Unbiased Analysis of the Functions and Interactions of Human Inflammasome Mutations

by

Lindsey Spiker

B.S. Biology, SUNY Plattsburgh, 2018

Submitted to the Graduate Faculty of the
Graduate School of Public Health in partial fulfillment
of the requirements for the degree of

Master of Science

University of Pittsburgh

2021
This thesis was presented

by

Lindsey Spiker

It was defended on

March 26, 2021

and approved by

Committee Member: Dr. F. Yesim Demirci, MD, Associate Professor of Human Genetics, University of Pittsburgh

Committee Member: Dr. Zsolt Urban, PhD, Associate Professor of Human Genetics, University of Pittsburgh

Thesis Advisor: Dr. Scott Canna, M.D, Assistant Professor of Pediatrics and of Immunology, University of Pittsburgh
Abstract

Background: Gain-of-Function (GOF) inflammasome mutations helped define the field of autoinflammation. Genotype/phenotype correlations both found within and between mutated genes that cause inflammasomopathies. We sought to identify features unique to NLRC4 GOF mutations and to develop a system for the unbiased assessment of Protein-Protein Interactions (PPi) made/lost by introduction of NLRC4 or NLRP3 inflammasome mutations in a myelomonocyte cell line.

Methods: Part 1: Transduction of cell lines with WT or Mutant NLRC4 alone was preformed, followed by messenger RNA-sequencing and analysis. Part 2: Transduction of cell lines with WT or various of mutated forms of NLRP3 and NLRC4 was preformed to represent a diversity of clinical autoinflammatory phenotypes directly conjugated to the BirA biotinylase. Several experiments were conducted to assess biotinylation and confirm proper transduction. Interactome was established through the use of mass spectrometry.

Results: Part 1: The most differentially expressed genes were evaluated. Unbiased GSEA analysis identified a strong IFN-induced gene signature among the most differentially expressed genes. Part 2: Inborn errors of immunity may be an initiating factor in the hyperactivation of the inflammasome.
Public Health Relevance: The assessment of various mutations within the same gene or different genes thought to have similar functions may help to diagnose and investigate phenotypic outcomes of disease.

Conclusion: *NLRC4* may directly induce IFN production, potentially through a non-canonical pathway. Screening for interacting partners may identify the mechanism by which *NLRP3* causes most IL-1β mediated disease and *NLRC4* drives both Interleukin-1β (IL-1β) and Interleukin-18 (IL-18) production.
# Table of Contents

Preface .......................................................................................................................................................... ix

1.0 Background .......................................................................................................................................... 1
  1.1 Previous Research ............................................................................................................................... 4
  1.2 Inflammasome Activation .................................................................................................................... 5
  1.3 Genes of Interest with Gain-of-Function Mutations ........................................................................... 6
  1.4 The Interactome and BirA miniTurbo .................................................................................................. 10
  1.5 Specific Aims ....................................................................................................................................... 11

2.0 Methods ................................................................................................................................................ 13

3.0 Results and Discussion ........................................................................................................................ 20
  3.1 Public Health Relevance ....................................................................................................................... 26
  3.2 Future Directions ................................................................................................................................ 26
  3.3 Limitations ......................................................................................................................................... 27

Appendix A Supplemental Material ........................................................................................................ 29

Bibliography ............................................................................................................................................... 34
List of Tables

Table 1. Genetic Mutations & Phenotypic Correlation................................................................. 9
List of Figures

Figure 1. Construct Protein Digestion ................................................................. 15
Figure 2. HEK 293T Transfection ...................................................................... 16
Figure 3. Transduction Efficiency ..................................................................... 18
Figure 4. CLC Genomics Data ......................................................................... 21
Figure 5. GSEA Analysis ................................................................................. 23
Figure 6. IL-1β Concentration ......................................................................... 24
Figure 7. THP1 Biotinylation ............................................................................ 25
Supplemental Figure 1. NLRC4 Chromosome and Domain Schematic .............. 29
Supplemental Figure 2. NLRP3 Chromosome and Domain Schematic ............... 29
Supplemental Figure 3. NLRC4 Plasmid Design ............................................... 30
Supplemental Figure 4. NLRP3 Plasmid Design ............................................... 31
Supplemental Figure 5. NLRP3 HEK GFP Flow Cytometry ............................... 32
Supplemental Figure 6. Unbiased Heat Map ..................................................... 33
Preface

Acknowledgements:

I’d like to thank the members of my lab for assisting me in every step of this process. I’d like to give extra recognition to my P.I., Dr. Scott Canna, who continued to believe in me and never lost hope of creating a great thesis presentation even during a global pandemic. Appreciation is needed for Dr. Bill Hawse and Richard Cattley from the University of Pittsburgh, Department of Immunology, who handled our mass spectrometry analysis. Additionally, to Bing Jiang from BioInnovatise who aided in creating our plasmid constructs.

Nomenclature used:

Monogenic autoinflammatory diseases (AIDs); pathogen-associated molecular patterns (PAMPs); damage-associated molecular patterns (DAMPs); Pattern Recognition Receptors (PRRs); Toll-like Receptor (TLR); Lipopolysaccharide (LPS); Gasdermin-D (GSDMD); Interleukin-1β (IL-1β); Interleukin-18 (IL-18); wildtype (WT); RNA-sequencing (RNA-seq); phorbol myristate acetate (PMA); gain-of-function (GOF); nucleotide-binding domain, leucine-rich repeat containing (NLR); NLR family CARD domain-containing protein 4 (NLRC4); NLR family pyrin domain containing 3 (NLRP3); Familial Cold-Induced Autoinflammatory Syndrome (FCAS); Macrophage Activation Syndrome (MAS); mitochondrial reactive oxygen species (mtROS); NLR family, apoptosis inhibitory proteins (NAIPs); Muckle Wells Syndrome (MWS); autoinflammation with infantile enterocolitis (AIEFC); neonatal onset multisystem inflammatory disease (NOMID); proximity labeling (PL); BirA miniTurbo (BirAMT); Human acute monocytic leukemia cell line (THP-1); protein-protein interactions (PPI); co-immunoprecipitation (Co-IP);
Gene Set Enrichment Analysis (GSEA); transcripts per kilobase million (TPM); interferon-response genes (IRG); SV40 large T antigen (HEK293T); interferon-α (IFNα); nominal molecular weight limit (NMWL); principal component analysis (PCA); empty vector (EV); Brilliant Violet 421 (BV421-A).
1.0 Background

Monogenic autoinflammatory diseases (AIDs) are caused by highly penetrant genetic variants in single genes and follow a Mendelian inheritance pattern.¹ A particular subset of this disease group centers around defects in the inflammasome and are subsequently referred to as inflammasomopathies.² Inflammasomes are multi-protein complexes that are formed as a host defense mechanism against intracellular danger or damage signals by the cell in the presence of infectious agents or harm to the body.³ Though the features of the inflammasomopathies are all in the autoinflammatory spectrum, the specific phenotypes consistently associated with certain mutations can vary substantially.⁴ Importantly, the in vitro consequences of inflammasome activation studied to date do not reflect this clinical heterogeneity: all inflammasome-activating mutations have a similar in vitro phenotype regardless of which inflammasome is involved or which specific mutation is causing a patient’s disease. Recent discoveries have led to a better understanding of how these pathways work but there is still more that needs to be understood in order to correctly diagnose these individuals.⁵ This study explores the mutations present in particular inflammasome genes that lead to the varied levels of inflammation found in patients with these diseases.

Our bodies have evolved to recognize what is dangerous and what is homeostatic. Inflammation is a defense mechanism initiated in response to infection and tissue damage to control the level of harm that is enacted on the body by infectious microbes and viruses.⁶ There are two main categories by which the innate immune system recognizes such threats and initiates inflammation. The first are pathogen-associated molecular patterns (PAMPs), which are conserved molecular features found in many different kinds of microbes and viruses that trigger innate
inflammatory responses, particularly when sensed in pathologic locations (e.g. LPS or methylated CpG DNA in a cell’s cytosol). The other type of molecules that drives innate inflammatory signals are damage-associated molecular patterns (DAMPs). DAMPs are molecules, or even disturbances in typical homeostasis, that derive not from infectious microbes but from danger or damage signals that are routinely caused by infection or trauma (e.g. extracellular molecules typically relegated to the intracellular space like ATP or HMGB1, or disturbances of homeostasis such as actin-cytoskeleton disruption or proteasome inhibition). These specialized features have advanced not only detect what is foreign to the body but also what type of foreign object it is, whether it be harmless non-self or pathogen-associated non-self. Both PAMPs and DAMPs are sensed by germ-line encoded Pattern Recognition Receptors (PRRs); the most famous of which is Toll-like Receptor (TLR) 4, which responds to the bacterial cell wall component Lipopolysaccharide (LPS) and whose discovery resulted in the 2011 Nobel Prize in Physiology/Medicine.

Inflammasomes are found in the innate immune system that regulates the release of soluble inflammatory mediators in response to cytosolic danger signals and leads to pyroptotic cell death. There are three elements that make up the inflammasome: a single or group of molecules that function as molecular Pattern Recognition Receptors (PRRs) and nucleate inflammasome formation; the adaptor ASC; and the protease caspase-1. Inflammasome activation triggers the activation of caspase-1 and thereby maturation of inflammasome substrates. The inflammasome substrates known to date are proIL-1β, proIL-18, and Gasdermin-D (GSDMD). Mature Interleukin-1β (IL-1β) and Interleukin-18 (IL-18) are proinflammatory cytokines that require plasma membrane pores produced by active Gasdermin-D for their release from the cell. IL-1β and IL-18 are important proinflammatory mediators that are often abundantly present in cells where inflammasome activation is initiated due to injury or immunological challenge. Production
of these inflammatory mediators is an important part of host defense, but their overproduction (as observed in the genetically-mediated inflammasomopathies) can lead to chronic, abnormal inflammation and resultant tissue damage.

Mutations in genes involved in inflammasome formation/regulation have been shown to cause monogenic autoinflammatory diseases with varying degrees of severity.\textsuperscript{9} Due to their heterogeneous nature and diversity of symptoms, the diseases associated with these inborn errors of immunity are rare and difficult to diagnose due to their infrequency and are often caught in the later stages of progression, which in turn makes it harder to treat them.\textsuperscript{11} Importantly, studying monogenic diseases like inflammasomopathies yields mechanistic insights generalizable to more common diseases. Herein, we are trying to determine what is causing inflammasomes to become hyperactive as a consequence of the mutations associated with inflammasomopathies in order to better understand how they result in these phenotypes.

My thesis was divided in two parts that spanned the scientific enterprise. Part one of this two-part study focused on the analysis of RNA-sequencing (RNA-seq) data generated to better understand the mechanisms by which gain-of-function (GOF) mutations in the inflammasome-nucleating protein NLRC4 cause disease. The second part of this study, initiated in part based on results found in Part 1, was intended to develop a system to facilitate the unbiased analysis of the protein-protein interactions (PPi’s) associated with both the NLRC4 and NLRP3 inflammasomes, both under normal circumstances and with several polymorphic GOF mutations.
1.1 Previous Research

In Weiss et al.\textsuperscript{12}, they determined that excess IL-18 was uniquely associated with NLRC4 hyperactivity. We hypothesized that macrophage-intrinsic transcriptional changes may be associated with the selective overproduction observed in NLRC4-related diseases. We are answering this by investigating the transcriptional effects of a GOF NLRC4 inflammasome mutation in phorbol myristate acetate (PMA)-stimulated THP1 human myelomonocytic cells that had been stably transduced to express either WT or an NLRC4 gain-of-function mutation (Thr 337 Ser). It was previously discovered that macrophages derived from NLRC4-transduced THP1 cells showed a 5-fold increase in secretion of IL-1β and IL-18 in the WT cells as compared to the empty vector (EV) and nearly a 10-fold increase in secretion of IL-1β and IL-18 in the NLRC4 GOF mutant cells. These findings support gain-of-function of the NLRC4 T337S-mutation through the increased secretion of the inflammatory cytokines in the mutant form.\textsuperscript{9} We are hoping to see a similar pattern in different mutations within the same gene as well as “common” mutations found in other genes to understand the mechanisms behind these diseases.

Though the first part of my thesis was to further explore previous research conducted on only NLRC4 wildtype (WT) and the GOF T337S mutation, the second part was to develop a novel strategy for assessing the effects of GOF inflammasome mutations and whether changes in their protein-protein interactome might help explain differences in phenotype and disease severity within and between NLRP3 and NLRC4 inflammasomopathies.
1.2 Inflammasome Activation

Inflammasomes are important complexes comprised of multiple proteins that are located in the cytosol of cells and are triggered in the presence of danger signals. The inflammasome-type activated is specialized based on the signal that is received by the innate immune system. These complexes are composed of many different types of proteins. The three categories of proteins are platform proteins, which involve our proteins of interest in this paper, are used for assembly of the inflammasome; adaptor proteins, called ASC; and effector proteins that are caspase-domain containing and include caspase-1, which is relevant to this study.

An essential protein that facilitates the secretion of IL-1β and IL-18, as well as pyroptosis downstream of inflammasome activation is Gasdermin-D (GSDMD). GSDMD is a known substrate of caspase-1, an inflammatory caspase that is fundamental to inflammasome-mediated pyroptosis, as well as the maturation of cytokines. It is known that pro-caspase-1 activation is important to mediate pyroptosis and facilitate the production of IL-1β and IL-18. Additional discoveries have uncovered that when macrophages are stimulated with lipopolysaccharides (LPS), GSDMD is recruited to NLRP3 in a similar fashion that pro-caspase-1 is, indicating its importance to the activation of inflammasomes. GSDMD has been proven important in both canonical (constantly active pathway) and non-canonical (signal-induced pathway) inflammasome signaling.

There are two cellular outcomes that may result from the activation of caspase-1. The first is among the most relevant to this study and our understanding of how this system works through means of concentration analysis. Cells stimulated with ATP, silica, and uric acid crystals maturate pro-inflammatory cytokines like IL-1β and IL-18, which lead to inflammation in the body. Over-production of these cytokines can be detected in the media of cells created to act this way in vitro.
The other result of the activation of caspase-1 is pyroptosis or proinflammatory cell death. This outcome plays a role in clearing pathogens from the cell, as well as providing immunity to the attacking infection.\textsuperscript{17} It is also evident that both pyroptosis and GSDMD are important in plasma membrane pore formation that releases cytoplasmic content, such as IL-1β and IL-18 out of the cell.\textsuperscript{18}

Although activation of these inflammasomes can be triggered due to infection or invasion, there are some genetic mutations or predispositions that cause these pathways to become constitutively active in the body. They are referred to as inflammasomopathies and are often very difficult to diagnose due their sparsity in populations and particular phenotypes. Due to the variability found in this subset of disease, there isn’t much known about the mechanisms behind them. This study aims to better understand the proteins that are interacting with specific disease-correlated mutations associated to better understand these internal mechanisms behind these inflammation pathways.

1.3 Genes of Interest with Gain-of-Function Mutations

The NLR (nucleotide-binding domain, leucine-rich repeat containing) family of proteins are crucial components of inflammasome activation. This study primarily focuses on two members of the NLR family of genes, which are further classified into subfamilies by their N-terminal effector domains.\textsuperscript{19} We are looking to investigate how different gain-of-function (GOF) mutations within \textit{NLRC4} (NLR family CARD domain-containing protein 4) and \textit{NLRP3} (NLR family pyrin domain containing 3)\textsuperscript{19} inflammasome activation can result in a wide array of phenotypes, ranging
from the milder Familial Cold-Induced Autoinflammatory Syndrome (FCAS) to the life-threatening Macrophage Activation Syndrome (MAS).

It is evident that NLRC4 and NLRP3 inflammasomes are affected by different interactors and regulators in order to control their functions. The presence of NEK7, a mitotic kinase, is thought to aid in the assembly, as well as the activation of the NLRP3 inflammasome. These kinases assist in the connection of neighboring NLRP3 subunits to create and activate the NLRP3 inflammasome. 

Ca^{2+} signaling is also important for the activation of the NLRP3 inflammasome. Although the mechanism behind this process is still unknown, it is thought that an overload of mitochondrial Ca^{2+} can lead to mitochondrial damage, which leads to mitochondrial reactive oxygen species (mtROS) production and this formation of species is thought to cause NLRP3 inflammasome activation by interacting with NEK7. Conversely, A20, an NF-κB inhibitor stops the spontaneous secretion of IL-1β by ubiquitinating the pro-IL-1β center K133. It is also noted that these processes are not necessary for NLRC4 activation.

Unlike the NLRP3 inflammasome, it is crucial that the NLRC4 inflammasome gets phosphorylated to initiate activation. PKC-δ is a kinase that is thought to be a contributing factor to this process. PKC-δ phosphorylates NLRC4 at Ser 533, which facilitates conformational changes that are necessary for NLRC4 inflammasome activity and how it plays a role in innate immunity. Additionally, it is known that specific NLR family, apoptosis inhibitory proteins (NAIPs) interact directly with flagellin or T3SS components like PrgJ, that in result, engage NLRC4.

NLRC4 and NLRP3 are triggered in different ways but their mutations result in nearly identical in vitro phenotypes: increased basal and stimulated inflammasome activation and IL-1β/IL-18 secretion. Clinically, the diseases associated with GOF defects overlap only a little. NLRP3
mutations cause a spectrum of mutations known as the Cryopyrin-Associated Periodic Syndromes (CAPS). CAPS are comprised of the relatively mild FCAS, the more severe Muckle Wells Syndrome (MWS) that includes progressive sensorineural hearing loss, and the potentially devastating Neonatal Onset Multisystem Inflammatory Disease (NOMID) that includes CNS inflammation, bony overgrowth of large joints, and severe neurological manifestations.\textsuperscript{25} For its part, the spectrum of \textit{NLRC4}-related diseases includes Cold-induced urticaria for some mutations in some families, and a high-fatality hyperinflammatory state with MAS and infantile enterocolitis caused by different mutations in other patients. We are interested in determining the molecular mechanisms underlying the genotype-phenotype correlations observed in monogenic autoinflammatory diseases caused by \textit{NLRC4} or \textit{NLRP3} mutations. In vitro, activation of either drives the production and release of the pro-inflammatory cytokines IL-1β and IL-18, as well as provoking pyroptotic death in cells through the activation of caspase-1.\textsuperscript{12}

It is important to note that inflammasomes often recruit various NLRs and effectors as a response to invasion in the body. Although a lot of the mechanism is still unknown about how these macromolecules interact, it is evident that they could be working together in some manner to enact the immune response seen in inflammation.\textsuperscript{26} Since \textit{NLRC4} and \textit{NLRP3} have relatively the same functional properties, we want to determine the molecules interacting with their mutated forms to identify the key regulators of this inflammatory pathway and potential disease modifiers.

The mutations targeted in this study were derived from previous literature as being associated with these inflammasomopathies. The point mutations that are commonly found in \textit{NLRC4} are a threonine to a serine change at amino acid 337 (T337S), which results in autoinflammation with infantile enterocolitis (AIEFC); a valine to alanine change at amino acid 341 (V341A) that results in AIEFC; a histidine to a proline change at amino acid 443 (H443P) that
results in familial cold autoinflammatory syndrome (FCAS); and a tryptophan to cysteine change at amino acid 655 (W655C), which results in macrophage activation syndrome (MAS).27 The point mutations that are commonly found in NLRP3 change an aspartic acid to an asparagine at amino acid 303 (D303N), which results in neonatal onset multisystem inflammatory disease (NOMID); a leucine to a proline at amino acid 353 (L353P), which results in FCAS; and an alanine to a valine at amino acid 495 (A495V), which results in Muckle-Wells Syndrome.27 The varied phenotypes are important to this study because we would like to identify differences in protein-protein interactions (PPI) partners that may help differentiate between the different diseases associated with these genes.

Table 1. Genetic Mutations & Phenotypic Correlation

<table>
<thead>
<tr>
<th>Gene/Bait</th>
<th>Mutation</th>
<th>Human Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>BirAmp alone</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NLRP3</td>
<td>WT</td>
<td>None</td>
</tr>
<tr>
<td>NLRP3</td>
<td>D303N</td>
<td>NOMID</td>
</tr>
<tr>
<td>NLRP3</td>
<td>L353P</td>
<td>FCAS</td>
</tr>
<tr>
<td>NLRP3</td>
<td>A495V</td>
<td>Muckle-Wells Syndrome</td>
</tr>
<tr>
<td>NLRC4</td>
<td>WT</td>
<td>None</td>
</tr>
<tr>
<td>NLRC4</td>
<td>T337S</td>
<td>AIFEC</td>
</tr>
<tr>
<td>NLRC4</td>
<td>V341A</td>
<td>AIFEC</td>
</tr>
<tr>
<td>NLRC4</td>
<td>H443P</td>
<td>FCAS</td>
</tr>
<tr>
<td>NLRC4</td>
<td>W665C</td>
<td>MAS</td>
</tr>
</tbody>
</table>

Table 1: Gene/mutation pairs to be fused to the BirAmp (MiniTurbo7) biotinylase. AIFEC: Autoinflammation with Infantile Enterocolitis; MAS: Macrophage Activation Syndrome; NOMID: Neonatal-Onset Multisystem Inflammatory Disease; FCAS: Familial Cold-Induced Autoinflammatory Syndrome
1.4 The Interactome and BirA miniTurbo

A protein’s “interactome” is a network of interactions that occur between different cellular components. The focus of Part 2 of this study was to evaluate protein-protein interactions (PPI) associated with particular mutations within our genes of interest. The use of proximity labeling due to biotinylation of the surrounding proteins interacting with each gene of interest is the method we chose to identify these PPI’s. We chose this method because other methods of identifying PPI, namely through co-immunoprecipitation (Co-IP), rely on lysing the cells in conditions in which interacting proteins remain attached. Such lysis conditions can vary widely depending on the interaction. Inflammasomes exist in large, insoluble complexes making them impossible to assess by classic IP. Likewise, the affinity and epitope of the antibody used greatly affect what interacting proteins are “pulled down.” The approach described herein overcomes many of these obstacles.

By investigating the interactome surrounding each of the genes of interest, NLRC4 and NLRP3, we hope to establish which protein interactions are affected by different mutations occurring in these genes. Once determined, these interactions intend to inform future efforts to establish an appropriate course of treatment that is specific to each variation of these inflammasomopathies.

In order for us to determine the interactome for these genes, we needed something in place in order to identify the interacting proteins. Enzyme-catalyzed proximity labeling (PL) is used to do just that. The enzyme we chose was the attachment of BirA, a mutant of E. coli enzyme biotin ligase with the MiniTurbo, a series of mutations in the original E.coli BirA that make it more efficient. This is placed following our established protein constructs for both NLRC4 and NLRP3. We chose this combination because of BirA’s affinity to bind to biotin and for miniTurbo’s ability to biotinylate endogenous proteins at a quicker rate than BioID. In the presence of exogenous
biotin, proteins that are within a radius of 2Å of the gene will be marked with biotin. Cells are then lysed under harsh, fully degrading conditions that can disassemble insoluble complexes. Streptavidin-coated magnetic beads are used to isolate biotinylated proteins, which are then digested, and the peptide fragments further analyzed through mass spectrometry. Mapping of the detected peptide fragments onto the known amino acid sequences of murine proteins will then determine which proteins are interacting with our genes of interest.

This method has been validated by Li et al.\cite{31}, where they created a workflow protocol in order to achieve success in these types of experiments. They found that this method of PPI detection had a high sensitivity and low false positive rates for many different levels of protein expression. The BirA is attached to the c-terminus of the plasmid and determines proteins that are within 2Å of the insert.\cite{31} This will be useful to our study in order to determine the interactome that is surrounding our genes of interest and will hopefully increase our understand behind the mechanisms of inflammasomopathies.

1.5 Specific Aims

Hypothesis: If inflammasome activation is more so present in gain-of-function mutations, then we can identify novel regulators of this key inflammatory pathway through the protein interactions gained or lost between wild-type and mutant forms of \textit{NLRP3} and \textit{NLRC4}.

Study Objectives: 1) Investigate hyperactive NLRC4 by analyzing the RNA-seq data from the monocytes with wild-type or mutant \textit{NLRC4} RNA-seq data. 2) Conduct a collaborative pilot
project to develop and use a novel and scalable PPI mapping method for assessing the effects of individual mutations in inflammatory genes.

**AIM 1:** Analyze RNA-seq data to further understand the effects of gain-of-function mutations by comparing the results from cell lines with wild-type or mutant *NLRC4*.

Filter data files in order to investigate patterns found in interferon activation caused by GOF mutations in *NLRC4*. Bioinformatic analysis conducted on multiple databases to enact informative conclusions about our wild-type or mutant *NLRC4* cell lines.

**AIM 2:** Create plasmid constructs containing the *NLRP3* or *NLRC4* insert. Generate stable human THP1 monocytic cell lines expressing *NLRP3* or *NLRC4* (WT or known GOF mutants) c-terminally fused to the BirA miniTurbo (BirA\textsuperscript{MT}) biotin ligase\textsuperscript{7}.

Plasmid constructs generated on retro-viral backbones were used in order to assess the differing interactomes associated with each mutation in our genes of interest. Thp1 cells were then transduced with the MigR1 retrovirus containing just miniTurbo (our empty vector) or the WT *NLRC4* and *NLRP3* or mutant retroviruses for each gene in order to test the levels of proinflammatory proteins associated with each disease-causing mutation. We exposed our cells to excess biotin in their media in order for the biotin to attach to the BirA insert and the surrounding proteins, within close proximity of the mutation of interest. Mass spectrometry was used to identify which proteins became biotinylated due to close proximity to the gene of interest.
2.0 Methods

RNA-Sequencing and Analysis

Human acute monocytic leukemia cell line (THP-1)\textsuperscript{32} monocytes transduced with a retroviral construct containing either WT or T337S mutant \textit{NLRC4} were stimulated as in Chae et al\textsuperscript{33} with PMA. Sample timepoints were collected at 0, 6, 24 and 72 with 3 samples for the WT and mutant at each timepoint. The RNA quality for one of the WT 6-hour technical replicates was below threshold and therefore was not sequenced, leaving only 2 replicates for this timepoint.

The materials and methods of this current study have/have not been modified from (Weiss et al., 2018). For RNA-seq, “RNA integrity was analyzed with the Agilent 2200 Tapestation.”\textsuperscript{12} “Messenger RNA purification and fragmentation, complementary DNA synthesis and target amplification were performed with Illumina TruSeq RNA Sample Preparation kit. Pooled complementary DNA libraries were sequenced on Illumina HiSequation 2000 platform.”\textsuperscript{12} Single-end sequencing was conducted and RAW FASTq files were generated.

Analysis was performed using CLC Genomics Workbench 20 (v20.0.4). FASTq files were uploaded, the adapter reads were imported and trimmed. Data was compared to Homo_sapiens_hg19_sequence. Gene expression data was filtered for a transcripts per kilobase million (TPM) \( \geq 1.5 \) and exported to an excel file for further analysis using different programs. That file was uploaded to Gene Set Enrichment Analysis (GSEA). The gene set database used to analyze this data set (HALLMARK_INTERFERON_ALPHA_RESPONSE.gmt) and the chip platform (Human_SYMBOL_with_Remapping_MsigDB.v7.1.chip) were used for further exploration of the RNA-seq data. Additionally, heatmaps were generated using the exported excel file using the Broad Institutes online software Morpheus.
Samples were compared in the heatmap to a 28 gene interferon score list generated from previous literature on interferon-response genes (IRG).33

**Plasmid Constructs**

We started with partially established plasmids used in Weiss et al.12 and confirmed the correct sequence of these plasmids against NCBI (https://www.ncbi.nlm.nih.gov) cDNA sequences (NLRC4: NM_001199139 and NLRP3: NM__001079821.3). Plasmid maps of the goal plasmids were generated using SnapGene 5.1.7 and sent to Bioinnovatise, Inc (Rockville, MD) to create the plasmids. From there, they inserted the specific construct we wanted the plasmid to have. Plasmid constructs were generated to contain a MigR1 retrovirus backbone with the addition of WT NLRC4 or NLRP3, along with the BirA_MT attached to the c-terminus. It is important to note that these constructs are retroviral vectors used for their ability to convert a single stranded RNA genome into a double stranded molecule of DNA that is able to incorporate into the genome of the target cell.34 The mutations of interest were generated using site-directed mutagenesis and complementary primers containing the specific mutation.

**Generation of stably-transduced THP1 myelomonocytic cells**

Once the finished plasmids were received, the bacterial transformation started. We used electroporation in order for the plasmids to pass through the bacterial cell wall and membrane. Our plasmids were placed into cuvettes along with electrocompetent cells and pulsed using the 1.8Kv bacterial setting. The bacteria were then streaked out on to 10cm dishes containing LB and ampicillin-treated agar and allowed to grow at 37ºC overnight. Single colonies were selected after growth and placed into 35ml of LB media spiked with ampicillin overnight. The following
morning, the media was spun down to separate out the bacteria and a Qiagen Rapid Vacuum Midi Prep (Cat No./ID: 12943) was used following the high-yield, high-copy plasmid protocol. DNA concentration was determined using a NanoDrop spectrophotometer. Plasmid DNA was confirmed using a restriction enzyme digest (Figure 1 a & b). For *NLRC4*, the restriction enzymes used were SalI and Sac1 and for NLRP3, the restriction enzymes were NotI and Sac1.

![Figure 1](image.png)

**Figure 1. Construct Protein Digestion**

(a) Digest of plasmid DNA from NLRC4 to confirm correct plasmid. The restriction enzymes used are SalI and Sac1. The expected products for each of the plasmids is EV: 3.6, 2.1, 1.2 & 0.4 kb, N4 WT: 5.5, 3.6 & 0.4 kb, N4 WT miniT and mutants: 4.3, 3.6, 2.0 & 0.4 kb. N4 WT miniT contains an extra band that may be due to incomplete protein digestion. Ladder values are in kilobase (Columns labeled WTminiT NES and V341A were added but do not contain information relevant to these findings). (b) Digest of plasmid DNA from NLRP3 to confirm correct plasmid. The restriction enzymes used are NotI and Sac1. The expected products for each of the plasmids is EV: 3.7 & 3.6 kb, N3 WT: 5.9 & 3.6 kb, N3 WT miniT and mutants: 4.2, 3.6 & 2.6 kb. Ladder values are in kilobase. Mutants (last three columns) also contain miniT.

Transfection is the process of physically introducing foreign DNA into target cells to express proteins encoded on that DNA. Human Embryonic Kidney cells that express a mutant
version of the SV40 large T antigen (HEK293T) cells were passaged in D10 media at a concentration of \( \sim 3 \times 10^6 \) cells in a 10cm dish and left to double in number and adhere to the plate overnight at 37°C to reach 70-90% confluency for the transfection. Appropriate concentrations of DNA, both MigR1 and pCL-ampho plasmids, are combined with Lipofectamine 3000 and optimem media and added to the 10cm dish and put back into the incubator for 3 days. It is important to note the need for pCL-ampho. This plasmid is a packaging vector that expresses high levels of Gag-Pol-4070A ENV to allow for our retroviral DNA to be inserted in these macrophages produced by these cells.

Figure 2. HEK 293T Transfection (a) GFP levels found in HEK 293T cells post transfection with \textit{NLRC4} plasmid. Images captured using EVOS FL Cell image transmitted and the GFP fluorescent channel. Scale bar measure is 200um (b) GFP levels found in HEK 293T cells post transfection with \textit{NLRP3} plasmid. Images captured using EVOS FL Cell image transmitted and the GFP fluorescent channel. Scale bar measure is 200um.
After those 72 hours, the HEK cells have generated virus in which the NLRC4-BirA insert can be integrated into the genome of our THP1 cells: a process called transduction. The media from the HEK cells is pulled off and spun gently to separate out any HEK cells that may have died and detached. The HEK cells are imaged to determine GFP intensity post transfection (Figure 2 a & b) using an EVOS FL Cell Imaging System under transmitted and the GFP fluorescent channel. The 2ml of the media containing the virus is placed straight on top of Thp1 cells with additional clean media, while the remaining media concentrated using centrifuge concentrators. These Amicon® Ultra-4 Centrifugal Filter Unit (UFC801008) filter 10 KDa nominal molecular weight limit (NMWL), which allows us to have a higher abundance of macrophages in the small amount of media that is place on top of the Thp1 cells. The virus media, concentrated or not, are placed gently over Thp1 cells in 16 well plates and are spinoculated, to increase viral infection efficiency, in a centrifuge for 1.5 hours at 2200 rpm and placed in 32°C overnight and let to grow up in 37 °C. After the cells had time to grow, they are sorted for having green fluorescent protein (GFP) using a BD FACS Aria II system (Figure 3). We are aiming for 90% GFP and above but the higher the percent, the better.
Figure 3. Transduction Efficiency a) GFP levels found in Thp1 cells post transduction and sorting with NLRC4 plasmid. Cells were flowed using BD FACSARia II system. GFP is indicated as a percent (%) (b) GFP
levels found in Thp1 cells post transduction and sorting with NLRP3 plasmid. Cells were flowed using BD FACSaria II system. GFP is indicated as a percent (%).

**Investigating the Interactome**

Sorted cells (WT, EV and each mutation) were tested to determine their interactome via biotinylation. First, they were exposed to LPS at 100ng/mL, to generate activated THP1 “macrophages”. Then the cells were incubated with exogenous biotin and 100μl samples of each cell type were taken 12 and 24 hours later. An enzyme-linked immunosorbent assay (ELISA) was performed to indicate concentration of certain cytokines in the transduced cells, particularly IL-1β. Analysis was carried out using GraphPad Prism v9.

Additionally, cells were prepped for mass spectrometry. Mass spectrometry is used to quantitatively identify protein fragments found in our cells. Our biotinylated cells were isolated and prepped for mass spectrometry by Richard Cattley from Dr. Bill Hawse’s Lab in the Department of Immunology at the University of Pittsburgh. They sent these prepped cells to the Notre Dame Mass Spectrometry and Proteomics Facility. The cells were lysed under denaturing conditions and the lysate was incubated with streptavidin-coated magnetic beads. The proteins that were biotinylated are attach to the coated beads and the non-adherent proteins were washed out thoroughly. The streptavidin-adherent proteins are refined, and tryptic fragments are generated through filter-aided preparation method. The proteins were then analyzed using a Dionix reverse phase liquid chromatography system with a New Objective PicoChip nanospray column in line with a ThermoFisher Q Exactive mass spectrometer. Output was compiled into excel and analysis was conducted.
3.0 Results and Discussion

Part 1:

RNAseq Analysis

According to our findings, we determined that each timepoint clustered together in the principal component analysis (PCA), which indicates that each timepoint was most similar to each other and not like those of other times (Figure 4a). It appears that there is a clear separation between the timepoints in this graph indicating that each sample, whether WT or mutant, is similar to others at that same time point. The only one where this did not hold true is at the 72-hour mark where the WT and mutant samples are distinct from one another based on genotype but still grouped by timepoint. This could be caused by the fact that the interleukin response in cells comes first and causes them to express a multitude of genes at earlier timepoints. It could also be that the NLRC4 mutation causes changes that accumulate over time as well as secondary/compensatory effects that themselves accumulate.

Figure 4b depicts a heatmap of our RNA-seq data at all 4 timepoints in both wildtype and mutant. The output was filtered to only include genes with a TPM ≥ 1.5. The figure shows that at the 24-hour timepoint, a majority of the interferon associated genes are hyperactive in the mutant as compared to the WT and even more so increased from the 0-hour timepoint (Figure 4b). An additional heatmap (supplemental figure 6) was generated in order to create an unbiased view of the dataset. The conclusion that can be made is that cells stably transduced with the patient-derived T337S mutant NLRC4 show a more substantial upregulation of IRGs after 24 hours of stimulation than their WT counterparts or cells at other timepoints. A different set of IRGs, as well as IL-18, are upregulated at 72 hours primarily in NLRC4-T337S transduced cells.
Figure 4. CLC Genomics Data (a) Principal component analysis (PCA) plot derived from CLC Genomics Workbench 20 of *NLRC4* WT and mutant. Time point is listed in hours. (b) Heatmap generated from RNA-
seq analysis with TPM ≥ 1.5. Gene set derived from previous literature of a 28-IFN response gene scoring system. Blue indicates a low value; red indicates a high value.

Further analysis done through GSEA shows at the 24-hour timepoint that there is a positive correlation associated with the NLRC4 mutant cells as compared to the WT NLRC4 cells for genes included in the hallmark interferon-α (IFNα) response (Figure 5a). Comparing figures 5a and 5b, we can also see that there are more black lines located around the NLRC4-Mutant side of the chart in the 24-hour timepoint only than there are in the GSEA plot that compares all of the timepoints (Figure 5b). This indicates that at the 24-hour timepoint, the comparison of NLRC4-T337S-transduced to wtNLRC4-transduced cells is more strongly enriched for IFNα response genes. Additionally, the enrichment score for all of the timepoints appears to be lower than just the 24-hour timepoint, displaying the significance of how densely packed the enrichment is at that particular timepoint in the mutant cells versus the WT cells as compared to overall. This confirms the observation indicated above and would advance our previous findings from the heatmap by also depicting a hyperactivity of interferon-associated genes at the 24 hours of PMA stimulation. From this was concluded that NLRC4 might interact with interferon stimulated genes.
Figure 5. GSEA Analysis
(a) GSEA results from the 24-hour timepoint only, table below image is an additional output of statistical analysis
(b) GSEA results from all timepoints, table below image is an additional output of statistical analysis

Part 2:
Inflammasome Activity

Inborn errors of immunity may be an initiating factor in the hyperactivation of the inflammasome. It has been previously shown that in vitro PMA-stimulation, IL-1β and IL-18 production are increased in NLRC4-T337S-transduced cells more so than in WT-NLRC4 transduced cells\(^5\), and we were trying to replicate these results within our own samples which have the addition of BirA\(^\text{MT}\). Through the mutations generated into our plasmids (Supplemental Figures
1 & 2), confirmed via restriction enzyme digestion (Figure 1) and assessed for overall GFP (Figure 2 & 3 and Supplemental Figure 3), we were able to replicate the types of cells in vitro to understand what could be occurring in vivo. We detected an increase in IL-1β more so in the mutant forms of *NLRC4* as compared to the WT, with the exception of mutant V341A (Figure 6). However, we would have liked for our normal, untransduced Thp1 cells and empty vector (EV) to be lower than they are depicted. This is something we are still focusing on to confirm that these mutated cells are in fact showing increased inflammasome activity in the mutant versus WT-transduced cells.

![Image of IL-1β Concentration](image)

**Figure 6. IL-1β Concentration** An ELISA was done in order to determine levels of IL-1β. Samples used were diluted 4-fold but data accounts for the dilution.

Nevertheless, to determine whether the THP1 monocytes stably transduced with the *NLRC4*-BirA-miniTurbo fusion proteins (WT or mutants of *NLRC4*-BirA<sup>MT</sup>), we treated these cells with biotin in culture with and without PMA and assessed them for global increase in biotinylation using intracellular staining with Streptavidin conjugated to the fluorophore BV421 (SA-BV). As this was simply a global pilot, we stimulated cells for 16 hours: far longer than should...
be necessary for BirA\textsuperscript{MT}. In the presence of biotin, the Brilliant Violet 421 (BV421-A) is increased in both of the biological replicates of the \textit{NLRC4} T337S mutant as compared to the WT (Figure 7). We hope that future replications of this experiment will continue to show us this data. Mass spectrometry data will assess which proteins are in close proximity to our genes of interest when in these aggravative states.

![Figure 7. THP1 Biotinylation Assessment of biotinylation via flow cytometry of Thp1 cells after 16hr PMA stim at 2ng/ml with 500mMolar Biotin Tryp LE added for 24 hours.](image)

25
3.1 Public Health Relevance

The most significant relevance to public health in this study is the assessment of disease to diagnose and investigate phenotypic outcomes based on various mutations within the same gene. We are studying these genetic mechanisms in order to obtain an understanding as to why these mutations result in the particular phenotype in patients. With the increasing accessibility of genetic sequencing and the power of modern genomic and proteomic analysis, health care providers have the opportunity to explore precision medicine by looking at differentially expressed genes when deciding on the best therapies. Our study aims to stimulate new ideas brought about by the more hopeful results that could lead to a strategy to better diagnose these patients and potentially stem into a targeted treatment in the future.

3.2 Future Directions

For the RNA-seq portion of this study, there is additional data that are being analyzed investigating why MAS has excess levels of IL-18. Further experiments may still be needed in order to fully understand the reasons why there are higher interferon numbers at the 24-hour timepoint present in the mutated cells. Although it We hope to establish a better understanding of these outcomes in order to enact a more solid mechanism behind these inflammasome mutations in NLRC4.

Though we have fairly exhaustively worked to confirm that our plasmid constructs are correct, confirmation of correct mutations generated in the plasmid constructs may need to be
investigated in more detail, in addition to if the protein is folding correctly given the c-terminal fusion. There appears to be some complication in terms of the ELISA assay that needs to be further investigated as well. The mutant versions are expected to generate more IL-1β than the WT miniT and both significantly higher than regular THP1 or THP1 transduced with EV.

Mass spectrometry data are in the process of being analyzed and we hope to see some proteins that are strongly related to inflammation present. This will hopefully help us determine the interactome surrounding these mutated genes and be able to tell us more about the proteins interacting and potentially causing the systemic inflammation found in these diseases. We hope that this projects will continue to progress so we can increase the scope of which interactomes we can map via similar methods. In addition to the genes we have already begun investigating, *NLRC4* and *NRLP3*, we would like to be able to investigate other genes with gain-of-function mutations to create a much larger database in order to categorize these differences among mutations in autoinflammation as a standard method for investigating gain-of-function IEIs. There is potential for this project to progress more smoothly once we establish the basic techniques, which will also help us to further understand the interactomes that are associated with other diseases similar to those included in this study.

### 3.3 Limitations

The cell type used in the transduction may be considered a limiting factor in this project. Specifically, in the *NLRP3* model, plain THP1 cells spontaneously express *NLRP3*, which created an obstacle for us. In this study, we decided to use THP1 cells in which *NLRP3* had been knocked-out to try to avoid this obstacle as a whole.
The largest factor that contributes to limitations of this study is time. I was a part of this lab for 1 year and 3 months, in which we were able to get things started but not a sufficient amount of time in order to get significant results. Another contributor was the fact that the COVID-19 pandemic occurred just a few months into my research, which created the need for the two-part project that is presented. Unsure of how long we would be out of the lab so remote work needed to be done in order further along my thesis project. The data mining journey ended up being beneficial not only for myself but also other projects within the lab. I was able to learn valuable analytical skills while furthering my thesis research project. I would have also liked more time to tackle this project. Just over a year was not enough time for me to see this develop into anything more and I plan on staying update to date in order to see this creation come to life and help others.
Appendix A Supplemental Material

Supplemental Figure 1. NLRC4 Chromosome and Domain Schematic Depicts location of gene on chromosome 2, the domains present in that gene and the location of the common disease mutations present in this study (copyright of Canna et al., 2014)

Supplemental Figure 2. NLRP3 Chromosome and Domain Schematic Depicts location of gene on chromosome 1, the domains present in that gene and the location of the common disease mutations present in this study
Supplemental Figure 3. NLRC4 Plasmid Design (a) A pMigR1 backbone containing an inserted wildtype NLRC4, a V5 tag in order to locate the inserted plasmid once the construct is made, followed by the BirA\textsuperscript{MT} insert to identify the interactome (b) sequence alignment of WT NLRC4 and a NLRC4 with a point mutation of threonine to serine at amino acid 337 (T337S) (c) sequence alignment of WT NLRC4 and an NLRC4 with a point mutation of valine to alanine at amino acid 341 (V341A) (d) sequence alignment of WT NLRC4 and an NLRC4 with a point mutation of histidine to proline at amino acid 443 (H443P) (e) sequence alignment of WT NLRC4 and an NLRC4 with a point mutation of tryptophan to cysteine at amino acid 665 (W665C)
Supplemental Figure 4. NLRP3 Plasmid Design (a) A pMigR1 backbone containing an inserted wildtype NLRP3, a V5 tag in order to locate the inserted plasmid once the construct is made, followed by the BirA<sup>MT</sup> insert to identify the interactome (b) sequence alignment of WT NLRP3 and a NLRP3 with point mutation of aspartic acid to asparagine at amino acid 303 (D303N) (c) sequence alignment of WT NLRP3 and a NLRP3 with a point mutation of leucine to proline at amino acid 353 (L353P) (d) sequence alignment of WT NLRP3 and a NLRP3 with a point mutation of alanine to valine at amino acid 495 (A495V)
Supplemental Figure 5. NLRP3 HEK GFP Flow Cytometry GFP levels found in HEK 293T cells post transfection with NLRP3 plasmid. Cells were flowed using BD FACSaria II system.
Supplemental Figure 6. Unbiased Heat Map In order to get an unbiased view of our original heat map, the top 49 differentially expressed genes in our data set and had an FDR p-value of less than 0.99 were extracted and analyzed under the same parameters as the original heatmap. Only 14 were detected with a TPM ≥ 1.5.


