UNIVERSITY OF CALIFORNIA

Santa Barbara

Discovery of disease-associated antibody biomarkers and their binding targets using bacterial displayed peptide libraries

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemical Engineering

by

Serra Eren Elliott

Committee in charge:

Professor Patrick S. Daugherty, Chair

Professor M. Scott Shell

Professor Samir Mitragotri

Professor Jamey Marth

Alex R. Soffici, M.D.

December 2014

The dissertation of Serra Eren Elliott is approved.

_____
Alex R. Soffici

_____
Jamey Marth

_____
Samir Mitragotri

_____
M. Scott Shell

_____
Patrick S. Daugherty, Committee Chair

October 2014

Discovery of disease-associated antibody biomarkers and their binding targets using bacterial displayed peptide libraries

ACKNOWLEDGEMENTS

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny…'" -Isaac Asimov

To all of my colleagues at UC Santa Barbara and Santa Barbara friends, thank you for helping to make these past five years truly memorable. This journey not only taught me about science and research, but through our shared experiences, friendships, and discussions, I learned important lessons on life and keeping things in perspective.

I would like to thank my advisor, Patrick Daugherty, for supporting me throughout my doctoral studies and providing the necessary tools and environment for my growth as an independent researcher. You provided overall goals and the freedom to explore these, enabling my development as a scientist. Over the past five years, we have made some exciting discoveries of which I am very proud.

I also want to thank our collaborator and my doctoral committee member Alex Soffici. You have been a tremendous source of support, and I enjoyed learning from your medical expertise in pre-eclampsia and hearing about all of your exciting stories and experiences. I have truly appreciated our friendship. Additionally, I want to express my gratitude for our collaborators at the University of Texas Medical School at Houston. The samples provided and scientific discussions have been incredibly helpful for the development of this work. Furthermore, to the rest of my committee members, Professors Scott Shell, Samir Mitragotri, and Jamey Marth, thank you for all of the interesting scientific discussions over the past few years.

I want to give a huge thanks to all of the members of the Daugherty lab, past and present. You all kept things interesting both in and out of the lab. Jen Getz, you provided so much helpful guidance right from the start and you and Ian together always helped keep things in perspective. Jack, Jen Guerrero, and Kelly, thank you for all of the discussions about science and life and of course helping me edit my dissertation. Also, thanks to Jen Guerrero for taking over the FACS duties!

I also want to thank my family who has provided so much support and love during this time. To Anne and Baba, I am happy to join the ranks as another Ph.D. in the family. I am so lucky to have such supportive parents who always provided a learning environment at home, instilled in me the importance of hard work, and gave me so many unique experiences. Eileen, you are the most amazing little sister and your visits here throughout the years have always brought a lot of fun and excitement.

Finally, Brian, there are no words to express how thankful I am to have you in my life. You are my best friend and favorite study buddy. You have always been there to listen to me, remaining patient, despite my obsessive perfectionism. Together, we made it through this journey in pursuit of our doctorates. I am so excited for the future we have in store as husband and wife!

VITA OF SERRA EREN ELLIOTT

October 2014

EDUCATION

University of California, Santa Barbara Santa Barbara, CA 2009-2014
Doctor of Philosophy in Chemical Engineering

The Ohio State University Columbus, OH 2005-2009
Bachelor of Science in Chemical Engineering
Graduated *Magna Cum Laude* with Distinction

PROFESSIONAL EXPERIENCE

Graduate Student Researcher UCSB Winter 2010-Present
Dissertation: "Discovery of disease-associated antibody biomarkers and their binding targets using bacterial displayed peptide libraries"
Academic Advisor: Patrick S. Daugherty

Undergraduate Student Researcher OSU Fall 2007-Spring 2009
Honors Thesis: "Cell damage due to hydrodynamic stress in fluorescence activated cell sorters"
Academic Advisor: Jeffrey J. Chalmers

R&D - Products Research Intern Procter & Gamble June 2007 to Sep. 2007
Cincinnati, OH
Project: Established three lead improvement ideas for Charmin® Ultra Soft based on focus group discussions and an ideation session and began groundwork development

Research Engineering Intern M&G Polymers June 2006 to Sep. 2006
Sharon Center, OH
Project: Enhanced and created process and instrumentation drawings (P&IDs) for the pilot plant and worked as a technician in analytical and applications labs

PUBLICATIONS

- **Elliott SE**, Parchim NF, Liu C, Yang X, Kellems RE, Soffici AR, and Daugherty PS. "Characterization of antibody specificities associated with preeclampsia." *Hypertension* (2014)
- **Elliott SE**, Parchim NF, Yang X, Kellems RE, Soffici AR, and Daugherty PS. "A pre-eclampsia associated Epstein-Barr virus antibody cross-reacts with placental GPR50." (2014, in preparation)
- Pantazes R, Reifert J, **Elliott SE**, and Daugherty PS. "IMUNE: Identifying Motifs Using Next-generation sequencing Experiments." (in preparation)

PATENTS

- Daugherty PS, **Elliott S**, and Soffici A. "Peptides that bind to pre-eclampsia specific antibodies, pre-clampsia specific peptides and uses thereof" U.S. Provisional Patent Application No: 61/926,774 (January 13, 2014)

PRESENTATIONS

- Oral presentation, "Discovery of antibody biomarkers present in pre-eclampsia" at AIChE Annual Meeting in San Francisco, CA (2013)
- Oral presentation, "Discovery of antibody biomarkers present in pre-eclampsia" at 6th Annual Amgen-Clorox Graduate Student Symposium at UCSB (2013)
- Oral presentation, "Discovery and characterization of antibody biomarkers of pre-eclampsia and reagents for their detection" at the 18th World Congress for International Society for the Study of Hypertension in Pregnancy (ISSHP) in Geneva, Switzerland (2012)
- Poster presentations, "Discovery and characterization of antibody biomarkers in autoimmune disease and reagents for their detection"
  - 12th Annual UC Systemwide Bioengineering Conference in Santa Barbara, CA (2011)
  - Graduate Student Symposium at UCSB (2011, 2012)
- Co-author on oral presentation, "Cell damage of human and industrial cell Lines in bioprocess equipment," at AIChE Annual Meeting in Nashville, TN (2009)

HONORS and AWARDS

- Women's Initiatives Committee Travel Award for the AIChE Annual Meeting in San Francisco, CA (2013)
- Best Oral Presentation at 6th Annual Amgen-Clorox Graduate Student Symposium at UCSB (2013)
- Travel Bursary Award for the World Congress for the International Society for the Study of Hypertension in Pregnancy (ISSHP) in Geneva, Switzerland (2012)
- UCSB Doctoral Student Travel Grant (2012)
- Best Poster in Devices category at 12th Annual UC Systemwide Bioengineering Symposium (2011)
- OSU FEH Nanotechnology Project – Most Innovative Project Award (Spring 2006)

MENTORING and TEACHING EXPERIENCE

- Facilitator for Graduate Student Showcase at UCSB — Spring 2014
- Mentor for EUREKA! Program — Summer 2013
  - Early Undergraduate Research and Knowledge Acquisition - 10 week summer research program with individual undergraduate
- Mentor for SIMS Program — Summer 2012
  - Summer Institute in Mathematics and Science – incoming undergraduate students
- Mentor for Summer Doctoral Research Institute (SDRI) — Summer 2010
- "Ask a Scientist" Contributor to UCSB's ScienceLine — Winter 2010-2012
- Guest Lecturer for ChE 170: Molecular and Cell Biology for Engineers — Fall 2011

SKILLS and TECHNIQUES

- Multi-color and single-color flow cytometry/cell sorting
- Bacterial display culture and cloning
- Display library construction and screening/directed evolution
- Peptide motif characterization
- Protein database searches: BLASTp and ScanProsite
- Immunoassay development
- Machine learning classification algorithms
- Mammalian cell culture and transient transfections
- Affinity-based protein purification
- Next-generation sequencing
- Microarrays
- IHC, SDS-PAGE, Western Blot
- Programming: R, Python, MATLAB

ABSTRACT

Discovery of disease-associated antibody biomarkers and their binding targets using bacterial displayed peptide libraries

By

Serra Eren Elliott

Discovery of biologic molecules specific to a diseased state, or biomarkers, can lead to diagnostic development, therapeutic target identification, and improved understanding of disease pathogenesis. Antibodies remain an attractive class of biomarkers given their amplification by the immune system, stability, and current clinical use. While the antibody repertoire represents a rich source for biomarker discovery, it has been difficult to impartially identify which molecules from this repertoire are associated with disease. This work demonstrates three molecular discovery processes centered on the utility of bacterial displayed peptide libraries and fluorescence activated cell sorting (FACS) for identifying novel antibody biomarkers and their targets. We applied these methods to discover and characterize disease-associated antibodies for pre-eclampsia (PE), a condition with unknown etiology that affects 5-8% of pregnancies, using an unbiased approach.

Applying three quantitative screening strategies against a set of PE and healthy-outcome pregnancies (HOP) identified unique disease-associated antibody binding peptides from a fully random 15 amino acid peptide library. With a two-color screening method, we used antibody fractions enriched from plasma to isolate

significantly PE cross-reactive and specific peptides distinct from a previously identified PE-associated antibody specificity. We used a panel of these antibody-detecting peptides to train and validate an Adaptive Boosting classification algorithm that achieved high specificity (95%) and a validated overall 80% diagnostic accuracy. To more closely replicate the native antibody binding environment, a second screening method used unprocessed, diluted plasma. This approach sequentially enriched peptides binding to PE antibodies and removed HOP antibody binders, resulting in a strong consensus motif that we further expanded through directed evolution. Importantly, we linked this motif to a region of a common viral antigen, Epstein-Barr virus nuclear antigen 1, and a human G protein-coupled receptor, GPR50, presenting a novel case for molecular mimicry. Thus, this method enabled unbiased identification of a disease-associated antibody and characterization of its targets. Finally, we developed and applied a unique methodology that combines bacterial displayed library screening with next-generation sequencing to profile the antibody repertoire of individual PE patients and HOP samples. This analysis re-identified the viral antigen-linked motif among several distinct PE- and HOP-associated antibody specificities, providing broader insights into alterations to the immune repertoire in PE.

This work demonstrates the utility of screening bacterial displayed peptide libraries to profile the antibody repertoire and identify new markers of disease. These disease-associated antibody-detecting peptide reagents enable development of molecular diagnostics and discovery of antibody binding target(s) to improve understanding of disease etiology and potentially elucidate therapeutic targets.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Background

The circulating human antibody repertoire catalogs the immune response to a variety of different targets, including pathogens, environmental allergens, and human proteins. Determining and characterizing the targets of specific antibodies associated with a disease can lead to diagnostic development, improved understanding of disease pathogenesis, and elucidation of potential therapeutic targets. However, it remains particularly difficult to unambiguously identify these disease-associated antibodies and their targets. Through this research, we demonstrate methods that do not require sophisticated knowledge of disease pathogenesis to identify molecular diagnostic tools and elucidate antibody binding targets, which may enable future therapeutic development. To emphasize the importance of an unbiased approach, we apply these methods to pre-eclampsia (PE), a disease with unknown etiology that affects 5-8% of pregnancies.[1] While previous studies reported the presence of antibodies against a specific region of the angiotensin II type 1 receptor in PE patients,[2,3] these have been difficult to detect and vary in prevalence. This chapter highlights the need for the development and application of unbiased antibody biomarker discovery tools, providing motivation for this work and describing the techniques involved.

A. Antibodies as disease-specific biomarkers

In clinical proteomics, the proteins in diseased patients' samples (i.e., blood, tissue) are compared with normal samples to determine biologic molecules unique to a diseased state, or biomarkers.[4] Work in this area has been applied to a broad spectrum of diseases, such as cardiovascular disease,[5] cancer,[6] Alzheimer's,[7] and

systemic lupus erythematosus,[8] and has involved a variety of techniques, from mass spectroscopy[9,10] and protein microarrays[11] to gene transcription profiling.[12]

In particular, the circulating antibody repertoire represents a rich source for biomarker discovery, due to easy access (blood sample), stability, and current clinical utility. Antibodies, or immunoglobulins (Ig), can be divided into several different subtypes (IgG, IgM, IgA, IgE, IgD), of which IgG is the most prevalent (~75%) circulating class. The Ig core structure consists of two heavy and two light chains comprising a constant fragment (Fc) linked to two regions responsible for binding (Fab), resulting in multivalency. These multivalent antibodies bind to antigenic targets, which include proteins, carbohydrates, lipids, and nucleic acids. The strength of the binding interaction between antibody (Ab) and antigen (Ag) is termed affinity, usually described by the equilibrium dissociation constant (K_D):

$$Ab + Ag \underset{k_d}{\overset{k_a}{\rightleftharpoons}} Ab \cdot Ag \quad (1.1)$$

$$K_D = \frac{k_d}{k_a} = \frac{[\mathrm{Ab}][\mathrm{Ag}]}{[\mathrm{Ab} \cdot \mathrm{Ag}]} \quad (1.2)$$

The affinity of the antibody-antigen interaction plays a role in the ability to detect antibody markers. A typical antibody affinity is 10^{-6} to 10^{-9} M, and this can be thought of as the concentration of antibody at which half of the antigen is bound. Thus, increased affinity (lower K_D) results in a lower detection limit. Specificity is another important consideration for a diagnostic assay. Using an antigen that binds multiple nonspecific antibodies to detect a disease-associated antibody increases the number of falsely diagnosed healthy samples.

Since antigenic targets are typically quite large, antibodies only recognize a small section of the antigen, termed an epitope. Similarly, a small region, paratope, of the antibody composed of complementary determining regions mediates the majority of binding. Epitopes can represent linear and discontinuous, conformational sections of a given antigen, or even complex macromolecular structures.[13] Typical linear epitopes comprise 6 to 13 amino acids[14] while conformational epitopes include stretches of linear regions typically resulting in 9 to 22 amino acids interacting with the antibody.[15] These discontinuous epitopes are often difficult to precisely identify, resulting in large sections of proteins deemed important for antibody binding.[16] Additionally, certain protein structures, such as β-turns, are often observed in antigenic sites,[17] further impressing the importance of structural context for antibody binding.

Characterizing the precise epitope of interaction can clarify important differences associated with a specific disease. For example, in celiac disease (CD), an immune response to gluten, a protein found in wheat, barley, and rye, leads to production of antibodies against the human protein, tissue transglutaminase (tTG). Although rare, individuals with other autoimmune diseases demonstrate tTG antibody positivity in certain enzyme-linked immunosorbent assays (ELISAs).[18] However, characterizing a discontinuous epitope for tTG antibodies in CD patients clarified differences between CD and these other disease patients with anti-tTG positivity.[19]

Regardless of the precise mechanism of antigen interaction, many examples of antibody biomarker based diagnostics exist (**Table 1-1**). Sensitivity and specificity are common metrics for assessing overall accuracy of a diagnostic assay. In this

respect, sensitivity represents the number of true positives (correctly diagnosed as disease) out of the total number of disease tested. Conversely, specificity indicates the number of true negatives (correctly classified as healthy) out of the total number of healthy evaluated.

B. Aberrant antibody production and molecular mimicry

By interacting specifically with antigenic targets, antibodies perform a diverse set of functions and are necessary for healthy survival and protection against foreign molecules. B-cells secrete antibodies upon binding of the appropriate antigen to the B-cell receptor and the necessary stimulation in a T-cell independent or dependent manner.[20] Once appropriately stimulated, the B-cell undergoes an affinity maturation process, which fine-tunes the binding interaction by selecting for variants with increased affinity to the antigen after somatic hypermutation of the B-cell receptor binding region. Focusing on the T-cell dependent process, antigens bound to the B-cell receptor are internalized, digested, and displayed as fragments on MHC II, which must then interact with the appropriate T-cell receptor. This two-step process helps regulate antibody production, reducing the likelihood of self-reactive antibodies.

Despite this two-step process, various diseases are associated with production of unusual, often pathological antibodies, including chronic hypertension,[21] Alzheimer's,[22] chronic graft versus host disease[23] and other autoimmune conditions. Therefore, sometimes a breakdown in self-reactive antibody production control can occur. This breakdown could be due to aberrant epitope spreading as hypothesized for the generation of autoantibodies against tTG in CD.[24] In this case, gliadin, a component of gluten, forms a complex with tTG as it acts upon the glutamine

residues present. Thus, a tTG specific B-cell may endocytose the complex but present a gliadin peptide to a gliadin specific T-cell that stimulates the tTG specific B-cell to start producing antibodies.[25] This epitope spreading phenomenon is associated with other diseases and has been shown to precede clinical presentation of rheumatoid arthritis.[26] Furthermore, if the specific residues involved in the antibody-antigen binding are structurally similar to additional proteins, molecular mimicry between proteins can increase or trigger autoantibody production.[27] For example, three autoantibodies associated with lupus patients have been shown to cross-react with regions of a common viral protein, Epstein-Barr virus nuclear antigen 1 (EBNA-1).[28] Molecular mimicry mechanisms have been proposed for several autoimmune diseases (**Table 1-1**): Guillain-Barré syndrome and *Campylobacter jejuni*;[29] myasthenia gravis and herpes simplex virus 1;[30] multiple sclerosis (MS) and Epstein-Barr virus (EBV), measles and HHV-6;[30] type-1 diabetes (T1D) and Coxsackie B virus, rubella, and rhinovirus;[30] and scleroderma and human cytomegalovirus.[30] In certain cases, such as MS and T1D, the molecular mimicry mechanism is thought to involve cross-reactive T-cells.[31]

As demonstrated by these examples, the molecular mimicry mechanism appears to play a significant role in generating autoimmune activity. Discovering new molecular mimicry cases that drive autoantibody activity can improve our understanding of disease etiology and lead to new diagnostics and potentially therapeutics. A variety of approaches have been used to discover and characterize disease-associated antibodies.

Table 1-1: Examples of autoantibody markers of disease and hypothesized molecular mimicry in disease pathogenesis

Disease	Target	Sens	Spec	Molecular Mimicry Proposed	Refs
Myasthenia Gravis	Acetylcholine receptor	0.96	0.97-0.99	Herpes simplex 1 virus, gpD	30,32,33
Celiac disease	Tissue transglutaminase	0.89[a]	0.98[a]	Gluten, Hepatitis C virus	30,34
Type-1 diabetes	Islet cells	0.81[b]	0.96[b]	Enteroviruses, rotaviruses	30,35,36
Systemic lupus erythematosus	Nuclear antigens (ANA)	0.98	0.92[c] 0.88[d]	Epstein-Barr virus, EBNA-1	28,37
	dsDNA	0.37	1[c] 0.97[d]		
Systemic sclerosis	Integrin-NAG-2 complex (endothelial cells)	0.93[e]	1[c,e]	Cytomegalovirus, UL94	30,38,39
	Topoisomerase1 (Scl-70)	0.43[c,f]	1 [c,f]		
Graves' disease	Thyroid stimulating hormone receptor	0.988	0.996[c]	*Yersinia enterocolitica*	40,41,42
Rheumatic fever	Cardiac myosin	0.87	1[d]	Streptococcal	43,44
Autoimmune liver disease	Cytochrome P4502D6	0.81	0.98	Hepatitis C virus, cytomegalovirus, herpes simplex virus	45,46
Rheumatoid arthritis	Cyclic citrullinated peptide	0.48	0.96	*Proteus mirabilis*	47,48
	Fc domain of IgG (IgM-Rheumatoid factor)	0.54	0.91		

Sens – Sensitivity, Spec – Specificity, a – pooled estimate, b – median from review, c – compared to healthy controls, d- compared to disease control patients, e – using a mimicking library-isolated peptide, f – measured by ELISA

C. Library screening to characterize disease-associated antibodies

For the majority of diseases, a sensitive/specific antibody biomarker for diagnosis remains unidentified. Furthermore, the push for more personalized medicine requires

new discovery efforts to help typify different diseases or predict and monitor response to a specific therapeutic.[49] These discovery efforts for new antibody biomarkers involve a diverse set of techniques. While there has been some work to develop techniques to characterize unique complementary determining regions of antibodies using mass spectrometry,[50] it remains difficult to apply and distinguish disease-associated antibody markers. More recently, with enhanced sequencing techniques, studies have been able to sequence the antibody repertoire itself and identify disease-associated antibodies.[51,52,53] However, in any effort to identify disease-specific antibody markers, it is important to characterize what the antibody binds. Thus, a common technique is to screen a set of candidate antigen targets for binding to disease-associated antibodies.

i. Protein microarrays

Protein microarrays enable high-throughput evaluation of proteins of interest for binding to disease-associated antibodies in a multiplexed manner.[54,55] Up to 10^3-10^4 proteins are immobilized to a substrate, such as a glass slide, sometimes through affinity tags.[56] By modifying substrate surfaces, proteins attach by noncovalent (e.g., hydrophobic or positively charged surfaces) or covalent (e.g., chemically activated surfaces such as epoxy or aldehyde esters) approaches.[57] Alternatively, protein arrays can be generated using cDNA microarrays followed by *in situ* transcription and translation directly before probing for binding.[11,58] While 10^3-10^4 proteins represents a fair amount of candidate targets, the human proteome alone may range from 10^5 to several million molecules, due to post-transcriptional control and post-translational modifications.[57] This does not include the potential myriad of environmental

organisms involved in disease. Thus, selecting a set of proteins to evaluate with a microarray often requires pre-existing knowledge of the disease. Utilizing these previously gained insights into disease etiology can lead to new discoveries, albeit in a specific direction. For example, based upon the knowledge that lupus patients possess antibodies against certain "serum factors," such as cytokines, chemokines, and growth hormones, a protein microarray was generated.[59] Assaying this microarray for binding to serum antibodies in lupus patients and healthy controls identified B-cell activating factor (BAFF) as an autoantigen target among others for lupus-associated antibodies.

In addition to the difficulty and bias introduced by selecting a set of proteins to analyze, generating the microarrays requires production, folding, and purification of these proteins, which adds significant complexity. Given the solubility requirements for protein microarray analysis, membrane proteins, such as G protein-coupled receptors (GPCRs), must often be excluded from analysis.[56] Some work has been done to print GPCRs and their associated lipids into microarrays using modified surfaces,[60] but the requirement of detergents limits throughput. Furthermore, with full-length protein microarrays, one cannot refine the important epitopes mediating the observed binding signal.

ii. Peptide library screening

Peptide library screening identifies short binding ligands for a variety of targets, such as disease-associated antibodies, out of a large pool of candidates. These peptides can be used as diagnostic reagents themselves and serve as antigen surrogates, enabling a more refined understanding of the epitope involved in an

antibody-antigen interaction. Although libraries of peptoids, synthetic peptide mimetics, have shown promise for disease detection,[7,61] these are difficult to relate to the native antigen of interest. Peptide libraries can be displayed on a cell surface or assayed as a solid-phase ELISA,[17] microarray,[62] or nitrocellulose membrane (SPOT).[63] In one application, a peptide microarray consisting of 7,446 overlapping 15 amino acid peptides representing 61 *M. tuberculosis* proteins was probed against tuberculosis positive and negative individuals.[64] This analysis demonstrated that positive and negative individuals produced antibodies against different epitope specificities from the same protein while certain shared epitopes exhibited differential activity. The SPOT synthesis methodology has been shown to map discontinuous epitopes of a known antigen; however, it requires high peptide density and protein concentration to identify these low affinity binders.[65] By incorporating knowledge of the distance between discontinuous regions in a solved protein structure, these discontinuous regions were linked and substitution analysis led to improved binding affinity for the antibody.[66] Although the SPOT and peptide microarray techniques can be useful, these approaches often require pre-existing knowledge of the antigen(s) of interest because peptide synthesis cost and efficiency limits the library diversity probed. Since cell surface display uses the natural machinery of the cell to produce and display these libraries, the diversity probed can be significantly larger, 10^9-10^{11} members.[67] These libraries of peptides can be fully randomized or represent linear fragments of protein(s) of interest derived from the human proteome,[68] a specific virus,[14,69] or a particular cell line.[70] Peptides identified from a randomized library may represent a linear sequence or mimic a discontinuous, conformational epitope

(i.e., mimotopes[71]). In fact, certain identified mimotopes can mimic the binding site of antibodies targeting double-stranded DNA.[13] In contrast, libraries composed of linear overlapping fragments from a specific protein of interest typically cannot detect conformational epitopes. The gene fragment library approach, which consists of different sized fragments from proteins of interest, can identify short linear epitopes and large sections of a protein that contain the residues of a discontinuous epitope.[72] In some cases, the known discontinuous regions can be used in conjunction to identify specific residues involved in binding.[73] However, these approaches are more complex and require *a priori* knowledge of the disease to develop these fragment libraries in contrast to random peptide libraries.

iii. Library screening by cell-surface display

Peptide display libraries are a major tool for screening and identifying disease-associated antibody binding peptides. Commonly used display platforms include phage, bacteria, and yeast. Phage, or bacteriophage, displayed libraries typically use M13 or T7 phage and have shown great utility for antibody repertoire profiling. Phage display screening against disease-associated antibodies has been applied to a variety of conditions, including infectious disease (e.g., Lyme disease[74]), inflammatory disease (e.g., ankylosing spondylitis[75]), and cancer (e.g., prostate,[76,77] breast,[70] and non-small cell lung[78]), using random[74-76] and cDNA libraries.[70,77,78] Furthermore, screening random phage display peptide libraries against autoimmune diseases has yielded interesting observations in multiple sclerosis,[79] celiac disease,[80] and type-1 diabetes.[81] In MS, four identified mimotopes bound antibodies in the cerebrospinal fluid in a higher fraction of MS patients compared to disease control,

and these shared similarities with envelope proteins of the MS-associated retrovirus.[79] Screening IgA from CD patients with high anti-gliadin titers identified the PEQ sequence as important for binding to gliadin.[80] Additionally, a monoclonal antibody from a T1D patient was screened against a randomized peptide library, which identified a common motif among isolated peptides.[81] Similar cluster patterns were found on the surface of islet antigen 2 for which antibody specificities in type 1 diabetes are hypothesized to be conformational. However, in each case, prior knowledge of disease-related proteins, made these discoveries possible. Additionally, phage cannot be quantitatively sorted because their small size (24-200 nm) prevents flow cytometric analysis.[82] Instead, a simple selection process, termed panning, is used to enrich binding phage. Here, antibodies are typically immobilized to a surface or beads and phage are panned for binding to the targets. Furthermore, each panning step is followed by an amplification step hosted in bacteria, which greatly reduces library diversity independently from peptide function selection.[83]

Bacterial displayed peptide libraries utilize the processing capabilities and fast replication of bacteria to display and screen a diverse collection of peptides for binding. Bacteria provide a convenient, more easily accessible link between the gene and protein binding phenotype. Moreover, bacterial display enables analysis by flow cytometry and fluorescence activated cell sorting (FACS) to quantitatively assess the degree of binding based on cell fluorescent intensities,[84] discussed in detail later in this chapter. Bacteria can also be selected in a similar fashion as phage display but without the biasing amplification process, such as with the flagellar display system, FliTrx.[85] This system has been applied to identify antibody binding specificities

present in scleroderma,[38] Sjögren's syndrome,[86] Cogan's syndrome,[87] and autoimmune pancreatitis.[88] Of these, the antibody binding specificity discovered for autoimmune pancreatitis exhibited the strongest diagnostic performance with binding activity observed in 18 of 20 patients and only 4 of 40 with pancreatic cancer. The identified peptide sequence was linked to a protein of *Helicobacter pylori*, and the corresponding peptide exhibited similar diagnostic utility in a validation set.

In addition to the FliTrx system, another bacterial display system of random peptide libraries utilizes FACS to identify peptides of interest. The library displays using a circularly permuted version of the outer membrane protein X in *E. coli*,[89] for which the display efficiency was further enhanced (eCPX).[90] The scaffold enables peptide expression on both the N- and C-terminus, which provides more peptide flexibility than simple loop insertion into the native transmembrane protein. This permits construction of a library of cells expressing unique peptides that behave more like in solution. To demonstrate the capabilities of this display system, a 15-mer library was screened against a pseudo disease consisting of a monoclonal antibody spiked into an immunoglobulin pool.[91] Multi-parameter cell sorting and differently labeled "disease" and control pools enabled selection of "disease" specific peptides, which were enhanced with directed evolution.

While yeast display platforms can be screened by flow cytometry, these require longer growth times and due to poor transformation efficiency require multiple transformations to achieve $>10^7$ diversity. A benefit to yeast display involves its ability to correctly fold and express larger proteins, such as fragments of epidermal growth factor receptor.[92] Conformational and linear antibody epitopes were

distinguished using these fragments. Similarly, fragment display enabled refinement of the binding target for an H1N1 neutralizing antibody.[93] These examples highlight the benefit of yeast display to express large structural domains from proteins of interest, but this approach again requires pre-existing knowledge of potential disease targets to characterize antibody binding.

It is important to note that multiple copies (~10,000 with eCPX) of the peptide are expressed on the cell surface. Therefore, binding interactions to multivalent antibodies not only depend on the individual affinity of the peptide for the target but also the avidity contributions based on the number of copies displayed on the cell surface. The peptides on the surface can be modeled as multivalent ligands and the avidity is determined by the inter- and intramolecular binding affinities. The distance between two peptides depends upon the overall density of peptide expression on the cell surface. With bivalent targets, such as IgG, if two peptides are displayed close to each other, two unbinding events must occur in rapid succession for the antibody to completely dissociate from the cell surface. Thus, these avidity interactions enhance apparent affinities and, importantly, increase the ability to detect the presence of antibodies with low affinity or titer.[94,95]

iv. Identifying antibody-detecting peptides from a bacterial display library

With the tools described above, the disease-associated antibody-binding peptides can be selected from a random peptide library using a quantitative molecular separation process through flow cytometry (see Techniques section). Importantly, as mentioned above, random peptide libraries do not necessitate prior knowledge of disease, unlike cDNA libraries generated from tissues or cells previously linked to

disease.[14,70,77,78] Some studies used few individual patient samples[76,79] or a single pool of samples[38,86-88] for discovery. However, this approach biases antibody identification to a subset of patients, deterring identification of widely prevalent antibodies. By incorporating FACS, peptides can be quantitatively screened for cross-reactivity among different patient samples. These peptides represent antigen surrogates, which enables antigenic target characterization based upon sequence alignments of identified peptides. Previous studies with random peptide libraries often relied upon few peptide sequences (<10) [38,79,81,86-88] and pre-existing knowledge of disease-related antigens[74,79-81,86,88] to relate motif similarities to potential antigenic targets. However, greater sequence depth facilitates stronger motif characterization to understand the variability in certain positions, which would benefit a directed evolution strategy for unbiased antigen discovery. Programs, such as PILEUP,[96] Clustal series,[97] and MEME,[98] are available to perform these sequence alignments. These tools relate amino acids using similarity matrices, such as the blocks substitution matrix (BLOSUM)[99] and the Tudos matrix generated using the idea of "neighborhood selectivity."[100] Highly enriched populations of short (~15 amino acids) peptide sequences can often be aligned manually based upon these similarity matrices and amino acid properties. Thus, bacterial displayed peptide library screening enables profiling of the immune repertoire to identify aberrant antibody specificities involved in a disease, potentially developed in response to an environmental trigger, such as a virus (**Figure 1-1**). The following section describes the motivation behind applying this method to profile the immune response in the pregnancy-related disease, pre-eclampsia.

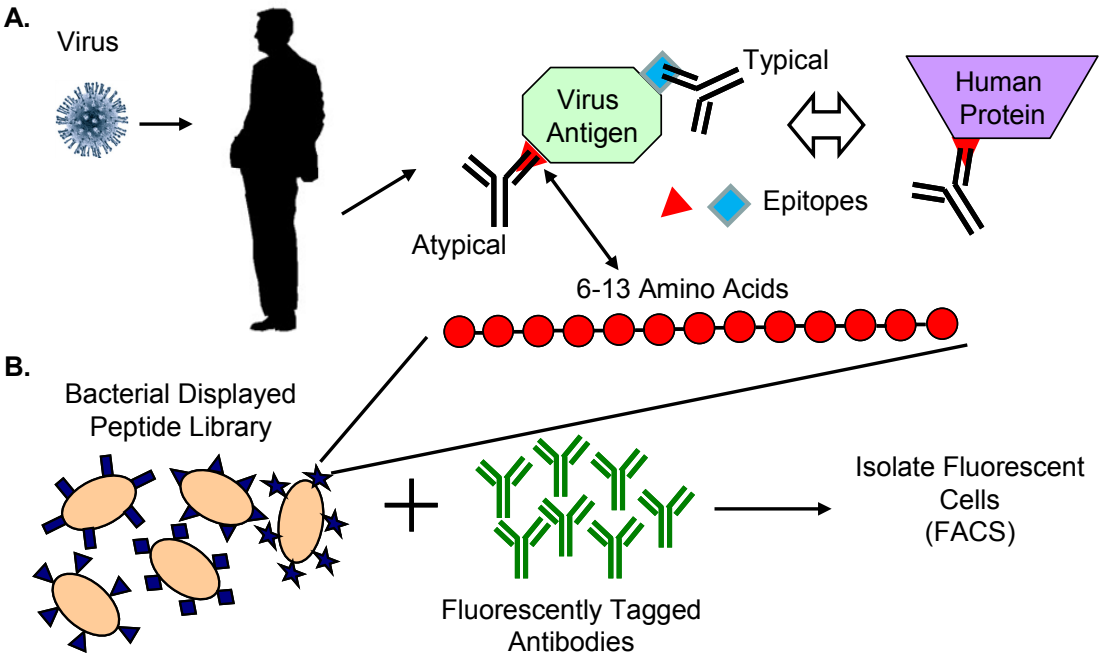

Figure 1-1: Bacterial displayed peptide library screening to profile the antibody repertoire. (**A**) An immunologic response to a particular pathogen, such as a virus, initiates production of antibodies against specific proteins, or antigens. These can result in binding to antigen regions, or epitopes, typically observed or an atypical response. When these atypical specificities cross-react with regions of human proteins, a process known as molecular mimicry, these aberrant antibodies can lead to various problems. (**B**) Since these linear epitopes are typically 6-13 amino acids in length a 15 amino acid peptide library can be used to represent these important binding regions. With fluorescently tagged antibodies, a quantitative separation process, known as FACS, can be used to select the relevant disease-associated specificities.

D. Pre-eclampsia

Pre-eclampsia (PE) affects 5-8% of pregnancies[1] and is a major cause of maternal mortality (15-20% in developed countries) and morbidities.[101] Furthermore, it is responsible for an estimated 500,000 fetal deaths worldwide each year. PE is marked by placental abnormalities, such as poor placentation[102,103] leading to intrauterine growth restriction (IUGR), as well as a maternal inflammatory response and endothelial dysfunction, which are involved with the mother's clinical symptoms.[103] Established guidelines exist for diagnosing PE upon presentation with maternal symptoms, such as hypertension and proteinuria after 20 weeks' gestation;[104]

however, the disease pathogenesis remains poorly understood and presentation involves a broad clinical spectrum besides hypertension and proteinuria.[105] It has recently been proposed that PE may represent a combination of several disease subtypes and future studies should recognize this possibility.[106] Alongside the complicated nature of PE presentation and management, the average cost of a PE case is estimated to be £9000 ($14,000).[107] Given the global impact of PE, a huge void remains to be filled with respect to advancing diagnostic methods, understanding disease pathogenesis, and developing an eventual therapeutic.

i. Current understanding of disease

Previous work has demonstrated that numerous contributing factors may lead to the maternal PE condition. Several molecules, including cystatin C, beta-trace protein, and beta-2-microglobulin, increased in healthy pregnancy show further elevation in women with PE.[108] Notably, disturbances to angiogenic/antiangiogenic protein levels have been observed in PE. Elevated levels of antiangiogenic proteins, soluble fms-like tyrosine kinase 1 (sFlt-1) and soluble endoglin (sEng), findings attributed to increased shedding of placental microparticles,[109] and lowered levels of placental growth factor (PlGF) are often present in PE patients.[110,111,112] sFlt-1 binds vascular endothelial growth factor (VEGF), thereby inhibiting VEGF's normal proangiogenic activity. The placenta-derived protein sEng induces vascular permeability and hypertension *in vivo* and is present in the sera of all pregnant women but elevated in PE patients.[111] Although high sFlt-1 and low PlGF levels have been demonstrated in a large number of patients, not all women with these altered levels develop PE and many PE patients present without these differences,[113,114] indicating

the existence of other factors that contribute to PE. One factor contributing to the altered renin-angiotensin system may come from oxidative stress due to placental dysfunction. A study showed renin more readily cleaved the oxidized form of angiotensinogen to form angiotensin, which angiotensin converting enzyme (ACE) eventually converts to angiotensin II.[115] Furthermore, the ratio of reduced to oxidized angiotensinogen was lowered in a small group of PE patients. Another effect of placental dysfunction in PE involves decreased levels of catechol-o-methyltransferase (COMT) leading to lower levels of the hypoxia-inducible factor-1α inhibitor protein 2-methoxyoestradiol (2-ME), normally elevated during pregnancy.[116]

Besides the alterations of these various proteins already affected in normal pregnancies, evidence suggests an immunological response associated with PE. Alterations in inflammatory cytokines, such as increased levels of pro-inflammatory TNFα, IL-6, and IL-8 have been observed in women with PE.[117] In addition, the complement system, potently activated by antibody binding, appears dysregulated in PE,[118,119] since levels of several complement proteins are increased in PE.[119,120,121] Mutations in complement regulatory proteins have been associated with an increased risk for PE.[118] While PE placental tissue showed increased presence of complement protein C4d, this was associated with higher levels of mRNA for complement regulatory proteins.[119] Furthermore, placental vessels in a transgenic rat model of PE showed increased C3 deposits, and supernatant from PE placental explants stimulates C3 expression in rat vascular smooth muscle cells.[122] In addition to increased complement, patients with PE exhibit decreased circulating[123] and placental[124] levels of $CD4^{+}CD25^{+}$FoxP3+ regulatory T-cells. These regulatory T-cells help control self-

reactive antibody production; thus, this reduction can lead to increased autoantibody production.[125] Furthermore, PE patients possess expanded levels of circulating $CD19^{+}CD5^{+}$ B-cells,[126] which produce natural antibodies and are similarly increased in autoimmune diseases, including Sjögren's syndrome and rheumatoid arthritis.[127] Taken together, these observations provide evidence for an immunological response consistent with aberrant antibody production in PE.

Several studies implicate autoantibody binding activity in patients with PE. PE-associated antibodies have been shown to bind human proteins, such as β1, β2 and α1 adrenoreceptors,[128] cardiolipin,[129] and prothrombin.[130] Most notably, PE patients produce agonistic autoantibodies against the angiotensin II AT_1 receptor $(AT_1\text{-}AA)^2$ that can induce PE symptoms, such as hypertension, proteinuria, and raised sFlt-1 levels, in pregnant mice.[3] Other studies have further demonstrated these antibodies precede and induce sFlt-1[131,132] and sEng production.[132] It was later shown that the AT_1-AA crosses the placenta and contributes to the IUGR often observed in PE.[133] Through competition experiments, scientists identified a seven amino acid epitope (AFHYESQ) on the second extracellular loop of the AT_1 receptor using antibodies isolated from PE patients[2] as well as a transgenic rat model.[134] This epitope was noted to resemble a section on the parvovirus B19 capsid protein VP2 and a human antibody against this protein demonstrated activity in the bioassay used to determine the presence of AT_1-AA, suggesting molecular mimicry as a source for generating these AT_1-AAs.[135] Importantly, co-injection into pregnant mice of the epitope and the antibodies isolated from PE patients prevented the manifestation of clinical symptoms.[3] Although AT_1-AAs are found as early as 18 weeks' gestation, one

deterrent to utility as a biomarker for PE is the presence of these antibodies in sera of women with abnormal placental perfusion but no PE.[136] Addressing the specificity of AT_1-AAs and resolving whether they are present in all PE cases has been hindered because the low antibody titer or affinity impedes ELISA analysis, forcing dependence on complex bioassays to determine the presence of AT_1-AAs.[137,138] Further investigation of the AT_1-AA's presence in normal pregnancies and PE patients will benefit from a method to increase affinity through avidity effects,[94,95] such as bacteria displaying multiple copies of the epitope. Additionally, attempts to isolate other autoantibody biomarkers that may be more specific than the AT_1-AA are required to advance this hypothesis of an abnormal immune response in PE.

ii. Necessity of a novel diagnostic assay

Current diagnosis of PE relies upon presentation of maternal clinical symptoms, including hypertension and proteinuria.[104] However, 10-15% of pregnant women who progress to hemolysis, elevated liver enzymes, and low platelet count (HELLP) syndrome and 20-25% of those that develop eclampsia never present with hypertension or proteinuria.[139] Various groups have suggested using the altered levels of sFlt-1, PlGF and sEng as predictors for pre-eclampsia. Elevated levels of sFlt-1 can occur five weeks before clinical manifestations, and decreased levels of PlGF can be observed as early as the first trimester of those that develop PE.[140] Several small studies have shown the diagnostic/predictive abilities of these angiogenic factors,[141] and recently diagnostic aids were constructed to detect levels of sFlt-1 and PlGF in patient samples. Studies on these assays demonstrated that the ratio of the proteins may be helpful for diagnosis.[142,143] However the PE sample sizes

in these two studies were small (15 and 48, respectively) and other studies have shown great variability in the sensitivity (62-100%) and specificity (51-85%) of the sFlt-1/PlGF ratio.[105] As can be expected due to the normal increase in sFlt-1 concentration during the third trimester, the assay performs better for diagnosing early-onset PE[105] (before 34 weeks), which only occurs in 10% of PE cases.[1] Additionally, one study evaluating alterations of sFlt-1, PlGF, and sEng in PE patients with high risk pregnancies, such as diabetes and chronic hypertension, demonstrated only modest differences and odds ratios close to or below one.[144]

Ideally, one would be able to predict future PE presentation early on during the first or second trimester. Several studies have evaluated certain clinical risk factors (e.g., family history, obesity, smoking, etc.).[145] However, one recent study evaluated 47 different biomarkers alongside various risk factors for predictive performance at 14-16 weeks' gestation.[146] Most of these markers showed only modest differences between PE and healthy controls in a large cohort (n=5623). Constructing and evaluating diagnostic algorithms revealed that only PlGF and cystatin C protein levels enhanced predictive performance over clinical characteristics alone for early-onset PE. Thus, while prediction may be ideal, this may be difficult to achieve due to disease complexity. However, enhanced understanding of disease-associated molecules is essential for any future predictive diagnostic and potentially eventual therapeutic development.

E. Techniques

i. Flow cytometry and fluorescence activated cell sorting

Through flow cytometry, cells are forced into a single-file line to be interrogated by a laser(s). Following interrogation by a blue laser (488 nm), the side scatter (SSC), which measures refracted or reflected light 90° from the laser line, and forward scatter (FSC), diffracted light, are used to characterize the relative complexity and size, respectively, of the cell population. By gating the appropriate population, the analysis focuses on healthy and singlet cells, instead of cell aggregates. A variety of additional fluorescent parameters can be measured simultaneously using the blue laser or others, such as a red laser (633 nm). Thus, a given cell population can be incubated with multiple fluorescently tagged reagents (e.g., antibodies) and quantitatively assessed for the presence of these tags. Importantly, based on the degree of fluorescence, an individual cell, for our purposes bacteria, can be collected or discarded (**Figure 1-2**) in a process known as fluorescence activated cell sorting (FACS). In this case, a gate is created to eliminate low or non-fluorescent cells based upon the background fluorescence of a negative control. The binding reagents of interest can either be directly conjugated to a tag, or a fluorophore-conjugated secondary reagent may be used to bind specifically to the protein of interest amongst an assortment of irrelevant proteins. Commonly used fluorophores include the Invitrogen Alexa series, such as Alexa488, and R-phycoerythrin. Both Alexa488 and R-phycoerythrin excite at 488 nm; however, their spectra differ, enabling distinct emission detection with 530 nm and 576 nm photomultiplier tubes, respectively.

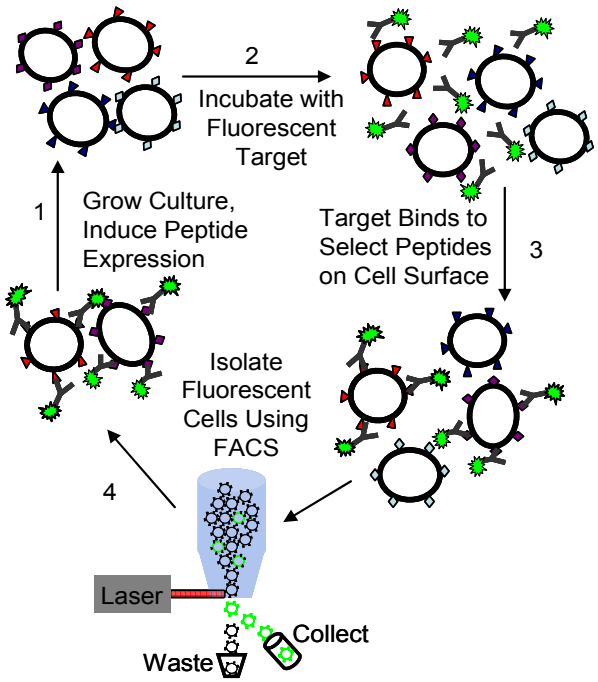

Figure 1-2: Schematic of fluorescence activated cell sorting (FACS) with a bacterial-displayed peptide library.

ii. Bacterial displayed library design

Traditional libraries are often linear, fully randomized amino acid sequences at a user-specified length. Although the length of a typical linear region of a protein that interacts with an antibody, or epitope, is 6-13 amino acids long,[14] a longer peptide of 15 amino acids on the surface may provide structural context with the flanking amino acids. Building upon this idea of conformational significance and taking influence from the highly selective conotoxin proteins in cone snail venom,[147,148] a multiply-constrained (MC) library was constructed with a network of four cysteines, which form disulfide bonds that impart structural constraints. Since protein structure is a significant aspect of binding, the MC library may yield ligands with high affinity and specificity for disease-associated antibodies.

To test this hypothesis, we screened traditional libraries (linear and singly-constrained) and the MC library against a model therapeutic antibody, Herceptin.

Utilizing stringent conditions (low nanomolar concentrations) ensured selection of high affinity binding ligands. The isolated sequences and a previously identified linear mimotope (H98)[149] cloned for bacterial surface expression were compared for apparent affinity (**Table 1-2**). Analysis by flow cytometry determined fluorescent signals associated with peptide binding at varied concentrations of the target antibody ([T]). After subtracting the background (Fl_{bkgd}) from the mean cell fluorescence (MCF), the data was fit to Equation 1.3, derived from Equation 1.2 above. This analysis established the maximum fluorescence (Fl_{max}) and apparent affinity ($K_{D,app}$) for each displayed peptide.

$$(MCF - Fl_{bkgd}) = \frac{(Fl_{max} - Fl_{bkgd})[T]}{K_{D,app} + [T]} \qquad (1.3)$$

Table 1-2: Properties of Herceptin binding peptides

Clone	Sequence	$K_{D, App}$ (nM)	Fl_{max}	Percent Deviation
MC-1	GCCLYGTCDLDSCG	4	5,800	24.2
MC-2	GCHSNCAFSCELDCG	5	18,000	6.3
MC-3	GCCDKNTCDLDHCTCG	24	15,300	12.7
MC-4	GCFQSGCSEGSSGCTRQWCG	27	9,950	13.2
Trad-1	RFP-TQVDTNRICCFVM	17	8,400	14.9
Trad-2	GIFACGQVWSESCGSKE	142	21,000	17.7
H98	LLGPYELWELSH [149]	9	23,500	11.6

To evaluate the peptides' apparent affinities, Fl_{max} must be considered because this relates to the number of peptides on the cell surface and thus, the contribution of avidity effects to enhance apparent affinity. Therefore, a low Fl_{max} and $K_{D,app}$ represents the best overall apparent affinity. MC-2 performed marginally better than the H98 peptide, and MC-1 exhibited the lowest Fl_{max} and $K_{D,app}$ (**Figure 1-3A**), providing preliminary evidence that the MC library yields higher affinity peptides than traditional libraries. Normalizing the fitted curves by the maximum fluorescence

highlighted the similar apparent affinities observed for different peptides, such as MC-1 and MC-2, and the weaker affinity for the Trad-2 peptide (**Figure 1-3B**). Additional off-rate analysis might provide further insights into the binding kinetics of the different peptides.

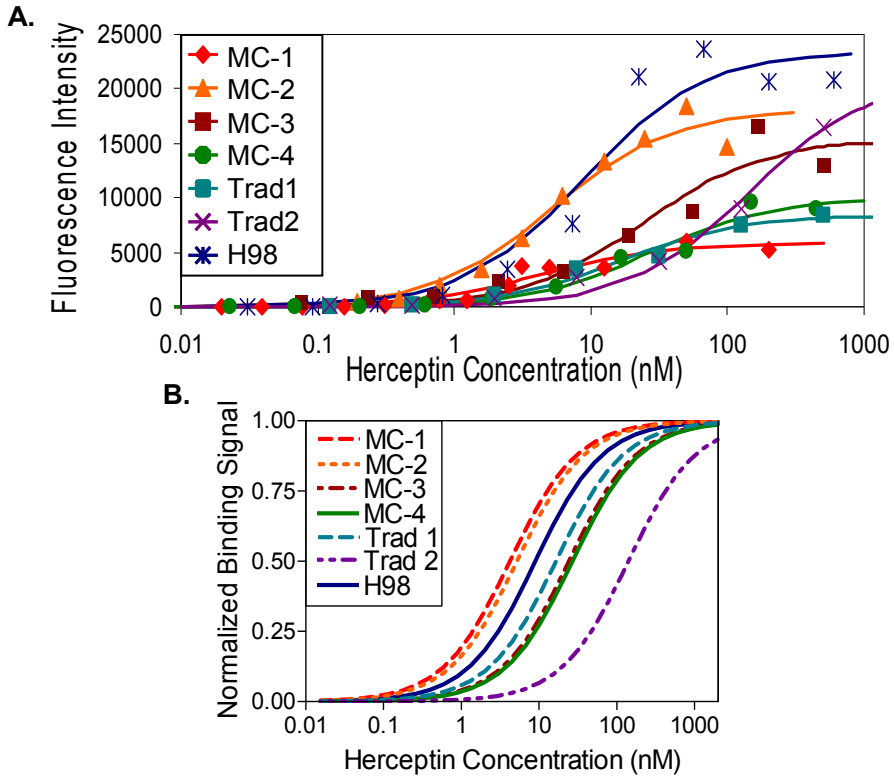

Figure 1-3: Apparent affinity analysis of Herceptin binding clones. (**A**) Experimental data (points) paired with fitted curves shows variability of maximum fluorescence and apparent affinity. (**B**) Normalized fitted curves highlight similarities and differences in apparent affinity.

While identifying high affinity disease-associated antibody binders could benefit diagnostic development, we also sought to identify specific antibody binding peptide epitopes. Library-designed cysteines could confound protein database searches using the generated motifs or possibly result in high affinity mimotopes that show no primary sequence resemblance to the original antigen. For example, although the motif highlighted among three of the MC peptides (**Table 1-2**) resembles one of the three known regions (PEADQ) of interaction between Herceptin and HER2,[150] the

upstream cysteine in the peptide motif is a proline in the native sequence. Thus, although the motif bears resemblance, this dissimilarity might result in a reduced score for a hit in database searches when the native antigen recognition sequences have not been determined. Therefore, we selected a traditional, fully randomized linear library for use in disease-associated antibody specificity profiling.

iii. Sequencing isolated peptides

To identify the sequences corresponding to isolated peptides, the vector encoding the peptides must be sequenced and the region of interest identified. Traditionally this proceeds by Sanger sequencing, which incorporates chain-terminating polymerase inhibitors.[151] These can be attached to a dye and evaluated to effectively piece together the sequence of oligonucleotide residues. Importing the obtained sequence files to analysis software, Geneious, enables easy translation of DNA to the amino acid sequences encoded. By selecting individual colonies of bacteria isolated from screening, each peptide sequence is matched directly to the clone that encodes it. Therefore, following motif analysis, the individual peptides of interest can be easily studied further. However, this process remains relatively low throughput, typically yielding only ~100s of sequences.

In comparison, massively parallelized sequencing processes, next-generation sequencing (NGS), permit high-throughput analysis. Several different instruments now exist with varied data output. For example, the 454 sequencer yields 200,000 reads of 250 bases, while tens of millions of shorter reads (35-50 bases) can be obtained on Illumina's Genome Analyzer.[152] Since these NGS methods return short sequence reads, the region of interest corresponding to the peptide sequence must first

be amplified from a large library pool of vectors obtained from multiple bacterial clones each expressing a unique peptide. Certain adapter regions are often added for the DNA to be sequenced on a specific instrument. Applying a specific barcode to distinct samples enables pooling, thereby increasing throughput. Although this results in a larger depth of sequencing, the specific clones displaying these peptides are not directly matched to the sequence. Thus, sequences of interest must be reconstructed for expression.

iv. Sample handling for diagnostic discovery

Consented individuals provided samples of blood plasma for use in this work. In the case of whole blood samples in EDTA, which prevents coagulation, the blood was centrifuged at 1000g and the plasma fraction isolated and aliquoted (50-200 μL) for storage at -20 or -80°C. In other cases, collaborators provided plasma, which was then aliquoted and stored. To prepare for diagnostic discovery, samples were either enriched for the antibody fractions, using ammonium sulfate precipitation reactions[2] or diluted 1:50. In both cases, samples were depleted of *E. coli* binding antibodies.[91] Using antibody enriched fractions enabled direct conjugation of a labeling reagent, while the diluted plasma samples were labeled using an immunoglobulin G (IgG) specific secondary reagent. While the directly labeled antibody enriched fractions facilitated multi-parameter cell sorting, the diluted plasma more closely replicates the binding conditions in the body and is anticipated to better support future diagnostic development.

v. Peptide reactivity analysis

Bacterial library screening typically reduced library diversity from $8x10^9$ members to ~1000 peptides. From these unique peptides, we must identify the one(s) showing the highest sensitivity and specificity. To characterize peptide reactivity with patient antibodies, an individual bacterial clone is incubated with several different PE and HOP antibody samples followed by fluorescent labeling of these cell-bound antibodies and FACS analysis. This FACS-based analysis recaptures the environment used in screening; however, it lacks throughput. Previous work in the lab identified that high density (10^{11} cells/mL) printing of cells could be probed by a monoclonal antibody.[153] Thus, to increase throughput, we investigated the utility of a bacterial microarray for analyzing antibody fractions from human patients. After printing, the array was heat treated at 65°C for 45 minutes before proceeding to a two hour blocking step with 5% milk and 1.5% bovine serum albumin buffer. A secondary labeling step with an antibody binding reagent, Protein A conjugated to Alexa647, followed incubation with 1 μM of unlabeled antibodies from ten PE and two HOP samples. In comparing the heat map data obtained for the peptides through FACS (**Figure 1-4A**) and microarray analysis (**Figure 1-4B**), the results only weakly correlated qualitatively, and further analysis demonstrated this discrepancy to be true. Therefore, we concluded that while the bacterial microarray would enable higher-throughput analysis, this could not accurately replace FACS based peptide reactivity analysis.

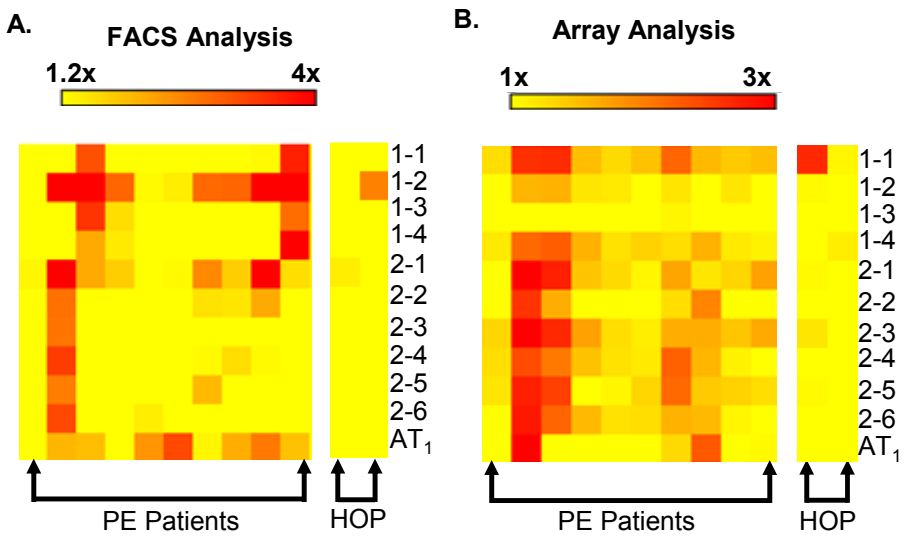

Figure 1-4: Different results observed for FACS and microarray-based peptide reactivity analysis. Peptides were analyzed for antibody binding activity to PE and HOP samples on the (**A**) FACS and using a (**B**) bacterial microarray.

vi. Developing diagnostic assays

A diagnostic assay must achieve high accuracy (sensitivity and specificity) in distinguishing healthy from disease samples. To optimize the performance of any diagnostic assay, varied activity cutoffs are plotted for sensitivity and specificity, in a receiver operating characteristic curve (ROC). The area underneath this curve (AUC) is an additional metric indicating the overall performance of the assay, where 0.5 indicates random chance and 1 represents a perfect diagnostic.

In some cases, one to two antigens or antibody-detecting peptides representing a disease-associated antigen can achieve the sensitivity and specificity to serve as stand alone diagnostic tests. Alternatively, an array of multiple reagents may be required to account for the variance in immune responses among patients that can lead to antibodies with differing epitope specificities, as seen in systemic lupus erythematous, where more than 150 putative autoantigens have been documented,[13] and myasthenia gravis.[154] A phage-displayed peptide microarray strategy has been

developed for diagnosing prostate[77] and breast cancer,[78] achieving 88% and 84% specificity in a validation set, respectively. Configuring the peptide array in a manner that achieves a high surface area to volume ratio, such as with bacterial display, increases sensitivity of detection as demonstrated with polyvalent nanoparticles.[155]

When a diagnostic assay is composed of multiple reagents and properties, a robust algorithm must be developed to combine these properties, optimizing for diagnostic accuracy. In particular, various machine learning algorithms, which modify parameters based upon data provided, have been developed and implemented for constructing accurate classification algorithms for disease diagnosis.[156] These algorithms are constructed, or trained, on a set of observed data from clinical samples. During the training, sampling methods, such as cross-validation and bootstrapping,[157] facilitate accuracy estimation and parameter optimization. Cross-validation divides the training set into different groups, and trains the algorithm on all but one group (or single sample), which is used to test performance, and repeats for all groups. Single-sample testing, known as leave-one-out cross-validation, often overestimates the true accuracy, while another common method using ten approximately equal-sized groups, ten-fold cross-validation, can pessimistically underestimate the true accuracy.[157] The support vector machine (SVM) algorithm, which maps data to a high-dimensional space that enables a linear separation of classes,[158] remains a common choice for classifier development and has been used in a variety of studies.[159,160] Another algorithm known as Adaptive Boosting, or AdaBoost, applies additive regression on reweighted versions of the training set to focus on misclassified samples after each round.[161] AdaBoost has been shown to outperform SVM[162] and is less prone to over-

fit the training data than other algorithms.[163] Additionally, if over-fitting becomes apparent, methods exist to overcome this problem.[163]

vii. Identifying antigenic targets

A consensus region represents the important residues mediating the mimicked antibody-antigen interaction. Therefore, a protein database search for similar/identical sequences to peptide motif(s) using NCBI BLASTp[164] and/or ExPasy ScanProsite[165] identifies candidate antigens. These searches can be limited to the human proteome to identify potential autoantigens, a specific organism known to be involved in disease, or left broad to identify all possible hits. A broad search enables identification of potential trigger antigens, which may have led to cases of molecular mimicry for aberrant autoantibody binding. Conformational epitopes resulting from random peptide libraries complicate the discovery of potential antigenic targets, since the native protein would have long stretches separating the amino acids closely linked together in the mimotope. Since conformational epitopes make up a significant portion (possibly as high as 90%) of antibody specificities,[166] certain algorithms exist to assist with their prediction.[166] Additionally, one algorithm in particular, incorporates peptide sequences generated by library screening for alignment with discontinuous portions of the antigen.[167] However, this only works for known target proteins with characterized structures. Thus, the main focus of this system of identifying antigenic targets through random peptide library screening focuses on linear epitope identification or linear sections composing a larger discontinuous epitope.

2. Characterization of antibody specificities associated with pre-eclampsia

The presence of maternal autoantibodies has been previously associated with pre-eclampsia, although the composition of the antibody repertoire in pre-eclampsia has not been well characterized. Given this, we applied a bacterial display peptide library to identify peptides that preferentially react with plasma antibodies from patients with pre-eclampsia (n=15) versus healthy-outcome pregnancies (n=18). Screening using fluorescence-activated cell sorting identified 38 peptides that preferentially bind to antibodies from individuals with pre-eclampsia. These pre-eclampsia-specific peptides possessed similar motifs of $R^{G}/_{S}{}^{G}/_{-}WW^{G}/_{S}$, $RWW^{G}/_{S}$, or $WGWGXX^{R}/_{K}$ distinct from the angiotensin II type 1 receptor epitope AFHYESQ. Seven library-isolated peptides and a cell surface-displayed angiotensin II type 1 receptor epitope were used to construct a diagnostic algorithm with a training set of 18 new pre-eclamptic and 22 healthy-outcome samples from geographically distinct cohorts. Cross-validation within the training group resulted in averaged areas underneath a receiver operating characteristic curve of 0.78 and 0.72 with and without the known receptor epitope, respectively. In a small validation set (12 pre-eclamptic; 8 healthy), the algorithm consisting only of library-isolated peptides correctly classified 10 pre-eclamptic and 6 healthy, using a predefined cutoff that achieved 61% sensitivity (95% confidence interval, 36-83%) at 95% specificity (95% confidence interval, 77-100%) in training set (n=40) cross-validation. Our results indicate that antibodies with specificities other than anti-angiotensin II type 1 receptor are prevalent in pre-eclampsia patients and may be useful as diagnostic biomarkers.

A. Introduction

Pre-eclampsia (PE) is a serious disorder that affects 5% to 8% of pregnancies[1] and causes 15% to 20% of maternal mortalities and morbidities in developed countries.[101] Despite serious global effect, the primary method of PE diagnosis continues to rely on presentation of maternal symptoms, including hypertension and proteinuria, after 20 weeks of gestation. However, present diagnostic approaches are inadequate to identify patients likely to experience adverse outcomes since 10% to 15% of women who experience hemolysis, elevated liver enzymes, and low platelet levels (HELLP syndrome) and 20% to 25% who progress to eclampsia do not present with either hypertension or proteinuria.[139] Consequently, there remains a need for noninvasive diagnostics that can accurately and reliably identify patients who develop PE and those at risk for adverse outcomes.

Diagnostic development efforts have focused on the identification of protein biomarkers with unique presentation in PE. Multiple proteins exhibit altered serum levels in PE and have been pursued as candidate biomarkers and therapeutic targets, including soluble vascular endothelial growth factor receptor (sFlt-1), placental growth factor (PlGF),[113] soluble endoglin,[111] placental protein 13,[168] and angiotensin II type 1 receptor (AT_1) autoantibodies (AT_1-AAs).[2] In particular, much effort has focused on evaluating the diagnostic utility of the ratio of the elevated sFlt-1 to lowered PlGF levels or the PlGF level alone.[169] These biomarkers have yielded high diagnostic accuracy for detecting early-onset PE and predicting adverse outcomes.[170] However, these biomarkers are less effective after 34 weeks' gestation,[169] the period where 90% of PE cases present.[1] Finally, they do not enable accurate prediction of

PE during the first trimester,[171] since sFlt-1 levels only become significantly altered ≈5 weeks before PE onset.[112] Therefore, despite the use of these biomarkers for early-onset PE detection, additional biomarkers are needed.

Circulating antibodies represent a rich source for additional biomarker discovery, and several observations link the immune system to PE pathogenesis. Most prominently, patients with PE have been found to produce agonistic immunoglobulin (Ig) AT_1-AAs[2] as early as 18 weeks' gestation.[136] Several *in vivo* and *in vitro* studies have demonstrated a potential pathological role for these antibodies. Injection of AT_1-AAs or total IgG isolated from PE patients into pregnant mice induced the hallmark PE symptoms (hypertension, proteinuria, and increased sFlt-1[3] and fetal growth restriction[133]), whereas coinjection with an antibody-blocking peptide epitope attenuated these effects. Interestingly, placental ischemia-stimulated AT_1-AAs similarly contribute to hypertension in an independent PE rat model.[172] AT_1-AAs increase complement protein C3 deposition in the placenta and kidney of pregnant mice,[173] while mutations within complement system regulatory proteins seem to be a risk factor for PE.[118] Complement activation has been further implicated in PE with increased C3 deposits in placental vessels from a transgenic PE rat model, and supernatant from PE placental explants stimulate C3 expression in rat vascular smooth muscle cells.[122] Furthermore, isolated $CD19^+CD5^+$ B-cells are elevated in PE and produce AT_1-AAs in culture upon addition of PE serum.[126] At the same time, individuals with PE exhibit significantly reduced levels of $CD4^+CD25^+$ regulatory T-cells,[123] a finding consistent with increased autoantibody production.[125] Collectively,

these previous studies indicate that immunological alterations are a conserved feature in PE.

Despite a demonstrated role of AT_1-AAs in the pathology of PE, their efficacy for PE diagnosis has not been established. Existing assays for AT_1-AAs that rely on cardiomyocyte beat rate[2] or a luciferase reporter[138] lack throughput and are unsuitable for point-of-care diagnostics. More importantly, AT_1-AA prevalence varies significantly in different studies (70%[135]-95%[138]), and AT_1-AAs are not specific to PE because they have been observed in individuals with healthy-outcome pregnancy (HOP),[135] chronic hypertension,[21] and renal allograft rejection.[174] Given these problems, we investigated whether additional PE-specific antibodies exist that could serve as biomarker(s) for PE diagnosis and further implicate a pathophysiological role for an altered immune system. To simultaneously identify antibody biomarkers and peptide reagents for their detection, we screened a bacterial display peptide library[91] against antibodies enriched from the plasma of individuals with PE and HOPs. Our results demonstrate the existence of PE-specific plasma antibodies, other than AT_1-AAs, that may be useful for PE diagnosis.

B. Materials and Methods

i. Patient Samples

Whole blood samples were obtained from pregnant women as aliquots of samples taken for routine blood work during clinical assessments at the Santa Barbara Cottage Hospital (cohort 1). The study was approved by the Santa Barbara Cottage Hospital review board. To qualify as affected with PE, subjects fulfilled ≥2 of the following criteria: (1) 2 documented blood pressures (BPs) with readings >140/90mm Hg ≥4

hours apart, with documented normal BPs in the first half of the pregnancy; (2) proteinuria as defined by ≥30 mg/dL on a spot urine check, ≥1+ dipstick reading, or ≥300 mg/24 hr; (3) central nervous system (CNS) symptoms (visual disturbances or unremitting headaches); (4) epigastric pain associated with elevated liver enzymes unrelated to other abdominal pathology; or (5) thrombocytopenia with platelet counts <100000 U/mL. This ensured that PE samples met the American College for Obstetricians and Gynecologists criteria for mild or severe PE diagnosis. Pre-existing hypertension and lupus patients were excluded from cohort 1. Samples were divided into a discovery set (n=33) for initial peptide identification and a training set (n=20) for testing diagnostic ability of isolated peptides. Additional deidentified samples provided from University of Texas Medical School at Houston (cohort 2) were used in either the training set (n=20) or a validation set (n=20). These PE samples were diagnosed by clinical assessments based on the National High Blood Pressure Education Program Working Group Report.

This study did not distinguish between early- and late-onset PE and did not discriminate based on parity. Therefore, these cohorts represent a mix of presentation times and parities. All subjects provided informed consent, and samples were collected according to institutional guidelines. Blood samples for both cohorts were obtained near the time of delivery. In cohort 1, BPs were recorded at the time of presentation whereas cohort 2 recorded maximum BP before delivery. In addition, while cohort 1 mainly used the spot urine check, cohort 2 diagnosis used 24-hour analysis and the dipstick test (n=7).

ii. Bacterial display and library screening

The AT_1 epitope AFHYESQ was displayed on *Escherichia coli* MC1061[175] with flanking glycines as a fusion to the N-terminus of the eCPX (enhanced circularly permuted OmpX) scaffold[90] along with a C-terminal peptide tag (P2x) that binds a fluorescent reporter (YPet-Mona) of scaffold expression.[176] A 15-mer random peptide library displayed on the N terminus of the eCPX scaffold was screened for peptides binding to PE-specific antibodies. All cultures for screening and analysis were subcultured (1:50), grown to an OD_{600} of 0.4-0.6, and induced for one hour with 0.04% arabinose. Library screening used antibody fractions in PBS (0.1% BSA) prepared by ammonium sulfate precipitation of patient plasma and depleted of *E. coli* binding antibodies. Magnetic selection enriched for peptides that bind pooled PE (n=9) antibodies (5 μmol/L total concentration) labeled with the FluoReporter Mini-Biotin-XX Protein Labeling Kit (Invitrogen) while outcompeting an unlabeled pool of HOP (n=12) antibodies (5 μmol/L total). Streptavidin (SA) binding peptides were removed from the library using MyOne SA-coated magnetic beads. To favor cross-reactivity, 2 pools of PE antibodies with distinct fluorophores were prepared: group 1 (n=4) labeled with Alexa Fluor 488 (Invitrogen; green) and group 2 (n=5) biotinylated (red) to enable detection with streptavidin-conjugated R-phycoerythrin (Invitrogen). Cells were coincubated with excess unlabeled HOP antibodies and labeled PE antibodies, and those cells exhibiting both red and green fluorescence were recovered by fluorescence-activated cell sorting (FACS) (**Figure 2-1A**). The first FACS round utilized 2.5 μmol/L from each PE group and five-fold excess HOP pool, while the subsequent two rounds lowered the PE concentration to 1 μmol/L and

increased normal antibodies to 10 and 15-fold excess, respectively. Peptides with specificity for PE antibodies were favored in separate sorts using a PE antibody pool (n=9) labeled with Alexa Fluor 488 and a biotinylated HOP antibody (n=12) pool. After recovering nonfluorescent cells that did not capture HOP antibodies (1 μmol/L), cells exhibiting green, not red, fluorescence were collected after labeling with PE and HOP antibody pools (1 μmol/L, each) followed by streptavidin-conjugated R-phycoerythrin (**Figure 2-1B**). Screening continued with a set of new PE and HOP (n=6 each) samples to enhance PE cross-reactivity and specificity. After completing rounds six and seven for cross-reactivity sorting with 1 μmol/L of each disease group and unlabeled 15-fold excess and 20-fold excess of HOP, respectively, two specificity rounds were performed to select peptides binding only disease samples as described above. Another negative sorting round enriched bacteria displaying peptides that did not bind the biotinylated HOP antibody pool. Finally, three additional rounds of peptide enrichment were performed towards binding pooled PE (1 μmol/L) but not HOP (2.5 or 1 μmol/L) antibodies, using the HOP concentration that showed the highest binding activity.

iii. Peptide sequence analysis and down-selection

Plasmid DNA from ≈200 bacterial colonies was sequenced from the final sorting round, from which 83 unique sequences were identified using Geneious. Three additional peptides that demonstrated PE reactivity and specificity in an earlier screening round were incorporated into the motifs identified by inspection (**Figure 2-1C**). The binding activity, or fold fluorescence over a negative control (empty scaffold), of each library peptide and the AT_1 epitope clone was measured in

duplicate with discovery set PE and HOP (n=6 each) pools. After incubation with each antibody pool (1 µmol/L), cells were labeled using biotinylated anti-human IgA + IgG + IgM (Jackson ImmunoResearch), followed by streptavidin-conjugated R-phycoerythrin. On average, peptides exhibited 1.4-fold increased PE antibody-binding activity compared with HOP. Therefore peptides demonstrating 50% higher PE reactivity over HOP (1.5-fold) were selected as the most specific and ranked according to the PE activity quotient defined as the PE-binding activity multiplied by the ratio of PE activity to HOP (ie, dynamic range).

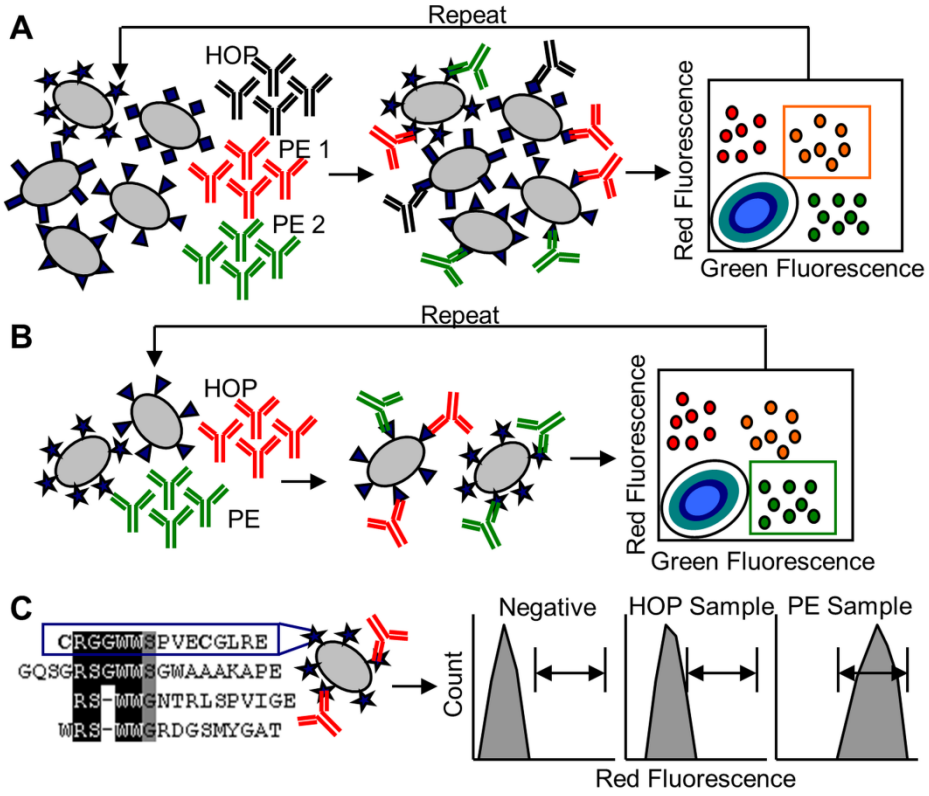

Figure 2-1: A 2-color screening methodology isolated a pool of pre-eclampsia (PE)-specific antibody-detecting peptides, enabling further characterization of individual peptides. **A,** After coincubating a bacterial displayed peptide library with 2 distinctly labeled pools of PE (red or green fluorophore) and unlabeled pool of healthy-outcome pregnancy (HOP) antibodies enriched from plasma, cells expressing peptides that bind antibodies present in both groups with PE were isolated. **B,** Subsequently, the library was coincubated with a labeled HOP antibody pool (red) and a PE pool (green). Bacteria exhibiting only green fluorescence were isolated, ensuring disease specificity. **C,** Finally, sequence analysis of the enriched pool identified unique peptides for motif characterization and individual peptide PE- and HOP-binding activities (fluorescent intensity) were assessed.

iv. Patient antibody reactivity assays

The AT_1 epitope and seven down-selected peptides were assayed in duplicate against biotinylated antibodies (1 μmol/L) enriched from individual samples of a validation set. This set included 10 PE and 10 HOP from cohort 1 and 20 PE and 20 HOP from cohort 2. The mean cell fluorescence associated with each clone was divided by that of a negative control for each sample. Although the class (PE or HOP) was known for cohort 1, cohort 2 samples' classes were revealed only after testing peptide reactivity. To account for the varied peptide activity range, each clone's binding activity was standardized by the average reactivity and standard deviation.

v. Testing for mimicry of the known AT_1 epitope

To determine whether library-isolated peptides were mimicking the known AT_1 epitope, *E. coli* displaying the AFHYESQ epitope were used to deplete binding antibodies, and reactivity with the indicated library-isolated clones was measured. The PE antibody pool (1 μmol/L) used for down-selection was incubated with AT_1 epitope expressing cells ($1x10^8$ cells/μL). Cells were centrifuged, and the AT_1-AA depleted supernatant was retained and assayed against library-isolated peptides and the AT_1 epitope clone.

vi. Identifying candidate antigens

The motifs identified by inspection were used to perform protein database searches using NCBI blastp and ScanProsite of the human proteome. After applying

a similarity scoring metric, the top 80% of extracellular protein hits from each search were compiled into a list of candidate antigens with the corresponding fragment.

vii. Statistical Analysis

A subset (20 cohort 1 and 20 randomly selected cohort 2 samples) of the 60 samples not used for discovery was used to train a classification algorithm followed by testing with a validation set (20 remaining cohort 2 samples) to verify peptide panel diagnostic accuracy. To reduce overfitting to the training set, an Adaptive Boosting (AdaBoost) classification algorithm, which applies the algorithm successively to reweighted versions of the training set,[161] was implemented through the ada package in R[177] using a 4-split tree. Diagnostic algorithms were generated with and without the AT_1 epitope included in the peptide panel. Three trials of ten-fold cross-validation across a combined set from both cohorts (n=40) yielded each sample's averaged probability for PE classification, which was used to perform receiver operating characteristic (ROC) curve analysis with the pROC R package and assess algorithm accuracy. Additional algorithms were trained using cohort 1 or 2 separately with or without the AT_1 epitope. Subsequently, the performance of the 6 algorithms was assessed by generating ROC curves and determining the PE, HOP, and overall classification accuracy at a 0.5 probability cutoff in the final validation set.

Separately, ROC analysis was performed with Prism 4 software (GraphPad Software Inc.) for each library-isolated peptide across the entire set of 60 validation samples based upon binding activity. ROC analysis of the AT_1 epitope was performed on the same 60 samples for comparison to library-isolated peptides

followed by analysis across the entire set of 45 PE and 48 HOP samples described in this study. Excluding Figures 3 and 4, data are presented as mean±SEM. Statistical significance using a Student *t*-test or Mann-Whitney *U*-test and Spearman correlation were assessed with Prism 4 software. For all analyses, $P<0.05$ was considered significant.

C. Results

i. Identification of peptides binding to plasma antibodies from PE patients

The 45 PE samples used in this study reflect the heterogeneity of PE presentation including early- and late-onset PE cases and atypical severe cases without proteinuria (**Table 2-1**). To identify PE-specific antibody-detecting peptide reagents within this diverse group, a bacterial display library was screened for peptides that bind antibodies present in multiple PE patients but not HOP subjects. This screening strategy used fluorescence-activated cell sorting to quantitatively measure the enrichment of PE antibody-binding peptides from 7% to 87% (**Figure 2-2 A,B**) of the bacterial cell population and reduction of HOP antibody-reactive peptides (**Figure 2-2 C,D**). From this enriched pool, DNA sequencing identified 83 unique peptides, enabling further characterization of each member alongside 3 peptides from an earlier screening round. None of the library-isolated peptides were similar to the known AT_1 receptor epitope AFHYESQ, but several different motifs were identified (**Table 2-2**). Three motifs (a-1 to a-3) were similar, and 4 additional motifs were distinct.

Table 2-1: Clinical characteristics of patients

	Discovery Samples (Cohort 1)		Training Samples (Cohort 1)		Training Samples (Cohort 2)		Validation Samples (Cohort 2)	
Patient Characteristics	HOP (n=18)	PE (n=15)	HOP (n=10)	PE (n=10)	HOP (n=12)	PE (n=8)	HOP (n=8)	PE (n=12)
Age, yr	29.8 (1.6)	29.9 (2.0)	27.9 (2.7)	30.1 (2.6)	26.6 (1.7)	25.1 (2.1)	25.5 (2.2)	26.5 (1.3)
GAD, wk	38.0 (0.7)	35.3 (0.8)	38.9 (0.4)	36.8 (0.9)	39.1 (0.5)	34.8 (1.2)	39.6 (0.3)	35.7 (0.6)
Highest SBP, mm Hg	112.0 (3.3)	164.7* (5.0)	114.4 (3.3)	161.1* (6.8)	113.1 (3.4)	156.3* (6.1)	124.1 (2.7)	152.3* (5.4)
Highest DBP, mm Hg	65.4 (2.2)	100.2* (2.2)	64.4 (1.9)	101.7* (1.7)	66.5 (2.8)	94.3* (4.2)	73.9 (2.0)	89.8† (3.9)
Proteinuria, n (%)	3 (17%)	13 (87%)‡	0 (0%)	8 (80%)‡	0 (0%)	6 (75%)	1 (13%)	7 (58%)
AST > 70 IU/mL, n (%)	ND	6 (40%)	ND	2 (20%)	ND	0	ND	0
ALT > 70 IU/mL, n (%)	ND	6 (40%)	ND	1 (10%)	ND	0	ND	1 (8%)
Platelets < 10^5 U/mL, n (%)	0	4 (27%)	0	2 (20%)	1 (8%)	0	0	1 (8%)
CNS Symptoms, n (%)	1 (6%)	7 (47%)	0	4 (40%)	1 (8%)	4 (50%)	0	4 (33%)
Epigastric Pain, n (%)	0	3 (20%)	0	2 (20%)	0	1 (13%)	0	0
Birth weight, kg	ND	ND	ND	ND	3.2 (0.16)	2.5 (0.32)	3.5 (0.15)	2.9* (0.22)
Pre-existing Conditions, n (%)								
Diabetes T1	ND	ND	ND	ND	0	2 (25%)	0	0
Diabetes T2	ND	ND	ND	ND	0	0	0	1 (8%)
Lupus	0	0	0	0	0	0	0	0
Hypertension	0	0	0	0	0	0	0	2 (17%)
Ethnicity, n (%)								
White	ND	ND	ND	ND	4 (33%)	2 (25%)	4 (50%)	4 (33%)
Hispanic	ND	ND	ND	ND	0	1 (13%)	2 (25%)	1 (8%)
Black	ND	ND	ND	ND	8 (67%)	5 (63%)	2 (25%)	7 (58%)

Data are given as the mean (SEM) unless otherwise indicated. PE, pre-eclampsia; HOP, healthy-outcome pregnancy; GAD, gestational age at delivery; SBP, systolic blood pressure; DBP, diastolic blood pressure; AST, aspartate aminotransferase; ALT, alanine aminotransferase; CNS, central nervous system.

* and † indicate a significant difference (P<0.01 and P<0.05, respectively) compared to HOP using a Mann-Whitney U-test; ‡PE subjects without proteinuria met the American College for Obstetricians and Gynecologists criteria for severe PE.

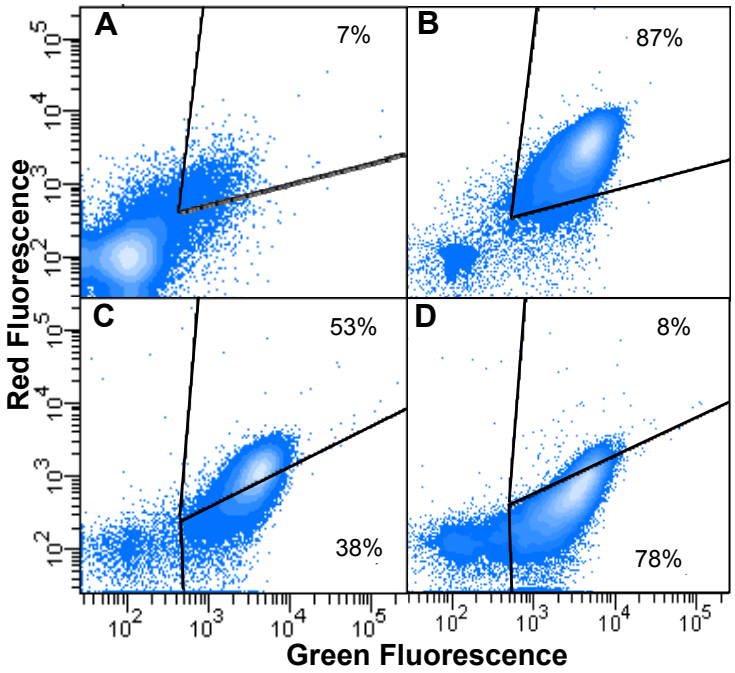

Figure 2-2: Library screening resulted in a PE cross-reactive and specific peptide population. Cross-reactivity sorts enriched for peptides that simultaneously bind antibodies present in two PE groups separately labeled with a red or green fluorophore while competing with an unlabeled HOP pool. **A,** Prior to FACS, only 7% of the peptide library population showed cross-reactive binding; however, **B,** after nine rounds of sorting, 87% bound antibodies present in two PE groups. Further sorting rounds focused on enhancing PE specificity using a red labeled HOP pool and green PE pool. **C,** While initially 53% of the PE cross-reactive population also bound antibodies in the HOP pool, **D,** the final library population demonstrated high specificity with 78% binding only to the PE antibody pool and 8% also binding the HOP antibody pool. Gates were set according to single color and negative control populations.

The library peptide consensus sequence motifs were not sufficiently strong to enable identification of antigens that elicited the antibodies. Roughly 48 candidate autoantigens were identified by rank-ordering proteins with high similarity scores obtained using blastp and ScanProsite searches (**Table 2-3**); however, a much larger number of non-self proteins in the entire database also carried these motifs.

Table 2-2: Peptide binding motifs

Motif	Sequence	Name
a-1	CRGGWWSPVECGLRE	33
	GQSGRSGWWSGWAAAKAPE	
	RS-WWGNTRLSPVIGE	22
	WRS-WWGRDGSMYGAT	
	R-GWWGVLHGQQAEHI	34
	MGR-GWWYGPGRVTND	20
	VNRM-WWAQGRLDGQK	15
	GRN-WWYPSTRVQGDG	
	GGWWYWANG-RAKPQE	
	ERVESKRCMWWCERS	27
	TDRWR-PWWMGHQRQE	35
	TTSGQGWWSQYGMKS	
	LRGYWFGNGAKSTPV	
	GRWGYYWLGQGGHAG	13
	SRWGWWSLKDLEAPP	
	YKQ-WWGSTGSWREQR	11
	VSRNVGWWGTRAPVS	5
	WWSYRGGVPAVPGLP	28
	RHWLSQWQTDRGGNF	
	TRS-WWTWRERPAMTW	
	GRQ-WWTATRPAPVLK	
	RGGCIGCTWMTEWGG	38
	STSGQFGGRTNWWNL	36
	FFWWNPMQSVSGRGS	37

Consensus: $R^{G}/_{S}{}^{G}/_{-}WW^{G}/_{S}$

Motif	Sequence	Name
a-2	ARWWGGRMDDTRMDVR	25
	VAKWWGRPENNYDRY	
	ARWWHGRDAHVGIGV	
	GEMARWWHTGMGVSR	
	MGKWWTRPNDGGRGL	21
	SPKWWVRPGQLVTNK	
	VPVRWWGNTAWRTPA	2
	WRWWGQKYDSNVTGG	
	YRWWGTQVPELSVTR	
	YRWWGAGTSGQEREATR	
	SPQWRWWERGEGRAC	
	WRWW-RPGPDNGVLPV	19
	RWWGYGSTAHQESIN	
	RWWGWGRTEAVDENT	
	GGWSGYE-RWWSRTDQ	31
	N-RWWFQRGDSMIPDM	
	RWYSWGQGMHSRVGA	6

Consensus: $RWW^{G}/_{S}$

Motif	Sequence	Name
a-3	KWWWGRAPQDLQQV	
	RWWWPARAPVE	16
	RWWWGANKDVGVAEK	
	RWWWGQERPALVLQS	
	ERWWWSHTGRTEGNA	
	GRWFWPKDPKLGNHP	
	HFWWTPRPTPPGSAK	

Consensus: $RWWW^{G}/_{P}$

Motif	Sequence	Name
b	NGRWLGWGMSRGSEG	12
	WGLRWLEPRAHFEGM	3
	RWLSWGSSANALGNK	
	RWLGQSWRTGPVMDG	
	NKWLGYGSQRGTIAA	
	NRWVSWAYPHSQMPS	
	RHWLSQWQTDRGGNF	17

Consensus: $RWL^{G}/_{S}$

Motif	Sequence	Name
c	MWGWMQ-KPTAFRVSE	10
	WGWGSWRPTRETVTL	
	WGWGVWRPDSKTGER	
	EMRSWGWGSGWHGNK	
	MRSWGWMSERETLAR	
	AR-WGWGHQQNAIGPR	
	YGYGTRKWPYWCYWG	14
	K-WIWGGGRSTGSTEK	
	K-WCWG--RPLNSRAEQG	4
	SWLYGWGPSRGQTEV	30
	TSKIWTWTG-KTSEHK	1

Consensus: $WGWGXX^{R}/_{K}$

Motif	Sequence	Name
d	WWAWRKAPQNDESG	26
	SWPGWKGGASRLAAE	23
	LGHWWGVRLPEVESA	29
	QGWWGWRNSGAGAVA	
	GCW-WRVCSGVAQEDG	18
	YW-WKPAGRDSGTVEV	
	GGYNS-W-WKSAPLISQ	
	LGHYSYW-WRVPAVREGGSGDS	
	SMRIWKWSTTSGDIS	
	HMGSW-WRQPAINGSV	

Consensus: $WW^{G}/_{-}W^{R}/_{K}$

Motif	Sequence	Name
e	MRYWGSWSAQKREPS	
	WGYWSG-YKGGSASEK	
	WFVWS---KPESRAGCGG	
	R-WGAWY--WRPTSALDK	9
	WGAWASGWRAQGVVG	
	RFWGNWS---RPAEVV	7
	WGWWS-G-RDMGNVAGR	8
	YGWWGKP-RGGWDAGE	32
	GGYGRNWFGGYPYLR	
	WGAWWAHSAPRNDQG	
	WGSLWG--RPMVASETQ	24

Consensus: $WGXW^{S}/_{-}XXX^{R}/_{K}$

The respective motif name and consensus sequence is shown. Names indicated correspond to peptides with 1.5-fold higher PE activity than HOP

Table 2-3: Candidate autoantigens mimicked by library peptides

Human Protein	Sequence Similarity
Motif a-1 and a-2: $R^{G}/_{S}{}^{G}$/-$WW^{G}/_{S}$ and $RWW^{G}/_{S}$	
Mucin-3A	GVRAV**RSGWWG**GQRRGR
Angiopoietin-related protein 6	CALYQ**RGGWWY**HACAH
Angiopoietin-related protein 1	CAHFH**KGGWWY**NACAH
Angiopoietin-related protein 2	AHYQ**KGGWWY**NACAH
Carboxypeptidase Z	SADGS**K**PW**WWS**YFTSL
Atrial natriuretic peptide receptor 2	YSGAE**K**QI**WWT**GRPIP
Apolipoprotein L1	TMDYG**KK**W**WT**QAQAH
Scavenger receptor class F, member 1	GRCAC**RPGWWG**PECQQ
Zona pellucida glycoprotein 2	MACRQ**RGG**S**WS**PSGWF
Skeletal muscle chloride channel protein 1 (several isoforms)	RSQQ**RGG**EQS**WWG**SDPQY
Angiopoietin-like 3	CPEGYS**GGWWW**HDECG
Tubulointerstitial nephritis antigen-like 1	QGC**RGG**RLDGA**WW**FLRR
cadherin, EGF LAG seven-pass G-type receptor 3	DQQCP**RGWW**GSPTCG
Voltage-dependent L-type calcium channel subunit alpha-1S	NGSEC**RGGW**P**G**PNHGI
Claudin 16	VSTKC**RG**L**WW**ECVTNA
Solute carrier family 22, member 3	APLVPC**RGGW**RYAQA
G protein-coupled receptor 113	DHSLFQG**RGGW**SKEGCQ
Proline-rich protein 24	**RGG**GGA**WWG**RGLCGLR
Immunoglobulin heavy chain variable region	**RGW**SSINGGPVECG
Glutamate [NMDA] receptor subunit epsilon-4	HDGLD**GGWW**APPPPP
Mucin 3B, cell surface associated	GVRAV**RS**E**WW**G GQRRG**RSW**DQDRK
Cadherin, EGF LAG seven-pass G-type receptor 1	**GWW**GNPVCG
SH3 and PX domain-containing protein 2A	GISF**RGG**QKAEVIDKNS**GGWW**YV
T-cell receptor beta chain VJ region	SGQA**RGSW**CSVSAGCWG
MRC2 mannose receptor, C type 2	LQSYEGQS**RG**A**W**LGMNFN
A disintegrin and metalloproteinase with thrombospondin motifs 7	QPCPA**RWWA**GEWQ
Extracellular calcium-sensing receptor	SFH**R**K**WWG**LNLQ
Proto-oncogene Wnt-1	AANSS**GRWWG**IVNVA
Poliovirus receptor-related protein 1	AGAA**GRWWG**LALGL
Bestrophin-2	TLVVN**RWWS**QYLCM
G protein-coupled receptor 139	**WWS**PGSACGL
A disintegrin and metalloproteinase with thrombospondin motifs 17	WSP**WG**A**WS**MCS**R**TCG
ADAMTS-like protein 1	WDA**WG**P**WS**ECS**R**TCG
ADAMTS-like protein 3	WDA**WG**D**WS**DCS**R**TCG

Table 2-3: Continued

Human Protein	Sequence Similarity
Motif c,d, e: WGWGXX$^{R}/_{K}$,WW$^{G}/_{-}$W$^{R}/_{K}$, and WGXW$^{S}/_{-}$XXX$^{R}/_{K}$	
Properdin	WDS**WG**E**WS**PCI**R**RNM
	PPCPVA**GGWGPWGP**VS
SCO-spondin	WGP**WG**P**WS**HCS**R**SCG
Thrombospondin type-1 domain-containing protein 4	GVWGA**WG**P**WS**ACS**R**SCSG
Anion exchange protein 3 (Gene SLC4A3)	APPHA**WG**R**WS**PGE**K**PEAA
Semaphorin-5B	AS**WGSWS**KCSS
	A**WGPWS**SCS**R**DCE
Brain-specific angiogenesis inhibitor 2	WEE**WG**S**WS**LCS**R**SCV
Brain-specific angiogenesis inhibitor 3	NQ**WG**H**WS**GCS**K**SCD
Catechol-O-methyl transferase	LLRH**WGWG**LCLIGWNE
placenta apolipoprotein B48 receptor type 2	SAVEQT**WGWG**DGSSHGS
Potassium voltage-gated channel isoform 1	AERKR**WGWG**RLPGAR
Scavenger receptor cysteine rich domain containing, group B	PQLDEKR**WGW**RLGDGSAA
Anaplastic lymphoma receptor tyrosine kinase	CPQAMKK**WGW**ETRGGFG
RCC1 domain containing 1	RTGELYT**WGWG**KYGQLGH
von Willebrand factor C domain-containing protein 2-like	GD**WWK**PAQCSKRE

In total, 38 of 86 novel peptides and the AT_1 epitope exhibited an average 1.5-fold increased PE-binding activity compared with HOP, and these peptides were ranked according to their PE-specific antibody-binding activity (ie, activity quotient; **Figure 2-3**). The most PE-specific and reactive peptides exhibited an activity quotient of ≥6, and the most highly represented motifs among these peptides were determined (**Figure 2-4**). The greatest fraction (10/22) of peptides with a high PE activity quotient represented the a-1 motif ($R^{G}/_{S}{}^{G}/_{-}WW^{G}/_{S}$), which was also comprised of the greatest number of unique peptides. Individual peptides from these motifs did not exhibit reduced PE antibody binding after depletion with the AT_1 epitope AFHYESQ (**Figure 2-5**), indicating that library-isolated peptides did not mimic the AT_1 epitope.

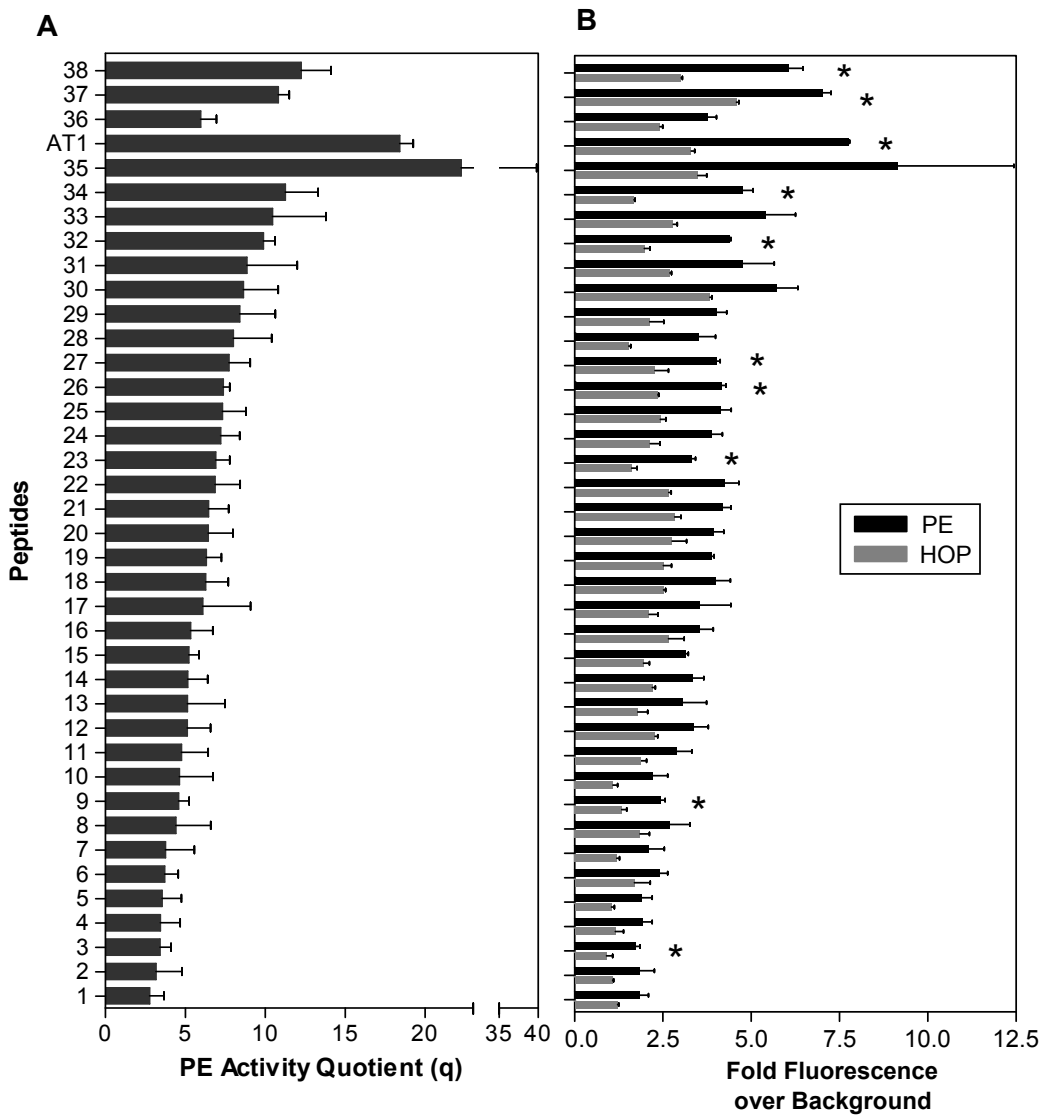

Figure 2-3: The AT_1 epitope and 38 library-isolated peptides showed PE reactivity and specificity. **A,** Activity quotient, q=(PE Activity)^2/HOP Activity, is shown for all peptides exhibiting 1.5-fold increased PE binding activity over HOP with **B,** the corresponding PE and HOP binding activity. Peptides from an earlier screening round separated above the AT_1 epitope. * $p < 0.05$ one-tailed *t*-test

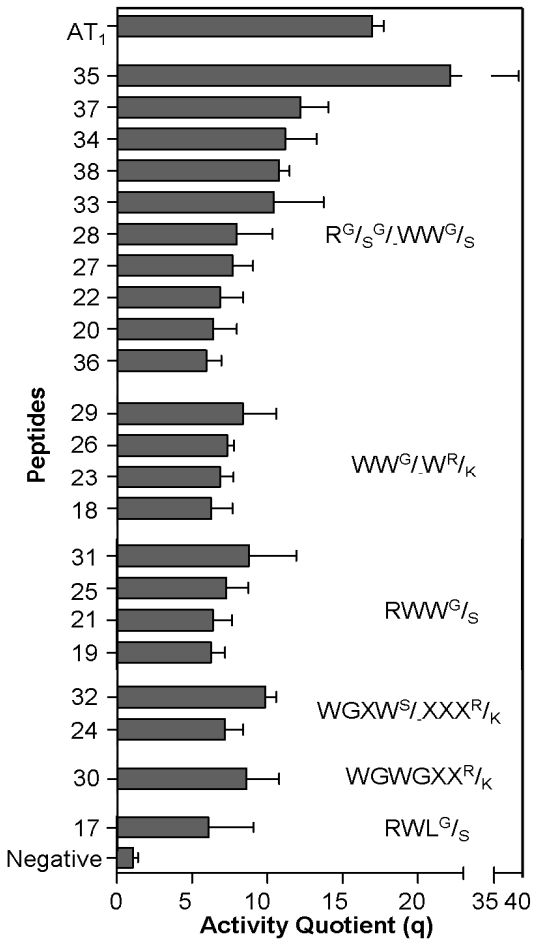

Figure 2-4: Consensus library peptides recognized by pre-eclampsia (PE)-specific antibodies. Twenty-two peptides exhibited a high-reactivity quotient (q>6) and were grouped according to consensus families, indicated on the right. q=(PE Activity)^2/HOP Activity

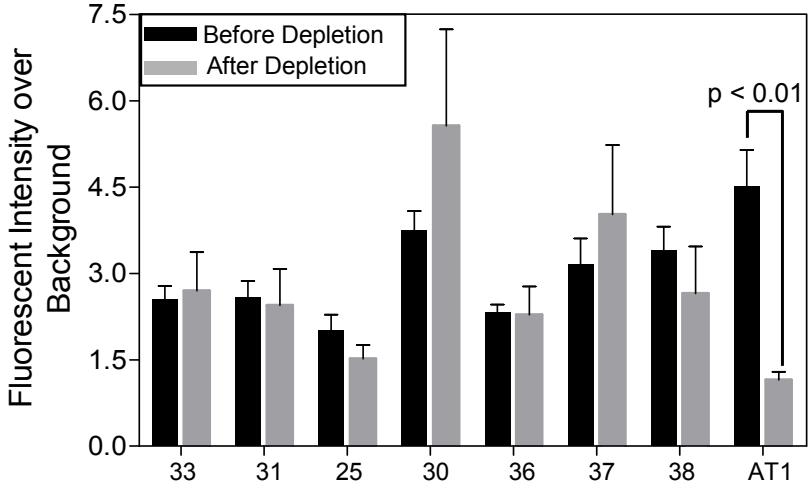

Figure 2-5: Comparison of antibody reactivity before and after AT_1-AA depletion indicates distinct specificities. The reactivity of each of the library-isolated peptides analyzed was not significantly decreased by AT_1-AA depletion. However, the significant loss in binding activity to the AT_1 epitope confirmed appropriate depletion of AT_1-AAs from the sample (n=6).

ii. Peptides demonstrate diagnostic ability on a validation set

To assess the diagnostic efficacy of these PE-specific peptides, seven library-isolated peptides with PE activity quotients >6 and the AT_1 epitope were tested for reactivity against 30 new PE and 30 HOP patients. Together, the panel of library-isolated peptides performed well, achieving 100% accuracy within the cohort 1 validation set (**Figure 2-6**), whereas the AT_1 epitope alone accurately classified 6 of 10 PE and 9 of 10 HOP in cohort 1.

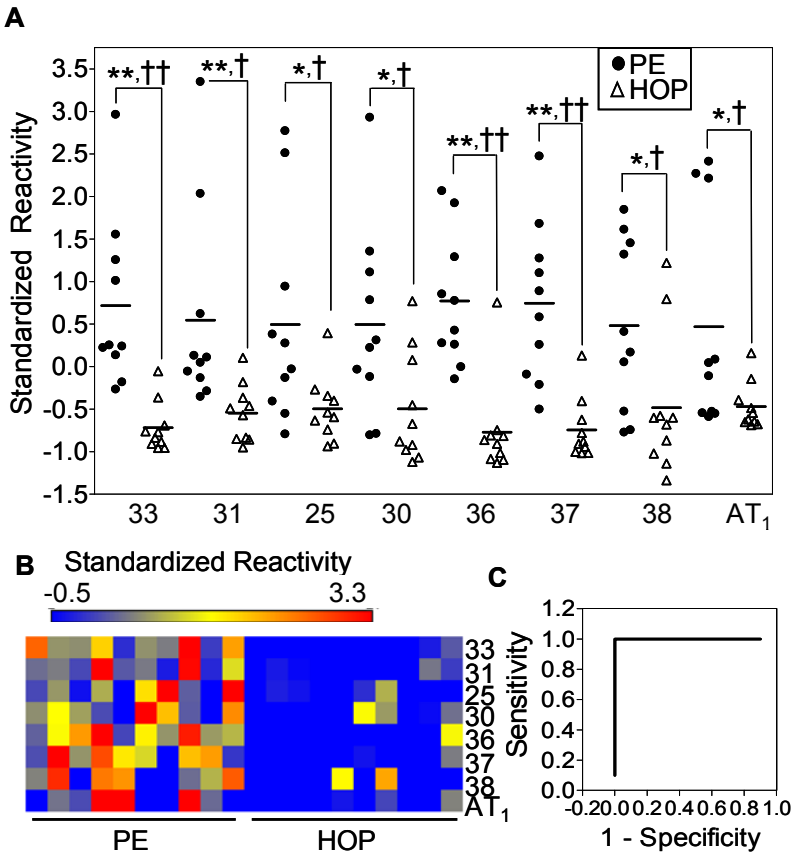

Figure 2-6: The library-isolated peptides demonstrated strong classification accuracy in a new set of PE and HOP cohort 1 samples. **A,** All peptides, including the AT_1 epitope demonstrated significantly higher PE binding activity than the controls using the Mann-Whitney U-test (*) or Student's *t*-test (†). **B,** The heat map shows that these peptides are highly cross-reactive with PE patients while remaining specific. **C,** Summing the standardized activity of the library peptides enables 100% diagnostic accuracy. * or † $p<0.05$; ** or †† $p<0.001$.

Cross-validation trials across the combined set of both cohorts (n=40) using library-isolated peptides with and without the AT_1 epitope yielded averaged areas

under the curve (AUC) of 0.78 and 0.72, respectively (**Figure 2-7**), achieving 61% (95% confidence interval, 36-83%) sensitivity at 95% (95% confidence interval, 77-100%) specificity (**Table 2-4**). Comparatively, the AT_1 epitope alone demonstrated an AUC of 0.65 in this set of 40 samples and at 61% sensitivity exhibited 55% (95% confidence interval, 32-76%) specificity.

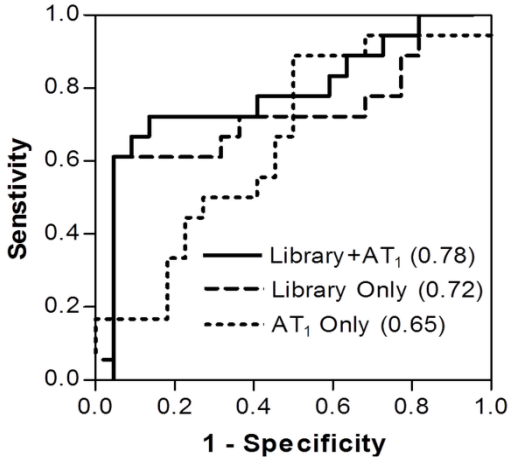

Figure 2-7: Receiver operating characteristic (ROC) curves are shown for algorithm predictions using library peptides with and without the angiotensin II type 1 receptor (AT_1) epitope and the AT_1 epitope alone across the combined training sample set (n=40). The respective areas underneath the ROC curves are also indicated in parantheses.

Next, the classification accuracy of algorithms trained using cohort 1, cohort 2, or the combined set with and without the AT_1 epitope was assessed using an external validation set (20 remaining cohort 2 samples). The ROC curves of all 6 predictive algorithms revealed that use of samples from both cohorts in training resulted in the highest AUC, especially when including the AT_1 epitope (0.83; **Figure 2-8**). The algorithm trained using only the library peptides with the combined set attained a similar AUC (0.78), and both algorithms achieved comparable overall accuracy, 75% and 80%, respectively, at the prescribed cutoff. Although the algorithm including the AT_1 epitope yielded only 1 false-positive, it misclassified 2 more PE samples than the algorithm using only library peptides (**Table 2-5**).

Table 2-4: Average algorithm predictions

	With AT_1 Epitope		Without AT_1 Epitope	
Sample ID	Average	SEM	Average	SEM
PE 1-1	0.10	0.03	0.16	0.13
PE 1-2	0.84	0.09	0.89	0.07
PE 1-3	0.70	0.13	0.75	0.11
PE 1-4	0.81	0.08	0.98	0.01
PE 1-5	0.74	0.04	0.74	0.11
PE 1-6	0.78	0.10	0.85	0.05
PE 1-7	0.39	0.06	0.32	0.19
PE 1-8	0.75	0.11	0.83	0.05
PE 1-9	0.79	0.05	0.81	0.08
PE 1-10	0.71	0.08	0.63	0.02
HOP 1-1	0.13	0.11	0.06	0.03
HOP 1-2	0.20	0.07	0.05	0.03
HOP 1-3	0.03	0.01	0.03	0.01
HOP 1-4	0.25	0.06	0.05	0.04
HOP 1-5	0.36	0.04	0.31	0.06
HOP 1-6	0.06	0.04	0.09	0.03
HOP 1-7	0.26	0.08	0.26	0.10
HOP 1-8	0.16	0.07	0.34	0.07
HOP 1-9	0.06	0.02	0.19	0.08
HOP 1-10	0.38	0.18	0.49	0.08
PE 2-1	0.82	0.08	0.88	0.03
PE 2-2	0.65	0.10	0.72	0.01
PE 2-3	0.36	0.12	0.34	0.13
PE 2-4	0.13	0.08	0.08	0.05
PE 2-5	0.73	0.06	0.87	0.08
PE 2-6	0.17	0.09	0.22	0.08
PE 2-7	0.30	0.07	0.07	0.03
PE 2-8	0.24	0.13	0.13	0.05
HOP 2-1	0.36	0.05	0.29	0.19
HOP 2-2	0.31	0.10	0.40	0.05
HOP 2-3	0.26	0.10	0.38	0.01
HOP 2-4	0.45	0.25	0.38	0.04
HOP 2-5	0.11	0.04	0.19	0.03
HOP 2-6	0.30	0.10	0.30	0.06
HOP 2-7	0.91	0.05	0.91	0.08
HOP 2-8	0.34	0.03	0.29	0.11
HOP 2-9	0.34	0.07	0.39	0.06
HOP 2-10	0.08	0.03	0.24	0.04
HOP 2-11	0.16	0.04	0.33	0.09
HOP 2-12	0.32	0.14	0.24	0.02
Sensitivity	61%		61%	
Specificity	95%		95%	
Accuracy	80%		80%	

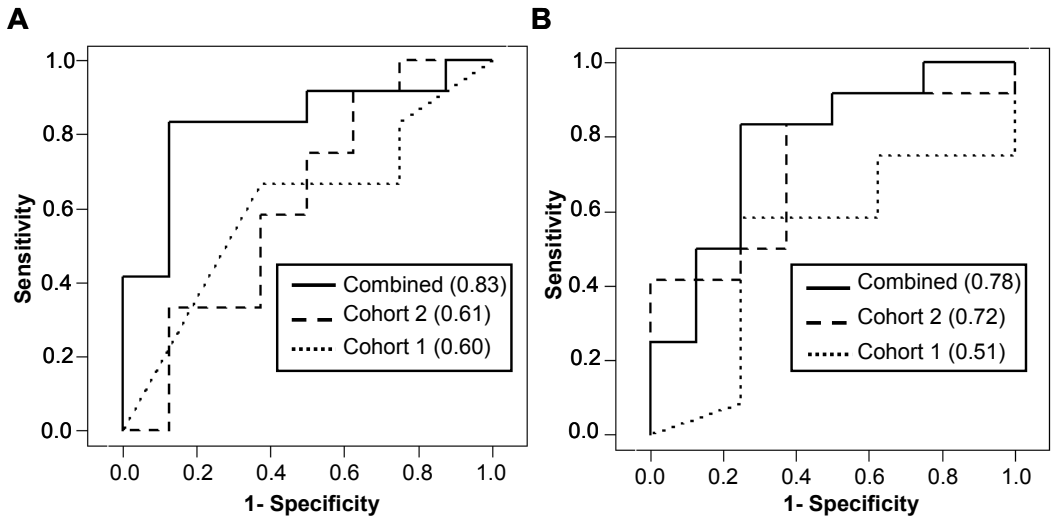

Figure 2-8: Receiver operating characteristic (ROC) curves demonstrated improved performance using the combined set of samples in a validation set. ROC curves were generated for algorithms **A,** with or **B,** without the AT_1 epitope in the peptide panel against the final 20 samples from cohort 2. The algorithms were trained using samples from cohort 1, cohort 2, or the combined set. The areas underneath the curves (AUC) are also indicated.

Finally, classification accuracy was not significantly different when analysis was restricted to cases of PE (n=19) identified strictly by proteinuria and no pre-existing hypertension. The array algorithm excluding the AT_1 epitope correctly detected 12 (63%) subjects of this PE subgroup, compared to 70% in the full, more heterogeneous group (n=30).

Table 2-5: Classification algorithm performance

Training Set	PE (n/N)	HOP (n/N)	Overall (n/N)
Library Peptides and AT_1 epitope			
Cohort 1	4/12 (33%)	7/8(88%)	11/20 (55%)
Cohort 2	8/12 (67%)	5/8 (63%)	13/20 (65%)
Combined	8/12 (67%)	7/8 (88%)	15/20 (75%)
Library Peptides			
Cohort 1	7/12 (58%)	5/8 (63%)	12/20 (60%)
Cohort 2	5/12 (42%)	7/8 (88%)	12/20 (60%)
Combined	10/12 (83%)	6/8 (75%)	16/20 (80%)

In addition to evaluating overall diagnostic performance, the antibody-detecting peptide algorithm was assessed for adverse outcome detection use. Because of the

inclusion criteria used in this study, 6 of 30 patients with PE used in training and validation exhibited severe symptoms, such as central nervous system disturbances, thrombocytopenia, and elevated liver enzymes, without proteinuria. Importantly, the antibody-detecting peptide panel identified 5 of these 6 atypical patients with PE. Furthermore, the peptide panel detected 5 of 6 nonproteinuric patients with PE that delivered before 37 weeks' gestation (n=18; **Table 2-6**). This could help stratify patients requiring more timely delivery thereby complementing severe symptom detection. Overall, combining clinical criteria of high BP and proteinuria along with the antibody-detecting peptide panel achieved the highest sensitivity for severe symptoms and early delivery.

Table 2-6: Association of antibody-detecting peptide panel reactivity with adverse outcomes.

Detection method	Severe symptoms (n=16)	Early delivery (<37 weeks) (n=18)
Proteinuria	10 (63%)	12 (61%)
Peptide panel (+AT_1 epitope)	12 (75%)	12 (67%)
Peptide panel	13 (81%)	13 (72%)
Proteinuria + Peptide panel	15 (94%)	17 (94%)

Severe symptoms include central nervous system disturbances, elevated liver enzymes, and/or thrombocytopenia. Proteinuria refers to patients with pre-eclampsia with positive proteinuria. AT_1 indicates angiotensin II type 1 receptor

iii. Statistical analysis of individual peptide performance

Individual peptides constituting the panel exhibited differing diagnostic efficacy. The AT_1 epitope exhibited significantly ($P<0.05$) higher PE antibody binding when evaluated across the entire sample set (45 PE and 48 HOP; **Figure 2-9**). Here, the AT_1 epitope detected binding antibodies in 78% of PE and 44% of HOP, resulting in an AUC of 0.66. However, binding of antibodies from subjects with PE to the AT_1 epitope was not significantly increased in the validation set composed of 60 samples. In contrast, 5 library-isolated peptides exhibited significantly ($P<0.05$) higher

reactivity with PE samples than with HOP samples (**Figure 2-10A**). In addition, library-isolated peptides achieved comparable or higher AUCs than the AT_1 epitope. Peptides cross-reacted with multiple PE patient antibodies, and antibodies from PE that reacted strongly with one peptide tended to bind multiple peptides (**Figure 2-10B**). Similarly, HOP antibodies that reacted with one peptide also tended to bind multiple peptides including the AT_1 epitope. Nevertheless, the 7-member panel exhibited stronger diagnostic efficacy than any individual peptide. Interestingly, peptide binding activity, especially peptide 36 (r_s=-0.62), inversely correlated with PE patient platelet count in this set (**Figure 2-11**). Summing the standardized binding activity of the 7 library-isolated peptides and AT_1 epitope yielded the overall correlation (r_s=-0.56) with platelet count. Analysis of other patient characteristics (ie, BP or proteinuria) did not reveal strong correlations with peptide standardized reactivity.

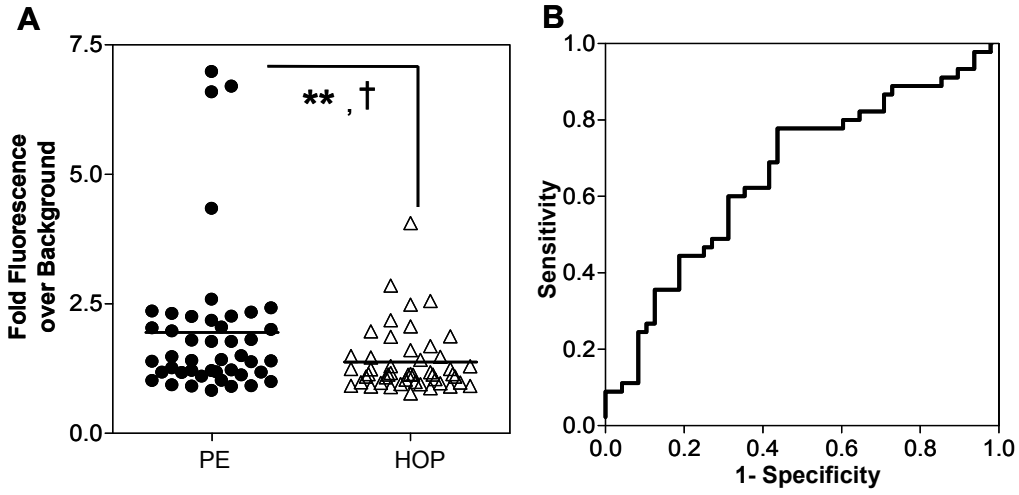

Figure 2-9: AT_1-AA detection in the entire set of 45 PE and 48 HOP samples. **A.** Binding activity of the cell-surface expressed AT_1 epitope and **B**. the ROC analysis which yielded an AUC of 0.66. Mann-Whitney U-test (**), Student's *t*-test (†) where † p<0.05, ** p<0.01

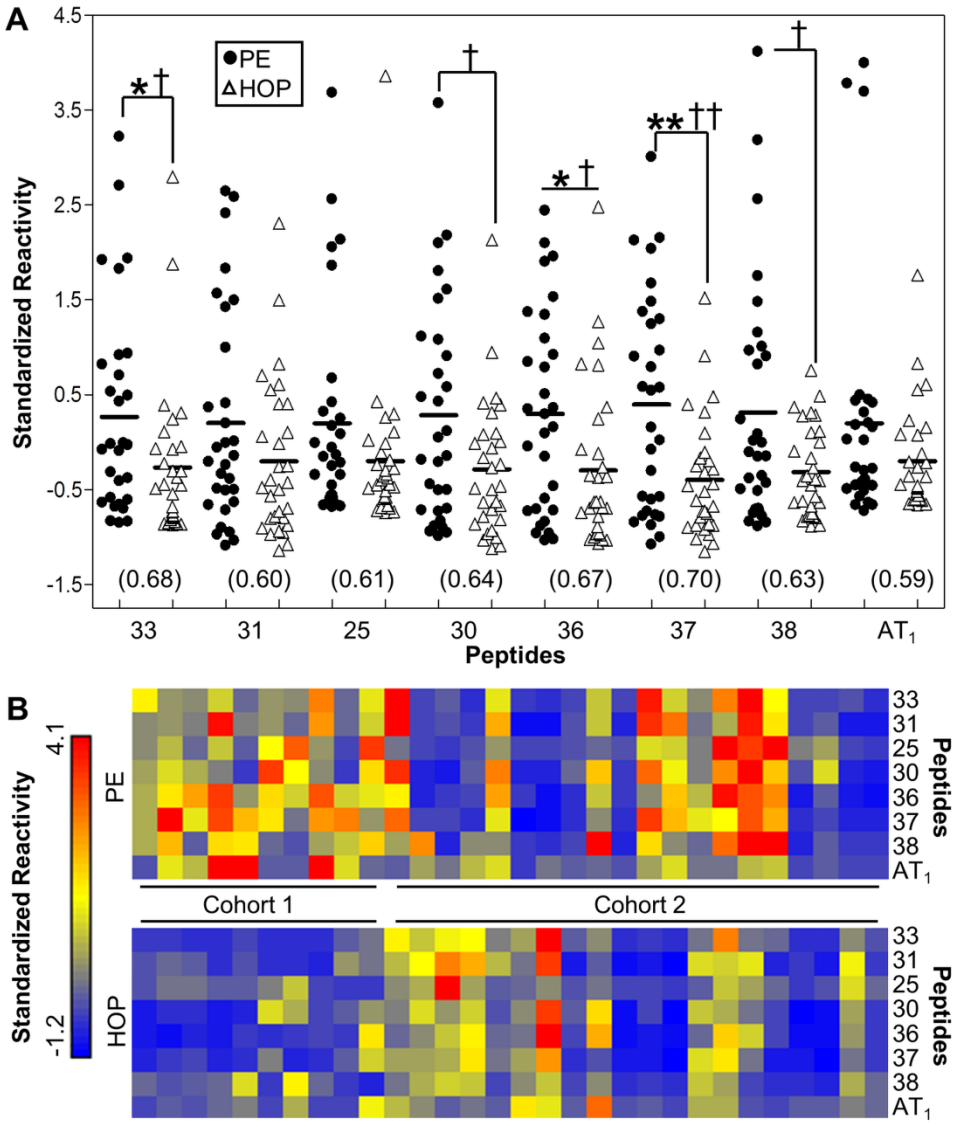

Figure 2-10: Library-isolated peptides and the angiotensin II type 1 receptor (AT_1) epitope demonstrate cross-reactivity with pre-eclampsia (PE) patients and decreased healthy-outcome pregnancy (HOP) binding. **A,** Peptide reactivity with 30 samples each of PE and HOP in a dot plot, where lines represent the mean of the population. Area underneath the ROC curve is also indicated for each peptide in parantheses. **B,** A heat map shows cross-reactivity and application as a peptide panel. Statistical significance indicated for Mann-Whitney U-test (*), Student *t*-test (†) where applicable. * or † P<0.05, ** or †† P<0.01 Standardized reactivity=(sample fluorescence–average fluorescence)/SD of fluorescence.

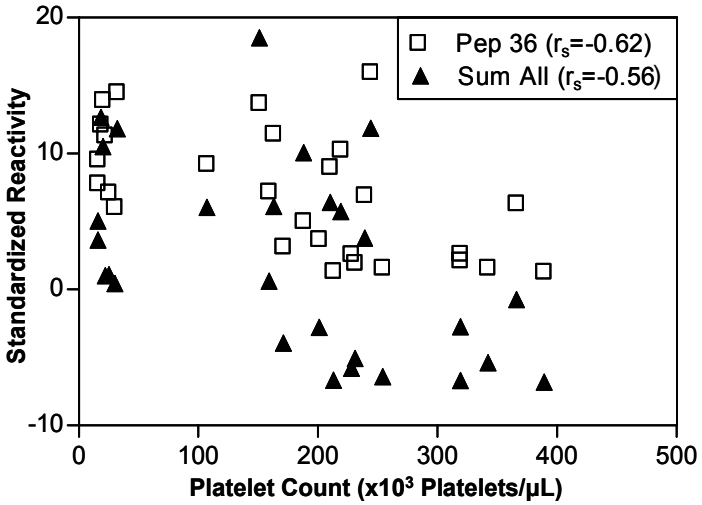

Figure 2-11: Peptide binding activity is inversely correlated with the platelet counts in PE patients. The strongest correlation observed with peptide 36 and the summed standardized reactivity for all eight peptides, including the AT_1 epitope, are shown. r_s=Spearman correlation coefficient

D. Discussion

In this study, we present evidence that PE is associated with a distinctive signature of antibody-binding specificities. This signature was represented in a PE antibody-detecting peptide panel composed of multiple epitope specificities. One such specificity corresponds to a 7-mer epitope of the AT_1 receptor.[2] The pathobiological significance of these AT_1-AAs is now supported by multiple independent studies, demonstrating their ability to increase BP, proteinuria, and sFlt-1[3] and complement deposition.[173] Despite this increased complement deposition in the placenta and kidney of pregnant mice, blocking the AT_1 receptor did not fully reduce deposition to HOP levels, leaving open the possibility that other antibodies could play a role in PE pathology. The pursuit for additional PE antibody markers has identified an association between PE and antibodies binding various autoantigens, including β1, β2 and α1 adrenoreceptors,[128] cardiolipin,[129] and prothrombin.[130] However, the use of unbiased discovery approaches to characterize the complete

antibody repertoire has not been described. Furthermore, the antigens responsible for eliciting these autoantibodies have not been determined, and their binding epitope specificities remain uncharacterized. Using an unbiased antibody repertoire analysis method, we observed a unique epitope reactivity pattern and used representative peptides to develop a PE antibody-detecting panel for diagnosis. In spite of the strong evidence for AT_1-AAs in PE, we did not identify peptides with similarity to the AT_1 epitope AFHYESQ. This result might be explained by an insufficient antibody affinity or titer, or the high frequency of AT_1-AA in HOP subjects (used for subtraction) in this and previous studies[135,136] and overall increased activity over nonpregnant normotensive subjects.[178] Although the mechanism responsible for the PE antibody signature is unclear, the reduced prevalence of T-regulatory cells in PE[123] may contribute to elevated autoantibody production. In addition, the observation that several nonself proteins also carry these motifs raises the possibility that the antibodies may originally be responding to an environmental trigger as previously proposed for AT_1-AAs.[135] Regardless of the production mechanism, this study supports an altered immune response in PE pathophysiology by verifying the presence of PE-specific antibodies in addition to AT_1-AAs that could be useful for diagnosis.

Despite the demonstrated role of AT_1-AAs in PE and potential for use in diagnosis and guiding therapy, their detection has proven exceptionally difficult. Current detection techniques rely upon complex biologic function assays,[2,138] that are unlikely to be effective for point-of-care diagnosis. To address this problem, we developed a unique binding assay for AT_1-AA detection using the 7-mer epitope

displayed in a high-avidity format on the bacterial cell surface. One recent study demonstrated higher AT_1-AA titers in PE patients (n=13) over HOPs (n=30) using the 27-aa second extracellular loop in an ELISA;[178] however, here the cell-displayed 7-mer was sufficient to detect epitope specific antibodies in 78% of PE patients and 44% of HOP (45 PE and 48 HOP). Thus, AT_1-AA detection can be performed with the 7-mer and does not require the entire loop. Assays based upon the minimal 7-mer may be important for developing therapeutics designed to block this antibody specificity[3] and differentiate this from specificities present in other diseases such as renal allograft rejection.[174] For comparison, a biologic function assay identified AT_1-AAs in ≈70% of patients with PE and 20% of HOPs[135] in addition to 62% of HOPs with abnormal perfusion.[136] The percentage of HOP subjects in our study that experienced abnormal perfusion during pregnancy is not known because the test is not part of routine practice. Regardless, these results demonstrate the use of an epitope-specific binding assay for AT_1-AA detection in patients with PE, for which researchers have previously relied upon complex biologic function assays.

The 7-member PE antibody-detecting panel demonstrated a potential use for detecting PE cases that is comparable with current protein biomarkers. The peptide panel exhibited 100% accuracy in the validation set from cohort 1 and maintained strong accuracy (80%) despite the inclusion of a second cohort from a distinct geographic location, which affects a population's antibody repertoire.[179] Despite this effect, the panel exhibited comparable diagnostic efficacy to biomarkers in clinical development, such as sFlt-1 and PlGF. Studies evaluating these markers have primarily focused on early-onset PE in which they show the strongest accuracy;

however, the accuracy drops in late-onset (≥34 weeks) PE.[180] Commercially available diagnostic kits for the sFlt-1/PlGf ratio and PlGF alone exhibited 59% sensitivity at 100% specificity and 77% sensitivity at 95% specificity, respectively, across all gestational onsets.[169] Similarly, this study did not discriminate between early- and late-onset and the antibody-detecting peptide panel achieved 61% sensitivity at 95% specificity within the cross-validated training set. It remains to be determined whether the PE-specific peptide panel can perform similarly in larger cohorts. Nevertheless, our results indicate that a PE antibody-detecting peptide panel can effectively discriminate PE and HOP samples and demonstrate that PE patients possess a distinctive antibody repertoire signature.

This PE-specific antibody signature was associated with an increased detection of adverse outcomes. A substantial percentage of patients who develop HELLP syndrome (10%-15%) or eclampsia (20%-25%) are not detected by the current clinical criteria of hypertension and proteinuria.[139] Thus, there exists an unmet need to identify those at risk of developing these adverse outcomes. The antibody-detecting peptide panel outperformed (81%) clinical criteria (63%) in detecting patients (n=16) with severe PE characterized by symptoms of central nervous system disturbances, elevated liver enzymes, and thrombocytopenia, suggesting that this antibody signature is more strongly associated with these severe PE symtpoms. Furthermore, peptide binding activity inversely correlated with platelet levels. Continued investigation of this association may link a pathophysiological role for these antibodies to these severe symptoms.

E. Perspectives

Here, we demonstrated the existence of antibody biomarkers present in PE patients distinct from known AT_1-AAs that achieve strong diagnostic accuracy (80%) for PE. Thus, our results provide supporting evidence for an altered immune response in PE. Identifying the antigen(s) mimicked by our library-isolated peptides may enable characterization of the antibody's contribution to PE pathogenesis while elucidating potential therapeutic targets. Furthermore, use of the whole or partial antigen(s) mimicked by library peptides may further increase sensitivity/specificity of the assay. In addition, the ability to detect PE early in pregnancy before clinical presentation would aid patient management. A time-course study using the PE antibody-detecting peptides described here could identify when these antibodies first present to assess their potential use in the early diagnosis of PE. Finally, because this methodology does not require purified antibodies, screening can be conducted using unprocessed, diluted plasma to identify peptide diagnostic reagents that bind PE specific antibodies without the additional purification step.

Novelty and Significance

What is New?

- This study identified antibody specificities that occur in pregnant women with pre-eclampsia (PE) that are distinct from known angiotensin II type 1 receptor (AT_1) autoantibodies (AT_1-AAs) with use for PE diagnosis.
- A binding assay for AT_1-AA detection was developed using bacterial cells that display on their surface the 7-mer AT_1 epitope.

What is Relevant?

- This peptide panel diagnostic assay demonstrated a high specificity (95%) at 61% sensitivity in the training set and maintained 80% accuracy in an independent validation set.
- Further characterization may identify antigen(s) mimicked by these PE-specific peptides potentially leading to novel therapeutic targets.
- The improved throughput of the binding assay for AT_1-AA detection will enable larger cohort studies to improve understanding of their prevalence in PE and HOP.

Summary

PE-specific antibody-detecting peptides isolated from a bacterial display library exhibited strong diagnostic efficacy within independent sample sets from distinct geographic locations. Using a unique AT_1-AA detection assay the prevalence of AT_1-AA in subjects with PE and HOP was comparable with that observed in previous studies using more complex biological function based bioassays.

3. A pre-eclampsia associated Epstein-Barr virus antibody cross-reacts with placental GPR50

While the human antibody repertoire reflects immune responses to diverse prior environmental exposures, it remains difficult to identify disease-associated antibodies and the environmental antigens eliciting these antibodies. To discover environmental antigens associated with pre-eclampsia (PE), bacterial display peptide library screening and evolution was applied to identify peptide epitopes recognized by antibodies occurring in women with PE. This strategy revealed that women with PE near delivery exhibit elevated IgG1 titers directed towards motif KXXXC[VIL]GCK present in the Epstein-Barr virus nuclear antigen (EBNA-1). Interestingly, the EBNA-1 epitope specific antibodies cross-reacted with a similar epitope within the extracellular N-terminus of the human G protein-coupled receptor, GPR50. GPR50 was strongly expressed in human placental tissue and found in immortalized placental trophoblast cells. Peptide epitopes derived from EBNA-1 and GPR50 exhibited significantly higher antibody binding activity among PE patients near delivery compared to that observed for healthy-outcome pregnancies at term and nulligravid samples. The EBNA-1 peptide potently blocked (IC50≈60-80 pM) binding of the PE-associated antibody to the peptide epitope within GPR50. These results reveal the existence of sequence level molecular mimicry between EBNA-1 and placental GPR50, indicating a mechanism for IgG1 deposition in the placenta, as a prevalent immunological feature of pre-eclampsia.

A. Introduction

Autoimmune diseases are thought to involve environmental factors that trigger disease onset, propagate disease, or influence disease severity.[30] However, the definitive identification of specific environmental factors has remained elusive in most cases. Nevertheless, the identification of specific environmental factors that drive pathological immune responses would greatly aid in the development of improved molecular diagnostics and efficacious therapeutic interventions.

Since the immune system archives environmental antigen exposures within the receptor repertoire of memory B- and T-cells, the identification of disease-associated antibodies and T-cell receptors (TCRs) and their corresponding antigen preferences has been pursued as a means to identify environmental factors. Antigen discovery for antibodies has been investigated using protein microarrays,[59,181] fragment libraries from proteins of interest,[69,182,183] and cDNA libraries from tissues or cells involved in pathology.[70,77] TCR epitope profiling methodologies include enzyme-linked immunospot[184] and peptide MHC class II tetramer assays.[185] These methods have enabled identification of both validated antigens[59,182-185] and novel candidate autoantigens.[59,70,77,181,183] However, because such methods rely upon assaying antibody repertoires against a predefined, and often incomplete, proteome they have not enabled an impartial identification of environmental antigens. Massively parallel DNA sequencing of antibody or TCR variable domains has proven to be a powerful approach to characterize alterations in the B-[52,186] and T-cell[187,188] repertoires. Characteristic repertoire changes have been observed in multiple sclerosis,[186] rheumatoid arthritis,[52] leukemia,[187] and severe aplastic anemia.[188] Recent advances

have enabled autoantigenic target identification for sequenced antibodies using microarray analysis;[52] however, in general, these repertoire sequencing approaches have not revealed environmental antigens that may elicit these molecules. Consequently, there remains a distinct need for methods which could identify the environmental antigens giving rise to pathological immune repertoires.

Random peptide library screening does not require *a priori* knowledge of disease mechanisms and has been applied to a variety of diseases. In particular, phage display screening has been applied to inflammatory diseases,[75] cancers,[76] and autoimmune diseases.[79-81] A phage displayed library screened against CD patients discovered a common epitope of gliadin.[80] In multiple sclerosis (MS), phage display identified peptides that bound a higher fraction of cerebrospinal fluid antibodies in MS than in disease control.[79] However, these screens either used few individual patient specimens[76,79] or a single pool/patient-derived antibody for discovery.[75,80,81] This approach inhibits selective isolation of widely prevalent antibodies in multiple individuals. Furthermore, some relied on few peptide sequences[79-81] to identify similarities with candidate viral and human antigens and used pre-existing knowledge to focus searches for candidate targets. However, recent application of random peptide library screening impartially identified a CD-specific epitope corresponding to one of the hallmark environmental antigens, deamidated gliadin, through directed evolution.[189] Since random peptide library screening requires no prior knowledge of the disease, we sought to apply this methodology to identify disease-associated antibodies and their binding targets for pe-eclampsia (PE), a disorder affecting 5-8% of pregnancies.[1] PE causes 15-20% of maternal mortalities and morbidities in

developed countries[101] and approximately a half million fetal deaths worldwide each year. Despite serious global impact, the etiology of PE remains largely unknown.

While the mechanisms behind PE pathology remain unclear, several immunological findings have been observed in patients with PE. Evidence suggests a further intensified inflammatory condition with increased levels of pro-inflammatory cytokines, IL-6 and IL-8, among PE patients compared to pregnant and non-pregnant control patients.[190] Additionally, a variety of autoantibodies[128,129,130] have been associated with PE. Notably, PE patients produce agonistic antibodies against the angiotensin II type 1 receptor (AT_1-AAs)[2] that appear as early as 18 weeks' gestation.[136] These AT_1-AAs induced symptoms of PE in an adoptive transfer mouse model[3] and amplified complement protein C3 deposition in the placenta and kidney of pregnant mice.[173] Increased C3 deposits have also been observed in placental vessels from a transgenic PE rat model, and addition of supernatant obtained from cultured PE placental explants enhanced C3 expression in rat vascular smooth muscle cells.[122] Furthermore, mutations within complement system regulatory proteins have been implicated in PE.[118] In addition to complement system disturbances, PE patients exhibit significantly reduced levels of circulating[123] and placental[124] CD4+CD25+FoxP3+ regulatory T-cells, which can lead to increased autoantibody production.[125] Moreover, CD19+CD5+ B-cells, a class associated with the production of autoreactive antibodies,[127] are elevated in PE patients.[126] These isolated cells have been shown to produce AT_1-AAs in culture upon addition of PE serum. Taken together, these prior studies support an important role for the immune system in the pathogenesis of PE.

Despite their apparent involvement in PE, previously characterized AT_1-AAs are difficult to detect and vary in prevalence among studies, from 70%[135] to 95%.[138] Patients with PE were recently shown to possess additional disease-associated antibodies distinct from these AT_1-AAs, although these antibodies could not be conclusively related to any self or environmental antigens (**Chapter 2**).[191] Here, we sought to identify antigens that may be recognized by PE-associated antibodies, and thereby elucidate the immune mechanisms of pathogenesis. Our results point to a molecular mimicry mechanism operating in PE women near delivery, wherein a viral antigen specific antibody cross-reacts with an abundant placental antigen. More generally, our study further supports the idea that bacterial display peptide library screening may be useful to identify environmental factors involved in human disease in an unbiased fashion.

B. Materials and Methods

i. Patient Samples

The majority of patient samples used in this study have been previously described (**Chapter 2**).[191] Whole blood samples from pregnant women with PE or healthy-outcome pregnancy (HOP) were provided as aliquots of samples taken for routine blood work during clinical assessments at the Santa Barbara Cottage Hospital. The Santa Barbara Cottage Hospital Institutional Review Board approved this study. PE subjects fulfilled at least two of the following criteria: i) two documented blood pressures (BP) with readings greater than 140/90mm Hg at least 4 hours apart, with documented normal BPs in the first half of the pregnancy, ii) proteinuria using a spot urine check (≥30 mg/dL), dipstick reading (≥1+), or 24 hour analysis (≥300 mg/24)

iii) central nervous system (CNS) symptoms (visual disturbances or unremitting headaches), iv) epigastric pain associated with elevated liver enzymes unrelated to other abdominal pathology, v) or thrombocytopenia with platelet counts less than 100,000 U/mL. Additional plasma aliquots from PE, HOP and nulligravid individuals and tissue samples were provided from the University of Texas, Houston Medical School. These PE samples were diagnosed by clinical assessments based on the National High Blood Pressure Education Program Working Group Report. Superimposed PE cases (those with a previous history of hypertension) were not included, but this study did not discriminate between early-and late-onset PE or based on parity. Subjects provided informed consent, and samples were collected according to institutional guidelines. Blood samples were obtained near the time of delivery. Recorded BPs represent the maxima observed upon PE presentation or prior to delivery.

ii. Screening a randomized peptide library against dilute plasma

A 15-mer random peptide library was displayed on the surface of *Escherichia coli* MC1061[175] as a fusion to the N-terminus of the eCPX scaffold.[90] Overnight incubation of patient derived diluted plasma (1:50) with empty library scaffold expressing cells removed *E. coli* binding antibodies from the retained supernatants. Using these depleted plasma samples, library screening sequentially removed peptides binding antibodies present in HOP samples using magnetic selection and enriched for peptides that bind PE plasma antibodies with fluorescence activated cell sorting (FACS). Groups of three PE and three to four HOP plasma samples were pooled together to create six pools of each class. After incubating with either PE or

HOP plasma diluted (1:100 or 1:200) in phosphate buffered saline (PBS), labeling proceeded with a biotinylated anti-human IgG specific secondary (Jackson ImmunoResearch) diluted (1:500) in PBS with 0.1% bovine serum albumin (BSA). Streptavidin (SA)-conjugated magnetic beads were used for magnetic depletion while SA conjugated to R-phycoerythrin (SA-PE) (Invitrogen) diluted (1:333) in PBS with 0.1% BSA fluorescently labeled cells for FACS. Incubations with plasma or labeling reagents were conducted at 4°C. All PE and HOP pools were quantitatively assessed for binding to the library population at each round of FACS to determine which pool to use for enrichment or depletion. The screening was performed in duplicate using the same sample pools, but in different order of depletion/enrichment. Bacterial colonies (~130) were randomly selected for sequencing from different rounds of the duplicate screens. Peptide binding motifs were separately determined by inspection of the unique sequences identified using the Geneious software package.

To further evolve the peptide binding motif, a focused bacterial display peptide library of the form XXXKXXXC[VIL]GCXXXX was constructed. Screening against this library proceeded as described above but involved three new pools each of PE and HOP using further diluted plasma (1:200 and 1:500). From the focused library, about 100 colonies were selected for sequencing from screening rounds (primarily from the final round) to assess the impact of increased screening stringency upon the consensus motif.

iii. Identifying the native antigen corresponding to the peptide motif

Unbiased searches using NCBI BLASTp and ScanProsite identified candidate antigens and the corresponding source organism. Three proteins were chosen as

candidate environmental trigger antigens. These fulfilled the requirements that they shared identities with the most common amino acids of the motif, specifically KXX[NSTG]C[VIL]GCK, and weren't hypothetical proteins. Individual 15-mer fragments derived from these proteins were cloned onto the N-terminus of the eCPX scaffold along with a C-terminal peptide tag (P2x) that binds a fluorescent reporter (YPet-Mona) of scaffold expression.[176] The human proteome was separately searched to identify candidate autoantigens. To qualify, the protein shared at least five identities to the searched motif. Two of the six human protein fragments only differed by one amino acid, so one was selected for autoantigen assays. These cell-surface expressed antigen fragments and library-isolated peptides were evaluated for significantly increased PE binding over HOP and dynamic range for at least 15 PE and 15 HOP. After incubating with diluted plasma (1:200), cells were washed twice with cold PBS and resuspended with biotinylated anti-human IgG (Jackson ImmunoResearch) diluted (1:500) in PBS with 0.1% BSA for secondary labeling. Subsequently, the cells were washed with cold PBS and resuspended in SA-PE diluted (1:333) in PBS with 0.1% BSA for fluorescent labeling and flow cytometric analysis. The fluorescent intensity measured for each peptide was divided by the background intensity of the negative control, scaffold without an N-terminal peptide. Statistical evaluation using Student's *t*-test or the Mann-Whitney U-test and receiver operating characteristic (ROC) curve analysis was conducted using Prism 4 or 6 software (GraphPad Software Inc.). Epitopes with a statistically significant ($p < 0.05$) difference between PE and HOP and highest dynamic range were down-selected for further analysis against 42 PE and 43 HOP samples.

iv. Analyzing full-length EBNA-1 protein activity

PE (n=36) and HOP (n=39) samples were assayed in duplicate for antibody binding to full-length EBNA-1 using a commercial IgG ELISA (GenBio ImmunoWell) following the manufacturer's protocol. The Spearman correlation coefficient between the binding activity of bacterial displayed peptides and the ELISA, statistical significance (one-way U-test) of PE activity in the ELISA, and distribution normality (Kolmogorov-Smirnov) was evaluated in Prism 4.

To confirm a relationship between EBNA-1 fragment (EB15) activity and the ELISA, EB15 binding antibodies were depleted from ten reactive PE samples, one reactive HOP, and nine nonreactive HOP samples and subsequently evaluated by ELISA. Depletions were carried out by incubating plasma samples diluted (1:50) in the specimen diluent provided in the commercial kit with ~$1x10^7$ cells/μL. The depleted supernatant was retained after centrifugation and evaluated for ELISA activity. This was repeated for duplicate measurement. As a control, the eCPX scaffold with just the C-terminal tag peptide was used to "deplete" five reactive PE and five nonreactive HOP samples and test for ELISA activity. Prism 4 analysis determined any statistically significant differences between the ELISA activity observed among depleted samples.

v. Antibody blocking activity of the EBNA-1 synthetic peptide

The 15-mer EBNA-1 fragment (EB15) was synthesized, including a disulfide bond between the cysteine residues. An unrelated synthetic peptide (NCP) was used as a negative control. Competition assays were conducted by pre-incubating pools of three PE patients for one hour at room temperature with varied concentrations of the

synthesized EB15. Selected patients included in these pools exhibited high binding activity to GPR50. These pools were then assayed for antibody binding to the GPR50 fragment to determine an IC50 value using Prism 4.

To evaluate antibody binding activity to the full-length GPR50, transient transfections were conducted using HEK293T cells. Cells were cultured in DMEM with 10% fetal bovine serum (FBS) and 1% antibiotic (Pen/Strep) prior to splitting for transfection, for which the Pen/Strep was removed. The commercially supplied vector (OriGene, pCMV-AC-GPR50-GFP) included a C-terminal GFP tag. Cells were harvested on the third day following transfection and pre-blocked for 1.5 hours at room temperature in PBS with 3% BSA. After confirming GFP expression, cells were evaluated for GPR50 expression, using a positive control monoclonal antibody (R&D Systems), and binding to pools of PE patients and HOPs plasma (1:200) with and without EB15 (20 nM) by flow cytometry. An anti-mouse IgG conjugated to Alexa647 (1:500) fluorescently tagged cells labeled with the monoclonal anti-GPR50 (1:400), while the biotinylated anti-human IgG followed by SA-PE fluorescently labeled human plasma antibodies bound to cells. Labeling reagents were diluted in PBS with 1% BSA. Washes between labeling steps used PBS with 0.05% Tween20 and 1% BSA but the final wash and resuspension used PBS with 1% BSA. Activity was compared to untransfected cells and an empty vector transfection control.

Additionally, antibody binding activity was evaluated using HTR-8/SVneo cells cultured in RPMI (10% FBS and 1% Pen/Strep). The immunoglobulin (Ig) fraction was purified from separate pools of PE and HOP plasma, exhibiting positive or negative antibody binding to the GPR50 fragment, respectively. Testing these Ig

pools (500 nM) with and without EB15 (20 nM) for binding to HTR-8/SVneo cells proceeded by flow cytometry similar to above except using an anti-human IgG secondary directly conjugated to R-phycoerythrin (Jackson ImmunoResearch) (1:100). Furthermore, the GPR50 specific monoclonal antibody detected GPR50 expression in this trophoblast cell line, as with HEK293T.

vi. Immunohistochemistry

Paraffin-embedded tissue sections were cleared using the xylene substitute, SafeClear II and dehydrated using 100%, 90%, and 70% ethanol solutions. Antigen retrieval was conducted by microwaving the tissue sections for ten minutes in citrate buffer (pH 6). All incubations with tissue slides were conducted in a humid chamber. Washes were performed two to three times each for five minutes, either with dH2O or PBS with 0.05% Tween20 (PBST). In addition to a native peroxidase block, a blocking step using PBS with 5% normal goat serum (blocking buffer) preceded overnight incubation at 4°C with primary antibody either anti-GPR50 (BMA Biomedical) or an isotype control, mIgG1 (Southern Biotech) at 10 μg/mL diluted into blocking buffer. Tissues were incubated with a biotinylated goat anti-mouse IgG secondary (1:200) (Vector Labs) diluted in blocking buffer at room temperature for one hour. During this time, the ABC (avidin/biotinylated enzyme complex) reagent (Vector Labs) was mixed and allowed to sit for 30 minutes at room temperature. Following the secondary, tissues were incubated with the ABC reagent for 45 minutes at room temperature. After adding the freshly prepared ImmPACT NovaRed HRP substrate (Vector Labs), the reaction continued for five minutes until stopped with dH2O. Mayer's hematoxylin counterstained the tissues, followed by dehydration

with 95% and 100% ethanol, clearing and mounting with DPX mountant. Slides were viewed on an Olympus BX51 microscope using the Q-Capture Pro 7 software.

C. Results

i. Library screening reveals an antibody response to a viral target

To identify antigen targets of antibodies associated with PE, we screened a bacterial display peptide library for antibody binding using diluted plasma. This study used samples from a heterogeneous set of women with PE (n=44), HOP (n=47), and nulligravid women (n=21) (**Table 3-1**). The diverse ($8x10^9$ members) peptide library underwent six depletion and five enrichment rounds against a subset of the PE (n=18) and HOP (n=20). This screening yielded a peptide population with increased binding activity to PE pools and reduced binding to HOP pools (**Figure 3-1A**).

Table 3-1: Clinical characteristics of patients

Patient Characteristics	PE (n=44)	HOP (n=47)	Nulligravid (n=21)
Age, yr	28.9 (1.2)	28.0 (1.0)	26.2 (5.6)
GAD, wk	35.4 (0.5)*	38.8 (0.3)	NA
Highest SBP, mm Hg	159.7 (2.9)*	114.9 (1.8)	118 (11)
Highest DBP, mm Hg	97.1 (1.6)*	67.0 (1.3)	75 (7)
Proteinuria, n (%)	35 (79%)	3 (7%)	ND
ALT > 70 IU/mL, n	11 (24%)	ND	ND
AST > 70 IU/mL, n	10 (21%)	ND	ND
CNS symptoms, n	21 (48%)	2 (5%)	ND
Platelets < 10^5 U/mL, n	6 (14%)	1 (2%)	ND

Abbreviations: SBP – systolic blood pressure, DBP – diastolic blood pressure, GAD – gestational age at delivery, ALT – alanine transaminase, AST – aspartate transaminase, CNS – central nervous system, ND- not determined, NA – not applicable; * $p < 0.0001$ Mann-Whitney U-test

Sequence analysis identified a binding motif (KXXXC[VIL]GC) that comprised 86% of a total of 91 unique sequences (**Figure 3-2A**). This same motif was observed in each of two replicate screens performed in parallel, confirming reproducibility.

Further screening of a second generation (focused) library of the form XXXKXXXC[IVL]GCXXXX using new pools of plasma from PE and HOP cases led to enhanced PE cross-reactivity and specificity (**Figure 3-1B**).

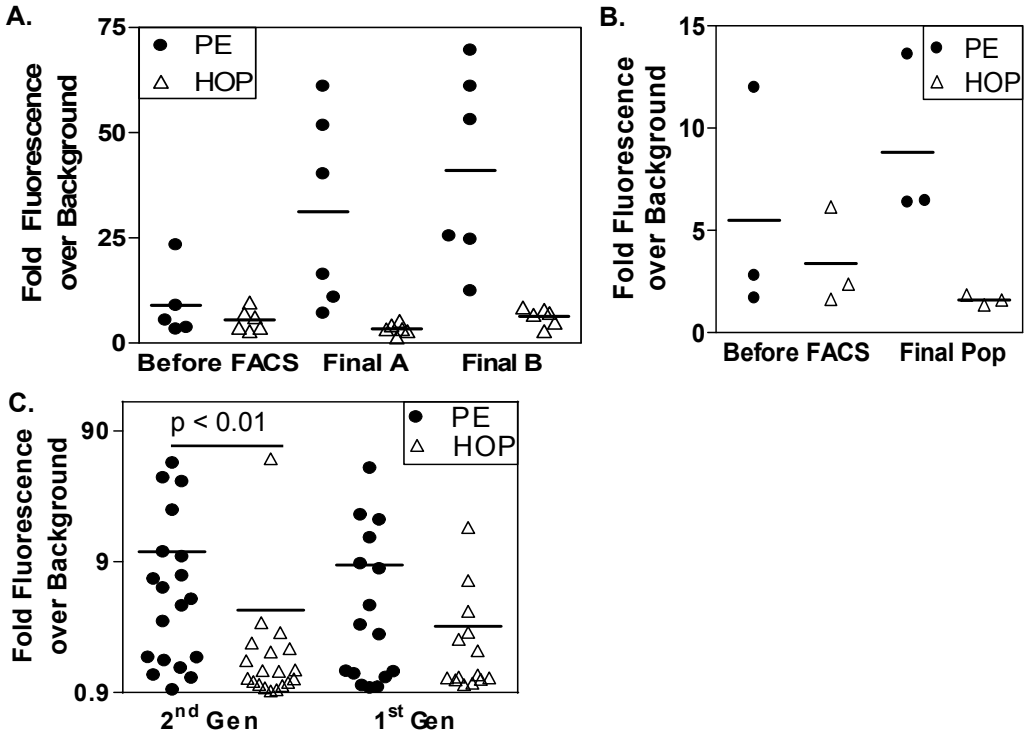

Figure 3-1: Library screening and evolution against dilute plasma identified cross-reactive and specific peptides. (**A**) Duplicate library screens against distinct pools of PE and HOP enhanced activity to antibodies present in PE in the first screen, and (**B**) further screening with new samples evolved the 2nd generation library. (**C**) Furthermore, a representative peptide from the 2nd generation library showed significantly increased PE antibody binding activity. p-value indicates Mann-Whitney U test

Although three pools of PE and HOP were evaluated at each round only two pools of PE (n=6) and HOP (n=6) were used to select for disease-associated peptides through five rounds of enrichment and depletion each. The other evaluated PE pool consistently exhibited strong antibody binding activity, while the remaining HOP pool demonstrated low binding activity. Sequencing individual clones isolated from the focused library identified 43 unique sequences from the final population. These sequences exhibited increased prevalence of valine in the middle position and lysine following the second cysteine (i.e., KXXXC[VLI]GCK **Figure 3-2A**). Comparison

of representative peptides from the first and second generation screens confirmed that directed evolution resulted in enhanced PE-specific antibody binding (**Figure 3-1C**).

Searches of the nonredundant protein database with variations of this extended motif yielded several hits, including an N-terminal epitope within the Epstein-Barr virus (EBV) nuclear antigen 1 (EBNA-1) KRPSCIGCK that exhibited high similarity. To investigate whether the identified motif mimicked the EBNA-1 epitope, bacterial displayed 15-mer fragments from EBNA-1 and two other candidate antigens showing high similarity (**Table 3-2**) were assayed for binding to antibodies present in PE and HOP samples. The EBNA-1 fragment (EB15) demonstrated significantly increased binding to antibodies present in PE over HOP (**Figure 3-2B**). Although another candidate antigen fragment, eFB, differentially bound antibodies present in PE compared to HOP samples, the PE antibody binding was significantly ($p < 0.001$) lower than that observed with EB15. Furthermore, EB15 exhibited a greater dynamic range for antibody binding activity between PE and HOP samples. While average antibody binding to EB15 was 4.8-fold higher among PE than HOP, eFB only showed a 2.3-fold difference. Additionally, to correctly differentiate 91% of HOP from PE, EB15 detected 73% of PE (n=23, each), while the eFB fragment only distinguished 43% of PE. These observations highlighted the enhanced antibody binding to EB15 among PE patients compared to eFB.

In addition to the PE-specific antibody binding to the EB15 epitope, a commercial ELISA evaluated whether a differential trend in antibody binding continued against the full-length protein. In a set of 36 PE and 39 HOP, the overall ELISA activity was increased ($p < 0.05$, one-tailed) among women with PE (**Figure 3-2C**). The PE

samples also exhibited a skewed distribution towards higher activity, failing the Kolmogorov-Smirnov normality test ($p < 0.03$). The observed ELISA activity correlated well with antibody binding to the bacterial displayed EB15 (r_s=0.77)(**Figure 3-2D**). For comparison, the second generation library-derived peptide showed a similar correlation (r_s=0.75), improved from the first generation peptide (r_s=0.63), highlighting the gain of function obtained via directed evolution.

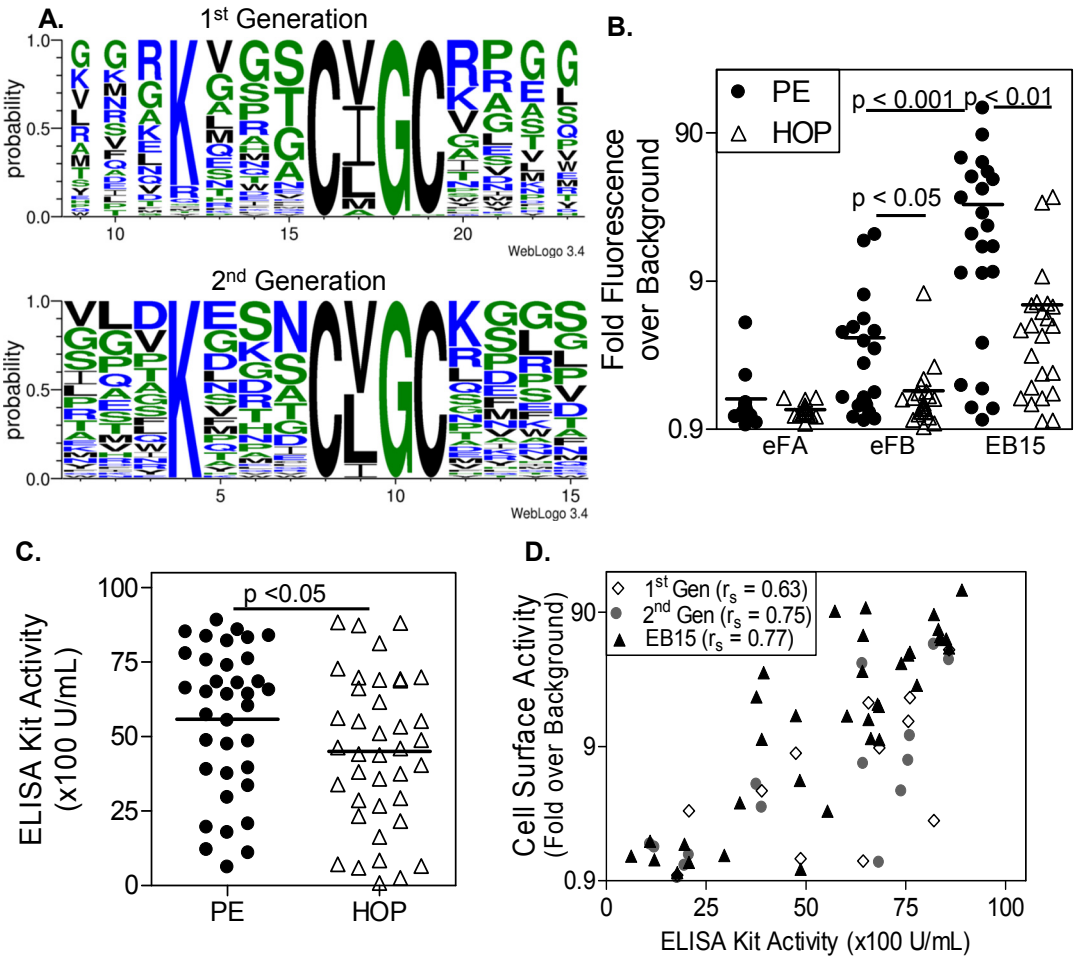

Figure 3-2: Library screening reveals presence of antibodies directed towards an epitope of the Epstein-Barr virus EBNA-1 protein. (**A**) Peptides from the 1st generation random library or a focused 2nd generation library were enriched for the KXXXC[VIL]GC[KR] motif. (**B**) Reactivity of three candidate antigen epitopes with PE and HOP antibodies is shown. (**C**). An ELISA for EBNA-1 antibodies exhibits a skewed distribution violating the Kolmogorov-Smirnov normality test ($p < 0.5$). PE samples exhibited significantly ($p<0.05$) higher activity according to a one-tailed U test. (**D**) ELISA signals correlated with binding activity to the surface displayed EB15, 1st, and 2nd generation peptides. r_s – Spearman correlation coefficient

Antibody depletion experiments further demonstrated that the motif identified from random peptide library screening corresponded to antibody binding to the EB15 epitope. Removing antibodies binding to the displayed EB15 resulted in a complete loss of binding to the library peptide (**Figure 3-3A**), indicating these peptides bind the same antibody species. Additionally, the contribution of the targeted EB15 epitope to the commercial ELISA signal observed for the full-length EBNA-1 was assessed. The ELISA evaluated the residual activity of plasma samples depleted of antibodies binding the displayed EB15 or scaffold protein only. Using EB15 reactive samples, bacterial displayed EB15 depleted plasma exhibited significantly reduced EBNA-1 binding activity by ELISA compared to scaffold depleted plasma (**Figure 3-3B**). In contrast, the EBNA-1 ELISA signals for plasma that did not show antibody binding activity to EB15 were not significantly different. Finally, the EB15 specific antibody was determined, using appropriate secondary reagents, to be of the IgG1 subtype (**Figure 3-4**), indicating that it is capable of fixing complement. Taken together, these results suggested that an IgG1 antibody prevalent in women with PE near delivery recognized an N-terminal linear epitope of EBNA-1 from the Epstein-Barr virus.

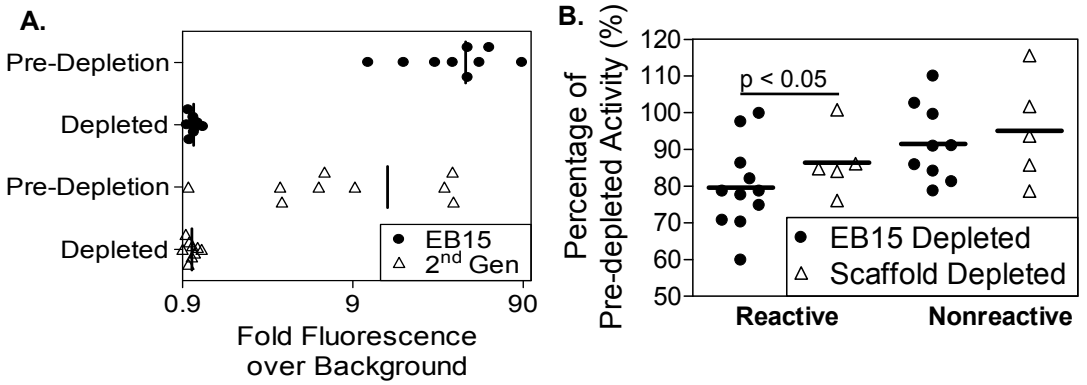

Figure 3-3: The EB15 linked library motif contributes to EBNA-1 ELISA signal. (**A**) By removing antibodies binding to EB15, the 2nd generation library peptide no longer exhibited antibody binding activity. (**B**) Furthermore, depleting EB15 binding antibodies from fragment reactive samples significantly reduced the overall ELISA activity observed compared to scaffold only incubations. p-value according to paired t-test

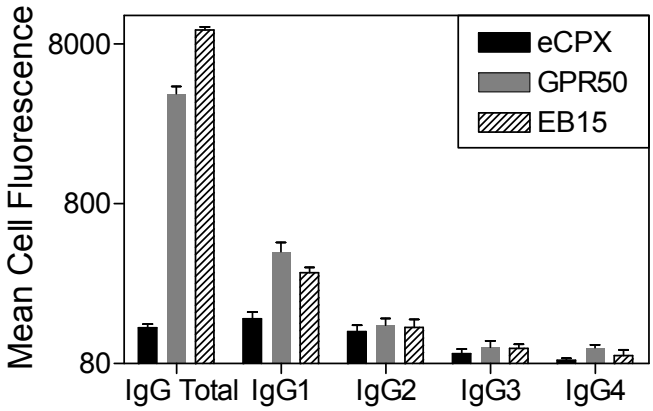

Figure 3-4: IgG subtyping experiments demonstrate only the IgG1 specific secondary reagent exhibited signal above background.

ii. The EB15 binding antibody cross-reacts with a region of GPR50

In an effort to identify human protein antigens that could be recognized by these epitope specific antibodies, BLASTp was applied to identify a set of candidate human antigens that exhibit the highest similarity to the library peptide motif (**Table 3-2**).

Table 3-2: Candidate antigen fragments

ID	Protein	Organism	Sequence
EB15	EBNA-1	Epstein-Barr Virus	RPQ**K**RP**SCIGCK**GTH
eFA	Iron-Sulfur Protein	Salmonella enterica	VFINEA**NCVGCK**LCV
eFB	Protein-Pyridine nucleotide disulfide oxidoreductase	Clostridium sp.: CAG:127	DGD**K**VK**NCIGCK**SCS
GPR50	GPR50	Homo sapiens	AVPTPYG**CIGCK**LPQ
hF2	Greb1	Homo sapiens	LLGFSG**NCVGC**GKKG
hF3	Tesmin	Homo sapiens	MCSSICK**CIGCK**NYE
hF4	Coiled-coil domain containing protein 18	Homo sapiens	FSN**K**EDR**CIGC**EANK
hF5	Zinc finger protein 501	Homo sapiens	TGE**K**PYE**CVGC**GKSF

Although five 15-mer fragments derived from candidate autoantigens were evaluated for antibody binding, only one fragment from an N-terminal region of a G protein-coupled receptor, GPR50, exhibited significantly increased binding to antibodies present in PE patients compared to HOP (**Figure 3-5A**). Furthermore, the EB15 peptide efficiently competed with PE antibody binding to the bacterial

displayed GPR50 fragment with an IC50 value of ~60-80 pM (**Figure 3-5B**). In contrast, a negative control peptide (NCP) did not affect antibody binding to the bacterial displayed GPR50 fragment (**Figure 3-6**).

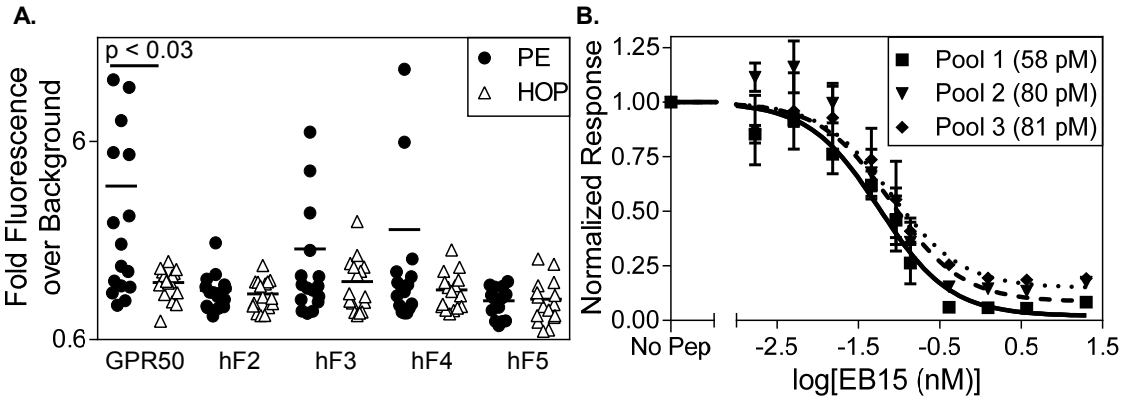

Figure 3-5: An EBNA-1 epitope (EB15) specific antibody cross-reacts with a GPR50 epitope. (**A**) Binding activity of PE and HOP antibodies to candidate autoantigen epitopes measured by flow cytometry. (**B**) Inhibition of antibody binding to GPR50 epitope by EB15 peptide for pools of diluted plasma from three PE patients. (IC50 values indicated in parentheses.) p-values represent Mann-Whitney U test results

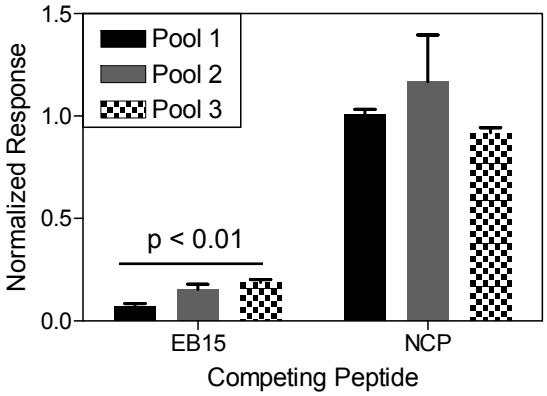

Figure 3-6: A synthetic negative control peptide (NCP) did not affect antibody binding to surface displayed GPR50. However, the same concentration of EB15 significantly reduced binding to near background levels. p-value indicates one-sample t-test

To investigate antibody binding to the full-length GPR50, cells expressing GPR50 were investigated by flow cytometry. In transiently transfected cells, GPR50 expression at the cell surface was confirmed using an anti-GPR50 monoclonal antibody and flow cytometry (**Figure 3-7A**). In this system of apparently strong GPR50 presence, GPR50 expressing cells exhibited increased binding to antibodies

present in pooled PE plasma compared to that observed with untransfected cells (**Figure 3-7B**). Furthermore, pre-incubating plasma with EB15 blocked antibody binding to the full-length GPR50 protein.

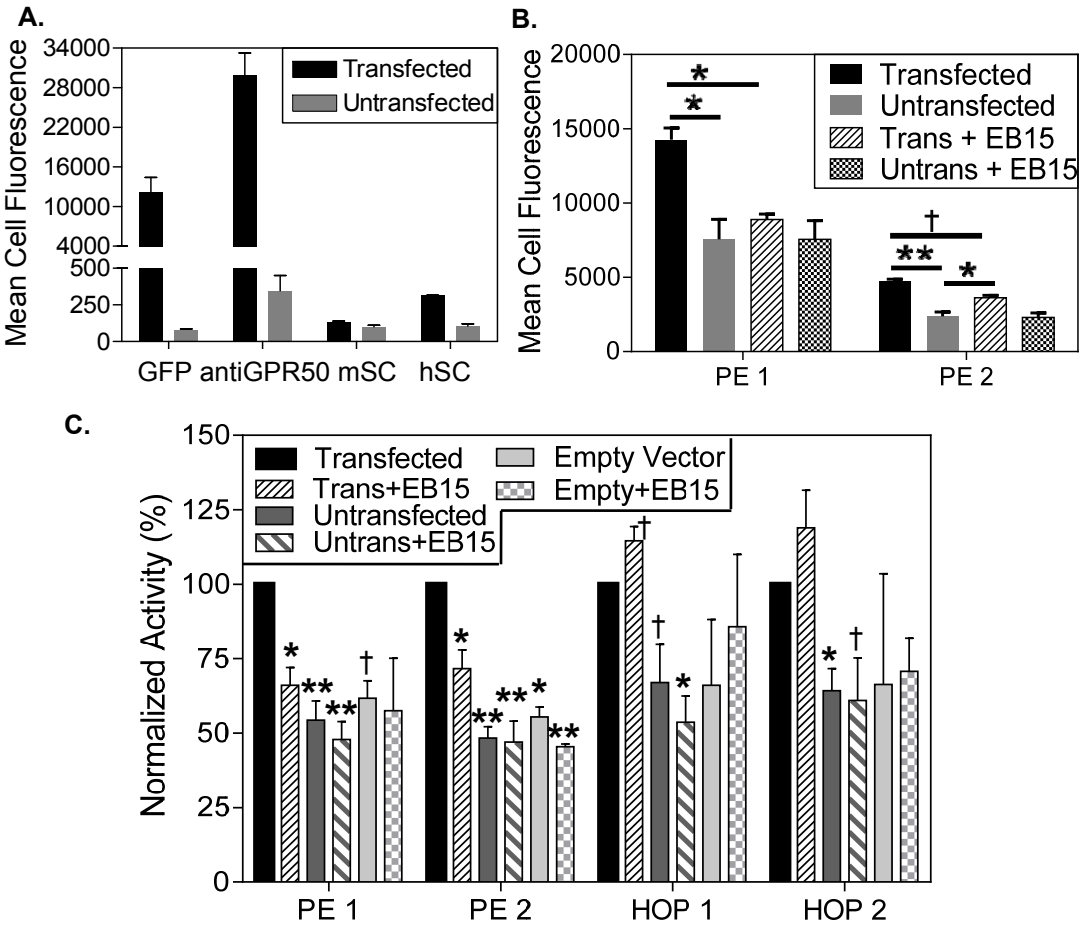

Figure 3-7: The EB15 epitope inhibits antibody binding to HEK293T cells expressing GPR50. (**A**) The GFP signal confirmed transfection of vector, while the anti-GPR50 monoclonal antibody verified specific expression of GPR50. Furthermore, the negative controls for secondary binding using anti-mIgG (mSC) for GPR50 monoclonal and anti-hIgG (hSC) for plasma samples showed minimal background binding activity. (**B**) GPR50 transfected HEK293T cells exhibited increased antibody binding to PE pools over the untransfected cells, which was significantly reduced by EB15. (n=5) * $p < 0.01$, ** $p < 0.001$, † - $p < 0.05$ for ANOVA and Tukey's multiple comparison analysis. (**C**) The empty vector transfection control performed similarly to the untransfected cells with the PE pools, showing reduced antibody binding compared to GPR50 transfected cells. However, while the evaluated HOP pools demonstrated increased antibody binding to GPR50 transfected cells compared to untransfected, the signals from the transfection control appeared much more variable. Importantly incubation with EB15 did not reduce the antibody binding signal to GPR50 transfected cells * $p < 0.01$, ** $p < 0.001$, † - $p < 0.05$ for one-sample t-test deviation from 100% (transfected cells).

The empty vector transfected cells also demonstrated significantly reduced antibody binding to the PE pools, similar to the untransfected cells (**Figure 3-7C**). In contrast, empty vector transfected cells exhibited high variability in antibody binding to the HOP pools, but compared to untransfected cells, HOP-associated plasma antibodies exhibited increased binding to GPR50 transfected cells. However, importantly, incubation with EB15 did not reduce HOP plasma antibody binding to the GPR50 transfected cells, indicating this variably increased activity was mediated through different antibody specificities than in the PE pools.

In addition to transient transfection analysis, an immortalized trophoblast cell line, HTR-8/SVneo, was assayed for binding to antibody fractions purified from GPR50+ PE and GPR50- HOP pools with or without EB15. First, GPR50 expression on HTR-8/SVneo cells was confirmed by flow cytometry using a GPR50 specific monoclonal antibody compared to a secondary only control (mSC) (**Figure 3-8A**). The EB15 peptide significantly reduced binding of PE antibodies to HTR-8/SVneo cells to levels comparable to HOP-associated antibodies as measured by flow cytometry (**Figure 3-8B**). Thus, the EB15 epitope specific antibodies present in PE were capable of binding to GPR50 expressed on the cell surface in transfected cells and a trophoblast-derived cell line.

To assess expression of GPR50 in human placentas,[192] placental tissue sections from PE patients were assayed by immunohistochemistry (IHC). Syncytiotrophoblast cells were strongly stained by GPR50 monoclonal antibody, but not by an isotype matched control (**Figure 3-9**). Thus, IHC confirmed the presence of the putative autoantigen GPR50 in the placenta.

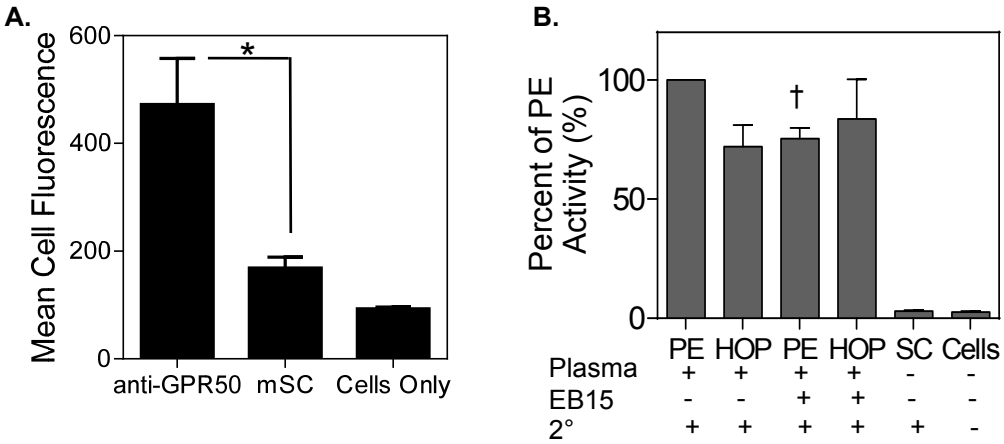

Figure 3-8: HTR-8/SVneo cells express GPR50 and EB15 regulates antibody binding to this trophoblast model cell line. (**A**) Cells incubated with the GPR50 specific monoclonal antibody demonstrated binding above background. (**B**) Additionally, EB15 reduced PE Ig binding to the natively expressed GPR50 in HTR-8/SVneo cells to a similar level as HOP Ig. * - $p < 0.03$ for U-test, † - $p < 0.05$ for one-sample t-test compared to 100%.

iii. EB15 and GPR50 fragments exhibit increased antibody binding activity in PE patients

The 15-mer fragments exhibited significantly higher binding activity among PE patients in a subset of the cohort. Antibody binding activity to bacterial displayed peptides in an expanded set of PE (n=42) and HOP (n=43) further assessed this differential activity. Additionally, 21 nulligravid samples were assayed for binding activity. EB15 and GPR50 fragments demonstrated significantly increased antibody binding to PE patients compared to HOP and nulligravid samples (**Figure 3-10**). Interestingly, no significant difference existed between nulligravid antibody binding activity and HOP. The GPR50 fragment exhibited a lower activity range than EB15 and reduced HOP antibody binding activity. Antibody binding activity appeared evenly distributed across early- and late-onset PE and adverse outcomes. These results highlighted that the PE condition is associated with significantly increased antibody binding activity to EB15 and GPR50.

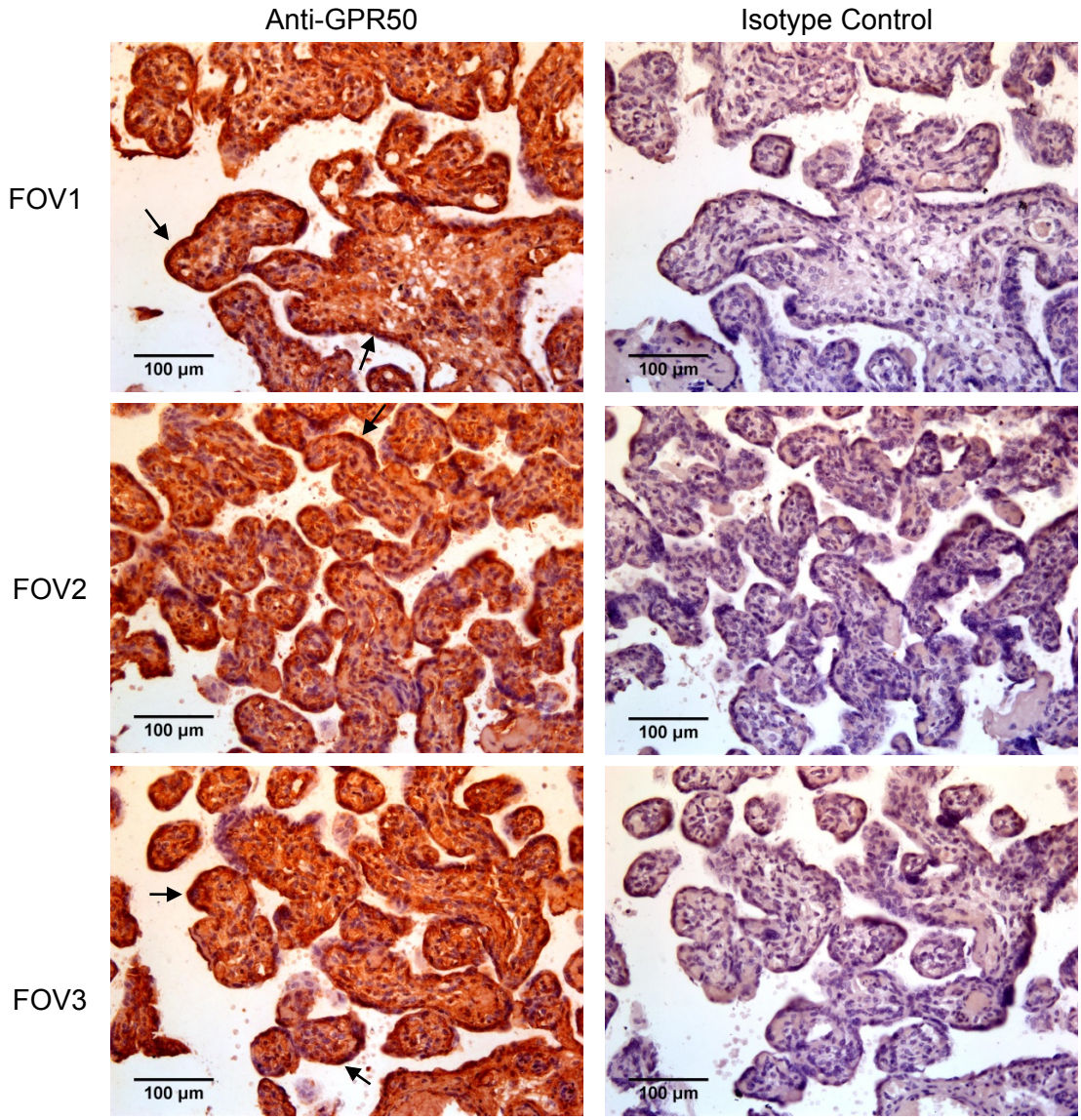

Figure 3-9: PE placental tissue expresses GPR50. Using the monoclonal antibody against GPR50 resulted in strong staining, with higher signal around syncytiotrophoblasts (arrows). This staining pattern was confirmed using three different fields of view (FOV) of the same tissue section

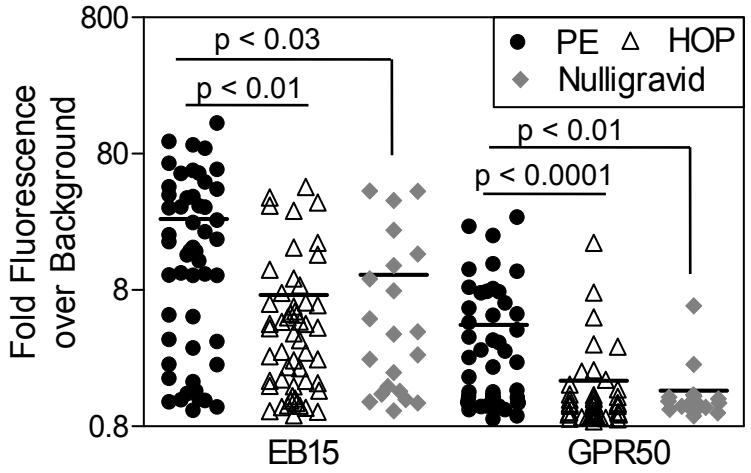

Figure 3-10: PE patients exhibit significantly increased antibody binding to EB15 and GPR50 compared to HOP and nulligravid. p-values represent Mann-Whitney U-test analysis

D. Discussion

Here, a PE-associated IgG1 antibody was identified that binds with high affinity to EBV protein EBNA-1 and cross-reacts with human GPR50. Previous studies have suggested a role for antibodies in the pathogenesis of PE. Most notably, IgG autoantibodies that agonize AT_1 (AT_1-AA) have been shown capable of inducing blood pressure elevation,[3] complement deposition[173] and other hallmarks of PE. In addition to AT_1-AAs, several other PE-associated antibodies have been reported, including those binding to β1, β2 and α1 adrenoreceptors,[128] cardiolipin,[129] and prothrombin.[130] We previously employed an unbiased discovery approach to investigate whether additional antibody specificities exist (**Chapter 2**)[191] and identify their preferred epitope binding specificities. This approach, using antibody fractions, indicated the presence of PE-associated antibodies distinct from AT_1-AAs, but did not reveal either autoantigens or environmental antigens. Given this problem, we applied a substantially modified discovery method termed Antibody Diagnostics via

Evolution of Peptides (ADEPt),[189] which has previously revealed specific environmental antigens in an unbiased fashion. The ADEPt method revealed a specific epitope within EBNA-1 as a target of the immune response in PE, and demonstrated that the same antibody species cross-reacts with placental GPR50. The cross-reactivity of a disease-associated antibody to regions of EBNA-1 and a human protein, GPR50, potentially holds further implications for PE pathology.

Our results provide evidence for a novel case of molecular mimicry, a mechanism proposed for a variety of other diseases. For example, in myasthenia gravis anti-acetylcholine receptor antibodies have been proposed to associate with herpes simplex virus glycoprotein D.[30] Peptides derived from the acetylcholine receptor and glycoprotein D inhibited antibody binding to the acetylcholine receptor.[33] Rheumatic fever/carditis patients possess cross-reactive antibodies to streptococcal *N*-acetyl-βD-glucosamine and cardiac myosin.[44] Furthermore, three regions of EBNA-1, one of which borders and partially overlaps with our EB15 fragment, bind lupus-associated antibodies.[28] An additional case for molecular mimicry in PE was proposed for the generation of AT_1-AAs, and a human IgG directed against parvovirus B19 VP2 demonstrated positivity in the AT_1-AA detection assay.[135] Thus, molecular mimicry represents a proposed mechanism of etiology for autoantibody activity in several diseases.

The Epstein-Barr virus (EBV), in spite of being ubiquitous in humans world-wide, has been associated with a variety of diseases. For example, EBV reactivation is linked to the development of Burkitt's lymphoma and nasopharyngeal carcinoma.[193] Interestingly, 30%[194] to 36%[195] of pregnant women show signs of

EBV reactivation based on serum antibody panels. Furthermore, the placental gene expression of the Epstein-Barr virus induced gene 3 (EBI3) appears increased in PE.[196] EBI3 comprises part of IL-27, which inhibits CD4+CD25+ regulatory T-cell development,[197] and regulatory T-cell levels are decreased in PE.[123,124] Therefore, reactivation could result in increased viral activity leading to a myriad of effects, including amplified antibody titers in a group of women that possess this antibody specificity ultimately leading to aberrant binding activity to human proteins, such as GPR50.

While the function of GPR50 is poorly understood, several observations relate to PE pathology. First, the presence of GPR50 in the model trophoblast cell line, HTR-8/SVneo cells, agrees with a previous finding of GPR50 presence in the placenta,[192] which we further confirmed. Second, since the PE-associated antibody represents an IgG1 subclass, which activates the complement system, antibody binding to this placental GPR50 could contribute to the increased complement deposits observed in the PE placenta.[119,122,173] Through dimerization, GPR50 inhibits the melatonin receptor 1,[198] and while melatonin exhibits a protective effect in the placenta[199] and helps regulate blood pressure,[200] the levels of melatonin and melatonin receptors 1 and 2 are reduced in PE.[201] Furthermore, mutations in GPR50 have been linked to increased triglyceride levels,[202] and hyperlipidemia often precedes PE presentation.[203] Taken together, these observations suggest that EBV-induced autoantibodies to GPR50 may be pathogenic in PE, and future studies should be designed to investigate this question.

Our finding that the GPR50 epitope was reactive with the largest number of PE specimens does not rule out the possibility that other, less similar human proteins might also bind to the EB15 directed antibody. Although we focused on cross-reactivity to GPR50 due to its increased binding activity compared to the other evaluated fragments, we observed moderate activity in some PE to other candidate targets. Therefore, cross-reactive binding to these and potentially other less similar autoantigens may occur. For example, kininogen-1 appears dysregulated in PE[204,205] and shares the CVGC sequence, potentially leading to low activity binding. Thus, aside from GPR50, other autoantigens could be recognized by this antibody species and contribute to the heterogeneity in presentation and severity of PE.

Bacterial displayed random peptide library screening enabled the impartial identification of an EBV directed antibody capable of cross-reacting with a human placental protein. Therefore, this study provides support for an aberrant immune response in PE through a mechanism of molecular mimicry involving a viral antigen (EBNA-1) and human proteins, especially the placental protein GPR50. Although this study focused on PE, this approach can be extended to a variety of diseases to impartially profile disease-associated antibody-antigen interactions.

4. Using next-generation sequencing to characterize individual patient antibody binding specificities

Next-generation sequencing (NGS) has significantly increased throughput (>5 million sequences) compared to traditional Sanger sequencing methods. Thus, NGS enables in depth analysis for a variety of applications, including evaluation of isolated sequences from peptide library display screening. In particular, multiple individual screens can be probed simultaneously without excessive numbers of sorting/selection rounds to reduce diversity. We sought to develop and apply a unique method of bacterial displayed peptide library screening with NGS to profile individual antibody repertoires from pre-eclampsia (PE) patients (n=4) and healthy-outcome pregnancies (n=4). Validating this methodology, we re-identified the previous viral antigen-linked PE-associated motif (KXXXC[VIL]GC). An additional PE-enriched motif (GXXGAGGG) supports an amplified immune response to the Epstein-Barr virus nuclear antigen 1. This unique approach simultaneously identified PE- and healthy pregnancy-associated motifs, highlighting an altered antibody response among women with PE.

A. Introduction

The advent of next-generation sequencing (NGS) has enabled high-throughput evaluation of millions of sequences, providing a rich source of data. This tool has tremendous potential to change genomics and clinical diagnostics as we quickly approach the $1000 personal genome.[206] Thus far, current diverse applications of this

technology include evaluation of phage display library panning,[207,208] gene mutation analysis,[209] influenza inhibitor design,[210] and DNA aptamer selection.[211] The ability to sequence >400 billion oligonucleotide bases enables full coverage of T- and B-cells from small organisms, such as zebrafish.[212] Although the full repertoire in humans has yet to be completely understood, certain insights have been gained. Profiling the antibody repertoire in humans by sequencing circulating B-cells provided evidence for B-cell memory recall.[53] Immune repertoire sequencing has revealed alterations in multiple sclerosis,[186] leukemia,[187] and severe aplastic anemia.[188] Furthermore, NGS has provided insights into ongoing immune responses through specific isolation and sequencing of antibody-producing plasmablasts.[51,52,213] NGS enables an enhanced understanding of the immune repertoire diversity, and it has been used to estimate the naïve T-cell repertoire diversity as 3-4x10^6 (three to four-fold higher than previously expected).[214] As these studies demonstrate, NGS represents an invaluable tool for probing the immune repertoire.

Given this useful tool, we sought to develop and apply a method to profile antibody binding specificities using NGS and bacterial displayed peptide library screening. In previous work, peptide library screening incorporated pools of disease and healthy samples and required multiple rounds of sorting to reduce library diversity from 8x10^9 to ~1000 members for traditional sequencing (**Chapters 2 and 3**).[153,189,191] However, one study demonstrated that the most abundant sequences identified after four phage selection rounds were among the most abundant in the 1st or 2nd round, but the high diversity at these rounds prevents identification by traditional sequencing (~50 clones).[215] Thus, we aimed to characterize antibody

binding peptides isolated from early screening rounds against individual samples at higher library diversity (10^4-10^5 members). We hypothesized that this approach would greatly reduce the number of sort rounds while providing an in-depth profile of the antibody binding repertoire for each individual. Instead of focusing on the highly enriched sequences that bind antibodies present in disease plasma,[68,208] a computational algorithm developed in the Daugherty Group at UCSB (Pantazes R, Reifert J, Elliott SE, et al. 2014, unpublished), compares the disease- and control-associated sequences. This algorithm evaluated the presence of enriched patterns of amino acid sequences instead of full-length peptides,[68,183] which may reduce the number of unique contributors when identifying motifs. Applying this methodology to the pregnancy-related disease, pre-eclampsia (PE), confirmed the previously characterized viral antigen-linked motif (**Chapter 3**) and elucidated additional PE- and healthy-outcome pregnancy (HOP)-associated motifs. Our results show that NGS-enabled antibody binding motif analysis yields broad insights into disease-related alterations to the antibody repertoire.

B. Materials and Methods

i. Patient samples

A heterogeneous set of PE and HOP (n=4, each) were selected from the same cohorts previously described (**Chapters 2 and 3**). In addition to clinical parameters, the samples showed a range of average antibody binding activity to the previously identified and tested antigen fragments from Chapter 3 (**Table 4-1**).

Table 4-1: Characteristics of samples for next-generation sequencing analysis

	Average Antibody Binding Activity		Max Blood Pressure		Proteinuria	WGA	Elevated Enzyme Levels	CNS
ID	EB15	GPR50	SBP	DBP	Y or N	wk	Y or N	Y or N
PE-1	26	5.3	152	99	Y	37.5	Y	N
PE-2	1.8	1.3	155	90	ND	36	N	Y
PE-3	15	6.3	148	102	N	32	N	Y
PE-4	1.7	0.89	167	90	Y	36	Y	Y
HOP-1	1.3	0.90	120	55	N	40.1	ND	N
HOP-2	1.5	1.2	134	70	ND	38.6	ND	N
HOP-3	31	1.3	111	66	N	38.5	ND	N
HOP-4	6.2	1.0	98	59	ND	38	ND	N

Average antibody binding of samples from previous studies expressed as fold fluorescence over background. Abbreviations: ND – not determined, SBP – systolic blood pressure, DBP, diastolic blood pressure, WGA – Weeks' gestational age at delivery, CNS – central nervous system.

ii. Library screening

A fully randomized, 15-mer peptide library displayed on eCPX[90] was screened against diluted plasma from individual PE and HOP. Samples were diluted (1:50) and depleted of *E. coli* binding antibodies prior to library incubation as previously described (**Chapter 3**). A magnetic selection step quickly removed irrelevant peptides in preparation for fluorescence activated cell sorting (FACS). A biotin-conjugated anti-human IgG secondary (1:500) (Jackson ImmunoResearch) was used to detect bound IgG following incubation with diluted plasma (1:100). Finally, cells were labeled with streptavidin (SA)-conjugated magnetic beads and application of a magnet enabled isolation of antibody bound cells. Two rounds of FACS further enriched the peptide library for each individual sample. In this case, two concentrations of diluted plasma (1:100 and 1:500) were used. The same biotinylated anti-human IgG secondary (1:500) was used, but SA conjugated to R-phycoerythrin (SA-PE) fluorescently labeled the cells expressing peptides bound to antibodies. Labeling reagents were diluted in PBS with 0.1% BSA and incubations conducted at

4°C. Sort gates were set according to a similarly treated negative control expressing only the display scaffold. The first round of FACS (F1) collected the top 4% of binders at 1:100 plasma dilution, the second round (F2) isolated the higher fluorescing cells from the total binding population observed with a 1:500 plasma dilution. To compare this strategy to a higher stringency single FACS sort, one population following magnetic selection (PE-1) was sorted against 1:500 diluted plasma. Incubating the library with secondary and SA-PE only (no plasma) evaluated the presence of false-positives binding to secondary or SA alone.

iii. Library preparation for sequencing

Following library screening, the plasmids were isolated from cells using Zyppy™ Plasmid Miniprep kit (Zymo Research). In addition to the four PE and HOP after two rounds of FACS, populations from other sort rounds with PE-1, PE-3, HOP-2, and HOP-4 were included for sequencing. The peptide encoding regions of the plasmid were amplified and prepared for sequencing (**Figure 4-1**) on the Illumina platform, MiSeq. Specifically, a two-step PCR process modified from an existing protocol[216] ensured amplification of the peptide region and addition of the required flanking adapter and barcoding sequences. The initial primers (For1 and Rev1) amplified the peptide encoding section of the plasmid while adding annealing regions for the second set of primers (**Table 4-2**). In the second PCR, primers from the Nextera® XT Index Kit (Illumina) added the adapter sequences and barcodes (i5 and i7) for dual-indexing. The first PCR underwent 25 cycles while the second used 8 cycles with annealing temperatures of 65°C and 62°C, respectively. This two-step PCR resulted in 12 different samples, multiplexed by 12 unique i7 barcodes with the same i5 index.

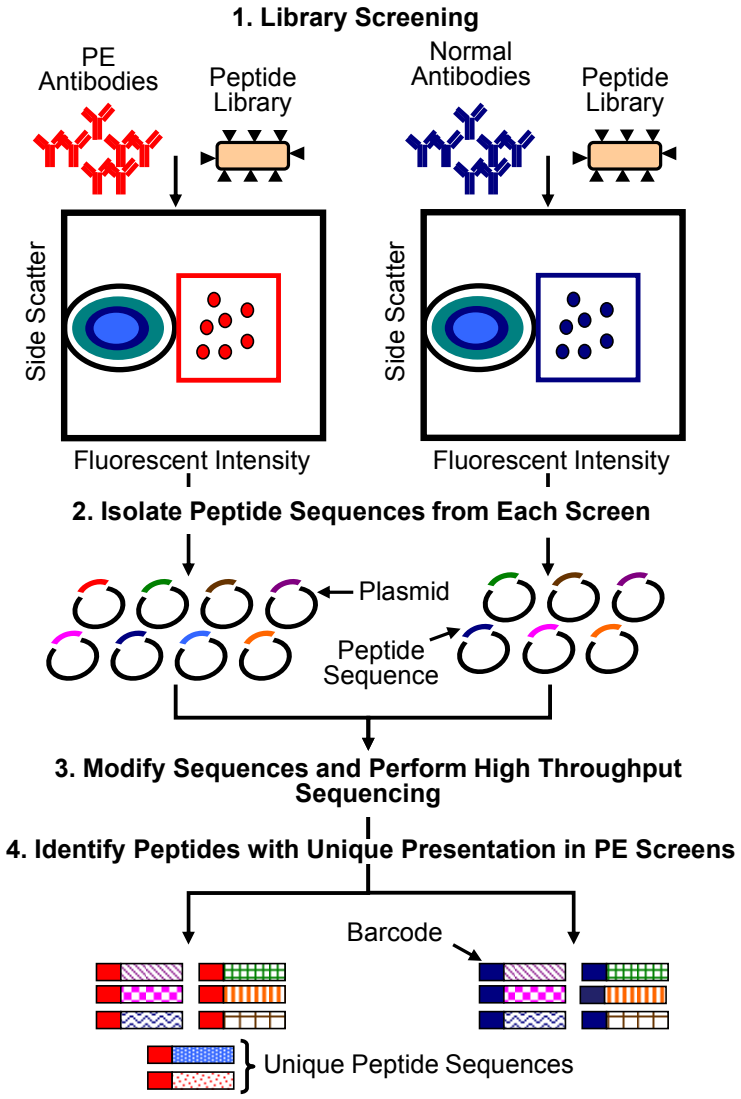

Figure 4-1: Library screening and preparation for next-generation sequencing. After initial library screening, peptide encoding regions are amplified and modified to include adapter sequences, barcodes, and sections required for sequencing. Finally by de-multiplexing and comparing the individual PE and HOP we can identify unique peptide sequences and patterns among PE.

The 12 unique samples were pooled together in equal amounts following quantification by Qbit, using an estimated length of 214 bases and Equation 4.1. After diluting each sample to 4 nM, 5 µL of each was removed and pooled together. An Agilent Bioanalyzer run evaluated the size and purity of the final pool.

$$\frac{\text{concentration (ng/}\mu\text{L)}}{660\,\text{(g/mol)}\times 200\,\text{bp}}\times 10^{6} = \text{concentration (nM)} \qquad (4.1)$$

Table 4-2: Primers used for library preparation

Primer	Sequence
For1:	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCCAGTCTGGCCAG
Rev1:	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGCCACCAGAACCGCC

Shaded region represents the complementary sequence for annealing to the desired regions of the plasmid for peptide encoding section amplification.

iv. Motif characterization

A computational algorithm, termed Identifying Motifs Using Next-generation sequencing Experiments (IMUNE), developed in the Daugherty Group at UCSB (Pantazes R, Reifert J, Elliott SE, et al. 2014, unpublished) was applied to the de-multiplexed samples to identify patterns of amino acids significantly enriched in the PE samples compared to the HOPs. Different enrichment conditions were evaluated across the 8 samples. These conditions include: 4_0, 4_1, 3_0, 3_1, 2_0, and 2_1, indicating the number of PE and HOP, respectively, with a given statistically enriched pattern. Related patterns identified with these different enrichment conditions were combined to identify common motifs. A Python code (MotifSearcher) enabled searches for sequences comprising a specific motif from the list of unique peptides to generate a traditional sequence logo. This analysis determined the number of peptides significantly (≥95% of observations) associated with PE samples out of the total number of sequences containing the searched motif. Specifically, if a given peptide sequence was observed 19 out of 20 times in one of the four PE patients, it was considered PE-associated. At the same time, the control- or HOP-specific motifs were similarly characterized.

For additional motif analysis, the web-based MEME algorithm[98] was employed. Comparing sequence observations after two FACS rounds (F2) to those in the first round (F1) identified the enriched sequences, or those representing a higher fraction

of the total sequences in F2 than F1. Cases in which a sequence appeared >10 times in F2 but never observed in F1 were also included. These enriched sequences were ranked according to the number of observations in the F2 population. The MEME algorithm characterized five motifs from the top 1000 ranked sequences. To characterize the effect of sort stringency, the single FACS round against PE-1 with 1:500 diluted plasma was compared to the population that underwent a less stringent F1 (1:100 dilution) followed by a second sort with 1:500 diluted plasma.

v. Candidate antigen identification

To relate generated motifs to candidate antigens containing the same amino acid sequence, a variety of search algorithms were employed. Primarily, NCBI BLASTp and ScanProsite determined the proteins with similarities/identities to variations of motifs. Additionally, the Immune Epitope Database was searched for sequences with positive antibody binding activity. Where appropriate, hits from organisms, such as western clawed frog, fruit fly, or extreme thermophiles, that would be rarely involved in the human immune system were removed from search results. Additionally, uncharacterized or putative protein hits were not included.

C. Results

i. Screening peptide display library for patient-specific binders

The first round of magnetic selection reduced the library population diversity from $8x10^9$ to ~10^6 members. This reduced diversity enabled FACS-based sorting for the subsequent rounds. The different PE and HOP samples enriched for binders at different rates (**Figures 4-2 and 4-3**). Only one sample (PE-2) exhibited less than

75% binding after two rounds of FACS (F1 and F2) at 63.4% binding, despite exhibiting the highest percent binding after the initial magnetic selection. Interestingly, library screening against PE-1 and PE-3 enriched for the highest percent binding out of the four PE samples and these samples previously showed increased binding to antigen fragments EB15 and GPR50. Additionally, peptide library screening with HOP-3, which showed the highest antibody binding activity against EB15, enriched the fastest. For the other samples, less than 30% of the population bound to sample antibodies after one round of FACS; however, HOP-3 exhibited 73.5% binding. Comparatively, the stringent F1 sort against PE-1 (1:500), resulted in a higher percentage (56%) of binders post F1 than when sorted using a 1:100 dilution (21.5%). This indicated that increased stringency enhanced enrichment of strong binders; however, both populations exhibited a bimodal distribution of binding peptides.

Importantly, the majority of peptide sequences in each library population following two FACS rounds bound to the specific sample being screened against, reducing the background noise of nonbinding peptides during sequence analysis. The lack of streptavidin binders was also confirmed. Estimated diversities of ~10^4 members enabled pooling of the individual screens for multiplexed sequencing by MiSeq. After separate PCR amplification using a two-step process, the DNA concentration of each pool was measured to facilitate equimolar pooling. The Agilent Bioanalyzer confirmed the correct size and purity of the pooled DNA.

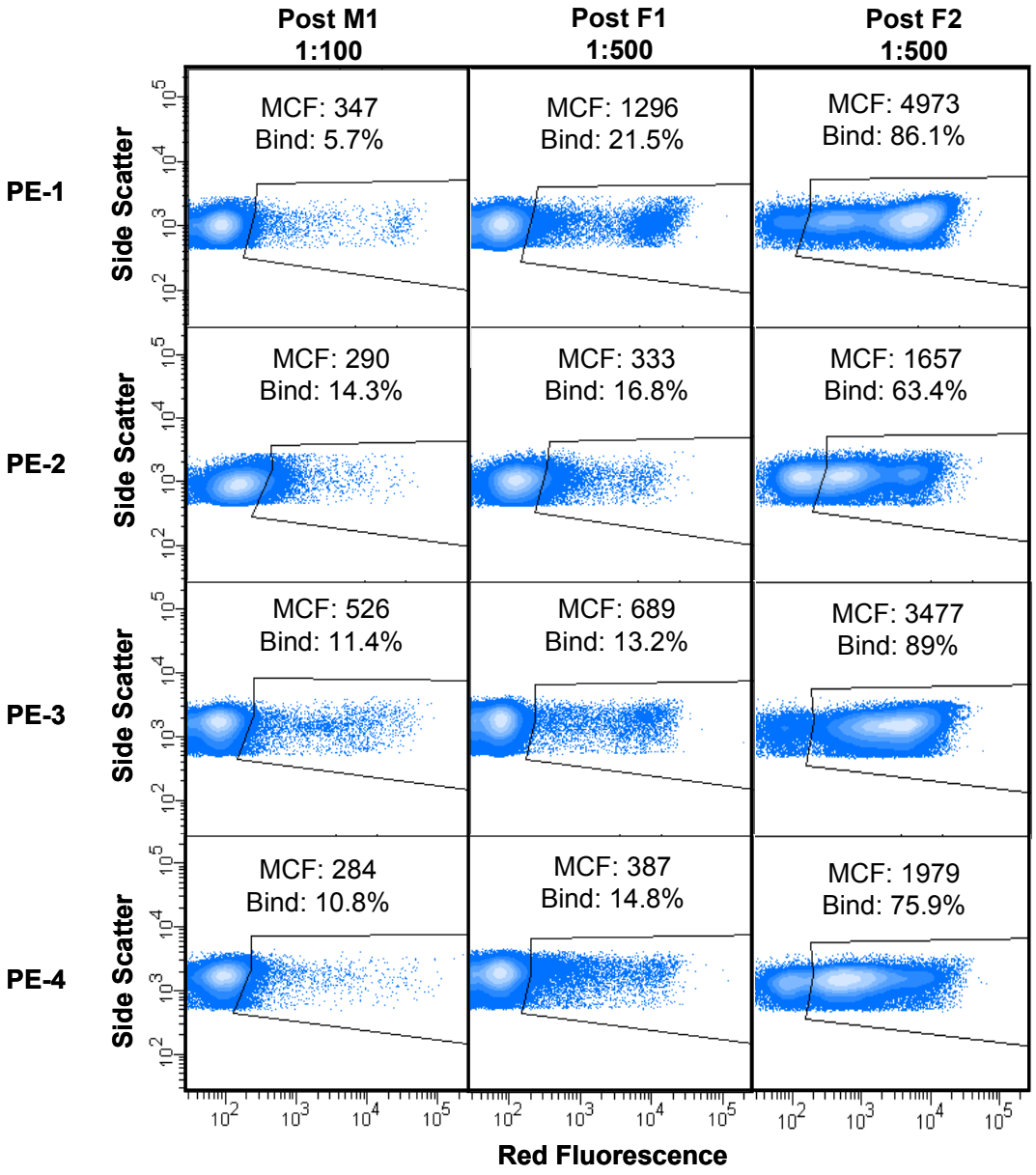

Figure 4-2: Individual screens enriched the library against different PE samples. Each distinct screen resulted in a majority population of cells expressing peptides that bind to unique PE.

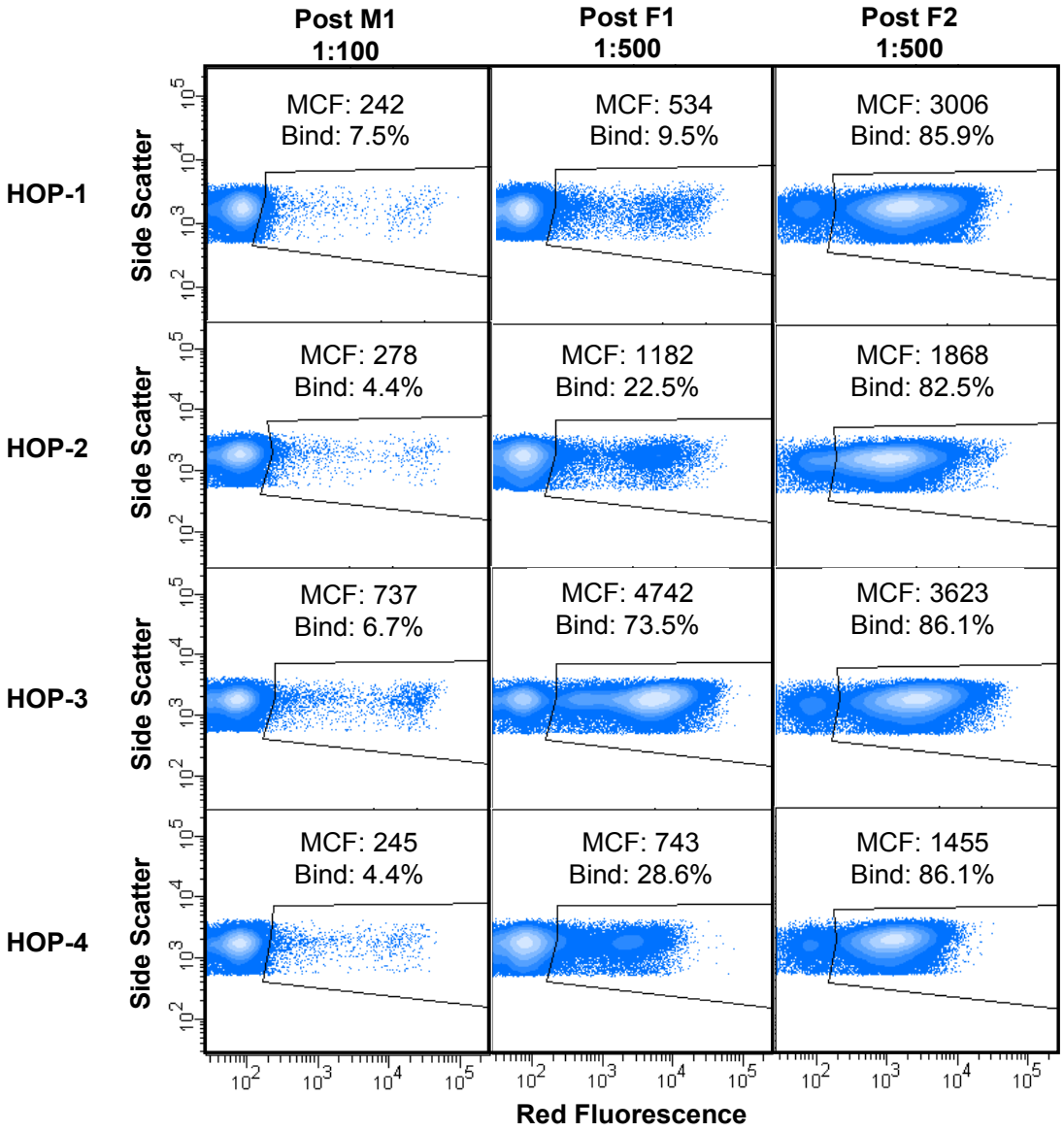

Figure 4-3: Individual screens enriched the library against different HOP samples. Each distinct screen resulted in a majority population of cells expressing peptides that bind to unique HOP.

ii. Sequence analysis identifies the viral antigen-linked motif among other PE-associated motifs

From MiSeq sequencing, a total of ~7x10^6 usable reads were identified. The redundant set of peptides was normalized to ~7.3x10^5 unique sequences across all twelve samples. The sequence counts for individual screens varied (**Table 4-3**). Despite the reduced number of total usable reads in the post F1 populations, these all

exhibited higher unique sequences than the corresponding post F2 populations. This reflects the higher diversity among the post F1 populations. Due to the reduced coverage (fold oversample), a significant portion (~40-70%) of sequences from post F2 was not observed in the post F1 populations.

Table 4-3: Number of sequence reads for each sample

Sample ID	Unique	Total	Fold Oversample
PE-1: Post F2	48,464	337,992	6.97
PE-2: Post F2	77,657	715,905	9.22
PE-3: Post F2	29,434	1,046,613	35.6
PE-4: Post F2	40,877	171,159	4.19
HOP-1: Post F2	34,900	658,011	18.9
HOP-2: Post F2	136,517	677,586	4.96
HOP-3: Post F2	69,488	466,980	6.72
HOP-4:Post F2	61,179	942,076	15.4
PE-1: Post F1 (1:500)	58,474	222,265	3.8
PE-3: Post F1	62,390	257,594	4.13
HOP-2: Post F1	177,324	510,591	2.88
HOP-4: Post F1	105,199	191,298	1.82

Applying the IMUNE algorithm identified a redundant list of statistically enriched patterns for a variety of comparisons. In these patterns, an X indicated that no amino acid was significantly enriched at that position. Combining similar patterns led to a general motif for different sets of patterns. Observed in different enrichment conditions, one dominating pattern set (>50 unique patterns) contained the previously identified viral antigen (EBNA-1)-linked motif from Chapter 3, KXXXC[VLI]GCK (**Figure 4-4**). Aligning these enriched patterns resulted in a general binding motif similar to that initially identified using sequential enrichment and depletion of plasma pools (**Chapter 3**). Based upon this pattern alignment, a search of KXXXCXGC in the list of unique sequences yielded the full set of 15 amino acid peptides containing motif 1. A majority of these peptides (80%) was identified as significantly associated with PE samples and used to generate a sequence logo. Although left undefined

during the search, amino acids valine, isoleucine, and leucine showed increased representation in the middle position, similar to the first generation motif previously identified (**Figure 3-2**). Also similar to this first generation motif is the lack of a C-terminal lysine that showed enrichment after directed evolution (**Chapter 3**).

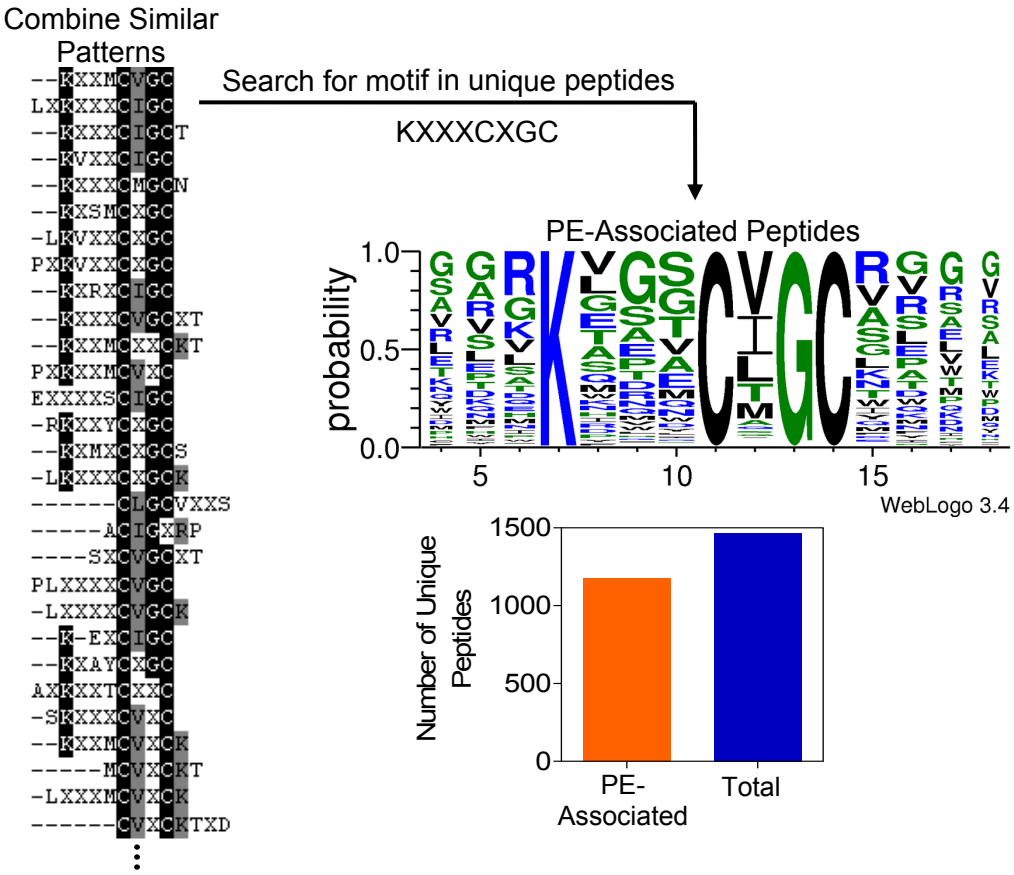

Figure 4-4: Identifying the PE-associated peptides containing the viral antigen-linked motif. Several patterns observed to be significantly enriched in PE patients screened represented the previously identified motif. After searching the unique peptides for this motif, the number of PE-associated peptides were identified and aligned to generate a sequence logo.

In evaluating the other highly represented patterns, another set was identified as enriched in three PE and only one HOP. Evaluating the similarities from these ~35 patterns elucidated a second PE-associated motif ([LME][YW]X[WFY]DX[RK]) for searching the unique sequence list (**Figure 4-5**). Interestingly, this shared some similarities with motif c, WGWGXX[RK], identified in a previous antibody profiling study using antibody fractions (**Chapter 2**).[191]

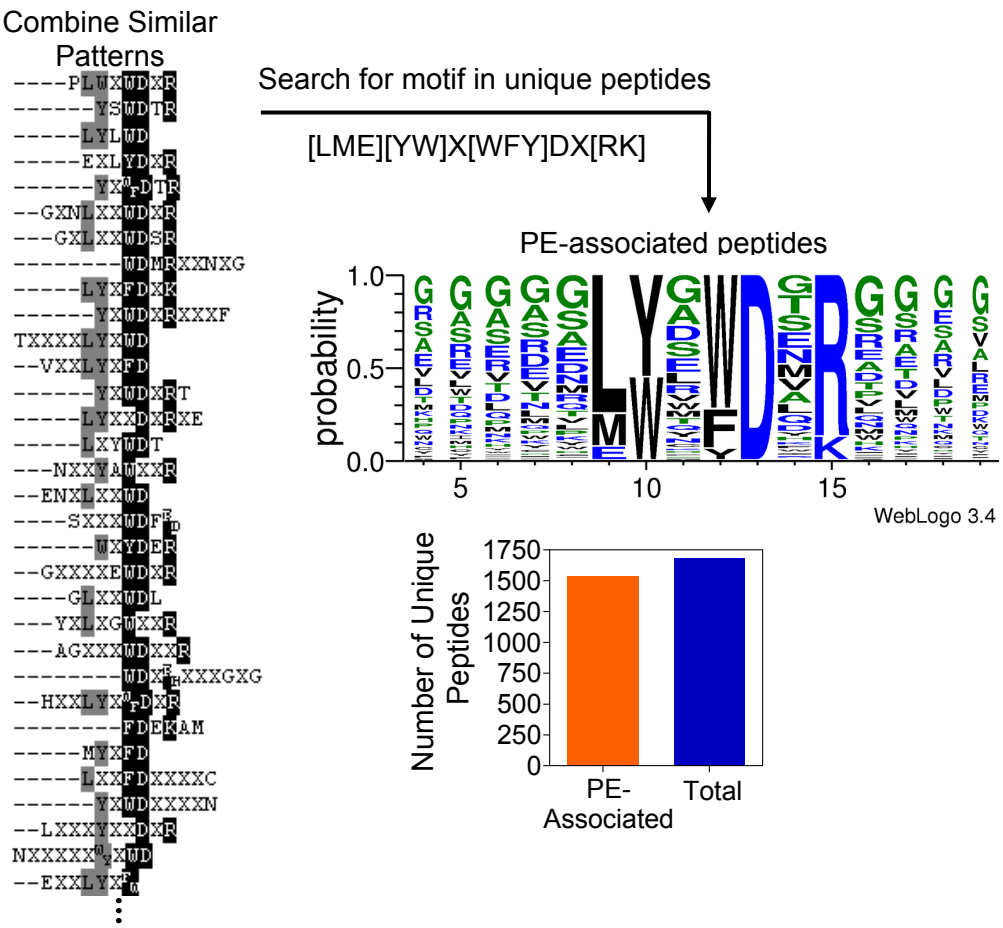

Figure 4-5: Characterizing a new PE-associated motif. Similar patterns found to be statistically enriched in three PE were combined to characterize the motif. Searching the set of unique peptides identified the PE-enriched sequences for logo generation and determined the number of PE-associated peptides.

Motif 1 comprised four fixed residues, while motif 2 shared five positions but only one was completely fixed. Despite the increased number of unique sequences containing motif 2 (1683) than motif 1 (1465), motif 2 exhibited a higher percentage of PE-associated peptides (91%). This aligned with the observation that these sequences predominantly stemmed from 3_1 enriched patterns. Additionally, since two of the HOP samples used in this analysis possessed antibodies capable of binding to EB15, these contribute to the reduced percentage of PE-associated peptides observed in motif 1 compared to motif 2.

Another motif initially consisted of three defined positions (GAG). Almost 25,000 sequences possessed this motif, but only 75% of these occurred ≥95% of the time in PE samples. However, using the generated sequence logo (**Figure 4-6A**) to define additional residues (GXXGAGGG), the search returned fewer peptides but these exhibited a higher enrichment among PE patients (**Figure 4-6B**). Thus, motif 3 consisted of 723 disease-enriched sequences out of 793 total (91%). This highlighted the importance of refining motifs to understand PE-associated epitopes.

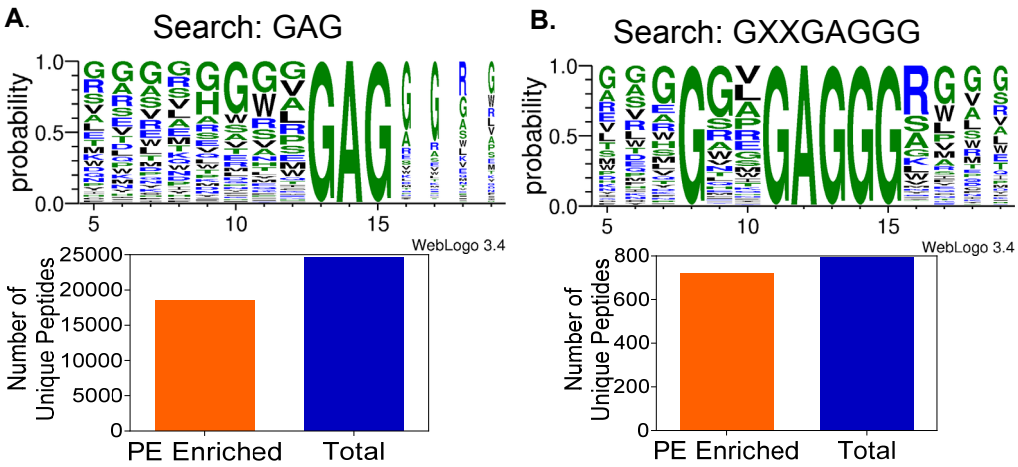

Figure 4-6: Defining additional motif positions increases PE-associated enrichment. (**A**) Using only three defined positions identified many sequences; however, a substantial percentage of these were not enriched in PE. (**B**) Using six defined positions decreased the total sequence count to ~800, but 91% of these were highly enriched in PE.

The motifs identified by patterns significantly enriched among PE samples according to the IMUNE algorithm exhibited a high percentage of disease-associated unique peptides. For comparison, a number of patterns identified by the algorithm from the 2_1 comparisons with reduced enrichment in PE compared to HOP were evaluated for PE-associated peptides. The patterns with low enrichment according to IMUNE exhibited similar numbers of disease and control peptides (**Figure 4-7**), indicating the significance of IMUNE-identified enriched patterns. In fact, one motif (ASXTXW) consisted of more HOP-associated peptides.

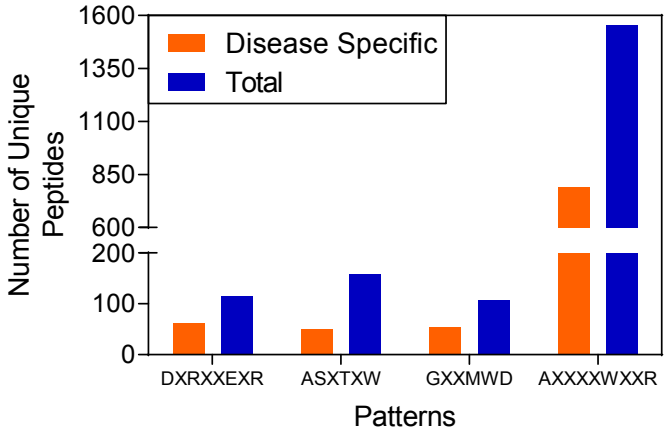

Figure 4-7: Number of peptides containing patterns with low PE enrichment compared to HOP according to IMUNE.

iii. Candidate antigens determined from disease-associated motifs

Focusing on the most represented amino acids as indicated by the sequence logo (L[YW]XWDXR) resulted in reduced total sequences (720) but increased specificity (93%). Searching this motif in ScanProsite identified a number of interesting candidate antigen hits. Initially searching the human proteome returned five unique protein hits (**Table 4-4**). Among nonhuman hits, several represent homologues to these human proteins in mice and rats. Interestingly, a prominent reoccurring hit (~580 strains) represented a protein from *Acinetobacter baumannii*, an opportunistic pathogen associated with hospital-derived infection.

Table 4-4: Human protein hits for prominent next-generation sequencing motif L[YW]XWDXR

Human Protein	Sequence
Calmodulin-binding transcription activator 1	AAVVLYKWDRRAISI
Protein cordon-bleu.	GIKELYAWDNRRETF
Elongation factor G, mitochondrial	FLPLLWNWDRRSGSQ
Insulin-like growth factor 1 receptor.	NLQQLWDWDHRNLTI
Prostaglandin F2-alpha receptor Isoform 5	LQMRLWTWDFRVNAL

Motif 3 (GXXGAGGG) corresponded to a common sequence found in a variety of viruses and pathogens, including the human adenovirus, Epstein-Barr virus (EBV), and torque teno virus. Searching the Immune Epitope Database returned 36 positive

hits for antibody-antigen binding. Many of these hits represented the same organism or multiple studies identifying similar sequences. Thus, these hits were condensed to six unique organisms, including EBV and a section of EBNA-1 that represents the immunodominant region[217] (**Table 4-5**). Although this motif may be present among the HOP samples it appeared significantly enriched among these PE.

These IMUNE-identified PE-enriched motifs enabled candidate human and environmental antigen discovery. Importantly, this in-depth antibody repertoire profiling approach distinguished two distinct motifs (KXXXC[VIL]GC and GXXGAGGG) from the same viral antigen, EBNA-1. This discovery highlights an altered antibody response to this antigen among PE patients.

Table 4-5: Positive antibody binding hits in the Immune Epitope Database for motif GXXGAGGG

Sequence	Protein	Organism
GAGGGAGAGGAGAGGGGRGR GAGGGAGGAGAGGGAGGAGA AGGAGAGGGAGAGGA	Epstein-Barr nuclear antigen 1 (EBNA-1)	Human herpesvirus 4
SSSSAGGGGGGAGGGGGGGG	early phosphoprotein P34	Human herpesvirus 4
MTSVNSAEASTGAGGGGSNSVK	Probable coat protein VP1	Human parvovirus B19
TGAGGGGSNSVKSMWSEGATFS	VP2	Human parvovirus B19
SGGGAGGGSSGSGQSGVDLSPV	Adhesin P1 precursor (Cytadhesin P1) (Attachment protein)	Mycoplasma pneumoniae
LKESGVKPGQFAAIVGAGGGL	alcohol dehydrogenase	Curvularia lunata
GLSLSAPSTGAGGGLPGPG	Myelin basic protein	Cavia porcellus
LGGGAGGGGDHADKLSLYSETDS	tripartite motif-containing 67	Homo sapiens

iv. Identifying control-specific motifs and related candidate antigens

In addition to discovery of PE-associated motifs, this methodology of individual screening followed by cross-examination of PE- and HOP-enriched patterns, enabled

identification of control- or HOP-specific motifs. To focus on truly HOP-specific motifs, the patterns enriched in four HOP and zero or one PE samples were evaluated. This analysis led to three main motifs identified as highly enriched among HOP samples (**Figure 4-8 A-C**). The HOP-specific sequences comprising these motifs represented 95.7%, 97.7%, and 92.9%, respectively, of total sequences. These control-associated specificities may be present among PE but reduced compared to the PE-associated motifs identified. Control-specific motifs highlighted a skewed antibody repertoire among the PE samples.

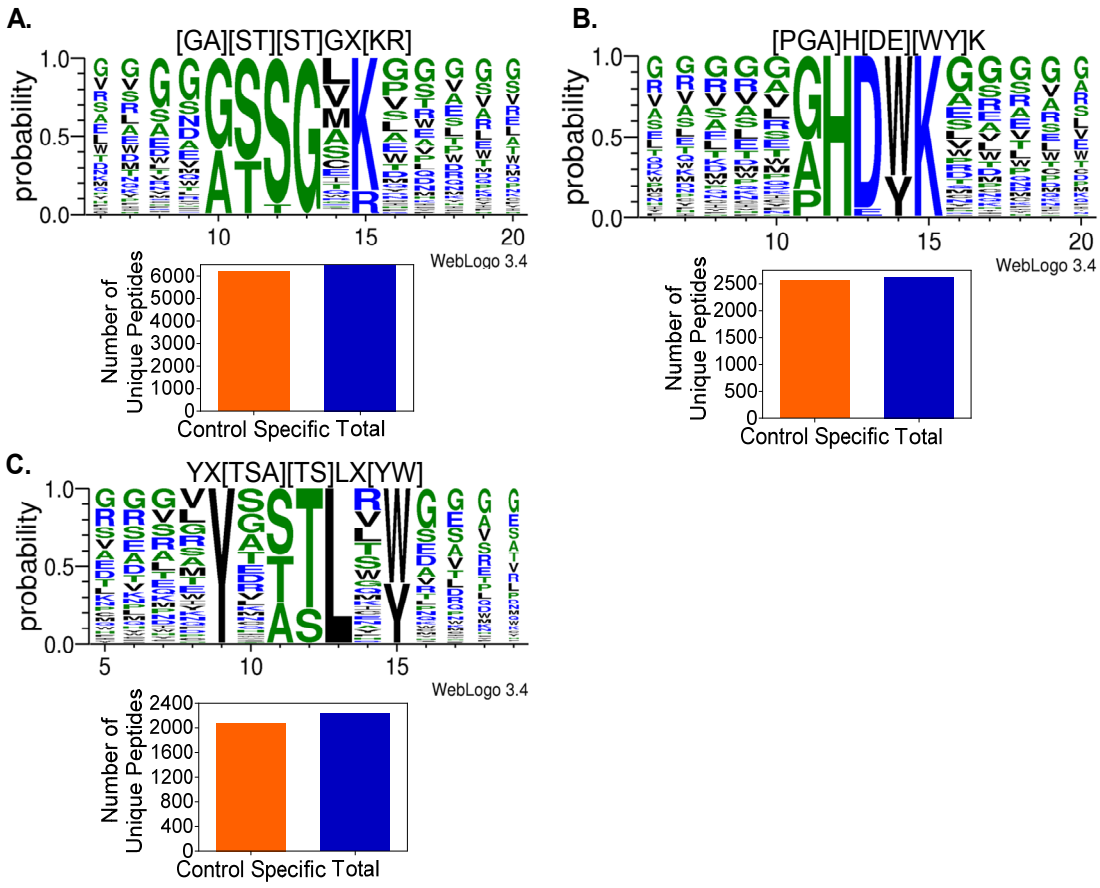

Figure 4-8: Control-specific motifs identified using the IMUNE algorithm. (**A**) One motif comprised the largest number of sequences; however, (**B**) (**C**) other motifs exhibited higher sequence complexity. All motifs exhibited high specificity (>92%) for control samples.

One motif (**Figure 4-8A**) comprised more sequences than the others; however, it mainly consisted of the highly represented amino acids glycine and serine. Therefore,

searches of the protein database using BLASTp did not return any significant hits and ScanProsite identified greater than 900 hits. For the other two motifs, the set of protein hits and corresponding organisms were evaluated and summarized. The control motif [PGA]HD[WY]K exhibited the greatest sequence complexity and resulted in the fewest hits (**Table 4-6**). Many of these hits include pathogenic viruses, such as human cytomegalovirus, or commonly encountered organisms like *Saccharomyces cerevisiae*. No human protein hits were identified for this motif using ScanProsite. In contrast, motif YX[TSA][TS]LX[WY] resulted in a large variety of hits, including several human proteins (**Table 4-7**) and various pathogens (**Table 4-8**). For some of these human hits, homologous proteins were identified in other organisms, such as primates, horses, rabbits, and mice.

Table 4-6: Candidate antigens for HOP-specific motif [PGA]HD[WY]K

Sequence	Protein	Organism
PHDYK	Cellobiose 2-epimerase	*Ruminococcus albus*
AHDYK	D-alanine--D-alanine ligase	*Clostridium novyi*
PHDWK	Envelope glycoprotein H	*Human cytomegalovirus*
AHDYK	Glycerol kinase	*Francisella philomiragia*
PHDYK	Hyaluronan synthase	*Streptococcus pyogenes*
AHDYK	Apolipoprotein N-acyltransferase	*Salmonella typhimurium*
PHDWK	Protein LST4	*Kluyveromyces lactis*
AHDYK	Nucleotide exchange factor SIL1	
GHDYK	Ribonuclease alpha-sarcin	*Aspergillus giganteus*
AHDYK	Ribonuclease H	*Saccharomyces cerevisiae*
GHDWK	Exosome complex exonuclease DIS3	
AHDYK	Nucleotide exchange factor SIL1	
GHDYK	Ribonuclease mitogillin	*Neosartorya fumigata*
GHDYK	Ribonuclease mitogillin	*Aspergillus restrictus*
AHDYK	Ubiquitin-like protein-NEDD8-like protein RUB3	*Oryza sativa subsp. japonica*
AHDYK	ATP-dependent RNA helicase SrmB	*Haemophilus influenzae*
AHDYK	TonB-dependent heme receptor A	*Haemophilus ducreyi*
GHDYK	UvrABC system protein B	*Ureaplasma parvum serovar 3*
AHDWK	Glycoprotein G	*Rabies virus*

Table 4-7: Candidate antigens from the human proteome for HOP-specific motif YX[TSA]TLX[WY]

Sequence	Protein
YISTLPY	A-kinase anchor protein 1, mitochondrial
YNSTLTW	Wolframin
YIATLLY	Myelin and lymphocyte protein
YLSTLRW	Cystic fibrosis transmembrane conductance regulator.
YFSTLDY	Cation channel sperm-associated protein 3
YLATLTW	EGF-like module-containing mucin-like hormone receptor-like 2
YKSTLGY	DnaJ homolog subfamily C member 18
YTTTLDY	Protein FAM101B
YLTTLAW	Olfactory receptor 5A1
YLSTLLY	Bardet-Biedl syndrome 12 protein
YSSTLRW	UPF0378 protein KIAA0100
YQSTLPW	Sushi domain-containing protein 5
YASTLGY	Transcription factor SOX-14
YLSTLYY	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
YVATLDY	Protein transport protein Sec24D
YVTTLLY	5'-AMP-activated protein kinase subunit beta-1

Table 4-8: Candidate environmental antigens for HOP-specific motif YX[TSA]TLX[WY]

Sequence	Protein	Organism
YVATLVY	Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase	*Emericella nidulans*
YETTLTY	ATP-dependent helicase/deoxyribonuclease subunit B	*Bacillus thuringiensis, B. weihenstephanensis, B. cereus, B. anthracis*
YGATLRY	Protein FimA	*Bordetella pertussis*
YWATLSY	Major cardiolipin synthase ClsA	*Bacillus subtilis*
YNTTLKY	Envelope glycoprotein B	*Human cytomegalovirus*
YLSTLWY	Spike glycoprotein	*Human coronavirus HKU1*
YKTTLEY	Mitochondrial Rho GTPase 1	*Candida albicans*
YKTTLAY	Mitochondrial Rho GTPase 1	*Schizosaccharomyces pombe*
YSSTLLY	Secretory component protein psh3	
YRATLNW	Glutamyl-Q tRNA(Asp) synthetase	*Yersinia pestis, Y. pseudotuberculosis*
YMSTLTY	Alpha-hemolysin	*Staphylococcus aureus*
YFSTLNY	Vacuolar ATPase assembly integral membrane protein VMA21	*Cryptococcus neoformans*
YGSSLQW	A1 cistron-splicing factor AAR2	*Saccharomyces cerevisiae*
YTTTLTY	Cell wall protein AWA1	
YGATLAY	tRNA-dihydrouridine(16/17) synthase [NAD(P)(+)]	

Table 4-8: Continued

Sequence	Sequence	Sequence
YTTTLTY	Haze protective factor 1	*Saccharomyces cerevisiae*
YRSTLNY	Cysteine-tRNA ligase 1	*Tropheryma whipplei*
YESTLNY	PAB-dependent poly(A)-specific ribonuclease subunit PAN2	*Scheffersomyces stipitis*
YFTTLTY	Tetraacyldisaccharide 4'-kinase	*Bacteroides vulgatus*
YEATLAY	Isopentenyl-diphosphate delta-isomerase	*Lactobacillus casei*
YTSTLNW	Cellulose synthase catalytic subunit	*Escherichia coli*
YETTLAW	Cellulose synthase catalytic subunit [UDP-forming]	*Xanthomonas axonopodis*
YVSTLLY	TP-dependent dethiobiotin synthetase BioD	*Fusobacterium nucleatum*
YQTTLSY	Vitamin B12 transporter BtuB	*Vibrio vulnificus*
YLSTLFY	CD2 homolog	*African swine fever virus*
YGSTLDW	UPF0061 protein YdiU	*Salmonella enteritidis PT4*
YTSTLNW	Cellulose synthase catalytic subunit [UDP-forming]	*Salmonella typhimurium*
YQATLDY	Adenylosuccinate synthetase	*Serratia proteamaculans*
YDSTLSY	FO synthase	*Mycobacterium tuberculosis, M. bovis, M. leprae, M. paratuberculosis*
YPSTLGY	Methionine aminopeptidase 2	*Mycobacterium tuberculosis, M. bovis*
YPATLEY	Queuine tRNA-ribosyltransferase	*Haemophilus ducreyi*
YFSTLYY	L-rhamnonate dehydratase	*Burkholderia ambifaria, B. phymatum, B. phytofirmans, B. xenovorans*
YVTTLTY	DNA-directed RNA polymerase subunit beta	*Arcobacter butzleri*
YDSTLDY	FO synthase	*Nocardia farcinica*
YSSTLDY	FO synthase subunit 1	*Methanosarcina acetivorans*
YSSTLEY	FO synthase subunit 1	*Methanosarcina barkeri, M. mazei*
YPSTLRY	Xylanolytic transcriptional activator xlnR	*Aspergillus terreus*
YGSTLVY	Endo-1,4-beta-xylanase A	*Butyrivibrio fibrisolvens*
YYSTLLY	Protein transport protein SEC23	*Encephalitozoon cuniculi*
YLSTLYY	L-rhamnonate dehydratase	*Delftia acidovorans*
YLTTLEY	RNA-directed RNA polymerase L	*Newcastle disease virus*
YGSTLSW	Probable glutathione S-transferase GSTF2	*Oryza sativa subsp. Japonica (rice)*
YNTTLLY	Bifunctional lycopene cyclase/phytoene synthase	*Phaeosphaeria nodorum*

v. Enriched sequences from different sorting rounds with the same sample reflected IMUNE-identified motifs

Although MEME represents an excellent tool for motif development, the algorithm cannot handle the entire sequence list generated by NGS. Therefore, only the top 1000 sequences enriched in the second sort round (F2) compared to a single stringent sort (PE-1) or the first FACS screen (F1) (PE-3, HOP-2, and HOP-4) were evaluated by MEME. Analysis of the two PE sample comparisons (PE-1 and PE-3) demonstrated enrichment of motifs already identified using the IMUNE algorithm (**Figure 4-9**). The F2 enriched sequences from PE-1 comprised the EBNA-1 linked motif previously identified by ADEPt screening (**Chapter 3**) and the IMUNE algorithm (**Figure 4-4**). The less stringently sorted population indicated a strong enrichment for PE-associated motifs, demonstrating that sort stringency affected isolation of important disease-related peptides. Furthermore, the first motif identified by MEME among the F2 enriched sequences for PE-3 represented another major IMUNE-identified PE-associated motif (L[YW]XWDXR) (**Figure 4-5**) except for the C-terminal arginine. In contrast, this motif did not appear among enriched sequences for PE-1. However, this sample contributed fewer sequences to the motif than PE-2 and PE-3. Additionally, two clearly related motifs (XGAG and HXXXGAG) from the PE-1 analysis and a motif from PE-3 (KGXGG[AG]Q) shared similarities to the predominant PE-associated GXXGAGGG motif (**Figure 4-6B**).

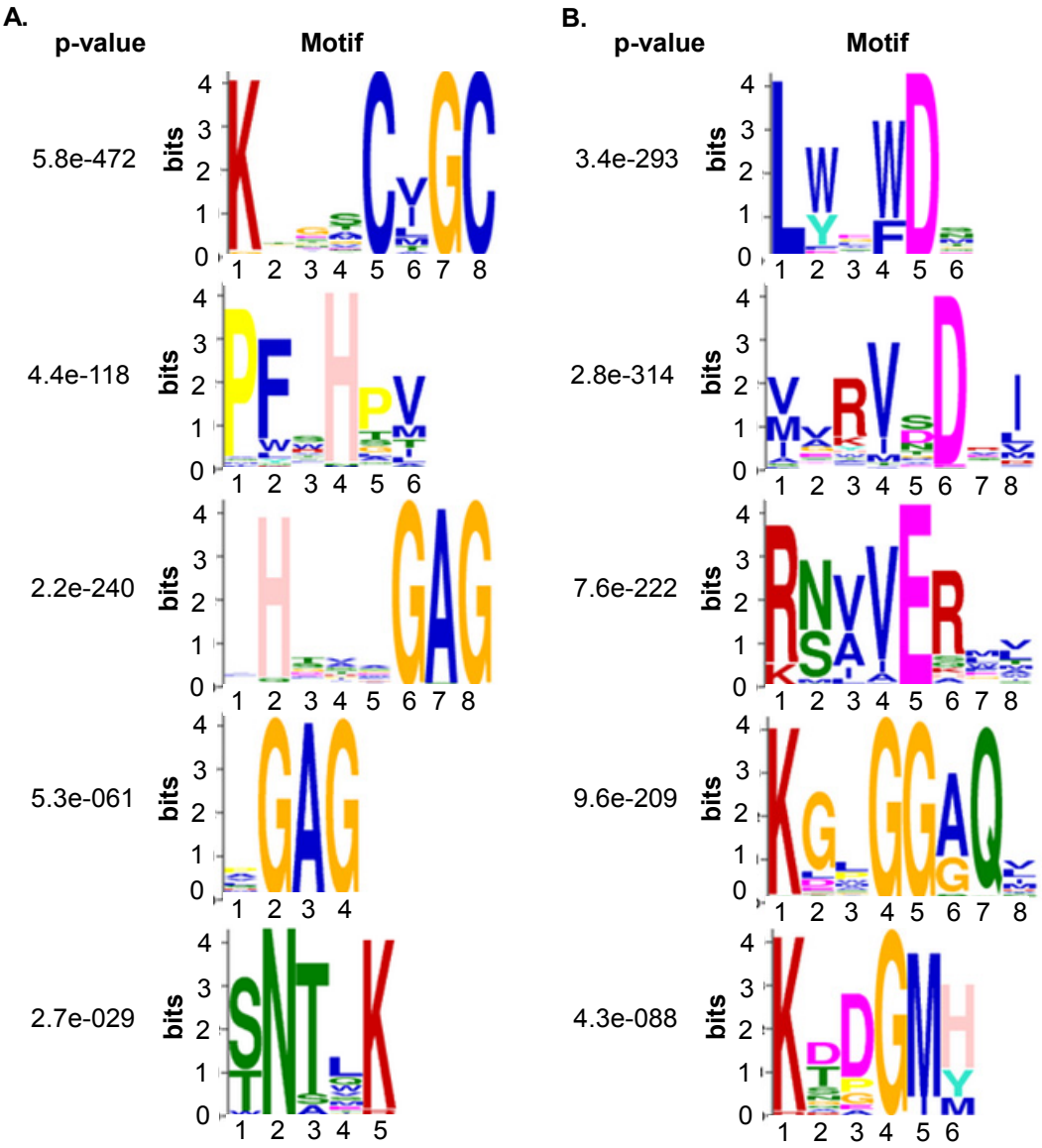

Figure 4-9: Enriched motifs after two rounds of FACS for PE samples. (**A**) Evaluating enrichment in PE-1 post F2 versus PE-1 post F1 with stringent sort conditions and (**B**) the PE-3 post F2 population compared to the F1 population yielded PE-enriched motifs.

Similar to the disease-associated motifs, the algorithm-identified control-specific motifs were represented among the F2 enriched sequences from the HOP samples (**Figure 4-10**). The first motifs from HOP-2 ([RK]AXHEX) and HOP-4 (XXHD[WFY]K) appeared related and shared similarities with the [PGA]HD[WY]K motif (**Figure 4-8B**). Additionally, the third motif ([GA][ST]SGXK) from the HOP-2 analysis was near identical to the HOP-specific motif comprised of the largest

number of sequences ([GA][ST]SGXK) (**Figure 4-8A**). The second motif (Y[DSN]T[SA]PR) from the HOP-4 analysis exhibited weak similarities to the third algorithm-identified control motif (YX[STA]TLX[WY]) (**Figure 4-8C**).

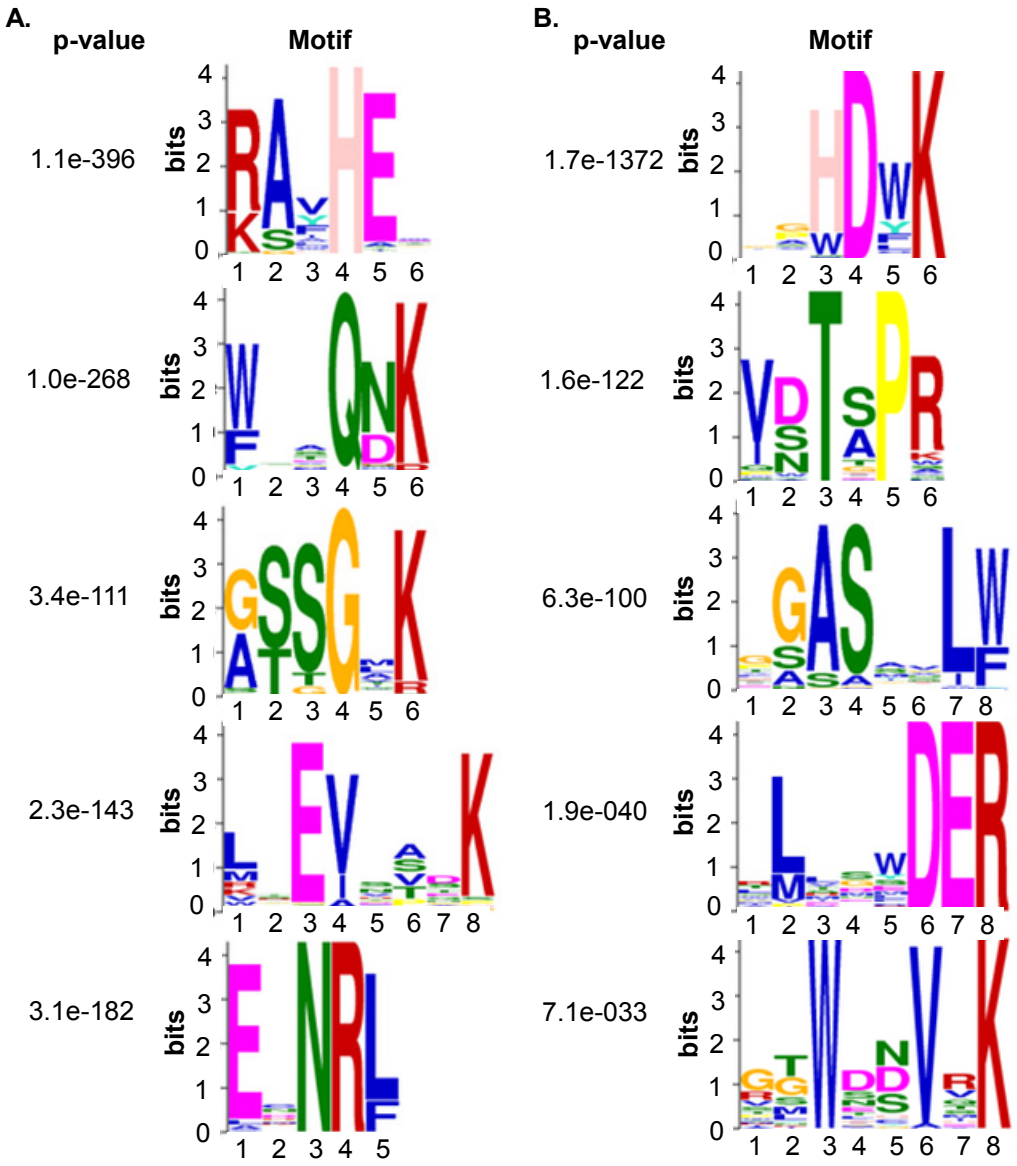

Figure 4-10: Enriched motifs after two rounds of FACS for HOP samples. Direct comparisons of peptide observations were made between the second round and the first round of FACS for (**A**) HOP-2 and (**B**) HOP-4.

This comparison analysis between two rounds of sorting the same sample did not take into account disease- versus control-associated motifs. Therefore, some of the different motifs identified may be ubiquitously represented and enriched across PE

and HOP. This explains the presence of certain motifs among enriched sequences (e.g., EXNRL, WD[NDS]VXK, [ST]NTXK) that did not appear as highly PE- or HOP-enriched when using IMUNE. However, these results demonstrated that multiple approaches revealed similar motifs enriched in PE and HOP samples.

D. Discussion

In this study, we demonstrated that NGS enables in-depth characterization of an individual's antibody binding repertoire. Specifically, we profiled the binding specificities of the circulating antibody repertoire in PE and HOP samples. By identifying PE- and HOP-associated motifs, we provide further evidence of an altered immune response among PE patients. Importantly, in this analysis, we re-identified the viral antigen (EBNA-1)-linked motif from previous work (**Chapter 3**). This serves to co-validate the reproducibility of this motif using a different method and the ability of the IMUNE algorithm to identify disease-associated motifs.

We applied multiple methodologies to characterize differences in antibody binding specificities between PE and HOP individuals. Other similar studies have focused on the most repeated sequences identified in the library population.[208] However, different conditions unrelated to binding could account for increased representation among library members. For example, it has been shown that the amplification process between selection rounds with phage greatly reduces the diversity.[207] Instead, for IMUNE, we normalized all observations of the specific peptides to one for generating the statistically enriched patterns, which prevents any inappropriately higher sequence repeats from skewing the motifs due to undesired issues, such as PCR bias or improved cell growth/expression. We hypothesized that a

given sequence need only be considered once, since numerous unique peptide sequences should comprise important motifs. A similar extension of this was applied to identify human-derived peptides associated with disease from a phage displayed library.[183] To characterize the enriched peptides, we compared two rounds of sorting from the same patient sample. Thus, the inherent biases, such as cell growth and PCR, should be present in both rounds, accounting for these undesired issues. Evaluating enriched sequences between a second and third round identified low nanomolar affinity binders from a DNA aptamer library.[211] Similarly, the most highly enriched peptides corresponded to the motifs identified from the IMUNE algorithm, highlighting the skewed antibody response.

While we re-identified the previous EBNA-1 linked motif (KXXXC[IVL]GCK) (**Chapter 3**), we also discovered further differences between the antibody repertoires present in PE and HOP individuals. One of these motifs (GXXGAGGG) represents a previously identified immunodominant region of EBNA-1.[217] This provides further support for an increased antibody response towards this viral antigen among PE patients compared to HOP (**Chapter 3**). Moreover, the previously described methods that depleted peptides binding to HOP-associated antibodies (**Chapters 2 and 3**) would hinder identification of this motif due to the likely presence of this specificity among HOP. This motif was also identified in a region of capsid proteins from parvovirus B19, which was previously implicated in PE for a possible molecular mimicry mechanism leading to the agonistic antibodies against the angiotensin II type 1 receptor (AT_1-AAs).[135] Another strong, PE-associated motif (L[YW]XWDXR) shared similarities with a number of human proteins, including insulin-like growth

factor receptor 1 and prostaglandin FP2α receptor. Interestingly, the insulin-like growth factor system appears to play a role in fetal growth, while alterations are associated with growth restriction[218] and PE patients exhibit reduced levels of insulin-like growth factor.[219] Additionally, evidence suggests the prostaglandin FP2α receptor is expressed in the umbilical vein and potentially mediates the contraction effect through the agonist, prostaglandin FP2α.[220] Umbilical endothelial cells produce increased prostaglandin FP2α when exposed to plasma from women with PE compared to healthy pregnancies.[221] Effectively linking the candidate autoantigen(s) to this motif will require further investigation of antibody binding activity, as shown in Chapter 3. However, the additional PE-associated motifs identified through this new methodology enhance our understanding of the altered antibody response in PE patients.

To complement the PE-associated motifs, the discovery of HOP-associated motifs provides additional evidence of an altered immune response. These motifs exhibit similarities with a variety of viruses and human pathogens. Interestingly, several hits with the YX[TSA][TS]LX[WY] motif included protein homologs (e.g., mitochondrial Rho GTPase 1 and FO synthase) found in different organisms, potentially indicating a ubiquitous specificity. The large diversity of candidate hits, especially for the YX[TSA][TS]LX[WY] motif, highlights the need for directed evolution (**Chapter 3**) to expand the motifs and reduce the number of similar proteins. While these antibody specificities may still be present among PE patients, their presence appears reduced with respect to the HOP samples or relevant disease-associated motifs. A significant reduction in antibody binding activity to these HOP-

associated peptides accompanied by increased binding to PE-associated peptides could yield an accurate diagnostic. A previous study included a panel of serum peptides with increased and decreased levels in PE compared to HOP that exhibited strong diagnostic accuracy.[205] Furthermore, understanding the relevant antigens these PE and HOP-associated motifs represent may reveal a pathological role for these alterations to the antibody repertoire.

Our results demonstrate the utility of NGS to characterize antibody binding peptides isolated by bacterial display library screening; however, this study evaluated a small set of samples. As with many diseases, PE presentation is very heterogeneous and the existence of multiple disease subtypes has been proposed.[106] Thus, additional samples would provide a more accurate representation of disease diversity and enable antibody repertoire profiling for various subtypes (i.e., early-onset, nonproteinuric, etc.). Furthermore, supplemental analysis of enriched peptides between sort rounds, such as post F1 compared to single magnetic selection, may indicate the point at which the sharpest enrichment occurs. However, especially important for this analysis, a higher degree of oversampling of unique sequences must be achieved to guarantee complete sequence coverage. Despite these potential areas of improvement, the in-depth antibody binding profile obtained through NGS identified PE- and HOP-associated motifs, providing further insights into the altered antibody response in PE.

5. Additional applications of the newly developed AT_1 epitope binding assay for antibody detection

While agonistic antibodies against the angiotensin II type 1 receptor appear to play a role in pre-eclampsia pathology, characterizing the presence of these antibodies has been limited by assay throughput. However, we recently developed a high-avidity assay that enabled higher-throughput fluorescence-based detection of antibody binding to the AT_1 epitope. Here, we demonstrate extensions of this epitope specific binding assay to evaluate antibody presence in non-pregnant women and a new pre-eclampsia mouse model.

A. Introduction

Pre-eclampsia (PE) is associated with an intensified inflammatory condition[190] and increased autoantibody activity,[128-130] including antibodies that bind GPR50 (**Chapter 3**). Additionally, agonistic antibodies against the angiotensin II type 1 receptor (AT_1-AAs) appear to play a role in PE pathology.[3] However, detecting these antibodies remains difficult, since complex, biologic function based assays such as cardiomyocyte beat rate[2] and luciferase reporter signal,[138] show the greatest sensitivity. As previously discussed (**Chapter 2**),[191] we developed a novel assay to detect antibodies binding to the epitope characterized for these AT_1-AAs. This assay enables higher-throughput analysis for the presence of AT_1 epitope binding antibodies. In comparison, recently developed ELISAs using AT_1 expressing cells[222] or the second extracellular loop[178] do not indicate the specific binding sequence

responsible for the signal. This is especially important for discerning the PE-associated AFHYESQ binding antibodies from distinct AT_1 receptor binding antibody specificities found in other diseases, such as renal allograft rejection.[174]

We sought to extend the applications of this assay to determine the presence of AT_1-AAs in women who have never been pregnant (nulligravid) and a new mouse model of PE. It has been previously shown that non-pregnant normotensive samples exhibit reduced levels of AT_1 receptor binding antibodies compared to HOP, while PE samples demonstrate further increased activity.[178] Thus, we aimed to further characterize this relationship using the epitope specific assay. Furthermore, recently it was shown that injecting pregnant and non-pregnant mice with LIGHT, or tumor necrosis factor superfamily member 14, induced symptoms of PE.[223] LIGHT is a pro-inflammatory cytokine that has been associated with autoimmune and inflammatory conditions.[224] The results of this study and others[190] provide support for an intensified inflammatory condition associated with PE. Thus, we sought to evaluate the presence of AT_1 epitope binding antibodies in this new mouse model. Our results confirm a relative increase in AT_1 epitope antibody binding activity from nulligravid to HOP and furthermore to PE samples. Additionally, LIGHT-injected pregnant and non-pregnant mouse samples exhibited increased AT_1 epitope binding activity.

B. Materials and Methods

The previously described AT_1 epitope binding assay (**Chapter 2**)[191] was applied in this study. A fragment of the characterized AT_1 epitope (AFHYESQ) flanked by glycine residues to increase flexibility was displayed on bacteria using the eCPX

scaffold.[90] Samples from women who had never been pregnant (nulligravid) (n=21) and pregnant and non-pregnant mice were provided from the University of Texas, Houston Medical School. Mouse model development and experiments were conducted by collaborators at the University of Texas, Houston Medical School. To evaluate binding to antibodies present in human plasma samples, antibody fractions were enriched through ammonium sulfate precipitations and labeled with biotin. Incubating antibody fractions (1 μM) with the bacterial displayed fragment followed by diluted (1:333) streptavidin-conjugated R-phycoerythrin (SA-PE) (Invitrogen) enabled binding analysis by flow cytometric fluorescence. Binding of the nulligravid samples was compared to previously analyzed PE (n=45) and healthy-outcome pregnancy (HOP) (n=48) samples (**Chapter 2**).[191] Binding to antibodies present in mouse model samples utilized unprocessed diluted plasma (1:100) in PBS. After incubating AT_1 epitope expressing bacteria with mouse plasma, labeling steps proceeded with reagents diluted in PBS with 0.1% BSA. A biotin-conjugated mouse IgG specific secondary (1:200) (Vector Labs) was followed by incubation with SA-PE to fluorescently tag the antibodies bound to the AT_1 epitope, enabling analysis by flow cytometry. All antibody binding activity was measured in duplicate and normalized by background activity with a negative control, eCPX scaffold not expressing peptide. Statistical analysis of Mann-Whitney U test and ANOVA with Tukey's multiple comparisons test was conducted using Prism4 or Prism6

C. Results and Discussion

Using the AT_1 epitope assay to detect antibodies demonstrated that nulligravid (n=21) samples exhibited similar binding activity as background. In comparison to

previously evaluated PE (n=45) and HOP (n=48) samples (**Chapter 2**),[191] antibody binding activity in nulligravid samples was significantly reduced (**Figure 5-1**). This agrees with previous results using the second extracellular loop of the AT_1 receptor in an ELISA.[178] In comparison, the nulligravid binding activity observed with the EB15 and GPR50 fragments appeared similar to HOP but significantly lower than the PE activity (**Figure 3-10**). Our results indicate that healthy pregnancy is associated with increased AT_1-AA production, while patients with PE experience further escalations in antibody production. Previous studies have demonstrated increased levels of pro-inflammatory cytokines and decreased anti-inflammtory cytokine (IL-10) in PE compared to control.[190] While reduced compared to PE, the pregnant control samples exhibited increased inflammatory cytokines compared with non-pregnant samples. Thus, inflammation could lead to increased autoantibody production with further intensity in PE resulting in aberrant levels.

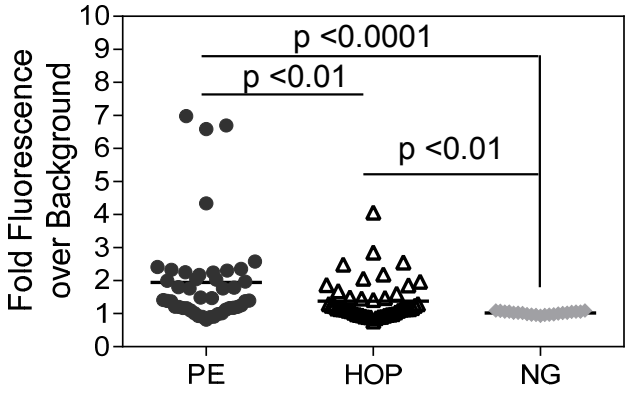

Figure 5-1: Antibody binding to the AT_1 epitope is significantly reduced in nulligravid (NG) samples. Binding activity compared to previously analyzed PE and HOP. p – values represent results from Mann-Whitney U test

Evaluating antibody binding to the AT_1 epitope in non-pregnant mouse samples revealed a dynamic increase in binding activity upon infusion of LIGHT for 14 days (n=8) compared to saline (n=7) (**Figure 5-2**). In contrast, a five day infusion (n=8)

did not significantly differ from saline injection. Interestingly, injecting a tissue transglutaminase (tTG) blocker (n=5), cystamine, prevented this signal of antibody production, indicating tTG involvement. This builds upon previous observations indicating a role for tTG in PE pathology.[225] PE patients exhibited increased levels of tTG while cystamine reduced antibody mediated hypertension and proteinuria in the adoptive transfer mouse model. Furthermore, a mouse knockout of the pro-inflammatory cytokine IL-6 (IL-6-/-) infused with LIGHT (n=3) did not exhibit enhanced binding activity, instead showing similar levels to saline injection. Similarly, a conditional knockdown of hypoxia-inducible factor 1 alpha, HIF-1a, in endothelial cells (HIF-1a v-cad cre) did not exhibit increased antibody production upon LIGHT infusion (n=6). These results indicated that increased autoantibody production by LIGHT was mediated by IL-6 and HIF-1a.

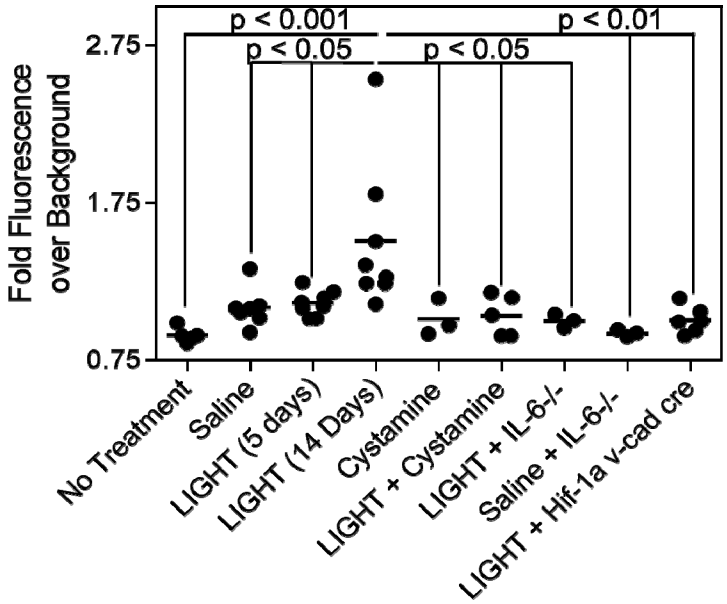

Figure 5-2: LIGHT affects plasma antibody binding activity to the bacterial displayed AT_1 epitope in non-pregnant mice. p-values indicate results from Tukey's multiple comparisons test following ANOVA

While non-pregnant mice experiencing a 14 day infusion exhibited increased AT_1 epitope binding activity, pregnant mice could not be subjected to the same level of

LIGHT injection without pregnancy loss. Instead, pregnant mice had been injected with LIGHT at gestational day 13.5 and 14.5 of pregnancy. Despite this reduced injection rate, the mice treated with LIGHT (n=13) demonstrated increased AT_1 epitope binding activity compared to saline injections (n=8) (**Figure 5-3**). This apparent antibody binding activity mediated by LIGHT injection indicates potential placental involvement to initiate a similar response as the 14 day infusion in non-pregnant mice. Similarly, treatment with cystamine (n=4) reduced the antibody binding activity observed, although this result was not statistically significant due to increased variance. However, mice treated with HIF-1a siRNA (small interfering or silencing RNA) and LIGHT (n=4) exhibited significantly reduced AT_1 epitope binding. Thus, results from pregnant mice reflect similar AT_1 antibody production responses to LIGHT as the non-pregnant mice despite reduced injection.

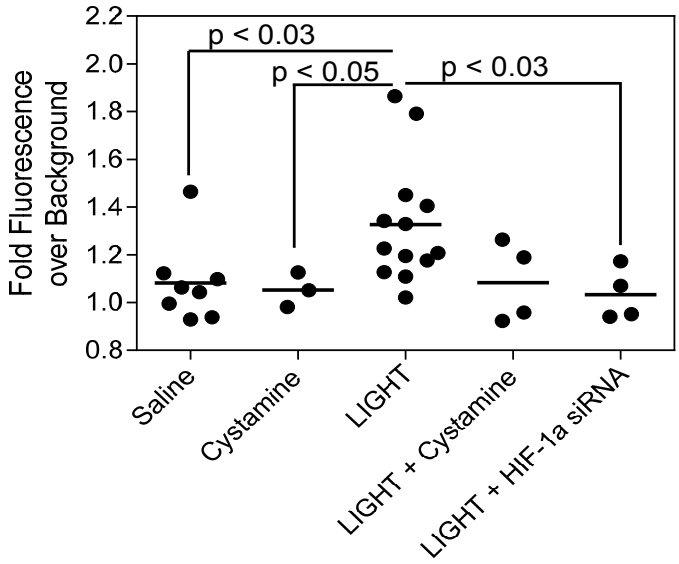

Figure 5-3: In a pregnant mouse model, LIGHT induces plasma antibody binding to the bacterial displayed AT_1 epitope. p-values indicate Mann-Whitney U test results

Our results demonstrate the utility of this AT_1 epitope binding assay for the detection of AT_1-AAs in various samples. The significantly reduced antibody binding to nulligravid samples confirms an epitope specific change in antibody

production previously shown with the full-length second extracellular loop.[178] Furthermore, we have shown increased antibody binding to the AT_1 epitope in response to LIGHT in non-pregnant and pregnant mice that involves tTG, IL-6, and HIF-1a. Despite reduced LIGHT injection compared to non-pregnant mice, the pregnant mice exhibited a similar pattern of increased antibody production in response to LIGHT, pointing towards placental involvement. Overall, these mouse model results indicate a relationship between the inflammatory condition associated with PE and autoantibody production.

6. Conclusions

A. Perspectives

i. Developing an antibody-detecting diagnostic for pre-eclampsia

While various biomarkers have been proposed for pre-eclampsia (PE) diagnosis, few have demonstrated accuracies required for clinical utility. In an effort to fill this need, we focused on the circulating antibody repertoire and characterized antibody specificities in PE (**Chapter 2**). Here, bacterial displayed peptide library screening identified PE-specific antibody-detecting peptides. Using differently labeled antibody fractions enriched from plasma samples of PE and healthy-outcome pregnancies (HOP), multi-parameter cell sorting isolated PE cross-reactive and specific peptides. We also developed a peptide display-based assay sensitive enough to detect antibody binding to the angiotensin II type 1 receptor (AT_1) epitope without using complex biologic function based assays.[2,138] This enables higher-throughput evaluation of AT_1-AA activity, which was utilized to perform AT_1 epitope specific analysis of antibodies present in a mouse model of PE (**Chapter 5**). Furthermore, depleting antibodies binding this AT_1 epitope verified that library screening identified antibody biomarkers present in PE patients distinct from known AT_1-AAs. A panel of antibody-detecting peptides, representing different antibody specificities, achieved strong diagnostic accuracy (80%) with high specificity (95%) in a set of cross-validated training samples. Importantly, we verified the 80% overall diagnostic accuracy from training in a full validation set of new samples, which differs from previous studies in the lab using machine learning algorithms.[153,160] This represents

an important step because many classification algorithms over-fit the training data. Combining the peptide panel with traditional diagnostic criteria of proteinuria and hypertension resulted in improved detection of PE patients that experienced severe symptoms or delivered early (<37 weeks' gestation). Furthermore, peptide reactivity negatively correlated with platelet levels in PE patients, indicating a relationship between antibody binding and PE symptoms. This study provided evidence for an altered antibody response in PE that resulted in the development of an antibody-detecting peptide array to differentiate PE from HOP. However, the characterized motifs did not enable unambiguous discovery of autoantigen or potential environmental trigger antigen.

ii. Identifying disease-associated antibody binding targets

While discovery of disease-associated antibodies enabled development of a PE classifier, the identity of the corresponding antigen(s) could improve understanding of the antibody's role in PE and elucidate potential therapeutic targets. Furthermore, rather than the mimicking library peptides, utilizing whole or partial antigen(s) to detect antibody presence may enhance diagnostic accuracy of the assay. Instead of antibody fractions, unprocessed, diluted plasma may better replicate the native antibody binding environment leading to improved motifs that more closely relate to protein targets. Therefore, a second study screened against unprocessed plasma to discover a PE-associated binding motif using a significantly altered screening method and characterized antibody binding targets (**Chapter 3**). Importantly, our results demonstrate a novel case for molecular mimicry operating near the time of delivery by which a PE-associated IgG1 antibody cross-reacts with a viral antigen (EBNA-1)

and a human G protein-coupled receptor (GPR50). While not a new concept, discovery of novel molecular mimicry mechanisms without depending on prior knowledge of disease pathology remains difficult. Moreover, this marks our first successful venture to link a library peptide motif to previously uncharacterized antibody binding protein targets using bacterial display. In characterizing this molecular mimicry interaction, we provided additional evidence suggesting that directed evolution of disease-associated motifs improves immune activity profiling, as previously shown in a celiac disease study.[189] Through directed evolution, PE-associated peptide binding activity increased and the correlation improved between library peptide reactivity and the ELISA signal observed for antibody binding to the identified target antigen, EBNA-1. The EBNA-1 derived fragment (EB15) sequestered antibody binding activity to the GPR50 fragment and the full-length protein expressed in cells. Furthermore, results indicated GPR50 expression in a trophoblast model cell line and placental tissue. As an IgG1 subtype, this antibody can activate the complement system; therefore, antibody binding to this placental protein may contribute to the increased complement deposition observed.[119,122,173] While additional findings, such as inhibition of the melatonin receptor 1,[198] may link GPR50 to PE-related disturbances, it remains unclear if aberrant antibody binding to this placental protein or other similar proteins causes any PE-related effects. However, since the EB15 peptide inhibited antibody binding to GPR50 *in vitro*, it might effectively block any aberrant antibody binding *in vivo* as well. An adoptive transfer model, similar to the study on AT_1-AAs in pregnant mice,[3] could provide

insights into these questions of PE pathology and effective blocking activity for this GPR50 and EBNA-1 cross-reactive antibody.

iii. Applying next-generation sequencing to profile individual antibody repertoires highlights alterations in PE

In addition to the EBNA-1 linked specificity (**Chapter 3**) we sought to identify new motifs by characterizing individual antibody repertoires of PE and HOP (**Chapter 4**). Using next-generation sequencing (NGS) enables higher resolution of individuals' antibody specificities that can be cross-examined between healthy and disease samples. Therefore, we developed and applied a unique methodology to profile individual patients' antibody-binding repertoire using NGS to analyze sequences obtained from bacterial displayed library screening with PE and HOP samples (n=4, each). Distinct individuals' peptide encoding DNA sequences were isolated, PCR amplified, and uniquely barcoded to enable pooling. After NGS, a computational algorithm (IMUNE) developed in the Daugherty Group at UCSB identified peptide patterns enriched separately in PE and HOP by cross-examining the individual repertoires. Analysis of algorithm-derived patterns led to PE- and HOP-associated motifs, which were confirmed using a second method of evaluating enriched sequences from different rounds of screening. Thus, multiple approaches revealed similar alterations to the antibody repertoire. Importantly, the previously characterized viral antigen (EBNA-1)-linked motif (KXXXC[VLI]GCK) was identified by this new method. Re-identifying this motif validates the algorithm's ability to discover disease-associated motifs and further demonstrates the motif's reproducibility using various methods. Another highly represented PE motif

(L[YW]XWDXR) shared similarities with a number of human proteins. Interestingly, a third PE-associated motif shared strong similarities to the immunodominant region of EBNA-1,[217] providing support for increased antibody binding activity against this viral antigen in PE. Only one of the HOP-associated motifs yielded a reasonable number of hits (<20) for antigen fragment analysis; however, in general, these HOP-associated motifs shared similarities with a number of human pathogens. The new motifs identified by this approach would benefit from directed evolution to characterize the candidate antigens they represent. In general, this NGS-enabled in-depth motif analysis by cross-examination and peptide enrichment provides further evidence of a skewed antibody response in PE patients compared to HOP.

iv. Overall conclusions

Taken together, the results from these studies demonstrate changes in the circulating antibody repertoire in patients with PE compared to HOP. These alterations enabled development of an antibody-detecting peptide-based classifier, achieving a validated accuracy of 80% for this heterogeneous disease. Moreover, motif characterization through three methodologies revealed a skewed antibody response, resulting in PE- and HOP-associated binding motifs. Thus, we have provided supporting evidence for an altered immune response in PE. Several observations point to an aberrant immune response, including dysregulated complement system,[118,119,122,173] altered levels of B-cells[126] and T-cells,[123,124] and an intensified inflammatory condition.[190] Interestingly, the AT_1-AA detection results from the LIGHT injected mouse model of PE provide a link between the increased

inflammatory condition and antibody production. As researchers continue to investigate the immunological alterations in PE, these different components, including aberrant autoantibody binding activity, may reveal insights into PE etiology.

A significant contribution of this work is the demonstration of unbiased antibody binding target identification. Specifically, library screening, directed evolution, and candidate antigen fragment activity analysis resulted in the characterization of a new example of molecular mimicry leading to antibody binding to human proteins. While previous antigen discovery efforts were aided by pre-existing knowledge of disease etiology,[189] our results emphasize the utility of peptide library screening for *de novo* identification of antigenic targets. Here, we applied this process to a PE-associated antibody specificity; however, this can be extended to a variety of diseases. Furthermore, we have demonstrated the powerful depth of sequence coverage provided by NGS, which will be an important tool for future applications of antibody repertoire profiling. Combining individual repertoire profiling by NGS with directed evolution strategies will enable investigation of multiple disease-associated motifs in parallel, revealing broad insights into disease etiology.

B. Future Directions

A major impact of these discoveries involves the new investigation opportunities they enable. Our results indicate a molecular mimicry mechanism acting in PE near the time of delivery in which an antibody cross-reacts with the viral antigen EBNA-1 and human protein GPR50. Since mice also express GPR50 in the placenta,[192] an adoptive transfer model, similar to the AT_1-AAs investigation by our collaborators,[3]

can investigate how this antibody activity might play a role in pathology. To focus on this antibody specificity, GPR50/EBNA-1 binding antibodies could be purified from PE patients. Purified antibody would be injected into pregnant mice with or without the blocking peptide EB15. The minimum concentration of peptide required to fully block antibody binding to the GPR50 fragment was ~$4x10^{-10}$ M, which means the antibody binding sites were largely saturated at this point. Based on a bivalent antibody and the 1:200 plasma dilution performed for this assay, this yields roughly $4x10^{-8}$ M of antibody in circulation. Since we don't know the precise affinity of the antibody for GPR50 or EB15, this serves as a rough estimate of the antibody concentration. Based on a mouse blood volume of 2 mL, ~12.5 μg of purified antibody should be injected to obtain a comparable concentration. In comparison, the AT_1-AA adoptive transfer model injected ~20 μg of purified antibodies into pregnant mice.[3] Taking several measurements, including blood pressure, urinary protein, and levels of melatonin, sFlt-1, and PlGF, will determine potential antibody effects and whether EB15 administration attenuates these effects. Furthermore, evaluating the kidney or placenta for complement deposition[173] might demonstrate this antibody's role in the complement system dysregulation[118-122,173] already associated with PE. Therefore, extending the insights gained from our results into an *in vivo* mouse model could provide evidence of a pathological role for this newly identified PE-associated antibody specificity.

In addition to this viral antigen-linked antibody specificity, NGS analysis of individual antibody repertoires identified PE and HOP-associated motifs. However, these motifs often yielded a large number of hits from BLASTp or ScanProsite

searches. Building upon previous success with directed evolution (**Chapter 3**),[189] these motifs can be further evolved to expand the epitope while increasing cross-reactivity and specificity. Applying this directed evolution strategy to the PE-associated motif L[YW]XWDXR may enable identification of new PE-related antibody binding targets. Especially interesting would be the discovery of an antibody specificity that is significantly reduced or lacking in PE patients. A similar condition has been shown in children with recurrent infections that exhibit normal levels of total IgG but reduced activity towards *Haemophilus influenzae* type b, despite immunization.[226] Combining measurements of differential antibody binding activities (reduced and enhanced) may improve diagnostic accuracy of an antibody-detecting peptide panel. For example, a profile of 19 increased/decreased serum-derived peptides yielded a classifier with 93% and 90% accuracy in training and validation sets, respectively.[205] Evolving these motifs will enable discovery of additional PE-associated alterations in antibody activity towards specific antigens. Characterizing these specificities could improve understanding of disease etiology and diagnostic performance of an antibody-detecting panel.

Although we have demonstrated the benefit of a diagnostic tool near the time of delivery to detect PE, especially those with non-traditional PE presentation (i.e., no proteinuria), a predictive test is highly sought after. Recently, 47 biomarkers were evaluated for PE identification during the first trimester with a large cohort of 5623 pregnant women.[146] However, the study showed that molecular biomarkers did not significantly improve predictive performance compared to clinical risk factors alone for all PE. Combining the ratio of cystatin C/PlGF with clinical risk factors resulted

in moderate predictive accuracy for early-onset PE in the validation set with 44% sensitivity at 95% specificity (AUC=0.73). The low prevalence of early-onset PE (0.5% in the study) limits this model's clinical application because for every one correctly diagnosed early PE there would be 23 false-positives.[146] Thus, accurate prediction of PE during the first trimester remains elusive. In an effort to fill this unmet need, samples obtained during the first trimester from women destined for PE and HOP can be screened for antibody markers. After dividing the samples into discovery and validation sets, the discovery samples will be screened by individual antibody repertoire analysis with NGS followed by motif evolution. Reactivity analysis with library peptides and antigen fragments in the validation set will provide preliminary performance statistics. Given the overall success of our bacterial displayed library screening in PE near delivery[191] and celiac disease,[153,160,189] we hypothesize that first trimester screening of PE and HOP samples will yield predictive antibody biomarkers for PE detection.

In addition to these PE-related future studies, different methodological advances, such as discontinuous or structural epitope discovery, remain unexplored. Our studies on antibody epitope profiling have primarily focused on identifying linear epitopes or potentially, a linear section of a discontinuous epitope. A substantial portion of all antibody epitopes are rather discontinuous, comprising residues from sections of greater than 40 amino acids.[227] In fact, some estimate that peptides of 7-15 amino acids in length from linear sections of candidate antigens recover all major functional residues only ~50% of the time. Although it may not be important to capture all functional residues to retain activity, an understanding of the complete

epitope may be desired. Other studies have utilized pre-existing knowledge of discontinuous epitopes to refine the residues involved from discontiguous protein sections through co-incubation[73] or fusion.[66] However, the biterminal eCPX scaffold enables conveniently linked co-expression of different peptides. Therefore, an identified disease-related motif or specific peptide can be expressed on one terminus with a peptide library expressed on the other. Screening for cooperatively enhanced binding may reveal additional functional residues stemming from a discontinuous section. Identifying important functional residues of a discontinuous epitope may help boost antibody detection sensitivity. Moreover, in combination with discontinuous epitope prediction tools[167] different candidate targets can be investigated for alignment with these discontinuous motifs. In addition to probing for linear sections of discontinuous epitopes, we can search for structural mimotopes with different library constructions. As previously discussed, a library designed to include multiple disulfide constraints affected the apparent affinity of identified antibody binders (**Chapter 1**). We hypothesized that the structure imparted by these disulfide bonds enables selection of high affinity structural mimotopes. While these might be difficult to relate to an antigenic target, the efficacy of these peptides as diagnostic reagents has not been investigated. For direct comparison, the same patients used for screening in the ADEPt[189] method (**Chapter 3**) that yielded the EBNA-1/GPR50 motif can be screened against this multiply-constrained peptide library. Since screening the fully randomized library naturally identified a disulfide-constrained motif, it will be interesting to compare the motif results from screening a library that

already includes such constraints. These different library construction approaches could yield novel discontinuous and structural epitopes/mimotopes.

While applying NGS to isolated peptides from library screening profiles the binding repertoire, this high-throughput technology has also provided insights into the antibody sequences themselves. Specifically, sorting and sequencing individual antibody producing plasmablasts provides a profile of the current immune response. This group of cells typically represents a small percentage (<1%) of circulating B-cells. However, recent exposures (e.g., infection) or chronic immune responses (e.g., autoimmune diseases) increase the presence of these plasmablasts (>30% in infections).[213] Methodological advances have enabled sequencing of the entire variable region of the antibodies produced by these plasmablasts.[51,52,213] Combining these variable regions with constant domains facilitates production of human derived full-length antibodies in the lab. However, the binding specificities of these antibodies must then be determined. Thus far, studies of plasmablast-derived antibodies have relied upon pre-existing knowledge to probe protein microarrays. However, our technology represents an invaluable tool for profiling antibody binding specificities. Specifically, by screening bacterial displayed peptide libraries against these human-derived antibodies, their individual binding specificities can be characterized. As previously demonstrated, searching the identified binding motifs in protein databases yields candidate antigens. While these candidate hits can be evaluated individually by bacterial displayed fragment analysis, protein/peptide microarrays based on these unbiased search results can also be constructed and probed. Therefore, combining NGS-based antibody sequencing with bacterial

displayed peptide library antibody profiling will provide a direct link between the antibody species and the specific epitope to which it binds.

Through this work, we have established the powerful utility of bacterial displayed peptide library screening to profile the antibody repertoire. The results from this work demonstrate this tool's application for diagnostic development and the ability to gain insights into disease etiology through antibody target identification. Thus, this impactful technology provides new ways to expand our knowledge of the changes to the circulating antibody repertoire in response to disease.

References

[1] Lain, K.Y. & Roberts, J.M. Contemporary Concepts of the Pathogenesis Management of Preeclampsia. *The Journal of the American Medical Association* **287**, 3183-3186 (2002).

[2] Wallukat, G. et al. Patients with Preeclampsia Develop Agonistic Autoantibodies against the Angiotensin AT1 Receptor. *The Journal of Clinical Investigation* **103**, 945-952 (1999).

[3] Zhou, C.C. et al. Angiotensin Receptor Agonistic Autoantibodies Induce Pre-eclampsia in Pregnant Mice. *Nature Medicine* **14**, 855-862 (2008).

[4] Petricoin, E.F., Zoon, K.C., Kohn, E.E., Barrett, J.C. & Liotta, L.A. Clinical Proteomics: Translating Benchside Promise into Bedside Reality. *Nature Reviews Drug Discovery* **1**, 683-695 (2002).

[5] Gerszten, R.E. & Wang, T.J. The Search for New Cardiovascular Biomarkers. *Nature* **451**, 949-952 (2008).

[6] Aebersold, R. et al. Perspective: A Program to Improve Protein Biomarker Discovery for Cancer. *Journal of Proteome Research* **4**, 1104-1109 (2005).

[7] Reddy, M.M. et al. Identification of Candidate IgG Biomarkers for Alzheimer's Disease via Combinatorial Library Screening. *Cell* **144**, 132-142 (2011).

[8] Price, J. V. *et al.* Protein microarray analysis reveals BAFF-binding autoantibodies in systemic lupus erythematosus. *The Journal of Clinical Investigation* **123,** 5135–5145 (2013).

[9] Sigdel, T. K. *et al.* Shotgun proteomics identifies proteins specific for acute renal transplant rejection. *Proteomics Clin. Appl.* **4,** 32–47 (2010).

[10] Rifai, N., Gillette, M.A. & Carr, S.A. Protein Biomarker Discovery and Validation: the Long and Uncertain Path to Clinical Utility. *Nature Biotechnology* **24**, 971-983 (2006).

[11] Anderson, K. S. *et al.* Protein Microarray Signature of Autoantibody Biomarkers for the Early Detection of Breast Cancer. *Journal of Proteome Research* **10,** 85–96 (2011).

[12] Kohrt, H. E. *et al.* Identification of gene microarray expression profiles in patients with chronic graft-versus-host disease following allogeneic hematopoietic cell transplantation. *Clinical Immunology* **148,** 124–135 (2013).

[13] Mahler, M. & Fritzler, M.J. Epitope Specificity and Significance in Systemic Autoimmune Diseases. *Annals of the New York Academy of Sciences* **1183**, 267-287 (2010).

[14] Christmann, A., Wentzel, A., Meyer, C., Meyers, G. & Kolmar, H. Epitope Mapping and Affinity Purification of Monospecific Antibodies by Escherichia coli Cell Surface Display of Gene-Derived Random Peptide Libraries. *Journal of Immunological Methods* **257,** 163–173 (2001).

[15] Berglund, L., Andrade, J., Odeberg, J. & Uhlen, M. The epitope space of the human proteome. *Protein Sci* **17,** 606–613 (2008).

[16] Nakachi, K. *et al.* Epitopes recognised by tissue transglutaminase antibodies in coeliac disease. *Journal of Autoimmunity* **22,** 53–63 (2004).

[17] Di Pisa, M. *et al.* Epitope mapping of the N-terminal portion of tissue transglutaminase protein antigen to identify linear epitopes in celiac disease. *J. Pept. Sci.* **20,** 689–695 (2014).

[18] Sárdy, M. *et al.* Tissue transglutaminase ELISA positivity in autoimmune disease independent of gluten-sensitive disease. *Clinica Chimica Acta* **376,** 126–135 (2007).

[19] Simon-Vecsei, Z. *et al.* A single conformational transglutaminase 2 epitope contributed by three domains is critical for celiac antibody binding and effects. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 431–436 (2012).

[20] Sompayrac, L. *How the Immune System Works*. (Wiley-Blackwell, 2012).

[21] Liao Y-H, Wei Y-M, Wang M, Wang Z-H, Yuan H-T, Cheng L-X. Autoantibodies against AT1-receptor and alpha1-adrenergic receptor in patients with hypertension. *Hypertension Res*. 2002;25:641–646.

[22] Gruden, M. A. *et al.* Differential neuroimmune markers to the onset of Alzheimer's disease neurodegeneration and dementia: Autoantibodies to Aβ(25–35) oligomers, S100b and neurotransmitters. *Journal of Neuroimmunology* **186,** 181–192 (2007).

[23] Pidala, J., Sarwal, M., Roedder, S. & Lee, S. Biologic markers of chronic GVHD. *Bone Marrow Transplantation* **49,** 324–331 (2014).

[24] Alaedini, A. & Green, P.H.R. Autoantibodies in Celiac Disease. *Autoimmunity* **41**, 19-26 (2008).

[25] Sollid, L.M., Molberg, O., McAdam, S. & Lundin, K.E.A. Autoantibodies in Coeliac Disease: Tissue Transglutaminase-Guilt by Association? *Gut* **41**, 851-852 (1997).

[26] Sokolove, J. *et al.* Autoantibody epitope spreading in the pre-clinical phase predicts progression to Rheumatoid Arthritis. *PLOS One* **7,** e35296 (2012).

[27] Davies, J.M. Molecular Mimicry: Can Epitope Mimicry Induce Autoimmune Disease? *Immunology and Cell Biology* **75**, 113-126 (1997).

[28] Poole, B. D., Scofield, R. H., Harley, J. B. & James, J. A. Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity* **39,** 63–70 (2006).

[29] Wim Ang, C., Jacobs, B. C. & Laman, J. D. The Guillain–Barré syndrome: a true case of molecular mimicry. *Trends in Immunology* **25,** 61–66 (2004).

[30] Cusick, M., Libbey, J. & Fujinami, R. Molecular Mimicry as a Mechanism of Autoimmune Disease. *Clinic Rev Allerg Immunol* **42,** 102–111 (2012).

[31] Davidson, A. & Diamond, B. Autoimmune Diseases. *New England Journal of Medicine* **345,** 340–350 (2001).

[32] Benatar, M. A systematic review of diagnostic studies in myasthenia gravis. *Neuromuscular Disorders* **16,** 459–467 (2006).

[33] Schwimmbeck, P. L., Dyrberg, T., Drachman, D. B. & Oldstone, M. B. Molecular mimicry and myasthenia gravis. An autoantigenic site of the acetylcholine receptor alpha-subunit that has biologic activity and reacts immunochemically with herpes simplex virus. *J Clin Invest* **84,** 1174–1180 (1989).

[34] Van der Windt, D. A. W. M., Jellema, P., Mulder, C. J., Kneepkens, C. M. F. & van der Horst, H. E. Diagnostic Testing for Celiac Disease Among Patients with Abdominal Symptoms. *Journal of American Medical Association* **303,** 1738–1746 (2010).

[35] Pihoker, C., Gilliam, L. K., Hampe, C. S. & Lernmark, Å. Autoantibodies in diabetes. *Diabetes* **54,** S52–S61 (2005).

[36] Belle, T. L. V., Coppieters, K. T. & Herrath, M. G. V. Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies. *Physiological Reviews* **91,** 79–118 (2011).

[37] Wichainun, R. *et al.* Sensitivity and specificity of ANA and anti-dsDNA in the diagnosis of systemic lupus erythematosus: a comparison using control sera

obtained from healthy individuals and patients with multiple medical problems. *Asian Pac. J. Allergy Immunol.* **31,** 292–298 (2013).

[38] Lunardi, C. *et al.* Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. *Nat Med* **6,** 1183–1186 (2000).

[39] Ho, K. T. & Reveille, J. D. The clinical relevance of autoantibodies in scleroderma. *Arthritis Research & Therapy* **5,** 80 (2003).

[40] Costagliola, S. *et al.* Second Generation Assay for Thyrotropin Receptor Antibodies Has Superior Diagnostic Sensitivity for Graves' Disease. *The Journal of Clinical Endocrinology & Metabolism* **84,** 90–97 (1999).

[41] Corapçioğlu, D. *et al.* Relationship between thyroid autoimmunity and Yersinia enterocolitica antibodies. *Thyroid* **12,** 613–617 (2002).

[42] Benvenga, S., Santarpia, L., Trimarchi, F. & Guarneri, F. Human thyroid autoantigens and proteins of Yersinia and Borrelia share amino acid sequence homology that includes binding motifs to HLA-DR molecules and T-cell receptor. *Thyroid* **16,** 225–236 (2006).

[43] Zabriskie, J. B., Hsu, K. C. & Seegal, B. C. Heart-reactive antibody associated with rheumatic fever: characterization and diagnostic significance. *Clin Exp Immunol* **7,** 147–159 (1970).

[44] Galvin, J. E., Hemric, M. E., Ward, K. & Cunningham, M. W. Cytotoxic mAb from rheumatic carditis recognizes heart valves and laminin. *J. Clin. Invest.* **106,** 217–224 (2000).

[45] Vitozzi, S., Lapierre, P., Djilali-Saiah, I. & Alvarez, F. Autoantibody detection in type 2 autoimmune hepatitis using a chimera recombinant protein. *Journal of Immunological Methods* **262,** 103–110 (2002).

[46] Bogdanos, D.-P., Choudhuri, K. & Vergani, D. Molecular mimicry and autoimmune liver disease: virtuous intentions, malign consequences. *Liver* **21,** 225–232 (2001).

[47] Schellekens, G. A. *et al.* The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis & Rheumatism* **43,** 155–163 (2000).

[48] Wilson, C., Tiwana, H. & Ebringer, A. Molecular mimicry between HLA-DR alleles associated with rheumatoid arthritis and Proteus mirabilis as the aetiological basis for autoimmunity. *Microbes and Infection* **2,** 1489–1496 (2000).

[49] Meyer, J. M. & Ginsburg, G. S. The path to personalized medicine. *Current Opinion in Chemical Biology* **6,** 434–438 (2002).

[50] Dekker, L. J. M. *et al.* An antibody-based biomarker discovery method by mass spectrometry sequencing of complementarity determining regions. *Anal Bioanal Chem* **399,** 1081–1091 (2011).

[51] Lu, D. R. *et al.* Identification of functional anti-staphylococcus aureus antibodies by sequencing patient plasmablast antibody repertoires. *Clinical Immunology* (2014). doi:10.1016/j.clim.2012.2.010

[52] Tan, Y.-C. *et al.* Barcode-enabled sequencing of plasmablast antibody repertoires in rheumatoid arthritis. *Arthritis and Rheumatism* (2014). doi:10.1002/art.38754

[53] Vollmers, C., Sit, R. V., Weinstein, J. A., Dekker, C. L. & Quake, S. R. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *PNAS* **110,** 13463–13468 (2013).

[54] Hueber, W. *et al.* Proteomic analysis of secreted proteins in early rheumatoid arthritis: anti-citrulline autoreactivity is associated with up regulation of proinflammatory cytokines. *Ann Rheum Dis* **66,** 712–719 (2007).

[55] Sundaresh, S. *et al.* Identification of Humoral immune Responses in Protein Microarrays Using DNA Microarray Data Analysis Techniques. *Bioinformatics* **22,** 1760–1766 (2006).

[56] Talapatra, A., Rouse, R. & Hardiman, G. Protein Microarrays: challenges and promises. *Pharmacogenomics* **3,** 527–536 (2002).

[57] Kersten, B., Wanker, E. E., Hoheisel, J. D. & Angenendt, P. Multiplex approaches in protein microarray technology. *Expert Rev Proteomics* **2,** 499–510 (2005).

[58] Ramachandran, N. *et al.* Self-Assembling Protein Microarrays. *Science* **305,** 86–90 (2004).

[59] Price, J. V. *et al.* Protein microarray analysis reveals BAFF-binding autoantibodies in systemic lupus erythematosus. *The Journal of Clinical Investigation* **123,** 5135–5145 (2013).

[60] Fang, Y., Frutos, A. G. & Lahiri, J. Membrane Protein Microarrays. *J. Am. Chem. Soc.* **124,** 2394–2395 (2002).

[61] Quan, J. *et al.* Discovery of biomarkers for systemic lupus erythematosus using a library of synthetic autoantigen surrogates. *Journal of Immu* **402,** 23–34 (2014).

[62] Hecker, M. *et al.* Computational analysis of high-density peptide microarray data with application from systemic sclerosis to multiple sclerosis. *Autoimmunity Reviews* **11,** 180–190 (2012).

[63] Frank, R. & Overwin, H. in *Methods in Molecular Biology* **66,** 149–169 (Humana Press Inc., 1996).

[64] Gaseitsiwe, S. *et al.* Pattern Recognition in Pulmonary Tuberculosis Defined by High Content Peptide Microarray Chip Analysis Representing 61 Proteins from M. tuberculosis. *PLoS ONE* **3,** e3840 (2008).

[65] Reineke, U., Sabat, R., Volk, H.-D. & Schneider-Mergener, J. Mapping of the interleukin-10/interleukin-10 receptor combining site. *Protein Science* **7,** 951–960 (1998).

[66] Reineke, U. *et al.* A synthetic mimic of a discontinuous binding site on interleukin-10. *Nat. Biotechnol.* **17,** 271–275 (1999).

[67] Protein Engineering and Design. (CRC Press, 2010).

[68] Larman, H. B. *et al.* Autoantigen Discovery with a Synthetic Human Peptidome. *Nature Biotechnology* 1–7 (2011). doi:10.1038/nbt.1856

[69] Chabierski, S. *et al.* Antibody responses in humans infected with newly emerging strains of West Nile virus in Europe. *PLOS One* **8,** e66507 (2013).

[70] Fernandez-Madrid, F. et al. Autoantibodies to Annexin XI-A and Other Autoantigens in the Diagnosis of Breast Cancer. *Cancer Research* **64**, 5089-5096 (2004).

[71] Geysen, H.M., Rodda, S.J. & Mason, T.J. A Priori Delineation of a Peptide which Mimics a Discontinuous Antigenic Determinant. *Molecular Immunology* **23**, 709-715 (1986).

[72] Hudson, E. P., Uhlen, M. & Rockberg, J. Multiplex epitope mapping using bacterial surface display reveals both linear and conformational epitopes. *Sci. Rep.* **2,** (2012).

[73] Dolleweerd, C. J. van, Kelly, C. G., Chargelegue, D. & Ma, J. K.-C. Peptide Mapping of a Novel Discontinuous Epitope of the Major Surface Adhesin from Streptococcus mutans. *J. Biol. Chem.* **279,** 22198–22203 (2004).

[74] Hamby, C. V., Llibre, M., Utpat, S. & Wormser, G. P. Use of Peptide Library Screening To Detect a Previously Unknown Linear Diagnostic Epitope: Proof of Principle by Use of Lyme Disease Sera. *Clin. Diagn. Lab. Immunol.* **12,** 801–807 (2005).

[75] Wang, M. et al. Screening and Evaluating the Mimic Peptides as a Useful Serum Biomarker of Ankylosing Spondylitis Using a Phage Display Technique. *Rheumatology International* 1-8 (2010).doi:10.1007/s00296-010-1403-8

[76] Mintz, P.J. et al. Fingerprinting the Circulating Repertoire of Antibodies from Cancer Patients. *Nature Biotechnology* **21**, 57-63 (2003).

[77] Wang, X. et al. Autoantibody Signatures in Prostate Cancer. *The New England Journal of Medicine* **353**, 1224-1235 (2005).

[78] Zhong, L. et al. Using Protein microarray as a Diagnostic Assay for Non-Small Cell Lung Cancer. *American Journal of Respiratory and Critical Care Medicine* **172**, 1308-1314 (2005).

[79] Jolivet-Reynaud, C. *et al.* Specificities of Multiple Sclerosis Cerebrospinal Fluid and Serum Antibodies against Mimotopes. *Clinical Immunology* **93,** 283–293 (1999).

[80] Osman, A. A. *et al.* B cell epitopes of gliadin. *Clinical & Experimental Immunology* **121,** 248–254 (2000).

[81] Dromey, J. A. *et al.* Mapping of Epitopes for Autoantibodies to the Type 1 Diabetes Autoantigen IA-2 by Peptide Phage Display and Molecular Modeling: Overlap of Antibody and T Cell Determinants. *J Immunol* **172,** 4084–4090 (2004).

[82] Daugherty, P.S., Iverson, B.L. & Georgiou, G. Flow Cytometric Screening of Cell-Based Libraries. *Journal of Immunological Methods* **243**, 211-227 (2000).

[83] Derda, R. *et al.* Diversity of Phage-Displayed Libraries of Peptides during Panning and Amplification. *Molecules* **16,** 1776–1803 (2011).

[84] Wittrup, K.D. Protein Engineering by Cell-Surface Display. *Current Opinion in Biotechnology* **12**, 395-399 (2001).

[85] Lu, Z. *et al.* Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. *Biotechnology (N.Y.)* **13,** 366–372 (1995).

[86] Navone, R. *et al.* Identification of tear lipocalin as a novel autoantigen target in Sjögren's syndrome. *Journal of Autoimmunity* **25,** 229–234 (2005).

[87] Lunardi, C. *et al.* Autoantibodies to inner ear and endothelial antigens in Cogan's syndrome. *The Lancet* **360,** 915–921 (2002).

[88] Frulloni, L. *et al.* Identification of a Novel Antibody Associated with Autoimmune Pancreatitis. *New England Journal of Medicine* **361,** 2135–2142 (2009).

[89] Rice, J. J., Schohn, A., Bessette, P. H., Boulware, K. T. & Daugherty, P. S. Bacterial Display Using Circularly Permuted Outer Membrane Protein OmpX Yields High Affinity Peptide Ligands. *Protein Science* **15,** 825–836 (2006).

[90] Rice, J. J. & Daugherty, P. S. Directed Evolution of a Biterminal Bacterial Display Scaffold Enhances Display of Diverse Peptides. *Protein Engineering, Design & Selection* **21,** 435–442 (2008).

[91] Hall, S.S. & Daugherty, P.S. Quantitative Specificity-Based Display Library Screening Identifies Determinants of Antibody-Epitope Binding Specificity. *Protein Science* **18.** 1926-1934 (2009).

[92] Cochran, J. R., Kim, Y.-S., Olsen, M. J., Bhandari, R. & Wittrup, K. D. Domain-level antibody epitope mapping through yeast surface display of epidermal growth factor receptor fragments. *J. Immunol. Methods* **287,** 147–158 (2004).

[93] Shembekar, N. *et al.* Isolation of a high affinity neutralizing monoclonal antibody against 2009 pandemic H1N1 virus that binds at the 'Sa' antigenic site. *PLoS ONE* **8,** e55516 (2013).

[94] Krishnamurthy, V.M., Semetey, V., Bracher, P.J., Shen, N. & Whitesides, G.M. Dependence of Effective Molarity on Linker Length for an Intramolecular Protein-Ligand System. *Journal American Chemical Society* **129**, 1312-1320

[95] Kane, R.S. Thermodynamics of Multivalent Interactions: Influence of the Linker. *Langmuir* **26**, 8636-8640 (2010).

[96] Davies, J.M. et al. Multiple Alignment and Sorting of Peptides Derived from Phage-Displayed Peptide Libraries with Polyclonal Sera Allows Discrimination of Relevant Phagotopes. *Molecular Immunology* **36**, 659-667 (1999)

[97] Chenna, R. et al. Multiple Sequence Alignment with the Clustal Series of Programs. *Nucleic Acids Research* **31**, 3497-3500 (2003).

[98] Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2,** 28–36 (1994).

[99] Henikoff, S. & Henikoff, J.G. Amino Acid Substitution Matrices from Protein Blocks. *Proc Natl Acad Sci* **89**, 10915-10919 (1992).

[100] Tudos, E., Cserzo, M. & Simon, I. Prediciting Isomorphic Residue Replacements for Protein Design. *International Journal of Peptide and Protein Research* **36**, 236-239 (1990).

[101] Sibai, B., Dekker, G. & Kupferminc, M. Pre-eclampsia. *Lancet* **365**, 785-799 (2005).

[102] Redman, C.W. & Sargent, I.L. Latests Advances in Understanding Preeclampsia. *Science* **308**, 1592-1594 (2005)

[103] Thadhani, R. Inching Towards a Targeted Therapy for Preeclampsia. *Hypertension* **55**, 238-240 (2010).

[104] Gilstrap, L.C.I. & Ramin, S.M. Diagnosis and Management of Preeclampsia and Eclampsia. *ACOG Practice Bulletin: Clinical Management Guidlines for Obstetrician-Gynecologists* **33**, 1-9 (2002).

[105] Lindheimer, M.D., Roberts, J.M. & Cunningham, F.G. *Chesley's Hypertensive Disorders in Pregnancy*. (Academic Press: 2009).

[106] Myatt, L. *et al.* Strategy for standardization of preeclampsia research study design. *Hypertension* **63,** 1293–1301 (2014).

[107] Meads, C.A. et al. Methods of Prediction and Prevention of Pre-eclampsia: Systematic Reviews of Accruacy and Effectiveness Literature with Economic Modeling. *Health Technology Assessment* **12**, (2008).

[108] Kristensen, K. *et al.* Cystatin C, beta-2-microglobulin and beta-trace protein in pre-eclampsia. *Acta Obstet Gynecol Scand* **86,** 921–926 (2007).

[109] Guller, S., Tang, Z., Ma, Y.Y., Di Santo, S. & Schneider, H. Protein Composition of Microparticles Shed from HUman Placenta during Placental Perfusion: Potential Role in Angiogenesis and Fribrinolysis. *Placenta* **32**, 63-69 (2011).

[110] Kumasawa, K. et al. Pravastatin Induces Placental Growth Factor (PGF) and Amliorates Pre-eclampsia in a Mouse Model. *Proc Natl Acad Sci* **108**, 1451-1455 (2011).

[111] Venkatesha, S. et al. Soluble Endoglin Contributes to the Pathogensis of Preeclampsia. *Nature Medicine* **12**, 642-649 (2006).

[112] Maynard, S.E. et al. Excess Placental Soluble FMS-Like Tyrosine Kinase 1 (sFlt1) May Contribute to Endothelial Dysfunction, Hypertension and Proteinuria in Preeclampsia. *The Journal of Clinical Investigation* **111**, 649-658 (2003).

[113] Levine, R.J. et al. Circulating Angiogenic Factors and the Risk of Preeclampsia. *The New England Journal of Medicine* **350**, 672-683 (2004).

[114] Solomon, C.G. & Seely, E.W. Preeclampsia - Searching for the Cause. *The New England Journal of Medicine* **350**, 641-642 (2004).

[115] Zhou, A. et al. A Redox Switch in Angiotensinogen Modulates Angiotensin Release. *Nature* **468**, 108-111 (2010).

[116] Kanasaki, K. et al. Deficiency in Catechol-O-Methyltransferase and 2-Methoxyoestradiol is Associated with Pre-Eclampsia. *Nature* **453**, 1117-1123 (2008).

[117] Lamarca, B. The role of immune activation in contributing to vascular dysfunction and the pathophysiology of hypertension during preeclampsia. *Minerva Ginecol* **62,** 105–120 (2010).

[118] Salmon, J. E. *et al.* Mutations in the Complement Regulatory Proteins Predispose to Preeclampsia: A Genetic Analysis of the PROMISSE Cohort. *PLoS Medicine* **8,** e1001013 1–9 (2011).

[119] Buurma, A. *et al.* Preeclampsia is Characterized by Placental Complement Dysregulation. *Hypertension* **60,** 1332–1337 (2012).

[120] Lynch, A. M. *et al.* Alternative complement pathway activation fragment Bb in early pregnancy as a predictor of preeclampsia. *American Journal of Obstetrics & Gynecology* **198,** 385.e1–385.e9 (2008).

[121] Burwick, R. M., Fichorova, R. N., Dawood, H. Y., Yamamoto, H. S. & Feinberg, B. B. Urinary excretion of C5b-9 in severe preeclampsia: Tipping the balance of complement activation in pregnancy. *Hypertension* **62**, 1040–1045 (2013).

[122] Hering, L. et al. Trophoblasts Reduce the Vascular Smooth Muscle Cell Proatherogenic Response. *Hypertension* **51[part 2]**, 554–559 (2008).

[123] Darmochwal-Kolarz, D. *et al.* Apoptosis Signaling is Altered in CD4+CD25+FoxP3+ T Regulatory Lymphocytes in Pre-eclampsia. *International Journal of Molecular Sciences* **13,** 6548–6560 (2012)..

[124] Wallace, K. *et al.* CD4+ T cells are important mediators of oxidative stress that cause hypertension in response to placental ischemia. *Hypertension* **64,** (2014).

[125] Sakaguchi, S. *et al.* Immunologic tolerance maintained by CD25+ CD4+ regulatory T cells: their common role in controlling autoimmunity, tumor immunity, and transplantation tolerance. *Immunological Reviews* **182,** 18–32 (2001).

[126] Jensen, F. *et al.* CD19+CD5+ Cells as Indicators of Preeclampsia. *Hypertension* **59,** 861–868 (2012).

[127] Yuling, H. *et al.* CD19+CD5+ B Cells in Primary IgA Nephropathy. *JASN* **19,** 2130–2139 (2008).

[128] Ma, G. *et al.* Association between the Presence of Autoantibodies against Adrenoreceptors and Severe Pre-Eclampsia: A Pilot Study. *PLoS ONE* **8,** e57983 (2013)..

[129] Do Prado, A. D., Piovesan, D. M., Staub, H. L. & Horta, B. L. Association of Anticardiolipin Antibodies with Preeclampsia: A Systematice Review and Meta-Analysis. *Obstetrics & Gynecology* **116,** 1433–1443 (2010).

[130] Marozio, L. *et al.* Anti-Prothrombin Antibodies are Associated with Adverse Pregnancy Outcome. *American Journal of Reproductive Immunology* **66,** 404–409 (2011).

[131] Zhou, C.C. et al. Autoantibody from Women with Preeclampsia Induces Soluble Fms-Like Tyrosin Kinase-1 Production via Angiotensin Type 1 Receptor and Calcineurin/Nuclear Factor of Activated T-Cells Signalling. *Hypertension* **51**, 1010-1019 (2008).

[132] Parrish, M.R. et al. The Effect of Immune Factors, Tumor Necrosis Factor-alpha, and Agonistic Autoantibodies to the Angiotensin II Type I Receptor on Soluble fms-like Tyrosine and Soluble Endoglin Production in Response to Hypertension During Pregnancy. *American Journal of Hypertension* **23**, 911-916 (2010).

[133] Irani, R.A. et al. The Detrimental Role of Angiotensin Receptor Agonistic Autoantibodies in Intrauterine Growth Restriction Seen in Preeclampsia. *Journal of Experimental Medicine* **206**, 2809-2822 (2009).

[134] Dechend, R. et al. Agonistic Autoantibodies to the AT1 Receptor in a Transgenic Rat Model of Preeclampsia. *Hypertension* **45**, 742-746 (2005).

[135] Herse, F. et al. Prevalence of Agonistic Autoantibodies Against the Angiotensin II Type I Receptor and Soluble fms-Like Tyrosine Kinase 1 in a Gestational Age-Matched Case Study. *Hypertension* **53**, 393-398 (2009).

[136] Walther, T. et al. Angiotensin II Type 1 Receptor Agonistic Antibodies Reflect Fundamental Alterations in the Uteroplacental Vasculature. *Hypertension* **46**, 1275-1279 (2005).

[137] Dechend, R. et al. Agonistic Antibodies Directed at the Angiotensin II, AT1 Receptor in Preeclampsia. *Journal of the Society for Gynecologic Investigation* **13**, 79-86 (2006).

[138] Siddiqui, A. H. *et al.* Angiotensin Receptor Agonistic Autoantibody is Highly Prevalent in Preeclampsia. *Hypertension* **55,** 386–393 (2010).

[139] Sibai, B.M. Biomarker for Hypertension-Preeclampsia: Are We Close Yet? *American Journal of Obstetrics & Gynecology* **197**, 1-2 (2007).

[140] Cerdeira, A.S. & Karumanchi, S.A. Angiogenic Proteins as Aid in the Diagnosis and Prediction of Preeclampsia. *Scandinavian Journal of Clinical &Laboratory Investigation* **70**, 73-78 (2010).

[141] Salahuddin, S., Lee, Y., Vadnais, M., Sachs, B.P. & Karumanchi, S.A. Diagnostic Utility of Soluble Fms-like Tyrosine Kinase 1 and Soluble Endoglin in Hypertensive Diseases of Pregnancy. *American Journal of Obstetrics & Gynecology* **197**, 28.e1-28.e6 (2007).

[142] Schiettecatte, J. et al. Multicenter Evaluation of the First Automated Elecsys sFlt-1 and PlGF Assays in Normal Pregnancies and Pre-eclampsia. *Clinical Biochemistry* **43**, 768-770 (2010).

[143] Sunderji, S. et al. Automated Assays for sVEGF R1 and PlGF as an Aid in the Diagnosis of Preterm Preeclampsia: a Prospective Clinical Study. *American Journal of Obstetrics & Gynecology* **202**, 40.e1-7 (2010).

[144] Powers, R.W. et al. Soluble fms-Like Tyrosine Kinase 1 (sFlt1), Endoglin and Placental Growth Factor (PlGF) on Preeclampsia among High Risk Pregnancies. *PLoS ONE* **5**, e13263 1-12 (2010).

[145] North, R. A. *et al.* Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ* **342,** d1875–d1875 (2011).

[146] Kenny, L. C. *et al.* Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: The Screening for Pregnancy Endpoints (SCOPE) International Cohort Study. *Hypertension* **64,** 644–652 (2014).

[147] Craik, D.J. & Adams, D.J. Chemical Modification of Conotoxins to Improve Stability and Activity. *ACS Chemical Biology* **2**, 457-468 (2007).

[148] Hu, S.-H. The 1.1 A Resolution Crystal Structure of [Tyr15]EpI, a Novel α-Conotoxin from Conus episcopatus, Solved by Direct Methods. *Biochemistry* **37**, 11425-11433 (1998).

[149] Jiang, B. et al. A Novel Peptide Isolated from a Phage Display Peptide Library with Trastuzumab can Mimic Antigen Epitope of HER-2. *The Journal of Biological Chemistry* **280**, 4656-4662 (2005).

[150] Cho, H.-S. et al. Structure of the Extracellular Region of HER2 Alone and in Complex with the Herceptin Fab. *Nature* **421**, 756-760 (2003).

[151] Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *PNAS* **74,** 5463–5467 (1977).

[152] Von Bubnoff, A. Next-Generation Sequencing: The Race Is On. *Cell* **132,** 721–723 (2008).

[153] Spatola, B. N. Antibody repertoire profiling using bacterial display random peptides for biomarker discovery. (2013).

[154] Vincent, A. et al. Antibody Heterogeneity and Specificity in Myasthenia Gravis. *Annals of the New York Academy of Sciences* **505**, 106-120 (1987)

[155] Giljohann, D.A. & Mirkin, C.A. Drivers of Biodiagnostic Development. *Nature* **462**, 461-464 (2009).

[156] Sajda, P. Machine Learning for Detection and Diagnosis of Disease. *Annual Review of Biomedical Engineering* **8,** 537–565 (2006).

[157] Kohavi, R. A study of cross-validation and bootstrap accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence* **2,** 1137–1143 (1995).

[158] Cortes, C. & Vapnik, V. Support-vector networks. *Mach Learn* **20,** 273–297 (1995).

[159] Furey, T. S. *et al.* Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16,** 906–914 (2000).

[160] Spatola, B. N., Murray, J. A., Kagnoff, M., Kaukinen, K. & Daugherty, P. S. Antibody Repertoire Profiling Using Bacterial Display Identifies Reactivity Signatures of Celiac Disease. *Anal. Chem.* **85,** 1215–1222 (2013)

[161] Friedman, J., Hastie, T. & Tibshirani, R. Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics* **28,** 337–407 (2000).

[162] Ulintz, P. J., Zhu, J., Qin, Z. S. & Andrews, P. C. Improved Classification of Mass Spectrometry Database Search Results Using Newer Machine Learning Approaches. *Molecular & Cellular Proteomics* **5,** 497–509 (2006).

[163] Hu, W., Hu, W. & Maybank, S. AdaBoost-Based Algorithm for Network Intrusion Detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **38,** 577–583 (2008).

[164] Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

[165] De Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34,** W362–W365 (2006).

[166] Zhao, L., Wong, L., Lu, L., Hoi, S. C. & Li, J. B-cell epitope prediction through a graph model. *BMC* **13,** S20 (2012).

[167] Moreau, V., Granier, C., Villard, S., Laune, D. & Molina, F. Discontinuous Epitope Prediction Based on Mimotope Analysis. *Bioinformatics* **22,** 1088–1095 (2006).

[168] Cowans, N. J., Stamatopoulou, A., Khalil, A. & Spencer, K. PP13 as a Marker of Pre-eclampsia: A Two Platform Comparison Study. *Placenta* **32,** S37–S41 (2011).

[169] Benton, S. J. *et al.* Angiogenic factors as diagnostic tests for preeclampsia: a performance comparison between two commercial immunoassays. *American Journal of Obstetrics & Gynecology* **205,** 469.e1–8 (2011).

[170] Rana, S. *et al.* Angiogenic factors and the risk of adverse outcomes in women with suspected preeclampsia. *Circulation* **125,** 911–919 (2012).

[171] Schneuer, F. J. *et al.* First trimester screening of serum soluble fms-like tyrosine kinase-1 and placental growth factor predicting hypertensive disorders of pregnancy. *Pregnancy Hypertension: An Internation Journal of Women's Cardiovascular Health* (2013). doi:http://dx.doi.org/10.1016/j.preghp.2013.04.119

[172] LaMarca, B. *et al.* Autoantibodies to the Angiotensin Type I Receptor in Response to Placental Ischemia and Tumor Necrosis Factor α in Pregnant Rats. *Hypertension* **52,** 1168–1172 (2008).

[173] Wang, W. *et al.* Autoantibody-Mediated Complement C3a Receptor Activation Contributes to the Pathogenesis of Preeclampsia. *Hypertension* **60,** 712–721 (2012).

[174] Dragun, D. *et al.* Angiotensin II Type 1-Receptor Activating Antibodies in Renal Allograft Rejection. *The New England Journal of Medicine* **352,** 558–569 (2005).

[175] Casadaban, M. J. & Cohen, S. N. Analysis of gene control signals by DNA fusion and cloning in Escherchia coli. *Journal of Molecular Biology* **138,** 179–207 (1980).

[176] Kenrick, S. A. & Daugherty, P. S. Bacterial Display Enables Efficient and Quantitative Peptide Affinity Maturation. *Prot Eng Des Sel* **23,** 9–17 (2010).

[177] Culp, M., Johnson, K. & Michailidis, G. ada: An R Package for Stochastic Boosting. *Journal of Statistical Software* **17,** 1–27 (2006).

[178] Rossitto, G. *et al.* Elevation of Angiotensin-II Type-1-Receptor Autoantibodies Titer in Primary Aldosteronism as a Result of Aldosterone-Producing Adenoma. *Hypertension* **61,** 526–533 (2013).

[179] Rabel, P. O., Planitzer, C. B., Farcet, M. R. & Kreil, T. R. Tick-Borne Encephalitis Virus-Neutralizing Antibodies in Different Immunoglobulin Preparations. *Clinical and Vaccine Immunology* **19,** 623–625 (2012).

[180] Chaiworapongsa, T. *et al.* Maternal plasma concentrations of angiogenic/antiangiogenic factors in the third trimester of pregnancy to identify the patient at risk for stillbirth at or near term and severe late preeclampsia. *American Journal of Obstetrics & Gynecology* **208,** 287.e1–287.e15 (2013).

[181] Quintana, F. J. *et al.* Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis. *PNAS* **105,** 18889–18894 (2008).

[182] Blüthner, M., Schäfer, C., Schneider, C. & Bautz, F. A. Identification of Major Linear Epitopes on the sp100 Nuclear PBC Autoantigen by the Gene-Fragment Phage-Display Technology. Autoimmunity 29, 33–42 (1999).

[183] Larman, H. B. *et al.* PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. *Journal of Autoimmunity* **43,** 1–9 (2013).

[184] Anderson, R. P., Degano, P., Godkin, A. J., Jewell, D. P. & Hill, A. V. S. In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. *Nat Med* 6, 337–342 (2000).

[185] Öling, V. *et al*. GAD65- and proinsulin-specific CD4+ T-cells detected by MHC class II tetramers in peripheral blood of type 1 diabetes patients and at-risk subjects. *Journal of Autoimmunity* **25**, 235–243 (2005).

[186] Von Büdingen, H.-C. *et al*. B cell exchange across the blood-brain barrier in multiple sclerosis. *Journal of Clinical Investigation* **122**, 4533–4543 (2012).

[187] Clemente, M. J. *et al*. Deep sequencing of the T-cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. *Blood* **122**, 4077–4085 (2013).

[188] Krell, P. F. I. *et al*. Next-generation-sequencing-spectratyping reveals public T-cell receptor repertoires in pediatric very severe aplastic anemia and identifies a β chain CDR3 sequence associated with hepatitis-induced pathogenesis. *Haematologica* **98**, 1388–1396 (2013).

[189] Ballew, J. T. *et al.* Antibody biomarker discovery through in vitro directed evolution of consensus recognition epitopes. *Proc Natl Acad Sci U S A* **110,** 19330–19335 (2013).

[190] Sharma, A., Satyam, A. & Sharma, J. B. Leptin, IL-10 and Inflammatory Markers (TNF-α, IL-6 and IL-8) in Pre-Eclamptic, Normotensive Pregnant and Healthy Non-Pregnant Women. *American Journal of Reproductive Immunology* **58,** 21–30 (2007).

[191] Elliott, S. E. *et al.* Characterization of antibody specificities associated with preeclampsia. *Hypertension* **63,** 1086–1093 (2014). Wolters Kluwer Health Lippincott Williams & Wilkins

[192] Cox, B. *et al.* Translational analysis of mouse and human placental protein and mRNA reveals distinct molecular pathologies in human preeclampsia. *Molecular & Cellular Proteomics* **10,** M111.012526 (2011).

[193] Pagano, J. S. Epstein-Barr Virus: The First Human Tumor Virus and its Role in Cancer. *Proceedings of the Association of American Physicians* **111,** 573–580 (1999).

[194] Haeri, S. *et al.* Maternal Depression and Epstein-Barr Virus Reactivation in Early Pregnancy. *Obstetrics & Gynecology* **117,** 862–866 (2011).

[195] Costa, S. *et al.* Detection of active Epstein-Barr infection in pregnant women. *Eur. J. Clin. Microbiol.* **4,** 335–336 (1985).

[196] Enquobahrie, D. A. *et al.* Differential placental gene expression in preeclampsia. *American Journal of Obstetrics & Gynecology* **199,** 566.e1–566.11 (2008).

[197] Huber, M. *et al.* IL-27 inhibts the development of regulatory T cells via STAT3. *International Immunology* **20,** 223–234 (2008).

[198] Levoye, A. *et al.* The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. *EMBO J* **25,** 3012–23 (2006).

[199] Lanoix, D., Lacasse, A.-A., Reiter, R. J. & Vaillancourt, C. Melatonin: The watchdog of villous trophoblast homeostasis against hypoxia/reoxygenation-induced oxidative stress and apoptosis. *Molecular and Cellular Endocrinology* **381,** 35–45 (2013).

[200] Paulis, L. & Simko, F. Blood pressure modulation and cardiovascular protection by melatonin: potential mechanisms behind. *Physiol. Res.* **56,** 671–684 (2007).

[201] Lanoix, D., Guerin, P. & Vaillancourt, C. Placental melatonin production and melatonin receptor expression are altered in preeclampsia: new insights into the role of this hormone in pregnancy. *Journal of Pineal Research* **53,** 417–425 (2012).

[202] Bhattacharyya, S. *et al.* Sequence variants in the melatonin-related receptor gene GPR50 associate with circulating triglycride and HDL levels. *Journal of Lipid Research* **47,** 761–766 (2006).

[203] Gallos, I. *et al.* Pre-eclampsia is associated with, and preceded by, hypertriglyceridaemia: a meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology* **120,** 1321–1332 (2013).

[204] Heitner, J. C. *et al.* Differentiation of HELLP patients from healthy pregnant women by proteome analysis—On the way towards a clinical marker set. *Journal of Chromatography B* **840,** 10–19 (2006).

[205] Wen, Q. *et al.* Peptidomic Identification of Serum Peptides Diagnosing Preeclampsia. *PLoS ONE* **8,** e65571 (2013).

[206] Dewey, F. E. *et al.* Clinical interpretation and implications of whole-genome sequencing. *JAMA* **311,** 1035–1045 (2014).

[207] Matochko, W. L., Chu, K., Lee, S. W., Whitesides, G. M. & Derda, R. Deep Sequencing analysis of phage libraries using Illumina platform. *Methods* **58,** 47–55 (2012).

[208] Liu, X. *et al.* Serum Antibody Repertoire Profiling Using In Silico Antigen Screen. *PLoS ONE* **8,** e67181 (2013).

[209] Hietpas, R., Roscoe, B., Jiang, L. & Bolon, D. N. A. Fitness analyses of all possible point mutations for regions of genes in yeast. *Nature Protocols* **7,** 1382–1396 (2012).

[210] Whitehead, T. A. *et al.* Optimization of affinity,specifcity, and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol* **30,** 543–548 (2012).

[211] Cho, M. *et al.* Quantitative selection fo DNA aptamers through microfluidic selection and high-throughput sequencing. *PNAS* **107,** 15373–15378 (2010).

[212] Benichou, J., Ben-Hamo, R., Louzoun, Y. & Efroni, S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* **135,** 183–191 (2012).

[213] Tan, Y.-C. *et al.* High-throughput sequencing for natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clinical Immunology* **151,** 55–65 (2014).

[214] Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* **114,** 4099–4107 (2009).

[215] 't Hoen, P. A. C. *et al.* Phage display screening without repetitious selection rounds. *Analytical Biochemistry* **421,** 622–631 (2012).

[216] Illumina. 16S Metagenomic Sequencing Library Preparation. (2013). at <support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf>

[217] Cheng, H. M., Foong, Y. T., Sam, C. K., Prasad, U. & Dillner, J. Epstein-Barr virus nuclear antigen 1 linear epitopes that are reactive with immunoglobulin A (IgA) or IgG in sera from nasopharyngeal carcinoma patients or from healthy donors. *J Clin Microbiol* **29,** 2180–2186 (1991).

[218] Randhawa, R. & Cohen, P. The role of the insulin-like growth factor system in prenatal growth. *Molecular Genetics and Metabolism* **86,** 84–90 (2005).

[219] Halhali, A. *et al.* Preeclampsia Is Associated with Low Circulating Levels of Insulin-Like Growth Factor I and 1,25-Dihydroxyvitamin D in Maternal and Umbilical Cord Compartments. *The Journal of Clinical Endocrinology & Metabolism* **85,** 1828–1833 (2000).

[220] Errasti, A. E. *et al.* Expression and functional evidence of the prostaglandin F2alpha receptor mediating contraction in human umbilical vein. *Eur. J. Pharmacol.* **610,** 68–74 (2009).

[221] De Groot, C. J. M., Murai, J. T., Vigne, J.-L. & Taylor, R. N. Eicosanoid secretion by human endothelial cells exposed to normal pregnancy and preeclampsia plasma in vitro. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **58,** 91–97 (1998).

[222] Dechend, R. *et al.* Activating autoantibodies against the AT1-receptor in vascular disease. *Transplantationsmedizin* **24,** 20–26 (2012).

[223] Wang, W. *et al.* Excess LIGHT Contributes to Placental Impairment, Increased Secretion of Vasoactive Factors, Hypertension, and Proteinuria in Preeclampsia. *Hypertension* **63,** 595–606 (2014).

[224] Ware, C. F. NETWORK COMMUNICATIONS: Lymphotoxins, LIGHT, and TNF. *Annual Review of Immunology* **23,** 787–819 (2005).

[225] Liu, C. *et al.* Tissue transglutaminase contributes to the pathogenesis of preeclampsia and stabilizes placental angiotensin receptor type 1 by ubiquitination-preventing isopeptide modification. *Hypertension* **63,** 353–361 (2014).

[226] Ambrosino, D. M. *et al.* Selective defect in the antibody response to Haemophilus influenzae type b in children with recurrent infections and normal serum IgG subclass levels. *Journal of Allergy and Clinical Immunology* **81,** 1175–1179 (1988).

[227] Sivalingam, G. N. & Shepherd, A. J. An analysis of B-cell epitope discontinuity. *Mol Immunol* **51,** 304–309 (2012).