# ABSTRACT

Application of Genomic Technology
in the Study of Human Disease

Esperanza Anguiano, Ph.D.

Mentor: M. Virginia Pascual, M.D.

Genomic technologies are helping advance our understanding of human diseases at a very fast pace, especially in the fields of cancer, auto inflammatory diseases as well as immunodeficiencies. High throughput technology permits to survey the entire genome and assess individual genetic variation. This application has proven successful in identifying causal genes in Mendelian diseases and has provided insights into the genetic etiology of complex diseases.

In the past, we have applied blood transcriptional profiling to identify biomarkers and therapeutic targets such as IL-1B for patients with systemic onset juvenile idiopathic arthritis (sJIA), one of the major causes of chronic inflammatory arthritis in children. Indeed, clinical trials to evaluate the effect of blocking IL-1B in this disease have proven successful. This discovery supports that sJIA belongs to the recently described category of autoinflammatory diseases, which share similar clinical phenotypes

and respond to IL1 blockade. Most autoinflammatory diseases have been described as Mendelian diseases affecting different genes in the IL-1B pathway. In the work presented here, we investigated a potential genetic basis for sJIA, which is a sporadic disease. We applied high throughput next generation sequencing technology to survey the known coding regions of the human genome, the exome, in 6 sJIA trios (individual patients and parents) and in an additional group of 11 patients. In the analysis of trios we identified several potentially pathogenic variants in genes involved in biological functions associated with IL1; we also found mutations in genes shared by two or more sJIA patients. In addition, a rare variant was found in the S100A12 gene, which encodes a biomarker and potential pathogenic factor in the disease, in one family.

Overall, our study revealed potential genetic models of compound heterozygous and/or multigenic inheritance in this complex and heterogeneous disease.

Application of Genomics Technology in the Study of Human Disease

by

Esperanza Anguiano, B.S.

A Dissertation

Approved by the Institute of Biomedical Studies

_____

Robert R. Kane, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

_____

M. Virginia Pascual, M.D., Chairperson

_____

A. Karolina Palucka, M.D., Ph.D.

_____

Venuprasad Poojary, Ph.D.

_____

Derek Blankenship, Ph.D.

_____

Myeongwoo Lee, Ph.D.

Accepted by the Graduate School

August 2015

_____

J. Larry Lyon, Ph.D., Dean

*Page bearing signatures is kept on file in the Graduate School.*

TABLE OF CONTENTS

LIST OF FIGURES ..... *viii*

LIST OF TABLES ..... *x*

ACKNOWLEDGMENTS ..... *xii*

DEDICATION ..... *xiv*

Chapter One ..... 1

Introduction to Genomics Technology: Applications, Development and Implementation ..... 1

Background ..... 1

Overview of Genomics Technology ..... 3

Application of Genomics Technology to Understand Immune MediatedDisease in a Genetically Defective Background ..... 10

Development and Validation of Genomics Technology Applications ..... 14

Chapter Two ..... 323

Introduction to Systemic Onset Juvenile Idiopathic Arthritis (sJIA) ..... 32

Background ..... 32

Pathogenesis of sJIA ..... 34

sJIA as an Autoinflammatory Disorder ..... 44

Treatment Strategies for sJIA Patients ..... 45

Chapter Three ..... *52*

Materials and Methods ..... 52

Patient Samples ..... 52

Sample Preparation ..... 52

Exome Data Processing and Alignment ..... 55
Variant Calling and Annotation ..... 56
Selection of Candidate Variants ..... 57
Prioritization of Candidate Variants ..... 63
Validation of candidate variants ..... 65
Chapter Four ..... *67*
Results and Discussion ..... 67
Clinical Characteristics of sJIA Cohort ..... 67
Overall Filtering Strategy in Individual Trios ..... 68
Family 1 ..... 68
Family 2 ..... 72
Family 3 ..... 76
Family 4 ..... 80
Family 5 ..... 84
Family 6 ..... 88
Variants Affecting the Same Gene in Multiple sJIA Patients ..... 91
Chapter Five ..... 102
Functional Implications for S100A12 Deletion Identified in sJIA Proband 1 ..... 102
Background ..... 102
Introduction to S100 Proteins and Association in sJIA ..... 102
Characterization of Mutant S100A12 ..... 105
Chapter Six ..... 112
General Conclusions and Future Direction ..... 112
General Conclusions ..... 112

Future Direction ..... 115

REFERENCES ..... 117

LIST OF FIGURES

Figure 1.1 Transcriptional profile of fibroblasts treated with or without Poly(I:C) ..... 13

Figure 1.2 Transcriptional profile of PBMC treated with or without Poly(I:C) for 2- or 8-hours..... 14

Figure 1.3 Integrity and Purity of Purified DNA. ..... 26

Figure 1.4 Coverage plot for exome sequence data - Nimblegen..... 27

Figure 3.1 Generalized discrete filtering strategy employed in the selection of candidate variants..... 62

Figure 4.1 Family 1 pedigree. ..... 69

Figure 4.2 Family 2 pedigree. ..... 72

Figure 4.3 Family 3 pedigree. ..... 76

Figure 4.4 Family 4 pedigree. ..... 80

Figure 4.5 Family 5 Pedigree. ..... 84

Figure 4.6 Family 6 pedigree. ..... 88

Figure 4.7 Overlap of candidate genes with sJIA exome published data. ..... 96

Figure 4.8 Canonical pathways enriched using the identified sJIA candidate gene..... 97

Figure 4.9 Gene Network 1..... 99

Figure 4.10 Gene Network 2..... 100

Figure 4.11 Gene Network 3..... 101

Figure 5.1 S100 Protein Amino Acid Sequence Alignment..... 103

Figure 5.2 Family 1 S100A12 read sequence alignment in IGV Browser ..... 105

Figure 5.3 Normal and Mutant cDNA sequence for S100A12..... 106

Figure 5.4 Amino acid sequence for normal and mutant S100A12..... 106

Figure 5.5 Alignment of paired-end mRNAseq reads over chromosomal coordinate containing S100A12 deletion..... 108

Figure 5.6 Family 1 Pedigree..... 109

LIST OF TABLES

Table 1.1 NanoString Targeted Assay Accuracy Estimate ..... 17

Table 1.2 NanoString Targeted Assay Precision Estimate for Tester 1 ..... 18

Table 1.3 NanoString Targeted Assay Precision Estimate for Tester 2 ..... 18

Table 1.4 NanoString Targeted Assay Reproducibility Estimate ..... 19

Table 1.5 NanoString Targeted Assay Stability Estimate ..... 20

Table 1.6 NanoString Targeted Assay Estimate of Ruggedness ..... 20

Table 1.7 NanoString Targeted Assay Selectivity Estimate ..... 21

Table 1.8 NanoString Assay Performance Stability Over 28-Month Period ..... 22

Table 1.9 NanoString Assay Performance in Endpoint Measurement for Clinical Study 23

Table 1.10 Characterization of Discordant calls between ACE and ..... 29

Table 4.1 Clinical Characteristics of sJIA Cohort. ..... 67

Table 4.2 Summary of discrete filtering strategy in family trios ..... 68

Table 4.3 Discrete Filtering Strategy for Family 1 ..... 70

Table 4.4 Discrete Filtering Strategy for Family 2 ..... 73

Table 4.5 Discrete filtering strategy for Family 3 ..... 77

Table 4.6 Discrete filtering strategy for Family 4 ..... 81

Table 4.7 Discrete filtering strategy for Family 5 ..... 85

Table 4.8 Discrete filtering strategy for Family 6 ..... 89

Table 4.9 List of  Potentially Interesting Candidate Genes in sJIA ..... 92

Table 4.10 Genes affected in more than 1 proband ..... 92

Table 4.11 Genes with Missense or Loss-of-Function Mutations in More Than One sJIA Patient ..... 93

ACKNOWLEDGMENTS

First, I would like to thank Dr. Virginia Pascual, my mentor, for providing me with the opportunity to partake in the privilege of higher education. She has been fundamental in both my career and academic development. Equally important and fundamental in my development were Dr. Damian Chaussabel and Dr. Jacques Banchereau. They both provided me the opportunity to join the Translational Genomics Group at the Baylor Institute for Immunology Research 7.5 years ago. Dr. Chaussabel instilled in me an appreciation for genomics technology; he introduced me to the fascinating world of transcriptional profiling and provided me the opportunity to get involved in very important work with world renowned scientists. I also want to thank Dr. Karolina Palucka. She provided tremendous encouragement and support in my career development. Her support and encouragement in my involvement in special programs, such as the Human Immunology Project Consortium (HIPC), provided me with a stage where I could enhance my understanding of basic immunology. Special thanks also to Dr. Jeanine Baisch, a colleague and dear friend, who was very fundamental in my molecular biology training which started over 15 years ago; her continuous support and encouragement has made many of my goals attainable.

I want to also thank collaborators and mentors, Drs. Lynn Punnaro, Carol Wise and Jonathan Rios. Dr. Lynn Punaro and staff at TSRHC provided access to clinical samples as well as clinical insight regarding sJIA and patients in our study cohort. Drs. Carol Wise and Jonathan Rios mentored me and provided feedback in the analysis and interpretation of exome data.

I want to also acknowledge the people who were fundamental in my accomplishing my thesis work. First and foremost, the entire BIIR's Genomics Core team. In particular, Benjamin Lemoine; he was fundamental in helping me with the generation of exome data and lightened my workload by helping me with the oversight of the genomics staff and operation. The Pascual Lab members for taking care of all the clinical details including patient recruitment, sample collection, and for teaching me and helping me with in vitro experiments.

I want to thank the Poojary lab for assisting with preliminary analysis of the S100A12 protein. A special thanks to Dr. TaeWhan Kim for teaching me how to run Western blots and for the many useful discussions. Also fundamental were Drs. Gerard and Sandy Zurawski for their support and cooperation in the expression of S100A12 mutant protein. I want to thank Drs. Nicole Baldwin's and Derek Blankenship's and respective teams for their support in data processing and data analysis. A special thanks to Dr. Brandi Cantarel who was fundamental in exome data processing and analysis.

Last but not least, I am very thankful to our patients; the administrative staff at the Baylor University and Baylor Research Institute; and the Baylor, Scott & White Healthcare System for generous financial support of some of this work.

# DEDICATION

In memory of my parents, who always supported any path I took

To my brothers, sisters and friends, who have encouraged me every step of the way

To my daughter Victoria, *the wind beneath my wings...*

# CHAPTER ONE

## Introduction to Genomics Technology: Applications, Development and Implementation

### *Background*

It can be said that the inception of genome technology took place in the 1950s. In 1953 James Watson and Francis Crick described the DNA structure consisting of a double helix of purine and pyrimidine pairings that suggested a possible mechanism for replication of the genetic material. This work earned them the Nobel Prize in 1962. In 1957, Arthur Kornberg discovered the first DNA polymerase and elucidated the mechanism of DNA replication. At the same time, Severo Ochoa elucidated the mechanism of RNA synthesis. They were both awarded the Nobel Prize in 1959.

Har Gobind Khorana made significant contributions to the description of the genetic code and was awarded a Nobel Prize in 1968. In this work, Khorana pioneered basic techniques required to make and use synthetic DNA oligonucleotides that ultimately led to synthesis of a functional human gene in 1976. This work laid the foundation for the prospect and methods of the synthesis and amplification of DNA. Indeed, in 1977 Sanger et al. published the first DNA sequencing method using chain-terminator chemistry, now commonly known as Sanger sequencing (Sanger et al. 1992). Subsequently, in 1984 Kary Mullins generated proof of concept data for targeted amplification of DNA while working for Cetus Corporation using the polymerase chain reaction (PCR) for which a patent was issued in 1987 and the method finally published in 1990. PCR and sequencing technology revolutionized the field of molecular genetics. Together, these led to the conception of the human genome project in the early 1990s

and, with a myriad of improvements and automation of methodologies (Meldrum 2000a; Meldrum 2000b); the human genome sequence was finally completed and published by Venter et al. in 2001 (Venter et al. 2001).

The vast amount of genomic information generated by the human genome project led to rapid development of high throughput genome technology, such as DNA microarrays first introduced in 1997 (Shalon et al. 1996; Pioch et al. 2008). This technology allowed interrogation of hundreds of thousands of single nucleotide polymorphisms (SNPs) identified in the human reference or other catalogs. These assays can assess single point mutations as well as structural variations such as insertions, deletions, transversions and copy number variation. Most genome wide association studies (GWAS) were carried out in the field of cancer, neurological, cardiovascular as well as other complex diseases. These led to successful identification of causal variants and association markers. Similarly, this technology has facilitated analysis of massively parallel high throughput gene expression analysis (Shimkets 2004). Application of this technology in transcriptional profiling studies has significantly improved our understanding of molecular mechanisms driving several disease states such as cancer, cardiovascular, neurological, and other complex diseases like lupus and autoinflammatory disorders (Shimkets 2004).

Today, various high throughput sequencing technology applications are employed in research settings for the purpose of discovering diagnostic, prognostic, treatment response as well as patient stratification biomarkers (Shendure et al. 2011). The high reliability and sensitivity of sequencing technology has facilitated application of the same

goals in the clinical setting. Further, significantly improved methodologies have made single cell genomics commonplace in today's research setting.

The aims of this body of work are to demonstrate (1) how high throughput gene expression technology can be applied to discover underlying molecular mechanisms in the background of immune mediated diseases, (2) how technology development can facilitate and advance scientific investigation, and (3) how we applied newly implemented technology to investigate the potential genetic basis of sJIA.

## *Overview of Genomics Technology*

### *Microarray Technology*

DNA microarrays make use of probes complementary to different DNA or RNA sequences specific to an organism to detect and measure the relative abundance of such targets. Microarrays have been widely applied in the study of genetic variation, gene expression profiling, and assessment of protein-DNA interactions as well as methylation studies (Bumgarner 2013). More recently microarrays have been used in the enrichment process of targeted sequencing (Ng et al. 2009). The technology allows for simultaneous detection of tens or hundreds of thousands of targets. The most commonly used arrays consist of spotted arrays, in-situ synthesized arrays and self-assembled arrays. Spotted arrays make use of poly-lysine coated glass microscope slides where a needle is used to spot DNA probes (DeRisi & Iyer 1999; DeRisi et al. 1996). One of the in-situ synthesized arrays utilizes light directed chemical synthesis combining photolabile protecting groups with photolithography to perform oligo synthesis on a solid substrate (Fodor et al. 1991); this technology is used in Affymetrix arrays. Another method similar

to this technology made use of inkjet printing technology for delivery of the four different nucleotides required for the oligo synthesis to a glass slide pre-patterned with hydrophilic and hydrophobic regions (Blanchard et al. 1996). This is utilized by Agilent Technologies in the commercialization of custom arrays. In both the Affymetrix and Agilent chips, the delivery of reagents follows the order of sequence information loaded to an automated system and the arrangement is spatially controlled. The self-assembled array synthesizes DNA on microscopic polystyrene beads. Hundreds of thousands of different beads are then combined and ultimately placed at the end of a fiber optic array where the ends of the fibers are pre-etched providing a well for placement of the bead (Ferguson et al. 2000; Michael et al. 1998). The array arrangement is random. This technology is commonly known as BeadChip array technology and is commercialized by Illumina. The decoding of the randomly arranged beads is done by hybridization and detection of a number of short, fluorescently labeled oligos (Gunderson et al. 2004) which facilitates an enormous number of different types of beads to be multiplexed on a single array and also allow for functional testing of the end product.

The manufacturing of arrays onto a solid matrix allows for the use of fluorescently labeled samples and as such, fluorescence intensity is the measure of target detection. The sample preparation process is specific to the different applications and technologies. An example of how I applied gene expression microarray technology to answer a scientific question is presented later in this Chapter.

*NanoString Technology*

The NanoString nCounter System employs capture and reporter oligonucleotide probes specific for a given gene transcript sequence. The capture probe is tagged with

biotin to facilitate its binding to a streptavidin-coated slide while the reporter probe is tagged with a combination of four fluorophores; this combination serves as molecular barcode for its target gene transcript. The design of capture and reporter probes includes a 3' and 5' repeat sequence, respectively, that facilitates affinity purification. Subsequently, purified complexes are prepared for binding, electrophoresis and immobilization onto solid matrix in preparation for signal detection. Binding onto the surface takes place via biotin tagged capture probe, voltage is applied for the purpose of elongation and uniformed alignment of molecules. Immobilization takes place via biotin and addition of anti-5' oligonucleotides. Finally, immobilized reporters are imaged using a CCD camera; image processing and code counting take place using specialized software. The code count is representative of targeted gene expression levels.

The combination and sequence arrangement of fluorophores allows for a very high number of uniquely labeled capture probes that can be used in a multiplexed approach to analyze expression levels of multiple gene transcripts (up to 800) in a single sample, all in a single reaction. The system is free of enzymatic reactions and its sensitivity is similar to that of real-time PCR (Geiss et al. 2008).

*Next Generation Sequencing Technology*

Library preparation, signal detection of incorporated nucleotides, and data analysis are components of several methods grouped broadly under the heading of sequencing technologies. A combination of unique methods employed in this process makes each technology different from each other. Next generation sequencing technology (NGS) was commercially introduced in 2005. Tremendous advancements have been made in terms of throughput, speed, read length, sample input requirements,

sequencing applications, and cost since its inception (Metzker 2010; Fox et al. 2009; Pabinger et al. 2014). These applications allow for whole genome sequencing, targeted sequencing of genomic regions of interest such as protein coding regions, the exome, small RNA sequencing, sequencing of methylated regions of the genome to determine transcriptional control mechanisms, and sequencing of genomic regions associated with proteins such as histone and/or transcription factor binding site (ATACSeq & ChipSeq, respectively). Data processing is a large and critical component in the application of NGS technology, but is not the focus of this work.

*Template preparation*. Most NGS methods employ fractionation of template sample from which library templates are subsequently prepared. This includes using platform specific adapter sequences as well as unique template barcode sequences that facilitate multiplexing of samples for sequencing (Clark et al. 2011; Liu et al. 2012). These constructs are subsequently spatially separated on a solid surface where they are clonally amplified. The most common method utilizes solid-phase amplification to produce randomly distributed clonally amplified clusters on a flow cell (Metzker 2010; Shendure & Ji 2008). This process occurs in two steps; (1) priming and extending of the single-stranded, single molecule template, and (2) bridge amplification of immobilized template with immediately adjacent primers to form clusters. Illumina sequencing technology uses this approach in their template preparation protocol. Another solid-phase template preparation approach makes use of emulsion PCR. In this reaction, a mixture consisting of an oil-aqueous emulsion is created to form droplets containing bead-DNA complexes. PCR takes place within these droplets to create beads containing thousands of copies of the same template sequence. Roche 454 , SOLiD and Ion Torrent

sequencing technologies employ this approach (Morozova & Marra 2008; Liu et al. 2012; Quail et al. 2012). Single-molecule template preparation methods do not require PCR. Single-molecule templates are immobilized on solid supports using either spatially distributed individual primer molecules that are covalently attached to the solid support, or spatially distributed single-molecule templates covalently attached to the solid support by priming and extending single-stranded, single-molecule templates from immobilized primers – both of these are employed by Helicos BioSciences. A third approach takes advantage of spatially distributed single polymerase molecules attached to a solid support to which a primed template molecule is bound – this approach is employed by Pacific Biosciences (Carneiro et al. 2012). The latter approach allows for real-time reads and longer read lengths (Chaisson et al. 2014).

*Sequencing and signal detection*. The key feature in Illumina sequencing is sequencing by synthesis (SBS) employing reversible terminator chemistry (Bentley et al. 2008; Metzker 2010). Briefly, a universal sequencing primer is hybridized to clonally amplified libraries to initiate the sequencing reaction then four-color reversible terminator-bound dNTPs are added and allowed to incorporate into respective templates; fluorescence is imaged and the bound terminator is then cleaved. Subsequent cycles of this process occur simultaneously for millions of clusters generating millions of short sequence reads (sequencing by synthesis) representative of the library fragments prepared from the original DNA sample (Metzker 2010). Image analysis and base calling takes place in a HiSeq using instrument control software (ICS). This software uses raw images to identify cluster position and intensity; this information along with noise estimates are used to determine the sequence of bases read from each cluster.

*RNASeq Technology*

RNAseq is an NGS sequencing application that allows sequencing of RNA molecules via random cDNA libraries with the aim to identify the transcriptional state of a given sample. Briefly, cDNA libraries are prepared from (1) sheared double-stranded cDNA, (2) adaptor-tagged random hexamer or (3) serial ligation of adaptors to fragmented RNA populations. After massively parallel sequencing, short reads are aligned to an appropriate reference genome to provide a global assessment of transcriptome activity (Cloonan & Grimmond 2008).

The advantages of RNAseq over microarrays are its quantitative output and high degree of reproducibility compared to microarrays (Cloonan & Grimmond 2008). In addition, the dynamic range of RNAseq is relatively unlimited as read counts are used to determine transcript abundance, while microarrays use the relative fluorescence intensity of an image to determine relative abundance. Further, the detection limitation of transcripts in RNAseq is limited to those present in each sample, whereas in microarrays this is limited to manufacturer chosen assay content. Moreover, cross-hybridization events are relatively high in microarray and this is resolved in RNAseq, as sequence identity is used for mapping to the correct location in the transcriptome.

The use of RNAseq has enabled researchers to conduct genome-wide surveys of the transcriptome and has resulted in detection of novel transcriptional complexity. For example, long non coding RNAs have been identified and associated to transcriptional regulatory mechanisms. Specific isoform expression and gene fusion events can be discerned from RNAseq data. In addition, it is also possible to assess post transcriptional modifications such as RNA editing as well as genomic variation (Han et al. 2014).

There are several challenges that remain to be addressed in RNAseq. One limitation in RNAseq is that reads are mapped to a known reference sequence. Introduction of sequencing platforms that allow for very long reads are useful in meeting this challenge as longer reads are more amenable to de novo assemblies. In addition, our ability to interpret transcriptome complexity is dependent on previous annotation builds. Alignment algorithms are also not optimal – there is poor consensus on aligned reads generated by two different methods (Martin & Wang 2011).

Nevertheless, RNAseq is already considered the gold standard for measuring gene expression. As part of this work, RNAseq technology was implemented to assess its utility in a variety of backgrounds; from single cell mRNAseq profiling to profiling different tissue samples, including whole blood.

*Whole Exome Sequencing (WES)*

The proportion of the human genome consisting of protein coding regions is approximately 1% spanning 180,000 exons – this constitutes a size of ~30Mb and is commonly called the "exome". Introduction of next generation sequencing technology capable of generating up to 300Gb of reads in a single run makes this, as well as whole genome sequencing, feasible in translational research and clinical settings (Priya et al. 2012; Ng et al. 2009; Sulonen et al. 2011). While technically feasible, the cost of whole genome sequencing is not practical for most budgets; as such, most genome studies have relied in sequencing the coding regions of the genome, the exome, with the rational that aberrations in these regions are likely disease causing.

The most common approach for enrichment of the "exome" is a hybridization capture approach which utilizes oligonucleotide sequences that are complementary to

known exons in the human genome. The application of whole exome sequencing has facilitated discovery of variants associated with Mendelian diseases, and today there are several commercially available kits that differ in probe design and regions targeted (Bamshad et al. 2011; Ng et al. 2010; Ng et al. 2009). As part of this thesis work, two exome capture approaches were evaluated and results are also presented and discussed later in this Chapter.

### *Application of Genomics Technology to Understand Immune Mediated Disease in a Genetically Defective Background*

The content of this section was included in the following publication:

Y Guo, M Audry, M Ciancanelli, L Alsina, J Azavedo, M Herman, **E Anguiano**, V Sancho-Shimizu, L Lorenzo, E Pauwels, P Bastard Philippe, R Perez de Diego, A Cardon, G Vogt, C Picaard, Z Zo Andrianirina, F Rozenberg, P Lobon, S Plancoulaine, M Tardieu, V Doireau, E Jouanguy, D Chaussabel, F Geissmann, L Abel, J L Casanova, and S Y Zhang. 2011. Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. JEM 208:2083-2098

### *Microarray Gene Expression Profiling in the Study of Immune Perturbations in a Patient with TLR3 Deficiency*

*Background.* Autosomal dominant TLR3 deficiency has been identified as a genetic risk factor in childhood herpes simplex virus 1 (HSV-1) encephalitis (HSE). In this study, our collaborators identified a compound heterozygous loss of function mutation in TLR3 in a patient with HSE. TLR3 is a non-specific receptor of double stranded (ds) RNA intermediates generated during the process of viral replication, including HSV-1. Commercially available Poly(I:C) mimics dsRNA. The group had already identified that TLR3 function was impaired in the patient's fibroblasts, as INF-alpha, - beta, and –gamma were not induced in response to Poly(I:C); the function was

restored upon expression of normal TLR3. They were interested in identifying TLR-3 dependent genes targeted during stimulation with Poly(I:C); as such, they engaged the BIIR team for transcriptional profiling work. I was responsible for generating gene expression microarray data and providing data mining support, including production of figures and figure legends for publication.

*Experimental design.* Cell lysates prepared from cultures of PBMC and fibroblasts from the patient (TLR3-/-), UNC-93B-/-, AD TLR3, MyD88-/-, and healthy controls treated with or without Poly(I:C) for 2h or 8h were submitted for microarray analysis. This design included autosomal recessive TLR3 deficiency (patient), autosomal dominant (AD TLR3), autosomal recessive UNC-93B deficiency, and autosomal recessive MyD88 deficiency. All three defects are known to impair TLR3-dependent induction of IFN-alpha, -beta, and –gamma in dermal fibroblast in response to stimulation with Poly(I:C). Given the limitation in sample size, statistical analysis was restricted to calculating fold change and magnitude of change respective to the healthy controls. In order to quantify the global transcriptional response, cumulative fold change for each sample was calculated and represented as magnitude of response in the form of a bar graph generated using (GraphPad Prism Software) and placed over expression heatmaps generated using GeneSpring v7.3 (Agilent) gene expression analysis software. The top most differentially expressed transcripts were mined for functional associations using Ingenuity Pathway Analysis software (Ingenuity).

*Results and discussion.* Using microarray technology and our analysis approach we were able to elucidate an impaired response to Poly(I:C) in the TLR3-/- patient. Thus,

the patient's blood transcriptional profile was strikingly similar to the profile of a patient with UNC-93B deficiency, an autosomal recessive disease that alters the recruitment of endosomal TLRs from the ER (Figure 1.1). Cells from a MyD88-/- patients and an autosomal dominant TLR3 deficient patient show normal delayed response (8h compared to 2h) compared to healthy controls. This transcriptional response is driven by interferon inducible genes with highest activation of STAT1, IRF7, and IRF1 in the healthy controls, autosomal dominant and non-TLR3 defective controls; whereas it is completely ablated in the TLR3-/- and UNC-93B-/- patient. In contrast to fibroblasts, a response similar to that of healthy controls is observed in PBMC of TLR3-/- patient treated with Poly(I:C) (Figure 1.2). Both sample types have an interferon inducible gene signature driven by IRF7 and STAT1.

*Remarks.* Genome wide transcriptional profiling methods and microarray technology are robust at assessing transcriptional changes that can unravel molecular mechanisms underlying genetic defects and their consequences on immune response mechanisms as well as tissue specific responses. By applying this technology we were able to elucidate that fibroblasts respond to Poly(I:C) in a TLR3-dependent manner and that PBMC respond to Poly(I:C) in a TLR3-independent manner. In this study, transcriptional changes were very robust. Inclusion of appropriate control samples with deficiencies in related pathways (as well as healthy) was paramount for characterizing normal and abnormal responses that have important correlates in vivo as defense mechanisms against infections.

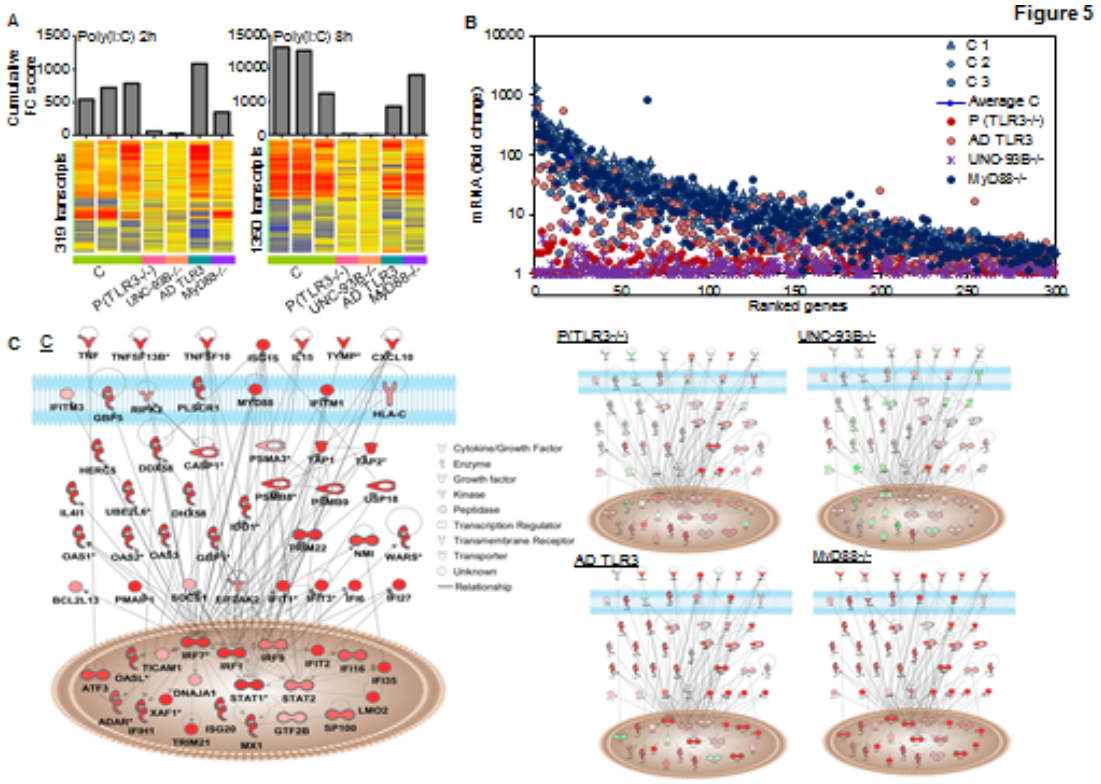


Figure 1.1 Transcriptional profile of fibroblasts treated with or without Poly(I:C) (A) Heatmap of 319 transcripts with 1.5 fold-change difference; red indicates up regulation, blue indicates down regulation and yellow indicates no change. Bar graph represents magnitude of FC responses. (B) Scatter plot of ranks for 319 genes and fold change expression for all samples analyzed. (C) Gene interaction and predicted cellular localization for healthy control samples (left panel) and patient samples (right panel). The figure is the same in all panels; however, the gene expression as indicated by red (up regulation) and green (down regulation) are specific for each individual patient.

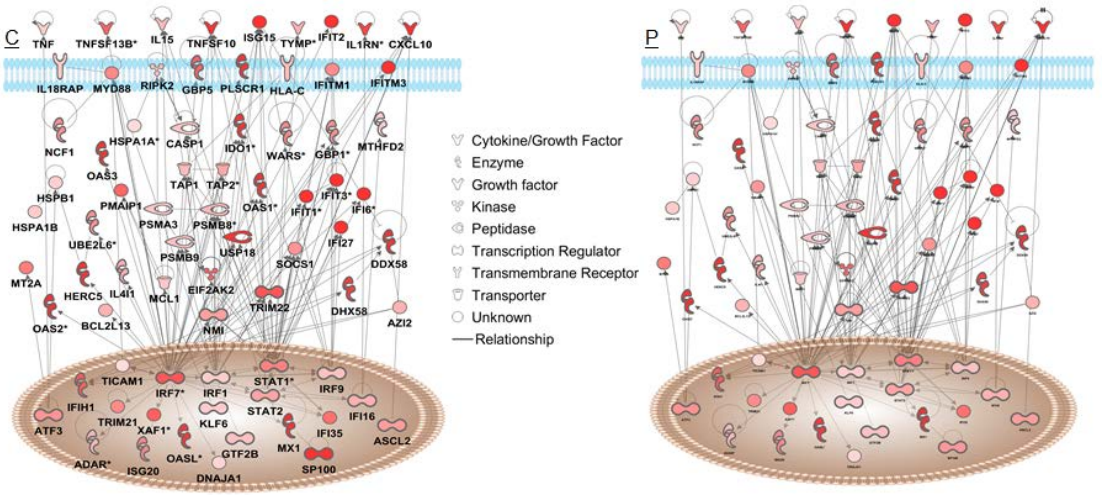


Figure 1.2 Transcriptional profile of PBMC treated with or without Poly(I:C) for 2- or 8-hours Gene interaction and predicted cellular localization for healthy control samples (left panel) and patient samples (right panel). The figure is the same in all panels; however, the gene expression as indicated by red (up regulation) and green (down regulation) are specific for each individual patient.

*Development and Validation of Genomics Technology Applications*

In the previous section we demonstrated how high throughput gene expression profiling was useful in unraveling molecular mechanisms underlying immune mediated disease under different genetic backgrounds. In addition, we demonstrated how the application of sequencing technology can elucidate new information and how the nucleotide sequence level information provided clarity where the technical limitations of microarrays provided us with noise. Technology development is paramount to scientific discovery. The work presented in the next section highlights two technology development projects. One of the projects makes use of the transcriptional clues provided from high throughput gene expression data. The second project takes on the assessment of new methodology to be implemented in the whole exome sequencing project, which is the focus of the next sections of this dissertation.

*Development and Validation of a NanoString Targeted Gene Expression Assay to Measure Interferon-alpha Inducible Gene Expression*

*Background.* Transcriptional gene modules have been described by our group and widely applied in the analysis of high throughput gene expression data (Chaussabel et al. 2008; Chaussabel & Baldwin 2014). Specifically, interferon signatures and modules have been associated with disease activity in Systemic Lupus Erythematosus (SLE) (Bennett et al. 2003; Chiche et al. 2014; Guiducci et al. 2010). A module consists of genes coordinately expressed across different datasets and is associated to a biological function. Our group has identified three sets or modules of Interferon-inducible transcripts (M1.2, M3.4 and M5.12). A drug developed by a private biotechnology company targeting the interferon pathway in SLE patients was to be tested for safety and efficacy in a clinical study. To this end, I developed and validated a targeted gene expression assay using the content of relevant interferon modules M1.2 and M3.4 using NanoString technology (Nanostring Technologies). This assay was utilized as clinical and pharmacokinetic (PK) endpoint for a compound used in the treatment of SLE patients. Only results from the assay development and validation as well as the assay performance over the course of the patient enrollment period and clinical study are presented as part of this body of work.

*Experimental design.* Selection of genes to test for each of the two interferon modules was based on their Spearman correlation values. These values were available from the module generation scores (Obermoser et al. 2013) and the top 15 correlated gene transcripts were selected. In addition, a total of 10 housekeeping genes were chosen in order to control for sample input variation. Criteria for selection of housekeeping

genes included high, medium, and low expression gene transcripts that displayed minimal variation across conditions selected from the BIIR gene expression database. Gene content of this panel is available upon request. A set of positive and negative assay controls was also generated. Positive assay control consisted of a pool of 24 SLE pediatric patients (SLE Pool) for whom gene expression data were available and an interferon signature was verified. Negative assay control consisted of a pool of 24 healthy subjects (Healthy Pool) also with known low interferon scores. Validation included four experiments making the following assessments: (1) accuracy, precision, and stability, (2) accuracy and precision, (3) ruggedness, and (4) selectivity. The experimental design for the validation was reviewed and approved by quality assurance experts employed by our collaborator.

Data analysis was carried out as follows: data were normalized by calculating a normalization factor using counts for internal spiked-in controls to adjust for variation in hybridization efficiency. First, the sum of each positive control (ERCC) was taken for each sample and set to X, the mean of positive controls (ERCC) for the Healthy Pool controls only was calculated and set to Y; and then the absolute difference of Y-X was set to Z. A normalization factor for each sample was calculated by dividing the smallest Z value by each sample's corresponding Z value; and then all count values for each sample were adjusted using their corresponding normalization factor. Background was subtracted following NanoString's recommended protocol. Further normalization to housekeeping (HK) genes was carried out to adjust for technical variation in RNA input. First, the mean of counts for each HK gene across Healthy Pool controls was calculated for each HK gene, and then the geometric mean of these values was calculated and set to

A. Subsequently, the geometric mean for HK genes for each sample was also calculated and set to B. A normalization factor consisting of A/B for each sample was calculated and used to adjust each sample's gene counts. Finally, the geometric mean of the normalized counts was taken for all module genes in each sample and then log2 transformed. This value was reported as the sample's/individual's interferon inducible score.

*Assessment of accuracy.* Accuracy of this assay was determined against a nominal value for SLE and healthy pool interferon scores, 8.1836 and 10.8246. These nominal values were obtained from historical performance of the assay for each control pool sample. Bias from respective nominal interferon score values was calculated using the following formula: absolute value of (observed value – nominal value)/nominal value X 100. Five technical replicates of each of the controls were tested to assess accuracy and precision and summary of results is represented in Table 1.1. It was determined that the accuracy of the results was well a within acceptable minimum criterion which was set of 15%.

Table 1.1 NanoString Targeted Assay Accuracy Estimate

| | SLE Pool | Healthy Pool |
|---|---|---|
| Average | 10.22 | 7.61 |
| Bias (%) | 5.57 | 7.05 |

*Assessment of precision.* Precision was measured as intra-assay precision (repeatability) and intermediate precision (reproducibility). Precision testing is designed to measure the degree of scatter between series of measurements obtained from multiple samplings from the same homogeneous sample. In our repeatability assessment SLE and Healthy pool samples were tested multiple times (N=5) by two independent testers, results are summarized in Tables 1.2 and 1.3. Mean standard deviation and coefficient of variation (CV) for each sample from each trial were calculated.

Table 1.2 NanoString Targeted Assay Precision Estimate for Tester 1

| Tester 1 | SLE Pool | Healthy Pool |
|---|---|---|
| Average | 10.16 | 7.55 |
| Std. Dev. | 0.02 | 0.03 |
| %CV | 0.17 | 0.36 |

Table 1.3 NanoString Targeted Assay Precision Estimate for Tester 2

| Tester 2 | SLE Pool | Healthy Pool |
|---|---|---|
| Average | 10.28 | 7.66 |
| Std. Dev. | 0.02 | 0 |
| %CV | 0.18 | 0.05 |

Reproducibility measures variability in measuring the same sample in multiple, independent trials and assesses variation within a single laboratory setting. Here, variability was measured between Tester 1 and Tester 2 runs on different days by

different lab members. The mean, standard deviation and CV was calculated from the 10 replicate samples tested by the two tester and results are summarized in Table 1.4.

Table 1.4 NanoString Targeted Assay Reproducibility Estimate

| | SLE Pool | Healthy Pool |
|---|---|---|
| Average | 10.22 | 7.61 |
| Std. Dev. | 0.06 | 0.06 |
| %CV | 0.61 | 0.78 |

*Assessment of stability.* This analytical method is dependent on the stability of the analyte, in our case the purified RNA. Here, stability of the analyte was tested to measure the extent to which handling of samples could compromise or alter the analyte being quantified. For the stability assay, we identified sample/freeze thaw cycles as a potential factor that can impact test results. Freezing and thawing cycles are known to induce sample instability. Therefore, stability of the assay was assessed using an SLE RNA pool that had undergone one-freeze/thaw cycle and a total of three-freeze/thaw cycles. Temperature variation went from -80C to 4C. Interferon scores were generated and compared to their expected return value through calculation of bias defined as the absolute value of (value of freeze/thaw sample – norminal value)/nominal value X 100.

Table 1.5 NanoString Targeted Assay Stability Estimate

| Number of Freeze/Thaws | 1 | 2 |
|---|---|---|
| Interferon Score | 10.2 | 10.19 |
| Bias from Nominal (%) | 6.13 | 5.81 |

*Assessment of ruggedness.* The degree of reproducibility of test results obtained by the analysis of the same sample under a variety of normal test conditions is referred to as ruggedness. Here, we tested ruggedness by comparing the performance of the assay on the same blood sample from which RNA was purified using two different RNA extraction kit lot numbers. Bias between the observed interferon score values was calculated as (interferon value from Kit1 – interferon value from Kit2)/average of interferon value from Kit1 and Kit2 X 100 and results are summarized in Table 1.6.

Table 1.6 NanoString Targeted Assay Estimate of Ruggedness

| | Sample 1 | Sample 2 |
|---|---|---|
| Kit 1 | 9.292 | 9.999 |
| Kit 2 | 9.345 | 10.012 |
| Bias (%) | 0.57 | 0.13 |

*Selectivity.* Design of this experiment aimed to assess the ability of the assay to differentiate the analyte of interest from endogenous components in the matrix. Here, a total of eleven individual sources of RNA derived from healthy volunteers were mixed in

equal part to RNA from the SLE pool. Interferon scores from each was generated and compared to assess any differences in the IFN score. As in the previous analyses, standard deviation and CV were calculated. Additionally, the Grubbs' outlier test with 0.05 significance level, two-tailed T-test was performed using GraphPad Prism web calculator (graphpad.com/quickcalcs). Results of these analyses are summarized in Table 1.7.

Table 1.7 NanoString Targeted Assay Selectivity Estimate

| Sample | Unspiked | Spiked |
|---|---|---|
| Mean | 7.742 | 9.707 |
| StdDev | 0.25 | 0.192 |
| %CV | 3.23 | 1.98 |
| Outlier | No | No |

Assay performance throughout patient stratification phase. This assay was used to test over 170 subjects for enrollment into the clinical study over a period of 28 months. The Healthy and SLE pools were used to assess the assay and lab performance over time as well as during the actual testing of the clinical study samples. A total of 65 assay runs were carried out during the 24-month period. Each time the assay was run, a minimum of duplicate healthy and SLE pool samples were included as controls. Assessment of variation as a measure of CV was performed and results are summarized in Table 1.8. Assay performance as measured by the standard deviation and coefficient of variation for interferon scores generated from the Healthy and SLE pool controls was very stable throughout the course of the study.

Table 1.8 NanoString Assay Performance Stability Over 28-Month Period

| | SLE Pool | Healthy Pool |
|---|---|---|
| Average | 10.778 | 8.136 |
| Std. Dev. | 0.622 | 0.556 |
| %CV | 5.8 | 6.8 |

*Performance of technical controls in clinical study.* A total of 437 total RNA samples purified from whole blood were tested as part of the clinical study assessing safety and efficacy of treatment. In an effort to control for technical variation inherent in gene expression profiling, samples were randomized into 16 batches reaching maximum throughput in sample processing over a period of 4 weeks. Each batch included a minimum of two SLE pool controls, three Healthy pool controls, and a non-template control. Following the established standard operating procedure, samples were hybridized and scanned using the NanoString nCounter System and data were analyzed as previously described. Technical noise was assessed as the standard deviation and coefficient of variation across all technical replicate controls tested with all study samples, Table 1.9.

These results demonstrate that this methodology is very stable and robust. Historical data for the entire study shows less than 7% variation in intermediate precision of the assay. Technical variation for samples in the clinical study is negligible, less than 1% CV, suggesting that the endpoint result would be highly reliable.

Table 1.9 NanoString Assay Performance in Endpoint Measurement for Clinical Study

| | Control | Mean IFN Score | Std. Dev. | %CV |
|---|---|---|---|---|
| Historical Data (n=77) | Hpool | 7.999 | 0.585 | 7.313 |
| | SLEpool | 10.653 | 0.66 | 6.198 |
| ClinicalStudy (n=45) | Hpool | 8.519 | 0.022 | 0.257 |
| | SLEpool | 11.429 | 0.022 | 0.194 |

*Remarks.* Although mandatory in clinical and translational research settings, assay validations are not commonplace in an academic research environment. This is peculiar, as published results are generally trusted. Validation of the method described in our study revealed that it is accurate, of high precision and very robust. Further, control samples produced for the purpose of validation and as assay controls proved very useful in assessing performance throughout both the validation process and testing of clinical samples. Validated methods and processes as well as inclusion of appropriate assay controls present the user with results that are reliable and reproducible. It seems appropriate that, at the very least, high throughput assays used in research studies should be assessed for reproducibility, precision and specificity.

*Development and Implementation of a Whole Exome Sequencing Pipeline for Implementation in Genetic Studies*

*Background.* Historically at BIIR, investigators have focused on transcriptional profiling and sample collection methods for research studies. However, emphasis and

interest in the broad scientific community is shifting to genetic analysis, which could be directly applicable to some of the most studied diseases at BIIR, but requires development and implementation of new technology. In order to establish a standard methodology and analysis pipeline, we designed a proof-of-concept study to identify genetic variants in a small experimental sample set. Specifically, I developed a method for purifying DNA and total RNA such that samples that had been previously collected with only RNA studies in mind could be repurposed for genetic studies or genomic analysis. As such, precious patient samples that have been collected longitudinally over a disease course or treatment regimen can now be integrated into investigators' study design as technology advances. Moreover, this development process included comparison of a commercially available exome capture kit, Nimblegen (Roche), with a non-commercially available optimized capture system, ACE Enhanced Exom (Personalis).

*Experimental design.* Whole blood was collected in Tempus RNA tubes from consented sJIA patients and parents. DNA was purified using my modified protocol described in Chapter Three. DNA purity and concentration were assessed using a NanoDrop 1000 and integrity was assessed using LabChip GX and Genomic DNA Chip (Caliper). Suitability and stability of both DNA purified from whole blood collected in Tempus tubes for WES and our WES workflow were assessed by comparing variant calls extracted from our WES data against SNP array data generated from the same DNA samples. SNP data was generated at the UTSW Genomics Core using the Illumina 500K OmniSNP Array. Purity and quality of purified DNA was compared against a commercially available DNA control obtained from the Coriell Institute. In addition, a

total of eighteen samples corresponding to family trios were sequenced using an enhanced exome capture (ACE exome capture) method available through sequencing services provided by Personalis. Design of the capture oligonucleotide targeting genomic regions of interest and process of capture can significantly impact our ability to detect relevant genetic variation. Data obtained was used not only for assessment of exome capture methods but also for validation of variants identified in sJIA exome sequencing project presented in the next several chapters.

*Assesment of DNA quality.* Integrity of DNA purified from whole blood collected in Tempus RNA tubes was assessed using Genomic DNA LabChip Kit and Caliper GX. Figure 1.6A is a representative image of the distribution of fragment sizes for DNA recovered; one can appreciate the distribution is more intense for fragments >10kb as marked by the DNA ladder. This suggests that the integrity of the DNA is of high quality and should be suitable for genotyping applications. Purity and concentration as assessed by absorbance measurements at 260 and 280 (Figure 1.3B and 1.3C), also demonstrates the purification process is efficient.

*Library quality and quantity.* Quality of the DNA libraries generated from the purified DNA was assessed by measuring the yield and size fragment distribution of material recovered using the Agilent DNA Chip 1000 and Bioanalyzer 2100. Observed concentration values and electrophoretic profiles met the expected values (data not shown). In addition, observed pre-captured average library fragment was 400bp and the library yield ranged between 10nM-30nM. Average pooled post-capture library fragment

size was similar to the pre-capture libraries and library concentrations were on average 20nM.

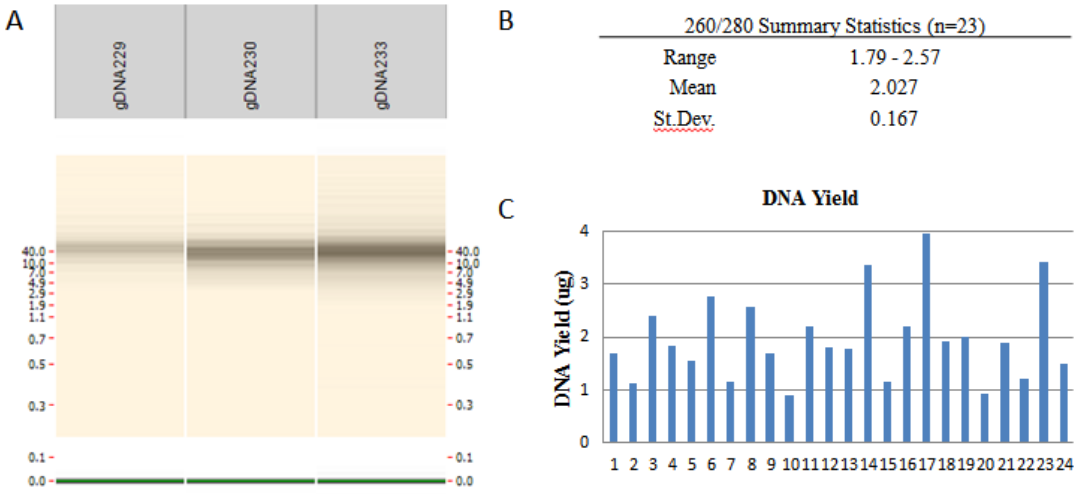


Figure 1.3 Integrity and Purity of Purified DNA. Quality assurance assessment of DNA purified from whole blood collected in RNA Tempus tubes. Representation of fragment size distribution depicting integrity (A), the summary statistics of the 260/280 ratio values observed (B) and the distribution of the amount of DNA recovered from the equivalent of 200ul of blood (C).

*Coverage and depth.* In this study and for comparison purposes we refer to coverage as the percent of the genome (targeted exome) covered. The depth here is referred to as the number of times a specific region was sequenced. The genome size and the number of sequencing reads are two important factors to consider in a re-sequencing experiment as not having sufficient coverage and depth can significantly impact results (Figure 1.4).

Using Nimblegen exome capture we achieved coverage of 80% of the targeted regions at a depth of 20X for most samples. There were a total of 6 samples where we only reached 50% coverage at a depth of 20X. The main difference in these samples was that a lower input amount (50ng) of DNA was used for library prep compared to for the

other samples (1ug). The average coverage using the ACE exome capture resulted in similar coverage, 80% of the targeted regions were covered at a depth of 20X (data not shown). The average coverage was 45% at a depth of 50, which is lower than that observed for Nimblegen's with average coverage depth of 50% at a depth of ~60X.

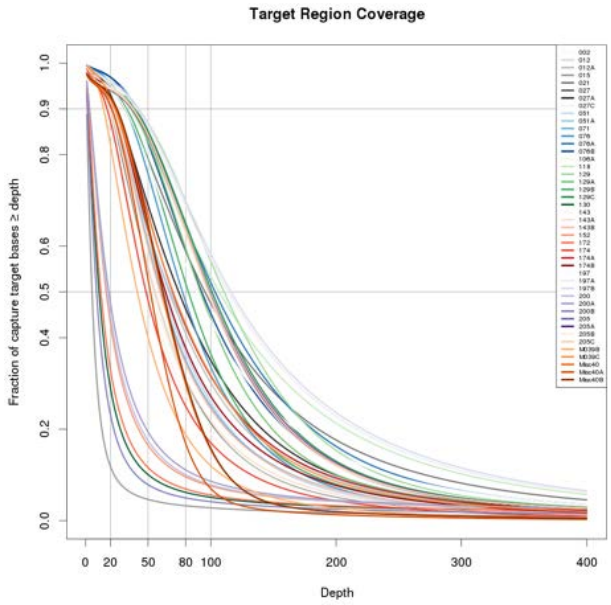


Figure 1.4 Coverage plot for exome sequence data - Nimblegen

*Assay performance.* The performance of the Nimblegen exome capture was assessed by comparing SNP calls made from the sequence data to SNP calls generated from same samples using SNP Arrays. In this analysis, a total of 576,812 calls overlapping both datasets were used. We found 92.5% correlation in the SNP calls. The proportion of discordant calls between heterozygous and homozygous calls was 2.4%. The proportion of calls with discordant allele calls (Reference vs. SNP) was 5.1%. These data suggest that the utility of both systems for genotyping purposes is very good and that the exome data generated are reliable. The estimated percent error as determine by the

discordant calls is considered acceptable given that these estimates are probably impacted by error in annotation and reference used in each technology, not to mention the limitations and strengths for each technology. For instance, small insertions and deletions are better detected using NGS; however large insertions and deletions as well as structural variation are better detected using SNP arrays.

Regarding assessment of the two different exome capture approaches used in this study, we compared the variants identified from the ACE enhanced capture method (Personalis) to those identified by the Nimblegen capture method. In the comparison of the variants identified from ACE enhanced capture following the strategy described in Chapter Three, we identified a total of 686 variants in the 6 families analyzed. To reiterate, the variants selected had to meet the following criteria: (1) read depth >10X, (2) genotype quality >20 and (3) rare in general population as determine by allele frequency. A query in the corresponding VCFs for corresponding chromosomal coordinates for the 686 variants revealed that only 481 of the coordinates were sequenced by both methods, representing 65% overlap. Seemingly, the region around a total of 205 variants was not enriched by the Nimblegen exome captuture or the sequence targeted was difficult to capture. Of the 481 overlapping variants, the genotype for a total of 372 variants was concordant between the two, representing a proportion of ~77%. The distribution of the 110 discordant calls is depicted in Table 1.10.

Allele drop-out or allele drop-in can result from instances where the template material is limiting or from PCR artifacts, respectively. PCR artifacts are random events and these are generally removed by selecting heterozygous calls with read proportions between 0.15 and 0.75 which we did in our selection of heterozygosity. Therefore, it can

be assumed that the heterozygous calls from ACE are likely correct. However, 35 of the 80 heterozygous calls by ACE are de novo and these rarely validate. Even with a liberal 10% error rate; together, the number of discordant calls at both the homozygous and heterozygous calls suggests poor enrichment of those regions in the Nimblegen method. Alternatively, the discordance could be a problem in the alignment process and variant call tools used – each of these is different in each case and the lack of concordance between these tools is well known (O'Rawe et al. 2013). These need further investigation.

Table 1.10 Characterization of discordant calls between ACE and Nimblegen exome capture

| SNV | ACE | Nimblegen |
|---|---|---|
| Heterozygous | | |
| de novo (n=35) | 80 | 31 |
| CompHet (n=45) | | |
| Homozygous | 27 | 75 |
| Indels | 3 | 4 |

*Remarks.* Purity and integrity of DNA purified from Tempus Blood RNA tube using a modified protocol was demonstrated to be good. SNP genotypes obtained from Nimblegen Capture and NGS technology were concordant with the SNP array data, >95% accuracy was demonstrated. There was also a good degree of concordance between SNP genotypes obtained using two different exome capture approaches. However, the number and type of variants identified as potentially interesting from each of the two different capture approaches was significantly different. There were 686 variants identified in the six different trios tested. Of these, a total of 371 variants were

concordant between the two capture methods, 205 variants were not detected or called in the Nimblegen capture samples, and 110 variant genotypes were discordant between the two.

Our data suggest that DNA purified from whole blood collected in Tempus RNA is adequate for sequencing and genotyping applications. Selection of targeted approaches, design of capture probes, and capture conditions must be carefully examined to ensure appropriate detection of targets of interest. In addition, off-the-shelf commercial kits need optimization in the laboratory to maximize reliability of the information obtained and to identify their limitations. Nevertheless, results are accurate for variants detected in most genomic regions. Complex genomic regions need enhancements to enable adequate coverage and detection of variants in those regions. Discordances observed in this study highlight how validation of variants using alternate technology (i.e. Sanger sequencing) and/or functional molecular biology work is paramount in this field of work.

*General Remarks*

This chapter provides a general overview of genomics technology. I also presented the reader with my background expertise in the application of gene expression profiling and how these can be applied to unravel molecular clues that when combined with other experimental data help corroborate molecular mechanisms underlying immune mediated disease in genetic defective backgrounds. Lastly, I presented background expertise in developing and implementing new technology. Together, this work helped lay the foundation and technical feasibility for a new investigative frontier that is the

genetic etiology of sJIA; this effort is presented in the next several chapters of this dissertation.

# CHAPTER TWO

## Introduction to Systemic Onset Juvenile Idiopathic Arthritis (sJIA)

### *Background*

Juvenile idiopathic arthritis (JIA) is a heterogeneous group of childhood disorders defined by chronic arthritis of unknown origin with disease onset under the age of 16 years. Systemic onset juvenile idiopathic arthritis (sJIA) is one of seven subtype disorders categorized under JIA. sJIA presents with a quotidian fever that persists for at least two weeks, a characteristic erythematous, evanescent skin rash and either lymphadenopathy, pericarditis, pleuritis, or hepatosplenomegaly (Ravelli & Martini 2007). When sJIA symptoms appear after age 16$^{th}$ it is called Still's disease of the adult. sJIA can occur in children of any age; however peak age of onset is typically between 1 and 5 years of age (Mellins et al. 2011)(Behrens 2008). Unlike other subtypes of JIA where a higher frequency of females are affected, sJIA occurs with similar frequency in both girls and boys. Patients with sJIA make up about 10-15% of JIA. Up to 35% of patients develop macrophage activation syndrome (MAS), a life threatening complication associated with cytokine storm syndrome with cytopenia, disseminated intravascular coagulation, hyperferritinemia, and hepatitis (Ravelli et al. 2012). sJIA has a variable prognosis, where 40%-50% of patients have either a monophasic or a polycyclic course. In a monophasic course patients experience symptoms over a single phase lasting up to 24 months; whereas in the polycyclic disease course several flare-cycles take place separated by inactive disease which may last months or years (Nigrovic 2014)(Woo 2006). The remaining 50%-60% of patients suffer from chronic, persistent arthritis

requiring extended treatment. The latter group of patients experiences the worst outcome, including growth failure, radiographically evident joint injury, and long-term disability.

The high potential for life-threatening complications compounded with the possibility for a physically damaging disease course makes sJIA a serious childhood disease. Many studies have been unable to identify predictive biomarkers that can identify, at disease onset, which patients will progress into the most damaging disease course (Nigrovic 2014). As such, markers for sJIA could be useful in predicting risk for disease as well as treatment response before or at time of disease onset. Studies reported involving sJIA patients have identified many types of biomarkers associated with the disease. These biomarkers can be categorized into serologic, cellular and transcriptional as well as genetic markers. Specific markers in each of these categories will be discussed in detail in the next several sections of this chapter.

The systemic nature of sJIA along with the accessibility to blood, make this an obvious biological source for discovering molecular markers. As discussed below, many serum biomarkers have been associated with disease activity and prognosis. There are also cellular biomarkers such as monocyte and NK cell unique phenotypes, as well as increased levels of precursor cells. Among transcriptional biomarkers, our group reported IL-1B overexpression in sJIA. This observation allowed us to identify this cytokine as a successful therapeutic target. In addition, genetic markers have been identified via genetic association studies; however, most of these studies involved genotyping a small number of genetic variants rather than a genome-wide approach. Most recently, a whole exome sequencing study was reported focusing on sJIA patients who developed MAS.

Findings from these various proteomic, genomic, transcriptomic and cellular studies on sJIA patients support an underlying molecular perturbation affecting innate immune responses. Such perturbations might be explained by underlying genetic variations. It is the aim of this body of work to identify candidate causal variants that could explain specific phenotypic characteristics of sJIA, as well as disease subtypes based on treatment response and age of disease onset.

## *Pathogenesis of sJIA*

### *Interleukin 1β (IL-1)*

This cytokine is an important mediator of inflammatory responses and is involved in a variety of cellular activities, including cell proliferation and apoptosis. IL-1 is produced as a pro-protein by activated cells and subsequently proteolytically processed to its active form by caspase 1 (CASP1/ICE). It has been identified as an endogenous pyrogen and reported to stimulate release of prostaglandin and collagenase from synovial cells. Discovery of the importance of IL-1 in the pathogenesis of sJIA came from gene expression studies showing induction of innate immune genes, including IL-1, upon stimulation of healthy peripheral blood mononuclear cells (PBMCs) with serum from sJIA patients ((Pascual et al. 2005). Studies assessing treatment response of sJIA patients with three modalities of IL1 antagonists: i) recombinant soluble IL-1 receptor antagonist, ii) IL1 receptor fusion protein, and iii) an anti-IL1B monoclonal antibody, have all reported significant amelioration of disease activity in the majority (60-70%) of patients treated (Pascual et al. 2005)(Grom 2014)(Verbsky & White 2004) (Autmizguine et al. 2015; Ruperto, Brunner, et al. 2012; Ruperto, Quartier, et al. 2012; Goh et al.).

*S100A8, S100A9 and S100A12 Proteins*

S100 proteins are calcium binding proteins that are exclusively expressed in vertebrates. They interact with enzymes, cytoskeletal proteins, receptors, transcription factors, and nucleic acids to regulate a wide variety of pathways including proliferation, differentiation, apoptosis, inflammation, cell migration, energy metabolism and calcium ion homeostasis (Tong et al. 2014)(Halawi et al. 2014). These proteins are secreted during neutrophil and monocyte activation. S100A8/A9 (MRP8/MRP14) form a complex that can serve as a TLR agonist triggering TLR4 signaling, resulting in production of IL-1B (Frosch et al. 2009)(Holzinger et al. 2012). S100A12 has also been shown to activate human monocytes via TLR4 (Foell et al. 2013). High levels of these proteins have been associated with active sJIA and may distinguish this disease from other febrile conditions such as systemic infection, leukemia, and Kawasaki disease (Wittkowski et al. 2008)(Ling et al. 2010a). Increased secretion of S100 proteins is a feature shared with other autoinflamamtory diseases such as Familial Mediterranean Fever (FMF) and Cryopyrin Associated Periodic Syndromes (CAPS) (Wittkowski et al. 2008). Moreover, serum concentrations for S100A8/9 are used to monitor response to drug treatment and disease activity. In fact, measurement of these proteins in serum has been shown to predict disease flares with 92% sensitivity and 88% specificity (Holzinger et al. 2012).

*Interleukin 18 (IL-18)*

This protein belongs to the interleukin 1 family of cytokines and is constitutively expressed in keratinocytes, epithelial cells and blood monocytes (Dinarello 2010). As

described for IL1B, IL18 production is dependent on caspase-1/inflammasome activation. IL-18 induces production of interferon gamma by NK cells and T cells as well as TNFa and chemokines secreted by macrophages (Avau et al. 2014). Levels of IL-18 in some sJIA patients can be up to 10-fold higher than in healthy individuals, whereas it is typically only 3-fold higher in other autoinflammatory conditions (Nirmala et al. 2014). Indeed, levels of IL-18 in serum of sJIA patients correlate with disease activity, decrease during clinical remission and can be predictive of MAS (Shimizu et al. 2014)(Shigemura et al. 2011). As shown for IL6, IL-18 levels correlate well with active versus inactive sJIA (Shimizu et al. 2014). Defective phosphorylation of interlukin-18 receptor has been reported in patients with active sJIA; this defect, however, is reversible (de Jager et al. 2009).

*Interleukin 6 (IL-6)*

Interleukin 6 is a pro-inflammatory cytokine with many functions. It is primarily produced at sites of acute and chronic inflammation and is subsequently secreted into the serum, acting on target cells through the IL-6R alpha. Target cells for IL-6 include B cells, T cells, hepatocytes, hematopoietic progenitor cells and cells of the central nervous system (CNS). In addition to the membrane receptor, a soluble form of IL-6R (sIL-6R) has been described and it binds IL-6 with the same affinity as the cognate receptor. The sIL-6R/IL-6 complex can activate cells via interaction with membrane bound gp130. This is an important feature of the complex as it acts as agonist for cell types expressing gp130 but would not normally respond to IL-6 alone; thereby expanding the repertoire of cells responsive to IL-6 (Jones et al. FASEB J 2001). IL-6 is relevant in sJIA, as shown by the effectiveness of anti-IL-6 therapy in clinical trials. IL-6 is highly elevated in sJIA

during peak of fever, and expression level is correlated with extent and severity of joint involvement, as well as platelet counts. Markedly microcytic anemia in sJIA patients characterized by a striking defect in the iron supply for erythropoiesis, with normal growth of erythroid colonies and production of erythropoietin, can be explained by IL-6 effects (Cazzola et al. 1996).

It has been reported that IL-6 stimulates both hypoxia-induced erythropoietin production and erythroid progenitor proliferation and, by increasing ferritin expression and hepatic uptake of serum iron, causes reticuloendothelial iron block (Ravelli et al. 2004). In addition, the iron malabsorption that has been observed in sJIA patients can be explained by the induction of hepcidin by IL-6. Hepcidin is an inhibitor of both iron uptake in small intestines and release of recycled iron in macrophages (Prieur & Stephan 1994)(Sawhney et al. 2001). Moreover, growth impairment is a well-known feature of sJIA occurring independently of steroid treatment. Studies have shown that IL-6-transgenic mice with high serum levels of IL-6 since birth show a decreased rate of growth, reaching only half the size of their wild-type counterparts as adults (Ravelli 2002). Similar to sJIA patients, these mice have normal levels of growth hormone production with low levels of insulin-like growth factor 1 (IGF-1) and IGF binding protein 3 (IGFBP3) (Ravelli 2002). These observations led to the hypothesis that sJIA is an IL-6 driven disease (Ravelli et al. 2005). This notion is further supported by treatment strategies successfully applied to patients using antibodies against IL-6 (discussed below).

*Ferritin*

Ferritin is the molecular form of iron that is stored in tissues. Iron can generate free radicals that are detrimental to cell membranes and other molecules; whereas ferritin stores iron in a soluble, non-toxic, readily available form. Ferritin has been shown to have both anti-inflammatory and pro-inflammatory properties. In sJIA ferritin levels can be extremely high and exceed 100ng/ml. Ferritin levels are typically used in the diagnosis of sJIA. Patients without MAS may still have very high levels of ferritin suggesting that this could very well be in and of itself a distinct pro-inflammatory marker unique to the disease (Nirmala et al. 2014). In the presence of MAS, ferritin levels are significantly higher. High ferritin levels have not been associated with FMF or CAPS.

*Macrophage Activation Syndrome (MAS) in sJIA*

Around 30% of sJIA patients are susceptible to developing a life-threatening complication, MAS. MAS is a form of reactive hemophagocytic lymphohistiocytosis (HLH) characterized by sudden onset of sustained fever, pancytopenia, hepatomegaly, liver insufficiency, a coagulopathy with hemorrhagic manifestations, and neurological symptoms (Behrens 2008)(Ravelli et al. 2012)(Martini 2012). HLH includes a variety of disease processes characterized by accumulation of well-differentiated mononuclear cells with a macrophage phenotype exhibiting hemophagocytic activity (Kaufman 9). HLH results from various genetic defects affecting cytolytic activity of NK cells and cytotoxic CD8+ lymphocytes (CTLs) (Kaufman 9). Affected genes include perforin (PRF1) (Stepp et al. 1999), Munc13-14 (Feldmann et al. 2003), Syntaxin 11 (STX11) (zur Stadt et al. 2005), and Syntaxin bidning protein 2 (STXBP2 or Munc 18-2) (zur Stadt et al. 2009). Other defective genes affecting NK cell function and associated with HLH disorders

include Rab27a (Menasche et al. 2000), LYST (Barbosa et al. 1996), SH2D1A (Coffey et al. 1998), and BIRC4 (Marsh et al. 2010).

*Role of Cellular Immunity in sJIA Pathogenesis*

Phenotypic characteristics of blood cell subsets in sJIA patients have been extensive. Two published reports have provided insight on immune cell subsets consisting of monocytes, dendritic cells, NK cells, α/β and γ/δ T cells and B cells in active and inactive disease states - identifying a T and B cell relative abundance lower in active sJIA compared to age-matched healthy controls (Macaubas et al. 2012a)(Macaubas et al. 2010). Reports assessing the frequency of Th1 and Th17 cells in sJIA patients are controversial. Omoyinmi et al. (Omoyinmi et al. 2012) reported an increased frequency of Th1 and Th17 in these patients regardless of disease state (active vs. inactive) compared to age-matched controls. Lasiglie et al. (Lasigliè et al. 2011), on the contrary, did not find such differences but reported that Th17 cells were more abundant in CAPS patients prior to treatment with an IL-1 antagonist.

Reports on NK cell frequency in sJIA are also controversial (de Jager et al. 2009)(Muller et al. 1998)(Macaubas et al. 2010), although a number of studies have shown decreased numbers of NK cells in these patients (Fall et al. 2007)(Wouters et al. 2002)(Zhou et al. 2013). In addition, frequency of NK cells has been negatively correlated with disease activity (Wouters et al. 2002). Moreover, cytotoxic activity of NK cells has been shown to be suppressed in sJIA patients (de Jager et al. 2009)(Villanueva et al. 2005)(Zhou et al. 2013)(Grom et al. 2003) and may explain the risk of MAS. Some have proposed a combination of genetic and environmental factors leading to NK dysfunction in sJIA (de Jager et al. 2009)(Grom et al. 2003)(Wulffraat

2003). In fact, perforin is under expressed in sJIA patients afflicted by MAS – this observation identifies a common pathway between sJIA, MAS and Familial Histiocytic Lymphocytosis (FLH) (Grom et al. 2003).

It has been proposed that the cytokine environment in sJIA may be responsible for the observed decreased cytotoxicity and might lead to a "helper" phenotype in NK cells rather than the canonical effector/cytotoxic phenotype (Avau et al. 2014). This hypothesis suggests a deficient pathway leading to induction of proliferation and cytotoxicity of NK cells by IL-2, IL-15 and IL-21 (Avau et al. 2014). Although it remains to be tested, genetic defects in these cytokines, receptors, and/or downstream signaling molecules could be responsible for observed NK cell dysfunction in sJIA patients.

An in-depth investigation of monocyte subsets carried out by Macaubas et al. (Macaubas et al. 2010) identified a significant increase of CD14+ monocytes in active patients compared to healthy controls or inactive patients. While the frequency of CD14++ or CD14+CD16+ monocytes was not different between patients and controls, expression levels of CD16 and CD14 were higher in patients (Macaubas et al. 2010)(Macaubas et al. 2012b). It is believed that CD16 is expressed by monocytes or macrophages that have a more inflammatory (M1) phenotype. CD14 is up regulated on monocytes with a more anti-inflammatory (M2) gene expression profile (Rey-Giraud et al. 2012). Macaubas et al (Macaubas et al. 2012b) reported increased expression of the prototypic M1 markers CD40 and CD80 on CD14++CD16- and CD14+CD16+ monocyte populations in patients with active disease compared to inactive patients and healthy controls. Moreover, they reported that an "M1-like" monocyte population in active

patients also expressed CD163 and CD209, two surface markers that are typically associated with an M2 phenotype. These observations suggest that monocytes from sJIA patients have a mixed inflammatory and anti-inflammatory phenotype. This mixed monocyte phenotype was confirmed in another report where expression of the M2-specific transcription factor Kruppel-like factor-4 was correlated with active sJIA disease (Y. Zhang et al. 2014). A study by Hinze et al. described a high frequency of immature cell populations in sJIA, including CD34+ cells that correlated strongly with an erythropoiesis gene signature. They also observed the same pattern in FLH and in active sJIA; however, the pattern was not present in inactive sJIA (Hinze et al. 2010). In these studies, pro-IL-1B was higher in sJIA samples compared to healthy controls, but release of mature IL-1B was lower, suggesting that sJIA monocytes may counteract the action of inflammatory mediators secreted by other cell types such as neutrophils, lymphocytes or endothelial cells.

Another potential cause for this peculiar mixed monocyte phenotype could be an underlying genetic defect affecting regulation of gene expression. Moreover, the high frequency of immature cell types observed in sJIA and FLH may be the result of both ineffective cellular development and differentiation as well as excessive phagocytosis of mature cells, which can be a consequence of the cytokine dysregulation.

Regarding IL6, differentiation of ATDC5 chondrogenic progenitor cells have been shown to be inhibited by over-expression of IL-6 in sJIA patients and could represent a potential cause of impaired growth in these patients (Nakashima & Takayanagi 2009). Taken together, these data support that pro-inflammatory cells play a

role in the pathogenesis of sJIA and that a quiescent disease state is in fact compensated inflammation.

*Gene Expression Profiles in sJIA*

Most gene expression studies involving sJIA patients have used PBMCs or whole blood. Interestingly, many of the genes differentially expressed in sJIA are up regulated instead of down regulated compared to either healthy controls or inactive patients. Studies carried out by our group pinpointed that factor(s) in sJIA serum up regulate IL-1 in healthy PBMCs, and went on to prove that IL1 is a key player and therapeutic target in the disease. Ex-vivo PBMC signatures of sJIA patients revealed a significant perturbation of innate immunity genes (Pascual et al. 2005)(Allantaz, Chaussabel, Banchereau, et al. 2007)(Allantaz, Chaussabel, Stichweh, et al. 2007). Genes comprising the inflammasome, as well as downstream of inflammasome activation, such as IL-1 and IL-18 as well as TLR signaling pathways are over expressed in sJIA compared to healthy controls (Barnes et al. 2009). Moreover, up regulation of neutrophil-related transcripts (Allantaz, Chaussabel, Banchereau, et al. 2007; Ogilvie et al. 2007) have been reported. A list of genes found differentially expressed in sJIA is included in Appendix A. Pathways enriched with over-expressed genes include IL-10 signaling and innate immunity pathways (Barnes et al. 2009). NK cell, T cell and antigen-presentation pathways are down regulated in sJIA (Avau et al. 2014). In a comparative study also carried out by our group using samples from patients with various infectious diseases, autoimmune diseases, and other autoinflammatory diseases, a total of 12 genes were uniquely expressed in sJIA patients (Allantaz, Chaussabel, Stichweh, et al. 2007). The putative gene function for 6 of the 12 genes was calcium transport and nuclear mRNA

splicing among others. Comparison studies of differentially expressed genes in sJIA show significant overlap with CINCA and CAPS relative to SLE, polyarticular JIA and Kawasaki disease (Ogilvie et al. 2007)(Balow et al. 2013)(Aubert et al. 2012)(Allantaz, Chaussabel, Stichweh, et al. 2007). Interestingly, a majority of genes overlapping between sJIA and CAPS (i.e. ALAS2, HB1, HB2, HBG and AQP9) are associated with immature cells (Nirmala et al. 2014). Furthermore, Hinze et al. (Hinze et al. 2010) reported that early myeloid progenitors and erythropoiesis differentiate sJIA from other types of JIA and a variety of febrile diseases. It is possible that the extreme inflammatory state in in sJIA, as it has already been demonstrated for CAPS, is the outcome of a genetic defect in cell cycle and/or differentiation process/mechanism of these cells.

*Genetic Markers Associated in sJIA*

Most studies investigating genetic risk factors for sJIA have applied targeted genotyping technology using small sample sizes. These studies have interesting findings in small portions of the samples studied. Most consistently, these studies described SNPs in the promoter and coding regions of the following genes: TNF, IL-6 (Woo & Humphries 2013), IL-10 (Fife et al. 2006), MIF, IL-1 family (Stock et al. 2008a), SLC26A2 (Lamb et al. 2007), P2X7 , MEFV, Munc13-4, PRF1 (Vastert et al. 2010), and IRF5 (Mellins et al. 2011). Others have made an attempt to identify genetic risk factors in sJIA through genome wide association studies (GWAS). Here, they associated IL6, IL18 and the IL1 family of genes to sJIA with some statistical significance (Stock et al. 2008b). Another GWAS found a weak association with a 3Mb interval containing genes involved in both innate and adaptive immunity, including BTNL2 (Nirmala et al. 2014). More recently, whole exome sequencing of 13 sJIA families where the probands had

developed MAS identified overlapping genetic markers with familial hemophagocytic lymphohistioctosis (FHLH) (Kaufman et al. 2014); however, this study did not report on genetic markers for sJIA as a population. A second exome sequencing study analyzed 13 sJIA patients from 5 consanguineous families and found a homozygous missense mutation in the enzyme laccase domain containing 1 (LACC1) in all 5 families studied (Wakil & Monies 2014). The clinical features presented in this particular study are not of those presented in bona fide sJIA.

### *sJIA as an Autoinflammatory Disorder*

Several studies have provided evidence suggesting that innate immunity plays a prominent role in sJIA (Vastert et al. 2009). Gene expression profiling studies have shown that sJIA is distinguished from other subtypes of JIA by up-regulation of innate immune pathways including IL1, TLR/IL1R, IL-6 and PPAR-gamma signaling pathways associated with down-regulation of gene networks associated with natural killer (NK), T cell and major histocompatability complex (MHC) antigen-related biological processes (Pascual et al. 2005; Allantaz, Chaussabel, Banchereau, et al. 2007; Fall et al. 2007; Ogilvie et al. 2007; Wittkowski et al. 2008; Barnes et al. 2009; Frosch & Roth 2008). Abnormalities in innate immunity pathways are associated with autoinflammatory syndromes (Hull et al. 2003; Church et al. 2008). An emerging consensus in the field is that sJIA should be classified as an autoinflammatory disorder. Compelling evidence supporting this notion include a large overlap of clinical manifestations such as fever, multisystem involvement (joints, skin, GI tract, eyes, etc.), recurrent episodes of inflammation, along with a predominance of monocytes and neutrophils as effector cells.

In addition, and as in autoinflammatory conditions, sJIA patients are at risk of amyloidosis (Vastert et al. 2009).

The general consensus in the field is that sJIA is a polygenic disease unlike other hereditary monogenic autoinflammatory diseases such as familial Mediterranean fever (FMS), cryopyrin associated periodic syndroms (CAPS), TNF-receptor associated periodic syndrome (TRAPS), hyper-Ig-D syndrome (HIDS) also known as mevalonate kinase deficiency (MKD), and PAPA syndrome. These rare genetic disorders share a common clinical phenotype and are for the most part highly responsive to IL-1 blockade. Overlap in clinical phenotypes and response to therapy between sJIA and autoinflammatory disorders makes a compelling case to search for an underlying genetic predisposition in innate immunity pathways in sJIA patients.

### *Treatment Strategies for sJIA Patients*

The overall treatment approach in JIA is designed to control disease and minimize disability while balancing excessive immunosuppression and side effects from drug treatment. A consensus treatment plan for new onset sJIA cases was published in 2012 (DeWitt et al. 2012) and for the first time includes IL-1 and IL-6 inhibitors as the first line of therapy in conjunction with disease modifying anti-rheumatic drugs (DMARDs), as well as non-steroidal anti-inflammatory drugs (NSAIDs) (Harris et al. 2013). Effectiveness of these drugs in clinical trials is measured by a core set of six variables defined by the American College of Rheumatology (ACR) (Giannini et al. 1997). The six measures include: i) physician global assessment of disease activity, ii) parent/patient global assessment of wellbeing, iii) functional assessment, iv) active joint count, v) restricted joint count, and vi) erythrocyte sedimentation rate (ESR). Either the ACR Pedi

30 or the ACR Pedi 50 and 70 determine clinical endpoint. The ACR Pedi 30 is defined as a minimum of 30% improvement from baseline in a minimum of 3 out of 6 measures, with a worsening of 30% in no more than one measure. Accordingly, the ACR Pedi 50 and 70 responses are defined as a minimum of 50% and 70% improvement, respectively, in at least 3 out of 6 measures, with a worsening of 30% in no more than one measure (Giannini et al. 1997). Most clinical studies evaluating safety and efficacy of biologic treatments for sJIA patients have achieved an ACR Pedi 30 response (Harris et al. 2013).

Biologics targeting IL-1 have been effective in the treatment of sJIA. These biologics take advantage of the naturally occurring agonists and antagonist receptor functions. The activity of IL-1 is mediated by binding to the cell surface receptor IL-1RI (the agonist receptor); this interaction induces a conformational change in the extracellular domain of IL-1RI enabling binding to IL-1RAP (IL-1R accessory protein) which is required for intracellular signaling. IL-1RII is an antagonist receptor in its membrane bound form or soluble form. IL-1RII acts as a decoy receptor on the cell surface or in the extracellular environment after cleavage of the extracellular domain. IL-1RI and IL-1RAP also exist as soluble forms produced by cleavage of extracellular domain or alternative splicing, respectively. sIL-1RAP can form an inactive complex with cell surface IL-1RI bound to IL-1 and can also increase the antagonistic potency of sIL-1RII (Jesus & Goldbach-Mansky 2014a).

*Anakinra*

Anakinra is the best-studied IL-1 inhibitor. It is a soluble recombinant IL-1 receptor antagonist. IL-1 signaling is inhibited by a naturally occurring soluble receptor-blocking molecule, IL-1 receptor antagonist (IL-1RII). This type of inhibitor prevents

interaction of the agonist receptor with IL-1 to inhibit an inflammatory response. Use of anakinra for sJIA was first suggested by our group (Pascual et al. 2005) and was followed by a randomized, double-blind, placebo-controlled trial (Quartier et al. 2011). In the randomized study, 8 of 12 patients receiving anakinra achieved an ACR Pedi 30 response compared to 1 out of 12 in the placebo group; subsequently 9 out of 10 patients in the placebo group were treated with the drug and became responders. A larger multicenter observational study assessing 46 patients reported 60% of patients reaching complete normalization of labs and no need for further therapy (Nigrovic et al. 2011). In this study, systemic symptoms (fever and rash) resolved in more than 95% of the patients within 30 days and no active arthritis was reported in 61% of the patients at average follow-up time of 14.5 months.

*Canakinumab*

Canakinumab is a humanized monoclonal antibody against IL-1B (Goh et al.). A multi-center open-label study of 25 sJIA patients treated with canakinumab reported that 61% of the patients achieved an ACR Pedi 30 result (Ruperto, Quartier, et al. 2012). Another phase III trial on sJIA patients with active disease reported an ACR Pedi 30 response in 84% of the treatment group with 10% in the placebo group after a single dose (Ruperto, Brunner, et al. 2012). Responders from this study were randomized for a second trial that followed 32 weeks of open-label treatment and measured time to flare as primary outcome. Ruperto et al. reported 74% of the canakinumab group had no flares after the withdrawal phase compared to 25% of patients in the placebo group.

*Rilonacept*

Rilonacept blocks IL-1 signaling via a fusion protein consisting of the IL-1 receptor (R1) and the Fc portion of IgG. This construct blocks the interaction of IL-1 by competing with the IL-1R1 on the cell surface (Jesus & Goldbach-Mansky 2014b). Preliminary results from a double-blind, placebo-controlled study of sJIA patients shows improvement in all ACR core variables (Ilowite 2014). Just recently, a multicenter, randomized, placebo-controlled study of rilonacept efficacy and safety in sJIA treatment was published. They assessed time to response as the primary end point using the PED30 criteria and found the time to respond was shorter in the rilonacept arm compared to placebo. Specifically, 57% of patients had a response at week 4 compared to 27% of patients in the placebo arm (Ilowite et al. 2014; Autmizguine et al. 2015).

*Tocilizumab*

Tocilizumab is anti-IL-6 soluble receptor monoclonal antibody (Tanaka et al. 2014). This biologic therapy is recommended with or without glucocorticoids as one of four consensus treatments plans for patients with new-onset sJIA (DeWitt et al. 2012). While IL-6 is produced by different types of cells, the IL-6R is mainly expressed in haemopoietic cells such as T-cells, monocytes, activated B-cells and neutrophils. IL-6 carries out its biological activities through IL-6R and gp130. When IL-6 binds to membrane bound IL-6R homodimerization of gp130 is induced forming a high affinity functional receptor complex consisting of IL-6, IL-6R and gp130. IL-6 receptor can become soluble after cleavage of its intracytoplasmic portion by ADAM17 or alternative splicing. sIL-6R can also bind IL-6 and form a complex with soluble gp130 resulting in

expanded activity of IL-6 to cells lacking IL-6R, also known as IL-6 tans-signaling (Mihara et al. 2012).

*Non-Steroidal Anti-Inflammatory Drugs*

Up until the development of biologic therapies, such as those described above, disease-modifying anti-rheumatic drugs (DMARDs) like methotrexate and sulfasalazine were the nonsteroidal medications of choice for JIA therapy (Harris et al. 2013). Nonsteroidal anti-inflammatory drugs (NSAIDs) have been the most widely used medications in arthritis. NSAIDs have both pain relieving and anti-inflammatory effects. These drugs block prostaglandin formation via inhibition of cyclooxygenease-1 and cyclooxygenase-2. Two such drugs, Naproxen and Meloxicam, have undergone a randomized, double-blind clinical trial each satisfying the ACR Pedi 30 with no statistically significant differences in efficacy when compared to each other (Ruperto et al. 2005).

*Steroids*

Glucocorticoid joint injections are also employed in the treatment of JIA. The use of systemic glucocorticoids is limited to sJIA patients rather than other JIA subtypes because of significant side effects with long-term steroid use (DeWitt et al. 2012)(Schaller 1977). Another drug part of the consensus treatment plan for new-onset sJIA is methotrexate (DeWitt et al. 2012). Methotrexate is a folic acid analogue that competitively inhibits dihydrofolate reductase interfering with purine biosynthesis and DNA production (Mediero et al. 2014). This non-biologic DMARD was found to be an

effective therapy for extended oligoarticular and sJIA patients in two randomized clinical trials (Giannini et al. 1992)(Woo et al. 2000).

*Autologous Stem Cell Transplant*

While the now conventional biologic therapeutic strategies and mainstay NSAIDs have resulted in efficacious treatment of sJIA, there remains a relatively sizeable fraction of patients who do not respond to any of these treatments due to toxicity or ineffectiveness. T cell depleted autologous stem cell transplantation (ASCT) has shown promise in inducing disease remission in children with refractory progressive systemic or polyarticular juvenile idiopathic arthritis (Brinkman et al. 2007). However, required pre-transplant immunosuppressive drugs and parallel treatment can be debilitating and sometimes fatal. While T cells have not been shown to play a critical role in sJIA pathogenesis, this treatment approach came about after two reports describing synovial CD4+ T cells with activation markers that secrete Th1-type cytokines with restricted TCR heterogeneity (Grom & Hirsch 2000)(de Kleer et al. 2004). Indeed, in a multicenter, prospective, phase II clinical trial where 22 patients with progressive refractory JIA were followed up for a median periods of 80 months after intensive immunosuppression and T cell depleted ASCT resulted in sustained complete remission in 15 out of 22 patients (68%) (Brinkman et al. 2007). One patient had a late relapse correlating with onset of puberty while the other five relapsed and had progressive disease, two of whom died at a later date. More recently, a study assessing T cell receptor (TCR) clonal diversity by TCRVb CDR3 length spectratyping in five sJIA patients before and after ASCT reported an association of remission from severe arthritis with an immune system composed of re-emerging T-cell clones that were present before

transplant as well as newly generated clones (Wu et al. 2014). In one patient who relapsed the presence of full length TCR CDR3 diversity, as assessed by TCR sequencing early during immune reconstitution, was suggested to be indicative of treatment response. Taken together, this treatment approach supports that, in addition to innate immunity, an antigen-driven T cell-mediated component plays a role in the pathogenesis of sJIA.

# CHAPTER THREE

## Materials and Methods

### *Patient Samples*

A total of 19 patients with Systemic-onset Juvenile Idiopathic Arthritis (sJIA) were recruited for this study through the pediatric rheumatology clinic at the Texas Scottish Rite Hospital for Children. All patients were consented and enrolled under an appropriate IRB-approved study protocol. Of the 19 SJIA cases, seven included samples from both of the patient's parents (trios). Participating patients were selected by the treating physician based on an extreme clinical phenotype determined by 1) response to anakinra treatment (responder vs. non-responder) and 2) age of disease onset (<2yo vs. >2yo).

### *Sample Preparation*

#### *DNA Extraction*

Whole blood (3ml) was collected via venipuncture in Tempus RNA Tubes (LifeTechnologies). A method for purifying genomic DNA from blood collected in an RNA stabilizing reagent, PAXgene Blood RNA Tubes (PreAnalytiX), was previously described (Kruhøffer et al. 2007). We followed a similar approach using up to 600ul of whole blood from Tempus tubes for DNA purification using a modified protocol for the QIAamp DNA Blood Kit (Qiagen). Samples were treated with equal volumes of Buffer ATL (Qiagen), 20U of RNase A (Qiagen) and 20µl of Proteinase K (Qiagen). After a 10 minute incubation at 70°C, DNA was precipitated using ethanol and subsequently bound

to and purified using QIAamp silica membrane columns. DNA yield and purity were assessed using the Qubit (Invitrogen) and NanoDrop 1000 (ThermoScientific), respectively. Integrity, as measured by size distribution, was assessed using the Genomic DNA Kit and Caliper Lab Chip GX capillary electrophoresis system (Perkin Elmer). By means of a proprietary algorithm the LabChip GX software assigns a quality score in the rage of 0 to 5, with 5 representing highest integrity for samples with fragment sizes greater than 30Kb.

*DNA Library Preparation*

Barcoded DNA libraries were generated using the TruSeq DNA Library Prep Kit (Illumina) following the manufacturer's instructions. A total of 1μg of genomic DNA was subjected to fractionation by sonication using the Covaris S2 (Covaris) with a resulting fragment average size of 300bp. 3' overhangs resulting from shearing were removed using an endonuclease while 5' overhangs were filled in with polymerase activity. Following end repair, adenylation of 3' blunt ends was carried out to prevent fragments from ligating to each other during adapter ligation. This single 3'A of the genomic DNA fragment is complementary to the 3' T end of indexed paired end adapters, and these are joined in a ligation reaction. Resulting DNA library constructs (Figure XX) are purified and size selected using AMPure XP beads (Invitrogen). Amplification by PCR was used to selectively enrich for library constructs with adapter sequences on both ends. Purified, enriched library products were subsequently qualified by measuring yield and assessing size distribution using the Qubit (Invitrogen) and Agilent Bioanalyzer and DNA 1000 Chip, respectively.

*Exome Capture*

NimbleGen SeqCap EZ Human Exome Library v3 (Roche) was used to enrich for targeted genomic regions from prepared DNA libraries following manufacturer's recommendations. This is a solution-based method for targeted DNA capture-sequencing that is directed to 64 Mb of the known human DNA coding and non-coding regions (Bainbridge et al. 2010)(Clark et al. 2011). High-density overlapping baits (capture probes) designed with sequence complementarity to targeted genomic regions were combined with a total of 1ug of DNA library pool (4 DNA libraries/pool) for enrichment of targeted regions during a 72-hour incubation. Capture efficiency was assessed as recommended by the manufacturer using the ViiA7 qPCR instrument (LifeTechnologies) and was found satisfactory for all capture experiments. Integrity of the captured exome library and size distribution was assessed using the Bioanalyzer and High Sensitivity DNA Chip (Agilent) following manufacturer's recommendations. The concentration for each pool of captured libraries was measured using a KAPA Library Quantification Kit (KAPA Biosystems) and the ViiA7 qPCR instrument (LifeTechnologies). In addition, exome capture was carried out using ACE clinical exome (Personalis) and this was perform by a commercial company.

*Cluster Generation and Sequencing*

Exome captured libraries were sequenced either at the UT Southwestern Sequencing Core (Dallas, TX), the BIIR Genomics Core, and/or Personalis (Menlo Park, CA). A pool consisting of four exome captured libraries (16pM) was loaded per lane in a high throughput Illumina Flow Cell v2 for cluster generation in cBot (Illumina). Briefly, oligonucleotides complementary to the P7 or P5 sequence portion of the adapter sequence

are covalently bound to the surface of the flow cell at high density and are used to physically immobilize single stranded library templates to the surface of the flow cell. Solid-phase clonal amplification via bridge PCR occurs resulting in approximately 1000 copies of each fragment template in one cluster (Mitra 1999) –cluster density on the flow cell ultimately determines the sequence run throughput and can significantly impact data quality. Therefore, an optimization experiment was conducted to determine optimal library input that results in maximum cluster density with highest per base quality scores.

Sequencing of exome-captured libraries was carried out using TruSeq SBS Kit v3 sequencing reagents in the HiSeq2000 or HiSeq25000 (Illumina). The sequencing run targeted 2X100bp paired-end reads. The key feature in Illumina sequencing is sequencing by synthesis (SBS) employing reversible terminator chemistry (Bentley et al. 2008; Metzker 2010). Briefly, a universal sequencing primer is hybridized to clonally amplified libraries to initiate the sequencing reaction then four-color reversible terminator-bound dNTPs are added and allowed to incorporate into respective templates; fluorescence is imaged and the bound terminator is then cleaved. Subsequent cycles of this process occur simultaneously for millions of clusters generating millions of short sequence reads (sequencing by synthesis) representative of the library fragments prepared from the original DNA sample (Metzker 2010).

### *Exome Data Processing and Alignment*

Image analysis and base calling takes place in a HiSeq using instrument control software (ICS). This software uses raw images to identify cluster position and intensity; this information along with noise estimates are used to determine the sequence of bases read from each cluster. A confidence level for each base is calculated and the same

software also determines whether or not the read (cluster) passed filter, also known as the Chastity threshold. Instrument output files, .bcl, were converted to .fastq.gz files using CASAVA v1.8.2 (Illumina) during which pooled sequenced samples were also de-multiplexed. In addition, the FastQC package (Illumina) was used to assess quality of the FastQ files prior to alignment.

Reads that passed the Chastity filter were aligned to the human genome reference B37 using the BWA-MEM aligner (Li & Durbin 2009)(Li 2014). Subsequently, aligned reads were furthered processed by Picard and Samtools to ensure removal of sequence duplicates and other library/sequencing artifacts as well as to ensure proper formatting of SAM to BAM conversion. Alignment quality scores were then recalibrated and realigned using GATK as previously described by De Pristo and Mc Kenna (DePristo et al. 2011; McKenna et al. 2010). The resulting output .csv file from this process was used as input for variant calling.

### *Variant Calling and Annotation*

GATK2, Atlas SNP V1.4.3 (Challis et al. 2012), Samtools V0.1.19 (Li & Durbin 2009) and FreeBays V0.9.7 (Garrison & Marth 2012) software programs were used to detect genomic variants. Integration of predicted SNPs from these algorithms was carried out using BAYSIC (Cantarel et al. 2014). BAYSIC calculates a posterior probability which represents likelihood in the accuracy of a variant call made by a combination of two or more variant callers. A posterior probability score is used to filter variants that meet a set threshold. Variants passing this filter are included in the variant call file (VCF) and it includes sequence quality information, read count, genotype call,

and genotype quality score as well as other useful information for evaluating the reliability of any variant call.

Annotation of variants was integrated using snpEff software (Cingolani et al.), which uses the GRCh37.69 database. This software annotates variants based on their genomic locations and predicts variant effects on corresponding genes. In addition snpEff, VarSeq (GoldenHelix) was also used for variant annotation. Allele frequency to the general population was determined by comparison to the 1000G (The G1K Consortium 2012) and Exon Varian Server (EVS) (http://evs.gs.washington.edu/EVS) databases. In addition, variants were annotated using SnpSift (Cingolani et al.), dbSNP (http://www.ncbi.nim.nih.gov/snp/), ANNOVAR (Wang et al. 2010), CLINVAR (Landrum et al. 2014), dbNSFP (Liu et al. 2011; Liu et al. 2013) and ENCODE (Bernstein et al. 2012; Pazin 2015) software and databases. Useful annotation integrated various software included genotoxic chemical mutational information, common variant frequency information, gene base annotation, region-based information, phylogenetic conservation scores, membership in clinically relevant databases, functional DNA elements, as well as other quantitative scores that predict "deleteriousness" for non-synonymous variants.

### *Selection of Candidate Variants*

sJIA is a sporadic, rare disease. There are no documented cases of bonafide sJIA occurring in families. Therefore, selection of candidate variants for this project was limited to inheritance models which could result in expression of a phenotype in a child birthed from two parents that did not express the phenotype themselves. Hence, de Novo, homozygous recessive, and compound heterozygous modes of inheritance were

the focus of our study. In the analysis of trio data we employed a discrete filtering strategy using either Excel (Microsoft), JMP Genomics (SAS), and/or VarSeq (Golden Helix).

*de Novo Variants*

In contrast to germ line mutations, *de novo* mutations are either acquired mutations in one cell as a result of replication errors or induced by environmental conditions. Somatic mutations that occur in germ cells are not present in any other cell of the individual's body but will be present in every cell of his/her offspring. Somatic mutations may also occur at the time of fertilization, immediately afterward just prior to first cell division, or during early stages of development. Somatic mutations that occur in a single cell in the early stages of development lead to mosaicism. Mosaic individuals display genetic variation is some cells or tissue types but not others. Mosaicism has been associated with disease states (Ng et al. 2013). We did not include a mosaic model in this work due to limitations in sample collection defined by the original IRB approved study design.

Variant selection from exome sequencing data has relied heavily on discrete filtering strategies. Such an approach has been successful in identifying causal variants in Mendelian diseases (Ng et al. 2010)(Ng et al. 2009)(Choi et al. 2009)(Erlich et al. 2011). Our specific strategy in the selection and prioritization of de novo candidate variants is described in Figure 3.1. This strategy aims to arrive at candidate variants or genes by selectively depleting variants in a process that can be parsed into three categories: (1) quality, (2) frequency, and (3) zygosity.

On average, 20 to over 100 novel mutations are identified per individual (Bamshad et al. 2011). Identification of somatic variants remains a challenge in the genome technology field, particularly in cancer studies. In non-cancer studies, the challenge in identifying novel mutations is due mostly to technical error introduced in the process. DNA damage can lead to inaccurate information by either inhibiting a reaction (deamination of DNA) or through acquired nucleotide transitions. Additionally, sequencing errors are known to occur as well as alignment error. Our quality filter strategy aims to minimize false positive novel variants.

To enhance confidence in our selection, we applied a quality filter that removed variants for which 1) the total read count at the specified genomic coordinate was less than 10 and 2) the read count for the alternate allele was less than 5. This approach essentially selects variants within regions that had been sequenced with a minimum depth of 10X coverage, thereby, minimizing false positive de novo candidates. In addition to depth, we selected for variants with a quality score >30 which enriches for variants with high confidence calls.

Next we selected variants with alternate allele frequency less than or equal to 1% in the general population using frequency values from the 1,000 Genomes database. Allele frequencies applied were specific for the patient's ethnic background. We made use of sequence data generated from non-affected controls and removed any variant detected in those samples. Because this study was done in collaboration with Drs. Wise and Rios from Texas Scottish Rite Hospital for Children (TSRH) in Dallas, we were able to further remove variants present in their private exome databse consisting of 117

subjects. This further enhanced variants that were unique to each subject. Results from this process are detailed below (Chapter Three).

*Homozygous Recessive Variants*

A homozygous recessive mode of inheritance is another mechanism by which a genetic trait can be expressed in a child but not in either of his/her parents. In this inheritance model, each parent carries two different copies of the same gene (i.e., both parents are heterozygous for a given genetic loci). One of the copies is normal, or reference and the other is a mutant. In terms of biological impact, this heterozygous condition results in reduced expression of a fully functional gene product and consequences of the mutation could vary depending on the importance/relevance of the gene product for a given biological function. On the other hand, a homozygous recessive genotype results from inheritance of two of the same copies of a gene. Such inheritance can be homozygous to either the reference or alternate allele. For this study, we selected for homozygous genotypes to the alternate allele.

The discrete filtering strategy followed for this inheritance model was the same as that for the de novo model for quality and frequency cut-off thresholds. Accuracy of the zygosity call for each family member was optimized by selecting variants with read call ratio of 1 (100% alternate allele) for homozygous alleles; and the read call ratio for both parents was between 0.15 and 0.85 for heterozygous alleles. This strategy ensures that the child is truly homozygous – all reads are for the one variant called; while reference to alternate ratio of reads for each parent ensures their heterozygous genotypes. Results of this process are detailed below (Chapter Three).

*Compound Heterozygous Variants*

Compound heterozygosity is another mode of inheritance whereby a genetic trait could be observed in progeny from two parents who do not express the same trait. In this model each parent carries one or more heterozygous alleles in the same gene at different loci. After fertilization the child inherits one chromosome from their mother containing the heterozygous allele at one of the loci and the second chromosome from their father with the other heterozygous allele at a different locus for the same gene. It is imperative to be able to determine that at least one of the two required heterozygous alleles is contributed either paternally or maternally, but not both. Therefore, parental sequence data are required for identifying compound heterozygote candidate variants.

A discrete filtering strategy for selection of candidate compound heterozygous variants followed the same strategy as the inheritance models described above. The strategy for selection of compound heterozygote variants was: (a) depletion of variants for which either parent was homozygous for the alternate allele; (b) depletion of variants for which both parents were heterozygous; (c) depletion of genes with no multiple hits; and (d) depletion of genes where multiple variants in the same gene were contributed by only one parent. This strategy enhanced final selection of variants meeting the definition of compound heterozygote alleles. Results of this process are detailed in Chapter Three. The general discrete filtering strategy applied is described in figure 3.1 and was adopted from a similar strategy described by Kim et al. (Kim et al. 2015).

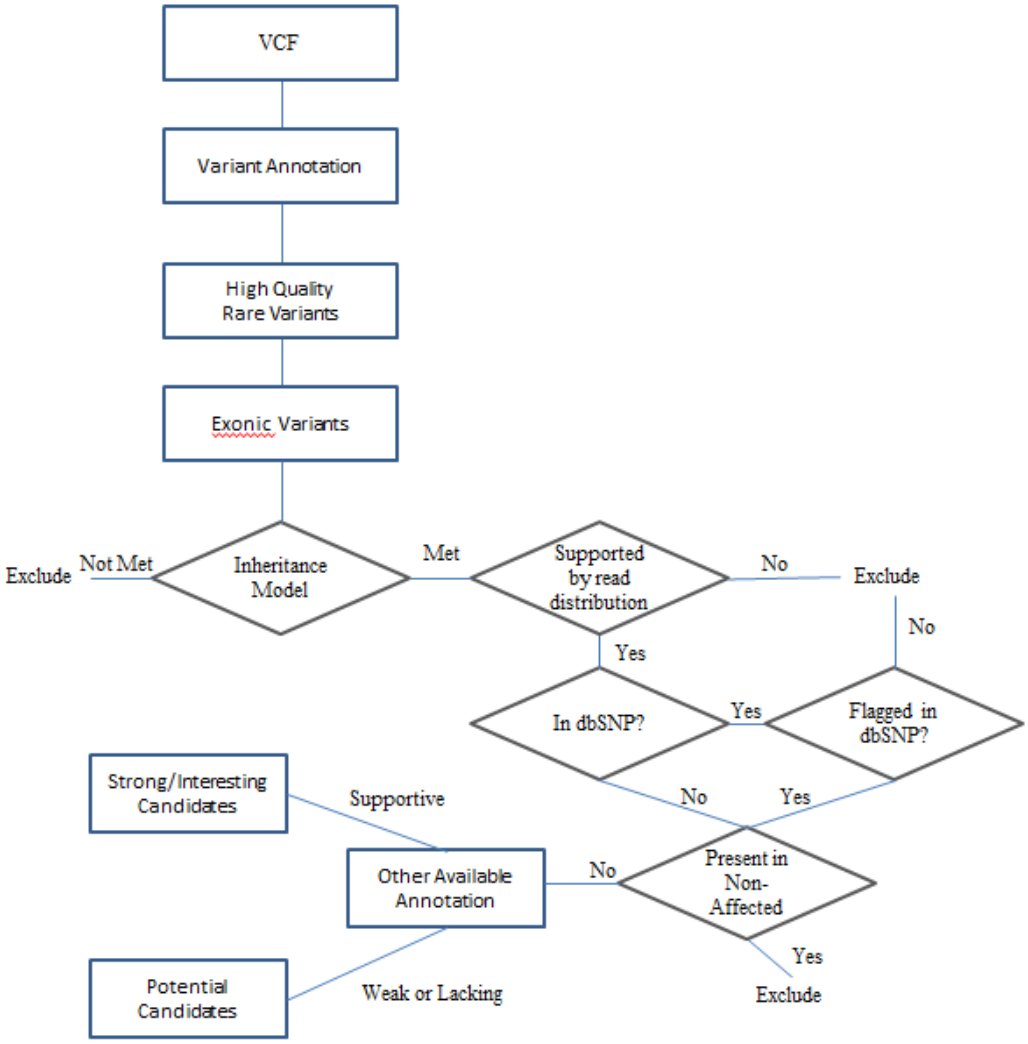


Figure 3.1 Generalized discrete filtering strategy employed in the selection of candidate variants VCF file is annotated using ANNOVAR. Using VarSeq software variants with GQ>20, read depth >10 and allele frequency in 1KG and ESP6500 <1% ware selected. After inheritance model, homozygous variants ware verified by read distribution (0 reads for reference) and 0.15-0.85 ratio for heterozygous calls. For de novo variants only, variants were screened against dbSNP 137 (UCSC). GERP and functional prediction scores, gene function, literature and other database searches are carried out to identify strong and interesting candidates.

*Prioritization of Candidate Variants*

In general, the number of variants identified in a single whole exome (coding region) is over 20,000, and this varies according to ethnic background (Bamshad et al. 2011). Sizable lists resulting from discrete filtering strategies must be judiciously reduced to a practical size which can then be carefully investigated through an intensive literature search and subsequent functional validation.

The aim of such a prioritization process is to narrow candidate variant lists to include only those variants that are potentially pathogenic. This process is dependent on identifying variants predicted to have a damaging effect on protein function or variants predicted to disrupt functional non-coding regions of the genome. In this thesis work, we have stratified variants within the coding region only. There is no specific consensus in the field as to the best approach or strategy to take in the prioritization of variants. Variables weighted in the prioritization process include those commonly used in published exome studies and incorporate mutational effect annotation, functional prediction scores, conservation scores, and gene function as evidenced by published literature.

It is difficult to determine potential pathogenicity for any given variant based on resulting changes in nucleotide and protein sequences alone. Although limited, available variant annotations and algorithms that can quantitatively predict the effect of these changes have proven to be particularly useful in the process of variant prioritization (Wang et al. 2010). Annotation for mutational effect was achieved using ANNOVAR as described previously in the 'variant calling and annotation' section. Non-synonymous and loss-of-function variants were prioritized. Loss of function variants included

frameshift, stop gain/loss, start gain/lost, and those occurring at splice sites. Loss of function variants would be functionally more deleterious compared to non-synonymous variants and were weighted accordingly.

There are many algorithms that have been developed for the purpose of predicting conservation scores. Available phylogenetic conservation annotation algorithms arose from comparative genome sequence analysis studies. All of these algorithms are based on claims of the theory of natural selection. The theory of natural selection assumes that most of the evolutionary changes at a molecular level are random and selectively neutral and are maintained in a species by balance between new variation and random extinction (Kimura 1991). This theory is in sharp contrast to the Darwinian Theory which claims mutational changes are selective for the survival of a species, or selective fitness. Sequencing technology has provided compelling evidence in support of the theory of natural selection (Kimura 1991). Further, phylogenetic analysis of orthologous sequences has characterized regions under purifying selection, or constraint, which corresponded with functional elements (Cooper & Brown 2008). Further, these conservation annotations have been successfully weighted in selection of causal variants in studies of Mendelian and complex diseases (Cooper & Shendure 2011)(Bamshad et al. 2011).

Genomic Evolutionary Rate Profiling (GERP) score identifies constrained elements within exonic and non-exonic regions that correspond well with functional elements (Cooper et al. 2005). Therefore and according to the neutral theory, high scoring variants are predicted to be highly, evolutionarily constrained genomic elements – likely to modify a phenotype, in other words. The Polymorphism Phenotyping

(PolyPhen) scoring method applies an algorithm which estimates conservation with less resolution compared to GERP. However, PolyPhen incorporates information about known biochemical properties for the affected amino acid sequence and estimates probability of impact given the amino acid substitution (Ritchie & Flicek 2014; Adzhubei et al. 2013). PolyPhen is optimized for non-synonymous mutations. Like GERP, the PolyPhen estimate is selective for SNPs that are more likely to modify a phenotype. Both GERP and PolyPhen scores were used for prioritization of variants in this study – each weighing in for evolutionary conservation and functional prediction, respectively.

There are limitations in both GERP and PolyPhen algorithms. For instance, scores generated by GERP are limited to inclusion of available orthologous sequences and could introduce a certain amount of bias. Moreover, there are concerns in the field regarding prioritization of species to be included in comparative genome analyses for the purpose of determining conservation status as defined in these algorithms. On the other hand, PolyPhen is limited to protein coding regions, and the biochemical properties of the vast majority of proteins are neither well studied nor well annotated. As a result, many variants do not receive a score. Therefore, in our final candidate gene list, we have included variants in the coding regions of genes known to be associated with sJIA or other autoinflammatory diseases irrespective of their conservation and/or functional prediction scores.

### *Validation of Candidate Variants*

Validation of variants was carried out using available RNASeq data for respective patients, re-sequencing of exome by 3$^{rd}$ party or targeted re-sequencing using Sanger sequencing. PCR primer design was to the variant flanking regions using Lasergene

software (DNASTAR®) or the NCBI Primer-Blast Website. Amplification and cycle sequencing was carried out using BigDye® Direct Cycle Sequencing Kit (Life Technologies) following manufacturer's instructions. Cycle sequencing products were analyzed using an AB3130 analyzer (LifeTechnologies) at the UTSW Sequencing Core. Sanger sequence data were analyzed using Lasergene software (DNASTAR). Third party sequenced data were processed using their proprietary method and analysis pipeline. VCF variant files were annotated and analyzed using VarSeq (GoldenHelix).

# CHAPTER FOUR

## Results and Discussion

### *Clinical Characteristics of sJIA Cohort*

DNA from a total of 19 sJIA patients and 21 non-affected controls, including six family trios, was subjected to exome sequencing. The patients were selected by the treating clinicians on the basis of severe phenotype, the patient's age at disease onset (early vs. late) and/or response to IL1 blockade (responder vs. non-responder). The details of these criteria are summarized in Table 4.1:

Table 4.1 Clinical Characteristics of sJIA Cohort. In the Gender column, M stands for male and F stands for female. In the Anakinra Response column, R stands for responder, NR stands for non-responder, and PR stands for partial responders

| Family ID | Proband ID | Gender | Race | Ethnicity | Age Group Disease Onset | Anakinra Response |
|---|---|---|---|---|---|---|
| Family1 | SYS_076 | M | White | Hispanic | 2-6yo | PR |
| Family2 | SYS_129 | M | White | Hispanic | 2-6yo | R |
| Family3 | SYS_143 | F | White | Non-Hispanic | 2-6yo | PR |
| Family4 | SYS_174 | F | White | Non-Hispanic | >6yo | NR |
| Family5 | SYS_197 | F | White | Non-Hispanic | <2yo | NR |
| Family6 | SYS_205 | F | White | Non-Hispanic | 2-6yo | R |
| | SYS_002 | F | White | Non-Hispanic | <2yo | PR |
| | SYS_012 | F | White | Non-Hispanic | <2yo | PR |
| | SYS_015 | F | White | Hispanic | <2yo | PR |
| | SYS_021 | F | Asian | Non-Hispanic | <2yo | NR |
| | SYS_027 | M | AA | Non-Hispanic | 2-6yo | NR |
| | SYS_051 | F | White | Non-Hispanic | <2yo | PR |
| | SYS_071 | F | White | Hispanic | >6yo | R |
| | SYS_106 | F | White | Hispanic | >6yo | R |
| | SYS_118 | M | AA | Non-Hispanic | >6yo | R |
| | SYS_130 | M | White | Non-Hispanic | <2yo | R |
| | SYS_157 | M | White | Non-Hispanic | <2yo | R |
| | SYS_172 | F | White | Non-Hispanic | <2yo | NR |
| | SYS_200 | M | White | Hispanic | >6yo | R |

*Overall Filtering Strategy in Individual Trios*

Table 4.2 Summary of discrete filtering strategy in family trios

| Variants Retained | Family 1 | Family 2 | Family 3 | Family 4 | Family 5 | Family 6 |
|---|---|---|---|---|---|---|
| Total # of Variants | 309,778 | 310,564 | 297,874 | 298,774 | 305,169 | 336,077 |
| Depth >10 | 227,698 | 212,330 | 226,633 | 212,492 | 236,051 | 248,845 |
| GT Quality >20 | 222,093 | 206,547 | 221,772 | 207,541 | 230,759 | 243,127 |
| Alternate Allele Frequency <1% | 25,084 | 23,689 | 24,060 | 22,069 | 25,332 | 27,937 |
| Effect | | | | | | |
| Non-Coding | 16,840 | 15,573 | 16,263 | 14,906 | 17,493 | 18,553 |
| Synonymous | 6,277 | 6,173 | 6,165 | 5,524 | 6,050 | 7,250 |
| Missense | 1,830 | 1,810 | 1,551 | 1,520 | 1,659 | 1,979 |
| Loss of Function | 137 | 133 | 111 | 119 | 130 | 155 |
| Model | | | | | | |
| Potentially Novel | 33 | 42 | 37 | 46 | 36 | 127 |
| Recessive | 18 | 25 | 8 | 5 | 4 | 5 |
| Compound Heterozygous | 60 | 76 | 51 | 44 | 80 | 14 |
| Hemizygous | 25 | 28 | 18 | 4 | 8 | 1 |

*Family 1*

The number of Family 1 participants in our exome analysis included the proband and both parents. The proband is a Hispanic male with disease onset between the ages of 2 and 6. The clinical course of disease for this patient include severe arthritis with difficult to control systemic features. He was treated with anakinra and, originally, the systemic features ameliorated and the severe joint features improved. After several months of positive, although not complete, response to anakinra treatment the patient flared and was found from then on non-responsive to IL1 blockade including both systemic and joint features.

*Family 1 Pedigree*

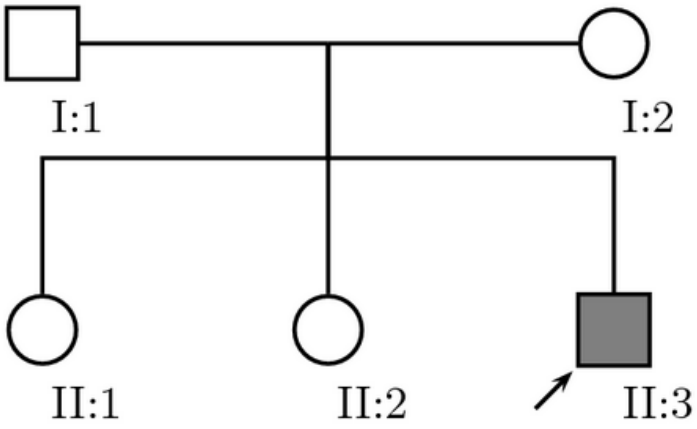


Figure 4.1 Family 1 pedigree. Squares designate males, circles designate females. Shaded square with arrow designate affected proband.

*Candidate Variants for Family 1 Proband*

A total of eight variants found in Proband 1 can be classified as strong candidates given their functional prediction or high conservation scores in addition to their associated gene function and a priori knowledge of association in sJIA. These include 7 missense variants and 1 splice-donor site variant affecting a total of 7 genes: LOXHD1, SDK1, ARHGEF10, A2M, ACAD11, KRT2, and ASPSCR1. The remaining variants did not have assigned functional prediction or conservation scores; therefore, available gene function, pathway, and interactions annotations were used for their selection. Specifically, genes associated with immune function, predicted to interact with known immune relevant genes and/or signaling molecules up stream or downstream of immune relevant pathways/networks. We found several genes that met these criteria and include IL1RAPL2 (aka IL1R9), TEX13A, FOXC1, ATXN2, MST1L, VPS13C, FCGBP, CCBL2, RBMXL1, IGFN1, TRRAP, TFAM, ITGA7, HERC2, among others.

Table 4.3 Discrete Filtering Strategy for Family 1

| Inheritance Model | Discrete Filtering Strategy | Candidates* |
|---|---|---|
| Autosomal Recessive | | |
| Homozygous | AF less than 1% in 1KG/EVS | 18 |
| Compound heterozygous | AF less than 1% in 1KG/EVS | 60 |
| Autosomal Dominant | | |
| De novo | AF less than 1% in 1KG/EVS, Not in dbSNP or Flagged in dbSNP | 33 |
| Hemizygous | | |
| Mitochondrial | AF<1% in 1KG/EVS | 2 |
| X-linked | AF<1% in 1KG/EVS | 23 |

**A complete list of these variants is available upon request.*

We prioritized as a strong candidate variant the Alpha-2-macroglobulin precursor (A2M). Proband 1 is homozygous for a missense mutation (LEU>261>PRO) with a high GERP score (5.2). A2M is a key member of the alpha macroglobulin superfamily with a variety of complex functions. A2M is primarily known for its ability to bind to all four classes of proteinases for subsequent delivery to an endocytotic clearance pathway. A2M regulates binding of transferrin to its surface receptor. It also binds defensin and myelin, NGF, PDGF, TNF-a, VGEF, IL-1B, IL-6 and modify their biological activities. Overall, A2M has been shown to have a protective role, but could also be pathogenic. High levels of A2M have been associated with liver fibrosis (Reehan et al. Cellular Phys 2012). Most importantly, A2M has been shown to be significantly over expressed in plasma from flaring sJIA patients compared to quiescent patients and is included in a

panel of 7 proteins (A2M, APO-A1, CRP, HP, S100A8/S100A9, SAA, and SAP) that demonstrated high specificity distinguishing sJIA flare from poly JIA and acute febrile illness (Ling et al. 2010b). These data suggest that A2M dysregulation could contribute to the plasma biomarker phenotype observed in sJIA.

In a survey of potentially dominant alleles assuming low penetrance we also identified a frameshift deletion in the S100A12 gene. This deletion was not found in any of the publically available databases, but it was also detected in the patient's father and as such this made our priority list. Given the co-expression of A2M and the recessive nature of the variant found on that gene and the relevance of S100A12 in sJIA, the frameshift deletion observed in the S100A12 was further examined and is discussed in Chapter Five.

Another interesting finding was an x-linked frameshift variant in the IL-1 receptor accessory protein-like 2 (IL1RAPL2), also known as IL-9R. This protein was just recently described and has been associated with cognitive function (Jin et al. 2000). However, the protein sequence contains the same functional domains as the IL1RAP and is ~80% homologous. IL1RAP is important in the IL-1 signaling pathway – the central mediator involved in acute and chronic inflammatory responses. IL-1 alpha and IL-1 beta (IL-1) induce cellular responses via binding of its receptor composed of two subunits, IL-1RI and IL1RAP. IL1RAP is essential for IL-1 induced activation of IRAK and stress-activated protein kinases (SAP) (Wesche et al. 1997).

We also identified a homozygous recessive 3-base disruptive in-frame insertion in Forkhead box protein C1, FOXC1. FOXC1 has been shown to regulate FOX01 through binding of conserved element in its promoter region, which suggests a common pathway with FOX01. FOX01 transcription factor is a major PI3K-AKT downstream effector

molecule. The P13K-AKT pathway is important in translation, metabolism, cell growth, proliferation, and autophagy. FOXO1 has been shown to play a critical role in cell differentiation, including cell cycle, apoptosis, oxidative stress response and DNA damage repair. FOXO1 deficiency impairs B-cell development (reduced expression of EBF1, IL7R, RAG1, RAG2, AID, L-selecting, BLNK) (Szydłowski et al. 2014). The functional consequence of the amino acid insertion in the patient's FOXC1 gene remains to be addressed.

*Family 2*

The number of Family 2 participants in our exome analysis included the proband, both parents and an unaffected female sibling. The proband is a Hispanic male with disease onset between the ages of 2 and 6 years. The clinical course of disease for this patient was of long term responder. That is, the patient's clinical features were ameliorated by treatment with anakinra and remain stable.

*Family 2 Pedigree*

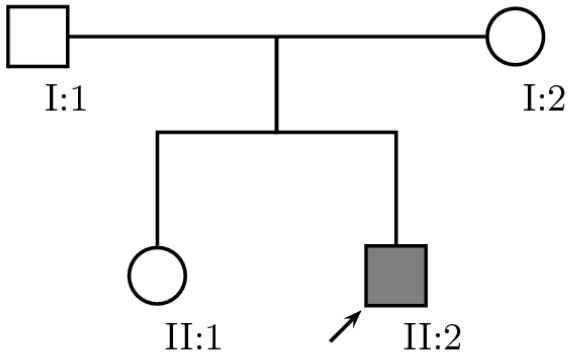


Figure 4.2 Family 2 pedigree. Squares designate males, circles designate females. Shaded square with arrow designate affected proband

Table 4.4 Discrete Filtering Strategy for Family 2

| Inheritance Model | Discrete Filtering Strategy | Candidates* |
|---|---|---|
| Autosomal Recessive | | |
| Homozygous | AF less than 1% in 1KG/EVS | 25 |
| Compound heterozygous | AF less than 1% in 1KG/EVS | 76 |
| Autosomal Dominant | | |
| De novo | AF less than 1% in 1KG/EVS, Not in dbSNP or Flagged in dbSNP | 42 |
| Hemizygous | | |
| Mitochondrial | AF<1% in 1KG/EVS | 6 |
| X-linked | AF<1% in 1KG/EVS | 22 |

**A complete list of these variants is available upon request.*

*Candidate Variants for Family 2 Proband*

A total of 14 variants found in Proband 2 can be classified as strong candidates based on their functional prediction or conservation scores. These include 13 missense variants and one splice-donor variant affecting a total of 13 genes. These genes include METTL21A, EVC, TRPM1, ZDHHC15, ZFHX3, MRC2, BCOR, TTC40, FASTKD2, PLEC, DZANK1, ZNF443, and NBPF1.

The most interesting were three compound heterozygous variants identified in the Zinc finger homebox 3 gene, ZFHX3 (also known as ATBF1). Two of these variants are missense variants, including one with damaging functional prediction score; and the third variant is a disruptive in-frame deletion. ZFHX3 has been associated through GWAS to Kawasaki disease (Burgner et al. 2009), which shares clinical characteristics with sJIA.

A study assessing the function of ZFHX3 in HL-1 cells via knockdown and overexpression experiments reported enhanced activation of STAT3; whereas overexpression had the opposite effect (Jiang et al. 2014). STAT3, signal transducer and activator of transcription 3, mediates cellular responses to interleukins, including IL6, KITLG/SCF and other growth factors. These data suggest potential interactions between ZFHX3 and that of the IL-6 signaling pathway. IL-6 is an important serum biomarker associated with sJIA pathogenesis as well as a therapeutic target. ZFHX3 has also been shown to induce PDGFRB, which has been shown to protect cerebellar neurons from oxidative stress (Kim et al.). PDGFRs are catalytic receptors with intracellular tyrosine kinase activity resulting in signaling through the MAPK, PI3K and PKCgamma pathways.

Another strong candidate variant is a homozygous variant with a high GERP score resulting in ALA870ASP amino acid substitution within the C-type mannose receptor 2 precursor (MRC2). MRC2 functions in the internalization of glycosylated ligands from the extracellular space into an endosomal compartment via clathrin-mediated endocytosis. It may regulate protease activity at the cell surface as it has been shown to be involved in cellular uptake, remodeling and degradation of extracellular matrices in cooperation with matrix metallopreteinases. In fact, in a study assessing tissue-remodeling processes, it was demonstrated that genetic ablation of MRC1 and MRC2 impaired an intracellular collagen degradation pathway observed in M2-like macrophages, Col1a1-expressing fibroblasts and CX3CR1-expressing macrophages (Madsen et al. 2012). Moreover, MRC2 has been shown to be a major player in determining the balance between collagen deposition and degradation resulting in fibrosis

protection. In the cited study, it was reported that MRC2 deficient mice exhibited larger collagen deposits after induction of liver fibrosis compared to WT mice (Madsen et al. 2012). This suggests that deficiency of MRC2 may also contribute to the pathogenesis of sJIA and/or its fibrotic complications, such as pulmonary fibrosis.

Variants for which no conservation scores or functional prediction information were available were found in potentially interesting genes, including TBL1Y, BMP8B, LILRB3, APOBEC3B, PRG4, KMT2C, EP400. TBL1Y (transducing (beta)-like1, Y-linked) is involved in the recruitment of ubiquitin/19S protesome complex to the nuclear receptor-regulated transcription units leading to the subsequent proteosomal degradation of transcription repressor complexes. BMP8B, bone morphogenic protein 8b, induces cartilage and bone formation and also participates in the TGF-beta signaling pathway, among others.

Of note, in addition to sJIA, this patient has history of osteopenia thought to be due to vitamin D deficiency. He has been followed in the metabolic clinic and treated with vitamin D3 supplement and diet modifications. Although 25-OH vitamin D levels have normalized, he continues to have osteopenia on bone density scan. ZFXH3 interacts with the IL-6 signaling pathway. IL-6 is known to induce RANKL expression on surface of osteoblasts. RANKL then interacts with RANK on osteoclast progenitors, which induces osteoclast differentiation. This is significant because osteoclasts function in bone resorption, which could contribute to this patient's osteopenia. Additionally, deficiency in BMP8 (which is involved in bone formation) could also contribute to osteopenia.

*Family 3*

The number of Family 3 participants in our exome analysis included the proband and both parents. The proband is a Caucasian female with disease onset between the ages of 2 and 6. This patient responded "partially" to anakinra treatment. This patient's clinical course has been complicated by recurrent Macrophage Activation Syndrome. She is now well-controlled on increased doses of Anakinra and Methotrexate.

*Candidate Variants for Family 3 Proband*

A total of 9 variants found in Proband 3 can be classified as strong candidates based on their functional prediction or conservation scores and available gene level annotation. These include 7 missense variants, 1 frameshift, and 1 stop-gained variant individually affecting a total of 8 genes. This gene list includes PDPR, DGKK, SH2D5, RFNG, ITGAD, CROCC, and MXRA5.

*Family 3 Pedigree*

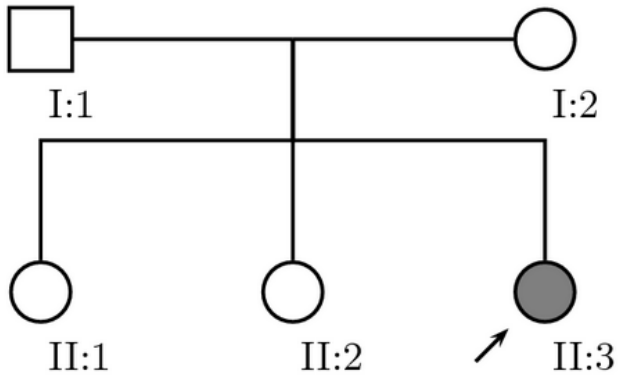


Figure 4.3 Family 3 pedigree. Squares designate males, circles designate females. Shaded circle with arrow designate affected proband

An interesting finding in this patient is compound heterozygosity in the integrin, alpha-D (ITGAD). One allele results in a stop-gained at position 246 of the amino acid sequence and the second allele is a missense variant with a VAL/ALA amino acid substitution at position 986. This gene is also known as CD11D and is a known receptor for ICAM3 and VCAM1. This protein is expressed on myelomonocytic cell lines and subsets of peripheral blood leukocytes. It is also highly expressed in specialized macrophages found in aortic fatty streaks that may develop into atherosclerotic lesions (Van der Vieren et al. 1995). Of relevance, ITGAD has been reported constitutively expressed by synovial macrophages and macrophage-like lining cells.

Table 4.5 Discrete filtering strategy for Family 3

| Inheritance Model | Discrete Filtering Strategy | Candidates* |
|---|---|---|
| Autosomal Recessive | | |
| Homozygous | AF less than 1% in 1KG/EVS | 8 |
| Compound heterozygous | AF less than 1% in 1KG/EVS | 51 |
| Autosomal Dominant | | |
| De novo | AF less than 1% in 1KG/EVS, Not in dbSNP or Flagged in dbSNP | 37 |
| Hemizygous | | |
| Mitochondrial | AF<1% in 1KG/EVS | 2 |
| X-linked | AF<1% in 1KG/EVS | 16 |

**A complete list of variants is available upon request.*

Particularly, a more intense coexpression of this protein and its ligand, ICAM-3, has been reported in the inflammatory infiltrates in rheumatoid synovitis (el-Gabalawy et al. 1996). It has also been shown that the cellular interaction between neutrophils and NK cells is mediated by ICAM3 and CD11d, respectively; and that dendritic cells potentiate the production of IFN gamma by NK cells via CD11a/CD11d (Costantini et al. 2011). Taken together, these data suggest that CD11d (ITGAD) could be potentially implicated not only in the arthritis phenotype but also the NK cell dysfunction observed in sJIA patients. Indeed, NK cell dysfunction has been associated with macrophage activation syndrome in these patients (Grom et al. 2003); this is highly relevant because the clinical course of this patient has been complicated with recurrent MAS.

Another strong candidate is the SH2 domain containing 5 (SH2D5) gene. There are two missense variants contributing to compound heterozygous state. One variant results in a LEU163SER amino acid substitution and the other variant sits at the 5'UTR resulting in a premature start codon gain. Very recently, SH2D5 has been described as a signaling protein responsible for regulating the levels of Rac1-GTP (Gray et al. 2014). Rac1 is a major player of the Rho family of small GTPases that controls different molecular pathways including organization of cytoskeleton (adhesion and motility), cell proliferation, apoptosis and activation of immune cells (D'Ambrosi et al. 2014). Interestingly, compound heterozygous variants were also found in a second gene, rootletin (CROCC), which is also associated with cytoskeleton organization. These observations are relevant as they reproduce the observation recently published from an exome study of a cohort of 13 sJIA patients (Kaufman et al. 2014) where they identified

variants in genes associated with cytoskeleton organization segregated in sJIA with MAS, a prominent feature in this patient.

Other strong candidate variants were found in the pyruvate dehydrogenase phosphatase regulatory subunit, mitochondrial precursor gene (PDRP) and the diacylglycerol kinase kappa gene (DGKK). The PDPR gene contains missense compound heterozygous variants and the DGKK gene contains an X-linked frameshift variant; both of these genes function in metabolic pathways; pyruvate and lipid metabolism, respectively. These suggest potential mitochondrial dysfunction.

Other potentially interesting candidate variants selected for which no functional prediction score or conservation scores were available include ITGA7, VOPP1, HNF1B, RFNG, MADCAM1, NOTCH2, CLEC18, DYRK1B, ALMS1, CACNA1F, IL1RAPL2, and TEX13.

The observed variant in NOTCH2 is potentially de novo in this proband. Notch2 is a very important membrane receptor for B-cell function. Mutations in Notch2 enhance cell proliferation through activation of the NF-kB signaling pathway in lymphoma (X. Zhang et al. 2014). This suggests that NF-kB could be constitutively expressed when Notch2 is under expressed and mutations diminishing or ablating the activity of Notch2 could explain the enhanced inflammation characteristic of sJIA patients. Two other missense variants were identified in a compound heterozygous state in the beta-1,3-N-acetylglucosamintransferase radical fringe precursor (RFNG) gene. A study assessing the expression level of three different Fringe homologs in naïve CD4+ T cells, demonstrated that these molecules are essential in modulating naïve CD4+ T cell differentiation, also through Notch signaling (Gu et al. 2012). Specifically, they reported

that Rfng was overexpressed in naïve CD4+ T cells in asthmatic rats compared to the other O-fucose glycans (Lfng and Mfng). When Rfng was underexpressed via siRNA treatment, they observed enhanced Th1 and diminished Th2 subset lineages. It is possible that deficiency in both Notch2 and RFNG could have a composite heterozygous effect in this sJIA patient resulting in a potentially inflammatory and/or MAS predisposing phenotype.

Additional variants were identified in relevant genes involved in cell adhesion, MADCAM1 and CLEC18. Finally, an x-linked frameshift variant was identified in this patient in the IL1RAPL1 and a de novo variant in VOPP1, both associated with the IL-1 signaling pathway already described above.

*Family 4*

The number of Family 4 participants in our exome analysis included the proband, and both parents. The proband is a Caucasian female with disease onset after the age of 6 years. The clinical course of disease was characterized by a lack of response to IL1 blockade.

*Family 4 Pedigree*

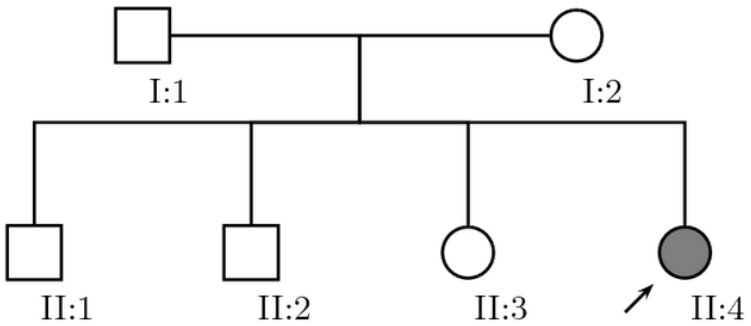


Figure 4.4 Family 4 pedigree. Squares designate males, circles designate females. Shaded circle with arrow designate affected proband

*Candidate Variants for Family 4 Proband*

In proband 4 we found a total of 7 strong candidate variants based on functional prediction or conservation scores. These are all component of compound heterozygous missense variants in a total of 4 genes. These genes are PDE4DIP, NEB, MYH7B, and IK.

Table 4.6 Discrete filtering strategy for Family 4

| Inheritance Model | Discrete Filtering Strategy | Candidates* |
|---|---|---|
| Autosomal Recessive | | |
| Homozygous | AF<1% in 1KG/EVS | 5 |
| Compound heterozygous | AF<1% in 1KG/EVS | 44 |
| Autosomal Dominant | | |
| De novo | AF<1% in 1KG/EVS; Not in dbSNP and/or Flagged in dbSNP | 46 |
| Hemizygous | | |
| Mitochondrial | AF<1% in 1KG/EVS | 4 |
| X-linked | AF<1% in 1KG/EVS | 0 |

**A complete list of variants is available upon request*

One of the compound heterozygous variants in the IK cytokine, down-regulator of HLA-II (IK) gene, results in Ile322Thr amino acid substitution and the other results in a Lys310Arg substitution. IK has been shown to inhibit both interferon gamma-induced and constitutive HLA-DR (HLA-Class II) expression (Krief et al. 1994). Further, IK has

also been shown to down regulate MHC class II expression on B cells via induction of cAMP during Coxsackievirus B (CVB3) infection (Park et al. 2013).

Major histocompatibility (MHC) class II proteins specialize in the presentation of exogenous antigens to the T cell receptor (TCR) of CD4+ T cells. Therefore, disruptions of this process could result in defects in the immune response; either impaired activation of T cells by antigen presenting cells or potentially exaggerated responses due to lack of regulatory T cell function. A down regulation of genes associated with antigen presentation has been reported in sJIA (Nirmala et al. 2014). Another interesting feature of IK is its implication in CD34+ hematopoietic progenitor cell proliferation and differentiation via its modulating effect on HLA-DR expression (Cao et al. 1997). In addition, IK has been shown to interact with MAD1 resulting in its kinetochore localization, mitotic progression, and activation of the spindle assembly checkpoint (Yeh et al. 2012). These data suggest that defects in IK could result in lymphocyte-mediated antigen-specific immune response defects and potentially cell differentiation defects; both of which are phenotypes reported in sJIA (Frosch & Roth 2008).

Two other genes containing strong candidate variants are the myosin heavy chain 7B, Cardiac Muscle, Beta (MYH7B) and nebulin (NEB). The variants identified in both of these genes are missense compound heterozygous variants. Three variants were identified in MYH7B and a total of four missense variants were identified in NEB. Functional annotation for MYH7B is lacking; however, digenic inheritance of variants in this gene together with ITGA7 has been reported to result in congenital myopathy and cardiac myopathy (Esposito et al. 2013). Further, NEB has been described as a giant protein component of the cytoskeletal matrix that interacts with the sarcomeres of skeletal

muscle. Autosomal recessive mutations in this gene are associated with various myopathies, including childhood-onset myopathies. It is peculiar to find these potentially pathogenic variants in a sJIA patient, as pericarditis, but not myopathies or cardiomyopathies, is one of the characteristic presentations of the disease. Interestingly, mutations in NEB were also reported to segregate in sJIA with MAS by Kaufman et al. (Kaufman et al. 2014).

Other interesting variants identified for which functional prediction or conservations scores were not available include FLG, ERMP1, APOBR, CD177, ATP6, SLC22A3, and UNC79. Thus, we identified a total of 5 missense compound heterozygous variants in the flaggrin (FLG) gene. This is an intermediate filament-associated protein that aggregates keratin intermediates in mammalian epidermis. Mutations in this gene are associated with dominant ichthyosis vulgaris and atopic dermatitis. In addition, mutations in this gene were reported to segregate to sJIA with MAS in the exome study reported by Kaufman et al.

Apolipoprotein B receptor (APOBR) and CD177 are cell surface receptors in macrophages and neutrophils, respectively. Both of these cell populations are highly relevant to the pathogenesis of sJIA. We identified a total of 3 compound heterozygous missense mutations in APOBR and a single missense homozygous variant in CD177. APOBR is associated with dietary lipid absorption and metabolism; CD177 is associated with adhesion and neutrophil interactions with platelets and endothelium, both of which are important in sJIA pathogenesis.

*Family 5*

The number of Family 5 participants in our exome analysis included the proband, and both parents. The proband is a Caucasian female with disease onset before the age of 2 years. The clinical course of disease for this patient was difficult to treat with no response to IL1 blockade.

*Candidate Variants for Family 5 Proband*

In proband 5 we found a total of 10 strong candidate variants based on functional prediction or conservation scores. These are all component of compound heterozygous missense variants identified in a total of 6 genes. These genes are IL-1A, NLRC5, NUMA1, RECQL4, and SCN7A.

*Family 5 Pedigree*

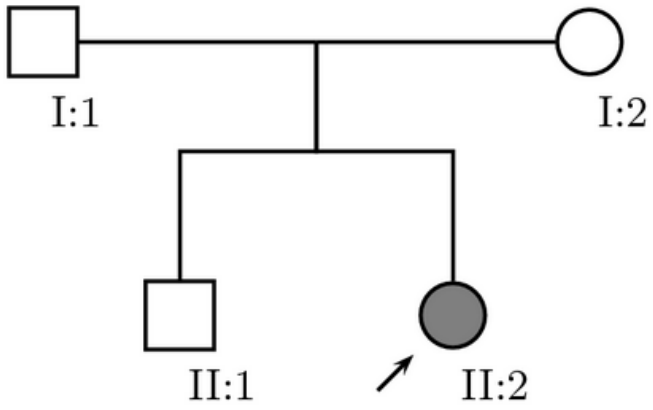


Figure 4.5 Family 5 Pedigree. Squares designate males, circles designate females. Shaded circle with arrow designate affected proband.

Table 4.7 Discrete filtering strategy for Family 5

| Inheritance Model | Discrete Filtering Strategy | Candidates* |
|---|---|---|
| Autosomal Recessive | | |
| Homozygous | AF less than 1% in 1KG/EVS | 4 |
| Compound heterozygous | AF less than 1% in 1KG/EVS | 80 |
| Autosomal Dominant | | |
| Potentially de novo | AF less than 1% in 1KG/EVS, Not in dbSNP or Flagged in dbSNP | 36 |
| Hemizygous | | |
| Mitochondrial | AF<1% in 1KG/EVS | 7 |
| X-linked | AF<1% in 1KG/EVS | 0 |

**Complete variant list available upon request*

The compound heterozygous missense variants in Interleukin 1A (IL1A) consist of ARG85GLN and GLY46ALA amino acid substitutions. The IL-1 family is a major proinflammatory family of cytokines that act mainly through their mediators and via expression of integrins on leukocytes and endothelial cells. Along with IL-1B, IL1-A is one of the two major agonistic members of this family. IL-1A, unlike IL-1B, is constitutively expressed and is active in its precursor as well as in a calpain-processed mature form. It acts as a transcription regulator inside the nucleus, whereas it affects inflammation and immunity outside the cell. IL-1A expressed by neutrophils early in the immune response process is followed by overexpression of IL-1B in later stages of the inflammatory response. Further, IL-1A has been shown to induce neutrophil infiltration

whereas IL-1B recruits macrophages (Rider et al. 2011). These data suggest that alterations in IL-1A function could result in defective regulatory mechanism within the nucleus and potentially proinflammatory effector function if the mutations result in increased protein stability or gain of function through increased affinity to the IL1 receptor. These hypotheses remain to be investigated.

The variants identified in the NLR Family, CARD domain containing 5 (NLRC5) are missense compound heterozygous variants resulting in ALA155THR and GLN1696PRO amino acid substitutions. The nucleotide-binding domain leucine-rich repeat-containing proteins (NLRs) are involved in the intracellular sensing of pathogen-associated molecular (PAM) patterns as well as damage-associated molecular (DAM) patterns. NLRC5 is mainly expressed in hematopoietic cells. NLRC5 has been shown to biochemically associate with NLRP3 in an inhibitory fashion, suggesting NLRC5-dependent inactivation of the inflammasome (Davis et al. 2011). NLRC5 plays a role in cytokine response against intracellular pathogens through inhibition of NF-kB activation and negative regulation of type I interferon signaling pathways. Thus, overexpression of NLRC5 in HECK293T cells resulted in down regulation of NFkB, AP1, and type I interferon-dependent signaling. Conversely, knockdown of NLRC5 expression in murine macrophages led to significant upregulation of proinflammatory responses to IFN-gamma and LPS resulting in increased secretion of TNF, IL-6 and IL-1beta (Benko et al. 2010). These data suggests that NLRC5 defects could lead to hyper immune responses to DAMPs and PAMPs.

Compound heterozygous missense mutations were also found in the nuclear mitotic apparatus protein 1 (NUMA1) resulting in ARG1154TRP and ARG972GLN

amino acid substations. NUMA1 is a component of the nuclear matrix and is required for maintenance and establishment of the mitotic spindle poles as well as in chromatin regulation in the nucleus during interphase. NUMA binds p53 and is required to modulate p53-mediated transcription (Ohata et al. 2013). The tumor suppressor gene p53 controls various cellular processes including cell cycle arrest, DNA repair, apoptosis, and cell proliferation – p53 is critical in cellular and chromosomal stability.

Variants identified in the Rec Q protein-like 4 (RECQL4) are also missense compound heterozygous; resulting in SER1154TYR and GLN1145HIS amino acid substitutions. RECQL4 is a DNA helicase and plays a role in DNA repair and maintenance of replication fork. The level of RecQL4 expression has been associated with mtDNA copy number; knockdown expression experiments have revealed decreased number of mtDNA copy number. Further, significantly decreased repair capacity for oxidative DNA damage repair was observed in cells deficient in RecQL4 (Chi et al. 2012). These data suggest that defects in RECQL4 could result in mitochondrial dysfunction.

Other potentially interesting candidate variants in this patient include MAP3K4, NOTCH2, NBPF1, FCGBP, and TPSAB1. A homozygous disruptive in-frame deletion was identified in the mitogen-activated protein kinase kinase kinase 4 (MAP3K4) gene. This protein activates CSBP2, P38 and JNK/MAPK signaling pathways. It has been reported that expression of IL-1A is dependent on the expression of MAP3K4. This interaction has been shown defective in Crohn's disease where chronic inflammation is part of the clinical phenotype (van der Pouw Kraan et al. 2012). This variant, together

with those identified in the IL-1A gene, suggest a potential digenic or multigenic inheritance at the core of disease pathogenesis in this patient.

*Family 6*

The number of Family participants in our exome analysis included the proband, both parents, and unaffected female sib. The proband is a Caucasian female with disease onset between the age of 2 and 6 years. The clinical course of disease for this patient was characterized by a good response to IL1 blockade.

*Family 6 Pedigree*

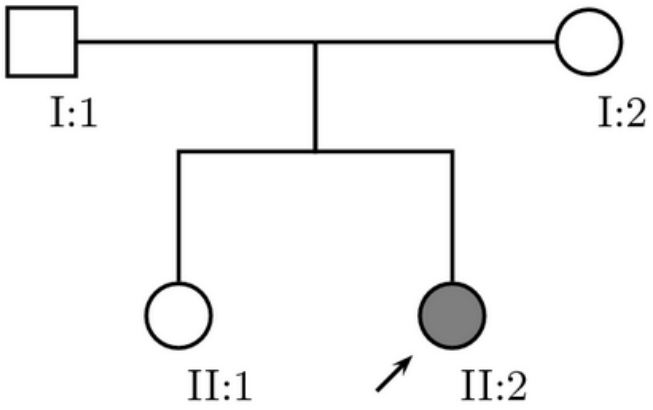


Figure 4.6 Family 6 pedigree. Squares designate males, circles designate females. Shaded circle with arrow designate affected proband

A large number of variants that were not identified in either parent were detected in this patient (potentially de novo variants) compared to the other families sequenced in this study. Two independent sequencing approaches by two different institutions resulted in similar results.

Table 4.8 Discrete filtering strategy for Family 6

| Inheritance Model | Discrete Filtering Strategy | Candidates* |
|---|---|---|
| Autosomal Recessive | | |
| Homozygous | AF less than 1% in 1KG/EVS | 5 |
| Compound heterozygous | AF less than 1% in 1KG/EVS | 14 |
| Autosomal Dominant | | |
| De novo | AF less than 1% in 1KG/EVS, Not in dbSNP or Flagged in dbSNP | 127 |
| Hemizygous | | |
| Mitochondrial | AF<1% in 1KG/EVS | 1 |
| X-linked | AF<1% in 1KG/EVS | 0 |

**A complete list of variants is available upon request.*

*Candidate Variants for Family 6 Proband*

In proband 6 we identified a total of 7 strong candidate variants based on functional prediction or conservation. These include TPSAB1, ATP6, CASP10, TNFRSF10C, CTSW, BAZ2A, and CFHR5.

A total of 5 compound heterozygous missense variants were identified in tryptase alpha/beta 1 (TPSAB1) resulting in, VAL29ALA, VAL205ILE, THR215SER, ARG216GLN, and GLN221LYS. Tryptase is the major neutral protease present in mast cells and is secreted after activation of the degranulation response. Mast cells are derived from hematopoietic progenitor cells. They play an important role in the host defense

against pathogens. It has been shown that variation of a single amino acid at the substrate binding cleft of a given tryptase can result in significant consequences in the regulation of its enzymatic activity and/or specificity for a given substrate. Asp215 dominantly restricts substrate specificity of tryptase alpha (Huang et al. 1999). Thus, the number of variants found in this gene suggests a potentially significant effect on its enzymatic activity and specificity.

Another strong potentially de novo candidate variant is a 5'UTR variant resulting in a premature start codon for caspase 10 (CASP10). CASP10 is involved in the activation of caspases responsible for apoptosis. It is recruited to both Fas- and TNFR-1 receptors in a FADD-dependent manner. It has been shown to cleave and activate CASP-3, -4, -6, -7, -8 and -9. CASP10 dimerizes as a pro-caspase and then is cleaved into active protease as a heterotretamer. It is possible that the biochemical property changes introduced with the early start codon could destabilize the binding and subsequent activation of CASP10 and its ability to activate other caspases. This finding is significant because monocytes have been reported to be resistant to apoptosis in sJIA (Srivastava et al. 2010). Interestingly, another potentially de novo frameshift variant was observed in another apoptosis related gene, tumor necrosis factor receptor superfamily, member 10c (TNFRSF10C) also known as TRAIL-R3. The variant results in a deletion of 1 nucleotide in exon 3/5. This gene serves as a decoy receptor for the cytotoxic ligand TRAIL; it lacks a cytoplasmic death domain and hence is not capable of inducing apoptosis. It is thought to protect cells against TRAIL mediated apoptosis by competing with the other R1 and R2 receptor ligands. The single nucleotide deletion changes the

amino acid sequence of the second half of the protein and therefore potentially changes its protection against TRAIL mediated -apoptosis.

Additional de novo frameshift variants were identified in cathepsin W (CTSW), and complement factor H-related 5 (CFHR5). CTSW has been shown to be expressed exclusively in CD8+ T cells and NK cells and is secreted during target cell killing; however, it is not clear whether it is required in the process of cytotoxicity (Stoeckle C et al. Exp Hematol 2009) (Wex T et al. FEBS Lett 2003). NK cell function is altered in sJIA and might play a role in the predisposition of these patients towards MAS. As for CFHR5, it forms a complex with C3b and is recruited to tissues damaged by C-reactive protein. Its immunological functions make it a potentially interesting candidate.

*Variants Affecting the Same Gene in Multiple sJIA Patients*

A list of strong or potentially interesting candidate variants was compiled from those identified in the trio analyses described above. The resulting 79 genes are listed in Table 4.9. Although some of these mutations were also found in the same variants in the non-affected controls, only the patients were compound heterozygous for these genes.

In an effort to segregate genes that may be relevant in sJIA, patients carrying variants in any of the listed genes were identified (Table 4.10). Searching within the 6 families only, 6 genes were identified as displaying variants affecting more than one patient. This alone was significant as it represents ~33% of the trio cohort.

An expanded search within the entire cohort of 19 sJIA patients revealed a total of 6 genes containing rare missense or loss of function variants in more than one patient. The variants found in non-affected individuals were in a heterozygous state; whereas in

the patients these were compound heterozygous or homozygous. The genes and list of patients is depicted in Table 4.10.

Table 4.9 List of Potentially Interesting Candidate Genes in sJIA

| Gene | Gene | Gene | Gene | Gene |
|---|---|---|---|---|
| A2M | CFHR5 | HNF1B | MXRA5 | SH2D5 |
| ACAD11 | CLEC18 | IGFN1 | MYH7B | SLC22A3 |
| ALMS1 | CROCC | IK | NBPF1 | TBL1Y |
| APOBEC3B | CTSW | IL1A | NEB | TPSAB1 |
| APOBR | DGKK | IL1RAPL2 | NLRC5 | TRPM1 |
| ARHGEF10 | DYRK1B | ITGA7 | NOTCH2 | TRRAP |
| ASPSCR1 | DZANK1 | ITGAD | NUMA1 | TTC40 |
| ATP6 | EP400 | KMT2C | PDE4DIP | UNC79 |
| ATXN2 | ERMP1 | KRT2 | PDPR | VOPP1 |
| BAZ2A | EVC | LILRB3 | PLEC | VSP13C |
| BCOR | FASTKD2 | LOXHD1 | PRG4 | ZDHHC15 |
| BMP8B | FCGBP | MADCAM1 | RBMXL1 | ZFHX3 |
| CACNA1F | FCGBP2 | MAP3K4 | RECQL4 | ZNF443 |
| CASP10 | FLG | METTL21A | RFNG | |
| CCBL2 | FOXC1 | MRC2 | SCN7A | |
| CD177 | HERC2 | MST1L | SDK1 | |

Table 4.10 Genes affected in more than 1 proband

| Gene | P1 | P2 | P3 | P4 | P5 | P6 |
|---|---|---|---|---|---|---|
| IL1RAP2 | √ | | √ | | | |
| ITGA7 | √ | | √ | | | |
| NBPF1 | | √ | | | √ | |
| NOTCH2 | | | √ | | √ | |
| ATP6 | | | | √ | | √ |
| TPSAB1 | | | | | √ | √ |

Table 4.11 Genes with Missense or Loss-of-Function Mutations in More Than One sJIA Patient

| Patient ID | FOXC1 | IL-1A | LOXHD1 | NLRC5 | TNFRSF10C | ZFHX3 |
|---|---|---|---|---|---|---|
| SYS_002 | √ | | | | | |
| SYS_012 | √ | | | | | √ |
| SYS_015 | | √ | | | | √ |
| SYS_021 | √ | | | | | |
| SYS_027 | | √ | √ | √ | | √ |
| SYS_051 | | | | | | |
| SYS_071 | | | √ | | | |
| SYS_076 | √ | | √ | | | |
| SYS_106 | √ | | √ | | | √ |
| SYS_118 | | | | √ | √ | |
| SYS_129 | | | | | | √ |
| SYS_130 | | | | | | √ |
| SYS_143 | | | | | | |
| SYS_157 | | | | | √ | √ |
| SYS_172 | | √ | | | | |
| SYS_174 | | | | | | |
| SYS_197 | | √ | | √ | | |
| SYS_200 | | | | | | √ |
| SYS_205 | | | | | √ | |

ZFHX3 (ATBF1) transcription factor was described in the Family 2 section. A total of 8 out of 19 sJIA patients have missense or loss of function mutations in this gene compared to 5/23 non-affected individuals (patient relatives). The protein product from this gene is 300kD protein that binds AT-motif elements in promoter regions. It is possible that the large size of this protein justifies that more mutations could occur by chance compared to smaller genes. As previously mentioned however, this gene has a GWAS association to Kawasaki disease. Indeed, our gene expression studies have demonstrated some definite similarities in blood gene signatures between sJIA and

Kawasaki patients. FOXC1 – was described in the results section for Family 1. A total of 5 patients out of 19 are affected compared to 4/23 non-affected individuals sequenced as part of this study which includes related individuals.

IL-1A – a total of 4 out of 19 sJIA patients have missense or loss-of-function mutations in this gene compared to 2/23 non-affected related individuals sequenced as part of this study. As mentioned, IL-1 signaling is relevant in the pathogenesis of sJIA as established in published works and by the fact that biologics targeting the IL-1 pathway make an improvement in up to 80% of patients. One interesting observation is that 3 out of 5 sJIA patients in our cohort who are classified as non-responders to anakinra treatment have mutations in this gene, along with a fourth patient classified as a partial responder. Anakinra is the recombinant form of IL-1Ra and is an IL-1R1 agonist, thereby blocking IL-1A and IL-1B signaling. IL-1A has dual functionality, extracellular in response to injury and intracellular in the nucleus as a transcription regulator. IL-1A is constitutively expressed in epithelial cells and serves a role in maintenance of the skin barrier. Production of IL-1A can be induced in most immune cells types, including macrophages, monocytes, T-cells, B-cells, and NK cells. IL-1A work synergistically with TNF and act to promote fever and inflammation. The functional implication of the observed mutations in these patients needs further investigation.

NLRC5 – a total of 3 sJIA patients out of 19 patients have missense or loss-of-function mutations in this gene compared to 2/23 non-affected related individuals sequenced as part of this study.

Tumor necrosis factor receptor superfamily, member 10C (TNFRSF10C) is a known receptor for the cytotoxic ligand TRAIL. It is a decoy receptor that lacks a

cytoplasmic death domain and hence is not capable of inducing apoptosis. Variants were identified in only two patients but not in other non-affected control samples.

Lipoxygenase homology domains 1 (LOXHD1) is a highly conserved protein consisting of polycystin/lipoxygenase/alpha-toxin domains (PLAT); its exact function is unknown, but it is postulated to target proteins to the plasma membrane. Mutations in this gene are associated with hearing loss and corneal dystrophy, Fuchs dystrophy. Fuchs' endothelial corneal dystrophy affects ~5% of the USA population over age of 40 and is major cause for corneal transplantation. The pathophysiology of Fuchs dystrophy involves deficiency of corneal endothelium pump function and induction of the unfolded protein response (Jalimarada et al. 2014). The unfolded protein response (UPR) is the intracellular transduction signaling that results from the imbalance (ER stress) between the amount of unfolded proteins entering the endoplasmic reticulum and the capacity of cellular machinery that handles the load. In this response, signaling aims to restore normal function by inhibiting protein synthesis, increasing degradation of misfolded proteins and enhancing synthesis of chaperons. If the balance is not restored, the UPR leads to apoptosis. Stress responses are important in the pathogenesis of sJIA as they lead to apoptosis and activation of immune response pathways. Additionally, unfolded protein response is a pathway that was reported enriched in sJIA with MAS by Kaufman et al. (Kaufman et al. 2014) .

*Comparison of Results to Published sJIA Exome Data*

A recent study published by Kaufman et al. (Kaufman et al. 2014) investigated the genetic etiology of MAS in sJIA patients. They did WES in 14 affected individuals and their parents and in 29 sJIA patients with no history of MAS. They published a total

of 111 genes that were enriched in sJIA and we used this list to compare with our results. Figure 4.7 depicts the overlap identified.

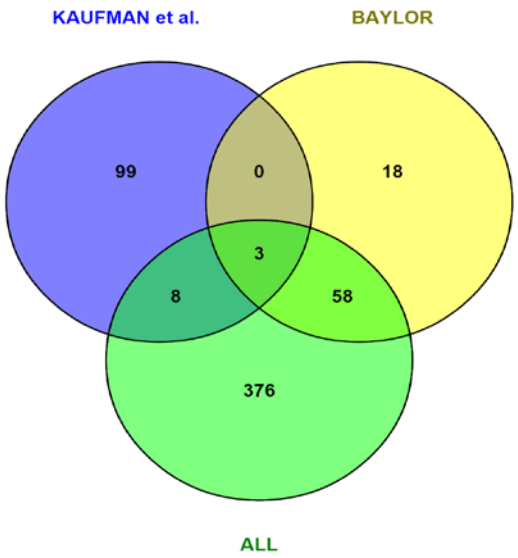


Figure 4.7 Overlap of candidate genes with sJIA exome published data. Venn diagram showing overlap of published sJIA exome data (Kaufman et al.), 79 strong candidates selected from family trio analysis, and all candidates from the discrete filtering process for each trio analysis. List of genes on left side pane includes the eight genes overlapping between Kaufman et al. all and all genes mutated in our trio dataset. List of genes on right side pane includes the three genes overlapping Kauffman and strong candidate genes in our trio dataset.

*Pathway Analysis Results*

Ingenuity pathway analysis (IPA) software (Qiagen) was used to identify canonical pathways enriched using the 79 gene candidates identified as potentially interesting. IPA is a tool that aids in the identification of pathways and/or molecular functions which are likely to explain an observed phenotype.

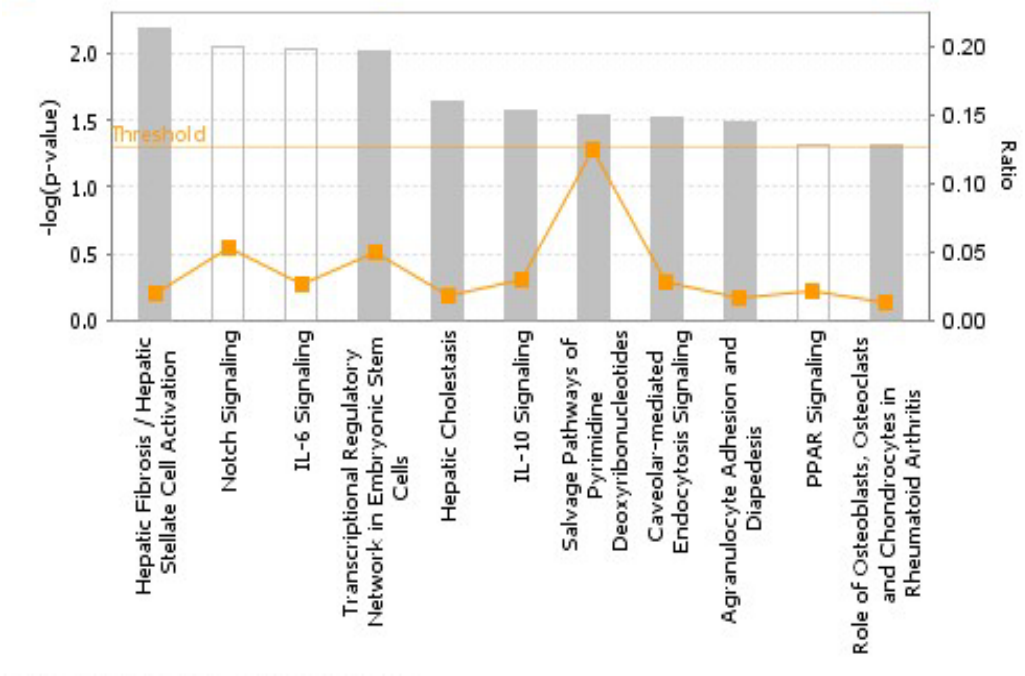


Figure 4.8 Canonical pathways enriched using the identified sJIA candidate gene Canonical pathways derived from 79 genes using IPA software. The statistical significance as calculated by Fisher's Exact Test is depicted in left Y-axis and threshold line defines p=0.05. The proportion of genes identified from total known in pathway is depicted in the right Y-axis (Ratio) and connected orange squares and represents strength of association.

Among the eleven canonical pathways derived from our gene list are Notch Signaling, IL-6 Signaling, and IL-10 Signaling (Figure 4.8). As already mentioned, IL-6 is implicated in the pathogenesis of sJIA (Woo & Humphries 2013). IL-10 is a cytokine with both immune-regulatory and anti-inflammatory functions. It can suppress proinflammatory cytokines including IL1, TNFa, and IL-6. IL-10 has also been identified as a potential contributor to sJIA pathogenesis through a GWAS study (Fife et al. 2006). It is possible that deficiency of this pathway can lead to the overexpression of IL-1 and IL-6 reported in sJIA patients in addition to over reactive immune responses. Notch signaling has not been directly implicated in sJIA pathogenesis; however, Notch

supports the development of blood cells including cytokine producing NK cells (Suresh & Irvine 2015) that may be critical in MAS (Avau et al. 2014; Grom et al. 2003).

A total of three gene networks were also generated in this analysis (Figures 4.9-4.11). The networks interactions are calculated using Ingenuity Knowledge Base, a curated database of published gene information. The first network (Figure 4.9) is enriched with cell signaling molecules including IL-1B, TLR4, TNF, JUN and IFN-gamma. They are highly relevant in immune response mechanism and therefore to sJIA disease pathogenesis.

The second network (Figure 4.10) is associated with biological functions in cell cycle, DNA repair, and cell death and survival. MYC, TP53 and SMAD3 suggest similar gene interactions in networks derived using candidate gene list of sJIA with MAS published by Kaufman et al. (data not shown). Finally, the third network (Figure 4.11) is associated with cellular development and lipid metabolism biological functions. Of importance is NFkB which shows direct and indirect interactions with other major players such as P38MAPK, ERK1/2, Akt, and Histone h3; similarly these molecules are found in the network interactions derived using the Kaufman et al. dataset.

*Remarks*

Although sJIA is a sporadic disease with no known familial cases, here we focused our analysis in textbook approaches to identifying causal variants in inherited diseases and encountered interesting findings. We identified a compound heterozygous model predominant in most of the variants detected in the list of 79 potentially interesting genes. In addition, we saw evidence of digenic and potentially multigenic events in most

of the probands in the family trios; as in, for instance, Proband 5 carrying mutations in IL-1A and MAP3K4 and Proband 3 carrying mutations in Notch and RFNG.

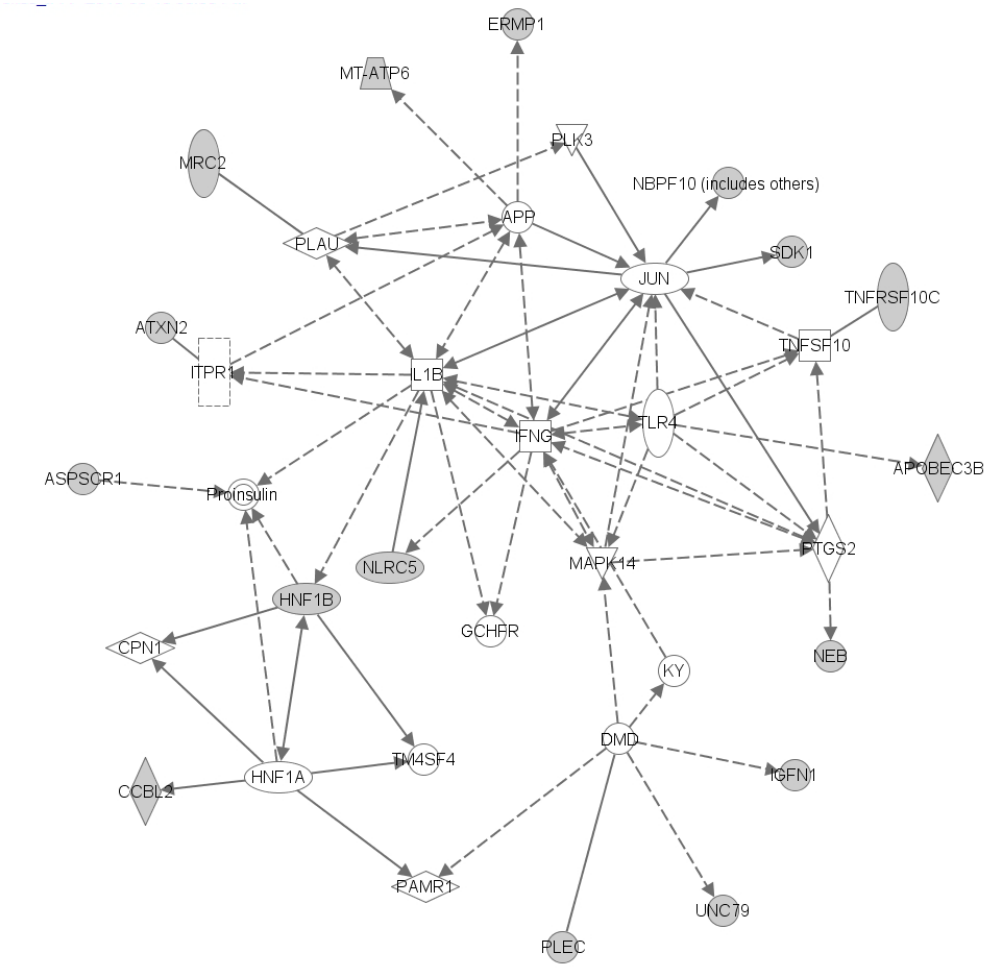


Figure 4.9 Gene Network 1. Gene network depicting direct (solid lines) and indirect (dotted lines) gene interactions. Gray filled geometric figures depict genes with mutations in our dataset.

Our pathway and network analysis results further suggest a multigenic model at the core of sJIA disease pathogenesis. These approaches provided us with interesting genes and pathways to test functionally in the future. These findings also suggest alternate analysis approaches accounting for gene interactions would be very useful. In

addition, new analysis tools facilitating the integration of linkage analysis in family pedigrees are arising and will provide us with the ability to examine groups rather than individual patients.

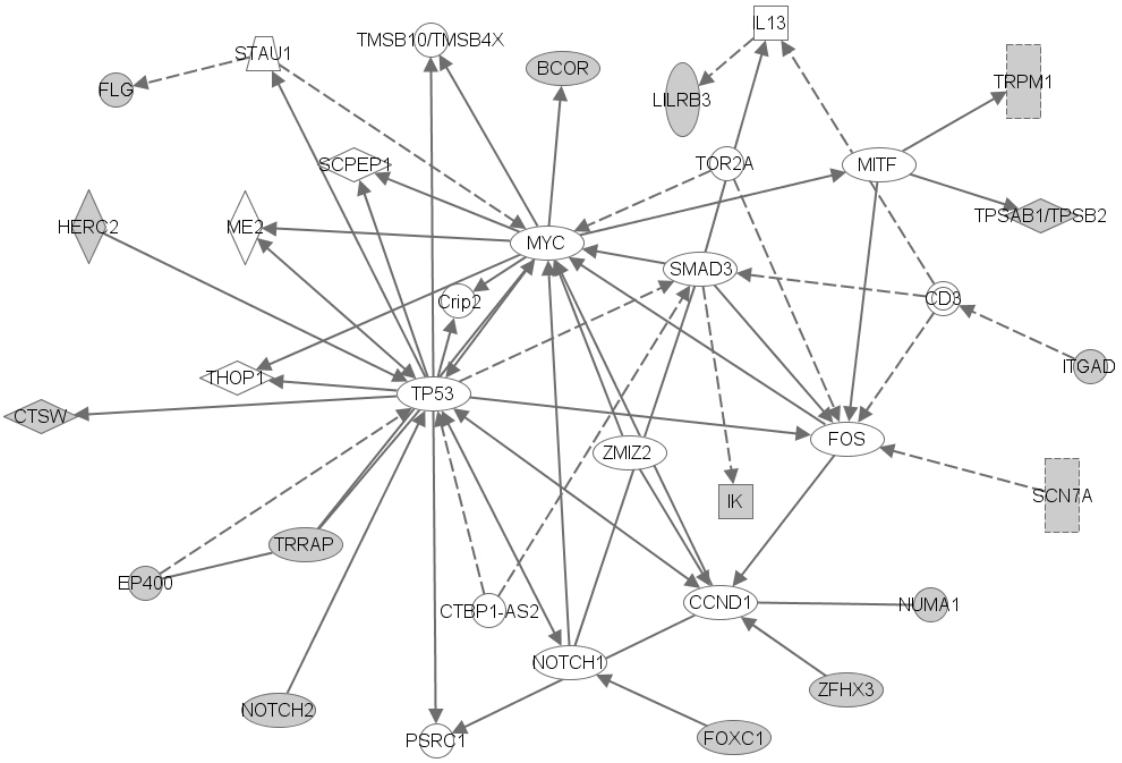


Figure 4.10 Gene Network 2 Gene network depicting direct (solid lines) and indirect (dotted lines) gene interactions. Gray filled geometric figures depict genes with mutations in our dataset.

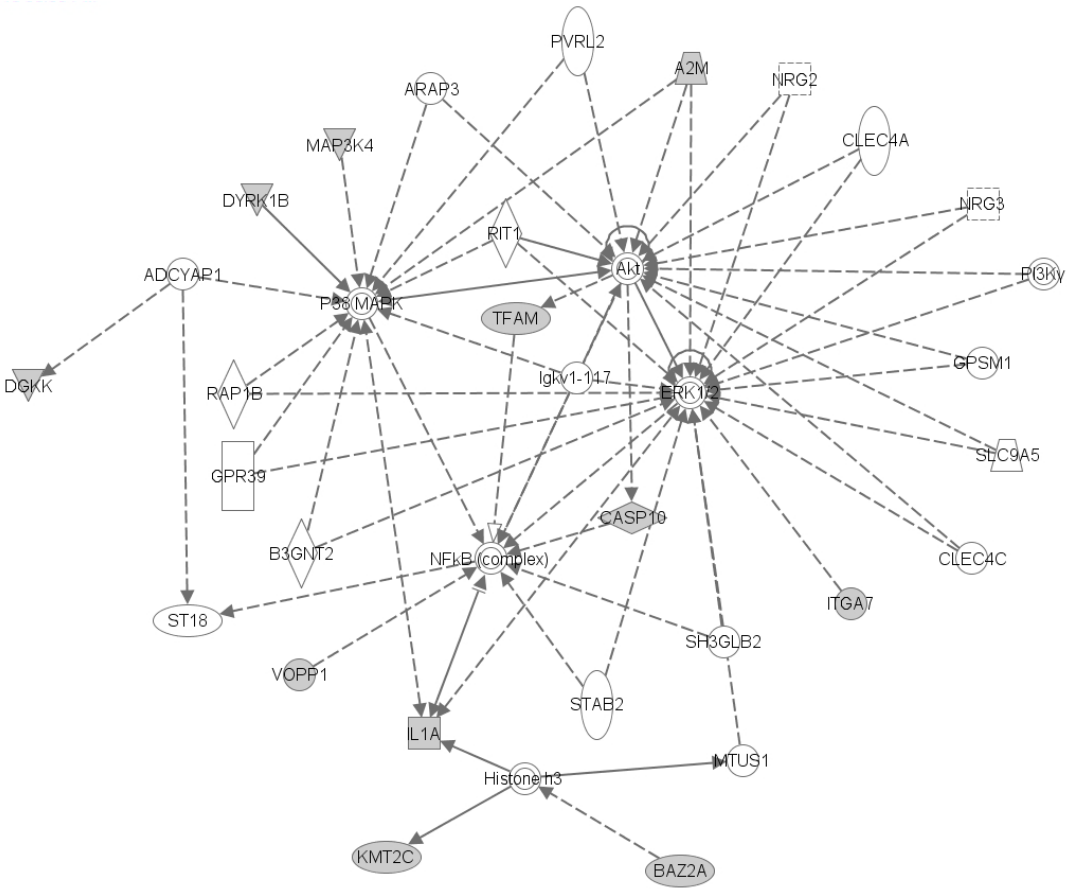


Figure 4.11 Gene Network 3 Gene network depicting direct (solid lines) and indirect (dotted lines) gene interactions. Gray filled geometric figures depict genes with mutations in our dataset.

# CHAPTER FIVE

## Functional Implications for S100A12 Deletion Identified in sJIA Proband 1

### *Background*

In the identification of autosomal dominant alleles, we focused our attention on those that were expressed in a de novo form. The reasoning behind this was that we only collected samples for two generations, i.e. parents and their children. Dominant inheritance results in expression of a phenotype whenever a heterozygous genotype is inherited, under the assumption of complete penetrance. Because there are no known cases of familial sJIA, a dominant inheritance model outside a de novo event is not expected. Therefore, our efforts did not include the assessment of this model. However, in the examination of gene variants associated with sJIA we identified a frameshift variant in the S100A12 gene in Proband 1 for which both the Proband and his Father carried the same mutation. As described in Chapter Two, S100A12 has been associated in the pathogenesis of sJIA. The protein, along with S100A8/A9, is highly expressed in the serum of patients and correlates with disease activity (Frosch & Roth 2008).

### *Introduction to S100 Proteins and Association in sJIA*

S100 proteins are known to participate in different cellular functions including, growth, differentiation and survival through intracellular events, cytoskeletal organization, and extracellular signaling via different cell membrane receptors such as RAGE and integrins.

S100 proteins are low molecular weight acidic peptides ranging from 10-12kDa. There are many members classified as S100A proteins, (S100A-S100A16) and others such as S100B, S100G, S100P, and S100Z. S100A proteins exist as monomers, homo- or heterodimers or multimeric forms within cells and their extracellular matrices (Donato et al. 2013; Gross et al. 2014). Protein sequence homology ranges from 16 to 98%. A feature common to all the S100 proteins is the presence of calcium-binding helix-loop-helix domain (EF-hand), where calcium binding regions are towards either end of the protein and separated by a hinge region (Figure 5.1). The canonical EF-hand motif is composed of 12 amino acids on the C-terminal end has a calcium affinity 10-50 times higher than the N-terminal pseudo EF-hand motif.

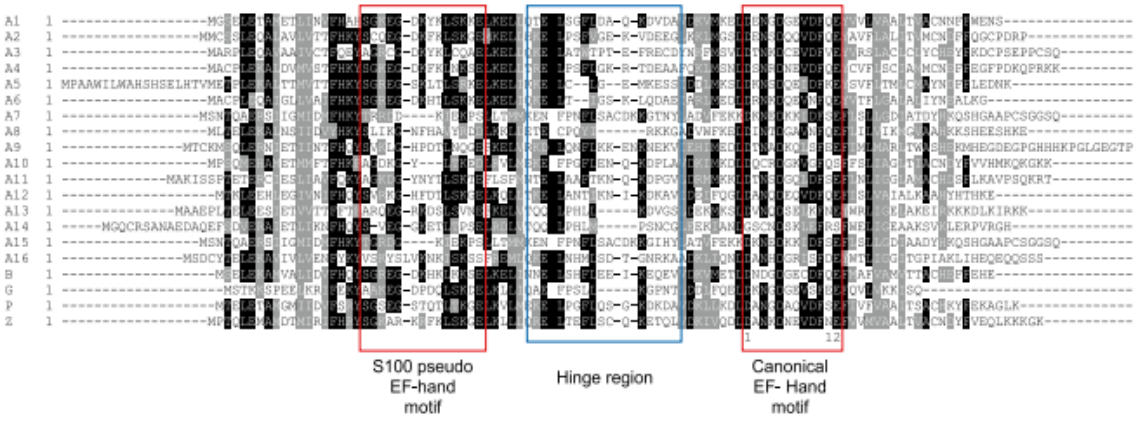


**Fig. 1** S100 protein amino acid sequence alignment. Amino acid sequences of S100 proteins were aligned with the EF-hands and central regions indicated (number 1–12 in the canonical EF-hand motif refers to the position of essential amino acids for the formation of the calcium-binding loop). All sequences are human, and the accession numbers are S100A1, AAH14392.1; S100A2, EAW53305.1; S100A3, EAW53306.1; S100A4, CAG29341.1; S100A5, EAW53317.1; S100A6, EAW53326.1; S100A7, EAW53327.1; S100A8, EAW53330.1; S100A9, EAW53334.1; S100A10, NP002957.1; S100A11, NP005611.1; S100A12, EAW53332.1; S100A13, CAA68188.1; S100A14, AAM19206.1; S100A15, AAO40033.1; S100A16, EAW53304.1; S100B, NP006263.1; S100G, EAW98916.1; S100P, EAW82384.1 and S100Z, EAW95784.1. Sequences were aligned using the multalin sequence comparison program (http://multalin.toulouse.inra.fr/multalin/) and the resulting data *shaded* and presented using the *boxshade* integrated programme (http://www.ch.embnet.org/software/BOX_form.html)

Figure 5.1 S100 Protein Amino Acid Sequence Alignment (Gross et al. Cell Mol Life Sci 2014). With permission from Springer

Two motifs show the highest level of amino acid conservation throughout the S100 proteins (Gross et al. 2014) and the protein undergoes conformational changes when it binds calcium, which in turn exposes a hydrophobic region. All S100 proteins lack a leader sequence required for endoplasmic reticulum entry, and subsequent externalization is therefore carried out independent of the canonical secretory system via endoplasmic reticulum and Golgi complex. Further, the C-terminal sequence beyond the canonical EF-hand motif of each of the S100 proteins determines their binding partners. S100A8/S100A9 and S100A12 have been implicated in the pathogenesis of sJIA. They interact with enzymes, cytoskeletal proteins, receptors, transcription factors, and nucleic acids to regulate a wide variety of pathways including proliferation, differentiation, apoptosis, inflammation, cell migration, energy metabolism and calcium ion homeostasis (Tong et al. 2014; Halawi et al. 2014). S100A12 is present in the human myeloid cell lineage, but its counterpart in not found in mice. These proteins are secreted during neutrophil and monocyte activation. S100A8/A9 (MRP8/MRP14) form a complex that can serve as a TLR4 agonist triggering production of IL-1B (Frosch et al. 2009; Holzinger et al. 2012). S100A12 has also been shown to activate human monocytes via TLR4 (Foell et al. 2013). High levels of these proteins have been associated with active sJIA and may distinguish this disease from other febrile conditions such as systemic infection, leukemia, and Kawasaki disease (Wittkowski et al. 2008)(Ling et al. 2010a). Increased secretion of S100 proteins is a feature shared with other autoinflamamtory diseases such as Familial Mediterranean Fever (FMF) and Cryopyrin Associated Periodic Syndromes (CAPS) (Wittkowski et al. 2008). Moreover, serum concentrations for S100A8/9 are used to monitor response to drug treatment and disease activity. In fact,

measurement of these proteins in serum has been shown to predict disease flares with 92% sensitivity and 88% specificity (Holzinger et al. 2012). Thus, mutations affecting one or more of the biological functions of S100A12 could be linked to sJIA pathogenesis.

## *Characterization of Mutant S100A12*

### *Identification of Frameshift Deletion in S100A12*

Aligned reads to the S100A12 were generated from BAM files creates as described in Chapter Three and visualized using the Integrative Genome Viewer (http://www.broadinstitute.org/igv/) software (Figure 5.2). To validate our findings, Sanger sequencing was conducted as described in Chapter Three.

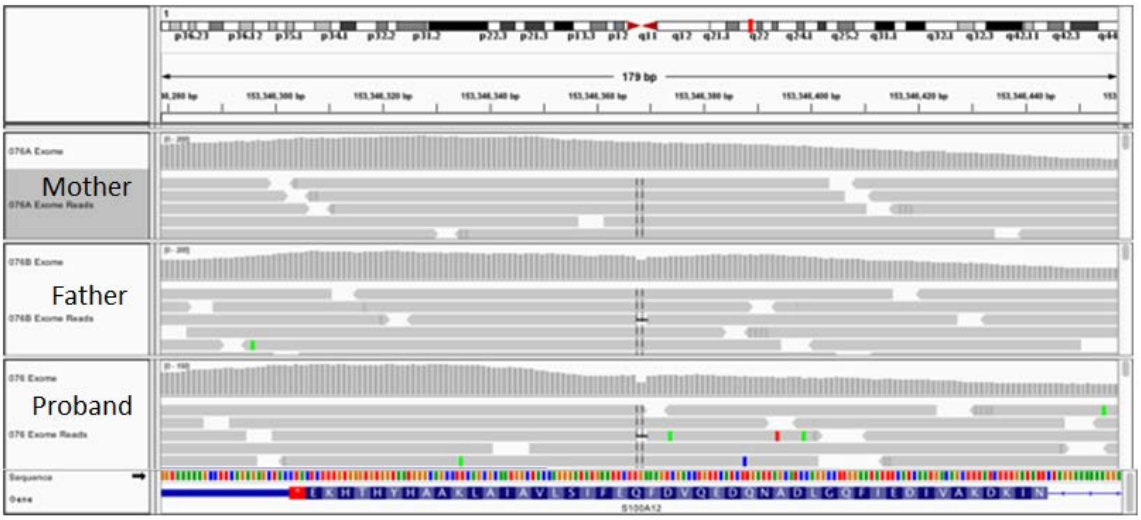


Figure 5.2 Family 1 S100A12 read sequence alignment in IGV Browser Visualization of NGS reads in IGV browser demonstrating the identification of reference reads only in mother (top panel), reads with 2bp deletion as well as reads matching reference in father (middle panel) and proband (bottom panel).

A 2-bp deletion (GA|--) was observed at the chromosomal coordinate Chr1:153346368 resulting in a frameshift that moved stop codon 96 bases downstream.

Resultant mutant cDNA and protein sequences ware generated using DNASTAR software and both the normal and mutant cDNA sequences are provided in Figure 5.3.

>Normal

ATGACAAAACTTGAAGAGCATCTGGAGGGAATTGTCAATATCTTCCACCAAT
ACTCAGTTCGGAAGGGGCATTTTGACACCCTCTCTAAGGGTGAGCTGAAGCA
GCTGCTTACAAAGGAGCTTGCAAACACCATCAAGAATATCAAAGATAAAGCT
GTCATTGATGAAATATTCCAAGGCCTGGATGCTAATCAAGATGAACAGGTCG
ACTTTCAAGAATTCATATCCCTGGTAGCCATTGCGCTGAAGGCTGCCCATTAC
CACACCCACAAAGAGTAG

>Mutant

Atgacaaaacttgaagagcatctggagggaattgtcaatatcttccaccaatactcagttcgg
aaggggcattttgacaccctctctaagggtgagctgaagcagctgcttacaaaggagcttgcaaacaccatcaagaatatcaaa
gataaagctgtcattgatgaaatattccaaggcctggatgctaatcaagatgaacaggtcgacttaagaattcatatccctggtagc
cattgcgctgaaggctgcccattaccacacccacaaagagtaggtagctctctgaaggctttttacccagcaatgtcctcatgagg
gtcttttctttccctcaccaaaacccagccttgcccgtggggagtaagagt

Figure 5.3 Normal and Mutant cDNA sequence for S100A12


Figure 5.4 Amino acid sequence for normal and mutant S100A12 Sequence alignment of amino acids in both normal (top) and mutant (bottom) sequence. The different colors indicate binding motifs and stars indicate amino acids required for optimal calcium binding affinity.

This frameshift variant results in the loss of the last amino acids in the 3 C-terminal domain canonical EF-hand motif and a completely different protein sequence in the C terminal domain that is 32 amino acids longer than its normal counterpart.

*Expression of Mutant mRNA in Whole Blood of Proband 1*

We wondered if the patient would result in a functionally transcribed gene. To that end, we sequenced purified RNA obtained from whole blood of the patient following Illumina TruSeq Stranded mRNA Library Sample Prep Kit. Briefly, mRNA enrichment was carried out using oligodT coated beads followed by chemical fragmentation of mRNA molecules. Subsequently, first strand cDNA was synthesized using random hexamer priming with actinomycin D to retain strandedness; this was followed by removal of RNA template using RNase H and incorporation of dUTP in place of dTTP to generate blunt ended ds cDNA using polymerase. The library constructs were prepared for paired-end sequencing by Illumina HiSeq2500 as described in Chapter Three.

After data processing and alignment, reads were visualized using an IGV browser (Figure 5.5). This permitted us to identify reads spanning the coordinate where the deletion was observed (blue arrow in figure 5.5) as well as reads without the mutation, as expected given allelic heterozygosity. Read count spanning the two transcripts did not appear to be different, suggesting allelic expression was not biased at the transcriptional level.

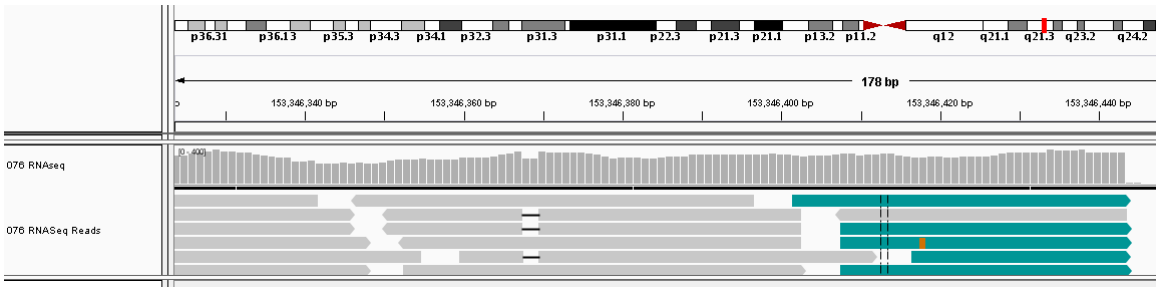


Figure 5.5 Alignment of paired-end mRNAseq reads over chromosomal coordinate containing S100A12 deletion. Visualization of aligned mRNAseq paired-end reads in IGV browser. Blue arrow illustrates the region where reads (gray horizontal bars) with deletion (dark line) can be visualized along with reads aligning to reference.

*Detection of Mutant S100A12 Protein in Serum of Proband 1*

Immunoblot analysis of the patient's serum collected during four different clinic visits revealed the presence of two bands around the 10kDa ladder mark (data not shown), suggesting that the mutant protein is indeed expressed. At the time this specific experiment was conducted, a protein control for normal S100A12 was not available. Recombinant S100A12 has been now purchased and expression experiments for mutant S100A12 protein are underway.

*Genotyping of S100A12 Deletion in Extended Family Members*

Sanger sequencing was carried out on a total of 8 family members and the results of the status of the deletion is represented in the following pedigree (Figure 6.7). Out of 7 family members tested in three generations, two additional individuals (the patient's father and sister) were found to be carriers of the same mutation.

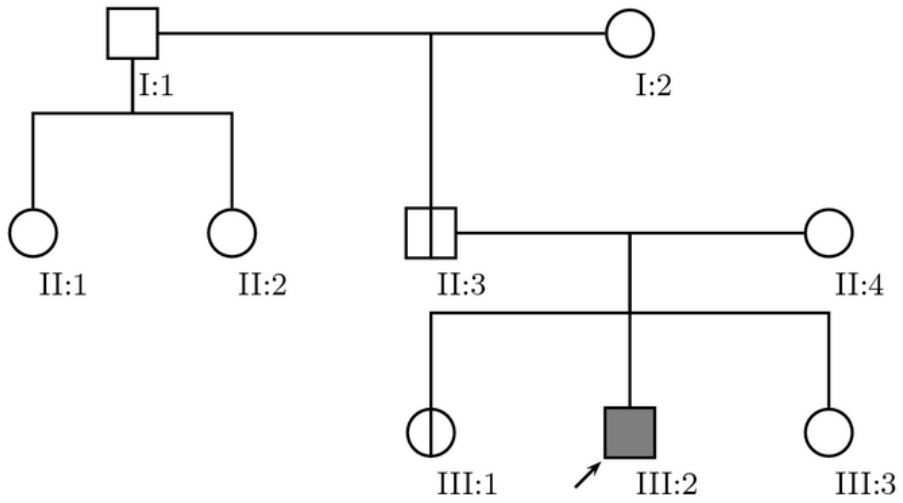


Figure 5.6 Family 1 Pedigree Family members subjected to sanger sequencing for status of S100A12 deletion. Squares represent males, circles represent females. Lines crossing cirles/squares represent carriers of the mutation (II:3 and III:1). The dark square represents the proband. I:1 was not tested but was included to illustrate relationship of II:1 and II:2 and potential genotype for I:1.

*Inflammatory Effect of Mutant S100A12*

In a preliminary *in vitro* experiment, the inflammatory effects of mutant S100A12 were assessed by measuring IL-1 and IL-6 mRNA expression. We exposed the patient's neutrophils to LPS in order to activate TLR4 signaling which results in activation of NFkB and induction of proinflammatory signals and secretion of cytokines, chemokines and our protein of interest, S100A12 which we expect to be in the culture supernatant. Subsequently, we exposed healthy monocytes to normal and mutant supernatants and collected cell lysates for microarray analysis. We observed that IL-1b expression was significantly higher in the healthy monocytes treated with both the mutant and normal activated supernatant compared to all other conditions (data not shown). The expression of IL-1b induced by mutant supernatant was not significantly different from that induced by the normal supernatant and LPS alone. We also saw that the inhibiting effect of PMB

was more pronounced in the patient samples compared to the LPS+PMB control culture. Therefore, the participation of LPS in the observed response cannot be ruled out. We also observed a significantly higher magnitude of expression of IL-6 by monocytes exposed to the mutant supernatant compared to normal (data not shown). However, as for the IL-1b expression, the participation of LPS in the observed response could not be ruled out.

Our *in vitro* data are inconclusive but given that this patient was not a complete responder to anakinra, our observations suggests the mutant S100A12 protein does not induce higher expression of IL-1B by monocytes but may preferentially induce expression of IL-6. This observation supports our hypothesis that an underlying genetic mechanism in this heterogenous disease would not be expected to be monogenic; rather multiple hits in genes within a pathway might be responsible for the phenotype (Esposito et al. 2013). Thus, two additional family members with same mutation do not suffer from autoinflammatory symptoms. Further, our data also suggest environmental interactions (i.e. LPS) together with the mutant S100A12 could induce stronger inflammatory responses. These patterns are starting to arise in other genetic studies.

*Future Direction*

We intend to continue to pursue expression of the mutant protein to facilitate *in vitro* studies. Our collaborators Drs. Gerard and Sandy Zurawski have successfully expressed both the normal and mutant proteins using an E. *coli* expression vector. Mutant S100A12 expression from a eukaryotic vector has been difficult to obtain so far, but we will continue to pursue this route. These products will eventually be used in *in vitro* experiment to analyze their direct effects of monocyte activation. We also plan to

measure the effects of the mutation on calcium binding through a calcium flux assay using flow cytometry technology. In addition, Immuno precipitation studies will permit us to identify potential binding partners for the mutant S100A12.

# CHAPTER SIX

## General Conclusions and Future Direction

### *General Conclusions*

The focus of my dissertation was the development and implementation of high throughput technology and its application to the study of gene expression and genetic variation underlying human immune-mediated diseases. Genomic technology has proven to be robust for assessing genome wide transcriptional profiling as well as for determining whole genome nucleotide sequence variations. While robust, both the technology and the analysis of results have limitations. Further, the number of high throughput applications that are commercially available is rapidly growing, and identifying the most suitable method(s) to address specific scientific questions is critical for the success of any research project or clinical study. Proper assessment of methodologies and subsequent validation can help us optimize the amount of reliable information obtained from such methods and also can help us identify instances when the information is not dependable; especially given the enormous amount of data generated.

The first aim of this study was to assess the performance of high throughput methodologies in preparation for their application to research and clinical studies. We determined that our targeted assay developed for measurement of expression of type I interferon inducible genes was robust and reliable. Data generated from the NanoString targeted assay validation experiments showed that this methodology is accurate and precise. Further, control samples produced for the purpose of validation and assay

controls proved very useful in assessing not only the method performance but also the laboratory performance.

Assessment of microarray and RNAseq methodologies demonstrated that both technologies are robust and highlighted limitations and benefits for each - for example, detection of a partial gene transcript by microarray for a gene that was essentially not transcribed, and the elucidation of such deficiency by RNAseq methodology. Nevertheless, each of these technologies confirmed the same underlying biological mechanism in the transcriptional network identified to be of biological significance. Next, we implemented exome capture methodologies and demonstrated that the accuracy for SNP detection for the Nimblegen exome capture was high, 95%, based on SNP array genotype calls. In addition, we established that the depth of coverage achieved using the Nimblegen capture was significantly higher than the ACE capture method using less sequencing, ~60M compared to 100M reads per sample, respectively. Nevertheless, ACE was seemingly superior at enriching targeted regions compared to Nimblegen based on the number of discordant heterozygous calls between the two methods, 80 and 31, respectively. This is most likely the result of supplemental probes used in ACE to overcome the problem of enrichment for known hard-to-sequence regions that are of clinical relevance. All in all, both methodologies generated accurate results. By comparing the two, we were made aware of their limitations and the importance of optimization and performance assessments in order to maximize reliability of the information obtained and be cognoscente of information missed.

The second aim of this study was to apply high throughput gene expression technology to gain insight into pathogenic mechanisms underlying human immune

deficiencies. Using microarray technology we elucidated the TLR-3 dependent signature elicited by stimulation with Poly(I:C) of fibroblasts and PBMCs from a *TLR3-/-* patient and healthy controls. Our data demonstrated that Poly(I:C) activation of skin fibroblast is fundamental for inducing an anti-viral transcriptional response in a TLR3-dependent manner. On the contrary, the transcriptional response observed in PBMCs is TLR-3 independent. Therefore, our collaborators concluded that fibroblast TLR3 is fundamental to elicit protection against HSV-1 encephalitis and suggested a redundant function for TLR3 in hemopoietic cells.

The third aim of this study was to apply whole exome sequencing technology to identify a potential genetic etiology for sJIA. We collected blood samples from a total of six family trios plus thirteen individual sJIA patients and generated WES data. The experimental design included selection of extreme phenotypes, and this proved very useful given the heterogeneous nature of the disease. In our analysis of each family trio, we identified several strong candidates in genes associated with the IL-1 signaling pathway, which is very relevant to sJIA pathogenesis and an approved therapeutic target. We combined all strong candidate and potentially interesting genes from each family and identified 72 unique genes. We examined how these genes segregated in sJIA as a cohort and identified a total of 7 genes segregating in at least 2 of the 19 patients examined. Of high potential, we identified mutations in IL-1A in 4 out of 5 patients that were classified as partial or non-responders to anakinra treatment. In addition, we observed that 8 out of 19 sJIA patients carried mutations in the ZFHX3 gene, which has a GWAS association to Kawasaki disease. We observed predominance of compound heterozygous variants in most genes that are highly relevant to sJIA. Our pathway and network analysis also

suggest a multigenic inheritance model at the core of sJIA disease pathogenesis. Moreover, we identified a heterozygous frameshift mutation in S100A12 in one of our six families. Preliminary work assessing the functional implications of S100A12 mutation suggests that both the mutant and normal proteins are expressed in the patient. We are pursuing cloning and expression of both proteins for further analysis.

*Future Direction*

Validation of identified variants was carried out through re-sequencing and/or RNASeq. Over half the variants identified were validated using these approaches. We intend to continue with validating the remaining candidate identified variants. This will include Sanger sequencing of variants that have not been validated through NGS re-sequencing or by RNAseq. Further, we intend to run targeted re-sequencing for 72 candidate genes using a larger cohort of sJIA patients. The DNA extraction method identified as part of this work facilitates the potential use of over 200 sJIA patient samples present in our repository. This will help calculate the frequency of mutations in these genes in sJIA compared to non-affected controls, which will enable us to compile a more complete report and eventually submit these data for publication. Further analysis of the SNP array data will also be done to identify copy number variants as well as structural variants.

We will also investigate other inheritance mechanisms that are starting to be recognized in the background of complex diseases such as *cis* and *trans* alleles, imprinting, low-penetrance disease predisposition with genetic modifier effect, and others (van Heyningen & Yeyati 2004). These analyses require complex algorithms and computational knowledge and therefore collaborations with experts in this area will be

arranged. Moreover, we will utilize current analysis tools such as Variant Analysis Annotation and Search Tool (VAAST) as well as pedigree-VAAST (Yandell et al. 2011; Hu et al. 2013; Kennedy et al. 2014; Hu et al. 2014). This analysis tool facilitates the analysis of grouped samples as well individual trios. The algorithm uses functional prediction scores, allele frequency and penetrance information to estimate and rank variants that are more likely to cause a phenotype. In addition, pVAAST integrates linkage analysis and resulting scores to assist in further prioritizing variants.

We also intend to select individual candidates and further investigate the functional implications of mutations observed. Where appropriate, we will profile effects using CRISPR-Cas9 technology as well as standard transfection models. In addition and as already stated, we will continue to pursue the expression work for the S100A12 variant to isolate and determine potential effects of the mutant protein in sJIA pathogenesis.

# REFERENCES


Adzhubei, I., Jordan, D.M. & Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, Chapter 7, p.Unit7.20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23315928 [Accessed October 23, 2014].

Allantaz, F., Chaussabel, D., Stichweh, D., et al., 2007. Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. *The Journal of experimental medicine*, 204(9), pp.2131–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2118700&tool=pmcentrez&rendertype=abstract [Accessed November 9, 2014].

Allantaz, F., Chaussabel, D., Banchereau, J., et al., 2007. Microarray-based identification of novel biomarkers in IL-1-mediated diseases. *Current opinion in immunology*, 19(6), pp.623–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2747489&tool=pmcentrez&rendertype=abstract [Accessed December 2, 2014].

Aubert, P. et al., 2012. Homeostatic tissue responses in skin biopsies from NOMID patients with constitutive overproduction of IL-1β. *PloS one*, 7(11), p.e49408. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3511496&tool=pmcentrez&rendertype=abstract [Accessed December 28, 2014].

Autmizguine, J. et al., 2015. Rilonacept pharmacokinetics in children with systemic juvenile idiopathic arthritis. *The Journal of Clinical Pharmacology*, 55(1), pp.39–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25079592 [Accessed March 9, 2015].

Avau, A. et al., 2014. Cytokine balance and cytokine-driven natural killer cell dysfunction in systemic juvenile idiopathic arthritis. *Cytokine & growth factor reviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24948570 [Accessed November 11, 2014].

Bainbridge, M.N. et al., 2010. Whole exome capture in solution with 3 Gbp of data. *Genome biology*, 11(6), p.R62. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2911110&tool=pmcentrez&rendertype=abstract [Accessed December 30, 2014].

Balow, J.E. et al., 2013. Microarray-based gene expression profiling in patients with cryopyrin-associated periodic syndromes defines a disease-related signature and IL-1-responsive transcripts. *Annals of the rheumatic diseases*, 72(6), pp.1064–70. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4174357&tool=pmcentrez&rendertype=abstract [Accessed December 28, 2014].

Bamshad, M.J. et al., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews. Genetics*, 12(11), pp.745–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21946919 [Accessed July 10, 2014].

Barbosa, M. et al., 1996. Identification of the homologous beige and Chediak-Higashi syndrome genes. *Nature*, (382), pp.262–265.

Barnes, M.G. et al., 2009. Subtype-specific peripheral blood gene expression profiles in recent-onset juvenile idiopathic arthritis. *Arthritis and rheumatism*, 60(7), pp.2102–12. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2782469&tool=pmcentrez&rendertype=abstract [Accessed December 2, 2014].

Behrens, E.M., 2008. Macrophage activation syndrome in rheumatic disease: what is the role of the antigen presenting cell? *Autoimmunity reviews*, 7(4), pp.305–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18295734 [Accessed November 17, 2014].

Benko, S. et al., 2010. NLRC5 limits the activation of inflammatory pathways. *Journal of immunology (Baltimore, Md. : 1950)*, 185(3), pp.1681–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20610642 [Accessed February 17, 2015].

Bennett, L. et al., 2003. Interferon and Granulopoiesis Signatures in Systemic Lupus Erythematosus Blood. *Journal of Experimental Medicine*, 197(6), pp.711–723. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2193846&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

Bentley, D.R. et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2581791&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Bernstein, B.E. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3439153&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Blanchard, A., RJ, K. & LE, H., 1996. High-density oligonucleotide arrays. *Biosensors and Bioelectronics*, 1, pp.687–690.

Brinkman, D.M.C. et al., 2007. Autologous stem cell transplantation in children with severe progressive systemic or polyarticular juvenile idiopathic arthritis: long-term follow-up of a prospective clinical trial. *Arthritis and rheumatism*, 56(7), pp.2410–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17599770 [Accessed December 29, 2014].

Bumgarner, R., 2013. Overview of DNA microarrays: types, applications, and their future. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Chapter 22, p.Unit 22.1. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4011503&tool=pmcentrez&rendertype=abstract [Accessed February 16, 2015].

Burgner, D. et al., 2009. A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease. *PLoS genetics*, 5(1), p.e1000319. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2607021&tool=pmcentrez&rendertype=abstract [Accessed March 7, 2015].

Cantarel, B.L. et al., 2014. BAYSIC: a Bayesian method for combining sets of genome variants with improved specificity and sensitivity. *BMC bioinformatics*, 15(1), p.104. Available at: http://www.biomedcentral.com/1471-2105/15/104 [Accessed January 9, 2015].

Cao, L.X. et al., 1997. Implication of a new molecule IK in CD34+ hematopoietic progenitor cell proliferation and differentiation. *Blood*, 89(10), pp.3615–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9160666 [Accessed March 8, 2015].

Carneiro, M.O. et al., 2012. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC genomics*, 13, p.375. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3443046&tool=pmcentrez&rendertype=abstract [Accessed February 9, 2015].

Cazzola, M., Ponchio, L. & De Benedetti, F., 1996. Defective iron supply to erythropoiesis and adequate endogenous erythropoietin production in the anemia associated with systemic onset juvenile chronic arthritis. *Blood*, 87, pp.4824–30.

Chaisson, M.J.P. et al., 2014. Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, 517(7536), pp.608–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25383537 [Accessed November 10, 2014].

Challis, D. et al., 2012. An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC bioinformatics*, 13(8).

Chaussabel, D. et al., 2008. A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus. *Immunity*, 29(1), pp.150–164. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2727981&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

Chaussabel, D. & Baldwin, N., 2014. Democratizing systems immunology with modular transcriptional repertoire analyses. *Nature Reviews Immunology*, 14(4), pp.271–280. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4118927&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

Chi, Z. et al., 2012. RecQL4 cytoplasmic localization: implications in mitochondrial DNA oxidative damage repair. *The international journal of biochemistry & cell biology*, 44(11), pp.1942–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3461334&tool=pmcentrez&rendertype=abstract [Accessed March 8, 2015].

Chiche, L. et al., 2014. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. *Arthritis & rheumatology (Hoboken, N.J.)*, 66(6), pp.1583–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24644022 [Accessed February 9, 2015].

Choi, M. et al., 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45), pp.19096–101. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2768590&tool=pmcentrez&rendertype=abstract [Accessed November 3, 2014].

Church, L.D., Cook, G.P. & McDermott, M.F., 2008. Primer: inflammasomes and interleukin 1beta in inflammatory disorders. *Nature clinical practice. Rheumatology*, 4(1), pp.34–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18172447 [Accessed January 21, 2015].

Cingolani, P. et al., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2), pp.80–92. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3679285&tool=pmcentrez&rendertype=abstract [Accessed December 22, 2014].

Clark, M.J. et al., 2011. Performance comparison of exome DNA sequencing technologies. *Nature biotechnology*, 29(10), pp.908–14. Available at: http://dx.doi.org/10.1038/nbt.1975 [Accessed July 10, 2014].

Cloonan, N. & Grimmond, S.M., 2008. Transcriptome content and dynamics at single-nucleotide resolution. *Genome biology*, 9(9), p.234. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2592708&tool=pmcentrez&rendertype=abstract [Accessed November 11, 2014].

Coffey, A. et al., 1998. Host response to EBV infection in X-linked lymphoproliferative diseae results from mutations in an SH2-domain encoding gene. *Nat Genet*, (20), pp.129–135.

Cooper, G.M. et al., 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome research*, 15(7), pp.901–13. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1172034&tool=pmcentrez&rendertype=abstract [Accessed January 7, 2015].

Cooper, G.M. & Brown, C.D., 2008. Qualifying the relationship between sequence conservation and molecular function. *Genome research*, 18(2), pp.201–5. Available at: http://genome.cshlp.org/content/18/2/201.long [Accessed January 29, 2015].

Cooper, G.M. & Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature reviews. Genetics*, 12(9), pp.628–40. Available at: http://dx.doi.org/10.1038/nrg3046 [Accessed January 10, 2015].

Costantini, C. et al., 2011. On the potential involvement of CD11d in co-stimulating the production of interferon-γ by natural killer cells upon interaction with neutrophils via intercellular adhesion molecule-3. *Haematologica*, 96(10), pp.1543–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3186317&tool=pmcentrez&rendertype=abstract [Accessed March 7, 2015].

D'Ambrosi, N. et al., 2014. Rac1 at the crossroad of actin dynamics and neuroinflammation in Amyotrophic Lateral Sclerosis. *Frontiers in cellular neuroscience*, 8, p.279. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4157560&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2015].

Davis, B.K. et al., 2011. Cutting edge: NLRC5-dependent activation of the inflammasome. *Journal of immunology (Baltimore, Md. : 1950)*, 186(3), pp.1333–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3669680&tool=pmcentrez&rendertype=abstract [Accessed March 9, 2015].

DePristo, M. et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 43(5), pp.491–498.

DeRisi, J. et al., 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature genetics*, 14(4), pp.457–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8944026 [Accessed March 23, 2015].

DeRisi, J.L. & Iyer, V.R., 1999. Genomics and array technology. *Current opinion in oncology*, 11(1), pp.76–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9914882 [Accessed March 23, 2015].

DeWitt, E.M. et al., 2012. Consensus treatment plans for new-onset systemic juvenile idiopathic arthritis. *Arthritis care & research*, 64(7), pp.1001–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3368104&tool=pmcentrez&rendertype=abstract [Accessed December 28, 2014].

Dinarello, C. a, 2010. IL-1: discoveries, controversies and future directions. *European journal of immunology*, 40(3), pp.599–606. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20201008 [Accessed October 11, 2014].

Donato, R. et al., 2013. Functions of S100 proteins. *Current molecular medicine*, 13(1), pp.24–57. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3707951&tool=pmcentrez&rendertype=abstract [Accessed December 3, 2014].

el-Gabalawy, H. et al., 1996. Synovial distribution of alpha d/CD18, a novel leukointegrin. Comparison with other integrins and their ligands. *Arthritis and rheumatism*, 39(11), pp.1913–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8912515 [Accessed March 7, 2015].

Erlich, Y. et al., 2011. Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome research*, 21(5), pp.658–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3083082&tool=pmcentrez&rendertype=abstract [Accessed January 29, 2015].

Esposito, T. et al., 2013. Digenic mutational inheritance of the integrin alpha 7 and the myosin heavy chain 7B genes causes congenital myopathy with left ventricular non-compact cardiomyopathy. *Orphanet journal of rare diseases*, 8, p.91. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3695851&tool=pmcentrez&rendertype=abstract [Accessed March 8, 2015].

Fall, N. et al., 2007. Gene expression profiling of peripheral blood from patients with untreated new-onset systemic juvenile idiopathic arthritis reveals molecular heterogeneity that may predict macrophage activation syndrome. *Arthritis and rheumatism*, 56(11), pp.3793–804. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17968951 [Accessed November 7, 2014].

Feldmann, J. et al., 2003. Munc13-4 is essential for cytolytic granules fusion and is mutated in a form of familial hemophagocytic lymphohistiocytosis (FHL3). *Cell*, 115(4), pp.461–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14622600 [Accessed December 3, 2014].

Ferguson, J.A., Steemers, F.J. & Walt, D.R., 2000. High-density fiber-optic DNA random microsphere array. *Analytical chemistry*, 72(22), pp.5618–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11101240 [Accessed March 23, 2015].

Fife, M.S. et al., 2006. Novel IL10 gene family associations with systemic juvenile idiopathic arthritis. *Arthritis research & therapy*, 8(5), p.R148. Available at: http://arthritis-research.com/content/8/5/R148 [Accessed December 28, 2014].

Fodor, S.P. et al., 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science (New York, N.Y.)*, 251(4995), pp.767–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1990438 [Accessed March 23, 2015].

Foell, D. et al., 2013. Proinflammatory S100A12 can activate human monocytes via Toll-like receptor 4. *American journal of respiratory and critical care medicine*, 187(12), pp.1324–34. Available at: http://www.atsjournals.org/doi/abs/10.1164/rccm.201209-1602OC?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub=pubmed&#.VH4LMIsbDow [Accessed December 2, 2014].

Fox, S., Filichkin, S. & Mockler, T.C., 2009. Applications of ultra-high-throughput sequencing. *Methods in molecular biology (Clifton, N.J.)*, 553, pp.79–108. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19588102 [Accessed March 15, 2015].

Frosch, M. et al., 2009. The myeloid-related proteins 8 and 14 complex, a novel ligand of toll-like receptor 4, and interleukin-1beta form a positive feedback mechanism in systemic-onset juvenile idiopathic arthritis. *Arthritis and rheumatism*, 60(3), pp.883–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19248102 [Accessed November 11, 2014].

Frosch, M. & Roth, J., 2008. New insights in systemic juvenile idiopathic arthritis--from pathophysiology to treatment. *Rheumatology (Oxford, England)*, 47(2), pp.121–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17971384 [Accessed November 11, 2014].

Garrison, E. & Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. *arXivorg*.

Geiss, G.K. et al., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature biotechnology*, 26(3), pp.317–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18278033 [Accessed July 10, 2014].

Giannini, E.H. et al., 1992. Methotrexate in resistant juvenile rheumatoid arthritis. Results of the U.S.A.-U.S.S.R. double-blind, placebo-controlled trial. The Pediatric Rheumatology Collaborative Study Group and The Cooperative Children's Study Group. *The New England journal of medicine*, 326(16), pp.1043–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1549149 [Accessed December 28, 2014].

Giannini, E.H. et al., 1997. Preliminary definition of improvement in juvenile arthritis. *Arthritis and rheumatism*, 40(7), pp.1202–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9214419 [Accessed December 28, 2014].

Goh, A.X.H. et al., A novel human anti-interleukin-1β neutralizing monoclonal antibody showing in vivo efficacy. *mAbs*, 6(3), pp.765–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24671001 [Accessed December 28, 2014].

Gray, E.J. et al., 2014. Src homology 2 domain containing protein 5 (SH2D5) binds the breakpoint cluster region protein, BCR, and regulates levels of Rac1-GTP. *The Journal of biological chemistry*, 289(51), pp.35397–408. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25331951 [Accessed March 7, 2015].

Grom, A.A., 2014. Canakinumab for the treatment of systemic juvenile idiopathic arthritis. *Expert review of clinical immunology*, 10(11), pp.1427–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25269376 [Accessed December 3, 2014].

Grom, A.A. et al., 2003. Natural killer cell dysfunction in patients with systemic-onset juvenile rheumatoid arthritis and macrophage activation syndrome. *The Journal of pediatrics*, 142(3), pp.292–6. Available at: http://www.jpeds.com/article/S0022347602403605/fulltext [Accessed December 28, 2014].

Grom, A.A. & Hirsch, R., 2000. T-cell and T-cell receptor abnormalities in the immunopathogenesis of juvenile rheumatoid arthritis. *Current opinion in rheumatology*, 12(5), pp.420–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10990179 [Accessed December 29, 2014].

Gross, S.R. et al., 2014. Joining S100 proteins and migration: for better or for worse, in sickness and in health. *Cellular and molecular life sciences : CMLS*, 71(9), pp.1551–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23811936 [Accessed January 9, 2015].

Gu, W. et al., 2012. Fringe controls naïve CD4(+)T cells differentiation through modulating notch signaling in asthmatic rat models. *PloS one*, 7(10), p.e47288. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3468515&tool=pmcentrez&rendertype=abstract [Accessed March 7, 2015].

Guiducci, C. et al., 2010. TLR recognition of self nucleic acids hampers glucocorticoid activity in lupus. *Nature*, 465(7300), pp.937–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2964153&tool=pmcentrez&rendertype=abstract [Accessed January 17, 2015].

Gunderson, K.L. et al., 2004. Decoding randomly ordered DNA arrays. *Genome research*, 14(5), pp.870–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=479114&tool=pmcentrez&rendertype=abstract [Accessed March 23, 2015].

Halawi, A., Abbas, O. & Mahalingam, M., 2014. S100 proteins and the skin: a review. *Journal of the European Academy of Dermatology and Venereology : JEADV*, 28(4), pp.405–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23924267 [Accessed December 3, 2014].

Han, L. et al., 2014. Alternative applicationns for distinct RNA sequencing strategies. *Briefings in bioinformatics*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25246237 [Accessed September 25, 2014].

Harris, J.G., Kessler, E.A. & Verbsky, J.W., 2013. Update on the treatment of juvenile idiopathic arthritis. *Current allergy and asthma reports*, 13(4), pp.337–46. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3729726&tool=pmcentrez&rendertype=abstract [Accessed December 28, 2014].

Van Heyningen, V. & Yeyati, P.L., 2004. Mechanisms of non-Mendelian inheritance in genetic disease. *Human Molecular Genetics*, 13(2), pp.R225–R233.

Hinze, C.H. et al., 2010. Immature cell populations and an erythropoiesis gene-expression signature in systemic juvenile idiopathic arthritis: implications for pathogenesis. *Arthritis research & therapy*, 12(3), p.R123. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2911917&tool=pmcentrez&rendertype=abstract [Accessed November 17, 2014].

Holzinger, D. et al., 2012. The Toll-like receptor 4 agonist MRP8/14 protein complex is a sensitive indicator for disease activity and predicts relapses in systemic-onset juvenile idiopathic arthritis. *Annals of the rheumatic diseases*, 71(6), pp.974–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22267331 [Accessed December 2, 2014].

Hu, H. et al., 2014. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. *Nature biotechnology*, 32(7), pp.663–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4157619&tool=pmcentrez&rendertype=abstract [Accessed November 3, 2014].

Hu, H. et al., 2013. VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genetic epidemiology*, 37(6), pp.622–34. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3791556&tool=pmcentrez&rendertype=abstract [Accessed August 21, 2014].

Huang, C. et al., 1999. Human tryptases alpha and beta/II are functionally distinct due, in part, to a single amino acid difference in one of the surface loops that forms the substrate-binding cleft. *The Journal of biological chemistry*, 274(28), pp.19670–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10391906 [Accessed March 9, 2015].

Hull, K.M. et al., 2003. The expanding spectrum of systemic autoinflammatory disorders and their rheumatic manifestations. *Current opinion in rheumatology*, 15(1), pp.61–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12496512 [Accessed March 14, 2015].

Ilowite, N.T., 2014. Randomized, Double-Blind, Placebo-Controlled Trial of the Efficacy and Safety of Rilonacept in the Treatment of Systemic Juvenile Idiopathic Arthritis. *Arthritis & rheumatology (Hoboken, N.J.)*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25511916 [Accessed December 28, 2014].

Ilowite, N.T. et al., 2014. Randomized, double-blind, placebo-controlled trial of the efficacy and safety of rilonacept in the treatment of systemic juvenile idiopathic arthritis. *Arthritis & rheumatology (Hoboken, N.J.)*, 66(9), pp.2570–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24839206 [Accessed March 9, 2015].

De Jager, W. et al., 2009. Defective phosphorylation of interleukin-18 receptor beta causes impaired natural killer cell function in systemic-onset juvenile idiopathic arthritis. *Arthritis and rheumatism*, 60(9), pp.2782–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19714583 [Accessed December 2, 2014].

Jalimarada, S.S., Ogando, D.G. & Bonanno, J.A., 2014. Loss of ion transporters and increased unfolded protein response in Fuchs' dystrophy. *Molecular vision*, 20, pp.1668–79. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4265779&tool=pmcentrez&rendertype=abstract [Accessed March 11, 2015].

Jesus, A.A. & Goldbach-Mansky, R., 2014a. IL-1 blockade in autoinflammatory syndromes. *Annual review of medicine*, 65, pp.223–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4178953&tool=pmcentrez&rendertype=abstract [Accessed March 12, 2015].

Jesus, A.A. & Goldbach-Mansky, R., 2014b. IL-1 blockade in autoinflammatory syndromes. *Annual review of medicine*, 65, pp.223–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24422572 [Accessed December 19, 2014].

Jiang, Q. et al., 2014. Down-regulation of ATBF1 activates STAT3 signaling via PIAS3 in pacing-induced HL-1 atrial myocytes. *Biochemical and biophysical research communications*, 449(3), pp.278–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24857987 [Accessed March 7, 2015].

Jin, H. et al., 2000. Two novel members of the interleukin-1 receptor gene family, one deleted in Xp22.1-Xp21.3 mental retardation. *European journal of human genetics : EJHG*, 8(2), pp.87–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10757639 [Accessed March 9, 2015].

Kaufman, K.M. et al., 2014. Whole-exome sequencing reveals overlap between macrophage activation syndrome in systemic juvenile idiopathic arthritis and familial hemophagocytic lymphohistiocytosis. *Arthritis & rheumatology (Hoboken, N.J.)*, 66(12), pp.3486–95. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4321811&tool=pmcentrez&rendertype=abstract [Accessed March 14, 2015].

Kennedy, B. et al., 2014. Using VAAST to Identify Disease-Associated Variants in Next-Generation Sequencing Data. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, 81(April), pp.6.14.1–6.14.25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4137768&tool=pmcentrez&rendertype=abstract [Accessed November 11, 2014].

Kim, N.K. et al., 2015. Whole-exome sequencing reveals diverse modes of inheritance in sporadic mild to moderate sensorineural hearing loss in a pediatric population. *Genetics in Medicine*.

Kim, T.-S. et al., The ZFHX3 (ATBF1) transcription factor induces PDGFRB, which activates ATM in the cytoplasm to protect cerebellar neurons from oxidative stress. *Disease models & mechanisms*, 3(11-12), pp.752–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20876357 [Accessed March 7, 2015].

Kimura, M., 1991. The neutral theory of molecular evolution: a review of recent evidence. *Idengaku zasshi*, 66, pp.367–386.

De Kleer, I.M. et al., 2004. CD4+CD25bright regulatory T cells actively regulate inflammation in the joints of patients with the remitting form of juvenile idiopathic arthritis. *Journal of immunology (Baltimore, Md. : 1950)*, 172(10), pp.6435–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15128835 [Accessed December 29, 2014].

Krief, P. et al., 1994. A new cytokine (IK) down-regulating HLA class II: monoclonal antibodies, cloning and chromosome localization. *Oncogene*, 9(12), pp.3449–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7970704 [Accessed March 8, 2015].

Kruhøffer, M. et al., 2007. Isolation of microarray-grade total RNA, microRNA, and DNA from a single PAXgene blood RNA tube. *The Journal of molecular diagnostics : JMD*, 9(4), pp.452–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1975097&tool=pmcentrez&rendertype=abstract [Accessed December 29, 2014].

Lamb, R. et al., 2007. Positive association of SLC26A2 gene polymorphisms with susceptibility to systemic-onset juvenile idiopathic arthritis. *Arthritis and rheumatism*, 56(4), pp.1286–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17393463 [Accessed December 28, 2014].

Landrum, M.J. et al., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research*, 42(Database issue), pp.D980–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965032&tool=pmcentrez&rendertype=abstract [Accessed January 14, 2015].

Lasigliè, D. et al., 2011. Role of IL-1 beta in the development of human T(H)17 cells: lesson from NLPR3 mutated patients. *PloS one*, 6(5), p.e20014. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3102666&tool=pmcentrez&rendertype=abstract [Accessed March 14, 2015].

Li, H., 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics (Oxford, England)*, 30(20), pp.2843–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24974202 [Accessed July 9, 2014].

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), pp.1754–60. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2705234&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Ling, X.B. et al., 2010a. Plasma profiles in active systemic juvenile idiopathic arthritis: Biomarkers and biological implications. *Proteomics*, 10(24), pp.4415–30. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3517169&tool=pmcentrez&rendertype=abstract [Accessed November 11, 2014].

Ling, X.B. et al., 2010b. Plasma profiles in active systemic juvenile idiopathic arthritis: Biomarkers and biological implications. *Proteomics*, 10(24), pp.4415–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21136595 [Accessed April 17, 2012].

Liu, L. et al., 2012. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology*, 2012, p.251364. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22829749 [Accessed July 11, 2014].

Liu, X., Jian, X. & Boerwinkle, E., 2013. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Human mutation*, 34(9), pp.E2393–402. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4109890&tool=pmcentrez&rendertype=abstract [Accessed January 26, 2015].

Liu, X., Jian, X. & Boerwinkle, E., 2011. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Human mutation*, 32(8), pp.894–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3145015&tool=pmcentrez&rendertype=abstract [Accessed February 11, 2015].

Macaubas, C. et al., 2012a. Alternative activation in systemic juvenile idiopathic arthritis monocytes. *Clinical immunology (Orlando, Fla.)*, 142(3), pp.362–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3288602&tool=pmcentrez&rendertype=abstract [Accessed December 2, 2014].

Macaubas, C. et al., 2012b. Alternative activation in systemic juvenile idiopathic arthritis monocytes. *Clinical immunology (Orlando, Fla.)*, 142(3), pp.362–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22281427 [Accessed March 24, 2012].

Macaubas, C. et al., 2010. Distribution of circulating cells in systemic juvenile idiopathic arthritis across disease activity states. *Clinical immunology (Orlando, Fla.)*, 134(2), pp.206–16. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2818241&tool=pmcentrez&rendertype=abstract [Accessed November 11, 2014].

Madsen, D.H. et al., 2012. Endocytic collagen degradation: a novel mechanism involved in protection against liver fibrosis. *The Journal of pathology*, 227(1), pp.94–105. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22294280 [Accessed March 7, 2015].

Marsh, R. et al., 2010. XIAP deficiency: a unique primary immunodeficiency best classified as X-linked familial hemophagocytic lymphohistiocytosis and not as X-linked lymphoproliferative disease. *Blood*, (116), pp.1079–1082.

Martin, J. a & Wang, Z., 2011. Next-generation transcriptome assembly. *Nature reviews. Genetics*, 12(10), pp.671–682. Available at: http://dx.doi.org/10.1038/nrg3068.

Martini, A., 2012. Systemic juvenile idiopathic arthritis. *Autoimmunity reviews*, 12(1), pp.56–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22884552 [Accessed December 3, 2014].

McKenna, A. et al., 2010. The genome analysis tool kit: a MapReduce framework for analyzign next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Mediero, A. et al., 2014. Methotrexate prevents wear particle-induced inflammatory osteolysis via activation of the adenosine A2A receptor. *Arthritis & rheumatology (Hoboken, N.J.)*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25533750 [Accessed December 28, 2014].

Meldrum, D., 2000a. Automation for genomics, part one: preparation for sequencing. *Genome research*, 10(8), pp.1081–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10958626 [Accessed March 15, 2015].

Meldrum, D., 2000b. Automation for genomics, part two: sequencers, microarrays, and future trends. *Genome research*, 10(9), pp.1288–303. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10984447 [Accessed March 15, 2015].

Mellins, E.D., Macaubas, C. & Grom, A.A., 2011. Pathogenesis of systemic juvenile idiopathic arthritis: some answers, more questions. *Nature reviews. Rheumatology*, 7(7), pp.416–26. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4180659&tool=pmcentrez&rendertype=abstract [Accessed December 3, 2014].

Menasche, G. et al., 2000. Mutations in RAB27A causes Griscelli syndrome associated with haemophgaocytic syndrome. *Nat Genet*, (25), pp.173–176.

Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11(1), pp.31–46. Available at: http://dx.doi.org/10.1038/nrg2626 [Accessed July 9, 2014].

Michael, K.L. et al., 1998. Randomly ordered addressable high-density optical sensor arrays. *Analytical chemistry*, 70(7), pp.1242–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9553489 [Accessed March 23, 2015].

Mihara, M. et al., 2012. IL-6/IL-6 receptor system and its role in physiological and pathological conditions. *Clinical science (London, England : 1979)*, 122(4), pp.143–59. Available at: http://www.clinsci.org/cs/122/0143/cs1220143.htm [Accessed January 29, 2015].

Mitra, R., 1999. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Research*, 27(24), p.34e–34. Available at: http://nar.oxfordjournals.org/content/27/24/e34 [Accessed December 30, 2014].

Morozova, O. & Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5), pp.255–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18703132 [Accessed July 10, 2014].

Muller, K. et al., 1998. Inflammatory cytokines and cytokine antagonists in whole blood cultures of patients with systemic juvenile chronic arthritis. *Rheumatology*, 37(5), pp.562–569. Available at: http://rheumatology.oxfordjournals.org/content/37/5/562 [Accessed December 28, 2014].

Nakashima, T. & Takayanagi, H., 2009. Osteoclasts and the immune system. *Journal of bone and mineral metabolism*, 27(5), pp.519–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19455385 [Accessed November 17, 2014].

Ng, B.G. et al., 2013. Mosaicism of the UDP-galactose transporter SLC35A2 causes a congenital disorder of glycosylation. *American journal of human genetics*, 92(4), pp.632–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3617373&tool=pmcentrez&rendertype=abstract [Accessed February 28, 2015].

Ng, S.B. et al., 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*, 42(1), pp.30–5. Available at: http://www.nature.com/ng/journal/v42/n1/full/ng.499.html#close [Accessed July 15, 2014].

Ng, S.B. et al., 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261), pp.272–6. Available at: http://dx.doi.org/10.1038/nature08250 [Accessed July 9, 2014].

Nigrovic, P.A. et al., 2011. Anakinra as first-line disease-modifying therapy in systemic juvenile idiopathic arthritis: report of forty-six patients from an international multicenter series. *Arthritis and rheumatism*, 63(2), pp.545–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21280009 [Accessed December 28, 2014].

Nigrovic, P.A., 2014. Review: is there a window of opportunity for treatment of systemic juvenile idiopathic arthritis? *Arthritis & rheumatology (Hoboken, N.J.)*, 66(6), pp.1405–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24623686 [Accessed December 3, 2014].

Nirmala, N., Grom, A. & Gram, H., 2014. Biomarkers in systemic juvenile idiopathic arthritis: a comparison with biomarkers in cryopyrin-associated periodic syndromes. *Current opinion in rheumatology*, 26(5), pp.543–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25050926 [Accessed November 11, 2014].

O'Rawe, J. et al., 2013. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome medicine*, 5(3), p.28. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3706896&tool=pmcentrez&rendertype=abstract [Accessed July 14, 2014].

Obermoser, G. et al., 2013. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity*, 38(4), pp.831–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3681204&tool=pmcentrez&rendertype=abstract [Accessed February 8, 2015].

Ogilvie, E.M. et al., 2007. Specific gene expression profiles in systemic juvenile idiopathic arthritis. *Arthritis and rheumatism*, 56(6), pp.1954–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17530721 [Accessed December 28, 2014].

Ohata, H. et al., 2013. NuMA is required for the selective induction of p53 target genes. *Molecular and cellular biology*, 33(12), pp.2447–57. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3700099&tool=pmcentrez&rendertype=abstract [Accessed March 8, 2015].

Omoyinmi, E. et al., 2012. Th1 and Th17 cell subpopulations are enriched in the peripheral blood of patients with systemic juvenile idiopathic arthritis. *Rheumatology (Oxford, England)*, 51(10), pp.1881–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3448884&tool=pmcentrez&rendertype=abstract [Accessed November 26, 2014].

Pabinger, S. et al., 2014. A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in bioinformatics*, 15(2), pp.256–78. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3956068&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Park, H.-L. et al., 2013. IK induced by coxsackievirus B3 infection transiently downregulates expression of MHC class II through increasing cAMP. *Viral immunology*, 26(1), pp.13–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23409929 [Accessed March 8, 2015].

Pascual, V. et al., 2005. Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade. *The Journal of experimental medicine*, 201(9), pp.1479–86. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2213182&tool=pmcentrez&rendertype=abstract [Accessed August 26, 2014].

Pazin, M.J., 2015. Using the ENCODE Resource for Functional Annotation of Genetic Variants. *Cold Spring Harbor protocols*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25762420 [Accessed March 14, 2015].

Perry, W.L. et al., 1998. The itchy locus encodes a novel ubiquitin protein ligase that is disrupted in a18H mice. *Nature genetics*, 18(2), pp.143–6. Available at: http://dx.doi.org/10.1038/ng0298-143 [Accessed March 12, 2015].

Pioch, D., Schweder, T. & Jürgen, B., 2008. Novel developments for improved detection of specific mRNAs by DNA chips. *Applied microbiology and biotechnology*, 80(6), pp.953–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18784921 [Accessed March 13, 2015].

Van der Pouw Kraan, C.T.M. et al., 2012. Defective IL-1A expression in patients with Crohn's disease is related to attenuated MAP3K4 signaling. *Human immunology*, 73(9), pp.912–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22732089 [Accessed March 8, 2015].

Prieur, A. & Stephan, J., 1994. Macrophage activation syndrome in children with joint diseases. *Rev. Rheum Engl Ed*, 61, pp.385–388.

Priya, R.R. et al., 2012. Exome sequencing: capture and sequencing of all human coding regions for disease gene discovery. *Methods in molecular biology (Clifton, N.J.)*, 884, pp.335–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22688718 [Accessed December 2, 2014].

Quail, M.A. et al., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics*, 13, p.341. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3431227&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Quartier, P. et al., 2011. A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial). *Annals of the rheumatic diseases*, 70(5), pp.747–54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3070271&tool=pmcentrez&rendertype=abstract [Accessed November 11, 2014].

Ravelli, A., 2002. Macrophage activation syndrome. *Current opinion in rheumatology*, 14, pp.548–552.

Ravelli, A. et al., 2012. Macrophage activation syndrome as part of systemic juvenile idiopathic arthritis: diagnosis, genetics, pathophysiology and treatment. *Genes and immunity*, 13(4), pp.289–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22418018 [Accessed November 17, 2014].

Ravelli, A. et al., 2005. Patients with anitnuclear antibody-positive juvenile idiopathic arthritis constitute a homogeneous subgroup irrespective of the course of joint disease. *Arthritis & Rheumatology*, 52(3), pp.826–32.

Ravelli, A. et al., 2004. Preliminary diagnostic guidelines for macrophage activation syndrome complicating systemic juvenile idiopathic arthritis. *The Journal of pediatrics*, 146(5), pp.598–604.

Ravelli, A. & Martini, A., 2007. Juvenile idiopathic arthritis. *Lancet*, 369(9563), pp.767–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17336654 [Accessed December 3, 2014].

Rey-Giraud, F., Hafner, M. & Ries, C.H., 2012. In vitro generation of monocyte-derived macrophages under serum-free conditions improves their tumor promoting functions. *PloS one*, 7(8), p.e42656. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3412794&tool=pmcentrez&rendertype=abstract [Accessed November 19, 2014].

Rider, P. et al., 2011. IL-1α and IL-1β recruit different myeloid cells and promote different stages of sterile inflammation. *Journal of immunology (Baltimore, Md. : 1950)*, 187(9), pp.4835–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21930960 [Accessed January 17, 2015].

Ritchie, G.R. & Flicek, P., 2014. Computational approaches to interpreting genomic sequence variation. *Genome Medicine*, 6(10), p.87. Available at: http://genomemedicine.com/content/6/10/87 [Accessed November 20, 2014].

Ruperto, N., Quartier, P., et al., 2012. A phase II, multicenter, open-label study evaluating dosing and preliminary safety and efficacy of canakinumab in systemic juvenile idiopathic arthritis with active systemic features. *Arthritis and rheumatism*, 64(2), pp.557–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21953497 [Accessed April 22, 2012].

Ruperto, N. et al., 2005. A randomized, double-blind clinical trial of two doses of meloxicam compared with naproxen in children with juvenile idiopathic arthritis: short- and long-term efficacy and safety results. *Arthritis and rheumatism*, 52(2), pp.563–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15692986 [Accessed December 28, 2014].

Ruperto, N., Brunner, H.I., et al., 2012. Two randomized trials of canakinumab in systemic juvenile idiopathic arthritis. *The New England journal of medicine*, 367(25), pp.2396–406. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23252526 [Accessed December 28, 2014].

Sanger, F., Nicklen, S. & Coulson, A.R., 1992. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology (Reading, Mass.)*, 24, pp.104–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1422003 [Accessed January 23, 2015].

Sawhney, S., Woo, P. & Murray, K., 2001. Macrophage activation syndrome: a potentially fatal complication of rheumatic disorders. *Arch Dis Child*, 85, pp.421–426.

Schaller, J.G., 1977. Corticosteroids in juvenile rheumatoid arthritis. *Arthritis and rheumatism*, 20(2 Suppl), pp.537–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/263922 [Accessed December 28, 2014].

Shalon, D., Smith, S.J. & Brown, P.O., 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome research*, 6(7), pp.639–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8796352 [Accessed March 14, 2015].

Shendure, J. & Ji, H., 2008. Next-generation DNA sequencing. *Nature biotechnology*, 26(10), pp.1135–1145.

Shendure, J.A. et al., 2011. Overview of DNA sequencing strategies. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Chapter 7, p.Unit7.1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21987056 [Accessed January 13, 2015].

Shigemura, T. et al., 2011. Monitoring serum IL-18 levels is useful for treatment of a patient with systemic juvenile idiopathic arthritis complicated by macrophage activation syndrome. *Pediatric rheumatology online journal*, 9(1), p.15. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3155148&tool=pmcentrez&rendertype=abstract [Accessed December 3, 2014].

Shimizu, M. et al., 2014. Serum interleukin 18 as a diagnostic remission criterion in systemic juvenile idiopathic arthritis. *The Journal of rheumatology*, 41(11), pp.2328–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25362722 [Accessed December 2, 2014].

Shimkets, R.A., 2004. Gene expression quantitation technology summary. *Methods in molecular biology (Clifton, N.J.)*, 258, pp.7–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14970454 [Accessed March 13, 2015].

Srivastava, S. et al., 2010. Monocytes are resistant to apoptosis in systemic juvenile idiopathic arthritis. *Clinical immunology (Orlando, Fla.)*, 136(2), pp.257–68. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2902595&tool=pmcentrez&rendertype=abstract [Accessed November 11, 2014].

Zur Stadt, U. et al., 2009. Famlial hemophagocytic lymphohistiocytosis type 5 (FHL-5) is caused by mutations in Munc18-2 and impaired binding to syntaxin 11. *Am J Hum Genet*, (85), pp.482–92.

Zur Stadt, U., Schmidt, S. & Kasper, B., 2005. Linkage of Familial hemophagocytic lymphohistiocytosis type 4 to chromosome 6q24 and identification of mutations in syntaxin 11. *Hum Mol Genet*, 14, pp.827–34.

Stepp, S.E. et al., 1999. Perforin gene defects in familial hemophagocytic lymphohistiocytosis. *Science (New York, N.Y.)*, 286(5446), pp.1957–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10583959 [Accessed December 3, 2014].

Stock, C.J.W. et al., 2008a. Comprehensive association study of genetic variants in the IL-1 gene family in systemic juvenile idiopathic arthritis. *Genes and immunity*, 9(4), pp.349–57. Available at: http://dx.doi.org/10.1038/gene.2008.24 [Accessed December 28, 2014].

Stock, C.J.W. et al., 2008b. Comprehensive association study of genetic variants in the IL-1 gene family in systemic juvenile idiopathic arthritis. *Genes and immunity*, 9(4), pp.349–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18418395 [Accessed November 17, 2014].

Sulonen, A.-M. et al., 2011. Comparison of solution-based exome capture methods for next generation sequencing. *Genome biology*, 12(9), p.R94. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3308057&tool=pmcentrez&rendertype=abstract [Accessed October 12, 2014].

Suresh, S. & Irvine, A.E., 2015. The NOTCH signaling pathway in normal and malignant blood cell production. *Journal of cell communication and signaling*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25711903 [Accessed March 7, 2015].

Szydłowski, M., Jabłońska, E. & Juszczyński, P., 2014. FOXO1 Transcription Factor: A Critical Effector of the PI3K-AKT Axis in B-Cell Development. Available at: http://informahealthcare.com/doi/abs/10.3109/08830185.2014.885022 [Accessed March 6, 2015].

Tanaka, T., Narazaki, M. & Kishimoto, T., 2014. IL-6 in Inflammation, Immunity, and Disease. *Cold Spring Harbor perspectives in biology*, 6(10), p.a016295. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25190079 [Accessed October 2, 2014].

Tong, L. et al., 2014. S100A proteins as molecular targets in the ocular surface inflammatory diseases. *The ocular surface*, 12(1), pp.23–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24439044 [Accessed December 3, 2014].

Vastert, S.J. et al., 2010. Mutations in the perforin gene can be linked to macrophage activation syndrome in patients with systemic onset juvenile idiopathic arthritis. *Rheumatology (Oxford, England)*, 49(3), pp.441–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20019066 [Accessed November 11, 2014].

Vastert, S.J., Kuis, W. & Grom, A.A., 2009. Systemic JIA: new developments in the understanding of the pathophysiology and therapy. *Best practice & research. Clinical rheumatology*, 23(5), pp.655–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2774820&tool=pmcentrez&rendertype=abstract [Accessed December 3, 2014].

Venter, J.C. et al., 2001. The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), pp.1304–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11181995 [Accessed July 10, 2014].

Verbsky, J.W. & White, A.J., 2004. Effective use of the recombinant interlukin 1 receptor antagonist anakinra in therapy resistant systemic onset juvenile rheumatoid arthritis. *Journal of Rheumatology*, 31(10), pp.2071–5. Available at: www.ncbi.nlm.nih.gov/pubmed/15468378.

Van der Vieren, M. et al., 1995. A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-3. *Immunity*, 3(6), pp.683–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8777714 [Accessed March 7, 2015].

Villanueva, J. et al., 2005. Natural killer cell dysfunction is a distinguished feature of systemic onset juvenile rheumatoid arthritis and macrophage activation syndrome. Available at: http://arthritis-research.com/content/pdf/ar1453.pdf [Accessed December 28, 2014].

Wakil, S. & Monies, D., 2014. Mutation of LACC1 is associated with a Monogenic Form of Systemic Juvenile Idiopathic Arthritis. *Arthritis & …*, DOI 10.100. Available at: http://onlinelibrary.wiley.com/doi/10.1002/art.38877/abstract [Accessed December 28, 2014].

Wang, K., Li, M. & Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16), p.e164. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2938201&tool=pmcentrez&rendertype=abstract [Accessed July 15, 2014].

Wesche, H. et al., 1997. The interleukin-1 receptor accessory protein (IL-1RAcP) is essential for IL-1-induced activation of interleukin-1 receptor-associated kinase (IRAK) and stress-activated protein kinases (SAP kinases). *The Journal of biological chemistry*, 272(12), pp.7727–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9065432 [Accessed March 9, 2015].

Wittkowski, H. et al., 2008. S100A12 is a novel molecular marker differentiating systemic-onset juvenile idiopathic arthritis from other causes of fever of unknown origin. *Arthritis and rheumatism*, 58(12), pp.3924–31. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2680303&tool=pmcentrez&rendertype=abstract [Accessed November 25, 2014].

Woo, P. et al., 2000. Randomized, placebo-controlled, crossover trial of low-dose oral methotrexate in children with extended oligoarticular or systemic arthritis. *Arthritis and rheumatism*, 43(8), pp.1849–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10943876 [Accessed December 28, 2014].

Woo, P., 2006. Systemic juvenile idiopathic arthritis: diagnosis, management, and outcome. *Nature clinical practice. Rheumatology*, 2(1), pp.28–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16932649 [Accessed November 11, 2014].

Woo, P. & Humphries, S.E., 2013. IL-6 polymorphisms: a useful genetic tool for inflammation research? *The Journal of clinical investigation*, 123(4), pp.1413–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3613933&tool=pmcentrez&rendertype=abstract [Accessed December 3, 2014].

Wouters, C., Ceuppens, J. & Stevens, E., 2002. Different circulating lymphocyte profiles in patients with different subtypes of juvenile idiopathic arthritis. *Clinical and Experimental Rheumatology*, 20(2), pp.239–248.

Wu, Q. et al., 2014. Immunological characteristics and T-cell receptor clonal diversity in children with systemic juvenile idiopathic arthritis undergoing T-cell-depleted autologous stem cell transplantation. *Immunology*, 142(2), pp.227–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24405357 [Accessed November 11, 2014].

Wulffraat, N.M., 2003. Reduced perforin expression in systemic juvenile idiopathic arthritis is restored by autologous stem-cell transplantation. *Rheumatology*, 42(2), pp.375–379. Available at: http://rheumatology.oxfordjournals.org/content/42/2/375 [Accessed December 28, 2014].

Yandell, M. et al., 2011. A probabilistic disease-gene finder for personal genomes. *Genome research*, 21(9), pp.1529–42. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3166837&tool=pmcentrez&rendertype=abstract [Accessed July 15, 2014].

Yeh, P.-C. et al., 2012. RED, a spindle pole-associated protein, is required for kinetochore localization of MAD1, mitotic progression, and activation of the spindle assembly checkpoint. *The Journal of biological chemistry*, 287(15), pp.11704–16. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3320919&tool=pmcentrez&rendertype=abstract [Accessed March 8, 2015].

Zhang, X. et al., 2014. The truncate mutation of Notch2 enhances cell proliferation through activating the NF-κB signal pathway in the diffuse large B-cell lymphomas. *PloS one*, 9(10), p.e108747. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4196756&tool=pmcentrez&rendertype=abstract [Accessed February 24, 2015].

Zhang, Y. et al., 2014. A160: Role of Interleukin-1 in Abnormal Monocyte Phenotype in Systemic Onset Juvenile Idiopathic Arthritis. *Arthritis & Rheumatology*, 66, pp.S207–S208. Available at: http://doi.wiley.com/10.1002/art.38586 [Accessed December 2, 2014].

Zhou, J. et al., 2013. Natural killer cell activity and frequency of killer cell immunoglobulin-like receptors in children with different forms of juvenile idiopathic arthritis. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology*, 24(7), pp.691–6. Available at: http://www.readcube.com/articles/10.1111/pai.12130?r3_referer=wol&tracking_action=preview_click&show_checkout=1 [Accessed December 28, 2014].