INVESTIGATING THE FUNCTIONAL CONSEQUENCES OF HETEROGENEITY WITHIN IMMUNE CELL POPULATIONS

A Dissertation

Presented to

The Academic Faculty

by

Linda Elisabeth Kippner

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the

School of Biomedical Engineering

Georgia Institute of Technology

Emory University

May 2017

INVESTIGATING THE FUNCTIONAL CONSEQUENCES OF HETEROGENEITY WITHIN IMMUNE CELL POPULATIONS

Approved by:

Dr. Melissa L Kemp, Advisor
School of Biomedical Engineering
Georgia Institute of Technology

Dr. Peng Qiu
School of Biomedical Engineering
Georgia Institute of Technology

Dr. Greg Gibson
School of Biology
Georgia Institute of Technology

Dr. Rabindra Tirouvanziam
School of Medicine
Emory University

Dr. Manu O Platt
School of Biomedical Engineering
Georgia Institute of Technology

Date Approved: March 29, 2017

For Rick and Katja

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr Melissa Kemp, for her guidance and support, and for giving me the opportunity to pursue this degree. She has been a wonderful PI and mentor and I am very grateful. I would also like to thank my committee members, Dr Greg Gibson, Dr Manu Platt, Dr Peng Qiu, and Dr Rabindra Tirouvanziam, for their expertise and valuable input. I would like to thank the members of the Kemp lab, past and present, for contributing to a scientific and non-scientific environment that have made the past years very enjoyable. In addition, I would like to thank members of the Gibson lab, in particular Jinhee Kim and Dalia Arafat, as well as the staff in the Petit institute core facilities, especially Nadia Boguslavsky, Sommer Durham, and Aaron Lifland, for their technical assistance. I would also like to thank Shannon Sullivan and Shannon Barker, for their administrative support and help throughout my time in grad school.

I am thankful that I have had a good role model in my mother, who led by example and showed me that anything is possible as long as you are willing to put in the work. I would also like to thank my father, who told me that things always work out in the end, though not necessarily the way you planned them. That wisdom holds true for both life and science. Finally, I would like to thank my husband and daughter for making it possible for me to pursue my dreams and for the support they have provided me. I could not have completed this work without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AIC	Akaike Information Criterion
AUC	Area Under the Curve
Ct	Cycle Threshold
Cs	Surface Complex
DAPI	4',6-diamidino-2-phenylindole
ERK	Extracellular-regulated Kinase
GFP	Green Fluorescent Protein
GUI	Graphical User Interface
HBSS	Hanks' Balanced Salt Solution
IL-2	Interleukin 2
IL-2Rα	Interleukin 2 Receptor Subunit Alpha
IL-2Rβ	Interleukin 2 Receptor Subunit Beta
IL-2Rγ	Interleukin 2 Receptor Gamma Subunit
JAK	Janus Kinase
k	Cluster Count
LOD	Limit of Detection
NFκB	Nuclear Factor Kappa-light-chain-enhancer of Activated B Cells
NK	Natural Killer (cell)
ODE	Ordinary Differential Equation
PCR	Polymerase Chain Reaction
PDMS	Polydimethylsiloxane
psi	pounds per square inch
PVDF	Polyvinylidene fluoride
STAT5	Signal Transducer and Activator of Transcription 5
WGA	Wheat Germ Agglutinin

SUMMARY

A well-functioning immune system is the result of the combined efforts of a multitude of different cell types, performing separate but coordinated tasks. The development of new techniques that enable the study of individual cells in greater detail ever before has shown that cellular heterogeneity exists such that prior classifications of immune cell types simply represent coarse binning across a spectrum of phenotypes. Diversity within these historically categorized "cell types" has functional consequences and deeper characterization of cell heterogeneity within immune cell populations is crucial for a better understanding of immune functionality.

The challenges of obtaining such a detailed understanding can be divided into two problems, namely how to obtain single-cell data in a high-throughput manner and how to analyze the data in order to correctly interpret the biological reality. In this work, we approached both these questions by taking a multilayered approach to single-cell analysis.

In two studies, we took advantage of novel microfluidics based techniques to collect data of gene expression and cell response. Gene expression analysis was performed on primary neutrophils and T cells, populations representing the innate and adaptive parts of the immune system. Through parallelized microfluidics-based assays, we collected information on a wide range of gene targets, including not only traditional surface markers but also intracellular signaling components and other functional markers. Our results show that using single-cell gene expression analysis, it is possibly to identify subgroups within cell populations and that the prevalence of these subgroups differs between individuals.

A microfluidic cell trap allowing for the capture and immobilization of non-adherent cells was coupled with time lapse fluorescent microscopy and tightly controlled fluctuations of extracellular cytokine concentrations. This allowed us to investigate single-cell T cell response to IL-2 under physiologically relevant conditions. In combination with computational modeling, this allowed us to interrogate the existence of functional subgroups within immune cell populations and the functional consequences of single-cell heterogeneity.

The question of how to correctly process and analyze single-cell data is as of yet unanswered, despite the fact that such analysis is crucial to correct interpretation of the biological information that such data can reveal. Established methods for bulk sample analysis, such as normalization to housekeeping genes, are unsuited to single-cell analysis due to the inherent noisiness of single-cell gene expression data. In this work, we took a systematic approach to analysis of single-cell gene expression data from primary neutrophils and T cells. Our results affirm the importance of choosing an analytical method that is well suited to the data, especially the method for data normalization.

In summary, we took advantage of emerging techniques for single-cell analysis to investigate the existence and functional consequences of single-cell heterogeneity within immune cell populations. By using a multidisciplinary approach incorporating molecular biology, statistics, computational modeling, and microfluidics, we were able to detect the existence of different subgroups within neutrophil and T cell populations, adding new insight into the effects of single-cell heterogeneity within the immune system.

CHAPTER 1 INTRODUCTION

The response of our immune system is produced by a combination of many different functional subsets of cells, working together to enable a systems response against a vast range of threats[1]. The development of new techniques for single-cell analysis has shown that even within populations of the same immune cell type, variability at the single-cell level exists and has functional consequences. With this in mind, it becomes clear that detailed studies of single-cell heterogeneity are essential for complete understanding of population-level functions. The importance of interleukin 2 (IL-2) as a regulation of T cells is well documented at the population level, but has been less studied in the context of single-cell response on short time scales. Furthermore, predictions that the IL-2 receptor system can enable cells to respond to physiologically relevant fluctuating levels of cytokine pose an intriguing possibility that a single-cell approach can help to verify. The main objectives of this research were to use a combined approach to investigating heterogeneity in immune cell populations by 1) taking advantage of emerging techniques, 2) combining this with systems modeling, and 3) investigating heterogeneity in cell responses to time dependent immunological stimulus under physiological conditions. To address these research objectives, this dissertation has three specific aims, which are:

1.1 Research Objectives and Specific Aims

Aim 1: Investigate the effect of data analysis methods on single-cell gene expression data from primary immune cells. This aim sought to develop a protocol for obtaining single-cell gene expression data from primary immune cells, and furthermore to evaluate the effect of data analysis methods on single-cell gene expression data interpretation. Finally, it aimed to assess whether distribution of cellular subtypes as defined by gene expression varies among individuals. Data was collected using microarray quantitative real-time PCR on single T cells and neutrophils from healthy donors. Once obtained, the

data were analyzed using a combination of analytical methods, and the results compared. *The working hypothesis was that subclasses of immune cells would be defined not only by traditional surface markers but also intracellular signaling components and other functional markers.* This work is presented in Chapter 3.

Aim 2: Develop a computational tool to investigate the effect of time dependent IL-2 stimulus on T cell response. This aim sought to use computational modeling to understand the dynamic response of T cells to IL-2 input. By using computational tools, we could investigate this in a setting that simulates biologically relevant conditions regarding cytokine concentration, in an approach that allows for the incorporation of the complexity of cytokine receptor subunit heterogeneity and its contributions to cytokine uptake and degradation. This work is presented in Chapter 4.

Aim 3: Quantify the variability of phenotypic responses to time-dependent immunological stimulus. In this aim, the predictions made in Aim 2 were tested experimentally by stimulating cells with cytokine and tracking their response. By using a microfluidic cell trap in combination with fluorescent microscopy, individual cells could be tracked over time and their responses quantified. In addition, the cytokine input that the cells received could be tightly regulated, allowing for matching experimental conditions to simulated conditions from Aim 2. The work from this aim is presented in Chapter 5.

1.2 Significance of results

Single-cell heterogeneity has recently emerged as an area of interest due to the development of new tools that enable analysis of individual cells in a high-throughput fashion. The resulting information has brought with it the understanding that cellular heterogeneity can have important functional consequences, both for the function of healthy biological systems and in disease states. Within the immune system, the

discovery of new functional subgroups could further our understanding of how this intricate network of cells functions to ensure our safety from external threats.

In this work, we took a multilayered approach to investigating the functional consequences of single-cell heterogeneity within immune cell populations. We made use of single-cell gene expression to collect data from primary immune cells and systematically evaluated methods for analyzing these single-cell data. Our results show that using high-throughput single-cell gene expression techniques, it is possible to identify subgroups within immune cell populations based on markers that are not limited to the traditional surface markers. In addition, our systematic testing of data analysis methods for inclusion and normalization on the acquired data highlight the importance of choosing analytical methods in order to correctly interpret the biological information from single-cell data. In particular, our results show that the choice of data normalization method will have significant impact on data interpretation. As traditional bulk sample methods of data normalization are unsuitable for single-cell data due to the inherent noisiness of such data, our conclusion adds further weight to the importance of determining the best alternative.

In order to explore the functionality aspect of single-cell heterogeneity, we developed a computational model that enabled us to explore single-cell response to pulsatile stimulus. We used this model to investigate the effects of reported variability in expression of IL-2 receptor subunits within T cell populations. By using a computational approach, we were able to remove effects of receptor competition and focus on effects of variability in the levels of subunits that are shared by other receptors. By using pre-primed Jurkat cells as our model organism, we also circumvented the effect of IL-2 stimulus on initializing receptor upregulation. Our model predicted that in for primed T cells, the shared IL-2Rβ and IL-2Rγ subunits become limiting factors in cellular response to IL-2. In addition to controlling specific aspects of the IL-2 receptor, we also used our model system to interrogate cellular response to pulsatile IL-2 stimulus in a

physiologically relevant concentration range. Our model predicted that cell-to-cell variability in receptor subunit levels would have an effect on cellular response under fluctuating conditions, suggesting ranges of pulsatile input and recovery times that would be of particular interest to investigate experimentally.

Finally, we used a microfluidic cell trap capable of capturing suspension cells to experimentally investigate single-cell response to pulsatile IL-2 stimulation, using the results from the computational model to inform the choice of input. As with our computational model, we used Jurkat cells as our experimental cell type, confirming experimentally that these cells exist in a pre-primed state under normal culture conditions such that initial activation of receptor expression with IL-2 was unnecessary for cell response. By combining the microfluidic cell trap with carefully controlled IL-2 input, we replicated our model-predicted pulsatile IL-2 ranges of interest in an *in vitro* setting. Using fluorescent microscopy, we analyzed individual cells longitudinally to determine translocation of STAT5, an indication of cellular response that is directly downstream in the IL-2 receptor signaling pathway. Our results show variability of response to IL-2 within Jurkat populations and suggest both that T cells exist in different states of response readiness prior to IL-2 stimulation and that T cells have favored ranges of pulsatile IL-2 input where the population response is stronger. Future work can further investigate this ideal frequency regime for T cell response to IL-2 in the context of biological relevance. Furthermore, the concept of preexisting cell states within immune cell populations could be explored by employing single-cell gene and protein expression analysis. This would allow for further characterization of cellular subgroups to define preexisting cellular states within the population that result in the observed behavior.

CHAPTER 2 BACKGROUND

2.1 Cell Heterogeneity in the Immune System

The hypothesis that populations of cells contain functional subpopulations has generated interest in various biological and pharmacological contexts [2-4]. Such subgroups can be defined by cell-to-cell heterogeneity at the genetic level, but also by differences in gene expression or epigenetic modifications, and result in functional differences both within healthy cell populations and in disease. Within the immune system, cell-to-cell heterogeneity enables the system as a whole to fulfil its role of protecting the organism from both external and internal threats. The immune response mounted at the population level is the result of a combination of many different functional subsets of cells, working together to enable a systems response against a vast range of threats [1]. The immune system also needs to maintain immunological memory as well as tolerance towards self-antigens in order to prevent autoimmunity. In order to fulfil these many demands, diversity within immune cell populations is crucial. It enables a wide range of recognition and response, both by the existence of immune cells with different basic functions, and by maintaining a repertoire of cells with different antigen recognition properties. Heterogeneity in cellular expression can also enable complex population-level behaviors to emerge from simple single-cell decisions. For example, placing different thresholds on cellular activation can translate a digital decision at the single-cell level to a graded analog response at the population level [5, 6]. While classification of immune cell subpopulations has traditionally been performed based on differences in surface markers (a straightforward but incomplete system of classification), recent development of single-cell analysis techniques has enabled the discovery of hitherto unknown subgroups based on more detailed analysis.

2.2 Cytokines and Cell-To-Cell Communication

Cell-to-cell communication within the immune system is essential to enable transmission of information between diverse cell populations such that individual cell responses can be propagated or suppressed. This enables coordination of systemic responses, as well as to tolerance and protection against autoimmunity and immune deficiency [7-10]. One of the ways in which immune cells communicate is through the use of cytokine signaling molecules (Table 1), small proteins which are produced and secreted by a variety of cell types. Cytokines bind to surface receptors to mediate intracellular signaling cascades in the target cell, inducing subsequent up- or downregulation of genes. This in turn results in further production of cytokines and expression of surface receptors and regulation of the cell's cytokine response by feedback inhibition. Cytokines play an essential role as regulators of immune function and cell proliferation and can act in endocrine, paracrine, and autocrine fashion to modulate cell behavior. They include the protein family of interleukins, the majority of which are produced by helper T cells, among other cells.

Table 1: Cytokines in cell-cell communication within the immune system

Consumer / **Producer**	**CD4 + T cells**	**CD8 + T cells**	**B cells**	**Neutrophils**	**Dendritic cells**	**Macrophages/ Monocytes**	**NK cells**	**Eosinophils**	**Mast cells**
CD4+ T cells	IL-2 IL-4 IL-10 IFNγ TNFβ TGFβ	IL-2 IL-4 IFNγ TNFβ TGFβ	IL-4 IL-5 IL-6 IL-13 IL-20 TNFβ IFNγ	IL-6 IL-17A IL-17F TNFβ IFNγ		IL-2 IL-3 IL-4 IL-10 IFNγ TNFα TGFβ	IFNγ	IL-5	IL-4 IL-13

Table 1 continued.

Consumer / Producer	CD4 + T cells	CD8 + T cells	B cells	Neutrophils	Dendritic cells	Macrophages/ Monocytes	NK cells	Eosinophils	Mast cells
CD8+ T cells		IL-2		TNFα IFNγ					
B cells	IL-5 IL-6 IL-13 IL-20 TNFβ IFNγ								
Neutrophils	IL-1 TNFβ	TNFα IFNγ	IL-1	IL-2 IL-6 IL-8		IL-1 IL-2 IL-6 IL-8			
Dendritic cells	IL-6 IL-12 TGFβ	IL-1 IL-15					IL-15		
Macrophages/ Monocytes	IL-1 IL-6 IL-10 IL-12	IL1 IL-6 IL-12 TNFα IL−15	IL-4 IL-6	IL-6 IL-8 TNFα		IL-6 IL-10	IL-15		
NK cells	IFNγ				IL-15				
Eosinophils	IL-1 IL-5 IL-6	IL-1 IL-6	IL-6 IL-8						

2.2.1 The Role of IL-2 in T cell Functionality

Interleukin-2 (IL-2) was discovered in 1976 [11] and first characterized as the growth factor responsible for T cell proliferation [12-16]. It has been extensively studied in the four decades since and shown to have a wide range of additional functions on both T cells and other immune cells (Table 1). As the first interleukin discovered, IL-2 ushered in a shift in immunological research towards the understanding of how small molecules such as interleukins function as major transmitters of signals in intercellular communication. Produced and secreted by T cells themselves, IL-2 is an essential part of a functioning immune system, most notably through its role as a regulator of T cell responses driving immunity and protecting against autoimmunity [7, 17-21]. IL-2 regulates T cell differentiation into effector and memory T cells in both an autocrine and paracrine fashion [7, 22] and induces proliferation upon interaction with IL-2 specific receptors expressed on the cell surface [23]. In addition to inducing upregulation of its own receptor, IL-2 functions as a control for other cytokines through modulation of receptor levels. While IL-2 affects the expression levels of its own receptor, it also increases receptor expression for IL-4 and IL-12 [24], and decreases gp130 [25]. Because of this, IL-2 modulates not only cellular responses to itself but also to IL-4, IL-12, and IL-6, other interleukins that affect the differentiation of T helper cell subsets, which allow the immune system to mount effective antigen-specific responses. By repressing expression of the IL-7 receptor, IL-2 suppresses cell survival signals in activated T cells, thus contributing to regulation of activation-induced cell death [26]. Through its many roles in intercellular communication, IL-2 has shown to be a major regulator of T cell functionality; however, the contributions of single-cell heterogeneity to IL-2 response in T cells have yet to be fully elucidated. One such source of variability lies in the nature of the IL-2 receptor [27], which is a heterotrimeric complex where the different subunits (the IL-2 specific IL-2Rα subunit and the shared IL-2Rβ and IL-2Rγ subunits) exist in varying levels on individual cells. This variability in the expressed levels of subunits

indicate that differences in IL-2 response capabilities exist between T cells within a population. The shared nature of two of the three subunits in a biological system introduces receptor competition which adds further complexity to the contributions of receptor-level variability to functional differences at the single-cell level. By studying the variability at a subunit level in a system devoid of receptor competition, we can investigate the effects that the individual IL-2 subunits have on T cell function. As IL-2 contributes to T cell function, modulation of IL-2 response is of considerable therapeutic interest, and renewed interest has been seen in investigating aspects of IL-2 functionality for this purpose. This has been attempted by the use of antibodies targeted against IL-2Rα, which can damped undesired IL-2 effects [28-30], and more recently by the use of IL-2 analogues such as superkines with increased affinity for IL-2Rb (Levin 2012). Such superkines can be used to activate downstream response to IL-2 without the need for IL-2Rα expression. The use of IL-2 superkine takes advantage of the heterotrimeric nature of the IL-2 receptor, where the IL-2Rα subunit does not have any effect on downstream signaling events as IL-2Rβ and IL-2γ do. In addition to IL-2 binding events, the subunits are subject to different intracellular trafficking fates upon internalization, a feature that contributes further opportunities for modulation of IL-2 function.

Cellular Response to Ligand Fluctuations

In order to mount a finely tuned response to extracellular cues, cells have the ability to respond not just in a binary fashion to bulk input of stimulus, but also in a more tunable manner to subtle fluctuations. Examples of physiologically relevant stimuli that are released or interpreted in a pulsatile manner include neurotransmitters [31], hormones [32], and cytokines [33]. Cells are able to detect rapid dynamic changes in concentration of stimulus in their microenvironment, and respond to time-varying ligands below saturation levels. Growing interest has developed in studying cellular responses to

pulsatile delivery of stimulus input, which can result in downstream temporal effects, such as different timing of transcription factor activation [34]. It has been suggested that pulsatile stimulation plays a role in a range of biological systems, such as TGF-β signaling during embryonic development [35] and growth factor-induced ERK signaling [36]. Such time-dependent stimulus can result in different cellular response profiles compared to continuous input as well as different responses for different pulsatile input metrics, such as stimulus pulse length and recovery time [37]. In a biological system, this can enable cells to distinguish between variations of input in a relevant dynamic range. Studies also indicate that such behavior enables cells to act as bandpass filters, filtering out fluctuations outside of this range as noise that does not induce cellular response. For T cells, such behavior has been seen in response to extracellular fluctuations of H_2O_2, which induces downstream shifts in the intracellular messenger Ca^{2+} in a manner showing T cell sensitivity to stimulus frequency [38]. Of particular interest is the response in systems with characteristics that enable cells to distinguish between fluctuations in extracellular cues at below equilibrium levels, namely downstream processes that are transient and faster than the time required to reach equilibrium for the upstream receptor-ligand interaction. This has been proposed as allowing for greater sensitivity to rapid changes in the extracellular environment [39]. The IL-2 receptor system exhibits such characteristics [40], and has been suggested as a potential system where such pre-equilibrium sensing takes place [39]. How T cells respond to oscillatory input of IL-2 at the single-cell level remains an important unanswered question. Of particular interest is exploring the effect of features of IL-2 fluctuation, such as the duration of input pulses, the length of recovery time, and if changes in these metrics correspond to altered cell behavior.

2.3 Techniques for Analyzing Single-Cell Heterogeneity

While bulk analysis methods can generate insights into population-level behavior of cells, single-cell techniques allow for more detailed knowledge of the complexity of biological systems by elucidating the effects of cell-to-cell heterogeneity. In the past few years, development of new technologies has dramatically increased the tools available for single-cell analysis at the genomics, transcriptomics, and proteomics levels. These new techniques enable the collection of information on more targets such as surface markers [41] or intracellular [42, 43] or secreted proteins [44]. These expanded capabilities present exiting benefits for research in healthy and diseased systems, as well as for drug development.

2.3.1 Single-Cell Gene Analysis Techniques

Gene expression analysis was previously limited to population and sample averages though the use of bulk samples. The development of techniques that allow for gene expression analysis and genome sequencing in single cells now lets us expand our understanding of the contributions of variability at the gene level to cellular heterogeneity [45]. This allows for the exploration of cell variability in populations, revealing low-abundance cell types and transitional cell states, and enabling the investigation of functional consequences of cellular heterogeneity both in healthy cell populations and in diseases [46]. The PCR based RNA-seq method which allows for exploration of gene expression heterogeneity is rapidly expanding the capability in both target and cell numbers [47]. An alternate quantitative method, CytoSeq, was recently developed by Fan et al. and relies on isolation of individual cells into microwells, allowing for interaction with barcoded beads labeled with oligonucleotide primers [48]. This allows for the amplification procedure for each cell to occur separately in order to limit contamination. Two microdrop-based methods, DropSeq [49] and InDrop [50], instead attempt cell sample isolation by separating out individual cells into droplets though the use of

microfluidic flow control. Although there has been a rapid development of new and improved techniques for acquisition of data, the question of how to best analyze single-cell data remains unresolved. Single-cell data analysis methods need to capture biological information while accounting for technical noise. A bulk sample approach which assumed every cell to be an average representation of the population is unsuitable for this purpose, as it does not adequately take into account the innate cell-to-cell fluctuations of both target and housekeeping genes.

2.3.2 Computational Modeling of Single-Cell Cytokine Response

Computational modeling is an attractive approach to studying biological systems is, as it allows for researcher-determined manipulation of system properties. In the study of single-cell functionality, computational models enables targeted investigation of characteristics of single-cell dynamics, and this has been applied to a range of questions in the area of single-cell cytokine response, including digital signal processing, molecular mechanisms and signaling pathway features of cytokine response, and the diffusion of cytokine signals during cell-cell communication. A historically rich pathway for computational modeling of signaling regulation is the NF-kB pathway. A model developed by Tay et al. to study downstream localization of NFκB in response to TNFα stimulation found that, contrary to what a population average indicated, not all cells in a population responded to the stimulus [5]. While analog dose dependency could be detected at the population level, the single-cell model also allowed for the reproduction of digital signaling features made possible by single-cell variability. A different model developed for the NFκB signaling network in macrophages predicted molecular mechanisms that contribute to the single-cell variability upon integration of upstream MyD88 and TRIF inputs at NF-kB activation [51]. By incorporating variability of cellular mechanisms at the single-cell level, this model was able to quantitatively match observed ranges of *in vitro* responses.

Single-cell modeling has recently been extended to the IL-2 signaling system. A computational modeling study identified how IL-2 feedback loops at the single-cell level can result in a scalable population-level response, allowing for an immune response that varies with antigen dose [52]. Other modeling approaches have focused on the addition of three-dimensional spatial description to a model of cytokine response and modulating extracellular IL-2 levels to elucidate the importance of IL-2 gradients in cell-cell communication in a physiological setting [53]. The ability to precisely manipulation single-cell properties and stimulus input makes computational modeling an inviting tool for studying single-cell behavior in the context of response to pulsatile IL-2 stimulus.

2.3.3 Methods for Investigating Single-Cell Functionality

For experimental investigation of single-cell functionality, a proteomics approach is useful for detecting heterogeneity through collection of quantifiable single-cell data. Fluorescence microscopy is a commonly used technique for single-cell analysis in both live and fixed cells. It enables the detection of both cell surface and intracellular targets using a variety of methods such as antibody staining and fluorescently labeled proteins. In addition, it has the benefit of allowing for quantitative analysis of images, enabling both spatial and time-dependent analysis. While the benefits are many, single-cell microscopy works best for adherent cell types, which can be easily immobilized. Non-adherent cells such as T cells are commonly studied using alternative techniques such as flow cytometry, which allows for detection of intra- and extracellular targets in a high-throughput fashion. This enables single-cell measurements without the need for precise cell localization; however, it provides only a single snapshot in time of each cell and not dynamic information of individual cells. Single-cell mass cytometry assays [41] allows for the collection of greatly expanded single-cell information but is likewise limited to one time point per cell. In order to overcome the difficulty of obtaining single-cell time

course measurements for non-adherent cells, these technical difficulties need to be overcome, and the use of microfluidic cell traps offers a solution for collecting quantitative single-cell information from live cells in a dynamic setting. A microfluidic device capable of immobilizing non-adherent cells allows for capture of dynamic information through fluorescent imaging and can be used to deliver desired stimulus in order to investigate single-cell response in a direct fashion. In contrast with macro-scale systems, the use of microfluidics enables precise manipulation of the cellular microenvironment [54]. Whereas the large volumes of macroscale systems hamper rapid and uniform changes extracellular concentrations, microfluidic devices can be engineered to enable simultaneous delivery of stimulus to cells. The microscale volumes and tight control of flow possible microfluidics allow for rapid change of concentration of ligands of interest in the extracellular space, mimicking physiologically relevant dynamic changes [55]. This make a microfluidics approach well suited for delivery of time-varying stimuli.

2.4 Motivations for Research

The motivation for this research was to explore cell-to-cell variability between phenotypic subpopulations existing within immune cell populations by combining emerging single-cell techniques such as single-cell gene expression analysis, computational modeling, and microfluidic techniques. The goal was to enable a deeper understanding of functional cell heterogeneity within the immune system.

CHAPTER 3 SINGLE-CELL TRANSCRIPTIONAL ANALYSIS AS A TOOL FOR DETECTING IMMUNE CELL SUBGROUPS

*This chapter was adapted from Kippner, L.E., et al., Single cell transcriptional analysis reveals novel innate immune cell types. PeerJ, 2014. **2**: p. e452. [56]*

3.1 Introduction

It is becoming evident that established methods, whereby averaging population data essentially assumes that all cells within a population are equivalent, are vastly oversimplifying cell functionality and obscuring the presence of cellular subtypes [57]; however, a more detailed analysis has been hindered by technical limitations. Previously, transcription analysis has been constrained to population averages, due to the inability to quantify single-cell levels of mRNA with existing techniques, such as such as Northern blotting or classical qRT-PCR [58-62]. Major technical advances in single-cell measurement systems have now enabled the investigation of such cell-level information [63-66]. These advances include high-throughput nanoscale real-time PCR, which allows for mapping of transcriptional profiles by highly parallelized assays enabled by microfluidics.

Standard methods for processing qRT-PCR data are well established; however these methods are based on population-averaged data and it cannot be taken for granted that the same approaches are optimal for single-cell data. Indeed, single-cell gene transcripts have been shown to follow log normal distribution curves [67]; thus, mean population averages are heavily influenced by a few cells showing relatively high expression levels. As single-cell data is inherently noisy, this must be taken into account when choosing analytical methods. For example, housekeeping genes show considerable variability of expression at the single-cell level such that standard methods of data normalization based on such genes should not be used [68]. In addition, single-cell

measurements exhibit noise due to technical variability and this must ideally be accounted for without losing variability due to biological function, which is often at comparable levels. A particularly important consideration is whether the complete absence of signal is due to lack of expression, or to stochastic technical failure. All analytical approaches make assumptions regarding this issue that could have a major impact on the conclusions derived from different modes of analysis [68].

The biological motivation for the current study was to assess gene expression variability among single leukocytes, and whether the prevalence of distinct sub-types (as defined by gene expression) varies among individuals. Neutrophils and T cells were selected as representatives of the innate and adaptive branches of the immune system, respectively. Recent studies have revealed a close correlation of functional phenotype to transcriptional profile [69-71], and we hypothesized that our results would yield immune cell subclasses separated not only by traditional surface markers, but also by intracellular signaling components, as well as other functional markers. As bimodality in expression of individual transcripts may be an indicator of functional heterogeneity [72], we further asked whether cellular subclasses were defined by shared bimodality of multiple transcripts between cells. To that end, we performed gene expression pattern analysis and hierarchical clustering of our cell populations. We found that genes exhibiting bimodal distribution patterns were preferentially assigned to the same cell clusters in our data sets.

In overcoming the technical challenges of analyzing single-cell data, we found that the decisions made in data processing can have dramatic consequences for the interpretation of cellular subpopulations. We systematically explore and recommend approaches that can be used in order to consistently analyze multiple single cells from multiple donor individuals across multiple genes. Nine alternate methods of data exclusion and normalization are considered, and their effect on secondary data analyses, such as hierarchical clustering, is assessed. Our results show that analysis and correct

interpretation of single-cell gene expression data is dependent on the method chosen for primary data analysis, specifically on the method chosen for data normalization [56].

3.2 Materials and Methods

To investigate single-cell gene expression profiles, single cell qRT-PCR was performed on primary cells from healthy donors. The workflow for data collection is outlined in Figure 1. This was followed by data processing in three steps, which are outlined in Figure 2.

3.2.1 Data Acquisition from Primary Immune Cells

Extraction and single-cell sorting of primary T cells and neutrophils from donor blood

Neutrophils and T cells were extracted from 5ml whole blood from 6 healthy donors and isolated based on surface marker expression by negative selection using antibody-coated magnetic beads (EasySep neutrophil extraction kit, Stem Cell Technologies, or Dynabeads for untouched T cells, Life Technologies). One donor's neutrophil count was too low for further processing, therefore all results presented for neutrophils correspond to n = 5. Negative selection was chosen so as to avoid cellular activation due to receptor cross-linking. For each purified cell type, flow cytometry sorting with a BD FACS Aria II gated by forward- and side scatter was utilized to deposit single cells into a 96-well PCR plate preloaded with 5 µl of lysis buffer with 0.05U Superase RNase inhibitor (Life Technologies) per well. The plates were centrifuged for 1 min at 1000 g in 4ºC and immediately frozen and stored at -80 ºC. All donors were individuals enrolled in The Center for Health Discovery and Well-Being at Emory Midtown Hospital and provided written consent for participation in the study. The protocol for blood collection was approved by the Georgia Tech Institutional Review Board (approval #H09364).

Single-cell Quantitative Real-Time PCR

The cellular lysates were converted to cDNA and 96 target genes per cell type were pre-amplified with a pool of 96 primer pairs targeting genes representing pattern recognition, cell-type markers, intracellular signaling, transcription, and immune response (Table 5**Error! Reference source not found.** and Table 6**Error! Reference source not found.**). For each donor, amplified cDNA samples from 48 cells of each type were then randomized and re-plated across 5 Fluidigm 96x96 microfluidic arrays, in order to avoid any plate effects confounding the analysis of single donors. Gene-specific quantitative real-time PCR reactions were performed using the Fluidigm BioMark I nano-scale platform. Negative controls (without cDNA) and samples of 10 and 100 cells were used as controls for single-cell loading. The mean difference in Ct value between 1 and 10 cells and between 10 and 100 cells per sample was determined in independent assays, providing a measurable control for single-cell loading of each sample. To enable reproducible comparison of gene expression between qRT-PCR samples, data is usually normalized with respect to data obtained for an internal or endogenous reference gene. Housekeeping genes such as β-actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) are most often used because their expression levels are expected to remain constant. Unfortunately, single cells exhibit large heterogeneity in housekeeping gene expression levels, and this method cannot be used as control for reproducible comparison of gene expression between single cell samples [73, 74].

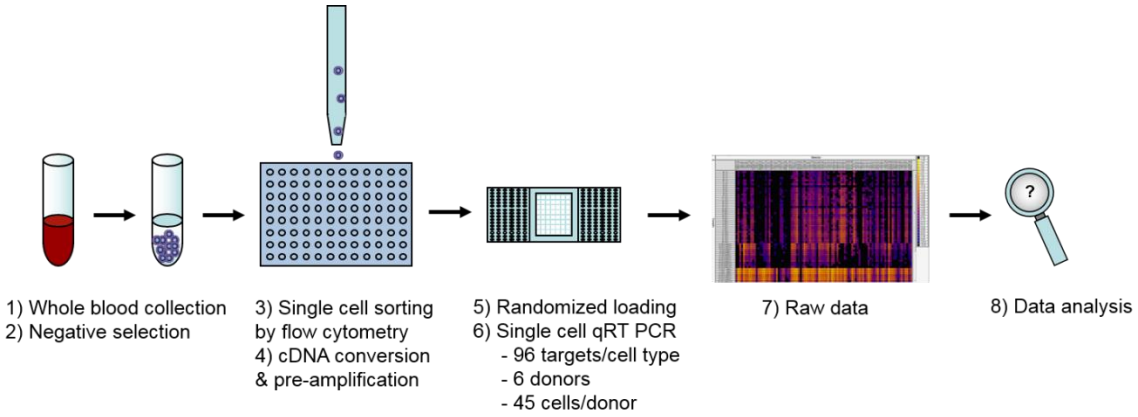

Figure 1. Data acquisition workflow for collection of single-cell gene expression data. Whole blood was collected from healthy donors, and negative selection used to isolate T cell and neutrophil populations. Single-cell sorting was then used to deposit one cell per well into 96-well plates, pre-loaded with lysis buffer. Following this, cDNA conversion and pre-amplification was done in plate, and resulting cDNA samples randomly loaded onto microfluidic arrays. qRT-PCR reactions were run simultaneously against 96 gene targets per cell. Raw data was obtained as Ct values.

3.2.2 Data Processing

Quality Control and Data Exclusion

Raw data for gene expression were obtained as Ct values between 1 and 40, with lower Ct value indicating higher abundance of a given gene-specific product. Missing data points were coded as Ct values of 999; such values can either be due to null or very low expression of the target gene in question or due to a failed reaction (truly missing data). Single missing measures may indicate technical failures, but consistent absence of a similar set of lowly expressed transcripts is more likely to imply coordinated loss of expression. Downstream methods differ largely with respect to how the missing data is handled. Three different sets of criteria were used for data exclusion for each of the two (neutrophil and T cell) data sets.

A) *Supervised Data Exclusion.* For the neutrophil data set, an empirical cutoff was set to transcripts present in at least 70% of cells, and subsequently to cell samples expressing at least 70% of these most uniformly expressed genes. We reasoned that the absence of the same set of genes in a common set of cells would imply true absence of expression, and used hierarchical clustering to provide a preliminary indication of such clusters of non-expressed genes. 23 such co-regulated low-abundance genes were identified, for which missing values were re-assigned a Ct value of 40 (the maximum number of cycles). Subsequently, for 36 genes, sporadic missing data was assumed to represent technical error and these values were reassigned to the average Ct for the gene in question in the data set. 34 genes were excluded in their entirety. Expression was evaluated for 59 genes in 202 cells. Because the T cell data set did not contain a natural cutoff for transcript presence, this method of analysis was not implemented for the T cell data.

B) *Data exclusion based on median standard deviation cutoff*. All missing data values were initially set to Ct 40, and the mean Ct and number of missing data points were calculated for all genes. The second and third highest expressed genes in the data set were selected and their mean Ct and standard deviation calculated. Note that the highest expressed gene in both the neutrophil and T cell data sets were treated as outliers and ignored for the purposes of calculating mean Ct, due to expression levels far higher than all other genes. Any gene whose average expression was within a cutoff of three standard deviations of the mean Ct value for the two chosen genes was included. All cells expressing less than half of these genes were then excluded. A plot of the maximum Ct across all cells for all 96 genes in the neutrophil data set showed a bimodal distribution of maximum Ct values, with a second peak starting at Ct 37 that corresponds to cells deemed not to express the target gene. The limit of detection (LOD) was thus set to Ct 37 for neutrophils, and the LOD Ct was set to 38 for T cells by the same methodology. All data values above LOD Ct, including Ct 999, were replaced with 37, and the LOD Ct value was then subtracted from all other Ct values, according to the Log_2EX method

(Log_2EX = LOD Ct – Ct [gene]). Consequently, the adjusted expression measure for this method is inverted and ranges from 0 to LOD Ct, with more highly expressed transcripts having higher values, more in line with intuition and with microarray or RNA-Seq data analysis. For the T cell data set, Ct values above LOD were interpreted and analyzed in two alternate ways; either as representing no expression of the target gene, with Ct values set to 0 or, alternatively, as missing data points due to technical error, with missing values replaced with average Ct for the gene (analogous to the supervised data analysis method used for neutrophils). Subsequently, entire cells were excluded, if the two gene targets with highest expression in our data set were more than three standard deviation units lower than the median. Additionally, any genes that were not expressed in any cell sample were excluded from the data set. For the T cell data set, 2 genes that were only expressed in one cell were also excluded from analysis. This resulted in the exclusion of 12 neutrophils and 31 genes in the neutrophil data set, and 7 T cells and 63 genes in the T cell data set. Expression was evaluated for 62 genes in 208 cells in the neutrophil data set and just 29 genes in 244 cells in the T cell data set.

C) *Inclusion of all data points*. All data points were initially included in analysis, with the exception of genes not expressed in any of the control samples (cDNA, tRNA, 10-cell samples). This excluded 12 genes from analysis in the neutrophil data set and 13 in the T cell data set. In addition, any transcripts missing from all samples in an array were excluded. This excluded 3 genes in the neutrophil set. LOD Ct was subtracted according to the Log_2EX method as described above. Expression was evaluated for 81 genes in 220 cell for the neutrophil data set and 85 genes in 247 cells for the T cell data set.

Data Normalization

Three different sets of criteria were used for data normalization for each of the data sets generated from the three methods for data exclusion.

1) *Mean centering*. The mean Ct value for each cell was calculated and subtracted from each data point for the same cell. This approach removes the dependence of magnitudes, allowing for easier visualization and comparison of relative differences in expression levels.

2) *Quantile normalization*. Gene expression data for each cell was re-ordered by raw Ct value, mean Ct values for each cell were calculated, and the original data was replaced by the average quantile. This method of rank-order analysis eliminates cell-to-cell differences in data density.

3) *Standardization of the genes.* Gene expression data were mean-centered for each cell, and then the values for each gene were standardized (converted to z-scores) by mean-centering and dividing by the standard deviation. Residuals from an ANOVA with Plate as the main effect were extracted. This method adjusts the distribution only of targets whose expression differs among plates. A further centering of residual expression values to a mean of zero for each cell ensures that no cells have artificially low or high expression of all genes.

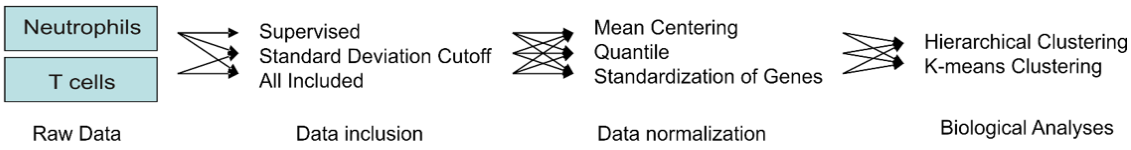

Figure 2. Data processing workflow for single cell gene expression data. For data processing, three methods were tested for data inclusion in combination with three methods for data normalization. Following this, the resulting nine data sets were analyzed for biological information by gene expression pattern analysis, detection of cellular subtypes by hierarchical clustering, and comparison of individual donor subtype representation.

3.2.3 Analysis of Gene Expression Patterns

The single-cell transcript abundance distribution for each gene was determined using SAS JMP Pro 10 (Cary, NC). Modality was assessed by Akaike information criteria (AIC) score, which was further verified by calculating deltaic, comparing scores of bimodality, trimodality and unimodality, as well as visual observation. Genes exhibiting bimodality were tracked and cluster membership was determined in the raw data set, as well as after data exclusion and normalization methods deemed most suitable, using the criteria above, namely exclusion by missing data cutoff and normalization by standardization of the genes. In addition, the number of cells included in each cluster was determined for each donor. Known gene product functionality was obtained from three databases: ToppFun [75] DAVID v6.7 [76, 77], and KEGG Pathway [78].

3.2.4 Analysis of Donor-to-donor Variability

For each donor, the cell count was determined for each of the cell clusters defined within the overall population. Following this, the observed frequencies were compared to expected frequency by Chi-square test comparison of the number of cells of each class in each of the five individuals relative to the expectation, assuming equivalent proportions.

3.2.5 Comparison of Primary Analysis Methods by Concordance of Cell Clusters

Combining the methods for data exclusion and normalization generated nine alternate sets of processed data for each of the two cell types. Each data set was organized by hierarchical clustering as well as k-means clustering by cell, resulting in cell clusters based on shared gene expression patterns. Concordance, defined as the percentage of cells ascribed to the same cluster, was compared between all combinations of analysis methods for both methods of clustering. For hierarchical clustering, data was clustered using Ward's minimum variance method [79], which minimizes the total within-cluster

variance. The k-means method of clustering aims to sort data into a pre-defined number of clusters, k, with each data point belonging to the cluster with the nearest mean [80]. K-means clustering was performed on all data sets with k values of 2 or 3 for both neutrophils and T lymphocytes. The k values were evaluated using Cubic Clustering Criteria (CCC) with external cluster validation. All computations were performed in SAS JMP-Genomics v5.0 (Cary, NC).

3.3 Results

3.3.1 Gene Expression Pattern Analysis

Bimodal Gene Expression Patterns in Neutrophils

Gene expression analysis of the raw neutrophil data revealed the existence of different expression patterns for genes, such as unimodal distribution of expression (Figure 3A), bimodal distribution (Figure 3B), and trimodal distribution (Figure 3C). Of these genes, all but one exhibiting bimodal distribution contained clear on/off expression patterns amongst the cells. The exception was DDX58, which showed one subgroup with higher and one group with lower expression, but no non-expressing cells. The existence of such non-expressing cells poses the problem of how to define these data points. One approach is to assign all such values the maximum Ct of 40, but this assumes that these data represent true missing expression; they could also result from technical errors due to failed PCR reactions. If the latter is the case, apparent bimodality with on/off expression patterns would in reality represent unimodal distribution with missing data points being technical artifacts instead of biologically relevant information. An alternative approach for addressing this issue is to look at patterns of missing data within the sets. If missing data points from the same genes tend to correlate within the cells, the cause is likely to be biological, suggesting that the populations contain cellular subgroups.

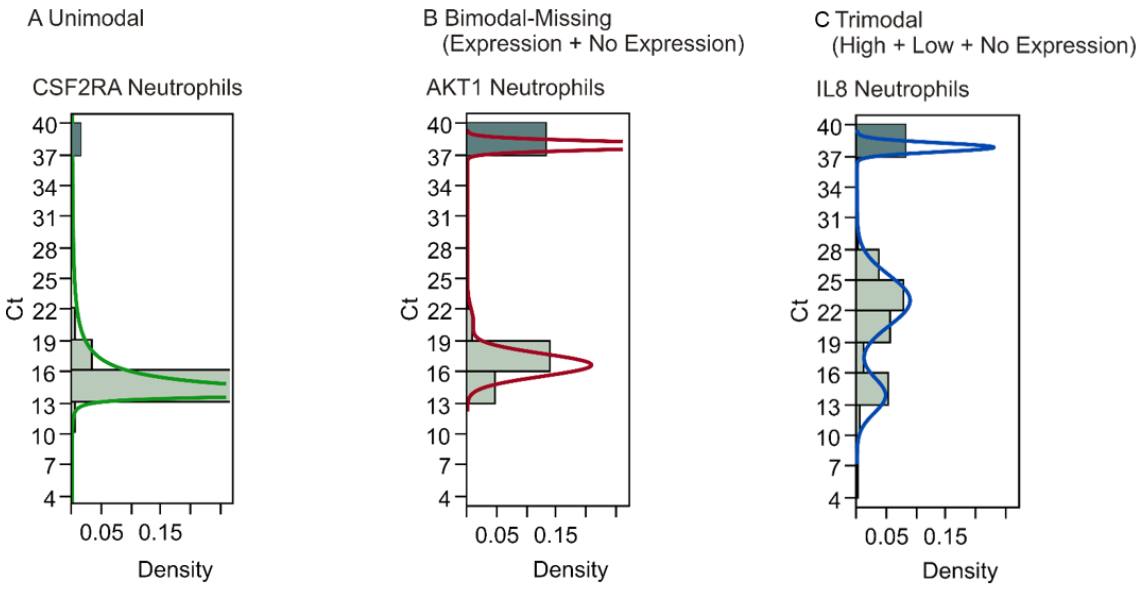

Figure 3. Gene expression analysis show bimodal expression patterns. Analysis of gene expression revealed varying patterns in both the neutrophil and T cell data sets. Examples from the neutrophil data set show A) Unimodal distribution, B) Bimodal distribution, and C) Trimodal distribution. A peak at Ct 40 indicates the existence of cells showing no expression of the gene.

Bimodal Gene Expression as an Indicator of Cellular Subclasses

In order to determine whether the existence of genes with bimodal expression patterns signaled the existence of cellular subclasses, the data was clustered based on shared gene expression patterns. Clustering showed that, for neutrophils, bimodal genes exhibiting on/off pattern tended to be off in the same cells, although they clustered together with unimodal genes implying that the differential expression between cell types is not restricted to bimodality (Figure 4A). Another potential cause for missing data points is low initial concentration of RNA in the sample, owing to inefficient RNA extraction, leading to complete loss of signal for the lowest-abundance genes that share the technical

inefficiency. In order to address this, we controlled for overall abundance of RNA by normalizing the data sets.

Bimodal Gene Expression Patterns in T cells

Similarly to the neutrophils, the T cell data set contained genes exhibiting bimodal gene expression. As seen in neutrophils, T cell genes with bimodal on/off expression patterns also tended to be interspersed with unimodally expressed transcripts (Figure 4B).

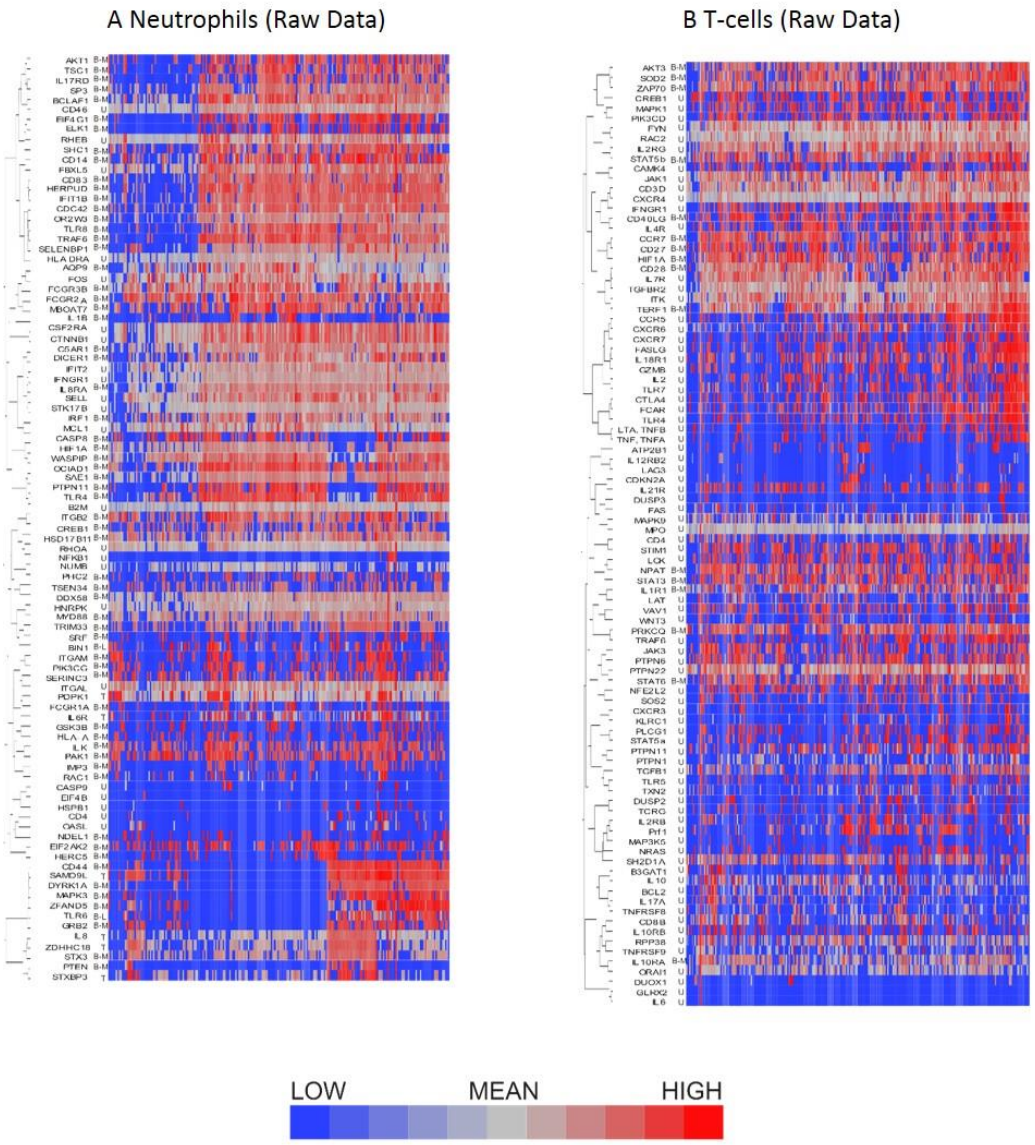

Figure 4. Hierarchical clustering of neutrophil and T cell data showed distinct sub-populations of cells characterized by shared patterns of gene expression**.** Hierarchical clustering of the pre-processed A) neutrophil and B) T cell data prior to data exclusion and normalization show bimodal genes preferentially clustering together. Bimodal genes are indicated by B-M, unimodal genes by U.

3.3.2 Detection of Neutrophil Subtypes

Hierarchical clustering was applied to the datasets using Ward's method, which has been shown to discriminate clusters efficiently on gene expression datasets [81, 82]. Figure 5 shows the results of hierarchical clustering with nine different methods combining the three methods for data exclusion and three methods for data normalization. The color coding (purple, green and orange) shows the degree of concordance of clustering relative to the method based on supervised data removal with mean centering (top left). Employing exclusion with any of the three methods, followed by either mean centering or quantile normalization, three clusters of neutrophils were observed consistently, with notable separation of the orange, and most of the green, clusters from the purple one. Concordance, defined as the percentage of cells assigned to the same cluster, ranged from 75% to 100%, *prima facie* supporting the presence of three cell types in our samples.

However, when a plate effect was fit to the standardized gene expression z-scores, only two major clusters were observed regardless of the data exclusion method (Figure 5G), and concordance of the two-way classification of orange versus green/purple cells was perfect. This analysis implies that a plate effect caused the splitting of the large purple/green clusters observed with the mean-centering and quantile normalization methods. That is to say, very low abundance gene expression led to loss of signal on one of the plates, generating an artificial signature of co-regulation of some cells. However, the orange cluster remains robustly detected by all methods. We conclude that there are two main clusters of cell types in neutrophils. There is also a hint of a sub-type within the orange cells defined by differential expression of a half-dozen genes, but a larger sample will be required to validate this inference.

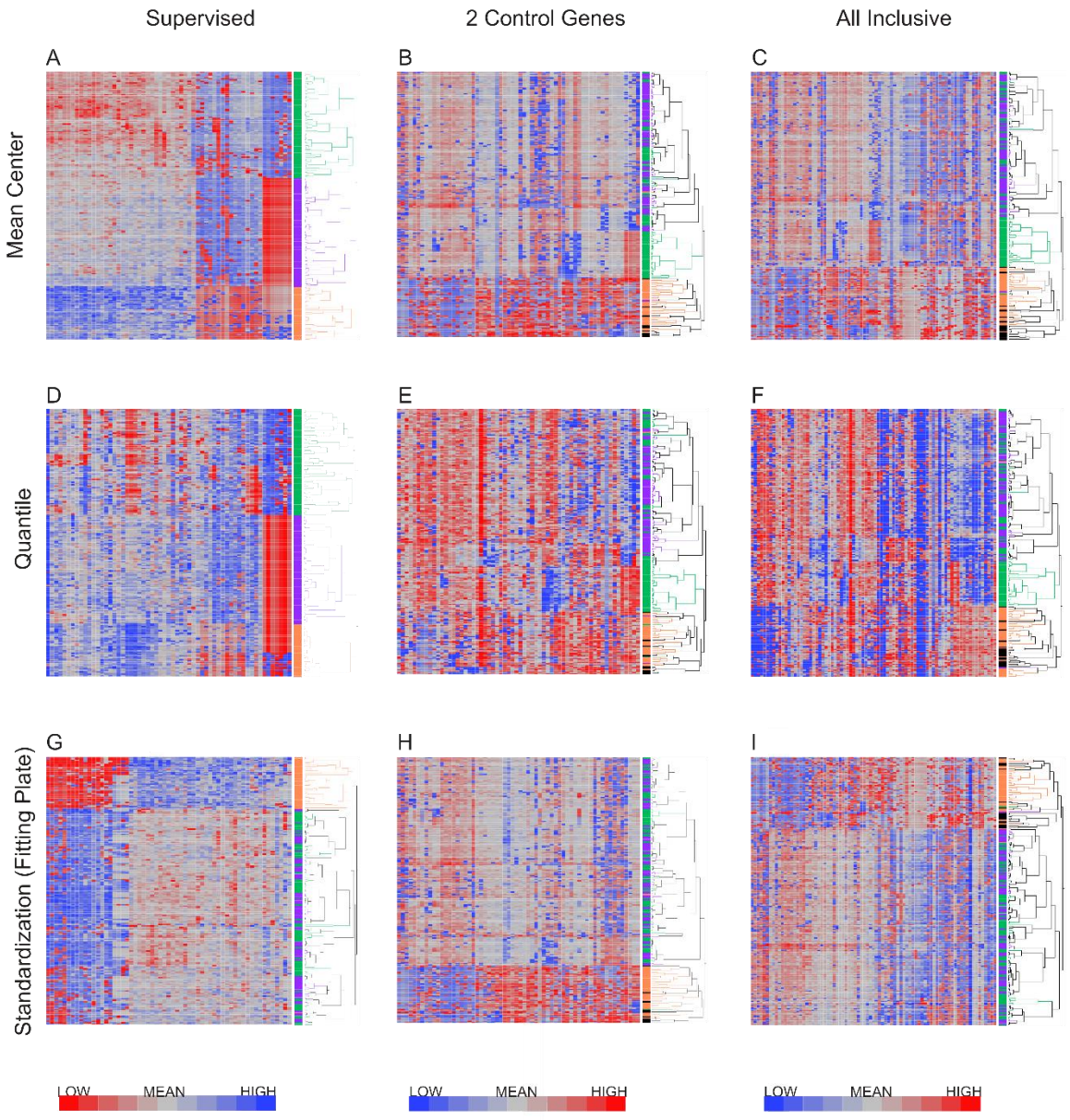

Figure 5. Hierarchical clustering of neutrophil data after nine combinations of primary analysis. Data was processed by 3 alternate methods of data exclusion (columns) and 3 methods of data normalization (rows). Following this, all resulting data sets were subjected to hierarchical clustering by Ward's minimum variance method. The results illustrate the effect of primary analysis method on data interpretation. Cells are colored by cluster for data analyzed by exclusion by exclusion based on the supervised method and normalization by mean centering (top left heatmap).

3.3.3 Hierarchical Clustering Verified the Existence of Cellular Subgroups

Having compared methods for data exclusion and normalization, we opted to focus on the analysis method using a two standard deviation cutoff for exclusion with normalization by standardization of the genes (Figure 5H). Hierarchical clustering revealed 2 major subclasses in both neutrophils (Figure 6A) and T cells (Figure 6B). The more clear definition of neutrophil subgroups, as compared to T cells, could be due to different levels of bimodality in the gene sets, such that more bimodality in the neutrophil data set gives rise to more distinct cellular subclasses. Alternatively, the two data sets could incorporate the same level of overall bimodality but differ in the level of co-variation of bimodally expressed genes. Since the expression of many genes on the T cell array was too low to detect consistently, the analysis is based on fewer genes which also reduces the power to detect clusters.

More refined clustering of the T cell data was also heavily impacted by the decision as to how to handle missing data. Including genes in the analysis according to the 2 standard deviation cutoff, setting missing data to null expression resulted in 6 clusters of cells irrespective of the data normalization procedure. In contrast, when missing data was assumed to be due to technical error and thus assigned the mean value for that transcript, the number of cellular subgroups observed after clustering differed: mean centering resulted in 2 large and 6 small clusters, quantile normalization in 7 clusters, and standardization of the genes in 6 clusters. The all-inclusive method of data selection also resulted in differing numbers of cell clusters depending on the normalization method, with mean centering indicating 4 cellular subgroups, while quantile normalization showed five groups, and standardization of the gene resulted in six groups after hierarchical clustering by visual observation. Concordance of cell clustering across methods ranges between 70% and 80%, arguing that there are multiple cell states

despite the high degree of heterogeneity. Concordance of the two-state clustering indicated in Figure 7B was 95%.

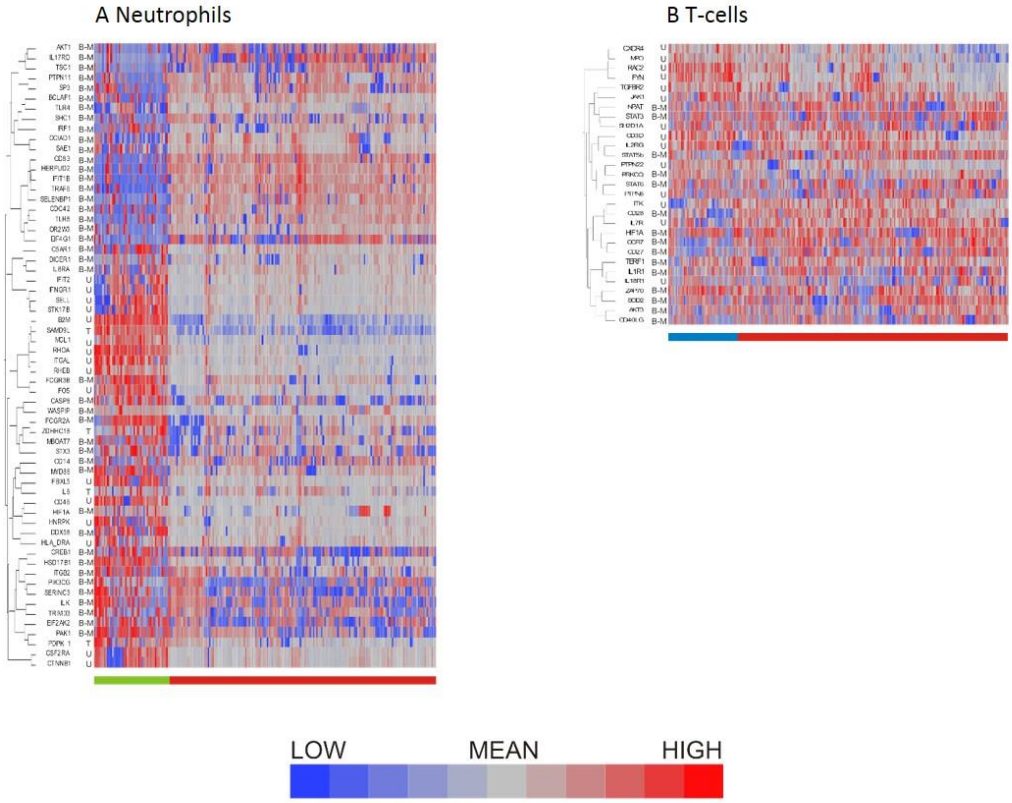

Figure 6. Distribution of bimodal genes in hierarchical clusters after primary analysis. Hierarchical clustering of processed A) neutrophil and B) T cell data after data exclusion by standard deviation cutoff and normalization by standardization of genes resulted in cell clusters defined by shared gene expression patterns. Genes observed to exhibit bimodal expression patterns in pre-processed data are indicated by *.

3.3.4 Individual Differences in Donor Representation in Cellular Subgroups

We next turned to analysis of differences in cellular abundance among donors, and asked whether cells from all donors were equally distributed among the observed clusters. The results show that the frequency of cells in each neutrophil cluster differed between donors (Table 2A, B) with donor 3 having a significantly lower than expected proportion of cells

in cluster A, whereas donor 4 has the inverse profile. Setting the number of clusters to 2 in the analysis of the standardized data following supervised normalization, the χ^2 value for differences in cell type abundance is 24.5 ($p = 6\times10^{-5}$, 4 degrees of freedom) (Table 2A).

As the intrinsic variability of T cell populations is greater than that of neutrophil populations, it is perhaps not surprising that we found donor-to-donor variability to be larger for T cells than for neutrophils. Compared to neutrophils, T cells had considerable variability in cell distribution between subgroups. The counts associated with the smallest subgroup were not large enough to establish whether the donors differ, but they do suggest divergence for the other clusters. Setting the number of clusters to 2 following data normalization, the χ^2 value for differences in cell type abundance is 36.8 ($p = 7x10^{-7}$, 5 degrees of freedom) (Table 2C). Sampling of more cells in more donors will be required to establish whether these differences correlate with physiological and immunological attributes of the individuals.

Table 2: Donor representation in cellular subgroups.

A: Neutrophils with 2 Clusters (Supervised STD)

Cell Count	Donor 1	Donor 2	Donor 3	Donor 4	Donor 5	Total Cell Count per Cluster
Cluster A	4	10	2	17	6	39
Cluster B	40	31	38	21	33	163
Total Cell Count per Donor	44	41	40	38	39	202

Table 2 continued.

B: Neutrophils with 2 Clusters (2G STD)

Cell Count	Donor 1	Donor 2	Donor 3	Donor 4	Donor 5	Total Cell Count per Cluster
Cluster A	40	30	39	21	33	163
Cluster B	5	10	1	17	12	45
Total Cell Count per Donor	45	40	40	38	45	208

C: T cells with 2 Clusters (2G missing STD)

Cell Count	Donor 1	Donor 2	Donor 3	Donor 4	Donor 5	Donor 6	Total Cell Count per Cluster
Cluster A	17	33	44	34	28	42	198
Cluster B	9	10	1	2	20	4	46
Total Cell Count per Donor	26	43	45	36	48	46	244

3.4 Discussion

To determine whether variation in gene expression correlated with variation in cellular phenotype, gene expression data for all genes were analyzed across all cells for expression patterns, such as unimodality and bimodality. Patterns differing from the prevalent, long-tailed, log-normal distribution may reflect active processes that contribute to cell-to-cell variation, which may reflect functional subclasses of cells [71]. Bimodality, in particular, can be expected in immune cell populations, due to the possibility that cells within such a population may be in states of either pre- or post-activation, with the changes in gene expression that this would entail. While bimodal behavior is a potentially important feature of gene expression in a population and can reflect true differences between subpopulations [72], not all bimodal distributions are likely to reflect biological reality in an unprocessed single-cell data set. The risk of excluding true bimodality by setting the cutoff too low must be weighed against the risk of including artificial bimodality by inclusion of all data points and thus more measurement-derived noise. In

addition, it is desirable to differentiate between bimodality due to high versus low expression of a given gene and bimodality due to a gene being expressed or not expressed. Finally, technical artifacts such as plate effects can also induce apparent bimodality if expression of low-abundance transcripts drops out completely in one plate.

The occurrence of bimodality of gene expression in both neutrophils and T cells leads us to conclude that the cell populations tested contain specific cellular sub-types. The results show unambiguous evidence for two cellular subtypes in both the neutrophil and T cell populations, possibly with additional subtypes that will require larger datasets to validate. The nature of the bimodal genes involved, however, hint at the functional nature of the cellular subgroups. For example, the neutrophil cluster represented by low TLR4/8, high PAK1, high ITGB2 (subunit of LFA-1) profiles would likely occur when extravasation and cell motility is more essential than direct microbial phagocytosis. Techniques for collecting single-cell gene expression data have developed rapidly, with recent additions of droplet based technologies such as DropSeq [49] and InDrop [50], and the use of RNA-seq in particular has expanded rapidly. While the techniques for collecting single-cell gene expression data have developed further, no consensus has yet been found for how to best analyze the resulting data sets. These large treasure troves of information offer enormous potential for new insight, making correct interpretation of utmost importance. In this work, we attempted to contribute to this area by systematically testing different combinations of data analysis methods in order to compare the resulting sets of information. Methods for analyzing population-level data are well established; however these are not optimal for single-cell data due to the high variability of gene expression between individual cells and the intrinsic noise in single-cell data sets. Technical variability obtained during sample processing cannot be fully avoided and for single-cell analysis where each cell represents one unreproducible batch, such technical noise proposes a challenge for data interpretation regardless of experimental platform [83]. Gene expression levels, even of housekeeping genes, can differ 1000-fold between

individual cells [67], and analysis of individual single-cell PCR calibration curves do not produce reliable values [74]. Comparison of the outputs from the different methods of primary analysis tested illustrates the impact of analysis method on subsequent interpretation of biologically relevant information such as cellular subtypes within a population. Our recommendation is to use standardization methods that allow for fitting of technical effects, such as the plate effect that generated two sub-types in the mean-centering and quantile normalization strategies. Data exclusion should be aware of the possibility that missing data reflects technical failure, but for the most part it seems to be due to very low and possibly missing expression. Replacement of missing data with average expression did not unduly impact our clustering at the 2-cell type level, and does not appear to be justified.

Cluster analysis is a natural choice for interpretation of qRT-PCR data. We employed two hierarchical clustering methods in order to quantitatively assess the robustness of our primary data processing methods. The results obtained by both methods of clustering were then compared, and the concordance between clusters, as defined by shared cluster assignment for cells, showed that k-means and hierarchical clustering approaches influence the conclusions but to a lesser extent than the data normalization strategy. The two approaches disagreed as more sub-types were added to the analysis, but were in good agreement at k=2 cell types for both neutrophils and T cells.

An additional question we addressed was whether or not the type of cell would have an effect on the concordance, in other words, whether different cell types would require different methods of data exclusion and normalization for optimized analysis outcome. It should be noted that although the trends are similar in both cell types, neutrophils show an overall lower heterogeneity than T cells. While neutrophils can be expected to be largely unimodal in *in vivo* populations, based on traditional cell surface markers, T cells can be expected to cluster into known subsets e.g. Treg, Th1, Th2 groups on the same basis. The observed higher stability of concordance of neutrophil clusters

when compared to T cell clusters is likely affected by these inherent properties of neutrophil and T cell populations. It is thus important to consider not only cluster robustness when choosing analysis methods, particularly when data represents a heterogeneous population, such as the T cell population investigated here.

In conclusion, this study shows that using single-cell analysis we can potentially detect functional subclasses not previously appreciated within immune cell populations. Bimodal patterns of gene expression within the cell populations suggested cellular subclasses, and this was confirmed by hierarchical clustering of cells. Emerging techniques enabling the study of single-cell transcription levels have made clear the need for insight into the appropriate methods of analyzing the data generated. Our systematic testing of different methods of single-cell data analysis clearly illustrates the differences in subsequent interpretation of the processed data. Importantly, our results highlight the necessity of using a method that adjusts for any defined technical effects, and that failure to do this will affect the inference of biological properties. Our main conclusion regarding the necessity of proper data handling is not platform dependent but should be extended to other methods of collecting single-cell gene expression data. Only after ensuring appropriate data handling can we be confident that the vast amount of new information offered to us by single cell data is correctly interpreted.

CHAPTER 4 A COMPUTATIONAL TOOL TO INVESTIGATE THE EFFECT OF TIME DEPENDENT IL-2 STIMULUS ON T CELL RESPONSE

4.1 Introduction

The cytokine IL-2 is an essential part of a functional immune system, playing a vital part in promoting tolerance and immunity. Its main role is through a wide-ranging impact on the function of immune cells, most notably on T cells, both as a growth factor [11] and as a regulator of T cell function [84, 85].

The IL-2 receptor (IL-2R) is comprised of three polypeptide subunits, α, β, and γ [86, 87]. Individually, the three subunits bind IL-2 with low to intermediate affinity [88] [89, 90], but upon the stepwise formation of the heterotrimeric receptor complex, their combined properties enable efficient ligand capture and subsequent cell response [40, 88, 91-95]. While the IL-2 specific α subunit contributes a strong affinity for the ligand, the β and γ subunits with their membrane-spanning domains allow for the initiation of an intracellular signaling transduction in response to ligand binding. Receptor-ligand interaction results in activation of cytosolic protein tyrosine kinases (PTK), such as members of the JAK family [96, 97]. In Jurkat cells, JAK1 and JAK3 associate with receptor subunits β and γ, and initialize a signaling cascade. Downstream of JAKs, phosphorylation of cytosolic STAT5 allows for its dimerization and import into the nucleus [98-100], where it acts as a transcription factor. The three subunits of the IL-2 receptor are all expressed in varying numbers among cells of a population [6, 56]; the number of receptors available to capture extracellular IL-2 and transduce signal will differ between individual cells, which in turn will lead to varying cellular response. Consequently, it is to be expected that a population average will not be sufficient to capture the range of responses in a cell population. In order to address this issue, I

developed a model that enables simulation of cellular response with different levels of receptor subunits. In order to experimentally capture the variability in response at the single-cell level and collect quantitative data, I employed a microfluidic device capable of trapping and arraying non-adherent cells for time-series imaging. This allowed me to precisely replicate the modeled cytokine exposure on live cells and to longitudinally track protein translocation at the single-cell level using live-cell imaging.

IL-2 stimulation of T cells causes *de novo* synthesis of IL-2Rα and subsequent IL-2 production [101], suggesting that *in vivo* the main contributor to cell response heterogeneity is the α subunit. In this model system, we remove this as the limiting factor by basing our model on Jurkat cells (which constitutively produce IL-2), such that the cells are primed with all three subunits being present at the cell surface. Under these conditions, we explored the rate-limiting steps of receptor complex formation and investigated kinetics and expression levels of the IL-2R subunits critical for signaling response under fluctuating IL-2 conditions.

4.2 Materials and Methods

We used a computational model in order to investigate the effect of pulsatile IL-2 on T cell response. An initial model of T cell proliferation in response to IL-2 was modified to represent a system of interest and expanded to include further receptor-level detail, as well as to allow for the addition of cytokine input in a pulsatile fashion and the interrogation of the downstream response.

4.2.1 A Model of Jurkat Cell Response to IL-2

The model simulates cellular response to periodic cytokine stimulus, using Jurkat T cells as the cell type of interest and IL-2 as the cytokine. The initial framework for our model was a previously published model of cellular proliferation in response to IL-2 [102]. Our model was constructed in the MatLab platform, SimBiology, and was adapted to

incorporate subunit-level detail of the IL-2 receptor, as well as downstream translocation of STAT5 from the cytosol to the nucleus. The volume of the model was customized to correspond to and experimental microfluidics setup rather than a cell culture system. The time scale was likewise altered to reflect a short ($\leq$ 1 hour) timescale of interest, rather than the two to five days that would be of interest for cellular proliferation. Transcriptional processes, both constitutive and IL-2 induced were included in the model; however, the longer time scale needed for these processes meant that their impact on the model outcome was minimal. The initial steps of the model describe binding of IL-2 by the three IL-2 receptor subunits at the cell surface, and the sequential assembly of the ligand-receptor complex at the cell membrane (Cs). This is followed by the resulting downstream cellular response to signaling by the ligand-receptor complex in the form of STAT5 translocation from the cytoplasm to the nucleus. In our model, the signaling steps between receptor-ligand interactions and downstream STAT5 translocation are represented in a simplified fashion by a logistic delay function (1), where translocation of STAT5 is dependent on the number of receptor-ligand complexes present at the cell surface.

$$Sr = y + b * Cs/(d + e^{-ku1*(t-ti)}) \quad (1)$$

IL-2 is bound to the three subunits of the receptor by stepwise assembly with IL-2α strengthening the bond to the ligand allowing first IL-2β and then IL-2γ to bind to the complex [92]. This step-wise formation of the receptor-ligand complex is represented by two steps in our model, with the initial step being the capture of IL-2 by IL-2α and the subsequent addition of IL-2β and IL-2γ simplified into one step. The IL-2Rβ and IL-2Rγ subunits of the receptor were modeled as one combined species, making the assumption that they were co-localized prior to complex formation. According to Feinerman et al,

[6], adding the association of first IL-2Rβ, then IL-2Rγ did not result in model outcome different from that of adding the two subunits simultaneously. We thus considered our simplification justified. The numbers of receptor subunits per cell were determined based ranges reported in literature [6, 103-105], with both high and low initial levels simulated. Upon formation of the receptor-ligand complex at the cell membrane, the complex is internalized and the components undergo differential sorting [104]. The model incorporates this internalization of the ligand-receptor complex, as well as degradation of the IL-2Rβ and IL-2Rγ subunits and recycling of the IL-2Rα subunit and ligand [104, 106]. In order to mimic microfluidic experimental conditions, which occur under constant flow, recycled IL-2 was modeled as lost from the model system and not available to the cell.

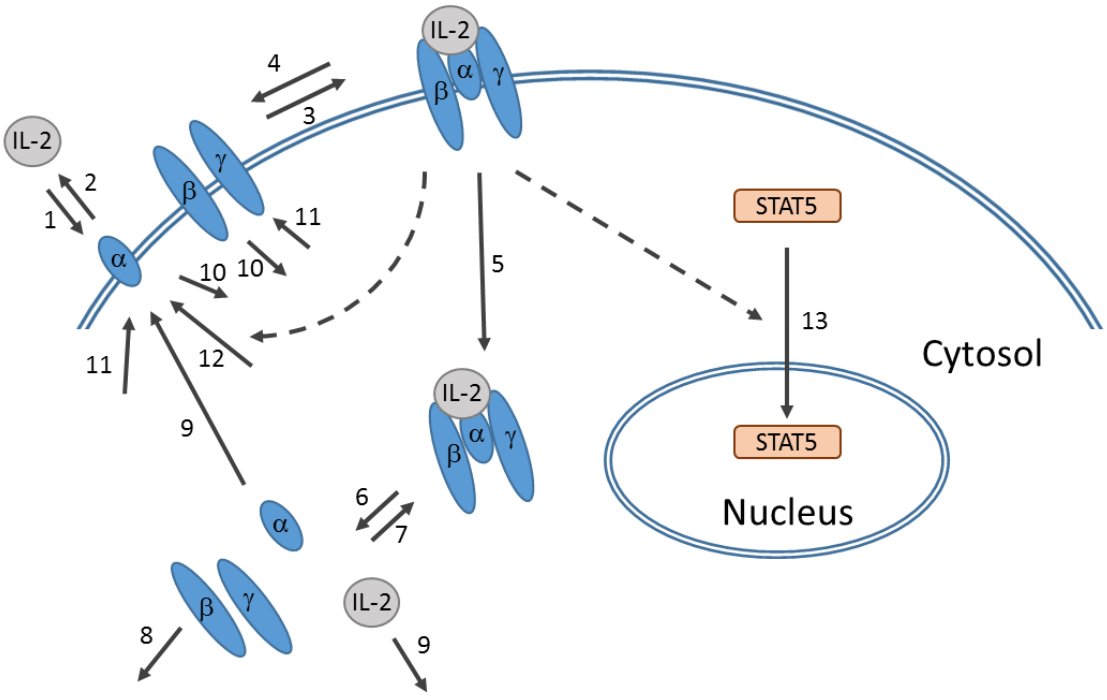

Figure 7. Overview of computational model. The model incorporated the formation of receptor-ligand complex from extracellular IL-2 and receptor subunits at the cell membrane (1-4). The complex is internalized (5) and its components sorted for intracellular degradation (8) or recycling (9). The binding of IL-2 by its receptor initiates

a downstream signaling cascade that results in translocation of STAT5 from the cytosol to the nucleus (13). This is represented by a delay function affected by the number of receptor-ligand complexes present at the cell membrane. The model also incorporated constitutive (11) and IL-2 induced (12) production of receptor subunits and consecutive internalization of unbound receptor (10). Dashed arrows indicate an effect of a species on a reaction. Numbers by reaction arrows correspond to process numbers in Table 3.

Downstream steps from surface complex formation to STAT5 translocation were modeled in a simplified fashion as a time delay function where the translocation of STAT5 was made dependent on the number of receptor-ligand complexes present on the cell surface. The rate of the delay was determined by fitting to experimentally obtained data from responding cells after constant stimulation with 100 pM IL-2 (Figure 8). Values for all model rates are shown in Table 3.

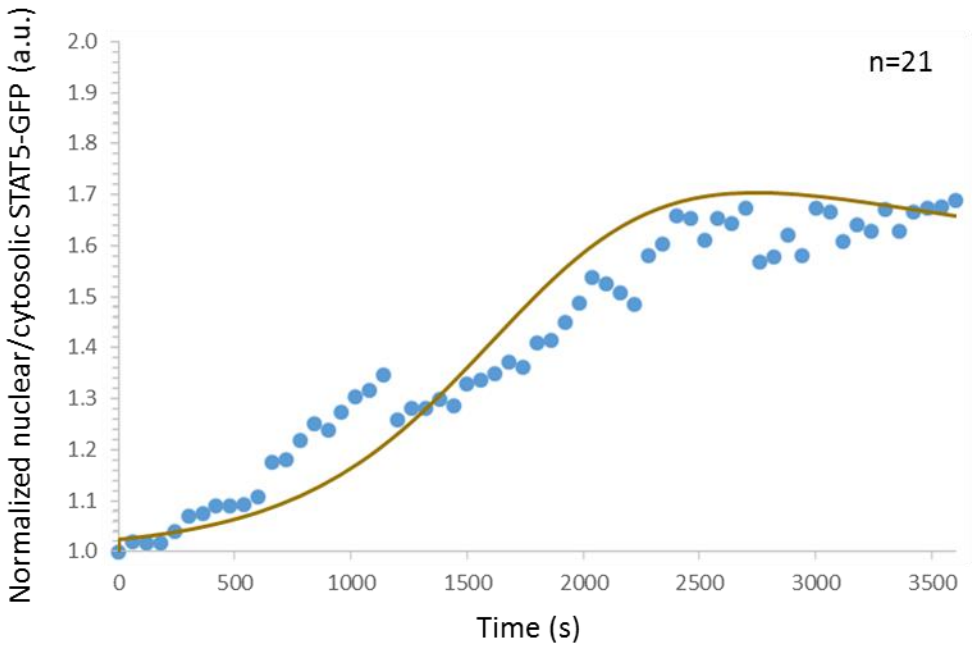

Figure 8. Fitting of delay function for STAT5 nuclear translocation. A logistic delay function was fitted to normalized experimental data form the mean of 21 responding cells (28 % of population).

Table 3: Parameter values for computational model. Process numbers correspond to number in model overview in Figure 7.

Parameter	Process	Definition	Value	Source
kon1	1	Association of IL2Rα and IL-2 at cell membrane	1.40E+07	[6]
koff1	2	Dissociation of IL2Rα-IL2 at cell membrane	0.4	[6]
kon2	3	Association rate for IL2Rα-IL2 to IL-2Rβ / IL-2Rγ	3.60E-04	Fitted to model parameters
kr	4	Cell surface dissociation of receptor and ligand	2.30E-04	[40]
ke	5	Internalization of receptor-ligand complex	6.67E-04	[102]
kre	6	Intracellular dissociation of ligand and receptor	0.00184	[102]
kfe	7	Intracellular association of ligand and receptor	1.84E-06	[102]
kh	8	Degradation of internalized receptor and ligand	5.83E-04	[107]
kx	9	Recycling	0.0025	[108]
kt	10	Constitutive internalization of unbound receptor	1.17E-04	[109]
Vs	11	Constitutive receptor subunit synthesis rate	0.183333	[107]
ksyn	12	Enhanced receptor synthesis due to receptor-ligand interaction	1.83E-05	[110]
ku1	13	STAT5 translocation delay steepness constant	2.50E-03	Fitted to experimental data
ti	13	STAT5 translocation delay inflection constant	2.00E+03	Fitted to experimental data
b	13	STAT5 translocation delay scaling constant	3.50E-03	Fitted to experimental data
d	13	STAT5 translocation delay constant	1.80E+00	Fitted to experimental data
NA		Avogadro's constant in pM	6.02E+11	[111]
Ve		Total endosomal volume	1.00E-14	[112]
Vol		Volume of system	2.01E-07	[113]

Modeled concentrations of IL-2

IL-2 concentrations of 10 pm and 100 pM were tested in the model, with 10 pM being the canonical threshold for triggering T cell response [114]. In a clinical setting, concentrations between 1 and 100 pM are reported as therapeutically relevant [115], while concentrations above 100 pM induce undesirable inflammatory responses [116, 117]. With this in mind, we considered a range of 10 to 100 pM to be biologically relevant and these two values were tested in our model.

Pulsatile input of IL-2

Model response to pulsatile input of IL-2 was investigated using custom MatLab code, which delivered a pre-defined set of IL-2 pulses interspersed by recovery pauses to the modeled system. IL-2 was delivered at pulse lengths of 0.5, 1, 2, 3, 4, or 5 minutes followed by pause lengths of the same durations. Total simulation time was set to one hour for all input settings.

Model outputs of interest

Two dynamic model outputs were collected: the total number of receptor-ligand complexes present at the cell membrane (Cs), and the nuclear localization of STAT5. Under the assumption that only complexes of trimeric IL-2 receptor and bound IL-2 present at the cell membrane are able to initiate downstream signaling, Cs was used as a first indicator of cell response to IL-2. The translocation of STAT5 from cytosol to nucleus is a direct downstream effect of IL-2-receptor interaction and was represented in the model as the ratio of nuclear to cytosolic STAT5 (Sr).

4.3 Results

4.3.1 Model Simulations Predict Differential Response to Oscillatory IL-2 Input Functions

Our model describes downstream responses of T cells upon binding IL-2, modifying a computational framework first developed by Fallon & Lauffenburger [102] to include features of interest to us associated with the earliest events of IL-2 receptor ligation. The IL-2 receptor is represented as individual subunits, which combine in a stepwise fashion upon binding IL-2 to form a heterotrimeric receptor. The receptor-ligand complex is assumed to initiate an intracellular signaling pathway, resulting in the translocation of STAT5 from the cytosol to the nucleus, which is used as a model output indicating cell response. In addition to this chain of events, the internalization and subsequent recycling and degradation of the receptor-ligand complex are included in the model, as this affects the number of cell surface receptor-ligand complexes available for initiation of cell response. The steps between receptor-ligand interaction to STAT5 translocation are represented in a simplified fashion by a delay function, where translocation of STAT5 is dependent on the number of receptor-ligand complexes present at the cell surface. The model allows for tight control of ligand input, simulating addition and removal of IL-2 to the extracellular environment. We simulated a range of input combinations in order to investigate cellular dynamics in response to pulsatile IL-2 stimulation at below equilibrium levels. Each simulation setting represented a one hour time course with pulses of 30 seconds to 5 minutes of IL-2 followed by recovery times (pauses) of 30 seconds to 5 minutes. The model under consideration represented an “average cell” with receptor subunit levels defined by values within the mean range reported in literature.

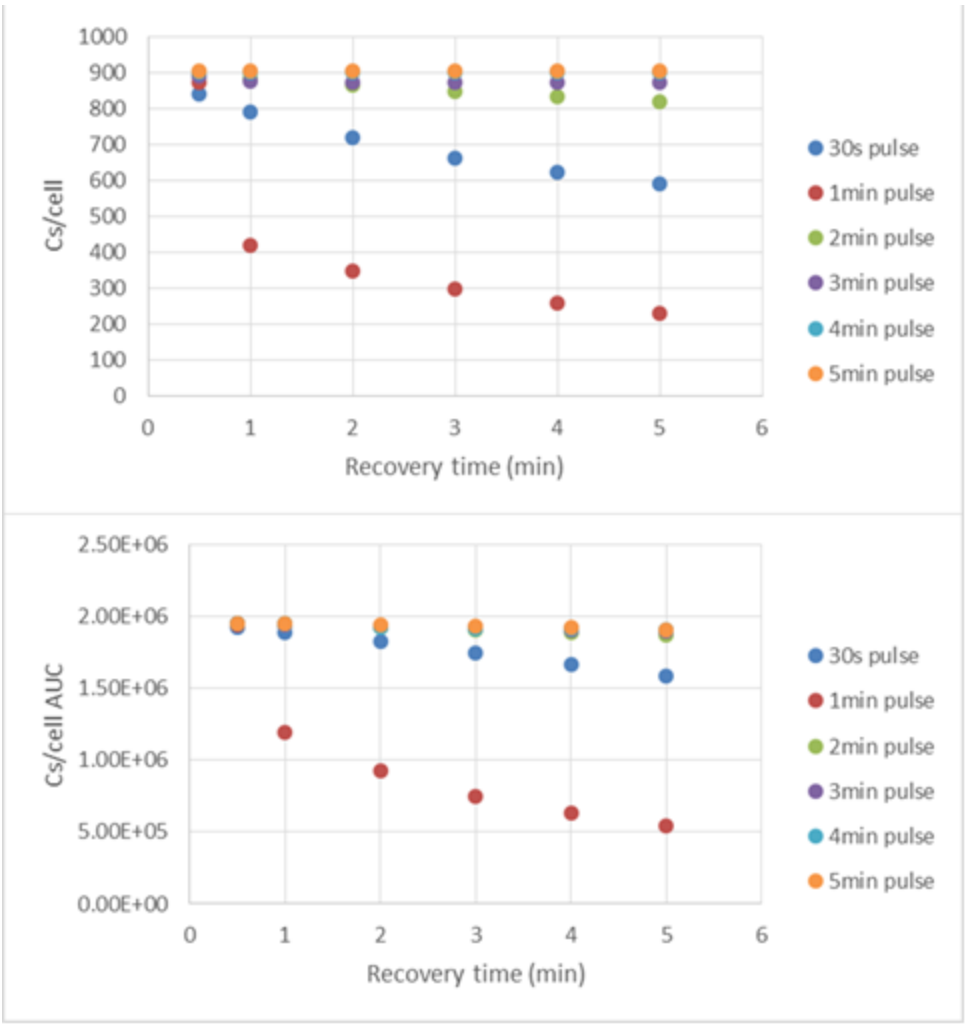

Figure 9. Model predicted receptor-ligand complex formation in response to pulsatile IL-2 stimulation.

The model predicted variation in maximum numbers of receptor-ligand complexes per cell (top) and area under curve for receptor-ligand complexes per cell (bottom) during a one hour simulation under different pulsatile IL-2 input conditions. Pulse length is indicated by legend and pause length on the x axis. All simulations were run using 100 pM IL-2.

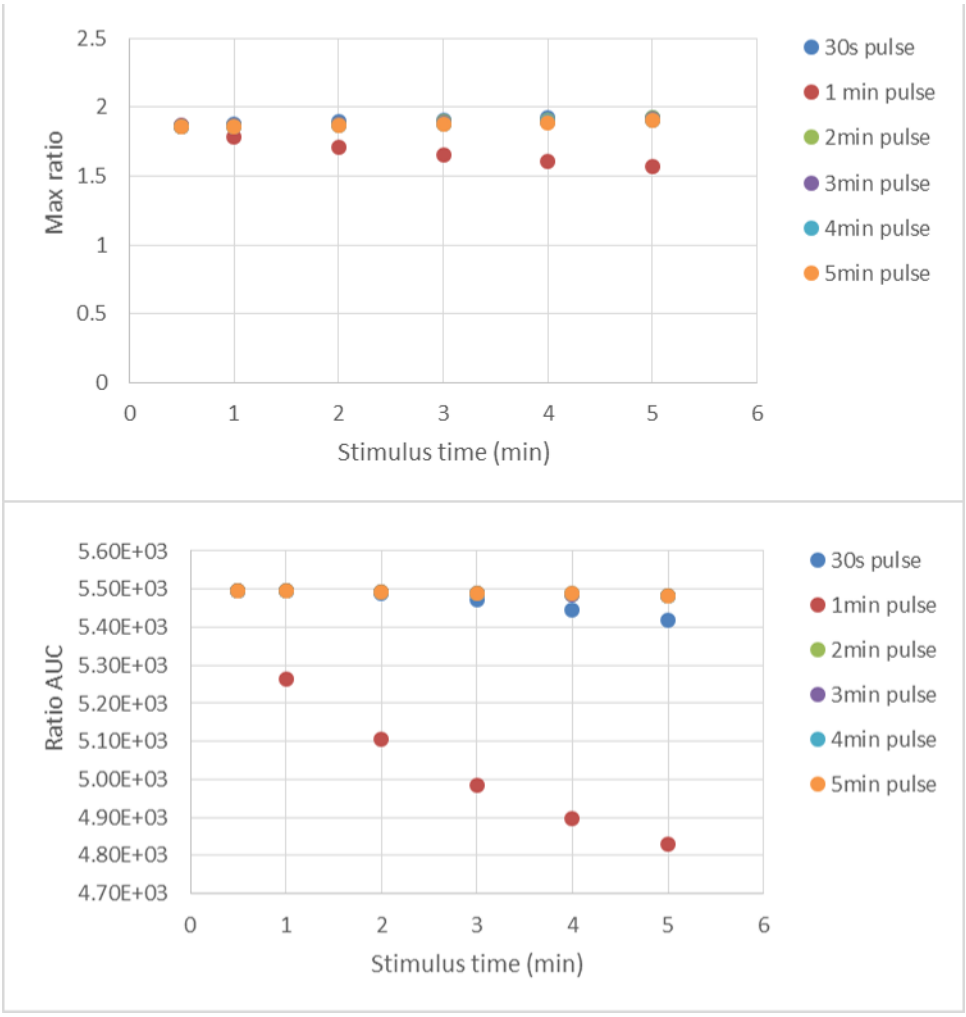

Figure 10. Model predicted ratio of nuclear to cytosolic STAT5 in response to pulsatile IL-2 stimulation.

The model predicted variation in maximum STA5 nuclear translocation per cell (top) and area under curve for the ratio of nuclear to cytosolic STAT5 (bottom) during a one hour simulation under different pulsatile IL-2 input conditions. Pulse length is indicated by legend and pause length on the x axis. All simulations were run using 100 pM IL-2.

The resulting cellular trajectories indicated a range of cellular response depending on the IL-2 input, as measured by the maximum number of receptor-ligand complexes present at the cell membrane (Figure 9, top), the maximum ratio of nuclear to cytosolic STAT5 (Figure 10, top), as well as the area under the curve for the trajectory of each of these

species (Figure 9, bottom, Figure 10, bottom). The maximum level of cellular response varied with length of stimulus pulse and recovery time. Both output metrics showed a decrease in peak value as recovery time between IL-2 pulses was increased, with the effect of recovery time being less influential as the length of stimulus pulses was increased. For both metrics, the response was most highly affected by changes in recovery time when pulse time was set to one minute. As pulse length decreased or increased from one minute, change in maximum response due to altered recovery time decreased. For shorter stimulus pulses, a decrease in maximum response and AUC was predicted with longer recovery times. Again, this effect was most marked when pulse length was one minute.

4.3.2 Model Results Show Variation in Maximum Cellular Response Depending on Receptor Subunit Heterogeneity

The three receptor subunits of the IL-2 receptor are present at the cell surface in numbers that vary between cells in a population. Activated T cells have been found to express the subunits of the IL-2 receptor in ranges up to two (IL-2Rβ) and three order of magnitude (IL-2Rα) [6, 56]. Due to this heterogeneity in expression, individual cell response to IL-2 can be assumed to vary with the availability of subunits. Our model addresses population heterogeneity by allowing for simulation of high and low levels of unbound receptor subunits at the resting level. Results for cells with 1000 or 1500 IL-2Rβ/γ indicate that the cells achieve different levels of maximum response, defined by both the maximum number of receptor-ligand complexes present at the cell membrane and the ratio of nuclear to cytosolic STAT5, to the same input of IL-2. The difference in dynamic response between simulated cells with low versus high initial levels of IL2Rβ/γ subunits showed a lower overall response level as an effect of lower IL-2Rβ/γ subunit levels across all input combinations, as compared to cells with higher subunit levels. This was

especially apparent for the input settings where the pulse time for extracellular IL-2 availability was 1 minute and the recovery time between pulses was varied (Figure 11). In this input range, cells showed the highest level of sensitivity to cell-to-cell variability in subunit levels, with a nine-fold difference in maximum response as recovery pulse time increased from 30 seconds to five minutes (Figure 11A). Likewise, response dynamics over a one hour simulation time showed a range of response profiles, with cells responding to one minute IL-2 pulses and one minute recovery pauses showing the largest shift in response curve slope as a result of differences in initial IL-2Rβ level (Figure 11B).

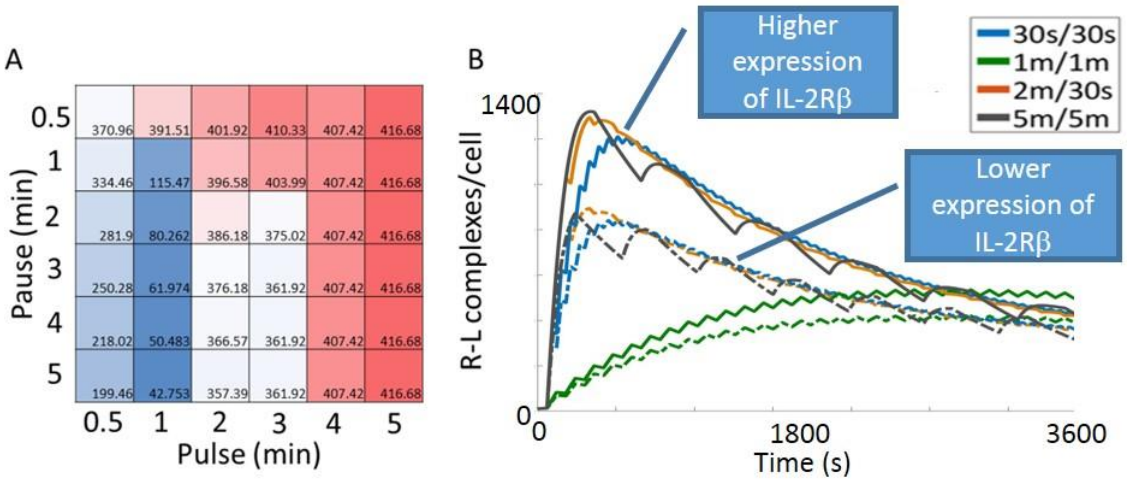

Figure 11. Difference in peak receptor-ligand complex formation due to cell-to-cell variability in IL-2Rβ/γ numbers. Model results showing A) the difference in maximum receptor-ligand complexes per cell and B) the change in receptor-ligand complexes per cell over time due to variability (high and low initial expression levels) of IL-2Rβ/γ numbers. Filled lines indicate high and dashed lines low initial subunit levels, respectively. The difference in initial subunit numbers affects cell response to pulsatile IL-2 input in varying degrees across all combinations of ligand pulse and recovery times. IL-2 concentration was 100 pM and total simulation time one hour for all pulsatile input combinations.

4.4 Discussion

Based on the reported ranges of subunit levels within T cell populations, including previously obtained results for gene expression of the IL-2 receptor subunits in primary T cells (Aim 1) [56], we predicted that there would be a range of cellular responses to IL-2 stimulation within the population, resulting from differences in the availability of unbound receptor subunits with which to form signal inducing receptor-ligand complexes. With our modeling approach we aimed to investigate this by simulating cells with variable levels of individual receptor subunits. A total of 25 combinations of time-dependent IL-2 stimulus were tested, with pulse time and recovery pauses varied from 30 seconds to five minutes. Our model predicted that an input of one minute IL-2 should result in the comparatively largest range of response profiles within a cell population based on initial IL-2Rβ numbers per cell. Within our modeled system, internalization of receptor-ligand complex removes receptor subunits from the cell surface, and while IL-2Rα is recycled, IL-2Rβ and IL-2Rγ are not. Despite the production of new subunits, this makes the two latter subunits rate limiting in the cell's response to IL-2. Our results pinpointed a one minute IL-2 pulse as the input range where this rate-limiting aspect of subunit availability had the greatest effect, a conclusion which is illustrated by the difference in maximum response between cells expressing high and low levels of IL-2Rβ and IL-2Rγ across input settings using a one minute pulse time. Likewise, a comparison of the shift in dynamic response for different input settings indicated that a one minute range should result in the widest range of intrapopulation response for Jurkat cells responding to 100 pM IL-2. Previous models of cellular interaction with IL-2 have simulated various aspects of interaction and downstream response, such as cell proliferation [102], or IL-2 induced cytokine production [52]. A common thread of the majority of studies has been the assumption that population averages of parameters and responses adequately represent biological reality. It is becoming more and more apparent that the variability of single-cell responses within populations have functional

consequences, and that the traditional approach of building models based on the average cell in a population fails to adequately capture this complex reality. The growing awareness of the importance of a single-cell approach has led to development of models focusing on aspects such as single-cell competition for IL-2 which can have population-level impact on immune function [6].

While the Fallon & Lauffenburger model used as our starting point incorporated post-binding internalization and trafficking of the receptor and ligand, it focused on the commonly used downstream responses of cell proliferation, which occurs on a time scale of days after IL-2 stimulation. In contrast, our model simulated an immediate response to IL-2, occurring within a time frame of minutes, which can be expected to be unaffected by IL-2 induced gene upregulation. This allowed us to make the assumption that the individual cell response would be dependent on the initial state of the cell. Our model attempts to capture aspects of this variability within a cell population by allowing for variability of receptor subunit levels. The added detail of receptor subunits allowed a model description that more closely adhered to the realities of differential subunit expression *in vivo*.

In conclusion, we have investigated single-cell response to variable pulsatile IL-2 input, using computational modeling to allow for detailed interrogation of the effects of receptor level variability as well as extracellular ligand fluctuations on Jurkat T cell line responses. We were limited in our goal of a single-cell approach by the scant availability of single-cell parameter values for the processes modeled, but by allowing for cell variability within ranges of subunit expression reported in literature, we were able to simulate cell heterogeneity at the single-cell level. An intriguing approach to incorporate single-cell information into computational models by single-cell parameter estimation was recently suggested by Yao et al. [118]. By using Bayesian parameter inference at the single-cell level, followed by inferred parameter clustering, these investigators detected existing cellular states within their population, explaining previously observed response

variability. Further development of our system using such an approach to incorporate our experimental information would allow for expanded interrogation of distinct cellular profiles. The differences seen in maximum response and in response profiles when using varying combinations of IL-2 pulse times and recovery times indicate that Jurkat cell response to IL-2 depends in part on these metrics of pulsatile IL-2 stimulus. Experimental testing using the model-predicted stimulus setting could shed light on whether this simulated behavior corresponds to actual *in vitro* and/or *in vivo* cell behavior.

CHAPTER 5 QUANTIFYING VARIABILITY OF PHENOTYPIC T CELL RESPONSES TO TIME DEPENDENT IL-2 STIMULUS

5.1 Introduction

Cell signaling systems often respond to extracellular ligand with exquisite sensitivity to minute changes in concentration, and it has been suggested that such pre-equilibrium sensing could occur in a system where the downstream response is faster than the time needed to reach equilibrium for receptor-ligand interaction at the cell surface, allowing the cells to distinguish between pre-equilibrium doses of ligand [39]. Pre-equilibrium sensing and signaling (PRESS) has been demonstrated to expand and shift the dynamic range of input ligand concentrations for orientation/polarization in chemotactic gradients. Based on the kinetics of its receptor-ligand interaction and the downstream processes, the IL-2 ligand-receptor system has recently been suggested as potentially regulated by pre-equilibrium sensing and signaling [39]. Here, we investigated the effects of rapidly fluctuating extracellular IL-2 levels on T cell response by subjecting cells to time-varying IL-2 input in a regime that is below equilibrium levels and at physiologically relevant concentrations. A previously developed microfluidic device allowed for the capture of Jurkat cells and the precise delivery of time-dependent IL-2 stimulus to the extracellular environment. The model results from Aim 2 were used to determine pulsatile input settings for stimulus delivery to the cells and fluorescent imaging was used to track fluorescently labeled STAT5-GFP translocation in the cell as a downstream indicator of IL-2 response.

5.2 Materials and Methods

5.2.1 Creating a Jurkat Cell Line with Stable Expression of STAT5-GFP

Cell culture

Jurkat cells and Human Embryonic Kidney (HEK) 293T cell lines were purchased from ATCC and maintained in RPMI-1640 medium without phenol red (Lonza), supplemented with 10 % FBS (Sigma), L-glutamine (Fisher Scientific), 1 % MEM non-essential amino acids (Mediatech), 10 mM HEPES (Mediatech), 1 mM sodium pyruvate (Mediatech), 50 U/mL penicillin and 50μg/mL streptomycin (Fisher Scientific) at 37^{o}C 5% CO_2.

Transfection

For transfection of plasmid containing GFP-labeled STAT5 (Origene Technologies), HEK 293T cells were transfected using the Neon transfection system (Thermo Fisher). After 2 days of culture following transfection, supernatant containing retrovirus was collected and used to transfect Jurkat cells by spinoculation. Following spinoculation and recovery, GFP-positive Jurkat cells were collected from the population by sorting using a BD FacsAria Fusion cell sorter. STAT5-GFP expressing Jurkat cells were maintained in culture as above.

5.2.2 Cell Response Assay using Microfluidic Devices

Manufacturing Microfluidic Devices

Two layer microfluidic devices were fabricated as previously described [113]. Briefly, the devices were molded from 10% polydimethylsiloxane (PDMS), assembled, and plasma bonded to glass slides to enable imaging. The two-layer design of the device allows for immobilization of suspension cells by horizontal flow in one layer and for delivery of stimulus to all cells in the trap simultaneously by perpendicular flow from the top layer (Figure 12).

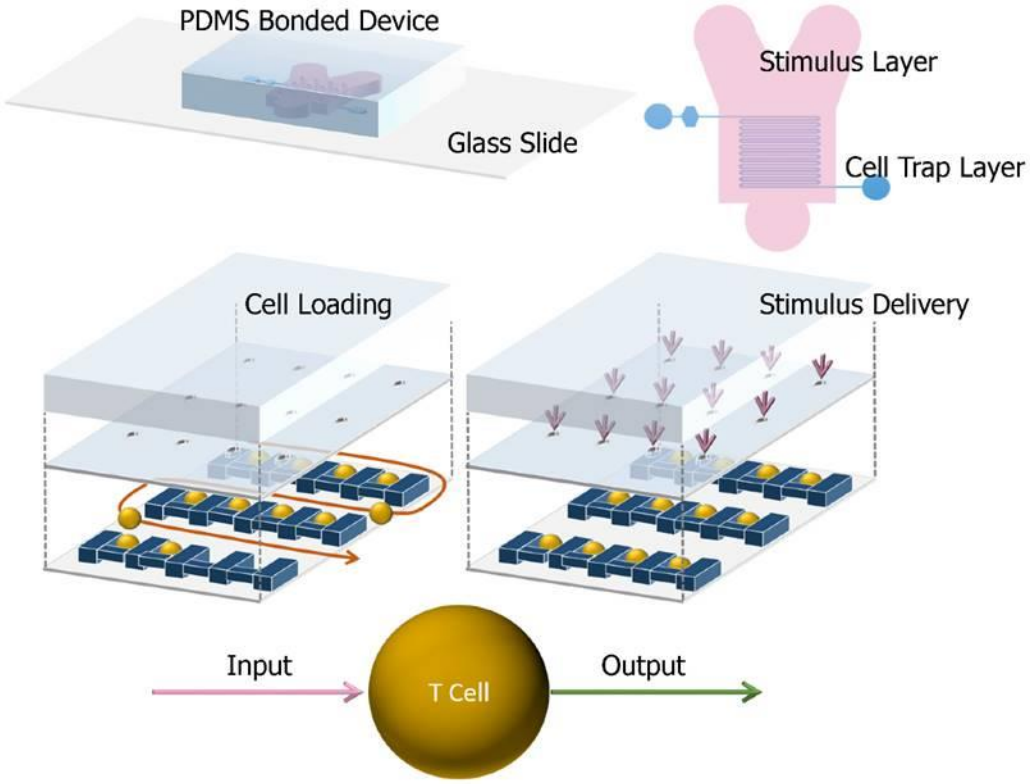

Figure 12. Overview of microfluidic cell trap. The two layer design allows for immobilization of non-adherent cells by horizontal flow and for delivery of stimulus to all cells in the trap simultaneously by perpendicular flow from the top layer. Reproduced with permission from [38]

Cell Response Assay in Microfluidic Devices

Before loading onto the microfluidic device, STAT5-GFP Jurkat cells were resuspended in HBSS with 0.5 % FBS and incubated with Hoechst 33258 nucleic acid dye (Sigma) and Wheat Germ Agglutinin (WGA) Alexa Fluor 647 membrane stain (Fisher Scientific) for 10 minutes at 37^{o}C. Following incubation, cells were washed, resuspended in growth medium, and loaded into the device using gravity flow. After loading, device inlet tubing for the two inlets was connected to growth medium with and without IL-2, and delivery was controlled using a custom built pressure box with inlet pressure set to 1 psi. A

custom GUI was used to control pinch valves enabling precise delivery of stimulus and medium to the trapped cells. Prior studies with fluorescein labeled buffer have characterized that the pulsatile properties of this platform are well-matched to the desired input function [38]. Cells were exposed to either a constant flow of medium containing 10 pm or 100 pM IL-2 or to a pulsatile input of 100 pM IL-2, alternating with growth medium without IL-2. Model predictions from Aim 2 (Figure 11) were used to inform the choice of pulse and recovery times. Time lapse images (Figure 13) were taken in 60 second intervals in the GFP and DAPI ranges at 20x magnification using a PerkinElmer UltraVIEW VoX spinning disk confocal microscope with a Nikon Ti-E camera. WGA labeling of the cell membrane was used to visually identify cells during imaging setup in order to minimize photobleaching of GFP. Total experimental time was set to one hour, to minimize possible effects of IL-2 mediated gene upregulation [119].

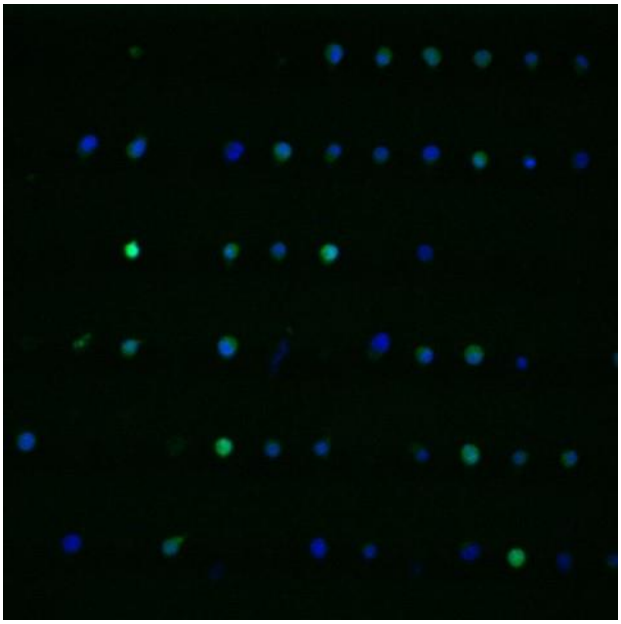

Figure 13. Jurkat cells expressing STAT5-GFP trapped in microfluidic deviceFluorescent image of Jurkat cells trapped in the microfluidic device. The image was taken at 20x magnification, using a GFP filter to capture the intracellular location of STAT5-GFP (green) and a DAPI filter to capture Hoechst staining (blue). Hoechst staining was define the nuclear area in each cell.

Quantification of STAT5-GFP translocation

Images taken during cell stimulation were analyzed using custom Matlab code (see Appendix A.7). Images were converted to tif format and then to binary format. In order to distinguish between the whole cell and the nuclear compartment, masks were created from the binary images to capture areas of interest (cells) for each image field and to create areas of interest (cytosol and nucleus) for each cell. Cells of interest were manually selected from the masks based on image quality over the course of the experiment. The masks where then used to define cells for further analysis in the original image and GFP intensity was calculated for the masked areas. Local background GFP intensity was subtracted from each individual whole cell and its nucleus at each time point, and the ratio of GFP intensity for nuclear versus cytosolic area was then calculated for each cell

at each time point. For time points where image quality was deemed too low, data was imputed by the mean of the two time points immediately preceding and following for that individual cell. The resulting STAT5-GFP ratio data were plotted over time, creating dynamic single-cell traces of STAT5-GFP translocation. Cells where the normalized ratio of nuclear to cytosolic GFP reached 1.5 or higher for two or more time points during the one hour time course were categorized as responders to IL-2 stimulus.

Statistical analysis

The mean of the maximum ratio of nuclear to cytosolic STAT5-GFP for responding cells from all fluctuating IL-2 input settings were tested using one-way ANOVA.

5.2.3 Comparison of STAT5 and STAT5-GFP in Transfected Jurkat Cells

Cytoplasmic and nuclear lysates were obtained from transfected and sorted cells. A Western blot was run using 20 µg of total protein sample from each fraction on a 10% SDS-PAGE gel followed by transfer to a PVDF membrane. The membrane was blocked with Near Infra-Red Blocking Buffer (Rockland Immunochemicals) and probed using anti-STAT5 antibody (BioLegend) at 1:1000 dilution in 4^{o}C overnight. This was followed by secondary anti-mouse antibody (IR dye 680CW donkey anti-mouse, Li-Cor Biosciences) at 1:10000 for 1 hour at room temperature. The membrane was imaged using a Licor Odyssey CLx Imaging System and analyzed using Image Studio software to quantify the relative intensities of the bands for STAT5 and STAT5-GFP in the two cell compartments.

5.2.4 Relative levels of IL-2Rα and IL-2Rβ within the Jurkat population.

In order to investigate relative levels of IL-2Rβ, cells were incubated with antibody against IL-2Rβ. 100,000 cells were resuspended in fresh medium and incubated with primary antibody against IL-2Rβ (Novus Biologicals) at 1:10 dilution for one hour at

room temperature. The cells were then washed three times and incubated for one hour with an Alexa 647-labeled secondary antibody (Life Technologies) at 1:200 dilution for one hour at room temperature, followed three additional washes. Cells were analyzed in two ways: by flow cytometry using a BD LSRFortessa, and by fluorescent microscopy using a PerkinElmer UltraVIEW VoX spinning disk confocal microscope with a Nikon Ti-E camera. The relative intensity of the Alexa 647 signal was quantified for each cell and used to assess the heterogeneity of available IL-2Rβ within the population. In order to investigate relative levels of IL-2Rα and IL-2Rβ on the same cell, cells were incubated with FITC-labeled antibody against IL-2Rα (BioLegend) and APC-labeled antibody against IL-2Rβ. Cells were analyzed using flow cytometry, as above.

5.3 Results

5.3.1 Jurkat Cells in Culture are Pre-primed to Express IL-2Rα

Jurkat cell constitutively produce STAT5, which means that tracking STAT5-GFP translocation does not reflect the total STAT5 in the system due to the presence of endogenous protein. In order to determine the proportion of GFP-labeled STAT5 in the transfected cells, lysates of the cytosolic and nuclear fractions from transfected and sorted cells were analyzed by Western blot. The quantified results show that the mean ratio of cytosolic GFP-labeled STAT5 to endogenous STAT5 in the population was 0.14 (Figure 14). The level of nuclear GFP-labeled STAT5 was too low for detection by Western blot, but endogenous nuclear STAT5 was present at a ratio of 0.017 to endogenous cytosolic STAT5. In addition, individual cell traces of GFP-labeled STAT5 showed its presence in both cytosol and nucleus of cells prior to stimulation. Thus, we can consider these cells to be “primed” by IL-2 and consequently expressing IL-2Rα at the cell surface, enabling

them to respond to IL-2 without the need for experimentally induced IL-2Rα upregulation.

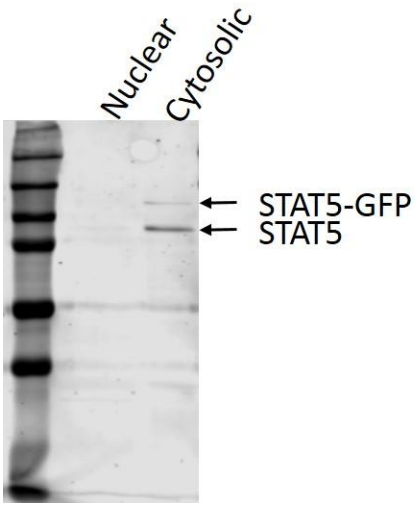

Figure 14. Relative amounts of STAT5 and STAT5-GFP in nucleus and cytosol.Western blot showing STAT5 and STAT5-GFP in the nuclear and cytosolic compartments of Jurkat cells prior to IL-2 stimulation.

5.3.2 Relative Numbers of Available IL-2Rα and IL-2Rβ Subunits Vary within the Jurkat Population

The relative numbers of available IL2-Rβ receptor were compared using flow cytometry and fluorescence imaging. We observed a range of subunit expression levels within the Jurkat cell population (Figure 15), indicating cell-cell variability spanning three orders of magnitude in the number of subunits available for IL-2 interaction.

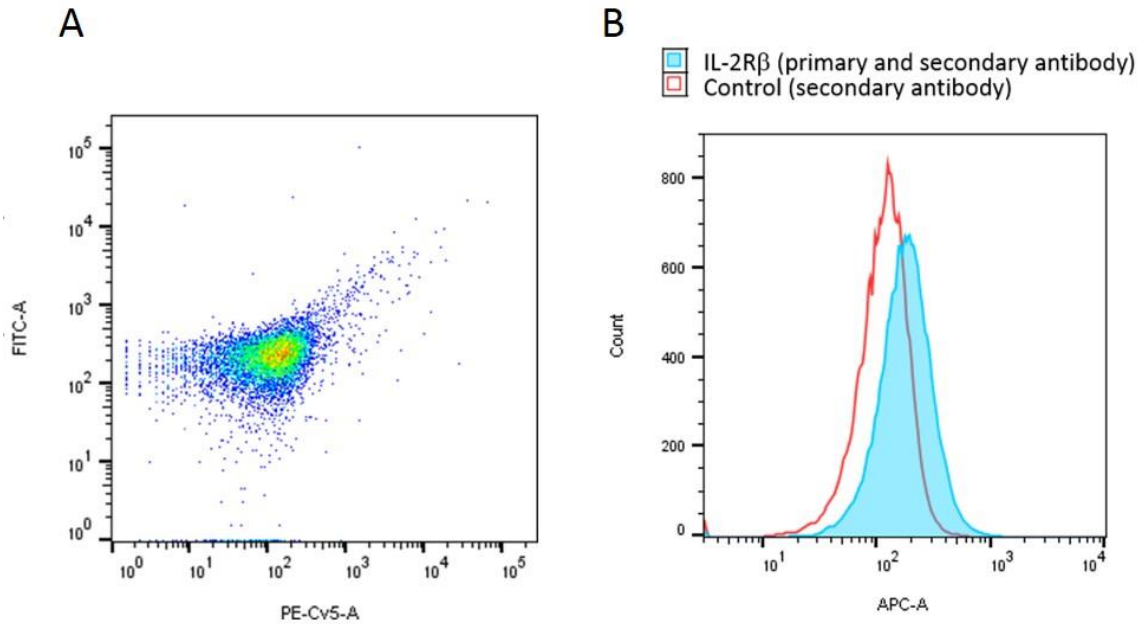

Figure 15. Expression range of IL-2Rα and IL-2Rβ protein at the cell membrane. Jurkat cells were stained with antibody against human IL-2Rα and IL-2Rβ . A) The relative intensity levels of FITC and Pe_Cy5 indicate the levels of IL-2Rβ and IL-2Rα, respectively, expressed on the same cells. B) The relative Alexa 647 intensity of cells indicate the expression range of available IL-2Rβ within the unstimulated Jurkat cell population.

5.3.3 Cellular Response to Constant IL-2 Stimulation Shows Variability within Jurkat Cell Population

Given the numerous combinations of pulses and recovery times possible to explore experimentally, we used the model simulations in Aim 2 to guide the selection of stimulatory conditions. To first explore how constant IL-2 stimulation (i.e. no recovery) would yield population-level responses, we subjected trapped cells to a steady flow of 10 pM and 100 pM IL-2, values that the model indicated would yield a below-saturation receptor-ligand complex formation. Cells subjected to 10 pM of IL-2 did not show any quantifiable translocation of STAT5-GFP *in vitro* in the one hour time scale of interest (results not shown) and further experiments were thus conducted using 100 pm IL-2. When exposed to a constant input of 100 pM, 33 % of cells showed STAT5-GFP translocation to the nucleus during the one hour course of the experiment, as indicated by increased ratio of nuclear to cytosolic GFP. These results highlight the importance of studying single-cell responses rather than relying on population average.

5.3.4 Cellular Response to Time-varying IL-2 Input Shows Heterogeneity in an *in vitro* Setting

In order to investigate live-cell response to pulsatile IL-2 inputs, Jurkat cells were trapped in a microfluidic device capable of delivering tightly controlled stimulus to all trapped cells simultaneously [113]. Cells were then exposed to varying IL-2 inputs and STAT5-GFP translocation was tracked over the course of an hour using fluorescent imaging. Pulsatile input settings were informed by the modeled response to inputs of varying lengths. Of the modeled responses, pulse and recovery time combinations of 30 seconds/30 seconds, one minute/one minute, two minutes/30 seconds, and five minutes/five minutes were experimentally tested. The model predicted similar maximum response for three of these four combinations as indicated by the number of receptor-

ligand complexes present at the cell membrane, leading us to ask whether the dynamic response of the cells would also be similar (Figure 9 and Figure 10, Chapter 4). After exposing Jurkat cells to the same pulsatile IL-2 stimulus *in vitro*, the resulting cell response, as indicated by STAT5-GFP translocation to the nucleus, show distinct profiles. We observed that cells exposed to IL-2 pulses of intermediate length (one and two minutes) followed by short recovery times (one minute and 30 seconds, respectively) (Figure 17 and Figure 18) responded more quickly than either cells exposed to short pulses with short recovery times (Figure 16) or cells exposed to long pulses and long recovery times (Figure 19). No statistical significance was found between the mean maximum ratio of nuclear to cytosolic STAT5-GFP for the responding cells between the four IL-2 input groups. The means of the AUC for the responding cells showed statistical significance using a one-way ANOVA with *post hoc* Tukey test ($p = 0.0211$), with significant difference of the mean seen between the 1 minute pulse 1 minute pause group and the 2 minute pulse 30 second pause group ($p = 0.213$) and the 5 minute pulse 5 minute pause group ($p = 0.0417$). Collectively, the results suggest that individual cell responses to pulsatile IL-2 stimulus vary within Jurkat cells, and that the response profiles are affected by variation in pulse length of recovery time, with a stronger and faster response occurring at intermediate pulse times compared to both short and long pulse and recovery times.

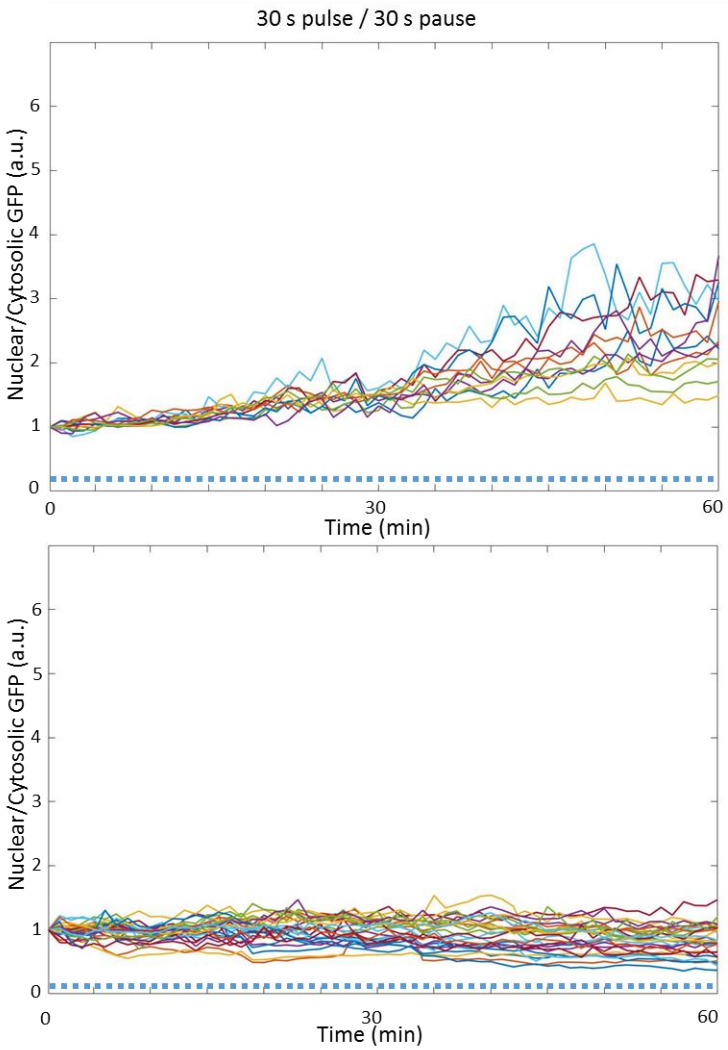

Figure 16. Nuclear translocation of STAT5-GFP upon short pulses of IL-2. Nuclear translocation of STAT5-GFP in Jurkat cells exposed to pulsatile input of 100 pM IL-2 with 30 s pulses and 30 s recovery pauses. Responding cells (top) show an increased ratio of nuclear to cytosolic GFP over to course of one hour. Cell traces were normalized to the initial time point. Pulses (blue) and pauses (white) are indicated along the x axis.

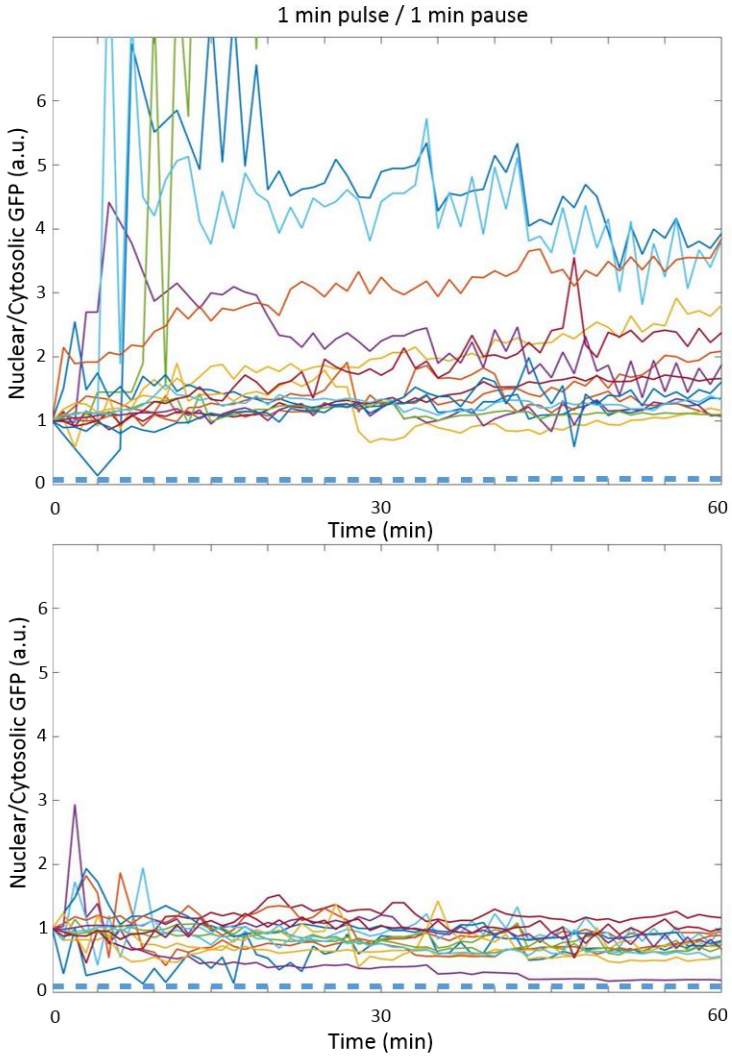

Figure 17. Nuclear translocation of STAT5-GFP upon 1 minute pulses of IL-2. Nuclear translocation of STAT5-GFP in Jurkat cells exposed to pulsatile input of 100 pM IL-2 with 1 minute pulses and 1 minute recovery pauses. Responding cells (top) show an increased ratio of nuclear to cytosolic GFP over to course of one hour. Cell traces were normalized to the initial time point. Pulses (blue) and pauses (white) are indicated along the x axis.

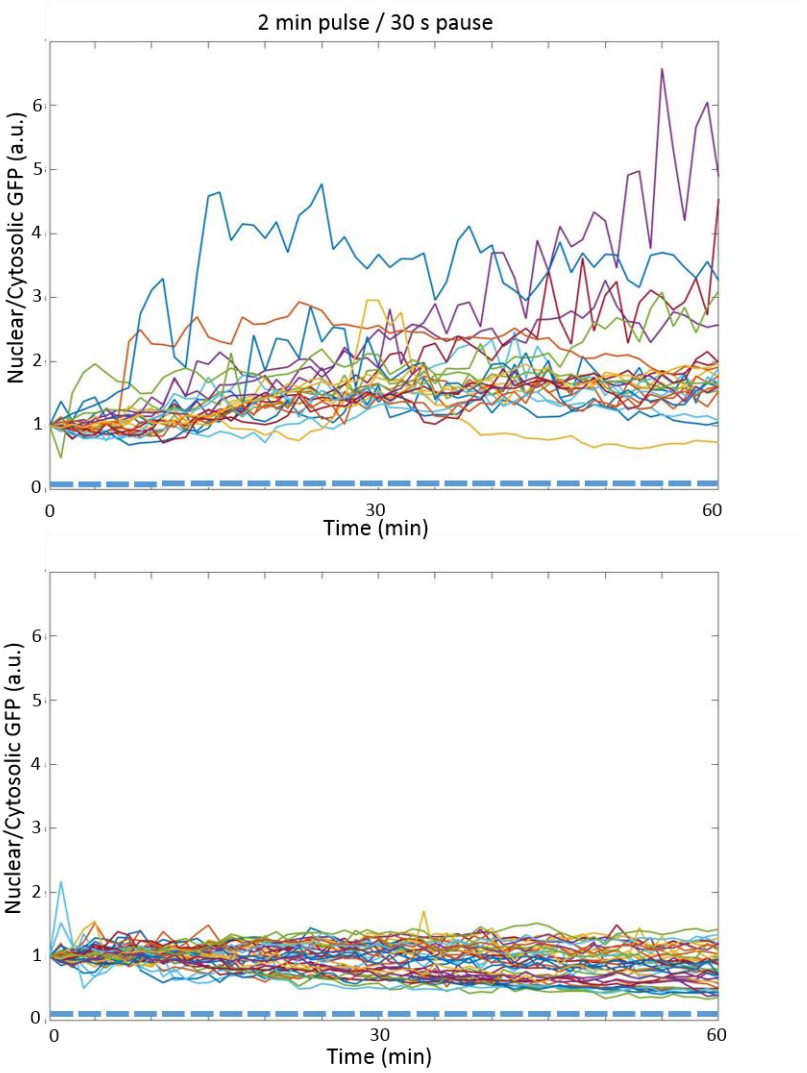

Figure 18. Nuclear translocation of STAT5-GFP upon 2 minute pulses of IL-2. Nuclear translocation of STAT5-GFP in Jurkat cells exposed to pulsatile input of 100 pM IL-2 with 2 minute pulses and 30 second recovery pauses. Responding cells (top) show an increased ratio of nuclear to cytosolic GFP over to course of one hour. Cell traces were normalized to the initial time point. Pulses (blue) and pauses (white) are indicated along the x axis.

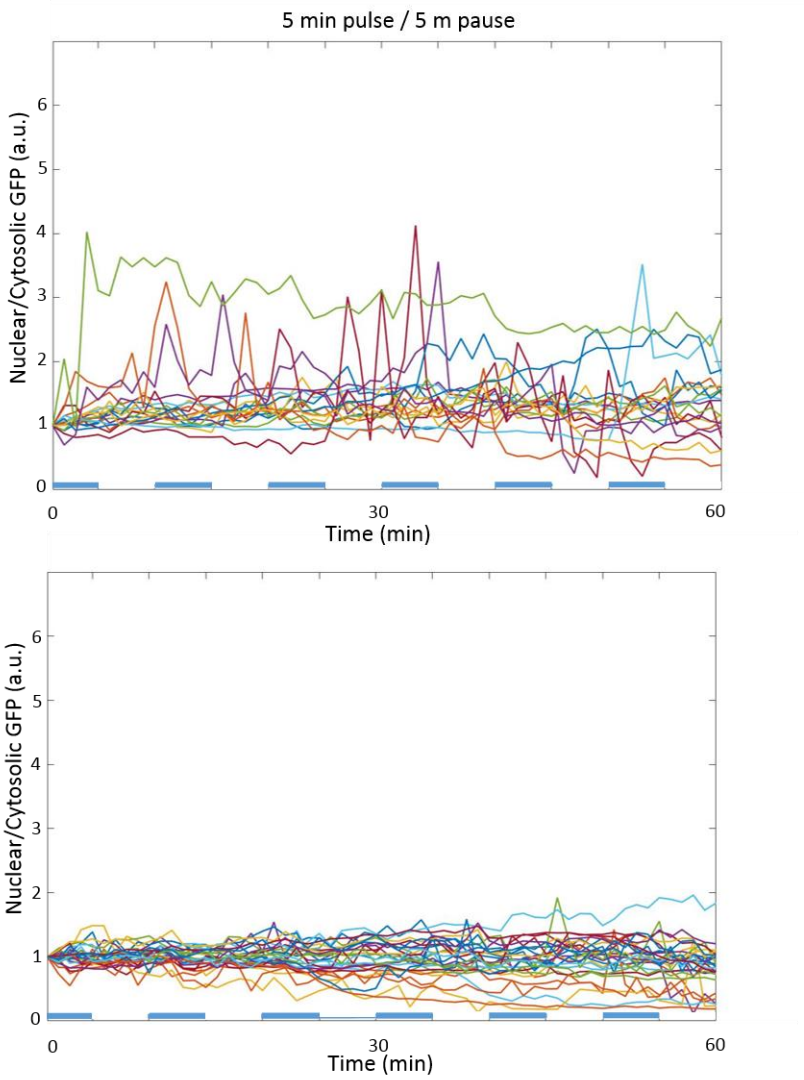

Figure 19. Nuclear translocation of STAT5-GFP upon long pulses of IL-2. Nuclear translocation of STAT5-GFP in Jurkat cells exposed to pulsatile input of 100 pM IL-2 with 5 minute pulses and 5 minute recovery pauses. Responding cells (top) show an increased ratio of nuclear to cytosolic GFP over to course of one hour. Cell traces were normalized to the initial time point. Pulses (blue) and pauses (white) are indicated along the x axis.

Table 4: Responding cells and average nuclear/cytosolic STAT5-GFP maximum and area under the curve for four time-dependent IL-2 stimulus settings.

Condition	30s/30s	1m/1m	2m/30s	5m/5m
Responding cells (%)	30	53	37	45
Average max norm ratio responders	2.7 ± 0.76	1.8 ± 1.02	2.1 ± 1.40	2.3 ± 0.88
AUC non-responders	56.2	50.72	57.2	56.2
AUC responders	98.5 ± 13.2	152.12 ± 159.3	71.8 ± 32.0	81.1 ± 21.5

5.4 Discussion

Our experimental microfluidic setup permitted investigation of single-cell responses to IL-2 in a controlled setting, by simultaneously collecting response data from multiple cells within a population through the use of arrayed cell traps, orthogonal buffer flow, and longitudinal imaging. When subjected to a constant input of 100 pM IL-2, the cells in the population showed a range of responses, with one third of cells showing a clear translocation of GFP-labeled STAT5 from the cytosol to the nucleus over the one hour time course. This range of responses to a uniform ligand condition highlights the importance of studying single-cell responses within populations rather than relying on population averages in order to fully understand the functionality of immune cell populations. The heterogeneity observed here suggests that varying levels of IL-2 receptor subunits at the single-cell level can have functional consequences with regards to filtering dynamic IL-2 cues in the extracellular environment.

The initial resting ratios of nuclear to cytosolic STAT5-GFP in Jurkat cells indicate that STAT5 is present in their nucleus prior to experimental IL-2 stimulation, albeit at levels markedly lower than in the cytosol. This presence of nuclear STAT5 is presumably due to the effects of IL-2 exposure in culture, where Jurkat cells constitutively produce and release IL-2. This constitutive response to IL-2 results in a low level of downstream response with phosphorylation and translocation of cytosolic STAT5. Our population-averaged probing of both unlabeled and labeled STAT5 shows

that the vast majority of STAT5 exists in the cytosol prior to IL-2 stimulation. We note also that based on our fluorescent images, the relative amount of STAT5-GFP in the nucleus varied between cells at resting state. No correlation (R^2 = 0.15) was found between resting ratio of nuclear to cytosolic STAT5-GFP and the level of the IL-2 induced response for individual cells.

Cell-to-cell variability in signaling is commonly thought of as the result of accumulation of variation in protein at each level of a biochemical cascade. An alternate hypothesis to explain this variability states that rather than being caused by random variations in gene expression, it is caused by cellular convergence to specific attractor states within cell state space [118]. Such clustered heterogeneity could indicate the existence of distinct kinetic profiles within the cell population, allowing for the fulfillment of different functional needs. The possibility that distinct cellular response states coexist in genetically identical cell populations prior to stimulation is an intriguing pendant to the discussion of cell-to-cell heterogeneity within immune cell populations. Our results show distinct phenotypic subgroups within our cell population, with different response profiles to identical inputs. High responder cells showed initiation of quantifiable STAT5-GFP translocation rapidly after the initial IL-2 pulse for the two intermediate pulse times of one and two minutes, indicating that these pulse times in combination with their respective recovery periods of one minute and 30 seconds provide a favorable input to fast cell response to extracellular IL-2. Interestingly, both shorter (30 seconds) and longer (five minutes) pulse and recovery times resulted in a slower initiation of STAT5-GFP translocation for high-responding cells. This raises the possibility that the existence of different cell states within the population enables the cells act as filters to initiate response within a preferred range of cytokine fluctuation while filtering out IL-2 input pulses that fall outside this range.

The dynamics of receptor-ligand interaction in the IL-2 system, coupled with the relative speed of the downstream processes, suggests that this system could allow cells to

distinguish between fluctuating pre-equilibrium (below steady-state) levels of ligand [39]. Our experimental design let us test this prediction by enabling us to tightly control extracellular IL-2, in a manner that let us investigate the effects of a pulsatile IL-2 input on the receptor-ligand complex dynamics in Jurkat cells. We exposed cells from the same population to IL-2 stimulus at varying range of combinations of stimulus pulse and recovery times, while simultaneously tracking downstream response for individual cells using fluorescent imaging of STAT5-GFP. The results provide a distribution of different response profiles within the population for all input settings. Interestingly, our results also show differences in the onset of cell response and the spread of response strength as the pulsatile input was varied. Notably, short IL-2 pulses of 30 seconds followed by short 30 second recovery times showed a population profile similar to that seen when both IL-2 pulses and recovery times were long (5 minutes). In contrast, intermediate pulse and pause times showed a faster onset of downstream response in the population, accompanied by a wider range of single-cell response profiles. This indicates that Jurkat cells do indeed respond to pulsatile IL-2 input in a manner that is able to distinguish between pulsatile variations in extracellular IL-2 levels at below steady-state concentrations. This is consistent with the prediction that the kinetics of the IL-2 receptor-ligand pair enables increased cellular sensitivity to fluctuating cytokine levels.

In an in vivo context, T cell response to IL-2 happens in a system subject to many control switches, such as *a priori* upregulation of IL-2Rα, and ligand competition both within the T cell population and between other cell types. By using a cell line which is pre-primed for IL-2Rα expression, we did not have to account for the need for IL-2Rα recruitment in order to initiate IL-2 response in the cells. In addition, our experimental setup allowed for precise control of IL-2 exposure while minimizing the number of biological control switches. Thus, our results indicate cell response to IL-2 stimulation in a context where the main contributors to cells response were IL-2 delivery and preexisting cell state of the responder cell.

Previously published data report a range of expression of IL-2 receptor subunits within T cell populations [6, 56]. While the expression levels of IL-2Rα are known to be affected by IL-2 due to priming of cells which initiated IL-2Rα upregulation, we used a cell line that is constitutively primed to express IL-2Rα. This allowed us to ask how levels of other IL-2 receptor subunits might alter cell responsiveness under primed conditions. In order to investigate the effects of initial cell-to-cell variability in receptor level, we modeled cells with high or low levels of these subunits and found that under primed conditions, the level of IL-2Rβ and IL-2γ became an indicator of cellular responsiveness to IL-2 as measured by downstream translocation of STAT5.

In conclusion, we have investigated single-cell responses to variable pulsatile IL-2 input, using a microfluidic platform in combination with time-lapse fluorescent imaging. Our results indicate that cell response profiles vary with varied pulsatile input at pre-equilibrium levels. By using microfluidics, we were able to investigate both specified cell-cell variability and the spatiotemporal effects of controlled IL-2 stimulation. This approach helped to increase our understanding of how cell-to-cell variability affects the range of cytokine responses within an immune cell population. Our experimental setup allowed us to focus on the cellular response at the single-cell level, in a manner that complemented the model developed in Chapter 4, by highlighting the variability of the *in vitro* cell response within a population using model-informed input settings. The difference in response onset time seen in populations responding to varying combinations of IL-2 pulse times and recovery times indicate that Jurkat cells have a preferred range of extracellular IL-2 fluctuations in which a cellular response is initiated quickly, while cells are slower to respond and show lower response levels outside this fluctuation range. Further investigation into this filtering behavior could increase our understanding of how variability within immune cell populations enable a systems response within preferred fluctuation ranges and whether these ranges correspond to *in vivo* conditions.

CHAPTER 6 CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

In this work, we have explored cellular heterogeneity in immune cell populations and the dynamics responses to stimuli based upon this heterogeneity. Our approach incorporated both computational and experimental approaches, taking advantage of technical progress in the area of single-cell gene expression analysis and microfluidics. In Aim 1, we showed that gene expression variability within immune cell populations can be used as an indicator of novel subgroups. Our data also illustrated the importance of the choice of analytical methods in interpreting single-cell transcriptomic data. While gene expression variability can predict functional differences between immune cell subgroups, cellular function is often defined by a cell's interactions with other cells. In order to address this, we studied single-cell responses to IL-2, one of the most important signaling molecules in the context of T cell functionality. Through computational modeling, we sought to investigate the effects of cellular heterogeneity on response to IL-2. A microfluidic device enabling the capture and imaging of suspension cells in combination with tightly regulated IL-2 stimulus allowed us to test model predicted ranges of interest for pulsatile IL-2 input. This combinatorial approach allowed us to achieve a more detailed view of T cell response to biologically relevant fluctuations in extracellular cytokine.

Single-Cell Gene Expression Analysis of Primary Immune Cells

Our initial study (Aim 1) examined heterogeneity in primary immune cell populations, and asked the question whether single-cell gene expression analysis could help in the discovery of distinct subgroups. Our results indicate that this approach can enable the detection of distinct cellular subgroups within populations, enabling grouping based on both cell surface markers and intracellular targets such as transcription factors and signaling proteins.

Interpretation of Single-Cell Data

Our systematic testing of analysis methods for single-cell gene expression data (Aim 1) showed the effect that methods choice can have on interpretation of this data type. This illustrates the importance of choosing a data analysis method that is suitable for single-cell data. Since the completion of our study which was carried out using the Fluidigm qRT-PCR system, newer technologies have been developed, allowing for larger sets of single-cell gene expression data to be collected in parallel. The expanding capabilities of RNA-seq in particular provide an improved dynamic range as compared to microarray-based techniques and RNA-seq has become a leading technology of choice [120, 121]. Despite the rapid progression of technologies that enable collection of such data, to date there is no consensus on the best analytical method. More work is needed in this area in order to establish standards for data analysis, taking into account that a one-size-fits-all approach may not be attainable for this type of data. As illustrated by our two data sets, representing two different populations of primary immune cells, characteristics, such as expected population variability based on previously known subgroups and function may be important factors to consider. Comparative studies incorporating more cell types and data sets of larger size would be useful for establishing standard data analysis methods. Although our study was carried out using data from the Fluidigm platform, the importance of establishing proper data analysis methods for single-cell gene expression data holds true across technologies. Technical variability is unavoidable and the as each cell in single-cell data represents one unique batch, bulk-based methods for minimizing technical noise such as implementation of technical replicates cannot be employed. This means that technical variability must be taken into account for sample processing of all single-cell gene expression data in order to ensure correct data interpretation.

Computational Modeling of T cell Response to IL-2

We developed a model to investigate the dynamic response of T cells to pulsatile IL-2 input. The model expanded on previously developed models by incorporating subunit level detail of the IL-2 receptor, as well as by interrogating the initial response to cytokine at the single-cell level during the course of one hour, a time scale that minimized effects of IL-2 induced upregulation and enabled us to investigate what effect preexisting cell states within a population had on IL-2 response. Furthermore, the model focused on a cellular environment scaled to a microfluidic experimental setup, allowing us to replicate the model input *in vitro* at the single-cell level, rather than using a more conventional bulk population approach. The model also enabled us to specify both cell-to-cell variability and pulsatile input of extracellular IL-2 available to the cells. The results offered new insights into the effects of cell-to-cell variability within immune cell populations on the response to fluctuations of extracellular cytokine levels. By incorporating subunit detail and the stepwise formation of the receptor ligand complex, we added more information that corresponded to the kinetic steps involved in mediating cell signaling *in vitro*. It could be argued that even more steps need to be added in order to fully model all known steps of complex formation. However, we feel that the simplification we included with regards to the IL-2Rβ and IL-2γ subunits was justified based on previously published models which compared the effect of including or excluding the individual IL-2Rγ binding step and found the different negligible. The first of our two model outputs, receptor-ligand complex number, was not experimentally verified as methods for such verification such as antibody binding were thought to interfere with the binding steps. A downstream response was added to our model in the form of STAT5 nuclear translocation, which was represented by a single delay function. STAT5 translocation provided a model output which could be experimentally validated; however it is an indirect measurement of upstream receptor-ligand ligation events. While the modeled STAT5 translocation step was fitted to experimental data, this was a simplification of a multistep phosphorylation cascade, and additional steps based on

dynamic experimental data from the intermediary steps could be useful in order to further improve the timing aspect of modeled predictions of translocation.

By taking a deterministic approach to incorporating cell-cell variability, we could specify initial receptor subunit settings at the single-cell level and predict downstream effects. An alternate approach would be to allow for setting initial single-cell values in a stochastic manner. This aspect could be added to the model by allowing for sampling within a known range of expression for model species of interest, e.g., by experimentally determining the numbers of receptor subunits within the cell population of interest and allowing for sampling within this dataset.

Using Microfluidics to Investigate Single-Cell Response to IL-2

We next sought to explore T cell response to IL-2 input experimentally. By using a previously developed microfluidic cell trap along with a tightly controlled stimulus delivery setup, we were able to capture the dynamic behavior of individual cells from the same population. By using this method, we were able to stimulate cells with biologically relevant concentrations of IL-2, rather than large doses, eliminating the risk of cell toxicity. It also allowed us to tightly regulate the time of stimulus delivery, rather than rely on diffusion in a larger culture volume.

Heterogeneity in T cell Response to IL-2

The Jurkat T cell line was transfected to stably express a fluorescently labeled target protein downstream of the IL-2 receptor. Fluorescent microscopy allowed us to track cellular response to the stimulus in real time and custom Matlab code for image analysis let us quantify this response to compare individual cell behavior. This approach enabled us to gain new insight into the variability of cell behavior within an immune cell population, hinting at the possibility of preexisting cellular states within the population that enable variability in response to an identical stimulus. Our measured levels of IL-

2Rβ expression within the Jurkat cell population prior to stimulation indicated a preexisting variability in subunit levels available for ligand binding; however a direct correlation between subunit levels and downstream response could not be verified on the individual cell level due to technical limitations. Our experimental platform allowed for measurement of subunit level by on-chip antibody staining, but this method was not optimal due to the risk of antibody binding interfering with receptor-ligand formation. An alternate method for directly quantifying receptor subunit levels in the responding cells would be useful in order to experimentally verify our conclusion that IL-2Rβ is a rate-limiting factor in the response of pre-primed T cells to IL-2. Cells expressing fluorescently labeled protein or a comparison of cells with up- or downregulated receptor subunits could provide such a method.

T cell response to Pulsatile IL-2 Stimulus

By combining the microfluidic cell trap with customizable input pulses, we investigated Jurkat T cell response to pulsatile IL-2 input under model-informed stimulus settings. Our results indicated that the cells had a preferred range of pulsatile IL-2 stimulus input in which downstream response was increased. While we did not test other stimuli, a previously published study found a preference for a frequency of 2.78 mHz in Jurkat cell response to fluctuating H_2O_2 [38]. Together with our results showing a preference for input pulses of one minute, this indicates that Jurkat cells exhibit different bandpass filtering characteristics for different stimuli. Our results also show not only a frequency preference but also an effect on cell response to IL-2 due to different lengths of cell recovery between stimulus pulses.

6.2 Future Directions

6.2.1 Single-Cell Gene Expression as a Tool for Detecting Distinct Subgroups of Immune Cells

In our work, we showed that high throughput gene expression analysis can be employed to discover distinct subgroups of immune cells that are separated by markers other than the traditional surface markers. With the rapid development of more sophisticated techniques for single cell genomics and transcriptomics [47-50], future work using this approach is sure to include even greater level of detail. Our comparison of cellular subgroup representation between individuals showed a difference between healthy donors. The inclusion of more cell types in the search could further our understanding of individual differences in immune system functionality and a complementary consideration when choosing gene targets makes it possible to infer systems level information regarding the communication between different cellular subgroups. While we compared only healthy individuals, this approach could also be useful to determine causes of dysregulation in disease states, which would be beneficial for the development of accurate disease models. In addition this approach should of interest for drug development research due to the implications of single-cell heterogeneity in differential response to treatment.

6.2.2 The Existence of Cellular States within T cell Populations

The results from our study of T cell response to IL-2 hint at the possibility of distinct cellular states existing in T cell populations. Future work should aim to characterize pre-existing cell states within T cell populations and how these affect immune response. Single-cell genomics [122], mass cytometry [123], and kinetic parameters search [118] have recently been employed to defined single-cell states, each method capturing a different aspect of such subclassification of populations. A future computational modeling approach could be taken to assess the existence of attractor states within T cell

populations such as developed for calcium responses in ATP-stimulated epithelial cells. Following the methods of Wollman and colleagues [118], deterministic models of IL-2 signaling could be individually optimized for parameters that yield single cell dynamics by inferring the full distribution for each parameter through Bayesian approaches. Clustering of the resulting parameter sets could reveal groupings of numerical values that explain responder/non-responder behavior. Finally, the addition of primary cell information would serve to further compare the functional consequences of such cellular states within the adaptive immune system where direct comparisons between parameters and proteomic information are possible.

6.2.3 Preferred Range of Pulsatile IL-2 Stimulus for Optimal Cell Response

Our computational model predicted different cellular responses to pulsatile IL-2 based on the duration of stimulus pulses and recovery times that the cells were exposed to. Our experimental testing of these predictions confirmed that cell response will vary with pulsatile IL-2 input and also indicated the existence of a preferred range of IL-2 fluctuations in which the cells exhibited more rapid and stronger downstream response to IL-2, independent of concentration. While we did not explore a wide range different IL-2 concentrations, this aspect could be investigated using our experimental platform to test the latter prediction. Future work should also explore this observed behavior to get a more detailed view of this effect and to determine its biological relevance. Questions of interest include whether the observed range of IL-2 fluctuation is of physiological importance, a question that could be investigated by determining IL-2 fluctuations in vivo. Whether cell secretion of IL-2 by IL-2 producing cells matches the preferred fluctuation seen in our responding cells could be experimentally tested using an experimental platform such as one developed by the Tay group which allows single cell measurement of dynamic cytokine secretion [54]. Additionally, the downstream effects of the observed response could be investigated to determine whether up- or down regulation

of genes influenced by IL-2 response will differ in correlation with the heterogeneous cell behavior seen, and what implications this will have for immune system functionality on a longer time scale. An experimental platform that allows for collection of both cell response and single-cell transcriptomics data from individual cells would be a useful tool for investigating this downstream effect.

APPENDIX

A.1. Appendix for Chapter 3

A.2. Target Genes for Single Cell qRT-PCR

Table 5: Target genes for qRT-PCR of primary T cells.

Gene name	Gene Product
AKT3	Akt3
ATP2B1	PMCA
B3GAT1	CD57
BCL2	Bcl-2
CAMK4	Calcium/calmodulin-dependent protein kinase type IV
CCR5	C-C chemokine receptor type 5
CCR7	C-C chemokine receptor type 7
CD27	CD27
CD28	CD28
CD3D	CD3 delta
CD4	CD4
CD40LG	CD40 ligand
CD8B	CD8, beta
CDKN2A	Cyclin-dependent kinase inhibitor 2A, p16ink4A
CREB1	CAMP responsive element binding protein 1
CTLA4	CD152
CXCR3	Chemokine (C-X-C motif) receptor 3
CXCR4	Chemokine (C-X-C motif) receptor 4, CXCR4
CXCR6	Chemokine (C-X-C motif) receptor 6
CXCR7	Chemokine (C-X-C motif) receptor 7
DUOX1	Dual oxidase 1
DUSP2	Dual specificity protein phosphatase 2
DUSP3	Dual specificity protein phosphatase 3
FAS	CD95
FASLG	Fas ligand
FCAR	CD89
FYN	Proto-oncogene tyrosine-protein kinase Fyn
GLRX2	Glutaredoxin 2
GZMB	Granzyme B
HIF1A	Hypoxia-inducible factor 1-alpha, HIF-1alpha
IFNGR1	Interferon gamma receptor, IFN gamma receptor
IL10	Interleukin-10, IL-10
IL10RA	Interleukin-10 receptor, alpha, CDw210
IL10RB	Interleukin-10 receptor, beta
IL12RB2	Interleukin 12 receptor, IL-12 receptor
IL17A	Interleukin 17, IL-17

IL18R1	Interleukin 18 receptor 1, IL-18 receptor 1

Table 5 continued.

IL1R1	CD121a (IL-1RI and IL-1RII)
IL2	Interleukin-2, IL-2
IL21R	Interleukin-21 receptor, IL-21 receptor
IL2RA	CD25
IL2RB	Interleukin-2 receptor, beta, Interleukin-15 receptor
IL2RG	Interleukin-2 receptor, gamma, CD132
IL4R	Interleukin-4, IL-4
IL6	Interleukin 6, IL-6
IL7R	Interleukin-7 receptor, alpha, IL-7 receptor
ITK	IL2-inducible T cell kinase
JAK1	Janus kinase 1
JAK3	Janus kinase 3
KLRC1	Killer cell lectin-like receptor subfamily C, member 1
LAG3	CD223
LAT	Linker for Activation of T cells, LAT
LCK	Lymphocyte-specific protein tyrosine kinase
LTA, TNFB	Tumor Necrosis Factor beta, TNF beta
MAP3K5	Apoptosis signal-regulating kinase 1
MAPK1	Mitogen-activated protein kinase 1, ERK2
MAPK9	Mitogen-activated protein kinase 9, JNK2
MPO	Myeloperoxidase
NFE2L2	Nuclear factor (erythroid-derived 2)-like 2
NPAT	CD245
NRAS	N-Ras
ORAI1	Calcium release-activated calcium channel protein 1
PIK3CD	PI3K, catalytic, delta polypeptide
PLCG1	PLC gamma
Prf1	Perforin 1
PRKCQ	PKC theta
PTEN	Phosphatase and tensin homolog, PTEN
PTPN1	Protein-tyrosine phosphatase 1B, PTP1B
PTPN11	Tyrosine-protein phosphatase non-receptor type 11, SHP-2
PTPN22	Protein tyrosine phosphatase, non-receptor type 22, LYP
PTPN6	Src homology region 2 domain-containing phosphatase-1, SHP1
RAC2	Ras-related C3 botulinum toxin substrate 2, Rac2
RPP38	p38
SH2D1A	SH2 domain-containing protein 1A, sphingolipid activator protein-1, SAP
SOD2	Superoxide dismutase , SOD2
SOS2	SOS-2
STAT3	Signal transducer and activator of transcription 3, STAT3
STAT5a	Signal transducer and activator of transcription 5A, STAT5a
STAT5b	Signal transducer and activator of transcription 5B, STAT5b
STAT6	Signal transducer and activator of transcription 6, STAT6
STIM1	Stromal interaction molecule , STIM1
TCRG	Tcell Receptor, gamma, TCR gamma
TERF1	Telomeric repeat-binding factor 1, TERF1

TGFB1	TGF beta 1

Table 5 continued.

TGFBR2	TGF beta receptor 2
TLR4	Toll-like receptor 4, TLR4
TLR5	Toll-like receptor 5, TLR5
TLR7	Toll-like receptor 7,TLR7
TNF, TNFA	Tumor Necrosis Factor alpha, TNF alpha
TNFRSF8	CD30
TNFRSF9	4-1BB, CD137
TRAF6	TNF receptor associated factor 6, TRAF6
TXN2	Thioredoxin 2, Trx2
VAV1	VAV1
WNT3	WNT3
ZAP70	Zeta-chain-associated protein kinase 70, Zap70

Table 6: Target genes for qRT-PCR of primary neutrophils.

Gene name	**Gene Product**
AKT1	V-akt murine thymoma viral oncogene homolog 1 (AKT1)
AQP9	Aquaporin 9
B2M	MHCI beta microglobulin
BCLAF1	BCL2-associated transcription factor 1
BIN1	Bridging integrator 1
BTK	Bruton agammaglobulinemia tyrosine kinase
C5AR1	Complement component 5a receptor 1
CASP8	Caspase 8
CASP9	Caspase 9
CD14	CD14
CD4	CD4
CD46	CD46
CD83	CD83
CDC42	CDC42
CSF2RA	CD116
CTNNB1	Beta-catenin
DDX58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58
DICER1	Dicer 1, ribonuclease type III
DYRK1A	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A
EIF2AK2	Eukaryotic translation initiation factor 2-alpha kinase, PKR
EIF4B	Eukaryotic translation initiation factor 4B
EIF4G1	Eukaryotic translation initiation factor 4 gamma 1
ELK1	ETS domain-containing protein Elk-1
FBXL5	F-box and leucine-rich repeat protein 5
FCGR1A	CD64
FCGR2A	CD32
FCGR3B	CD16b
FOS	c-Fos (FBJ murine osteosarcoma viral oncogene homolog)

GRB2	Growth factor receptor-bound protein 2

Table 6 continued

GSK3B	Glycogen synthase kinase 3 beta
HERC5	Hect domain and RLD 5
HERPUD2	Homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 2 protein
HIF1A	Hypoxia-inducible factor 1-alpha
HLA-A	MHCI alpha chain
HLA-DRA	HLA class II histocompatibility antigen, DR alpha chain
HNRPK	Heterogeneous nuclear ribonucleoprotein K
HSD17B11	Hydroxysteroid (17-beta) dehydrogenase 11
HSPB1	Heat shock protein beta-1, Heat shock protein 27 (Hsp27)
IFIT1B	Interferon-induced protein with tetratricopeptide repeats 1B
IFIT2	Interferon-induced protein with tetratricopeptide repeats 2
IFNGR1	Interferon gamma receptor 1
IL17RD	Interleukin 17 receptor D
IL1B	Interleukin 1, beta
IL2RB	Interleukin 2 receptor, beta
IL6R	IL-6 receptor
IL8	IL-8 (CXCL8)
IL8RA	IL-8 receptor, alpha (CXCR1)
ILK	Integrin-linked kinase
IMP3	U3 small nucleolar ribonucleoprotein protein IMP3
IRAK1	Interleukin-1 receptor-associated kinase 1
IRF1	Interferon regulatory factor 1
CD44	CD44
ITGAL	CD11a (Integrin alpha L chain0
ITGAM	Integrin Alpha-M (CD11b)
ITGB2	CD18 (Integrin, beta 2)
MAPK3	Mitogen-activated protein kinase 3
MBOAT7	Membrane bound O-acyltransferase domain containing 7
MCL1	Myeloid cell leukemia sequence 1 (BCL2-related)
MYD88	Myeloid differentiation primary response gene (88)
NDEL1	NudE nuclear distribution gene E homolog (A. nidulans)-like 1
NFKB1	NF Kappa B
NUMB	NUMB homolog
OASL	2'-5'-oligoadenylate synthetase-like
OCIAD1	OCIA domain containing 1
OR2W3	Olfactory receptor, family 2, subfamily W, member 3
PAK1	Serine/threonine-protein kinase PAK 1
PDPK1	3-phosphoinositide dependent protein kinase-1
PHC2	Polyhomeotic homolog 2
CREB	cAMP response element-binding protein
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma isoform
PTEN	Phosphatase and tensin homolog
PTPN11	Tyrosine-protein phosphatase non-receptor type 11
RAC1	Ras-related C3 botulinum toxin substrate 1

RHEB	GTP-binding protein Rheb (Ras homolog enriched in brain)

Table 6 continued.

RHOA	Ras homolog gene family, member A
SAE1	SUMO1 activating enzyme subunit 1
SAMD9L	Sterile alpha motif domain containing 9-like
SELENBP1	Selenium binding protein 1
SELL	CD62 ligand
SERINC3	Serine incorporator 3
SHC1	SHC-transforming protein 1
SP3	Sp3 transcription factor
SRF	Srf serum response factor
STK17B	Serine/threonine kinase 17b
STX3	Syntaxin 3
STXBP3	Syntaxin binding protein 3
TLR4	Toll-like receptor 4
TLR6	Toll-like receptor 6
TLR8	Toll-like receptor 8
TRAF6	TNF receptor-associated factor 6
TRIM33	Tripartite motif containing 33
TSC1	Tuberous sclerosis protein 1
TSEN34	tRNA splicing endonuclease 34 homolog
WASPIP	WAS/WASL interacting protein family, member 1
ZDHHC18	Zinc finger, DHHC-type containing 18
ZFAND5	Zinc finger, AN1-type domain 5

A.3. Gene and Cell Exclusion by Primary Analysis Method

Table 7: The numbers of cells and genes excluded for each combination of primary analysis methods.

	Neutrophils		T Lymphocytes
Raw Data Cells Genes	 220 93	Raw Data Cells Genes	 251 94
Supervised Cells Genes	 202 59	2 Control Genes: Average Cells Genes	 244 38
2 Control Genes * Cells Genes	 208 62	2 Control Genes*: Missing Cells Genes	 244 38
All Inclusive Cells Genes	 220 81	All Inclusive Cells Genes	 247 94

A.4. Hierarchical Clustering of T cell Data

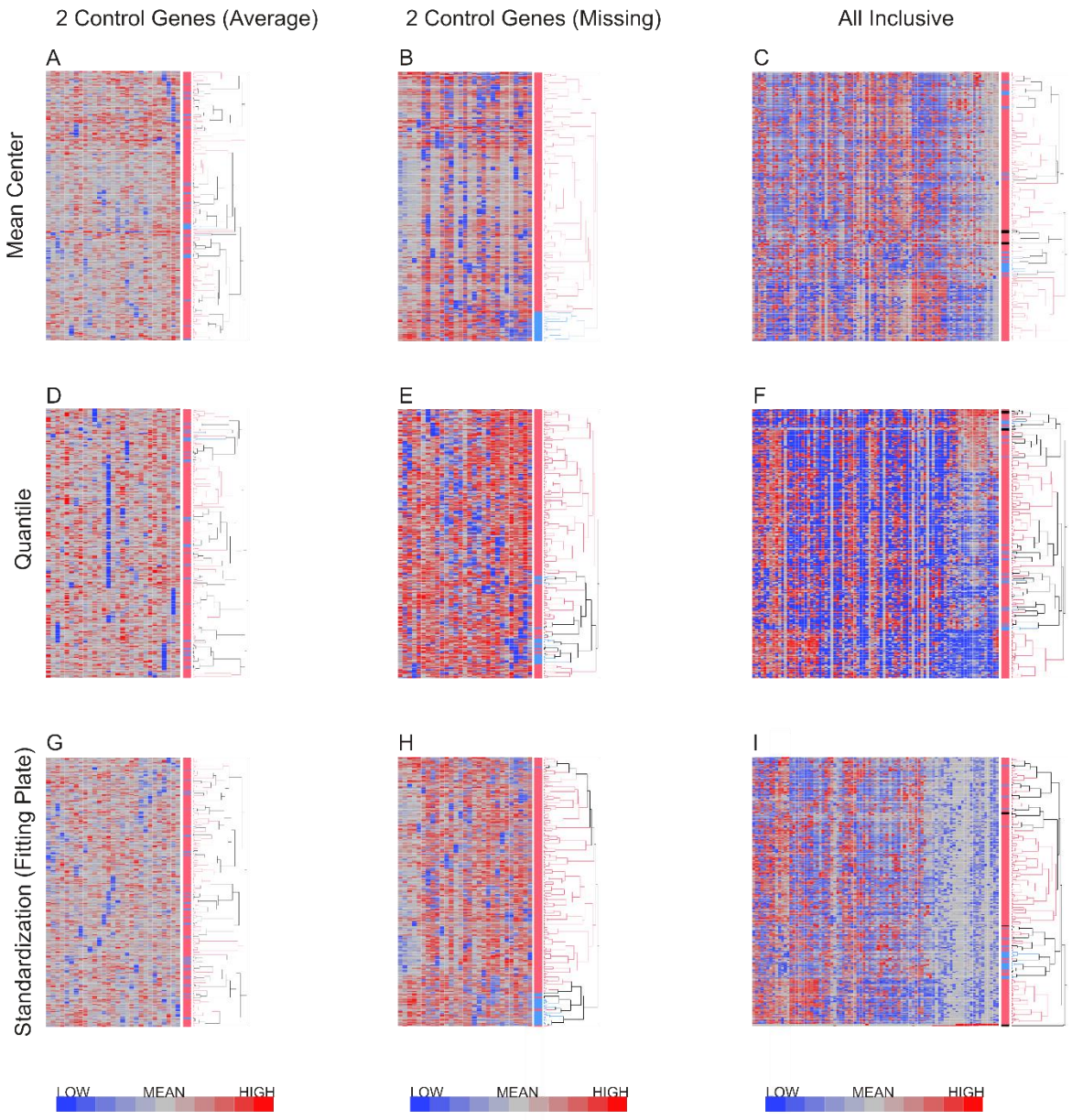

Figure 20. Hierarchical clustering of T cell data after nine combinations of primary analysis.

The primary data analysis method chosen affects the subsequent results for T cells.

A.5. Appendix for Chapter 4

A.6. Computational Model

Table 8: Model species.

Name	Initial value	Definition
L	100	Ligand
Cs	10	Receptor-ligand complex at cell surface
Ri	300	Intracellular receptor (lumped parameter)
Li	0.01	Intracellular ligand
Ci	1	Internalized Cs
Ld	1	Degraded ligand
Y	4.22E+09	System volume
A	14996.3	IL-2Ralpha
BG	1500	IL-2RBetaGamma (limped parameter)
LA	1	A binding L

Table 9: Model reactions.

Process	Equation
null -> L	(koff1 * LA) * Y * Vol / NA
null -> Cs	(kon2 * LA * BG)
null -> Ri	(kt * BG + kre * Ci)
null -> Li	((kre * Ci) / (Ve *NA))
null -> Ci	(kfe * Li * Ri + ke * Cs)
L -> null	kon1 * L * A * Y / NA
Cs -> null	(kr * Cs) + (ke * Cs)
Ri -> null	(kfe * Li * Ri + kh * Ri)
Li -> null	((kfe * Li * Ri) / (Ve *NA)) + (kx * Li)
Ci -> null	(kre + kh) * Ci
null -> A	(koff1 * LA) + (ksyn * Cs) + Vs + (kx * Li)
A -> null	(kon1 * L * A) + (kt * A)
null -> LA	(kon1 * L * A) + (kr * Cs)
LA -> null	(koff1 * LA) + (kon2 * LA * BG)
BG -> null	(kon2 * LA * BG) + (kt * BG)
null -> BG	kr * Cs + Vs

A.7. Appendix for Chapter 5

A.8. Code for Image Analysis Step 1 Data Acquisition

```
%%% Linda's code for analyzing gfp in both whole cell and nuclei
%%%%STEP1 - ENTER FILE NAME FOR RESULTS FILE
%Indicate what filename you would like to save the results as:
clear
Filename =
['20170208_100pM_1min1min_Jurkat_gfp_XY2_',datestr(now,30),'.mat'];

%%% Tell the code what images to use for analysis.
if ~exist('NbImages', 'var')

    %%%%STEP2 - ENTER NUMBER OF IMAGES IN YOUR SERIES
    NbImages = 62;                              %number of images
    Im = cell(1,NbImages);    %cell(n) returns an n-by-n cell array of
empty matrices for the gfp images
    Im2 = cell(1,NbImages);   %returns array for the DAPI images

    %%%%STEP3 - CHANGE PIXEL RANGE IF NECESSARY
    % This defines the size of indivudual cells to be detected
    pixelarearange = [30,900];              %Range of pixel area
values, used below with bwareafilt


%%%%**********************************************************************
******
    %%%If unsure how to filter sizes by pixel range, uncomment below,
and the cells' sizes will be graphed:
%     Area1 = regionprops(bw,'Area');
%     Area1e = extractfield(Area1,'Area');
%     figure()
%     hist(Area1e,100)
%     title('Histogram of ALL Cell Sizes')
%     xlabel('Cell Area Size (in pixels)')
%     ylabel('Count')

%**********************************************************************
********


    %%%%STEP4 - MAKE SURE FILES ARE READ IN CORRECTLY NAMED
        % Put a %d in place of the changing number in the file names
e.g. trapt01xy1.tif becomes 'trapt0%dxy1.tif'
    FNAMEFMT1_1 = 'T0000%dC01Z001.tif'; %Change these to reflect tif
images from Volocity export. First 9 images in gfp set
    FNAMEFMT1_2 = 'T0000%dC02Z001.tif'; % First 9 images in DAPI set

    FNAMEFMT2_1 = 'T000%dC01Z001.tif'; % Image 10-99 in gfp set
    FNAMEFMT2_2 = 'T000%dC02Z001.tif'; % Image 10-99 in DAPIset

    FNAMEFMT3_1 = 'T00%dC01Z001.tif'; % Image 10-99 in gfp set
    FNAMEFMT3_2 = 'T00%dC02Z001.tif'; % Image 10-99 in DAPIset
```

```
    for i=1:9                                       % For the first 9
images
      Im{i} = imread(sprintf(FNAMEFMT1_1,i));      % Fluorescent images
of first color (eg gfp)
      Im2{i} = imread(sprintf(FNAMEFMT1_2,i));     % Second color (eg
DAPI)
    end

    for i = 10:NbImages                           % For images 10-99
      Im{i} = imread(sprintf(FNAMEFMT2_1,i));   % Gfp images
      Im2{i} = imread(sprintf(FNAMEFMT2_2,i)); % DAPI images
    end

%     for i = 100:NbImages                          % For images 100-999
%       Im{i} = imread(sprintf(FNAMEFMT3_1,i));  % Gfp images
%       Im2{i} = imread(sprintf(FNAMEFMT3_2,i)); % DAPI images
%     end

end

% Adjust contrast to help pick out cells in binary image
Imadjust = imadjust(Im{2});              %For images in fist color (gfp)
%Imadjust = imadjust(Im{2},[0 1],[]);   %Additional sensitivity
definition
Imadjust2 = imadjust(Im2{2});            %For images in the second color
(DAPI)

[level EM1] = graythresh(Imadjust);      %Get the threshold from
graythresh function
bw = im2bw(Imadjust,level);              %Convert to binary
bwclearborder = imclearborder(bw);       %Clear any objects on the
border of image e.g. half cells
bwclear = bwareafilt(bwclearborder,pixelarearange);            %Remove
spots not in pixelrange

[level2 EM2] = graythresh(Imadjust2);     %Get the threshold from
graythresh function of the second set of images
bw2 = im2bw(Imadjust2,level2);             %Convert to binary
bwclearborder2 = imclearborder(bw2);      %Clear any objects on the
border of image
bwclear2 = bwareafilt(bwclearborder2,pixelarearange);          % Do the
same for second set of images


figure %Comment out after printing
imshow(bwclear) % Plot the whole cells found
title('Binary Img w Sizes Filtered (whole cell gfp)')

figure %Comment out after printing
imshow(bwclear2) % Plot the nuclei found using DAPI
title('Binary Img w Sizes Filtered (nuclei)')

% You must convert a binary image into a label matrix before calling
% regionprops to get the stats of interest. Here, we use the bwlabel
function to do this.
```

```
% L returns label matrix with labels for objects (bright areas)
% num returns the number of connected components
[L, num] = bwlabel(bwclear);    %Original code, gfp full cells (as
opposed to gfp for just nuclei)
[L2, num2] = bwlabel(bwclear2); %Areas defined as nuclei

for ii=1:NbImages
    % Boundingbox: smallest square that contains whole area of interest
    % Note: now the two defined areas are applied to the same image
set!
    stats =
regionprops(L,Im{ii},'MeanIntensity','MinIntensity','MaxIntensity','Cen
troid','BoundingBox','Extrema','Area');   %intensities in full cell for
gfp i.e whole cell
    stats2 =
regionprops(L2,Im2{ii},'MeanIntensity','MinIntensity','MaxIntensity','C
entroid','BoundingBox','Extrema','Area'); % intensities in the nuclear
area for gfp i.e. nucleus
end

%%% End here to print initial masks and manually choose the cells to
%%% analyze before continuing

%%% STEP 5 MANUALLY UPDATE THE MASKS TO REMOVE UNWANTED CELLS AND
CREATE NEW MASKS

% Step 5a Open first mask (bwclear) and choose all cells to analyze
hh = figure
imshow(imcomplement(bwclear)) %Invert colors to make it easier to see
crosshairs
[x,y] = ginput(11); %Specify how many areas will be chosen from the
bwclear mask
for k = 1:length(x)
 for i = 1:length(stats)
    distance(i) = sqrt((x(k)-stats(i).Centroid(1))^2+(y(k)-
stats(i).Centroid(2))^2); %find cell closest to clicked pixel
 end
  [val ind] = (min(distance))
  indmat1(k) = ind;
end
close(hh) %Close the image after the specified number of bad points are
chosen

%Now do the same for the nuclei (bwclear2), choosing desired areas
manually
hh = figure
imshow(imcomplement(bwclear2))
[x,y] = ginput(11); %Specify how many points will be chosen from the
bwclear2 mask
for k = 1:length(x)
for i = 1:length(stats2)
    distance(i) = sqrt((x(k)-stats2(i).Centroid(1))^2+(y(k)-
stats2(i).Centroid(2))^2);
    end
    [val ind] = (min(distance))
    indmat2(k) = ind;
```

```
end
close(hh)

%%% Step 5b Update masks to exclude the chosen areas
numcells1 = length(stats); %For whole cell
numcells2 = length(stats2); %For nuclei

goodimg1 = ismember(bwlabel(L),indmat1); %Define updated mask as only
goodimg1 for whole cells
goodimg2 = ismember(bwlabel(L2),indmat2); %Define updated mask as only
goodimg2 for nuclei

% Optional: make figures to show new masks and overlays
% figure
% imshow(goodimg1)
% figure
% imshow(goodimg2)

figure
imshow(imfuse(goodimg1,goodimg2)) % QC: Image of overlaid masks of good
cells in gfp with good nuclei
title('Overlaid masks')

% Optional: Apply new masks to first images as test
% testimg1 = Im{1,1};    %Grabbing the original gfp image (i.e. non-
binary)
% testimg1(goodimg1==0)=0; %Applying the mask of only good cells
(goodimg1)
 [L3, num3] = bwlabel(goodimg1); %Get updated label from mask 2
% stats3 =
regionprops(L3,testimg1,'MeanIntensity','MinIntensity','MaxIntensity','
Centroid','BoundingBox','Extrema','Area');   %do a regionprops on
original gfp image using mask 2

% testimg2 = Im{1,1};    %Grabbing the original first gfp image (i.e.
non-binary)
% testimg2(goodimg2==0)=0; %Applying the mask of only good nuclei
(goodimg2)
 [L4 num4] = bwlabel(goodimg2); %Get updated label from mask with
chosen objects
% stats4 =
regionprops(L4,testimg1,'MeanIntensity','MinIntensity','MaxIntensity','
Centroid','BoundingBox','Extrema','Area');   %do a regionprops on
original gfp image using mask 2

L5 = (L3-L4); %Create a mask of cells' cytosol by subtracting nuclear
area from whole cell

%%% STEP6 APPLY UPDATED MASKS TO IMAGE SERIES. SUBTRACT BACKGROUND

% for ii=1:NbImages    % Optional: If interested in whole cell area
%     stats5 =
regionprops(L3,Im{ii},'MeanIntensity','MinIntensity','MaxIntensity','Ce
ntroid','BoundingBox','Extrema','Area'); %intensities in full cell for
gfp i.e whole cell
```

```
%     MeanIntensityCell(:,ii) = extractfield(stats5,'MeanIntensity')';
%This takes the value from the structure and makes a vector
%     BoundingBox(:,ii) = extractfield(stats5,'BoundingBox')'; %
Boundingbox: smallest square that contains whole area of interest
%     MinIntensityCell(:,ii) = extractfield(stats5,'MinIntensity');
%     MaxIntensityCell(:,ii) = extractfield(stats5,'MaxIntensity');
%     CentroidCell(:,ii) = extractfield(stats5,'Centroid');
%     ExtremaCell(:,ii) = extractfield(stats5,'Extrema');
%     AreaCell(:,ii) = extractfield(stats5,'Area');
%
%     % Subtract background for each cell
%     for i = 0:num3-1
%         j = i*4+1;                          %Bounding box has 4
values for every object: 1. x coordinate 2. y coordinate 3. x width 4.
y width
%         ycoord(i+1) = round(BoundingBox(j+1));   %Find the y value
for all bounding boxes
%         xcoord(i+1) = round(BoundingBox(j));     %Find the x value
for all bounding boxes
%         Back(i+1,ii) = mean(mean(Im{ii}(ycoord(i+1)-
2:ycoord(i+1),xcoord(i+1)-2:xcoord(i+1)))); %Take the average of pixel
intensities from the top left, a 3x3 pixel square
%         BackSubMeanIntensityCell(i+1,ii) = MeanIntensityCell(i+1,ii)
- Back(i+1,ii); % subtract background for each cell
%      end
% end

 for ii=1:NbImages  %For cytosol
    stats5 =
regionprops(L5,Im{ii},'MeanIntensity','MinIntensity','MaxIntensity','Ce
ntroid','BoundingBox','Extrema','Area'); %intensities in cytosolic gfp
    MeanIntensityCyt(:,ii) = extractfield(stats5,'MeanIntensity')';
%This takes the value from the structure and makes a vector
    BoundingBox(:,ii) = extractfield(stats5,'BoundingBox')'; %
Boundingbox: smallest square that contains whole area of interest
    MinIntensityCyt(:,ii) = extractfield(stats5,'MinIntensity');
    MaxIntensityCyt(:,ii) = extractfield(stats5,'MaxIntensity');
    CentroidCyt(:,ii) = extractfield(stats5,'Centroid');
    ExtremaCyt(:,ii) = extractfield(stats5,'Extrema');
    AreaCyt(:,ii) = extractfield(stats5,'Area');

    % Subtract local background for each cell
    for i = 0:num3-1
        j = i*4+1;                          %Bounding box has 4 values
for every object: 1. x coordinate 2. y coordinate 3. x width 4. y width
        ycoord(i+1) = round(BoundingBox(j+1));   %Find the y value for
all bounding boxes
        xcoord(i+1) = round(BoundingBox(j));     %Find the x value for
all bounding boxes
        Back(i+1,ii) = mean(mean(Im{ii}(ycoord(i+1)-
2:ycoord(i+1),xcoord(i+1)-2:xcoord(i+1)))); %Take the average of pixel
intensities from the top left, a 3x3 pixel square
        BackSubMeanIntensityCyt(i+1,ii) = MeanIntensityCyt(i+1,ii) -
Back(i+1,ii); % subtract background for each cell
     end
end
```

```
 for ii=1:NbImages %For nuclei
    stats5 = regionprops(L5,Im{ii},'BoundingBox'); %intensities in full
cell for gfp i.e whole cell
    stats6 =
regionprops(L4,Im{ii},'MeanIntensity','MinIntensity','MaxIntensity','Ce
ntroid','BoundingBox','Extrema','Area'); %intensities in the nuclear
area for gfp i.e. nucleus
    MeanIntensityNucleus(:,ii) = extractfield(stats6,'MeanIntensity')';
%This takes the values from the structure and makes a vector
    BoundingBox(:,ii) = extractfield(stats5,'BoundingBox')'; %Note:
using box for Cell here so background area will be same
    MinIntensityNucleus(:,ii) = extractfield(stats6,'MinIntensity');
    MaxIntensityNucleus(:,ii) = extractfield(stats6,'MaxIntensity');
    CentroidNucleus(:,ii) = extractfield(stats6,'Centroid');
    ExtremaNucleus(:,ii) = extractfield(stats6,'Extrema');
    AreaNucleus(:,ii) = extractfield(stats6,'Area');

    % Subtract background using SAME background coords as for
corresponding cell!
    for i = 0:num4-1
        j = i*4+1;                          %Bounding box has 4 values
for every object: 1. x coordinate 2. y coordinate 3. x width 4. y width
        Back(i+1,ii) = mean(mean(Im{ii}(ycoord(i+1)-
2:ycoord(i+1),xcoord(i+1)-2:xcoord(i+1)))); %Take the average of pixel
intensities from the top left, a 3x3 pixel square
        BackSubMeanIntensityNucleus(i+1,ii) =
MeanIntensityNucleus(i+1,ii) - Back(i+1,ii); % use same areas of
backgound for nuclei as for whole cell background subtraction
    end

 end

 % QC: Print the matrix sizes before continuing to ensure they are the
same
 %size(BackSubMeanIntensityCell)
 size(BackSubMeanIntensityCyt)
 size(BackSubMeanIntensityNucleus)

 %%% STEP 7 GO TO LK_Step2 file to normalize data and create figures

save(Filename)  % Save results with the name specified at the beginning
of this file
```

A.9. Code for Image Analysis Step 2 Data Processing

```
%Linda's code for data processing of data from image analysis
clear all; close all; clc

%Insert file name to load from previous step of image analysis, but do
not include '.mat'
File = '20170222_100pM_1m1m_Jurkat_gfp_XY1_20170223T174840',;
```

```
ExpLength = 62; % Enter length of the experiment in minutes
Fs = 1;     % Enter sampling rate in minutes
FileToLoad = [File,'.mat']; %Identify the variables you want to bring
in. File will have saved all variables from Step1

Matobj = matfile(FileToLoad);   %Load in specific file

BackSubMeanIntensityCyt = Matobj.BackSubMeanIntensityCyt;     %
Intensity for cytosol
%BackSubMeanIntensityCell = Matobj.BackSubMeanIntensityCell;     %
Intensity for whole cells
BackSubMeanIntensityNucleus = Matobj.BackSubMeanIntensityNucleus;   %
Intensity for nuclei
BackSubMeanIntensityRatio2 =
Matobj.BackSubMeanIntensityNucleus./Matobj.BackSubMeanIntensityCyt;%Get
ratio of gfp intensity for nucleus vs cytosol
%Labeled with period listed followed by date BackSubMeanIntensityCorr
was
%created and finally, the date in which the file was created
OutputFilename = [File,'_',datestr(now,30),'.txt'];
SaveFilename = [File,'_',datestr(now,30),'.mat'];

Time = 0:Fs:ExpLength;  %Create your time vector

cellnumber = size(BackSubMeanIntensityCyt,1);
length = size(BackSubMeanIntensityCyt,2);
Cminmaxeachnorm2 = zeros(cellnumber ,length);

for i = 1:cellnumber
   M = BackSubMeanIntensityRatio2(i,1:length); %take the
nuclear/cytosol ratios for all the time point
   Mnorm = minmaxnorm(M);
   Cminmaxeachnorm2(i,:) = (Mnorm); % Normalized
end

NbImages = size(BackSubMeanIntensityRatio2,2);
for i = 1:NbImages

NBackSubMeanIntensityRatio2(:,i)=BackSubMeanIntensityRatio2(:,i)./BackS
ubMeanIntensityRatio2(:,2);
end

% Make figures. Plotting nuclear and cytosolic gfp over time allows for
QC

figure
plot(BackSubMeanIntensityCyt')
xlabel('Frame')
ylabel('gfp intensity')
title('Cyt gfp 20170222 100pM IL2 1min pulse 1min pause XY1')

figure
plot (BackSubMeanIntensityNucleus')
xlabel('Frame')
ylabel('gfp intensity')
```

```
title('Nuclear GFP Intensity 20170222 100pM IL2 1min pulse 1min pause
XY1')

figure
plot(BackSubMeanIntensityRatio2')
xlabel('Frame')
ylabel('Nuclear/Cyt gfp Ratio')
title('Nuclear/Cyt GFP Intensity Ratio 20170222 100pM IL2 1min pulse
1min pause XY1')

figure
plot(NBackSubMeanIntensityRatio2')
xlabel('Frame (min)')
ylabel('Nuclear/Cyt gfp Ratio')
title('Normalized Nuclear/Cyt GFP Intensity Ratio 20170222 100pM IL2
1min pulse 1min pause XY1')

save(SaveFilename)

%%%function to accompany this file. Make a file with it in the same
folder.
% function Mnorm = minmaxnorm(M)
% a = min(min(M));
% b = max(max(M));
% Mnorm = (M-a)/(b-a);
```

REFERENCES

1. Germain, R.N., *Maintaining system homeostasis: the third law of Newtonian immunology.* Nat Immunol, 2012. **13**(10): p. 902-6.
2. Niepel, M., S.L. Spencer, and P.K. Sorger, *Non-genetic cell-to-cell variability and the consequences for pharmacology.* Curr Opin Chem Biol, 2009. **13**(5-6): p. 556-61.
3. Spencer, S.L., et al., *Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis.* Nature, 2009. **459**(7245): p. 428-32.
4. Spencer, S.L. and P.K. Sorger, *Measuring and modeling apoptosis in single cells.* Cell, 2011. **144**(6): p. 926-39.
5. Tay, S., et al., *Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing.* Nature, 2010. **466**(7303): p. 267-71.
6. Feinerman, O., et al., *Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response.* Mol Syst Biol, 2010. **6**: p. 437.
7. Malek, T.R. and I. Castro, *Interleukin-2 receptor signaling: at the interface between tolerance and immunity.* Immunity, 2010. **33**(2): p. 153-65.
8. Alugupalli, K.R. and R.M. Gerstein, *Divide and conquer: division of labor by B-1 B cells.* Immunity, 2005. **23**(1): p. 1-2.
9. Littman, D.R. and A.Y. Rudensky, *Th17 and regulatory T cells in mediating and restraining inflammation.* Cell, 2010. **140**(6): p. 845-58.
10. Wing, K. and S. Sakaguchi, *Regulatory T cells exert checks and balances on self tolerance and autoimmunity.* Nat Immunol, 2010. **11**(1): p. 7-13.
11. Morgan, D.A., F.W. Ruscetti, and R. Gallo, *Selective in vitro growth of T lymphocytes from normal human bone marrows.* Science, 1976. **193**(4257): p. 1007-8.
12. Gillis, S., et al., *T cell growth factor: parameters of production and a quantitative microassay for activity.* J Immunol, 1978. **120**(6): p. 2027-32.
13. Baker, P.E., et al., *The effect of T cell growth factor on the generation of cytolytic T cells.* J Immunol, 1978. **121**(6): p. 2168-73.
14. Smith, K.A., et al., *T cell growth factor-mediated T cell proliferation.* Ann N Y Acad Sci, 1979. **332**: p. 423-32.
15. Larsson, E.L., N.N. Iscove, and A. Coutinho, *Two distinct factors are required for induction of T cell growth.* Nature, 1980. **283**(5748): p. 664-6.
16. Smith, K.A., et al., *The functional relationship of the interleukins.* J Exp Med, 1980. **151**(6): p. 1551-6.
17. Barthlott, T., et al., *CD25+ CD4+ T cells compete with naive CD4+ T cells for IL-2 and exploit it for the induction of IL-10 production.* Int Immunol, 2005. **17**(3): p. 279-88.
18. Pandiyan, P., et al., *CD4+CD25+Foxp3+ regulatory T cells induce cytokine deprivation-mediated apoptosis of effector CD4+ T cells.* Nat Immunol, 2007. **8**(12): p. 1353-62.

19. Busse, D., et al., *Competing feedback loops shape IL-2 signaling between helper and regulatory T lymphocytes in cellular microenvironments.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 3058-63.
20. Sakaguchi, S., et al., *Regulatory T cells: how do they suppress immune responses?* Int Immunol, 2009. **21**(10): p. 1105-11.
21. Cacalano, N.A. and J.A. Johnston, *Interleukin-2 signaling and inherited immunodeficiency.* Am J Hum Genet, 1999. **65**(2): p. 287-93.
22. Liao, W., J.X. Lin, and W.J. Leonard, *IL-2 family cytokines: new insights into the complex roles of IL-2 as a broad regulator of T helper cell differentiation.* Curr Opin Immunol, 2011. **23**(5): p. 598-604.
23. Cantrell, D.A. and K.A. Smith, *The interleukin-2 T cell system: a new cell growth model.* Science, 1984. **224**(4655): p. 1312-6.
24. Liao, W., et al., *Priming for T helper type 2 differentiation by interleukin 2-mediated induction of interleukin 4 receptor alpha-chain expression.* Nat Immunol, 2008. **9**(11): p. 1288-96.
25. Liao, W., et al., *Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages.* Nat Immunol, 2011. **12**(6): p. 551-9.
26. Xue, H.H., et al., *IL-2 negatively regulates IL-7 receptor alpha chain expression in activated T lymphocytes.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13759-64.
27. Smith, K.A., *The interleukin 2 receptor.* Adv Immunol, 1988. **42**: p. 165-79.
28. Vincenti, F., et al., *Interleukin-2-receptor blockade with daclizumab to prevent acute rejection in renal transplantation. Daclizumab Triple Therapy Study Group.* N Engl J Med, 1998. **338**(3): p. 161-5.
29. Beniaminovitz, A., et al., *Prevention of rejection in cardiac transplantation by blockade of the interleukin-2 receptor with a monoclonal antibody.* N Engl J Med, 2000. **342**(9): p. 613-9.
30. Bielekova, B., et al., *Humanized anti-CD25 (daclizumab) inhibits disease activity in multiple sclerosis patients failing to respond to interferon beta.* Proc Natl Acad Sci U S A, 2004. **101**(23): p. 8705-8.
31. Dunant, Y., et al., *Sustained oscillations of acetylcholine during nerve stimulation.* Nature, 1974. **252**(5483): p. 485-6.
32. Meier, J.J., J.D. Veldhuis, and P.C. Butler, *Pulsatile insulin secretion dictates systemic insulin delivery by regulating hepatic insulin extraction in humans.* Diabetes, 2005. **54**(6): p. 1649-56.
33. Ashall, L., et al., *Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription.* Science, 2009. **324**(5924): p. 242-6.
34. Sumit, M., et al., *Band-pass processing in a GPCR signaling pathway selects for NFAT transcription factor activation.* Integr Biol (Camb), 2015. **7**(11): p. 1378-86.
35. Warmflash, A., et al., *Dynamics of TGF-beta signaling reveal adaptive and pulsatile behaviors reflected in the nuclear localization of transcription factor Smad4.* Proc Natl Acad Sci U S A, 2012. **109**(28): p. E1947-56.
36. Ryu, H., et al., *Frequency modulation of ERK activation dynamics rewires cell fate.* Mol Syst Biol, 2015. **11**(11): p. 838.

37. Tomida, T., et al., *The temporal pattern of stimulation determines the extent and duration of MAPK activation in a Caenorhabditis elegans sensory neuron.* Sci Signal, 2012. **5**(246): p. ra76.
38. Kniss-James, A.S., et al., *Single-cell resolution of intracellular T cell Ca2+ dynamics in response to frequency-based H2O2 stimulation.* Integr Biol (Camb), 2017.
39. Ventura, A.C., et al., *Utilization of extracellular information before ligand-receptor binding reaches equilibrium expands and shifts the input dynamic range.* Proc Natl Acad Sci U S A, 2014. **111**(37): p. E3860-9.
40. Wang, H.M. and K.A. Smith, *The interleukin 2 receptor. Functional consequences of its bimolecular structure.* J Exp Med, 1987. **166**(4): p. 1055-69.
41. Bendall, S.C., et al., *Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum.* Science, 2011. **332**(6030): p. 687-96.
42. Shi, Q., et al., *Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells.* Proc Natl Acad Sci U S A, 2012. **109**(2): p. 419-24.
43. Irish, J.M., et al., *Single cell profiling of potentiated phospho-protein networks in cancer cells.* Cell, 2004. **118**(2): p. 217-28.
44. Lu, Y., et al., *Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands.* Proc Natl Acad Sci U S A, 2015. **112**(7): p. E607-15.
45. Gawad, C., W. Koh, and S.R. Quake, *Single-cell genome sequencing: current state of the science.* Nat Rev Genet, 2016. **17**(3): p. 175-88.
46. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. **472**(7341): p. 90-4.
47. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell.* Nat Methods, 2009. **6**(5): p. 377-82.
48. Fan, H.C., G.K. Fu, and S.P. Fodor, *Expression profiling. Combinatorial labeling of single cells for gene expression cytometry.* Science, 2015. **347**(6222): p. 1258367.
49. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-14.
50. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-201.
51. Cheng, Z., et al., *Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation.* Sci Signal, 2015. **8**(385): p. ra69.
52. Tkach, K.E., et al., *T cells translate individual, quantal activation into collective, analog cytokine responses via time-integrated feedbacks.* Elife, 2014. **3**: p. e01944.
53. Thurley, K., et al., *Three-Dimensional Gradients of Cytokine Signaling between T Cells.* PLoS Comput Biol, 2015. **11**(4): p. e1004206.
54. Junkin, M., et al., *High-Content Quantification of Single-Cell Immune Dynamics.* Cell Rep, 2016. **15**(2): p. 411-22.

55. Duncombe, T.A., A.M. Tentori, and A.E. Herr, *Microfluidics: reframing biological enquiry.* Nat Rev Mol Cell Biol, 2015. **16**(9): p. 554-67.
56. Kippner, L.E., et al., *Single cell transcriptional analysis reveals novel innate immune cell types.* PeerJ, 2014. **2**: p. e452.
57. Sachs, K., et al., *Causal protein-signaling networks derived from multiparameter single-cell data.* Science, 2005. **308**(5721): p. 523-9.
58. Kalisky, T. and S.R. Quake, *Single-cell genomics.* Nat Methods, 2011. **8**(4): p. 311-4.
59. Kurimoto, K., et al., *Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis.* Nat Protoc, 2007. **2**(3): p. 739-52.
60. Flatz, L., et al., *Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines.* Proc Natl Acad Sci U S A, 2011. **108**(14): p. 5724-9.
61. Shi, X., et al., *Real-time PCR of single bacterial cells on an array of adhering droplets.* Lab Chip, 2011. **11**(13): p. 2276-81.
62. White, A.K., et al., *High-throughput microfluidic single-cell RT-qPCR.* Proc Natl Acad Sci U S A, 2011. **108**(34): p. 13999-4004.
63. Janes, K.A., et al., *Identifying single-cell molecular programs by stochastic profiling.* Nat Methods, 2010. **7**(4): p. 311-7.
64. Rajan, S., et al., *The living microarray: a high-throughput platform for measuring transcription dynamics in single cells.* BMC Genomics, 2011. **12**: p. 115.
65. Zhang, Y., et al., *Nanolitre droplet array for real time reverse transcription polymerase chain reaction.* Lab Chip, 2011. **11**(8): p. 1545-9.
66. Morris, J., J.M. Singh, and J.H. Eberwine, *Transcriptome analysis of single cells.* J Vis Exp, 2011(50).
67. Bengtsson, M., et al., *Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels.* Genome Res, 2005. **15**(10): p. 1388-92.
68. McDavid, A., et al., *Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments.* Bioinformatics, 2013. **29**(4): p. 461-7.
69. Mucida, D., et al., *Transcriptional reprogramming of mature CD4(+) helper T cells generates distinct MHC class II-restricted cytotoxic T lymphocytes.* Nat Immunol, 2013. **14**(3): p. 281-9.
70. Hoshida, Y., et al., *Gene expression in fixed tissues and outcome in hepatocellular carcinoma.* N Engl J Med, 2008. **359**(19): p. 1995-2004.
71. Dalerba, P., et al., *Single-cell dissection of transcriptional heterogeneity in human colon tumors.* Nat Biotechnol, 2011. **29**(12): p. 1120-7.
72. Shalek, A.K., et al., *Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells.* Nature, 2013. **498**(7453): p. 236-40.
73. Suzuki, T., P.J. Higgins, and D.R. Crawford, *Control selection for RNA quantitation.* Biotechniques, 2000. **29**(2): p. 332-7.
74. Liss, B., et al., *Tuning pacemaker frequency of individual dopaminergic neurons by Kv4.3L and KChip3.1 transcription.* EMBO J, 2001. **20**(20): p. 5715-24.
75. Chen, J., et al., *ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W305-11.

76. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
77. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
78. Nakao, M., et al., *Genome-scale Gene Expression Analysis and Pathway Reconstruction in KEGG.* Genome Inform Ser Workshop Genome Inform, 1999. **10**: p. 94-103.
79. Ward, J.H., Jr, *Hierarchical Grouping to Optimize an Objective Function.* Journal of the American Statistical Association, 1963. **58**(301): p. 236-244.
80. MacQueen, J.C. *Some methods for classification and analysis of multivariate observations*. in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967. Statistical Laboratory of the University of California, Berkeley.
81. Ferreira, L. and D. Hitchcock, *A Comparison of Hierarchical Methods for Clustering Functional Data.* Communications in Statistics: Simulation and Computation, 2009. **38**(9): p. 1925-1949.
82. Ma, N. and Z.G. Zhang, *Evaluation of clustering algorithms for gene expression data using gene ontology annotations.* Chin Med J (Engl), 2012. **125**(17): p. 3048-52.
83. Stegle, O., S.A. Teichmann, and J.C. Marioni, *Computational and analytical challenges in single-cell transcriptomics.* Nat Rev Genet, 2015. **16**(3): p. 133-45.
84. Swain, S.L., *Lymphokines and the immune response: the central role of interleukin-2.* Curr Opin Immunol, 1991. **3**(3): p. 304-10.
85. O'Shea, J.J., A. Ma, and P. Lipsky, *Cytokines and autoimmunity.* Nat Rev Immunol, 2002. **2**(1): p. 37-45.
86. Takeshita, T., et al., *Cloning of the gamma chain of the human IL-2 receptor.* Science, 1992. **257**(5068): p. 379-82.
87. Sugamura, K., et al., *The IL-2/IL-2 receptor system: involvement of a novel receptor subunit, gamma chain, in growth signal transduction.* Tohoku J Exp Med, 1992. **168**(2): p. 231-7.
88. Minami, Y., et al., *The IL-2 receptor complex: its structure, function, and target genes.* Annu Rev Immunol, 1993. **11**: p. 245-68.
89. Waldmann, T., Y. Tagaya, and R. Bamford, *Interleukin-2, interleukin-15, and their receptors.* Int Rev Immunol, 1998. **16**(3-4): p. 205-26.
90. Fehniger, T.A. and M.A. Caligiuri, *Interleukin 15: biology and relevance to human disease.* Blood, 2001. **97**(1): p. 14-32.
91. Taniguchi, T. and Y. Minami, *The IL-2/IL-2 receptor system: a current overview.* Cell, 1993. **73**(1): p. 5-8.
92. Stauber, D.J., et al., *Crystal structure of the IL-2 signaling complex: paradigm for a heterotrimeric cytokine receptor.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2788-93.
93. Sugamura, K., et al., *The common gamma-chain for multiple cytokine receptors.* Adv Immunol, 1995. **59**: p. 225-77.
94. Teshigawara, K., et al., *Interleukin 2 high-affinity receptor expression requires two distinct binding proteins.* J Exp Med, 1987. **165**(1): p. 223-38.

95. Arima, N., et al., *Pseudo-high affinity interleukin 2 (IL-2) receptor lacks the third component that is essential for functional IL-2 binding and signaling.* J Exp Med, 1992. **176**(5): p. 1265-72.
96. Hatakeyama, M., et al., *A restricted cytoplasmic region of IL-2 receptor beta chain is essential for growth signal transduction but not for ligand binding and internalization.* Cell, 1989. **59**(5): p. 837-45.
97. Merida, I. and G.N. Gaulton, *Protein tyrosine phosphorylation associated with activation of the interleukin 2 receptor.* J Biol Chem, 1990. **265**(10): p. 5690-4.
98. Shin, H.Y. and N.C. Reich, *Dynamic trafficking of STAT5 depends on an unconventional nuclear localization signal.* J Cell Sci, 2013. **126**(Pt 15): p. 3333-43.
99. Kohler, M., et al., *Evidence for distinct substrate specificities of importin alpha family members in nuclear protein import.* Mol Cell Biol, 1999. **19**(11): p. 7782-91.
100. Kohler, M., et al., *Cloning of two novel human importin-alpha subunits and analysis of the expression pattern of the importin-alpha protein family.* FEBS Lett, 1997. **417**(1): p. 104-8.
101. Kim, H.P., J. Imbert, and W.J. Leonard, *Both integrated and differential regulation of components of the IL-2/IL-2 receptor system.* Cytokine Growth Factor Rev, 2006. **17**(5): p. 349-66.
102. Fallon, E.M. and D.A. Lauffenburger, *Computational model for effects of ligand/receptor binding properties on interleukin-2 trafficking dynamics and T cell proliferation response.* Biotechnol Prog, 2000. **16**(5): p. 905-16.
103. de Jong, J.L., et al., *Interaction of IL-15 with the shared IL-2 receptor beta and gamma c subunits. The IL-15/beta/gamma c receptor-ligand complex is less stable than the IL-2/beta/gamma c receptor-ligand complex.* J Immunol, 1996. **156**(4): p. 1339-48.
104. Hemar, A., et al., *Endocytosis of interleukin 2 receptors in human T lymphocytes: distinct intracellular localization and fate of the receptor alpha, beta, and gamma chains.* J Cell Biol, 1995. **129**(1): p. 55-64.
105. Levin, A.M., et al., *Exploiting a natural conformational switch to engineer an interleukin-2 'superkine'.* Nature, 2012. **484**(7395): p. 529-33.
106. Duprez, V., V. Cornet, and A. Dautry-Varsat, *Down-regulation of high affinity interleukin 2 receptors in a human tumor T cell line. Interleukin 2 increases the rate of surface receptor decay.* J Biol Chem, 1988. **263**(26): p. 12860-5.
107. Duprez, V. and A. Dautry-Varsat, *Receptor-mediated endocytosis of interleukin 2 in a human tumor T cell line. Degradation of interleukin 2 and evidence for the absence of recycling of interleukin receptors.* J Biol Chem, 1986. **261**(33): p. 15450-4.
108. Ghosh, R.N., D.L. Gelman, and F.R. Maxfield, *Quantification of low density lipoprotein and transferrin endocytic sorting HEp2 cells using confocal microscopy.* J Cell Sci, 1994. **107 (Pt 8)**: p. 2177-89.
109. Hemar, A., et al., *Endocytosis of the beta chain of interleukin-2 receptor requires neither interleukin-2 nor the gamma chain.* Eur J Immunol, 1994. **24**(9): p. 1951-5.

110. Smith, K.A. and D.A. Cantrell, *Interleukin 2 regulates its own receptors.* Proc Natl Acad Sci U S A, 1985. **82**(3): p. 864-8.
111. A., A., *Essai d'une maniere de determiner les masses relatives des molecules elementaires des corps, et les proportions selon lesquelles elles entrent dans ces combinaisons.* Journal de Physique, 1811. **73**: p. 58-76.
112. French, A.R., et al., *Intracellular trafficking of epidermal growth factor family ligands is directly influenced by the pH sensitivity of the receptor/ligand interaction.* J Biol Chem, 1995. **270**(9): p. 4334-40.
113. Chingozha, L., et al., *A Generalizable, Tunable Microfluidic Platform for Delivering Fast Temporally Varying Chemical Signals to Probe Single-Cell Response Dynamics.* Anal Chem, 2014.
114. Smith, K.A., *Interleukin-2: inception, impact, and implications.* Science, 1988. **240**(4856): p. 1169-76.
115. Rao, B.M., et al., *Interleukin 2 (IL-2) variants engineered for increased IL-2 receptor alpha-subunit affinity exhibit increased potency arising from a cell surface ligand reservoir effect.* Mol Pharmacol, 2004. **66**(4): p. 864-9.
116. Smith, K.A., *Lowest dose interleukin-2 immunotherapy.* Blood, 1993. **81**(6): p. 1414-23.
117. Jacobson, E.L., F. Pilaro, and K.A. Smith, *Rational interleukin 2 therapy for HIV positive individuals: daily low doses enhance immune function without toxicity.* Proc Natl Acad Sci U S A, 1996. **93**(19): p. 10405-10.
118. Yao, J., A. Pilko, and R. Wollman, *Distinct cellular states determine calcium signaling response.* Mol Syst Biol, 2016. **12**(12): p. 894.
119. Depper, J.M., et al., *Interleukin 2 (IL-2) augments transcription of the IL-2 receptor gene.* Proc Natl Acad Sci U S A, 1985. **82**(12): p. 4230-4.
120. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.
121. Saliba, A.E., et al., *Single-cell RNA-seq: advances and future challenges.* Nucleic Acids Res, 2014. **42**(14): p. 8845-60.
122. Trapnell, C., *Defining cell types and states with single-cell genomics.* Genome Res, 2015. **25**(10): p. 1491-8.
123. Spitzer, M.H., et al., *IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system.* Science, 2015. **349**(6244): p. 1259425.