Anti-microbial Antibody Signature in IBD Patients and Common Autoantibodies Shared by Healthy Individuals

by

Mahasish Shome

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2021 by the
Graduate Supervisory Committee:

Joshua LaBaer, Chair
Chad Borges
Nicholas Stephanopoulos

ARIZONA STATE UNIVERSITY

December 2021

ABSTRACT

Antibodies are the immunoglobulins which are secreted by the B cells after a microbial invasion. They are stable and stays in the serum for a long time which makes them an excellent biomarker for disease diagnosis. Inflammatory bowel disease is a type of autoimmune disease where the immune system mistakenly attacks the commensal bacteria and leads to inflammation. We studied antibody response of 100 Crohn's disease (CD), 100 ulcerative colitis (UC) and 100 healthy controls against 1,173 bacterial and 397 viral proteins. We found some anti-bacterial antibodies higher in CD compared to controls while some antibodies lower in UC compared to controls. We were able to build biomarker panels with AUCs of 0.81, 0.87, and 0.82 distinguishing CD vs. control, UC vs. control, and CD vs. UC, respectively. Subgroup analysis based on the Montreal classification revealed that penetrating CD behavior (B3), colonic CD location (L2), and extensive UC (E3) exhibited highest antibody reactivity among all patients. We also wanted to study the reason for the presence of autoantibodies in the sera of healthy individuals. A meta-analysis of 9 independent biomarker study was performed to find 77 common autoantibodies shared by healthy individuals. There was no gender bias; however, the number of autoantibodies increased with age, plateauing around adolescence. Molecular mimicry likely contributed to the elicitation of a subset of these common autoantibodies as 21 common autoantigens had 7 or more ungapped amino acid matches with viral proteins. Intrinsic properties of protein like hydrophilicity, basicity, aromaticity, and flexibility were enriched for common autoantigens. Subcellular localization and tissue expression analysis indicated the sequestration of some autoantigens from circulating autoantibodies can explain the absence of autoimmunity in these healthy individuals.

DEDICATION

I dedicate this dissertation to my parents for their unconditional support and love.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Joshua LaBaer for his valuable guidance throughout my PhD journey. He has motivated me to think about different research ideas which helped me to develop as an independent researcher. His guidance on writing had a very positive impact on me. Other professors from our center, Dr. Ji Qiu, Dr. Yunro Chung, and Dr. Jin G. Park who have helped me in broadening my understanding of science. I would also like to thank my committee members, Dr. Chad Borges, and Dr. Nicholas Stephanopoulos, for their time and advice. Members of LaBaer lab with whom I had the pleasure of working, particularly Dr. Lusheng Song, Kailash Karthikeyan, Jennifer Hesterman and Inam Ridha.

And finally, friends who became like family during my PhD journey, Subhadeep Purkayastha, Neha Sagar, Abhishek Debnath, Sayalee Joshi, Soma Chaudhary, Tamojit Chakraborty, Abesh Banerjee, Mohammed Imtiazur Rahman, Mohammed Alamin and Ankush Tyagi without whom my PhD would not be complete.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

Abbreviation	Full name
ACPA	Anti-citrullinated protein antibodies
AIEC	Adhesion-invasive E. coli
AMCA	Aminomethylcoumarin Acetate
ANA	Antinuclear antibodies
anti-OmpC	anti-Outer membrane protein C
AP	Alkaline phosphatase
APC	Antigen presenting cell
APTES	(3-Aminopropyl) triethoxysilane
ASCA	Anti-Saccharomyces cerevisiae antibodies
ATG16L1	Autophagy-related 16-like 1
AUC	Area under the ROC curve
BCR-ABL	Breakpoint cluster region-Abelson
BLAST	Basic Local Alignment Search Tool
BRAF	v-Raf murine sarcoma viral oncogene homolog B
BRCA1	Breast Cancer 1
BS3	Bis (sulfosuccinimidyl) suberate
BSA	Bovine serum albumin
CD	Crohn's Disease
CD4	Cluster of differentiation 4

CH	Constant heavy chain
CL	Constant light chain
CNS	Central nervous system
CRP	C-reactive protein
DC	Dendritic cells
DLA	Dog leukocyte antigen
DM	Diabetes mellitus
EBV	Epstein-Barr virus
EDC	Ethyl (dimethylaminopropyl) carbodiimide
ELISA	Enzyme Linked Immunosorbent Assay
F. prausnitzii	Faecalibacterium prausnitzii
Fab	Fragment antigen-binding
Fc	Fragment crystallizable
FDR	False discovery rate
FITC	Fluorescein Isothiocyanate
GAD65	Glutamic acid decarboxylase 65
GADA	Glutamic acid decarboxylase antibodies
GSEA	Gene set enrichment analysis
GST	Glutathione S-transferases
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association studies
H. pylori	Helicobacter pylori
HER-2	Human epidermal growth factor receptor 2

HLA	Human leukocyte antigen
HPA	hypothalamo-pituitary-adrenal
HRP	Horseradish peroxidase
IA2	Insulinoma associated protein 2
IAA	Insulin autoantibodies
IBD	Inflammatory Bowel Disease
ICA	Islet cells antibodies
IEDB	Immune Epitope Database
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IL-1β	Interleukin-1 beta
iTOL	interactive Tree of Life
IVTT	*in-vitro* transcription and translation
K_D	Equilibrium dissociation constant
LCMV	Lymphocytic choriomeningitis virus
LDL	Low-Density Lipoproteins
MBP	Myelin basic protein
MES	4-morpholinoethanesulfonic acid
MHC	Major Histocompatibility Complex
MNI	Median normalized intensity
NAbs	Natural antibodies
NAPPA	Nucleic Acid Programmable Protein Array
NCBI	National Center for Biotechnology Information

NHS	*N*-Hydroxysuccinimide
NLR	Nod-like receptor
NOD2	Nucleotide-binding oligomerization domain containing 2
OR	Odds ratio
pANCA	Perinuclear anti-neutrophil cytoplasmic antibodies
PRR	Pattern recognition receptors
PSA	Prostate specific antigen
RA	Rheumatoid arthritis
RAPID	Rapid Antigenic Protein in Situ Display
ROC	Receiver-operating characteristics
SCFA	Short-chain fatty acids
SDS	Sodium dodecyl sulfate
SIgA	Secretory IgA
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SPR	Surface Plasmon Resonance
T1D	Type 1 diabetes
TAA	Tumor-associated antigens
TCR	T cell receptors
TNF	Tumor necrosis factor
TPM	Transcripts per million
TRITC	Tetramethyl rhodamine
UC	Ulcerative colitis

VH	Variable heavy chain
VL	Variable light chain
ZNT8	Zinc transporter 8

CHAPTER 1

1. INTRODUCTION AND MOTIVATION

1.1 Immune Response to Foreign Invasion

Antibodies are the immunoglobulins which are secreted by the B cells in response to an invasion by an antigen. When an infection occurs, both parts of the immune system (innate and adaptive) get activated and they work closely to eradicate the infection. Innate immune system is the first line of defense against germs entering the body. It comprises the cells of innate immune system like macrophages and neutrophils, which act fast and recognize common microorganisms to eradicate. However, they cannot always eliminate infectious organisms, especially pathogens which they can't recognize. This is when a more specialized part of the immune response (adaptive) is initiated. Adaptive immune system is induced by antigen-presenting cell (APC) that activates pathogen-specific lymphocytes in the lymph node. These pathogen-specific lymphocytes can bind to the pathogen and trigger a cascade of actions that clears off the pathogen.

B cells and T cells are the part of adaptive immune system which secretes antibodies and cytokines, respectively. Once the B cell is activated by APC, it matures and differentiates into plasma cells and memory B cells. Plasma cells can secrete antibodies for years while memory B cells gets activated once the same antigen is detected in future. This property of B cells to remember past infections make them behave like a "log-book" which registers all the pathogens it encountered. As a new pathogen is encountered, the pathogen-specific antibody increases in level and then fades

out as the infection is cleared out. This increase and decrease in the antibody level make them as a good biomarker.

A biomarker is a measurable substance which is indicative of disease, infection, or environmental exposure (Strimbu & Tavel, 2010). Tumor size (radiographic), blood pressure (physiologic), blood glucose (molecular) are different examples of biomarkers. Histologic and radiographic biomarkers require invasive procedure or expensive apparatus while blood based molecular biomarkers are non-invasive and cheap to analyze. Common examples of blood based biomarkers are C-reactive protein (CRP) for inflammation, P53 gene for cancer, and low-density lipoproteins (LDL) for cholesterol. Antibody based biomarkers are also a part of the blood based biomarkers and they can be easily detected using an immunoassay.

1.2 Early Detection of Disease and Risk Management

Early detection of disease is key for the proper risk management and survival of patients with dangerous disease like cancer. A woman with ovarian cancer detected in stage I has 90% 5-year survival rate while detection in stage III reduces the 5-year survival rate to 28% (Horner MJ, 2008). Similarly, colon cancer detected early has a 91% 5-year survival rate, vs. an only 11% survival rate if it is detected late and has spread to other organs (Horner MJ, 2008). In addition, autoimmune disease like Inflammatory Bowel Disease (IBD) can lead to colorectal cancer if not controlled in the right time. These data demonstrate the importance of an easy and accessible diagnostic technique which do not require advanced medical apparatus. Antibody based biomarkers have this potential to accurately predict the disease condition with minimal invasion.

1.3 Antibodies as Biomarker of Chronic Disease

Antibodies can predict the state of chronic diseases like Crohn's disease (CD), Ulcerative colitis (UC), multiple sclerosis, and rheumatoid arthritis. The chronic diseases usually require long-term immunosuppressive therapies with undesirable side effects (Katsanos et al., 2017). During the period of therapy, it is important to check the effect of the therapy and decide on continuing or trying a new therapy based on side effects. Using an invasive biopsy sampling can be of more burden and risk to the patient which can be avoided by using a non-invasive or minimally invasive biomarker from the bodily fluid. Antibodies as biomarker can help in the decision making process during the disease therapy course. Often, a single biomarker is not enough to understand the prognosis and a set of biomarkers provide better picture of the disease status. For instance, ASCA, pANCA, anti-OmpC, anti-flagellin, anti-I2 are some of the known biomarkers used for CD diagnosis.

Inflammatory Bowel Disease (IBD) is an umbrella term which represents two disease types: CD and UC. The disease is caused when the immune system mistakenly attacks the commensal microbes of our gut leading to inflammation in the gut. It is directly linked with the composition of the gut microbes, and they play a crucial role in priming and inducing tolerance to innocuous antigens. Our gut is colonized by around 1000 bacterial species (Eisenstein, 2020) and therefore a lot of the bacterial species role in the disease are still unknown. Looking at the scale of the number of proteins, we need a high-throughput method to screen through various bacterial species proteins which were never studied.

The antibodies in sera are detected using different techniques like Enzyme Linked Immunosorbent Assay (ELISA), Western blot, and Luminex Beads assay. However, all the above-mentioned techniques are of low-throughput and lacks upscaling capabilities. Therefore, we used an innovative protein microarray named Nucleic Acid Programmable Protein Array (NAPPA) which circumvents the issue of low-throughput and is highly scalable. For comparison, our microarrays can analyze up to 24,000 proteins simultaneously within 12 hours while the above-mentioned techniques can analyze around hundred proteins in the same time (Takulapalli et al., 2012).

1.4 Autoantibodies in Healthy Individuals

Autoantibodies are the immunoglobulins that targets self-antigens. The B cell clones which are self-reactive are removed during the clonal deletion mechanism. However, some B cell clones escape this mechanism and secrete autoantibodies. It was earlier believed that autoantibodies are only found in people with autoimmune diseases. However, it is now known that they are not specific for autoimmune patients but are frequently available in healthy individuals as well. Their presence in healthy individuals raises the question of false discovery of antibody based biomarkers which are ubiquitously available in human. This issue demands the documentation of all autoantibodies which are frequently found in healthy individuals to avoid false positives during biomarker discovery.

The presence of these autoantibodies also begs the questions like: What mechanism elicits these autoantibodies? Why don’t these autoantibodies cause any disease? How are they effected by age and gender? The answer to these questions will

provide a holistic understanding of these autoantibodies in healthy individuals and thereby helping in finding more robust biomarkers for disease prognosis.

1.5 Structure of Dissertation

The work presented in this dissertation demonstrates the use of versatile platform of NAPPA to discover the antibody signature found in IBD patients, healthy individuals, and canine diabetes. These antibodies can reveal the aspect of IBD pathogenesis, the role of autoantibodies in healthy individuals and assist in early detection of canine diabetes.

In chapter 2, I introduce the background and literature review on antibodies, how they are elicited, natural antibodies and their distinction from adaptive antibodies, antibodies as biomarker of disease, various techniques to determine the antibody level in serum, IBD and its causes, phenotypes of IBD and canine diabetes.

In chapter 3, I introduce the in-depth study of IBD to discover microbial signatures in the sera. A first of its kind study to analyze proteins from different bacteria and viruses and study their antibody response in IBD patients. The multi-antibody panels to distinguish CD vs. controls, UC vs. controls and CD vs. UC using immunoassay data is explained. The association of number of markers high in different IBD phenotype is accessed. Lastly, the correlation between the occurrence of autoantibodies and microbial antibodies in CD samples is explained.

In chapter 4, the concept of common autoantibodies in healthy individuals is introduced. The effects of age and gender on these autoantibodies in healthy is explained. The correlation of occurrence of these common antibodies among themselves was accessed. The reason for the elicitation of these autoantibodies is explained, both from a

molecular mimicry standpoint and intrinsic property standpoint. Lastly, the possible reason for not causing any autoimmune disease is analyzed.

In chapter 5, I introduce the autoantibodies found in canine diabetes. A protein microarray fabrication having human proteins relevant to diabetes is explained. The antibody reactivity profile was used to build multi-antibody panels to distinguish diabetic vs. controls. Identity of individual markers found, and their role is explained.

CHAPTER 2

2. BACKGROUND AND LITERATURE REVIEW

2.1 Function and Structure of Antibody Molecule

Antibodies are immunoglobulins that bind specifically to antigens and initiate a cascade of immune response to fight pathogens (Sela-Culang et al., 2013). The ability of antibodies to bind virtually to any foreign molecule with high specificity and high affinity makes them a key part of our immune system (Sela-Culang et al., 2013). The antibody molecule has two different parts: fragment antigen-binding (Fab) and fragment crystallizable (Fc) region (**Fig. 2.1**). Fab region of the antibody binds to the antigen while Fc region is the tail region of the antibody that interacts with the cell surface receptors, and because this part of the antibody is common regardless of the target, it can be exploited (Chiu et al., 2019).

When a pathogen is encountered, antibodies can bind to the pathogen in a fashion which blocks the physical binding of the pathogen with its receptor. This coating of the pathogen with antibodies makes it physically nonfunctional and the process is known as neutralization. The neutralized pathogen can now be eradicated by cytotoxic cells, phagocytosis, or complement-mediated lysis. Cytotoxic T cells look for infected cells and check for the pieces of protein presented on the surface of the cell through class I major histocompatibility complex proteins (MHC class I). Cytotoxic T cells have T cell receptors (TCR) on their surface which can recognize the presented protein as foreign and then kill the infected cell using cytotoxic factors. In phagocytosis, the neutralized

pathogen binds to Fc receptors on the surface of phagocytic cells, which then engulfs the neutralized pathogen and destroy it.

Antibody structure comprises of four polypeptides – two heavy chains and two light chains joined together to form Y-shaped molecule. Both heavy chains and light chains have variable (VL, VH) and constant (CL, CH1, CH2, CH3) domains (Schroeder & Cavacini, 2010). The amino acid sequence at the variable domains varies greatly among antibodies while constant domains have essentially the same amino acid sequence in all antibodies of the same isotype (IgG, IgM, IgA, IgD, IgE). Fab region at the tip of the antibody comprises variable domains from both heavy and light chain.

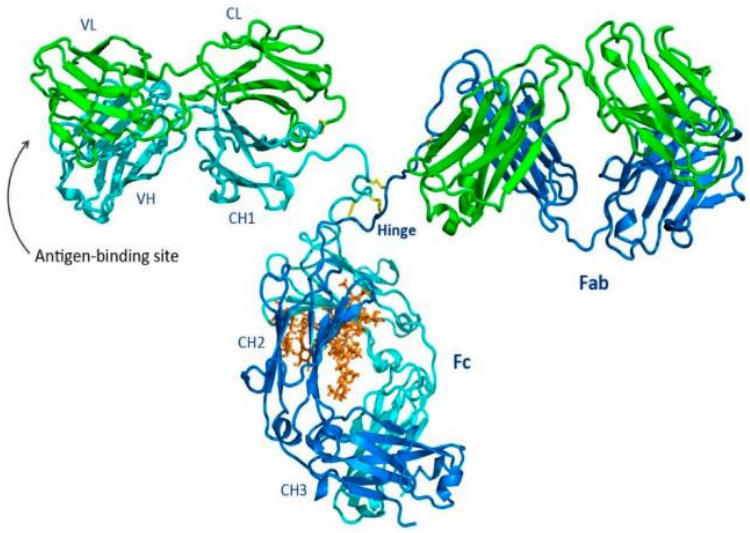


Fig. 2.1 Ribbon Representation of IgG Antibody. The light chains are represented as green color while the heavy chains are represented as cyan and blue color. Fab is the antigen binding region while Fc region interacts with the cell surface receptor. Image taken from (Chiu et al., 2019).

There are five isotypes of antibodies with different function and relative abundance. IgG isotype has the highest relative abundance of 70-85% in normal human serum and the longest half-life of 20-24 days among antibody isotypes (Cruse & Lewis, 2010). IgM is pentamer isotype predominant in primary immune response and accounts for 5-10% relative abundance in human blood (Cruse & Lewis, 2010). On the other hand, IgA is dimeric isotype found mostly in mucous secretion such as saliva, tears, milk, and intestinal juices. They are also prevalent in mucus layer of human gut.

Another clinically significant property of antibodies is their binding affinity. The equilibrium dissociation constant (K_D) between the antigen and antibody denotes the binding affinity. Typical antibody K_D value lies in low micromolar (10^{-6}) to nanomolar (10^{-9}) range (Landry et al., 2015). There are also antibodies with very high affinity and their K_D value lies in picomolar (10^{-12}) range (Landry et al., 2015). IgG isotype has the highest binding affinity while IgM has the highest binding avidity. The high abundance and strong affinity of IgG makes it more clinically significant over other isotypes for biomarker study.

2.2 Mechanism of Antibody Elicitation

Antibodies are elicited by our immune system as a defense mechanism to tag the pathogen for the further process of complete eradication. When a pathogen enters our body, the innate immune system tries to eradicate the pathogen by recruiting the APC like dendritic cells, macrophages and B cells (Gaudino & Kumar, 2019). APC can detect, engulf, and phagocytose the pathogen leading to many fragments of the pathogenic antigen. Antigen fragments are then transported to the surface of the APC and presented

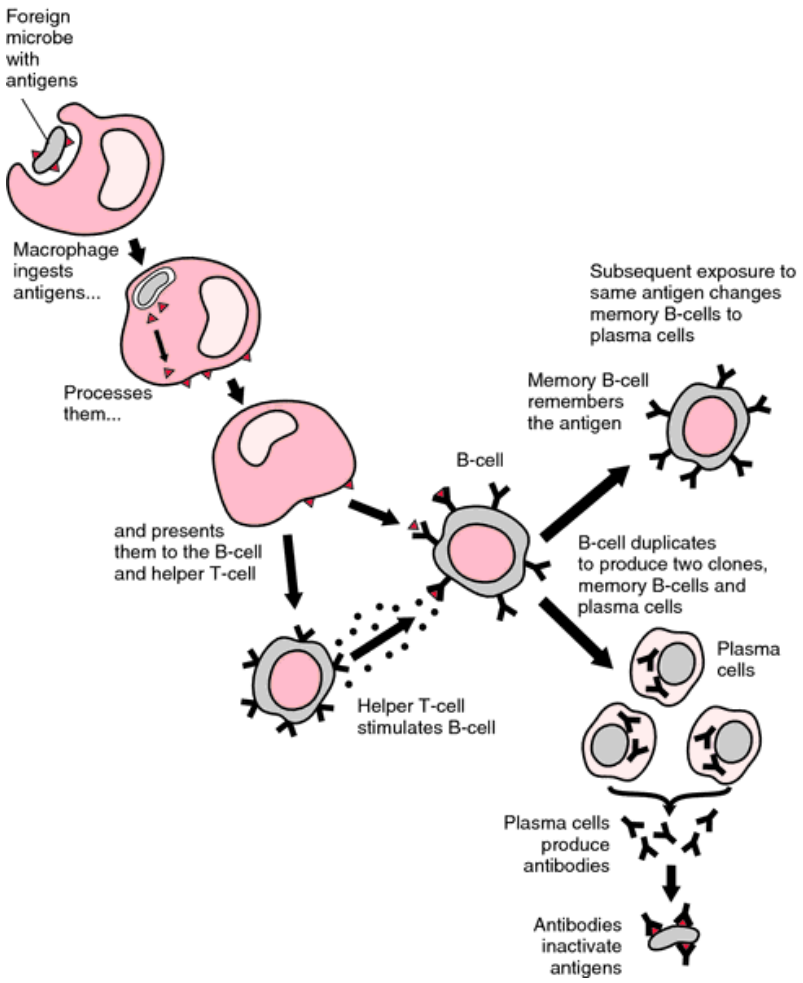


Fig 2.2 Mechanism of Antibody Elicitation During a Pathogen Invasion. Taken from Applegate, 2000.

in conjunction with class II major histocompatibility complex (MHC class II) molecule where it can interact with T cell receptors (Cruse et al., 2004). Helper T cells ($CD4^+$) stimulates B lymphocytes to produce antibodies (**Fig 2.3**). This type of activation occurs for proteins as T cells can only recognize peptides derived from proteins. The antibodies produced through this manner undergo isotype switching, and affinity maturation. The

activated B cell then differentiates into long-lived plasma cells and memory B cells. The plasma cells can produce antibodies in large amount.

2.3 Antibody Response to Intestinal Microbiota

Intestinal microbiota plays an important role in intestinal homeostasis, immune regulation, and disease pathogenesis. The microbiota and the immune system co-evolves right from birth with a symbiotic relationship (Hooper et al., 2012). The host provides the nutrients to the microbiota while the microbiota provides essential enzymes to digest the food. Due to environmental exposure, the intestinal microbiota rapidly expands and in turn also develops the immune system by priming with the commensal bacteria. As the microbiota expands, it reaches up to 100 trillion bacteria which is 10 times more than the number of cells in human body (Yu, 2015). Because of the huge density of commensal bacteria, the pathogenic bacteria have to compete to colonize. This competitive colonization plays a protective role to the host. Even if they successfully colonize, they have to compete for nutrients to survive.

This huge number of microbes is separated from our blood vessels by a single layer of epithelial cells. As the epithelial barrier is also responsible for the absorption of nutrients, hence occasional penetration of microbes along with nutrients is unavoidable. If the microbe is commensal then it is phagocytosed and destroyed by macrophages without strong antibody response. However, if it is pathogenic then it avoids phagocyte biocidal activity, and thereby initiates inflammation. Commensals live within intestinal lumen or mucus coat barrier, while pathogens either binds to the epithelial surface or penetrates it (Macpherson et al., 2005).

Usually microbes do not penetrate through the epithelial barrier and they do not directly produce a wide range of antibodies, they rather sensitize the immune system to produce antibodies if the microbes get into the bloodstream (Bern, 2020). However, in the case of an infection, the microbes can get into the bloodstream which will initiate a strong antibody response. The antibodies generated due to microbial invasion in bloodstream are of IgG isotype which are class switched and affinity matured. During the process, APCs engulf the invaded pathogen, digest the proteins into peptides which are presented to T cell and B cells for activation. As the process involves digestion of proteins, both cytoplasmic and surface proteins are equally likely to be presented. In fact, a study focused on B cell repertoire found when germ-free mice were exposed to microbiota intravenously; it led to the generation of diverse repertoire of B cells which secreted IgG to both cytoplasmic and cell-surface proteins (Li et al., 2020). When the same germ-free mice were exposed to microbiota orally, the IgA response was predominately against cell-surface antigens. These observations show that IgA repertoire do not diversify a lot and is located predominately in mucus layer while IgG is mainly found in serum which diversifies extensively to bind to target antigens.

After an infection, activated B cells differentiate into plasma cells and memory B cells. Long-lived plasma cells in the bone marrow can secrete antibodies for years while memory B cells can start producing antibodies if the same pathogen re-infects (Ahuja et al., 2008). For instance, long-lived plasma cells secrete *H. pylori*-specific antibodies even 2 years after bacterial eradication (Miernyk et al., 2007). The mean *H. pylori*-specific IgG declined 43% in participants in 2 years, but they were still seropositive. Participants who got re-infected with *H. pylori* during the study were removed from the study to segregate

the effect of memory B cell activation. There are also examples of antibodies which persist throughout life, like anti-EBV nuclear antigen (EBNA). Epstein-Barr virus (EBV) infects B cells and can survive in the infected cell by expression of few selected viral genes. The expression of these genes leads to the immortalization of the B cells and their transformation into proliferating blast (De Paschale & Clerici, 2012). The lifelong survival of EBV guarantees the persistent level of anti-EBV antibodies in the serum due to continuous exposure to the antigen. In contrast to long-lived plasma cells, memory B cells can survive for decades without the need for continuous interaction with the antigen (Crotty et al., 2003).

2.4 Antibodies as Biomarker for Disease Diagnosis

A biomarker is any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease (Strimbu & Tavel, 2010). They can provide an indication of medical state observed outside the patient. The advantage of biomarker lies in its easier and less expensive measurement of the clinical endpoint. A good biomarker should be easily measurable and correlated well with the state of the clinical outcome (Aronson & Ferner, 2017). They help epidemiologists, physicians, and scientists to understand the prediction, cause, diagnosis, progression, or outcome of the treatment of the disease (Mayeux, 2004). Biomarkers can have molecular (plasma, cerebrospinal fluid), histologic (staging of cancer), radiographic (tumor size), or physiological (blood pressure) characteristics. They can be divided into various categories based on application: diagnostic, predictive, monitoring, risk, prognostic, and safety. A diagnostic biomarker detects or confirms the

presence of a disease or a condition. Predictive biomarkers are genomic markers which indicate the likelihood of an individual to progress to serious disease condition if exposed to medical product or an environmental agent (Burckart & Green, 2018). A monitoring biomarker is measured serially for assessing the status of the disease.

The advantage of molecular biomarkers (serum) over other biomarker (tumor size) lies in the non-invasive or minimally invasive nature. Other biomarkers require imaging techniques to measure the composition and structure of the body part, which is expensive and requires advanced medical equipment. Serum based biomarkers can be a normal protein, a piece of genetic material or an immunoglobulin. Renal tubular proteins, and inflammatory proteins can reveal the functioning of kidney and bladder (Koyner et al., 2010). Several cerebrospinal fluid biomarkers can reveal the brain physiology (Limbrick Jr et al., 2017) while some of the blood based genomic biomarkers of cancer are: BRCA1 / BRCA2 (breast and ovarian cancer), BRAF V600E (melanoma), HER-2 (breast cancer), PSA (prostate cancer), BCR-ABL (leukemia) (Bedi et al., 1994; Long et al., 2011; Mehrgou & Akouchekian, 2016; Nordström et al., 2018).

Antibodies can be classified into autoantibodies and anti-microbial antibodies based on the target antigen. If the target antigen is a self-protein, then it is an autoantibody while if the target antigen is foreign microbial antigen, then it is an anti-microbial antibody. Antibodies are mainly produced in autoimmune diseases, infectious diseases, and cancers. In autoimmune disease, the autoantibodies target self-antigens which can be found in multiple location (systemic autoimmunity) in the body or can be found in a single organ (organ-specific autoimmunity). In infectious disease, the antibodies are produced against the pathogen while in cancer the antibodies are produced

against tumor-associated antigens (TAA). Apart from the above mentioned diseases, antibodies are also found in cardiovascular diseases and neurodegenerative diseases (Aziz et al., 2018). However, it is not yet clear if these antibodies play a direct role in the pathology, or they are merely symptomatic of disease. Irrespective of a causative or symptomatic role of circulating antibodies, it is worth exploring their clinical potential as diagnostic biomarker.

There are some clear advantages of antibodies as biomarker. Compared to other proteins, antibodies are long lived with limited proteolysis and the amount of antibodies produced is huge even though the target antigen is small in quantity providing a form of natural amplification of signal (Pedersen & Wandall, 2011).

2.5 Detection of Antibodies Using Immunoassays

An immunoassay is a bioanalytical technique to determine the presence or concentration of an analyte using antibody or antigen as recognition agent (**Fig. 2.5**). The analyte is usually a protein, although other molecules can also be used as long as a proper antibody is there for recognition. The importance of immunoassays comes from the ability of an antibody to recognize its target antigens from a pool of other antigens with high specificity. There are two types of immunoassays: labeled and unlabeled. In labeled immunoassay, a detectable label in the form of enzyme, fluorophore, radioactive substance or a chemiluminescent is used while in unlabeled immunoassay the change in an intrinsic property due to binding is detected.

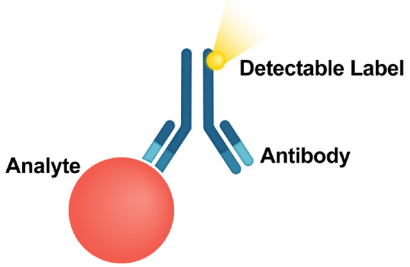


Fig 2.3 Components of Immunoassay. An analyte, an antibody, and a label for detection.

In an enzyme-based label, enzyme such as horseradish peroxidase (HRP), alkaline phosphatase (AP) or glucose oxidase is used to oxidize the substrate (usually 3,3',5,5'-tetramethylbenzidine (TMB)) which shows a color change (Maekawa, 1995). Chemiluminescent on the other hand, depends on generating light due to chemical reaction. Enzymes such as AP and HRP are used to oxidize the substrate (usually 5-amino-2,3-dihydro-1,4-phthalazinedione (Luminol)) forming an intermediate (Kricka, 1996). The intermediate then returns to a ground stable state by releasing photons which are detected by a luminescent signal instrument. The advantage of this technique over enzyme-based assay is the linear relationship between luminous intensity and the concentration of the measured substance.

Fluorophore based immunoassay is an easy and highly sensitive technique where typical fluorophore used are Fluorescein Isothiocyanate (FITC), Rhodamine-(Red™-X), Tetramethyl rhodamine (TRITC) and Aminomethylcoumarin Acetate (AMCA), Alexa Fluor, DyLight and Cy. These fluorophores absorb light at certain wavelength range (excitation) and emits light at another wavelength range (emission). One key consideration while choosing a panel of fluorophores for multiplexed experiment is their

spectral interference. The excitation wavelength of fluorophores should not overlap or be close to each other. In radioactive immunoassay, a radioactive label like iodine isotope ^{125}I is attached to tyrosine of the antigen. The concentration of antibody bound to the radioactively labeled antigen is detected using a gamma counter. Although this technique is very sensitive and extremely specific, it is still not used widely as it requires special precautions and licensing as radioactive probes are used.

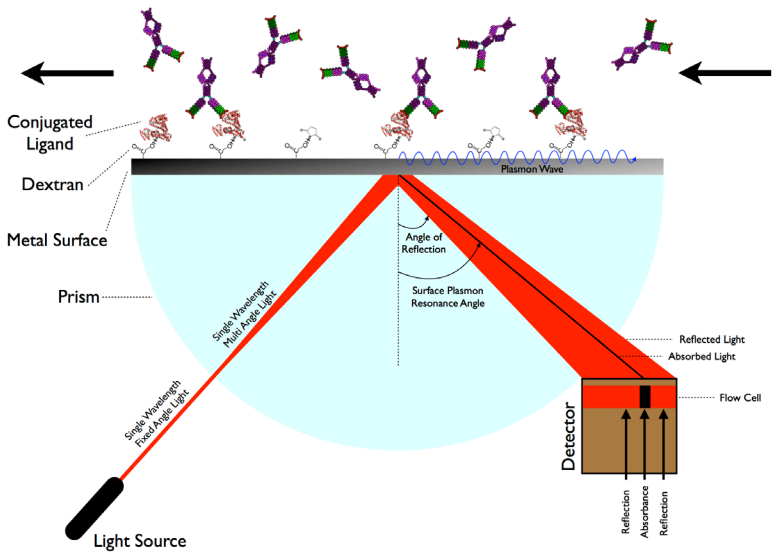


Fig 2.4 Principle of Surface Plasmon Resonance. The binding of antibodies to the conjugated ligand is determined in real-time. Taken from Sari Sabban, 2011 (PhD thesis), The University of Sheffield, CC BY-SA 3.0

The label-free immunoassay is a technique where the antibody is not labeled and a change in physical or chemical property is measured using an apparatus. The two most common label-free techniques are Surface Plasmon Resonance (SPR) and electrochemical based biosensors. SPR is a phenomenon which occurs when a polarized

light hits a surface of metal (typically gold) at the interface of media with different refractive indices (**Fig 2.6**) (Nguyen et al., 2015). At a certain angle of incidence, a portion of the light energy couples with metal surface electrons, which then moves due to excitation (Nguyen et al., 2015). The electron movement is called plasmon, which ceases to form when there is a slight change in the reflective index (due to the binding of biomolecule). A detector measures the changes in reflected light obtained from the surface which can be interpreted as the real-time kinetics of binding and unbinding of the ligand and antibodies.

Electrochemical biosensor on the other hand has an electrode which can show changes in electrochemical signals as the analyte binds to the antibody. The primary antibody is coated on the surface of the electrode and the electro-active residues of the antibody molecule gives a specific electric response. When an antigen binds to the primary antibody there is a change in electric signal which is registered as a change in electrode potential (Vestergaard et al., 2007).

2.5.1 Enzyme Linked Immunosorbent Assay

ELISA is a label based bioanalytical technique which is considered to be the gold standard of immunoassays. The technique is very sensitive and used to detect and quantify antibodies, antigen, proteins, glycoproteins, and hormones (Alhajj & Farhana, 2021). It is extensively used in clinical settings to diagnose HIV testing, pregnancy test and blood typing, among others. They are typically performed in a 96-well polystyrene plate that can bind to proteins with high affinity due to the hydrophobic interactions between the polystyrene and the non-polar protein residues (Nowak et al., 2014). The

basic principle of ELISA involves the immobilization of the antigen directly on the surface of multi-well plate or via the use of capture antibody itself immobilized on the surface of multi-well plate. The antigen then forms an immune complex with the detection antibody coupled with an enzyme. There are four major types of ELISA: Direct ELISA, Indirect ELISA, Sandwich ELISA, Competitive ELISA (**Fig 2.7**).

In Direct ELISA, the antigen is directly immobilized to the surface of the multi-well plate and detected with an antibody specific for the antigen. The antibody is directly coupled with HRP or other similar enzyme. The advantage of direct ELISA lies in the quick and simple steps involved while disadvantage lies in the adverse effect of immunoreactivity due to the interference of label in the primary antibody. In Indirect ELISA, the antigen is immobilized to the surface of the multi-well plate, then a primary antibody binds to the antigen. A labeled secondary antibody against the host species of the primary antibody binds to it for detection.

The advantage of Indirect ELISA lies in the signal amplification of the assay as there are multiple epitopes on each primary antibody where the labeled secondary antibodies can bind. The main disadvantage of this type of ELISA lies in the possibility of cross-reactivity due to secondary antibody. In Sandwich ELISA, two antibodies specific for different epitopes of the antigen is used. One antibody is immobilized on the multi-well plate surface to which the antigen binds and is termed as capture antibody. A detection antibody then binds to another epitope of the antigen. The labeled secondary antibody then binds to the primary antibody. This type of ELISA is highly sensitive and specific for target antigen as two antibodies are used for capture and detection. However, the protocol is lengthy and challenging to optimize for this type of ELISA. In competitive

ELISA, the antigen is either too small or has only one epitope to sandwich between two antibodies. Instead of using a conjugated detection antibody, a conjugated antigen is used for competitive binding with actual antigen in the sample. If there are more actual antigen in the sample, then the conjugated antigen will be less bound to the capture antibody and vice-versa. The signal generated in this type of ELISA is inversely proportional to the amount of protein present in the sample. The advantage of competitive ELISA lies in the ability to quantitate small molecules while the disadvantage lies in the requirement of conjugated antigen. All the four types have some advantages and disadvantages.

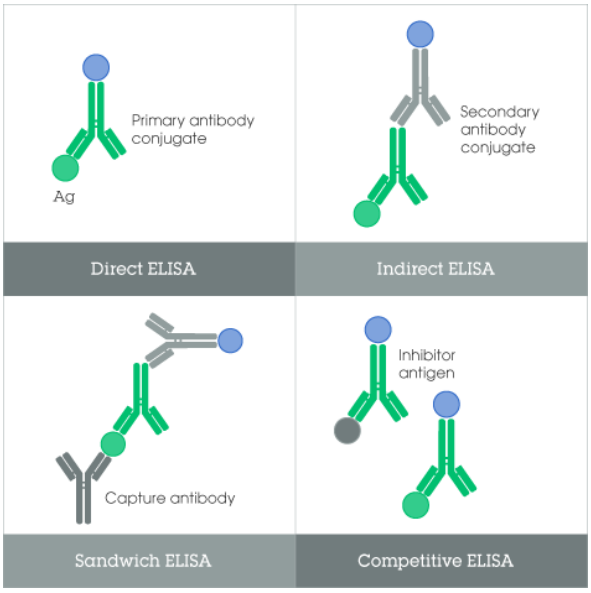


Fig 2.5 Different Types of Enzyme Linked Immunosorbent Assay. They are useful for specific application and have various advantages and disadvantages. Taken from (abcam.com)

Considering that biomarker discovery requires highly sensitive and specific assay, hence sandwich ELISA is the preferred type. RAPID (Rapid Antigenic Protein *in Situ*

Display) ELISA is a modification of sandwich ELISA where the antigen is produced from cDNA using cell free expression system just in time for testing (Rauf et al., 2020). This features aids in processing a large number of antigens simultaneously in ELISA settings, thereby increasing the throughput.

2.5.2 Western Blot and Beads-based Immunoassay

Western blot (or immunoblotting) is a common bioanalytical technique to detect specific proteins in a mixture (Najafov & Hoxhaj, 2017). It allows researchers to detect and quantify specific proteins from a mixture using an antibody probe. In this technique, the mixture of proteins is separated based on molecular weight using gel electrophoresis. Sodium dodecyl sulfate (SDS) unfolds the proteins and binds to all the positive charges on the protein, effectively coating the protein in negative charge (Mahmood & Yang, 2012). Upon applying electric field, the proteins migrate towards positive terminal and separated based on the molecular weight. Then they are transferred from the gel to a nitrocellulose membrane which is known as blotting. The membrane is blocked to prevent non-specific binding of the antibodies to the surface of the membrane. It is then probed with suitable primary and secondary antibodies, after which it is washed to remove unbounded antibodies. As the antibodies only bind to the protein of interest, hence a single band should be visible in final membrane. The secondary antibody is conjugated with an enzyme which when combined with appropriate substrate produces detectable signal. The thickness of the band corresponds to the amount of protein present in the original mixture.

The decision to choose ELISA or Western blot lies in the purpose of the experiment. ELISA tells us if a specific antibody is present in the sera while Western blot tells us if a specific antigen is present in the mixture. Clinically, ELISA provide information about past infection while Western blot provide information related to current state of infection. Also, ELISA can measure folded protein in solution whereas western blot always measures denatured proteins.

Beads assay or beads-based immunoassay is another type of immunoassay where the analyte is immobilized on the surface of the beads. Several color-coded microbeads with pre-coated analyte-specific capture antibodies are suspended with the analytes (van der Wal et al., 2014). The analytes bind to their respective capture antibody on beads. Antibodies from sera or other fluid then binds to these analytes which are detected using a secondary antibody conjugated with Phycoerythrin (Djoba Siawaya et al., 2008). The beads are then passed through flow-based detector with two lasers, one for the identification of the beads and determination of the analyte while other laser determines the magnitude of the phycoerythrin signal. Different analytes can be tested in a single well of 96-well plate by adding the respective beads and the sera. The capability of this assay to multiplex the experiment helps in reducing the amount of sample required and time required.

2.5.3 Protein Microarrays and NAPPA

Protein microarrays are a high throughput miniatured immunoassay on a microscopic slide in which thousands of protein are immobilized in organized rows and columns. This technique became popular in last decade when scientist realized the need

for analyzing more proteins simultaneously for comprehensive result. The capability of microarrays has reached to a point where genome-wide protein-protein interactions can be performed. For comparison, a genome-wide assay on an ELISA can take months to complete while microarrays can complete it within days.

Before microarrays were invented, techniques like phage display and yeast display were used. In phage display, the gene of interest is inserted into the phage coat-protein gene, causing the phage to display the protein of interest on its surface when it expresses the coat-protein. The protein displayed on the surface can then be screened against other proteins, peptides, or DNA sequences for the determination of possible interactions. However, the issue with phage display lies in the expression bias, misfolding of proteins and labor intensive. The phage uses the prokaryote machinery of bacteria to express and fold complex human proteins. Bacteria can have bias to certain amino acids, like preferring proteins rich in methionine and lysine. Also, the prokaryote mechanism is not enough to fold the proteins correctly with post-translational modifications.

To address the shortcomings, scientists came up with other methods of detecting protein-protein interactions which led to invention of protein microarrays. The applications of microarrays have diversified over years ranging from protein-protein interaction, protein-DNA interaction, protein-peptide interaction, phosphorylation, and immunoassay. The first proteome microarray was constructed with >5800 individually purified yeast proteins which constitutes 85% of the yeast proteome (Zhu et al., 2001). These microarrays were then used to study protein-protein interactions. Microarrays can also help to discover DNA binding proteins, which are important for studying pathways involving transcription factors. Several groups have studied transcription factor

homologs, isoforms and complexes using protein-DNA binding microarrays (Andrilenas et al., 2015). Also, they are frequently used in immunoassays to determine antibody targets and biomarkers (Wang et al., 2015b).

There are various advantages of protein microarrays as compared to other immunoassays. First, the possibility of analyzing thousands of proteins simultaneously with high sensitivity. Second, the property of equal chance of display for less abundant and weakly binding proteins as opposed to in-vivo or mass spectrometry where the less abundant are masked. Third, they are highly customizable as the protein targets can be changed based on requirements.

To construct a protein microarray, the proteins of interest are purified and printed on the slides in high density format using a contact sprayer or non-contact microarrayer (Chen & Zhu, 2006). The microscopic slide can be made up of glass, plastic, or silicon. The slides are functionalized with aldehyde and epoxy groups for attachment through amine (Kusnezow et al., 2003; MacBeath & Schreiber, 2000), nickel for attachment through His-tagged proteins (Angenendt et al., 2002), and nitrocellulose membrane for non-covalent attachment of DNA and proteins (Stillman & Tonkinson, 2000). Another different approach is used where the slides are coated with (3-Aminopropyl) triethoxysilane (APTES), an aminosilane group that can bind to the SiO_2 groups on the glass slide covalently (**Fig. 2.8**). The terminal amino group from aminosilane can be covalently attached to a activated carboxy group of protein using EDC/NHS chemistry (Olde Damink et al., 1996). Briefly, 1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) reacts with carboxylic acid group of protein to form an active O-acylisourea intermediate that is prone to nucleophilic attack by primary amino acid

group. The primary amino acid group from the aminosilane can attack the intermediate and form a covalent bond. This reaction is however very sensitive to aqueous medium, so non-aqueous buffer like MES buffer (4-morpholinoethanesulfonic acid) is used for the reaction purpose. The issue with this type of functionalization is the unoriented binding of proteins / antibodies to the surface of the microarray. The proteins / antibodies have epitopes through which they recognize their binding target. Due to unoriented binding on slide surface, some of proteins / antibodies will have their epitopes inaccessible for the immunoassay.

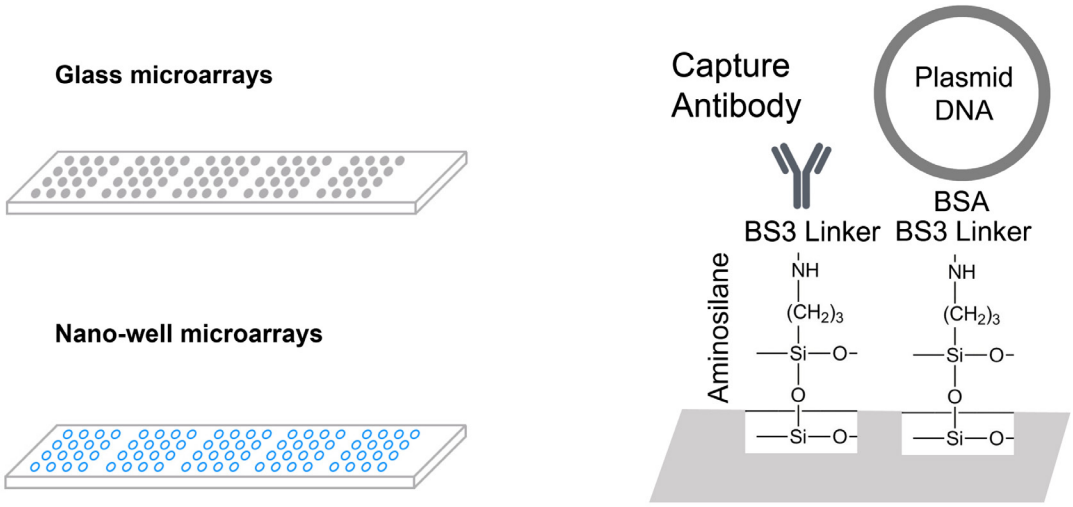


Fig. 2.6 Distinction Between Glass Microarray and Nano-well Microarray. The proteins are printed on the flat surface of glass microarray while for nano-well microarray, it is printed in the wells. The chemistry of binding is shown.

There are some disadvantages of traditional protein microarrays as well. Protein expression and purification is a tedious job which requires huge amount of time and manual labor. The gene of interest is transfected into the bacterial cell, followed by

culturing of the bacterial cells in a suitable nutrients-rich medium for 12-14 hours. The cells are then lysed, and the suspended proteins in the solution are passed through a column for purification. On top of this, proper protein storage is also critical without which the proteins might get denatured.

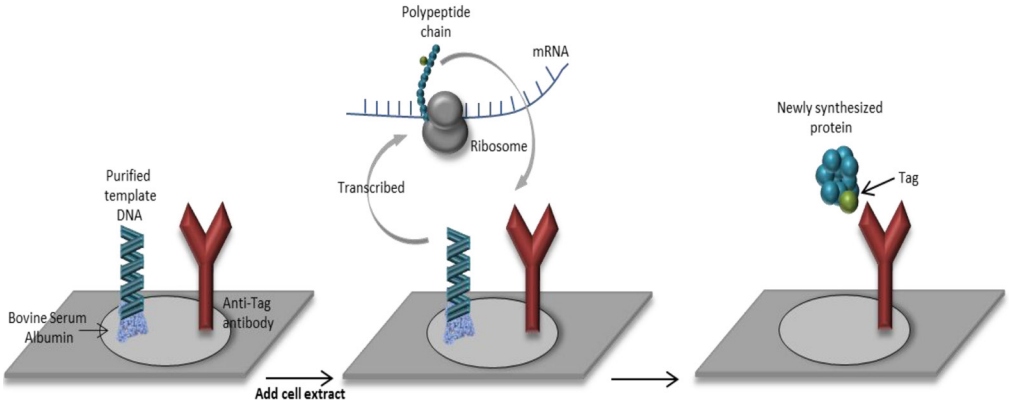


Fig. 2.7 Steps of Nucleic Acid Programmable Protein Array (NAPPA) Fabrication. In each spot of the microarray, plasmid DNA is co-spotted with BSA, BS3 linker and anti-GST antibody. Cell-free expression system is added to express the protein on the spot. The newly synthesized protein is captured by the anti-GST antibody. Image taken from (Díez et al., 2015).

In order to circumvent the above-mentioned issues, Nucleic acid programmable protein array (NAPPA) was designed. Instead of using purified proteins, NAPPA uses plasmid having the gene of interest, T7 promoter, and Glutathione (GST) tag (Ramachandran et al., 2004). The plasmid is co-spotted with capture antibody specific for GST tag. The protein along with the GST tag is expressed from the plasmid right at the time of experiment using a cell-free expression system (Ramachandran et al., 2008). The

expressed protein is captured by the anti-GST antibody in each microarray spot. This approach abrogates the concern of tedious work and time required for protein expression and purification. Also, protein stability during storage is no longer a concern as DNA are much more stable than proteins.

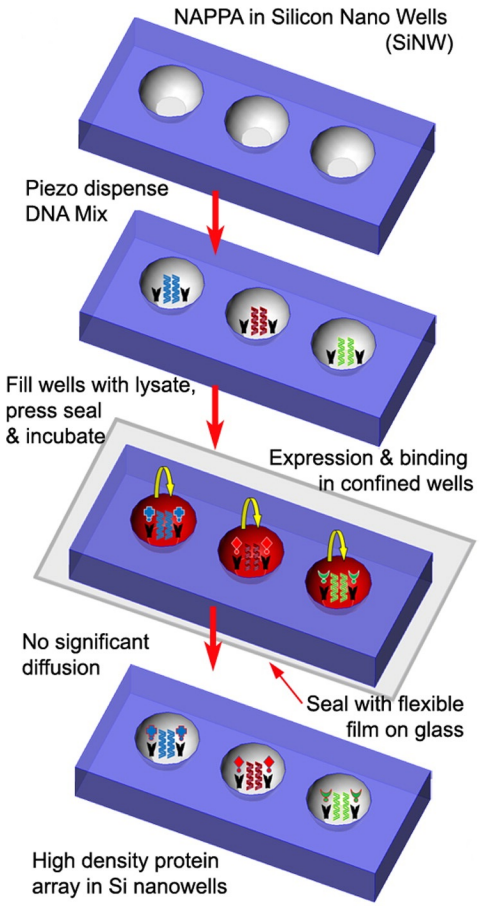


Fig. 2.8 Workflow of High Definition NAPPA (HD-NAPPA). Reactive ion etching using CHF_3–O_2 plasma is used to make the nanowells. Piezo dispenser is used to add the DNA mix into each well followed by NAPPA assay procedure. Image adapted from (Takulapalli et al., 2012).

The traditional and commercially available glass microarrays have mediocre density (< 4,500 target proteins). One of the key challenges with increasing the density of the microarray lies in the problem of diffusion of freshly expressed protein into adjacent spots. This can lead to false positive and thus the microarray spots need to be far enough so that their signals don't interfere. In order to increase the density further, high definition NAPPA (HD-NAPPA) was created which had etched nanowells on a silicon slide (**Fig. 2.10**) (Takulapalli et al., 2012). This modification increased the density of the microarray significantly, making it capable of up to 24,000 target proteins in a single slide (Takulapalli et al., 2012). In addition to increasing density, the nanowells separates the reaction mixture which eliminates the chances of cross-reactivity among adjacent spots.

In terms of actual values, glass microarrays had center-to-center distance between adjacent spots as 750 μm. When the spot distance was reduced to 375 μm, there was significant diffusion of signals (Takulapalli et al., 2012). Silicon microarrays with semispherical nanowells on the other hand had approximately 250 μm diameter and 75 μm deep with 375 μm center-to-center distance among spots. There was no observable diffusion in silicon microarrays.

2.6 Inflammatory Bowel Disease and Its Pathogenesis

Inflammatory Bowel Disease (IBD) is a chronic and life-threatening intestinal disorder characterized by inflammation in the gastroenteric tissues. The pathogenesis of IBD is complex as thousands of different bacterial species resides and interacts in the gastroenterological tract. The common symptoms of IBD includes episodes of abdominal pain, diarrhea, bloody stool, weight loss, inflammation and ulceration (Guan, 2019).

There are two subtypes of the disease namely Crohn's disease (CD) and Ulcerative colitis (UC). In UC, the inflammation is localized in colon with continuous pattern while CD can affect any part of the digestive tract often in discontinuous pattern. CD exhibits histologically thickened submucosa, transmural inflammation, fissuring ulceration, and granulomas while UC exhibits inflammation only in mucosa and submucosa with cryptitis and crypt abscesses (Abraham & Cho, 2009; Gajendran et al., 2018; Khor et al., 2011). IBD can occur at any age, although there are higher incidence found in early adults (Khor et al., 2011). The cause of IBD is attributed to continuous inflammatory response to commensal bacteria in a genetically predisposed individual. The global healthcare burden of IBD is rising with highest age-standardized prevalence in United States (Alatab et al., 2020). The age-standardized prevalence rate increased from 79.5 per 100,000 population in 1990 to 84.3 per 100,000 population in 2017 (Alatab et al., 2020).

IBD is a chronic disease which can be controlled using surgery, antibiotics, anti-inflammatory and immunosuppressive drugs. There are periods of remission when the disease is not active but after that, the inflammation starts again. This type of chronic disease requires a continuous monitoring biomarker which can provide information about the effect of various treatment and how the disease is progressing or regressing.

2.6.1 Role of Genetics in IBD

Genetic predisposition is one of the causes of the disease as various IBD susceptible genes heterogeneity are correlated with occurrence of the disease. Therefore, the identification of IBD susceptible genes might provide key understanding about the pathophysiology of IBD. The technological advancement in DNA sequencing and genetic

testing has allowed large-scale genome-wide association studies (GWAS) in various cohorts of IBD sample. The analysis of the susceptible genes revealed that they play role in several pathways critical for epithelial barrier function, innate mucosal defense, immune regulation, cell migration, autophagy, and metabolic pathways associated with cellular homeostasis (Guan, 2019). The advancement of genetic research using GWAS and next generation sequencing revealed 242 susceptibility loci in IBD (Mirkov et al., 2017).

Nucleotide-binding oligomerization domain containing 2 (NOD2) was the first susceptibility gene for CD discovered in 2001 (Loddo & Romano, 2015). This gene is a member of the cytosolic Nod-like receptor (NLR) family which encodes for an intracellular receptor to sense microbial invaders (Yamamoto & Ma, 2009). NOD2 can recognize the bioactive fragments of peptidoglycans found in the cell wall of gram-negative and gram-positive bacteria, called muramyl dipeptide (Guan, 2019). NOD2 binds to its ligand muramyl dipeptide which then activates NF-κB. The activation of NOD2 with muramyl dipeptide also induces autophagy in dendritic cells (DC). The DCs with NOD2 mutation are deficient in inducing autophagy and also show reduced localization of bacteria in autophagolysosomes (Cooney et al., 2010).

ATG16L1 (autophagy-related 16-like 1) is another gene which is important for autophagy and clearance of intracellular bacteria. The mutation T300A in *ATG16L1* increases the susceptibility of the protein ATG16L1 to caspase-3 cleavage and decreases it function (Murthy et al., 2014). ATG16L1 is also reported to be important for antigen specific T-cell responses (Conway et al., 2013).

2.6.2 Role of Environment in IBD

Studies on IBD epidemiology have revealed the role of environment factors on disease pathogenesis. There are multiple factors that influence the development of IBD like diet, smoking, intake of drugs, geographical and social status, stress, breast feeding and appendectomy.

Food intake plays an important role in the development of IBD. It is reported that intake of more fruits and vegetables is associated with decreased risk of CD as dietary fiber have a protective effect due to their antioxidant properties (Dolan & Chang, 2017). On the other hand, intake of fast food with sugar-rich food may elevate the risk of CD (Hibi & Ogata, 2006). In addition, artificial food additives promote intestinal inflammation by interfering with barrier function in the gut (Dolan & Chang, 2017). Epidemiology study based on Japan have revealed that increased dietary intake of animal protein and long-chain omega-6 polyunsaturated fatty acids may lead to increased risk of CD (Shoda et al., 1996). Wu *et. al* demonstrated that long-term intake of diet rich in fat and low in fiber is associated with high quantities of *Bacteroides* and low quantities of *Prevotella* (Wu Gary et al., 2011).

The association of smoking and CD is well established, and it is found that patients of CD consume more tobacco products than general population (Dam et al., 2013). In addition, heavy smokers have worse clinical outcome compared to light smokers (Seksik et al., 2009). Women smokers are more susceptible to CD than men smokers (Cosnes, 2004).

Prolonged antibiotics consumption can alter the composition of gut microbiota by decreasing taxonomic diversity (Dethlefsen et al., 2008). It is found that previous

exposure to antibiotics increased the risk of CD (odds ratio (OR) = 1.74, 95% CI: 1.35-2.23) but not UC (OR = 1.08, 95% CI: 0.91-1.27) (Abegunde et al., 2016). In another study, it is found that people diagnosed with IBD are more likely to have prescribed antibiotics 2-5 years before their diagnosis (Shaw et al., 2011). It is found that western countries have higher incidence of IBD compared to the rest of the world. One reason is because western diets are high in fat and low in fiber. The heterogeneity of IBD occurrence with geographical location can also be attributed to environmental exposure of certain microorganism which are prevalent in that location (Piovani et al., 2019).

Stress is how the brain and body respond to a certain event or situation. The mechanism that connects the brain and our immune system during a homeostatic response is hypothalamo-pituitary-adrenal (HPA) axis (Abegunde et al., 2016). When a stressful event is encountered, our immune system activates the HPA axis by producing cytokines that results in the production of anti-inflammatory agents (Sternberg et al., 1992). However, if the HPA axis and immune system feedback loop is disrupted, then it can lead to inflammatory diseases when applied to a stressful situation (Abegunde et al., 2016). In animal models, it is observed that acute stress leads to change in mucosal inflammation, intestinal permeability, and bacterial-host relationship (Dam et al., 2013; Mawdsley & Rampton, 2006). In humans, studies have been less convincing but the result of disruption of HPA axis during stressful events is well accepted.

Breast feeding provides early exposure for the development of immune system and gut microbiome (Koloski et al., 2008). Breast milk is a rich source of 1) protective IgA which inhibits the binding of enteric bacteria to the surface of gut (Goldman, 2007), 2) live *Lactobacillus* which promote immune tolerance and epithelial barrier

maintenance, 3) lactoferrin that prevents the multiplication of bacteria and possess anti-inflammatory properties (Abegunde et al., 2016). The gut microbiome is regulated by breast feeding as it enhances the growth of certain bacterial species like Bifidobacterium and Lactobacillus while its anti-bacterial components prevent the growth of pathogenic bacteria (Koloski et al., 2008). A meta-analysis on the effects of breast feeding on IBD showed a significant protective role in CD with an odds ratio of 0.67 (95% CI: 0.52-0.86) (Klement et al., 2004). In another study, breast feeding prevented CD (OR = 0.71, 95% CI: 0.59–0.85), UC (OR = 0.78, 95% CI: 0.67–0.91), and IBD (OR = 0.74, 95% CI: 0.66–0.83) (Piovani et al., 2019).

Appendectomy is surgery to remove appendix when it is infected. The association of appendectomy and CD is conflicting and not conclusive. Some studies show that appendectomy is a risk factor for CD while some other study reports no significant association among appendectomy and CD. Kaplan *et. al* found an increased risk of CD following appendectomy (relative risk = 1.61, 95% CI 1.28-2.02) (Kaplan et al., 2008; Kaplan et al., 2007). The risk of CD largely increased in the first year after appendectomy and was no longer significant after 5 years (Kaplan et al., 2008).

2.6.3 Role of Immune System in IBD

The intestinal immune system plays a very crucial and delicate task of rapid and effective immune response against pathogenic bacteria, while maintaining tolerance towards food and commensal bacteria. This is achieved through the presence of efficient epithelial barrier and complex immune system in the gut. The balance between tolerance to commensal bacteria and effective immune response to pathogenic bacteria is disturbed

and dysregulated immune response contribute to the aberrant intestinal inflammatory response. Abnormalities like epithelial damage, increase in inflammation due to large number of T cells, B cells, macrophages, dendritic cells and neutrophils infiltrating into the mucosa and the failure of the immune system to control the elevated inflammation lead to immune dysregulation (Guan, 2019). The activated mucosa cells produce high level of pro-inflammatory cytokines likes TNF, IFN-γ, IL-1β.

The tight junctions between epithelial cells acts as a selective barrier for the entry of microbial antigens for inducing tolerance to commensal bacteria. This barrier also aids in selective passing of nutrients and fluids from food while preventing the influx of antigens and invasion by pathogens and commensal bacteria. Given its vast surface area which is in constant exposure to thousands of bacterial species, the epithelial barrier is susceptible to damage by pathogens, ischemia, and environmental toxins (Ni et al., 2017a). An impaired epithelial barrier leads to increased penetration of pathogenic bacteria which is often observed in IBD patients. Goblet cells of intestinal epithelium performs two important task of producing mucus and transporting and presenting luminal antigens to tolerogenic dendritic cells (Parikh et al., 2019). Goblet cell WFDC2 gene is downregulated in patients with mucosal inflammation, which is an anti-protease molecule that has the ability to maintain the barrier integrity of intestinal epithelial cells (Parikh et al., 2019). Particularly in UC patients, there has been association between the disease and single nucleotide polymorphism (SNP) at three different loci. *HNF4A*, a transcription factor that regulates the assembly of cell junctions. *CDH1*, encodes E-cadherin, which the main component of adherent junctions. *LAMB1*, is a laminin which is expressed in the basal membrane of intestinal epithelium.

Besides the physical barrier, epithelial cells can secrete various bactericidal agents like defensins (Geremia et al., 2014). Defensins are endogenous anti-microbial peptides that acts against bacteria, fungi, viruses, and protozoa. They are secreted constitutively or due to the recognition of bacterial components by pattern recognition receptors (PRR) like Toll like receptors (TLR) and nucleotide oligomerization domain receptors (NOD) (Choy et al., 2017). A defective expression of anti-microbial peptides is found in CD compared to UC. α-defensins (HD5 and HD6) levels are found to be reduced in ileal CD patients demonstrating a functionally deficient antimicrobial barrier (Wehkamp et al., 2005). Interestingly, the reduction is even more prominent in those with NOD2 frameshift susceptibility variants (Wehkamp et al., 2004). β-defensins (HBD1, HBD2, HBD3, and HBD4) are found deficient in colonic CD patients (Wehkamp et al., 2003). Unlike CD, UC is associated with increased expression of HBD2, HBD3 and other anti-microbial peptides (Klag et al., 2013). The mucosal layer shows increased anti-microbial activity while the inner mucous layer is not sterile and thus shows the inefficiency of anti-microbial barrier (Klag et al., 2013; Swidsinski et al., 2007).

Dendritic cells (DC) are professional antigen presenting cells (APC) specialized in antigen capture, process, and presentation to T cells. Due to TLR, DCs can recognize certain molecular patterns on the surface of bacteria, and it enables them to distinguish very similar microorganisms (Silva et al., 2016). This action of DC determines whether immunogenic or tolerogenic responses are formed. DC and Macrophages can promote inflammation. $CD14^+$ macrophages are found in abundance in CD patients and they produce proinflammatory cytokines like IL-23, TNF, and IL-6 (Kamada et al., 2008).

The dysregulation in innate immune system causes functional abnormalities in adaptive immune system. IBD is strongly T cell mediated, and abnormally activated T cells lead to inflammation through the release of inflammatory cytokines. Antigen presenting cells like dendritic cells, macrophages and B cells provide a peptide from the microbial antigen to naïve T cells. Th-0 cells become activated and differentiates into Th-1 (clearance of intracellular pathogens), Th-2 (parasite and allergic responses), or Th-17 (extracellular bacteria and fungi) cells (Geremia et al., 2014). Th-1 cells are induced by IL-12 and produce interferon-gamma (IFN-γ), interleukin 2 (IL-2), and tumor necrosis factor-beta (TNF-β) (Romagnani, 1994). Abnormal Th-1 response triggered by an increased level of IL-18 and IL-12, is thought to be the cause of intestinal inflammation in CD (Monteleone et al., 1997; Monteleone et al., 1999). Th-2 response produces IL-4, IL-5, IL-6, IL-9, IL-10, and IL-13, of which IL-5 and IL-13 are thought to be critical for UC pathogenesis (Di Sabatino et al., 2012; Romagnani, 1994). It was demonstrated that activated T cells from CD mucosa release more IFN-γ than T cells from UC patients or controls (Breese et al., 1993).

B cells produce and secrete a deregulated amount of antibodies, especially IgG, IgA, and IgM (MacDermott et al., 1981). In CD patients, IgG1, IgG2 and IgG3 levels are high in both serum and in intestinal mucosa, compared to healthy mucosa (Scott et al., 1986). IgA is the second most abundant antibody type and is critical for mucosal immunity. Intestinal B cells differentiate into plasma cells and secrete IgA (Cerutti & Rescigno, 2008). TGF-β triggers the class switching of B cells, which then differentiates into plasma cells. The plasma cells secrete polymeric IgA which is transcytosed through the epithelial membrane into the lumen as a secretory IgA (SIgA). SIgA limits bacterial

access to the epithelial cells of the host and blocks invading bacteria by recognizing pathogenic epitopes (Alexander et al., 2014). SIgA also plays a vital role in maintaining immune tolerance to commensal bacteria. They form complexes with commensal bacteria and subsequently crosses the lumen to the mucosa by binding to specialized IgA receptor (Mantis et al., 2002).

2.6.4 Dysbiosis in IBD

Dysbiosis is a condition when there is a compositional and metabolic change in intestinal microbiota (**Fig. 2.11**). This change reduces the diversity of the gut microbiota owing to a shift in the balance between commensal and potentially pathogenic microorganisms (Ni et al., 2017b). More than 99% of the gut microbiota belongs to four phylum – Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria. Firmicutes and Bacteroidetes dominate the intestinal microbiota in healthy adults. In IBD, the most consistent change is that of Firmicutes whose abundance decreases (Frank et al., 2007; Manichanh et al., 2006a; Walker et al., 2011). Bacteroidetes are also reported to be reduced in abundance, but increase is also reported (Frank et al., 2007; Manichanh et al., 2006a; Walker et al., 2011). *F. prausnitzii* which belongs to *Clostridium* cluster IV along with *Blautia faecis*, *Roseburia inulinivorans*, *Ruminococcus torques*, and *Clostridium lavalense* were decreased in CD patients when compared to healthy individuals (Fujimoto et al., 2013; Takahashi et al., 2016).

The commensal bacteria can produce short-chain fatty acids (SCFA) by fermenting resistant starch or indigestible carbohydrates (Nishida et al., 2018). SCFA are mainly anions in the colon as acetate, propionate, and butyrate. They help in the

expansion of Treg cells which in turn release anti-inflammatory cytokines which is important to keep inflammation in check. For instance, butyrate is one of the SCFA which has anti-inflammatory effects, and it is found in lower abundance in IBD patients. When the butyrate-producing bacteria like *F. prausnitzii* and *Clostridium* cluster IV, XIVa, XVIII (Takahashi et al., 2016) are reduced in abundance then SCFAs also decreases in amount, which is not enough to activate enough Treg cells. This leads to the expansion of Teff cells which then release pro-inflammatory cytokines.

In contrast, the relative abundance of Proteobacteria, mainly adhesion-invasive *E. coli* (AIEC) increases in CD patients compared to healthy individuals (Darfeuille-Michaud et al., 2004). These pathogenic bacteria have the capability to adhere to intestinal epithelium which increases the permeability of the membrane leading to inflammation. These disbalance in relative abundance of microbiota has a detrimental impact on the intestinal homeostasis.

At species level, it is found that some of them exert a protective role while some other exert inflammatory effects. For examples, *Bacteroides* and *E. coli* species are increased in abundance in IBD while *Bifidobacteria* species are decreased in abundance (Tamboli et al., 2004). There are evidences from experimental models which suggest that gut bacteria often drive immune activation, while chronic inflammation to some bacterial species in turn shapes the gut microbiota and contributes to dysbiosis (Ni et al., 2017a).

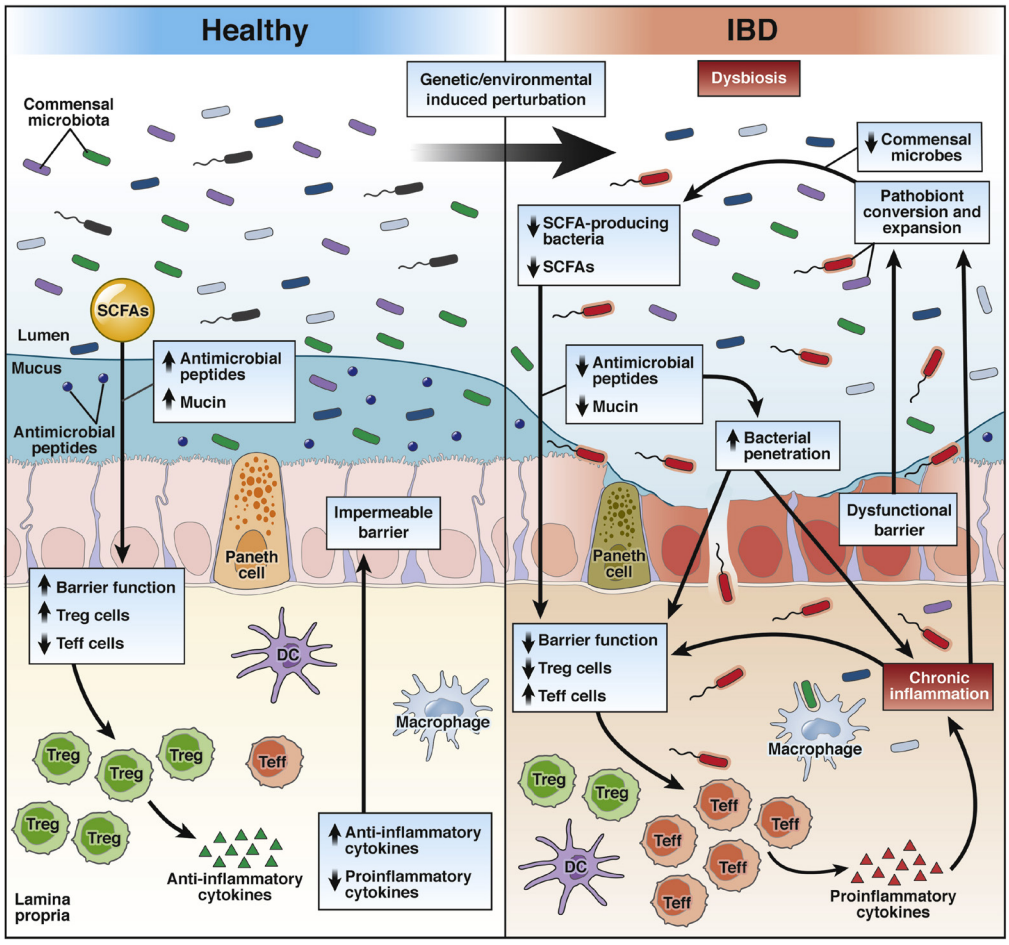


Fig. 2.9 Gut Dysbiosis Observed in IBD Patients. The left panel shows the healthy gut where the commensal bacteria are in correct proportion. SCFA helps in expansion of Treg cells which releases anti-inflammatory cytokines. The right panel shows the IBD gut where the balance between commensal and pathogenic bacteria is broken. Due to less relative abundance of commensal bacteria SCFA reduces which increases the release of proinflammatory cytokines. Image taken from (Lee & Chang, 2021).

2.6.5 Montreal Classification of IBD

As CD is highly heterogenous from a clinical point of view, hence a disease classification for CD was proposed by Montreal Working Party in 2005 (Torres et al.,

2010). The advantage of having a proper classification of the disease is related to accurate patient counselling, assessing disease prognosis, and choosing the most appropriate therapy for each disease subtype (Satsangi et al., 2006). Montreal classification categorizes age of onset of the CD into three categories (A1: below 16 years old, A2: between 17 years to 40 years old, A3: above 40 years old). Based on location of the disease, it categorizes CD into four categories (L1: ileal, L2: colonic, L3: ileocolonic, L4: upper gastrointestinal tract) while the behavior of the disease is categorized into four categories (B1: non-stricturing, non-penetrating, B2: stricturing, B3: penetrating, B3p: perianally penetrating).

For UC, the Working Party felt the relevance of disease extent and disease severity classification (Satsangi et al., 2006). The disease extent for UC was categorized into three categories (E1: Ulcerative proctitis, E2: left sided UC, E3: extensive UC) while the disease severity for UC was categorized into four categories (S0: clinical remission, S1: mild UC, S2: moderate UC, S3: severe UC).

IBD is a dynamic disease which changes its phenotype over time. For instance, it is found that CD patients with predominately inflammatory disease at diagnosis are likely to develop either fistulizing or stricturing complications within 5, 10, and 20 years (Louis et al., 2001). Similarly, extent-based classification of UC is highly unstable as the phenotype of disease changes with time. For instance, the proximal extension of proctitis over 10 years is estimated to be between 41% to 54% (Silverberg et al., 2005). The contrary observation is also valid, where the disease regress over time, with regression rate estimated from a crude rate of 1.6% to an actual rate of 71% after 10 years

(Silverberg et al., 2005). This observation clarifies the need to reevaluate the classifications frequently as the disease keeps on changing.

2.7 Autoantibodies and Autoimmune Disease

Autoantibodies are the immunoglobulins that reacts with individual's own molecules. They are produced by the immune system when it fails to distinguish between self and non-self molecules (Elkon & Casali, 2008a). Usually, the immune system is able to discriminate between individual's own antigens and foreign antigens (bacteria, virus). Thus, it produces antibodies only when it senses the presence of foreign antigens in our body. However, when one of the person's own protein is ceased to be recognized as self-antigen, then antibodies are produced against that self-antigen leading to inflammation and damage to the tissues and organs.

Normally, self-reactive lymphocytes are removed during self-regulatory process of clonal deletion (**Fig. 2.4**) (Janeway, 2001). During development, different lymphocytes originate from a single progenitor cell with diverse antigen receptor specificity. Some of these lymphocytes will be reactive to individual's self-proteins. Those lymphocytes which receives too much or too little signal from their antigen receptor during development are eliminated by the process of apoptosis (Janeway, 2001; Pelanda & Torres, 2012). This process of elimination of self-reactive lymphocytes makes our immune system immunologically unresponsive, or tolerant to self-proteins.

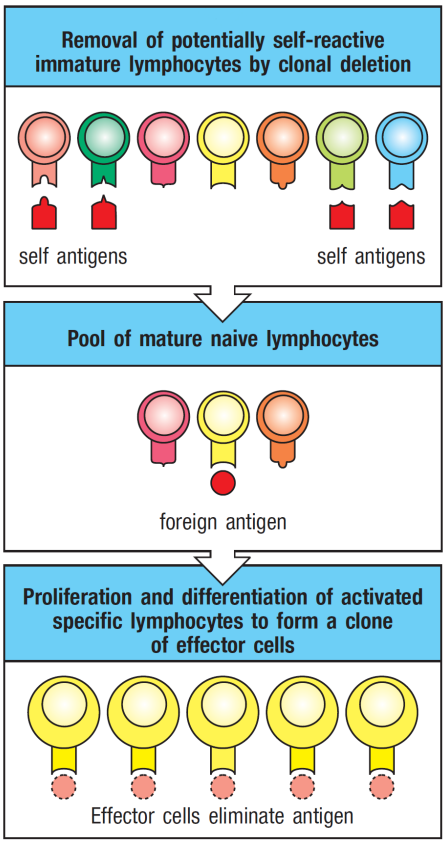


Fig. 2.10 Clonal Deletion of Self-reactive Lymphocytes. Lymphocytes with receptors that bind to self-proteins are removed before maturation. The mature lymphocytes passed through this regulatory process should have tolerance to self-proteins. Image adapted from (Murphy & Weaver, 2016).

Due to genetic predisposition and environmental effect, some self-reactive lymphocytes pass through the clonal deletion process without getting eliminated (Xiao et al., 2021). These lymphocytes then secrete antibodies which can target tissues and organs. The secreted autoantibodies can cause one of the classifications of autoimmune disease: organ-specific or systemic. Organ-specific autoimmune diseases are those in

which a particular tissue or organ is targeted by the immune system. For instance, beta cells of the endocrine pancreas are targeted in type 1 diabetes and skin cells in vitiligo (Roep & Peakman, 2012). On the other hand, systemic autoimmune diseases are those where the autoantigen is found on cells of multiple tissues and organs. For instance, the ribonucleoproteins in the cells are targeted in systemic lupus erythematosus (SLE) (Riemekasten & Hahn, 2005) and collagen type II in rheumatoid arthritis (RA) (Panayi & Corrigall, 2002).

Several autoantibodies are well-known biomarkers of autoimmune diseases. In type 1 diabetes, islet cells antibodies (ICA), glutamic acid decarboxylase antibodies (GADA), and insulin autoantibodies (IAA) are common biomarkers. In rheumatoid arthritis, anti-citrullinated protein antibodies (ACPA) are common biomarkers.

2.8 Natural Autoantibodies and Their Role in Cellular Homeostasis

Antibodies that react with self-molecules in healthy individuals are referred to as natural autoantibodies (Elkon & Casali, 2008b). It was earlier believed that autoantibodies are found only in patients with autoimmune disease. However, it is now known that they are also found in most healthy individuals and in germ-free mice (DeMarshall et al., 2015). In literature, most studies focused on natural autoantibodies of IgM isotype, however, in our own study and in another study the focus was on IgG isotype in healthy individuals (Nagele et al., 2013).

It is believed that adaptive antibodies are very specific for the antigen while natural autoantibodies are less specific and can bind to a range of autoantigens and infectious agents. Because of this property, they act as a first line of defense against

infection and plays a role in cellular homeostasis to clear off cellular debris (Reyneveld et al., 2020a). They are produced by a subpopulation of B lymphocytes, mainly B1 lymphocytes and B lymphocytes of marginal zone (Palma et al., 2018a). Natural autoantibodies have germline encoded VH and VL regions that restricts their binding to phylogenetically conserved epitopes (Binder, 2012). Most of the natural autoantibodies bind to intracellular and membrane expressed autoantigens, circulating macromolecules and haptens which are conserved during evolution (Reyneveld et al., 2020b). As they bind to intracellular autoantigens which are released during apoptosis, that can mitigate the development of autoimmune diseases (Grönwall & Silverman, 2014; Lobo, 2016).

The mechanism underlying the secretion of natural autoantibodies is not fully understood and is complicated. Fetal and neonatal natural autoantibodies are of IgM isotype which are produced steady state without the need of foreign antigen exposure (Baumgarth et al., 2015). However, natural autoantibodies of IgG isotype requires antigenic stimulation to undergo hypermutation and isotype switching from IgM, which doesn't fit into with the classical definition of natural antibodies (Reyneveld et al., 2020b). Hence, only IgM isotype can be referred as natural autoantibodies while IgG isotype found in healthy individuals can be referred as common autoantibodies. Natural autoantibodies also play a protective role of removing the cellular debris after apoptosis, so that autoimmunity is not initiated against the cellular debris (Reyneveld et al., 2020a).

2.9 Molecular Mimicry and Intrinsic Properties of the Autoantigens

The common autoantibodies are elicited either due to the cross-reactivity with microbial protein or due to the intrinsic property of the autoantigen itself. When a

microbial protein shares sequence or structural similarities with a self-protein, it can activate autoreactive T or B cells (Rojas et al., 2018). These activated B cells can secrete antibodies which can cross-react with human proteins. Essentially, the common autoantibodies observed in healthy individuals may be cross-reactive antibodies to microbial proteins. For instance, the protein 2C of coxsackie virus B4 and the autoantigen glutamic acid decarboxylase (GAD65) have sequence homology which is highly conserved. This homology might be responsible for the pathogenesis of insulin-dependent diabetes mellitus (Vreugdenhil et al., 1998). Another work by Fujinami, Oldstone, and colleagues found that mouse antibodies to measles and herpes simplex virus were reactive to both intermediate filaments of normal cells and viral antigen. Further work by the same group using computer analysis found six consecutive amino acids match between myelin basic protein (MBP) and hepatitis B virus polymerase (HBVP). The rabbits when sensitized with either MBP or HBVP peptides, had antibodies in the sera which reacted with MB (Cusick et al., 2012).

Apart from molecular mimicry, intrinsic properties of autoantigens can also be responsible for the elicitation of common autoantibodies. Several properties like foreignness, chemical nature, physical form etc. can influence immunogenicity. Antigen which is evolutionary and phylogenetically far from the host is more likely to be immunogenic. More complex the substance is chemically, the more immunogenic it will be. For instance, proteins are more immunogenic than polysaccharide and nucleic acids. It is presumed that presence of aromatic residue is essential for rigidity and antigenicity of a substance. Also, the molecular weight of an antigen plays a role in its immunogenicity. Most active immunogens are found to have a molecular mass of 14,000

to 600,000 Da. Other properties like hydrophilicity and flexibility also plays a role (Berzofsky, 1985).

2.10 Biostatistics and Bioinformatics Techniques

Biostatistics and bioinformatics are the science of designing, conducting, analyzing, and interpreting studies aimed at improving diagnosis, health, and medicine. Biostatistics is the branch of statistics that determines factors which impact health using programming and mathematical problem-solving. Bioinformatics on the other hand involves developing algorithms for genetic sequences, patterns, enrichment, and pathway analysis.

2.10.1 Receiver Operating Characteristics

Receiver-operating characteristics (ROC) analysis was originally developed during World War II to evaluate the accuracy in differentiating signal from noise in radar detection (Lusted, 1971). This methodology has been adapted for clinical purpose, in particular radiology, diagnostics test, and epidemiology (Zou et al., 2007). In clinical purpose, usually the subjects are classified into one of two categories, diseased or non-diseased. A threshold (sometimes called a cutoff) value is selected to classify subjects as diseased or non-diseased. Graphically, ROC curve shows the tradeoff between the sensitivity and 1- specificity as one changes the threshold (Mandrekar, 2010). Sensitivity (also known as true positive rate) and specificity (also known as true negative rate) are fundamental measures of diagnostic accuracy (Zou et al., 2007).

After a diagnostic test, there are four possible outcomes, true positive, true negative, false positive, and false negative. True positive is defined as when the diagnostic test is able to predict correctly that the subject is diseased. False positive is defined as when the diagnostic test gives positive result, but the subject actually is not diseased. False negative is defined as when the diagnostic test gives negative result even though the subject is diseased. True negative is defined as when the diagnostic test accurately predicts that the subject is non-diseased.

Table 2.1 Possible Test Outcomes of a Diagnostic Test.

Disease status / Test result	Diseased	Non-diseased
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

$$\text{Sensitivity (True positive rate)} = \frac{\text{TP}}{\text{TP + FN}},$$

$$1 - \text{Sensitivity (False negative rate)} = \frac{\text{FN}}{\text{TP + FN}},$$

$$\text{Specificity (True negative rate)} = \frac{\text{TP}}{\text{TP + FN}},$$

$$1 - \text{Specificity (False positive rate)} = \frac{\text{FP}}{\text{TN + FP}}$$

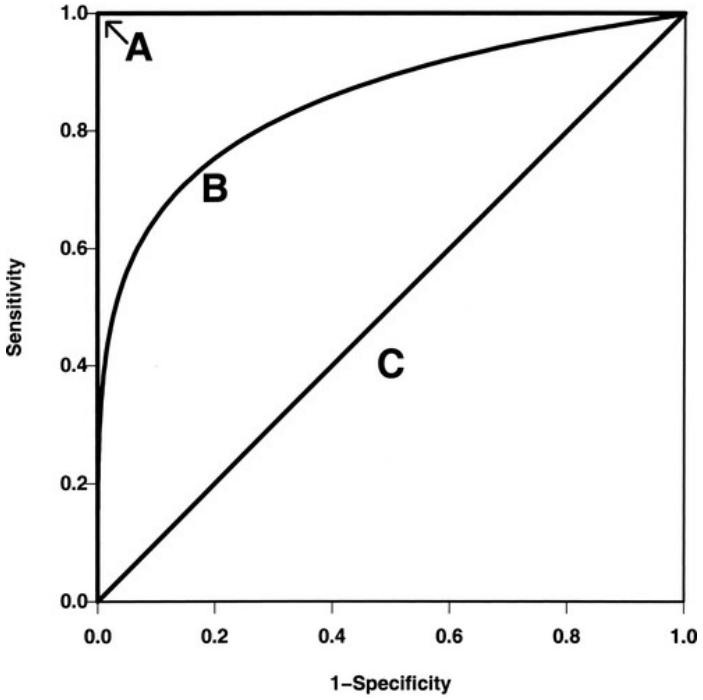


Fig. 2.11 Receiver Operating Characteristics Curves. Three hypothetical curves showing the area under the curve (AUC) are shown. A typical ROC curve is shown in curve B (AUC = 0.80), and the diagonal curve C is corresponding to random chance (AUC = 0.5), while the curve A represents a perfect discrimination (AUC = 1). Image taken from (Zou et al., 2007).

Area under the ROC curve (AUC) is an effective way to summarize the overall diagnostic accuarcy pf the test. The value of AUC ranges from 0 to 1, where 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test (Mandrekar, 2010). A value of 0.5 suggests no discrimination (inability of the test to distinguosh between diseased and non-diseased), 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding (Hosmer Jr et al., 2013). When a diagnostic test of biomarker has an AUC of 0.83, it suggests that there is

83% chance that the biomarker is high only in cases and not in controls. A parametric estimation method would yield smoothed curve while a non-parametric estimation method would yield a jagged curve (Zou et al., 2007).

2.10.2 Basic Local Alignment Search Tool

Basic Local Alignment Search Tool (BLAST) is a sequence similarity search tool developed by National Center for Biotechnology Information (NCBI). BLAST enables one to decipher the biological significance of the parts of sequence by providing statistical information like “expect” value, bitscore, query covered and percentage identity of the search. There are four types of searches available for BLAST: Nucleotide, Protein, Translated, and Genomes. All the sequences must be provided as FASTA sequence or GenBank accession number.

A BLAST alignment consists of a pair of sequences, in which every amino acid / nucleotide is paired to another amino acid / nucleotide or a gap in the other. The program tries to find short matches between two sequences and attempts to start alignments from these “hot spots” (McGinnis & Madden, 2004). For each pair of identical amino acid / nucleotide, BLAST uses a reward while a non-identical amino acid / nucleotide leads to a penalty. The summation of rewards and penalty is computed throughout the sequence to determine the statistical parameters of the search. For nucleotide alignments, identical nucleotides lead to a +2 reward while a nonidentical nucleotide lead to a -3 penalty.

Score	Expect	Method	Identities	Positives	Gaps
55.5 bits(132)	1e-18	Compositional matrix adjust.	25/32(78%)	27/32(84%)	0/32(0%)

```
Query  1   GPDGKRARTAYTRYQTLELEKEFHFNRYLTRR  32
           GP G+R R  YT YQTLELEKEFH+NRYLTRR
Sbjct  8   GPSGRRGRQTYTEYQTLELEKEFHYNRYLTRR  39
```

Fig. 2.12 Alignment View of Two Proteins Using BLAST. The query protein is aligned against the subject protein. The position of the query protein aligned are from 1 to 32 while the position of the subject protein is from 8 to 39.

2.11 Early Detection of Canine Diabetes Markers

Diabetes is a chronic disease that can affect dogs, as well as humans. Insulin-deficient diabetes occurs when the immune cells attack the pancreatic beta cells, which are responsible for synthesizing and secreting insulin. Due to the loss of pancreatic cells, there is a decrease in circulating insulin in the blood. Insulin is a hormone which is responsible for promoting the absorption of glucose from blood to the organs. When there isn't enough insulin, the organs become starved due to the lack of the primary source of energy, i.e., glucose. There is another type of diabetes which is insulin-resistant, where the pancreas produces some insulin, but the dog's body isn't responding to insulin signal to absorb glucose from bloodstream. The insulin-deficient diabetes is more common while insulin-resistant diabetes especially occurs in older, obese dogs (Catchpole et al., 2008).

It is logical to check the immune response to the genes which are upregulated in human diabetes and see if the same set of genes show strong antibody response. The

discovery of biomarkers specific for canine diabetes will help veterinary doctors to take appropriate medication and control the disease way in advance.

CHAPTER 3

3. SEROLOGICAL PROFILING OF CROHN'S DISEASE AND ULCERATIVE COLITIS PATIENTS REVEALS ANTI-MICROBIAL ANTIBODY SIGNATURES

3.1 Introduction

IBD represents a group of intestinal disorders that causes chronic inflammation in the digestive tract. The two main clinical phenotypes are UC and CD. The public health burden of IBD is rising globally (Alatab et al., 2020). The key to reduce the burden is through early and accurate diagnosis. Physicians often use a combination of invasive techniques like histopathology and endoscopy to determine the state of the disease. There is a need for serological biomarkers that can reveal the disease state non-invasively.

The cause of IBD is a combination of genetic predisposition, faulty immune response, and environmental factors (de Souza & Fiocchi, 2016). It is estimated that trillions of microbes of more than 1000 different species, including bacteria, viruses, and fungi, reside in human intestine, which are tenfold the total number of human cells (Ley et al., 2006; Qin et al., 2010). The interaction of microbes with gut mucosa in a genetically susceptible individual and the corresponding immune response plays a pivotal role in the initiation and progression of IBD (Fakhoury et al., 2014). After birth, a relatively less diverse microbial community develops into a complex community due to the influence of diet and environmental factors (Zuo & Ng, 2018). During the second or third decade of life, a dysbiosis is observed in IBD patients which leads to a disbalance between commensal and potentially pathogenic microorganisms (Ni et al., 2017a). The healthy gut microbiota predominately comprises of Firmicutes and Bacteroidetes, and a

lesser extent of Actinobacteria and Proteobacteria (Frank et al., 2007; Sartor, 2008). In IBD, dysbiosis is observed with lower abundance of Firmicutes, which are responsible for the production of short-chain fatty acids critical for the maintenance of barrier integrity (Parada Venegas et al., 2019; Shin et al., 2015). *Faecalibacterium prausnitzii* is known to produce butyrate, which is the primary energy source of colonic epithelial cells and has anti-inflammatory effects, is depleted in IBD patients (Nishida et al., 2018). On the other hand, adherent-invasive *Escherichia coli* have a heightened ability to adhere to gastrointestinal epithelial cells and are found increased in abundance in IBD patients (Lee & Chang, 2021). Besides compositional changes, genetic alterations also contribute to gut dysbiosis that leads to disease initiation and progression. For example, NOD2 mutation is found in 20% – 40% of European and American CD patients (Ahmad et al., 2002; Hugot et al., 2001). NOD2 encodes an intracellular receptor for the bacterial peptidoglycan muramyl dipeptide, which helps in maintaining the balance of commensal bacterial flora (Ramanan et al., 2014).

Immune response to microbes results in the production of antibodies to microbial antigens (Elkadri et al., 2013). Anti-*Saccharomyces cerevisiae* antibodies (ASCA) are associated with CD patients, with sensitivities and specificities ranging between 55% to 65% and 80% to 95%, respectively (Vermeire et al., 2001). Perinuclear antineutrophil cytoplasmic antibodies (pANCA) are associated with UC patients, with sensitivities and specificities ranging between 50% to 71% and 75% to 98%, respectively (Kuna, 2013). Outer membrane protein of *Escherichia coli* (OmpC) and flagellin (CBir1) antibodies are prevalent in CD patients, with prevalence ranging between 24% – 55% and 50% – 56% (Kuna, 2013). The number and response magnitude of anti-microbial antibodies have

previously been shown to indicate the presence of IBD, its severity and its clinical course; however, clinical utility of available antibodies in diagnosis and clinical management of IBD patients has been limited. The techniques used to discover the known anti-microbial antibodies associated with IBD are of low throughput, outdated, and have only been tested on small number of candidate microorganisms or microbial antigens (Lodes et al., 2004). We have performed a large-scale comparative profiling of anti-microbial antibodies in CD and UC patients and healthy controls using an innovative protein microarray technology, namely Nucleic Acid Programmable Protein Array (NAPPA). A total of 1,570 microbial proteins from 50 bacteria and 33 viruses were freshly displayed on the microarrays and probed against 100 CD, 100 UC and 100 healthy control samples. The data obtained from the microarrays provided a unique opportunity to identify diagnostic markers and to understand the association of microbial infection with IBD development.

3.2 Materials and Methods

3.2.1 Patients and Samples

All the serum samples were acquired from Serum Biobank at Mayo clinic with approval from institutional review board. The samples (100 CD, 100 UC and 100 controls) were divided evenly into two non-overlapping discovery and validation sets randomly (**Table 3.1**).

3.2.2 Microbial Protein Array Fabrication

We analyzed 1,570 microbial proteins, of which 1173 proteins were from 50 different species of bacteria, 397 proteins were from 33 different species of viruses and

the remaining proteins were autoantigens. These proteins were selected from our large collection of microbial antigens (DNASU.org) with reference to our anti-microbial antibody studies on other diseases (unpublished data). The NAPPA microarrays were fabricated as described earlier(Song et al., 2017; Takulapalli et al., 2012). Briefly, plasmids with genes of interest cloned in the pANT7_cGST expression vector were obtained from the DNASU plasmid repository, prepared, and printed into silicon nanowells using a piezoelectric dispensing system to produce microbial protein arrays. On the day of experiment, proteins were freshly expressed from printed plasmids using an *in-vitro* transcription and translation (IVTT) protein expression kit (Fisher Scientific) and captured by anti-GST antibody co-printed in each nanowell. After expression, the microarrays were incubated with 1:100 diluted serum samples. IgG and IgA anti-microbial protein antibodies were detected by Alexa-647 goat anti-human IgG (H+L) and Cy3 goat anti-human IgA (Jackson ImmunoResearch). After washing and drying, the microarrays were scanned in a Tecan PowerScanner and the raw fluorescence intensity data were extracted using the ArrayPro Analyzer Software. Raw fluorescence intensities of all the proteins on each microarray were divided by the median intensity for normalization. The median normalized intensity (MNI) was used for all the analysis. Seropositivity of antibody for a particular antigen was defined as MNI ≥ 2 as we have done for our other studies (Song, Song, Camargo, et al., 2021; Song, Song, Rabkin, et al., 2021).

3.2.3 Statistical Analysis

Pairwise comparisons of numbers of IgG or IgA antibodies for each bacterial species among the 3 subject groups were performed using Chi-squared tests to assess statistical significance (**Fig. 3.2**). For each pairwise comparison, the Chi-squared p-values were adjusted using the FDR (false discovery rate) method to reduce the likelihood of false positives. In addition to the multiple comparison adjustment at the antibody level, we performed adjustment at the species level.

For univariate analysis between two comparison groups, we calculated sensitivity for one group at the 96th percentile of the other group or the MNI of 2, whichever was larger. Antibodies with ≥ 14% sensitivity in the discovery set were selected as candidates for further validation. If an antibody had ≥ 14% sensitivity at 96% specificity in both discovery and validation sets, then it was considered as a "validated marker". Venn diagram for the overlap of microbial antigen targets were plotted using Venny (Oliveros, 2007).

We used a three-stage approach to build our classification panels. In the first stage, we selected all candidate biomarkers that passed the criteria above, *i.e.*, sensitivity was greater or equal than 14% at 96% specificity. Next, we applied the minimum redundancy maximum relevance algorithm to further select biomarkers that were possibly the most important and least correlated (Ding & Peng, 2005). In the third stage, we fit a logistic regression model using the selected biomarkers from the first two stages and generated its ROC curve and AUC value to evaluate the model's discriminatory performance between CD, UC, and healthy controls.

Pair-wise subgroup comparisons based on the Montreal classification were performed for the odds ratio (OR) of each antibody using the seropositivity threshold

defined as the maximum of MNI 2 and the 75th percentile of all samples. Chi-squared tests were used to test global significance between all groups with a slight modification by adding 0.5 to each cell of the table to avoid zero cell counts. P-values from the chi-squared method were adjusted for each pair of comparisons and for all candidate biomarkers.

Spearman's rank correlation analysis was performed to assess the correlation between autoantibody and anti-microbial antibody reactivity for CD patients and healthy controls. The R "pheatmap" package was used to generate the heatmap for correlation coefficients.

3.2.4 Bioinformatics Analysis

The NCBI Taxonomy browser was used to find the taxonomical details of all the bacteria and viruses used in our study. The taxa were downloaded as phylip tree file and was used as an input in interactive tree of life (iTOL) software. Two phylogenetic trees were created for bacteria and viruses with different colors distinguishing the phylum. For sequence homology analysis, a pair-wise BLAST analysis was carried out on the antigen protein sequences of validated antibodies for CD vs. control analysis on the Agave high performance cluster at Arizona State University. The E-value were used to generate a heatmap using Python Seaborn package.

To determine the presence of other flagellins on the microarray, BLAST was employed with N-terminal immunodominant region of A4-Fla2 as query and all proteins on the microarray as subject. The significant E-value of 1E-05 was used as a cutoff to

identify other proteins on the microarray that share homology to the region. Multiple sequence alignment was performed with ClustalX.

3.3 Results

3.3.1 Anti-microbial Antibody Profiling in IBD on Microbial Protein Arrays

We studied IgG and IgA anti-microbial antibody profiles of 100 CD and 100 UC patients and 100 age-gender matched healthy controls (**Table 1**) against 1,570 microbial antigens including 1,173 antigens from 50 different bacteria and 397 antigens from 33 different viruses using our protein microarray platform (**Fig. 3.1**). This study provided us a representative overview of anti-microbial antibody response in IBD (**Fig. 3.2**). Total number of IgG antibodies against bacterial proteins from *Bacteroidetes vulgatus* and *Citrobacter koseri* were significantly higher in CD patients compared with those in healthy controls (Chi-square test, $P < 0.01$) (**Fig. 3.2**). On the contrary, total number of IgG antibodies against proteins from several bacteria, such as *Streptococcus pneumoniae, Haemophilus influenzae*, *Staphylococcus aureus, Helicobacter pylori* and *Parvimonas micra* were significantly lower in CD and UC patients compared with those in healthy controls (Chi-square test, $P < 0.05$) (**Fig. 3.2**) Overall, fewer IgA anti-microbial antibodies were found than IgG antibodies. Total number of IgA antibodies against *S. pneumoniae*, *H. influenzae*, *S. aureus*, and *H. pylori* were significantly lower in UC patients compared with those in healthy controls (Chi-square test, $P < 0.01$). On the other hand, anti-viral IgG and IgA antibodies showed heterogenous reactivity with no clear trend of differences among CD, UC, healthy controls (data not shown). Therefore, we focused our analysis on anti-bacterial antibodies.

Table 3.1 Clinical Information of the Samples.

	Discovery set			Validation set		
	CD	UC	HC	CD	UC	HC
N	50	50	50	50	50	50
Gender [Female, Male]	29, 21	29, 21	29, 21	28, 22	28, 22	28, 22
Age (median ± SD)	41 ± 17.66	44 ± 17.25	42 ± 18.47	39.5 ± 17.49	44.5 ± 17.23	39.5 ± 16.02
Disease Behavior [B1/B2/B3]	9/10/6			16/8/2		
Disease Location [L1/L2/L3/L4]	12/6/7/0			12/7/7/0		
Disease Extent [E1/E2/E3]		0/32/18			0/34/16	
Surgery [Yes, No]	24, 25	8, 42		22, 27	7, 42	

CD: Crohn's disease; UC: Ulcerative colitis; HC: Healthy control

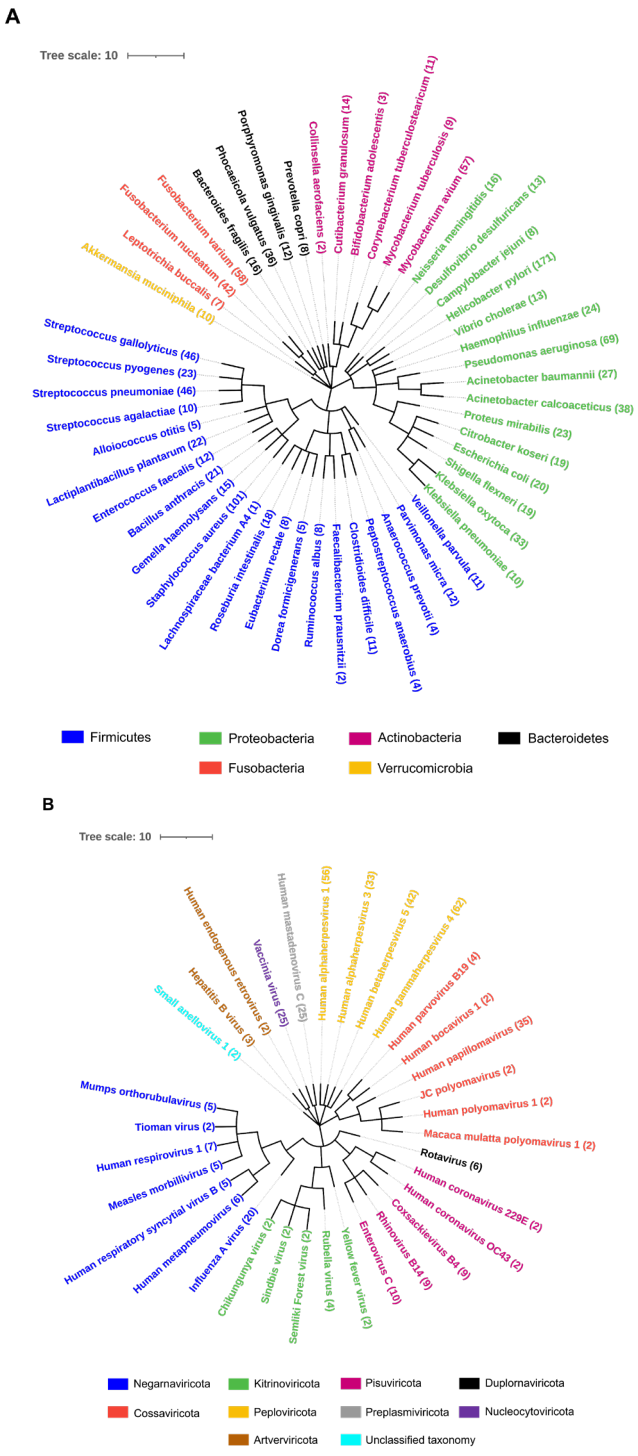


Fig. 3.1. Phylogenetic Tree of Microbes Studied. A. Fifty (50) species of bacteria with 1173 proteins were segregated into 6 phyla. **B.** Thirty-three (33) species of viruses with 397 proteins were segregated into 10 phyla.

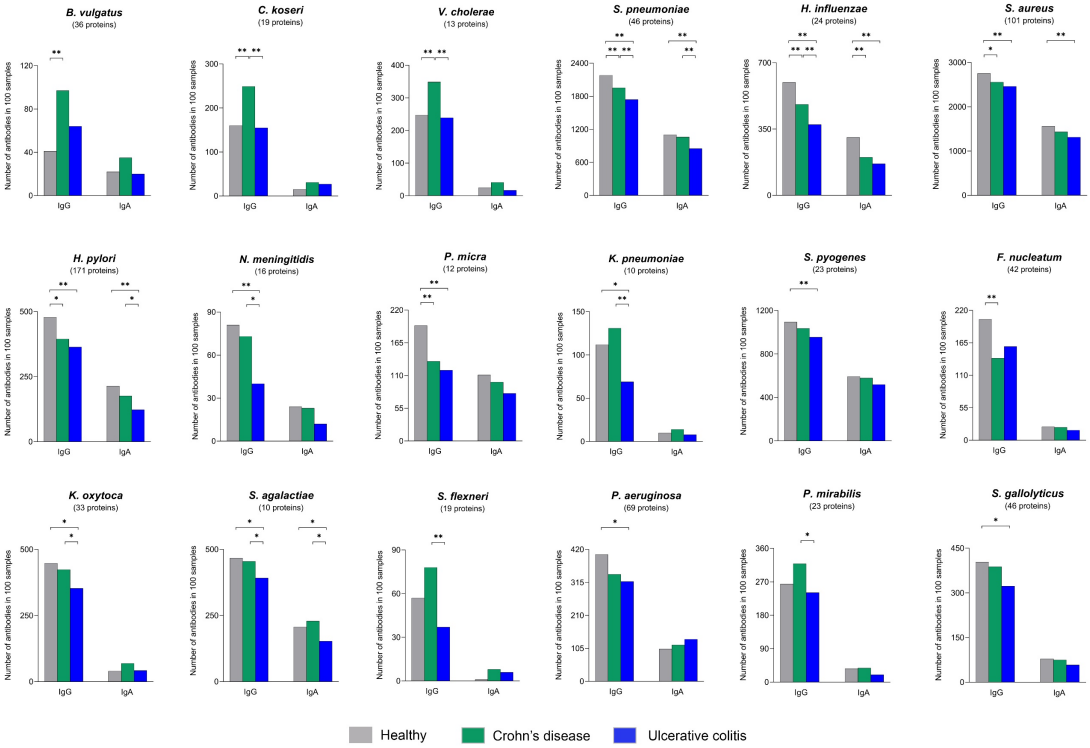


Fig 3.2. Total Number of Antibodies in Healthy Controls, CD and UC. The number of proteins displayed on the microbial protein arrays for each species is shown in parenthesis. The statistical significance of the difference in seroprevalence between groups were calculated using Chi-squared test, $*P < 0.05$, $**P < 0.01$.

3.3.2 Antibodies Distinguishing CD from Healthy Controls

We compared reactivity for individual anti-microbial antibodies between CD patients and healthy controls. We randomly split samples evenly into the discovery and the validation sets (**Table 3.1**). For antibodies with elevated reactivity in CD patients, 13 IgG antibodies passed the criteria (sensitivity ≥ 14% at 96% specificity) in both discovery and validation sets (**Table 3.2**). Anti-A4-Fla2_IgG, a well-studied anti-bacterial flagellin

antibody in CD, had the best performance with 47% sensitivity at 96% specificity in the full sample set (**Table 3.2**). It is interesting to note that the target antigens for 7 out the 13 validated IgG antibodies belonged to the flagellin family apart from A4-Fla2 (**Table 3.2**) and shared high sequence similarity among themselves and with the N-terminal conserved immunodominant region of A4-Fla2 (BLAST e-value < 1e05) (**Fig. 3.3A**). Besides these 7 bacterial flagellins, there were 9 additional proteins showing sequence homology with the N-terminal conserved immunodominant region of A4-Fla2 on the arrays with the percentage identity of the N-terminal conserved region varied from 28% for CJ_flaA to 52% for PMI_RS07910 with A4-Fla2 (**Fig. 3.3B**). All flagellins had higher reactivity in CD than in controls, but the level of difference varied (data not shown). Four target antigens of the validated antibodies, BVU_0562, SP_1992, PMI_RS06815, and SF_Lpp showed no significant homology to flagellins.

Table 3.2 Sensitivity of CD & UC vs. Control (IgG) Antibodies in Discovery & Validation.

		Antigen	Protein name	Organism	Discovery	Validation	Entire
Crohn's disease	Bacteria	HP_0115	Flagellin B	*H. pylori*	28	48	38
		BVU_0562	Uncharacterized protein	*B. vulgatus*	26	22	25
		CK_LafA	Lateral flagellin	*C. koseri*	20	22	21
		CK_LafA.1	Lateral flagellin	*C. koseri*	16	26	24
		A4-Fla2	Flagellin	*L. bacterium A4*	40	54	47
		PMI_RS06815	Hypothetical protein	*P. mirabilis*	14	16	15
		VC_flaD	Flagellin	*V. cholerae*	24	18	19
		VC_flaB	Flagellin	*V. cholerae*	28	22	24
		VC_flaE	Flagellin	*V. cholerae*	26	28	23
		VC_flaA	Flagellin	*V. cholerae*	20	22	21
		SF_Lpp	Outer membrane lipoprotein	*S. flexneri*	14	18	14
		SP_1992	Cell wall surface anchor	*S. pneumoniae*	20	16	18
	Virus	BILF2	Glycoprotein BILF2	Human herpesvirus 4	18	18	18
Ulcerative colitis	Bacteria	CK_flgG	Flagellar basal-body rod protein	*C. koseri*	14	16	15
		A4-Fla2	Flagellin	*L. bacterium A4*	22	16	18
	Virus	BVRF2	Capsid scaffolding protein	Human herpesvirus 4	14	16	14
		UL139	Membrane glycoprotein UL139	Human herpesvirus 5	14	20	17

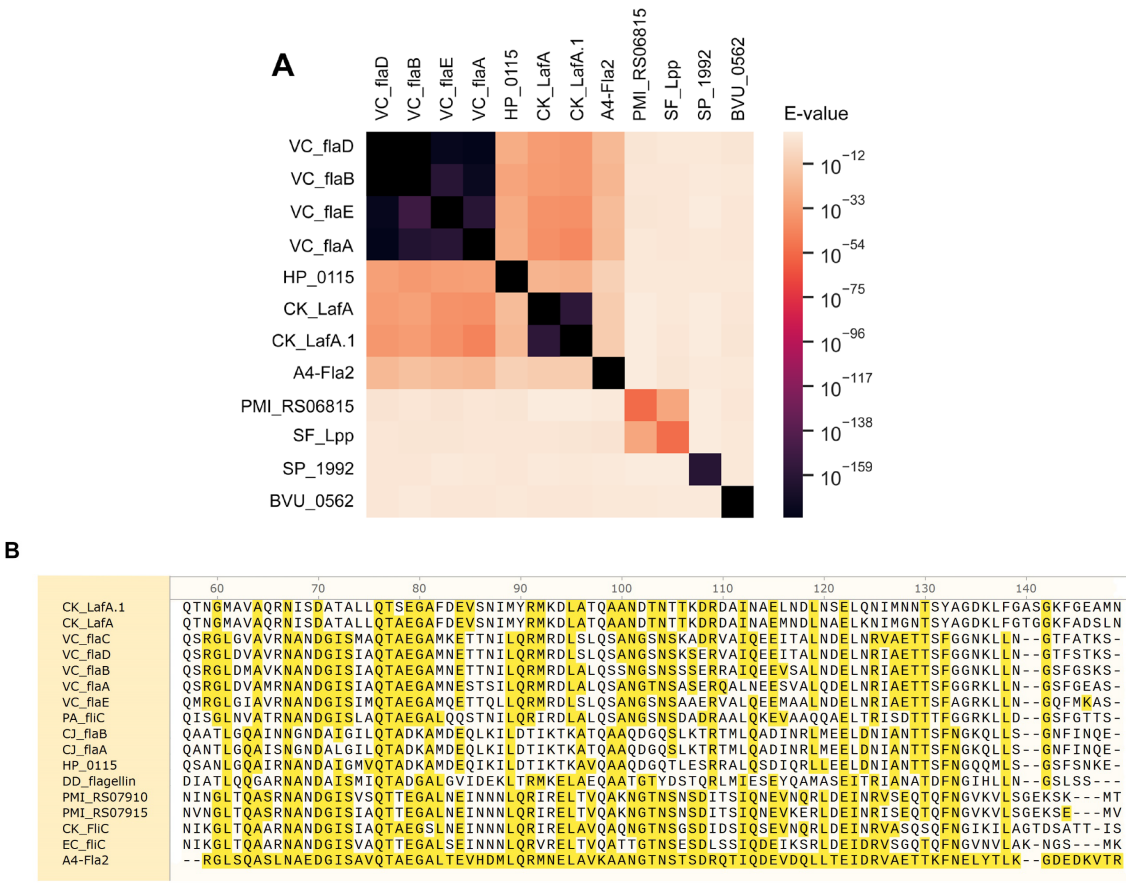


Fig 3.3. Sequence Homology of Proteins. A. Heatmap showing sequence homology among target antigens for antibodies with validated performance of ≥ 14% sensitivity at 96% specificity comparing CD patients with healthy controls. **B.** Multiple sequence alignment of proteins on the microbial protein arrays that had BLAST E-value < 1E-05 with N-terminal immunodominant region of A4-Fla2. Amino acids identical to A4-Fla2 amino acid sequence were highlighted in yellow.

To our surprise, 12 validated IgG antibodies showed elevated reactivity in healthy controls relative to CD patients (**Table 3.4**). Among these 12 antibodies, anti-bacterial antibodies performed better in differentiating CD patients from healthy controls than anti-

viral antibodies (**Table 3.4**). Antibody against SPy_2009, an anchoring protein located in the cell wall of *S. pyogenes*, had the highest sensitivity of 24% at 96% specificity in healthy controls relative to CD patients. Seven (7) validated IgA antibodies showed reactivity in healthy controls relative to CD patients (**Table 3.6**).

3.3.3 Antibodies Distinguishing UC from Healthy Controls

For anti-microbial antibodies with elevated reactivity in UC patients relative to healthy controls, 4 IgG antibodies passed the criteria in both discovery and validation sets (**Table 3.2**). Antibodies to A4-Fla2_IgG and a flagellin from *C. koseri* had a sensitivity of 18% and 15% respectively. For IgG antibodies with higher reactivity in healthy controls relative to UC patients, 32 antibodies got validated (**Table 3.5**). Source microorganisms for the target antigens of these 32 antibodies were enriched for *S. pneumoniae, S. aureus,* and *H. influenzae* (2-sample proportion test, $P < 0.05$). 2.7% of the proteins on the microbial protein arrays were from *S. pneumoniae* while 18.7% of antigens for validated antibodies were from *S. pneumoniae*, 6.1% of the proteins on the arrays were from *S. aureus* while 18.7% of antigens for validated antibodies were from *S. aureus,* and 1.4% of the proteins on the arrays were from *H. influenzae* while 12.5% of antigens for validated antibodies were from *H. influenzae*. Nine (9) validated IgA antibodies showed reactivity in healthy controls relative to UC patients (**Table 3.6**). Fewer anti-viral antibodies than anti-bacterial antibodies were validated comparing CD or UC patients with healthy controls (**Table 3.2**). Anti-viral antibodies to Rhinovirus B14, Enterovirus C, Influenza A virus, Human metapneumovirus had higher reactivity in healthy controls compared with CD and UC patients (**Table 3.4** and **3.5**).

3.3.4 Comparison of Anti-microbial Antibody Response Between CD and UC

We found 46 IgG and 22 IgA validated anti-microbial antibodies with higher reactivity in CD patients compared to UC patients while 28 IgG and 9 IgA validated anti-microbial antibodies with higher reactivity in UC patients compared to CD patients. There was minimal overlap of the target antigens of these validated IgG and IgA antibodies (**Fig. 3.4**).

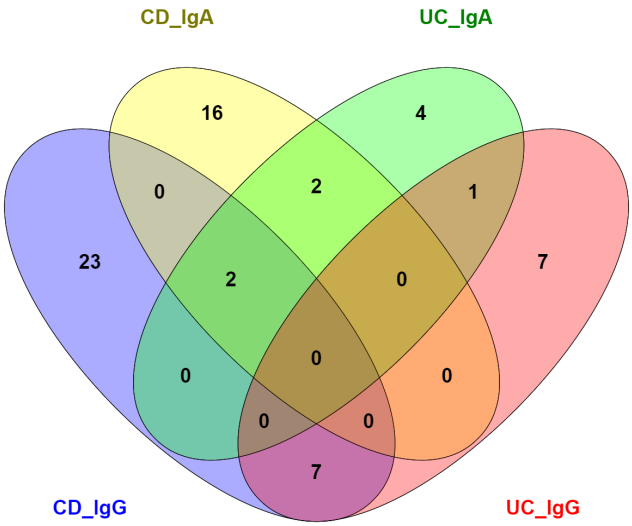


Fig 3.4. Overlap of Target Antigens for IgG and IgA Anti-microbial Antibodies. The Venn diagram shows the differential reactivity between CD and UC patients with ≥ 14% sensitivity at 96% specificity.

3.3.5 Multivariate Analysis to Distinguish CD, UC, and Healthy Controls

We built multi-antibody panels that could distinguish CD vs. control, UC vs. control, and CD vs. UC with an area under curve (AUC) of 0.81, 0.87, and 0.82

respectively. For CD vs. control, the novel flagellins (HP_0115, CK_LafA, CK_LafA.1, VC_flaD, VC_flaB, VC_flaE, VC_flaA) had an AUC of 0.73, the non-flagellin (BVU_0562, SP_1992, PMI_RS06815, and SF_Lpp) had an AUC of 0.75 and the combined AUC of novel flagellins and non-flagellins was 0.81 (**Fig. 3.5A**). For UC vs. control, combination of seven antibodies, four against *S. pneumoniae* and one each against *S. aureus*, *H. influenzae* and *B. vulgatus* had an AUC of 0.87 (**Fig. 3.5B**). For CD vs. UC, combination of seven antibodies, two against *H. pylori* and one each against *E. coli*, *S. pneumoniae*, *S. pyogenes*, *C. jejuni* and *L. bacterium A4* had an AUC of 0.82 (**Fig. 3.5C**).

3.3.6 Subgroup Analysis Based on the Montreal Classification

We investigated the association of UC disease extent (E1, E2, E3), CD disease behavior (B1, B2, B3) and disease location (L1, L2, L3) based on the Montreal classification with the anti-microbial antibody reactivity. We calculated the 4^{th} quartile odds ratio for each antibody between the two comparison groups and compared number of antibodies with significant odds ratio P-values with higher prevalence in each group. For UC disease extent, 39 antibodies had higher prevalence in E3 (extensive UC) than E2 (left sided UC), while only 11 antibodies had higher prevalence in E2 than E3 (**Table 3.3**). For disease behavior, significantly more antibodies had higher prevalence in patients with more severe disease. For examples, comparing B1 and B2, 32 antibodies had higher prevalence in B2 (stricturing) but 0 had higher prevalence in B1 (non-stricturing, non-penetrating). Comparing B2 and B3, 19 antibodies had higher prevalence in B3 (penetrating) than B2 but 0 had higher prevalence in B2 than in B3 (**Table 3.3**). For

disease location, 38 antibodies had higher prevalence in L2 (colonic) than in L1 (ileal), while 0 had higher prevalence in L1 than L2. Comparing L2 and L3 (ileocolonic), L2 was slightly higher than L3 (**Table 3.3**).

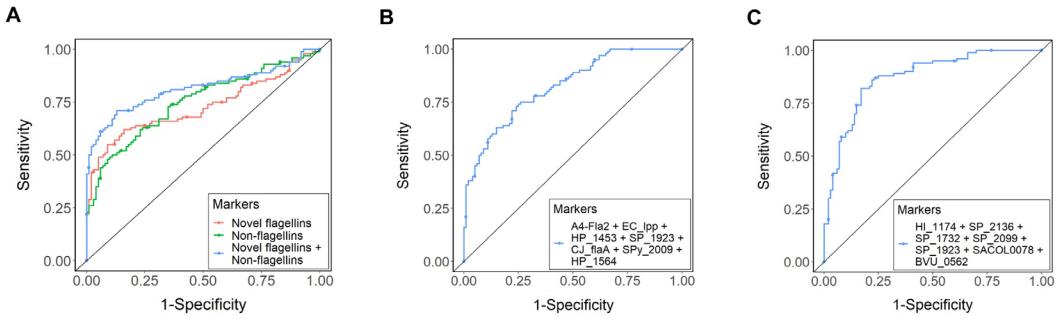


Fig 3.5. ROC Curves to Discriminate CD, UC, and Healthy Controls. **A**, ROC curve for CD vs. healthy controls. The AUC values of novel flagellins (HP_0115, CK_LafA, CK_LafA.1, VC_flaD, VC_flaB, VC_flaE, VC_flaA) and non-flagellins (BVU_0562, SP_1992, PMI_RS06815, SF_Lpp) was 0.73 and 0.75, respectively. The AUC value obtained with a combination of novel flagellins and non-flagellins was 0.81. **B**, ROC curve for UC vs. healthy controls. The AUC value obtained with a combination of 7 markers was 0.87. **C**, ROC curve for CD vs. UC. The AUC value obtained with a combination of 7 markers was 0.82.

Table 3.3 Subgroup Analysis of IBD Patients Based on the Montreal Classification. For each row, the markers with significant difference in reactivity between two classifications were counted based on odds ratio (OR) > 1 and OR < 1, which accounts for the number of markers with high reactivity in both the classification. The difference in number of markers between OR > 1 and OR < 1 was calculated by comparing it with the number of markers with non-significant (*P* >= 0.05) reactivity.

Montreal classification		Number of markers with		Fisher's exact test
		OR > 1	OR < 1	
Disease Behavior B1: non-stricturing, non-penetrating. B2: stricturing. B3: penetrating	B1 vs. B2 (*P* < 0.05)	0	32	*P* < 0.001
	B2 vs. B3 (*P* < 0.05)	0	19	*P* < 0.01
	B1 vs. B3 (*P* < 0.05)	2	41	*P* < 0.001
Disease Location L1: ileal. L2: colonic. L3: ileocolonic	L1 vs. L2 (*P* < 0.05)	0	38	*P* < 0.001
	L2 vs. L3 (*P* < 0.05)	9	5	*P* = 0.08
	L1 vs. L3 (*P* < 0.05)	5	16	*P* = 0.81
Disease Extent E2: left sided UC; E3: extensive UC	E2 vs. E3 (*P* < 0.05)	11	39	*P* = 0.16

3.3.7 Correlation of Anti-microbial Antibodies and Autoantibodies

We previously reported novel autoantibodies in CD patients using the same set of CD patients and healthy controls (Wang et al., 2017). We profiled both IgG and IgA for autoantibodies and anti-microbial antibodies in all 100 CD and 100 healthy controls. It is interesting to note the antibodies showing differences for autoantibodies were mostly

IgA, but the anti-microbial antibodies were mostly IgG (Wang et al., 2017). Anti-SNRPB_IgA had the highest sensitivity of 20% at 96% specificity among all autoantibodies compared with 47% sensitivity at 96% specificity for the best performing anti-microbial antibody, anti-A4-Fla2_IgG.

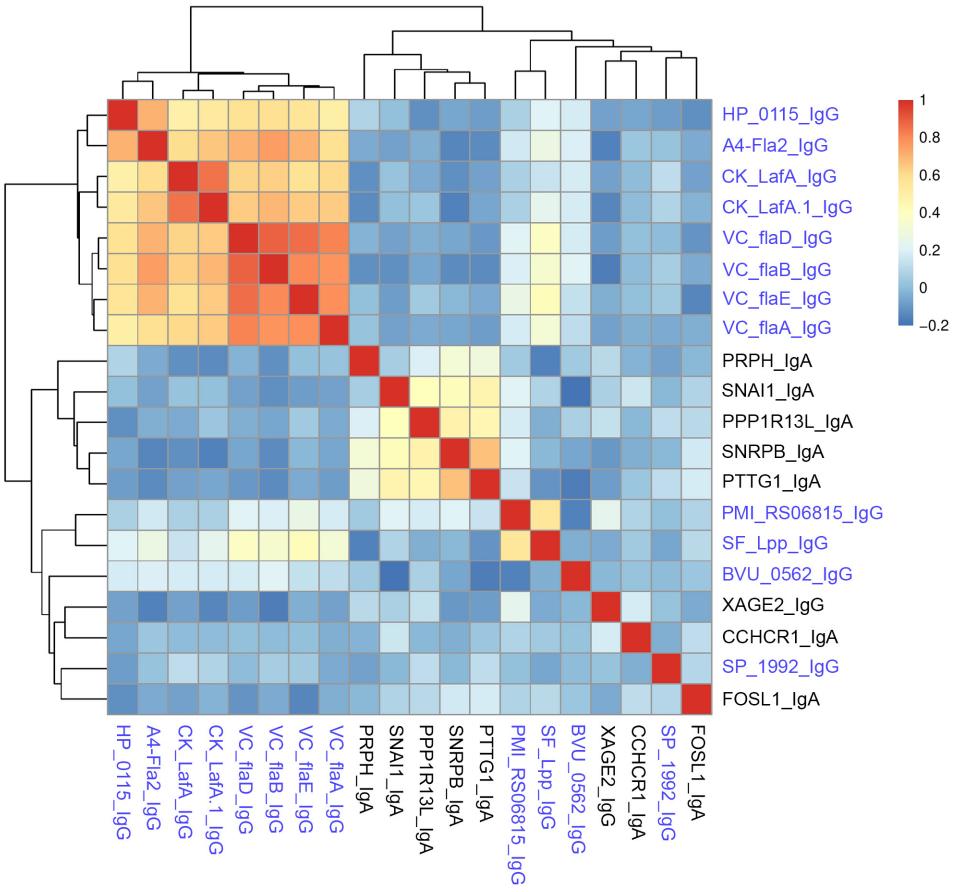


Fig 3.6. Correlation Heatmap of Anti-microbial Antibodies & Autoantibodies. Spearman's rank correlation in CD patients. The names of anti-microbial antibodies are colored in blue while autoantibodies are colored in black.

We compared the novel autoantibodies and validated anti-microbial antibody profiles to determine if correlation exists between their reactivity. Overall, we did not

observe high correlation between autoantibodies and anti-microbial antibodies in CD patients (**Fig. 3.6**). Anti-microbial antibodies formed two clusters, one with anti-flagellin antibodies, and the other with SF_Lpp_IgG and PMI_RS06815_IgG. Five autoantibodies, PRPH_IgA, SNAI1_IgA, PPP1R13L_IgA, SNRPB_IgA and PTTG1_IgA, formed a cluster. The rest antibodies had relatively unique reactivity patterns.

3.4 Discussion

We have performed a microbiomics study to understand the connection between microbial infections and IBD and to identify antibody signatures that can benefit accurate diagnosis of IBD. We studied antibody response to 1570 microbial antigens from 50 species of bacteria and 33 species of viruses in 100 CD and 100 UC patients. Besides flagellins, we identified several non-flagellin anti-microbial antibodies with elevated reactivity in CD patients compared with healthy controls. On the contrary, we observed many anti-microbial antibodies with lower reactivity in UC patients relative to healthy controls. We were able to build antibody panels that could distinguish CD vs. control, UC vs. control and CD vs. UC with AUC of 0.81, 0.87, and 0.82, respectively. We showed that both CD and UC patients with severer disease had stronger anti-microbial antibody response. We demonstrated that anti-microbial antibodies and autoantibodies had different reactivity patterns in CD patients.

Our comprehensive study of anti-microbial antibodies in IBD patients provided interesting insight into its pathogenesis. Antibody responses to *B. vulgatus* and *C. koseri* were shown to be elevated in CD patients. *B. vulgatus* has been reported to induce colitis in IBD-susceptible mice (Bloom et al., 2011; Glassner et al., 2020). Members of the

Citrobacter genus were previously shown to infect urinary tract, respiratory tract and intestine (Donnenberg, 2015). *C. rodentium* was also reported to induce colitis in mice (Koroleva et al., 2015). Our results suggest that *B. vulgatus* and *C. koseri* may also play a role in human CD development. We observed reduced antibody responses in UC patients to several genera of the Firmicutes phylum including *P. micra*, *S. pyogenes, S. aureus,* which were often reduced in abundance in UC patients' gut microbiota (Zakerska-Banaszak et al., 2021). For several genera belonging to Proteobacteria phylum, such as *H. influenzae*, *H. pylori*, *K. oxytoca,* we observed overall reduced antibody responses; however, their abundance in the gut microbiota of UC patients were reported to be either increased or remained the same compared with healthy controls (Frank et al., 2007; Manichanh et al., 2006b). Antibody response connotes exposure and functional immunological interaction between a microorganism and the host; however, anti-microbial antibody by itself does not prove causality. Source microorganisms whose antibodies showing significant changes between IBD patients and healthy controls warrant future confirmation and functional assessment in inducing IBD.

We identified several antibodies with higher reactivity in CD patients relative to healthy controls. Among them, most were against bacterial flagellins from *H. pylori, C. koseri,* or *V. cholerae*. Flagellin is a major structural protein of motile bacterial flagella, and it interacts with the pattern recognition receptor, Toll-like receptor 5 (TLR5). Flagellin is highly antigenic and has the potential to induce dendritic cell maturation, cytokine, and chemokine production. Antibodies against flagellins are well documented in IBD (Lodes et al., 2004; Targan et al., 2005). A4-Fla2 flagellin included in our study showed IBD specific reactivity with performance similar to that reported in the literature

(**Table 3.2**). Sixteen proteins included in this study were homologous to the N-terminal immunodominant region of A4-Fla2 (Lodes et al., 2004). There is some correlation of reactivity and degree of sequence homology with A4-Fla2; however, several flagellins from *C. koseri*, *E. coli*, and *D. desulfuricans* did not show high reactivity in CD patients but had high sequence homology to A4-Fla2. We may have a better understanding of the critical amino acids in the antigenicity of A4-Fla2 as we study more bacterial flagellins in the future.

Previous studies mostly focused on antibodies with higher reactivity in IBD patients (Mitsuyama et al., 2016). Our unbiased data driven approach revealed the existence of many anti-microbial antibodies with higher reactivity in healthy controls relative to CD and especially UC patients (**Table 3.4** and **3.5**). The reduction observed in CD and UC patients may be attributed to the dysbiosis and reduced diversity of gut microbiota in CD and UC patients (Furrie et al., 2004; Manichanh et al., 2006b). The higher number of antibodies having higher reactivity in CD patients compared with UC patients indicates stronger anti-microbial humoral immunity in CD than in UC. This agrees with reports in the literature that most known anti-microbial antibodies, such as ASCA, anti-OmpC, anti-Cbir1, and anti-I2, had higher prevalence in CD patients than in UC patients (Mitsuyama et al., 2016).

We compared serum anti-microbial antibody response in IBD patients with different disease behavior (B1, B2, and B3 for CD), disease location (L1, L2, L3 for CD) and disease extent (E1, E2, E3 for UC), according to the Montreal classification. Our results were consistent with previous reports that increasing diversity and magnitude of anti-microbial immune response is correlated with increased frequency of penetrating

and/or stricturing disease behavior (Dubinsky et al., 2006; Kuna, 2013).We found that penetrating disease behavior (B3) had the highest antibody response, followed by stricturing disease behavior (B2) while non-penetrating and non-stricturing disease behavior (B1) had the lowest antibody response. Anti-microbial antibody response was also associated with disease locations. Disease located in colon (L2) had the highest response followed by ileocolonic (L3) and ileum (L1). Similarly, extensive UC (E3) had higher anti-microbial antibody response compared to left-sided UC (E2). It is known that colon has microbial density of 10^{11}-10^{12} anaerobic bacteria/gram while ileum is colonized by 10^{7}-10^{8} anaerobic bacteria/gram (Sartor, 2008). Stronger anti-microbial immune response in patients with severer CD or UC disease suggests higher abundance of the source microorganisms for the target antigens of the differential antibodies and / or stronger more conducive immune microenvironment at the disease site in severer disease.

Both autoantibodies and anti-microbial antibodies associated with IBD have been reported (Mitsuyama et al., 2016). One popular hypothesis for the autoantibody elicitation is molecular mimicry, i.e., anti-microbial antibodies cross react with human proteins. However, we found minimal correlation between the anti-microbial antibody and the autoantibody profiles in the same set of CD samples (**Fig. 3.6**). The lack of correlation suggests that IBD specific autoantibodies and anti-microbial antibodies should be elicited independently through different underlying mechanisms, and cross-reactivity may play a less role in eliciting CD associated autoantibodies. The break of immune tolerance to human proteins might have occurred due to the damaged gut epithelial cells and the faulty immunological microenvironment partly caused by microbial infections. In addition, the elicitation of autoantibodies may be associated with

the infections of multiple microorganisms, and the correlation with individual anti-microbial antibodies may not be great.

Strengths of our study include the broadest analysis of IgG and IgA antibodies against individual antigens from many different microorganisms in both CD and UC patients and the use of a two-stage approach with discovery and independent validation of antibody markers. There are some limitations to our study. Except for a few microbes, the number of proteins studied for each species is small, which might limit our interpretation of antibody response in IBD at the species level. Furthermore, many samples used in studies were collected from patients with established disease. Future studies with samples collected from newly diagnosed patients will strengthen our ability to identify diagnostic markers and microbial connections for IBD development. In summary, we have demonstrated the power of a microbiomics study of anti-microbial antibodies in IBD for the identification of anti-microbial antibody signatures to improve early accurate diagnosis and help understand IBD etiology.

Supplementary tables and figures

Table 3.4 Sensitivity of Control vs. CD (IgG) Antibodies in Discovery & Validation.

	Antigen	Protein name	Organism	Discovery	Validation	Entire
Bacteria	HP_1564	ABC transporter substrate-binding protein	*H. pylori*	14	16	16
	SACOL0985	MAP domain-containing protein	*S. aureus*	16	14	14
	AUO97_RS08350	hypothetical protein	*A. baumannii*	14	14	13
	SACOL1164	complement convertase inhibitor Ecb	*S. aureus*	14	16	14
	SPy_2009	LPXTG-anchored fibronectin-binding protein FbpA	*S. pyogenes*	38	22	24
	HI_0162	hypothetical protein	*H. influenzae*	18	16	16
	PA_exoT	T3SS effector bifunctional cytotoxin exoenzyme T	*P. aeruginosa*	14	14	13
	AB185_RS23245	type VI secretion system effector Hcp	*K. oxytoca*	18	18	19
	AB185_RS19385	Hcp family type VI secretion system effector	*K. oxytoca*	20	16	19
Virus	null	capsid protein, partial	Rhinovirus B14	14	22	17
	null	nucleocapsid protein	Human coronavirus	32	16	18

Table 3.5 Sensitivity of Control vs. UC (IgG) Antibodies in Discovery & Validation.

	Antigen	Protein name	Organism	Discovery	Validation	Entire
Bacteria	SACOL0858	extracellular matrix protein-binding adhesin Emp	*S. aureus*	20	14	11
	SACOL1140	LPXTG-anchored heme-scavenging protein IsdA	*S. aureus*	16	20	16
	SACOL0078	phosphatidylinositol-specific phospholipase C	*S. aureus*	14	20	16
	SACOL2197	MAP domain-containing protein	*S. aureus*	16	16	14
	PMI_RS02875	peptidoglycan-associated lipoprotein Pal	*P. mirabilis*	14	16	15
	SPy_2191	lytic transglycosylase domain-containing protein	*S. pyogenes*	16	22	20
	SPy_cfa	CAMP factor pore-forming toxin Cfa	*S. pyogenes*	18	20	16
	HI_0256	outer membrane protein assembly factor BamC	*H. influenzae*	18	18	15
	HI_null	cell envelope integrity protein TolA	*H. influenzae*	14	18	15
	HI_0162	hypothetical protein	*H. influenzae*	16	20	18
	HI_1174	outer membrane beta-barrel protein	*H. influenzae*	16	16	16
	PM_null	InlB B-repeat-containing protein	*P. micra*	14	20	14
	SP_1732	Stk1 family PASTA domain-containing Ser/Thr kinase	*S. pneumoniae*	24	32	24
	SP_2136	choline-binding protein PcpA	*S. pneumoniae*	22	32	23
	SP_0785	membrane-fusion protein	*S. pneumoniae*	16	32	19
	SP_0366	oligopeptide ABC transporter, oligopeptide-binding protein AliA	*S. pneumoniae*	14	22	17
	SP_1923	pneumolysin	*S. pneumoniae*	32	34	33
	SP_0377	choline-binding protein CbpC	*S. pneumoniae*	20	28	22
	SACOL1869	serine protease SplA	*S. aureus*	16	18	16
	AB185_RS27465	peptidoglycan-associated lipoprotein Pal	*K. oxytoca*	16	14	15
	SACOL2291	CHAP domain-containing protein	*S. aureus*	24	18	15
Virus	PVgp1	capsid protein VP1	Enterovirus C	14	20	18
	null	capsid protein, partial	Rhinovirus B14	14	24	18
	null	polyprotein	Rhinovirus B14	34	36	28
	null	polyprotein	Coxsackievirus B4	22	24	18
	N	Nucleoprotein	Human metapneumovirus	14	28	18
	F	fusion glycoprotein	Human metapneumovirus	16	24	20
	null	fusion protein	Human respiratory syncytial virus B	14	18	16
	PA	Polymerase acidic protein	Influenza A virus	26	20	23
	PVgp1	genome polyprotein	Enterovirus C	20	32	28
	NP	nucleoprotein	Influenza A virus	14	20	17
	M1	matrix protein 1	Influenza A virus	22	32	25

Table 3.6 Sensitivity of Control vs. CD & UC (IgA) Antibodies in Discovery & Validation.

		Antigen	Protein name	Organism	Discovery	Validation	Entire
Crohn's disease	Bacteria	SACOL2509	fibronectin-binding protein FnbB	*S. aureus*	28	18	17
		SACOL2511	fibronectin-binding protein FnbA	*S. aureus*	18	22	19
		SACOL2476	staphylopine-dependent metal ABC transporter substrate-binding protein CntA	*S. aureus*	18	14	12
		SPy_2009	LPXTG-anchored fibronectin-binding protein FbpA	*S. pyogenes*	30	20	21
		HI_null	cell envelope integrity protein TolA	*H. influenzae*	18	18	17
		HI_oapA	opacity-associated protein OapA	*H. influenzae*	16	14	15
		SP_1479	polysaccharide deacetylase family protein	*S. pneumoniae*	18	20	20
Ulcerative colitis	Bacteria	SACOL1868	serine protease SplB	*S. aureus*	18	14	13
		SACOL2509	fibronectin-binding protein FnbB	*S. aureus*	22	18	18
		HI_oapA	opacity-associated protein OapA	*H. influenzae*	14	18	17
		SP_0366	oligopeptide ABC transporter, oligopeptide-binding protein AliA	*S. pneumoniae*	16	16	13
		SP_0346	capsular polysaccharide biosynthesis protein Cps4A	*S. pneumoniae*	20	16	18
		SP_0336	penicillin-binding protein 2X	*S. pneumoniae*	14	16	15
		SP_1479	polysaccharide deacetylase family protein	*S. pneumoniae*	18	18	14
		SP_0377	choline-binding protein CbpC	*S. pneumoniae*	20	16	18
		SACOL2194	hyaluronate lyase HysA	*S. aureus*	20	18	19

CHAPTER 4

4. SERUM ANTIBODYOME REVEALS HEALTHY INDIVIDUALS SHARE COMMON AUTOANTIBODIES

4.1 Introduction

Autoantibodies have been reported in individuals with autoimmune disease and cancer. They are believed to be absent in healthy individuals due to the immune tolerance mechanism (Nemazee, 2017); however, some have been found frequently in healthy individuals, which we call common autoantibodies. These common autoantibodies can confound the search for disease-linked autoantibodies and their documentation will simplify the identification of autoantibodies specific to certain diseases. Indeed, only a small fraction of the autoantibodies reported in the literature have been validated in independent cohorts (Wang et al., 2016), suggesting the classification performance for many reported autoantibodies requires further investigation.

A comprehensive documentation of common autoantibodies will facilitate the elucidation of the complex immunology underlying their elicitation. One class of common autoantibodies is referred to as natural antibodies (NAbs). Unlike adaptive antibodies, NAbs are synthesized by B1 lymphocytes (bearing $CD20^{+}CD27^{+}CD43^{+}CD70^{-}$) and marginal zone B cells (Griffin et al., 2011; Palma et al., 2018b), and do not undergo affinity maturation by antigen stimulation or extensive somatic mutation (Coutinho et al., 1995). Another class of common autoantibodies may arise from cross-reactive antibodies to infectious agent proteins when the similarity in foreign and self-peptides may activate self-reactive T or B cells. It has been

experimentally demonstrated that patients with either measles virus or herpes simplex virus type 1 produce antibodies against a viral phosphoprotein that cross-react with an intermediate filament protein of human cells (Fujinami et al., 1983). Additionally, transgenic mice infected with lymphocytic choriomeningitis virus (LCMV) may develop chronic inflammation in the central nervous system (CNS) due to epitopes shared between LCMV proteins and CNS antigens (Evans et al., 1996). Several bioinformatics techniques have been developed to discover potential mimicry candidates (Doxey & McConkey, 2013; Ludin et al., 2011; Venigalla et al., 2020).

The immunogenicity of a protein can be attributed to its intrinsic properties and extrinsic responses by the host (Berzofsky, 1985). Biochemical and structural properties like flexibility, hydrophilicity and beta-turns can promote antigenicity while hydrophobicity, alpha-helices and beta-sheets can suppress antigenicity. Those common autoantibodies do not cause evident autoimmune disease is intriguing. The presence of autoantibodies in serum reflects leakiness of central and/or peripheral tolerance mechanisms (Ludwig et al., 2017). However, their presence does not guarantee a causal role in autoimmune disease development. For autoantibody-induced pathology, the autoantibody needs to bind to the autoantigen to form an immune complex (Suurmond & Diamond, 2015). Sequestration of the autoantigen from autoantibodies can inhibit the autoantibody-induced pathology.

In this report, we have performed a meta-analysis of antibodyome data from 9 different case-control biomarker studies to identify common IgG autoantibodies in healthy individuals (Bian et al., 2016; Gibson et al., 2012; Katchman et al., 2017; Wang et al., 2017; Wang et al., 2015a). A total of 8,282 unique proteins were queried as

possible antigens on 587 subjects including 272 serum samples from healthy individuals and 315 from individuals with various diseases. We identified and compared autoantibodies with the highest prevalence in healthy and diseased individuals. Demographic information, as well as antigen biochemical and structural properties, were correlated to the autoantibody profiles. Sequence similarity between proteins from human viruses and common autoantigens was examined to understand factors that might contribute to the autoantibody elicitation. Tissue expression and subcellular localization were analyzed to explore whether these autoantibodies had access to their cognate antigens in healthy individuals.

4.2 Materials and Methods

4.2.1 Datasets

The healthy subjects included in this study were originally included in 9 different case-control studies (**Table 4.1**). The serum samples were collected from various parts of the USA and the UK. The goal of the original studies was to discover biomarkers of various cancers and autoimmune diseases by comparing the prevalence of antibodies present in diseased and healthy subjects. The presence of antibody was determined using protein microarrays that displayed thousands of human proteins as potential targets. Serum samples were probed on protein microarrays followed by a secondary antibody with a fluorophore tag specific for human IgG. Microarrays were scanned by a laser scanner. The microarray images from the 9 studies were qualitatively examined to identify protein targets that serum antibodies bound using Array-Pro Analyzer (Media Cybernetics) (Bian et al., 2016; Montor et al., 2009). All proteins were not probed by all

samples included in our analysis (**Fig. 4.1**). Several studies focused on female-associated disease and thus only employed samples from females. A table of 8,282 rows of unique proteins and 587 columns of subjects in the case and control groups with binary response data of protein microarrays was created for data and statistical analysis.

4.2.2 Weighted Prevalence

Due to the heterogeneous number of proteins and subjects being analyzed in each study, we computed the weight for the jth antibody as $\hat{p}_j = \sum_{i=1}^{k} w_{ij} p_{ij} / \sum_{i=1}^{k} w_{ij}$, where $p_{ij} = x_{ij}/n_{ij}$ is the prevalence, x_{ij} is the number of expressed samples, and n_{ij} is the number of samples for the jth antibody in the study i, and k is the number of studies. Here, w_{ij}=$(v_{ij} + \tau_j^2)^{-1}$ is the inverse variance-weighting which accounts for the heterogeneous effects between studies (Borenstein et al., 2010), where $v_{ij} = n_{ij}/(p_{ij}(1-p_{ij}))$, $\tau_j^2 = (Q_j - k + 1)/U_j$ if $Q_j > k-1$ or ${\tau_j}^2 = 0$ otherwise, $Q_j = \sum_{i=1}^{k} v_{ij}\left(p_{ij} - p_j\right)^2$, $U_j = (k-1)(\overline{v}_j - s_j^2/(kv_j))$, $p_j = \sum_{i=1}^{k} v_{ij} p_{ij} / \sum_{i=1}^{k} v_{ij}$, $s_j^2 = (\sum_{i=1}^{k} v_{ij}^2 - k\overline{v}_j^2)/(k-1)$, and $\overline{v}_j = \sum_{i=1}^{k} v_{ij}/k$. The same analysis was performed to calculate gender-specific weighted prevalence by splitting the dataset into male and female subsets.

4.2.3 Age and Gender Comparison

To understand the effect of age on autoantibody counts in healthy individuals, studies having both male and female subjects with age information were used (Studies I, II, IV, VI, VII, **Table 4.1**). A total of 160 subjects were divided into 4 groups based on

human development stages. The groups were 0 to 6 years old (infancy & early childhood), 6 to 12 years old (middle & late childhood), 12 to 18 years old (adolescence), and 18 years and above (adult). The number of autoantibodies in each subject was plotted using GraphPad Prism by age groups. The significance of increase in the autoantibody counts among the four age groups was calculated using the Welch's t-test.

To understand the effect of gender on autoantibody counts in healthy individuals, studies having both male and female subjects with matched age were used (Studies I, II, IV, VII, **Table 4.1**). The subjects were divided into male and female groups. The number of autoantibodies found in each subject was plotted using GraphPad Prism. The significance of difference in autoantibody counts between the male and female groups was calculated using a two-sample unpaired t-test. The weighted prevalence of each autoantibody was calculated for male and female separately. Prevalence values for the 77 most common autoantibodies were plotted as a population pyramid. A paired t-test was performed to determine the significance of the prevalence difference between genders. Pearson correlation of common autoantibodies reactivity in diseased and healthy cohorts were plotted using python seaborn package.

4.2.4 Correlation of Common Autoantibodies

As the presence of common autoantibodies were measured on a binary scale, a phi correlation coefficient (Cramér, 2016) was computed to measure associations between autoantibodies. Specifically, for each pair of antibodies, a phi correlation coefficient was computed for each study, and multiple phi correlation coefficients across different studies were combined into a single phi correlation coefficient using the R meta package. The R

"pheatmap" package was then used to produce correlation heatmap plots for both healthy and diseased cohorts (**Fig. 4.3a**). Here, phi correlation coefficient was not defined when one pair of antibodies showed no responses for all the samples, and these undefined phi correlation coefficients were colored as grey on the heatmap plots. Pairs of antibodies having correlation coefficient higher than 0.6 in both cohorts and have correlation in more than one study were validated.

4.2.5 Sequence Similarity with Viral Proteins

The proteomes of common viruses found in children of the US were downloaded from UniProt as a FASTA file. All the common human viruses were included except sexually transmitted ones as common autoantibodies that develop early in age and then plateau (**Table 4.5**). CD-HIT was employed to remove duplicate sequences in the file (sequence identity cut-off: 1) (Huang et al., 2010). The sequences were then segregated into 14-mer peptides using a Python script (sliding window: 1) and consecutive amino acid repeats (3 or more) were removed. The sequences of all the human proteins analyzed on microarrays were retrieved from DNASU (https://dnasu.org) and split into two sequence databases, "common autoantigens" and "unreactive proteins". The "unreactive proteins" database comprises proteins from the microarrays without any autoantibody responses. Repeats and low-complexity regions were masked using BLAST+ (Basic Local Alignment Search Tool, version 2.10.1) package "segmasker" (Galperin., 2003). A protein-protein BLAST was run with the following parameters, "-ungapped, -db_hard_mask 21, -comp_based_stats F, -evalue 10", between viral 14-mer peptides and "common autoantigens". Another protein-protein BLAST was run between viral 14-mer

peptides and “unreactive proteins” with similar parameters except adjusted “-evalue 593.89” to compensate for the bigger size of unreactive proteins database (Effective search space of “unreactive proteins” and “common autoantigens” databases were 15,970,464 and 268,912, respectively). The total number of amino acids matches higher than the cutoff (7 ungapped amino acids match) was calculated for both databases and compared with the total number of amino acids in each database using a chi-square test (Fig. 3a).

4.2.6 Biochemical and Structural Properties

Biopython (version 1.75) module Bio.SeqUtils.ProtParam for Python (version 3.7.6) was used to calculate the values of aromaticity, isoelectric point, hydrophobicity, the fraction of amino acids in sheets and turns for each protein. Secondary structure and antigenicity prediction methods from Immune Epitope Database (IEDB) were also used. Command-line tools from IEDB analysis resource (http://tools.iedb.org/bcell/download/) were employed to calculate the values of Chou & Fasman beta-turn, Emini surface accessibility, Karplus & Schulz flexibility, and Parker hydrophilicity across the proteins, which were then averaged for each protein. The computed biochemical property values were used for the enrichment analysis on the identified common autoantigens using Gene Set Enrichment Analysis (GSEA) pre-ranked package (version 4.0.3) (Subramanian et al., 2005).

4.2.7 Subcellular Localization and Tissue Expression

All 8,282 proteins were used to query the UniProt database for subcellular localization (downloaded in December 2020), among which 6,875 proteins had subcellular localization data available in the database. Some of the proteins were found simultaneously in more than one location, and hence, seven groups were created to segregate the proteins based on their subcellular localization profiles. Proteins that were found only in one subcellular location were put into “intracellular only”, “membrane only” and “secreted only” groups. Proteins that were found in two subcellular locations were put into “intracellular & membrane”, “membrane & secreted” and “secreted & intracellular” groups. Proteins that were found simultaneously inside the cell, attached to the membrane, and outside the cell were put into “intracellular, membrane and secreted” group. *P* value was calculated to assess the statistical significance of difference in fractions of “intracellular only” proteins for all proteins on the arrays and for common autoantigens using the proportion test.

All 8,282 proteins were mapped to the Ensembl IDs using “BiomaRt” package available for R (version 3.5.0). The Ensembl IDs were used to identify the protein of interest in the Genotype-Tissue Expression (GTEx, version 8) dataset. The gene expression levels in 52 human tissue types, measured in transcripts per million (TPM), were downloaded from GTEx. Expression values for tissue types belonging to the same organ were averaged. Differentially expressed genes for each organ/tissue were identified using edgeR package for R (version 3.6.2) with a cutoff of Log_2 (fold change) > 3 to determine organ/tissue specificity, where the fold change for each gene was calculated by dividing the TPM value in a particular organ/tissue by the mean TPM values in all other organs/tissues. The log2-scaled fold changes across the organs/tissues for each gene were

standardized to the Z scores for data visualization. The Z score profiles were displayed in a heatmap with correlation-based average-linkage clustering by using the seaborn Python package.

4.3 Results

4.3.1 Identity and Prevalence of Common Autoantibodies

Autoantibody profiles for 272 healthy subjects from 9 case-control studies were compiled (**Table 4.1**). There were more females than males, 195 vs. 67, because several studies focused on female-specific diseases such as breast and ovarian cancers. These studies were diverse in terms of subject ages, ranging from infancy to adulthood, with most above 50 years old. Antibodies against 8,282 unique human proteins were studied; although, the number of proteins studied for each subject varied by study (**Fig. 4.1**). For the healthy subjects, 6,030 proteins showed no reactivity, and 2,175 reactive proteins had a weighted prevalence less than 10%. The remaining 77 proteins, termed as common autoantibodies, occurred frequently and had a weighted prevalence between 10% - 47% (**Table 4.4**). Antibodies against STMN4, ODF2, RBPJ, AMY2A, EPCAM, and ZNF688 showed the highest prevalence (**Table 4.2**).

Table 4.1 Study Summary with Demographic Information. Each study was performed independently with age and gender matched case-control samples. Some studies focused on female-specific diseases without samples from male subjects.

Study	Healthy subjects	Diseased subjects	Healthy Male	Healthy Female	Age (median)	No. of proteins in the study
I	40	40	14	26	13	7,653
II	40	40	19	21	71.5	7,653
III	45	45	0	45	--	7,653
IV	40	40	21	19	51	1,666
V	10	20	--	--	--	1,666
VI	30	50	8	22	7.11	1,666
VII	10	21	5	5	51.5	1,985
VIII	30	30	0	30	--	7,854
IX	27	29	0	27	50	7,854
Total	272	315				8,282 (unique)

"--" represents "data not available".

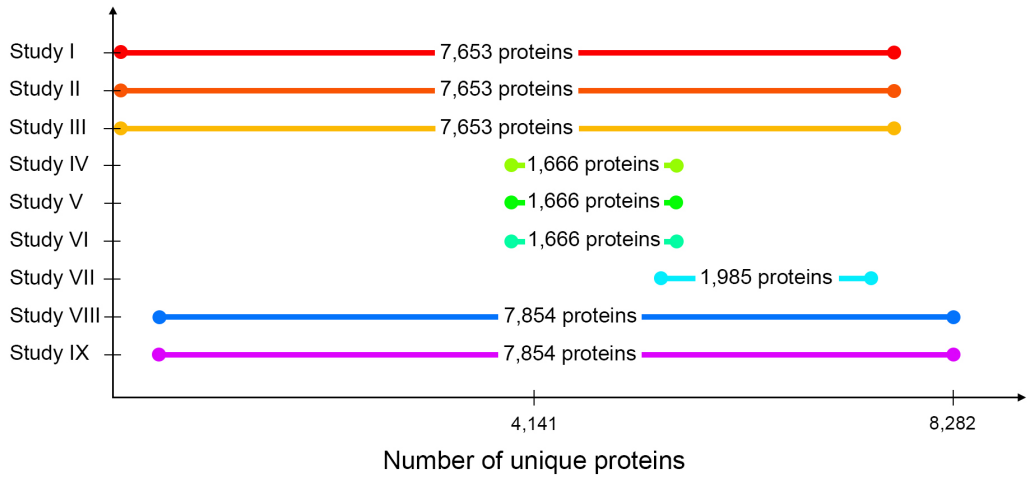


Fig. 4.1. Proteins Investigated by the 9 Studies Used for the Meta-analysis. All 8282 unique proteins are represented on the x-axis and the proteins analyzed in each study is shown as a line overlapping with the x-axis corresponding to the labels on the y-axis. There were 123 proteins studied by all studies and 7,242 proteins by studies I, II, III, VIII and IX.

Table 4.2 Common Autoantigens in Healthy Individuals.

Gene	UniProt ID	Protein	Weighted prevalence
STMN4	Q9H169	Stathmin-4	0.47
ODF2	Q5BJF6	Outer dense fiber protein 2	0.42
RBPJ	Q06330	Recombining binding protein suppressor of hairless	0.37
AMY2A	P04746	Pancreatic alpha-amylase	0.34
EPCAM	P16422	Epithelial cell adhesion molecule	0.31
ZNF688	P0C7X2	Zinc finger protein 688	0.29
CSF3	P09919	Granulocyte colony-stimulating factor	0.25
RAD51AP1	Q96B01	RAD51-associated protein 1	0.23
PSKH1	P11801	Serine/threonine-protein kinase H1	0.23
LENG1	Q96BZ8	Leukocyte receptor cluster member 1	0.22
S1PR3	Q99500	Sphingosine 1-phosphate receptor 3	0.21
LYSMD1	Q96S90	LysM and putative peptidoglycan-binding domain-containing protein 1	0.21
FAM76A	Q8TAV0	Protein FAM76A	0.20
CDR2L	Q86X02	Cerebellar degeneration-related protein 2-like	0.20
CCDC130	P13994	Coiled-coil domain-containing protein 130	0.20

To examine the time course of autoantibody development, we divided 160 healthy subjects from five studies that included age information (Studies I, II, IV, VI, VII, **Table 4.1**) into four age groups based on human development stages. The number of autoantibodies per individual showed an increasing trend during early development stages. The infant and early childhood age group (0-6 years) had the least number of

autoantibodies. The number increased in the middle and late childhood age group (6-12 years) and then plateaued (**Fig. 4.2a**, $P < 0.001$). To investigate whether the number or identity of autoantibodies showed a gender bias, we compiled four studies that included both male and female subjects with matched age (Studies I, II, IV, VII) and compared the counts and identities of the antibodies. The median numbers of autoantibodies for male and female subjects were similar (**Fig. 4.2b**, $P = 0.17$). The weighted prevalence of 77 common autoantibodies also had comparable distribution between male and female subjects (**Fig. 4.2c**, $P = 0.06$).

We reasoned that if these common autoantibodies observed in the healthy subjects were elicited through common non-pathogenic mechanisms, they should also occur at similar frequencies in their matched disease cohorts. Indeed, the 77 common autoantibodies occurred at similar frequencies in diseased cohorts to those in healthy cohorts (**Fig. 4.2d**, Pearson correlation coefficient $r = 0.975$).

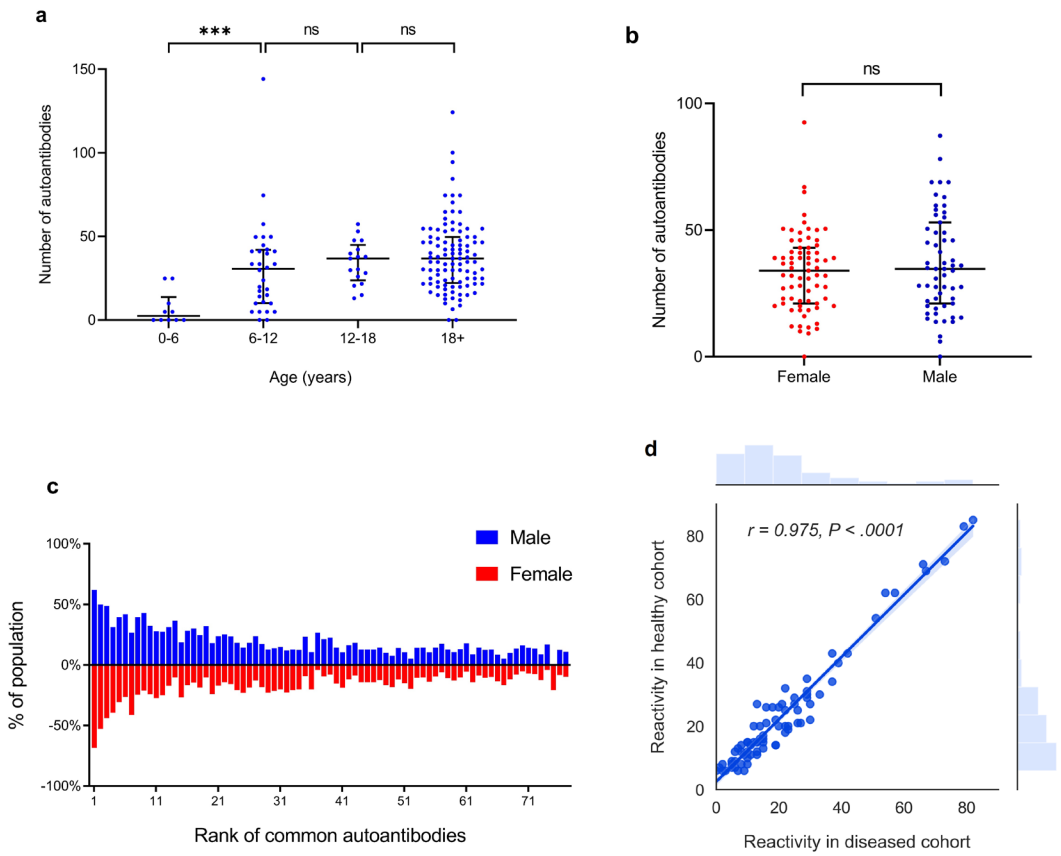
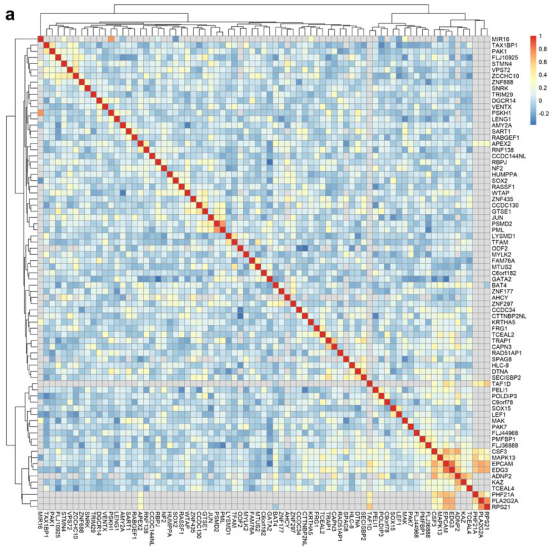


Fig. 4.2. Autoantibody development in healthy subjects. **a**, All subjects were divided into four groups based on human development stages. Each blue dot represents the number of autoantibodies found in a healthy subject belonging to that age group. The number of autoantibodies increased significantly over the age groups ($P < 0.001$). **b**, Comparison of number of autoantibodies in female and male healthy subjects. There are no significant differences between male and female for the number of autoantibodies (two-sample unpaired t-test, $P = 0.17$) **c**, Comparison of weighted prevalence of common autoantibodies in male and female healthy subjects. A blue bar represents the weighted prevalence of a common autoantibody in the male population while a red bar below the blue one represents the weighted prevalence of the same autoantibody in the female population. **d**, Pearson correlation of common autoantibodies reactivity in healthy and diseased cohorts ($r = 0.975$).

We wondered if any of these common autoantibodies were related to each other; that is, was there any concordance among them or were their occurrences independent. We analyzed the common autoantibodies pairwise to determine if any occur together in healthy individuals at frequencies greater than chance alone. We found the majority of them were independent of each other except several pairs: EDG3 and EPCAM (Phi correlation coefficient: 0.83), PML and PSMD2 (Phi correlation coefficient: 0.73), and EPCAM and CSF3 (Phi correlation coefficient: 0.67). Moreover, when we looked at these pairs in the diseased individuals, their concordance was also elevated (**Fig. 4.3**).



b

Alias1	Alias2	Phi correlation coefficient (healthy)	Phi correlation coefficient (diseased)	Number of studies having correlation
EDG3	**EPCAM**	**0.83**	**0.67**	**6**
RPS21	PLA2G2A	0.76	0.53	2
PML	**PSMD2**	**0.73**	**0.63**	**3**
PLA2G2A	EPCAM	0.69	0.44	2
PLA2G2A	EDG3	0.68	0.19	2
CSF3	**EPCAM**	**0.67**	**0.65**	**6**
RPS21	EDG3	0.66	0.58	3
PHF21A	EPCAM	0.63	0.31	2
RPS21	EPCAM	0.63	0.52	3

Fig. 4.3 Correlation of Co-occurrence of Common Autoantibodies in Healthy Cohort. **a**, Phi correlation coefficient was calculated for each pair of antibodies and shown as a heatmap. The grey color on the heatmap represents pairs of antibodies whose phi correlation coefficient was not defined. **b**, Pairs that have correlation coefficient higher than 0.6 in healthy cohort and that have correlation in more than one study were shown with their corresponding value in diseased cohort. Pairs that have correlation coefficient higher than 0.6 in both the cohorts were bolded.

4.3.2 Sequence Similarity with Viral Proteins

To understand the extent that common autoantibodies observed in our study resulted from cross-reactivity of antibodies induced by viral infection, we examined the sequence similarities between viral proteins and common autoantigens. As these autoantibodies developed early in age and did not change after adolescence, respiratory and common viruses found in children of the US were included in the analysis (**Table 4.5**). In order to avoid redundancy and false positives, duplicate proteins and consecutive amino acid repeats were removed from viral proteomes (**Fig. 4.4**). Similarly, human proteins were masked to avoid repeats and low-complexity regions (homopolymeric runs, short-period repeats and over representation of one or few residues) as potential hits. Using 7 ungapped amino acids match as the cutoff, we identified 28 instances of 7 ungapped amino acid matches and 1 instance of 8 ungapped amino acid match with viral proteins that were present in 21 common autoantigens (**Table 4.3**). Some of the matches were from the peptides of high-complexity regions like SYFGLRT, LRQEINA, WPEGYQL, ARCETQN. To assure that these matches were not due to random chance, we analyzed the frequency of 7 or more ungapped amino acids match for the unreactive proteins (i.e., proteins without any autoantibody response) against the same set of viral proteins. To control for increased chance of a match due to protein length, the results were normalized and expressed as frequency at amino acid level. There were 201 amino acids in matched peptides higher than the cutoff among 34,070 amino acids of the common autoantigens while 5,801 amino acids matched higher than the cutoff among 2,026,890 amino acids of the unreactive proteins (Chi-square test, $P < 0.00001$).

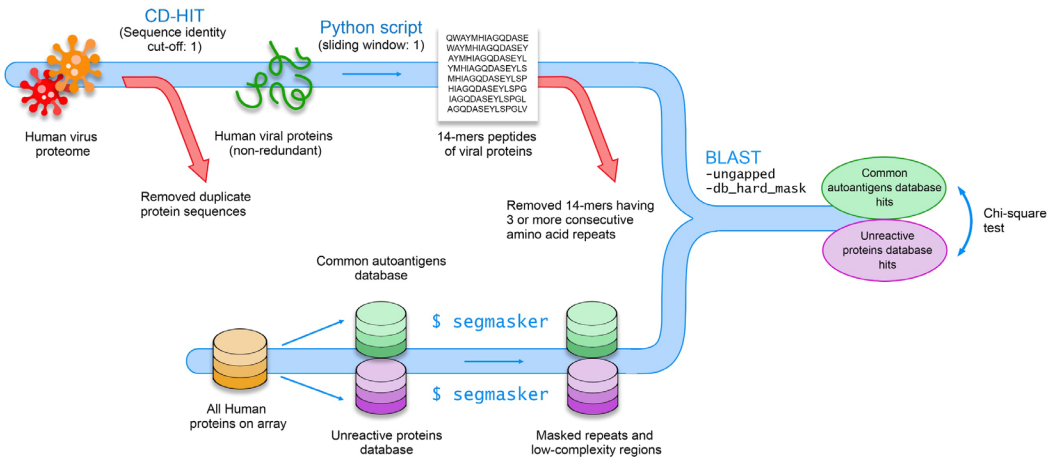


Fig. 4.4 Schematic Diagram for the Discovery of Sequence Similarity. The pipeline was used to find 7 or more ungapped amino acids matches between common autoantigens and viral proteins.

Table 4.3 Sequence Similarity Between Common Autoantigens and Viral Proteins.

S. No.	Autoantigen	Viral UniProt ID	Sequence similarity	Organism
	ADNP2	P16812	LPVPPGG LPVPPGG	Human herpesvirus 5
		H9C1C1	SYFGLRT SYFGLRT	Human rotavirus C
2	AHCY	F8WQQ3	GKLNVKL GKLNVKL	Human adenovirus 41
3	AMY2A	P16766	SAGTSST SAGTSST	Human herpesvirus 5
4	APEX2	M1JRT8	NRSGYSG NRSGYSG	Influenza A virus
		P09289	ALLAAGS ALLAAGS	Human herpesvirus 3
5	C9orf78	P16764	EDCLYEL EDCLYEL	Human herpesvirus 5
6	CTTNBP2NL	P52529	EQLRAKL EQLRAKL	Human herpesvirus 6A
		C4AL53	AKLNREE AKLNREE	Influenza A virus
		Q6SW92	SSNTVVA SSNTVVA	Human herpesvirus 5
7	FLJ36888	P52355	TIKRTLV TIKRTLV	Human herpesvirus 7
8	KAZ	O09800	ARCETQN ARCETQN	Human herpesvirus 1
9	MAK	P16793	GTSEVDE GTSEVDE	Human herpesvirus 5
		Q01350	WPEGYQL WPEGYQL	Human herpesvirus 6A
		Q69513	KSDSELS KSDSELS	Human herpesvirus 7
10	MAPK13	Q8QT31	VIGLLDV VIGLLDV	Human parainfluenza virus 1
11	MTUS2	P09284	IDQNTVV IDQNTVV	Human herpesvirus 3
		A0A0D5Z8N5	SPIKLSP SPIKLSP	Rotavirus B
12	MYLK2	Q6SWD0	TAEEGKNI KAEEGKNI	Human herpesvirus 5
13	PAK1	P24433	SVIEPLP SVIEPLP	Human herpesvirus 6A
14	PAK7	P16739	NATAQELL RATAQELL	Human herpesvirus 5
15	PELI1	Q9QJ30	LRQEINA LRQEINA	Human herpesvirus 6B
16	PML	A0MK42	TLGAVVP TLGAVVP	Human adenovirus 52
17	RABGEF1	I1V183	SPRKQEAE SPRKQEAE	Human adenovirus 7
18	SECISBP2	D3JIS2	ELTVAAR ELTVAAR	Human adenovirus 18
19	TAF1D	P09252	DATHLED DATHLED	Human herpesvirus 3
20	TRAP1	P0C723	ALIRKLRQ ALIRKLRD	Epstein-Barr virus
		P10200	AQLGPRR AQLGPRR	Human herpesvirus 1
21	ZNF688	Q1HVD1	GAQPPAP GAQPPAP	Epstein-Barr virus

4.3.3 Biochemical and Structural Properties

We asked whether any intrinsic biochemical and structural properties of the target antigens were responsible for common autoantibodies production. We examined aromaticity, hydrophobicity, isoelectric point, protein length, the fraction of amino acids

in β-sheets, Chou & Fasman beta-turn prediction, Emini surface accessibility prediction, Karplus & Schulz flexibility prediction, and Parker hydrophilicity prediction by comparing our list of common autoantigens to all 8,282 proteins using Gene Set Enrichment Analysis (GSEA). The 77 common autoantigens were significantly enriched with proteins having low aromaticity (NES or normalized enrichment score: -1.70, $P < 0.001$), low hydrophobicity (NES: -2.02, $P < 0.001$), high isoelectric point (NES: 1.61, $P = 0.018$), high fraction of amino acids in β-turns (NES: 1.99, $P = 0.04$), high Karplus & Schulz flexibility (NES: 4.35, $P < 0.001$), high Parker hydrophobicity (NES: 2.35, $P < 0.001$), and high Chou & Fasman β-turn score (NES: 2.59, $P < 0.001$) (**Fig. 4.5**). However, other biochemical properties such as protein length, the fraction of amino acids in β-sheets, and Emini surface accessibility showed no significant enrichment.

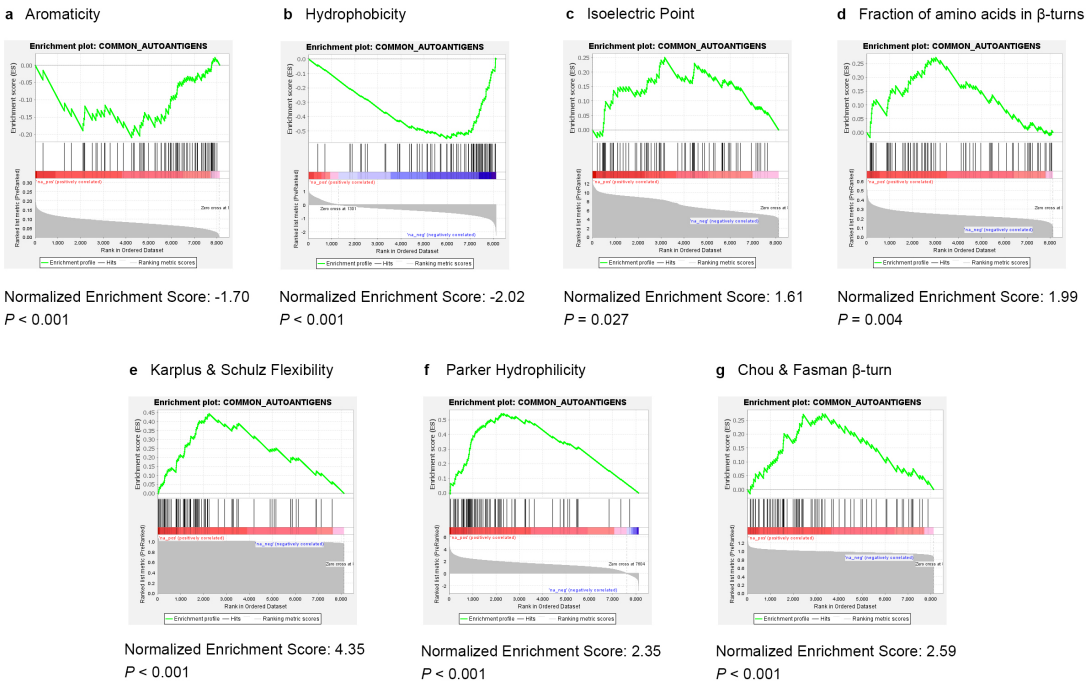


Fig. 4.5 Gene Set Enrichment Analysis of Common Autoantigens. **a**, **b**, **c**, **d** shows primary structure enrichment analysis as labeled. **e**, **f**, **g** shows antigenicity and secondary structure prediction method enrichment analysis as labeled. The grey colored curve on the graph represents the values of the property sorted in descending order for all the proteins studied. The black vertical lines on the graph show where the common autoantigens appear in the ranked list. The green line represents the enrichment score for the corresponding ranked list. Concentration of vertical lines on the graph towards a side signifies enrichment while randomly dispersion of vertical lines on the graph signifies no enrichment.

4.3.4 Subcellular Localization and Tissue Expression

The discovery of common autoantibodies in healthy individuals raised the question about why these antibodies do not lead to autoantibody-mediated pathology. A

primary requirement for autoantibody-mediated pathology is the formation of immune complexes. We examined the subcellular localization of the common autoantigens to see if they were antibody accessible. We divided them into three broad categories: "intracellular" having autoantigens located in the cytoplasm, nucleus, endoplasmic reticulum, Golgi apparatus, mitochondrion, and lysosome, "membrane" having autoantigens located on the cell surface and bound to the cell membrane, and "secreted" having autoantigens secreted from the cell. The localization of an autoantigen can belong to one or more of these 3 categories. We found 55 among 70 common autoantigens were located exclusively at intracellular sites. The percentage of common autoantigens with "intracellular" only subcellular localization was significantly higher than that for all the proteins studied on the arrays (78% vs. 49%, $P < 0.001$) (**Fig. 4.6a**).

Tissue-specific gene expression can impact autoantigen exposure to circulating autoantibodies and the potential to trigger autoimmune disease. In the GTEx dataset, transcripts encoding for 14 common autoantigens were organ/tissue-specific (defined as having $\log_2$ ((organ expression)/(mean expression in all other organs) > 3) (**Fig. 4.6b**). Among them, *PMFBP1, ODF2*, *RNF138,* and *CCDC34* were predominately expressed in testis while *STMN4* and *SOX2* were predominantly expressed in the brain. For instance, *PMFBP1* has 29.47 TPM (transcripts per million) in testis while the mean in other organs is 0.48 TPM. Similarly, *STMN4* has 77.23 TPM in the brain while the mean in other organs is 0.32 TPM. Other common autoantigens did not show tissue specificity.

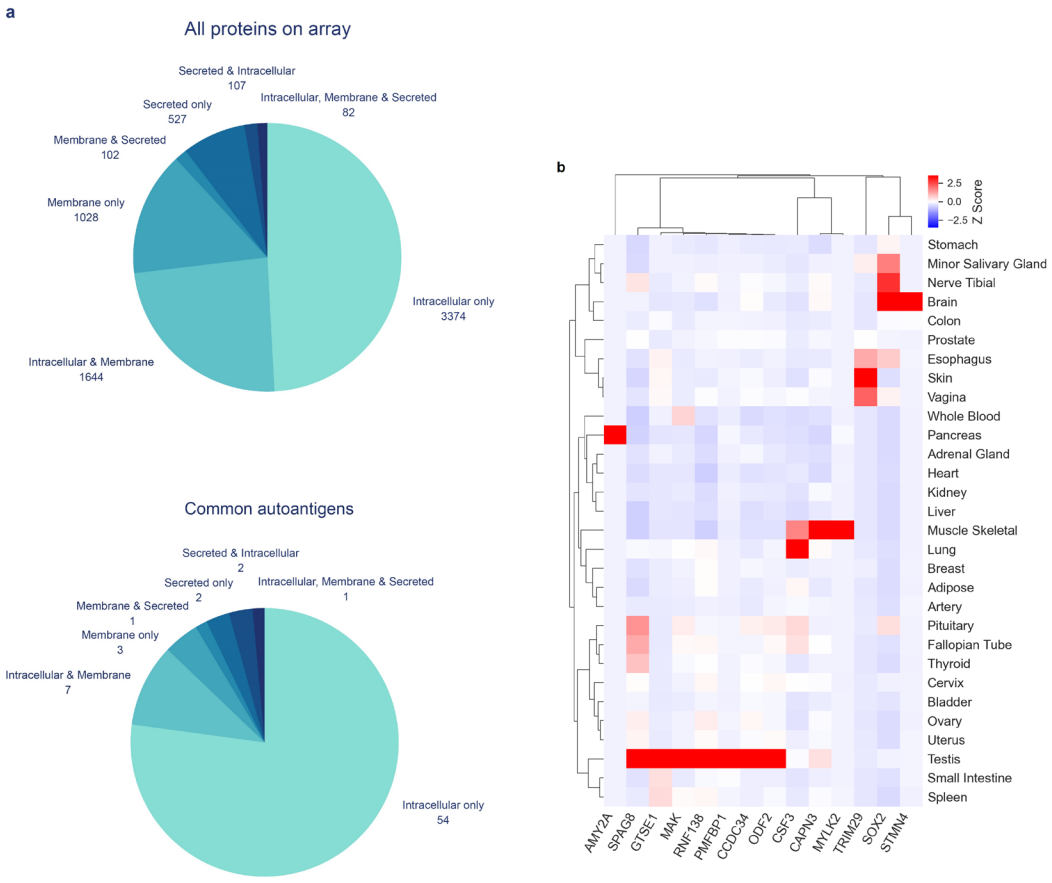


Fig. 4.6 Subcellular Localization and Gene Expression of Common Autoantigens. **a**, Subcellular localization of all proteins on the microarray and common autoantigens. **b**, Expression profiles of organ/tissue-specific common autoantigens. Each row represents an organ as labelled on the right and each column represents an autoantigen as labelled at the bottom. Gene expression in transcripts per million (TPM) from GTEx dataset was standardized to the Z scores for data visualization. Organs and autoantigens were clustered based on correlation-based average-linkage clustering.

4.4 Discussion

Autoantibodies have been well-established as useful biomarkers in autoimmune diseases and cancer (Leslie et al., 2001). Theoretically, autoantibodies should have minimal presence in healthy individuals because of clonal deletion during development; nonetheless, it is widely observed that healthy individuals often have autoantibodies. Some of these autoantibodies occur frequently enough to confound studies intended to find disease-related autoantibodies. To identify commonly found autoantibodies in healthy individuals, we performed a meta-analysis of 9 independent studies using protein microarray data that examined autoantibodies against 8,282 human proteins in 272 healthy individuals of different ages and geographic locations (Bian et al., 2016; Gibson et al., 2012; Katchman et al., 2017; Wang et al., 2015a) and further confirmed these findings on an additional 315 individuals, making this the largest analysis of its kind. We found 77 autoantibodies that occurred frequently in healthy subjects with a weighted prevalence above 10%.

The number of unique IgG autoantibodies in healthy individuals increased with age from infancy to early adulthood. Interestingly, this number did not increase significantly after childhood and remained constant (median 29 ± 6 among 8,282 possibilities) throughout adulthood. This observation suggests that while response to infectious agents (and maybe vaccines) might contribute to autoantibodies through molecular mimicry, this mechanism does not appear to continue to accumulate autoantibodies throughout life. Gender did not appear to play a role in autoantibody production in healthy individuals. The median number of autoantibodies present in healthy male and female were similar (median 35 vs. 34), which is in agreement with the results reported by Neiman *et al*. (Neiman et al., 2019). This stands in contrast to the fact

that autoimmune diseases disproportionally affect females compared to males because male-predominant autoimmune disease is associated with acute inflammation, whereas female-predominant autoimmune disease is associated with antibody-mediated pathology (Fairweather et al., 2008).

We looked for common autoantibodies whose presence correlated with each other in healthy individuals to determine if any were interrelated. This could occur if the same antibody recognized two different proteins that share a common epitope. Other possibilities include sharing common HLA haplotypes or playing similar biological roles that lead to escape from tolerance. The majority of the common autoantibodies showed no apparent relationship to each other; however, there were a number that had a high level of correlation. These observations in the healthy population also held true for the diseased population (**Fig. 4.3**). An evaluation of the target antigens of these correlated antibodies revealed no apparent sequence similarity, and examinations of common protein interaction databases (String, UniProt) did not reveal that any co-occurred in the same protein complexes. It is notable that the targets of several of the co-occurring antibodies play roles in stem cell proliferation and differentiation (EPCAM, EDG3 and CSF3) and two others play roles in DNA damage repair (PML and PSMD2). The meaning of this is not clear, but it occurred frequently enough (Phi correlation coefficient > 0.6) that it is worth further investigation. This observation supports the notion that the identity and level of autoantibodies found in healthy individuals are generally independent but there are subsets of related autoantibodies.

4.4.1 Sequence Similarity with Viral Proteins

Viral proteins with sequences similar to a human protein may initiate cross-reactive antibodies leading to autoimmunity. One theory for a cause of type 1 diabetes, which occurs seasonally in some locations, is that it is triggered by a virus infection, where an association has been found with coxsackievirus B4 (Gamble et al., 1973). Antibody to coxsackievirus B4 can cross-react to pancreatic β-cell antigens leading to diabetes. There are around 20 autoimmune diseases reported in literature where autoantibodies are generated due to cross-reactivity to infectious agent proteins (Cusick et al., 2012). We reasoned that some of the common autoantibodies may be a result from cross-reactivity from anti-viral antibodies, albeit without causing subsequent pathology. Among the targets of the 77 common autoantibodies, 21 common autoantigens had at least one or more 7 ungapped amino acids matches with viral proteins (**Table 4.3**). Some of the matches were from high-complexity regions like SYFGLRT, LRQEINA, WPEGYQL, ARCETQN, etc. The typical length of linear epitope of antibodies ranges from 7 to 9 amino acids and hence these specific matches have the potential to elicit cross-reactive antibodies (Buus et al., 2012; Dunn et al., 1999). The fact that these matches occur significantly more frequently between viral proteins and common autoantigens but less frequently for unreactive proteins on the arrays further suggests the role of molecular mimicry in common autoantibody elicitation.

We also tried another method by comparing full-length human proteins with full-length viral proteins using BLAST. The output predominately yielded conserved proteins like kinases and histones, having very significant E-values. However, we did not observe enrichment of kinases and histones in the common autoantigens relative to other proteins on the arrays (data not shown). A limitation of employing BLAST against the entire

sequence was that it concealed short ungapped alignment at the peptide level, and the output was over represented by low-complexity regions due to inefficient masking (Galperin., 2003). The pipeline demonstrated here is an important technique to discover linear epitopes that can contribute to autoantibody production through potential mimicry mechanism. Of course, the validity of these identified epitopes still awaits further experimental investigation. We also acknowledge that linear epitope matches are not always sufficient, and three-dimensional conformation might play a role. This may explain why there only exist a subset of common autoantigens having sequence similarity with viral proteins.

4.4.2 Biochemical and Structural Properties

The intrinsic properties of a protein, such as its chemical and structural complexity, can impact its antigenicity (Berzofsky, 1985). Based on our Gene Set Enrichment Analysis (GSEA), we found that common autoantigens tended to favor more basic, hydrophilic with fewer aromatic amino acids. In addition, common autoantigens were also found to be more flexible and have more beta-turns. Flexibility is a property that can help the polypeptide chain to bind easily to immunoglobulin compared to a stiff polypeptide chain (Berzofsky, 1985). Also, beta-turns can be a potential site for antibody binding as the peptide chain reverses its direction at beta-turns with side chain projected outwards (Rose et al., 1985). Protein length and the fraction of amino acids present in beta-sheets showed no effect on protein antigenicity. The observation that these B cell antigens favored hydrophilicity contrasts with a recent finding that hydrophobic residues within a peptide improved antigenicity and MHC presentation for $CD8^+$ T cells (Chowell

et al., 2015). This difference may suggest a point of specialization between the two branches of the adaptive immune system. This is certainly highlighted by the differences in antigen presentation, where antibodies are soluble and bind to their whole targets in solution, whereas T cells recognize peptides presented on the surfaces of cells surrounded by MHC molecules.

4.4.3 Subcellular Localization and Tissue Expression

Accessibility of autoantigens to circulating autoantibodies is critical to autoimmune disease pathology (Janeway, 2001). In systemic autoimmune diseases, a majority of the target antigens are intracellular molecules and therefore not normally accessible to the B cells or antibodies (Suurmond & Diamond, 2015). Only after excessive cell death or ineffective clearance of apoptotic debris do these intracellular autoantigens become available for immune complex formation. Antinuclear antibodies (ANA) that bind to nuclear antigens in systemic lupus erythematosus provide an example. In Wegener's granulomatosis, the autoantigen is an intracellular protease that becomes accessible to the autoantibodies only after an infection triggers translocation of the protease to the surface (Janeway, 2001). Similarly, the autoantigen in Goodpasture's syndrome, normally ensconced in the basal membranes of alveolar capillaries, becomes accessible to the antibodies after an environmental insult to the capillaries, leading to pulmonary hemorrhage (Janeway, 2001).

A majority of the common autoantigens we identified were located exclusively at intracellular sites, which make them inaccessible to circulating autoantibodies. We also found that some of the common autoantigens are organ/tissue-specific and predominately

expressed in the testis and brain, which are isolated from the immune system by the blood-testis or blood-brain barriers, respectively. No obvious form of sequestration was identified for the remaining autoantigens although this cannot be ruled out.

4.4.4 Autoantibodies as Disease Biomarkers

Thousands of studies over the past decade have investigated autoantibodies as potential biomarkers for disease risk assessment, diagnosis, and prognosis (Leslie et al., 2001; Yadav et al., 2019). Given the prevalence we observed for these common autoantibodies in healthy individuals, in some cases exceeding a quarter of all individuals, they will be frequently encountered in such studies and may confound them as false positives. An examination of the AAgAtlas & PubMed revealed that 20 of our 77 common autoantibodies have been reported as disease-related biomarkers (**Table 4.4**). Although membership among the common autoantibodies found here does not exclude the possibility that an antibody could not also be disease-specific, it would certainly be beneficial for authors to know which autoantibodies commonly occur in healthy individuals (Dervan et al., 2010; Frostegård et al., 2018). Our report on these common autoantibodies in hundreds of healthy individuals provides a good reference for future studies on disease-specific autoantibodies.

In summary, our comprehensive profiles of autoantibody-omes in healthy people greatly expand our knowledge about autoimmunity in the two genders and its development over ages. Their sequence similarity with viral proteins, biochemical, and structural properties provide novel insight of autoimmunity in healthy people and provide guidance to design future experiments to elucidate the underlying immunology.

Furthermore, disease-specific autoantibodies are important for many clinical applications including risk assessment and accurate early diagnosis, and this list of common autoantibodies will assist in the future discovery of more robust biomarkers.

Supplementary tables and figures

Table 4.4 Weighted Prevalence of Common Autoantibodies in Healthy & Diseased Cohorts. Availability of the autoantibodies in AAgAtlas / PubMed literature were reported as yes [Y] or no [N].

Gene name	No. of reactivity (Healthy)	Total samples (Healthy)	Weighted prevalence (Healthy)	No. of reactivity (Diseased)	Total samples (Diseased)	Weighted prevalence (Diseased)	Found in AAgAtlas / PubMed
STMN4	83	192	0.47	79	205	0.37	Y
ODF2	85	178	0.42	82	182	0.38	Y
RBPJ	69	178	0.37	67	182	0.33	Y
AMY2A	71	167	0.34	66	169	0.33	Y
EPCAM	72	205	0.31	73	235	0.28	Y
ZNF688	54	192	0.29	51	205	0.26	Y
CSF3	62	259	0.25	54	293	0.19	Y
RAD51AP1	30	178	0.23	33	182	0.24	N
PSKH1	15	140	0.23	12	140	0.18	N
LENG1	43	192	0.22	42	205	0.18	N
S1PR3	62	269	0.21	57	314	0.17	Y
LYSMD1	30	178	0.21	29	182	0.17	N
FAM76A	35	178	0.20	29	182	0.20	N
CDR2L	29	192	0.20	25	205	0.12	Y
CCDC130	27	142	0.20	21	143	0.16	N
SOX15	43	192	0.20	37	205	0.14	N
PLA2G2A	15	134	0.19	10	168	0.13	N
TAF1D	29	133	0.19	29	167	0.17	N
CEP57L1	25	178	0.19	26	182	0.16	N
RASSF1	26	140	0.18	20	140	0.14	Y
PHF21A	12	147	0.18	13	190	0.16	N
POLDIP3	40	182	0.18	39	183	0.20	N
ESS2	34	192	0.17	37	205	0.15	N
C17orf80	32	182	0.17	22	183	0.12	N

C9orf78	15	85	0.17	10	85	0.10	N
MTUS2	16	178	0.16	15	182	0.15	N
GDE1	9	140	0.16	6	140	0.12	N
PMFBP1	25	192	0.16	22	205	0.17	N
KAZN	22	125	0.16	19	125	0.14	N
SPAG8	14	182	0.16	8	183	0.10	Y
CCDC144NL	15	125	0.15	15	125	0.09	N
SNRK	21	140	0.15	26	140	0.18	N
CCDC34	26	182	0.15	18	183	0.11	N
MAP11	11	182	0.14	9	183	0.14	N
TRAP1	8	139	0.14	5	143	0.09	Y
SART1	21	182	0.14	16	183	0.09	Y
CTTNBP2NL	20	125	0.14	22	125	0.13	N
KRT35	27	182	0.13	25	183	0.14	N
WTAP	8	178	0.13	10	182	0.10	N
TCEAL4	21	182	0.13	27	183	0.16	N
C19orf47	8	182	0.13	2	183	0.04	N
GATA2	8	192	0.13	8	205	0.03	Y
ZNF177	13	192	0.12	15	205	0.08	N
PSMD2	18	140	0.12	22	140	0.16	N
PML	19	140	0.12	23	140	0.16	Y
SOX2	27	192	0.12	30	205	0.12	Y
MAK	20	140	0.12	14	140	0.09	N
FRG1	22	178	0.12	30	182	0.18	N
ZSCAN16	20	192	0.12	23	205	0.11	N
TRIM29	20	192	0.12	20	205	0.10	N
PAK5	20	192	0.12	12	205	0.08	N
PELI1	7	178	0.11	1	182	0.02	Y
GTSE1	14	179	0.11	19	183	0.13	N
MAPK13	31	259	0.11	29	293	0.11	N
APEX2	27	259	0.11	13	293	0.04	N
VPS72	11	192	0.11	13	205	0.05	N
MYLK2	12	140	0.11	6	140	0.04	N
TAX1BP1	6	140	0.11	7	140	0.12	Y
LEF1	26	192	0.11	16	205	0.06	N
AHCY	6	192	0.10	2	205	0.04	N
ADNP2	6	125	0.10	0	125	0	N
RPS21	12	147	0.10	8	190	0.13	N
TCEAL2	7	192	0.10	6	205	0.03	N
RABGEF1	16	125	0.10	14	125	0.11	N
TFAM	10	125	0.10	10	125	0.10	N
GPANK1	14	259	0.10	19	293	0.05	N
CAPN3	12	178	0.10	10	182	0.10	N

DTNA	6	178	0.10	3	182	0.05	N
ZCCHC10	6	178	0.10	9	182	0.07	N
VENTX	15	125	0.10	13	125	0.09	N
NF2	14	179	0.10	11	183	0.06	N
YJEFN3	15	182	0.10	13	183	0.07	N
SECISBP2	17	182	0.10	15	183	0.09	N
ZBTB22	13	140	0.10	7	140	0.06	N
RNF138	11	125	0.10	11	125	0.06	N
JUN	7	272	0.10	5	315	0.05	Y
PAK1	9	140	0.10	5	140	0.10	Y

Table 4.5 Viruses Used for Sequence Similarity Analysis. Proteins encoded by respiratory and common viruses found in children of the US as annotated in UniProt. The reference proteomes from UniProt were included for each virus.

Organism	Strains	Number of proteins
Influenza A virus	(A/Alaska/105/2015(H3N2)), (A/Boston/151/2009(H1N1)), (A/Boston/DOA29/2011(H3N2)), (A/Boston/YGA_01042/2012(H3N2)), (A/California/47/2016(H3N2)), (A/California/VRDL67/2009(H1N1)), (A/California/VRDL364/2009(mixed)), (A/Hawaii/67/2014(H1N1)), (A/Hawaii/74/2015(H3N2)), (A/Houston/JMM_42/2012(H3N2)), (A/Kentucky/16/2015(H1N1)), (A/Louisiana/13/2014(H3N2)), (A/New York/169/2000(H3N2)), (A/New York/441/2001(H1N1)), (A/New York/1144/2008(H3N2)), (A/New York/3052/2009(mixed)), (A/New York/WC-LVD-14-057/2014(H1N1)), (A/Oregon/29/2015(H1N1)), (A/South Carolina/09/2009(H1N1)), (A/Tennessee/F2019A/2011(H3N2)), (A/Utah/06/2016(H1N1)), (A/Virginia/43/2016(H3N2)), (A/Puerto Rico/8/1934 H1N1), (A/South Carolina/1/1918 H1N1)), (A/WS/1933 H1N1)), (swl A/California/04/2009 H1N1)	218
Influenza B virus	(B/Florida/66/2015), (B/Florida/78/2015), (B/Texas/14/1991), (B/Utah/15/2015), (B/Florida/78/2015), (B/Utah/31/2016), (B/Lee/1940)	69
Influenza C virus	(C/Ann Arbor/1/1950)	8
Herpes simplex virus 1	(strain 17) (HHV-1)	73
Varicella-zoster virus	(strain Dumas) (HHV-3), (strain Oka vaccine) (HHV-3)	140
Epstein-Barr virus	(strain AG876) (HHV-4), (strain B95-8) (HHV-4), (strain GD1) (HHV-4)	187
Human cytomegalovirus	(strain AD169) (HHV-5), (strain Merlin) (HHV-5)	525
Human B lymphotropic virus	(strain Uganda-1102) (HHV-6 variant A), (strain Z29) (HHV-6 variant B)	205
Human T lymphotropic virus	(strain JI) (HHV-7)	102
Human rhinovirus A	(strain 41467-Gallo) (HRV-89)	1
Human rotavirus A, B, C, G9P[8]		45
Human adenovirus 21, 21a, 26, 52, 55, 56, A, B, C, D, D, E, F	(HAdV-18), (HAdV-7), (HAdV-1), (HAdV-2), (HAdV-5), (HAdV-17), (HAdV-4), (HAdV-41)	599
Human parainfluenza virus 1, 2, 4a	(strain Washington/1964)	24
Human respirovirus 1, 3		10
Respiratory syncytial virus A, B	(strain A2), (strain S-2) (HRSV-S2), (strain B1)	44
Norwalk virus	(strain GI/Human/United States/Norwalk/1968)	3
Human Enterovirus	(strain USA/BrCr/1970) (EV71), (EV68) (EV-68), Human parechovirus 2 (strain Williamson) (HPeV-2), Coxsackievirus A16, Coxsackievirus B2 (strain Ohio-1), Coxsackievirus B3 (strain Nancy), Coxsackievirus B4 (strain JVB/Benschoten/New York/51)	11
Human metapneumovirus	(strain CAN97-83) (HMPV)	9
Total		**2,273**

CHAPTER 5

5. EXPLORATION OF AUTOANTIBODY RESPONSES IN CANINE DIABETES USING PROTEIN ARRAYS

5.1 Introduction

Diabetes mellitus (DM) is a common endocrine disorder in dogs with an increasing prevalence over time (Catchpole et al., 2005; Nelson & Reusch, 2014). The disease is characterized by insulin deficiency, necessitates lifelong therapy with exogenous insulin, and in some ways is similar to type 1 diabetes (T1D) in humans (Nelson & Reusch, 2014). Though much remains unknown regarding the pathogenesis of canine diabetes, contributing factors may include one or more of exocrine pancreatic disease, concurrent endocrinopathies such as hyperadrenocorticism leading to insulin resistance and secondary β cell dysfunction, or autoimmune destruction of the β cells (Catchpole et al., 2005; Gilor et al., 2016).

In human T1D, most cases are thought to result from β-cell directed autoimmunity leading to β-cell loss (Katsarou et al., 2017). While autoantibodies are not themselves thought pathogenic in T1D (i.e., destructive for β-cells), they are commonly used as either diagnostic biomarkers of T1D or those at increased risk for the disease (Mathieu et al., 2018). In terms of specific antigenic targets, they most commonly include antibodies targeting insulin, insulinoma associated protein 2 (IA2), glutamic acid decarboxylase 65 (GAD65), and zinc transporter 8 (ZNT8) (Katsarou et al., 2017). Indeed, one or more of these autoantibodies are detected months to years before symptomatic disease ensues in nearly all subjects and more than 90% of patients are

positive for at least one autoantibody at diagnosis (Gan et al., 2012; Katsarou et al., 2017).

In dogs, studies evaluating for the presence of these autoantibodies have, unfortunately, reported inconsistent results, with 0-13% of dogs testing positive for GAD65 antibodies (Ahlgren et al., 2014; Davison et al., 2008; Kim et al., 2016), 0-10% of diabetic dogs testing positive for IA2 antibodies (Davison et al., 2008; Kim et al., 2016), 3-12.5% of untreated diabetic dogs testing positive for insulin antibodies (Davison et al., 2003; Holder et al., 2015), and 0% of diabetic dogs testing positive for ZnT8 antibodies (Kim et al., 2016). Additionally, a small study evaluated autoantibodies against canine proinsulin, in which 53% of newly diagnosed diabetic dogs were positive (Davison et al., 2011). Although islet cell cytoplasmic antibodies (ICA) have yet to be detected in naïve diabetic dogs (Ahlgren et al., 2014; Haines, 1986), approximately 50% of dogs in one study were noted as positive for serum anti-β-cell antibodies using purified islets utilizing a rat insulinoma cell line as an antigen (Hoenig & Dawe, 1992); a situation not unlike humans who are positive for ICA yet negative for other known autoantibodies (Andersson et al., 2014).

Beyond these commonly reported autoantibodies, multiple studies of humans with T1D have identified other novel autoantigens using a variety of techniques (Bian et al., 2017; Miersch et al., 2013; Müller et al., 2018; Walther et al., 2016). Two of these studies used an innovative Nucleic Acid Programmable Protein Array (NAPPA) platform to identify novel candidate autoantigens (Bian et al., 2017; Miersch et al., 2013). Unlike traditional protein microarrays that use purified proteins, NAPPA uses cDNA-encoding plasmids that are transcribed and translated *in situ* to create protein microarrays (Miersch

et al., 2013; Sibani & LaBaer, 2011). This method avoids some limitations of traditional purified protein arrays such as the time and cost of purifying multiple proteins as well as limited shelf stability (Miersch et al., 2013).

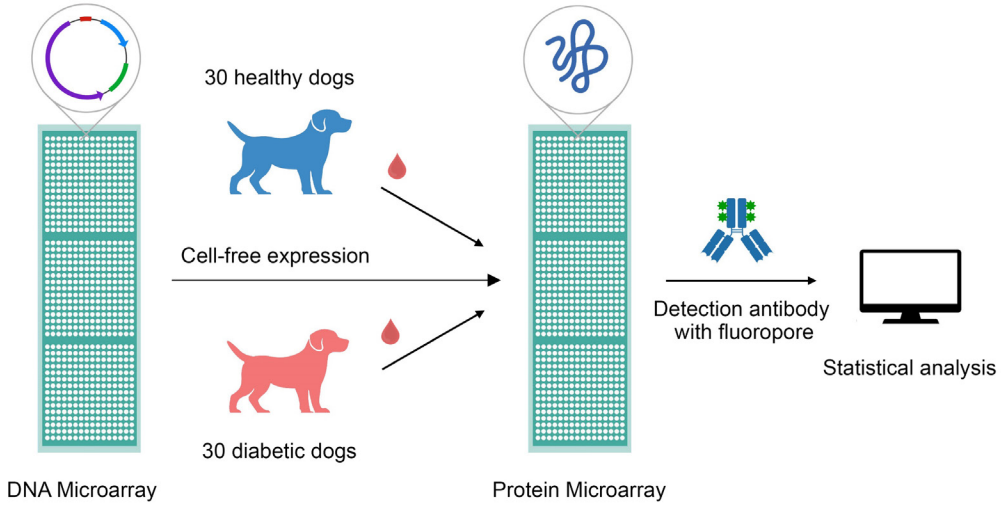


Fig. 5.1 Flowchart Explaining the Study Design. Initially, a DNA microarray is printed which is expressed using cell-free expression kit to make it a protein microarray. Sera from dogs were added followed by addition of detection antibodies.

One possible reason for the lack of consistent evidence for autoimmunity in canine diabetes is that the relevant autoantibodies, and thus autoantigens, have not been identified, and a large proteome-scale search for autoantibodies in diabetic dogs has yet to been published. Given the similarities in genes between humans and dogs (Lindblad-Toh et al., 2005), alongside the aforementioned quest to identify similarities between human T1D and canine diabetes, we elected to use a readily available human gene bank

and the established NAPPA assay. Specifically, the objective of the study is to compare autoantibody responses in diabetic and healthy control dogs using a NAPPA platform.

5.2 Materials and Methods

5.2.1 Dogs

Dogs were recruited from the client owned dog population from the University of Florida Small Animal Hospital. The study was approved by the Institutional Animal Care and Use Committee and the Veterinary Hospital Research Review Committee. Owners provided informed consent prior to study enrollment. Dogs were enrolled between May of 2016 and November of 2019. Diabetes was diagnosed by the attending clinician based on the presence of hyperglycemia, glucosuria, and compatible clinical signs of diabetes (i.e., polyuria, polydipsia, weight loss). Diabetic dogs were included if they were a minimum of 3 kg body weight, at least 1 year of age, and, if female, were spayed prior the diagnosis of diabetes. Diabetic dogs that had a history of pancreatitis or hyperadrenocorticism were excluded. Healthy control dogs were included if they were a minimum of 3 kg body weight, at least 1 year of age, if a female were spayed, and received no other medications other than routine flea/tick/heartworm preventatives. Control dogs were deemed healthy based on a history and physical exam and a lack of clinical evidence of concurrent disease. Blood samples were collected via routine venipuncture into red top vacutainer tubes. Serum was separated routinely within 30 minutes of collection and frozen immediately at -80 °C until analysis.

Table 5.1 Breed Distribution of Diabetic and Control Dogs.

Diabetic Group Breed	Number of Dogs	Control Group Breed	Number of Dogs
Mixed	8	Mixed	7
Labrador Retriever	4	Labrador Retriever	5
Dachshund	3	Dachshund	3
Miniature Pinscher	2	Shi Tzu	2
Miniature Schnauzer	2	Miniature Schnauzer	2
Cairn Terrier	1	Golden Retriever	1
Toy Poodle	1	Miniature Poodle	1
Cavalier King Charles Spaniel	1	Cavalier King Charles Spaniel	1
Australian Shepherd	1	Australian Shepherd	1
Miniature Australian Shepherd	1	Miniature Australian Shepherd	1
Yorkshire Terrier	1	Miniature Pinscher	1
Shih Tzu	1	Flat Coated Retriever	1
Pomeranian	1	Pomeranian	1
Pembroke Welsh Corgi	1	Pembroke Welsh Corgi	1
Beagle	1	Beagle	1
Pug	1	Pug	1

5.2.2 Gene Selection

Genes for NAPPA arrays were selected based on a literature search for human pancreatic genes, known genes important in human T1D screening, and candidate genes from an unpublished pilot study of diabetic dogs using NAPPA arrays. There were 1620 genes from literature search, 75 genes were known genes, and 5 genes from an unpublished pilot study for a total of 1700 genes.

5.2.3 NAPPA Arrays

NAPPA arrays were manufactured as previously described (Takulapalli et al., 2012), (Song et al., 2017). Briefly, bacterial clones having the genes of interest with a GST tag at the c-terminus, were obtained from the DNASU Plasmid Repository (DNASU.org). Plasmid DNA was purified using a mini-prep kit (Macherey-Nagel, #740499.50). DNA concentrations were then measured and normalized to 100 ng/ul for all 1700 genes. Silicon nanowell substrates were coated with (3-Aminopropyl) triethoxysilane (APTS) (Thermo Scientific, #80370) and then the plasmid DNA was printed using a piezo electric printer. At the time of usage, proteins were expressed from plasmid DNA using an *in-vitro* transcription and translation (IVTT) kit (Thermo Scientific, #88882). The printing quality of a batch was determined by expressing a random slide from the batch with the IVTT kit, followed by the detection of GST-tagged proteins with Mouse anti-GST antibody (Cell Signaling, #2624S) and Alexa 555 Goat anti-mouse IgG antibody (Invitrogen, #A-21422).

5.2.4 Serological Profiling on NAPPA

Proteins were expressed using the IVTT kit and displayed on NAPPA. Dog serum samples diluted at 1:200 in PBST with 5% milk were added to the microarrays, followed by overnight rocking at 4°C. After washing with PBST, dog autoantibodies were detected by 1:3000 diluted biotinylated anti-dog IgG (KPL, #16-19-06) followed by 1:2000 diluted Alexa 555 Streptavidin (Invitrogen, #S21381). Scanned microarray images were analyzed by the ArrayPro image analysis software. Antibody reactivity of each spot was normalized by division with the median spot intensity of each corresponding microarray.

This normalized intensity value is denoted as Median Normalized Intensity (MNI). The study design is summarized in **Fig. 5.1**.

5.2.5 Data and Statistical Analysis

Data recorded included age, breed, sex (and neutering status), body weight, duration of diabetes (if applicable), and concurrent medical conditions. Continuous data were tested for normality using the D'Agostino and Pearson test, and parametric or non-parametric tests used as indicated. Age and body weight were compared between diabetic and control groups using an unpaired T-test and Mann-Whitney U test, respectively, with sex distribution compared with a Chi Squared test.

We used the MNI values to analyze antibodies quantified on NAPPA. Seropositive proteins were defined as proteins whose antibodies having MNI values greater than an empirical cutoff of 1.5 on NAPPA. We compared the number of seropositive proteins between the diabetic and control groups using Wilcoxon rank-sum test.

Antibody sensitivity in the diabetic group at 90% specificity was calculated as follows. For each antibody, we calculated the threshold as the maximum between the 90th percentile of MNI values in the control samples and the empirical seropositivity cutoff of 1.5. We then computed the corresponding sensitivity as the percentage of diabetic samples higher than the threshold. Antibodies with sensitivity greater than 10% were selected, and a subset of them were further selected as a panel of diabetic biomarkers using lasso logistic regression. Its discriminatory performance between diabetic and controls was evaluated by sensitivity at 90% specificity, the area under the

receiver operating characteristics (ROC) curve (AUC), and their 95% confidence intervals (CI). GraphPad Prism (v8.0, San Diego, CA) and R (v4.1.0, Vienna, Austria) were used for the analyses.

5.3 Results

Thirty diabetic dogs and thirty healthy control dogs were included in this study. The breed distribution for each group is shown in **Table 5.1**. The diabetic dog group had a mean age of 8.5 +/- 3 years and were older than the control dog group, with a mean age of 6.7 +/- 2.8 years ($P = 0.015$). Body weight was not different between the groups, with a median body weight of 10.2 kg (range, 5.5-33.1 kg) in the diabetic group and 10.6 kg (range, 4.4-45.1 kg) in the control group ($P = 0.94$). The diabetic group consisted of 18 males neutered and 12 female spayed dogs, which was not significantly different than the control group, which included 16 males neutered and 14 female spayed dogs ($P = 0.60$). Diabetic dogs had a median (range) duration of disease of 3 months (0-36 months).

Table 5.2 Dog Diabetes Biomarkers with Sensitivity >10% at a Specificity of 90%.

Gene	Protein name	UniProt ID	Sensitivity (%)
TACSTD2	Tumor-associated calcium signal transducer 2	P09758	20
SCGB1C1	Secretoglobin family 1C member 1	Q8TD33	17
SUMO2	Small ubiquitin-related modifier 2	P61956	17
KANK2	KN motif and ankyrin repeat domain-containing protein 2	Q63ZY3	13
GLI1	Zinc finger protein GLI1	P08151	13
CPA4	Carboxypeptidase A4	Q9UI42	13

We evaluated the antibody profiles of these dogs against 1,700 human proteins relevant to diabetes and the pancreas. The number of autoantibodies with MNI values greater than 1.5 were 8.83 +/- 9.37 and 9.93 +/- 12.51 for the diabetic and control groups, respectively, a finding that was not significantly different ($P = 0.74$). At a specificity of 90%, six autoantibodies had sensitivity greater than 10%: anti-TACSTD2, Anti-SCGB1C1, anti-SUMO2, anti-KANK2, anti-GLI1, and anti-CPA4 (**Table 5.2** and **Fig. 5.2**). Using lasso logistic regression, a subset of the aforementioned autoantibodies (anti-KANK2, anti-GLI1, anti-SUMO2) had a sensitivity of 37% (95% CI: 0.17-0.67%) at 90% specificity and an AUC of 0.66 (95% CI: 0.52-0.80) (**Fig. 5.3**).

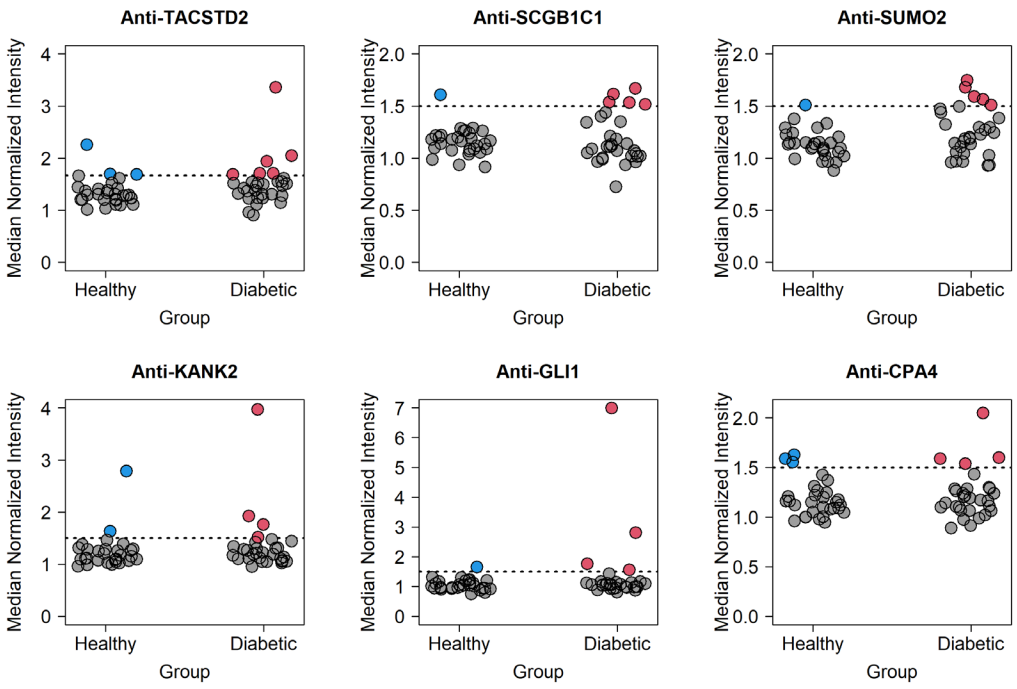


Fig. 5.2 Reactivity of Antibodies with Sensitivities >10% in Diabetic & Healthy Dogs. Each dot represents an individual dog and the reactivity to the respective antibody. The horizontal dashed line represents the maximum between 90% percentile of the control samples and 1.5, and sensitivity is the proportion of red dots in the diabetic samples.

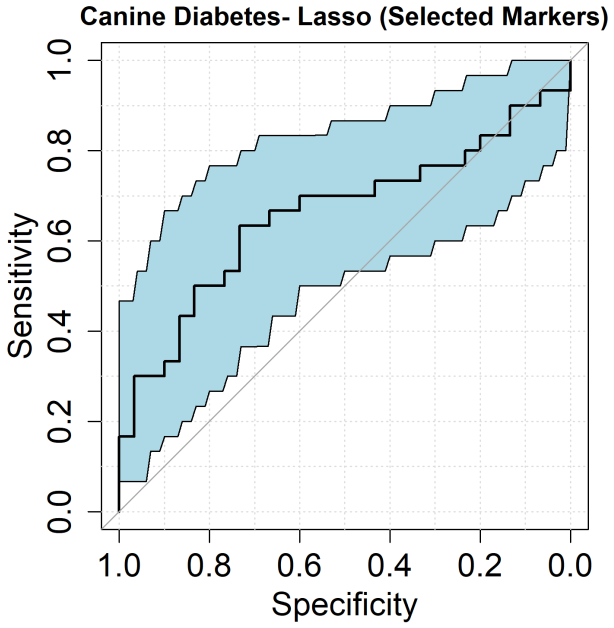


Fig. 5.3 ROC Curve. The antibody panel (Anti-KANK2, Anti-GLI1, Anti-SUMO2) was obtained from lasso logistic regression model with a sensitivity of 37% (95% CI: 0.17-0.67%) at 90% specificity and an AUC of 0.66 (95% CI: 0.52-0.80). The blue area represents the 95% CIs of sensitivities for each value of specificities, and the 45-degree straight line represents a useless biomarker having a sensitivity of 10% at 90% specificity and an AUC value of 0.5.

5.4 Discussion

This is the first study, to the authors' knowledge, to use a large proteomics-based approach to search for autoantibodies in canine diabetes. As noted, canine diabetes shares some features with human T1D, however the detection of the key autoantibodies found in the human disease has been inconsistent in dogs. Therefore, our goal was to broaden the scope of potential antigens screened in an attempt to find novel autoantibodies that may

be important for the disease in dogs. Using the NAPPA arrays, displaying 1700 human proteins, we identified several candidate autoantibody/autoantigen combinations, however these have low sensitivities for distinguishing between diabetic and control groups.

Previous NAPPA array results in humans have identified novel minor type 1 diabetes associated antigens such as MTIF3, PPIL2, and MLHI (Bian et al., 2017) in 7-24% of type 1 diabetic patients, along with small numbers of non-diabetic control patients, using a luciferase immunoprecipitation system (Müller et al., 2018). Tetraspanin 7 autoantibodies are present in 35% of auto-antibody positive type 1 diabetes patients, but do not provide additional diagnostic value over the other established autoantibodies (Walther et al., 2016). It has been suggested that these minor autoantigens may not be important for disease diagnosis but may shed light into pathogenesis (Müller et al., 2018). Of the proteins with a sensitivity of >10%, all of the genes have homologs or orthologs in dogs (*Gene [Internet]*, 2004), and none have been linked to diabetes in other species to the authors' knowledge. However, several genes have been associated with other pancreatic disease, or fat and glucose metabolism. Both the CPA4 and TACSTD2 proteins are overexpressed in pancreatic carcinoma in humans and are associated with decreased survival (Fong et al., 2008; Shao et al., 2020). Increased TACSTD2 gene expression is associated with increased fat mass in children (Groom et al., 2012). In addition, SUMO2 expression has been reported as increased in rat mesangial cells exposed to high glucose conditions (Zhou et al., 2014), while CPA4 is a negative modulator of adipogenesis and insulin sensitivity (He et al., 2016). The mechanisms

leading to development of these autoantibodies require further study given the associations with the pancreas and metabolism in other species.

There are several potential reasons that we did not observe autoantibody candidates with higher sensitivities, including that our efforts did not screen all possible proteins. Additionally, although gene predictions have estimated that most of the almost 20,000 canine genes have human homologues (Lindblad-Toh et al., 2005), the potential exists that some of the relevant canine proteins do not have human homologue or that the antigen binding sites on the canine antibodies do not recognize the epitopes on the human proteins.

Additionally, there is also growing evidence that canine diabetes, like human T1D, is a heterogenous disease. In dogs, strong breed predispositions suggest a genetic component contributes to disease risk (Catchpole et al., 2013). Denyer *et al* evaluated dog leukocyte antigen (DLA) (the canine equivalent to human leukocyte antigen) haplotypes in diabetic and control dogs (at least 20 in each group for each breed) in 12 different dog breeds (Denyer et al., 2020). They identified five dog breeds with DLA haplotypes associated with risk or protection, but other dog breeds, including 3 of the breeds at highest risk for diabetes, had no DLA associations with DM. This suggests that the disease may be heterogenous among breeds, especially with respect to immune related genes contributing to pathogenesis. Our study included small numbers of dogs of multiple dog breeds. Focusing screening on diabetic and control dogs in those dog breeds with DLA haplotype associations with diabetes risk may identify novel autoantibodies that are missed by testing a wide variety of breeds such as in the present study. Another limitation of the study is that dogs were evaluated at a single point in time, and it is

possible that some diabetic dogs may have had autoantibodies earlier in the disease process or that some control dogs may have gone on to develop diabetes in the future. Other studies performed with different techniques to evaluate for autoantibodies in canine serum have also reported some control dog reactivity when human assays were used (Davison et al., 2008; O'Kell et al., 2020). The development of NAPPA arrays using canine specific genes/proteins is necessary to address this limitation and will afford future studies allowing for assessment, over time.

In conclusion, we identified six candidate novel autoantibodies in canine diabetes, however sensitivity to distinguish from non-diabetic control dogs was somewhat limited. This study does not provide strong support for the role of autoimmunity in disease pathogenesis in dogs using this set of genes and proteins; however, the small numbers of dogs of a variety of breeds are an important limitation. Future studies should focus on larger numbers of breeds considered high risk for diabetes using canine specific genes and proteins.

REFERENCES

Abegunde, A. T., Muhammad, B. H., Bhatti, O., & Ali, T. (2016). Environmental risk factors for inflammatory bowel diseases: Evidence based literature review. *World journal of gastroenterology, 22*(27), 6296-6317. https://doi.org/10.3748/wjg.v22.i27.6296

Abraham, C., & Cho, J. H. (2009). Inflammatory bowel disease. *The New England journal of medicine, 361*(21), 2066-2078. https://doi.org/10.1056/NEJMra0804647

Ahlgren, K. M., Fall, T., Landegren, N., Grimelius, L., von Euler, H., Sundberg, K., Lindblad-Toh, K., Lobell, A., Hedhammar, Å., Andersson, G., Hansson-Hamlin, H., Lernmark, Å., & Kämpe, O. (2014). Lack of evidence for a role of islet autoimmunity in the aetiology of canine diabetes mellitus. *PLoS One, 9*(8), e105473. https://doi.org/10.1371/journal.pone.0105473

Ahmad, T., Armuzzi, A., Bunce, M., Mulcahy‐Hawes, K., Marshall, S. E., Orchard, T. R., Crawshaw, J., Large, O., De Silva, A., Cook, J. T., Barnardo, M., Cullen, S., Welsh, K. I., & Jewell, D. P. (2002, 2002/04/01/). The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology, 122*(4), 854-866. https://doi.org/https://doi.org/10.1053/gast.2002.32413

Ahuja, A., Anderson, S. M., Khalil, A., & Shlomchik, M. J. (2008). Maintenance of the plasma cell pool is independent of memory B cells. *Proceedings of the National Academy of Sciences, 105*(12), 4802. https://doi.org/10.1073/pnas.0800555105

Alatab, S., Sepanlou, S. G., Ikuta, K., Vahedi, H., Bisignano, C., Safiri, S., Sadeghi, A., Nixon, M. R., Abdoli, A., Abolhassani, H., Alipour, V., Almadi, M. A. H., Almasi-Hashiani, A., Anushiravani, A., Arabloo, J., Atique, S., Awasthi, A., Badawi, A., Baig, A. A. A., Bhala, N., Bijani, A., Biondi, A., Borzì, A. M., Burke, K. E., Carvalho, F., Daryani, A., Dubey, M., Eftekhari, A., Fernandes, E., Fernandes, J. C., Fischer, F., Haj-Mirzaian, A., Haj-Mirzaian, A., Hasanzadeh, A., Hashemian, M., Hay, S. I., Hoang, C. L., Househ, M., Ilesanmi, O. S., Jafari Balalami, N., James, S. L., Kengne, A. P., Malekzadeh, M. M., Merat, S., Meretoja, T. J., Mestrovic, T., Mirrakhimov, E. M., Mirzaei, H., Mohammad, K. A., Mokdad, A. H., Monasta, L., Negoi, I., Nguyen, T. H., Nguyen, C. T., Pourshams, A., Poustchi, H., Rabiee, M., Rabiee, N., Ramezanzadeh, K., Rawaf, D. L., Rawaf, S., Rezaei, N., Robinson, S. R., Ronfani, L., Saxena, S., Sepehrimanesh, M., Shaikh, M. A., Sharafi, Z., Sharif, M., Siabani, S., Sima, A. R., Singh, J. A., Soheili, A., Sotoudehmanesh, R., Suleria, H. A. R., Tesfay, B. E.,

Tran, B., Tsoi, D., Vacante, M., Wondmieneh, A. B., Zarghi, A., Zhang, Z.-J., Dirac, M., Malekzadeh, R., & Naghavi, M. (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology, 5*(1), 17-30. https://doi.org/10.1016/S2468-1253(19)30333-4

Alexander, K. L., Targan, S. R., & Elson Iii, C. O. (2014, 2014/07/01). Microbiota activation and regulation of innate and adaptive immunity [https://doi.org/10.1111/imr.12180]. *Immunological Reviews, 260*(1), 206-220. https://doi.org/https://doi.org/10.1111/imr.12180

Alhajj, M., & Farhana, A. (2021). *Enzyme Linked Immunosorbent Assay.* StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK555922/?report=classic

Andersson, C., Kolmodin, M., Ivarsson, S. A., Carlsson, A., Forsander, G., Lindblad, B., Ludvigsson, J., Kockum, I., Marcus, C., Samuelsson, U., Ortqvist, E., Lernmark, A., Elding Larsson, H., Törn, C., & Group, B. D. D. S. (2014, Aug). Islet cell antibodies (ICA) identify autoimmunity in children with new onset diabetes mellitus negative for other islet cell antibodies. *Pediatr Diabetes, 15*(5), 336-344. https://doi.org/10.1111/pedi.12093

Andrilenas, K. K., Penvose, A., & Siggers, T. (2015). Using protein-binding microarrays to study transcription factor specificity: homologs, isoforms and complexes. *Briefings in Functional Genomics, 14*(1), 17-29. https://doi.org/10.1093/bfgp/elu046

Angenendt, P., Glökler, J., Murphy, D., Lehrach, H., & Cahill, D. J. (2002). Toward optimized antibody microarrays: a comparison of current microarray support materials. *Analytical biochemistry, 309*(2), 253-260.

Aronson, J. K., & Ferner, R. E. (2017). Biomarkers−A General Review. *Current Protocols in Pharmacology, 76*(1), 9.23.21-29.23.17. https://doi.org/https://doi.org/10.1002/cpph.19

Aziz, F., Smith, M., & Blackburn, J. M. (2018). Autoantibody-based diagnostic biomarkers: technological approaches to discovery and validation. In *Autoantibodies and Cytokines*. IntechOpen.

Baumgarth, N., Waffarn, E. E., & Nguyen, T. T. T. (2015). Natural and induced B-1 cell immunity to infections raises questions of nature versus nurture. *Annals of the*

New York Academy of Sciences, 1362, 188-199. https://doi.org/10.1111/nyas.12804

Bedi, A., Zehnbauer, B. A., Barber, J. P., Sharkis, S. J., & Jones, R. J. (1994). Inhibition of apoptosis by BCR-ABL in chronic myeloid leukemia.

Bern, U. o. (2020). *Gut microbes shape our antibodies before we are infected by pathogens.* ScienceDaily. www.sciencedaily.com/releases/2020/08/200805124038.htm

Berzofsky, J. A. (1985). Intrinsic and extrinsic factors in protein antigenic structure. *Science, 229*(4717), 932. https://doi.org/10.1126/science.2410982

Bian, X., Wallstrom, G., Davis, A., Wang, J., Park, J., Throop, A., Steel, J., Yu, X., Wasserfall, C., Schatz, D., Atkinson, M., Qiu, J., & LaBaer, J. (2016). Immunoproteomic Profiling of Antiviral Antibodies in New-Onset Type 1 Diabetes Using Protein Arrays. *Diabetes, 65*(1), 285. https://doi.org/10.2337/db15-0179

Bian, X., Wasserfall, C., Wallstrom, G., Wang, J., Wang, H., Barker, K., Schatz, D., Atkinson, M., Qiu, J., & LaBaer, J. (2017, Jan 6). Tracking the Antibody Immunome in Type 1 Diabetes Using Protein Arrays. *J Proteome Res, 16*(1), 195-203. https://doi.org/10.1021/acs.jproteome.6b00354

Binder, C. J. (2012). Naturally occurring IgM antibodies to oxidation-specific epitopes. In *Naturally Occurring Antibodies (NAbs)* (pp. 2-13). Springer.

Bloom, S. M., Bijanki, V. N., Nava, G. M., Sun, L., Malvin, N. P., Donermeyer, D. L., Dunne, W. M., Jr., Allen, P. M., & Stappenbeck, T. S. (2011). Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell host & microbe, 9*(5), 390-403. https://doi.org/10.1016/j.chom.2011.04.009

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010, 2010/04/01). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111. https://doi.org/10.1002/jrsm.12

Breese, E., Braegger, C., Corrigan, C., Walker-Smith, J., & MacDonald, T. (1993). Interleukin-2-and interferon-gamma-secreting T cells in normal and diseased human intestinal mucosa. *Immunology, 78*(1), 127.

Burckart, G. J., & Green, D. J. (2018). Chapter 7 - Regulatory Aspects of Pediatric Biomarkers for Assessing Medication Response. In S. J. Szefler, F. Holguin, & M. E. Wechsler (Eds.), *Personalizing Asthma Management for the Clinician* (pp. 69-86). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-323-48552-4.00007-X

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

Catchpole, B., Adams, J. P., Holder, A. L., Short, A. D., Ollier, W. E., & Kennedy, L. J. (2013, Feb). Genetics of canine diabetes mellitus: are the diabetes susceptibility genes identified in humans involved in breed susceptibility to diabetes mellitus in dogs? *Vet J, 195*(2), 139-147. https://doi.org/10.1016/j.tvjl.2012.11.013

Catchpole, B., Kennedy, L. J., Davison, L. J., & Ollier, W. E. R. (2008, 2008/01/01). Canine diabetes mellitus: from phenotype to genotype [https://doi.org/10.1111/j.1748-5827.2007.00398.x]. *Journal of Small Animal Practice, 49*(1), 4-10. https://doi.org/https://doi.org/10.1111/j.1748-5827.2007.00398.x

Catchpole, B., Ristic, J. M., Fleeman, L. M., & Davison, L. J. (2005, Oct). Canine diabetes mellitus: can old dogs teach us new tricks? *Diabetologia, 48*(10), 1948-1956. https://doi.org/10.1007/s00125-005-1921-1

Cerutti, A., & Rescigno, M. (2008). The biology of intestinal immunoglobulin A responses. *Immunity, 28*(6), 740-750.

Chen, C.-S., & Zhu, H. (2006, 2006/04/01). Protein Microarrays. *BioTechniques, 40*(4), 423-429. https://doi.org/10.2144/06404TE01

Chiu, M. L., Goulet, D. R., Teplyakov, A., & Gilliland, G. L. (2019). Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies (Basel, Switzerland), 8*(4), 55. https://doi.org/10.3390/antib8040055

Chowell, D., Krishna, S., Becker, P. D., Cocita, C., Shu, J., Tan, X., Greenberg, P. D., Klavinskis, L. S., Blattman, J. N., & Anderson, K. S. (2015). TCR contact residue hydrophobicity is a hallmark of immunogenic CD8⁺ T

cell epitopes. *Proceedings of the National Academy of Sciences, 112*(14), E1754. https://doi.org/10.1073/pnas.1500973112

Choy, M. C., Visvanathan, K., & De Cruz, P. (2017). An Overview of the Innate and Adaptive Immune System in Inflammatory Bowel Disease. *Inflammatory Bowel Diseases, 23*(1), 2-13. https://doi.org/10.1097/MIB.0000000000000955

Conway, K. L., Kuballa, P., Song, J.-H., Patel, K. K., Castoreno, A. B., Yilmaz, O. H., Jijon, H. B., Zhang, M., Aldrich, L. N., Villablanca, E. J., Peloquin, J. M., Goel, G., Lee, I.-A., Mizoguchi, E., Shi, H. N., Bhan, A. K., Shaw, S. Y., Schreiber, S. L., Virgin, H. W., Shamji, A. F., Stappenbeck, T. S., Reinecker, H.-C., & Xavier, R. J. (2013). Atg16l1 is required for autophagy in intestinal epithelial cells and protection of mice from Salmonella infection. *Gastroenterology, 145*(6), 1347-1357. https://doi.org/10.1053/j.gastro.2013.08.035

Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., Ferguson, D. J. P., Campbell, B. J., Jewell, D., & Simmons, A. (2010, 2010/01/01). NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Medicine, 16*(1), 90-97. https://doi.org/10.1038/nm.2069

Cosnes, J. (2004, 2004/06/01/). Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. *Best Practice & Research Clinical Gastroenterology, 18*(3), 481-496. https://doi.org/https://doi.org/10.1016/j.bpg.2003.12.003

Coutinho, A., Kazatchkine, M. D., & Avrameas, S. (1995, 1995/12/01/). Natural autoantibodies. *Current Opinion in Immunology, 7*(6), 812-818. https://doi.org/https://doi.org/10.1016/0952-7915(95)80053-0

Cramér, H. (2016). *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press. https://doi.org/doi:10.1515/9781400883868

Crotty, S., Felgner, P., Davies, H., Glidewell, J., Villarreal, L., & Ahmed, R. (2003). Cutting Edge: Long-Term B Cell Memory in Humans after Smallpox Vaccination. *The Journal of Immunology, 171*(10), 4969. https://doi.org/10.4049/jimmunol.171.10.4969

Cruse, J. M., & Lewis, R. E. (2010). *Atlas of Immunology*. CRC Press. https://books.google.com/books?id=kNI5Lk2z37sC

Cruse, J. M., Lewis, R. E., & Wang, H. (2004). 7 - ANTIGEN PRESENTATION. In *Immunology Guidebook* (pp. 267-276). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012198382-6/50031-5

Cusick, M. F., Libbey, J. E., & Fujinami, R. S. (2012). Molecular mimicry as a mechanism of autoimmune disease. *Clinical reviews in allergy & immunology, 42*(1), 102-111. https://doi.org/10.1007/s12016-011-8294-7

Dam, A. N., Berg, A. M., & Farraye, F. A. (2013). Environmental influences on the onset and clinical course of Crohn's disease-part 1: an overview of external risk factors. *Gastroenterology & hepatology, 9*(11), 711-717. https://pubmed.ncbi.nlm.nih.gov/24764788

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3995193/

Darfeuille-Michaud, A., Boudeau, J., Bulois, P., Neut, C., Glasser, A.-L., Barnich, N., Bringer, M.-A., Swidsinski, A., Beaugerie, L., & Colombel, J.-F. (2004). High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn’ s disease. *Gastroenterology, 127*(2), 412-421.

Davison, L. J., Herrtage, M. E., & Catchpole, B. (2011, Aug). Autoantibodies to recombinant canine proinsulin in canine diabetic patients. *Res Vet Sci, 91*(1), 58-63. https://doi.org/10.1016/j.rvsc.2010.08.007

Davison, L. J., Ristic, J. M., Herrtage, M. E., Ramsey, I. K., & Catchpole, B. (2003, Jan). Anti-insulin antibodies in dogs with naturally occurring diabetes mellitus. *Vet Immunol Immunopathol, 91*(1), 53-60. http://www.ncbi.nlm.nih.gov/pubmed/12507850

Davison, L. J., Weenink, S. M., Christie, M. R., Herrtage, M. E., & Catchpole, B. (2008, Nov). Autoantibodies to GAD65 and IA-2 in canine diabetes mellitus. *Vet Immunol Immunopathol, 126*(1-2), 83-90. https://doi.org/10.1016/j.vetimm.2008.06.016

De Paschale, M., & Clerici, P. (2012). Serological diagnosis of Epstein-Barr virus infection: Problems and solutions. *World journal of virology, 1*(1), 31-43. https://doi.org/10.5501/wjv.v1.i1.31

de Souza, H. S. P., & Fiocchi, C. (2016, 2016/01/01). Immunopathogenesis of IBD: current state of the art. *Nature Reviews Gastroenterology & Hepatology, 13*(1), 13-27. https://doi.org/10.1038/nrgastro.2015.186

DeMarshall, C., Sarkar, A., Nagele, E. P., Goldwaser, E., Godsey, G., Acharya, N. K., & Nagele, R. G. (2015). Chapter One - Utility of Autoantibodies as Biomarkers for Diagnosis and Staging of Neurodegenerative Diseases. In M. J. Hurley (Ed.), *International Review of Neurobiology* (Vol. 122, pp. 1-51). Academic Press. https://doi.org/https://doi.org/10.1016/bs.irn.2015.05.005

Denyer, A. L., Massey, J. P., Davison, L. J., Ollier, W. E. R., Catchpole, B., & Kennedy, L. J. (2020, Oct 31). Dog leucocyte antigen (DLA) class II haplotypes and risk of canine diabetes mellitus in specific dog breeds. *Canine Med Genet, 7*(1), 15. https://doi.org/10.1186/s40575-020-00093-9

Dervan, E. W., Chen, H., Ho, S. L., Brummel, N., Schmid, J., Toomey, D., Haralambova, M., Gould, E., Wallace, D. M., Prehn, J. H. M., O'Brien, C. J., & Murphy, D. (2010). Protein Macroarray Profiling of Serum Autoantibodies in Pseudoexfoliation Glaucoma. *Investigative Ophthalmology & Visual Science, 51*(6), 2968-2975. https://doi.org/10.1167/iovs.09-4898

Dethlefsen, L., Huse, S., Sogin, M. L., & Relman, D. A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS biology, 6*(11), e280.

Di Sabatino, A., Biancheri, P., Rovedatti, L., MacDonald, T. T., & Corazza, G. R. (2012). New Pathogenic Paradigms in Inflammatory Bowel Disease. *Inflammatory Bowel Diseases, 18*(2), 368-371. https://doi.org/10.1002/ibd.21735

Díez, P., González-González, M., Lourido, L., Dégano, R. M., Ibarrola, N., Casado-Vela, J., LaBaer, J., & Fuentes, M. (2015). NAPPA as a Real New Method for Protein Microarray Generation. *Microarrays, 4*(2). https://doi.org/10.3390/microarrays4020214

Ding, C., & Peng, H. (2005, 2005/04/01). MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA. *Journal of Bioinformatics and Computational Biology, 03*(02), 185-205. https://doi.org/10.1142/S0219720005001004

Djoba Siawaya, J. F., Roberts, T., Babb, C., Black, G., Golakai, H. J., Stanley, K., Bapela, N. B., Hoal, E., Parida, S., & Van Helden, P. (2008). An evaluation of

commercial fluorescent bead-based luminex cytokine assays. *PLOS ONE, 3*(7), e2535.

Dolan, K. T., & Chang, E. B. (2017, 2017/01/01). Diet, gut microbes, and the pathogenesis of inflammatory bowel diseases [https://doi.org/10.1002/mnfr.201600129]. *Molecular Nutrition & Food Research, 61*(1), 1600129. https://doi.org/https://doi.org/10.1002/mnfr.201600129

Donnenberg, M. S. (2015). 220 - Enterobacteriaceae. In J. E. Bennett, R. Dolin, & M. J. Blaser (Eds.), *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)* (pp. 2503-2517.e2505). W.B. Saunders. https://doi.org/https://doi.org/10.1016/B978-1-4557-4801-3.00220-4

Doxey, A. C., & McConkey, B. J. (2013, 2013/08/15). Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence, 4*(6), 453-466. https://doi.org/10.4161/viru.25180

Dubinsky, M. C., Lin, Y.-C., Dutridge, D., Picornell, Y., Landers, C. J., Farrior, S., Wrobel, I., Quiros, A., Vasiliauskas, E. A., Grill, B., Israel, D., Bahar, R., Christie, D., Wahbeh, G., Silber, G., Dallazadeh, S., Shah, P., Thomas, D., Kelts, D., Hershberg, R. M., Elson, C. O., Targan, S. R., Taylor, K. D., Rotter, J. I., Yang, H., & Western Regional Pediatric, I. B. D. R. A. (2006). Serum immune responses predict rapid disease progression among children with Crohn's disease: immune responses predict disease progression. *The American journal of gastroenterology, 101*(2), 360-367. https://doi.org/10.1111/j.1572-0241.2006.00456.x

Dunn, C., O'Dowd, A., & Randall, R. E. (1999, 1999/04/22/). Fine mapping of the binding sites of monoclonal antibodies raised against the Pk tag. *Journal of Immunological Methods, 224*(1), 141-150. https://doi.org/https://doi.org/10.1016/S0022-1759(99)00017-4

Eisenstein, M. (2020). The hunt for a healthy microbiome. *Nature*(1476-4687 (Electronic)). https://doi.org/10.1038/d41586-020-00193-3

Elkadri, A. A., Stempak, J. M., Walters, T. D., Lal, S., Griffiths, A. M., Steinhart, A. H., & Silverberg, M. S. (2013). Serum Antibodies Associated with Complex Inflammatory Bowel Disease. *Inflammatory Bowel Diseases, 19*(7), 1499-1505. https://doi.org/10.1097/MIB.0b013e318281f2a1

Elkon, K., & Casali, P. (2008a). Nature and functions of autoantibodies. *Nature clinical practice. Rheumatology, 4*(9), 491-498. https://doi.org/10.1038/ncprheum0895

Elkon, K., & Casali, P. (2008b, 2008/09/01). Nature and functions of autoantibodies. *Nature Clinical Practice Rheumatology, 4*(9), 491-498. https://doi.org/10.1038/ncprheum0895

Evans, C. F., Horwitz, M. S., Hobbs, M. V., & Oldstone, M. B. A. (1996). Viral Infection of Transgenic Mice Expressing a Viral Protein in Oligodendrocytes Leads to Chronic Central Nervous System Autoimmune Disease. *Journal of Experimental Medicine, 184*(6), 2371-2384. https://doi.org/10.1084/jem.184.6.2371

Fairweather, D., Frisancho-Kiss, S., & Rose, N. R. (2008). Sex differences in autoimmune disease from a pathological perspective. *The American journal of pathology, 173*(3), 600-609. https://doi.org/10.2353/ajpath.2008.071008

Fakhoury, M., Negrulj, R., Mooranian, A., & Al-Salami, H. (2014). Inflammatory bowel disease: clinical aspects and treatments. *Journal of inflammation research, 7*, 113-120. https://doi.org/10.2147/JIR.S65979

Fong, D., Moser, P., Krammel, C., Gostner, J. M., Margreiter, R., Mitterer, M., Gastl, G., & Spizzo, G. (2008, Oct 21). High expression of TROP2 correlates with poor prognosis in pancreatic cancer. *Br J Cancer, 99*(8), 1290-1295. https://doi.org/10.1038/sj.bjc.6604677

Frank, D. N., St. Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences, 104*(34), 13780. https://doi.org/10.1073/pnas.0706625104

Frostegård, J., Hellström, C., Nilsson, P., Frostegård, A. G., & Ajeganova, S. (2018, 2018/09/01). Autoantibody profiling reveals four protein candidate autoantigens associated with systemic lupus erythematosus. *Lupus, 27*(10), 1670-1678. https://doi.org/10.1177/0961203318788153

Fujimoto, T., Imaeda, H., Takahashi, K., Kasumi, E., Bamba, S., Fujiyama, Y., & Andoh, A. (2013). Decreased abundance of F aecalibacterium prausnitzii in the gut microbiota of C rohn's disease. *Journal of gastroenterology and hepatology, 28*(4), 613-619.

Fujinami, R. S., Oldstone, M. B., Wroblewska, Z., Frankel, M. E., & Koprowski, H. (1983). Molecular mimicry in virus infection: crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. *Proceedings of the National Academy of Sciences, 80*(8), 2346. https://doi.org/10.1073/pnas.80.8.2346

Furrie, E., Macfarlane, S., Cummings, J. H., & Macfarlane, G. T. (2004). Systemic antibodies towards mucosal bacteria in ulcerative colitis and Crohn's disease differentially activate the innate immune response. *Gut, 53*(1), 91-98. https://doi.org/10.1136/gut.53.1.91

Gajendran, M., Loganathan, P., Catinella, A. P., & Hashash, J. G. (2018). A comprehensive review and update on Crohn's disease. *Disease-a-month, 64*(2), 20-57.

Galperin., E. V. K. a. M. Y. (2003). *Sequence - Evolution - Function: Computational Approaches in Comparative Genomics.* Kluwer Academic. https://www.ncbi.nlm.nih.gov/books/NBK20261/

Gamble, D., Taylor, K., & Cumming, H. (1973). Coxsackie viruses and diabetes mellitus. *Br Med J, 4*(5887), 260-262.

Gan, M. J., Albanese-O'Neill, A., & Haller, M. J. (2012, 2012 Nov-Dec). Type 1 diabetes: current concepts in epidemiology, pathophysiology, clinical care, and research. *Curr Probl Pediatr Adolesc Health Care, 42*(10), 269-291. https://doi.org/10.1016/j.cppeds.2012.07.002

Gaudino, S. J., & Kumar, P. (2019, 2019-March-06). Cross-Talk Between Antigen Presenting Cells and T Cells Impacts Intestinal Homeostasis, Bacterial Infections, and Tumorigenesis [Review]. *Frontiers in Immunology, 10*(360). https://doi.org/10.3389/fimmu.2019.00360

Gene [Internet]. (2004). National Library of

Medicine (US),

National Library of

Medicine (US), National Center for Biotechnology Information. Retrieved July 08 from

Geremia, A., Biancheri, P., Allan, P., Corazza, G. R., & Di Sabatino, A. (2014). Innate and adaptive immunity in inflammatory bowel disease. *Autoimmunity Reviews, 13*(1), 3-10.

Gibson, D. S., Qiu, J., Mendoza, E. A., Barker, K., Rooney, M. E., & LaBaer, J. (2012, 2012/04/17). Circulating and synovial antibody profiling of juvenile arthritis patients by nucleic acid programmable protein arrays. *Arthritis Research & Therapy, 14*(2), R77. https://doi.org/10.1186/ar3800

Gilor, C., Niessen, S. J., Furrow, E., & DiBartola, S. P. (2016, Jul). What's in a Name? Classification of Diabetes Mellitus in Veterinary Medicine and Why It Matters. *J Vet Intern Med, 30*(4), 927-940. https://doi.org/10.1111/jvim.14357

Glassner, K. L., Abraham, B. P., & Quigley, E. M. M. (2020, 2020/01/01/). The microbiome and inflammatory bowel disease. *Journal of Allergy and Clinical Immunology, 145*(1), 16-27. https://doi.org/https://doi.org/10.1016/j.jaci.2019.11.003

Goldman, A. S. (2007, 2007/12/01). The Immune System in Human Milk and the Developing Infant. *Breastfeeding Medicine, 2*(4), 195-204. https://doi.org/10.1089/bfm.2007.0024

Griffin, D. O., Holodick, N. E., & Rothstein, T. L. (2011). Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+ CD27+ CD43+ CD70. *The Journal of experimental medicine, 208*(1), 67-80. https://doi.org/10.1084/jem.20101499

Grönwall, C., & Silverman, G. J. (2014). Natural IgM: beneficial autoantibodies for the control of inflammatory and autoimmune disease. *Journal of clinical immunology, 34*(1), 12-21.

Groom, A., Potter, C., Swan, D. C., Fatemifar, G., Evans, D. M., Ring, S. M., Turcot, V., Pearce, M. S., Embleton, N. D., Smith, G. D., Mathers, J. C., & Relton, C. L. (2012). Postnatal growth and DNA methylation are associated with differential gene expression of the TACSTD2 gene and childhood fat mass. *Diabetes, 61*(2), 391-400. https://doi.org/10.2337/db11-1039

Guan, Q. (2019). A Comprehensive Review and Update on the Pathogenesis of Inflammatory Bowel Disease. *Journal of immunology research, 2019*, 7247238-7247238. https://doi.org/10.1155/2019/7247238

Haines, D. M. (1986, Mar). A re-examination of islet cell cytoplasmic antibodies in diabetic dogs. *Vet Immunol Immunopathol, 11*(3), 225-233. https://www.ncbi.nlm.nih.gov/pubmed/3515747

He, J., Chen, D. L., Samocha-Bonet, D., Gillinder, K. R., Barclay, J. L., Magor, G. W., Perkins, A. C., Greenfield, J. R., Yang, G., & Whitehead, J. P. (2016, 2016/11/01). Fibroblast growth factor-1 (FGF-1) promotes adipogenesis by downregulation of carboxypeptidase A4 (CPA4) ‐ a negative regulator of adipogenesis implicated in the modulation of local and systemic insulin sensitivity. *Growth Factors, 34*(5-6), 210-216. https://doi.org/10.1080/08977194.2017.1285764

Hibi, T., & Ogata, H. (2006). Novel pathophysiological concepts of inflammatory bowel disease. *Journal of gastroenterology, 41*(1), 10-16.

Hoenig, M., & Dawe, D. L. (1992, May). A qualitative assay for beta cell antibodies. Preliminary results in dogs with diabetes mellitus. *Vet Immunol Immunopathol, 32*(3-4), 195-203. http://www.ncbi.nlm.nih.gov/pubmed/1632063

Holder, A. L., Kennedy, L. J., Ollier, W. E., & Catchpole, B. (2015, Oct). Breed differences in development of anti-insulin antibodies in diabetic dogs and investigation of the role of dog leukocyte antigen (DLA) genes. *Vet Immunol Immunopathol, 167*(3-4), 130-138. https://doi.org/10.1016/j.vetimm.2015.07.014

Hooper, L. V., Littman, D. R., & Macpherson, A. J. (2012). Interactions between the microbiota and the immune system. *Science, 336*(6086), 1268-1273.

Horner MJ, R. L., Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK. (2008). *SEER Cancer Statistics Review*. National Cancer Institute. https://seer.cancer.gov/archive/csr/1975_2006/

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics, 26*(5), 680-682. https://doi.org/10.1093/bioinformatics/btq003

Hugot, J.-P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J.-P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J.-F., Sahbatou, M., & Thomas, G. (2001, 2001/05/01). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature, 411*(6837), 599-603. https://doi.org/10.1038/35079107

Janeway, C. A. J., Travers, P., Walport, M. & Shlomchik M. J. . (2001). *Immunobiology, 5th edition*. Garland Science.

Kamada, N., Hisamatsu, T., Okamoto, S., Chinen, H., Kobayashi, T., Sato, T., Sakuraba, A., Kitazume, M. T., Sugita, A., & Koganei, K. (2008). Unique CD14+ intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN- γ axis. *The Journal of clinical investigation, 118*(6), 2269-2280.

Kaplan, G. G., Jackson, T., Sands, B. E., Frisch, M., Andersson, R. E., & Korzenik, J. (2008). The risk of developing Crohn's disease after an appendectomy: a meta-analysis. *Official journal of the American College of Gastroenterology| ACG, 103*(11), 2925-2931.

Kaplan, G. G., Pedersen, B. V., Andersson, R. E., Sands, B. E., Korzenik, J., & Frisch, M. (2007). The risk of developing Crohn’ s disease after an appendectomy: a population-based cohort study in Sweden and Denmark. *Gut, 56*(10), 1387-1392.

Katchman, B. A., Chowell, D., Wallstrom, G., Vitonis, A. F., LaBaer, J., Cramer, D. W., & Anderson, K. S. (2017, 2017/07/01/). Autoantibody biomarkers for the detection of serous ovarian cancer. *Gynecologic Oncology, 146*(1), 129-136. https://doi.org/https://doi.org/10.1016/j.ygyno.2017.04.005

Katsanos, A. H., Kyriakidi, K., Karassa, F. B., Politis, D., Skamnelos, A., Christodoulou, D. K., & Katsanos, K. H. (2017). Biomarker Development in Chronic Inflammatory Diseases. *Biomarkers for Endometriosis: State of the Art*, 41-75. https://doi.org/10.1007/978-3-319-59856-7_3

Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B. J., Jacobsen, L. M., Schatz, D. A., & Lernmark, Å. (2017, Mar). Type 1 diabetes mellitus. *Nat Rev Dis Primers, 3*, 17016. https://doi.org/10.1038/nrdp.2017.16

Khor, B., Gardet, A., & Xavier, R. J. (2011). Genetics and pathogenesis of inflammatory bowel disease. *Nature, 474*(7351), 307-317.

Kim, J. H., Furrow, E., Ritt, M. G., Utz, P. J., Robinson, W. H., Yu, L., Eckert, A., Stuebner, K., O'Brien, T. D., Steinman, L., & Modiano, J. F. (2016). Anti-Insulin Immune Responses Are Detectable in Dogs with Spontaneous Diabetes. *PLoS One, 11*(3), e0152397. https://doi.org/10.1371/journal.pone.0152397

Klag, T., Stange, E. F., & Wehkamp, J. (2013). Defective Antibacterial Barrier in Inflammatory Bowel Disease. *Digestive Diseases, 31*(3-4), 310-316. https://doi.org/10.1159/000354858

Klement, E., Cohen, R. V., Boxman, J., Joseph, A., & Reif, S. (2004). Breastfeeding and risk of inflammatory bowel disease: a systematic review with meta-analysis. *The American Journal of Clinical Nutrition, 80*(5), 1342-1352. https://doi.org/10.1093/ajcn/80.5.1342

Koloski, N. A., Bret, L., & Radford-Smith, G. (2008). Hygiene hypothesis in inflammatory bowel disease: a critical review of the literature. *World journal of gastroenterology: WJG, 14*(2), 165.

Koroleva, E. P., Halperin, S., Gubernatorova, E. O., Macho-Fernandez, E., Spencer, C. M., & Tumanov, A. V. (2015, 2015/06/01/). Citrobacter rodentium-induced colitis: A robust model to study mucosal immune responses in the gut. *Journal of immunological methods, 421*, 61-72. https://doi.org/https://doi.org/10.1016/j.jim.2015.02.003

Koyner, J. L., Vaidya, V. S., Bennett, M. R., Ma, Q., Worcester, E., Akhter, S. A., Raman, J., Jeevanandam, V., O'Connor, M. F., Devarajan, P., Bonventre, J. V., & Murray, P. T. (2010). Urinary biomarkers in the clinical prognosis and early detection of acute kidney injury. *Clinical journal of the American Society of Nephrology : CJASN, 5*(12), 2154-2165. https://doi.org/10.2215/CJN.00740110

Kricka, L. J. (1996). 15 - CHEMILUMINESCENCE IMMUNOASSAY. In E. P. Diamandis & T. K. Christopoulos (Eds.), *Immunoassay* (pp. 337-353). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012214730-2/50016-9

Kuna, A. T. (2013). Serological markers of inflammatory bowel disease. *Biochemia medica, 23*(1), 28-42. https://doi.org/10.11613/bm.2013.006

Kusnezow, W., Jacob, A., Walijew, A., Diehl, F., & Hoheisel, J. D. (2003). Antibody microarrays: an evaluation of production parameters. *Proteomics, 3*(3), 254-264.

Landry, J. P., Ke, Y., Yu, G.-L., & Zhu, X. D. (2015). Measuring affinity constants of 1450 monoclonal antibodies to peptide targets with a microarray-based label-free assay platform. *Journal of immunological methods, 417*, 86-96. https://doi.org/10.1016/j.jim.2014.12.011

Lee, M., & Chang, E. B. (2021, 2021/01/01/). Inflammatory Bowel Diseases (IBD) and the Microbiome—Searching the Crime Scene for Clues. *Gastroenterology, 160*(2), 524-537. https://doi.org/https://doi.org/10.1053/j.gastro.2020.09.056

Leslie, D., Lipsky, P., & Notkins, A. L. (2001). Autoantibodies as predictors of disease. *The Journal of clinical investigation, 108*(10), 1417-1422. https://doi.org/10.1172/JCI14452

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006, 2006/02/24/). Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell, 124*(4), 837-848. https://doi.org/https://doi.org/10.1016/j.cell.2006.02.017

Li, H., Limenitakis, J. P., Greiff, V., Yilmaz, B., Schären, O., Urbaniak, C., Zünd, M., Lawson, M. A. E., Young, I. D., Rupp, S., Heikenwälder, M., McCoy, K. D., Hapfelmeier, S., Ganal-Vonarburg, S. C., & Macpherson, A. J. (2020, 2020/08/01). Mucosal or systemic microbiota exposures shape the B cell repertoire. *Nature, 584*(7820), 274-278. https://doi.org/10.1038/s41586-020-2564-6

Limbrick Jr, D. D., Castaneyra-Ruiz, L., Han, R. H., Berger, D., McAllister, J. P., & Morales, D. M. (2017). Cerebrospinal Fluid Biomarkers of Pediatric Hydrocephalus. *Pediatric Neurosurgery, 52*(6), 426-435. https://doi.org/10.1159/000477175

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J.,

Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A. C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J. P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A., & Lander, E. S. (2005, Dec). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature, 438*(7069), 803-819. https://doi.org/10.1038/nature04338

Lobo, P. I. (2016). Role of natural autoantibodies and natural IgM anti-leucocyte autoantibodies in health and disease. *Frontiers in Immunology, 7*, 198.

Loddo, I., & Romano, C. (2015). Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Frontiers in Immunology, 6*, 551-551. https://doi.org/10.3389/fimmu.2015.00551

Lodes, M. J., Cong, Y., Elson, C. O., Mohamath, R., Landers, C. J., Targan, S. R., Fort, M., & Hershberg, R. M. (2004, 05/01/). Bacterial flagellin is a dominant antigen in Crohn disease. *The Journal of clinical investigation, 113*(9), 1296-1306. https://doi.org/10.1172/JCI20295

Long, G. V., Menzies, A. M., Nagrial, A. M., Haydu, L. E., Hamilton, A. L., Mann, G. J., Hughes, T. M., Thompson, J. F., Scolyer, R. A., & Kefford, R. F. (2011).

Prognostic and clinicopathologic associations of oncogenic BRAF in metastatic melanoma. *Journal of Clinical Oncology, 29*(10), 1239-1246.

Louis, E., Collard, A., Oger, A. F., Degroote, E., Aboul Nasr El Yafi, F. A., & Belaiche, J. (2001). Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut, 49*(6), 777-782. https://doi.org/10.1136/gut.49.6.777

Ludin, P., Nilsson, D., & Mäser, P. (2011). Genome-Wide Identification of Molecular Mimicry Candidates in Parasites. *PLOS ONE, 6*(3), e17546. https://doi.org/10.1371/journal.pone.0017546

Ludwig, R. J., Vanhoorelbeke, K., Leypoldt, F., Kaya, Z., Bieber, K., McLachlan, S. M., Komorowski, L., Luo, J., Cabral-Marques, O., Hammers, C. M., Lindstrom, J. M., Lamprecht, P., Fischer, A., Riemekasten, G., Tersteeg, C., Sondermann, P., Rapoport, B., Wandinger, K.-P., Probst, C., El Beidaq, A., Schmidt, E., Verkman, A., Manz, R. A., & Nimmerjahn, F. (2017). Mechanisms of Autoantibody-Induced Pathology. *Frontiers in Immunology, 8*, 603-603. https://doi.org/10.3389/fimmu.2017.00603

Lusted, L. B. (1971). Signal detectability and medical decision-making. *Science, 171*(3977), 1217-1219.

MacBeath, G., & Schreiber, S. L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science, 289*(5485), 1760-1763.

MacDermott, R. P., Nash, G. S., Bertovich, M. J., Seiden, M. V., Bragdon, M. J., & Beale, M. G. (1981). Alterations of IgM, IgG, and IgA synthesis and secretion by peripheral blood and intestinal mononuclear cells from patients with ulcerative colitis and Crohn's disease. *Gastroenterology, 81*(5), 844-852.

Macpherson, A. J., Geuking, M. B., & McCoy, K. D. (2005). Immune responses that adapt the intestinal mucosa to commensal intestinal bacteria. *Immunology, 115*(2), 153-162. https://doi.org/10.1111/j.1365-2567.2005.02159.x

Maekawa, M. (1995). Properties of enzymes for enzyme immunoassay. *Nihon rinsho. Japanese journal of clinical medicine, 53*(9), 2154-2159.

Mahmood, T., & Yang, P.-C. (2012). Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences, 4*(9), 429-434. https://doi.org/10.4103/1947-2714.100998

Mandrekar, J. N. (2010, 2010/09/01/). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316. https://doi.org/https://doi.org/10.1097/JTO.0b013e3181ec173d

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., & Dore, J. (2006a). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut, 55*(2), 205-211. https://doi.org/10.1136/gut.2005.073817

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., & Dore, J. (2006b). Reduced diversity of faecal microbiota in Crohn’s disease revealed by a metagenomic approach. *Gut, 55*(2), 205. https://doi.org/10.1136/gut.2005.073817

Mantis, N. J., Cheung, M. C., Chintalacharuvu, K. R., Rey, J., Corthésy, B., & Neutra, M. R. (2002). Selective adherence of IgA to murine Peyer’s patch M cells: evidence for a novel IgA receptor. *The Journal of Immunology, 169*(4), 1844-1851.

Mathieu, C., Lahesmaa, R., Bonifacio, E., Achenbach, P., & Tree, T. (2018, 11). Immunological biomarkers for the development and progression of type 1 diabetes. *Diabetologia, 61*(11), 2252-2258. https://doi.org/10.1007/s00125-018-4726-8

Mawdsley, J. E., & Rampton, D. S. (2006). The Role of Psychological Stress in Inflammatory Bowel Disease. *Neuroimmunomodulation, 13*(5-6), 327-336. https://doi.org/10.1159/000104861

Mayeux, R. (2004). Biomarkers: potential uses and limitations. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics, 1*(2), 182-188. https://doi.org/10.1602/neurorx.1.2.182

McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research, 32*(Web Server issue), W20-W25. https://doi.org/10.1093/nar/gkh435

Mehrgou, A., & Akouchekian, M. (2016). The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. *Medical journal of the Islamic Republic of Iran, 30*, 369-369. https://pubmed.ncbi.nlm.nih.gov/27493913

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4972064/

Miernyk, K. M., Bruden, D. L., Bruce, M. G., McMahon, B. J., Hennessy, T. W., Peters, H. V., Hurlburt, D. A., Sacco, F., & Parkinson, A. J. (2007). Dynamics of Helicobacter pylori-specific immunoglobulin G for 2 years after successful eradication of Helicobacter pylori infection in an American Indian and Alaska Native population. *Clinical and vaccine immunology : CVI, 14*(1), 85-86. https://doi.org/10.1128/CVI.00253-06

Miersch, S., Bian, X., Wallstrom, G., Sibani, S., Logvinenko, T., Wasserfall, C. H., Schatz, D., Atkinson, M., Qiu, J., & LaBaer, J. (2013, Dec). Serological autoantibody profiling of type 1 diabetes by protein arrays. *J Proteomics, 94*, 486-496. https://doi.org/10.1016/j.jprot.2013.10.018

Mirkov, M. U., Verstockt, B., & Cleynen, I. (2017, 2017/03/01/). Genetics of inflammatory bowel disease: beyond NOD2. *The Lancet Gastroenterology & Hepatology, 2*(3), 224-234. https://doi.org/https://doi.org/10.1016/S2468-1253(16)30111-X

Mitsuyama, K., Niwa, M., Takedatsu, H., Yamasaki, H., Kuwaki, K., Yoshioka, S., Yamauchi, R., Fukunaga, S., & Torimura, T. (2016). Antibody markers in the diagnosis of inflammatory bowel disease. *World journal of gastroenterology, 22*(3), 1304-1310. https://doi.org/10.3748/wjg.v22.i3.1304

Monteleone, G., Biancone, L., Marasco, R., Morrone, G., Marasco, O., Luzza, F., & Pallone, F. (1997). Interleukin 12 is expressed and actively released by Crohn's disease intestinal lamina propria mononuclear cells. *Gastroenterology, 112*(4), 1169-1178.

Monteleone, G., Trapasso, F., Parrello, T., Biancone, L., Stella, A., Iuliano, R., Luzza, F., Fusco, A., & Pallone, F. (1999). Bioactive IL-18 expression is up-regulated in Crohn’ s disease. *The Journal of Immunology, 163*(1), 143-147.

Montor, W. R., Huang, J., Hu, Y., Hainsworth, E., Lynch, S., Kronish, J.-W., Ordonez, C. L., Logvinenko, T., Lory, S., & LaBaer, J. (2009). Genome-Wide Study of Pseudomonas aeruginosa Outer Membrane Protein Immunogenicity Using Self-Assembling Protein Microarrays. *Infection and Immunity, 77*(11), 4877. https://doi.org/10.1128/IAI.00698-09

Müller, D., Telieps, T., Eugster, A., Weinzierl, C., Jolink, M., Ziegler, A. G., & Bonifacio, E. (2018, Sep). Novel minor HLA DR associated antigens in type 1 diabetes. *Clin Immunol, 194*, 87-91. https://doi.org/10.1016/j.clim.2018.07.001

Murphy, K. M., & Weaver, C. (2016). *Janeway's Immunobiology: Ninth International Student Edition*. Garland Science, Taylor & Francis Group, LLC. https://books.google.com/books?id=DmebDwAAQBAJ

Murthy, A., Li, Y., Peng, I., Reichelt, M., Katakam, A. K., Noubade, R., Roose-Girma, M., DeVoss, J., Diehl, L., & Graham, R. R. (2014). A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature, 506*(7489), 456-462.

Nagele, E. P., Han, M., Acharya, N. K., DeMarshall, C., Kosciuk, M. C., & Nagele, R. G. (2013). Natural IgG Autoantibodies Are Abundant and Ubiquitous in Human Sera, and Their Number Is Influenced By Age, Gender, and Disease. *PLOS ONE, 8*(4), e60726. https://doi.org/10.1371/journal.pone.0060726

Najafov, A., & Hoxhaj, G. (2017). Chapter 1 - Introduction. In A. Najafov & G. Hoxhaj (Eds.), *Western Blotting Guru* (pp. 1-3). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-813537-2.00001-1

Neiman, M., Hellström, C., Just, D., Mattsson, C., Fagerberg, L., Schuppe-Koistinen, I., Gummesson, A., Bergström, G., Kallioniemi, O., Achour, A., Sallinen, R., Uhlén, M., & Nilsson, P. (2019, 2019/01/02). Individual and stable autoantibody repertoires in healthy individuals. *Autoimmunity, 52*(1), 1-11. https://doi.org/10.1080/08916934.2019.1581774

Nelson, R. W., & Reusch, C. E. (2014, Sep). Animal models of disease: classification and etiology of diabetes in dogs and cats. *J Endocrinol, 222*(3), T1-9. https://doi.org/10.1530/JOE-14-0202

Nemazee, D. (2017, 2017/05/01). Mechanisms of central tolerance for B cells. *Nature Reviews Immunology, 17*(5), 281-294. https://doi.org/10.1038/nri.2017.19

Nguyen, H. H., Park, J., Kang, S., & Kim, M. (2015). Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel, Switzerland), 15*(5), 10481-10510. https://doi.org/10.3390/s150510481

Ni, J., Wu, G. D., Albenberg, L., & Tomov, V. T. (2017a, 2017/10/01). Gut microbiota and IBD: causation or correlation? *Nature Reviews Gastroenterology & Hepatology, 14*(10), 573-584. https://doi.org/10.1038/nrgastro.2017.88

Ni, J., Wu, G. D., Albenberg, L., & Tomov, V. T. (2017b). Gut microbiota and IBD: causation or correlation? *Nature reviews. Gastroenterology & hepatology, 14*(10), 573-584. https://doi.org/10.1038/nrgastro.2017.88

Nishida, A., Inoue, R., Inatomi, O., Bamba, S., Naito, Y., & Andoh, A. (2018, 2018/02/01). Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clinical Journal of Gastroenterology, 11*(1), 1-10. https://doi.org/10.1007/s12328-017-0813-5

Nordström, T., Akre, O., Aly, M., Grönberg, H., & Eklund, M. (2018). Prostate-specific antigen (PSA) density in the diagnostic algorithm of prostate cancer. *Prostate cancer and prostatic diseases, 21*(1), 57-63.

Nowak, J., Watala, C., & Boncler, M. (2014). Antibody binding, platelet adhesion, and protein adsorption on various polymer surfaces. *Blood Coagulation & Fibrinolysis, 25*(1). https://journals.lww.com/bloodcoagulation/Fulltext/2014/01000/Antibody_binding,_platelet_adhesion,_and_protein.10.aspx

O'Kell, A. L., Wasserfall, C. H., Henthorn, P. S., Atkinson, M. A., & Hess, R. S. (2020, 2020/08/11). Evaluation for type 1 diabetes associated autoantibodies in diabetic and non-diabetic Australian terriers and Samoyeds. *Canine Medicine and Genetics, 7*(1), 10. https://doi.org/10.1186/s40575-020-00089-5

Olde Damink, L. H. H., Dijkstra, P. J., van Luyn, M. J. A., van Wachem, P. B., Nieuwenhuis, P., & Feijen, J. (1996, 1996/01/01/). Cross-linking of dermal sheep collagen using a water-soluble carbodiimide. *Biomaterials, 17*(8), 765-773. https://doi.org/https://doi.org/10.1016/0142-9612(96)81413-X

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. *http://bioinfogp. cnb. csic. es/tools/venny/index. html*.

Palma, J., Tokarz-Deptuła, B., Deptuła, J., & Deptuła, W. (2018a). Natural antibodies - facts known and unknown. *Central-European journal of immunology, 43*(4), 466-475. https://doi.org/10.5114/ceji.2018.81354

Palma, J., Tokarz-Deptuła, B., Deptuła, J., & Deptuła, W. (2018b, 2018). Natural antibodies ‐ facts known and unknown. *Central European Journal of Immunology, 43*(4), 466-475. https://doi.org/10.5114/ceji.2018.81354

Panayi, G. S., & Corrigall, V. M. (2002, 2002-08-01). Autoantigens and Immune Pathways in Rheumatoid Arthritis. *22*(4), 13. https://doi.org/10.1615/CritRevImmunol.v22.i4.30

Parada Venegas, D., De la Fuente, M. K., Landskron, G., González, M. J., Quera, R., Dijkstra, G., Harmsen, H. J. M., Faber, K. N., & Hermoso, M. A. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases [10.3389/fimmu.2019.00277]. *Frontiers in Immunology, 10*, 277. https://www.frontiersin.org/article/10.3389/fimmu.2019.00277

Parikh, K., Antanaviciute, A., Fawkner-Corbett, D., Jagielowicz, M., Aulicino, A., Lagerholm, C., Davis, S., Kinchen, J., Chen, H. H., Alham, N. K., Ashley, N., Johnson, E., Hublitz, P., Bao, L., Lukomska, J., Andev, R. S., Björklund, E., Kessler, B. M., Fischer, R., Goldin, R., Koohy, H., & Simmons, A. (2019, 2019/03/01). Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature, 567*(7746), 49-55. https://doi.org/10.1038/s41586-019-0992-y

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

Pelanda, R., & Torres, R. M. (2012). Central B-cell tolerance: where selection begins. *Cold Spring Harbor perspectives in biology, 4*(4), a007146-a007146. https://doi.org/10.1101/cshperspect.a007146

Piovani, D., Danese, S., Peyrin-Biroulet, L., Nikolopoulos, G. K., Lytras, T., & Bonovas, S. (2019). Environmental Risk Factors for Inflammatory Bowel Diseases: An Umbrella Review of Meta-analyses. *Gastroenterology, 157*(3), 647-659.e644. https://doi.org/10.1053/j.gastro.2019.04.016

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak,

S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Le Roux, K., Leclerc, M., Maguin, E., Melo Minardi, R., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., de Vos, W., Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S. D., Wang, J., & Meta, H. I. T. C. (2010, 2010/03/01). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59-65. https://doi.org/10.1038/nature08821

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau Albert, Y., Walter Johannes, C., & LaBaer, J. (2004, 2004/07/02). Self-Assembling Protein Microarrays. *Science, 305*(5680), 86-90. https://doi.org/10.1126/science.1097639

Ramachandran, N., Raphael, J. V., Hainsworth, E., Demirkan, G., Fuentes, M. G., Rolfs, A., Hu, Y., & LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. *Nature methods, 5*(6), 535-538. https://doi.org/10.1038/nmeth.1210

Ramanan, D., Tang, Mei S., Bowcutt, R., Loke, P. n., & Cadwell, K. (2014). Bacterial Sensor Nod2 Prevents Inflammation of the Small Intestine by Restricting the Expansion of the Commensal Bacteroides vulgatus. *Immunity, 41*(2), 311-324. https://doi.org/10.1016/j.immuni.2014.06.015

Rauf, F., Anderson, K. S., & LaBaer, J. (2020). Autoantibodies in Early Detection of Breast Cancer. *Cancer Epidemiology Biomarkers & Prevention, 29*(12), 2475. https://doi.org/10.1158/1055-9965.EPI-20-0331

Reyneveld, G. I., Savelkoul, H. F. J., & Parmentier, H. K. (2020a, 2020-September-10). Current Understanding of Natural Antibodies and Exploring the Possibilities of Modulation Using Veterinary Models. A Review [Review]. *Frontiers in Immunology, 11*(2139). https://doi.org/10.3389/fimmu.2020.02139

Reyneveld, G. I., Savelkoul, H. F. J., & Parmentier, H. K. (2020b). Current Understanding of Natural Antibodies and Exploring the Possibilities of Modulation Using Veterinary Models. A Review [10.3389/fimmu.2020.02139]. *Frontiers in Immunology, 11*, 2139. https://www.frontiersin.org/article/10.3389/fimmu.2020.02139

Riemekasten, G., & Hahn, B. H. (2005). Key autoantigens in SLE. *Rheumatology, 44*(8), 975-982. https://doi.org/10.1093/rheumatology/keh688

Roep, B. O., & Peakman, M. (2012). Antigen targets of type 1 diabetes autoimmunity. *Cold Spring Harbor perspectives in medicine, 2*(4), a007781-a007781. https://doi.org/10.1101/cshperspect.a007781

Rojas, M., Restrepo-Jiménez, P., Monsalve, D. M., Pacheco, Y., Acosta-Ampudia, Y., Ramírez-Santana, C., Leung, P. S. C., Ansari, A. A., Gershwin, M. E., & Anaya, J.-M. (2018, 2018/12/01/). Molecular mimicry and autoimmunity. *Journal of Autoimmunity, 95*, 100-123. https://doi.org/https://doi.org/10.1016/j.jaut.2018.10.012

Romagnani, S. (1994). Lymphokine production by human T cells in disease states. *Annual review of immunology, 12*(1), 227-257.

Rose, G. D., Glerasch, L. M., & Smith, J. A. (1985). Turns in Peptides and Proteins. In C. B. Anfinsen, J. T. Edsall, & F. M. Richards (Eds.), *Advances in Protein Chemistry* (Vol. 37, pp. 1-109). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-3233(08)60063-7

Sartor, R. B. (2008, 2008/02/01/). Microbial Influences in Inflammatory Bowel Diseases. *Gastroenterology, 134*(2), 577-594. https://doi.org/https://doi.org/10.1053/j.gastro.2007.11.059

Satsangi, J., Silverberg, M. S., Vermeire, S., & Colombel, J. F. (2006). The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut, 55*(6), 749-753. https://doi.org/10.1136/gut.2005.082909

Schroeder, H. W., Jr., & Cavacini, L. (2010). Structure and function of immunoglobulins. *The Journal of allergy and clinical immunology, 125*(2 Suppl 2), S41-S52. https://doi.org/10.1016/j.jaci.2009.09.046

Scott, M., Nahm, M., Macke, K., Nash, G., Bertovich, M., & MacDermott, R. (1986). Spontaneous secretion of IgG subclasses by intestinal mononuclear cells: differences between ulcerative colitis, Crohn's disease, and controls. *Clinical and experimental immunology, 66*(1), 209.

Seksik, P., Nion-Larmurier, I., Sokol, H., Beaugerie, L., & Cosnes, J. (2009). Effects of light smoking consumption on the clinical course of Crohn's disease. *Inflammatory Bowel Diseases, 15*(5), 734-741.

Sela-Culang, I., Kunik, V., & Ofran, Y. (2013, 2013-October-08). The Structural Basis of Antibody-Antigen Recognition [Review]. *Frontiers in Immunology, 4*(302). https://doi.org/10.3389/fimmu.2013.00302

Shao, Q., Zhang, Z., Cao, R., Zang, H., Pei, W., & Sun, T. (2020). CPA4 Promotes EMT in Pancreatic Cancer via Stimulating PI3K-AKT-mTOR Signaling. *Onco Targets Ther, 13*, 8567-8580. https://doi.org/10.2147/ott.s257057

Shaw, S. Y., Blanchard, J. F., & Bernstein, C. N. (2011). Association Between the Use of Antibiotics and New Diagnoses of Crohn's Disease and Ulcerative Colitis. *Official journal of the American College of Gastroenterology | ACG, 106*(12). https://journals.lww.com/ajg/Fulltext/2011/12000/Association_Between_the_Use_of_Antibiotics_and_New.16.aspx

Shin, N.-R., Whon, T. W., & Bae, J.-W. (2015, 2015/09/01/). Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in Biotechnology, 33*(9), 496-503. https://doi.org/https://doi.org/10.1016/j.tibtech.2015.06.011

Shoda, R., Matsueda, K., Yamato, S., & Umeda, N. (1996). Epidemiologic analysis of Crohn disease in Japan: increased dietary intake of n-6 polyunsaturated fatty acids and animal protein relates to the increased incidence of Crohn disease in Japan. *The American Journal of Clinical Nutrition, 63*(5), 741-745. https://doi.org/10.1093/ajcn/63.5.741

Sibani, S., & LaBaer, J. (2011). Immunoprofiling Using NAPPA Protein Microarrays. In C. J. Wu (Ed.), *Protein Microarray for Disease Analysis: Methods and Protocols* (Vol. 723). Springer.

Silva, F. A. R., Rodrigues, B. L., Ayrizono, M. d. L. S., & Leal, R. F. (2016, 2016/12/14). The Immunological Basis of Inflammatory Bowel Disease. *Gastroenterology Research and Practice, 2016*, 2097274. https://doi.org/10.1155/2016/2097274

Silverberg, M. S., Satsangi, J., Ahmad, T., Arnott, I. D. R., Bernstein, C. N., Brant, S. R., Caprilli, R., Colombel, J.-F., Gasche, C., Geboes, K., Jewell, D. P., Karban, A., Loftus, E. V., Peña, A. S., Riddell, R. H., Sachar, D. B., Schreiber, S., Steinhart, A. H., Targan, S. R., Vermeire, S., & Warren, B. F. (2005, 1900/01/01). Toward an Integrated Clinical, Molecular and Serological Classification of Inflammatory

Bowel Disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Canadian Journal of Gastroenterology, 19*, 269076. https://doi.org/10.1155/2005/269076

Song, L., Song, M., Camargo, M. C., Van Duine, J., Williams, S., Chung, Y., Kim, K.-M., Lissowska, J., Sivins, A., Gao, W., Karthikeyan, K., Park, J., Leja, M., Cohen, J. I., LaBaer, J., Qiu, J., & Rabkin, C. S. (2021, 2021/07/01). Identification of anti-Epstein-Barr virus (EBV) antibody signature in EBV-associated gastric carcinoma. *Gastric Cancer, 24*(4), 858-867. https://doi.org/10.1007/s10120-021-01170-z

Song, L., Song, M., Rabkin, C. S., Williams, S., Chung, Y., Van Duine, J., Liao, L. M., Karthikeyan, K., Gao, W., Park, J. G., Tang, Y., Lissowska, J., Qiu, J., LaBaer, J., & Camargo, M. C. (2021, 2021/01/01). Helicobacter pylori Immunoproteomic Profiles in Gastric Cancer. *Journal of proteome research, 20*(1), 409-419. https://doi.org/10.1021/acs.jproteome.0c00466

Song, L., Wallstrom, G., Yu, X., Hopper, M., Van Duine, J., Steel, J., Park, J., Wiktor, P., Kahn, P., Brunner, A., Wilson, D., Jenny-Avital, E. R., Qiu, J., Labaer, J., Magee, D. M., & Achkar, J. M. (2017). Identification of Antibody Targets for Tuberculosis Serology using High-Density Nucleic Acid Programmable Protein Arrays. *Molecular & cellular proteomics : MCP, 16*(4 suppl 1), S277-S289. https://doi.org/10.1074/mcp.M116.065953

Sternberg, E. M., Chrousos, G. P., Wilder, R. L., & Gold, P. W. (1992, 1992/11/15). The Stress Response and the Regulation of Inflammatory Disease. *Annals of Internal Medicine, 117*(10), 854-866. https://doi.org/10.7326/0003-4819-117-10-854

Stillman, B. A., & Tonkinson, J. L. (2000, 2000/09/01). FASTTM Slides: A Novel Surface for Microarrays. *BioTechniques, 29*(3), 630-635. https://doi.org/10.2144/00293pf01

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current opinion in HIV and AIDS, 5*(6), 463-466. https://doi.org/10.1097/COH.0b013e32833ed177

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545. https://doi.org/10.1073/pnas.0506580102

Suurmond, J., & Diamond, B. (2015, 06/01/). Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity. *The Journal of Clinical Investigation, 125*(6), 2194-2202. https://doi.org/10.1172/JCI78084

Swidsinski, A., Loening-Baucke, V., Theissig, F., Engelhardt, H., Bengmark, S., Koch, S., Lochs, H., & Dörffel, Y. (2007). Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut, 56*(3), 343. https://doi.org/10.1136/gut.2006.098160

Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., Inatomi, O., Bamba, S., Andoh, A., & Sugimoto, M. (2016). Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. *Digestion, 93*(1), 59-65.

Takulapalli, B. R., Qiu, J., Magee, D. M., Kahn, P., Brunner, A., Barker, K., Means, S., Miersch, S., Bian, X., Mendoza, A., Festa, F., Syal, K., Park, J. G., LaBaer, J., & Wiktor, P. (2012). High density diffusion-free nanowell arrays. *Journal of proteome research, 11*(8), 4382-4391. https://doi.org/10.1021/pr300467q

Tamboli, C. P., Neut, C., Desreumaux, P., & Colombel, J. F. (2004). Dysbiosis in inflammatory bowel disease. *Gut, 53*(1), 1-4. https://doi.org/10.1136/gut.53.1.1

Targan, S. R., Landers, C. J., Yang, H., Lodes, M. J., Cong, Y., Papadakis, K. A., Vasiliauskas, E., Elson, C. O., & Hershberg, R. M. (2005, 2005/06/01/). Antibodies to CBir1 Flagellin Define a Unique Response That Is Associated Independently With Complicated Crohn’s Disease. *Gastroenterology, 128*(7), 2020-2028. https://doi.org/https://doi.org/10.1053/j.gastro.2005.03.046

Torres, U. d. S., Rodrigues, J. O., Junqueira, M. S. G., Uezato, S., & Netinho, J. G. (2010). The Montreal classification for Crohn's disease: clinical application to a Brazilian single-center cohort of 90 consecutive patients. *Arquivos de gastroenterologia, 47*, 279-284.

van der Wal, F. J., Bergervoet, J. H. W., Achterberg, R. P., & Haasnoot, W. (2014). Bead-based immunoassays. In *Novel Approaches in Immunoassays* (pp. 52-71). Future Medicine Ltd. https://doi.org/doi:10.2217/ebo.13.560

10.2217/ebo.13.560

Venigalla, S. S. K., Premakumar, S., & Janakiraman, V. (2020, 2020/01/22). A possible role for autoimmunity through molecular mimicry in alphavirus mediated

arthritis. *Scientific Reports, 10*(1), 938. https://doi.org/10.1038/s41598-019-55730-6

Vermeire, S., Joossens, S., Peeters, M., Monsuur, F., Marien, G., Bossuyt, X., Groenen, P., Vlietinck, R., & Rutgeerts, P. (2001, 2001/03/01/). Comparative study of ASCA (Anti‐Saccharomyces cerevisiae antibody) assays in inflammatory bowel disease. *Gastroenterology, 120*(4), 827-833. https://doi.org/https://doi.org/10.1053/gast.2001.22546

Vestergaard, M. d., Kerman, K., & Tamiya, E. (2007). An Overview of Label-free Electrochemical Protein Sensors. *Sensors (Basel, Switzerland), 7*(12), 3442-3458. https://doi.org/10.3390/s7123442

Vreugdenhil, G. R., Geluk, A., Ottenhoff, T. H. M., Melchers, W. J. G., Roep, B. O., & Galama, J. M. D. (1998, 1998/01/01). Molecular mimicry in diabetes mellitus: the homologous domain in coxsackie B virus protein 2C and islet autoantigen GAD65 is highly conserved in the coxsackie B-like enteroviruses and binds to the diabetes associated HLA-DR3 molecule. *Diabetologia, 41*(1), 40-46. https://doi.org/10.1007/s001250050864

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G., & Petrovska, L. (2011, 2011/01/10). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiology, 11*(1), 7. https://doi.org/10.1186/1471-2180-11-7

Walther, D., Eugster, A., Jergens, S., Gavrisan, A., Weinzierl, C., Telieps, T., Winkler, C., Ziegler, A. G., & Bonifacio, E. (2016, Sep). Tetraspanin 7 autoantibodies in type 1 diabetes. *Diabetologia, 59*(9), 1973-1976. https://doi.org/10.1007/s00125-016-3997-1

Wang, D., Yang, L., Zhang, P., LaBaer, J., Hermjakob, H., Li, D., & Yu, X. (2016). AAgAtlas 1.0: a human autoantigen database. *Nucleic Acids Research, 45*(D1), D769-D776. https://doi.org/10.1093/nar/gkw946

Wang, H., Demirkan, G., Bian, X., Wallstrom, G., Barker, K., Karthikeyan, K., Tang, Y., Pasha, S. F., Leighton, J. A., Qiu, J., & LaBaer, J. (2017). Identification of Antibody Against SNRPB, Small Nuclear Ribonucleoprotein-Associated Proteins B and B’, as an Autoantibody Marker in Crohn’s Disease using an

Immunoproteomics Approach. *Journal of Crohn's and Colitis, 11*(7), 848-856. https://doi.org/10.1093/ecco-jcc/jjx019

Wang, J., Figueroa, J. D., Wallstrom, G., Barker, K., Park, J. G., Demirkan, G., Lissowska, J., Anderson, K. S., Qiu, J., & LaBaer, J. (2015a). Plasma Autoantibodies Associated with Basal-like Breast Cancers. *Cancer Epidemiology Biomarkers & Prevention, 24*(9), 1332. https://doi.org/10.1158/1055-9965.EPI-15-0047

Wang, J., Figueroa, J. D., Wallstrom, G., Barker, K., Park, J. G., Demirkan, G., Lissowska, J., Anderson, K. S., Qiu, J., & LaBaer, J. (2015b). Plasma Autoantibodies Associated with Basal-like Breast Cancers. *Cancer Epidemiology Biomarkers & Prevention, 24*(9), 1332. https://doi.org/10.1158/1055-9965.EPI-15-0047

Wehkamp, J., Harder, J., Weichenthal, M., Mueller, O., Herrlinger, K. R., Fellermann, K., Schroeder, J. M., & Stange, E. F. (2003). Inducible and constitutive β-defensins are differentially expressed in Crohn's disease and ulcerative colitis. *Inflammatory Bowel Diseases, 9*(4), 215-223.

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schäffeler, E., Schlee, M., Herrlinger, K., Stallmach, A., Noack, F., & Fritz, P. (2004). NOD2 (CARD15) mutations in Crohn’s disease are associated with diminished mucosal α-defensin expression. *Gut, 53*(11), 1658-1664.

Wehkamp, J., Salzman, N. H., Porter, E., Nuding, S., Weichenthal, M., Petras, R. E., Shen, B., Schaeffeler, E., Schwab, M., & Linzmeier, R. (2005). Reduced Paneth cell α-defensins in ileal Crohn's disease. *Proceedings of the National Academy of Sciences, 102*(50), 18129-18134.

Wu Gary, D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh Sue, A., Bewtra, M., Knights, D., Walters William, A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman Frederic, D., & Lewis James, D. (2011, 2011/10/07). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science, 334*(6052), 105-108. https://doi.org/10.1126/science.1208344

Xiao, Z. X., Miller, J. S., & Zheng, S. G. (2021, 2021/02/01/). An updated advance of autoantibodies in autoimmune diseases. *Autoimmunity Reviews, 20*(2), 102743. https://doi.org/https://doi.org/10.1016/j.autrev.2020.102743

Yadav, S., Kashaninejad, N., Masud, M. K., Yamauchi, Y., Nguyen, N.-T., & Shiddiky, M. J. A. (2019, 2019/08/15/). Autoantibodies as diagnostic and prognostic cancer biomarker: Detection techniques and approaches. *Biosensors and Bioelectronics, 139*, 111315. https://doi.org/https://doi.org/10.1016/j.bios.2019.111315

Yamamoto, S., & Ma, X. (2009). Role of Nod2 in the development of Crohn's disease. *Microbes and infection, 11*(12), 912-918.

Yu, L. C.-H. (2015). Commensal bacterial internalization by epithelial cells: An alternative portal for gut leakiness. *Tissue barriers, 3*(3), e1008895-e1008895. https://doi.org/10.1080/21688370.2015.1008895

Zakerska-Banaszak, O., Tomczak, H., Gabryel, M., Baturo, A., Wolko, L., Michalak, M., Malinska, N., Mankowska-Wierzbicka, D., Eder, P., Dobrowolska, A., Slomski, R., & Skrzypczak-Zielinska, M. (2021, 2021/01/25). Dysbiosis of gut microbiota in Polish patients with ulcerative colitis: a pilot study. *Scientific Reports, 11*(1), 2166. https://doi.org/10.1038/s41598-021-81628-3

Zhou, X., Gao, C., Huang, W., Yang, M., Chen, G., Jiang, L., Gou, F., Feng, H., Ai, N., & Xu, Y. (2014, 2014/05/27). High Glucose Induces Sumoylation of Smad4 via SUMO2/3 in Mesangial Cells. *BioMed Research International, 2014*, 782625. https://doi.org/10.1155/2014/782625

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean Ralph, A., Gerstein, M., & Snyder, M. (2001, 2001/09/14). Global Analysis of Protein Activities Using Proteome Chips. *Science, 293*(5537), 2101-2105. https://doi.org/10.1126/science.1062191

Zou, K. H., O' Malley, A. J., & Mauri, L. (2007, 2007/02/06). Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation, 115*(5), 654-657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929

Zuo, T., & Ng, S. C. (2018). The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory Bowel Disease [10.3389/fmicb.2018.02247]. *Frontiers in Microbiology, 9*, 2247. https://www.frontiersin.org/article/10.3389/fmicb.2018.02247