.00001382/0.00001419) = -0.03795. This result indicates that SDTG occurs slightly less often in the proteome than expected. With it occurring less often in the proteome or even with an average frequency as expected, finding SDTG in 45 of our top 100 co-recognized epitopes is quite unexpected. The Log2(0.45/0.00001382) = 15, consistent with a far higher than expected presence in the epitopes recognized by the immune sera from these 6 mice, and thereby suggesting it has biological significance in this setting.

When testing for overrepresentation of SDSG in the proteome, we found an overrepresentation of SDSG with an expected frequency of 0.00002240 and an observed frequency of 0.00003594, leading to a value of 0.68209 which indicates a small overrepresentation of the sequence. Thus, the appearance of SDSG in 14 of the 100 top epitopes is also greater than expected, but ~ 10-fold less so than SDTG (**Figure 3.1D**). For SDSG, the Log2(0.14/0.00003594)= 12.

Most, but not all SDTG motif containing peptides are recognized by some immune sera

We investigated if we would observe binding by immune serum to all peptides in the proteome containing SDTG. We filtered the whole proteome peptide array dataset for SDTG-containing probes and grouped those into regions, combining individual consecutive peptides into one region, and

removed all redundant sequences from separate splice variants. By this process we found 169 separate SDTG sequences in the proteome (**Figure 3.2A**). Not all SDTG peptides showed binding in the immune samples. Of these 169 sequences, 125 sequences were detected utilizing HERON with 48 being recognized in the restrictive category (light green in **Figure 3.2A**), 40 in the moderate category (dark green in **Figure 3.2A**) and 37 in the inclusive category (dark red in **Figure 3.2A**). 44 of these SDTG-regions were not recognized by HERON (light red in **Figure 3.2A**). Overall, very little reactivity to SDTG was found in naïve samples, while over half of all peptides showed significant binding in some immune samples and a quarter of peptides showed very high reactivity (88 regions were recognized in moderate or restrictive, 48 in the restrictive category by HERON) in immune samples.

To investigate what may be different between the SDTG-containing peptides with substantial immune serum binding versus the SDTG-containing peptides without substantial immune serum binding, we utilized a random forest model. This screened for the potential importance of many different properties of these peptides to help distinguish those peptides that were bound by immune sera from those that were not. The properties evaluated, and their importance in whether an individual SDTG containing peptide was bound by immune sera are shown in **Figure 3.2B**, and include: peptide charge, presence of specific amino acids in specific positions, polarity, groups they contain, their acidity, and other features.

The random forest model showed that charge, as well as amino acid positions adjacent to the SDTG motif do have some influence on whether a particular peptide might or might not be bound by immune serum, none of these parameters, nor any combination of these parameters can be used with certainty to determine whether a given peptide will or will not be bound by immune serum (**Figure 3.2B**). For example, the determinant of charge was identified as the most important determinant tested in **Figure 3.2B**. Peptides with extreme charges (< -4 or > +3) all show low or

moderate binding by immune serum (**Figure 3.2C**). Yet, peptides with moderate-low charges ranging from -4 to +3 show about equal distribution of peptides with low signal, with moderate signal and with high signal. In other words, within the moderate charge range, there is no charge difference between peptides that are recognized well vs. not recognized by immune sera (**Figure 3.2C**).

We then asked whether the single amino acid position immediately before the SDTG sequence influenced the ability of the peptide to be bound by immune sera (**Figure 3.2D**). However, for the 20 aa tested in the position before the SDTG, 17 showed a wide distribution of peptides with high, medium, and low signal average with immune sera; indicating that these 17 individual aa did not substantially influence immune serum binding. Two aa appeared to show only strong binding by immune sera (C and K), while one aa showed only low binding (W). However, for all 3 of these, the number of peptides found with that amino acid in the first position was low (4 for C, 9 for K and 5 for W). Further inspection showed that these peptides were all closely related stacked peptides from a single epitope for C and W and from 2 epitopes for K. Thus, it is possible that the presence of C or K in this position actually increased binding by immune sera, while a W may actually decrease binding. However, as the number of epitopes with these 3 aa in this position is so small, it may be that some totally different property is influencing the high or low binding of these, and by coincidence/chance alone, the C, K or W is showing up in this position in these few peptides. **Figure 3.2E & F** show a similar analysis as shown in **Figure 3.2D**, but for the importance of the specific aa (of 20 possible) seen in position 6 (**Figure 3.2E**) and position 7 (2 down from the SDTG) (**Figure 3.2F**). For the most part, most aa in this position 6 and 7 are associated with a range of peptides with high, moderate, or low binding by immune sera. The combination of all 3 of these positions (1-SDTG-6-7) can be seen in **Supplemental Figure 3.1**.

Antibody to the SDTG Motif is present in additional immune animals

For further investigation of the antibody response in immune mice to this SDTG motif, we tested whether additional animals that were cured of the same B78 melanoma utilizing the same treatment also developed antibodies against SDTG containing epitopes. We chose 26 separate peptides containing SDTG from those that showed high-binding by some immune serum samples from the original whole-proteome peptide array. The original whole proteome peptide array data using the original naïve and immune serum samples (Hoefges et al, 2023) are shown in heat-map format, against these 26 peptides, which are shown on the Y axis in **Figure 3.3A**. We tested these same 26 peptides via a separate peptide ELISA assay system (see Methods and Hoefges et al, 2023) on the same original naïve and immune serum samples (where they were available) (**Figure 3.3B**) and were able to reproduce somewhat similar binding patterns between the data from the whole proteome array and the peptide ELISA systems. When testing these same 26 SDTG-containing peptides in the peptide ELISA system using serum samples from 20, new, additional, independent mice that had successfully cleared their B78 melanoma tumor utilizing the same radiation in combination with the immunocytokine immunotherapy, and had never been tested on the whole proteome array, we were able to identify 9 additional mice that also responded to at least some of the SDTG-containing peptides (**Figure 3.3C**). With 9 out of 20 recognizing at least some SDTG-containing peptides, the ratio is similar to that seen with our original whole proteome sample set (3 out of 6 mice recognizing SDTG peptides). This therefore confirms that induction of antibody to some SDTG containing peptides is induced by the process of bearing B78 tumor and receiving this radio-immunotherapy treatment.

Why is antibody recognition of SDTG so prevalent?

One possible cause of SDTG reactivity could relate to the fact that the B78 tumor used here was derived from B78-H1 melanoma that was then transfected with GD2 synthase and GD3 synthase to induce membrane expression of the GD2 disialoganglioside, to enable recognition by the anti-GD2 immunocytokine used in the therapy (Haraguchi et al., 1994). To test this, we obtained the parental B78-H1 cells (B78 cells never transfected to express GD2) and implanted these into 20 C57BL/6 mice which we then treated with radiation and intratumoral IL2. Since these tumors did not express GD2, treatment with an anti-GD2 immunocytokine would not be advantageous. This treatment with intratumoral IL2 would not be as effective as the immunocytokine against a GD2 expressing tumor but should still provide some immunotherapeutic benefit. To help generate some tumor-free mice, we treated half of the cohort with anti-CTLA4 in addition to the radiation and IL-2, since we have previously shown that the addition of anti-CTLA4 to the treatment increases response and can cure larger tumors (Morris et al., 2016; Morris et al., 2018). Serum samples from all 20 of these mice were obtained on day 55 post-treatment (when some animals with progressive tumors needed euthanasia). These serum samples were all tested for reactivity to SDTG-expressing peptides using 13 of the previously most-reactive peptides in the ELISA system (as shown in **Figure 3.3B&C**). We included all mice here irrespective of tumor clearance or not. While 1 of the 20 mice showed very strong anti-SDTG antibody response to all of the SDTG containing peptides (mouse D5), 11 of the 19 mice tested had detectible antibody to at least one SDTG containing peptide, 2 mice had reactivity to 2 SDTG peptides and 2 mice had reactivity to 3 or more SDTG peptides (**Figure 3.4A**).

Tumor growth curves of these 20 mice treated either with radiation and IL-2 only or radiation in conjunction with IL2 and anti-CTLA4 are shown in **Figure 3.4B**, no obvious difference could be noted between the two separate groups. Overall, the complete response rate was low in this experiment: 2 mice getting the IL2 alone had complete response (CR) and 5 mice getting the IL2 with anti-CTLA4

had a complete response (CR), out of 10 in each group. When dividing the groups based on ELISA reactivity to SDTG-containing peptides and grouping them into 3 groups: a) did not recognize any peptide, b) recognized one peptide, or c) recognized more than 1 peptide, we also saw no obvious differences in the growth kinetics (**Figure 3.4C,** p-values 0.05 for 0 vs. 1, 0.06 for 1 vs. 2+ and 0.79 for 0 vs. 2+). The complete response rate and partial response rate were similar in **Figure 3.4C** for these mice with antibody detected to SDTG to **Figure 3.4B** where groups were divided based on treatment received.

Hypotheses regarding the frequent SDTG recognition

We established that the SDTG motif is recognized frequently by mice cured of B78 melanoma utilizing radiation and IC treatment, but we are still investigating why it is frequently recognized. One of our hypotheses is that somewhere in the B78 genome a mutation happened that caused an amino acid sequence change in a protein which led to a protein previously not containing SDTG, now containing SDTG. This could have happened via a deletion, insertion or via a base pair switch that resulted in a changed amino acid sequence (**Figure 3.5A**). Another hypothesis is that the protein already contained an SDTG, but due to a mutation in the gene the conformation of the protein was changed which now exposes the SDTG in the folded protein. Either of these scenarios could make the modified protein appear foreign and different to the immune system, in the exposed SDTG region, resulting in immune recognition of this mutated protein (**Figure 3.5B**). However, our search through DNA mutations in B78 versus normal mouse cells has not yielded any new proteins in B78 that contain SDTG and has not shown mutations that would cause conformational changes in a protein that contains SDTG. We did find 2 proteins (Ccna1f and Btbd8) with mutations that contain SDTG (**Supplemental Figure 3.2**), but the mutations seemed to be minor and not affect folding. Both proteins did show high binding in the whole proteome peptide array to the SDTG-containing region. However, further analyses utilizing alpha fold protein prediction is underway.

The third hypothesis is not based on DNA mutations, but rather differential expression between normal tissue and B78 melanoma either due to treatment or just in general (**Figure 3.5C**). Might there be a protein that contains SDTG that has very restricted expression in post-natal tissues, such that the immune system is not aware of it. If that protein were overexpressed in or on B78 tumors, the process of rejecting it immunotherapeutically may turn on a strong immune response, which could focus on the SDTG containing epitope. To investigate this further, we conducted scRNAseq of treated and untreated tumors taken 8 days post radiation, and immediately after receiving the 4 subsequent doses of the immunocytokine, to look at differential expression in SDTG-containing proteins at this time point in treated and untreated tumors. In these analyses of scRNAseq data, we used RNA expression of SDTG-containing proteins in the B78 tumor cells themselves and compared this to RNA expression from the non-malignant, non-melanoma cells within the tumor, consisting of cells from the tumor microenvironment, identified as epithelial cells, myeloid cells, B cells, stromal cells, NK cells and T cells (**Figure 3.6A**). We further split this up into treated and untreated tumor and other.

We found four SDTG-containing proteins that had significantly different expression in the treated vs. untreated tumor samples as well as in comparison to the non-tumor cells found in the tumor microenvironment, mainly T and NK cells, stromal cells, myeloid and B cells and endothelial cells (**Figure 3.6B**). The 4 genes with differential mRNA expression and SDTG-motif present in their protein were Junb, Pdia3, Bng and Rpl11 as identified via 2-way ANOVA and Tukey's multiple comparisons test with Junb and RpI11 higher in the treated than untreated, and the other 2 lower in the treated tumor than the untreated tumor. Bsg showed an overall higher expression in tumor than other independent of treatment. For Pdia3 we observed highest expression in the untreated groups vs. the treated groups.

Junb, Pdia3 and Rpl11 were also highly recognized via antibody responses as measured on the whole proteome peptide array (**Figure 3.3A**). We attempted ELISA analyses with these peptides but were not able to get good ELISA binding to any of these peptides even testing the original serum samples from 4 of the original 6 mice, of which at least 2 mice tested in this ELISA had shown strong binding to these peptides in the proteome array (**Figure 3.3A** vs. **Figure 3.3B**).

Discussion and conclusions

In this paper we identify a 4 aa SDTG motif, found in many peptides, that was strongly bound by immune sera from mice cured of B78 melanoma utilizing an ISV treatment consistent of radiation and intratumoral immunocytokine targeting GD2 on the tumor cell surface and delivering IL2. Antibody able to bind to this motif was identified in roughly half of all mice that have become immune to this B78 melanoma tumor via our treatment as shown via successful rejection of a rechallenge (**Figure 3.1**). This motif is not overrepresented on its own in the mouse proteome when utilizing an algorithm to determine expected vs. observed frequencies of specific motifs (Shen et al., 2006). To identify a motif that is this frequently recognized in the immunome of mice is especially striking when looking at the study of Han & Lotze who took two blood samples from the same individual, from the left arm in the morning and the right arm in the evening and divided each draw into 20 aliquots (Han & Lotze, 2020). They then went on to analyze all complementarity-determining regions (CDR)-3 regions found in T and B cells, via sequencing, to establish the immune repertoire that this one individual had at the 2 timepoints. The immune repertoire was defined as the sum-total of the individual clonotypes within one chain, including individual CDR3 sequences. They found over 5 million separate clonotypes with only 12,220 (0.24%) of these being shared by all 40 aliquots. Only about 50% were shared between any 2 aliquots with a higher percentage being shared at the morning draw (55%). About half of the identified clonotypes appeared in only one of the 40 aliquots. Seeing that only 0.24% of all clonotypes were shared between all aliquots taken within one

individual highlights the diversity present at any given time in clonotypes and the adaptome, where the adaptome is defined as the sum-total of expressed T and B cell receptor genes in a sample, composed of seven chains, including the alpha/beta and gamma/delta chains for T cells, and heavy/lambda or kappa chains for B cells (Han & Lotze, 2020).

Finding this 4 aa sequence motif in more than 50% of the top epitopes from a whole proteome antibody binding screen identifying epitopes co-recognized by at least 3 of 6 separate immune mice, but not by their naïve sera (**Figure 3.1**) in comparison to the fraction of shared clonotypes within a single individual, suggests this motif is highly recognized by an immune response triggered by bearing the B78 melanoma and receiving this curative immunotherapy.

We further characterized the SDTG-motif and its properties. We identified all SDTG-containing regions present on the whole proteome peptide array and kept only unique SDTG-containing regions. This revealed that of the 310 SDTG-mers only 169 were unique epitopes and not part of an isoform of the same gene (**Figure 3.2A**). Within these 169 SDTG-containing regions we found 125 being recognized by HERON, with 48 in the restrictive category, 40 in the moderate category and 37 in the inclusive category. 44 peptides were not recognized as having significant binding by HERON. While this highlights that most SDTG regions did show antibody binding, not all of them did. To determine what is required for antibody binding besides the presence of the four aa long sequence of SDTG we utilized a random forest model. While the random forest model showed some correlation to overall charge of the peptide as well as aa present right before or after the SDTG motif, it did not show a clear answer to what determines whether immune sera from a mouse that binds to at least some SDTG containing peptides will or won't bind to an individual SDTG containing peptide of interest (**Figure 3.2B**). One issue with this analysis could be that we were only able to use the data present from the 169 SDTG containing regions from the proteome used in the initial peptide array. However, for a thorough analysis of importance of which aa found just before or after

the SDTG aa (as in **Figue 3.2D&E**) it may be helpful to look at additional aa before or after the motif as well as all possible combinations of the 20 aa in all positions.

To further substantiate the significance of this motif, we tested sera from several separate immune mice, to validate that antibody against this SDTG motif could be confirmed in additional mouse serum samples. By testing for antibody binding to SDTG peptides in a validation set of 20 mice (compared to the original 6) we were able to identify an additional 9 mice with strong reactivity towards at least 1 SDTG-containing peptide (45%, **Figure 3.3C**), with several mice recognizing several SDTG peptides. We translated these results with the whole proteome peptide array system to an analysis in a separate ELISA system using a fraction of the SDTG containing peptides that were strongly positive in the array. Not all of these peptides showed the same pattern of binding on the same samples utilized in the whole proteome peptide array, but there were some clear similarities. In general, peptides that did not show the same pattern, i.e., lacked positive binding signal in the ELISA using the serum samples from the initial peptide array, also did not show binding to any of our validation serum samples via ELISA. This does not mean that these validation samples do not have reactivity towards these peptides, but rather hints at a possible issue with binding in the ELISA system. This could be caused by the different synthesis mechanisms used to generate these peptides, the different way of attaching peptides to the plate between the whole proteome array and the peptide ELISA, different incubation times and temperatures as well as the opposite direction in which the peptides are attached to the plate in the array vs. the ELISA (N-terminal attachment via Biotin in the ELISA system vs. C-terminal attachment via a Lysine in the peptide array).

Due to these issues with some of the peptides, we decided to choose peptides we had seen good ELISA results with to further screen for SDTG-reactive antibody presence.

One possible mechanism that might have induced antibody to SDTG could potentially be the plasmid used to integrate GD2 and GD3 synthase into the B78-D14 cell line to express GD2 on the cell surface. To test if this was the case, we implanted B78-H1 cells into mice and treated them with radiation and IL2 as well as radiation, IL2 and anti-CTLA4 to generate tumor-free immune mice. B78-H1 is the original B78 cell line that had never been transfected and does not express GD2. We were able to show that 11 of 20 mice generated antibody able to bind at least one of the 13 SDTG peptides tested, and one mouse generated a very strong antibody response to all 13 SDTG peptides (**Figure 3.4A**). However, this mouse did not clear its tumor. Nonetheless, the ability of these mice, never exposed to a plasmid containing tumor to recognize some SDTG containing peptides establishes that SDTG generation in these mice is not caused by possible changes in the genome due to integration of the plasmid enabling GD2 expression on the cell surface.

The strength of the anti-SDTG antibody response in this experiment was lower than those shown in **Figure 3.3**, based on the lower fraction of SDTG peptides recognized. Even so, the number of positive peptides recognized by these 19 serum samples from treated mice (**Figure 3.4A**) of these 13 selected SDTG containing peptides (27 positive reactions out of 247 tests involving 13 SDTG-containing peptides identified with separate mice) far exceeds ($p < 0.001$) the number of ELISA positive reactions seen with the 20 highly immune validation mice shown in **Figure 3.3C**, when they were tested on 10 randomly selected 16-mer peptides from the whole proteome array (1 weakly positive reaction out of 200 combinations in a similar ELISA, as reported previously (Hoefges et al., 2023). This could be further caused by differences in selection of mice tested via ELISA. For this experiment we tested all mice included in the experiment. We rechallenged all of these mice regardless of current tumor status 30 days after they finished treatment and harvested their serum 10 days after rechallenge. For all other experiments we only utilized rechallenge serum from mice

that had successfully rejected a rechallenge as well as their primary tumor with the primary tumor cured at least 30 days prior to rechallenge.

We wanted to assess if the development of antibodies targeting SDTG-containing regions was associated with tumor clearance or response to treatment. To assess this, we grouped the mice based on SDTG-recognition by their serum testing in ELISA. If only 1 SDTG-containing peptide of the 13 tested was recognized, they were evaluated separately than those mice that had serum that bound more than one SDTG-containing peptide. A third group contained animals that did not recognize any of the 13 tested SDTG-containing peptides. Analyses of tumor size in response to treatment between these 3 groups did not show any significant difference; nor was there any trend. These results indicate, in this small study, that the development of antibody to SDTG following treatment was not associated with a detectible impact on anti-tumor effect in the initial tumor response to treatment. The only trend visible was that the addition of anti-CTLA-4 to the radiation + immunocytokine treatment induced a slightly better overall response rate than radiation and IL-2 alone ($p = < 0.001$); all of the animals receiving the additional anti-CTLA4 showed a partial or complete response (**Figure 3.4B & C**). SDTG-reactivity and response correlation was difficult due to sample size. While this might be due to SDTG-detection not aiding response, it could also be because this cohort overall had a lower than previously reported response rate (for SDTG as well as treatment). All of this might again be due to the different cell line and treatment used than in previous reports as mentioned above. In a separate analysis of the curative immune response to these same B78-D14 tumors, in response to radiation, checkpoint blockade and toll-like-receptor activation, our collaborative team found that the involvement of B cells played a role in the primary anti-tumor response, and tumor-reactive antibody was also involved in the anti-tumor effect (Jagodinsky et al., 2022). It was not possible to obtain serum from these animals to evaluate anti-SDTG antibody.

When testing the different hypotheses proposed in **Figure 3.5A&B**, DNA sequencing only showed mutations affecting the amino acid sequence in two genes containing SDTG (Cacna1f and Btbd8). Both had these mutations in a separate area of the protein than the SDTG sequence is located, and the mutation present did not seem to affect protein folding properties (**Supplemental Figure 3.2**). We are further investigating the influence of the mutation on protein structure utilizing alpha fold predictions.

RNAseq analysis revealed differential expression of 4 genes between the tumor and other cell types found in the tumor microenvironment (**Figure 3.6**). One issue with the RNAseq data is that we were only able to match 115 gene names containing SDTG. Some of the missing genes were ones with high peptide array binding. These potentially could have a more differential expression between normal tissue and tumor. We are currently also conducting bulk RNAseq of tissues obtained from tumor-bearing mice, to compare B78 cells vs. radiated B78 cells vs. normal C57BL6 skin cells. We are hoping that this comparison will allow for better comparison between tumor cells and "normal" non-tumor cells (those found in skin) than we did with our initial analysis that used the non-tumor immune and stromal cells from the tumor microenvironment, to evaluate gene expression in non-tumor cells with data from the single cell RNAseq data obtained from in vivo tumors.

In summary, we identified an SDTG-based antibody binding motif that seems to have biological meaning, given the very strong and high frequency of its recognition by a large proportion of mice cured of the B78 tumor by this immunotherapy regimen. While the translatability of this exact motif to humans might not be given, the approach used to identify this motif could certainly be used as a template for motif identification in human samples. It will be important to investigate this motif further to elucidate the reason why it is recognized so strongly and so frequently, and whether its presence may somehow be involved in the primary, or memory, anti-tumor response.

Acknowledgements and clarification of collaborative efforts from others for the work presented in this chapter

This project started due to the somewhat accidental finding of the motif. When deciding which peptides we wanted to order for ELISA validation of our top co-recognized epitopes, we decided we should make sure that the peptides we were deciding to use were representing completely different antibody-binding epitopes. We therefore took the top peptides based on their high antibody binding and being recognized by antibodies from multiple mice and used a sequence alignment tool (MAFFT) to check for sequence overlap. Much to our surprise, we found that more than half of the peptides shared a four amino acid motif and our quest to solve the meaning of this motif began.

I would like to thoroughly thank Nick Mathers , a former Shapiro student who stayed on working with me throughout his time in medical school, who was the driving force behind the motif discovery. His expertise in tools to utilize for sequence analysis as well as his interest and motivation to dig into this motif helped ot drive this research forward.

Sean McIlwain and Irene Ong were instrumental in the bioinformatic investigation of the motif and their broad knowledge of bioinformatic tools and ways of analysis definitely were a driving force in this project. I would also like to thank Noah Tsarovsky, Alina Hampton, and Arika Feils for all their help with tumor implantations, measurements and serum harvesting. I really appreciate all their help I have received over the years. Another important person for this work was Kaitlin Tetreault who helped us with all the statistical analysis involved in this project. Her ability and knowledge in statistics helped to refine the analysis and resulting data.

Alvin Farrel and John Maris were instrumental in providing us with the DNA sequencing data of the murine B78-D14 and B16 melanoma cell lines. I would like to express my gratitude for their work and help with the DNA sequencing aspect of this project. Anqi Gao and Huy Dinh are analyzing the

single cell RNAseq data and have provided me with expression values for the SDTG-containing genes. I want to thank them for the great cooperation and willingness to help with this.

I would further like to extend my gratitude to Jigar Patel and Brad Garcia from NimbleTherapeutics, Inc. who have always had time for us to meet and discuss and give many helpful suggestions for other aspects to consider and further investigate.

In addition, I want to acknowledge the specific contribution to the manuscript associated with this chapter. I designed the experiments, collected the data, designed, and performed the analysis, interpreted the data, generated the figures, and drafted the manuscript for this study. Sean McIlwain designed and performed the analysis, generated figures, and assisted in writing and editing of the manuscript. Nicholas Mathers assisted in data collection and analysis, specifically the identification of the motif, and will assist in editing of the manuscript. Kaitlin Tetreault helped in the analysis of experiments and editing the manuscript. Richard Pinapati, Brad Garcia and Jigar Patel assisted in consulting about results and implications and will assist in editing of the manuscript. Alina Hampton, Arika Feils and Noah Tsarovsky assisted in collecting the data and will assist with editing the manuscript. Zach Morris assisted in designing the experiments and will assist in editing the manuscript. Amy Erbe assisted in designing the experiments and interpreting the data as well as editing the manuscript. She was also instrumental in the acquisition of the RNAseq data. Paul Sondel and Irene Ong assisted in designing the experiments and analysis, interpreting the data, and editing the manuscript. In addition, I want to thank Anqi Gao and Huy Dinh for their help with the analysis of the scRNAseq data.

The manuscript will be submitted to BioRxiv once we have updated some results with data from ongoing experiments and will be found as:

Hoefges A, Mathers N, McIlwain SJ, Tetreault K, Hampton A, Feils A, Tsarovsky N, Pinapati R, Garcia B, Patel J, Morris ZS, Erbe AK, Ong IM, Sondel PM. Prevalent binding motif in C57BL6 mice cured of B78 melanoma via immunotherapy. In preparation for submission; 2023.

Figures

A

Seq_ID	Probe Sequence	Naïve B2	Naïve A3	Immune PD1	Immune B2	Immune A3
Q99N03;221	PPSYKDVAQGHSSSDT	1.65	2.66	1.76	1.7	1.59
Q99N03;223	SYKDVAQGHSSSDTGR	2.00	3.11	12.51	127.05	1.46
Q99N03;225	KDVAQGHSSSDTGRAL	1.79	3.17	8541.39	13083.01	143.84
Q99N03;227	VAQGHSSSDTGRALAT	2.45	2.64	18115.8	16739.56	690.69
Q99N03;229	Q GHSSSDTGRALATSS	1.61	2.63	8556.49	7722.39	277.27
Q99N03;231	H SSSDTGRALATSSGL	1.85	3.03	7101.62	9103.33	265.77
Q99N03;233	SSDTGRALATSSGLLL	2.57	2.33	9434.99	6604.3	91.18
Q99N03;235	D TGRALATSSGLLLAS	1.78	2.67	1.71	2.59	2.41

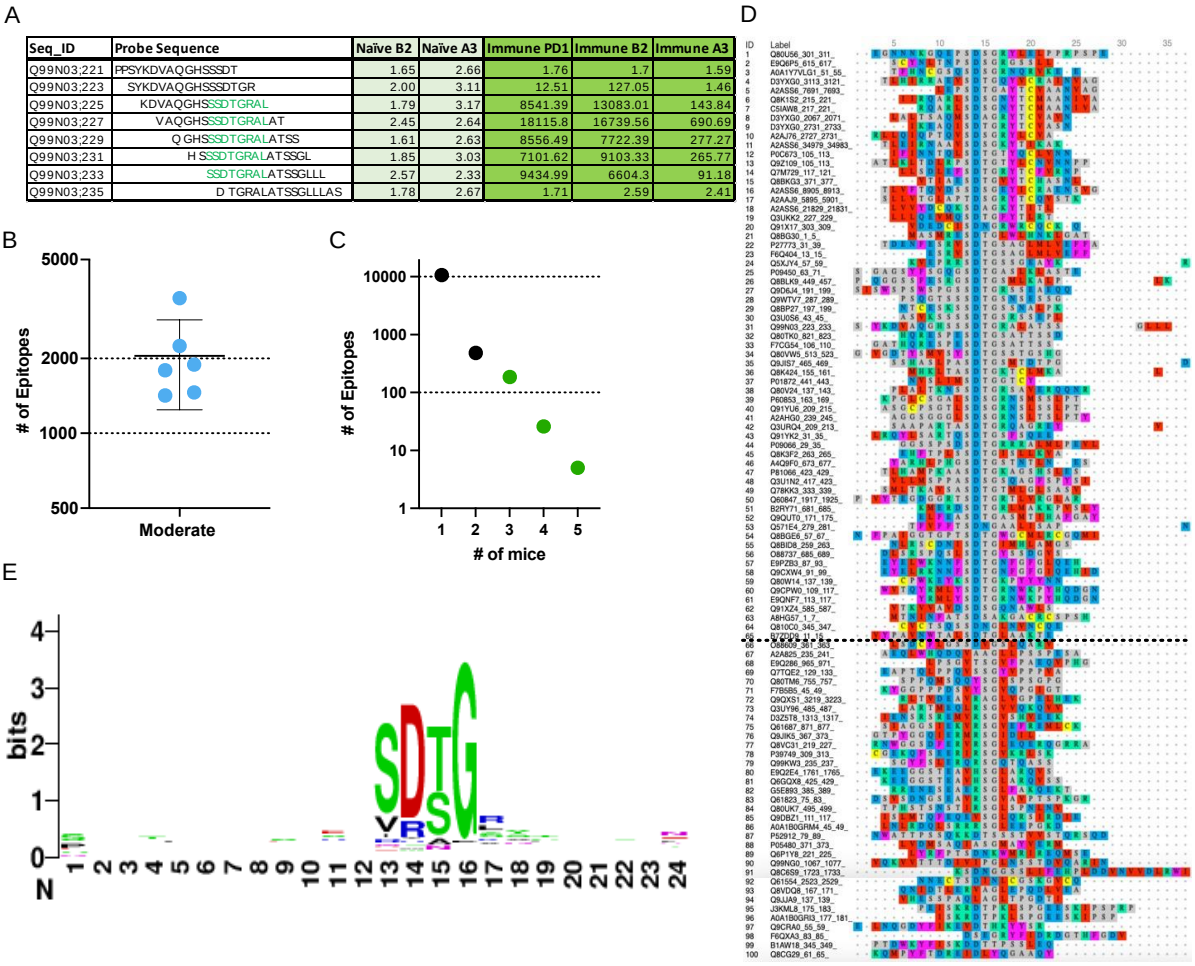


Figure 3.1: Motif identification.

A: Example of raw high-density peptide array binding data highlighting a predicted epitope, defined as a clustered and overlapping antibody binding region in the peptide microarray. A section of the membrane-spanning 4-domains subfamily A member 10 (Ms4A10) protein is shown, with 8 stacked 16-mer peptides, each shifted by 2 aa positions, starting sequentially at

aa position 221 through 235, and thereby collectively representing aa positions 221-251. Fluorescence intensity results are shown for each of these 8 16-mer peptides for separate serum samples from 2 naïve mice (naïve B2 and naïve A3) and 3 immune mice (PD1, B2 and A3). Five of the consecutive 16-mers show strong binding by 2 of the 3 immune sera, while the other 3 16-mers show very weak binding by all 5 sera shown. Immune A3 shows binding to the same 5 peptides but signal strength is significantly reduced in comparison to the other 2 immune mice. The 5 well recognized 16mers each share the 8 sequential aa shown in green font in the probe sequence section, indicating this 8 aa sequence, shared by these 5 peptides is the recognized epitope.

B: Number of epitopes identified in the Immune samples with significantly higher antibody binding (in the moderate threshold category) in immune serum than in Naïve serum samples. Each dot represents the number of epitopes recognized in the moderate category for each of the 6 separate immune mice tested.

C: Number of unique epitopes each recognized by any individual immune mouse, or co-recognized by 2, 3, 4, 5 or 6 immune mice (of 6 total mice) for the number of epitopes in the moderate category. The single dot plotted above the individual numbers plotted on the X axis indicates the number of total distinct epitopes recognized by exactly that number of mice. Epitopes shared by 3 or more mice are highlighted with green dots. For example, 186 epitopes are recognized by 3 mice. That means 186 epitopes total are shared by exactly 3 of the six immune mice tested, different epitopes out of this group of 186 are recognized by different combinations of 3 mice out of the 6 immune mice. For reference of epitopes recognized in the inclusive and restrictive categories, please refer to Hoefges et al 2023, Figure 3.

D: All epitopes co-recognized by at least 3 mice were filtered to include unique epitopes only representing one isoform of each protein. The average immune score of a specific epitope was used to represent the epitope. Epitopes were ranked by highest average immune signal and the top scoring 100 epitopes were aligned to identify amino acid sequences shared or similar between these using MAFFT alignment. MAFFT alignment revealed a 4-aa sequence (with 4 identical or similar amino acids) shared by 65 of the top 100 epitopes.

E: WebLogo representation of the top 100 epitopes aligned in **D**. WebLogo is a software for sequence logo generation which is a graphical representation of an amino acid multisequence alignment. It consists of a stack of symbols for each position in the sequence. Level of sequence conservation at a certain position is displayed by the overall height of the stack while a specific amino acid's conservation and frequency is indicated by the height and size of the specific letter. Length of the WebLogo is 24 positions due to MAFFT alignment aligning the original 16-mer sequences to start and end at different positions. The WebLogo clearly highlights the conserved common motif of SDTG within these 100 epitopes.

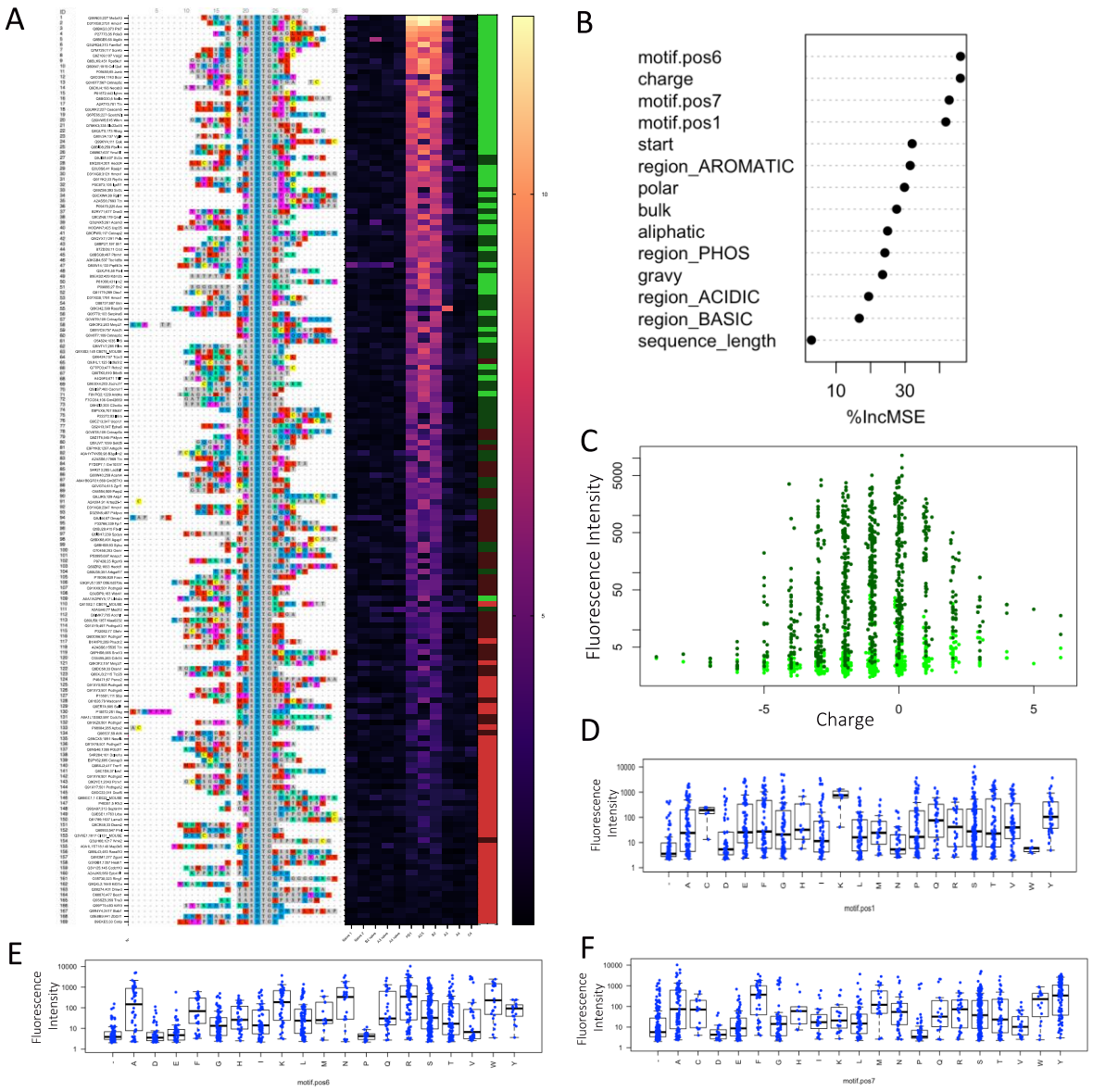


Figure 3.2: SDTG-Motif in the proteome: binding vs. no binding.

A: MAFFT alignment of all 169 unique SDTG-sites present on the whole proteome peptide array is shown on the left. To the right, are the corresponding measured log2-transformed fluorescence intensities for each of these unique SDTG peptides, as detected with all serum samples tested. These serum samples include samples from 5 naïve mice and 6 immune mice. The heat map color scale is linearly related to the strength of the fluorescent signal for that

serum sample tested, ranging from black (weakest) over purple to light yellow (strongest). Last column in the heatmap represents the level at which HERON identified these peptides as being bound, or not bound by the immune sera (light green: restrictive, dark green: moderate, dark red: inclusive, light red: not called).

B: Random Forest model significance matrix showing parameters with influence on signal strength for a specific peptide containing SDTG. Each line on the Y axis is a separate parameter tested for importance of influence on signal strength (see statistical methods). The X axis indicates % IncMSE, defined as % increase in mean standard error, the higher the MSE is, the more influence this parameter has on the model inferring importance of the parameter.

C: Charge as a significant influence on binding: Each light green dot is a fluorescent signal value for naïve serum being tested on 848 unique SDTG-containing peptides. Each dark green dot is a value for immune serum being tested on the same 848 peptides. The strength of the signal is indicated on the y axis, and the electric charge for that peptide is indicated, from -9 to +6 on the X-axis. Those few peptides with extreme charges < -5 or > +4 have low chances of SDTG-antibody binding. Peptides with charges ranging from -4 to +2 show a distribution of peptides with low, moderate, or high immune serum binding. None of the peptides have high binding with naïve sera, regardless of charge.

D: Amino acid position immediately before SDTG (position 1): each column along the X-Axis shows the dots for immune serum binding to all 848 SDTG-containing unique peptides, based on which of the 20 possible aa is in the position immediately before the SDTG. The vast majority of peptides have one of 17 aa in that 1st preceding position, that are associated with a similar distribution of high, medium or low values in each column. A small number of epitopes have a C

or K in this position, and they all have high binding, while a small number have a W in this position, and they all have low binding. These correspond to “motif.pos.1” in **Figure 3.2B**.

E: Amino acid position 6, one amino acid after SDTG: This analysis is similar to that in Fig. 2D, but instead is looking at those epitopes that have the indicated aa on the X-axis in position 6, 1 aa after SDTG in the epitope. Some amino acid combinations show a trend toward either high or low binding while others show a wide range of high, medium, and low binding possibilities. The analyses yield somewhat similar results as seen in **Figure 3.2D**, and this corresponds to the analysis of “motif.pos.6” in **Figure 3.2B**.

F: Amino acid position 7, two amino acids after SDTG: This analysis is similar to that in **Figure 3.2D&E**, but instead is looking at those epitopes that have the indicated aa on the X-axis in position 7, 2 aa after SDTG in the epitope. Some amino acid combinations show a trend toward either high or low binding while others show a wide range of high, medium, and low binding possibilities. The analyses yield somewhat similar results as seen in **Figure 3.2D&E**, and this corresponds to the analysis of “motif.pos.7” in **Figure 3.2B**.

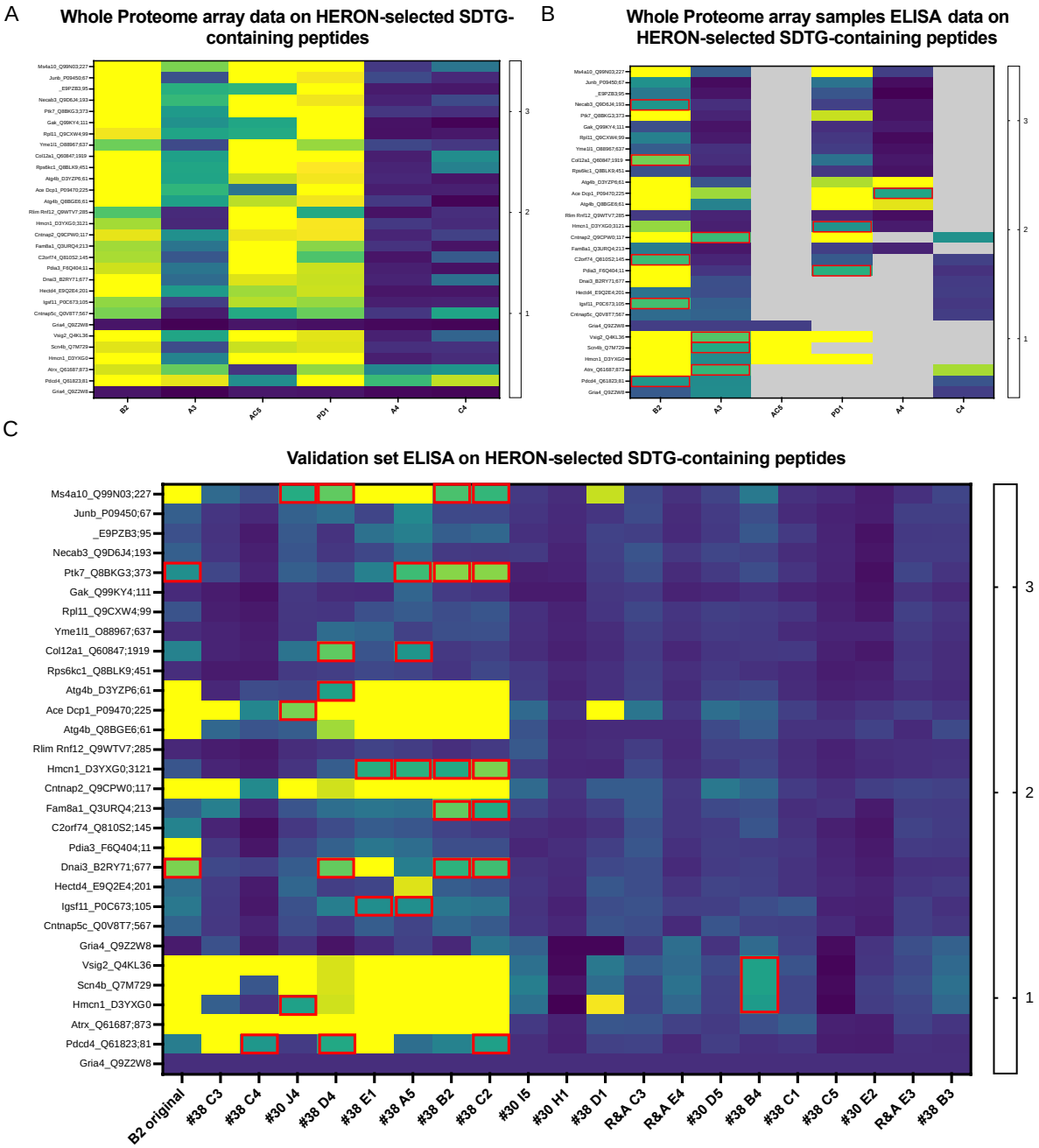


Figure 3.3: SDTG-containing peptides are seen by additional mice and can be validated utilizing ELISA.

A: Heatmap of 29 chosen peptides from the whole proteome peptide array displaying whole proteome peptide array data using the serum samples from the 6 individual immune mice used in

the original peptide array (Hoefges et al., 2023). 26 of these 29 peptides contain SDTG in their sequence and were chosen based on significant binding (restrictive or moderate category) by serum from at least 50% ($\geq$ 3 of the 6 mice tested) of the immune samples tested in the whole proteome array. Two peptides (Artx_Q61687;873 and Pdcd4_Q61823;81) did not contain SDTG in their sequence but were recognized by 5 of the 6 individual immune mouse serum samples in the whole proteome peptide array, were also included, as was 1 peptide that was chosen as a negative control peptide (Gria4_Q9Z2W8) as no binding was observed to it by any of the serum samples tested in the whole proteome peptide array (from naïve or immune mice). Data shown are log10 of the fluorescence units of the peptide array signal.

B: Heatmap of ELISA results using the same peptides and serum samples as in **Figure 3.3A**. Gray areas indicate the 4 serum samples that had been previously depleted. To enable testing on all 29 peptides, only some peptides were tested with these; the untested combinations are shown in gray. Serum samples B2, A3, PD1, A4 and C4 were cryopreserved serum samples from the same exact aliquot used in the proteome array (Fig. 3A). For some of the original samples no serum or almost no serum remained (AC5, PD1, A4 and C4, grey areas indicate peptide not tested). Data shown are normalized to the average value of the negative control peptide (Gria4) for each serum sample. Red boxes are applied to all values above 2 and below 3 to highlight significant increases over background to determine positive binding values.

C: ELISA data for the same peptides as in **Figure 3.3A&B** but using independent immune mouse serum samples never tested before from 20 separate mice that received the same treatment to cure their B78 cancer. Also included here is a repeat immune serum sample from one of the 6 immune mice used in the original whole proteome samples, as an internal control (mouse B2, also shown in in **Figure 3.3A&B.** Data shown are normalized to the average value of the negative control peptide for each serum sample. Red boxes are applied to all values above 2 and below 3 to highlight

significant increases over background to determine positive binding values. Values > =3 are displayed in yellow and easily recognizable as positive.

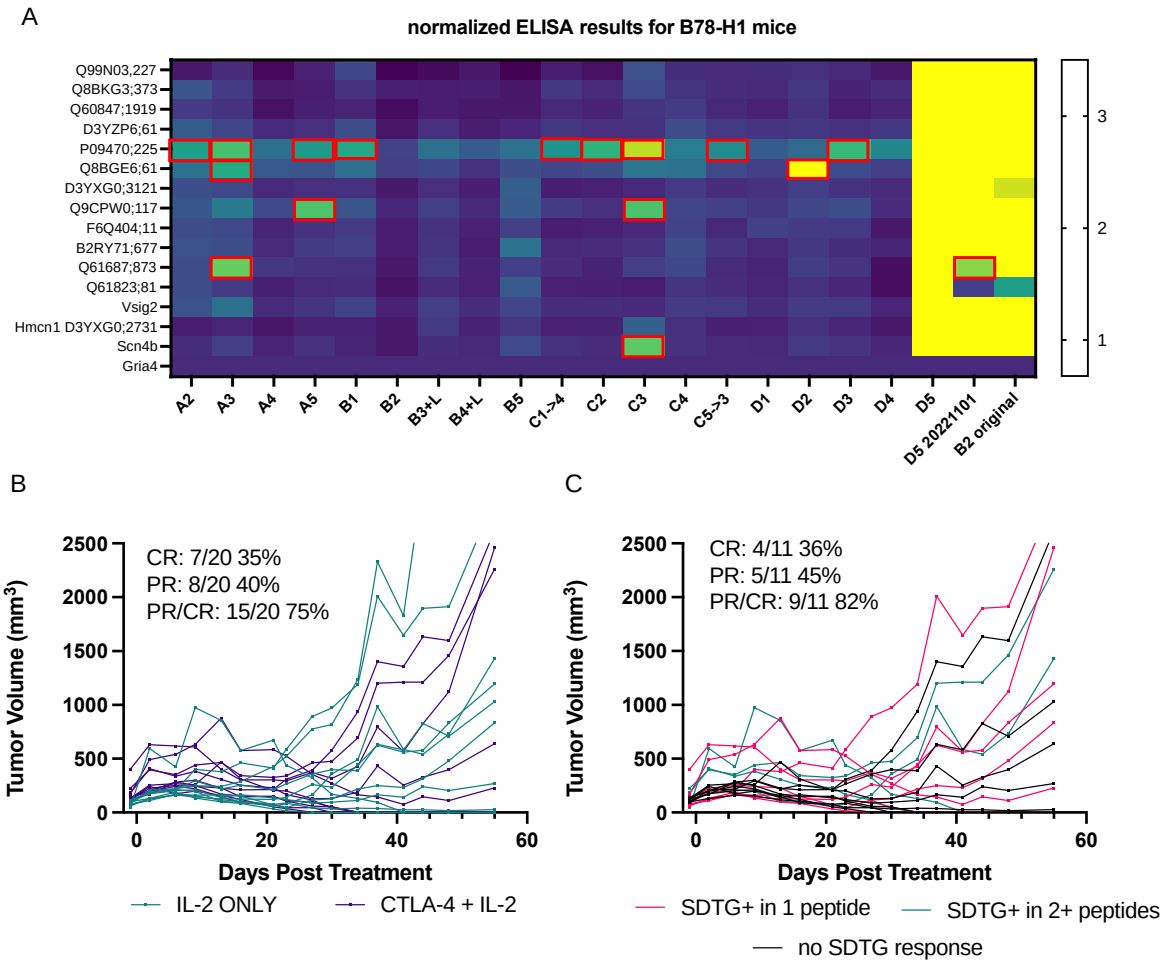


Figure 3.4: GD2 expression on the tumor is not required to induce antibody to SDTG.

A: ELISA data for 13 SDTG-containing peptides and for 2 peptides not expressing SDTG, but showing reactivity with 5 of the 6 initial serum samples used in the peptide array, as well as 1 negative control peptide (Gria4) (as described for **Figure 3.3**) are shown. The 20 mice tested here were injected with B78-H1, a B78-tumor cell line lacking GD2 expression. Post-treatment serum samples from 19 individual mice were available as collected on day 55, 9 treated with radiation and IL-2, and 10 treated with radiation, IL-2 and anti-CTLA4. One of the mice (D5), showed potent recognition of

all peptides tested, and a separate timepoint from this same mouse (labeled D5 20221101, was also tested in parallel and showed similar reactivity. Mouse D5 was treated with radiation and IL2 alone. Also included here is a repeat immune serum sample from one of the 6 immune mice used in the original whole proteome samples as an internal control (from the same timepoint from mouse B2, also shown in **Figure 3.3 A, B** and **C** labeled here as B2 original). Data shown are normalized to the average value of the negative control peptide (Gria4) for each serum sample. Red boxes are applied to all fold-change values above 2 and below 3 to highlight significant increases over background to determine positive binding values. Values > =3 are displayed in yellow and easily recognizable as positive.

B: Tumor growth curves for all 20 animals with B78-H1 flank tumors are shown starting the day before treatment. Serum was harvested at day 55 post treatment. Complete responses (CR), partial responses (PR) and the combination of partial and complete responses are shown as a fraction of the 20 mice evaluated as well as by percent.

C: Tumor growth curves of the same animals in B are shown divided into mice with reactivity to no SDTG peptides (8 mice) , reactivity to only one SDTG peptide (7 mice), or reactivity to more than one SDTG peptide (4 mice), as shown in **Figure 3.4A**. Complete, partial and a combination of complete and partial responses are noted as absolute numbers as well as percent for those 11 mice with at least 1 SDTG peptide detected.

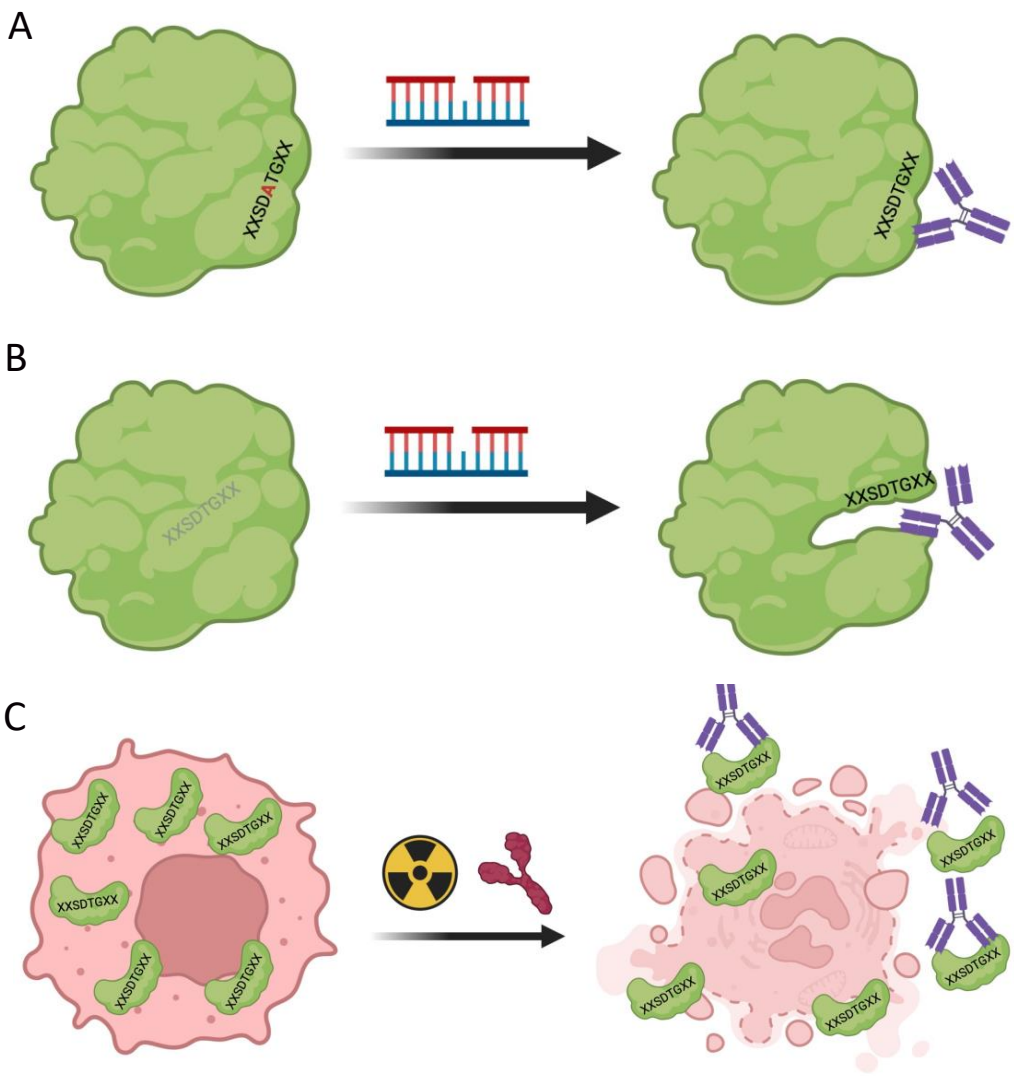


Figure 3.5: Hypotheses to elucidate strong and frequent SDTG-response:

A: A mutation in B78 created an additional SDTG-site in a protein that is now being recognized by the immune system as foreign.

B: A mutation in an SDTG-containing protein caused a conformational change which now makes the SDTG-site accessible and recognizable for the immune system.

C: SDTG-containing proteins are upregulated in B78 vs. normal cells and are being recognized through immunogenic cell death.

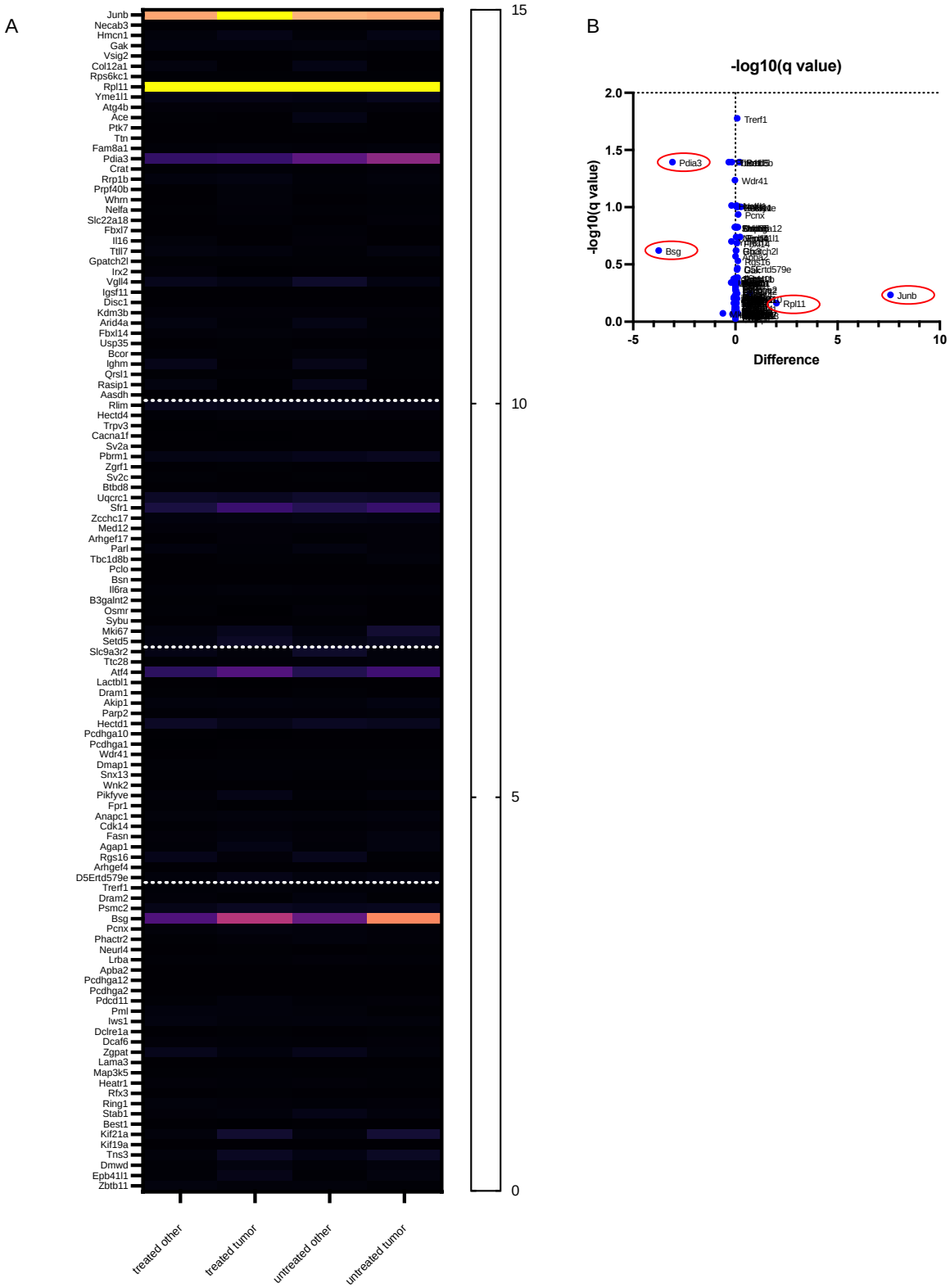


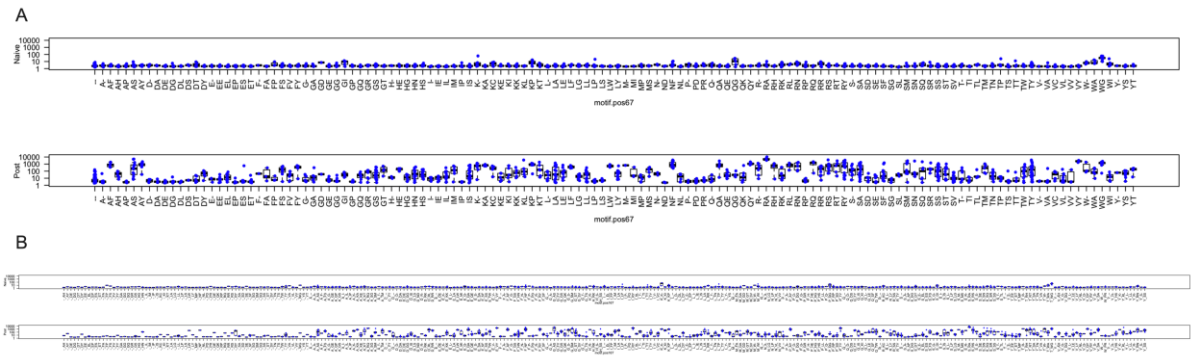
Figure 3.6: Differential expression values of mRNA measures of SDTG-containing proteins.

Figure 3.6: Differential expression values of mRNA measures of SDTG-containing proteins.

A: Heatmap of scRNAseq data from tumor homogenates of 2 treated and 2 untreated mice taken on treatment day 8 (RT given on day 0, intratumoral IC injections given on days 5-8). Populations are divided into malignant B78 tumor cells (tumor) and normal non-malignant normal stromal cells (other). The population labeled as other consists of epithelial cells, stromal cells, myeloid cells, B cells, NK cells, and T cells. Medians of each of these populations divided by treatment were used. X axis shows different cell types, the Y axis displays a list of 115 SDTG-containing genes ordered by epitope detection category based on HERON calls starting with restrictive (top 38, until first dotted white line), 24 in moderate category (until next white dashed line), 23 in the inclusive category and 30 that were not called by HERON. Expression values are shown ranging from 1 to 15 as a heatmap from black over purple to light yellow. Expression values above 15 are shown in bright yellow.

B: Volcano plot of the values displayed in **Figure 3.6A** for treated vs. untreated tumor samples. Red circles highlight genes with differential expression as identified using a two-way ANOVA utilizing data displayed in **6A**. The X axis plots the difference between means. A dotted grid line is shown at X=0, no difference. Y value plots the minus logarithm of the q ratio, using the method of FDR. A dotted grid line is shown at Y=-log(Q).

Supplemental Figures

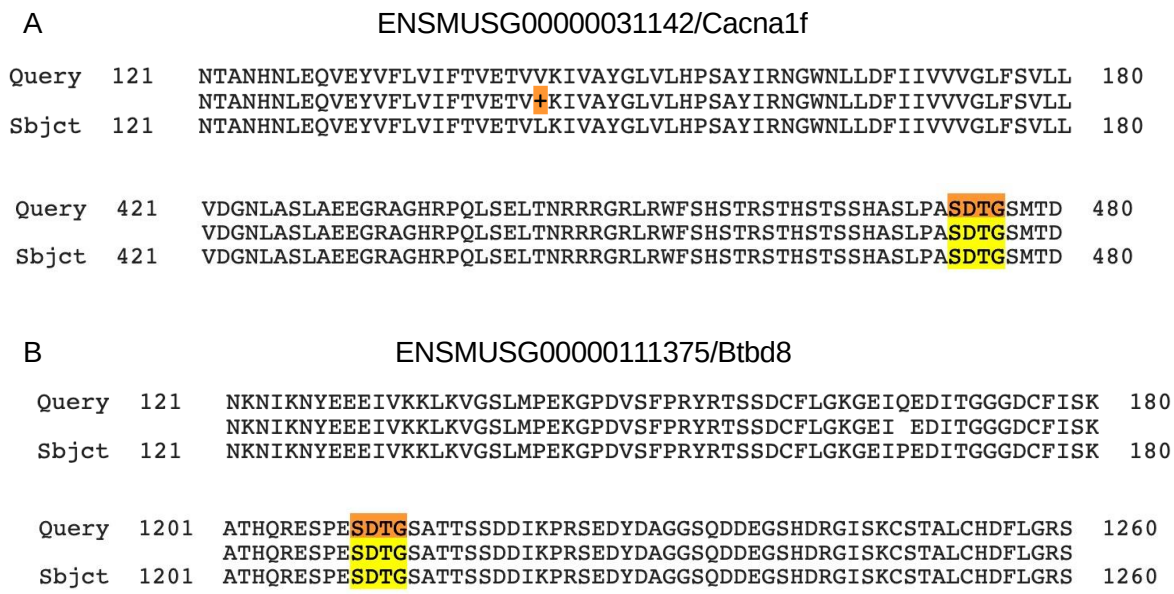


Supplemental Figure 3.1: Amino acid position surrounding SDTG:

A: aa position 6 & 7, the two positions immediately following SDTG. Each column along the X-Axis shows the dots for immune serum binding to all 848 SDTG-containing unique peptides, based on all represented combinations of which of the 20 possible aa is in the 1st position immediately after the SDTG and which is in the 2nd position immediately after the SDTG. The top graph shows naïve samples, the bottom graph shows all immune samples. Almost all naïve samples show low signal in all possible combinations. Immune samples show low or high binding depending on the combinations, with some showing high and low binding across different peptides with the same combination of aa in position 6 & 7 after the SDTG.

B: Positions right before as well as 1st and 2nd position after SDTG (position 1, 6, & 7) are shown in combination. Each column along the X-Axis shows the dots for immune serum binding to all 848 SDTG-containing unique peptides, based on all represented combinations of which of the 20 possible aa is in each of the 3 indicated positions, namely, the position immediately before SDTG, the 1st position immediately after the SDTG, and the 2nd position immediately after the SDTG. The top graph shows naïve samples, the bottom graph shows all peptides with mean values of immune samples. Almost all peptides with mean values based on naïve samples show low signal in all

possible combinations. Mean values of immune sample peptides show low or high binding depending on the amino acid combinations, with some showing high and low binding across different peptides with the same combination of aa in position 1, 6, & 7 surrounding the SDTG.



Supplemental Figure 3.2: DNA sequencing results with mutations in SDTG-containing proteins.

A: Protein sequence alignment of B78 DNA sequencing data translated to a protein sequence vs. C57BL/6 protein sequence for Cacnaf1 is shown. A single nucleotide variation caused a change at aa position 124 with a V (Valine) to L (Leucine), both are aliphatic. SDTG is located at aa position 472.

B: Protein sequence alignment of B78 DNA sequencing data translated to a protein sequence vs. C57BL/6 protein sequence for Btbd8 is shown. A single nucleotide variation caused the change of Q (Glutamine) to P (Proline) at aa position 166. Q is amidic, P is aliphatic.

Chapter 4: Discussion, Conclusions, & Future directions

Overview

In this chapter I summarize the findings and conclusions from chapters 2 and 3, discuss additional scientific questions that arise, as well as outline the major implications of this work. I also give a brief overview of additional findings I did not present in chapters 2 and 3. Finally, I discuss possible ways of addressing in future and ongoing experiments some questions so far left unanswered.

Summary of thesis findings

Antibody landscape of C57BL/6 mice cured of B78 melanoma via immunotherapy

Here we established a method to utilize a high-density overlapping stacked array of 16-mer peptides representing the entire C57BL/6 proteome to identify the linear "immunome" of epitopes recognized by antibody induced in mice after receiving curative immunotherapy that is associated with complete and durable eradication of B78 melanoma tumors and induces tumor-specific immune memory. We utilized the HERON analysis method (Appendix A), developed by us specifically to analyze these whole proteome peptide array datasets of immune and naïve mice, and isolated peptides, epitopes, and proteins that showed a significant increase in antibody binding compared to the naïve samples.

By utilizing epitopes, rather than single peptides, we increased the reliability and reproducibility (**Figure 2.2B & 2C**) as measured by repeat samples. HERON requires a degree of similar recognition of the related epitopes, but does not require that all probes that are part of the epitope are also recognized by another sample to indicate that the same epitope has been recognized in an additional sample. The same principle applies to protein data, where recognition of the same protein was given if the individual mouse sample showed antibody binding to any region of this

protein. By not requiring the same or similar region of the protein to be recognized, we achieved a higher number of proteins that were recognized by 4, 5 or even 6 mice than when comparing to epitope data (**Figure 2.3C & 4C**).

Separately, we showed that peptides that showed antibody binding in the whole proteome peptide array, in most cases, also showed antibody binding in a separate peptide ELISA as well as a smaller scale peptide array (**Figure 2.5**). Furthermore, utilizing additional mice that had not been evaluated using a peptide array, we found that 65% of the mice in the validation cohort recognized at least one of the peptides tested that was previously recognized by 3 of the original 6 mice (**Figure 2.7D**). Having any of these peptides recognized by additional mice supports the biological importance of these proteins being antibody targets by multiple mice in our system. To enhance the importance of these peptides being recognized by additional mice, we also selected random peptides from the whole proteome dataset and tested these on the same validation set of mice. In the randomly selected peptides we only achieved a 0.05% rate of positive reactions, while the HERON-selected peptides yielded 20% positive reactions. This indicates that the HERON method used to select peptides from the whole proteome peptide array data successfully identifies peptides that are also being recognized in validation assays with the same serum samples at a significantly increased rate over randomly chosen peptides ($p<0.001$).

More importantly, due to the random rearrangement of V-D-J immunoglobulin gene components, a large diversity in antibody repertoire, even in a group of genetically identical mice, was expected. The ability of HERON to identify peptides based on their recognition by an initial set of mice using the original dataset and the same peptides being subsequently recognized by serum from additional mice drives the point of immunologic importance of these proteins for mice of the same strain immunized to the same B78 tumor using the same immunotherapy regimen.

In summary, we described the immunologic methods used to obtain data and validate it using additional small-scale peptide arrays and peptide ELISA systems. We showed the spectrum of peptides, epitopes and proteins recognized and described what fraction of targets recognized by at least one immune mouse are also recognized by some other mice, despite the stochastic nature of each mouse's individual B-cell repertoires.

We showed that this peptide array technology and analysis method can be used to detect the linear antibody-recognized "immunome" of sera from mice immune to B78 (and B16) melanoma through RT + IC immunotherapy. Proteins recognized by sera from multiple immune mice may potentially be of importance in achieving immunity to the cancer or could potentially be used as a biomarker of a potent adaptive response to the cancer.

Prevalent binding motif in C57BL/6 mice cured of B78 melanoma

Here we described a four amino acid motif identified in many strongly recognized epitopes from immune sera of mice cured of B78 melanoma utilizing a RT + IC immunotherapy regimen that acts as an *in situ* vaccine.

The motif was identified when focusing on all epitopes that were recognized via antibody from at least three of the six original immune samples that were tested on the whole proteome. We took all epitopes recognized by at least three mice and ordered them based on average immune signal from highest signal to lowest. When using the top 100 epitopes (highest immune signals), 65 of these showed the four aa motif (**Figure 3.1**). This motif was found in over 50% of the top epitopes from a whole proteome antibody binding screening of epitopes that were co-recognized by at least 3 of the 6 separate immune mice, but not by the naïve sera. This observation suggests that this motif is recognized during an immune response to the B78 tumor for many B78-bearing mice. We identified 310 separate SDTG-containing regions of which 194 were unique epitopes that did not belong to an

isoform of the same gene (**Figure 3.2**). Within these 194, we found 114 of these with an average immune binding score of over 100 and 42 with an average immune binding score of under 20. This showed that while most SDTG-containing regions were recognized via an antibody response in some of these mice, some SDTG-containing epitopes were not. This generated the question: what besides the presence of SDTG determines binding? To elucidate this, we used a random forest model and discovered that charge, as well as the amino acid position before SDTG and two positions after SDTG, does play a role. However, none of these findings by themselves gave a clear answer (**Figure 3.2**) and we are investigating further.

We tested an additional validation set of 20 immune mice for reactivity to SDTG-containing peptides and found 9 mice with strong reactivity towards at least 1 SDTG-containing peptide (45%, **Figure 3.3**) and several of these 20 mice recognizing multiple SDTG-containing peptides. One possible cause of SDTG-reactivity could have been the transfection of the original GD2-negative line with a plasmid needed to integrate GD2 and GD3 synthase into the B78 tumor cell genome to cause expression of GD2 on the tumor cell surface. If these genes, or a translated portion of the plasmid, contained an SDTG sequence that was "foreign" to these mice, this could have induced antibody to the foreign proteins, that might have created some anti-SDTG antibody. To test for this, we aquired the "original" B78 cell line, referred to as B78-H1 that was not transfected to express GD2, and treated mice bearing B78-H1 tumors with radiation and IL-2 or with radiation, IL-2 and anti-CTLA4, in an effort to cure them of their B78-H1 tumors. We then harvested serum from all 20 treated mice and tested for reactivity against SDTG-containing peptides. We found 1 mouse with strong response to SDTG and a few with weaker but detectible SDTG-reactivity (**Figure 3.4**). With the ability to generate an SDTG-response in these mice, we have shown that the transfection of the plasmid needed to integrate GD2 and GD3 synthase into the B78 genome was not required to induce anti-SDTG antibody.

We performed DNA sequencing on B78 and B16 cells to search for mutations as possible mechanisms for a strong and frequent SDTG response. We hypothesized that a mutation could lead to an additional SDTG sequence being present in a protein or, that a protein that contained an SDTG naturally, could aquire a mutation that resulted in a conformational change leading to exposure of SDTG (**Figure 3.5**). However, we only found 2 genes (Cacna1f and Btbd8) containing SDTG that had mutations at different parts of the protein. These mutations were not predicted to have an influence on the protein structure and cause exposure of the SDTG motif.

We then went on to RNAseq analyses of the B78 tumor, where we found four genes that were differentially expressed between tumor and cell types found in the tumor microenvironment (**Figure 3.6**).

In summary, we identified a four aa motif that most likely has immunologic relevance; antibody against it is present in a large proportion of mice cured of B78 tumor via in situ vaccine treatment (RT + IC). The exact meaning and origin of this motif are yet to be fully elucidated, but the workflow used here to identify this motif could certainly be applied to human data.

Summary of additional findings not mentioned in the previous chapters

In addition to the work presented above, we have collaborated with researchers from the Children's Hospital of Philadelphia (CHOP) who have developed an analysis pipeline utilizing surface proteomics and target prioritization. The Maris lab at CHOP focuses on neuroblastoma research. While this thesis is focused on melanoma, and there are many fundamental differences between melanoma and neuroblastoma, both of these tumors are from neuroectodermal origin and generally express GD2 on their surface. Amber Weiner PhD, a postdoc in the lab of John M. Maris, generated a neuroblastoma surfacesome atlas. To do this, she performed plasma membrane protein extraction utilizing a sucrose density gradient methodology followed by nano-liquid chromatography

coupled to mass spectrometry (nLC-MS/MS) on 9 neuroblastoma cell lines as well as 12 patient-derived xenografts (PDX). She then integrated the MS data with RNA-sequencing data from 153 neuroblastomas and 7859 normal tissues to evaluate proteins that had an annotated extracellular domain that was differentially expressed in neuroblastoma compared to normal tissues. She identified 4826 unique membrane proteins and confirmed known neuroblastoma surface proteins in development as immunotherapeutic targets like ALK, GPC2, NCAM1, DLL3 and CD276. Utilizing her dataset of the human neuroblastoma surfaceome, we screened all of these neuroblastoma surfacome proteins to identify those that are human homologues of the mouse proteins that showed strong antibody binding by the whole proteome array data we generated using the sera from the 6 originally tested mice that had been immunotherapeutically cured of the B78 melanoma. Focusing on the mouse proteins with strong antibody binding (identified in our restrictive category), we were able to map the human homologues of 71 of them to human UniProt IDs. Of these, 52 were unique gene names. After intersecting these 52 proteins with the neuroblastoma surfaceome and target prioritization, we found five proteins that were also enriched in the neuroblastoma surfaceome data. These five proteins were Hectd4, Cntnap2, Ace, Ptk7 and Pdia3 (**Figure 4.1**).

The top hit based on abundance in neuroblastoma was Pdia3 (**Figure 4.1**), a protein disulfide-isomerase. It is well studied and showed some potential as a biomarker in multiple cancer types (Wang et al., 2017; Zhang et al., 2020; Zhang, Li, et al., 2022; Zhang, Wang, et al., 2022). Ptk7, also highly abundant on the neuroblastoma surfaceome (**Figure 4.1**), called the inactive tyrosine-protein kinase, is a tyrosine kinase involved in the Wnt signaling pathway. It functions in cell adhesion, cell migration, cell polarity, proliferation, actin skeleton reorganization, and apoptosis and also has a role in embryogenesis, epithelial tissue organization and angiogenesis (Golubkov et al., 2010; Meng et al., 2010; Prebet et al., 2010; Puppo et al., 2011). Recently, an anti-PTK7 antibody as well as CAR T-cells targeting PTK7 have been developed and are undergoing testing as possible therapeutics

(Damelin et al., 2017; Jie et al., 2021; Kim et al., 2022). Ace, angiotensin-converting enzyme, plays a key role in the regulation of blood pressure, electrolyte homeostasis or synaptic plasticity (Anthony et al., 2010; Yang et al., 1970) and is also very abundant in the neuroblastoma surfaceome where it is highly recognized via antibody as detected in the whole proteome peptide array of mice cured of B78 melanoma. Cntnap2, contactin-associated protein-like 2, is associated with several neurodivergent disorders, like autism spectrum disorder, epilepsy, ADHD, schizophrenia, and intellectual disability (Choe et al., 2022; de Jong et al., 2021; Penagarikano et al., 2011; Poot, 2015).

Hectd4, an E3 ubiquitin-protein ligase, showed relatively low abundance on the neuroblastoma surfaceome abundance scale, but was still enriched (**Figure 4.1**). Certain Hectd4 variants are associated with type 2 diabetes and neurodevelopmental disorders (Faqeih et al., 2023; Lee et al., 2022; Sun et al., 2021). Hectd4 is involved in Myc stability (Vatapalli et al., 2020). Components of this pathway could potentially be targeted and used as therapeutic targets for select cancers. This data showing that these human proteins are homologues of mouse proteins that are strongly recognized by antibody from mice cured of B78 further supports pursuit of these as potential biomarkers or therapeutic targets.

We also utilized a different approach from the Maris lab to identify possible new therapeutic targets. In this approach Mark Yarmarkovich PhD, a previous post-doc in the Maris lab who now has his own lab at NYU, developed a tumor discovery workflow to identify peptides presented via MHC in neuroblastoma tumors (Yarmarkovich et al., 2021). He purified MHC and eluted peptides from patient-derived xenografts and primary neuroblastoma tumors and characterized peptide sequences by tandem mass spectrometry. Exome sequencing of these tumors was also performed. Utilizing the exome sequencing data of the tumors, HLA types were inferred using PHLAT (Bai et al.,

2014). Predicted MHC binding affinities for eluted peptides were obtained using NetMHC. Differential gene expression was assessed via comparison of 153 neuroblastoma cases to the GTEx database of normal tissue gene expression. The differential gene expression was used to derive antigens and compare these against a database of MHC presented peptides from 190 healthy tissues. This resulted in the discovery of 83 neuroblastoma antigens that had not been observed in healthy tissue. With Mark's help, we utilized his MHC-peptidome dataset for neuroblastoma and filtered the mouse homologues of these to search for any that were identified as mouse proteins highly recognized by our immune mouse serum samples. We identified some overlapping genes that were recognized in his neuroblastoma MHC peptide dataset that were also recognized in our whole proteome peptide array dataset of antibody recognition by mice cured of their B78 melanoma. Within our restrictive dataset, we identified a total of 33 proteins that are also in his list of 1980 human proteins with peptides found presented by HLA on human neuroblastomas; we are further investigating these. One of the issues with the comparison of this dataset to ours is the gene translation from mouse to human as well as the different HLA subtypes present in humans that are not translatable to mouse data. However, we are looking further into the recognized proteins, because their dual recognition through these 2 detection systems would seem to increase the likelihood of their biological/immunological importance.

In addition to these collaborations, we looked more closely into several other proteins. One of these being Hemicentin 1 (Hmcn1, D3YXG0). Hmcn1 was strongly recognized by the 4 of the 6 mice in the original peptide array with another mouse with weak recognition. This protein has multiple antibody-recognized epitopes, 4 of them contain SDTG sites, 2 of them with high binding and 2 with moderate to low binding. We validated 2 of the Hmcn1 SDTG-containing peptides as being recognized via ELISA by additional mouse serum samples (**Figure 3.3**). We furthermore detected strong antibody staining for Hmcn1 in B78 and B16 cells (**Figure 4.2A**) as well as B78 tumors

removed from tumor-bearing mice and skin of C57BL/6 mice (**Figure 4.2B&C**), evaluated via IHC-IF (immunohistochemistry-immunofluorescence). Looking at RNAseq profiles for Hmcn1 expression across tissues, RNA for this protein is expressed in almost all tissues as well as most cancers. However, Hmcn1 is highly overexpressed in human melanoma tumors (**Figure 4.2D**). Hmcn1 has been found to have negative effects on outcome if cancers have a higher expression of Hmcn1 (Gong et al., 2022; Kikutake et al., 2018; Liu et al., 2019; Wen et al., 2022). We are further investigating whether the presence of antibodies against Hmcn1 could potentially be used as biomarkers.

Additional work in progress, future directions and scientific conclusions:

Work in Progress

While we have characterized the antibody landscape of C57BL/6 mice cured of B78 melanoma via *in situ* vaccine (RT + IC), we have a number of questions yet left unanswered. Here I discuss some of the ongoing and future experiments we are proposing to elucidate some of the unanswered questions.

One of the first things we plan to investigate further is whether there is an association of antibodies against select proteins that are only present if the animal has a complete response. So far we have focused our analysis on naïve and immune samples, i.e. samples that had never seen a tumor in comparison to serum samples of mice that successfully rejected a tumor and exhibited memory against the same or similar tumor type. Our next step will be to test the peptides that were recognized by the highest number of individual immune samples (combining original data with ELISA data from validation set mice) on mice that have a tumor but did not receive treatment, as well as on mice that did receive treatment but failed to clear the tumor. In chapter 3 we took a first look at mice that did not fully reject their initial B78 (in this case B78-H1) tumors; in that limited study we

were not able to see differences in anti-tumor response based on the pattern/level of antibody detected against SDTG peptides tested. However, with a larger number of mice and by using the original B78-D14 tumor (which does have GD2 surface expression), as well as focusing on select peptides, we might be able to discern differences in recognition frequency by antibody, based on whether the mice were immunotherapeutically cured, or not. One factor that needs to be controlled for in such studies is that mice that showed a response after receiving treatment, but were unable to clear their tumor fully, might still show the same antibody generation as seen in mice that were able to clear their tumor fully, as an immune response against the tumor took place, but in the end, was not strong enough to fully clear it. For this experiment, we are currently collecting more serum samples to test untreated, treated but not cured, as well as mice that were treated and cured; sera from these 3 groups will be tested and compared via ELISA on select peptides. I plan to conduct these experiments early in summer 2023.

We hope that the workflow we developed here to identify important cancer antigens can be transferred to humans and other mouse cancer models. While we don't necessarily expect that all our identified proteins will translate to human data, the workflow developed here could easily be applied to human serum samples of different cancer types and help identify additional targetable proteins or antibodies as potential biomarkers of response to therapy. We currently have an ongoing clinical trial for advanced melanoma at UW Madison (UW16134) investigating the same utilized *in situ* vaccine regimen (RT + IC) in melanoma patients that was used in the mice in this thesis. We are collecting their serum at multiple timepoints. So far this trial has enrolled seven patients. Due to the large diversity across humans, a larger sample size will be necessary to perform statistical analysis and possibly reach significance, when trying to see whether there may be any association with detectible antibody in the serum of these patients after treatment, and their likelihood of responding to the therapy with a clinical anti-tumor response.

Another question we are currently investigating is what caused the strong antibody response against SDTG-containing peptides in the first place?

To investigate this further, we are working on DNA sequencing analysis of B78, B16 and the normal C57BL/6 mouse genome. While we were not able to find genetic alterations leading to generation of SDTG so far (point mutations), we are still working on analysis concerning more complicated alterations like chromosomal inversions, deletions and duplications as well as copy number variations. To complement the DNA sequencing, we conducted RNAseq on a single cell level of *in vivo* extracted tumors as well as bulk RNAseq of B78 cells in culture. For the scRNAseq we used 2 tumors from untreated animals and 2 tumors from treated animals. For the bulk RNAseq data, we treated some B78 cells with radiation and left others untreated. We also included healthy C57BL6 skin lysates as our control. With these data we hope to see changes in genes that contain the SDTG motif in the tumors compared to the normal skin. Some of this work is detailed in **Figure 3.6**. However, we hope that the bulk RNAseq data will provide more information about B78 vs. healthy skin as well as the differences caused by radiation in B78 gene expression without the confounding effects of the presence of the cells (and their RNA) from the non-tumor cells in the tumor microenvironment.

Future Directions

Another possible place SDTG could have originated from to induce an antibody response is the microbiome; if so what we are actually detecting with the ELISA and peptide array may be cross-reactivity of microbial proteins and normal mouse proteins. If the bulk RNAseq as well as scRNAseq data do not show any significant differences in any SDTG-containing genes between tumor and normal tissue, we will look closer into the microbiome as a possible source of inducing antibody reactivity to SDTG. To do this, we will utilize a blast peptide search against all common known

bacterial species present in C57BL/6 mice. In conjunction, we will conduct 16S rRNA sequencing (Johnson et al., 2019) to capture the microbiome present in mice with SDTG-recognition in comparison to mice that did not develop antibodies recognizing SDTG (but have also been treated against B78 tumors and rejected the tumor with memory). We hypothesize that an increase in a specific bacterial population that contains the motif SDTG in one of its proteins might be the cause of SDTG recognition and subsequent SDTG-antibody generation.

It could also be possible that the original site of SDTG-recognition was from a viral protein. To investigate viral proteins and possible viruses that these mice could have been infected with, we would again start with a blast search for SDTG in mouse virus proteins and then move on to sequencing.

Another approach would be to focus on the antibody sequence and B cells. Characterizing the sequence of the antibodies generated, specifically the hyper variable regions, may help assess antibody targets. Contrasting the antibody repertoire sequencing of antibody producing cells from: a) mice before tumor implantation (as a baseline and to filter out tumor-unrelated antibodies), b) mice after tumor implantation but before treatment (to focus on antibodies associated with B78 melanoma) and c) after mice have been cured and exhibit memory (to compare to the other 2 groups to filter out differences associated with a positive anti-tumor response) might help us uncover the original target that the SDTG antibodies are directed against. The technology for antibody repertoire sequencing is rapidly improving and allows for ranking and robust detection of antibody sequences (Choi et al., 2023; Greiff et al., 2014). We might be able to screen these sequences for potential binding ability to SDTG-containing peptides. To answer the question why only some SDTG-containg proteins showed strong antibody binding and others did not, we could develop monoclonal antibodies that recognize the top 5 SDTG-containing peptides based on antibody binding scores from the whole proteome array and ELISA, and use as “controls” the

bottom 5 SDTG-containing peptides based on fluorescent signal from the whole proteome peptide array. We would hope to generate at least one antibody that strongly recognizes the top 5 peptides and not the bottom 5 peptides. This antibody and its interactions with the top 5 peptides, vs. the bottom 5 peptides could be studied using structural biology.

There are different ways to generate these antibodies: 1) We can use the peptides to isolate B cells from mice that are recognizing SDTG and generate a hybridoma cell line from these B cells.

2) We can introduce these peptides to mice or rabbits to generate B cells recognizing a specific set of peptides, then generate a hybridoma of the B cells from these mice (or rabbits), select for successful hybridoma cells and test via ELISA for specificity to the 5 peptides (Zarei et al., 2015).

3) We could use phage display to generate binders that bind well to our top 5 peptides and not the bottom 5 peptides. 4) We can use the sequencing information we got from the antibody repertoire sequencing and generate recombinant monoclonal antibodies which is proposed to be quicker and cheaper than the traditional generation of antibodies and allows for easy modifications (DeLuca et al., 2021). Once we have the separate antibodies, we will be able to test them against all SDTG-containing peptides we identified. With these antibodies we will have the possibility of conducting multiple experiments and reproducibly use the same source of antibody for characterization reasons. However, generation of monoclonal antibodies is laborious and time consuming. Instead, we could also utilize mice that developed anti-SDTG antibodies and pull out these specific antibodies from the serum utilizing the peptides we know are bound. After we have a somewhat “pure” anti-SDTG antibody sample, we can use this antibody as a tag. Tumor, skin, or microbiome lysates could be probed against the SDTG-targeting antibody. Everything bound to the SDTG-antibody can then be analyzed using mass spectrometry to identify the specific target. Furthermore, the antibody will help us in characterization of biophysical aspects including its affinity to its target and solvent

accessibility. If the affinity is very low, what we have identified is most likely not the primary target or its native antigen. We recognize this latter approach might not have the discriminatory capability, or analytic power of the mAb approach, and thus is a "2nd-line" option.

Utilizing mass spectrometry to identify the native antigen of the anti-SDTG antibody will also help us identify if the native target was a linear sequence containing SDTG, or if it actually targets a conformational epitope. Up until this point, i.e. without the "original" anti-SDTG antibody, we were unable to assess if the native target was a linear or conformational epitope. Predicting a conformational epitope and globally testing a whole proteome for conformational epitopes is not yet possible. Phage display may allow for restricted identification of some conformational epitopes.

After further investigation utilizing the above described assays, we might be able to answer the questions so far left unanswered concerning SDTG:

1) Why is it being recognized and what is its native antigen?

2) What determins binding besides the presence of SDTG as some SDTG-containing peptides were not recognized in the original dataset?

The answers to these questions might help us in determining the significance and usability of this motif and the workflow developed to detect it.

Another way to utilize the whole proteome peptide array data is to use it for identification of new CAR-T cell targets. For this approach, we are partnering with the DeKosky lab at MIT as well as with the Yarmarkovich lab at NYU. The DeKosky lab focuses on inventing and applying cutting-edge tools in translational molecular biology to study immune system interactions and focused on TCR and antibody sequencing. The Yarmarkovich lab focuses on developing and employing technologies at the intersection of genomics, proteomics, immunology, antibody engineering and computational biology and has developed a novel class of peptide-centric CAR T cells that potently eradicate

tumors in preclinical models. We will supply the DeKosky lab with tumor, lymph node and spleen cells from mice cured of B78 using our ISV regimen that showed immune memory via rejection of a rechallenge with the same or a related tumor type. We will first test their serum against some of the previously identified top peptides from the whole proteome peptide array dataset, then harvest B and T cells from these mice from their tumor, lymph node and spleen where the DeKosky lab will generate a TCR library of the most "frequent" TCRs present in the cells. In addition, they will generate a library of all "frequent" scFvs in the same samples utilizing yeast display and next generation sequencing (Banach et al., 2022; Fahad et al., 2021; Wang et al., 2018). These libraries will be transferred to the Yarmarkovich lab who will then clone these into CAR T cells and genetically modified TCR T cells and generate a library of these (Yarmarkovich et al., 2021). These will be further evaluated utilizing the Berkeley lights platform to screen for the most potent tumor-recognizing and killing TCR and CAR T cells with a high-throughput, single cell level resolution (Bunse et al., 2019; Green et al., 2019). This will help identify the most functionally potent MHC-presented peptides as well as the most potent surface antigens (not MHC-presented) for best CAR T cell recognition.

Conclusions

Returning to the main 3 hypotheses for this thesis, proposed at the end of chapter 1, we have shown in chapter 2 and chapter 3, that mice that reject a rechallenge with the same, or a related, tumor type have developed antibodies that recognize the cancer cells, and that some immunodominant antigens are recognized by antibodies in immune sera from multiple mice cured from the same tumor (Hypothesis A).

We furthermore provided evidence that a whole-proteome peptide array will be a useful tool in identification of antibody targets on cancer cells (Hypothesis B) as we were able to identify a number of candidates that are now under further investigation.

Our last hypothesis was that antibody targets co-recognized by the antibodies of multiple immune mice may potentially be useful as future therapeutic targets, and possibly as a way to monitor patient's antibody responses as a biomarker of anti-tumor immunity. We are currently still working on answering this hypothesis and have proposed a series of experiments to investigate this further, some of which are already underway and outlined above.

In summary, we have identified a list of tumor-associated antigens and a binding motif on B78 tumors that may potentially be used as biomarkers of response as well as possible new immunotherapeutic targets. Filtering these against serum samples from mice with large tumors that did not receive treatment as well as mice that received treatment but were unable to successfully reject their tumor will help narrow down the list of possible biomarkers and therapeutic targets.

Overall, this work could be extended to the identification of immunodominant tumor-associated antigens within other cold murine cancers. Ultimately, these findings and this methodology could be extended to the study of sera from human cancer patients, cured via immunotherapy, to similarly probe the antigens recognized by tumor-immune human sera.

Figures

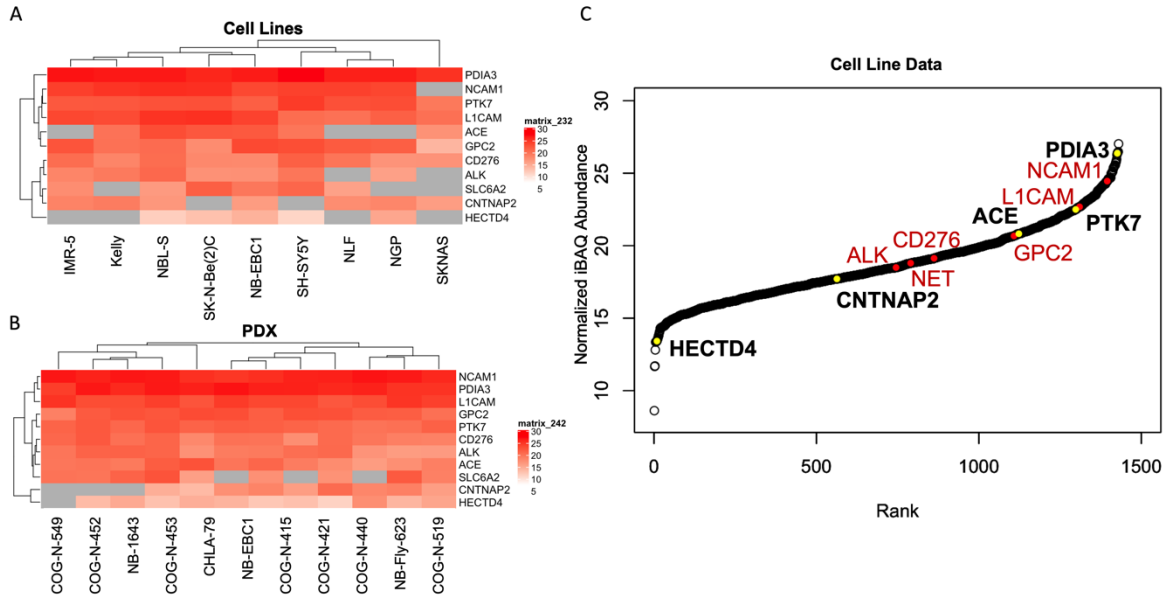


Figure 4.1: Identified enriched neuroblastoma surface proteins with highlighted antibody targets in whole proteome peptide array dataset.

A: Heatmap of relative abundance of cross-identified proteins together with known neuroblastoma surface targets in 9 neuroblastoma cell lines sorted by average abundance. **B**: Heatmap of relative abundance of cross-identified proteins together with known neuroblastoma surface targets in 11 neuroblastoma patient-derived xenografts sorted by average abundance. **C**: Protein abundance of all identified neuroblastoma surfaceome proteins with commonly known proteins highlighted in red and proteins identified in our dataset highlighted in yellow (with black font). Normalized iBAQ abundance is a measure of protein abundance where the iBAQ value is obtained by dividing protein intensities by the number of theoretically observable tryptic peptides between 6 and 30 aa and is on average highly correlated with protein abundance (Krey et al., 2014; Schwanhausser et al., 2011)

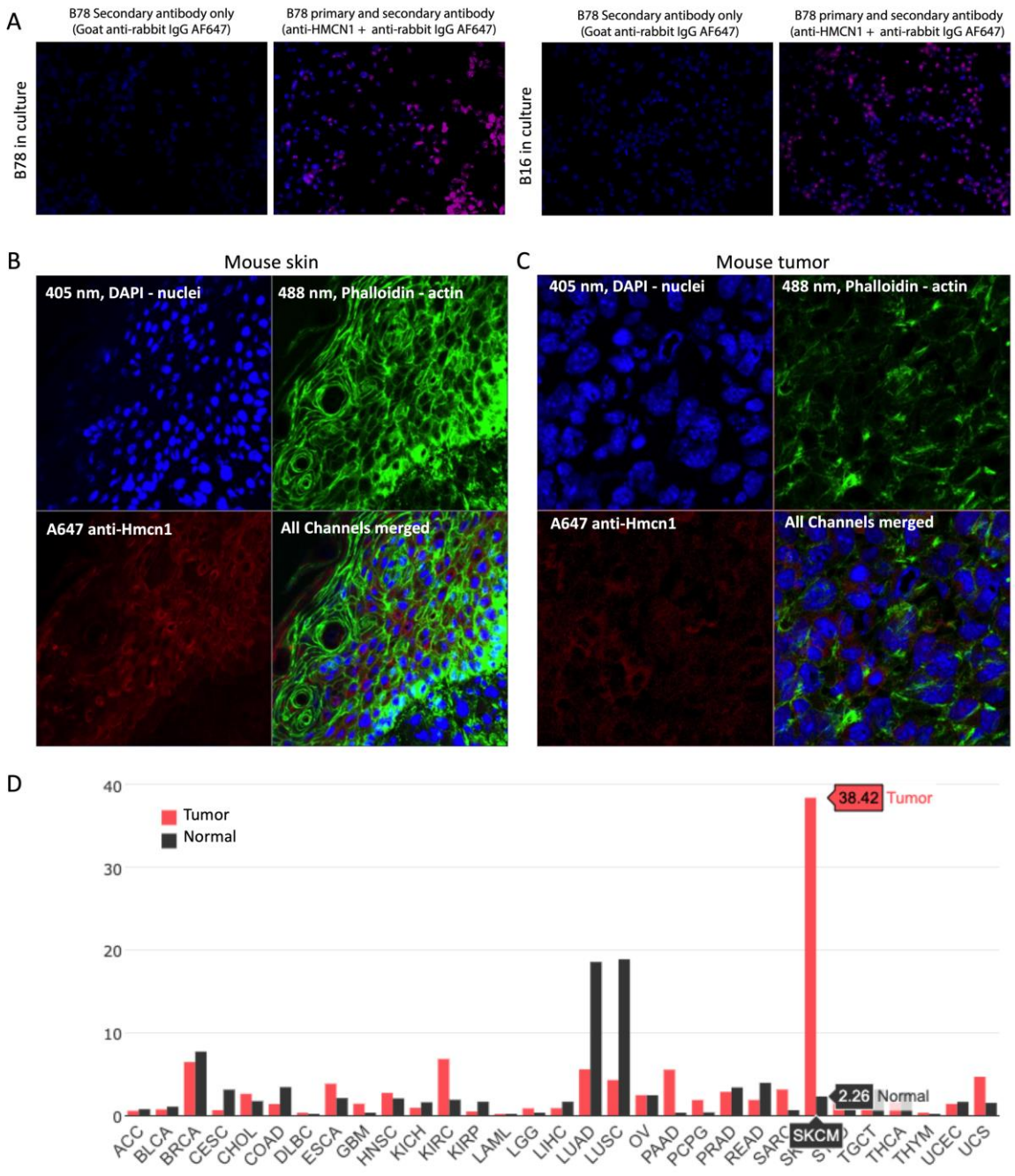


Figure 4.2: Hmcn1 is expressed in B78, B16 and normal mouse skin.

A: Immunohistochemistry staining of B78 and B16 cultured on slides with anti-HMCN1 antibody and anti-rabbit IgG AF647 as a secondary antibody. Strong staining was observed. Images were aquired using the EVOS FL imaging system from ThermoFisher at 20x. **B**: Fresh frozen sections of mouse skin

were stained with phalloidin (green), DAPI (blue), and anti-Hmcn1 in conjunction with a secondary anti-rabbit IgG AF647 antibody (red). Images were aquired utilizing a Zeiss LSM710 confocal microscope. Images were aquired with a 400x magnification. **C**: Fresh frozen sections of mouse tumor were stained with phalloidin (green), DAPI (blue), and anti-Hmcn1 in conjunction with a secondary anti-rabbit IgG AF647 antibody (red). Images were aquired utilizing a Zeiss LSM710 confocal microscope. Images were aquired with a 630x magnification. **D**: Hmcn1 mRNA expression in normal and matched tumor tissue as collected from Gepia: Gene Expression Profiling Interactive Analysis, http://gepia2.cancer-pku.cn/, doi: 10.1093/nar/gkz430.

Bibliography

Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., Kiso, W. K., Schmitt, D. L., Waddell, P. J., Bhaskara, S., Jensen, S. T., Maley, C. C., & Schiffman, J. D. (2015). Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA*, *314*(17), 1850-1860. https://doi.org/10.1001/jama.2015.13134

Albertini, M. R., Ranheim, E. A., Zuleger, C. L., Sondel, P. M., Hank, J. A., Bridges, A., Newton, M. A., McFarland, T., Collins, J., Clements, E., Henry, M. B., Neuman, H. B., Weber, S., Whalen, G., & Galili, U. (2016). Phase I study to evaluate toxicity and feasibility of intratumoral injection of alpha-gal glycolipids in patients with advanced melanoma. *Cancer Immunol Immunother*, *65*(8), 897-907. https://doi.org/10.1007/s00262-016-1846-1

Albittar, A. A., Alhalabi, O., & Glitza Oliva, I. C. (2020). Immunotherapy for Melanoma. *Adv Exp Med Biol*, *1244*, 51-68. https://doi.org/10.1007/978-3-030-41008-7_3

Alves, G., & Yu, Y.-K. (2014). Accuracy Evaluation of the Unified P-Value from Combining Correlated P-Values. *PLOS ONE*, *9*(3), e91225. https://doi.org/10.1371/journal.pone.0091225

Anderson, J., Majzner, R. G., & Sondel, P. M. (2022). Immunotherapy of Neuroblastoma: Facts and Hopes. *Clin Cancer Res*, *28*(15), 3196-3206. https://doi.org/10.1158/1078-0432.CCR-21-1356

Anthony, C. S., Corradi, H. R., Schwager, S. L., Redelinghuys, P., Georgiadis, D., Dive, V., Acharya, K. R., & Sturrock, E. D. (2010). The N domain of human angiotensin-I-converting enzyme: the role of N-glycosylation and the crystal structure in complex with an N domain-specific phosphinic inhibitor, RXP407. *J Biol Chem*, *285*(46), 35685-35693. https://doi.org/10.1074/jbc.M110.167866

Apetoh, L., Ghiringhelli, F., Tesniere, A., Obeid, M., Ortiz, C., Criollo, A., Mignot, G., Maiuri, M. C., Ullrich, E., Saulnier, P., Yang, H., Amigorena, S., Ryffel, B., Barrat, F. J., Saftig, P., Levi, F., Lidereau, R., Nogues, C., Mira, J. P., . . . Zitvogel, L. (2007). Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. *Nat Med*, *13*(9), 1050-1059. https://doi.org/10.1038/nm1622

Arima, S., Lin, J., Pecora, V., & Tardella, L. (2012). A Bayesian hierarchical model for identifying epitopes in peptide microarray data. *Biostatistics*, *13*(1), 101-112. https://doi.org/10.1093/biostatistics/kxr022

AssunÇão, R. M., Neves, M. C., Câmara, G., & Da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, *20*(7), 797-811. https://doi.org/10.1080/13658810600665111

Atkins, M. B., Lotze, M. T., Dutcher, J. P., Fisher, R. I., Weiss, G., Margolin, K., Abrams, J., Sznol, M., Parkinson, D., Hawkins, M., Paradise, C., Kunkel, L., & Rosenberg, S. A. (1999). High-dose recombinant interleukin 2 therapy for patients with metastatic melanoma: analysis of 270 patients treated between 1985 and 1993. *J Clin Oncol*, *17*(7), 2105-2116. https://doi.org/10.1200/JCO.1999.17.7.2105

Ayoglu, B., Haggmark, A., Khademi, M., Olsson, T., Uhlen, M., Schwenk, J. M., & Nilsson, P. (2013). Autoantibody profiling in multiple sclerosis using arrays of human protein fragments. *Mol Cell Proteomics*, *12*(9), 2657-2672. https://doi.org/10.1074/mcp.M112.026757

Baerga-Ortiz, A., Hughes, C. A., Mandell, J. G., & Komives, E. A. (2002). Epitope mapping of a monoclonal antibody against human thrombin by H/D-exchange mass spectrometry reveals selection of a diverse sequence in a highly conserved protein. *Protein Sci*, *11*(6), 1300-1308. https://doi.org/10.1110/ps.4670102

Bai, Y., Ni, M., Cooper, B., Wei, Y., & Fury, W. (2014). Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics*, *15*(1), 325. https://doi.org/10.1186/1471-2164-15-325

Baker, D. (2020). *pepMeld*. https://github.com/DABAKER165/pepMeld

Banach, B. B., Tripathi, P., Da Silva Pereira, L., Gorman, J., Nguyen, T. D., Dillon, M., Fahad, A. S., Kiyuka, P. K., Madan, B., Wolfe, J. R., Bonilla, B., Flynn, B., Francica, J. R., Hurlburt, N. K., Kisalu, N. K., Liu, T., Ou, L., Rawi, R., Schon, A., . . . DeKosky, B. J. (2022). Highly protective antimalarial antibodies via precision library generation and yeast display screening. *J Exp Med*, *219*(8). https://doi.org/10.1084/jem.20220323

Baniel, C. C., Heinze, C. M., Hoefges, A., Sumiec, E. G., Hank, J. A., Carlson, P. M., Jin, W. J., Patel, R. B., Sriramaneni, R. N., Gillies, S. D., Erbe, A. K., Schwarz, C. N., Pieper, A. A., Rakhmilevich, A. L., Sondel, P. M., & Morris, Z. S. (2020). In situ Vaccine Plus Checkpoint Blockade Induces Memory Humoral Response. *Front Immunol*, *11*, 1610. https://doi.org/10.3389/fimmu.2020.01610

Baniel, C. C., Sumiec, E. G., Hank, J. A., Bates, A. M., Erbe, A. K., Pieper, A. A., Hoefges, A. G., Patel, R. B., Rakhmilevich, A. L., Morris, Z. S., & Sondel, P. M. (2020). Intratumoral injection reduces toxicity and antibody-mediated neutralization of immunocytokine in a mouse melanoma model. *J Immunother Cancer*, *8*(2). https://doi.org/10.1136/jitc-2020-001262

Barnett, G. C., West, C. M., Dunning, A. M., Elliott, R. M., Coles, C. E., Pharoah, P. D., & Burnet, N. G. (2009). Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nat Rev Cancer*, *9*(2), 134-142. https://doi.org/10.1038/nrc2587

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1 - 48. https://doi.org/10.18637/jss.v067.i01

Battula, V. L., Shi, Y., Evans, K. W., Wang, R. Y., Spaeth, E. L., Jacamo, R. O., Guerra, R., Sahin, A. A., Marini, F. C., Hortobagyi, G., Mani, S. A., & Andreeff, M. (2012). Ganglioside GD2 identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest*, *122*(6), 2066-2078. https://doi.org/10.1172/JCI59735

Becker, B. J. (1994). Combining significance levels. In *The handbook of research synthesis.* (pp. 215-230). Russell Sage Foundation.

Becker, J. C., Varki, N., Gillies, S. D., Furukawa, K., & Reisfeld, R. A. (1996). An antibody-interleukin 2 fusion protein overcomes tumor heterogeneity by induction of a cellular immune response. *Proc Natl Acad Sci U S A*, *93*(15), 7826-7831. https://doi.org/10.1073/pnas.93.15.7826

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. http://www.jstor.org/stable/2346101

Berois, N., & Osinaga, E. (2014). Corrigendum: glycobiology of neuroblastoma: impact on tumor behavior, prognosis, and therapeutic strategies. *Front Oncol*, *4*, 193. https://doi.org/10.3389/fonc.2014.00193

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., Bruneval, P., Fridman, W. H., Becker, C., Pages, F., Speicher, M. R., Trajanoski, Z., & Galon, J. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*, *39*(4), 782-795. https://doi.org/10.1016/j.immuni.2013.10.003

Binnewies, M., Mujal, A. M., Pollack, J. L., Combes, A. J., Hardison, E. A., Barry, K. C., Tsui, J., Ruhland, M. K., Kersten, K., Abushawish, M. A., Spasic, M., Giurintano, J. P., Chan, V., Daud, A. I., Ha, P., Ye, C. J., Roberts, E. W., & Krummel, M. F. (2019). Unleashing Type-2 Dendritic Cells to Drive Protective Antitumor CD4(+) T Cell Immunity. *Cell*, *177*(3), 556-571 e516. https://doi.org/10.1016/j.cell.2019.02.005

Birnbaum, A. (1954). Combining Independent Tests of Significance. *Journal of the American Statistical Association*, *49*(267), 559-574. https://doi.org/10.2307/2281130

Bivand, R. S., Pebsema, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*. Springer.

Bivand, R. S., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, *27*(3), 716-748. https://doi.org/10.1007/s11749-018-0599-x

Bloy, N., Garcia, P., Laumont, C. M., Pitt, J. M., Sistigu, A., Stoll, G., Yamazaki, T., Bonneil, E., Buque, A., Humeau, J., Drijfhout, J. W., Meurice, G., Walter, S., Fritsche, J., Weinschenk, T., Rammensee, H. G., Melief, C., Thibault, P., Perreault, C., . . . Kroemer, G. (2017). Immunogenic stress and death of cancer cells: Contribution of antigenicity vs adjuvanticity to immunosurveillance. *Immunol Rev*, *280*(1), 165-174. https://doi.org/10.1111/imr.12582

Bolesta, E., Kowalczyk, A., Wierzbicki, A., Rotkiewicz, P., Bambach, B., Tsao, C. Y., Horwacik, I., Kolinski, A., Rokita, H., Brecher, M., Wang, X., Ferrone, S., & Kozbor, D. (2005). DNA vaccine expressing the mimotope of GD2 ganglioside induces protective GD2 cross-reactive antibody responses. *Cancer Res*, *65*(8), 3410-3418. https://doi.org/10.1158/0008-5472.CAN-04-2164

Bolotin, D. A., Poslavsky, S., Davydov, A. N., Frenkel, F. E., Fanchi, L., Zolotareva, O. I., Hemmers, S., Putintseva, E. V., Obraztsova, A. S., Shugay, M., Ataullakhanov, R. I., Rudensky, A. Y., Schumacher, T. N., & Chudakov, D. M. (2017). Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol*, *35*(10), 908-911. https://doi.org/10.1038/nbt.3979

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations [https://doi.org/10.1111/j.2517-6161.1964.tb00553.x]. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243. https://doi.org/https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Boyman, O., & Sprent, J. (2012). The role of interleukin-2 during homeostasis and activation of the immune system. *Nat Rev Immunol*, *12*(3), 180-190. https://doi.org/10.1038/nri3156

Breimer, L. H. (1988). Ionizing radiation-induced mutagenesis. *Br J Cancer*, *57*(1), 6-18. https://doi.org/10.1038/bjc.1988.2

Brodeur, G. M. (2003). Neuroblastoma: biological insights into a clinical enigma. *Nat Rev Cancer*, *3*(3), 203-216. https://doi.org/10.1038/nrc1014

Bromberg, J. F., Horvath, C. M., Wen, Z., Schreiber, R. D., & Darnell, J. E., Jr. (1996). Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. *Proc Natl Acad Sci U S A*, *93*(15), 7673-7678. https://doi.org/10.1073/pnas.93.15.7673

Brown, M. B. (1975). 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics*, *31*(4), 987-992. https://doi.org/10.2307/2529826

Broz, M. L., Binnewies, M., Boldajipour, B., Nelson, A. E., Pollack, J. L., Erle, D. J., Barczak, A., Rosenblum, M. D., Daud, A., Barber, D. L., Amigorena, S., Van't Veer, L. J., Sperling, A. I., Wolf, D. M., & Krummel, M. F. (2014). Dissecting the Tumor Myeloid Compartment Reveals Rare Activating Antigen-Presenting Cells Critical for T Cell Immunity. *Cancer Cell*, *26*(6), 938. https://doi.org/10.1016/j.ccell.2014.11.010

Bruno, T. C., Ebner, P. J., Moore, B. L., Squalls, O. G., Waugh, K. A., Eruslanov, E. B., Singhal, S., Mitchell, J. D., Franklin, W. A., Merrick, D. T., McCarter, M. D., Palmer, B. E., Kern, J. A., & Slansky, J. E. (2017). Antigen-Presenting Intratumoral B Cells Affect CD4(+) TIL Phenotypes in Non-Small Cell Lung Cancer Patients. *Cancer Immunol Res*, *5*(10), 898-907. https://doi.org/10.1158/2326-6066.CIR-17-0075

Bruschi, M., Cavalli, A., Moll, S., Candiano, G., Scapozza, L., Patel, J. J., Tan, J. C., Lo, K. C., Angeletti, A., Ghiggeri, G. M., & Prunotto, M. (2022). Discovery of anti-Formin-like 1 protein (FMNL1) antibodies in membranous nephropathy and other glomerular diseases. *Sci Rep*, *12*(1), 13659. https://doi.org/10.1038/s41598-022-17696-w

Bunse, L., Green, E. W., & Platten, M. (2019). High-throughput discovery of cancer-targeting TCRs. *Methods Enzymol*, *629*, 401-417. https://doi.org/10.1016/bs.mie.2019.10.009

Buus, S., Rockberg, J., Forsstrom, B., Nilsson, P., Uhlen, M., & Schafer-Nielsen, C. (2012). High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Mol Cell Proteomics*, *11*(12), 1790-1800. https://doi.org/10.1074/mcp.M112.020800

Carmi, Y., Spitzer, M. H., Linde, I. L., Burt, B. M., Prestwood, T. R., Perlman, N., Davidson, M. G., Kenkel, J. A., Segal, E., Pusapati, G. V., Bhattacharya, N., & Engleman, E. G. (2015). Allogeneic IgG combined with dendritic cell stimuli induce antitumour T-cell immunity. *Nature*, *521*(7550), 99-104. https://doi.org/10.1038/nature14424

Castino, G. F., Cortese, N., Capretti, G., Serio, S., Di Caro, G., Mineri, R., Magrini, E., Grizzi, F., Cappello, P., Novelli, F., Spaggiari, P., Roncalli, M., Ridolfi, C., Gavazzi, F., Zerbi, A., Allavena, P., & Marchesi, F. (2016). Spatial distribution of B cells predicts prognosis in human pancreatic adenocarcinoma. *Oncoimmunology*, *5*(4), e1085147. https://doi.org/10.1080/2162402X.2015.1085147

Catlett-Falcone, R., Landowski, T. H., Oshiro, M. M., Turkson, J., Levitzki, A., Savino, R., Ciliberto, G., Moscinski, L., Fernandez-Luna, J. L., Nunez, G., Dalton, W. S., & Jove, R. (1999). Constitutive activation of Stat3 signaling confers resistance to apoptosis in human U266 myeloma cells. *Immunity*, *10*(1), 105-115. https://doi.org/10.1016/s1074-7613(00)80011-4

Cavdarli, S., Groux-Degroote, S., & Delannoy, P. (2019). Gangliosides: The Double-Edge Sword of Neuro-Ectodermal Derived Tumors. *Biomolecules*, *9*(8). https://doi.org/10.3390/biom9080311

Chakravarty, P. K., Alfieri, A., Thomas, E. K., Beri, V., Tanaka, K. E., Vikram, B., & Guha, C. (1999). Flt3-ligand administration after radiation therapy prolongs survival in a murine model of metastatic lung cancer. *Cancer Res*, *59*(24), 6028-6032. https://www.ncbi.nlm.nih.gov/pubmed/10626784

Chang, M. H., & Chen, D. S. (2015). Prevention of hepatitis B. *Cold Spring Harb Perspect Med*, *5*(3), a021493. https://doi.org/10.1101/cshperspect.a021493

Chen, A., Kammers, K., Larman, H. B., Scharpf, R. B., & Ruczinski, I. (2022). Detecting and quantifying antibody reactivity in PhIP-Seq data with BEER. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btac555

Cheung, I. Y., Cheung, N. V., Modak, S., Mauguen, A., Feng, Y., Basu, E., Roberts, S. S., Ragupathi, G., & Kushner, B. H. (2021). Survival Impact of Anti-GD2 Antibody Response in a Phase II Ganglioside Vaccine Trial Among Patients With High-Risk Neuroblastoma With Prior Disease Progression. *J Clin Oncol*, *39*(3), 215-226. https://doi.org/10.1200/JCO.20.01892

Cheung, N. K., Lazarus, H., Miraldi, F. D., Abramowsky, C. R., Kallick, S., Saarinen, U. M., Spitzer, T., Strandjord, S. E., Coccia, P. F., & Berger, N. A. (1987). Ganglioside GD2 specific monoclonal antibody 3F8: a phase I study in patients with neuroblastoma and malignant melanoma. *J Clin Oncol*, *5*(9), 1430-1440. https://doi.org/10.1200/JCO.1987.5.9.1430

Chevrier, S., Levine, J. H., Zanotelli, V. R. T., Silina, K., Schulz, D., Bacac, M., Ries, C. H., Ailles, L., Jewett, M. A. S., Moch, H., van den Broek, M., Beisel, C., Stadler, M. B., Gedye, C., Reis, B., Pe'er, D., & Bodenmiller, B. (2017). An Immune Atlas of Clear Cell Renal Cell Carcinoma. *Cell*, *169*(4), 736-749 e718. https://doi.org/10.1016/j.cell.2017.04.016

Chiriva-Internati, M., & Bot, A. (2015). A new era in cancer immunotherapy: discovering novel targets and reprogramming the immune system. *Int Rev Immunol*, *34*(2), 101-103. https://doi.org/10.3109/08830185.2015.1015888

Choe, K. Y., Bethlehem, R. A. I., Safrin, M., Dong, H., Salman, E., Li, Y., Grinevich, V., Golshani, P., DeNardo, L. A., Penagarikano, O., Harris, N. G., & Geschwind, D. H. (2022). Oxytocin normalizes altered circuit connectivity for social rescue of the Cntnap2 knockout mouse. *Neuron*, *110*(5), 795-808 e796. https://doi.org/10.1016/j.neuron.2021.11.031

Choi, H. L., Yang, H. R., Shin, H. G., Hwang, K., Kim, J. W., Lee, J. H., Ryu, T., Jung, Y., & Lee, S. (2023). Generation and Next-Generation Sequencing-Based Characterization of a Large Human Combinatorial Antibody Library. *Int J Mol Sci*, *24*(6). https://doi.org/10.3390/ijms24066011

Christensen, W. F., & Zabriskie, B. N. (2022). When Your Permutation Test is Doomed to Fail. *The American Statistician*, *76*(1), 53-63. https://EconPapers.repec.org/RePEc:taf:amstat:v:76:y:2022:i:1:p:53-63

Christiansen, A., Kringelum, J. V., Hansen, C. S., Bogh, K. L., Sullivan, E., Patel, J., Rigby, N. M., Eiwegger, T., Szepfalusi, Z., de Masi, F., Nielsen, M., Lund, O., & Dufva, M. (2015). High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Sci Rep*, *5*, 12913. https://doi.org/10.1038/srep12913

Chung, T. W., Kim, S. J., Choi, H. J., Kim, K. J., Kim, M. J., Kim, S. H., Lee, H. J., Ko, J. H., Lee, Y. C., Suzuki, A., & Kim, C. H. (2009). Ganglioside GM3 inhibits VEGF/VEGFR-2-mediated angiogenesis: direct interaction of GM3 with VEGFR-2. *Glycobiology*, *19*(3), 229-239. https://doi.org/10.1093/glycob/cwn114

Cipponi, A., Mercier, M., Seremet, T., Baurain, J. F., Theate, I., van den Oord, J., Stas, M., Boon, T., Coulie, P. G., & van Baren, N. (2012). Neogenesis of lymphoid structures and antibody responses occur in human melanoma metastases. *Cancer Res*, *72*(16), 3997-4007. https://doi.org/10.1158/0008-5472.CAN-12-1377

Coley, W. B. (1893). The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clin Orthop Relat Res*(262), 3-11. https://www.ncbi.nlm.nih.gov/pubmed/1984929

Coronella, J. A., Spier, C., Welch, M., Trevor, K. T., Stopeck, A. T., Villar, H., & Hersh, E. M. (2002). Antigen-driven oligoclonal expansion of tumor-infiltrating B cells in infiltrating ductal carcinoma of the breast. *J Immunol*, *169*(4), 1829-1836. https://doi.org/10.4049/jimmunol.169.4.1829

Coronella-Wood, J. A., & Hersh, E. M. (2003). Naturally occurring B-cell responses to breast cancer. *Cancer Immunol Immunother*, *52*(12), 715-738. https://doi.org/10.1007/s00262-003-0409-4

Coughlin, C. M., Salhany, K. E., Gee, M. S., LaTemple, D. C., Kotenko, S., Ma, X., Gri, G., Wysocka, M., Kim, J. E., Liu, L., Liao, F., Farber, J. M., Pestka, S., Trinchieri, G., & Lee, W. M. (1998). Tumor cell responses to IFNgamma affect tumorigenicity and response to IL-12 therapy and antiangiogenesis. *Immunity*, *9*(1), 25-34. https://doi.org/10.1016/s1074-7613(00)80585-3

Couzin-Frankel, J. (2013). Breakthrough of the year 2013. Cancer immunotherapy. *Science*, *342*(6165), 1432-1433. https://doi.org/10.1126/science.342.6165.1432

Crocenzi, T., Cottam, B., Newell, P., Wolf, R. F., Hansen, P. D., Hammill, C., Solhjem, M. C., To, Y. Y., Greathouse, A., Tormoen, G., Jutric, Z., Young, K., Bahjat, K. S., Gough, M. J., & Crittenden, M. R. (2016). A hypofractionated radiation regimen avoids the lymphopenia associated with neoadjuvant chemoradiation therapy of borderline resectable and locally advanced pancreatic adenocarcinoma. *J Immunother Cancer*, *4*, 45. https://doi.org/10.1186/s40425-016-0149-6

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res*, *14*(6), 1188-1190. https://doi.org/10.1101/gr.849004

Damelin, M., Bankovich, A., Bernstein, J., Lucas, J., Chen, L., Williams, S., Park, A., Aguilar, J., Ernstoff, E., Charati, M., Dushin, R., Aujay, M., Lee, C., Ramoth, H., Milton, M., Hampl, J., Lazetic, S., Pulito, V., Rosfjord, E., . . . Dylla, S. J. (2017). A PTK7-targeted antibody-drug conjugate reduces tumor-initiating cells and induces sustained tumor regressions. *Sci Transl Med*, *9*(372). https://doi.org/10.1126/scitranslmed.aag2611

Dang, V. D., Hilgenberg, E., Ries, S., Shen, P., & Fillatreau, S. (2014). From the regulatory functions of B cells to the identification of cytokine-producing plasma cell subsets. *Curr Opin Immunol*, *28*, 77-83. https://doi.org/10.1016/j.coi.2014.02.009

Davar, D., Ding, F., Saul, M., Sander, C., Tarhini, A. A., Kirkwood, J. M., & Tawbi, H. A. (2017). High-dose interleukin-2 (HD IL-2) for advanced melanoma: a single center experience from the University of Pittsburgh Cancer Institute. *J Immunother Cancer*, *5*(1), 74. https://doi.org/10.1186/s40425-017-0279-5

Davies, R. W., Wells, G. A., Stewart, A. F., Erdmann, J., Shah, S. H., Ferguson, J. F., Hall, A. S., Anand, S. S., Burnett, M. S., Epstein, S. E., Dandona, S., Chen, L., Nahrstaedt, J., Loley, C., Konig, I. R., Kraus, W. E., Granger, C. B., Engert, J. C., Hengstenberg, C., . . . McPherson, R. (2012). A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet*, *5*(2), 217-225. https://doi.org/10.1161/CIRCGENETICS.111.961243

de Jong, J. O., Llapashtica, C., Genestine, M., Strauss, K., Provenzano, F., Sun, Y., Zhu, H., Cortese, G. P., Brundu, F., Brigatti, K. W., Corneo, B., Migliori, B., Tomer, R., Kushner, S. A., Kellendonk, C., Javitch, J. A., Xu, B., & Markx, S. (2021). Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder. *Nat Commun*, *12*(1), 4087. https://doi.org/10.1038/s41467-021-24358-4

De Vincenzo, R., Conte, C., Ricci, C., Scambia, G., & Capelli, G. (2014). Long-term efficacy and safety of human papillomavirus vaccination. *Int J Womens Health*, *6*, 999-1010. https://doi.org/10.2147/IJWH.S50365

DeFalco, J., Harbell, M., Manning-Bog, A., Baia, G., Scholz, A., Millare, B., Sumi, M., Zhang, D., Chu, F., Dowd, C., Zuno-Mitchell, P., Kim, D., Leung, Y., Jiang, S., Tang, X., Williamson, K. S., Chen, X., Carroll, S. M., Espiritu Santo, G., . . . Robinson, W. H. (2018). Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. *Clin Immunol*, *187*, 37-45. https://doi.org/10.1016/j.clim.2017.10.002

DeLuca, K. F., Mick, J. E., Ide, A. H., Lima, W. C., Sherman, L., Schaller, K. L., Anderson, S. M., Zhao, N., Stasevich, T. J., Varma, D., Nilsson, J., & DeLuca, J. G. (2021). Generation and diversification of recombinant monoclonal antibodies. *Elife*, *10*. https://doi.org/10.7554/eLife.72093

Demaria, S., & Formenti, S. C. (2012). Role of T lymphocytes in tumor response to radiotherapy. *Front Oncol*, *2*, 95. https://doi.org/10.3389/fonc.2012.00095

Dewey, M. (2022). *metap:meta-analysis of significance values*. In (Version 1.8) http://www.dewey.myzen.co.uk/meta/meta.html

Dobrenkov, K., & Cheung, N. K. (2014). GD2-targeted immunotherapy and radioimmunotherapy. *Semin Oncol*, *41*(5), 589-612. https://doi.org/10.1053/j.seminoncol.2014.07.003

Dudley, M. E., Wunderlich, J. R., Yang, J. C., Sherry, R. M., Topalian, S. L., Restifo, N. P., Royal, R. E., Kammula, U., White, D. E., Mavroukakis, S. A., Rogers, L. J., Gracia, G. J., Jones, S. A., Mangiameli, D. P., Pelletier, M. M., Gea-Banacloche, J., Robinson, M. R., Berman, D. M., Filie, A. C., . . . Rosenberg, S. A. (2005). Adoptive cell transfer therapy following non-myeloablative but lymphodepleting chemotherapy for the treatment of patients with refractory metastatic melanoma. *J Clin Oncol*, *23*(10), 2346-2357. https://doi.org/10.1200/JCO.2005.00.240

Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004). The three Es of cancer immunoediting. *Annu Rev Immunol*, *22*, 329-360. https://doi.org/10.1146/annurev.immunol.22.012703.104803

Durante, M., & Loeffler, J. S. (2010). Charged particles in radiation oncology. *Nat Rev Clin Oncol*, *7*(1), 37-43. https://doi.org/10.1038/nrclinonc.2009.183

Durgeau, A., Virk, Y., Corgnac, S., & Mami-Chouaib, F. (2018). Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Front Immunol*, *9*, 14. https://doi.org/10.3389/fimmu.2018.00014

Engmark, M., Andersen, M. R., Laustsen, A. H., Patel, J., Sullivan, E., de Masi, F., Hansen, C. S., Kringelum, J. V., Lomonte, B., Gutierrez, J. M., & Lund, O. (2016). High-throughput immuno-profiling of mamba (Dendroaspis) venom toxin epitopes using high-density peptide microarrays. *Sci Rep*, *6*, 36629. https://doi.org/10.1038/srep36629

Erdag, G., Schaefer, J. T., Smolkin, M. E., Deacon, D. H., Shea, S. M., Dengel, L. T., Patterson, J. W., & Slingluff, C. L., Jr. (2012). Immunotype and immunohistologic characteristics of tumor-infiltrating immune cells are associated with clinical outcome in metastatic melanoma. *Cancer Res*, *72*(5), 1070-1080. https://doi.org/10.1158/0008-5472.CAN-11-3218

Esaki, N., Ohkawa, Y., Hashimoto, N., Tsuda, Y., Ohmi, Y., Bhuiyan, R. H., Kotani, N., Honke, K., Enomoto, A., Takahashi, M., Furukawa, K., & Furukawa, K. (2018). ASC amino acid transporter 2, defined by enzyme-mediated activation of radical sources, enhances

malignancy of GD2-positive small-cell lung cancer. *Cancer Sci*, *109*(1), 141-153. https://doi.org/10.1111/cas.13448

Fahad, A. S., Timm, M. R., Madan, B., Burgomaster, K. E., Dowd, K. A., Normandin, E., Gutierrez-Gonzalez, M. F., Pennington, J. M., De Souza, M. O., Henry, A. R., Laboune, F., Wang, L., Ambrozak, D. R., Gordon, I. J., Douek, D. C., Ledgerwood, J. E., Graham, B. S., Castilho, L. R., Pierson, T. C., . . . DeKosky, B. J. (2021). Functional Profiling of Antibody Immune Repertoires in Convalescent Zika Virus Disease Patients. *Front Immunol*, *12*, 615102. https://doi.org/10.3389/fimmu.2021.615102

Faqeih, E. A., Alghamdi, M. A., Almahroos, M. A., Alharby, E., Almuntashri, M., Alshangiti, A. M., Clement, P., Calame, D. G., Qebibo, L., Burglen, L., Doco-Fenzy, M., Mastrangelo, M., Torella, A., Manti, F., Nigro, V., Alban, Z., Alharbi, G. S., Hashmi, J. A., Alraddadi, R., . . . Almontashiri, N. A. M. (2023). Biallelic variants in HECT E3 paralogs, HECTD4 and UBE3C, encoding ubiquitin ligases cause neurodevelopmental disorders that overlap with Angelman syndrome. *Genet Med*, *25*(2), 100323. https://doi.org/10.1016/j.gim.2022.10.006

Filatenkov, A., Baker, J., Mueller, A. M., Kenkel, J., Ahn, G. O., Dutt, S., Zhang, N., Kohrt, H., Jensen, K., Dejbakhsh-Jones, S., Shizuru, J. A., Negrin, R. N., Engleman, E. G., & Strober, S. (2015). Ablative Tumor Radiation Can Change the Tumor Immune Cell Microenvironment to Induce Durable Complete Remissions. *Clin Cancer Res*, *21*(16), 3727-3739. https://doi.org/10.1158/1078-0432.CCR-14-2824

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S., & Gottardo, R. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*, *16*, 278. https://doi.org/10.1186/s13059-015-0844-5

Fleurence, J., Fougeray, S., Bahri, M., Cochonneau, D., Clemenceau, B., Paris, F., Heczey, A., & Birkle, S. (2017). Targeting O-Acetyl-GD2 Ganglioside for Cancer Immunotherapy. *J Immunol Res*, *2017*, 5604891. https://doi.org/10.1155/2017/5604891

Flinterman, A. E., Knol, E. F., Lencer, D. A., Bardina, L., den Hartog Jager, C. F., Lin, J., Pasmans, S. G. M. A., Bruijnzeel-Koomen, C. A. F. M., Sampson, H. A., van Hoffen, E., & Shreffler, W. G. (2008). Peanut epitopes for IgE and IgG4 in peanut-sensitized children in relation to severity of peanut allergy. *Journal of Allergy and Clinical Immunology*, *121*(3), 737-743.e710. https://doi.org/10.1016/j.jaci.2007.11.039

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., & Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, *251*(4995), 767-773. https://doi.org/10.1126/science.1990438

Forsstrom, B., Axnas, B. B., Stengele, K. P., Buhler, J., Albert, T. J., Richmond, T. A., Hu, F. J., Nilsson, P., Hudson, E. P., Rockberg, J., & Uhlen, M. (2014). Proteome-wide epitope mapping of antibodies using ultra-dense peptide arrays. *Mol Cell Proteomics*, *13*(6), 1585-1597. https://doi.org/10.1074/mcp.M113.033308

Forte, G., Sorrentino, R., Montinaro, A., Luciano, A., Adcock, I. M., Maiolino, P., Arra, C., Cicala, C., Pinto, A., & Morello, S. (2012). Inhibition of CD73 improves B cell-mediated anti-tumor immunity in a mouse model of melanoma. *J Immunol*, *189*(5), 2226-2233. https://doi.org/10.4049/jimmunol.1200744

Frank, R. (2002). The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications. *J Immunol Methods*, *267*(1), 13-26. https://doi.org/10.1016/s0022-1759(02)00137-0

Fremd, C., Stefanovic, S., Beckhove, P., Pritsch, M., Lim, H., Wallwiener, M., Heil, J., Golatta, M., Rom, J., Sohn, C., Schneeweiss, A., Schuetz, F., & Domschke, C. (2016). Mucin 1-specific B cell immune responses and their impact on overall survival in breast cancer patients. *Oncoimmunology*, *5*(1), e1057387. https://doi.org/10.1080/2162402X.2015.1057387

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., Alnemri, E. S., Altucci, L., Amelio, I., Andrews, D. W., Annicchiarico-Petruzzelli, M., Antonov, A. V., Arama, E., Baehrecke, E. H., Barlev, N. A., Bazan, N. G., Bernassola, F., Bertrand, M. J. M., Bianchi, K., . . . Kroemer, G. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ*, *25*(3), 486-541. https://doi.org/10.1038/s41418-017-0012-4

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Abeni, D., Boyle, P., & Melchi, C. F. (2005). Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. *Eur J Cancer*, *41*(1), 28-44. https://doi.org/10.1016/j.ejca.2004.10.015

Garaud, S., Zayakin, P., Buisseret, L., Rulle, U., Silina, K., de Wind, A., Van den Eyden, G., Larsimont, D., Willard-Gallo, K., & Line, A. (2018). Antigen Specificity and Clinical Significance of IgG and IgA Autoantibodies Produced in situ by Tumor-Infiltrating B Cells in Breast Cancer. *Front Immunol*, *9*, 2660. https://doi.org/10.3389/fimmu.2018.02660

Garg, A. D., Galluzzi, L., Apetoh, L., Baert, T., Birge, R. B., Bravo-San Pedro, J. M., Breckpot, K., Brough, D., Chaurio, R., Cirone, M., Coosemans, A., Coulie, P. G., De Ruysscher, D., Dini, L., de Witte, P., Dudek-Peric, A. M., Faggioni, A., Fucikova, J., Gaipl, U. S., . . . Agostinis, P. (2015). Molecular and Translational Classifications of DAMPs in Immunogenic Cell Death. *Front Immunol*, *6*, 588. https://doi.org/10.3389/fimmu.2015.00588

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., Nair, V. S., Xu, Y., Khuong, A., Hoang, C. D., Diehn, M., West, R. B., Plevritis, S. K., & Alizadeh, A. A. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*, *21*(8), 938-945. https://doi.org/10.1038/nm.3909

Germain, C., Gnjatic, S., Tamzalit, F., Knockaert, S., Remark, R., Goc, J., Lepelley, A., Becht, E., Katsahian, S., Bizouard, G., Validire, P., Damotte, D., Alifano, M., Magdeleinat, P., Cremer, I., Teillaud, J. L., Fridman, W. H., Sautes-Fridman, C., & Dieu-Nosjean, M. C. (2014). Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *Am J Respir Crit Care Med*, *189*(7), 832-844. https://doi.org/10.1164/rccm.201309-1611OC

Geysen, H. M., Meloen, R. H., & Barteling, S. J. (1984). Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc Natl Acad Sci U S A*, *81*(13), 3998-4002. https://doi.org/10.1073/pnas.81.13.3998

Ghiringhelli, F., Apetoh, L., Tesniere, A., Aymeric, L., Ma, Y., Ortiz, C., Vermaelen, K., Panaretakis, T., Mignot, G., Ullrich, E., Perfettini, J. L., Schlemmer, F., Tasdemir, E., Uhl, M., Genin, P., Civas, A., Ryffel, B., Kanellopoulos, J., Tschopp, J., . . . Zitvogel, L. (2009). Activation of the NLRP3 inflammasome in dendritic cells induces IL-1beta-dependent adaptive immunity against tumors. *Nat Med*, *15*(10), 1170-1178. https://doi.org/10.1038/nm.2028

Gilbert, A. E., Karagiannis, P., Dodev, T., Koers, A., Lacy, K., Josephs, D. H., Takhar, P., Geh, J. L., Healy, C., Harries, M., Acland, K. M., Rudman, S. M., Beavil, R. L., Blower, P. J., Beavil, A. J., Gould, H. J., Spicer, J., Nestle, F. O., & Karagiannis, S. N. (2011). Monitoring the systemic human memory B cell compartment of melanoma patients for anti-tumor IgG antibodies. *PLoS One*, *6*(4), e19330. https://doi.org/10.1371/journal.pone.0019330

Gnjatic, S., Ritter, E., Buchler, M. W., Giese, N. A., Brors, B., Frei, C., Murray, A., Halama, N., Zornig, I., Chen, Y. T., Andrews, C., Ritter, G., Old, L. J., Odunsi, K., & Jager, D. (2010). Seromic profiling of ovarian and pancreatic cancer. *Proc Natl Acad Sci U S A*, *107*(11), 5088-5093. https://doi.org/10.1073/pnas.0914213107

Goldszmid, R. S., Dzutsev, A., & Trinchieri, G. (2014). Host immune response to infection and cancer: unexpected commonalities. *Cell Host Microbe*, *15*(3), 295-305. https://doi.org/10.1016/j.chom.2014.02.003

Golubkov, V. S., Chekanov, A. V., Cieplak, P., Aleshin, A. E., Chernov, A. V., Zhu, W., Radichev, I. A., Zhang, D., Dong, P. D., & Strongin, A. Y. (2010). The Wnt/planar cell polarity protein-tyrosine kinase-7 (PTK7) is a highly efficient proteolytic target of membrane type-1 matrix metalloproteinase: implications in cancer and embryogenesis. *J Biol Chem*, *285*(46), 35740-35749. https://doi.org/10.1074/jbc.M110.165159

Gómez Román, V. R., Murray, J. C., & Weiner, L. M. (2014). Chapter 1 - Antibody-Dependent Cellular Cytotoxicity (ADCC). In M. E. Ackerman & F. Nimmerjahn (Eds.), *Antibody Fc* (pp. 1-27). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-394802-1.00001-7

Gong, Z., Wu, X., Guo, Q., Du, H., Zhang, F., & Kong, Y. (2022). Comprehensive Analysis of HMCN1 Somatic Mutation in Clear Cell Renal Cell Carcinoma. *Genes (Basel)*, *13*(7). https://doi.org/10.3390/genes13071282

Green, E. W., Bunse, L., Bozza, M., Sanghvi, K., & Platten, M. (2019). TCR validation toward gene therapy for cancer. *Methods Enzymol*, *629*, 419-441. https://doi.org/10.1016/bs.mie.2019.10.010

Greiff, V., Menzel, U., Haessler, U., Cook, S. C., Friedensohn, S., Khan, T. A., Pogson, M., Hellmann, I., & Reddy, S. T. (2014). Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice. *BMC Immunol*, *15*, 40. https://doi.org/10.1186/s12865-014-0040-5

Groh, V., Wu, J., Yee, C., & Spies, T. (2002). Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature*, *419*(6908), 734-738. https://doi.org/10.1038/nature01112

Gu, Z. (2022). Complex heatmap visualization [https://doi.org/10.1002/imt2.43]. *iMeta*, *1*(3), e43. https://doi.org/https://doi.org/10.1002/imt2.43

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847-2849. https://doi.org/10.1093/bioinformatics/btw313

Gudmundsdottir, H., Wells, A. D., & Turka, L. A. (1999). Dynamics and requirements of T cell clonal expansion in vivo at the single-cell level: effector function is linked to proliferative capacity. *J Immunol*, *162*(9), 5212-5223. https://www.ncbi.nlm.nih.gov/pubmed/10227995

Guo, C., Manjili, M. H., Subjeck, J. R., Sarkar, D., Fisher, P. B., & Wang, X. Y. (2013). Therapeutic cancer vaccines: past, present, and future. *Adv Cancer Res*, *119*, 421-475. https://doi.org/10.1016/B978-0-12-407190-2.00007-1

Haj, A. K., Breitbach, M. E., Baker, D. A., Mohns, M. S., Moreno, G. K., Wilson, N. A., Lyamichev, V., Patel, J., Weisgrau, K. L., Dudley, D. M., & O'Connor, D. H. (2020). High-Throughput Identification of MHC Class I Binding Peptides Using an Ultradense Peptide Array. *J Immunol*, *204*(6), 1689-1696. https://doi.org/10.4049/jimmunol.1900889

Halperin, R. F., Stafford, P., & Johnston, S. A. (2011). Exploring antibody recognition of sequence space through random-sequence peptide microarrays. *Mol Cell Proteomics*, *10*(3), M110 000786. https://doi.org/10.1074/mcp.M110.000786

Hamanaka, Y., Suehiro, Y., Fukui, M., Shikichi, K., Imai, K., & Hinoda, Y. (2003). Circulating anti-MUC1 IgG antibodies as a favorable prognostic factor for pancreatic cancer. *Int J Cancer*, *103*(1), 97-100. https://doi.org/10.1002/ijc.10801

Hammerich, L., Binder, A., & Brody, J. D. (2015). In situ vaccination: Cancer immunotherapy both personalized and off-the-shelf. *Mol Oncol*, *9*(10), 1966-1981. https://doi.org/10.1016/j.molonc.2015.10.016

Han, J., & Lotze, M. T. (2020). The Adaptome as Biomarker for Assessing Cancer Immunity and Immunotherapy. *Methods Mol Biol*, *2055*, 369-397. https://doi.org/10.1007/978-1-4939-9773-2_17

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discov*, *12*(1), 31-46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57-70. https://doi.org/10.1016/s0092-8674(00)81683-9

Hanton, G., & Pastoret, P. P. (1984). [The reaction of antibody-dependent cell-mediated cytotoxicity (ADCC)]. *Ann Rech Vet*, *15*(4), 443-454. https://www.ncbi.nlm.nih.gov/pubmed/6529114 (La reaction de cytotoxicite a mediation cellulaire dependante des anticorps (ADCC).)

Haraguchi, M., Yamashiro, S., Yamamoto, A., Furukawa, K., Takamiya, K., Lloyd, K. O., Shiku, H., & Furukawa, K. (1994). Isolation of GD3 synthase gene by expression cloning of GM3 alpha-2,8-sialyltransferase cDNA using anti-GD2 monoclonal antibody. *Proc Natl Acad Sci U S A*, *91*(22), 10455-10459. https://doi.org/10.1073/pnas.91.22.10455

Harper, J. W., & Bennett, E. J. (2016). Proteome complexity and the forces that drive proteome imbalance. *Nature*, *537*(7620), 328-338. https://doi.org/10.1038/nature19947

Heffron, A. S., McIlwain, S. J., Amjadi, M. F., Baker, D. A., Khullar, S., Armbrust, T., Halfmann, P. J., Kawaoka, Y., Sethi, A. K., Palmenberg, A. C., Shelef, M. A., O'Connor, D. H., & Ong, I. M. (2021). The landscape of antibody binding in SARS-CoV-2 infection. *PLoS Biol*, *19*(6), e3001265. https://doi.org/10.1371/journal.pbio.3001265

Heffron, A. S., Mohr, E. L., Baker, D., Haj, A. K., Buechler, C. R., Bailey, A., Dudley, D. M., Newman, C. M., Mohns, M. S., Koenig, M., Breitbach, M. E., Rasheed, M., Stewart, L. M., Eickhoff, J., Pinapati, R. S., Beckman, E., Li, H., Patel, J., Tan, J. C., & O'Connor, D. H. (2018). Antibody responses to Zika virus proteins in pregnant and non-pregnant macaques. *PLOS Neglected Tropical Diseases*, *12*(11), e0006903. https://doi.org/10.1371/journal.pntd.0006903

Hersey, P., & Jamal, O. (1989). Expression of the gangliosides GD3 and GD2 on lymphocytes in tissue sections of melanoma. *Pathology*, *21*(1), 51-58. https://doi.org/10.3109/00313028909059531

Hibbs, J. B., Jr., Taintor, R. R., Chapman, H. A., Jr., & Weinberg, J. B. (1977). Macrophage tumor killing: influence of the local environment. *Science*, *197*(4300), 279-282. https://doi.org/10.1126/science.327547

Hirasawa, Y., Kohno, N., Yokoyama, A., Kondo, K., Hiwada, K., & Miyake, M. (2000). Natural autoantibody to MUC1 is a prognostic indicator for non-small cell lung cancer. *Am J Respir Crit Care Med*, *161*(2 Pt 1), 589-594. https://doi.org/10.1164/ajrccm.161.2.9905028

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., Akerley, W., van den Eertwegh, A. J., Lutzky, J., Lorigan, P., Vaubel, J. M., Linette, G. P., Hogg, D., Ottensmeier, C. H., Lebbe, C., . . . Urba, W. J. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*, *363*(8), 711-723. https://doi.org/10.1056/NEJMoa1003466

Hoefges, A., Mathers, N., McIlwain, S., Pinapati, R., Garcia, B., Patel, J., Erbe, A., Ong, I., & Sondel, P. (2022). 45 Prevalent binding motif in C57BL6 mice cured of B78 melanoma via immunotherapy. *Journal for ImmunoTherapy of Cancer*, *10*(Suppl 2), A48. https://doi.org/10.1136/jitc-2022-SITC2022.0045

Hoefges, A., McIlwain, S. J., Erbe, A. K., Mathers, N., Xu, A., Melby, D., Tetreault, K., Le, T., Kim, K., Pinapati, R. S., Garcia, B., Patel, J., Heck, M., Feils, A. S., Tsarovsky, N., Hank, J. A., Morris, Z. S., Ong, I. M., & Sondel, P. M. (2023). Antibody landscape of C57BL/6 mice cured of B78 melanoma via immunotherapy. *bioRxiv*. https://doi.org/10.1101/2023.02.24.529012

Horwacik, I., Golik, P., Grudnik, P., Kolinski, M., Zdzalik, M., Rokita, H., & Dubin, G. (2015). Structural Basis of GD2 Ganglioside and Mimetic Peptide Recognition by 14G2a Antibody. *Mol Cell Proteomics*, *14*(10), 2577-2590. https://doi.org/10.1074/mcp.M115.052720

Houseman, B. T., Huh, J. H., Kron, S. J., & Mrksich, M. (2002). Peptide chips for the quantitative evaluation of protein kinase activity. *Nat Biotechnol*, *20*(3), 270-274. https://doi.org/10.1038/nbt0302-270

Hsueh, E. C., & Morton, D. L. (2003). Antigen-based immunotherapy of melanoma: Canvaxin therapeutic polyvalent cancer vaccine. *Semin Cancer Biol*, *13*(6), 401-407. https://doi.org/10.1016/j.semcancer.2003.09.003

Hulett, T. W., Jensen, S. M., Wilmarth, P. A., Reddy, A. P., Ballesteros-Merino, C., Afentoulis, M. E., Dubay, C., David, L. L., & Fox, B. A. (2018). Coordinated responses to individual tumor antigens by IgG antibody and CD8+ T cells following cancer vaccination. *J Immunother Cancer*, *6*(1), 27. https://doi.org/10.1186/s40425-018-0331-0

Huo, Z., Tang, S., Park, Y., & Tseng, G. (2020). P-value evaluation, variability index and biomarker categorization for adaptively weighted Fisher's meta-analysis method in omics applications. *Bioinformatics*, *36*(2), 524-532. https://doi.org/10.1093/bioinformatics/btz589

Imholte, G., & Gottardo, R. (2016). Bayesian hierarchical modeling for subject-level response classification in peptide microarray immunoassays. *Biometrics*, *72*(72), 1206-1215. https://doi.org/https://doi.org/10.1111/biom.12523 %U https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12523

Imholte, G. C., Sauteraud, R., Korber, B., Bailer, R. T., Turk, E. T., Shen, X., Tomaras, G. D., Mascola, J. R., Koup, R. A., Montefiori, D. C., & Gottardo, R. (2013). A computational framework for the analysis of peptide microarray antibody binding data with application to HIV vaccine profiling. *J Immunol Methods*, *395*(1-2), 1-13. https://doi.org/10.1016/j.jim.2013.06.001

International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931-945. https://doi.org/10.1038/nature03001

Irving, M. B., Pan, O., & Scott, J. K. (2001). Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr Opin Chem Biol*, *5*(3), 314-324. https://doi.org/10.1016/s1367-5931(00)00208-8

Isaeva, O. I., Sharonov, G. V., Serebrovskaya, E. O., Turchaninova, M. A., Zaretsky, A. R., Shugay, M., & Chudakov, D. M. (2019). Intratumoral immunoglobulin isotypes predict survival in lung adenocarcinoma subtypes. *J Immunother Cancer*, *7*(1), 279. https://doi.org/10.1186/s40425-019-0747-1

Jabara, H. H., Boyden, S. E., Chou, J., Ramesh, N., Massaad, M. J., Benson, H., Bainter, W., Fraulino, D., Rahimov, F., Sieff, C., Liu, Z. J., Alshemmari, S. H., Al-Ramadi, B. K., Al-Dhekri, H., Arnaout, R., Abu-Shukair, M., Vatsayan, A., Silver, E., Ahuja, S., . . . Geha, R. S. (2016). A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency. *Nat Genet*, *48*(1), 74-78. https://doi.org/10.1038/ng.3465

Jackson, P. A., Green, M. A., Marks, C. G., King, R. J., Hubbard, R., & Cook, M. G. (1996). Lymphocyte subset infiltration patterns and HLA antigen status in colorectal carcinomas and adenomas. *Gut*, *38*(1), 85-89. https://doi.org/10.1136/gut.38.1.85

Jagodinsky, J. C., Bates, A. M., Clark, P. A., Sriramaneni, R. N., Havighurst, T. C., Chakravarty, I., Nystuen, E. J., Kim, K., Sondel, P. M., Jin, W. J., & Morris, Z. S. (2022). Local TLR4 stimulation augments in situ vaccination induced via local radiation and anti-CTLA-4 checkpoint blockade through induction of CD8 T-cell independent Th1 polarization. *J Immunother Cancer*, *10*(10). https://doi.org/10.1136/jitc-2022-005103

Janeway, C. A., Jr. (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol*, *54 Pt 1*, 1-13. https://doi.org/10.1101/sqb.1989.054.01.003

Jarboe, J. S., Jaboin, J. J., Anderson, J. C., Nowsheen, S., Stanley, J. A., Naji, F., Ruijtenbeek, R., Tu, T., Hallahan, D. E., Yang, E. S., Bonner, J. A., & Willey, C. D. (2012). Kinomic profiling approach identifies Trk as a novel radiation modulator. *Radiother Oncol*, *103*(3), 380-387. https://doi.org/10.1016/j.radonc.2012.03.014

Jensen, O. N. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol*, *8*(1), 33-41. https://doi.org/10.1016/j.cbpa.2003.12.009

Jiang, T., Zhou, C., & Ren, S. (2016). Role of IL-2 in cancer immunotherapy. *Oncoimmunology*, *5*(6), e1163462. https://doi.org/10.1080/2162402X.2016.1163462

Jie, Y., Liu, G., Feng, L., Li, Y., E, M., Wu, L., Li, Y., Rong, G., Li, Y., Wei, H., & Gu, A. (2021). PTK7-Targeting CAR T-Cells for the Treatment of Lung Cancer and Other Malignancies. *Front Immunol*, *12*, 665970. https://doi.org/10.3389/fimmu.2021.665970

Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*, *10*(1), 5029. https://doi.org/10.1038/s41467-019-13036-1

Kaplan, D. H., Shankaran, V., Dighe, A. S., Stockert, E., Aguet, M., Old, L. J., & Schreiber, R. D. (1998). Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. *Proc Natl Acad Sci U S A*, *95*(13), 7556-7561. https://doi.org/10.1073/pnas.95.13.7556

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*, *20*(4), 1160-1166. https://doi.org/10.1093/bib/bbx108

Kikutake, C., Yoshihara, M., Sato, T., Saito, D., & Suyama, M. (2018). Intratumor heterogeneity of HMCN1 mutant alleles associated with poor prognosis in patients with breast cancer. *Oncotarget*, *9*(70), 33337-33347. https://doi.org/10.18632/oncotarget.26071

Kim, J. H., Shin, W. S., Lee, S. R., Kim, S., Choi, S. Y., & Lee, S. T. (2022). Anti-PTK7 Monoclonal Antibodies Exhibit Anti-Tumor Activity at the Cellular Level and in Mouse Xenograft Models of Esophageal Squamous Cell Carcinoma. *Int J Mol Sci*, *23*(20). https://doi.org/10.3390/ijms232012195

Klapper, J. A., Downey, S. G., Smith, F. O., Yang, J. C., Hughes, M. S., Kammula, U. S., Sherry, R. M., Royal, R. E., Steinberg, S. M., & Rosenberg, S. (2008). High-dose interleukin-2 for the treatment of metastatic renal cell carcinoma : a retrospective analysis of response and survival in patients treated in the surgery branch at the National Cancer Institute between 1986 and 2006. *Cancer*, *113*(2), 293-301. https://doi.org/10.1002/cncr.23552

Knijnenburg, T. A., Wessels, L. F., Reinders, M. J., & Shmulevich, I. (2009). Fewer permutations, more accurate P-values. *Bioinformatics*, *25*(12), i161-168. https://doi.org/10.1093/bioinformatics/btp211

Ko, E. C., Benjamin, K. T., & Formenti, S. C. (2018). Generating antitumor immunity by targeted radiation therapy: Role of dose and fractionation. *Adv Radiat Oncol*, *3*(4), 486-493. https://doi.org/10.1016/j.adro.2018.08.021

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. R., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*, *16*(12), 1289-1296. https://doi.org/10.1038/s41592-019-0619-0

Kost, J. T., & McDermott, M. P. (2002). Combining dependent P-values. *Statistics & Probability Letters*, *60*(2), 183-190.

Krassowski, M., Arts, M., Lagger, C., & Max. (2022). *krassowski/complex-upset: v1.3.5*. In Zenodo. https://doi.org/10.5281/zenodo.7314197#.ZA94xT7GO74.mendeley

Krengel, U., & Bousquet, P. A. (2014). Molecular recognition of gangliosides and their potential for cancer immunotherapies. *Front Immunol*, *5*, 325. https://doi.org/10.3389/fimmu.2014.00325

Krey, J. F., Wilmarth, P. A., Shin, J. B., Klimek, J., Sherman, N. E., Jeffery, E. D., Choi, D., David, L. L., & Barr-Gillespie, P. G. (2014). Accurate label-free protein quantitation with high- and low-

resolution mass spectrometers. *J Proteome Res*, *13*(2), 1034-1044. https://doi.org/10.1021/pr401017h

Kringelum, J. V., Nielsen, M., Padkjaer, S. B., & Lund, O. (2013). Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol*, *53*(1-2), 24-34. https://doi.org/10.1016/j.molimm.2012.06.001

Kroeger, D. R., Milne, K., & Nelson, B. H. (2016). Tumor-Infiltrating Plasma Cells Are Associated with Tertiary Lymphoid Structures, Cytolytic T-Cell Responses, and Superior Prognosis in Ovarian Cancer. *Clin Cancer Res*, *22*(12), 3005-3015. https://doi.org/10.1158/1078-0432.CCR-15-2762

Kumar, A., Commane, M., Flickinger, T. W., Horvath, C. M., & Stark, G. R. (1997). Defective TNF-alpha-induced apoptosis in STAT1-null cells due to low constitutive levels of caspases. *Science*, *278*(5343), 1630-1632. https://doi.org/10.1126/science.278.5343.1630

Kurai, J., Chikumi, H., Hashimoto, K., Yamaguchi, K., Yamasaki, A., Sako, T., Touge, H., Makino, H., Takata, M., Miyata, M., Nakamoto, M., Burioka, N., & Shimizu, E. (2007). Antibody-dependent cellular cytotoxicity mediated by cetuximab against lung cancer cell lines. *Clin Cancer Res*, *13*(5), 1552-1561. https://doi.org/10.1158/1078-0432.CCR-06-1726

Ladanyi, A., Kiss, J., Mohos, A., Somlai, B., Liszkay, G., Gilde, K., Fejos, Z., Gaudi, I., Dobos, J., & Timar, J. (2011). Prognostic impact of B-cell density in cutaneous melanoma. *Cancer Immunol Immunother*, *60*(12), 1729-1738. https://doi.org/10.1007/s00262-011-1071-x

Ladenstein, R., Potschger, U., Valteau-Couanet, D., Luksch, R., Castel, V., Yaniv, I., Laureys, G., Brock, P., Michon, J. M., Owens, C., Trahair, T., Chan, G. C. F., Ruud, E., Schroeder, H., Beck Popovic, M., Schreier, G., Loibner, H., Ambros, P., Holmes, K., . . . Lode, H. N. (2018). Interleukin 2 with anti-GD2 antibody ch14.18/CHO (dinutuximab beta) in patients with high-risk neuroblastoma (HR-NBL1/SIOPEN): a multicentre, randomised, phase 3 trial. *Lancet Oncol*, *19*(12), 1617-1629. https://doi.org/10.1016/S1470-2045(18)30578-3

Larman, H. B., Zhao, Z., Laserson, U., Li, M. Z., Ciccia, A., Gakidis, M. A., Church, G. M., Kesari, S., Leproust, E. M., Solimini, N. L., & Elledge, S. J. (2011). Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol*, *29*(6), 535-541. https://doi.org/10.1038/nbt.1856

Ledeen, R., & Wu, G. (2018). Gangliosides of the Nervous System. *Methods Mol Biol*, *1804*, 19-55. https://doi.org/10.1007/978-1-4939-8552-4_2

Ledeen, R. W., & Yu, R. K. (1982). Gangliosides: structure, isolation, and analysis. *Methods Enzymol*, *83*, 139-191. https://doi.org/10.1016/0076-6879(82)83012-7

Lee, Y., Auh, S. L., Wang, Y., Burnette, B., Wang, Y., Meng, Y., Beckett, M., Sharma, R., Chin, R., Tu, T., Weichselbaum, R. R., & Fu, Y. X. (2009). Therapeutic effects of ablative radiation on local tumor require CD8+ T cells: changing strategies for cancer treatment. *Blood*, *114*(3), 589-595. https://doi.org/10.1182/blood-2009-02-206870

Lee, Y. J., Lee, H., Jang, H. B., Yoo, M. G., Im, S., Koo, S. K., & Lee, H. J. (2022). The potential effects of HECTD4 variants on fasting glucose and triglyceride levels in relation to prevalence of type 2 diabetes based on alcohol intake. *Arch Toxicol*, *96*(9), 2487-2499. https://doi.org/10.1007/s00204-022-03325-y

Leonard, W. J. (2001). Cytokines and immunodeficiency diseases. *Nat Rev Immunol*, *1*(3), 200-208. https://doi.org/10.1038/35105066

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph*, *20*(12), 1983-1992. https://doi.org/10.1109/TVCG.2014.2346248

Li, S., Marthandan, N., Bowerman, D., Garner, H. R., & Kodadek, T. (2005). Photolithographic synthesis of cyclic peptide arrays using a differential deprotection strategy. *Chem Commun (Camb)*(5), 581-583. https://doi.org/10.1039/b415578e

Lin, J., Bardina, L., Shreffler, W. G., Andreae, D. A., Ge, Y., Wang, J., Bruni, F. M., Fu, Z., Han, Y., & Sampson, H. A. (2009). Development of a novel peptide microarray for large-scale epitope mapping of food allergens. *Journal of Allergy and Clinical Immunology*, *124*(2), 315-322.e313. https://doi.org/10.1016/j.jaci.2009.05.024

Lin, J., Bruni, F. M., Fu, Z., Maloney, J., Bardina, L., Boner, A. L., Gimenez, G., & Sampson, H. A. (2012). A bioinformatics approach to identify patients with symptomatic peanut allergy using peptide microarray immunoassay. *Journal of Allergy and Clinical Immunology*, *129*(5), 1321-1328.e1325. https://doi.org/10.1016/j.jaci.2012.02.012

Liu, C. L., Pan, H. W., Torng, P. L., Fan, M. H., & Mao, T. L. (2019). SRPX and HMCN1 regulate cancer-associated fibroblasts to promote the invasiveness of ovarian carcinoma. *Oncol Rep*, *42*(6), 2706-2715. https://doi.org/10.3892/or.2019.7379

Liu, J., Zheng, X., Pang, X., Li, L., Wang, J., Yang, C., & Du, G. (2018). Ganglioside GD3 synthase (GD3S), a novel cancer drug target. *Acta Pharm Sin B*, *8*(5), 713-720. https://doi.org/10.1016/j.apsb.2018.07.009

Liu, Y., Wondimu, A., Yan, S., Bobb, D., & Ladisch, S. (2014). Tumor gangliosides accelerate murine tumor angiogenesis. *Angiogenesis*, *17*(3), 563-571. https://doi.org/10.1007/s10456-013-9403-4

Liu, Y., & Xie, J. (2020). Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures. *Journal of the American Statistical Association*, *115*(529), 393-402. https://doi.org/10.1080/01621459.2018.1554485

Lo, K. C., Sullivan, E., Bannen, R. M., Jin, H., Rowe, M., Li, H., Pinapati, R. S., Cartwright, A. J., Tan, J. C., Patel, J., Keystone, E. C., & Siminovitch, K. A. (2020). Comprehensive Profiling of the Rheumatoid Arthritis Antibody Repertoire. *Arthritis Rheumatol*, *72*(2), 242-250. https://doi.org/10.1002/art.41089

Lo Nigro, C., Macagno, M., Sangiolo, D., Bertolaccini, L., Aglietta, M., & Merlano, M. C. (2019). NK-mediated antibody-dependent cell-mediated cytotoxicity in solid tumors: biological evidence and clinical perspectives. *Ann Transl Med*, *7*(5), 105. https://doi.org/10.21037/atm.2019.01.42

Loeb, L. A., Loeb, K. R., & Anderson, J. P. (2003). Multiple mutations and cancer. *Proc Natl Acad Sci U S A*, *100*(3), 776-781. https://doi.org/10.1073/pnas.0334858100

London, C. A., Lodge, M. P., & Abbas, A. K. (2000). Functional responses and costimulator dependence of memory CD4+ T cells. *J Immunol*, *164*(1), 265-272. https://doi.org/10.4049/jimmunol.164.1.265

Lopez, P. H., & Schnaar, R. L. (2009). Gangliosides in cell recognition and membrane protein regulation. *Curr Opin Struct Biol*, *19*(5), 549-557. https://doi.org/10.1016/j.sbi.2009.06.001

Lu, H., Goodell, V., & Disis, M. L. (2008). Humoral immunity directed against tumor-associated antigens as potential biomarkers for the early diagnosis of cancer. *J Proteome Res*, *7*(4), 1388-1394. https://doi.org/10.1021/pr700818f

Lund, F. E., & Randall, T. D. (2010). Effector and regulatory B cells: modulators of CD4+ T cell immunity. *Nat Rev Immunol*, *10*(4), 236-247. https://doi.org/10.1038/nri2729

Lyamichev, V. I., Goodrich, L. E., Sullivan, E. H., Bannen, R. M., Benz, J., Albert, T. J., & Patel, J. J. (2017). Stepwise Evolution Improves Identification of Diverse Peptides Binding to a Protein Target. *Sci Rep*, *7*(1), 12116. https://doi.org/10.1038/s41598-017-12440-1

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022, 2022). *cluster: Cluster Analysis Basics and Extensions*. https://CRAN.R-project.org/package=cluster

Maier, R. H., Maier, C. J., Rid, R., Hintner, H., Bauer, J. W., & Onder, K. (2010). Epitope mapping of antibodies using a cell array-based polypeptide library. *J Biomol Screen*, *15*(4), 418-426. https://doi.org/10.1177/1087057110363821

Manenq, C., Lesesve, J. F., Dreumont, N., Massin, F., Salignac, S., Mansuy, L., Chastagner, P., Latger-Cannard, V., & Broseus, J. (2020). Combined use of multiparametric flow cytometry and cytomorphology to enhance detection of neuroblastoma metastatic cells in bone marrow. *Int J Lab Hematol*, *42*(1), 52-60. https://doi.org/10.1111/ijlh.13137

Marincola, F. M., Jaffee, E. M., Hicklin, D. J., & Ferrone, S. (2000). Escape of human solid tumors from T-cell recognition: molecular mechanisms and functional significance. *Adv Immunol*, *74*, 181-273. https://doi.org/10.1016/s0065-2776(08)60911-6

Maskos, U., & Southern, E. M. (1992). Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. *Nucleic Acids Res*, *20*(7), 1679-1684. https://doi.org/10.1093/nar/20.7.1679

Masur, P. K. S., Michael (2020). specr: Conducting and Visualizing Specification Curve Analyses (Version 1.0.0). https://CRAN.R-project.org/package=specr

Matzinger, P. (1994). Tolerance, danger, and the extended family. *Annu Rev Immunol*, *12*, 991-1045. https://doi.org/10.1146/annurev.iy.12.040194.005015

McGinnis, C. S., Murrow, L. M., & Gartner, Z. J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*, *8*(4), 329-337 e324. https://doi.org/10.1016/j.cels.2019.03.003

McIlwain, S. J., Hoefges, A., Erbe, A. K., Sondel, P. M., & Ong, I. M. (2023). Ranking Antibody Binding Epitopes and Proteins Across Samples from Whole Proteome Tiled Linear Peptides. *bioRxiv*, 2023.2004.2023.536620. https://doi.org/10.1101/2023.04.23.536620

Menendez, A., & Scott, J. K. (2005). The nature of target-unrelated peptides recovered in the screening of phage-displayed random peptide libraries with antibodies. *Anal Biochem*, *336*(2), 145-157. https://doi.org/10.1016/j.ab.2004.09.048

Meng, L., Sefah, K., O'Donoghue, M. B., Zhu, G., Shangguan, D., Noorali, A., Chen, Y., Zhou, L., & Tan, W. (2010). Silencing of PTK7 in colon cancer cells: caspase-10-dependent apoptosis via mitochondrial pathway. *PLoS One*, *5*(11), e14018. https://doi.org/10.1371/journal.pone.0014018

Meng, Y., Carpentier, A. F., Chen, L., Boisserie, G., Simon, J. M., Mazeron, J. J., & Delattre, J. Y. (2005). Successful combination of local CpG-ODN and radiotherapy in malignant glioma. *Int J Cancer*, *116*(6), 992-997. https://doi.org/10.1002/ijc.21131

Mergaert, A. M., Zheng, Z., Denny, M. F., Amjadi, M. F., Bashar, S. J., Newton, M. A., Malmstrom, V., Gronwall, C., McCoy, S. S., & Shelef, M. A. (2022). Rheumatoid Factor and Anti-Modified Protein Antibody Reactivities Converge on IgG Epitopes. *Arthritis Rheumatol*, *74*(6), 984-991. https://doi.org/10.1002/art.42064

Mishra, N., Huang, X., Joshi, S., Guo, C., Ng, J., Thakkar, R., Wu, Y., Dong, X., Li, Q., Pinapati, R. S., Sullivan, E., Caciula, A., Tokarz, R., Briese, T., Lu, J., & Lipkin, W. I. (2021). Immunoreactive peptide maps of SARS-CoV-2. *Commun Biol*, *4*(1), 225. https://doi.org/10.1038/s42003-021-01743-9

Morris, Z. S., Guy, E. I., Francis, D. M., Gressett, M. M., Werner, L. R., Carmichael, L. L., Yang, R. K., Armstrong, E. A., Huang, S., Navid, F., Gillies, S. D., Korman, A., Hank, J. A., Rakhmilevich, A. L., Harari, P. M., & Sondel, P. M. (2016). In Situ Tumor Vaccination by Combining Local Radiation and Tumor-Specific Antibody or Immunocytokine Treatments. *Cancer Res*, *76*(13), 3929-3941. https://doi.org/10.1158/0008-5472.CAN-15-2644

Morris, Z. S., Guy, E. I., Werner, L. R., Carlson, P. M., Heinze, C. M., Kler, J. S., Busche, S. M., Jaquish, A. A., Sriramaneni, R. N., Carmichael, L. L., Loibner, H., Gillies, S. D., Korman, A. J., Erbe, A. K., Hank, J. A., Rakhmilevich, A. L., Harari, P. M., & Sondel, P. M. (2018). Tumor-Specific Inhibition of In Situ Vaccination by Distant Untreated Tumor Sites. *Cancer Immunol Res*, *6*(7), 825-834. https://doi.org/10.1158/2326-6066.CIR-17-0353

Mose, L. E., Selitsky, S. R., Bixby, L. M., Marron, D. L., Iglesia, M. D., Serody, J. S., Perou, C. M., Vincent, B. G., & Parker, J. S. (2016). Assembly-based inference of B-cell receptor repertoires from short read RNA sequencing data with V'DJer. *Bioinformatics*, *32*(24), 3729-3734. https://doi.org/10.1093/bioinformatics/btw526

Mujoo, K., Cheresh, D. A., Yang, H. M., & Reisfeld, R. A. (1987). Disialoganglioside GD2 on human neuroblastoma cells: target antigen for monoclonal antibody-mediated cytolysis and suppression of tumor growth. *Cancer Res*, *47*(4), 1098-1104. https://www.ncbi.nlm.nih.gov/pubmed/3100030

Murtagh, F. (1985). *Multidimensional clustering algorithms*. Physica-Verlag.

Nagele, E. P., Han, M., Acharya, N. K., DeMarshall, C., Kosciuk, M. C., & Nagele, R. G. (2013). Natural IgG autoantibodies are abundant and ubiquitous in human sera, and their number is influenced by age, gender, and disease. *PLoS One*, *8*(4), e60726. https://doi.org/10.1371/journal.pone.0060726

Nahtman, T., Jernberg, A., Mahdavifar, S., Zerweck, J., Schutkowski, M., Maeurer, M., & Reilly, M. (2007). Validation of peptide epitope microarray experiments and extraction of quality data. *J Immunol Methods*, *328*(1-2), 1-13. https://doi.org/10.1016/j.jim.2007.07.015

Navid, F., Santana, V. M., & Barfield, R. C. (2010). Anti-GD2 antibody therapy for GD2-expressing tumors. *Curr Cancer Drug Targets*, *10*(2), 200-209. https://doi.org/10.2174/156800910791054167

Nazha, B., Inal, C., & Owonikoko, T. K. (2020). Disialoganglioside GD2 Expression in Solid Tumors and Role as a Target for Cancer Therapy. *Front Oncol*, *10*, 1000. https://doi.org/10.3389/fonc.2020.01000

Nedzi, L. A. (2008). The implementation of ablative hypofractionated radiotherapy for stereotactic treatments in the brain and body: observations on efficacy and toxicity in clinical practice. *Semin Radiat Oncol*, *18*(4), 265-272. https://doi.org/10.1016/j.semradonc.2008.04.009

Nelson, A. L., Dhimolea, E., & Reichert, J. M. (2010). Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov*, *9*(10), 767-774. https://doi.org/10.1038/nrd3229

Neri, D., & Sondel, P. M. (2016). Immunocytokines for cancer treatment: past, present and future. *Curr Opin Immunol*, *40*, 96-102. https://doi.org/10.1016/j.coi.2016.03.006

North, R. J. (1986). Radiation-induced, immunologically mediated regression of an established tumor as an example of successful therapeutic immunomanipulation. Preferential elimination of suppressor T cells allows sustained production of effector T cells. *J Exp Med*, *164*(5), 1652-1666. https://doi.org/10.1084/jem.164.5.1652

Nzula, S., Going, J. J., & Stott, D. I. (2003). Antigen-driven clonal proliferation, somatic hypermutation, and selection of B lymphocytes infiltrating human ductal breast carcinomas. *Cancer Res*, *63*(12), 3275-3280. https://www.ncbi.nlm.nih.gov/pubmed/12810659

Obeid, M., Panaretakis, T., Joza, N., Tufi, R., Tesniere, A., van Endert, P., Zitvogel, L., & Kroemer, G. (2007). Calreticulin exposure is required for the immunogenicity of gamma-irradiation and UVC light-induced apoptosis. *Cell Death Differ*, *14*(10), 1848-1850. https://doi.org/10.1038/sj.cdd.4402201

Ong, Y., Huey, C. W. T., & Shelat, V. G. (2022). Paraneoplastic syndromes in hepatocellular carcinoma: a review. *Expert Rev Gastroenterol Hepatol*, *16*(5), 449-471. https://doi.org/10.1080/17474124.2022.2085556

Osinska, I., Popko, K., & Demkow, U. (2014). Perforin: an important player in immune response. *Cent Eur J Immunol*, *39*(1), 109-115. https://doi.org/10.5114/ceji.2014.42135

Pace, J. L., Russell, S. W., Torres, B. A., Johnson, H. M., & Gray, P. W. (1983). Recombinant mouse gamma interferon induces the priming step in macrophage activation for tumor cell killing. *J Immunol*, *130*(5), 2011-2013. https://www.ncbi.nlm.nih.gov/pubmed/6403616

Paliard, X., de Waal Malefijt, R., Yssel, H., Blanchard, D., Chretien, I., Abrams, J., de Vries, J., & Spits, H. (1988). Simultaneous production of IL-2, IL-4, and IFN-gamma by activated human CD4+ and CD8+ T cell clones. *J Immunol*, *141*(3), 849-855. https://www.ncbi.nlm.nih.gov/pubmed/2969394

Palucka, A. K., & Coussens, L. M. (2016). The Basis of Oncoimmunology. *Cell*, *164*(6), 1233-1247. https://doi.org/10.1016/j.cell.2016.01.049

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, *40*(12), 1413-1415. https://doi.org/10.1038/ng.259

Pande, J., Szewczyk, M. M., & Grover, A. K. (2010). Phage display: concept, innovations, applications and future. *Biotechnol Adv*, *28*(6), 849-858. https://doi.org/10.1016/j.biotechadv.2010.07.004

Patel, S. A., & Minn, A. J. (2018). Combination Cancer Therapy with Immune Checkpoint Blockade: Mechanisms and Strategies. *Immunity*, *48*(3), 417-433. https://doi.org/10.1016/j.immuni.2018.03.007

Penagarikano, O., Abrahams, B. S., Herman, E. I., Winden, K. D., Gdalyahu, A., Dong, H., Sonnenblick, L. I., Gruver, R., Almajano, J., Bragin, A., Golshani, P., Trachtenberg, J. T., Peles, E., & Geschwind, D. H. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell*, *147*(1), 235-246. https://doi.org/10.1016/j.cell.2011.08.040

Perez Horta, Z., Goldberg, J. L., & Sondel, P. M. (2016). Anti-GD2 mAbs and next-generation mAb-based agents for cancer therapy. *Immunotherapy*, *8*(9), 1097-1117. https://doi.org/10.2217/imt-2016-0021

Petersen, G., Song, D., Hugle-Dorr, B., Oldenburg, I., & Bautz, E. K. (1995). Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragment phage display libraries. *Mol Gen Genet*, *249*(4), 425-431. https://doi.org/10.1007/BF00287104

Pitzalis, C., Jones, G. W., Bombardieri, M., & Jones, S. A. (2014). Ectopic lymphoid-like structures in infection, cancer and autoimmunity. *Nat Rev Immunol*, *14*(7), 447-462. https://doi.org/10.1038/nri3700

Poole, W., Gibbs, D. L., Shmulevich, I., Bernard, B., & Knijnenburg, T. A. (2016). Combining dependent P-values with an empirical adaptation of Brown's method. *Bioinformatics*, *32*(17), i430-i436. https://doi.org/10.1093/bioinformatics/btw438

Poot, M. (2015). Connecting the CNTNAP2 Networks with Neurodevelopmental Disorders. *Mol Syndromol*, *6*(1), 7-22. https://doi.org/10.1159/000371594

Potluri, H. K., Ng, T. L., Newton, M. A., & McNeel, D. G. (2022). GM-CSF elicits antibodies to tumor-associated proteins when used as a prostate cancer vaccine adjuvant. *Cancer Immunology, Immunotherapy*, *71*(9), 2267-2275. https://doi.org/10.1007/s00262-022-03150-3

Potluri, H. K., Ng, T. L., Newton, M. A., Zhang, J., Maher, C. A., Nelson, P. S., & McNeel, D. G. (2020). Antibody profiling of patients with prostate cancer reveals differences in antibody signatures among disease stages. *J Immunother Cancer*, *8*(2). https://doi.org/10.1136/jitc-2020-001510

Prebet, T., Lhoumeau, A. C., Arnoulet, C., Aulas, A., Marchetto, S., Audebert, S., Puppo, F., Chabannon, C., Sainty, D., Santoni, M. J., Sebbagh, M., Summerour, V., Huon, Y., Shin, W. S., Lee, S. T., Esterni, B., Vey, N., & Borg, J. P. (2010). The cell polarity PTK7 receptor acts as a modulator of the chemotherapeutic response in acute myeloid leukemia and impairs clinical outcome. *Blood*, *116*(13), 2315-2323. https://doi.org/10.1182/blood-2010-01-262352

Puppo, F., Thome, V., Lhoumeau, A. C., Cibois, M., Gangar, A., Lembo, F., Belotti, E., Marchetto, S., Lecine, P., Prebet, T., Sebbagh, M., Shin, W. S., Lee, S. T., Kodjabachian, L., & Borg, J. P. (2011). Protein tyrosine kinase 7 has a conserved role in Wnt/beta-catenin canonical signalling. *EMBO Rep*, *12*(1), 43-49. https://doi.org/10.1038/embor.2010.185

Qin, Z., & Blankenstein, T. (2000). CD4+ T cell--mediated tumor rejection involves inhibition of angiogenesis that is dependent on IFN gamma receptor expression by nonhematopoietic cells. *Immunity*, *12*(6), 677-686. https://doi.org/10.1016/s1074-7613(00)80218-6

R-Core-Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Renard, B. Y., Löwer, M., Kühne, Y., Reimer, U., Rothermel, A., Türeci, Ö., Castle, J. C., & Sahin, U. (2011). rapmad: Robust analysis of peptide microarray data. *BMC Bioinformatics*, *12*(1), 324. https://doi.org/10.1186/1471-2105-12-324

Reuschenbach, M., von Knebel Doeberitz, M., & Wentzensen, N. (2009). A systematic review of humoral immune responses against tumor antigens. *Cancer Immunol Immunother*, *58*(10), 1535-1544. https://doi.org/10.1007/s00262-009-0733-4

Rieber, M., & Rieber, M. S. (2008). Sensitization to radiation-induced DNA damage accelerates loss of bcl-2 and increases apoptosis and autophagy. *Cancer Biol Ther*, *7*(10), 1561-1566. https://doi.org/10.4161/cbt.7.10.6540

Rivera, A., Chen, C. C., Ron, N., Dougherty, J. P., & Ron, Y. (2001). Role of B cells as antigen-presenting cells in vivo revisited: antigen-specific B cells are essential for T cell expansion in lymph nodes and for systemic T cell responses to low antigen concentrations. *Int Immunol*, *13*(12), 1583-1593. https://doi.org/10.1093/intimm/13.12.1583

Robert, C., Long, G. V., Brady, B., Dutriaux, C., Maio, M., Mortier, L., Hassel, J. C., Rutkowski, P., McNeil, C., Kalinka-Warzocha, E., Savage, K. J., Hernberg, M. M., Lebbe, C., Charles, J., Mihalcioiu, C., Chiarion-Sileni, V., Mauch, C., Cognetti, F., Arance, A., . . . Ascierto, P. A. (2015). Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*, *372*(4), 320-330. https://doi.org/10.1056/NEJMoa1412082

Robert, C., Schachter, J., Long, G. V., Arance, A., Grob, J. J., Mortier, L., Daud, A., Carlino, M. S., McNeil, C., Lotem, M., Larkin, J., Lorigan, P., Neyns, B., Blank, C. U., Hamid, O., Mateus, C., Shapira-Frommer, R., Kosh, M., Zhou, H., . . . investigators, K.-. (2015). Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N Engl J Med*, *372*(26), 2521-2532. https://doi.org/10.1056/NEJMoa1503093

Robert, C., Thomas, L., Bondarenko, I., O'Day, S., Weber, J., Garbe, C., Lebbe, C., Baurain, J. F., Testori, A., Grob, J. J., Davidson, N., Richards, J., Maio, M., Hauschild, A., Miller, W. H., Jr., Gascon, P., Lotem, M., Harmankaya, K., Ibrahim, R., . . . Wolchok, J. D. (2011). Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med*, *364*(26), 2517-2526. https://doi.org/10.1056/NEJMoa1104621

Rockberg, J., Lofblom, J., Hjelm, B., Uhlen, M., & Stahl, S. (2008). Epitope mapping of antibodies using bacterial surface display. *Nat Methods*, *5*(12), 1039-1045. https://doi.org/10.1038/nmeth.1272

Rodriguez-Rocha, H., Garcia-Garcia, A., Panayiotidis, M. I., & Franco, R. (2011). DNA damage and autophagy. *Mutat Res*, *711*(1-2), 158-166. https://doi.org/10.1016/j.mrfmmm.2011.03.007

Rosenberg, S. A. (2014). IL-2: the first effective immunotherapy for human cancer. *J Immunol*, *192*(12), 5451-5458. https://doi.org/10.4049/jimmunol.1490019

Rossetti, R. A. M., Lorenzi, N. P. C., Yokochi, K., Rosa, M., Benevides, L., Margarido, P. F. R., Baracat, E. C., Carvalho, J. P., Villa, L. L., & Lepique, A. P. (2018). B lymphocytes can be activated to act as antigen presenting cells to promote anti-tumor responses. *PLoS One*, *13*(7), e0199034. https://doi.org/10.1371/journal.pone.0199034

Rubtsov, A. V., Rubtsova, K., Kappler, J. W., Jacobelli, J., Friedman, R. S., & Marrack, P. (2015). CD11c-Expressing B Cells Are Located at the T Cell/B Cell Border in Spleen and Are Potent APCs. *J Immunol*, *195*(1), 71-79. https://doi.org/10.4049/jimmunol.1500055

Sahin, U., Oehm, P., Derhovanessian, E., Jabulowsky, R. A., Vormehr, M., Gold, M., Maurus, D., Schwarck-Kokarakis, D., Kuhn, A. N., Omokoko, T., Kranz, L. M., Diken, M., Kreiter, S., Haas, H., Attig, S., Rae, R., Cuk, K., Kemmer-Bruck, A., Breitkreuz, A., . . . Tureci, O. (2020). An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. *Nature*, *585*(7823), 107-112. https://doi.org/10.1038/s41586-020-2537-9

Sahin, U., Tureci, O., Schmitt, H., Cochlovius, B., Johannes, T., Schmits, R., Stenner, F., Luo, G., Schobert, I., & Pfreundschuh, M. (1995). Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc Natl Acad Sci U S A*, *92*(25), 11810-11813. https://doi.org/10.1073/pnas.92.25.11810

Sahlstrom, P., Hansson, M., Steen, J., Amara, K., Titcombe, P. J., Forsstrom, B., Stalesen, R., Israelsson, L., Piccoli, L., Lundberg, K., Klareskog, L., Mueller, D. L., Catrina, A. I., Skriner, K., Malmstrom, V., & Gronwall, C. (2020). Different Hierarchies of Anti-Modified Protein Autoantibody Reactivities in Rheumatoid Arthritis. *Arthritis Rheumatol*, *72*(10), 1643-1657. https://doi.org/10.1002/art.41385

Sallusto, F., Mackay, C. R., & Lanzavecchia, A. (2000). The role of chemokine receptors in primary, effector, and memory immune responses. *Annu Rev Immunol*, *18*, 593-620. https://doi.org/10.1146/annurev.immunol.18.1.593

Sariola, H., Terava, H., Rapola, J., & Saarinen, U. M. (1991). Cell-surface ganglioside GD2 in the immunohistochemical detection and differential diagnosis of neuroblastoma. *Am J Clin Pathol*, *96*(2), 248-252. https://doi.org/10.1093/ajcp/96.2.248

Sasaki, K., Harada, M., Miyashita, Y., Tagawa, H., Kishimura, A., Mori, T., & Katayama, Y. (2020). Fc-binding antibody-recruiting molecules exploit endogenous antibodies for anti-tumor immune responses. *Chem Sci*, *11*(12), 3208-3214. https://doi.org/10.1039/d0sc00017e

Sautes-Fridman, C., Petitprez, F., Calderaro, J., & Fridman, W. H. (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat Rev Cancer*, *19*(6), 307-325. https://doi.org/10.1038/s41568-019-0144-6

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270*(5235), 467-470. https://doi.org/10.1126/science.270.5235.467

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., & Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A*, *93*(20), 10614-10619. https://doi.org/10.1073/pnas.93.20.10614

Schnaar, R. L., Suzuki, A., & Stanley, P. (2009). Glycosphingolipids. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). https://www.ncbi.nlm.nih.gov/pubmed/20301240

Schneider, T. D., & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, *18*(20), 6097-6100. https://doi.org/10.1093/nar/18.20.6097

Schoorl, R., Riviere, A. B., Borne, A. E., & Feltkamp-Vroom, T. M. (1976). Identification of T and B lymphocytes in human breast cancer with immunohistochemical techniques. *Am J Pathol*, *84*(3), 529-544. https://www.ncbi.nlm.nih.gov/pubmed/183507

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337-342. https://doi.org/10.1038/nature10098

Schwartzentruber, D. J., Lawson, D. H., Richards, J. M., Conry, R. M., Miller, D. M., Treisman, J., Gailani, F., Riley, L., Conlon, K., Pockaj, B., Kendra, K. L., White, R. L., Gonzalez, R., Kuzel, T. M., Curti, B., Leming, P. D., Whitman, E. D., Balkissoon, J., Reintgen, D. S., . . . Hwu, P. (2011). gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N Engl J Med*, *364*(22), 2119-2127. https://doi.org/10.1056/NEJMoa1012863

Sevilla, M. D., Becker, D., Kumar, A., & Adhikary, A. (2016). Gamma and Ion-Beam Irradiation of DNA: Free Radical Mechanisms, Electron Effects, and Radiation Chemical Track Structure. *Radiat Phys Chem Oxf Engl 1993*, *128*, 60-74. https://doi.org/10.1016/j.radphyschem.2016.04.022

Sharonov, G. V., Serebrovskaya, E. O., Yuzhakova, D. V., Britanova, O. V., & Chudakov, D. M. (2020). B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nat Rev Immunol*, *20*(5), 294-307. https://doi.org/10.1038/s41577-019-0257-x

Shen, L., Zhang, J., Lee, H., Batista, M. T., & Johnston, S. A. (2019). RNA Transcription and Splicing Errors as a Source of Cancer Frameshift Neoantigens for Vaccines. *Sci Rep*, *9*(1), 14184. https://doi.org/10.1038/s41598-019-50738-4

Shen, S., Kai, B., Ruan, J., Torin Huzil, J., Carpenter, E., & Tuszynski, J. A. (2006). Probabilistic analysis of the frequencies of amino acid pairs within characterized protein sequences. *Physica A*, *370*(2), 651-662. https://doi.org/10.1016/j.physa.2006.03.004

Shen, X., Jin, J., Ding, Y., Wang, P., Wang, A., Xiao, D., Xue, X., Zhu, S., Zhang, L., & Zhu, G. (2013). Novel immunodominant epitopes derived from MAGE-A3 and its significance in serological diagnosis of gastric cancer. *J Cancer Res Clin Oncol*, *139*(9), 1529-1538. https://doi.org/10.1007/s00432-013-1463-8

Shibuya, H., Hamamura, K., Hotta, H., Matsumoto, Y., Nishida, Y., Hattori, H., Furukawa, K., Ueda, M., & Furukawa, K. (2012). Enhancement of malignant properties of human osteosarcoma cells with disialyl gangliosides GD2/GD3. *Cancer Sci*, *103*(9), 1656-1664. https://doi.org/10.1111/j.1349-7006.2012.02344.x

Siebert, N., Seidel, D., Eger, C., Juttner, M., & Lode, H. N. (2014). Functional bioassays for immune monitoring of high-risk neuroblastoma patients treated with ch14.18/CHO anti-GD2 antibody. *PLoS One*, *9*(9), e107692. https://doi.org/10.1371/journal.pone.0107692

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA Cancer J Clin*, *72*(1), 7-33. https://doi.org/10.3322/caac.21708

Silagi, S. (1969). Control of pigment production in mouse melanoma cells in vitro. Evocation and maintenance. *J Cell Biol*, *43*(2), 263-274. https://doi.org/10.1083/jcb.43.2.263

Silagi, S., Beju, D., Wrathall, J., & Deharven, E. (1972). Tumorigenicity, immunogenicity, and virus production in mouse melanoma cells treated with 5-bromodeoxyuridine. *Proc Natl Acad Sci U S A*, *69*(11), 3443-3447. https://doi.org/10.1073/pnas.69.11.3443

Stewart, F. W. (1933). The Diagnosis of Tumors by Aspiration. *Am J Pathol*, *9*(Suppl), 801-812 803. https://www.ncbi.nlm.nih.gov/pubmed/19970115

Stockert, E., Jager, E., Chen, Y. T., Scanlan, M. J., Gout, I., Karbach, J., Arand, M., Knuth, A., & Old, L. J. (1998). A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *J Exp Med*, *187*(8), 1349-1354. https://doi.org/10.1084/jem.187.8.1349

Stone, H. B., Peters, L. J., & Milas, L. (1979). Effect of host immune capability on radiocurability and subsequent transplantability of a murine fibrosarcoma. *J Natl Cancer Inst*, *63*(5), 1229-1235. https://www.ncbi.nlm.nih.gov/pubmed/291749

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902 e1821. https://doi.org/10.1016/j.cell.2019.05.031

Sun, C., Kovacs, P., & Guiu-Jurado, E. (2021). Genetics of Body Fat Distribution: Comparative Analyses in Populations with European, Asian and African Ancestries. *Genes (Basel)*, *12*(6). https://doi.org/10.3390/genes12060841

Sun, R., & Lin, X. (2020). Genetic Variant Set-Based Tests Using the Generalized Berk-Jones Statistic with Application to a Genome-Wide Association Study of Breast Cancer. *J Am Stat Assoc*, *115*(531), 1079-1091. https://doi.org/10.1080/01621459.2019.1660170

Suresh, M., Whitmire, J. K., Harrington, L. E., Larsen, C. P., Pearson, T. C., Altman, J. D., & Ahmed, R. (2001). Role of CD28-B7 interactions in generation and maintenance of CD8 T cell memory. *J Immunol*, *167*(10), 5565-5573. https://doi.org/10.4049/jimmunol.167.10.5565

Svennerholm, L. (1963). Chromatographic Separation of Human Brain Gangliosides. *J Neurochem*, *10*, 613-623. https://doi.org/10.1111/j.1471-4159.1963.tb08933.x

Takeda, K., Smyth, M. J., Cretney, E., Hayakawa, Y., Kayagaki, N., Yagita, H., & Okumura, K. (2002). Critical role for tumor necrosis factor-related apoptosis-inducing ligand in immune surveillance against tumor development. *J Exp Med*, *195*(2), 161-169. https://doi.org/10.1084/jem.20011171

Tarp, M. A., Sorensen, A. L., Mandel, U., Paulsen, H., Burchell, J., Taylor-Papadimitriou, J., & Clausen, H. (2007). Identification of a novel cancer-specific immunodominant glycopeptide epitope in the MUC1 tandem repeat. *Glycobiology*, *17*(2), 197-209. https://doi.org/10.1093/glycob/cwl061

Tas, F., Keskin, S., Karadeniz, A., Dagoglu, N., Sen, F., Kilic, L., & Yildiz, I. (2011). Noncutaneous melanoma have distinct features from each other and cutaneous melanoma. *Oncology*, *81*(5-6), 353-358. https://doi.org/10.1159/000334863

Telenti, A., Pierce, L. C., Biggs, W. H., di Iulio, J., Wong, E. H., Fabani, M. M., Kirkness, E. F., Moustafa, A., Shah, N., Xie, C., Brewerton, S. C., Bulsara, N., Garner, C., Metzker, G., Sandoval, E., Perkins, B. A., Och, F. J., Turpaz, Y., & Venter, J. C. (2016). Deep sequencing of 10,000 human genomes. *Proc Natl Acad Sci U S A*, *113*(42), 11901-11906. https://doi.org/10.1073/pnas.1613365113

The UniProt, C. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, *45*(D1), D158-D169. https://doi.org/10.1093/nar/gkw1099

Thompson, L. H. (2012). Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat Res*, *751*(2), 158-246. https://doi.org/10.1016/j.mrrev.2012.06.002

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic [https://doi.org/10.1111/1467-9868.00293]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423. https://doi.org/https://doi.org/10.1111/1467-9868.00293

Tippett, L. H. C. (1931). *The Methods of Statistics*. London: Williams & Norgate Ltd.

Trott, K. R. (1982). Experimental results and clinical implications of the four R's in fractionated radiotherapy. *Radiat Environ Biophys*, *20*(3), 159-170. https://doi.org/10.1007/BF01325465

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L., & Ponten, F. (2010). Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*, *28*(12), 1248-1250. https://doi.org/10.1038/nbt1210-1248

Vaage, J. (1987). Local and systemic effects during interleukin-2 therapy of mouse mammary tumors. *Cancer Res*, *47*(16), 4296-4298. https://www.ncbi.nlm.nih.gov/pubmed/3496962

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., & Seeberger, P. H. (2022). Essentials of Glycobiology. In A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, & P. H. Seeberger (Eds.), *Essentials of Glycobiology* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/9781621824213

Vatapalli, R., Sagar, V., Rodriguez, Y., Zhao, J. C., Unno, K., Pamarthy, S., Lysy, B., Anker, J., Han, H., Yoo, Y. A., Truica, M., Chalmers, Z. R., Giles, F., Yu, J., Chakravarti, D., Carneiro, B., & Abdulkadir, S. A. (2020). Histone methyltransferase DOT1L coordinates AR and MYC stability in prostate cancer. *Nat Commun*, *11*(1), 4153. https://doi.org/10.1038/s41467-020-18013-7

Villagran-Garcia, M., Muniz-Castrillo, S., Ciano-Petersen, N. L., Vogrig, A., Farina, A., Villard, M., Psimaras, D., Alentorn, A., Goncalves, D., Fabien, N., Rogemond, V., Joubert, B., & Honnorat, J. (2023). Paraneoplastic neurological syndromes associated with renal or bladder cancer: case series and PRISMA-IPD systematic review. *J Neurol*, *270*(1), 283-299. https://doi.org/10.1007/s00415-022-11356-9

Vodnik, M., Zager, U., Strukelj, B., & Lunder, M. (2011). Phage display: selecting straws instead of a needle from a haystack. *Molecules*, *16*(1), 790-817. https://doi.org/10.3390/molecules16010790

Voeller, J., & Sondel, P. M. (2019). Advances in Anti-GD2 Immunotherapy for Treatment of High-risk Neuroblastoma. *J Pediatr Hematol Oncol*, *41*(3), 163-169. https://doi.org/10.1097/MPH.0000000000001369

von Reibnitz, D., Shaikh, F., Wu, A. J., Treharne, G. C., Dick-Godfrey, R., Foster, A., Woo, K. M., Shi, W., Zhang, Z., Din, S. U., Gelblum, D. Y., Yorke, E. D., Rosenzweig, K. E., & Rimner, A. (2018). Stereotactic body radiation therapy (SBRT) improves local control and overall survival compared to conventionally fractionated radiation for stage I non-small cell lung cancer (NSCLC). *Acta Oncol*, *57*(11), 1567-1573. https://doi.org/10.1080/0284186X.2018.1481292

Wang, B., DeKosky, B. J., Timm, M. R., Lee, J., Normandin, E., Misasi, J., Kong, R., McDaniel, J. R., Delidakis, G., Leigh, K. E., Niezold, T., Choi, C. W., Viox, E. G., Fahad, A., Cagigi, A., Ploquin, A., Leung, K., Yang, E. S., Kong, W. P., . . . Georgiou, G. (2018). Functional interrogation and

mining of natively paired human V(H):V(L) antibody repertoires. *Nat Biotechnol*, *36*(2), 152-155. https://doi.org/10.1038/nbt.4052

Wang, J., Feng, Y., Sun, T., Zhang, Q., & Chai, Y. (2022). Photolabile 2-(2-Nitrophenyl)-propyloxycarbonyl (NPPOC) for Stereoselective Glycosylation and Its Application in Consecutive Assembly of Oligosaccharides. *The Journal of Organic Chemistry*, *87*(5), 3402-3421. https://doi.org/10.1021/acs.joc.1c03006

Wang, K., Li, H., Chen, R., Zhang, Y., Sun, X. X., Huang, W., Bian, H., & Chen, Z. N. (2017). Combination of CALR and PDIA3 is a potential prognostic biomarker for non-small cell lung cancer. *Oncotarget*, *8*(57), 96945-96957. https://doi.org/10.18632/oncotarget.18547

Weide, B., Derhovanessian, E., Pflugfelder, A., Eigentler, T. K., Radny, P., Zelba, H., Pfohler, C., Pawelec, G., & Garbe, C. (2010). High response rate after intratumoral treatment with interleukin-2: results from a phase 2 study in 51 patients with metastasized melanoma. *Cancer*, *116*(17), 4139-4146. https://doi.org/10.1002/cncr.25156

Wen, D., Wang, L., Tan, S., Tang, R., Xie, W., Liu, S., Tang, C., & He, Y. (2022). HOXD9 aggravates the development of cervical cancer by transcriptionally activating HMCN1. *Panminerva Med*, *64*(4), 532-536. https://doi.org/10.23736/S0031-0808.20.03911-7

Werner, L. R., Kler, J. S., Gressett, M. M., Riegert, M., Werner, L. K., Heinze, C. M., Kern, J. G., Abbariki, M., Erbe, A. K., Patel, R. B., Sriramaneni, R. N., Harari, P. M., & Morris, Z. S. (2017). Transcriptional-mediated effects of radiation on the expression of immune susceptibility markers in melanoma. *Radiother Oncol*, *124*(3), 418-426. https://doi.org/10.1016/j.radonc.2017.08.016

Whelan, T. J., Pignol, J. P., Levine, M. N., Julian, J. A., MacKenzie, R., Parpia, S., Shelley, W., Grimard, L., Bowen, J., Lukka, H., Perera, F., Fyles, A., Schneider, K., Gulavita, S., & Freeman, C. (2010). Long-term results of hypofractionated radiation therapy for breast cancer. *N Engl J Med*, *362*(6), 513-520. https://doi.org/10.1056/NEJMoa0906260

Wilkinson, B. (1951). A statistical consideration in psychological research. *Psychol Bull*, *48*(3), 156-158. https://doi.org/10.1037/h0059111

Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proc Natl Acad Sci U S A*, *116*(4), 1195-1200. https://doi.org/10.1073/pnas.1814092116

Winkler, D. F., Andresen, H., & Hilpert, K. (2011). SPOT synthesis as a tool to study protein-protein interactions. *Methods Mol Biol*, *723*, 105-127. https://doi.org/10.1007/978-1-61779-043-0_8

Wondimu, A., Zhang, T., Kieber-Emmons, T., Gimotty, P., Sproesser, K., Somasundaram, R., Ferrone, S., Tsao, C. Y., & Herlyn, D. (2008). Peptides mimicking GD2 ganglioside elicit cellular, humoral and tumor-protective immune responses in mice. *Cancer Immunol Immunother*, *57*(7), 1079-1089. https://doi.org/10.1007/s00262-007-0439-4

Wrenshall, L. E., Stevens, R. B., Cerra, F. B., & Platt, J. L. (1999). Modulation of macrophage and B cell function by glycosaminoglycans. *J Leukoc Biol*, *66*(3), 391-400. https://doi.org/10.1002/jlb.66.3.391

Yan, Y., Sun, N., Wang, H., Kobayashi, M., Ladd, J. J., Long, J. P., Lo, K. C., Patel, J., Sullivan, E., Albert, T., Goodman, G. E., Do, K. A., & Hanash, S. M. (2019). Whole Genome-Derived Tiled Peptide

Arrays Detect Prediagnostic Autoantibody Signatures in Non-Small-Cell Lung Cancer. *Cancer Res*, *79*(7), 1549-1557. https://doi.org/10.1158/0008-5472.CAN-18-1536

Yang, A., Farmer, E., Lin, J., Wu, T. C., & Hung, C. F. (2017). The current state of therapeutic and T cell-based vaccines against human papillomaviruses. *Virus Res*, *231*, 148-165. https://doi.org/10.1016/j.virusres.2016.12.002

Yang, H. Y., Erdos, E. G., & Levin, Y. (1970). A dipeptidyl carboxypeptidase that converts angiotensin I and inactivates bradykinin. *Biochim Biophys Acta*, *214*(2), 374-376. https://doi.org/10.1016/0005-2795(70)90017-6

Yang, R. K., Kalogriopoulos, N. A., Rakhmilevich, A. L., Ranheim, E. A., Seo, S., Kim, K., Alderson, K. L., Gan, J., Reisfeld, R. A., Gillies, S. D., Hank, J. A., & Sondel, P. M. (2012). Intratumoral hu14.18-IL-2 (IC) induces local and systemic antitumor effects that involve both activated T and NK cells as well as enhanced IC retention. *J Immunol*, *189*(5), 2656-2664. https://doi.org/10.4049/jimmunol.1200934

Yarmarkovich, M., Marshall, Q. F., Warrington, J. M., Premaratne, R., Farrel, A., Groff, D., Li, W., di Marco, M., Runbeck, E., Truong, H., Toor, J. S., Tripathi, S., Nguyen, S., Shen, H., Noel, T., Church, N. L., Weiner, A., Kendsersky, N., Martinez, D., . . . Maris, J. M. (2021). Cross-HLA targeting of intracellular oncoproteins with peptide-centric CARs. *Nature*, *599*(7885), 477-484. https://doi.org/10.1038/s41586-021-04061-6

Yoshida, S., Fukumoto, S., Kawaguchi, H., Sato, S., Ueda, R., & Furukawa, K. (2001). Ganglioside G(D2) in small cell lung cancer cell lines: enhancement of cell proliferation and mediation of apoptosis. *Cancer Res*, *61*(10), 4244-4252. https://www.ncbi.nlm.nih.gov/pubmed/11358851

Yoshida, S., Kawaguchi, H., Sato, S., Ueda, R., & Furukawa, K. (2002). An anti-GD2 monoclonal antibody enhances apoptotic effects of anti-cancer drugs against small cell lung cancer cells via JNK (c-Jun terminal kinase) activation. *Jpn J Cancer Res*, *93*(7), 816-824. https://doi.org/10.1111/j.1349-7006.2002.tb01324.x

Yu, A. L., Gilman, A. L., Ozkaynak, M. F., London, W. B., Kreissman, S. G., Chen, H. X., Smith, M., Anderson, B., Villablanca, J. G., Matthay, K. K., Shimada, H., Grupp, S. A., Seeger, R., Reynolds, C. P., Buxton, A., Reisfeld, R. A., Gillies, S. D., Cohn, S. L., Maris, J. M., . . . Children's Oncology, G. (2010). Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. *N Engl J Med*, *363*(14), 1324-1334. https://doi.org/10.1056/NEJMoa0911123

Yu, A. L., Gilman, A. L., Ozkaynak, M. F., Naranjo, A., Diccianni, M. B., Gan, J., Hank, J. A., Batova, A., London, W. B., Tenney, S. C., Smith, M., Shulkin, B. L., Parisi, M., Matthay, K. K., Cohn, S. L., Maris, J. M., Bagatell, R., Park, J. R., & Sondel, P. M. (2021). Long-Term Follow-up of a Phase III Study of ch14.18 (Dinutuximab) + Cytokine Immunotherapy in Children with High-Risk Neuroblastoma: COG Study ANBL0032. *Clin Cancer Res*, *27*(8), 2179-2189. https://doi.org/10.1158/1078-0432.CCR-20-3909

Yu, R. K., Tsai, Y. T., Ariga, T., & Yanagisawa, M. (2011). Structures, biosynthesis, and functions of gangliosides--an overview. *J Oleo Sci*, *60*(10), 537-544. https://doi.org/10.5650/jos.60.537

Yui, M. A., Sharp, L. L., Havran, W. L., & Rothenberg, E. V. (2004). Preferential activation of an IL-2 regulatory sequence transgene in TCR gamma delta and NKT cells: subset-specific

differences in IL-2 regulation. *J Immunol*, *172*(8), 4691-4699. https://doi.org/10.4049/jimmunol.172.8.4691

Zandian, A., Forsstrom, B., Haggmark-Manberg, A., Schwenk, J. M., Uhlen, M., Nilsson, P., & Ayoglu, B. (2017). Whole-Proteome Peptide Microarrays for Profiling Autoantibody Repertoires within Multiple Sclerosis and Narcolepsy. *J Proteome Res*, *16*(3), 1300-1314. https://doi.org/10.1021/acs.jproteome.6b00916

Zarei, S., Bayat, A. A., Hadavi, R., Mahmoudi, A. R., Tavangar, B., Vojgani, Y., Jeddi-Tehrani, M., & Amirghofran, Z. (2015). Production and characterization of a peptide-based monoclonal antibody against CD44 variant 6. *Monoclon Antib Immunodiagn Immunother*, *34*(1), 36-43. https://doi.org/10.1089/mab.2014.0077

Zerweck, J., Reimer, U., Jansong, J., Pawlowski, N., Tersch, C., Eckey, M., & Knaute, T. (2016). High-Throughput Microarray Incubations Using Multi-Well Chambers. *Methods Mol Biol*, *1352*, 19-26. https://doi.org/10.1007/978-1-4939-3037-1_2

Zhang, H., Zhou, Y., Cheng, Q., Dai, Z., Wang, Z., Liu, F., Fan, F., Cui, B., & Cao, H. (2020). PDIA3 correlates with clinical malignant features and immune signature in human gliomas. *Aging (Albany NY)*, *12*(15), 15392-15413. https://doi.org/10.18632/aging.103601

Zhang, J., Li, H., Li, H., Lin, D., Wang, X., & Wang, K. (2022). Expression and Prognostic Significance of PDIA3 in Cervical Cancer. *Int J Genomics*, *2022*, 4382645. https://doi.org/10.1155/2022/4382645

Zhang, J., Wang, K., Hainisayimu, T., & Li, H. (2022). Pan-Cancer Analysis of PDIA3: Identifying It as a Potential Biomarker for Tumor Prognosis and Immunotherapy. *Oxid Med Cell Longev*, *2022*, 9614819. https://doi.org/10.1155/2022/9614819

Zheng, Z., Mergaert, A., Fahmy, L., Bawadekar, M., Holmes, C., Ong, I., Bridges, A., Newton, M., & Shelef, M. (2019). Disordered Antigens and Epitope Overlap Between Anti-Citrullinated Protein Antibodies and Rheumatoid Factor in Rheumatoid Arthritis. *Arthritis & Rheumatology*, *72*. https://doi.org/10.1002/art.41074

Zheng, Z., Mergaert, A. M., Ong, I. M., Shelef, M. A., & Newton, M. A. (2021). MixTwice: large-scale hypothesis testing for peptide arrays by variance mixing. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btab162

Zhou, P., Qiu, J., L'Italien, L., Gu, D., Hodges, D., Chao, C. C., & Schebye, X. M. (2010). Mature B cells are critical to T-cell-mediated tumor immunity induced by an agonist anti-GITR monoclonal antibody. *J Immunother*, *33*(8), 789-797. https://doi.org/10.1097/CJI.0b013e3181ee6ba9

Zhu, W., Germain, C., Liu, Z., Sebastian, Y., Devi, P., Knockaert, S., Brohawn, P., Lehmann, K., Damotte, D., Validire, P., Yao, Y., Valge-Archer, V., Hammond, S. A., Dieu-Nosjean, M. C., & Higgs, B. W. (2015). A high density of tertiary lymphoid structure B cells in lung tumors is associated with increased CD4(+) T cell receptor repertoire clonality. *Oncoimmunology*, *4*(12), e1051922. https://doi.org/10.1080/2162402X.2015.1051922

Appendix A: Collaborative biostatistical-bioinformatic methodology manuscript directly related to the work presented in this thesis

Preface

This manuscript accompanies chapter 2 of this thesis as it detailes the bioinformatics method developed to analyze the serum samples of naïve mice and mice cured of B78 melanoma with immunological memory. It was necessary to develop an analysis method that reliably filters results for peptides, epitopes and proteins that show antibody binding in immune mouse serum samples but not (or to a significantly lower degree) in naïve mouse serum samples. Over the course of 4 years we worked to perfect this algorithm to accurately analyze the raw data. Many iterations with larger or smaller changes were analyzed for precision in analysis by comparing random hits from the algorithm output with the raw data and vice versa.

For this manuscript I was a vital part in developing HERON. I helped test different versions of HERON for its capabilities and completeness in capturing epitopes and proteins of interest by comparing analysis results to the raw data and using the raw data to make informed decisions of what HERON should be picking up as positive hits vs. not and where and how to set thresholds. I was involved in the conceptualization, data curation, formal analysis, investigation, methodology, validation, review, and editing process for this paper.

The manuscript has been sent out for peer review for publication and has been deposited in BioRxiv and can be found online as:

Sean J. McIlwain, Anna Hoefges, Amy K. Erbe, Paul M. Sondel, Irene M. Ong; (2023) Ranking Antibody Binding Epitopes and Proteins Across Samples from Whole Proteome Tiled Linear Peptides, bioRxiv 2023.04.23.536620; doi: https://doi.org/10.1101/2023.04.23.536620

Ranking Antibody Binding Epitopes and Proteins Across Samples from Whole Proteome Tiled Linear Peptides

Abstract

Ultradense peptide binding arrays that can probe millions of linear peptides comprising the entire proteomes or immunomes of human or mouse, or numerous microbes, are powerful tools for studying the abundance of different antibody repertoire in serum samples to understand adaptive immune responses. There are few statistical analysis tools for exploring high-dimensional, significant and reproducible antibody targets for ultradense peptide binding arrays at the linear peptide, epitope (grouping of adjacent peptides), and protein level across multiple samples/subjects (i.e. epitope spread or immunogenic regions within each protein) for understanding the heterogeneity of immune responses. We developed HERON (Hierarchical antibody binding Epitopes and pROteins from liNear peptides), an R package, which allows users to identify immunogenic epitopes using meta-analyses and spatial clustering techniques to explore antibody targets at various resolution and confidence levels, that can be found consistently across a specified number of samples through the entire proteome to study antibody responses for diagnostics or treatment. Our approach estimates significance values at the linear peptide (probe), epitope, and protein level to identify top candidates for validation. We test the performance of predictions on all three levels using correlation between technical replicates and comparison of epitope calls on 2 datasets, which shows HERON's competitiveness in estimating false discovery rates and finding general and sample-level regions of interest for antibody binding. The code is available as an R package downloadable from http://github.com/Ong-Research/HERON.

Introduction

The technology for high-dimensional identification of specific antibody binding repertoire has significantly improved over the last decade, allowing for the determination of antibody binding to ~6 million peptides simultaneously with peptide array technology to probe every mouse or human protein using 16-mer peptides with 1, 2 or 4 amino acid (aa) tiling to identify antibody targets from serum or plasma (Lyamichev et al., 2017). This extremely high-dimensional, high-throughput method for antigen-specific immune profiling can enable detection of immune responses to infection or vaccines, and potentially study the precise targeting of tumors by one's own immune system with or without immunotherapy/combination therapy. There are many existing methods for analyzing microarray and peptide array data (Flinterman et al., 2008; Heffron et al., 2021; Heffron et al., 2018; Imholte & Gottardo, 2016; Imholte et al., 2013; Lin et al., 2009; Lin et al., 2012; Lo et al., 2020; Mergaert et al., 2022; Nahtman et al., 2007; Potluri et al., 2022; Renard et al., 2011; Zheng et al., 2019; Zheng et al., 2021), including pepStat, which was designed to analyze different viral strains for a single protein, and a few methods for analyzing these ultra-dense, high-dimensional array data (Chen et al., 2022; Mergaert et al., 2022; Zheng et al., 2021), but additional methods are necessary to advance understanding of immune response as antibody binding signals from arrays represent an aggregate of complex biological and biochemical interactions. Antibodies may bind to peptides via various mechanisms of interaction based on the antibody's antigen binding site and amino acid configuration of the peptides, affecting binding affinity, and each protein may have multiple epitopes that are bound by antibodies, amplifying the immune response to the protein. These considerations are important when studying immune responses involved in diseases, including infectious, auto-immune, or neoplastic. To the best of our knowledge, there are currently no existing methods that can simultaneously identify and rank antibody binding responses at different scales, i.e., to linear peptides, epitopes (defined here as antibody bound region of adjacent linear

peptides or probes tiled within a protein) and proteins, for diagnostics or treatment; or for understanding the heterogeneity of immune response within an individual or across a population.

Existing methods for analyzing antibody binding of peptide array data such as pepStat (Imholte et al., 2013), estimates the false discovery rate (FDR) at a peptide-level and reports subject-level statistics. Bayesian models such as pepBayes have been shown to be potentially superior to pepStat but is also designed to analyze a single protein from multiple related viral strains and does not take into account the sequential probes across the protein (Imholte & Gottardo, 2016). Another Bayesian model does include provision for handling sequential probe signals by using a latent autoregressive component, however the authors propose a limit in the problem size (300 peptides and 50 samples) when using their implementation. MixTwice (Zheng et al., 2021) is a statistical method that achieves more power to detect significant probes by using local FDR. None of these methods ensure a reproducible way to rank epitopes and score both the epitopes and proteins when analyzing peptide arrays with ~6 million unique sequence probes.

We developed an algorithm, HERON (**H**ierarchical antibody binding **E**pitopes and p**RO**teins from li**N**ear peptides), which is an R (R-Core-Team, 2022) package available on GitHub, for analyzing ultra-dense peptide arrays to identify and rank significantly bound linear peptides, epitopes, and proteins. The overall workflow is shown in **Figure A.1A**. Our approach builds on existing approaches (Imholte et al., 2013) including clustering methods to locate contiguous probes (i.e., epitopes) with high binding affinity, and meta-analysis methods from Fisher and others (Becker, 1994; Dewey, 2022; Liu & Xie, 2020; Poole et al., 2016; Sun & Lin, 2020; Wilkinson, 1951; Wilson, 2019) to: 1) allow for reliability and reproducibility with granularity in confidence level for making antibody binding calls at different thresholds, 2) ensure that we identify probes that are more highly bound in post-

(experimental) compared to pre-exposure (control), and that we give more weight to those with higher overall binding, 3) identify consecutive overlapping probes with high signal and categorize the shared amino acid (aa) sequences represented by those highly recognized probes as epitopes based on specified thresholds, and 4) identify proteins due to epitope spread (i.e., identify proteins recognized by distinct epitope binding in different regions of the protein by sera from multiple samples, not necessarily at the same place on the protein).

To ensure that probes that are more highly bound in post- (experimental or positive) compared to pre-exposure (control or negative) are identified, and that probes with higher overall binding are given more weight or a smaller estimated p-value, our algorithm calculates probe-level p-value scores for each post-sample by combining t-test scores for differential expression (one sided test above control) and global z-tests normalized for relative height compared to all of the data in the experiment (**Figure A.1B**).

In order to identify common and unique epitopes across individual samples we find regions of consecutive probes across the tiling of the proteins that are consistent with their calls across positive samples by making initial antibody binding calls on the probe-level adjusted p-values based on a threshold, then utilizes one of three approaches for identifying epitopes, which are consecutive probes tiled across a single protein. After finding the epitopes, we then calculate a score for all the epitopes using a respective meta p-value estimation method (Becker, 1994; Dewey, 2022; Liu & Xie, 2020; Poole et al., 2016; Sun & Lin, 2020; Wilkinson, 1951; Wilson, 2019) and adjust for false discoveries using Benjamini-Hochberg (BH) (Benjamini & Hochberg, 1995). A similar approach using meta-analysis is applied to epitope level calls at the protein level to calculate scores for all the proteins.

In this report, we measured the performance of our method on two datasets to illustrate its utility. The goal of the first dataset (COVID-19) was to identify diagnostic epitopes across the landscape of the SARS-CoV-2 viral proteome as detected by human serum samples from individuals following SARS-CoV-2 infection (Heffron et al., 2021). The goal of the second dataset (Melanoma) was to study the landscape of antibody responses to the whole mouse proteome of genetically identical mice before cancer and after curative immunotherapeutic *in situ* vaccine treatment (Morris et al., 2016) and rechallenge with a related tumor and to rank the responses for study and validation. For evaluation of the COVID-19 dataset, we show performance measures for finding significant epitopes by comparing them with the regions identified using a t-test method from the SARS-CoV-2 manuscript (Heffron et al., 2021) and for the Melanoma dataset, we measured correlation between technical replicates using different parameter settings.

Materials and Methods

Datasets

The COVID-19 dataset (Heffron et al., 2021) compares analyses of serum samples from individuals following proven infection with SARS-CoV-2 (COVID+) with serum samples from uninfected individuals (COVID-); this dataset was downloaded following the instructions from https://github.com/Ong-Research/UW_Adult_Covid-19. There are 60 (20 COVID-, 40 COVID+) samples in total, with 118,651 unique sequence probes mapped to 470,086 16-mer probes for 387 proteins with a tiling of 1 amino acid (aa) and up to 5 replicates on the array. The data were preprocessed using pepMeld (Baker, 2020) and the resulting sequence matrix was quantile normalized at the sequence probe level.

The melanoma dataset compares naïve mouse serum to mice that have successfully rejected a tumor after immunotherapy and also showed immune memory when presented with a rechallenge

of the same or a similar tumor type. These mice had an established B78 melanoma tumor, have been treated with a potent immunotherapy treatment consisting of 12 Gy external beam radiation followed 5 days later by 5 consecutive days of intratumor injections of an immunocytokine which targets the GD2-expressing tumor via a monoclonal antibody and delivers IL-2 directly to the tumor site to elicit a strong and long-lasting anti-tumor immune response.

The Melanoma dataset was collected by Hoefges et al. 2023 (Hoefges et al., 2023) and can be found on Zenodo under the following DOI: 10.5281/zenodo.7871566.

The dataset consists of 11 biological (serum) samples tested for recognition of 6,090,593 16-mer unique sequence probes mapped to 8,459,970 protein probes using a mixed tiling of either 2 aa or 4 aa with a total of 53,640 individual proteins. Of the 11 samples, 5 samples were from naïve mice, (referred to as “pre-treatment”; i.e., prior to tumor introduction or immunotherapy), and 6 were immune samples (referred to as “post-treatment”; i.e., they were tumor-bearing mice that underwent radiation and immunotherapy treatment to become tumor-free. Following tumor-clearance, the mice were rechallenged with the tumor line to test immune memory). Two of the immune samples had technical replicates: one (sample B2) had replicates of the same Immune serum sample tested on the same array (performed within 24 h of each other) while the other (sample PD1) had the replicates of the same immune serum sample on identical arrays but performed ~ 1 year apart from each other. In total, there were 13 serum samples, 5 from naïve mice, 8 from 6 immune mice, (including the 2 replicates for B2 and PD1). The data were pre-processed by Nimble Therapeutics, quantile normalized, and then smoothed using a sliding average mean window across the protein location of +/-8 aa

Algorithm

The R package, HERON, can be downloaded and installed from GitHub (http://github.com/Ong-Research/HERON). The overall workflow of the algorithm is illustrated in **Figure A.1** including the underlying approach for finding significant probes (**Figure A.1B**). Herein, we describe the methods for finding consecutive blocks of probes on the protein, which we call epitopes, as they can be used to identify the exact sequence of 2-16 consecutive amino acids, shared by consecutive probes, all recognized by the same antibody, and thereby corresponding to an epitope recognized by an antibody (Hoefges et al., 2023). Given the found epitopes, we then describe how to estimate the significant epitopes and corresponding proteins.

Estimating probe-level p-values for each positive sample

HERON was designed to analyze ultradense peptide binding arrays with multiple protein sequences, e.g., whole mouse or human proteomes, and the goal is to rank or calculate a p-value/false discovery rate (FDR) for each linear peptide probe for each positive sample and summarize ranking at different resolutions (at epitope and protein levels) for multiple positive samples. As shown in **Figure A.1B**, we first calculate a global p-value using a distribution function for all of the data provided in the experiment, then a differential p-value using a one-sided t-test for each positive sample is calculated assuming that the standard deviation is the same between groups using the pre- (control) or post- (experimental or positive) samples. Parameters for generating probe-level p-values can be adjusted and one can use either or both of the global and differential tests when calculating significance for peptides to identify probes that are more highly bound in post- (experimental or positive) compared to pre-exposure (control) samples, or give more weight to those with higher overall binding, respectively. Our approach calculates a combined p-value from the global p-value and differential p-value using Wilkinson's max meta p-value method (Becker,

1994; Birnbaum, 1954; Tippett, 1931; Wilkinson, 1951). **Supplementary Figure A.1** shows the estimated global p-values (**SA.1A**), differential p-values (**SA.1B**), combined global and differential p-values (**SA.1C**), and adjusted p-values (Benjamini-Hochberg, **SA.1D**) versus the peptide array signal after applying the HERON workflow on data for a positive sample from the Melanoma dataset, where a low p-value/FDR indicates an increased probability that both the global and differential hypotheses are rejected.

K of N Calls and One-Hit Filter

The linear probe-level p-values are then adjusted for false discovery rates using the Benjamini-Hochberg algorithm. Next the significance values for the linear peptides are calculated and an FDR threshold cutoff is used to identify the set of antibody-bound probes (which we will also refer to as "calls" or "called" probes) for each positive sample. HERON also reports the number of samples that have an adjusted p-value less than the specified threshold, i.e., reporting a number, K samples that indicated antibody binding for the called probes, out of a total of N samples. To filter out inconsistent calls due to spurious noise (or non-specific signals), the algorithm has a one-hit filter option for removing probes that are called as bound by antibodies that do not have a supporting consecutive probe call in the same serum sample and does not have a call for the same probe in another serum sample. The one-hit filtering is a soft procedure, where the raw and adjusted p-values are set to 1 for the filtered one-hit probe. These procedures are also applied to epitope- and protein- level probes after the steps described below.

Epitope Finding

Once the set of antibody-bound probes are identified, epitopes can then be found using three different segmentation algorithms: unique, hierarchical, or skater.

The unique method iterates through all the samples, finding consecutive runs of probes that were called on the same protein, then combines all the regions found across the samples into a single list, and reports the unique set of blocks/epitopes. The method permits overlapping blocks. **FigureA. 2A** illustrates the "unique" algorithm on an example set of called probes.

Both the hierarchical clustering (hclust call in base R) (Murtagh, 1985) and skater (spdep R package) methods (Bivand et al., 2013; Bivand & Wong, 2018) in HERON first finds regions within each protein where consecutive probes are called in any positive sample (**Figure A.2B**). These groups of called probes are then segmented using either the hierarchical (**Figure A.2C**) or skater (**Figure A.2D**) clustering method and the average silhouette score is used to determine the optimal number of clusters or graph cuts respectively.

To calculate similarity or dissimilarity between probes in relation to their significance scores or calls for the post samples, HERON provides the option to utilize a "binary" or a "z-score" method. The binary method uses the probe calls as labels (true/false for bound/unbound by an individual serum sample) for which a hamming distance can be calculated between the probes. The z-score method converts the probe p-values to a one-sided z-score before clustering using the Euclidean distance. After building a distance matrix for all possible pairs of probes on the protein, we then use a clustering algorithm to find groups of consecutive probes on the protein that are similar in their calls or significance values across the post-treatment samples. For both the hierarchical and skater segmentation implementation we use the maximum average silhouette score to determine the number of clusters (hierarchical clustering) or cuts (skater). Ties in the silhouette score are broken by selecting the clustering with the maximum number of clusters or cuts. Other different distance metrics can also be used for measuring the distance between elements; however, our algorithm used the Euclidean distance with the z-score clustering and an average hamming distance for segmenting using the binary calls.

Using the hierarchical clustering algorithm from hclust, HERON first calculates the distance matrix as usual. For each identified cluster, a distance matrix is derived by setting each i/j element to the maximum pairwise distance between probes starting at protein position i and ending at protein position j for the group of probes to be segmented. Complete hierarchical clustering is then performed using the distance matrix to find clusters that are contiguous probes within the protein (**Figure A.2C**).

The skater algorithm is used to find consecutive probe regions with similar signal patterns across the samples. Since skater allows graph constraints on the clustering process, we introduced a linear graph where an edge is introduced between each peptide that is the adjacent peptide within the protein tiling. Each iteration of the skater algorithm finds the best cut in the graph using the calculated distance metric between the probes. After iterating through all possible cuts to find the cut that gives the best average silhouette score, the epitope regions are then defined by the remaining connected nodes in the graph (**Figure A.2D**).

Epitope and Protein p-values

An epitope level p-value is calculated by using a meta p-value method with each positive sample's linear probe p-values for the probes that are contained in that epitope. Similarly, the protein level p-value is determined using a meta p-value method based on the epitope p-values found across the protein for each positive sample. Many meta p-value estimation methods exist (Alves & Yu, 2014; Becker, 1994; Huo et al., 2020; Liu & Xie, 2020; Poole et al., 2016; Sun & Lin, 2020; Wilkinson, 1951; Wilson, 2019), including Fisher, and each has its own characteristics depending on the type of underlying hypothesis to be tested.

For epitopes, we chose a meta p-value method that requires most, if not all, of the peptide probes within the epitope block to have significant values, such as Wilkinson's max (wmax) (Becker, 1994; Birnbaum, 1954; Tippett, 1931; Wilkinson, 1951). To relax the requirement that all probe p-values

within an epitope be significant, Wilkinson's max can be calculated on the n^{th} maximum p-value (I.e., wmax2 for the 2^{nd} highest p-value within the epitope). In cases where Wilkinson's max is used and the number of probes in the epitope is fewer than n peptides, HERON conservatively assumes that the p-value was 1. Due to the inherent possibility of dependency between the signals of adjacent probes, HERON also allows the use of the harmonic mean (hmp) (Wilson, 2019) and the Cauchy (cct) (Liu & Xie, 2020) meta p-value methods, which have been shown to be tolerant to inter-dependencies between the elements for which the p-values are combined.

At the protein level, our goal was to identify proteins where at least 1 (or more) epitope (s) are significant, a meta p-value method of choice would be the minimum + Bonferroni correction (min_bonf) or the Wilkinson's minimum (wmin) or Tippett's method (Becker, 1994; Birnbaum, 1954; Tippett, 1931; Wilkinson, 1951). Using the n^{th} minimum (I.e., wmin2 for the 2^{nd} smallest epitope p-value within the protein) would be a more stringent requirement, where at least n epitopes have to be significant. In the case where there are fewer than n epitopes with p-values, HERON tries to call Wilkinson's min with the $(n-1)^{th}$ p-value iteratively. If there is only one p-value, then Wilkinson's n^{th} min will just return the single p-value.

Results

The performance of the algorithm and optimal parameter settings were tested on two different datasets, COVID-19 and Melanoma. Performance was evaluated by testing the correlation between technical replicates for repeatability of calls, finding peptides that ensure reproducibility with a validation assay such as ELISA, and epitope boundary finding parameters as described in the next sections.

Comparison of HERON performance on COVID-19 Dataset

We compared HERON's performance on a COVID-19 dataset, using parameters that were similar to the t-test workflow used to identify diagnostic peptides (Heffron et al., 2021). We utilized the following significance parameters: an absolute shift on the differential t-test of one, where one would mean the difference between the current COVID+ sample's normalized fluorescent value from the Nimble system sample and the average normalized values from the COVID- samples is significantly different by more than 2-fold, an adjusted p-value threshold <0.01, one-hit filter, hierarchical clustering with binary and hamming distance scores, Williamson's max and Tippett's for epitope and protein meta p-values estimation respectively. Probe, Epitope, and Protein calls were made if 25% of the COVID+ samples were called at the adjusted p-value threshold of <0.01. We also use the differential t-test (not the global z-test) to resemble the processing done in the Heffron et al. 2021 paper.

Each segmentation result and subsequent calls at the probe, epitope and protein levels are compared to the previously reported regions (Heffron et al., 2021) (**Figure A.3A**). For the epitope-level, we converted the selected epitopes back to the list of probes that are contained within each epitope, and then compared the probes between the two methods. The overlaps and probes unique to the Heffron et al. 2021 and the HERON method are provided in **Supplementary Table A.1**. All but 7 probes previously reported were identified by HERON indicating good agreement between the probe calls. The differences between the probes called can be due to 1) relaxation for every COVID+ sample that gets called separately using the standard deviation from the COVID- samples, 2) the Heffron paper uses a strict 2-fold change cutoff whereas HERON is using a less stringent method that incorporates the 2-fold change as part of the differential p-value calculation. Looking at the probes, we found that of the 7 missed by HERON, 4 are on the edge of achieving the 25% cutoff (9 out of 40 COVID+ samples). Also, all 7 of the probes missed by HERON are part of short (1-2 probes)

epitopes. Conversely, of the 145 probes identified in HERON but not in Heffron et al., 56 are on the edge of not being called (exactly 10 out of 40 COVID+ samples).

Figure A.3B is a heatmap of the spike protein with line graphs showing the probe and epitope annotations of the SARS-CoV-2 spike protein from the heat map displayed across the top. The epitope and probe annotations also display the agreements and disagreements between HERON and the epitopes and probes as previously reported (Heffron et al., 2021). The differences between the epitope calls are most likely due to the method for finding epitope segments and the meta p-value method used to score them. The overlaps and epitope probes unique to the Heffron or HERON methods are provided in **Supplementary Table A.2**. Of the 61 epitope probes that were missed by HERON, 43 are due to an epitope extension, where HERON finds a longer epitope than the previous method, but results in a loss of significance. 3 of the 43 extensions are on the edge of 25% (9 out of 40 COVID+ samples). The remaining 18 have matching epitopes in HERON but are shifted or fragmented and also have a loss of significance. Subsequently, of the 79 epitope probes found in HERON but not in the previous method, 49 are extensions (with 9 on the edge, i.e., 10 out of 40 COVID+ samples called), and 21 are from new epitopes (with 6 of the 21 on the edge of being called. These results indicate that HERON can find epitopes that can be more specific to a subset of subjects, which warrants additional study to further improve the epitope segmentation process.

The protein level Venn diagram at the bottom in **Figure A.3A** shows HERON selected additional proteins than the previous method (Heffron et al., 2021). Since the requirement on the protein-level is to have one or more significant epitopes and many potential epitopes are detected on the proteins, the proteins selected in the COVID dataset will be permissive in the number of proteins called.

Repeatability of Probe, Epitope, and Protein Significance Values with the Melanoma Dataset

We studied the repeatability of the p-value on the probe, epitope, and protein level by looking at the correlation between the significance values obtained on technical replicate samples from the murine Melanoma immunotherapy dataset. The Pearson correlation of the $-\log_{10}$(FDR) between two technical replicates: one single immune serum sample was divided into two identical aliquots which were tested in parallel with parallel data sets collected in the same array assay (B2), or a separate single immune serum sample divided into two identical aliquots which were tested on similar arrays, independently with parallel datasets collected in two separate similar arrays that were performed approximately one year apart (PD1).

For the parameters specific to probe p-value calculations, we chose three different levels of significance on the global z-test p-value coupled with the FDR cutoff used (Inclusive – global with a standard deviation (sd) shift of 3 and adjusted p-value cutoff of <0.2, Moderate – global sd shift of 6 and adjusted p-value cutoff of <0.05, and Restrictive – global sd shift of 10 and adjusted p-value cutoff of <0.01) and investigated with or without the use of the one-hit filter. For the epitope p-value parameters, we chose five different ways of finding epitopes (unique, hierarchical clustering with binary calls and hamming distance (hbh) or z-score with Euclidean distance (hze), skater with binary calls and hamming distance (sbh), and skater with z-score and Euclidean distance (sbz). The epitope meta p-value was either Wilkinson's max (wmax1), Wilkinson's 2nd max (wmax2), Fisher (fisher), harmonic mean (hmp), or Cauchy (cct). For protein meta p-values, we choose either Fisher (fisher), Tippetts/Wilkinsons min (wmin1), Wilkinson 2nd min (wmin2), or min with a Bonferroni correction (min_bonf). Different adjusted p-value cutoffs could be used for the probe, epitope, and protein levels. Our results have tied these three parameters to the same value.

We calculate the Pearson correlation for the between the technical replicates of B2 and PD1 using the $-\log_{10}$(adjusted p-values) for the probes, epitopes, and proteins using an exhaustive search of the parameters mentioned above. The correlation results are presented in **Supplementary Table A.3**.

Looking at the average correlations between the probe-level p-values of the technical replicates of B2 and PD1 (**Supplementary Table A.4**) and marginalizing across the one-hit filter results, we found that using more stringent statistical filtering improves the technical replicate Pearson intercorrelation of the probe $-\log_{10}$(adjusted p-values) (Inclusive – 0.806, Moderate – 0.841, and Restrictive – 0.846). By marginalizing across the statistical parameters, we also found that using the one-hit filter slightly improves the average replicate intercorrelation (Without – 0.824, With – 0.837). To find a good setting with good repeatability across the probe level significance values used, we then averaged the average correlation for the probe, epitope, and protein across the inclusive, moderate, and restrictive statistical parameters. The overall correlation results are reported in **Supplementary Table A.5** and the top 10 parameter settings are summarized in **Figure A.4A**. Looking at the top 10, it appears that the Wilkinson's max on the 2nd highest p-value for epitopes and either Tippett's or min+Bonferroni for protein meta p-values using the skater or hierarchical clustering segmentation methods on the binary calls achieve the highest average correlation across the significance levels and the two technical replicates.

The choice of segmentation method can also determine the length of the epitope (**Figure A.4B**). It appears, on average, that the "unique" (uniq) segmentation method gives longer epitopes, while the skater or hclust segmentation methods using the binary calls obtains shorter epitopes. Looking at **Figure A.4A**, the shorter epitopes maybe indeed be more accurate since those methods are within the top 2 parameters for highest average technical correlation.

Supplementary Figure A.2 shows technical replicates as scatterplots for each of the probe, epitope, and proteins for the best overall average correlation using the Moderate global standard deviation shift and FDR cutoff values.

Probe, Epitope, and Protein Counts with Different Significant levels on the Melanoma Dataset

We used the best correlation parameters to make calls on the Melanoma dataset for which the technical replicates were averaged together at the inclusive, moderate, or restrictive significance levels to obtain calls at the probe, epitope, and protein level. The significance calls for each sample and level for the moderate significance values are displayed as a complex upset plot (Krassowski et al., 2022; Lex et al., 2014) in **Figures A.5A** – Probe-level, **A.5B** – Epitope-level, and **A.5C** – Protein-level. The number of probes, epitopes, and proteins recognized by 4, 5 or 6 of the 6 immune mice was substantially lower than the number of epitopes mutually recognized by multiple individuals. **Figure A.6** presents data for the LEM containing protein 3 (Lemd3), illustrating where two epitopes were found in two different positive samples (1 out of 6, ~16%), which resulted in a K of N of 2 (33%) at the protein level. **Supplementary Figure A.3** depicts the Hemicentin-1 (Hmcn1) protein, which had 11 epitopes called in different regions with different numbers of positive samples called.

Validation by ELISA from expert selected peptides

For the Melanoma dataset, the probe, epitope, and protein calls using the one-hit filter, unique segmentation method, the Wilkinson's max for epitopes, and min+bonf for proteins with the inclusive, moderate, and restrictive significance levels were used to select peptides for validation. Hoefges et al 2023 selected 16 peptides (16-mers that had been tested in the high-density array) to validate using ELISA; 14 were selected based on their strong signal > 6SD over the mean by at least 3 of the 6 immune serum samples tested in the high-density array. The other 2 were selected as

negative controls, based on signals < 3SD over the mean for all 6 immune serum samples in the high-density array. These were all then tested in a standard ELISA assay, with data reported out as optical density (O.D.) values for each serum sample (naïve or immune) tested, as detailed by Hoefges et al. 2023 (Hoefges et al., 2023).

The replicates for ELISA were first averaged together. To indicate a positive hit, a threshold O.D. value of greater than or equal to two was used on each ELISA data point. For each peptide, the fraction of positive hits was calculated for the immune samples in the original and validated ELISA set and for the pre-/Naïve samples in the validated set. A peptide was called validated for positive reactivity if 25% or more of the respective pre-/Naïve or post-/Immune samples were called.

The Internal Validation Cohort consisted of immune and naïve sera from the original 6 immune mice used in the original high-density array (and shown in **Figure A.5**). When these sera were tested in the ELISA on these 16 peptides (**Figure A.7A**), 10 out of 14 positively-selected peptides (71%) validated for positive reactivity on the immune samples, while 0 (0%) of the 14 peptides validated for positive reactivity on the naïve samples, as expected. In contrast, 0 out of 2 negatively-selected peptides (0%) had validated for positive reactivity on the immune samples, and similarly 0 out of 2 negatively-selected peptides (0%) had positive reactivity on the pre-/naïve samples, as expected.

The External Validation Cohort consisted of serum samples from 20 separate immune mice and naïve serum samples from 14 of those mice, that had not ever been tested before in any high-density array or any ELISA assay. When these 34 sera were tested in the ELISA on these 16 peptides (**Figure A.7B**), most (8 out of 14) positively-selected peptides (57%) validated for positive reactivity on the immune samples, as expected, while few (only 2 of the 14) peptides (14%) had positive reactivity on the naïve samples, as expected. In contrast, 0 out of 2 negatively-selected peptides (0%) validated for positive reactivity on the immune samples, and similarly 0 out of 2 negatively-

selected peptides (0%) validated for positive reactivity on the pre-/naïve samples, as expected. These results with this external, independent, validation cohort are showing a similar, but not quite identical, pattern as seen with the internal validation cohort; the only difference is some of the naïve sera of these 14 independent cohort mice tested were also detecting a fraction of the positive peptides, which was not seen with naïve sera from the original cohort of 6 immune mice. The naïve sera from the original 6 mice did recognize some peptides in the high density array, but these peptides were not chosen for this validation ELISA testing. Thus, it is not surprising that some of the naïve sera in the independent cohort of mice might be able to recognize some peptides not recognized by the naive sera of the original cohort (**Figure A.7**).

Discussion

Ultra-dense linear peptide binding arrays are useful for identifying the landscape of antibody binding to entire microbe or mouse proteomes for immune studies, however, many factors determine the interpretation of immune response in the form of antibody binding to linear peptides from peptide arrays. Defining antibody binding epitope boundaries can be challenging as regions of interest can be defined as regions with similar antibody binding responses across multiple samples, or a region from an individual or a small number of samples. We developed HERON, an algorithm and software package that can make antibody binding calls and estimate significance at the peptide, epitope, and protein level, while identifying global and individual features.

COVID-19 dataset performance

In the COVID-19 comparison between HERON and the method described in Heffron et al. 2021, we found good agreement between the probe- and protein- levels found by HERON. However, there were significant disagreements at the epitope-level. The discrepancies are due to differences in the way the epitopes are found and scored. The HERON method finds epitopes with a procedure that

tries to balance an individual sample and across-sample levels and reports K of N or fraction scores to allow the user to determine the number of samples that are needed for valid epitopes. In contrast, the t-test method presented in the Heffron et al. 2021 study is looking for a difference of means between the COVID- negative and positive samples. Looking at the figures of epitope annotations for the spike protein of SARS-CoV-2 depicted in **Figure A.3**, some of the epitope boundary differences between the Heffron et al. 2021 t-test method (Heffron et al., 2021) and HERON method are probably due to some samples with a high post score. The t-test method makes calls based upon the average across all post samples, where a few outliers from the mean could result in a higher overall mean. The HERON method makes a call on each post sample, which could separate high binding "outliers" from lower binding "inliers".

What parameters to use depends upon the type of question and goals of the study. In the case of virus epitope discovery, finding consistent epitopes across positive samples is desired to find a general biomarker for the development of vaccines and diagnostics. HERON also provides the ability to find positive sample specific epitopes and for finding epitopes of longer span using the "unique" segmentation method. Finding longer epitopes may be of interest to researchers when trying to find positive sample-specific epitopes.

Melanoma dataset performance

Our results using the different meta p-values on the Melanoma dataset gives an idea of what method to use when calculating an aggregate p-value for the epitope and protein level calls. Different methods need to be employed depending upon whether the desired result is for all p-values within the aggregate to be significant (epitope level) or at least one of the p-values within the aggregate needs to be significant (protein level). We have chosen a few different meta p-value methods to compare and contrast against each other. Using Wilkinson's max p-value is also

competitive for epitopes. In the literature, there are many more meta p-value methods and a comparison of those methods with the ones studied here would be a good next step. However, some of them, while modeling the dependencies between the p-values using correlation coefficient, require many more samples to be accurate (Kost & McDermott, 2002; Poole et al., 2016) Berk-Jones, HC, and others take a longer time to run (Sun & Lin, 2020).

From our results, the Fisher method seems competitive and the extensions to that method such as the empirical Brown (Poole et al., 2016) methods that use covariance between samples to estimate p-values might be helpful in finding consistent probes, epitopes, or proteins across all samples when the number of samples is high. As the antibody repertoire of mice shows stochastic variability between mice, based on VDJ recombination of genes determining antigen binding immunoglobulin regions during immune oncogeny, the number of probes, epitopes and proteins expected to be recognized by serum from all immune mice is expected to be small. Our initial results evaluating this, using the methods developed herein, confirms this prediction (Hoefges et al., 2023). However, we are finding that there are some probes, epitopes and proteins co-recognized by a substantial fraction of immune mice, suggesting some antigens may be of importance in a substantial fraction of mice. For epitope meta p-values, we will use methods that ensure that the underlying p-values are mostly significant for epitope scoring. The one-hit filter seems to increase the correlation results slightly at the cost of reduced number of hits found. Although the hits that were missed may be due to noise from the peptide array. The choice of segmentation algorithm depends upon the desired type of epitopes. The “unique” segmentation method seems to find on average longer runs of probes and epitopes, while the other clustering methods seem to find shorter and more consistent epitopes. In the case of the Melanoma dataset, longer runs of peptides for epitopes seem to be desired, whereas the COVID-19 and other virus type studies trying to find good vaccine targets would prefer the shorter and more consistent sequences.

Assumptions, Issues, and Calibration

There are several complexities with running HERON during processing peptide binding array datasets and we list the majority of them here. Starting at the unique sequence-level, HERON assumes that the data is normally distributed, or if the data has been smoothed, the smoothed probe-level is normally distributed.

Assumptions on the unique sequence or smoothed probe-level

For calculating the differential t-test p-values/scores, HERON assumes that the variance of the post-sample is the same as the estimated variance of the pre-sample and that the post-sample signal represents the mean of the post-group values. To help deal with the higher variance and the obvious breakdown of the assumptions, we use the degrees of freedom from the pre-samples to estimate the one-sided p-value from the t-distribution.

Upon calculating the global z-test p-values/scores, HERON assumes that the sequence or smoothed probes signals are comparable across different sequences and that the mean and standard deviations can be used to calculate the p-value against one sequence signal for a post sample.

When combining the differential and global z-test p-values, the Wilkinson's max meta p-value method assumes independence among the p-values to be combined. More study is needed to determine if HERON has violated this assumption and, if so, the ramifications when estimating the combined p-values/scores or if using a different meta p-value that is more tolerant to correlated p-values would improve HERON's estimation of significance on the probe-level.

Finally, in the case of copying the unique sequences to the probe-level, there is an assumption that it is fair to do so. The main assumption is that resulting probe-level p-values are still well behaved. In the case of the Melanoma dataset, since we are smoothing the data beforehand, each probe can be treated as an independent measurement, even though there are some peptide probes that come

from a sequence that maps to more than one protein. The COVID-19 dataset, however, is unsmoothed and contains many sequences shared between proteins from different strains of the coronavirus. To process unsmoothed data in HERON, we estimate the adjusted p-values on the unique sequences and then copy the result to the respective probe-level identifiers.

Assumptions on the epitope-level

For the epitopes, or group of consecutive probes across a protein, the meta p-value methods used assume that the p-values are accurate and well-calibrated. Several of them (Wilkinson, min+bonf, Fisher, etc.) assume independence among the p-values, which is directly violated as we group adjacent probes which overlap, as epitopes. Other methods such as the Cauchy combination test or the harmonic mean try to alleviate this assumption, while other meta p-value methods (Brown, Kosts, etc.) attempt to model the covariance between the p-values to improve the estimation.

There is also an assumption that the epitopes identified by the epitope finding methods are the only ones to be included in the analyses. For example, do we need to correct for the number of all possible epitopes per protein? Currently, HERON treats the list of epitopes as a separate list and just corrects using the Bonferroni-Hochberg algorithm when reporting the adjusted p-values for the epitopes and uses the uncorrected p-values for estimating the protein-level p-values.

Assumptions on the protein-level

Going up another level in the hierarchy of probes, epitopes, and proteins, the protein p-values calculations assume that the epitope regions and p-values/scores are well-calibrated and well-behaved and the further assumptions made by the meta p-value method used to estimate the protein p-value/scores from the epitope p-values. We also assume that the adjusted p-value correction only needs to be applied to the proteins for which at least one epitope region was found.

Calibration of p-values using permutation tests

One of the ways to alleviate some of the issues presented above is to re-calibrate the scores into actual accurate/well-behaved p-values. Permutation statistics, shuffling the sample labels, calculating the resulting p-values/score, and using the permutation scores to calibrate the raw p-values is a popular method for ensuring well behaved statistics.

While the use of permutation statistics to provide accurate p-values on the unique sequence or probe-level is seemingly straightforward (data not shown), it is unclear how to properly perform this calibration at the epitope, and protein levels. We are currently exploring this avenue to see if providing calibrated p-values to the unique/smoothed probe-level also equivalently gives well calibrated scores at the epitope and protein level, or if more complicated calibration is needed.

Furthermore, permutation statistics is also affected by the number of samples used in the experiment. For example, in the Melanoma experiment, there are 8,459,970 unique peptide probes (not 6,090,593 due to the smoothing across probes), and only 11 biological samples. Testing for all possible permutations (11! = 39,916,800) would achieve a minimum permutation estimated p-value of 2.505×10^{-8}, which after using a Bonferroni correction against all probes would achieve a corrected p-value ($2.505 \times 10^{-8} \times 8{,}459{,}970$) of 0.212 or 0.153 using 6,090,593 unique sequence probes. While there are methods to further increase the p-value accuracy in the lower range of p-values (Knijnenburg et al., 2009) with fewer permutations, there is a limitation with using permutation tests with millions of features and a small set of independent biological samples.

Finally, recent studies have shown that care must be taken when using the permutation test techniques (Christensen & Zabriskie, 2022). In this paper, we use the normalized log-transformed data, which is part of the suggested use of the Box-Cox transformation (Box & Cox, 1964). Other methods for analyzing high-throughput peptide binding array data perform a log-log transformation

before estimating the statistics (Mergaert et al., 2022; Zheng et al., 2021). While studying the application of permutation calibration, we will investigate the usefulness of the Box-Cox transformations when calculating the statistics and calibration procedures.

Proposed Extensions

With our initial success of the workflow for finding probes, epitopes, and proteins in a high-throughput protein data set, several avenues exist to further improve upon the HERON method. In the next few paragraphs, we discuss four ideas: 1) Improving the combination of probe-level p-values with copulas or trying other meta p-value methods for calculating epitope and protein p-values, 2) Enhancing the flexibility of K of N calls using a Binomial or Poisson Binomial, 3) Trying different segmentation approaches for finding epitopes such as bi-clustering, and 4) using Bayesian models to model the hierarchical nature of the calls made by the workflow.

HERON allows for the combination of probe-level p-values from the global and differential tests by using Wilkinson's max meta p-value method, which may have violated the independent p-value assumption. Other meta p-value methods exist that are tolerant to dependent p-values (Liu & Xie, 2020; Wilson, 2019) and could be used to estimate the combined p-value. Other meta p-value methods incorporate correlation to model the dependence (Brown, 1975; Poole et al., 2016), however, these methods' accuracy is determined by the number of samples for estimating the covariances (Alves & Yu, 2014).

As previously mentioned, HERON was not designed to handle cases with a high incidence where one amino acid sequence maps to many proteins and further study is needed to investigate how these redundancies affect the scores and how to adapt HERON to improve handling of datasets with this feature.

The current workflow calculates the K of N for each probe, epitope, and protein after the respective FDR cutoff is chosen. A user may want to infer a statistical test for selecting elements where $K > b$. Using the Poisson binomial distribution, we could estimate a probability for each level of K to calculate p-values with a null hypothesis of $K \leq b$. While another approach would be to just use a binomial with a set probability. More investigation is needed to determine the pros and cons of utilizing these statistical tests within HERON.

The segmentation methods tested could be extended with other clustering methods. Use of different distance calculation functions, clustering score methods or clustering methods could further improve discovery. The hierarchical and skater clustering mentioned here were only run using average Hamming or Euclidean distance. There are many more cluster distance metrics available (e.g., Manhattan) to compare and contrast. In this paper, HERON finds the optimal clustering by finding the number of clusters/cuts by optimizing the average silhouette score. Other clustering scoring metrics, such as the gap statistic (Maechler et al., 2022; Tibshirani et al., 2001) could also be used. The problem with the gap statistic is the increase in processing time to run the bootstraps for finding the optimal clustering. Using bi-clustering with the skater graph-like constraints would allow segments to be found that could overlap with other segments. Future iterations of HERON may include some of these and more study should be done to determine the advantages and disadvantages of using these different clustering approaches.

The provided workflow attempts to place significance values on epitopes, proteins, and probes using meta p-value methods. Another approach would be to estimate probabilities at each level jointly using all the provided data. Implementing a Bayesian hierarchical model, which can borrow information from the estimated probability of the different levels from a fitted Bayesian model, would be a natural extension of the workflow presented in this paper. For peptide antibody array analysis, Bayesian models have been proposed. The pepBayes implementation is designed to handle

one protein at a time from multiple strains and does not take into account the sequential probes across the protein (Imholte & Gottardo, 2016). Another Bayesian model does include provision of handling sequential probe signals by using a latent autoregressive component, however the authors propose a limit in the problem size (300 peptides and 50 samples) when using their implementation in WinBUGS (Arima et al., 2012). Additionally, neither consider cases for multiple proteins and the additional information that can be obtained and leveraged for estimating parameters on the protein level.

Conclusions

Ultra-dense peptide binding arrays are powerful tools for studying the abundance of different antibody repertoire in serum samples to understand adaptive immune responses. HERON, an R package for analyzing peptide binding array data, is a flexible and powerful tool for selecting groups of linear peptide probes with improved reliability and reproducibility when considering epitopes rather than single peptide probes due to several factors; first, there are many more probes than epitopes in the proteome, giving a larger number of possible mismatches. Second, an individual epitope can be a component of several overlapping probes; our HERON algorithm for detecting epitopes recognized by separate assessments of serum samples, requires a degree of similar recognition of the related epitope containing probes by the 2 samples, but does not require complete identity of probe recognition and signal. This enables higher reproducibility of epitopes recognized with high signals than peptides recognized with high signals when replicate chips are evaluated for separate aliquots of the same immune serum sample when evaluating proteins that are recognized, since a single protein might be recognized by different individuals at different regions.

With HERON, finding interesting linear peptide probes, epitopes, and proteins within experimental data from large (mouse proteome) and small (COVID-19 proteins) is possible. HERON exists in the Bioconductor environment and will be maintained to be of use to the scientific community. Future improvements to HERON are proposed and other approaches will be explored to provide versatile algorithms for analyzing peptide binding arrays in a multitude of possible experiments.

Figures

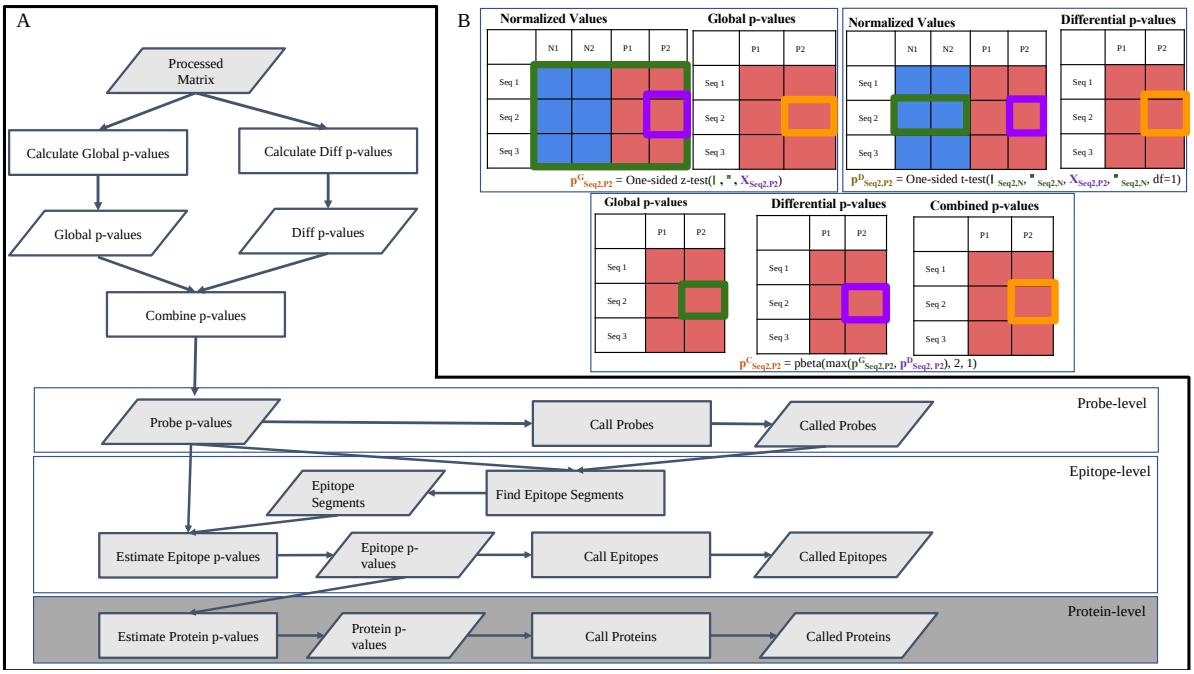


Figure A.1: Workflow and method for processing data.

A: Workflow illustrating algorithm from a normalized matrix of sequence probes to making probe-level, epitope-level, and protein-level differential antibody binding calls on each sample. After calculating the probe p-values, a threshold is used to identify the probes significantly differentially bound by antibodies (probe-level, purple box). The epitope regions are detected using clustering algorithms, scored using meta-analyses methods to combine p-values, and a threshold is used to identify significant differentially bound antibodies (epitope-level, yellow box). Finally, the protein

scores are determined from the associated protein's epitope p-value(s) using meta-analyses methods (protein-level, grey box). (**B**) The light-blue, pink, and green boxes from **Figure A.1A** correspond to the light-blue, pink, and green boxes in **Figure A.1B**. The light blue boxes illustrate the calculation of the global p-value, the pink box illustrate the calculations for the differential p-values, and the green box illustrate the calculation of the combined p-values. P-values or scores are calculated by combining p-values from a global z-test and a differential t-test using Wilkinson's max meta p-value method. The N1 and N2 columns (dark-blue colored) indicate the pre-treatment sample values and output and the P1 and P2 columns (dark-red colored) indicate the post-treatment sample values and output. The green and purple outlined boxes and highlighted (dark orange) text indicate the inputs for estimating the parameters of each separate test. The orange outlined box and colored text indicate the equation output and location in the results matrix of p-values.

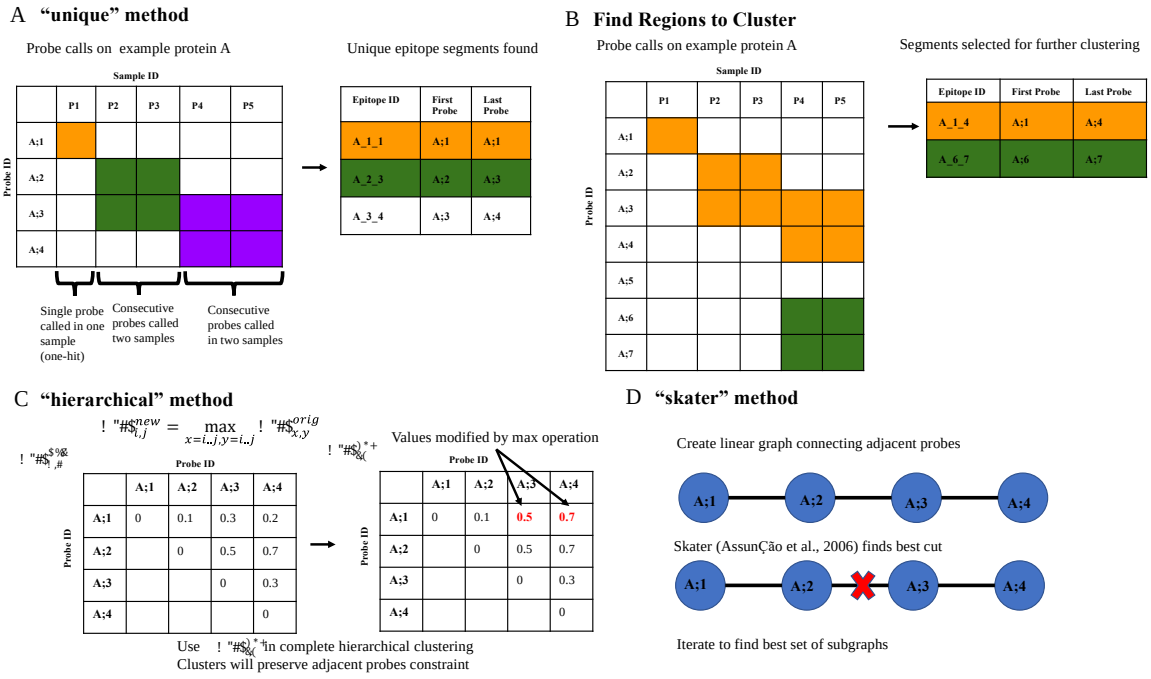


Figure A.2: Illustration and Example of epitope finding algorithms implemented in HERON.

Figure A.2: Illustration and Example of epitope finding algorithms implemented in HERON.

A: Illustration of the "unique" method, which finds the unique set of grouped probes called from each post-treatment sample. The cells highlighted in orange, green, or purple are the called probes. The table to the right of the illustration is filled in with the color corresponding to the outlined boxes in the illustration (orange, green and purple) to indicate the called epitope probe regions in the illustration. Epitope ID is defined by Protein_FirstProbe_LastProbe of the epitope.

B: Illustration of how regions are found to further cluster using either the hierarchical or skater method. The orange and green highlighted boxes mark the probes that are adjacent and the table on the right indicates the larger groups of probes, in the corresponding orange and green, that will be further segmented using the clustering methods.

C: Illustration of the hierarchical clustering method, which adapts hierarchical clustering to find consecutive probes with consistent call patterns across samples. The matrix on the left is the original distance matrix calculated from the dissimilarity of the probe calls or scores. The matrix on the right is the new matrix after applying the max operation. The two red numbers indicate the cells that were changed due to the max operation.

D: Illustration of the "skater" method which decomposes a linear graph of consecutive probes to find consistent call patterns across samples.

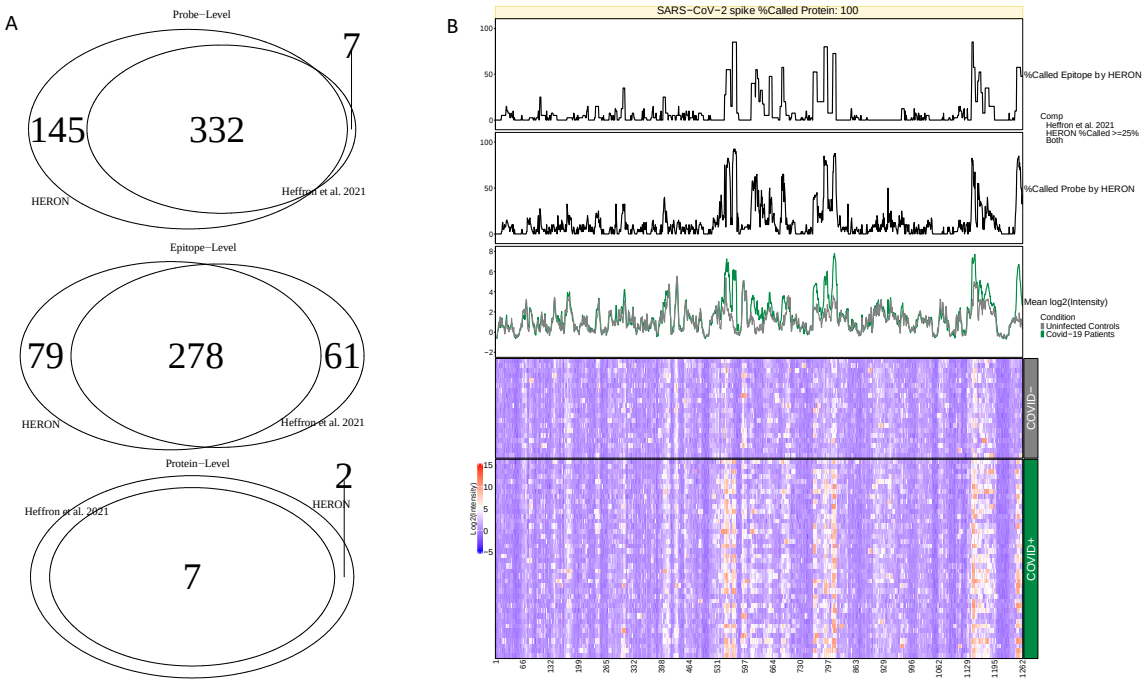


Figure A.3: Venn Diagram comparisons for COVID-19 dataset and Heatmap of SARS CoV-2 spike protein.

A: Venn diagrams of the linear peptide probe-level (top), epitope-level (middle), and protein-level (bottom) where the green circle represents the calls made by HERON and the blue circle represents the calls made by Heffron et al. 2021.

B: Annotated heatmap (Gu, 2022; Gu et al., 2016) of SARS CoV-2 spike protein. The heatmap (at the bottom) depicts the normalized intensity values for the probes (x-axis) tiled across the SARS-CoV-2 membrane protein and the individual patient serum samples (y-axis) with red representing high antibody (Ab) binding and blue representing low Ab binding. The top line plot indicates the percent of COVID-19 positive (COVID+) samples that were called as significantly differentially bound at least 2-fold over the average of the COVID-19 negative (COVID-) samples and after calculating the meta p-values on the epitope-level, the middle line plot indicates the percent of COVID+ samples called significantly 2-fold over the COVID- samples on the probe-level, and the bottom line plot shows the

average signal between control (COVID-) and positive samples. In these line plots, the % Called Probe by HERON and % Called Epitope by HERON are highlighted in red if both the HERON and Heffron et al. 2021 method called the same probes or group of probes, blue if only the Heffron et al. 2021 method called the probe or epitope, and light green if the probe or epitope was called by only the HERON method.

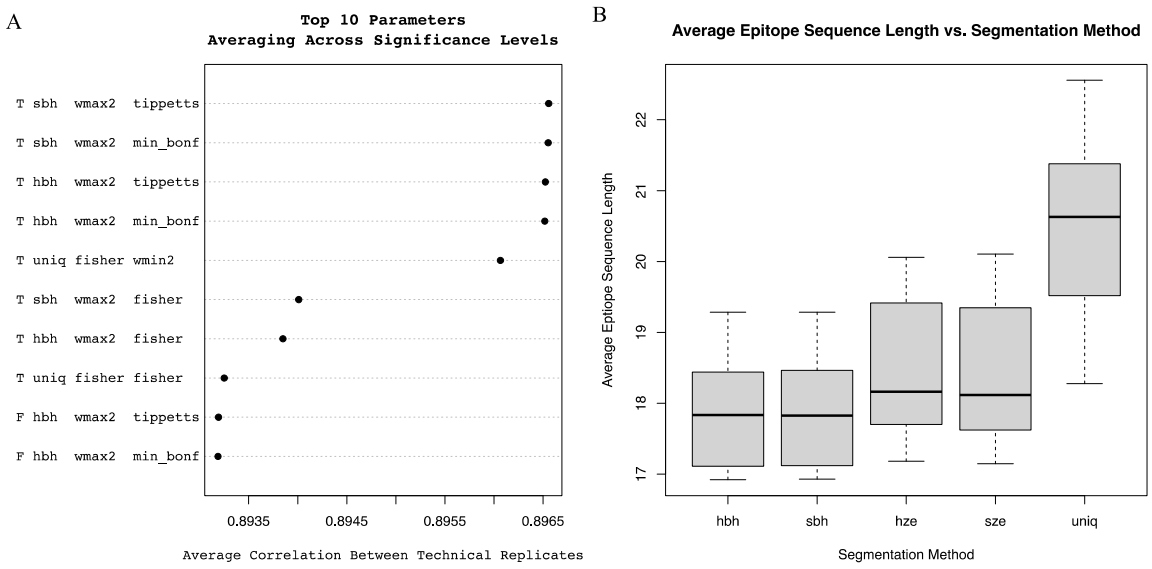


Figure A.4: Performance Results for Melanoma Dataset.

A: Average Correlation between technical replicates for the Top 10 algorithm parameter settings. Each row-label is as follows: One hit filter (T - True or F – False indicates whether filter is used or not), Segmentation method (*uniq* – unique set of epitopes found across all post samples, *hbh* - hierarchical clustering with binary calls with hamming distance, *sbh* - skater with binary calls and hamming distance, and *sbz* - skater with z-score and Euclidean distance), epitope meta p-value method (*wmax2* – Wilkinson's max on the 2nd largest p-value, *fisher* – Fisher's method), protein meta p-value method (*tippetts* – Tippett's or Wilkinsons's min on the 1st smallest p-value, *min_bonf*

– finds the minimum p-value then correct by the number of epitopes on the proteins using Bonferroni), *wmin2* – Wilkinson min on the 2nd smallest p-value).

B: Average Epitope Sequence Length vs. Segmentation Method. Box of the average sequence length that results from stitching the contained probe sequences for the detected epitope. For each segmentation method, the box is calculated using all of the average epitope sequence lengths for the remaining parameters, keep the epitopes that have at least one sample called.

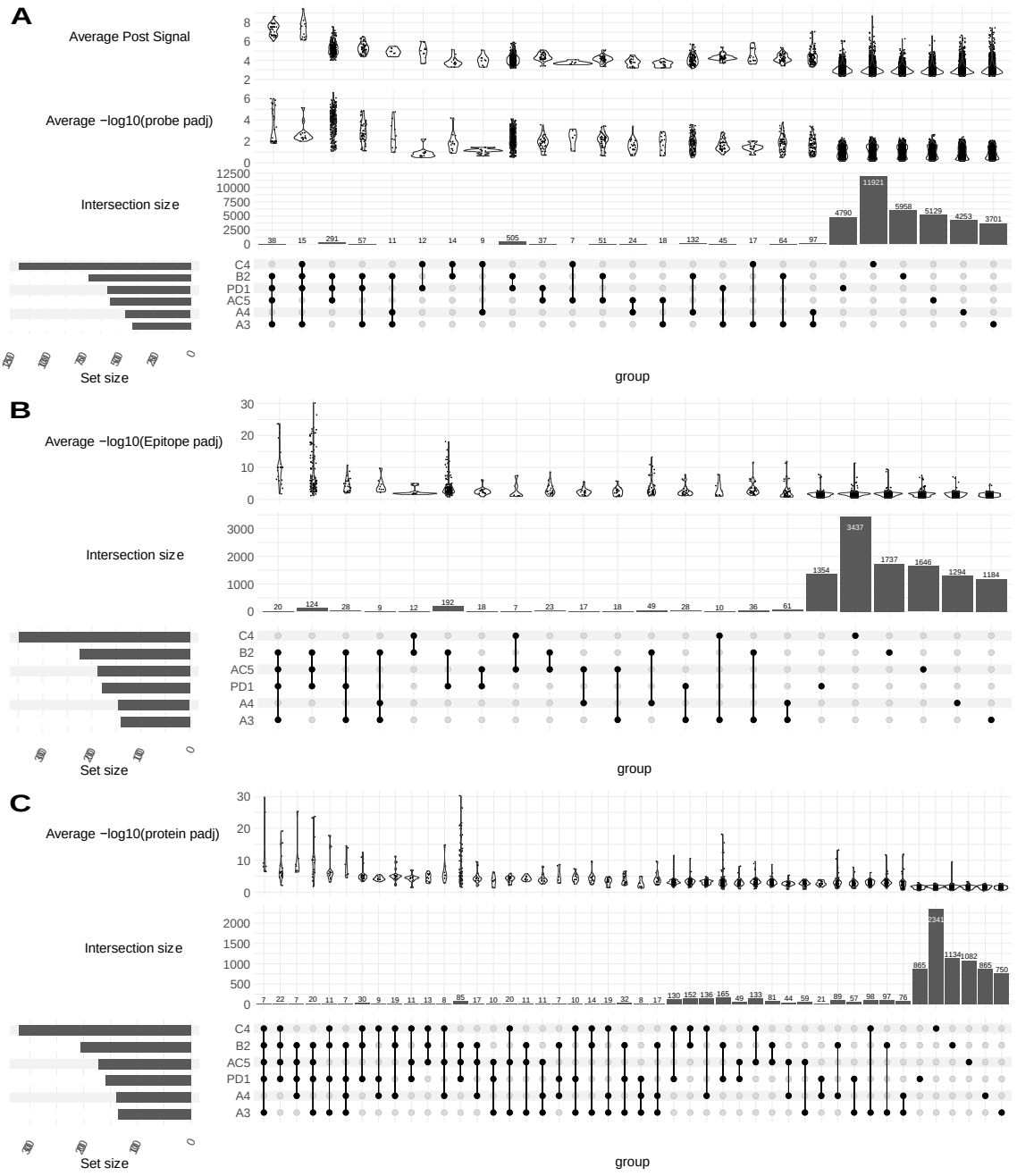


Figure A.5: Upset Plots with Average Post Signal and -log10 adjusted p-values.

Upset plot of the probes (**A**), epitopes (**B**), and proteins (**C**) called for each sample using the moderate significance level (min set size = X, sorted by the decreasing number of categories in the intersection set). AC5, A3, C4, A4, B2 and PD1 are the designations for the individual mice cured of melanoma, that provided the 6 separate immune serum samples tested here. For the probe plot (**A**),

the top row of violin plots shows the average post signal for each intersection, the middle row of violin plots shows the average of -$\log_{10}$(probe adjusted p-values) for each intersection. The epitope (**B**) and protein (**C**) violin plots are the respective -$\log_{10}$(epitope adjusted p-value) and -log10(protein adjusted p-values).

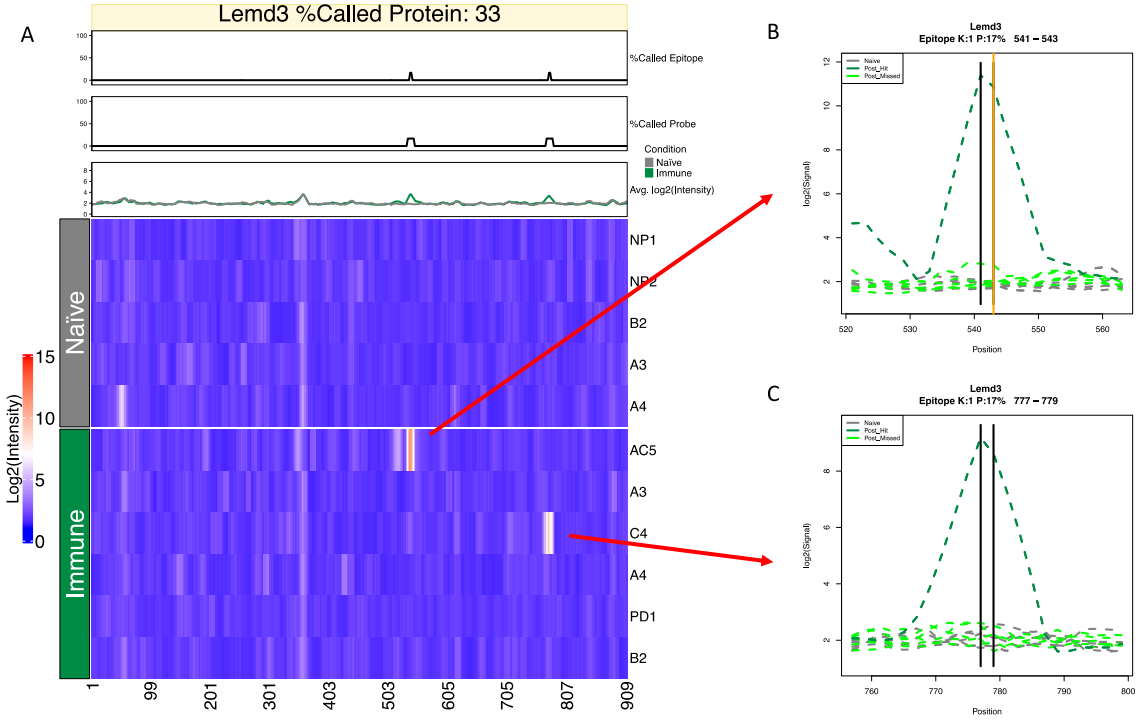


Figure A.6: Heatmap and line plots of Lemd3.

A: Line charts and Heatmap of Lemd3 using normalized and smoothed intensity values. The X-axis indicates the starting position of the probe within the protein and the Y-axis rows consist of the individual pre-treatment (Naïve) and post-treatment (Immune) samples. The first line chart above the heatmap indicates the percent of positive samples that were called on the epitope-level (Moderate Significance), the second line chart indicates the percent of positive samples called at the probe-level (Moderate Significance), and the third line chart shows the average signal between negative (Naïve) and positive (Immune) samples.

B: Line plot of epitope detected for the mouse (AC5) sample.

C: Line plot of the epitope detected in the mouse (C4) sample. For both (B) and (C), the Y-axis is the normalized intensity values and the X-axis is the starting position of the probe within the protein and the dark green dotted lines indicate the samples that were called within the epitope boundary, which are indicated by vertical black or orange lines. The orange line indicates a probe tested and validated by ELISA.

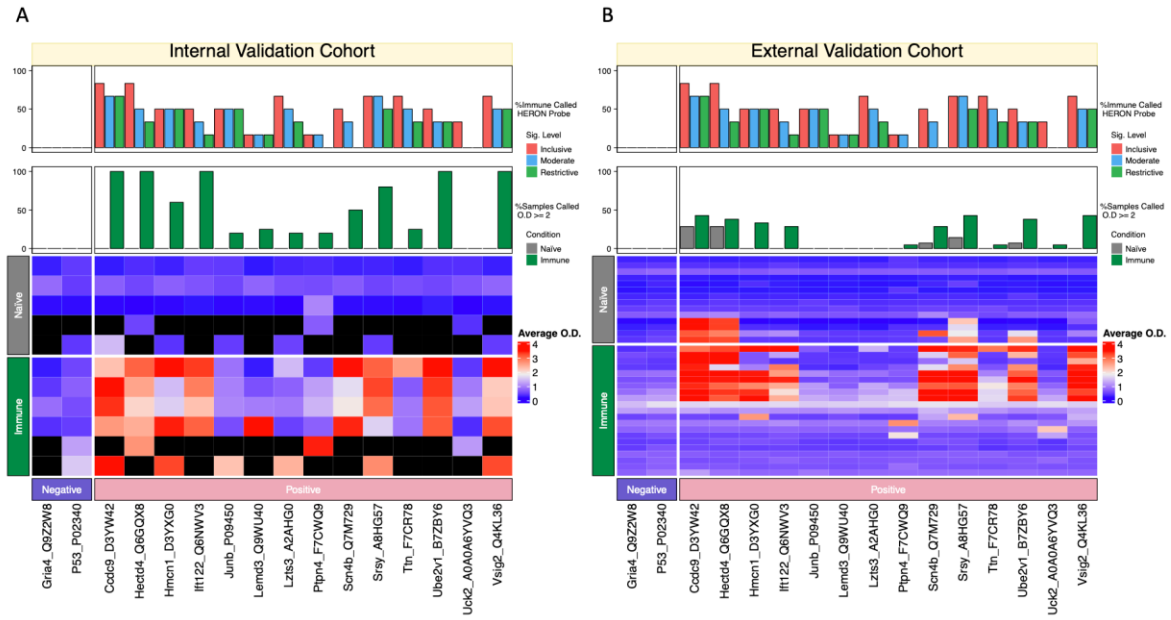
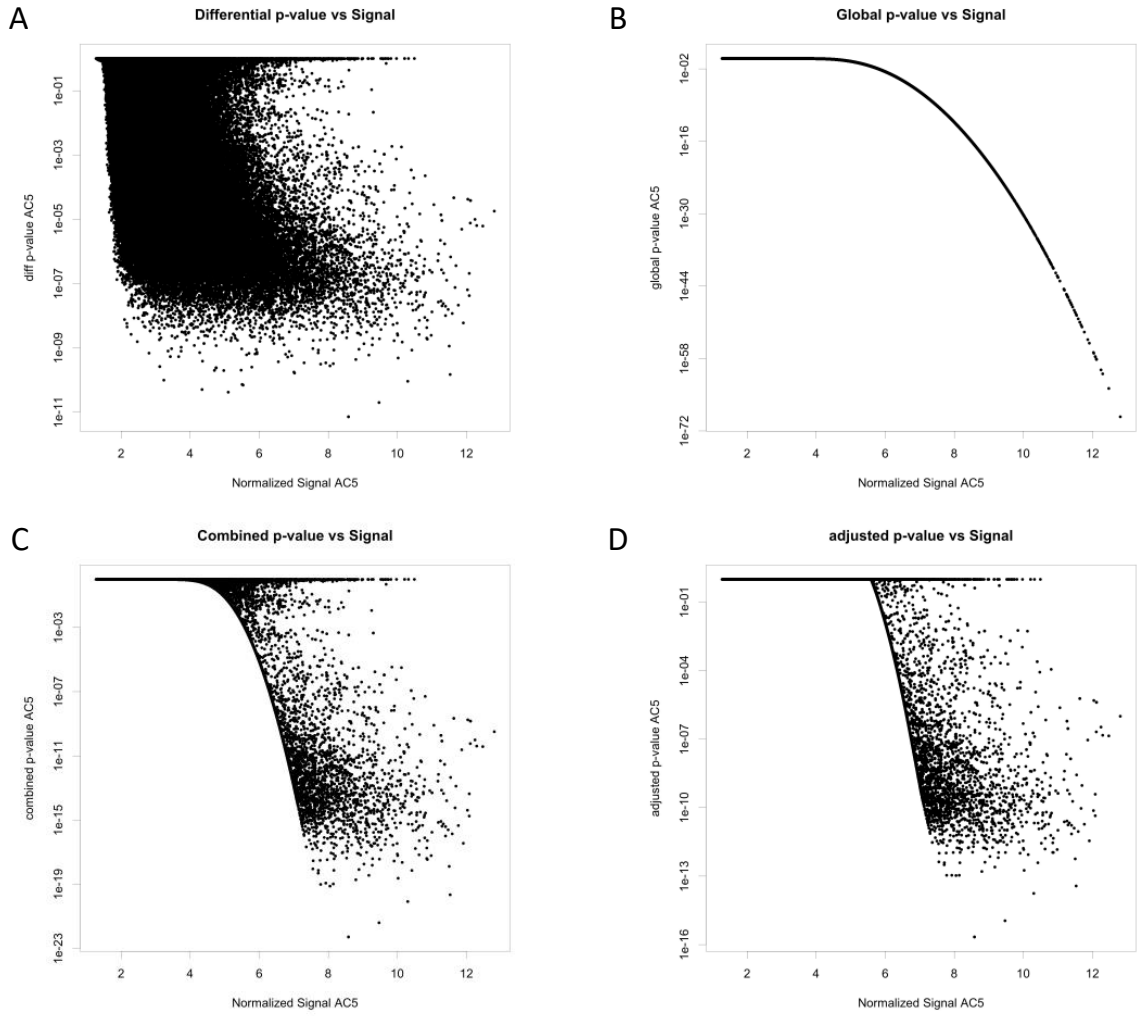


Figure A.7: Heatmaps of ELISA results.

Sixteen-mer peptides tested are shown on the columns and the individual naïve and immune serum samples are indicated on the rows. The top bar plots indicate the percentage of calls made on the probe-level by HERON in the Inclusive, Moderate, and Restrictive significance levels based on the original high density peptide array data (note these data are shown for comparison, and the graph of these high-density data are replicated for comparison's sake in the top panels for A and B). The lower bar plot indicates the percentage of Naïve (grey) and Immune samples (green) that had an

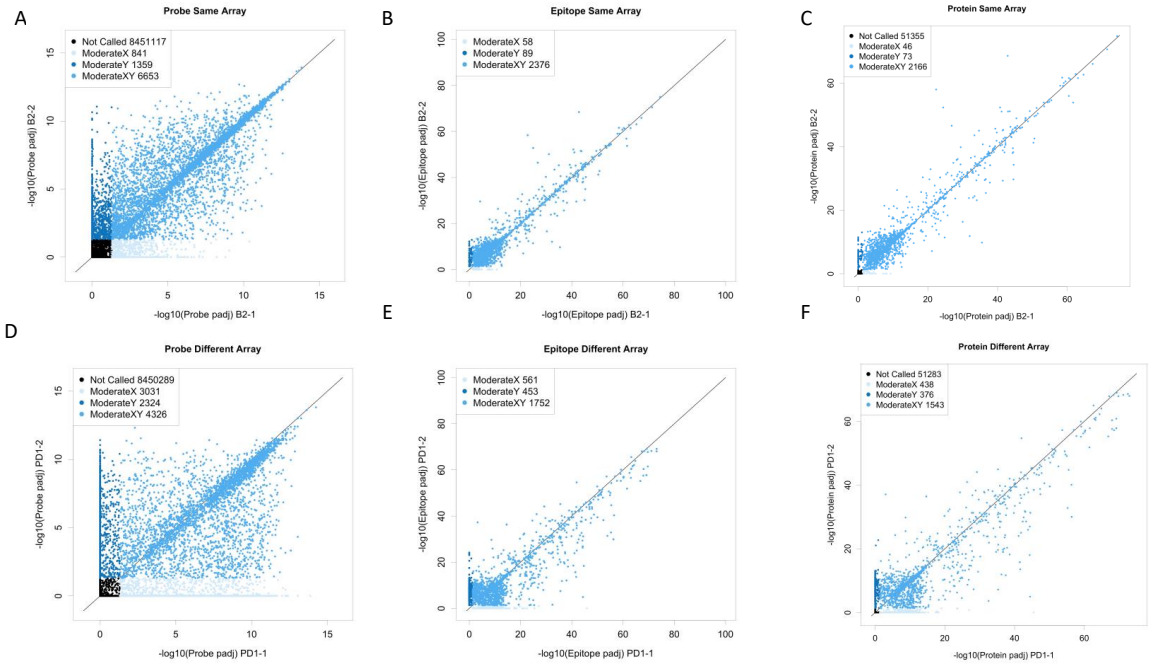
average ELISA value ≥ 2 O.D. intensity. Below these are the heat maps, showing the average across replicates O.D. values for each serum sample (naïve or immune) tested against each of the 16 peptides, with the 2 negative peptides shown at the far left and the 14 positive peptides at the right. (**A**) Internal Validation Cohort, (**B**) External Validation Cohort. Missing values (due to insufficient volume of serum available for that sample) are indicated as black values in the heatmaps

Supplemental Figures & Tables



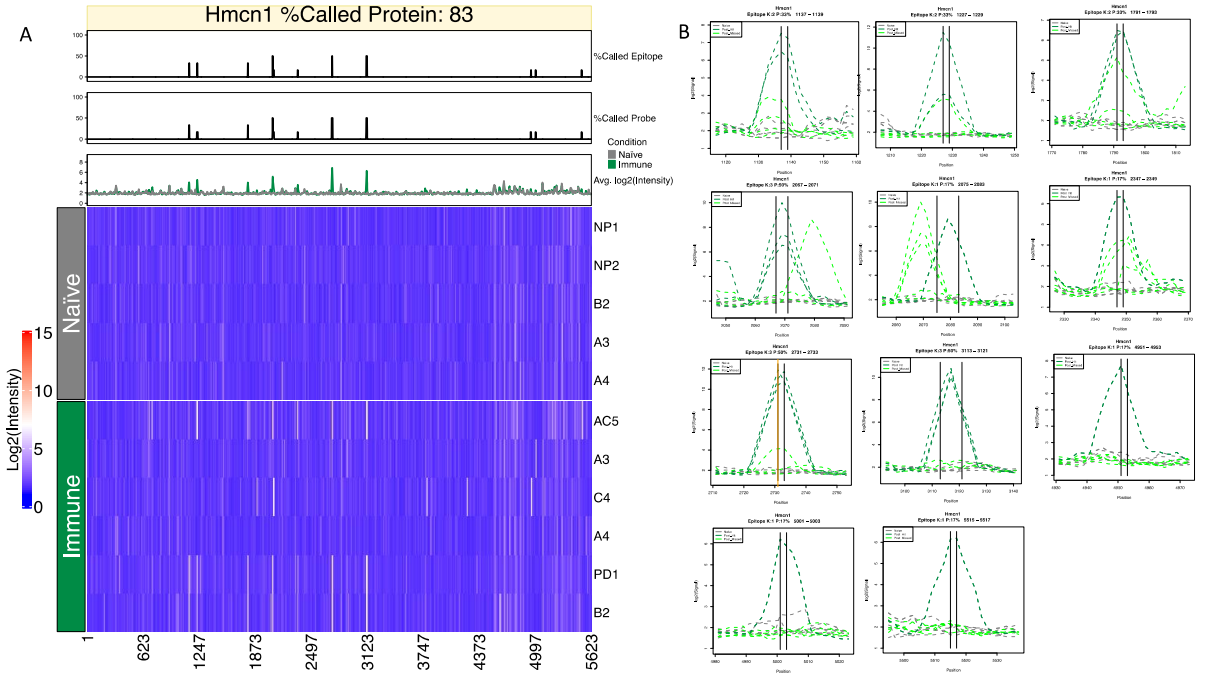
Supplementary Figure A.1: Probe estimated p-values vs. normalized signal for one representative serum sample from an immune mouse (AC5) included in the melanoma dataset using moderate

statistics parameters. In each panel, each dot represents the value obtained for this single serum sample on the ~8x10^6 peptide probes (16-amino acids each) tested in the Nimble peptide array system. **A**) result of the differential p-value estimation using t-test. **B**) result of global p-value using z-test. **C**) result of the combined p-value using Wilkinson's max. **D**) result of adjusted p-values (Benjamini-Hochberg).



Supplementary Figure A.2: Technical replicate scatterplots of -log$_{10}$ of adjusted p-values on probes, epitopes, and proteins using the moderate level statistics. **A**, **B**, and **C** are from the technical replicates collected on the same array (two separate replicate data sets obtained in the same assay using split serum aliquots from immune mouse B2). **D**, **E**, and **F** are from the technical replicates collected on arrays collected at different times (two separate replicate data sets obtained on replicate high density array chips in assays performed one-year apart, using split serum aliquots from immune mouse PD1). **A** and **D** are from the probe-level p-values, **B** and **E** are from the epitope-level p-values, and **C** and **F** are from the protein-level p-values. Light-blue are the probes, epitopes, or proteins that are called in just the 1st replicate shown on the X-axis (B2-1 or PD-1). Dark blue are

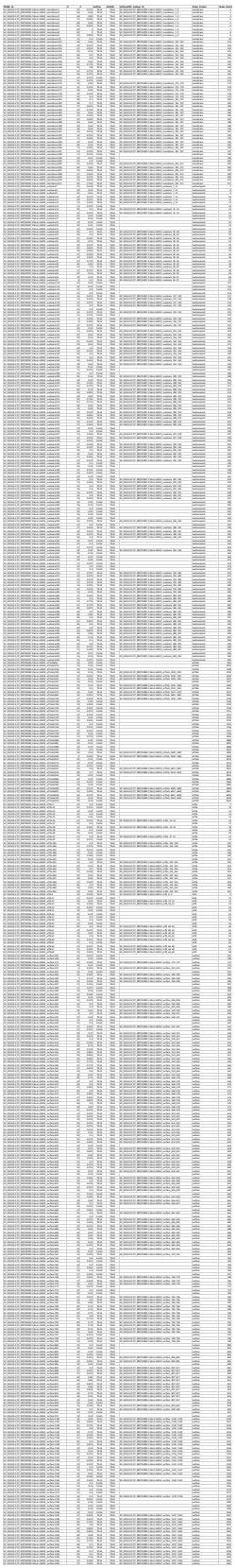
the probes, epitopes, or proteins that are called in the 2nd replicate (B2-2 or PD-2) shown on the Y-axis. Medium blue are the probes, epitope, or probes that are called in both replicates (and thus running more on the diagonal). For **A** and **D**, the black points indicate probes that were not called in either replicate.



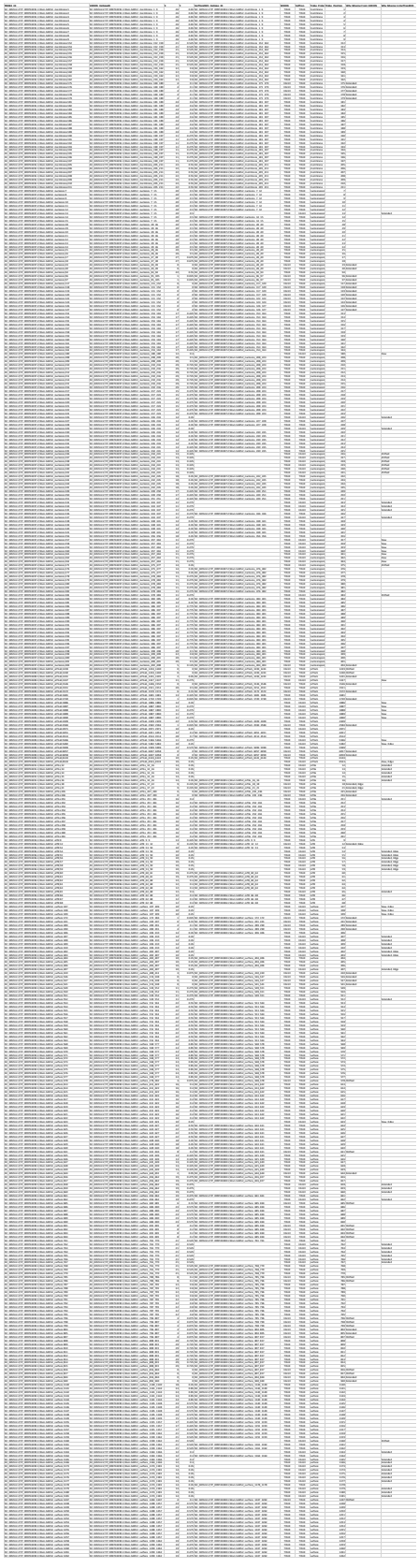
Supplementary Figure A.3: Heatmap and Lineplots of Hmcn1. (**A**) Line charts and Heatmap of Lemd3 using normalized and smoothed intensity values. The X-axis of indicates the starting position of the probe within the protein and the Y-axis is the individual pre- and post- samples. The first line chart above the heatmap indicates the percent of positive samples that were called on the epitope-level (Moderate Significance), the second line chart indicates the percent of positive samples called at the probe-level (Moderate Significance), and the third line chart shows the average signal between negative (Naïve) and positive (Immune) samples. (**B**) Line plots of 11 epitopes detected for the various mouse samples, Y-axis is the normalized intensity values and the X-axis is the starting position of the probe within the protein. The dark green dotted lines indicate the samples that were

called within the epitope boundary, which are indicated by vertical black or orange lines. The orange line indicates a probe tested and validated by ELISA.

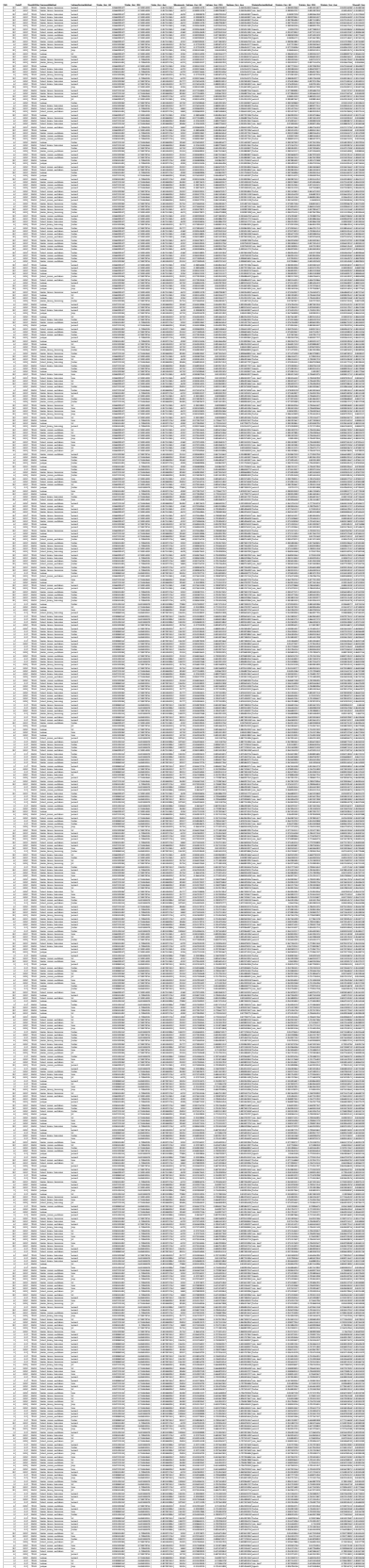
Supplemental Table 1: comparison of probe calls between HERON and previous method



Supplemental Table 2: Overlaps and epitope probes unique to the Heffron or HERON methods



Supplemental Table 3: Correlation for B2 and PD1 replicates for probe, epitope, and protein



Supplemental Table 4: Average correlation for B2 and PD1 replicates for probe-level p values for separate statistical filtering cutoffs

ZSD	Cutoff	OneHitFilter	Probe_Corr_B2	Probe_Corr_PD1	Probe_Corr_Avg
3	0.2	TRUE	0.918884554	0.696930351	0.807907453
3	0.2	FALSE	0.915105314	0.691380478	0.803242896
6	0.05	TRUE	0.952329284	0.738278754	0.845304019
6	0.05	FALSE	0.947274124	0.724462663	0.835868394
10	0.01	TRUE	0.966009537	0.749014392	0.857511965
10	0.01	FALSE	0.958341481	0.70960195	0.833971716

Supplemental Table 5: Correlation between technical replicates for tested algorithm parameter settings

OneHitFilter	SegmentMethod	EpitopePvalueMethod	ProteinPvalueMethod	Overall_Corr_Avg
TRUE	slater_binary_hamming	wmax2	tippets	0.896557242
TRUE	slater_binary_hamming	wmax2	min_bonf	0.89655204
TRUE	hclust_binary_hamming	wmax2	tippets	0.896523833
TRUE	hclust_binary_hamming	wmax2	min_bonf	0.896517675
TRUE	unique	fischer	wmin2	0.896065689
TRUE	slater_binary_hamming	wmax2	fischer	0.894010075
TRUE	hclust_binary_hamming	wmax2	fischer	0.893848916
TRUE	unique	fischer	fischer	0.893250866
FALSE	hclust_binary_hamming	wmax2	tippets	0.893192665
FALSE	hclust_binary_hamming	wmax2	min_bonf	0.89318728
FALSE	slater_binary_hamming	wmax2	tippets	0.893183642
FALSE	slater_binary_hamming	wmax2	min_bonf	0.893179121
FALSE	unique	fischer	wmin2	0.891991359
FALSE	slater_binary_hamming	wmax2	fischer	0.890393614
FALSE	hclust_binary_hamming	wmax2	fischer	0.890308299
TRUE	hclust_zscore_euclidean	wmax2	tippets	0.889890591
TRUE	hclust_zscore_euclidean	wmax2	min_bonf	0.88988621
TRUE	slater_zscore_euclidean	wmax2	fischer	0.889321421
TRUE	slater_zscore_euclidean	wmax2	tippets	0.889183554
TRUE	slater_zscore_euclidean	wmax2	min_bonf	0.889179074
TRUE	hclust_zscore_euclidean	wmax2	fischer	0.888894212
FALSE	unique	fischer	fischer	0.88884702
FALSE	hclust_zscore_euclidean	wmax2	tippets	0.886835501
FALSE	hclust_zscore_euclidean	wmax2	min_bonf	0.886826833
FALSE	slater_zscore_euclidean	wmax2	tippets	0.886121937
FALSE	slater_zscore_euclidean	wmax2	min_bonf	0.88611809
FALSE	slater_zscore_euclidean	wmax2	fischer	0.885953466
FALSE	hclust_zscore_euclidean	wmax2	fischer	0.885551495
TRUE	slater_binary_hamming	fischer	tippets	0.885410102
TRUE	slater_binary_hamming	fischer	min_bonf	0.88540482
TRUE	hclust_binary_hamming	fischer	tippets	0.884039657
TRUE	hclust_binary_hamming	fischer	min_bonf	0.884038349
TRUE	unique	wmax2	fischer	0.881528365
TRUE	hclust_zscore_euclidean	fischer	tippets	0.879939028
TRUE	hclust_zscore_euclidean	fischer	min_bonf	0.879931264
TRUE	unique	hmp	fischer	0.879890366
TRUE	unique	cct	fischer	0.879845829
FALSE	unique	wmax2	fischer	0.879461637
TRUE	slater_zscore_euclidean	fischer	tippets	0.879354751
TRUE	slater_zscore_euclidean	fischer	min_bonf	0.879350668
TRUE	slater_binary_hamming	fischer	fischer	0.878555548
TRUE	unique	wmax2	wmin2	0.878445197
TRUE	hclust_binary_hamming	fischer	fischer	0.877905359
TRUE	unique	cct	wmin2	0.876641963
TRUE	unique	hmp	wmin2	0.876628716
FALSE	unique	wmax2	wmin2	0.876577721
TRUE	slater_zscore_euclidean	fischer	fischer	0.876024242
TRUE	hclust_zscore_euclidean	fischer	fischer	0.876014798
TRUE	unique	fischer	tippets	0.874115034
TRUE	unique	fischer	min_bonf	0.874113311
FALSE	slater_binary_hamming	fischer	tippets	0.87325265
FALSE	slater_binary_hamming	fischer	min_bonf	0.873246643
FALSE	hclust_binary_hamming	fischer	tippets	0.872121447
FALSE	hclust_binary_hamming	fischer	min_bonf	0.872115436
FALSE	hclust_zscore_euclidean	fischer	tippets	0.87051832
FALSE	hclust_zscore_euclidean	fischer	min_bonf	0.870513753
FALSE	slater_zscore_euclidean	fischer	tippets	0.869598013
FALSE	slater_zscore_euclidean	fischer	min_bonf	0.869581659
FALSE	unique	fischer	tippets	0.869544569
FALSE	unique	fischer	min_bonf	0.869547224
TRUE	slater_zscore_euclidean	wmax2	wmin2	0.869166293
TRUE	unique	wmax2	tippets	0.867934947
TRUE	unique	wmax2	min_bonf	0.867832096
FALSE	slater_binary_hamming	fischer	fischer	0.867617998
FALSE	hclust_zscore_euclidean	fischer	fischer	0.867327002
FALSE	hclust_binary_hamming	fischer	fischer	0.86709844
FALSE	slater_zscore_euclidean	fischer	fischer	0.867013188
FALSE	unique	wmax2	tippets	0.866312226
FALSE	unique	wmax2	min_bonf	0.866310181
FALSE	slater_zscore_euclidean	wmax2	wmin2	0.866275307
TRUE	hclust_zscore_euclidean	wmax2	wmin2	0.864941784
FALSE	unique	hmp	fischer	0.863994705
FALSE	unique	cct	fischer	0.86392637
TRUE	hclust_binary_hamming	wmax2	wmin2	0.863533657
TRUE	slater_binary_hamming	wmax2	wmin2	0.863444112
FALSE	hclust_zscore_euclidean	wmax2	wmin2	0.861957635
FALSE	unique	hmp	wmin2	0.860878223
FALSE	unique	cct	wmin2	0.860872272
FALSE	hclust_binary_hamming	wmax2	wmin2	0.860388321
FALSE	slater_binary_hamming	wmax2	wmin2	0.8603264
TRUE	slater_zscore_euclidean	wmax1	tippets	0.856892914
TRUE	slater_zscore_euclidean	wmax1	min_bonf	0.85688805
TRUE	slater_binary_hamming	wmax1	tippets	0.856174476
TRUE	slater_binary_hamming	wmax1	min_bonf	0.856161879
TRUE	slater_zscore_euclidean	wmax1	fischer	0.856038101
TRUE	hclust_binary_hamming	wmax1	tippets	0.855686383
TRUE	hclust_binary_hamming	wmax1	min_bonf	0.855679391
TRUE	slater_binary_hamming	wmax1	fischer	0.855553709
TRUE	hclust_binary_hamming	wmax1	fischer	0.855282698
TRUE	hclust_zscore_euclidean	wmax1	tippets	0.854729723
TRUE	hclust_zscore_euclidean	wmax1	min_bonf	0.854724537
TRUE	hclust_binary_hamming	hmp	fischer	0.853930685
TRUE	hclust_binary_hamming	cct	fischer	0.853767263
TRUE	hclust_zscore_euclidean	wmax1	fischer	0.853291532
TRUE	unique	cct	tippets	0.853812624
TRUE	unique	cct	min_bonf	0.853797959
TRUE	unique	hmp	tippets	0.852768871
TRUE	unique	hmp	min_bonf	0.852754832
TRUE	slater_binary_hamming	hmp	fischer	0.851828843
TRUE	slater_binary_hamming	cct	fischer	0.8517474
TRUE	hclust_binary_hamming	cct	tippets	0.847412257
TRUE	hclust_binary_hamming	cct	min_bonf	0.847380184
TRUE	hclust_binary_hamming	hmp	tippets	0.847359353
TRUE	hclust_binary_hamming	hmp	min_bonf	0.847332992
TRUE	slater_binary_hamming	cct	tippets	0.847316295
TRUE	slater_binary_hamming	cct	min_bonf	0.847291835
TRUE	slater_binary_hamming	hmp	tippets	0.847187597
TRUE	slater_binary_hamming	hmp	min_bonf	0.84714457
FALSE	slater_zscore_euclidean	wmax1	tippets	0.846988028
FALSE	slater_zscore_euclidean	wmax1	min_bonf	0.84688223
FALSE	slater_zscore_euclidean	wmax1	fischer	0.846852819
TRUE	slater_zscore_euclidean	fischer	wmin2	0.846248935
FALSE	hclust_zscore_euclidean	wmax1	tippets	0.845031091
FALSE	hclust_zscore_euclidean	wmax1	min_bonf	0.845027165
FALSE	hclust_zscore_euclidean	wmax1	fischer	0.844319986
TRUE	hclust_zscore_euclidean	fischer	wmin2	0.84304593
TRUE	hclust_binary_hamming	fischer	wmin2	0.843618739
FALSE	slater_binary_hamming	wmax1	fischer	0.843488655
FALSE	hclust_binary_hamming	wmax1	fischer	0.843385896
FALSE	slater_binary_hamming	wmax1	tippets	0.84276992
FALSE	slater_binary_hamming	wmax1	min_bonf	0.842762042
TRUE	slater_binary_hamming	fischer	wmin2	0.842434562
FALSE	hclust_binary_hamming	wmax1	tippets	0.84241622
FALSE	hclust_binary_hamming	wmax1	min_bonf	0.842407818
TRUE	slater_zscore_euclidean	cct	tippets	0.841704029
TRUE	slater_zscore_euclidean	cct	min_bonf	0.841679928
TRUE	slater_zscore_euclidean	hmp	tippets	0.841533142
TRUE	slater_zscore_euclidean	hmp	min_bonf	0.841506917
TRUE	slater_zscore_euclidean	hmp	fischer	0.841155547
TRUE	hclust_zscore_euclidean	cct	fischer	0.840997386
TRUE	hclust_zscore_euclidean	cct	tippets	0.840497992
TRUE	hclust_zscore_euclidean	cct	min_bonf	0.840474942
TRUE	hclust_zscore_euclidean	hmp	fischer	0.840034905
TRUE	hclust_zscore_euclidean	hmp	tippets	0.840028792
TRUE	hclust_zscore_euclidean	hmp	min_bonf	0.840026375
TRUE	hclust_zscore_euclidean	cct	fischer	0.840043072
FALSE	slater_zscore_euclidean	fischer	wmin2	0.836521562
FALSE	hclust_zscore_euclidean	fischer	wmin2	0.834399375
TRUE	unique	wmax1	fischer	0.832583769
FALSE	hclust_binary_hamming	fischer	wmin2	0.832462567
FALSE	slater_binary_hamming	fischer	wmin2	0.831252413
FALSE	unique	cct	tippets	0.830084986
FALSE	unique	hmp	tippets	0.830063588
FALSE	unique	cct	min_bonf	0.83006275
FALSE	unique	hmp	min_bonf	0.830041377
FALSE	unique	wmax1	fischer	0.829010889
TRUE	slater_zscore_euclidean	wmax1	wmin2	0.82779148
TRUE	unique	wmax1	tippets	0.827248037
TRUE	unique	wmax1	min_bonf	0.82274532
FALSE	hclust_binary_hamming	hmp	fischer	0.825549061
FALSE	hclust_binary_hamming	cct	fischer	0.825395696
FALSE	unique	wmax1	tippets	0.823134687
FALSE	unique	wmax1	min_bonf	0.823130531
TRUE	hclust_zscore_euclidean	wmax1	wmin2	0.82255413
FALSE	slater_binary_hamming	hmp	fischer	0.82231566
FALSE	slater_binary_hamming	cct	fischer	0.822230981
TRUE	hclust_binary_hamming	wmax1	wmin2	0.821353503
TRUE	slater_binary_hamming	wmax1	wmin2	0.820941389
TRUE	hclust_binary_hamming	hmp	wmin2	0.820160627
TRUE	hclust_binary_hamming	cct	wmin2	0.820120129
TRUE	slater_binary_hamming	cct	wmin2	0.818451402
TRUE	slater_binary_hamming	hmp	wmin2	0.818367144
FALSE	slater_zscore_euclidean	wmax1	wmin2	0.813083328
FALSE	hclust_zscore_euclidean	wmax1	wmin2	0.813182918
FALSE	hclust_zscore_euclidean	hmp	fischer	0.812585668
FALSE	hclust_zscore_euclidean	cct	fischer	0.812392093
FALSE	slater_zscore_euclidean	hmp	fischer	0.812376403
FALSE	slater_zscore_euclidean	cct	fischer	0.812213601
FALSE	hclust_binary_hamming	cct	tippets	0.81129719
FALSE	hclust_binary_hamming	cct	min_bonf	0.811259494
FALSE	hclust_binary_hamming	hmp	tippets	0.811246735
FALSE	hclust_binary_hamming	hmp	min_bonf	0.811215238
FALSE	slater_binary_hamming	cct	tippets	0.810091257
FALSE	slater_binary_hamming	cct	min_bonf	0.810063057
FALSE	slater_binary_hamming	hmp	tippets	0.809987272
FALSE	slater_binary_hamming	hmp	min_bonf	0.809957887
FALSE	hclust_binary_hamming	wmax1	wmin2	0.809420925
FALSE	slater_binary_hamming	wmax1	wmin2	0.808937403
TRUE	slater_zscore_euclidean	cct	wmin2	0.80773527
TRUE	slater_zscore_euclidean	hmp	wmin2	0.807554214
FALSE	slater_zscore_euclidean	cct	tippets	0.806369487
FALSE	hclust_zscore_euclidean	cct	tippets	0.806364613
FALSE	slater_zscore_euclidean	cct	min_bonf	0.806339333
FALSE	hclust_zscore_euclidean	cct	min_bonf	0.806335257
FALSE	slater_zscore_euclidean	hmp	tippets	0.806223507
FALSE	hclust_zscore_euclidean	hmp	tippets	0.806220555
FALSE	slater_zscore_euclidean	hmp	min_bonf	0.806193914
FALSE	hclust_zscore_euclidean	hmp	min_bonf	0.806191406
TRUE	unique	wmax1	wmin2	0.80429594
TRUE	hclust_zscore_euclidean	cct	wmin2	0.804125463
TRUE	hclust_zscore_euclidean	hmp	wmin2	0.803988395
FALSE	unique	wmax1	wmin2	0.801461999
FALSE	hclust_binary_hamming	hmp	wmin2	0.791746792
FALSE	hclust_binary_hamming	cct	wmin2	0.791721445
FALSE	slater_binary_hamming	cct	wmin2	0.789042074
FALSE	slater_binary_hamming	hmp	wmin2	0.78897183
FALSE	slater_zscore_euclidean	cct	wmin2	0.779310655
FALSE	slater_zscore_euclidean	hmp	wmin2	0.779149884
FALSE	hclust_zscore_euclidean	cct	wmin2	0.776987657
FALSE	hclust_zscore_euclidean	hmp	wmin2	0.776861521

Appendix B: collaborative work that resulted in co-authorship during my PhD work

This appendix outlines additional work that was submitted for publication by members of the Sondel lab during my PhD for which I was a contributing author. Each following page provides the title of the work, the contributing authors, a short summary of the major findings as well as my contributions as a co-author. The following is a list of the included publications.

B.1: Rakhmilevich AL, Felder M, Lever L, Slowinski J, Rasmussen K, **Hoefges A**, Van De Voort TJ, Loibner H, Korman AJ, Gillies SD, Sondel PM. Effective Combination of Innate and Adaptive Immunotherapeutic Approaches in a Mouse Melanoma Model. J Immunol. 2017 Feb 15;198(4):1575-1584. doi: 10.4049/jimmunol.1601255. Epub 2017 Jan 6. PMID: 28062694

B.2: Erbe AK, Wang W, Carmichael L, **Hoefges A**, Grzywacz B, Reville PK, Ranheim EA, Hank JA, Kim K, Seo S, Mendonca EA, Song Y, Kenkre VP, Hong F, Gascoyne RD, Paietta E, Horning SJ, Miller JS, Kahl B, Sondel PM. Follicular lymphoma patients with KIR2DL2 and KIR3DL1 and their ligands (HLA-C1 and HLA-Bw4) show improved outcome when receiving rituximab. J Immunother Cancer. 2019 Mar 12;7(1):70. doi: 10.1186/s40425-019-0538-8. PMID: 30871628

B.3: Voeller J, Erbe AK, Slowinski J, Rasmussen K, Carlson PM, **Hoefges A**, VandenHeuvel S, Stuckwisch A, Wang X, Gillies SD, Patel RB, Farrel A, Rokita JL, Maris J, Hank JA, Morris ZS, Rakhmilevich AL, Sondel PM. Combined innate and adaptive immunotherapy overcomes resistance of immunologically cold syngeneic murine neuroblastoma to checkpoint inhibition. J Immunother Cancer. 2019 Dec 6;7(1):344. doi: 10.1186/s40425-019-0823-6.

B.4: Goldberg JL, Navid F, Hank JA, Erbe AK, Santana V, Gan J, de Bie F, Javaid AM, **Hoefges A**, Merdler M, Carmichael L, Kim K, Bishop MW, Meager MM, Gillies SD, Pandey JP, Sondel PM. Pre-

existing antitherapeutic antibodies against the Fc region of the hu14.18K322A mAb are associated with outcome in patients with relapsed neuroblastoma. J Immunother Cancer. 2020;8(1):e000590. doi:10.1136/jitc-2020-000590

B.5: Baniel CC, Heinze CM, **Hoefges A**, Sumiec EG, Hank JA, Carlson PM, Jin WJ, Patel RB, Sriramaneni RN, Gillies SD, Erbe AK, Schwarz CN, Pieper AA, Rakhmilevich AL, Sondel PM, Morris ZS. In situ Vaccine Plus Checkpoint Blockade Induces Memory Humoral Response. Front Immunol. 2020 Jul 24;11:1610. doi: 10.3389/fimmu.2020.01610. eCollection 2020. PMID: 32849544

B.6: Baniel CC, Sumiec EG, Hank JA, Bates AM, Erbe AK, Pieper AA, **Hoefges A**, Patel RB, Rakhmilevich AL, Morris ZS, Sondel PM. Intratumoral injection reduces toxicity and antibody-mediated neutralization of immunocytokine in a mouse melanoma model. J Immunother Cancer. 2020 Oct;8(2):e001262. doi: 10.1136/jitc-2020-001262. PMID: 33115944

B.7: Carlson PM, Mohan M, Patel RB, Birstler J, Nettenstrom L, Sheerar D, Fox K, Rodriguez M, **Hoefges A**, Hernandez R, Zahm C, Kim K, McNeel DG, Weichert J, Morris ZS, Sondel PM. Optimizing Flow Cytometric Analysis of Immune Cells in Samples Requiring Cryopreservation from Tumor-Bearing Mice. J Immunol. 2021 Jul 15;207(2):720-734. doi: 10.4049/jimmunol.2000656. Epub 2021 Jul 14. PMID: 34261667

B.8: Pieper AA, Zangl LM, Speigelman DV, Feils AS, **Hoefges A**, Jagodinsky JC, Felder MA, Tsarovsky NW, Arthur IS, Brown RJ, Birstler J, Le T, Carlson PM, Bates AM, Hank JA, Rakhmilevich AL, Erbe AK, Sondel PM, Patel RB, Morris ZS. Radiation Augments the Local Anti-Tumor Effect of In Situ Vaccine With CpG-Oligodeoxynucleotides and Anti-OX40 in Immunologically Cold Tumor Models. Front Immunol. 2021 Nov 15;12:763888. doi: 10.3389/fimmu.2021.763888. PMID: 34868010

B.9: Heaton AR, Rehani PR, **Hoefges A**, Lopez AF, Rakhmilevich AL, Erbe AK, Sondel PM, Skala MC. Single cell metabolic imaging of tumor and immune cells in vivo in melanoma bearing mice. Front Oncol., 20 March 2023; 13: 1110503. doi.org/10.3389/fonc.2023.1110503

B.1: Effective Combination of Innate and Adaptive Immunotherapeutic Approaches in a Mouse Melanoma Model.

Rakhmilevich AL, Felder M, Lever L, Slowinski J, Rasmussen K, **Hoefges A**, Van De Voort TJ, Loibner H, Korman AJ, Gillies SD, Sondel PM. Effective Combination of Innate and Adaptive Immunotherapeutic Approaches in a Mouse Melanoma Model. J Immunol. 2017 Feb 15;198(4):1575-1584. doi: 10.4049/jimmunol.1601255. Epub 2017 Jan 6. PMID: 28062694

In this paper we combine treatments that activate the adaptive and innate immune system to further enhance the anti-tumor effect. Using an anti-CD40 mAb and CpG in conjunction has been shown to activate macrophages and induce tumor cell killing. By combining anti-CD40 mAb and CpG treatment with anti-CTLA4 and an immunocytokine (IC) targeting GD2 and delivering IL-2 to the tumor site, we hypothesized to target macrophages using anti-CD40 and CpG and NK cells via IC would cause tumor cell death and increased presentation of tumor-associated antigens which then will be presented to T cells and cause an adaptive, T cell driven immune response which would be further augmented via anti-CTLA4 treatment. This combination treatment was able to reduce the amount of T regulatory cells in the tumor microenvironment and lead to a curative anti-tumor immune response with immunological memory in the majority of mice.

My involvement in this work was mainly focused on flow cytometric assays analyzing the tumor microenvironment and its immune components. I was also involved in part of the experimental planning and design of experiments as well as review of the manuscript.

B.2: Follicular lymphoma patients with KIR2DL2 and KIR3DL1 and their ligands (HLA-C1 and HLA-Bw4) show improved outcome when receiving rituximab.

Erbe AK, Wang W, Carmichael L, **Hoefges A**, Grzywacz B, Reville PK, Ranheim EA, Hank JA, Kim K, Seo S, Mendonca EA, Song Y, Kenkre VP, Hong F, Gascoyne RD, Paietta E, Horning SJ, Miller JS, Kahl B, Sondel PM. Follicular lymphoma patients with KIR2DL2 and KIR3DL1 and their ligands (HLA-C1 and HLA-Bw4) show improved outcome when receiving rituximab. J Immunother Cancer. 2019 Mar 12;7(1):70. doi: 10.1186/s40425-019-0538-8.

In this paper we describe that patients which have the KIR2DL2 gene and its ligand (HLA-C1) as well as KIR3DL1 and its ligand (HLA-Bw4) had improved outcome in comparison to patients without these genotypes. Furthermore, these patients showed an improved duration of treatment response as well as tumor shrinkage as long as they received maintenance treatment. No effect of maintenance treatment was seen for patients without this genotype. This shows that the efficacy of tumor-reactive monoclonal antibody treatment for some patients is influenced by their KIR genes present on NK cells. These findings could help for treatment strategy planning and aid in the decision if a patient might benefit from rituximab treatment or not.

My involvement with this work was the analysis and interpretation of genotyping results for the different KIRs and HLA genotypes of the patients. I also assisted in the editing of the manuscript.

B.3: Combined innate and adaptive immunotherapy overcomes resistance of immunologically cold syngeneic murine neuroblastoma to checkpoint inhibition.

Voeller J, Erbe AK, Slowinski J, Rasmussen K, Carlson PM, **Hoefges A**, VandenHeuvel S, Stuckwisch A, Wang X, Gillies SD, Patel RB, Farrel A, Rokita JL, Maris J, Hank JA, Morris ZS, Rakhmilevich AL, Sondel PM. Combined innate and adaptive immunotherapy overcomes resistance of immunologically cold syngeneic murine neuroblastoma to checkpoint inhibition. J Immunother Cancer. 2019 Dec 6;7(1):344. doi: 10.1186/s40425-019-0823-6.

In this manuscript we describe the generation of a new treatment regimen to cure immunologically cold tumors like 9464D-GD2 which is a good model for human neuroblastoma models as it has a low tumor mutation burden compared to other tumors, is N-MYC amplified and GD2 positive. The conventional in situ vaccine in combination with radiation previously developed in our lab as well as the addition of anti-CTLA4 was not able to cure these tumors even at small sizes which furthermore indicates that this could be a good model for high-risk neuroblastoma. To be able to cure mice of their 9464D-GD2 tumors, these tumors had to be treated with radiation, immunocytokine, anti-CTLA4 and the TRL agonist CpG and CD40. However, this treatment was not able to induce a curative memory response against rechallenge but did slow tumor growth in comparison to naïve mice suggesting that a large part of the immune response is coming from the innate immune populations like macrophages and monocytes which were increased in the tumor microenvironment as seen using flow cytometry.

My contribution to this work was assisting in the design and analysis as well as acquisition of the flow cytometry data characterizing the tumor immune infiltrate and planning and performing the IHC-IF staining and flow cytometry to characterize the 9464D-GD2 line vs. its parental GD2-negative

cell line *in vitro* and *in vivo*. I further assisted in figure generation of the flow and IHC data and editing of the manuscript.

B.4: Pre-existing antitherapeutic antibodies against the Fc region of the hu14.18K322A mAb are associated with outcome in patients with relapsed neuroblastoma.

Goldberg JL, Navid F, Hank JA, Erbe AK, Santana V, Gan J, de Bie F, Javaid AM, **Hoefges A**, Merdler M, Carmichael L, Kim K, Bishop MW, Meager MM, Gillies SD, Pandey JP, Sondel PM. Pre-existing antitherapeutic antibodies against the Fc region of the hu14.18K322A mAb are associated with outcome in patients with relapsed neuroblastoma. J Immunother Cancer. 2020;8(1):e000590. doi:10.1136/jitc-2020-000590

In this manuscript we descripe an interesting finding in pre-treatment sera of neuroblastoma patients. Patients treated with tumor-reactive humanized monoclonal antibody therapy can develop a human anti-human antibody (HAHA). During HAHA evaluations, we found that 9 of 38 patients treated in a phase 1 study of hu14.18K322A (a humanized anti-GD2 mAb) had anti-therapeutic antibody responses in their pre-treatment serum designated as PATA (pre-existing anti-therapeutic antibody). These patients had not been exposed to any mAb therapy prior. Only 4 patients of all 38 patients in the trial showed no disease progression for >2.5 years. All 4 of these patients were PATA positive. We were able to characterize binding of these PATA antibodies further. PATA antibodies showed reactivity against the Fc portion of hu14.18K322A, binding to Rituximab, Dinutuximab and mouse IgG2a isotype antibodies. However, binding was not observed to mouse IgG1 or the fully human Panitumumab.

My involvement with this paper focused on re-organization and formatting of figures, data interpretation, editing of the text and performing some confirmatory ELISA assays.

B.5: In situ Vaccine Plus Checkpoint Blockade Induces Memory Humoral Response.

Baniel CC, Heinze CM, **Hoefges A**, Sumiec EG, Hank JA, Carlson PM, Jin WJ, Patel RB, Sriramaneni RN, Gillies SD, Erbe AK, Schwarz CN, Pieper AA, Rakhmilevich AL, Sondel PM, Morris ZS. In situ Vaccine Plus Checkpoint Blockade Induces Memory Humoral Response. Front Immunol. 2020 Jul 24;11:1610. doi: 10.3389/fimmu.2020.01610. eCollection 2020. PMID: 32849544

This manuscript was the foundation for my thesis work as it demonstrated that the combination of radiation, intratumoral immunocytokine, and checkpoint blockade can induce a humoral antibody response that is targeted against antigens expressed by B78 melanoma cells. We were able to measure these tumor specific antibodies in the serum of disease-free mice for more than 100 days after the mice received treatment. However, in this manuscript, we were not able to show that these antibodies do have a significant involvement in the anti-tumor immune response generated by the in situ vaccine treatment. Furthermore, mice with depleted B cells were still able to generate a strong anti-tumor response that cleared these mice of their tumor.

I contributed to this work by assisting in the cytotoxicity assays as well as the flow cytometry assays to detect tumor-specific antibodies in the serum. I also assisted in the analysis of results and generation of figures as well as editing of the manuscript.

B.6: Intratumoral injection reduces toxicity and antibody-mediated neutralization of immunocytokine in a mouse melanoma model.

Baniel CC, Sumiec EG, Hank JA, Bates AM, Erbe AK, Pieper AA, **Hoefges A**, Patel RB, Rakhmilevich AL, Morris ZS, Sondel PM. Intratumoral injection reduces toxicity and antibody-mediated neutralization of immunocytokine in a mouse melanoma model. J Immunother Cancer. 2020 Oct;8(2):e001262. doi: 10.1136/jitc-2020-001262. PMID: 33115944

This manuscript highlights the benefits of local over systemic anti-tumor antibody injection in the B78 mouse melanoma model. We were able to show that repeated intravenous injections of a humanized immunocytokine increases the frequency of mouse anti-human antibodies (MAHA) that target the immunocytokine. In in vitro assays these MAHA antibodies were able to prevent the antibody dependent cytotoxicity potential of the immunocytokine when co-cultured. In vivo we were able to show that the ability of intravenously injected immunocytokine to bind to B78 tumor cells was negatively impacted. However, a presence of MAHA did not seem to influence the binding potential of the immunocytokine when injected intratumorally.

For this manuscript I helped with data generation using flow cytometry for binding inhibition assays as well as detection of bound immunocytokine to the tumor in vivo. I further assisted with data analysis, experiment design and review of the manuscript.

B.7: Optimizing Flow Cytometric Analysis of Immune Cells in Samples Requiring Cryopreservation from Tumor-Bearing Mice.

Carlson PM, Mohan M, Patel RB, Birstler J, Nettenstrom L, Sheerar D, Fox K, Rodriguez M, **Hoefges A**, Hernandez R, Zahm C, Kim K, McNeel DG, Weichert J, Morris ZS, Sondel PM. Optimizing Flow Cytometric Analysis of Immune Cells in Samples Requiring Cryopreservation from Tumor-Bearing Mice. J Immunol. 2021 Jul 15;207(2):720-734. doi: 10.4049/jimmunol.2000656. Epub 2021 Jul 14. PMID: 34261667

In this paper we developed a cryopreservation method for fresh single cell suspensions from mouse tumors to analyze via flow cytometry. Our results demonstrated that freezing the samples after all staining and fixation was complete most accurately matched the freshly analyzed samples. Cryopreservation before staining and fixation resulted in an altered PD1 expression pattern. This method is needed to safely analyze radioactive samples as those can only be run on a flow cytometer after radiation ahs reached background level and need to be cryopreserved until then.

My contribution to this work was as a consultant for experimental design and analysis of the flow cytometry data. I also assisted in figure design and editing of the manuscript.

B.8: Radiation Augments the Local Anti-Tumor Effect of In Situ Vaccine With CpG-Oligodeoxynucleotides and Anti-OX40 in Immunologically Cold Tumor Models.

Pieper AA, Zangl LM, Speigelman DV, Feils AS, **Hoefges A**, Jagodinsky JC, Felder MA, Tsarovsky NW, Arthur IS, Brown RJ, Birstler J, Le T, Carlson PM, Bates AM, Hank JA, Rakhmilevich AL, Erbe AK, Sondel PM, Patel RB, Morris ZS. Radiation Augments the Local Anti-Tumor Effect of In Situ Vaccine With CpG-Oligodeoxynucleotides and Anti-OX40 in Immunologically Cold Tumor Models. Front Immunol. 2021 Nov 15;12:763888. doi: 10.3389/fimmu.2021.763888. PMID: 34868010

In this paper we tested the efficacy of CpG and anti-OX40 as an *in situ* vaccine. While this treatment was effective in an A20 lymphoma model, it did not significantly improve survival or tumor response in a B78 melanoma or 4T1 breast cancer model. However, combining this treatment with radiation enabled a local anti-tumor immune response in B78 and 4T1 tumors. This treatment increased the ratio of tumor-infiltrating effector T cells to T regulatory cells as well as CD4+ and CD8+ T cell activation in tumor draining lymph nodes and spleen. Radiation increased the expression of OX40 on tumor-infiltrating CD4+ non-regulatory T cells.

I contributed to this work by analyzing the flow cytometry data as well as assisiting with panel design. I further helped with graphical representation of the data and editing of the manuscript.

B.9: Single cell metabolic imaging of tumor and immune cells in vivo in melanoma bearing mice.

Heaton AR, Rehani PR, **Hoefges A**, Lopez AF, Rakhmilevich AL, Erbe AK, Sondel PM, Skala MC. Single cell metabolic imaging of tumor and immune cells in vivo in melanoma bearing mice. Front Oncol., 20 March 2023; 13: 1110503. doi.org/10.3389/fonc.2023.1110503

In this manuscript we report an immunocompetent mCherry reporter mouse model for immine cells with CD4 expression either during differentiation or in their mature state and perfrom *in vivo* imaging o fimmune and tumor cells within a B78 melaonma model as well as a single cell segmentation algorithm of mCherry-expressing immune cells within *in vivo* images. We also performed flow cytometry of tumor and immune cell populations to complelent *in vivo* imaging data. Some of our findings were that immune cell size increased in tumor compared to spleens, immune cells within B78 tumors exhibited decreased FAD mean lifetime and an increased proportion of bound FAD compared to immune cells in the spleen. All of these findings are consistent with a more activated phenotype of immune cells in the tumor vs. the spleen. We propose that this approach can be used to monitor single cell metabolic heterogeneity in tumor cells and immune cells to study promising treatments for cancer in the native, label-free *in vivo* context.

My contribution to this manauscript was the design, execution and analysis of all flow cytometry data included. I designed the panel, dissociated tumors and spleens, stained the samples and ran them on the flow cytometer and also performed the analysis. I further helped generate the figures based on flow cytometry data and edited the manuscript.