2014

Exploring Novel System Biology Approaches to Understand the Molecular Mechanisms of Immune Responses

Cassandra Camille Garbutt
University of Alabama at Birmingham

EXPLORING NOVEL SYSTEM BIOLOGY APPROACHES TO UNDERSTAND THE MOLECULAR MECHANISMS OF IMMUNE RESPONSES

by

CASSANDRA GARBUTT

SHAHID MUKHTAR, COMMITTEE CHAIR
ASIM K. BEJ
ROBERT W. THACKER

A THESIS

Submitted to the graduate faculty of The University of Alabama at Birmingham, in partial fulfillment of the requirements for the degree of Master of Science

BIRMINGHAM, ALABAMA

2014

EXPLORING NOVEL SYSTEM BIOLOGY APPROACHES TO UNDERSTAND THE MOLECULAR MECHANISMS OF IMMUNE RESPONSES

CASSANDRA GARBUTT

MOLECULAR BIOLOGY

ABSTRACT

Entire new molecular worlds of immunity and autoimmunity have been unveiled through the lens of systems biology. Although many believe that vertebrates maintain the most complex immune system, a rival to this concept is arising due to a systems' biology perspective of plant immunity. There are various rising systems biology approaches that unveil this previously uncharted territory. The organization of subjects within systems biology including "-omes" and protein–protein interaction networks enhance such exploration. The field of plant immune network biology is growing alike its parts: prevailing computational modeling approaches of biological regulatory network dynamics, rising technologies and availing research avenues pertaining to the "-omics" approach. Systems biology approaches also pursue clues related to the molecular mechanisms of human autoimmunity, a current mystery.

Although the factors that cause the onset of systemic lupus erythematosus (SLE) are not fully understood, it is known to have several genetic risk factors. One factor relates to the fragment crystallizable receptor gene 2B, FCGR2B, that codes for the protein FcγRIIB. FcγRIIB is highly responsible for maintaining homeostasis within a cell by simultaneously triggering the activation or inhibition of receptors

related to undesired autoimmune responses. Systems biology provides effective approaches towards uncovering the role of human fragment crystallizable receptors (FCRs) in autoimmunity. In this thesis, three primary objectives were pursued to expand the knowledge of molecular human autoimmunity: the identification of novel interacting partners of FCRs' cytoplasmic domains; the finding of statistically overrepresented *cis*-regulatory elements in FCGR2B and identification of their cognate transcription factors; and the identification of FCGR2B CNV (Copy number variation) in SLE patients. The first objective entailed the application of a yeast-two hybrid assay, a high-throughput technology that identified protein-protein interactions and resulted in the generation of the first human autoimmune network. Bioinformatic tools that identify motifs, namely MEME and POBO, were utilized for the second objective. Lastly, the third objective entailed a revamped methodological approach that yielded a full-length RACE PCR product of the 1q23 gene cluster, which is the location of the FCGR2B gene. This full-length product enables the investigation for associations between FCGR2B CNV and SLE onset.

ACKNOWLEGEMENTS

My mentor, Dr. Shahid Mukhtar, taught me essentially all of my lab- and research- related skills. He fueled my spark for research and served as a fantastic mentor. I am grateful for all of the time and energy he has shared, as I would be a completely different person today if not for his generosity. Dr. Robert Kimberly was also an inspiring person to work with. Dr. Travis Ptacek also thoroughly described the copy number variation aspect of the project to me. Dr. Marc Vidal and the team at Dana Farber Cancer Institute in Boston, MA, including Nidhi Sahni inspired me to pursue research and be my best; they made me feel like a part of their team as they taught me a golden skillset. The Mukhtar lab has tremendously contributed to these projects. Dr. Elsa Gonzales, whom I worked with this summer in Lima, Peru on a public health research project polished my lab-skills, and taught me how to expand my skills to best suit my interests. Marshall Pollard, my mentor at the Birmingham Education Foundation, has taught me valuable leadership skills and is a great role model. The Huaorani tribe of the Ecuadorian Amazon, Judith Kimerling and Rusty Rushton have been amazing friends and family along my journey. Doug Baulos is also an inspirational awakening artist, friend and mentor. My parents, David R. Garbutt and Xochitl Escarsega, and siblings Alejandra and Johnny Garbutt have always been there for me. And my closest friends who have become family as they help share love and insight with me through my highs and lows.

TABLE OF CONTENTS

LIST OF TABLES

Table *Page*

UNDERSTANDING FC RECEPTROS THROUGH THEIR TRANSCRIPTIONAL REGULATION AND COPY NUMBER VARIATION

LIST OF FIGURES

Figure *Page*

COMMON ABBREVIATIONS

CNV	Copy Number Variation
ETI	Effector-Triggered Immunity
FcR	Fc Receptor
FCGR	Fc Gamma Receptor
FCGR2B	Fc Gamma Receptor 2B (gene)
FcγRIIB	Fc Gamma Receptor 2B (protein)
MAMPs	Microbial-Associated Molecular Patterns
ORF	Open Reading Frame
RA	Rheumatoid Arthritis
SLE	Systemic Lupus Erythematosus
SNP	Single Nucleotide Polymorphism

INTRODUCTION

The immune system is a versatile constellation of defense responses against pathogenic invading microorganisms and cancer in humans. The immune system consists of many cells and molecules that can specifically recognize and eliminate an innumerable variety of foreign invaders. Despite its complexity, it can functionally be categorized into two main activities in both plants and animals, recognition and response, as its aim is to protect the host from infectious disease. Recognition entails the ability to differentiate between self and non-self, and each species has a diverse array of responses to pathogenic disruptions. Some of these responses are pathogen-specific while others are general (Kindt, T. *et* al 2006). The key difference between human and plant immunity is that plants lack mobile immune cells (Spoel *et* al 2012). This feature comes with the complex adaptive immune system seen in vertebrates. In a nutshell, adaptive immunity enables vertebrates to have mechanisms that reduce self-reactivity while maintaining antigen-specific immune capacity and memory (Spoel *et* al 2012). Only recently have researches begun exploring a plant version of autoimmunity to explain instances in which offspring of healthy plant strains becomes ill (Katsnelson 2008). The term autoimmunity is characteristically reserved for animals since plants lack adaptive immune systems (Katsnelson 2008). A question arises then as to how plants can maintain an effective immune system without an adaptive immune system. Recently, with the uprising of systems biology, plants have been discovered to have complex, sophisticated, efficient and multifaceted innate immune responses that involve systemic signaling, cross-talk,

chromosomal changes and self-surveillance (Spoel *et* al 2012). Plants also exhibit pathogen-specific RNA silencing elicited from previous infections (Pumplin, N. *et* al 2013). RNA silencing is crucial for gene regulation in many eukaryotes as it acts on two levels: DNA methylation at the transcriptional level and direct mRNA interference via small RNAs at the post-transcriptional level (Pumplin, N. *et* al 2013). Both plants and animals have pattern-recognition receptors (PRRs), which detect microorganism-associated molecular patterns (MAMPs) (Spoel *et* al 2012). MAMPs include characteristic microbial features such as flagellin and peptidoglycans (Spoel *et* al 2012). The structural similarities between the PRRs of animals and plants are attributed to convergent evolution, which occurs when two organisms of different ancestry evolve similar advantageous traits (Spoel *et* al 2012). Another similarity between plant and animal immunity is the methods of study. Both systems can be analyzed through a systems' biology lens, which entails network biology and interactome maps. The subsequent sections within this introduction explore plant and human immunity, autoimmune diseases particularly systemic lupus erythmatosis (SLE), genetics associated with autoimmunity, and lastly branches of systems biology that serve as novel tools to understand immunity in both plants and animals. The first chapter of my thesis explores existing systems biology approaches towards understanding plant immunity, and the second chapter of my thesis is composed of three objectives that aim to understand the regulation of human fragment crystallizable receptors (FCRs) in autoimmunity: (1) the identification of novel interacting partners of FCRs' cytoplasmic domains, (2) the finding of statistically overrepresented *cis*-regulatory elements in FCGR2, the only classical FCR and the identification of their cognate transcription factors, and (3) the identification of

FCGR2B CNV (copy number variation) in SLE patients.

Plant Immunity

Plants also have inherent or induced defenses against pathogens. Pathogenic molecules are recognized by the plant's cell surface receptors, which initiate specific signaling cascades to protect the host. Upon pathogen recognition, the plant-microbe pathosystem undergoes an extensive transcriptional reprogramming in a highly dynamic and temporally regulated manner (Proietti, S. *et* al. 2013). Stimulation of these plant defenses involves complex signal transduction networks incorporating feedback and cross-talk controlled by largely unknown mechanisms (Mukhtar *et* al. 2009). Upon pathogen-recognition, plants rewire their cellular network to activate immune responses that entail the first line of defense, microbial-associated molecular patterns (MAMPs)- and the second line of defense, Effector-Triggered Immunity (ETI). MAMPs constitute a part of the plants non-self recognition signals, which also includes damage-associated molecular patterns (DAMPs), and pathogen-derived effectors. Despite the maintenance of a sophisticated plant immune system, certain pathogens have evolved suites of virulence proteins, or effectors, that influence the dynamics of the plant pathosystem to cause diseases and alter the host system for their benefit, as to acquire nutrients for instance (Robinson *et* al. 2006). The co-evolutionary arms race between pathogen and plant systems resulted in the development of two major classes of receptors that provide efficient surveillance. Pattern-recognition receptors (PRRs) and R proteins perceive MAMPs and DAMPs and activate MTI. This is the first line of defense against pathogen proliferation and dispersion within the host. Interestingly, pathogens have evolved specialized phytopathogens that they secrete to avoid detection, and they are known to

target MTI signaling components. These efforts are geared to diminish the plant's defenses and enhance the pathogen's virulence. Plants have a sophisticated immune system that has a counter defense to these pathogenic strategies that entail a major class of R proteins, NLR receptors (nucleotide-binding domain and leucine-rich repeat receptors), which directly bind to certain effectors or indirectly discover effector activities to trigger the second line of defense known as ETI (Sessa 2012). To evade ETI, pathogens continually alter their effector suites and modify specific effectors. This continual renewal of pathogen effectors enforces selective pressures onto the host's recognition and response abilities.

PAMP- and effector- triggered immunity are also major players in plant defense. PAMP (pathogen associated molecular pattern) are molecules recognized as foreign by the plant, and effectors are usually associated with parasitism and are recognized by host intracellular resistance proteins (Sessa 2012). Plant tolerance and resistance categorize the plant's reaction to pathogens. Tolerance is measured by the damage inflicted on the host by a pathogen load, whereas resistance is measured by the reduction of pathogen load and growth.

Human Immunity

When a healthy immune system encounters foreign molecules, it can discriminate them from the body's proteins and cells. The active molecules in immunoglobulin are called antibodies, which function to neutralize, precipitate and agglutinate toxins (Kindt *et* al, 2006). Upon identification of a foreign molecule, an effector response is mounted to eliminate or neutralize the entity. This immediate response is a part of innate immunity, which is a more primitive immune system among vertebrates, uniform within a species,

and consists of many circulating cell types (Kindt *et* al, 2006). Upon later exposure to the same foreign molecule, the memory response is launched, which is characterized by a heightened and rapid immune reaction that eliminates the pathogen and prevents disease (Kindt *et* al, 2006). The memory response can occur many times throughout a human's life. The memory response is a later evolved system that is acquired and more specific as well as a part of adaptive immunity. Lymphocytes, their antibodies and molecules they produce are the main players of adaptive immunity as through a maturation process, high affinity antigen specific receptors are made. Lymphocytes are formed in the bone marrow in mammals and create antibodies in response to a specific antigen. T- and B-lymphocytes are created in the bone marrow, but T- lymphocytes mature in the thymus whereas B- lymphocytes mature in the bone marrow. Innate immunity is the first line of defense against a pathogen whereas adaptive immunity takes more time to execute (Kindt *et* al, 2006). For any immune response, antigen-presenting cells in addition to T lymphocytes are crucial. Antigen presenting cells are specialized and regulate T lymphocyte responses by delivering a necessary co-stimulatory signal for T lymphocyte activation. T lymphocytes recognize epitope proteins bound with the major histocompatibility complex (MHC) that are displayed on host-cells, which are altered by viruses or cancer. In summary, T lymphocytes recognize a "non-self" target, such as a pathogen, only after antigens (small fragments of the pathogen) have been processed and presented in combination with a "self" receptor called a major histocompatibility complex (MHC) molecule. Whereas B lymphocytic antigen-specific receptors are antibody molecules located on the surface of B-lymphocytes, which recognize pathogens without any need for antigen processing (Kindt *et* al, 2006).

The innate immune system can also adapt through an existing pool of natural variation in pattern recognition and other immune receptors. Random mutations can also create new variants over time (Carvunis et al., 2013). Copy number variation (discussed in later section) can also alter the number of certain receptors present on lymphocytes. The advantage of a diverse collection of receptor alleles can enable the innate immune system to respond through natural selection when pathogens alter their antigens. The innate immune system changes occur through many generations and at the population level. Thus the spectrum of alleles associated with the innate immune system varies throughout geographical locations (Kindt *et* al, 2006).

Human Autoimmunity

The immune system can fail as a protector when it becomes an aggressor to its own host. The activation of the immune system against self-antigens is characterized as autoimmunity. Modern medicine encounters difficulty when treating patients with autoimmune diseases, because the culprit of the patient's deteriorating health is essentially themselves, specifically their body's own machinery. As a result, modern medicine can only hinder the deteriorating effects of these diseases, rather than cure them. Another reason why autoimmune diseases are difficult to treat is that they tend to be polygenic, which means that there are many genes that contribute to disease susceptibility. There are cellular mechanisms in place to prevent autoimmunity. An inappropriate T lymphocyte response to the host components can lead to fatal autoimmune consequences, which is why antigen-presenting cells are necessary before T lymphocyte activation. During lymphocyte development, the bone marrow employs two mechanisms to modify autoreactive B lymphocytes (Kindt *et* al, 2006). These

mechanisms of high importance are called central and peripheral tolerance. Central tolerance includes mechanisms that modify, destroy, or inactivate autoreactive cells. Central tolerance is a method intended to catch these defects early on in the development of B lymphocytes. If an autoreactive cell escapes this safety net, then there is another mechanism that limits the occurrence of autoreactivity in mature lymphocytes (Kindt *et* al, 2006). This is called peripheral tolerance. The failure of both peripheral and central tolerances leads to the activation of the immune system against self-antigens, or better known as autoimmunity. Autoimmunity can affect any organ and organ system in the human body.

Systemic Autoimmune Diseases

The human body can be characterized into major systems including muscular, skeletal, nervous, respiratory, cardiovascular, and immune systems. These systems are composed of groups of organs and molecules that together orchestrate one or several functions together. The involvements of several tissues and organs as well as a wide range of target antigens categorize systemic autoimmune diseases. In general, there is hyperactivity among T and B lymphocytes. The accumulation of immune complexes, auto-antibodies, and cell-mediated immune responses cause widespread tissue damage not exclusive to one system (Kindt *et* al, 2006). Over 11 million Americans suffer from rheumatic autoimmune diseases including systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), which are characterized by painful inflammation of the joints and muscles (Helmick *et* al, 2008). In order to understand SLE on a molecular level, the FCGR2B gene, which has been associated with the pathogenesis of SLE, was investigated. The role of this gene and the receptor it encodes will be elaborated in

subsequent sections.

Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a chronic inflammatory disease whose symptoms include fever, arthritis, skin rashes, kidney dysfunction and pleurisy. The lupus foundation of America estimates that 1.5 million Americans have a form of lupus. SLE appears more so in women then men, a 10:1 ratio, and within the 20 to 40 age group (Kindt *et* al, 2006). There are antibodies produced to a variety of tissue antigens including red blood cells, leukocytes, clotting factors, DNA and histones. Patients with severe SLE have excessive complement activation observable by high levels of C3a and C5a, which can be up to four times more than normal (Kindt *et* al, 2006). Complements enhance the clearing of pathogens by assisting phagocytes and antibodies. C5a increases type 3 complement receptor expression, which is located on neutrophils; this enhances neutrophil aggregation and attachment to vascular endothelium (Kindt *et* al, 2006). As neutrophils become attached to the vascular endothelium, blood vessels become blocked (vasculitis), which leads to widespread tissue damage. The diagnosis of SLE relies on indirect immunofluorescent staining that reveals SLE characteristic nucleus-staining patterns (Kindt *et* al, 2006). The staining utilizes SLE specific antinuclear antibodies that are directed to DNA, histones, nuclear RNA and nucleoprotein.

Although the onset of SLE is not completely understood, environmental and genetic factors are known causes. Variability in the clinical phenotype suggests an environmental influence on the disease phenotype. However, there is an intersection of genetics and environmental factors that can be seen in ethnic differences in the disease phenotype. For instance, SLE is more common among African Americans and they are

also more at risk for developing SLE-nephritis (Kindt *et* al, 2006). Moreover, certain genetic factors characterize SLE in certain populations while not in others. Specifically, the FcγRIIB I187T allele is known to cause an SLE risk among Asians, but not among those of European or African decent. The degree of genetic and environmental contributions to the onset of SLE is unknown, however, the genetics associated will be investigated in two projects of my thesis. Specifically, the focus will be on the classical low affinity Fcγ receptors, which is discussed thoroughly in subsequent section.

Rheumatoid Arthritis

Rheumatoid arthritis (RA) is characterized by chronic inflammation of the joints, and respiratory and cardiovascular systems are often affected. RA is most common among 40 to 60 year old women. A majority of patients with RA produce auto-antibodies known as rheumatoid factors that react with the Fc region of IgG. IgM antibody with such reactivity is most common and binds to IgG, which creates IgG-IgM complexes that are deposited in joints (Kindt *et* al, 2006). These IgG-IgM complex accumulations cause a hypersensitive reaction that causes chronic inflammation.

Fc Receptors Signaling in Autoimmune Diseases

Fc receptors are membrane glycoproteins that have an affinity for the Fc portion of an antibody. They are present in all of the immunoglobulin classes. Fc receptors are critical for the biological functions of antibodies as they are responsible for the regulation of IgG serum levels, the movement of antibodies across the cell membrane as well as the transfer of IgG across the placenta from the mother to child during gestation. Fc receptors are also essential for the passive transfer of antibodies for certain cell types including neutrophils, mast cells, macrophages, natural killer cells, and B and T lymphocytes.

Antibodies being the products of adaptive immunity, require Fc receptors to recruit essential elements of defense from innate immunity such as natural killer cells and macrophages, which enable phagocytosis of antigen-antibody complexes (Kindt *et* al, 2006). The crosslinking of FcR-bound antibodies can also generate immunoregulatory signals that induce cell activation, differentiation as well as the down-regulation of certain cellular responses. Fc receptor is often a part of a signal transduction complex that requires the participation of accessory polypeptide chains. An extracellular receptor association with intracellular signal transduction occurs in B lymphocytic receptors and is crucial in T lymphocytes. There is an entire suite of Fc receptors. The FcαR binds to IgA, the FcεR binds to IgE, the FcμR binds to IgM, the FcδR binds to IgD, and there are various Fcγ receptors that bind to IgG and its subclasses. The focus will be on the suite of Fcγ receptors, which is depicted in Figure 1.

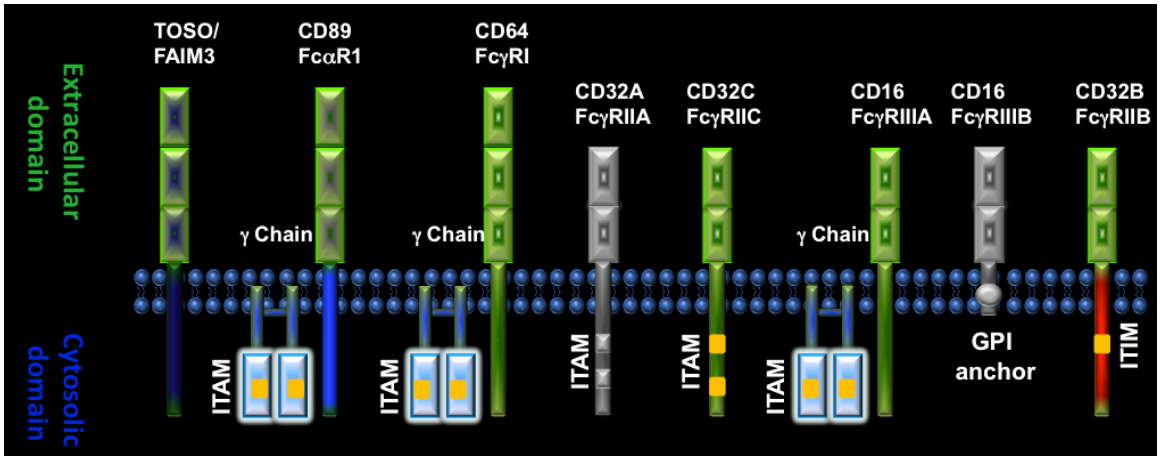


Figure 1. Suite of FC receptors and description of their expression location

Fcγ receptor 1 is a high-affinity receptor whereas FcγRIIA, FcγRIIC, FcγRIIIA and FcγRIIIB are a family of low-affinity receptors. FcγIIB is the only Fc receptor with inhibitory functions. For signaling, FcγRs rely on immunoreceptor tyrosine-based activation and inhibition motifs (ITAM/ITIM respectively). FcγRIIB is the only FcγRII with ITIM, all of the rest have an ITAM in the cytoplasmic domain of their alpha chains.

FcγRIIB, specifically, regulates immune and inflammatory responses. The suite of FcγRs have an alpha chain and two or three extracellular Ig-like domains, which bind to the Fc-domain of IgG. With exception to FcγRIIIB, the alpha chains on all of the receptors maintain single transmembrane and cytoplasmic domains. Variation in the gene, FCGR2B, encoding FcγRIIB have long been linked to autoimmune disease susceptibility, particularly SLE. Other FcRs of interest include: the FcαR1, a receptor for IgA that is present on myeloid cells in humans; the recently identified FcμR/TOSO, which binds IgM; and finally Fcα/μR, which binds IgA and IgM. Various immunoreceptors like FcγRI, FcγRIIIa and FcaRI associate with an adaptor protein called the common g-chain, which is necessary for signal transduction.

Genomic Perspective of FCGR gene evolution

FcγRIIA, FcγRIIB, and FcγRIIC are encoded by FCGR2A, FCGR2B, and FCGR2C, respectively (Figure 2). FCGR2A, FCGR2B, and FCGR2C share similarities within their genetic structure: there are two exons that code for a signal peptide followed by two exons that code for two extracellular Ig-like domains, then a single exon that codes for the transmembrane domain, and lastly two or three exons that code for the cytoplasmic domain (Figure 3).

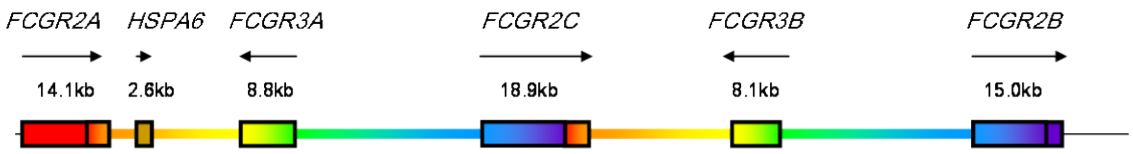


Figure 2: Structure of the 1q23 FCGR gene cluster

These three genes reside within a ~180kb gene cluster on 1q23, with FCGR2A and FCGR2B defining the ends of the gene cluster and FCGR2C falling within the middle (Figure 2). The boxes represent genes, the intersecting lines represent intergenic regions, and the regions of color correlation indicate ~98% sequence homology with exception to paralogs that are represented with identical colors. The arrows indicate the direction of transcription and gene size is written above each gene. FCGR2A, FCGR2B, and FCGR2C are transcribed in the same direction.

Note: From Travis Ptacek 2012. Reprinted with permission.

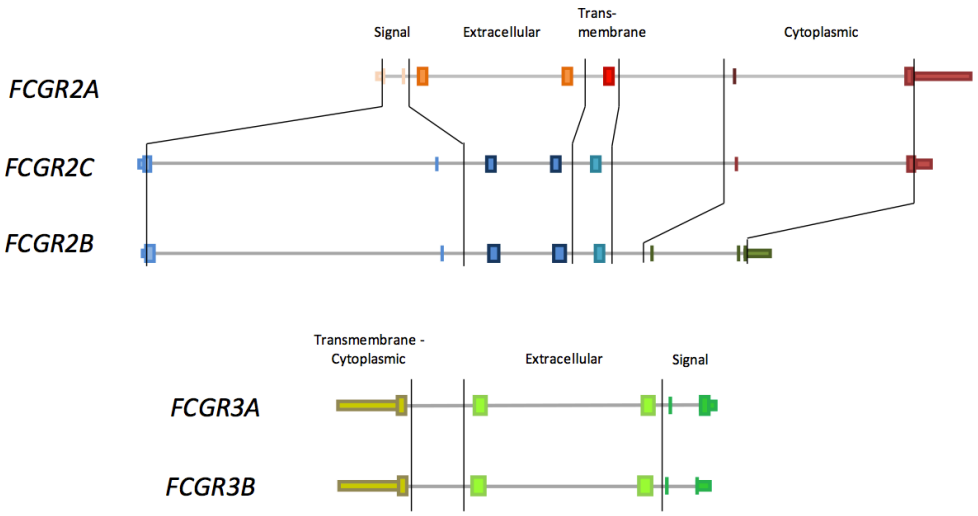


Figure 3: Structure of the Individual Genes within the 1q23 FCGR gene cluster

The boxes indicate exons, untranslated regions are represented by thin boxes, introns are represented by intersecting grey lines, and exons that code for specific domains are indicated by boundaries and labels. Colors indicate paralogous regions. The first and second halves of the 1q23 FCGR gene cluster are segmental duplications with ~98% similarity among the interval regions. As shown in the figure, FCGR -2A and -2C contain two exons for their cytoplasmic domains whereas FCGR2B maintains 3.

Note: From Travis Ptacek 2012. Reprinted with permission.

There are actually two existing cytoplasmic domains for FCGR2B. One variant has all three of the exons depicted and is called 2B-1, whereas the second is a splice variant that includes the last two cytoplasmic domain exons and is called 2B-2. SLE has been linked to all of the polymorphisms belonging to FcγRII genes. Moreover, among most people in the population, FCGR2C is a pseudogene that does not express a protein, but it does have an open reading frame allele (X13Q) that has a ~8-16% of frequency in the population. This allele initiates the expression of an activating receptor of B-cells and is associated with SLE. The FCGR3 genes, FCGR3A and FCGR3B, lay within the 1q23 gene cluster as well. However, the focus is particularly on FCGR2B.

Copy Number Variation

Prior to the full human genome sequence, a uniform manner of genetic inheritance was assumed. Many inherited genetic diseases have been associated with structural mutations of copy number variations in which DNA sequences of up to five million letters are either added or removed from an individual's genome (Conrad *et* al, 2010). Copy number variation (CNV) describes the occurrence of gene copy variability among individuals. In disease association studies, CNV usually refers to gene copy numbers that deviate from the most common copy number of that gene (Conrad *et* al, 2010). CNV of FCGR2B was investigated for its association with SLE. FCGRs are already associated with SLE, yet a novel investigation done by a colleague, Dr. Travis Ptacek investigated CNV within the 1q23 locus for any association with SLE. Alongside Dr. Ptacek, the identification of FCGR3B alleles associated with high SLE risk among certain ethnic races and a strong link between a FCGR2B variant and SLE onset were uncovered. These findings form a foundation for future studies of this gene cluster and its

role in autoimmunity. I redesigned the methodology in order to obtain the full-length product of the 180kb 1q23 gene cluster instead of two products (5' and 3') that were previously obtained. This will be elaborated further in the methods section of Chapter 2.

Copy Number Variation in FCGR and Autoimmunity

Previously, the 1q12 FCGR gene cluster was primarily studied for CNV and genetic abnormalities associated with SLE, which began in 1990 when one patient exhibited an abnormality on FCGR3B (Breunis *et* al, 2009). Now, researchers are aware of a plethora of CNV intervals within the FCGR gene cluster. The results of several studies reveal no apparent association between the CNV of FCGR2A and FCGR2B with SLE or RA (Breunis *et* al, 2009). Although breakpoint studies of FCGR2C and FCGR3B have demonstrated that FCGR CNV can emerge *de novo* during meiosis or somatically, and other studies show that FCGR CNV can be heritable (Breunis *et* al, 2009). A link with autoimmune diseases and CNV of FCGR2C has been studied but only in its role as a partner of FCGR3B. This could be due to its role as a pseudogene in the majority of the population. However, FcγRIIC proteins are associated with the immune response and its role in autoimmunity has yet to be investigated. Although many studies connect FCGR3B CNV with SLE, the actual mechanistic role it has on the onset of the disease is not yet known. All in all, the role of FCGR CNV in autoimmunity is not fully understood. To better understand the molecular mechanisms of autoimmunity, three approaches were taken, namely: (1) the identification of novel interacting partners of FCRs' cytoplasmic domains, (2) the finding of statistically overrepresented *cis*-regulatory elements in FCGR2, the only classical FCR and the identification of their cognate transcription factors, and (3) the identification of FCGR2B CNV (Copy number variation) in SLE

patients. These approaches are based in systems biology and strive to expand the molecular perspective of autoimmunity.

A Systems Biology lens on Immunity and Autoimmunity

Biological processes operate via precise interactions amongst thousands of molecules. Thus systems biology is a novel area of study for understanding disease pathologies as well as healthy genotypes at a genomic level. Systems biology strives to understand complex biological systems and their pathways that coordinate molecular events (Zhu *et* al, 2007). A systems perspective seeks to uncover the unpredictable and predictable intricacies of many different causal relations within diverse biological components. The recent invention of high-throughput technologies has been advantageous for studying macromolecular interactions as well as expression patterns. The yeast-two hybrid system is a high-throughput technology utilized to couple the first hypothesis of my thesis related to the identification of novel interacting partners of FCRs' cytoplasmic domains. Network biology, a branch of systems biology, translates the complexities of molecular interactions into a biological message.

Network Biology

In any eukaryotic cell, thousands of genes and their products orchestrate their transcriptional, translational, and metabolic activities to create cellular functions, phenotypic plasticity and organismal fecundity. Functional modules embedded within protein–DNA interactions, and protein–protein, and metabolite–substrate networks execute diverse cellular functions (Mitra *et* al., 2013). The dichotomous nature of network modules is beneficial to cells or organisms for adaptation to physiological perturbations, environmental cues, or pathological signals (Shmulevich *et* al, 2009). Any

given genotype has a sophisticated underlying network of macromolecular interactions that give rise to a phenotype. The idea behind systems biology is that cellular networks and biological systems are the bridges from genotype to phenotype (Carvunis *et* al, 2013). Typically in a network, physical and functional interactions between molecules are referred to as edges, and the molecules involved in the interactions are termed nodes. Nodes can correspond to nucleic acids, proteins, hormones, metabolites, or other macromolecules. Edges can be directed or undirected depending on the type of an interaction being illustrated in the graph (Seebacher *et* al, 2011).

Interactome Map: A Novel Network for Understanding Disease

An interactome map is a type of network that seeks to map a universe of molecular interactions that could portray key information towards understanding healthy and disease phenotypes. For this reason it is referred to as the gateway to systems biology. Interactions mediated by protein-protein, protein-RNA, protein-DNA, and protein-metabolite can for interactome networks. A major aim of systems biology is to create interactome maps as they facilitate studies of biological processes and systems. Static interactome maps can be generated through *in vitro* technologies and include many *in vivo* edges. On a node level, interactomes can facilitate the understanding of uncharacterized and well-known gene products, and at a systems level, interactomes provide insight into the properties of regulatory systems and global cellular networks as they relate to biological functions. Interactomes of human diseases have been created, and interestingly there is overlap between proteins among diseases with similar symptoms.

GETTING TO THE EDGE: PROTEIN DYNAMICAL NETWORKS AS A NEW FRONTIER IN PLANT-MICROBE INTERACTIONS

by

CASSANDRA C. GARBUTT, PURUSHOTHAM V. BANGALORE, PEGAH KANNAR, AND M.S. MUKHTAR

Frontiers in Plant Science

CHAPTER 1

GETTING TO THE EDGE: PROTEIN DYNAMICAL NETWORKS AS A NEW FRONTIER IN PLANT-MICROBE INTERACTIONS

Introduction

A systems perspective on diverse phenotypes, mechanisms of infection, and responses to environmental stresses can lead to considerable advances in agriculture and medicine. A significant promise of systems biology within plants is the development of disease-resistant crop varieties, which would maximize yield output for food, clothing, building materials, and biofuel production. A systems or "-omics" perspective frames the next frontier in the search for enhanced knowledge of plant network biology. The functional understanding of network structure and dynamics is vital to expanding our knowledge of how the intercellular communication processes are executed. This review article will systematically discuss various levels of organization of systems biology beginning with the building blocks termed "-omes" and ending with complex transcriptional and protein–protein interaction networks. The prevailing computational modeling approaches of biological regulatory network dynamics were highlighted. The latest developments in the "-omics" approach were discussed to underline and highlight novel technologies and research directions in plant network biology.

Systems Biology: A Paradigm Shift from Reductionism

Despite the progress of understanding phytopathogenic microbes and plant infectious

diseases, the arms race between hosts and pathogens fuels further scientific research (Boyd *et* al, 2013). Within the past decades, the molecular approaches to solve these crises entailed reductionism that seeks to explain a biological system through the summation of its isolated parts. While conceptual origins of systems biology date back almost 100 years, a shift from the reductionist approach to a more inclusive and integrative one started to occur at dawn of this millennium (Figure 1A; Arkin and Schaffer, 2011). This revolutionary, holistic approach is inspired by Aristotle's belief that "The whole is more than the sum of its parts." "Systems" has also been referenced as the "fifth fundamental requirement for Life" considering that biological structures and molecules never function in isolation, as is true for sociological structures (Carvunis *et* al, 2013). The limitations of reductionism with respect to medical science are widely recognized and systems biology offers a way of transcendence (Ahn *et* al, 2006). Extending this observation further, the tenets of systems biology certainly offer an alternative viewpoint for other biological research including plant biology. In addition, this holistic approach can be attributed to the scientific community's search for understanding the complexity and interconnectedness in a wide array of natural systems ranging from the microscale of a cell to the macroscale of socioecosystems. It has become strikingly evident that significant similarities exist at the structural organization levels among the extremes of these biological spectra (Keurentjes *et* al, 2011). Thus, systems biology yields models that analyze various changes in biological systems over time, and a systems perspective complements reductionism to facilitate innovative investigations and discoveries (Mitra *et* al., 2013). It also seeks to uncover the unpredictable and predictable intricacies of many different causal relations within diverse

biological components. Cumulatively, a systems approach to medical and agricultural research could guide new developments in techniques, knowledge, and ultimately therapeutics (Barabasi *et* al., 2011; Vidal *et* al., 2011).

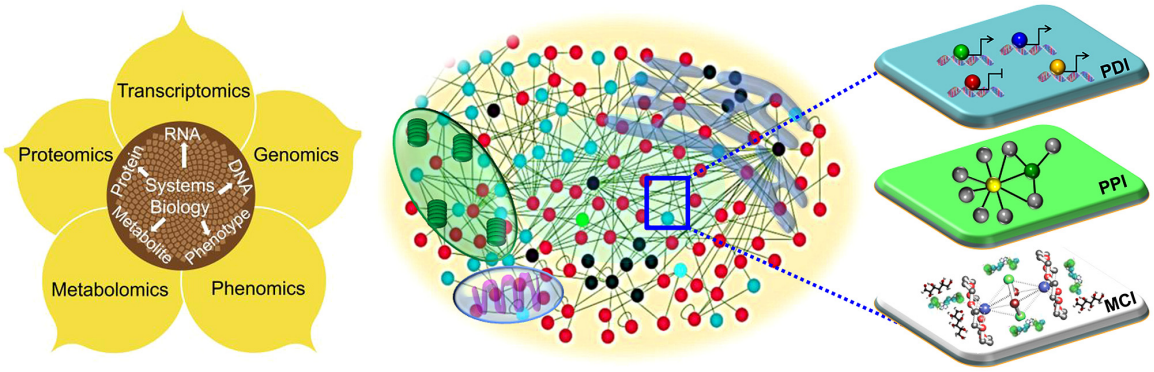


Figure 1. The systems biology approaches to understand plant immune systems.

(A) The diagrammatic overview of the integrative framework of multiple layers of "-omics" including genomics, transcriptomics, proteomics and phonemics. (B) Visualization of a cell as a complex web of macromolecular interactions that constitutes an "interactome." Functional modules, such as transcriptional (protein–DNA interactions; PDI), translational (protein–protein interactions; PPI) and metabolic (metabolite–compound interactions; MCI) are illustrated.

Here, various elements of systems biology were highlighted beginning with the level of the "-omes" (Figure 1A) and then elaborated on the scale of macromolecules and their interactions. The application and translation of any discovery teems with possibilities from improvements in medicinal therapeutics and plant biology to improvements in crop yield, quality, and pathogen resistance.

Network Biology: A useful tool in understanding Plant Immunity

In any eukaryotic cell, thousands of genes and their products orchestrate their transcriptional, translational, and metabolic activities to create cellular functions, phenotypic plasticity and organismal fecundity. Functional modules embedded within protein–DNA interactions, and protein–protein, and metabolite–substrate networks execute diverse cellular functions (Figure 1B; Mitra *et* al., 2013). The dichotomous (deterministic or stochastic) nature of network modules is beneficial to cells or organisms for adaptation to physiological perturbations, environmental cues, or pathological signals (Shmulevich and Aitchison, 2009). On the contrary, pathogens have evolved a suite of virulence proteins (effector molecules) that perturb the intracellular networks of their hosts to cause infection (Mukhtar, 2013). As with any host–pathogen conflict, plants and their pathogens are in an evolutionary "arms race," in which the host mounts defenses, the pathogen develops new strategies to thwart the defensive mechanisms, which in turn forces the host to adapt (Mukhtar *et* al., 2011; Pajerowska-Mukhtar *et* al., 2013). Network-based analysis is a holistic approach that can enable a detailed understanding of the relationships between phytopathogens and plants (Pritchard and Birch, 2011).

Network biology, a branch of systems biology, translates the complexities of molecular interactions into a biological message. Any given genotype has a sophisticated

underlying network of macromolecular interactions that give rise to a phenotype. The idea behind systems biology is that cellular networks and biological systems are the bridges from genotype to phenotype (Carvunis *et* al., 2013). Typically in a network, physical and functional interactions between molecules are referred to as edges, and the molecules involved in the interactions are termed nodes. Nodes can correspond to nucleic acids, proteins, hormones, metabolites, or other macromolecules. Edges can be directed or undirected depending on the type of an interaction being illustrated in the graph (Figure 2A, Seebacher and Gavin, 2011). Computational biologists and mathematicians have developed numerous algorithms to analyze the versatile relationships of nodes and understand the cellular organization of communication for a particular network. Research efforts to uncover potential universal laws that govern cellular networks are underway (Carvunis *et* al., 2013).

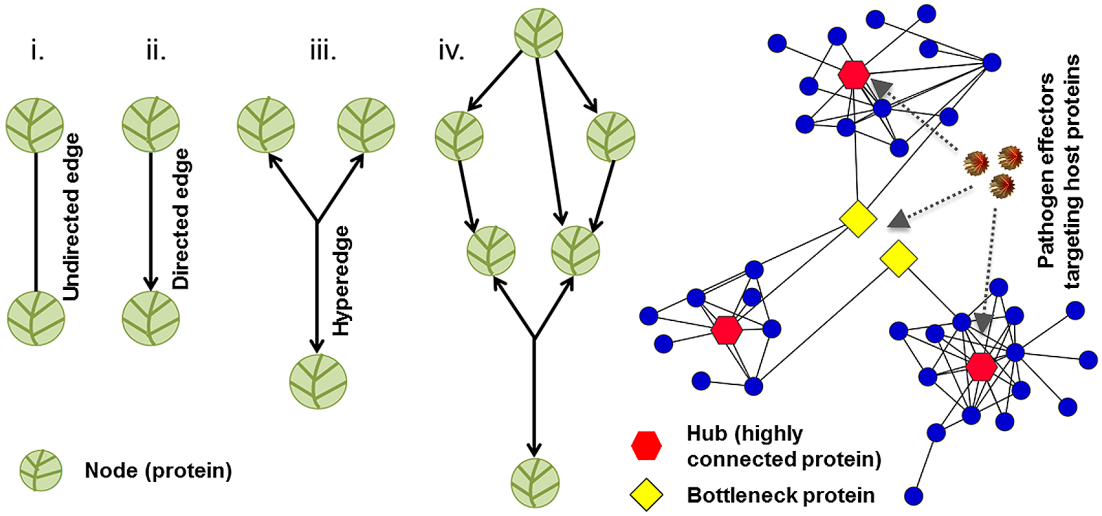


Figure 2. Network structure and topology.

(A) The organization of nodes and edges in a graph represents network structure. Vertices and links represent nodes and edges, respectively. Two nodes can be connected by undirected or directed edges. **(B)** A sub-network of plant–pathogen interactions is demonstrated. Hubs (highly connected proteins), bottlenecks (high betweenness nodes), and pathogen effectors (virulence factors) are depicted in red, yellow, and brown colors, respectively. Network with scale-free topology might be vulnerable to pathogen-mediated perturbations.

Dynamics of Transcriptional Regulatory Networks in Plant Defense

Upon pathogen recognition, the plant cell undergoes an extensive transcriptional reprogramming in a highly dynamic and temporally regulated manner. Stimulation of these plant defenses involves complex signal transduction networks incorporating feedback and cross-talk controlled by largely unknown mechanisms (Mukhtar *et* al., 2009). While transcriptomics has already uncovered hundreds of pathogen-responsive genes and key regulatory nodes, a large-scale immune transcriptional regulatory network is yet to be generated. In building a transcriptome, yeast-one-hybrid, and chromatin immunoprecipitation assays are especially useful as they can differentiate between indirect and direct gene regulation (Vidal *et* al., 2011). Experimental large-scale and static cellular networks give insight into biological systems at certain times and conditions. This data is often combined with protein localization data, protein–protein interactions (PPIs) and other temporal expression data (Uzoma and Zhu, 2013). Network dynamics modeling enables changes in transcriptional networks, interactomes, and signaling pathways to connect genotypic changes with plant defenses and disease phenotypes. Furthermore, immune-related subnetworks or modules help decode the complexities within biological systems (Riccione *et* al., 2012). By analyzing network dynamics from an evolutionary perspective, a phylogenetic relationship among molecules can also be identified (Morange, 2013; Soyer and O'Malley, 2013). Here, the number of existing computational tools and algorithms that can be exploited to predict, model and determine the dynamics of plant immune regulatory networks were explored.

Static plant transcriptional immune networks are usually inferred using both linear

and non-linear correlations as well as non-linear dimensionality reduction method (RepEdLEGG; Ernst *et* al., 2007; Sato et al., 2010). RepEdLEGG was employed on transcriptomic data obtained from diverse *Arabidopsis* immune mutants to model a static immune signaling network (Sato *et* al., 2010). Dynamic regulatory events miner (DREM) utilizes an input–output hidden Markov model and gene expression time series data to construct dynamic regulatory networks (Schulz *et* al., 2012). Recently, DREM was used to analyze ethylene transcriptional response in context of dynamic EIN3 binding data (Chang *et* al., 2013). Signaling and DREM (SDREM) extends DREM to address perturbation in the regulatory networks (Gitter *et* al., 2013). However, modeling dynamic interactions of the genes and generating meaningful perturbations requires a more expanded framework that must incorporate dynamic data as well as any environmental dependencies. To consider solutions available to address dynamic network perturbations, Shannon's mutual information was used to model dynamic relationships of genes and show that both linear and non-linear models could be incorporated while integrating dynamic and environment-dependent complexities of gene expression (Wang *et* al., 2013). NEXCADE is another interactive network perturbation program, that uses a graph theoretical approach and simulates single, multiple, and sequential perturbations (Yadav and Babu, 2012). However, it remains to be determined whether these algorithms are effective in modeling plant immune system network. Furthermore, standardized qualitative dynamical modeling suite (SQUAD) uses a binary decision diagram algorithm to identify all the stable steady states and then applies a qualitative dynamical systems approach to solve the resulting continuous dynamic system (Di Cara *et* al., 2007). Noteworthy, Naseem *et* al. (2012) recently utilized SQUAD to perform dynamic

modeling of the plant hormonal signaling network.

Computational cost is another essential aspect in modeling large-scale dynamic regulatory networks. Such cost can be drastically reduced by using software equipped with a deterministic model along with a heuristic algorithm, such as NetGenerator V2.0 (Weber *et* al., 2013). Collectively, in the light of aforementioned bioinformatics tools, an expanded computational framework is needed that incorporates expression data with multiple timescales, cellular compartments, host proteins–pathogen effector interactions and other environmental dependencies to model plant–pathogen interactions networks.

Host-Pathogen Protein-Protein Interaction Networks

Complementary proteomics analyses are essential to understand global virulence effects caused by pathogens' effector-mediated perturbations of the key nodes in the plant immune system. The first plant–pathogen interaction network-1 (PPIN-1) was constructed using effectors from two pathogens spanning the eukaryote–eubacteria divergence and three classes of Arabidopsis immune system proteins (Mukhtar *et* al., 2011). The resulting network contains 3,148 interactions among 926 proteins. The PPIN-1 also identified 165 effector-interacting proteins (effector targets), compared to only approximately 20 described previously. While a stringent yeast-two-hybrid (Y2H) system was employed for the above analyses, the common limitations of this heterologous system may still apply concerning both the false-positive and false-negative discovery rates. In addition, PPIN-1 revealed that pathogen effectors target highly interconnected host machinery to suppress effective host defenses and promote pathogen fitness (Figure 2B; Mukhtar *et* al., 2011). Several network biology hypotheses/premises have been

developed through interactome mapping. The centrality–lethality rule and local impact hypothesis are two examples that have been applied to human diseases (Arabidopsis Interactome Mapping Consortium, 2011; Barabasi *et* al., 2011; Gulbahce *et* al., 2012). According to the centrality–lethality rule, nodes that are central to many connections have the potential of dismantling the entire system if disabled, such as through a viral attack. The local impact hypothesis states that "products of disease susceptibility genes should reside in the network vicinity of the corresponding viral targets" (Gulbahce *et* al., 2012). For this study, the host interactome was developed by integrating different data sources. Epstein–Barr virus and human papillomavirus strains were selected to explore mechanisms of virally implicated diseases. These strains were found to target host proteins that were in proximity to other proteins associated with viral diseases as evidenced by significant shift in gene expression levels in corresponding disease implicated tissues (Gulbahce *et* al., 2012). Viral "neighborhoods" existed in the host interactome and were labeled as "viral disease networks." Similar demonstrations or contradictions need to be studied in the plant kingdom. Considering network components, hubs (highly connected proteins) and edges play an integral role in human and Arabidopsis immune systems. In humans, understanding the role of hub proteins requires differentiating between disease-related genes and essential genes. Given that human hubs frequently correspond to disease-related proteins (Barabasi *et* al., 2011), it would be interesting to extend this theory to plants for testing and verification in diverse natural populations. While several high-throughput technologies have been applied in plants, there still exist hundreds of thousands of unconstructed plant cell PPIs. Construction and access to reference PPIs can be achieved through computational and predictive methods.

Predicative capabilities are based on a wide range of protein and interactome characteristics (Fukunishi and Nakamura, 2008; Lee *et* al., 2010). The Protein Data Bank serves as a reference for proteins' three-dimensional structures and protein complexes (Velankar and Kleywegt, 2011; Velankar *et* al., 2012; Gutmanas *et* al., 2014). As a predictive tool, the Protein Data Bank is a methodological starting point for exploring experimentally determined protein interfaces, emphasizing particular features that can be used to predict domain–domain interactions in proteomes (Braun *et* al., 2013). Protein docking software provides another category of methods to infer protein–protein binding domains and interaction sites (Cai *et* al., 2013; Pencheva *et* al., 2013).

Comparing Human and Plant Immunity Through Differential and Three Dimensional Networks

While the development of comprehensive reference maps is one of the current challenges in the field, the future of omics-based research will integrate biological insights into networks to drive translational research. Creating comprehensive reference network maps is the first step toward developing dynamic, information-rich resources. To assist in these efforts, standardized experimental benchmarking and validation assays provide a mechanism to estimate the size and validity of the existing networks (Braun *et* al., 2013). In contrast to the highly dynamic and fluctuating endogenous conditions, under which biological systems normally operate, most physical interactome maps are developed from experiments conducted under static conditions. Differential network mapping takes into account the dynamic state to produce a cell-type and condition-specific interactome (Carvunis *et* al., 2013). Due to this characteristic, differential network mapping provides a more accurate description of the molecular and cellular

mechanisms within a living system. Although network biology provides a platform for decoding complexity, collapsing of networks to nodes and edges may lead to a significant loss of data. Current two-dimensional interactome maps do not consider either structure or conformation of the individual proteins within a network and ignore the spatial limitations of protein interactions. Because protein structure and function are highly interwoven, three-dimensional interactome maps that account for protein structure, interfaces, and even isoforms can greatly enhance the level of understanding of *in vivo* PPIs (Stein *et* al., 2011; Wang *et* al., 2012; Zhang *et* al., 2012).

Two-dimensional protein networks can be reconstructed with a third dimension to integrate protein structure, conformation, and spatial limitations. Atomic-level protein structure information was resolved for several large-scale human PPI networks to create the third dimension of analysis (Das *et* al., 2014a, b). Previously, a three-dimensional reconstruction of protein networks was conducted to elucidate the genetic and molecular mechanisms underlying human diseases; this investigation primarily focused on gene pleiotropy and locus heterogeneity (Wang *et* al., 2012). This type of network construction could also be applied to plant network maps in an effort to better understand plant disease and genotype–phenotype complexity. To assist with the creation of 3D interactome networks, the first iteration of interactome networks with structural information (INstruct; a database that houses current high quality, three-dimensional PPI networks that are structurally resolved to the atomic level) was built using several model organisms (Meyer *et* al., 2013). INstruct includes 37 Schizosaccharomyces pombe, 1273 Saccharomyces cerevisiae, 119 Mus musculus, 166 Drosophila melanogaster, 120 Caenorhabditis elegans, 644 Arabidopsis thaliana, and 6585 human interactions.

Node and Edgetic Investigations within Pathogen-Host Systems

Network components and topological properties provide novel avenues of investigation. Phenotypic variations due to total loss of a gene product (node-removal) emphasize the importance of node-centered investigations. Network topological properties of a node can be investigated to determine key proteins that are central to many interactions (Barzel and Barabasi, 2013). One topological property involves the degree of a node, which describes the number of edges a node has within a network. Hubs are central and critical to many edges within a network. In a scale-free network, most nodes possess few connections to other nodes while a handful of hubs essentially form the foundation of the network. This characteristic of scale-free networks is incorporated in PPI and metabolic network maps developed for organisms ranging from yeast to humans (Vidal *et* al., 2011). Recently, several independent studies confirmed the importance of hub proteins in pathogen virulence mechanisms. The results indicate that diverse pathogen proteins (spanning across viruses, bacteria, and fungi) target hub proteins in both humans and plants (Vidal *et* al., 2011; Braun *et* al., 2013). Thus far, two different categories of network hubs have been identified. Party and date hubs differ by their number of edges and the conditions that enable the interaction (Vidal *et* al., 2011). Party hubs are known for maintaining connections with all of their partners in all tested conditions. Date hubs tend to interact with different partners based on specific conditions. Node-removal can affect inter- and intra-network hub proteins or ensue on the periphery of a community of proteins. Other phenotypic variants can arise from edgetic perturbations (removal of a specific edge). In an edgetic disruption, a targeted interaction is disrupted while all other interactions (edges) remain unaffected. The consequence of

node-removal on the structure of the network might be greater because removing a node impacts more than one specific interaction (Zhong *et* al., 2009). Conversely, edgetic perturbations produce less significant network structure changes. At the molecular level, edgetic disruptions are characterized by in-frame point mutations that cause single amino acid substitutions and minute insertions, whereas truncating mutations and deletions reflect node-removal mechanisms. Given that about half of the ~50,000 known human diseases could be linked to edgetic disruptions (Zhong *et* al., 2009; Vidal *et* al., 2011), a similar application of the edgetic hypothesis to the plant kingdom can potentially shed light on disease and abiotic stress responses, yielding tools for crop improvement.

New methods such as forward and reverse edgetics aid in the analysis of phenotypic variation due to disturbances in specific molecular interactions (Charloteaux *et* al., 2011). Forward and reverse edgetics are complementary strategies of phenotypic investigation. Forward edgetics takes a mutated gene associated with a specific phenotype and uses Y2H to establish the interaction disruption. Reverse edgetics begins with a protein of interest and its corresponding set of interactions. Using reverse Y2H screens, reverse edgetics concerns the systematic separation of edgetic alleles that code for a protein defect (Charloteaux *et* al., 2011). These novel methods can differentiate between edgetic disruptions and node-removal mechanisms of phenotypic changes and pathogen infections. Edge direction is also essential to biological signaling systems/mechanisms and recent technology allows for the development of experimental methods to measure edgetic properties (Barzel and Barabasi, 2013). Continued exploration/experimentation will produce interactome network models on the proteome level that can integrate properties of edge strength, direction, and dynamics. Future

interactome maps will combine weighted and animated edgetic information (Carvunis *et* al., 2013). Clearly, a database of all possible protein interactions for each species will be the next milestone in systems research.

Conclusion

In summary, emerging technologies, resources, and research offer new opportunities to investigate unchartered territories in plant biology. Current interactome maps primarily reflect static states of time, internal conditions, and external influences. As such, today's interactome maps should be utilized as a scaffold to model in vivo conditions by coalescing other layers of functional "-omic" data, including: genomics, phenomics, transcriptomics, metabolomics, and epigenomics. Integrating diverse plant "-omics" data enables researchers to investigate and address plant processes and responses, such as development, signal transduction pathways, RNA processing, protein modifications, cell cycle, and plant immune responses. A global understanding of plant stress and disease responses and phenotypic diversity will promote investigations of network topological properties. Computational tools, databases, and other systems resources will continue to grow and facilitate functional analysis and integration of multiple heterogeneous data sources. This may lead to improvements in environmental resilience, pathogen resistance, and overall crop production.

Author Contributions

Cassandra C. Garbutt, Purushotham V. Bangalore, Pegah Kannar, and M. S. Mukhtar wrote the manuscript. Cassandra C. Garbutt and M. S. Mukhtar prepared the figures. M. S. Mukhtar prepared the manuscript outline and coordinated the contributions from all co-authors. Published online June 30, 2014.

Acknowledgments

The authors apologize to all authors whose work was not included within this review due to space restrictions. M. S. Mukhtar was supported through funds from the Department of Biology, The University of Alabama at Birmingham.

REFERENCES

Ahn, A. C., Tewari, M., Poon, C. S., and Phillips, R. S. (2006). The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med.* 3:e208. doi: 10.1371/journal.pmed.0030208

Arabidopsis Interactome Mapping Consortium. (2011). Evidence for network evolution in an *Arabidopsis* interactome map. *Science* 333, 601–607. doi: 10.1126/science.1203877

Arkin, A. P., and Schaffer, D. V. (2011). Network news: innovations in 21st century systems biology. *Cell* 144, 844–849. doi: 10.1016/j.cell.2011.03.008

Barabasi, A. L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918

Barzel, B., and Barabasi, A. L. (2013). Universality in network dynamics. *Nat. Phys.* 9, 673–681. doi: 10.1038/nphys2741

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., and Leung, H. (2013). Plant–pathogen interactions: disease resistance in modern agriculture. *Trends Genet.* 29, 233–240. doi: 10.1016/j.tig.2012.10.011

Braun, P., Aubourg, S., Van Leene, J., De Jaeger, G., and Lurin, C. (2013). Plant protein interactomes. *Annu. Rev. Plant Biol.* 64, 161–187. doi: 10.1146/annurev-arplant-050312-120140

Cai, B., Wang, H., Zheng, H., and Wang, H. (2013). Integrating domain similarity to improve protein complexes identification in TAP-MS data. *Proteome Sci.* 11:S2. doi: 10.1186/1477-5956-11-S1-S2

Carvunis, A. R., Roth, F. P., Calderwood, M. A., Cusick, M. E., Superti-Furga, G., and Vidal, M. (2013). "Interactome networks," in *Handbook of Systems Biology Concepts and Insights*, Chap.3, eds M. Walhout, M. Vidal, and J. Dekker (Amsterdam: Elsevier Inc.), 45–60. doi: 10.1016/B978-0-12-385944-0.00003-4

Chang, K. N., Zhong, S., Weirauch, M. T., Hon, G., Pelizzola, M., Li, H., et al. (2013). Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in *Arabidopsis*. *Elife* 2:e00675. doi: 10.7554/eLife.00675

Charloteaux, B., Zhong, Q., Dreze, M., Cusick, M. E., Hill, D. E., and Vidal, M. (2011). Protein–protein interactions and networks: forward and reverse edgetics. *Methods Mol.*

Biol. 759, 197–213. doi: 10.1007/978-1-61779-173-4_12

Das, J., Fragoza, R., Lee, H. R., Cordero, N. A., Guo, Y., Meyer, M. J., et al. (2014a). Exploring mechanisms of human disease through structurally resolved protein interactome networks. *Mol. Biosyst.* 10, 9–17. doi: 10.1039/c3mb70225a

Das, J., Lee, H. R., Sagar, A., Fragoza, R., Liang, J., Wei, X., et al. (2014b). Elucidating common structural features of human pathogenic variations using large-scale atomic-resolution protein networks. *Hum. Mutat.* 35, 585–593. doi: 10.1002/humu.22534

Di Cara, A., Garg, A., De Micheli, G., Xenarios, I., and Mendoza, L. (2007). Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics* 8:462. doi: 10.1186/1471-2105-8-462

Ernst, J., Vainas, O., Harbison, C. T., Simon, I., and Bar-Joseph, Z. (2007). Reconstructing dynamic regulatory maps. *Mol. Syst. Biol.* 3:74. doi: 10.1038/msb4100115

Fukunishi, Y., and Nakamura, H. (2008). Prediction of protein–ligand complex structure by docking software guided by other complex structures. *J. Mol. Graph. Model.* 26, 1030–1033. doi: 10.1016/j.jmgm.2007.07.001

Gitter, A., Carmi, M., Barkai, N., and Bar-Joseph, Z. (2013). Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome Res.* 23, 365–376. doi: 10.1101/gr.138628.112

Gulbahce, N., Yan, H., Dricot, A., Padi, M., Byrdsong, D., Franchi, R., et al. (2012). Viral perturbations of host networks reflect disease etiology. *PLoS Comput. Biol.* 8:e1002531. doi: 10.1371/journal.pcbi.1002531

Gutmanas, A., Alhroub, Y., Battle, G. M., Berrisford, J. M., Bochet, E., Conroy, M. J., et al. (2014). PDBe: protein data bank in Europe. *Nucleic Acids Res.* 42, D285–D291. doi: 10.1093/nar/gkt1180

Keurentjes, J. J., Angenent, G. C., Dicke, M., Dos Santos, V. A., Molenaar, J., Van Der Putten, W. H., et al. (2011). Redefining plant systems biology: from cell to ecosystem. *Trends Plant Sci.* 16, 183–190. doi: 10.1016/j.tplants.2010.12.002

Lee, K., Thorneycroft, D., Achuthan, P., Hermjakob, H., and Ideker, T. (2010). Mapping plant interactomes using literature curated and predicted protein–protein interaction data sets. *Plant Cell* 22, 997–1005. doi: 10.1105/tpc.109.072736

Meyer, M. J., Das, J., Wang, X., and Yu, H. (2013). INstruct: a database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics* 29, 1577–1579. doi: 10.1093/bioinformatics/btt181

Mitra, K., Carvunis, A. R., Ramesh, S. K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* 14, 719–732. doi: 10.1038/nrg3552

Morange, M. (2013). From experimental systems to evolutionary biology: an impossible journey? *Hist. Philos. Life Sci.* 35, 27–32.

Mukhtar, M. S. (2013). Engineering NLR immune receptors for broad-spectrum disease resistance. *Trends Plant Sci.* 18, 469–472. doi: 10.1016/j.tplants.2013.08.005

Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333, 596–601. doi: 10.1126/science.1203659

Mukhtar, M. S., Nishimura, M. T., and Dangl, J. (2009). NPR1 in plant defense: it's not over 'til it's turned over. *Cell* 137, 804–806. doi: 10.1016/j.cell.2009.05.010

Naseem, M., Philippi, N., Hussain, A., Wangorsch, G., Ahmed, N., and Dandekar, T. (2012). Integrated systems view on networking by hormones in *Arabidopsis* immunity reveals multiple crosstalk for cytokinin. *Plant Cell* 24, 1793–1814. doi: 10.1105/tpc.112.098335

Pajerowska-Mukhtar, K. M., Emerine, D. K., and Mukhtar, M. S. (2013). Tell me more: roles of NPRs in plant immunity. *Trends Plant Sci.* 18, 402–411. doi: 10.1016/j.tplants.2013.04.004

Pencheva, T., Jereva, D., Miteva, M. A., and Pajeva, I. (2013). Post-docking optimization and analysis of protein–ligand interactions of estrogen receptor alpha using AMMOS software. *Curr. Comput. Aided Drug Des.* 9, 83–94.

Pritchard, L., and Birch, P. (2011). A systems biology perspective on plant–microbe interactions: biochemical and structural targets of pathogen effectors. *Plant Sci.* 180, 584–603. doi: 10.1016/j.plantsci.2010.12.008

Riccione, K. A., Smith, R. P., Lee, A. J., and You, L. (2012). A synthetic biology approach to understanding cellular information processing. *ACS Synth. Biol.* 1, 389–402. doi: 10.1021/sb300044r

Sato, M., Tsuda, K., Wang, L., Coller, J., Watanabe, Y., Glazebrook, J., et al. (2010). Network modeling reveals prevalent negative regulatory relationships between signaling sectors in *Arabidopsis* immune signaling. *PLoS Pathog.* 6:e1001011. doi: 10.1371/journal.ppat.1001011

Schulz, M. H., Devanny, W. E., Gitter, A., Zhong, S., Ernst, J., and Bar-Joseph, Z. (2012). DREM 2.0: improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst. Biol.* 6:104. doi: 10.1186/1752-0509-6-104

Seebacher, J., and Gavin, A. C. (2011). SnapShot: protein–protein interaction networks. *Cell* 144, 1000:1001.e1. doi: 10.1016/j.cell.2011.02.025

Shmulevich, I., and Aitchison, J. D. (2009). Deterministic and stochastic models of genetic regulatory networks. *Methods Enzymol.* 467, 335–356. doi: 10.1016/S0076-6879(09)67013-67010

Soyer, O. S., and O'Malley, M. A. (2013). Evolutionary systems biology: what it is and why it matters. *Bioessays* 35, 696–705. doi: 10.1002/bies.201300029

Stein, A., Ceol, A., and Aloy, P. (2011). 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 39, D718–D723. doi: 10.1093/nar/gkq962

Uzoma, I., and Zhu, H. (2013). Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. *Genomics Proteomics Bioinformatics* 11, 18–28. doi: 10.1016/j.gpb.2012.12.005

Velankar, S., Alhroub, Y., Best, C., Caboche, S., Conroy, M. J., Dana, J. M., et al. (2012). PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* 40, D445–D452. doi: 10.1093/nar/gkr998

Velankar, S., and Kleywegt, G. J. (2011). The Protein Data Bank in Europe (PDBe): bringing structure to biology. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 324–330. doi: 10.1107/S090744491004117X

Vidal, M., Cusick, M. E., and Barabasi, A. L. (2011). Interactome networks and human disease. *Cell* 144, 986–998. doi: 10.1016/j.cell.2011.02.016

Wang, J., Chen, B., Wang, Y., Wang, N., Garbey, M., Tran-Son-Tay, R., et al. (2013). Reconstructing regulatory networks from the dynamic plasticity of gene expression by mutual information. *Nucleic Acids Res.* 41:e97. doi: 10.1093/nar/gkt147

Wang, X., Wei, X., Thijssen, B., Das, J., Lipkin, S. M., and Yu, H. (2012). Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* 30, 159–164. doi: 10.1038/nbt.2106

Weber, M., Henkel, S. G., Vlaic, S., Guthke, R., Van Zoelen, E. J., and Driesch, D. (2013). Inference of dynamical gene-regulatory networks based on time-resolved multi-stimuli multi-experiment data applying NetGenerator V2.0. *BMC Syst. Biol.* 7:1. doi: 10.1186/1752-0509-7-1

Yadav, G., and Babu, S. (2012). NEXCADE: perturbation analysis for complex networks. *PLoS ONE* 7:e41827. doi: 10.1371/journal.pone.0041827

Zhang, Q. C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C. A., et al. (2012). Structure-based prediction of protein–protein interactions on a genome-wide scale. *Nature* 490, 556–560. doi: 10.1038/nature11503

Zhong, Q., Simonis, N., Li, Q. R., Charloteaux, B., Heuze, F., Klitgord, N., et al. (2009). Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* 5:321. doi: 10.1038/msb.2009.80

CHAPTER 2

UNDERSTANDING FC RECEPTROS THROUGH THEIR TRANSCRIPTIONAL REGULATION AND COPY NUMBER VARIATION

Introduction

The immunoglobulin-binding molecules termed fragment crystallizable receptors (FCRs) are critical for linking antigen-specific recognition to effector cells, which facilitates antibody performance. FCRs are also implicated in the pathogenesis of various autoimmune diseases including diverse rheumatic diseases, such as SLE and RA. Painful swelling in joints or muscles are characteristic of rheumatic diseases. In contribution to the future improvement of existing diagnostic and therapeutic methods for rheumatic diseases, several molecular mechanisms of FCRs were studied. The primary FCRs of interest were the Fc gamma (Fcγ) suite, specifically the FcγIIB receptor. Fcγ receptors are known regulators of immune responses and are the FCRs for antibody Immunoglobulin G. Most of the FcγRs are activating receptors and include the high-affinity receptor FcγRI and a family of low affinity receptors, including FcγRIIa, FcγRIIc, FcγRIIIa and FcγRIIIb. Other FCRs of interest include: the FcαR1, a receptor for IgA that is present on myeloid cells in humans, the recently identified FcμR/TOSO, which binds IgM, and finally Fcα/μR, which binds IgA and IgM.

To understand the regulation of human FCRs in SLE and RA patients, three comprehensive approaches were undertaken and include: (1) the identification of novel

interacting partners of FCRs' cytoplasmic domains, (2) the finding of statistically overrepresented *cis*-regulatory elements in FCGR2, the only classical FCR and the identification of their cognate transcription factors, and (3) the identification of FCGR2B CNV (Copy number variation) in SLE patients. To identify the novel interacting partners of FCRs' cytoplasmic domains, the yeast-two hybrid system was applied. The yeast-two hybrid system enabled the pair-wise test of 48 clones of FCRs' cytoplasmic domains against ~14,000 full length human clones, provided by the Dana Farber Cancer Institute (DFCI) in Boston, Massachusetts. This extensive library from DFCI has been previously applied to understand the perturbed molecular pathways that cause cancer progression, neurological disorders, and type 2 diabetes (Vidal *et* al, 2013). After conducting the experiment twice at DFCI and once more at UAB as to enrich the number of interactions, the first human autoimmune disease network was generated. An interactome map inclusive of receptors associated with autoimmune diseases is a valuable tool for future SLE and RA diagnostic and treatment methods.

In the search of statistically overrepresented *cis*-regulatory elements in FCGR2 and the identification of their cognate transcription factors, Multiple Em for Motif Elicitation (MEME) and a promoter-bootstrapping program (POBO) were utilized. MEME is designed to find or predict motifs such as transcription factor binding sites and *cis*-regulatory elements, and POBO can screen and verify known cis-regulatory elements within a sequence. A *cis*-regulatory element is a region of DNA or RNA that regulates the expression of genes located on the same strand, and transcription factors are proteins that control gene expression. By finding overrepresented *cis*-regulatory elements, through MEME and POBO, the transcriptional binding factors that cause FcγIIB

expression, the only Fcγ inhibitory receptor, can be determined and thus mechanistically regulated for therapeutic treatment of some rheumatic diseases.

Dr. Travis Ptacek, a lab colleague, began the first novel study of the 1q23 FCGR gene cluster in association with SLE risk. He identified FCGR3B alleles associated with high SLE risk among certain races and a strong link between a FCGR2B variant and SLE onset. These findings form a foundation for future studies of this gene cluster and its role in autoimmunity. I redesigned the methodology in order to obtain the full-length product of the 180kb 1q23 gene cluster instead of two products (5' and 3') that were previously obtained. To identify the FCGR2B CNV (copy number variation) in SLE donor cohorts, the methodology was redesigned to yield a single full-length product through the application of SMART cDNA synthesis, 3' RACE PCR, a labor-intensive PCR Nesting performed 3 times, recombination based cloning, mini-prepping of 36 products per round, and then Sanger Sequencing. The full-length product contains the genes FCGR2A, FCGR2B, FCGR2C, FCGR3A and FCGR3B that are located within 180kb on 1q23. The full-length product can enable the discrimination between: FCGR2B and FCGR2C fused to FCGR2A, and at the 3' region, FCGR2B and FCGR2C. By understanding the CNV of various functional FCGR alleles, the efficacy of therapeutics can be determined.

Methods

Yeast-Two Hybrid System

Prior to the implementation of the yeast-two hybrid system, 48 clones coding for cytoplasmic domains of FCRs were generated. A receptor has extracytosolic, intramenbraneous and cytoplasmic domains. Although the extracytosolic domain binds with IgG in case of Fcγ, interactors within the cytoplasmic domain of the receptors is

expected. Clones were generated using the Gateway cloning and the LR reaction. A primer containing attB1 and attB2 sites was designed. PCR was done using a proof reading taq polymerase and these products were cloned into pDONR207 vector. The sequence was generated through the Heflin Genomic Center at UAB. Upon sequence confirmation, I performed an LR clonase reaction into a DB (bait) vector and an AD (prey) vector. For the yeast two-hybrid system, one strain contains a DNA-Binding domain, known as DB and the other contains an activation domain, known as AD. A known protein, called the bait, is fused to the DB, and an interacting protein, called the prey, is fused to the AD of a transcriptional activator. The mating of these two yeast strains enables the efficient combination of large collections of DNA constructs. When these yeast mate and their vectors join, the reporter gene is activated, which can be screened for on selective media containing HIS 3AT. Vectors have 3 functional regions: (1) an origin of replication (2) a drug-resistance gene (3) a region where DNA can be inserted without interfering with plasmid replication or expression of the drug-resistance gene.

There were a total of 48 clones, including the WT cytosolic domains, phospho-mimic/phospho-dead mutants (clones attained from Dr. Kimberly's lab that were used as DNA template for PCR amplifications), in addition to naturally occurring alleles. A pair-wise test of the 48 clones was done against ~14,000 full length human clones, provided by DFCI in Boston, Massachusetts.

To enrich the number of interactions, the yeast strains expressed the human full-length genes in both the forward direction (48 Fcγ receptor clones as baits) and reciprocal direction (Fcγ pool as prey) for a total of 96 clone-types. Prior to beginning, bait and prey

cloned vectors are transformed into their corresponding haploid Saccharomyces cerevisiae yeast strains: MATα/Y8930 for DB (bait vectors), and MATa/Y8800 for AD (prey vectors).

With a multichannel pipette, transformed DB yeast cells were spotted onto two agar plates of SD-Leucine, which were then grown for 1-2 nights at 30°C. The same was done for the AD yeast cells but with SD-Tryptophan. Afterwards, the DB and AD colonies were picked with a multichannel pipette and directly inoculated (per row) into 96-well plates filled with 120μl of Sc-Leu/Sc-Trp respectively and grown overnight at 30°C. Then, the cultures were mixed with a multichannel pipette before 3μl of culture per well were transferred into falcon tubes containing 7ml of Sc-Leu/Sc-Trp respectively. The caps were loosened and taped securely before being placed in a shaker at 30°C to grow for 1-2 nights.

Afterwards, the glycerol stocks of the cultures were prepared. Glycerol is required in a 1:1 ratio in order for the cells to survive in an environment of -80°C. In a solution basin, 7mL of each grown culture was mixed with 7mL of a 40% glycerol stock. With a multichannel pipette, 130μl-140μl of this glycerol/culture solution was aliquoted into 96-well costar plates then covered with aluminum tape to store in -80°C. This was repeated for all 48 baits and 48 preys.

For mating, 140μl of YEPD media is aliquoted into 96 96-well costar plates. Then, 5μl of the AD pool is pipetted into the 96 YEPD costar plates. Then 5μl of bait glycerol stock is pipetted into the properly labeled YEPD/AD pool costar plates. The mate-plates are covered with airport tape before being grown overnight at 30°C.

For diploid selection, 140μl of Sc-Leu-Trp media is aliquoted into 96 96-well costar

plates. The mated plates are then shaken before 5-7μl is aliquoted into the prepared 96-well Sc-Leu-Trp plates, which are covered with airport tape before being grown overnight at 30°C.

For spotting, the diploid selection plates are shaken before 4-5μl are spotted onto various selective media including: SD/-Leu-Trp-His and SD/-Leu-His with plates containing multiple concentrations of cyclohexamide and 3AT (explained in detail below). Then incubation at 30°C for 3-4 nights occurs. Afterwards, colony picking occurs using autoclaved toothpicks for the colonies with positive growth. At this point, colonies were assigned confidence scores, which categorize the strength of the interaction based upon the growth. These colonies undergo cell lysis and PCR. Cell Lysis is required to expose and obtain the vectors. The hybrid proteins (DB-X and AD-Y) are amplified directly from yeast transformants by PCR, and subsequently identified by Sanger sequencing.

Throughout the experiment, controls were always present. These controls consisted of a variety of genotypes and were a part of the collection at DFCI. I performed this pair-wise screen twice at DFCI and four times in Birmingham to enrich the number of interactions.

YEPD is a nonselective, enriched media that ensures growth for all cells and provides the proper environment to support ample interactions. YEPD allows for all mating to occur and the production of many daughter cells, regardless of their phenotype. Next, these diploid cells will be selected for positive vectors rather than the random ones. This is done by growing them on selective media grown overnight at 28°C. The positive interaction of a protein with the activating domain and the binding domain complex will

cause for the expression of a screening marker (Gal4). This gene can be selected for on media lacking His. Another Y2H condition applied to detect possible autoreactivity of the new vectors includes the spotting of Y8800/MATa AD encoding plasmid lacking an insert onto the corresponding selective media. The same should be done for the DB-X baits. Growth on selective medium identifies autoactivators.

Growth on SC-Leu-Trp-His+3AT media:

- Should theoretically yield all desired colonies (positive selection)
- Positives are screened for their ability to activate the His3 marker
- 3AT is an inhibitor of His and is added to reduce background growth.

Growth on SC-Leu-His+cyclohexamine+3AT media:

- Cyclohexamide is a part of the AD vector. And therefore, serves for negative selection because this media kills the prey.

By comparing these plates, the false positives can be identified. Autoreactivity will be elaborated in the discussion section.

Cytoscape to Visualize and Analyze Interaction Data

Data from the yeast-two-hybrid assay was uploaded as an excel document into Cytoscape 3.2.0. The integration of the data enabled the visualization of an autoimmune interactome network. A random network was generated using a cytoscape plug-in called "Randomnetworks" as well. This random network was analyzed in parallel with the autoimmune network to verify our findings. Cytoscape plug-ins were also applied for network analysis through the assignment of topological properties. The plug-ins include: "ShortestPath" to find the shortest paths between nodes, "ClustnSee" identifies clusters,

"CalculatenodeDegree" calculates degree distribution, and "CentiScaPe" calculates node centrality or betweenness.

FCGR Promoter analysis

One major question in systems biology attempts to answer how certain genes are controlled or regulated. In the case of this study, the regulation of the FCG2B gene is of primary interest. As previously described, MEME and POBO are tools utilized for the summation, verification and screening of predetermined *cis*-element motifs belonging to a set of sequences. When a particular sequence was considered to be statistically overrepresented, it was subject to a computer-based boot strapping experiment, which utilizes all transcription factors in human genome. CSL ("CBF-1, Suppressor of Hairless, Lag-2," after its mammalian, *Drosophila*, and *Caenorhabditis elegans* orthologs) is a DNA-binding transcription factor that binds to overregulated elements and NICD (intracellular domain of Notch) is a signaling protein. The NOTCH signal collaborates with CSL via NCID. NCID removes the repressor and binds the activator, thus leading to gene expression. The functions of these proteins can be analyzed through the use of online databanks.

Protocols for Investigating Copy Number Variation

A labor-intensive Nest PCR was applied 3 times as well as recombination based cloning, transformation, mini-prep of 36 products per round, sanger sequencing, and then sequence analysis. Later the 3' RACE PCR conditions were optimized to increase the mini-preps to 48 products per round.

RACE PCR Approach

The protocol for the RACE PCR approach was completely adapted from the SMART RACE cDNA Amplification Kit User Manual by Clontech (Clontech 2012) and is featured below; all credit is due to Clontech. This protocol is for generating RACE-ready cDNA.

1. Enough of the following Buffer mix was prepared for all of the 5'- & 3'-RACE-Ready cDNA synthesis reactions in addition to one extra reaction to ensure sufficient volume. For each 10 µl cDNA synthesis reaction, I mixed the following reagents and spun them briefly in a microcentrifuge, then set them aside at room temperature until Step 6:
    a. 2.0 µl of the 5X First-Strand Buffer
    b. 1.0 µl of DTT (20 mM)
    c. 1.0 µl of dNTP Mix (10 mM)
    d. This yields a total volume of 4.0 µl per reaction.
2. The following reagents were combined in a separate microcentrifuge tube:
    a. For preparation of 3'-RACE-ready cDNA
        i. 1.0-3.75 µl of RNA
        ii. 1.0 µl of 3'-CDS Primer A
3. Sterile water was added to the tubes from Step 2 for a final volume of 4.75 µl per reaction.
4. The contents were then mixed and spun the tubes briefly in a microcentrifuge.
5. Afterwards, the tubes were incubated at 72°C for 3 minutes, the cooled at 42°C for 2 minutes. After they were cooled, the tubes were spun briefly for 10 seconds at 14,000 g to collect the contents at the bottom.

6. The following master mix was prepared for the 3'-RACE-ready cDNA synthesis reactions. The following reagents were mixed at room temperature in the following order:
    a. 4.0 µl of the Buffer Mix from Step 1
    b. 0.25 µl of the RNase Inhibitor (40 U/µl)
    c. 1.0 µl of the SMARTScribe Reverse Transcriptase (100 U)
    d. This yields a total volume of 5.25 µl per reaction.
7. 5.25 µl of the Master Mix from Step 6 was added to the denatured RNA from Step 5 for a total volume of 10 µl.
8. The contents were mixed in the tubes by gentle pipetting, and then spun briefly to collect the contents at the bottom.
9. The tubes were incubated at 42°C for 90 minutes in a hot-lid thermal cycler.
10. The tubes were then heated at 70°C for 10 minutes.
11. The samples were stored at -20°C for up to three months. In this case, they were applied to Nest PCR.

Nested PCR Approach

Nest PCR is a type of PCR that gives way to more specific product by reducing non-specific binding of products that are from the amplification of off-target primer binding sites. Nested PCR also has 2 sets of primers, the second amplifies a target on the first round of product. The PCR was performed using FCGR2A and FCGR2B/C specific primers. The product from the first round of PCR was used for the second round of PCR, and the product of the second round of PCR was used for the third round.

Table 1. The first round of PCR:

Duration	Temperature (°C)
5 minutes	94
15 seconds	94
22 seconds	65
2 minutes	72
7 minutes	72
Infinity	4

What is in gray indicates that this stage was run 28 times before the infinity phase.

Table 2. Second and Third Rounds of PCR:

Duration	Temperature (°C)
5 minutes	94
15 seconds	94
15 seconds	65
2 minutes	72
Infinity	16

What is in gray indicates that this stage was run 50 times before the infinity phase

Recombination Based Cloning: BP Reaction

The following master mix was made (per reaction) for the BP Reaction, which would generate an entry clone.

1. 1.5 µl of the pDONR 207 vector
2. 3.5 µl of the PCR product
3. 2.0 µl of the BP clonase
4. 3.0 µl of autoclaved deionized water

5. The total volume is 10 µl per reaction.

Bacterial transformation

Transformations with competent *E. coli* cells were prepared with the BP reaction product. Below is the protocol:

1. One vial of chemically competent cells (about 1 mL) was thawed on ice. One box of appropriately sized pipette tips was placed in the freezer. The hot-water bath incubator was set to 42°C, and once at 42°C, a tube with a sufficient amount of LB broth was placed in the hot bath until used at Step 6.
2. In a separate tube, 1-5 µl of the BP reaction product was added to 50 µl of competent cells (per reaction). This was mixed gently without pipetting.
3. The vials were incubated on ice for 30 minutes.
4. The cells were then heat-shocked for 30 seconds in the hot-water bath at 42°C.
5. The vials were then removed from the hot-water bath and placed on ice for 2 minutes.
6. 250 µl of the pre-warmed LB broth was then added aseptically to the vials.
7. The vials were then shaken horizontally at 37°C for one hour at 225 rpm.
8. With glass beads, 20 to 200 µl of solution from each transformation was spread on a room-temperature selective (Gentamicin or Ampicillin) plate and incubated overnight at 37°C (the plates were inverted).
9. A select number of colonies were collected and then used to generate cultures that are then mini-prepped.

Bacterial growth cultures

1. 50 mL of LB medium was added to a falcon tube.

2. 50 µl of antibiotic (gent/amp) was added into the falcon tube from Step 1 and mixed.
3. 5 mL of the Step 2 mixture was aliquoted into individual appropriately labeled culture tubes.
4. A single colony from the plates of Step 8 from the transformation protocol detailed directly above, was selected with a pipette tip and then dropped into the culture tube.
5. The culture tubes were tapped securely to the incubator/shaker, which was set vigorously shaking at 37°C.
6. 36 cultures were prepared per plate.

Protocol for the Mini-Prep of 36 cultures per PCR product

The following protocol is from the Mini-Prep handbook from Qiagen.

1. The hot-water bath set at 42°C was prepared and sufficient Qiagen Buffer EB solution was added into a tube and placed in the bath to pre-warm for Step 11.
2. The majority of the culture fluid was discarded, and then 1.5 µl of the pelleted bacterial cells were added to a microcentrifuge tube. This tube was microcentrifuged for one minute. This step was repeated 3 times per culture.
3. Then, 250 µl of Solution 1 (kept at 4 °C) was added. RNase A was added to Buffer P1 previously.
4. 250 µl of Solution 2 was added and gently inverted 4–6 times.
5. 350 µl of Solution 3 was added and then the tubes were inverted immediately but gently 4–6 times.

6. Then, these tubes were centrifuged for 10 min at 13,000 rpm. A compact white pellet formed.
7. The supernatants were then transferred to the QIAprep spin column by pipetting.
8. These were then centrifuged for 30–60 seconds. The flow-through was then discarded.
9. The QIAprep spin column was washed by the addition of 0.75 ml Buffer PE before centrifuging for 30–60 seconds.
10. The flow-through was then discarded, before centrifuging for an additional 1 minute to remove any residual wash buffer.
11. The QIAprep column was then placed in a clean 1.5 ml microcentrifuge tube. 50 µl of pre-warmed Buffer EB (10 mM Tris·Cl) was added to elute the product. This was added to the center of each QIAprep spin column before sitting for 1 minute. This was then centrifuged for 1 minute.
12. The products were then subject to Sanger-sequencing and afterwards analysis.

Results

To develop the Fc Receptor interactome, FCRs that have been conventionally known for the past 30 years were the starting point. Cloning the cytosolic domains was the primary point of interest since they have been known to relay signals to the nucleus. Seven receptors were investigated namely: TOSO, gama-chain, Fc α RI, FcγRII, FcγRIIC, FcγRIIA, and FcγRIIB. Proteins are written with latin and genes are written in Arabic numerals. E.e. FCGR2B is a gene and FcγRIIB is its protein.

For the yeast two-hybrid system, one strain contains a DNA-Binding domain, known as DB and the other contains an activation domain, known as AD. A known

protein, called the bait, is fused to the DB, and an interacting protein, called the prey, is fused to the AD of a transcriptional activator. The mating of these two yeast strains enables the efficient combination of large collections of DNA constructs. When these yeast mate and their vectors join, the reporter gene is activated, which can be screened for on selective media containing HIS 3AT. Vectors have 3 functional regions: (1) an origin of replication (2) an antibiotic-resistance gene (3) a Gateway cassette; aregion where DNA can be inserted without interfering with plasmid replication or expression of the antibiotic-resistance gene.

Using above mentioned yeast two-hybrid strategy, a pair-wise test of 48 clones of cytoplasmic domains of Fc receptors against an extensive library of ~14,000 full length human clones. After identifying the yeast strains that passed phenotypic selectivity, the cells were lysed and PCR reactions were performed in a 96-well format. PCR products were purified and sequenced using Sanger-based sequencing technology. The sequence reads were Blast against NCBI databank to identify the identity of putative positive interactions. The positive interactions identified were subject to careful verification of all interacting pairs and validation tests using orthogonal assays, which are crucial to ensure the release of the FCRs interactome map of the highest possible caliber. Later, the sequences are analyzed and their corresponding proteins are identified. With the results, the first autoimmune network was generated. The network is composed of nodes (581) and edges (783). The generation of a random network using computer-based simulations in order to analyze side-by-side with our autoimmune interactome. Biological networks are very different than random networks in terms of organization and structure (Zhu *et* al,

2007). To understand the interactions of a network means to understand disease pathologies, SLE particularly in this case (Zhu *et* al, 2007).

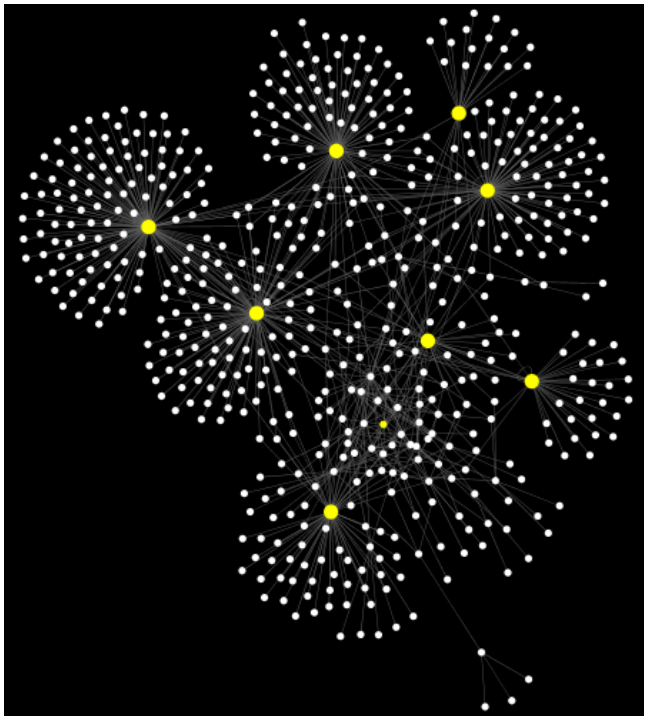


Figure 1. The Global View of the First Human Autoimmune Network

The white dots are nodes or interacting proteins. The lines depict physical and functional interactions between them called edges. The yellow nodes represent hub proteins. There are 581 nodes and 783 edges. The topological properties of this network are discussed in subsequent figures and reveal that this is a scale-free network.

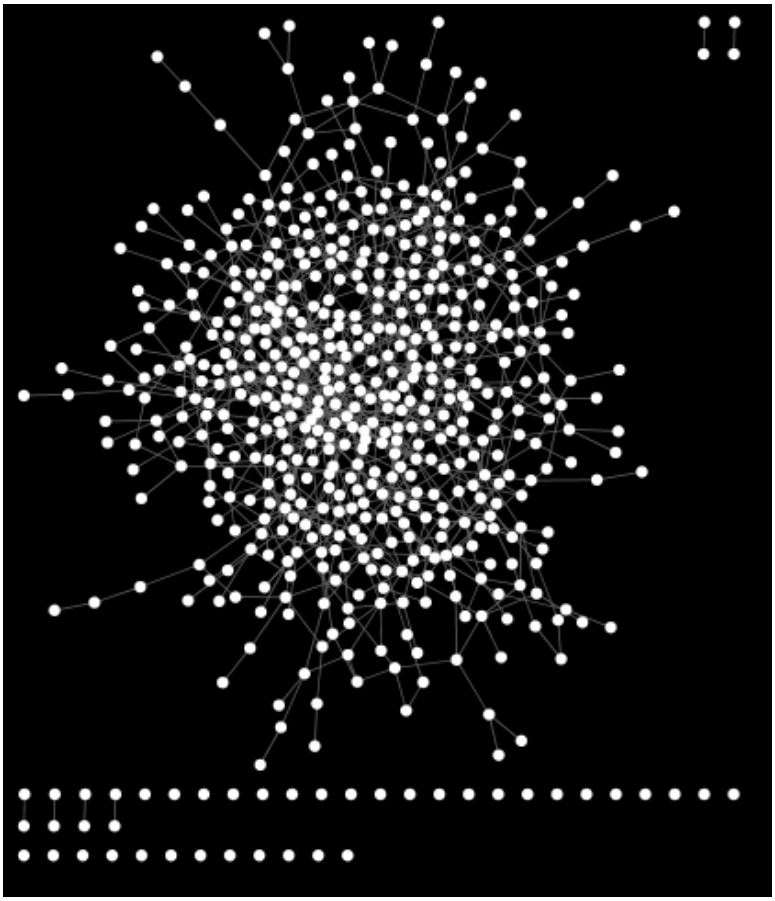


Figure 2. Random network generated for comparison with the human autoimmune interactome

This random network was generated with the same number of nodes (581) and edges (783) as the autoimmune network, yet displays a completely random arrangement. There are no hub proteins. And the nodes depicted on the bottom have either no interacting partners or one; this is an occurrence not observed in the autoimmune network.

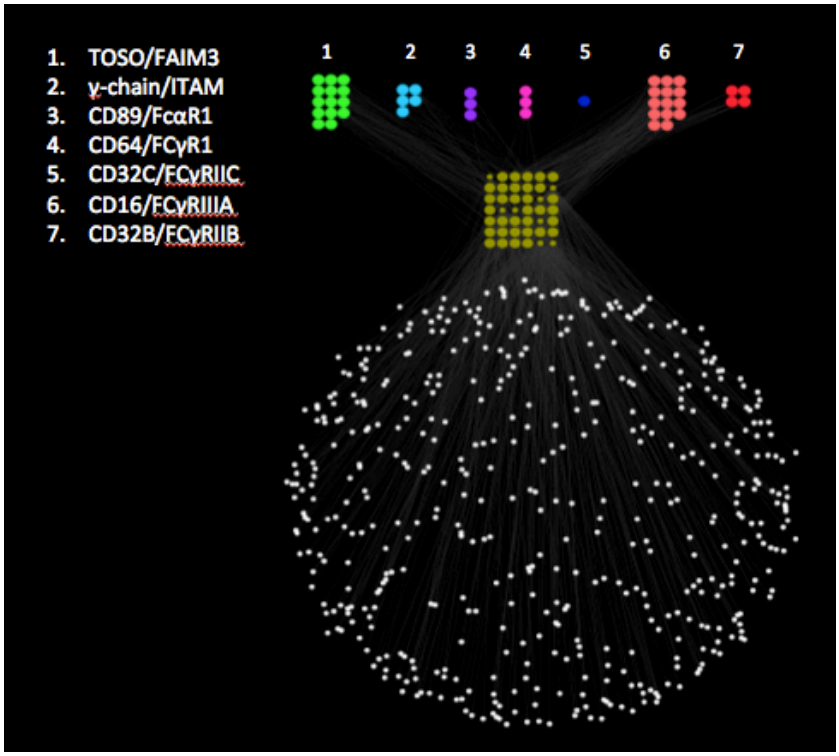


Figure 3. Fc receptor signalomics

The FCRs' network was organized into 3 layers. The first contains 7 colored nodes on the top that represent the 7 Fcγ receptors, which match the key by chronological order. The second layer is the most important, which is the mustard-green colored square of nodes that represent the direct protein interactors with the receptors. The third layer consists of white nodes, which represent the interactors of the positive interactions. A list of several key proteins analyzed from the second layer of Figure 3 were compiled in Table 3.

Table 3. Known functions of proteins identified through their affinity for binding to promoters

Identified Protein	Known Protein Function
SEC61G	Glioblastoma proto-oncogene SEC61gamma is required for **tumor cell survival** and response to endoplasmic reticulum stress."
TP53	"Tp53 gene mutation is associated with **malignant epithelioid angiomyolipoma** of the kidney with pulmonary metastases."
PADI4	The synovial expression of cyclic citrullinated peptide and the generation of anti-CCP antibodies are strongly associated with shared epitope alleles and/or certain PADI4 gene SNPs in **rheumatoid arthritis**."
S100A4	"S100A4 and its downstream factors play important roles in **pancreatic cancer** invasion, and silencing A100A4 can significantly contain the invasiveness of pancreatic cancer."

Network Visualization and Topology Analysis

Several bioinformatic analyses were performed to understand the properties of this interactome map. Among these properties, our analysis demonstrated that FcRs interactome exhibit scale-free properties *i.e.* a networks exhibiting a power-law degree of distribution (Figure 6). The number of edges that nodes have can be calculated through a negative exponent of the value for edges. This negative exponent is a parameter usually between 2 and 3. This random network serves to validate the autoimmune network as well as a useful comparison when assessing the degree of distribution, degree of betweeness, clustering coefficient, and the shortest path of the autoimmune network. Cytoscape plug-ins were utilized to learn the different topological properties of the graphs. Cytoscape is an open-source platform for analysis as well as visualization of networks (Salto *et* al, 2012). Plug-ins provide network-profiling analyses. Cytoscape plug-ins are extensions of the software that enable it to be a highly useful bioinformatics tool when analyzing large-scale data from a high-throughput experiment such as the yeast-two hybrid assay. This software enables the modification of nodes visible such as nodes with certain numbers of interactions. The degree of a node in a network is the number of connections or edges the node has to other nodes. The degree of distribution is the probability of distribution of these degrees over the whole network (Figure 6). The shortest paths identify the minimal distance between two nodes in an interactome (Figure 7). High degree of betweeness means that there are multiple shortest paths (Figure 8). This graph demonstrates the high betweeness of the interactors. Clustering means the proteins that have a high number of neighbors and are involved in clustering (Figure 9). The clustering coefficient is calculated by a N/M ratio, in which N is the number of edges

between the protein's neighbor and M is the maximum number of possible neighbors that could exist in a protein (Gu *et* al, 2011). A high clustering coefficient suggests that protein-protein interactions are highly modular.

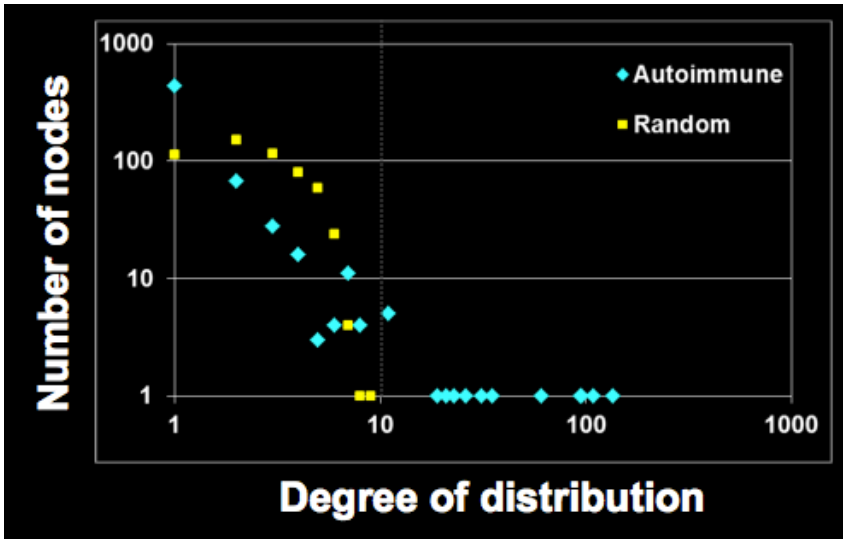


Figure 4. Immune interactors posses a higher degree of distribution in the autoimmune interactome

The random and autoimmune data were computed to analyze their frequency from 1 to 1000. The range of frequency for the random network is 1 to 10. The range of frequency for the autoimmune network is 1 to ~100. The degree of distribution for the autoimmune network is 1 to about 600. In contrast, the degree of distribution for the random network is 1 to ~100. The degree of distribution reveals valuable insight into a network's structure. For instance, the random network exhibits a small range of distribution, characteristic of a simple network. In contrast, the autoimmune network has a variety of degree of distribution. Most of the nodes maintain a small degree (around 50), while a few nodes have large degree (around 100) as they have high connectivity. These nodes of high connectivity are key nodes that could be hub proteins.

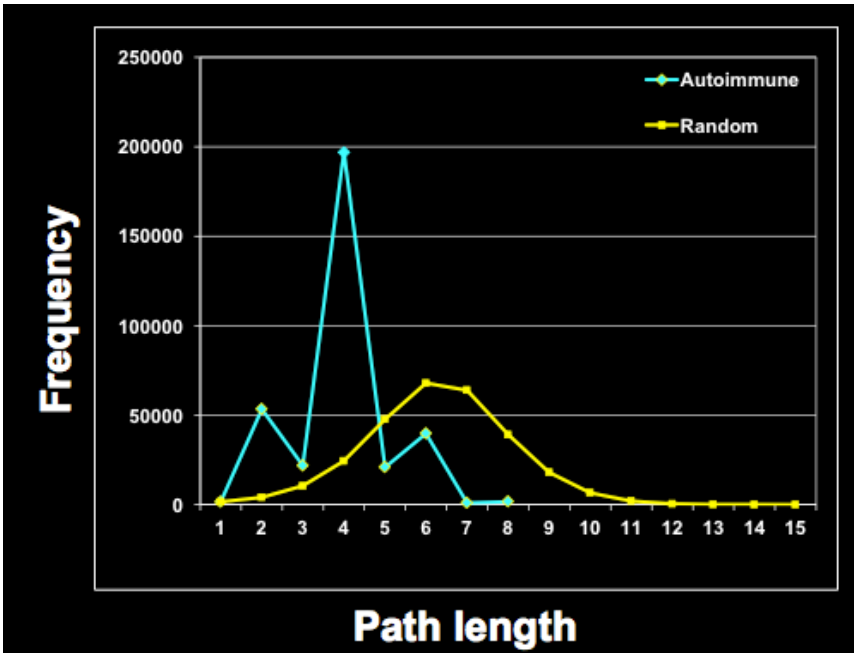


Figure 5. Immune interactors possess a shorter path in the autoimmune interactome

The frequency of the random network ranges from 1 to 15, whereas the frequency of the autoimmune network ranges from 1 to 8. The path length of the autoimmune network ranges from 0 to 200,000. The path length of the random network ranges from 0 to ~60,000. This data suggests that proteins in the FCR interactome exhibit the usage of minimal distances between nodes to reach a target destination. The high number of frequency or connectivity of nodes at the path length of 4 suggests that these nodes could be key nodes central to many connections, whereas most nodes vary in their connectivity or frequency and path lengths.

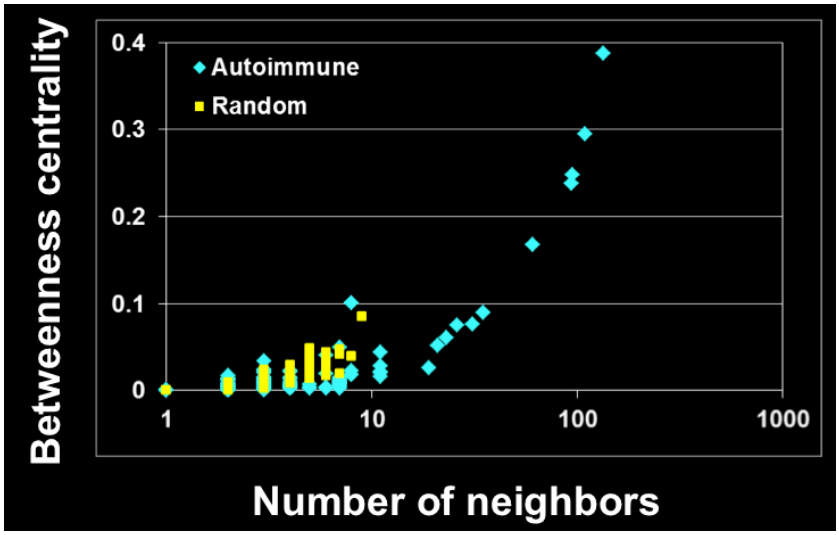


Figure 6. Immune interactors posses a high betweenness in the autoimmune interactome

The degree of betweeness ranges from 1 to ~100 for the autoimmune network whereas the degree of betweeness ranges from 1 to ~10 for the random network. The number of neighbors for the random network is between 0 and ~0.1. The number of neighbors for the autoimmune network ranges from 0 and ~200. This data suggests that proteins in FCRs maintain a complex array of interactions, in which some proteins have a high numbers of neighbors, which highlights their biological functional significance.

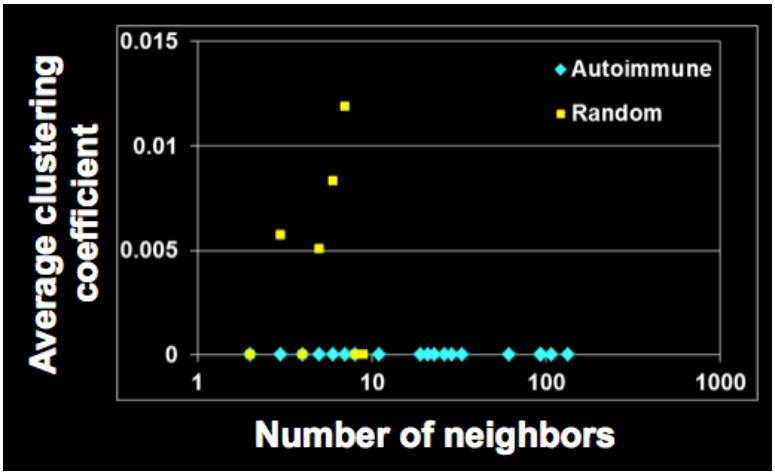


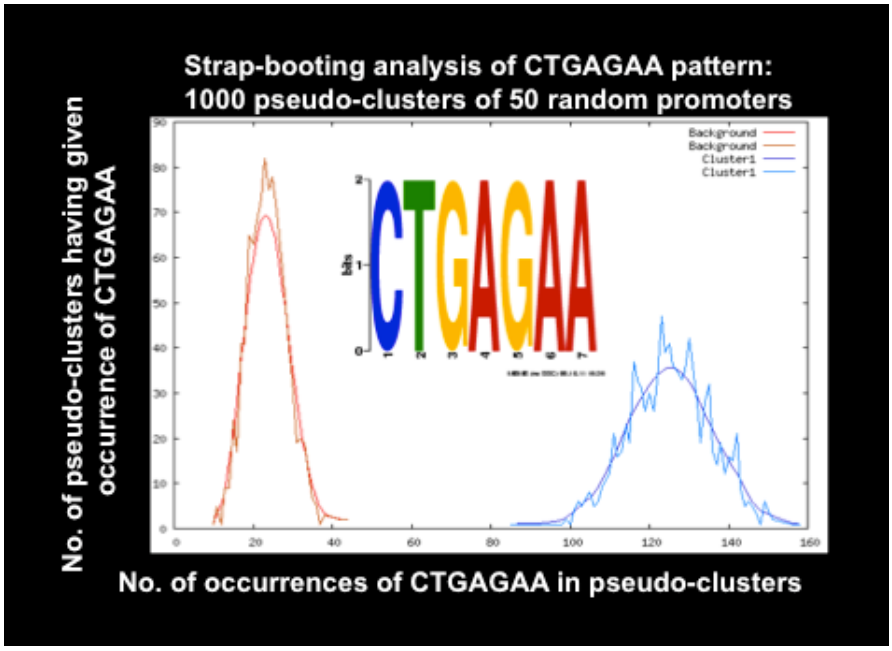
Figure 7. Immune interactors tend to form clusters in the autoimmune interactome

The average clustering coefficient for the random network ranges from ~3 to ~10. The average clustering coefficient for the autoimmune network ranges from ~3 to ~200. The number of neighbors of the autoimmune network maintains within the 0 range. The clustering coefficient at 0 corresponds to true clusters. The autoimmune network contains only true clusters, which validates our findings. In contrast, the number of neighbors for the random network ranges from 0 to ~0.0125. This data suggests that proteins in FCRs are highly connected and some are key to many connections.

FCGR2B promoter analysis identifies key regulatory cis-elements

A *cis*-regulatory element is a region of DNA or RNA that regulates the expression of genes located on that same molecule, and transcription factors are proteins that control gene expression. By finding overrepresented *cis*-regulatory elements, through the use of bioinformatics, the binding factors that cause FcγIIB expression can be determined. The overregulated *cis*-elements were analyzed using MEME and POBO. The software found 5 sites of CTGAGAA in 1.6 kilobites of data. To define the reason for this occurrence, a computer-based boot strapping experiment seemed like the perfect means to find the answer. 50 random promoters were applied 1000 times in the human genome.

A.



B.

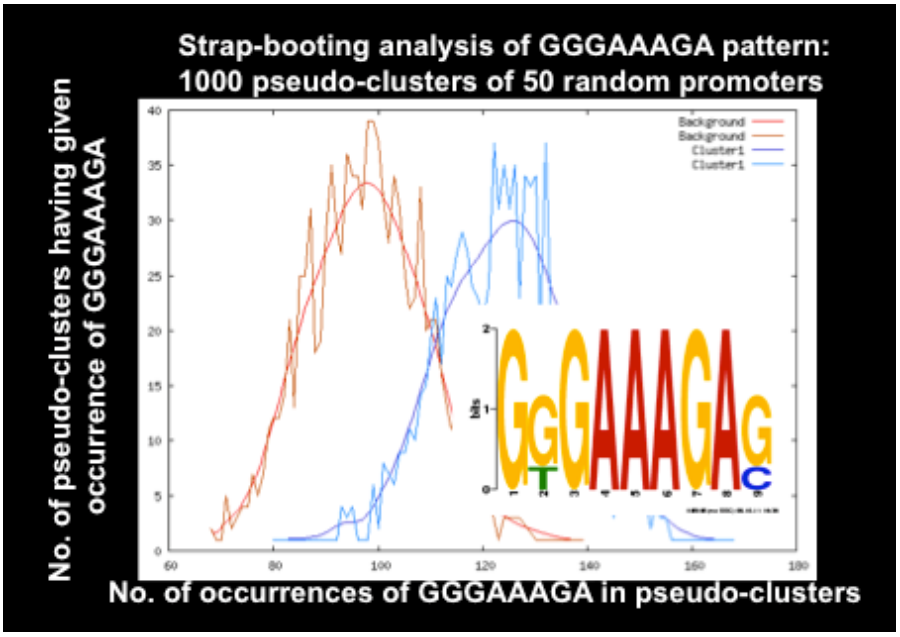


Figure 8: Overrepresented motifs using MEME and POBO in FCGR2B

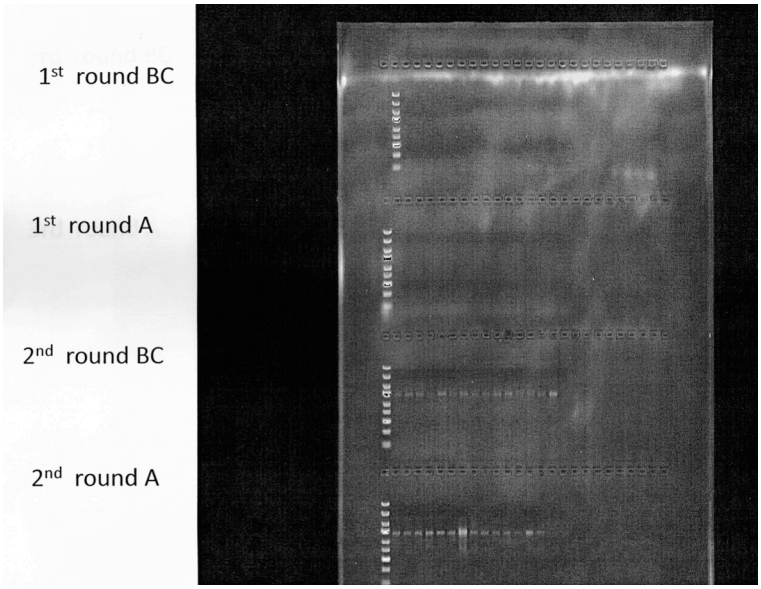
Overrepresentation of CTGAGAA (A) and GG/TGAAAGA (B):

A. Five sites in 1.6 kb of data were identified. The core binding site for CSL (CBF1/RBP-J, Su(H), Lag-1). For the random cluster (the red curve), the average number of pseudo-clusters for this sequence is ~70 and the average range is ~25. For the blue curve, which represents activated *cis*-regulatory elements of the CTGAGAA sequence, the average number of pseudo-clusters is ~40 whereas the average range is ~125.

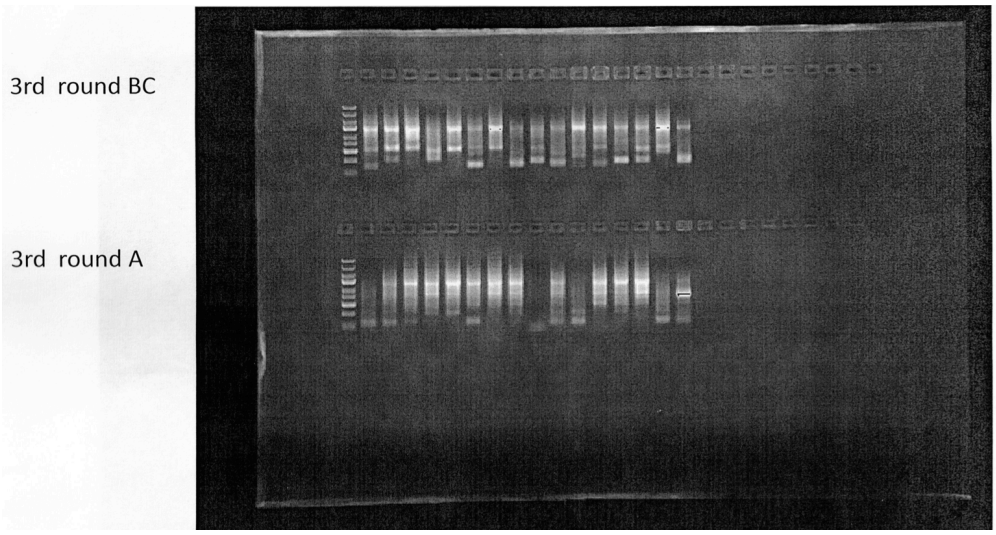
B. Five sites in 1.9 kb of data were identified. These are the core binding sequence for members of the NFkB family (RelA; p65 and NFkB1; p50). For the random cluster (the red curve), the average number of pseudo-clusters for this sequence is ~34 and the average range is ~100. For the blue curve, which represents activated *cis*-regulatory elements of the sequence, the average number of pseudo-clusters is ~30 whereas the average range is ~130.

The red curve represents random populations of gene expression and the blue curve is the CSL binding motif. The fact that the blue curve, which represents activated *cis*-regulatory elements of the sequences, is distinct from the random replication occurrences suggests that this *cis*-element is overregulated. CSL is a DNA binding protein that functions as a repressor or activator. Promoter analysis identified enriched elements in CSL. Further investigation needs to be done to determine whether these are not just elements, but binding partners.

A.


1st round BC
1st round A
2nd round BC
2nd round A

B.


3rd round BC
3rd round A

C.

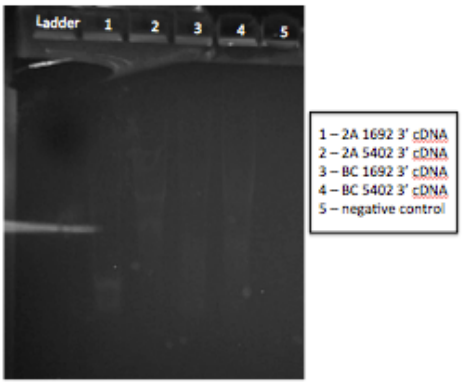


Figure 9. Optimization of RACE PCR for CNV of FCGR2- A, B & C

This figure demonstrates the quality of the products obtained before and after optimization of the protocol. **A.** This shows the results of the gel electrophoresis of the first and second round Nest PCR products. "BC" and "A" refer to the portions of the 1q23 gene cluster that contains the FCGR2B and FCGR2C (BC), and FCGR2A (A). **B.** This is an image of the gel electrophoresis of the third and last round of the Nest PCR. Evidently, there are more bands visible throughout the PCR rounds. **C.** This is an image of the gel electrophoresis of the products before optimization of the protocol. Evidently, there are no strong bands. The key details what cDNA is in the lanes corresponding numerically. The numbers correspond to donors and the "A" and "BC" reference the genes they contain (as explained above). The redesigned approach yielded stronger bands as well as the optimal product for analyzation.

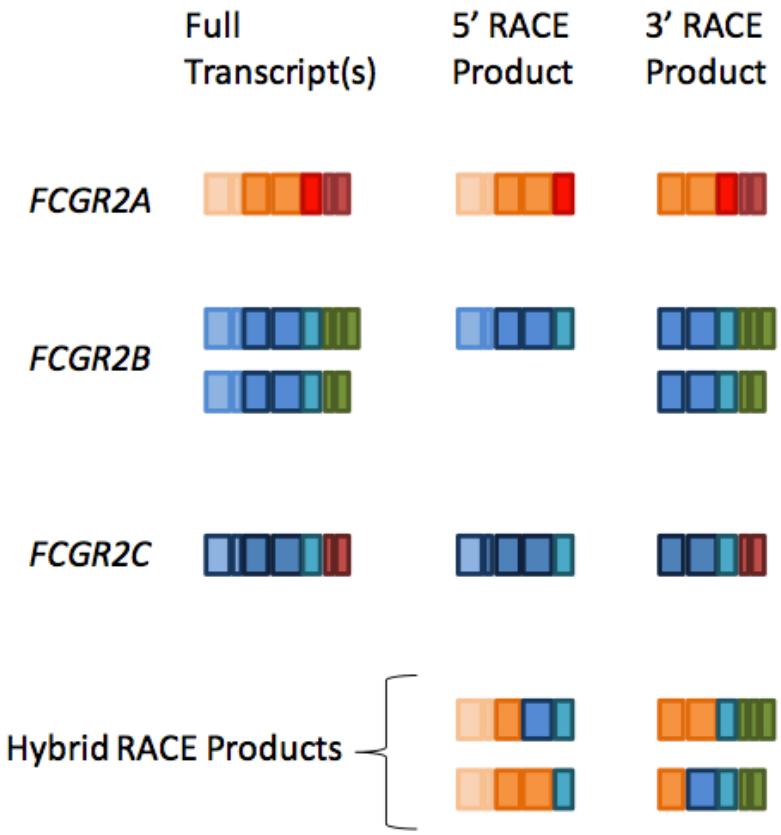


Figure 10. RACE Strategy resulting in FCGR2A/FCGR2B Hybrid Products

Color homology and shapes represent homology, thus the portions of the hybrid RACE product that match the wild type FCGR2A and FCGR2B sequences are the same. Paralogous sequencing of the variants alongside the FCGR2A and FCGR2B wild types can confirm the same FCGR2B origin as opposed to the FCGR2C sequence. Future aims include approaches to further understand this hybrid RACE product.

Note: From Travis Ptacek 2012. Reprinted with permission.

Discussion

Large-scale yeast two-hybrid screens generated most of the binary protein interaction data currently available. The positive interactions identified through the Y2H system could be biologically relevant and thus provide a foundation for future investigations of the mechanisms by which the interactions affect the physiology of an organism. Recent efforts to map interactions in model organisms and in humans illustrate the promise and some of the limitations of the two-hybrid approach. Although these maps are incomplete and include false positives, they serve as a framework around which to elaborate and model the in vivo interactome. False positives can entail "pseudo-interactions" which define the occurrences of positive interactions identified from the Y2H method, but *in vivo*, these macromolecules are either temporally or spatially separated throughout an organism's lifetime. Although there are several levels of control to identify false positives, high-throughput assays that generate large datasets can include a fraction of false positives. Thus, the dataset must be analyzed thoroughly for quality before the construction of a reliable interactome map. This includes the prefilteration of the datasets with the use of orthogonal data or assays. Weak to strong interactions can be identified from Y2H assay, although these interaction strengths may not be reflective of the *in vivo* interactions. Regardless, strong interactions are considered as higher quality and thought to have a higher potential to be involved in biophysical interactions. Confidence scores are assigned that suggest significant binary interactions upon analyzing the robustness or strength of the yeast growth on various selective media.

Since the first development of the Y2H system, controls and strategies to avoid misinterpreted or false information have been developed which has upgraded the Y2H

system to a highly reliable binary interaction assay. One of the strategies and its controls are aimed to identify autoactivators. Autoactivators occur when a DB causes for the expression of its downstream reporter gene, although it is haploid. The yeast-two hybrid system is highly stringent exhibiting 97 percent accuracy. AD-Y autoactivators are identified on cycloheximide containing media, which allows for growth of only DB-X containing yeast. Yeast strains that grow on cycloheximide are disregarded and classified as *de novo* autoactivators. To identify DB-X autoactivators, high 3AT concentrations in media can be incorporated, which act on the GAL1-HIS3 reporter gene of DB-X.

A system of interacting macromolecules can be abstracted into a mathematical framework, a graph, which can be further mathematically analyzed to obtain a comprehension of the system. The autoimmune interactome includes 581 nodes and 783 edges. Many of these interactions are indirect. To identify key proteins, including hubs, several topological properties are identified and include: the degree of distribution, degree of betweeness, clustering coefficient, and shortest path. The number of edges occurring at a node is called the degree. A high node degree would be characteristic of a macromolecule that interacts with many other molecules, such as ATP, which interacts with many proteins. Degree of betweenness among nodes can identify key proteins (even non-hub nodes) that coordinate interactions between many nodes. The clustering coefficient reveals the number of neighbors that are likely to be connected in a network. The least number of distinct connected nodes between two nodes defines the shortest path, which reveals how quickly information can be transported in a network. Additionally, topological properties can support hypotheses of perturbation affects among proteins. For instance the central-lethality relation describes that within small motifs,

certain proteins are indispensable due to its number of direct connections. Identifying key proteins including hub and non-hub proteins, can have major implications on modeling interactomes and evaluating drug design target protein-protein interactions. The random and autoimmune generated networks were both analyzed for their topological properties, and the data from the random network served to validate the significance of the findings within the autoimmune interactome.

Transcription factors can induce or repress gene expression upon binding to their cognate *cis*-regulatory DNA sequences. The discovery of an entire collection of transcription factor binding sites within an organism is one of the largest challenges in computational biology. Moreover, the understanding of the regulatory mechanisms of gene expression is a major role of genomics. FCGIIB promoter analysis is very important and done on a bioinformatic and systemic level. By finding overrepresented *cis*-regulatory elements, through the use of bioinformatics, the binding factors that cause FcγIIB expression can be determined. This is one of the first steps to analyzing the regulatory network. Several *cis*-regulatory sequences were found that require further investigation to determine that they are indeed binding partners and not just overregulated elements.

Several proteins were identified through their affinity for binding to promoters. Several proteins that have been linked to diseases in previously published literature were found within the autoimmune network. For example, PAD14 and SD100 have been associated with RA, and some proteins associated with immune diseases were also confirmed. These findings could be attributable to the location of FCG receptors on

immunoglobulin G, which is a major antibody against pathogens. Further investigations should be done for confirmation, however the link is promising.

With regards to the investigation of FCR CNV, the optimization of the protocol yielded advanced results in comparison to the initial conventional approach. Moreover, the hybrid RACE product creates more room for future investigation. Utilizing SMART cDNA synthesis and RACE PCR enabled us to obtain a full-length product as well as discriminate between FCGR2A, FCGR2B and FCGR2C (Figure 12). SMART cDNA synthesis eliminated the need to fuse two products, thus decreasing the potential margin for error. According to Clontech (2012), with SMART (Switching Mechanism at 5' End of RNA Template) technology there is no adaptor ligation necessary during the addition of known sequences to both ends of the cDNA strand during first strand synthesis, which is also shown in Figure 12. Rapid Amplification of cDNA Ends (RACE) is useful for the amplification of mRNA sequences between pre-defined internal sites and an unknown sequence that is located at either the 3' or 5' ends of the mRNA. SMART cDNA synthesis was utilized for the downstream applications of 5' and 3' RACE PCR amplification strategies. For the original RACE strategy, the primers were designed to amplify FCGR2B, FCGR2A and FCGR2C. Two gene-specific primers were required per RACE reaction: a FCGR2B-FCGR2C specific primer as well as pan-FCGR2 primer (which amplified mostly FCGR2A transcripts) as seen in Figure 11. There is individuality in 5' portion of FCGR2A, thus 3' RACE PCR method is beneficial for the obtainment of a full-length PCR product with the cytoplasmic domain that can also discriminate between FCGR2B and FCGR2C fused to FCGR2A. The 3' region can be analyzed to discriminate between FCGR2B and FCGR2C. As depicted in Figure 2 of the

introduction, there is high homology between the extracellular domain of FCGR2C and FCGR2B, but no homology at their cytosolic domains. With regards to FCGR2A and FCGR2C, there is high homology at their cytoplasmic domains only, and FCGR2A and FCGR2B exhibit no homology. There were several trial-and-error attempts with PCR before the optimal conditions were defined (shown in tables 1 and 2). After minor adjustments to the initial settings, the results were optimal.

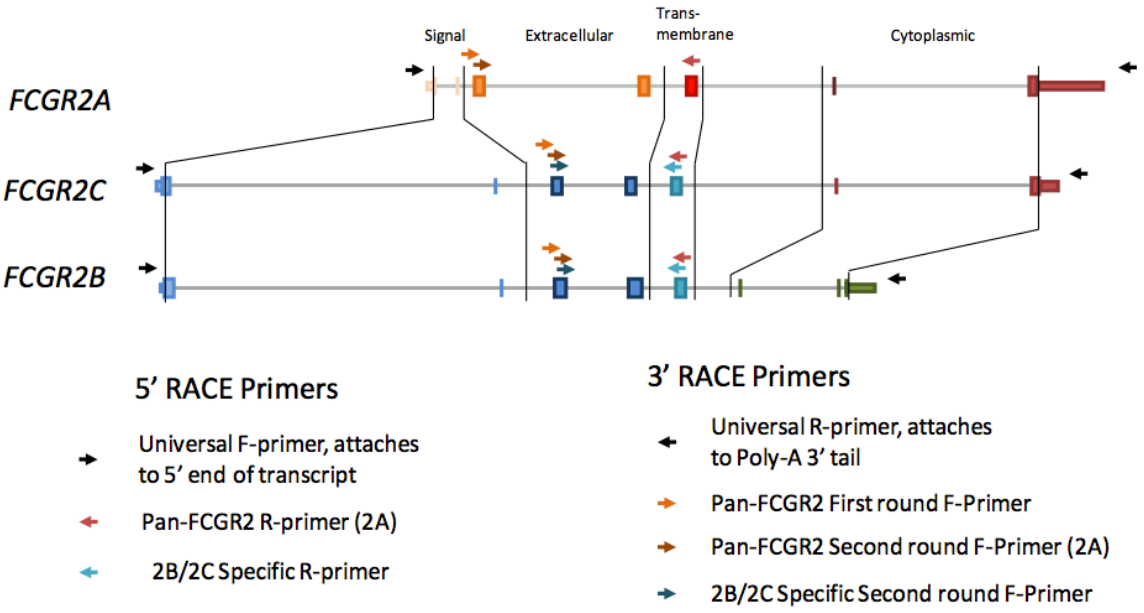


Figure 11. Previous RACE Strategy

Exons are color coded by domain and by sequence paralogy to other genes. Moreover, two exons located within different genes and share the same color maintain about 95-100% paralogy. The arrows show primer sites and their respective reactions are listed. As you can see, the 5' universal primer can anneal onto the 5' site, and there is a gene specific primer on the transmembrane domain. There is also a 3' universal primer on the poly-a-tail and the 3' forward primer on the extracellular domain. This set-up resulted in two independent products (5' and 3' products) that were cloned as opposed to sequencing, which is highly labor intensive.

Note: From Travis Ptacek 2012. Reprinted with permission.

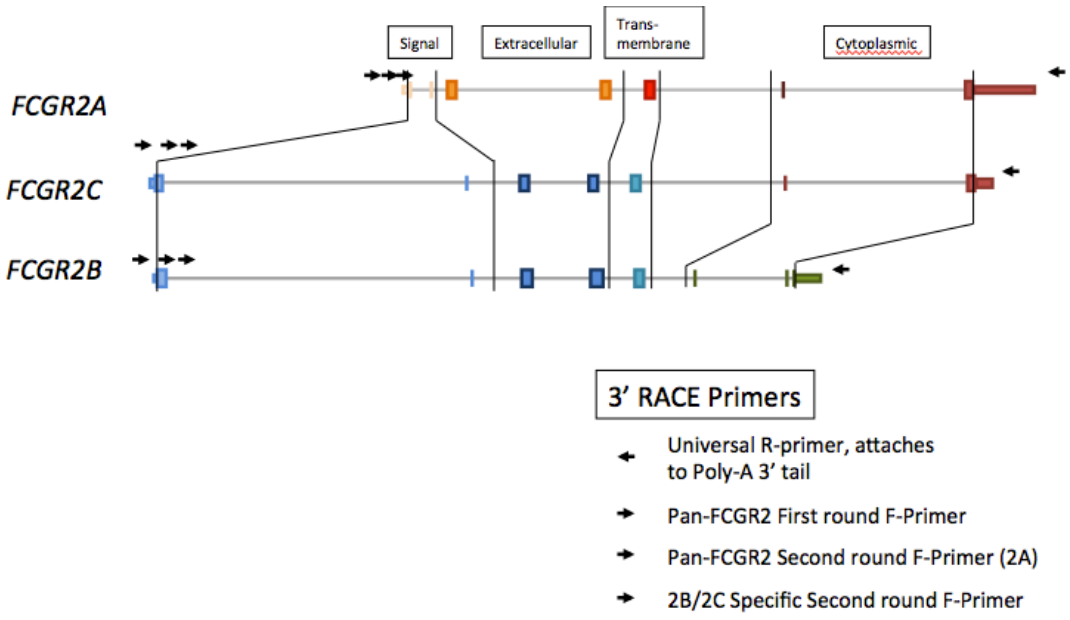


Figure 12. Modified RACE Strategy

With the 3' RACE the full-length product that can discriminate between FCG2B and FCG2C fused to FCG2A via the cytoplasmic domain was obtained. The 3' RACE primers bind to the 5' signaling domain and the 3' cytoplasmic domain.

Note: Edited from Travis Ptacek 2012. Reprinted with permission.

All in all, the results satisfy the 3 objectives of the investigation including: the identification of novel interacting partners of FCRs' cytoplasmic domains; the finding of statistically overrepresented *cis*-regulatory elements in FCGR2, the only classical FCR and the identification of their cognate transcription factors; and the identification of FCGR2B CNV in SLE patients.

REFERENCES

Breunis, W.B., van Mirre, E., Bruin, M., Geissler, J., de Boer, M., Peters, M., Roos, D., de Haas, M., Koene, H.R., Kuijpers, T.W. (2008). Copy number variation of the activating FCGR2C gene predisposes to idiopathic thrombocytopenic purpura. Blood 111,1029-1038.

Breunis, W.B., van Mirre, E., Geissler, J., Laddach, N., Wolbink, G., van der Schoot, E., de Haas, M., de Boer, M., Roos, D., Kuijpers, T.W. (2009) Copy number variations at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B. Hum. Mutat. 30,E640-650.

Bruder, C.E.G., Piotrowski, A., Gijsbers, A.A.C.J., Andersson, R., Erickson, S., de Stahl, T.D., Menzel, U., Sandgren, J., von Tell, D., et al. (2008). Phenotypicall concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am. J. Hum. Genet. 82,763-771.

Cantsilieris, S., and White, S.J. (2012). Correlating multiallelic copy number polymorphisms with disease susceptibility. Hum. Mutat. doi: 10.1002/humu.22172.

Carvunis, A. R., Roth, F. P., Calderwood, M. A., Cusick, M. E., Superti-Furga, G., and Vidal, M. (2013). “Interactome networks,” in *Handbook of Systems Biology Concepts and Insights*, Chap.3, eds M. Walhout, M. Vidal, and J. Dekker (Amsterdam: Elsevier Inc.), 45–60. doi: 10.1016/B978-0-12-385944-0.00003-4

Clark, M.R., Liu, L., Clarkson, S.B., Ory, P.A., and Goldstein, I.M. (1990). An abnormality of the gene that encodes neutrophil fc receptor IIII in a patient with systemic lupus erythematosus. J. Clin. Invest. 86, 341-346.

Clancy, S. (2008). Copy number variations (CNVs) have been linked to dozens of human diseases, but can they also represent the genetic variation that was so essential to our evolution? Nature Education.

Conrad DF, Pinto D, Redon, R, et al. (2010). Origins and functional impact of copy number variation in the human genome. Nature. PMID: 19812545.

Croker, J.A., Kimberly, R.P. (2005). SLE: challenges and candidates in human disease. Trends Immunol. 26,580-586.

Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., Harper, L., Kamesh, L., Heward, J.M., Gough, S.C.L., de Smith, A., Blakemore, A.I.F., et al. (2007). FCGR3B copy

number variation is associated with susceptibility to systemic, but not organ-specific autoimmunity. Nat. Genet. 39,721-723.

Feuk, L., Carson, A.R., Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. 7,85-97.

Gergely, J. and Sarmay, G. (1990). The two binding-site models of human IgG binding Fc gamma receptors. The FASEB Journal. Vol. 4. No. 15 3275-3283.

Haibin, Gu., Zhu P., Jiao Y., Meng Y., and Chen M. (2011) PRIN: a predicted rice interactome network. Doi: 10.1186/1471-2105-12-161

The International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley, J.B., Aarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L.

Karassa, F.B., Trikalinos, T.A., Ioannidis J.P.A. (2004). The role of FcγRIIA and IIIA polymorphisms in autoimmune diseases. Biomed. Pharmacother. 58,286-291.

Katsnelson, A. (2008). Autoimmunity in plants? The Scientist.

Koene, H.R., Kleijer, M., Roos, D., de Haas, M., von dem Borne, A.E. (1998). FcγRIIIB gene duplication: evidence for presence and expression of three distinct FcγRIIIB genes in NA(1+,2+)Sh(+) individuals. Blood. 91,673-679.

Kindt, T., Osborne, B., Goldsby, R. (2006). Kuby Immunology. 1429202114

Kissel, K., Hofmann, C., Gittinger, F.S., Daniels, G., and Bux, J. (2000). HNA-1a, HNA-1b, and HNA-1c (NA1, NA2, SH) frequencies in African American blacks and in Chinese. Tissue Antigens *56,143-148.*

Li, X., Ptacek, T.S., Brown, E.E., Edberg, J.C. (2009) Fcγ receptors: structure, function and role as genetic risk factors in SLE. Genes. Immun. 10,380-389.

Li, X., Wu, J., Ptacek, T., Redden, D.T., Brown, E.E., Alarcón, G.S., Ramsey-Goldman, R., Petri, M.A., Reveille, J.D., Kaslow, R.A., et al. 2013. Allelic dpendent expression of a novel activationg Fc receptor on B cells enchances humoral immune responses. In Preparation.

Lupus Foundation of America. (2013). Understanding lupus. www.lupus.org.

Mamtani, M., Anaya, J.M., He, W., Ahuja, S.K. (2010). Association of coyp number variation in the FCGR3B gene with risk of autoimmune diseases. Genes Immun. 11,155-160.

Matsuo, K., Rocter, J., and Stroncek, D. (2000). Variations in genes encoding neutrophil antigens NA1 and NA2. Transfusion. *40,645-653.*

McKinney, C., Broen, J.C.A., Vonk, M.C., Beretta, L., Hesselstrand, R., Hunzelmann, N., Riemekasten, G., Scorza, R., Simeon, C.P., Fonollosa, V., et al. (2012). Evidence that deletion ant FCGR3B is a risk factor for systemic sclerosis. Genes Immun. 12,458-460.

McKinney, C., Fanciulli, M., Merriman, M.E., Phipps-Green, A., Alizadeh, B.Z., Koeleman, B.P., Dalbeth, N., Gow, P.J., Harrison, A.A., Highton, J. et al. (2010). Association of variation in Fcgamma receptor 3B gene copy number with rheumatoid arthritis in Caucasian samples. Ann. Rheum. Dis. 69,1711–1716.

Mitra, K., Carvunis, A. R., Ramesh, S. K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* 14, 719–732. doi: 10.1038/nrg3552

Mockli, N., and Auerbach, D. (2004). Quantitative β-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. BioTechniques 36:872-876.

Molokihia, M., Fanciulli, M., Petretto, E., Patrick, A.L., McKeigue, P., Roberts, A.L., Vyse, T.J., and Aitman, T.J. (2011). FCGR3B copy number variation is associated with systemic lupus erythematosus risk in Afro-Carribeans. Rheumatology. 50,1206-1210.

Morris, D.L., Roberts, A.L., Witherden, A.S., Tarzi, R., Barros, P., Whittaker, J.C., Cook, T.H., Aitman, T.J., and Vyse, T.J. (2010). Evidence for both copy number and allelic (NA1/NA2) risk at the FCGR3B locus in systemic lupus erythematosus. Eur. J. Hum. Genet. 18, 1027-1031.

Mueller, M., Barrors, P., Witherden, A., Roberts, A., Zhang, Z., Schaschl, H., Yu, C.Y., Hurles, M., Schaffner C., Floto, A., et al. (2013). Genomic pathology of SLE-associated copy number variation at the FCGR2C/FCGR3B/FCGR2B locus. Am. J. Hum. Genet. 92,28-40.

Parrish, J., Gulyas, K., and Finley, R. (2006). Yeast two-hybrid contributions to interactome mapping. Current Opinion in Biotechnology. Vol. 17. doi: 10.1016/j.copbio.2006.06.006.

Prioietti S, Bertini L., Timperio AM, Zolla L, et al. (2013). Crosstalk between salicylic acid and jasmonate in Arabidopsis investigated by an integrated proteomic and transcriptomic approach. Doi: 10.1039/c3mb25569g.

Pumplin, N., and Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. Nature Reviews Microbiology. Doi:10.1038/nrmicro3120

Qiagen Miniprep Handbook. (2012).

Restrepo, J., Ott, E., and Hunt, B. (2007). Approximating the largest eigenvalue of network adjacency matrices. Phys Rev E, 76:056119.

Restrepo, J., Ott, E., and Hunt, B. (2006). Emergence of coherence in complex networks

of heterogeneous dynamical systems. Phys Rev Lett, 96:254103.

Robinson, J.I., Carr, I.M., Cooper, D.L., Rashid, L.H., Martin, S.G., Emery, P., Isaacs, J.D., Barton, A., BRAGGSS, Wilson, A.G., et al. (2012). Confirmation of association of FCGR3B but not FCGR3A copy number with susceptibility to autoantibody positive rheumatoid arthritis. Hum. Mutat.

Robinson, R. (2006). Plant Pathosystems. The Genetic Basis of Epidemics in Agriculture. Annals of the New York Academy of Sciences. DOI: 10.1111

Salto, R., Smoot, M., Ono, K., Ruscheinski, J., Wang, P., Lotia, S., Pico, A., Bader, G., and Ideker, T. (2012). Nature Methods. Vol. 9. doi:10.1038/nmeth.2212.

Seebacher, J., and Gavin, A. C. (2011). SnapShot: protein–protein interaction networks. *Cell* 144, 1000:1001.e1. doi: 10.1016/j.cell.2011.02.025

Sessa, G. (2012). Molecular Plant Immunity. John Wiley & Sons.

Shih, Y., and Parthasarathy, S. (2012). A single source k-shortest paths algorithm to infer regulatory pathways in a gene network. Bioinformatics. Doi: 10.1093.

Shmulevich, I., and Aitchison, J. D. (2009). Deterministic and stochastic models of genetic regulatory networks. *Methods Enzymol.* 467, 335–356. doi: 10.1016/S0076-6879(09)67013-67010

Smith, K.G.C., Clatworthy, M.R. 2010. FcγIIB in autoimmunity and infection: evolutionary and therapeutic implications. Nat. Rev. Immunol. 10,328-343.

Spoel, S., and Dong, X. (2012). How do plants achieve immunity? Defense without specialized immune cells. Nature Reviews Immunology. Vol. 12.

Su, K., Wu, J., Edberg, J.C., Li, X., Ferguson, P., Cooper, G.S., Langefeld, C.D., Kimberly, R.P. (2004a). A promoter haplotype of the immunoreceptor tyrosine-based inhibitory motif-bearing FcgammaRIIb alters receptor expression and associates with autoimmunity. I. Regulatory FCGR2B polymorphisms and their association with systemic lupus erythematosus. J. Immunol. 172,7186–7191.

Swanson, D. (1999). Smart Race to the finish: Clontech enters Smart Race for cDNA amplification. The Scientist.

Tong, Y., Jin, J., Yan, L., Neppert, J., Marget, M., and Flesch, B.K. (2003). FCGR3B gene frequencies and FCGR3 variants in a Chinese population from Zhejiang Province. Ann. Hematol. *82,574-578.*

Willcocks, L.C., Lyons, P.A., Clatworthy, M.R., Robinson, J.I., Yang, W., Newland, S.A., Plagnol, V., McGovern, N.N., Condliffe, A.M., Chilvers, E.R., et al. (2008). Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. J. Exp. Med. 205,1573-1582.

Van Binsbergen, E. (2011). Origins and breakpoint analyses of copy number variations: up close and personal. Cytogenet. Genome Res. 135,271-276.

Vidal, M., Cusick, M., Klitgord N., and Hill, D. (2005). Interactome: Gateway into Systems Biology. Human Molecular Genetics. Vol. 14. doi: 10.1093/hmg/ddi335.

Zhao, L., Beverlin, B., Netoff, T., and Nykamp, D. (2011). Synchronization from second order network connectivity statistics. Frontiers in Computational Neuroscience.

Zhou, X., Lv, J., Bu D., Yu, L., Yang, Y., Zhao, J., Cui, Z., Yang, R., Zhao, M., Zhang, H. (2009). Copy number variation of *FCGR3A* rather than *FCGR3B* and *FCGR2B* is associated with susceptibility to anti-GBM disease. Int. Immunol. 22,45-51.

Zhu, X., Gerstein, M., and Snyder, M. (2007). Getting connected: analysis and principles of biological networks. Genes & Dev. Cold Spring harbor Laboratory Press. 21:1010-1024.

CONCLUSIONS

The results provide a deeper understanding of the regulatory mechanisms of FCRs, particularly FcγRIIb, the only inhibitory receptor of its class. The results also satisfy the 3 objectives including (1) the identification of novel interacting partners of FCRs' cytoplasmic domains, (2) the finding of statistically overrepresented *cis*-regulatory elements in FCGR2, the only classical FCR and the identification of their cognate transcription factors, and (3) the identification of FCGR2B CNV (Copy number variation) in SLE patients. The human autoimmune network is a novel leap towards understanding the onset of systemic autoimmune diseases associated with FCRs, particularly FcγRIIB. The added control measures to the yeast two-hybrid assay enable the creation of a highly stringent network in addition to the orthogonal assays done to assure that the network was of the highest caliber. The topological properties give insight into the network's structure and the identification of potential key proteins within the autoimmune interactome. Upon analyzation, some of the proteins' functions are already published in existing literature. Some of these proteins have already been linked to autoimmune and immune diseases and were highlighted in the results section (Table 3). Moreover, to understand the regulation of the FCGR2B promoter, overrepresented *cis*-regulatory elements were sought using MEME and POBO. The results yielded the identification of several *cis*-regulatory elements that require future verification. In addition, the modification of the methods approach was successful and yielded a full-

length product of the 1q23 gene locus. The full-length product enables the analyzation of the FCR-coding genes within the donor cohort as well as discrimination between the contained genes including FCGR2B and FCGR2C fused to FCGR2A. The identification of copy number variations associated with SLE furthers our understanding of the disease onset and provides a gateway into diagnostic and treatment options. All in all, the results reveal further understanding of SLE on a genomic level that can contribute to diagnostic methods that enable therapeutic treatments that regulate the autoimmune response—of other rheumatic diseases as well—on a molecular level via FCRs, particularly FcγRIIB.

Significance

Systems biology is the next generation approach towards understanding immunity in plants and humans alike. There is a plethora of new technologies that contribute to this frontier. The development of interactome maps, for instance, enables further understanding of immunity. Interactome maps can mirror *in vivo* conditions through the incorporation of layers of functional "-omic" data, including: genomics, phenomics, transcriptomics, metabolomics, and epigenomics. The integration of this "-omics" data can answer questions regarding plant processes and responses, such as development, signal transduction pathways, RNA processing, protein modifications, cell cycle, and plant immune responses.

In relation to autoimmune diseases, systems biology can enable the realization of the ultimate goal that is the alleviation of symptoms. Patients with RA suffer from chronic pain derived from the inflammation, tenderness, deformation and stiffness of affected joints. At the onset of RA, fatigue, fever and weight loss are common symptoms as well as pain in smaller joints, such as those in the hands and feet. With the progression

of the disease, symptoms of joint pain and stiffness spread to the ankles, elbows, hips and shoulders. The severity of the symptoms can nearly disappear during remission phases of the disease. The symptoms of SLE vary among patients, but almost everyone experiences joint pain. Other symptoms include a butterfly shape rash on the face, fever, fatigue, arrhythmias (abnormal heart rhythms), Raynaud's phenomenon (change of color in phalanges from cold), arthritis and pleurisy (pain in the chest during deep breaths). Both of these diseases affect a significant subset of the population. The Lupus Foundation of America estimates that 1.5 million Americans, and at least five million people worldwide, have a form of lupus. The U.S. Department of Health and Human Services estimates that RA affects one percent of the adult population and approximately 2.5 million Americans. However, the trigger that causes the onset of RA and SLE is unknown. At present, there is no single SLE diagnostic test, but rather an inclusive testing approach that aims to minimize the possible margin of error from false positives and/or false negatives. Molecular medicine could give way to an efficient diagnostic test as well as a potential cure. The findings enable a molecular understanding of the receptors and their genes involved in the pathogenesis of SLE and RA as well as other rheumatic diseases. This understanding provides a foundation for the development of diagnostic methods and efficient therapeutic targets.

Future Directions

With regards to the autoimmune network, further investigations related to the proteins are necessary to fully understand the interactions. Future investigations include the identifications of the involved proteins' functions, which will contribute to the prediction of phenotypic affects on the perturbance of specific protein-protein

interactions. Moreover, further investigations of the identified *cis*-regulatory elements of the FCGR2B promoter are needed to verify that the results are binding partners rather than solely overregulated sequences. Experiential proof of the cis-elements CSL and NκB binding to FCGR2B is needed in addition to determining their roles in autoimmunity. Among all of the interacting partners, we know that some are key immune players thus we are confident that future investigations will illuminate their roles in autoimmunity. Regarding understanding the role of CNV of FCGR2B, although a hybrid product was identified using the RACE PCR strategy, there is much to learn about its genomic organization, such as how the hybrid is made.

LIST OF REFERENCES

Breunis, W.B., van Mirre, E., Geissler, J., Laddach, N., Wolbink, G., van der Schoot, E., de Haas, M., de Boer, M., Roos, D., Kuijpers, T.W. (2009) Copy number variations at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B. Hum. Mutat. 30,E640-650.

Carvunis, A. R., Roth, F. P., Calderwood, M. A., Cusick, M. E., Superti-Furga, G., and Vidal, M. (2013). "Interactome networks," in *Handbook of Systems Biology Concepts and Insights*, Chap.3, eds M. Walhout, M. Vidal, and J. Dekker (Amsterdam: Elsevier Inc.), 45–60. doi: 10.1016/B978-0-12-385944-0.00003-4

Conrad DF, Pinto D, Redon, R, et al. (2010). Origins and functional impact of copy number variation in the human genome. Nature. PMID: 19812545.

Katsnelson, A. (2008). Autoimmunity in plants? The Scientist.

Kindt, T., Osborne, B., Goldsby, R. (2006). Kuby Immunology. 1429202114

Mitra, K., Carvunis, A. R., Ramesh, S. K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* 14, 719–732. doi: 10.1038/nrg3552

Mukhtar, M. S., Nishimura, M. T., and Dangl, J. (2009). NPR1 in plant defense: it's not over 'til it's turned over. *Cell* 137, 804–806. doi: 10.1016/j.cell.2009.05.010

Prioietti S, Bertini L., Timperio AM, Zolla L, et al. (2013). Crosstalk between salicylic acid and jasmonate in Arabidopsis investigated by an integrated proteomic and transcriptomic approach. Doi: 10.1039/c3mb25569g.

Pumplin, N., and Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. Nature Reviews Microbiology. Doi:10.1038/nrmicro3120

Robinson, R. (2006). Plant Pathosystems. The Genetic Basis of Epidemics in Agriculture. Annals of the New York Academy of Sciences. DOI: 10.1111

Sessa, G. (2012). Molecular Plant Immunity. John Wiley & Sons.

Seebacher, J., and Gavin, A. C. (2011). SnapShot: protein–protein interaction networks. *Cell* 144, 1000:1001.e1. doi: 10.1016/j.cell.2011.02.025

Spoel, S., and Dong, X. (2012). How do plants achieve immunity? Defense without specialized immune cells. Nature Reviews Immunology. Vol. 12.

Shmulevich, I., and Aitchison, J. D. (2009). Deterministic and stochastic models of genetic regulatory networks. *Methods Enzymol.* 467, 335–356. doi: 10.1016/S0076-6879(09)67013-67010

Zhu, X., Gerstein, M., and Snyder, M. (2007). Getting connected: analysis and principles of biological networks. Genes & Dev. Cold Spring harbor Laboratory Press. 21:1010-1024.