UNIVERSITY OF CALIFORNIA, SAN DIEGO

Single-cell NF-kappaB signaling in macrophages

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Brooks Edward Taylor Jr.

Committee in Charge:

Professor Jeff Hasty, Chair
Professor Alexander Hoffmann, Co-Chair
Professor Christopher Glass
Professor Geert Schmidt-Shoenbein
Professor Kun Zhang

2015

The dissertation of Brooks Edward Taylor Jr. is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2015

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A.U. . . . . arbitrary units

BMDM . . . . . bone-marrow derived macrophage

CHO . . . . . Chinese hamster ovary

Ca^{2+} . . . . . calcium

DIC . . . . . differerential interference contrast microscopy

DNA . . . . . deoxyribonucleic acid

E. coli . . . . . *Escherichia coli*

EC_{50} . . . . . half maximal effective concentration

Eq. . . . . equation

EGFP . . . . . enhanced green fluorescent protein

EYFP . . . . . enhanced yellow fluorescent protein

fig. . . . . figure

hr. . . . . hours

IκB . . . . . inhibitor of NFκB

IKK . . . . . IκB kinase

kb . . . . . kilobase

LPS . . . . . lipopolysaccharide

mRNA . . . . . messenger ribonucleic acid

MyD88 . . . . . myeloid differentiation primary response 88

NF-κB . . . . . nuclear factor κB

NLS . . . . . nuclear localization signal

no. . . . . number

ODE . . . . . ordinary differential equation

PBS . . . . . phosphate-buffered saline

TLR . . . . . Toll-like receptor

TNF . . . . . tissue necrosis factor

TRIF . . . . . TIR-domain-containing adapter-inducing interferon-β

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am eternally thankful for the compassion shown me by my mentors, the love and patience granted me by my friends and family, and the inspiration and beauty evident in nature and all her designs.

Chapter 3 contains material originally published as "Accurate information transmission through dynamic biochemical signaling networks". Selimkhanov, J.*, Taylor, B.*, Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. ***Science*** (2014). Copyright permission to republish here was granted by AAAS.

Chapter 4 contains material originally published as "Distinct Dynamical Characteristics Conferred by the MyD88 and TRIF Pathways in TLR4 Signaling". Cheng, Z.*, Taylor, B.*, Ourthiague, D., and Hoffmann, A. ***Science Signaling*** (2015 - in review). Copyright permission to republish here was granted by AAAS.

VITA

2009	Bachelor of Science in Biomedical Engineering University of Virginia
2012	Candidate of Philosophy in Bioengineering University of California, San Diego
2015	Doctor of Philosophy in Bioengineering University of California, San Diego

PUBLICATIONS

Peer Reviewed Journal Articles

Accurate information transmission through dynamic biochemical signaling networks. Selimkhanov, J.*, Taylor, B.*, Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. ***Science*** (2014)

Distinct Dynamical Characteristics Conferred by the MyD88 and TRIF Pathways in TLR4 Signaling. Cheng, Z.*, Taylor, B.*, Rios, D., and Hoffmann, A. ***Science Signaling*** (2015 - in review)

(* indicates equal contribution)

FIELDS OF STUDY

Major Field: Bioengineering (Systems Biology)

Studies in Biological Dynamics
Professor Alexander Hoffmann
Professor Jeff Hasty

ABSTRACT OF THE DISSERTATION

Single-cell NF-kappaB signaling in macrophages

by

Brooks Edward Taylor Jr.
Doctor of Philosophy in Bioengineering

University of California, San Diego, 2015

Professor Jeff Hasty, Chair
Professor Alexander Hoffmann, Co-Chair

The transcription factor NF-κB is heavily involved in innate immunity, and shows complex dynamic patterns of nuclear translocation in response to many stimuli. These dynamics are hypothesized to contain information about a particular stimulus. Single cell studies of NF-κB have shown high variability across cells, raising questions about the reliability of biochemical information encoding. Additionally, the majority of work to date has only measured NF-κB activity in non-immune cell types, leaving the question of how dynamics function in native contexts, like macrophage activation by Toll-like receptor signaling. To answer these questions, I developed an automated image processing algorithm to accurately track individual macrophages over 12-24 hours. I use this algorithm to measure multiple dimensions of macrophage activation, including NF-κB activity, in fluorescently-tagged cells.

Thousands of NF-κB responses to lipopolysaccharide in macrophages were generated with this automated process, then applied in the development of an algorithm

that estimates the information transduction capacities of biochemical networks. I show that NF-κB dynamics, as well as response dynamics in the ERK and Ca^{2+} systems, all demonstrate enhanced information transmission compared to nondynamic responses. Theoretical analysis demonstrates that dynamics play a key role in overcoming extrinsic noise. Experimental measurements of information transmission in the ERK network under varying signal-to-noise levels confirm these predictions and show that signaling dynamics mitigate extrinsic noise-induced information loss. By reducing information loss from cell-to-cell variability, dynamic responses improve the accuracy of signaling networks.

Finally, I present a study that couples single cell measurements with an iteratively-developed computational model to examine the respective roles of the MyD88 and TRIF pathways in determining dynamic responses TLR4 stimulation. I show how each pathway encodes distinct features in NF-κB dynamics, and contributes uniquely to the high variability observed in single-cell measurements. In one pathway, the assembly of a macromolecular signaling platform dictates initial response timing and provides for a reliable NF-κB signal. In the other, ligand-induced receptor translocation and endosomal maturation combine to produce noisy yet sustained NF-κB signals via TRIF. Thus, I arrive at a predictive understanding of how these molecular mechanisms provide for ligand-dose and pathogen-specific signaling dynamics and information transduction capacities.

Chapter 1

Introduction: context-specific NF-κB dynamics

The transcription factor NF-κB is ubiquitous. This most obviously applies to its presence in almost all animal cell types[29], or its many roles in cellular development and acute responses[35;89;45;12], but it might also be said of biological and medicinal study worldwide. This family of closely-related proteins has been mentioned in some 40,000 publications since 2010. NF-κB is crucial in immunity and inflammation[61], and as such plays an obvious role in infectious disease[64]. It is also thought to play a role in hundreds of non-infections diseases and conditions, with cancer[12] and heart disease[6] prominent among them. Understanding the context-specific behavior of NF-κB will prove essential to advancing the treatment of inflammatory and immune disease.

The primary function of NF-κB is well-understood and straightforward: the protein, once freed from a inhibitor, translocates to the nucleus and binds certain sequences of DNA[69]. It can then can be recognized by other co-factors to help activate transcription[65]. There is dissonance, then, between that rather basic role, and the sheer number of states that mere NF-κB activity seems to inform. That is to say, NF-κB is certainly strongly triggered during bacterial infections[47], and likewise in viral ones[18]. Aberrant NF-κB activation, for too long, is typical of many cancer cells[12]. Activation,

in one manner, shapes lymphoid organ development[14], but in another, aids in rapid B-cell expansion[3]. All of these states are associated with their own independent set of gene products, decisions, and behaviors. Some of these differences are explained by the specific roles of particular NF-κB family members[91] (e.g. RelA vs cRel), and many are explained by the co-presence or absence of other transcription factors, be they lineage-determining (e.g. the macrophage-specific PU.1[43]) or signal-dependent (e.g. co-activation of IRF3[77]). Beyond this, though, it has been hypothesized that we make a crucial oversimplification by our use of the term "activation"[76].

The network controlling NF-κB contains of many distinct levels of negative feedback[35], all operating on separate time scales. The central activation hub, IKK, is stimulated by a diverse set of signals, from infectious agents to warning cytokines to ultraviolet light[11]. All of this gives rise to dynamic patterns: not mere steady-state turn-on, or a transient rise and fall, but a rich family of behaviors[111]. We can trace oscillations or persistence, timing and amplitude differences as we follow the nuclear translocation and subsequent export of NF-κB . These signals are thought to encode information[19] about the nature and degree of stimulus, such that a variety of states can be encoded by the changes in NF-κB behavior over time[10].

One can trace the "dynamic encoding" hypothesis to 2002, with work showing differential roles by the various IκB transcriptionally-activated negative regulators in controlling NF-κB activity[37]. The authors then show differential gene regulation in genotypes where the most crucial of these (IκBα) is knocked out. Subsequent studies began to suggest examples of, and mechanisms behind, stimulus specificity, particularly between the damped oscillatory NF-κB dynamics in populations of cells exposed to endogenous cytokine TNF, versus the persistent translocation patterns produced in cell populations in response to the exogenous bacterial signal LPS[110;22]. Both works hypothesized that a key distinction was the transcription and secretion of TNF in response to LPS, which was thought to serve as a feed-forward mechanism that could ultimately differentiate LPS responses.

In 2004, another group performed live-cell imaging on immortalized cell lines

transfected with fluorescently-tagged p65 (an NF-κB family member) to visualize and measure undamped NF-κB oscillations in single cells in response to TNF, showing some measure of heterogeneity in activation. This approach was immediately challenged[8] as potentially producing results that were artifacts of overexpression of the p65 protein. A transgenic mouse line expressing EGFP-p65 was introduced, with expression and biochemical behavior that seemingly matched the wild type[23]. Single MEFs from this mouse line were imaged soon thereafter[97], and also showed sustained, undamped oscillations in response to TNF(though were markedly more heterogeneous in behavior), demonstrating that the overexpression model could be a reasonable approximation to primary cell dynamics.

Further single cell studies recapitulated the stimulus specificity observed at a population level: though TNF, as elsewhere, induced decaying oscillations over time, LPS induced a separate class of NF-κB activity[53]. None of the LPS-treated cells showed oscillations, but rather half showed a single, roughly 1-hour window of translocation, while the other half of cells exposed to LPS showed windows of persistent NF-κB activity. We do note, however, that this laboratory has not been able to repeat these findings: LPS (and other stimuli) caused oscillatory, not persistent, NF-κB dynamics in two followup studies performed by the same group[31;38]. Other work has focused on differential dynamics (and resultant gene expression) induced by dynamic inputs: Ashall *et al.*[5] demonstrated a capacity for network "reset", allowing high-amplitude responses to a regular, pulsatile TNF stimulus. More recently, this phenomenon has been studied under the context of oscillator entrainment[50], which appears, in specific cases, to cause more efficient gene transcription.

Each of these signal dynamics studies leaned heavily upon computational modeling in their approach. Indeed, the mainstream rise of systems modeling not only occurred simultaneously, but several among these listed are considered seminal publications in the development of that field[37;66]. NF-κB dynamics provided a unique and rich source of data to parameterize accurate systems models that could easily make testable predictions, and these models, in turn, provided a new paradigm for insight. System

components and interactions were now to be quantitatively combined, measured and compared[113], bringing to bear network and dynamical systems analysis tools from engineering and physics[81;21]to aid the understanding of complex mammalian responses.

The more recent availability of single cell data has reframed systems modeling approaches. NF-κB responses are, to the best of our ability to measure them, extremely heterogeneous and unsynchronized[52 77], raising questions about the ability of biochemical networks to reliably transmit information. At the same time, we observe that populations are able to respond to stimuli in a reliable manner[73]. As our understanding is built upwards from the responses of individuals, we seek to understand the interplay between noise and predictability, variability and robustness. Theories springing from the earliest measurements of noise sources in prokaryotic cells[25] have been more recently applied to form predictive models of fractional apoptosis[93]. In the NF-κB system, research to date has focused upon stochastic elements intrinsic in receptor binding or initiation of transcription[5;99;50;113]. Other work has looked more generally at the capacity of the NF-κB system to reliably transmit input information[19;79], finding that in TNF stimulation, most of the input information loss is upstream, near the receptor, while IκB -NF-κB signaling is comparatively reliable.

What is missing from all of this is context. NF-κB is considered to be a "master regulator" of innate immunity and inflammation, which means that its activity can be far better understood in a cell such as the macrophage, whose major functions generally involve the activation of this particular transcription factor. Macrophages act as sentinel cells and primary mediators of immunuty in multiple tissue types[116;51], and as such, are exquisitely tuned[26] to activate defense responses[59]. Dysregulated inflammatory states are incredibly destructive, and are major factors in the pathogenesis of cancer[12] or metabolic disease[6]. Therefore, the process of sensitively encoding threats in the dynamics of NF-κB , and the various mechanisms of decoding these responses are all best understood in the circumstances closest to how they would occur *in vivo*.

This, finally, brings us to the purpose of this work. We know that individual macrophages, with networks and activity shaped and informed by their local microen-

vironment, initiate gene expression and behavioral changes in response to perceived threats. This is the ”what”, and we naturally begin by defining these particulars. But more than this, we seek to understand the ”how”: how consistency can be derived from inescapable variation, how underlying mechanisms control this variability, how information can best be transferred in biochemical networks, and how these cells, granted incredible power over their surroundings, regulate the balance between defense and destruction. The work presented here focuses on the development and application of computational and analytical tools designed to further these goals.

Chapter 2

Automated analysis of signal dynamics in macrophages

Introduction

A bottleneck in the field of single cell signaling, particularly in immunity, is the measurement and analysis of high-throughput imaging data. Zeiss, Nikon, and Canon all offer fully-automated epifluorescent or confocal platforms that allow stable cell culture and time-lapse imaging for a period of hours to days, but the conversion of tens to hundreds of gigabytes of imaging data, to connected trajectories of single cell behavior is not a trivial task. Generally, the solution to this problem can be broken down into three parts: (1) identification of nuclear (and cellular) boundaries in a given image, (2) updating boundaries as the next image in the full sequence is updated, or else matching a new set of boundaries to the cells identified in the prior image, and (3) appropriately quantifying the pixels contained in each individual's boundaries, generating a trajectory of that cell for each frame that it appears in.

Proposed solutions to this issue vary significantly by field: laboratories specializing in medical imaging or computer vision frequently publish cutting-edge analysis techniques based on advanced algorithms and machine learning[63;80;120]. These studies

are generally extremely limited in the scope of analysis, focusing on sample image sets, and are rarely, if ever, cited or used in an experimental context. By contrast, even the most recent, high-profile work published in biologically-focused labs employs rudimentary, often semi-manual techniques[96;93;77]. As a result, the total number of experimental conditions and individuals measured per condition are often very limited.

Compounding this issue is the highly heterogeneous nature of macrophages themselves. Prior efforts to measure single cell dynamics often relied on fibroblasts, or a fibroblast-derived cell line (*e.g.* 3T3 cells)[97;53;99]. These cells grow in a confluent monolayer, and as a result, individual cells move very little over the course of an experiment, making frame-to-frame tracking a trivial task. Additionally, most fibroblasts have similar sizes and thicknesses, such that simple, distance-based means of segmentation are sufficiently accurate. By contrast, not only do macrophages present in a bewildering variety of sizes and shapes, they often undergo extreme morphological changes as a direct consequence of stimulation (Fig. 2.1). Accurate algorithmic tracking of these cells, then, requires code that is extremely flexible, and utilizes as much information as possible in making assessments in where one cell ends, and another begins.

In addition to the premium placed on algorithmic flexibility in assessing what is, and what is not, a cell, we sought to create a tool that would prove provide the maximal experimental flexibility. First, in place of relying on a whole-cell fluorescent stain, we chose to design routines that would accurately identify cells in brightfield imaging modalities (either phase contrast or DIC). As a result, we preserve the maximal number of fluorescent channels for the study of other species, even if they are not uniformly distributed in the cytoplasm. Secondly, we looked to make this tool accessible to non-programming users. To that end, we designed visual tools that aid in tracking parameter selection. Third, we designed the program to be maximally modular, particularly in its handling of measurement: this allows ease of extension to different imaged conditions, channels and species. Finally, we created a related tool that allows for rapid exploration of tracked datasets. These sets are often messy, containing cells that may need to be filtered, scaled, and sorted for understanding. This ”CellQuery” tool was designed to

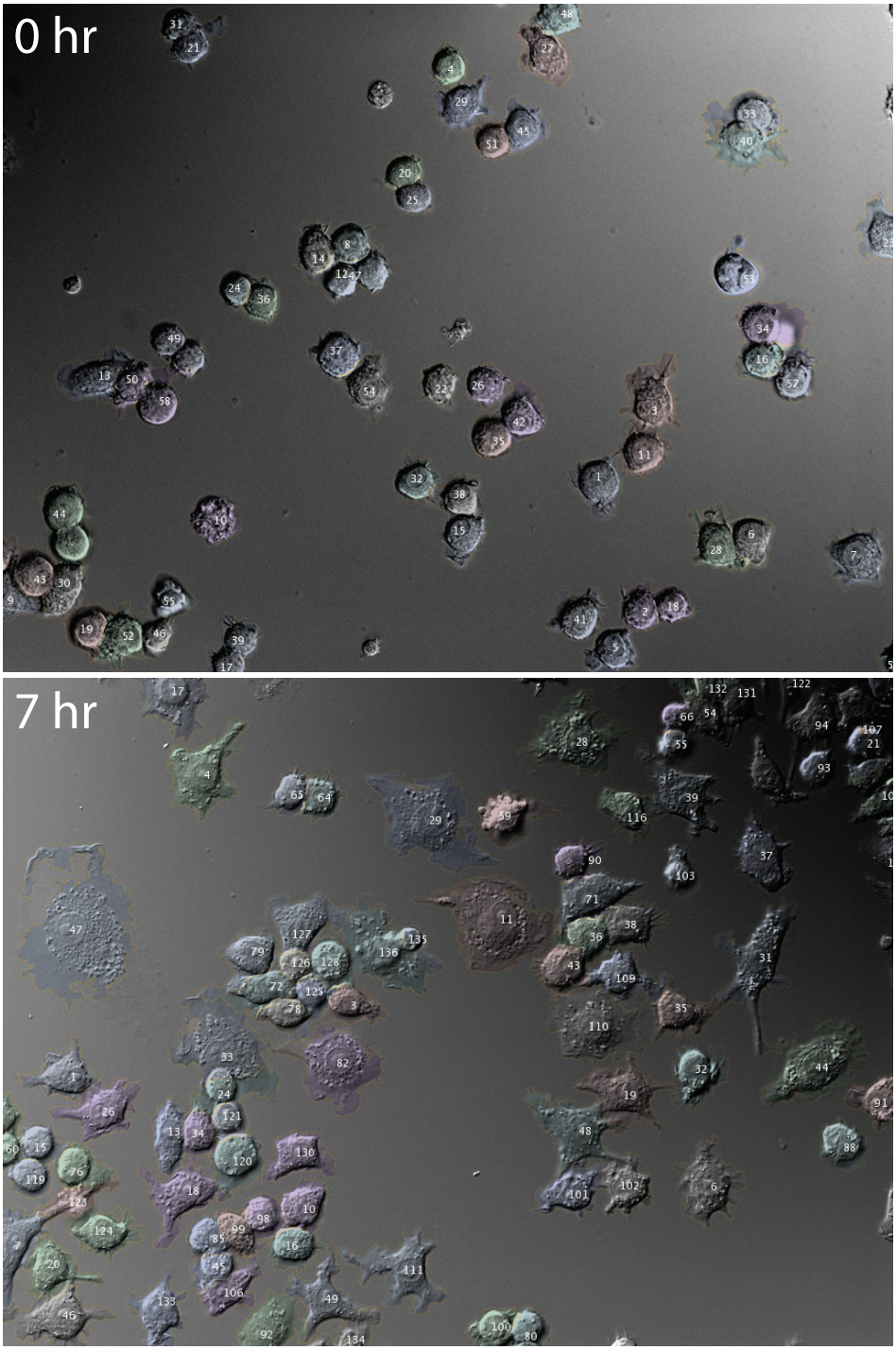

Figure 2.1: Heterogeneity and morphological changes in RAW 264.7 macrophages. Top and bottom images show the same set of cells at 0 hrs (before stimulation), and 7 hrs after stimulation with 500ng/mL LPS. Both primary and immortalized macrophages show large amounts of heterogeneity during activation, whether measured by motility or morphological change.

allow discovery and visualization of high-dimensional results. Our implemented image processing routine consists of 6 major steps:

1. Cell identification in phase contrast or DIC images
2. Nucleus identification
3. Nucleus tracking/error correction
4. Shape-based segmentation
5. Error correction
6. NF-κB /other measurement

I. Cell identification in phase contrast or DIC images

Cell identification forms the first step in the segmentation process. The task, at heart, typically relies on a thresholding method to distinguish foreground from background. This threshold can often be applied to the distribution of intensity values in the image, and there are many variants on thresholding algorithms in current use[85;71]. In the case of phase contrast or DIC microscopy images, however, we apply a vertical and horizontal Sobel edge transformation (Fig. 2.2B): we can then threshold on edge strength. The resultant distribution of edge values often has high outliers, which strongly skew the results of traditional thresholding methodologies: log-compression, an oft-applied technique to help correct this issue in intensity distributions[17], causes inconsistent results when applied directly to the transformed image.

We therefore employ a custom speckle-noise-based threshold of the combined edge magnitude image formed the basis for foreground-vs-background determination. Noting that optimal threshold placement is generally the most lenient value possible before accumulation of speckle noise in the image, we calculate speckle noise as a function of threshold, and directly compute this value by finding the inflection point on the resultant curve(Fig. 2.2C-D).

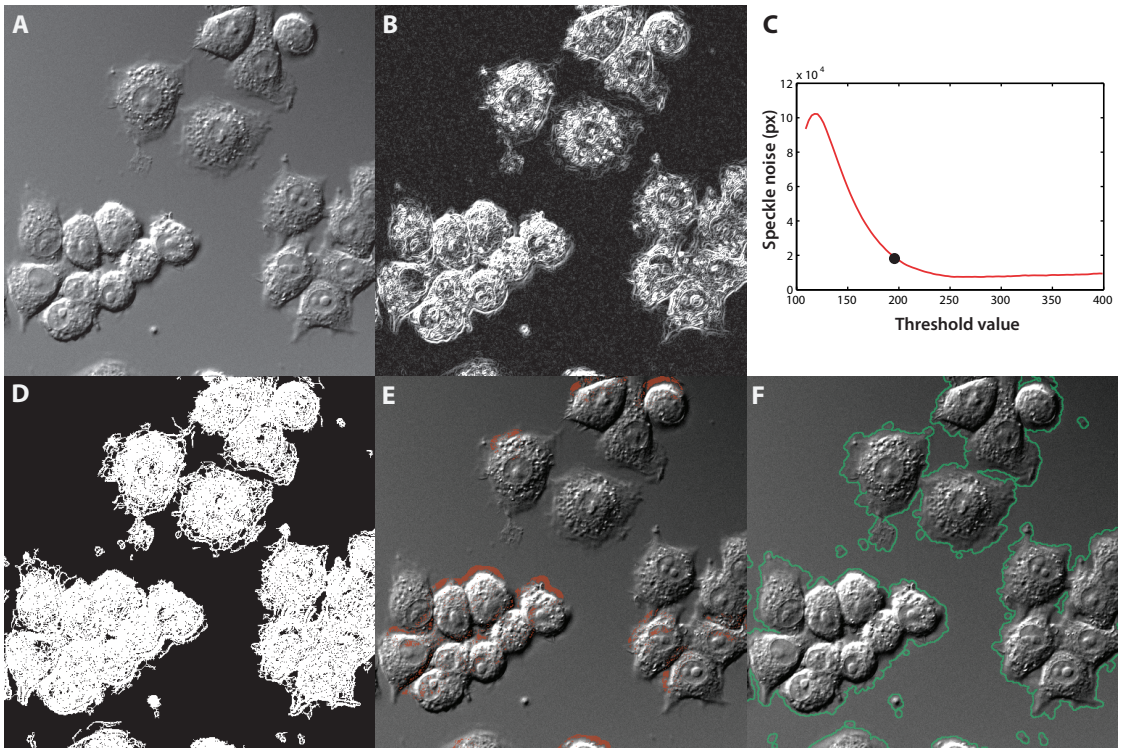

Figure 2.2: Identification of foreground in brightfield images. **(A)** Original image. **(B)** Total magnitude of horizontal and vertical Sobel edge transformation. **(C)** Edge threshold is placed at inflection point after calculating speckle noise. **(D)** Initial edge masking. **(E)** "Do not fill" areas are found after weak-edge calculation, and used to prevent overestimation of cell boundaries. **(F)** Final cell mask overlaid on original image.

To better identify spread cells with weak outer edges, we employ a variation upon classic two-step Canny edge detection[115]: we use the preliminary masked Sobel image to ”mark” edges in a Gaussian-smoothed image. A secondary ”weak” threshold is additionally calculated, using the Tsai threhold[103]: this result is then edge-thinned, and then unmarked edges are dropped. This smoothing and edge-finding operation is repeated at multiple scales of the Gaussian kernel, iterating from small to large values. We then fill gaps: marked edges are then dilated pixel-by-pixel, then skeletonized, thus checking for possible connectedness to nearby edges. We then intelligently fill holes: this method tends to overestimate foreground near strong edges, so we determine candidate ”do not fill” areas, and fill (based on size) otherwise (Fig. 2.2E). The final mask is shown in (Fig. 2.2F).

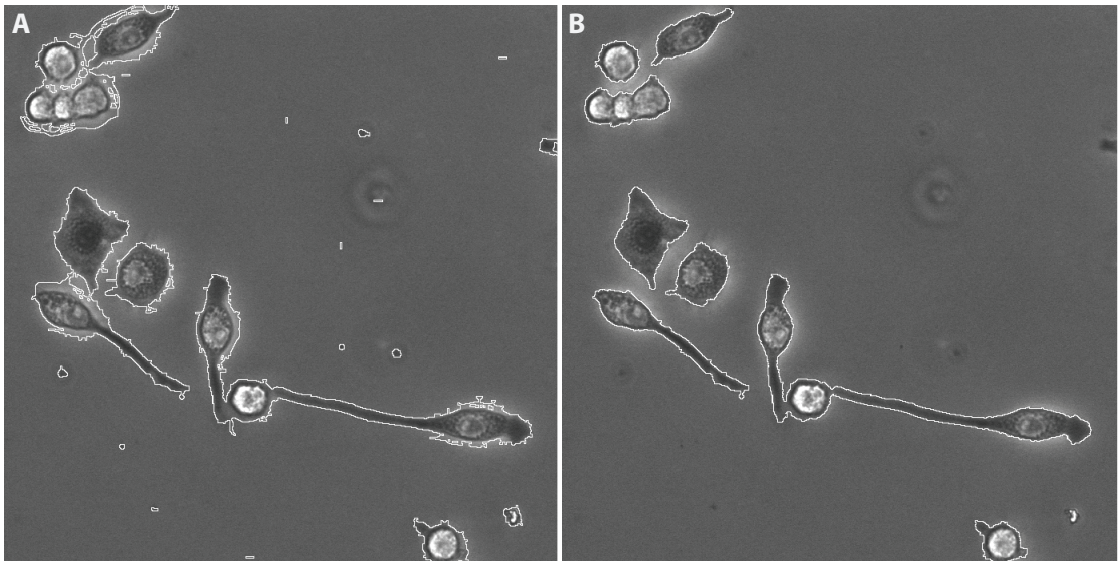

Figure 2.3: Phase-contrast halo correction. (A) Original mask in a phase contrast image. **(B)** ”Walked-in” mask no longer contains halo artifact.

For phase contrast images, we implement one additional step to counteract the bright ”halo” artifact surrounding each cell, typical of this imaging modality. The image mask is refined, as a final step, using a stepwise, gradient-directed erosion process, implemented similarly as a published algorithm[15]. This step effectively walks the outer boundary in until the halo is no longer included, providing accurate foreground calling

in these images (Fig. 2.3).

II. Two-stage nuclear identification

A ”seed” channel is required in this system in order to assess the number and position of cells in a given image. Nuclei are an attractive target, as their relatively homogeneous size and shape make separation of touching objects simple. Without this channel of information, accurate cluster separation is often difficult to achieve by an expert human, let alone an algorithm. In most cases, cell nuclei were marked using a genetic reporter (either a fluorophore with a nuclear localization sequence, or a fluorescently-tagged histone). The particular technical challenge here was dealing with objects of disparate brightness: a single threshold applied to nuclear intensity would vastly overestimate some objects, while entirely missing others. Additionally, fluorophore localization to the nucleus is often incomplete, leading to a cytoplasmic signal in a subset of cells that had to be discarded (Fig. 2.4A).

The search space for nuclei is constrained to foreground-identified regions from step I. Nuclei are then identified in a two-step process: first, using the Sobel edge-magnitude transformed image, we iteratively step candidate thresholds from high to low values. At each step, we fill any closed shape that was formed. If a suitably-sized object results from this process, then search in the affected local area is halted before moving to the next threshold value (this local area is identified using a watershed transformation on the smoothed nuclear image).

In the second stage, weaker nuclei are identified by searching and ranking all pixels in each remaining watershed region by percentile (Fig. 2.4C) - appropriately concentric regions are scored, isolated, and combined with the first set of nuclei. In the final step, we discern between ”pieces” of individual nuclei that need to be combined, versus touching nuclei that require separated, by merging all possible combinations of touching objects, and finding the combination that minimizes both the total morphological compactness (or ratio between object area and perimeter) and elliptical eccentricity of

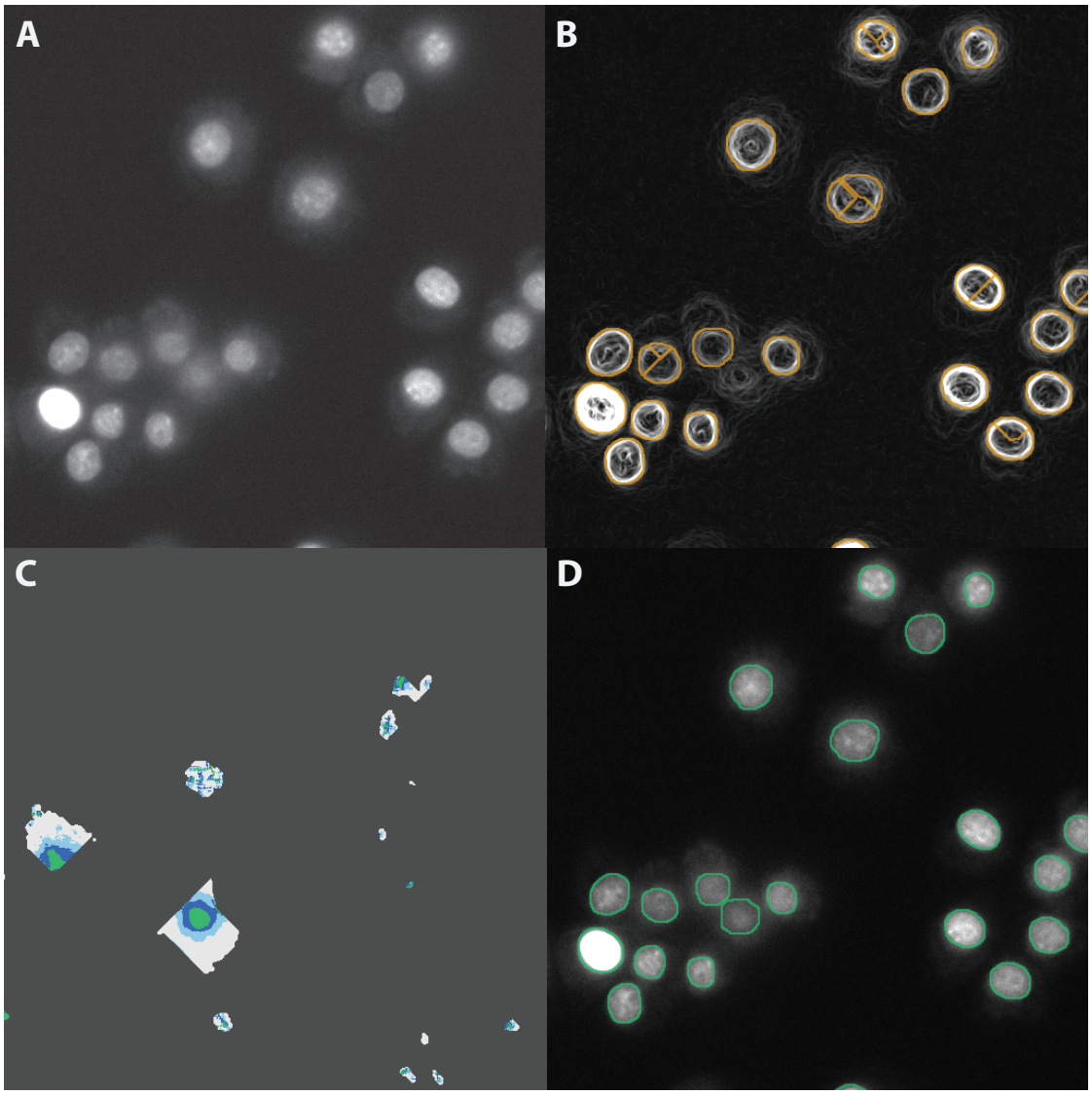

Figure 2.4: Identification of nuclei in fluorescently-labeled images. **(A)** Original nuclear image.**(B)** Primary (strong) nuclear identification, via iterative edge thresholding, then filling closed shapes. **(C)** Secondary (weak) nuclear identification, via identification of suitable concentricity, after ranking intensity of remaining pixel areas. **(D)** Final nuclear label, overlaid on original image.

each object. This procedure is motivated by existent work[80], and allows assembly of intact nuclei in cases where the object was inadvertently subdivided (Fig. 2.4B). Overall, this procedure benefits from reasonably accurate estimates of minimum and maximum nuclear size, but shows excellent performance in the range of expression levels typical of live-cell reporters.

III. Multiframe nuclear tracking

The task of tracking cells requires linking trajectories across hundreds (or even thousands) of frames, and making decisions about drift out of frame, division, and death. Even a highly accurate nuclear identification technique would still be expected to have measurable error rates of a few percent (which may be compounded by differential expression of the nuclear marker in an individual over time). We would expect trajectories to be destroyed at a rate of $1/\epsilon$, where ϵ is the average error rate of the nuclear classification algorithm. Even simple object recognition tasks (such as number recognition in cropped images), as performed by the the most current and powerful algorithms currently available, still struggle to achieve $\epsilon<0.04$[30], which would suggest trajectories that get broken, on average, after 25 frames.

To maintain cell lineages, we apply a tracking algorithm across 7 frames simultaneously. Label matricies, or the location of all putative nuclei in a given frame, are processed into a queue that is initially filled: all possible links are computed between each object in frame n, and all viable candidate objects in subsequent frames (frame n+1 to frame n+6), based on a projected drift distance between frames. These links are then resolved into high-confidence matches based on the unique frame-to-frame matches that minimize distance and morphological similarity (area and perimeter) of the objects. These matches are then combined into ”blocks”, the unambiguous list of each object’s likely location in each frame.

After the queue is filled, it is updated with every new frame: at the (new) end of the queue, incoming objects are processed into existing or new blocks. At each iteration

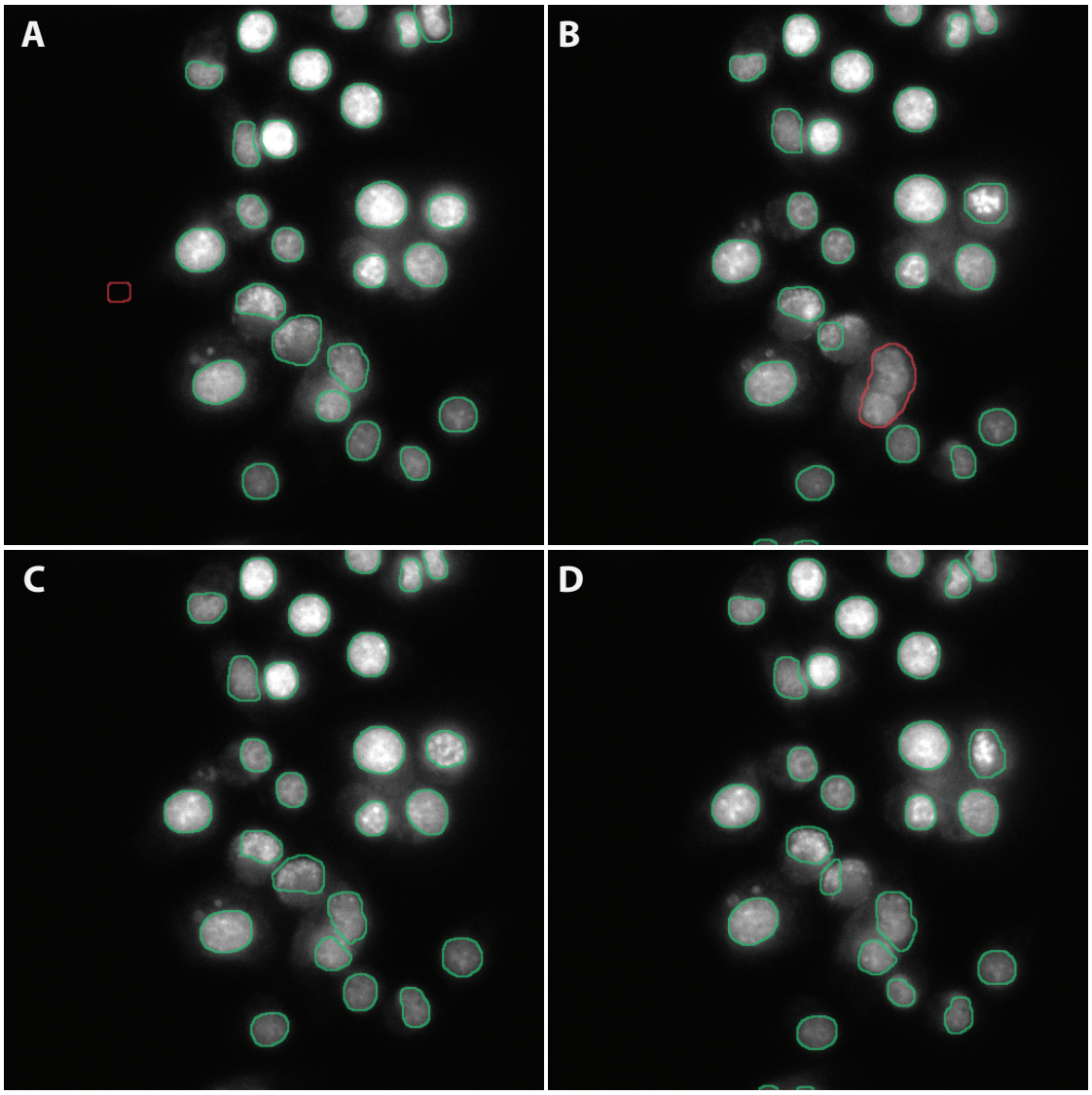

Figure 2.5: Tracking and error correction using voting among images in queue. (A) A false positive in frame 1 was not found in majority of subsequent frames and so can be ignored. **(B)** A falsely-joined set of nuclei in frame 2 can similarly be identified by voting. **(C-D)** Frames 3 and 4 in queue, used to help correct initial nuclear identification in frames 1-2.

of the ”update” loops, we make final decisions on the fate of blocks at the bottom of the queue: whether a particular block should be merged with another block in the list, whether a terminating block represents a false negative, or death (or cell drifting out-of-frame), and whether a new block represents a division (or drift-in). In these cases, we can use the full queue to vote on cell fate: for instance, false negatives on a single object (multiple skipped frames) can be corrected if a high-confidence match is found in a majority of later frames in the queue. Thus, we can correct common issues such as spurious positives (Fig. 2.5A) or falsely combined nuclei (Fig. 2.5B-D). Once this process is completed, we can drop the (old) bottom of the queue, and move to to the next stage. We perform this step prior to segmentation, assuring that the most accurate cell count and seed is handed off to that process.

IV. Shape-based object segmentation

Commonly-employed cell segmentation techniques often use distance and/or intensity information to heavily segmentation results in clusters[41;36], meaning they are ill-suited to drawing boundaries between clusters of brightfield-imaged macrophages, where edge-to-center distance is quite disparate between cells (Fig. 2.6A), and cell-depth variability leads to inconsistent fluorescent profiles at the edges of cells. We thus utilized cell shape information: though cell shape may vary dramatically, it is very often the case that an inflection point marks the boundary where two cells meet[107]. Therefore, to aid segmentation, we employ this shape-based approach ,which splits cells at potential inflection points between nuclei, in an effort to achieve accurate segmentation.

As a first step, to speed computation, the image is broken into subsets, with each subset holding precisely one cluster (*i.e.* an area of cell/foreground that contains more than one nucleus), and then downsampled. The morphological skeleton of the overall cell shape is then computed (Fig. 2.6B). The center of nuclei in the cluster is matched to its nearest point on the resultant skeleton. This skeleton is pruned to consist solely segments connecting all nuclei of cells within the cluster. A distance transform calcu-

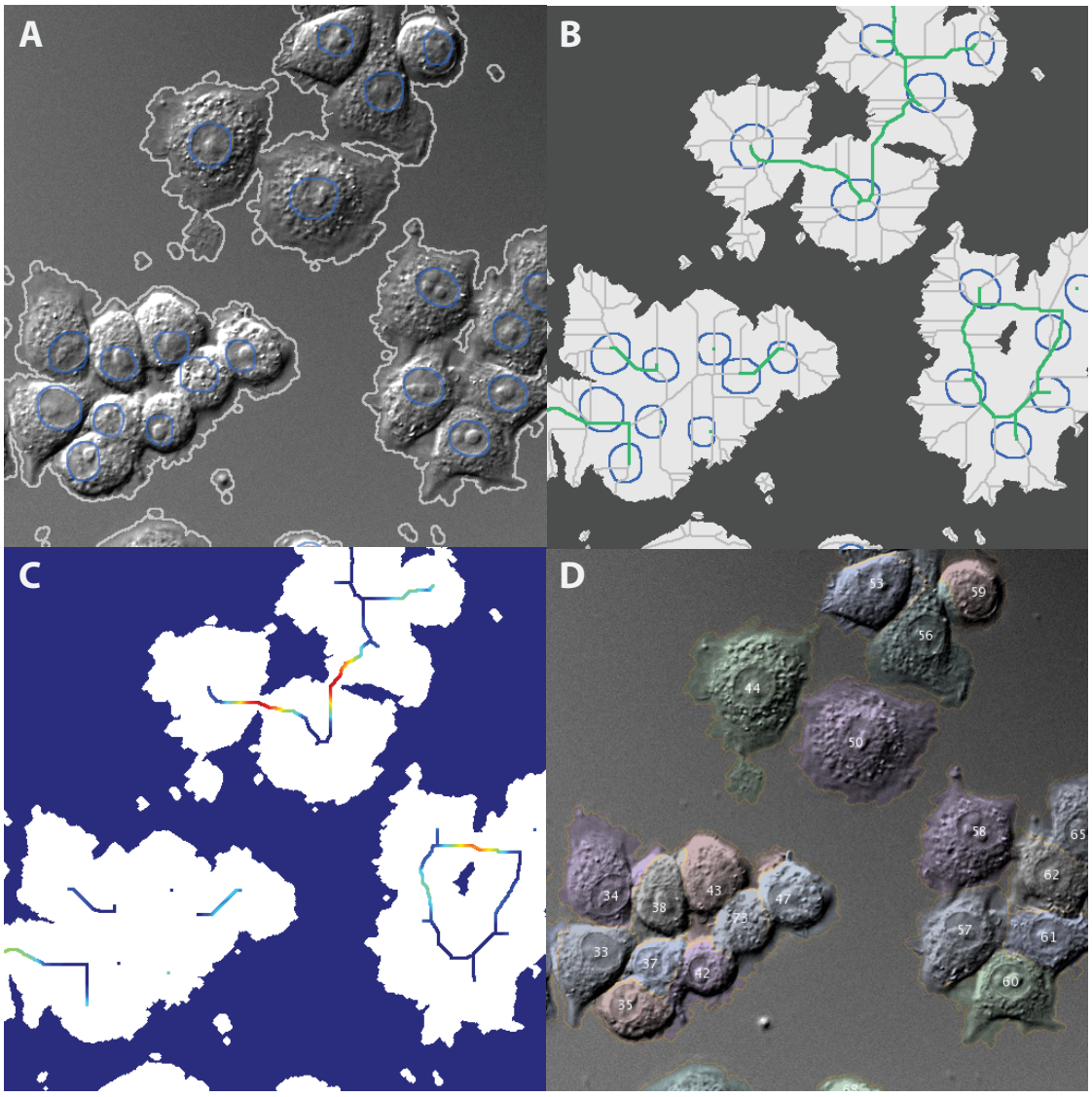

Figure 2.6: Shape-based object segmentation. **(A)** Segmentation begins using the nuclear and cellular boundaries identified in sections I and II. **(B)** The morphological skeleton is computed, then pruned to areas connecting each nuclei **(C)** We identify the local maxima of the distance transformation along the pruned skeleton, as candidate splitting points.**(C)**Final segmentation result, overlaid on original image.

lated along these pruned-skeleton segments shows distinct local maxima (Fig. 2.6C) that correspond to inflection points, or candidate points for the boundary between these two cells. We break the cell cluster mask at each of these candidate points (this operation is restricted to ”cuts” of a maximal length). The resultant mask then informs the a first step in the segmentation decision, which is ultimately made by two iterations of the ”propagate” algorithm employed in the open-source CellProfiler[41].

V. Error correction and efficacy of tracking

Following segmentation step, we have implemented a rearwards-looking correction routine, which seeks to enforce biological identity. This ”memoryCheck” algorithm is case-based, and looks, in order, for the following specific errors: they are all but impossible biologically, but reasonably common in any tracking algorithm.

- Cell jump/swap: tracking error leads to improper switching of nearby cells, or the identity of a tracked cell ”jumps” to an untracked one (Fig. 2.7A-B)
- Large area increase: cell experiences a 50% increase in total area in a single frame (Fig. 2.7C)
- Large area decrease: cell experiences a 25% decrease in total area in a single frame (Fig. 2.7D)

Detection of cell swapping/jumping calculates each cell’s frame-to-frame displacement, then checks whether the majority of that displacement line runs across another cell. If a cell nearby also experienced a jump/swap, we can resolve this error by switching back the cells. If no swap partner is found, we can assume that the problem was when a false negative event was paired with a nearby false positive; we therefore destroy the latter, and create a new nucleus in its last known position, before passing it to be resegmented.

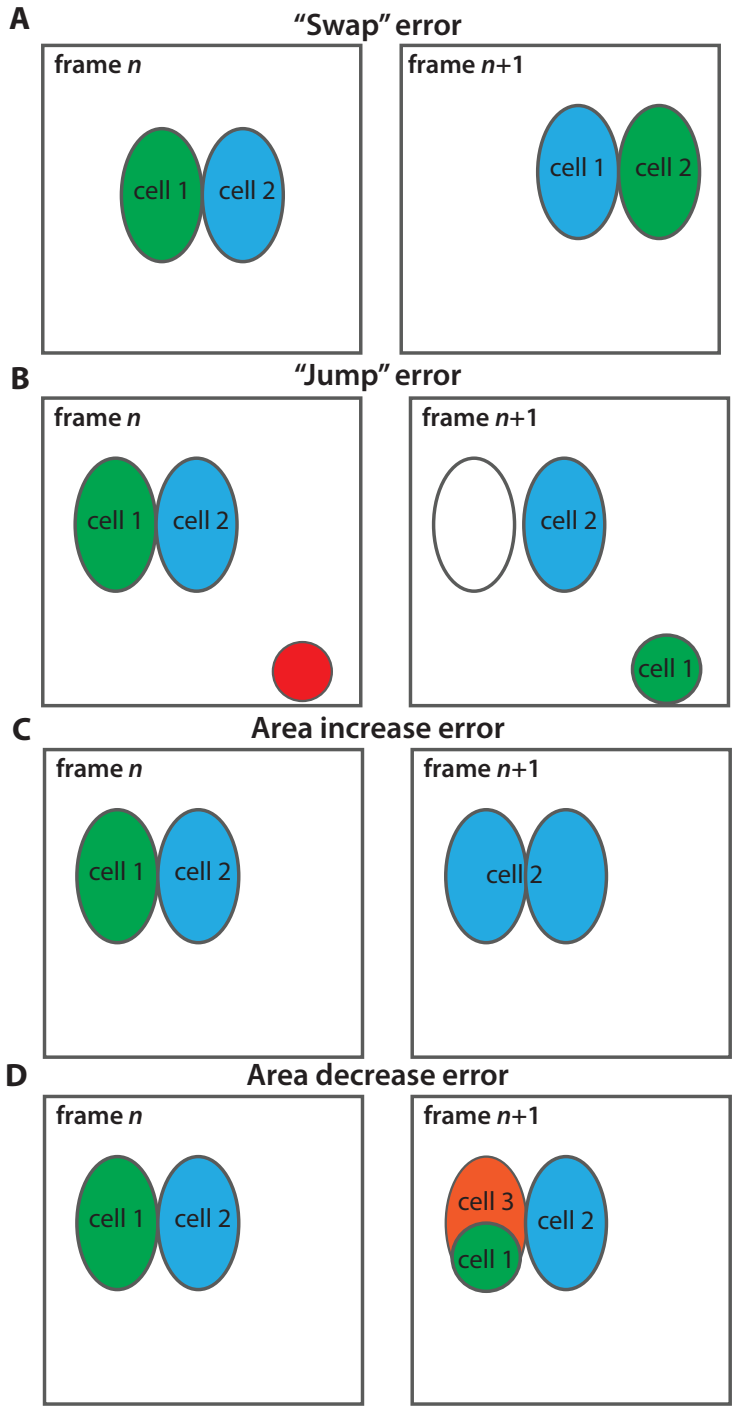

Figure 2.7: Automatically identified/corrected errors. **(A)** A swap error: typically caused by paired movement in cells, such that cell 2's position in frame *n+1* is close to cell 1's position in frame *n* **(B)** A nearby, simultaneous false positive and false negative can result in a cell jump.**(C)** Area increase error: sub-case where one cell may take area from a neighboring cell is shown.**(D)**Area decrease error: sub-case where a spurious false positive cell takes area from a preexisting one.

In the case of a large area increase, we seek first to categorize the source of the new area: it could be, in order: a false negative nucleus (*i.e.* what was 2 cells is now being called 1), a segmentation error with nearby cells (*i.e.* "stolen" area), or a background/foreground assignment error, which commonly happens with inadvertent changes in imaging conditions (*e.g.* bubbles on the media surface). Each of these cases triggers its own error-handling process. Area decreases are handled in a similar manner: the algorithm attempts to determine where the lost area went, and triggers separate error handling routines depending on that assessment.

Error correction and forward-looking tracking turn tracking across multiple frames, which would normally be a liability that introduces greater error, into a means for verification and validation. In the small time intervals (generally less than 10 minutes) needed to accurately quantify signal dynamics, the changes in any given cell are slight. Therefore, we can use simple biological realities (*e.g.* cells are relatively stable, and tend neither to disappear nor appear, nor change dramatically, in a short amount of time) to enhance the accuracy of the analysis result.

To quantify overall accuracy, we employ the commonly-used *F-score*, the geometric ratio of *precision* and *recall*. The formulas for each of these are based on the automated segmentation result for a given object (*i.e.* cell), and a ground-truth image (*i.e.* that same cell, manually outlined by an expert). We can thus calculate the intersection of the two sets (true positives or *tp*), and the set of pixels contained in only one object but not the other (false positives (*fp*) and false negatives (*fn*), respectively)

$$precision = \frac{tp}{tp + fp} \tag{2.1}$$

$$recall = \frac{tp}{tp + fn} \tag{2.2}$$

$$Fscore = 2 \times \frac{precision \times recall}{precision + recall} \tag{2.3}$$

Segmentation success rates in other published algorithms, performed upon on sample data sets, generally fall in a range of 0.80-0.90 (*e.g.* a median score of 0.80 in

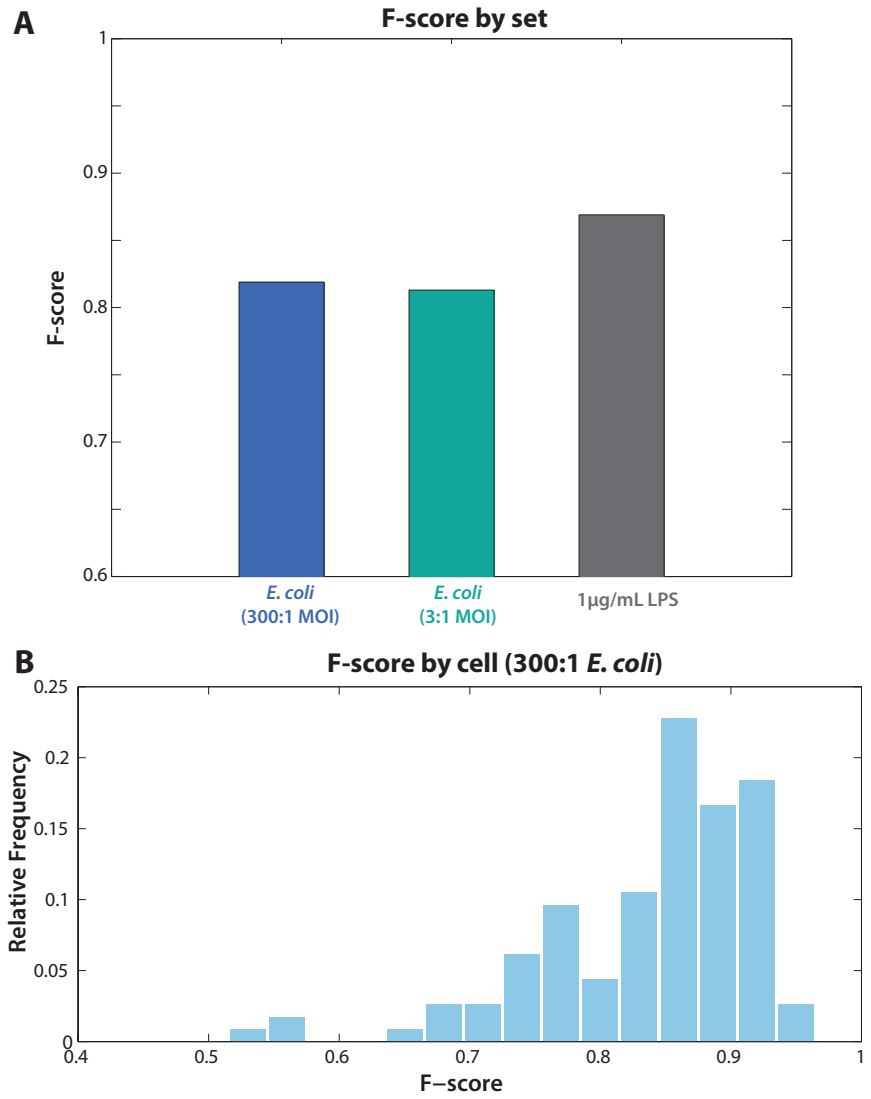

Figure 2.8: Verifying tracking accuracy. **(A)** Total F-scores from 3 different experimental conditions. Images were captured on separate days. **(B)** F-scores, by cell, from randomly chosen RAW 264.7 cells (at various timepoints) stimulated with *E. coli* (300:1 MOI).Median F-score for all cells measured is 0.89, and 70% of cells show an F-score of over 0.86.

macrophages[84], or a total F-score of 0.87 in CHO cells[63]). To assess accuracy of this algorithm, we delineated the boundaries for 100 cells from randomly-selected timepoints during three different days of experimentation, in three different conditions. We then compute F-score, as above, comparing the automatically tracked/segmented result with this ground-truth experiment. Total F-scores (*i.e.* ratios over all tracked pixels contained within indivdual cells) are shown in fig. 2.8A. Additionally, a breakdown of F-score by cell in cells stimulated with *E. coli* (300:1 MOI) is shown in in fig. 2.8B. The median F-score, in this case, is 0.89, with over 70% of cells showing an F-score of over 0.86. We are able to achieve, then, state-of-the-art results in actual experimental results, rather than just in the training sets of images that the algorithm was originally designed for.

VI. Measurement and of NF-κB and related quantities

In this section, we outline the novel techniques applied in service of accurate and robust measurement in alternate fluorescent channels: while tracking offers morphological and gross behavioral outputs (*e.g.* cell division or death), this fails to capture many relevant aspects of the macrophage response to immune signals. In particular, this work focuses predominantly on signaling events, particularly the translocation response of NF-κB .

Raw nuclear intensity traces could not be used directly, as macrophages experience strong morphological changes in response to immune signals. To account for these changes, we computed both the median value for the nucleus of each cell (Fig. 2.9A), and a value representing the cytoplasm - we chose the upper mode of a generally bimodal distribution, corresponding to the brighter portion of each cell, close to the nucleus. This cytoplasmic function is insensitive to nuclear translocation events (Fig. 2.9B), but does show pronounced changes that were assumed to be indicative of the overall morphological shape of each cell. We used the cytoplasmic trajectory to fit the changes in nuclear values of each cell - cells were assumed to be in an off/zero state at time t=0, and at extremely late timepoints (t=12+ hrs), so baseline values from these two points were

used in fitting the magnitude of the changes, such that we calculated a reliable baseline (Fig. 2.9C) from which nuclear translocation could be measured.

Finally, it has been strongly suggested that relative, not absolute, NF-κB signal is more informative in predicting downstream cellular responses[52]. We thus take the final step of normalizing each cell's NF-κB trajectory to its initial starting cytoplasmic value. This same morphological and normalization methodology can be generalized to other translocation events, and is used to track IRF3 activation in chapter 4.

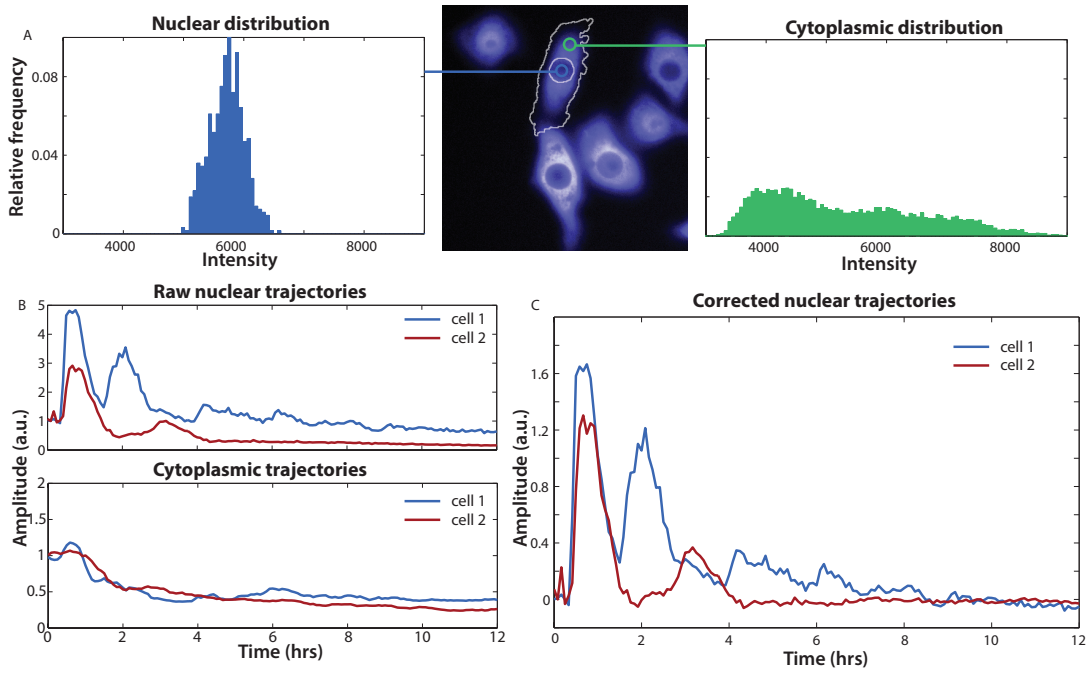

Figure 2.9: NF-κB measurement. (A) We calculate values that correspond to the nuclear (blue, median/mode) and cytoplasmic (green, higher mode) intensity distributions, which show identical decreasing trends over time (as a function of changing cell morphology). **(B)**Raw nuclear trajectories show cells are off before stimulation, and eventually stabilizes after a maximum 10-14 hrs. We use this information, along with the shape computed from each cell's cytoplasmic trajectory, to calculate a true baseline for each cell. **(C)** Final, corrected and normalized nuclear trajectories can be directly compared between cells.

User interfaces for intelligent parameter selection and data display

While parameter-free methods are becoming more prevalent, particularly with the rise of deep learning algorithms[30,63], efficient parameter selection can often be faster in handling changing conditions (*e.g.* cell type, microscope, or magnification) than generating a new training set of labeled images: in places where data acquisition is generally adapted and optimized over time, this can be advantageous. However, though GUI-based tracking/segmentation programs (CellProfiler) often provide detailed diagnostic output[17], they provide little or no guidance when choosing parameters, forcing users to use a guess-and-check methodology. This furthers the disconnect between the biologist and the programmer: in order to facilitate software learning, we seek to lower barriers to entry wherever possible. While certain parameters are abstract, and difficult to display directly, several parameter classes can be directly visualized:

- Noise-based thresholding search range: while the curve of speckle noise as a function of edge magnitude threshold can be automatically searched (and varies with image brightness), narrowing of search range can increase accuracy and speed. Graphic display allows the user to find the characteristic curve and set search bounds around either side (phase contrast/DIC), or the high start value in thresholding the nuclear image(nuclear fluorescence)(Fig. 2.10A-B, left).

- Size-based parameters: particularly in nuclear identification, size proves a valuable tool in shape identification. Unfortunately, it is difficult to have an intuitive feel for how object is (in pixels) in a given object. After image loading, we display objects defined by minimum and maximum radii superimposed on the image, allowing direct comparison and appropriate selection. Similarly, we show size-based hole-filling parameters on the transmitted-light image (Fig. 2.10A-B, right).

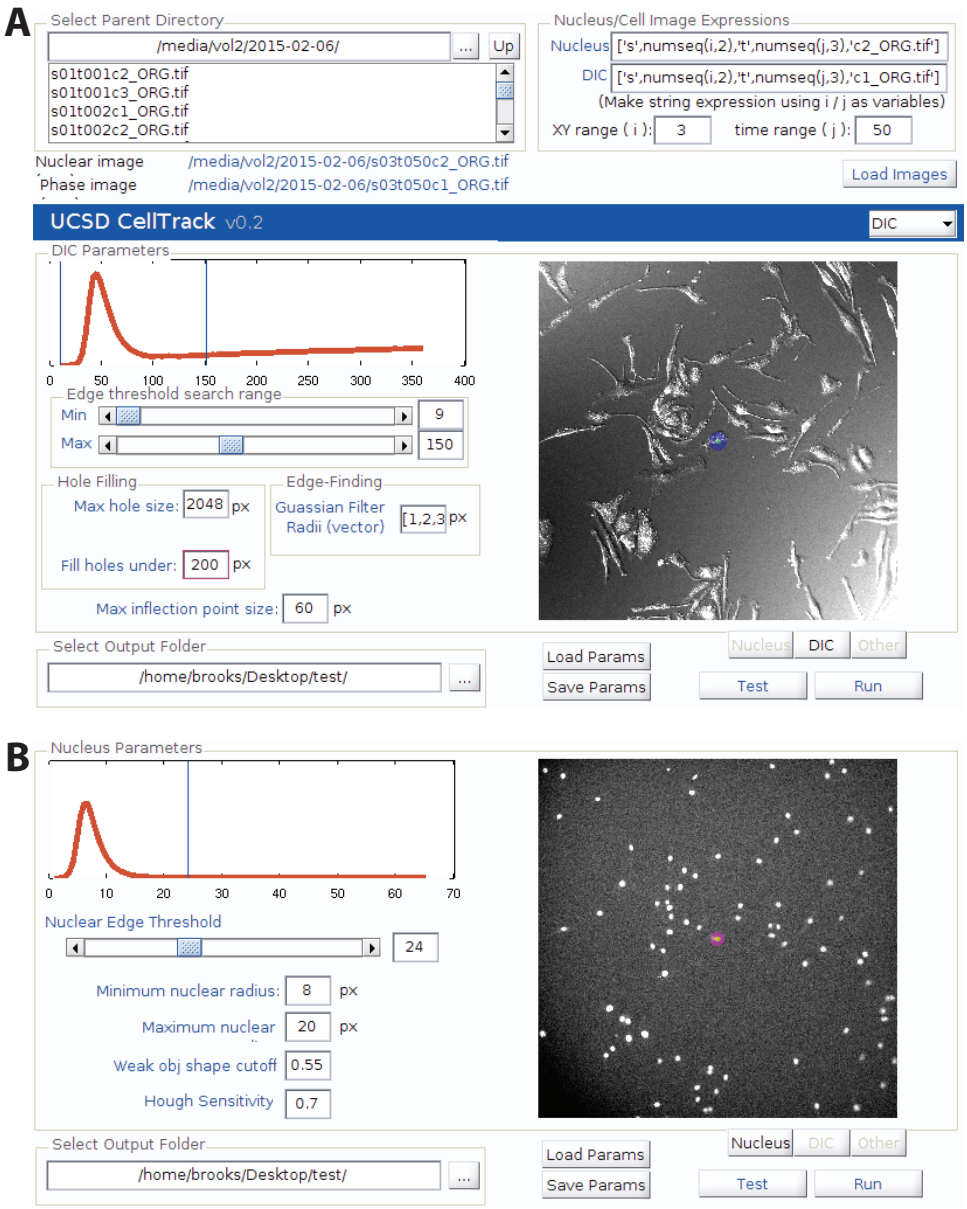

Figure 2.10: User interface for graphical parameter selection. **(A)** DIC image parameter selection. *(Left)* Choosing search range for speckle-noise algorithm in DIC (or phase) image. *(Right)* Hole-filling parameters overlaid on DIC image. **(B)** Nuclear image parameter selection. *(Left)* Choosing starting value for edge threshold in nucleus: though imaging modality is different, shape of speckle noise function is similar.*(Right)* Setting minimum and maximum size values for nuclei using overlaid shapes.

These tools are finally coupled with a general-purpose suite of routines that enable the user to rapidly collect high-quality trajectories, smooth and statistically model results, and display the data in a variety of ways (line plots of trajectories, heatmaps, histograms, or state-space representations). Additionally, we allow allow parsing of Zeiss metadata to collect positional information, and verification of results, overlaying nuclear and cytoplasmic boundaries on any original image channel, above, as appropriate, the object's measurement over time (Fig. 2.11). In the end, we considered the problem as a complex space, which could be studied from many angles: though we focus in this work on the activity of the transcription factor NF-κB , this toolset was designed to be readily extended to focus on other dynamic processes in this system.

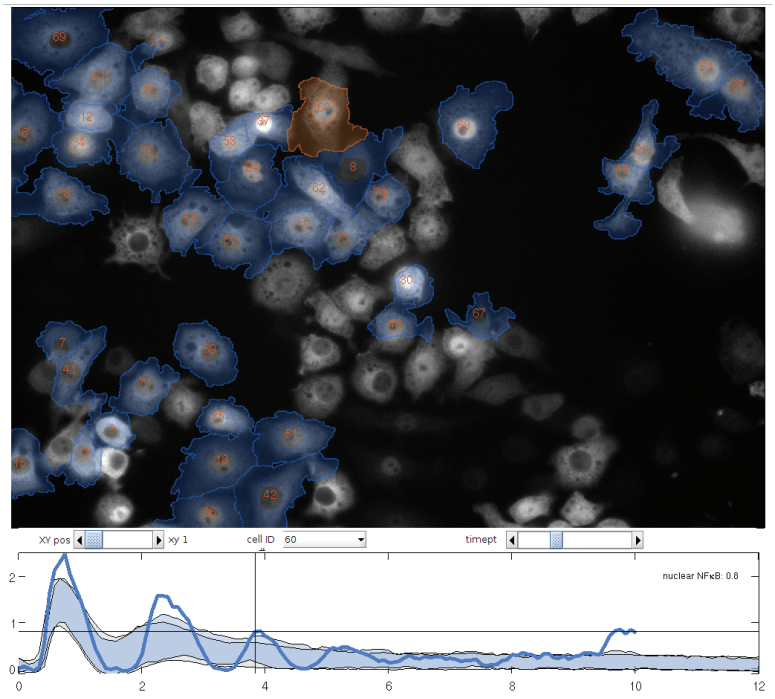

Figure 2.11: Visualizing image and measurement output. In order to facilitate data processing decisions, we created a tool to show a single cell's NF-κB behavior (orange cell; dark blue line)in the context of its neighbors (light blue) as well as all cells in the experiment (in light gray) in addition to the tracked image output at any given point in time.

Chapter 3

Faithful information transduction in dynamic signaling systems

Introduction

The question at the heart of the macrophage response to innate immune stimuli is one of information transfer. The vital function of these cells is read environmental cues, which can trigger appropriately strong genetic and behavioral defense responses. A century of biological research since macrophages were first discovered has carefully described these responses, and named many of the molecular players involved. However, the past decade has made it abundantly clear that merely knowing the names and natures of these interacting parts is not enough to gain significant understanding and predictive power, particularly in the clinic[20]. One emerging shortcoming, in particular, is that mean population-level behavior often masks individual reality: essential variation and distribution[4;104]. It is critical that our grasp of immune responses bridge to the level of the fundamental decision-making unit, the cell. We seek, therefore, to find meaningful frameworks for interpreting single cell responses, made in the context of both external information, and internal levels of variation and noise.

The role of biological signaling networks is to reliably transmit specific informa-

tion about the extracellular environment to multiple intracellular downstream effectors, allowing the cell to adjust its physiological state to changing conditions. The stochasticity of molecular interactions that underlies various forms of "noise" in biological systems[25] can interfere with signal transduction and degrade the transmitted information. How signaling networks perform their core functions in the presence of noise is a fundamental question in signal transduction. Information-theoretic approaches allow estimating the information transmission capacity of noisy biochemical networks[78;24;118;11]. Previous applications of such methods to the analysis of signaling networks suggested that due to noise, cells lose most of the information about the concentration of ligands and only reliably identify the presence or absence of an activating ligand[19;106]. The suggestion that noise degrades most of the information about the activating ligand was surprising given the demonstrated ability of key signaling hubs to differentially respond to multiple classes of ligands[37;75;33], which indicates higher than binary information transmission capacity.

Measurement of dynamics in multiple systems

The recent application of an optogentic tool to measure single cell dose responses has shown that high precision in cellular ERK response persists over multiple hours[101]. Therefore, it is unclear to what extent noise actually limits information transmission capacity of biochemical signaling networks. One possible resolution of these conflicting observations is related to the types of responses used in these observations. Thus far, the information-theoretic analyses of signaling networks have been based on scalar measurements performed at a single time point. However, the information on activating ligands is often encoded using a dynamic signal represented by a multivariate vector that contains a single cell's response at multiple time points[37;33;75]. To test the hypothesis that dynamic vector responses contain more information than static scalar responses, we performed dynamical single cell measurements of three key signaling pathways (Fig. 3.1): ERK activity (Fig. 3.1C) as reported by the EKARev FRET

biosensor[2] in MCF10A cells in response to Epidermal Growth Factor (EGF), Ca^{2+} levels (Fig. 3.1D) in MCF10A cells in response to ATP, and NF-κB activity (Figure 1E) as reported by the dynamic nuclear-cytoplasmic shuttling of EYFP-p65 in RAW 264.7 cells in response to LPS. Fully automated computational image analysis allowed us to measure the response of 910,121 live cells (see "Methods" section). The large sample size was instrumental for analyzing high-dimensional multivariate dynamic responses. In all three pathways there was substantial variability within the dynamic vector (Fig. 3.1C-E) and static scalar (Fig. 3.1F) single cell responses across multiple levels of activating ligands.

Dynamics recover information lost to extrinsic noise

To analyze the implications of noise on information loss we utilized an information-theoretic approach to calculate the information transmission capacity of a dynamic signaling network. The information transmission capacity (also referred to as channel capacity[21]) is measured as the maximal mutual information between the measured response and the activating ligand level. To calculate the mutual information between a multivariate dynamic response (a vector) and the activating ligand (a scalar), we developed a new estimation algorithm. The algorithm uses continuous multi-dimensional response data and a k-nearest-neighbor approach to estimate the conditional probability density for each cell's response.

In order to estimate information transfer in signal dynamics, we first considered the type of experimental data that we have acquired, which guided our general approach. In our signal transduction networks, the input signal S is defined by m discrete levels of extracellular ligand concentration ($S = [s_1, s_2, ..., s_m]$). For each input signal s_i we have n_i output protein trajectories ($R_i = [r_{i1}, r_{i2}, ..., r_{in_i}]$), with each trajectory occupying a single point in continuous Euclidean space of dimension d, where d is the number of time points in each output trajectory. Combined, we have $N = \sum_i^m n_i$ trajectories in our response R array. The general breakdown of the data is as follows:

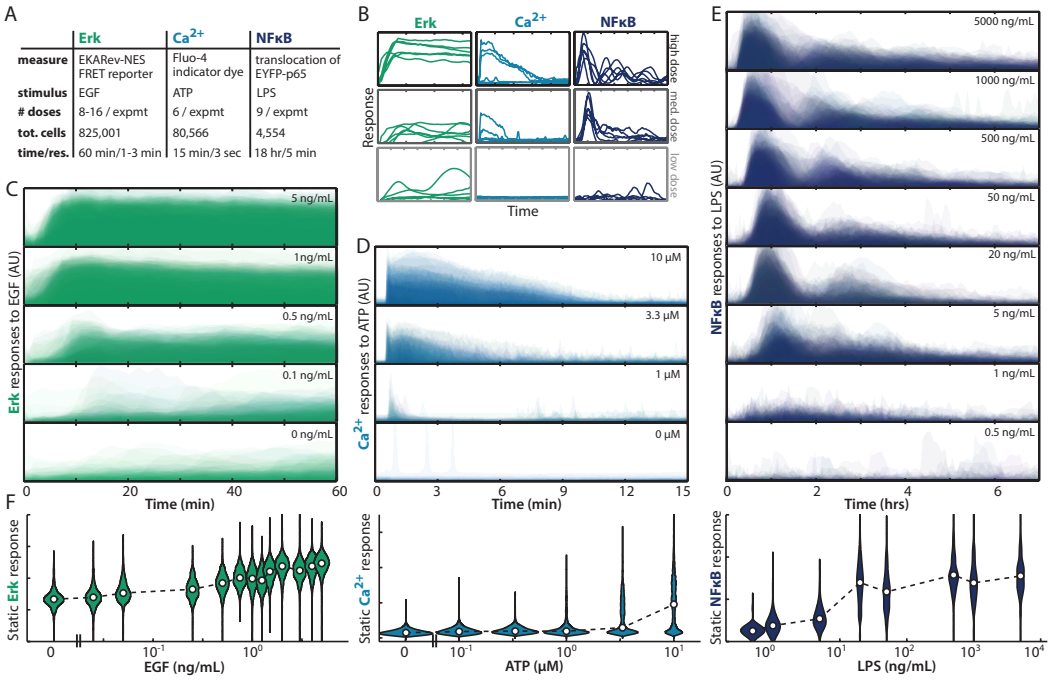

Figure 3.1: Single cell measurement of of ERK, Ca2+ and NF-κB dynamics. (A) Overview of single cell data analyzed in this work. **(B)** Examples of single cell dynamic response trajectories. **(C-E)**. Temporal histograms of several representative dosages for ERK **(C)** Ca^{2+} **(D)** and NF-κB **(E)**. Color intensity reflects the probability density of a cellular response magnitude at each time point. Y-axis in **(B-E)** is the same for each pathway and is of arbitrary units representing the FRET/CFP ratio reported by the EKARev ERK biosensor **(C)**, intensity of Ca^{2+} indicator dye Flou-4 **(D)**, and ratio of nuclear to cytoplasmic localization of an EYFP-p65 reporter **(E)**. **(F)**.Violin plot of the maximally separable static response in the three signaling pathways. Shape width shows response distribution (areas are equal), and point is the median response in each condition. All ERK and Ca^{2+} data was collected by Roy Wollman, Jason Yao, and Anna Pilko.

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_i \\ s_m \end{bmatrix}, R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_i \\ \vdots \\ R_m \end{bmatrix}, s_i \rightarrow R_i = \begin{bmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{ij} \\ \vdots \\ r_{in_i} \end{bmatrix}, r_{ij} = [r_{ij,1}, r_{ij,2}, \ldots, r_{ij,d}]$$

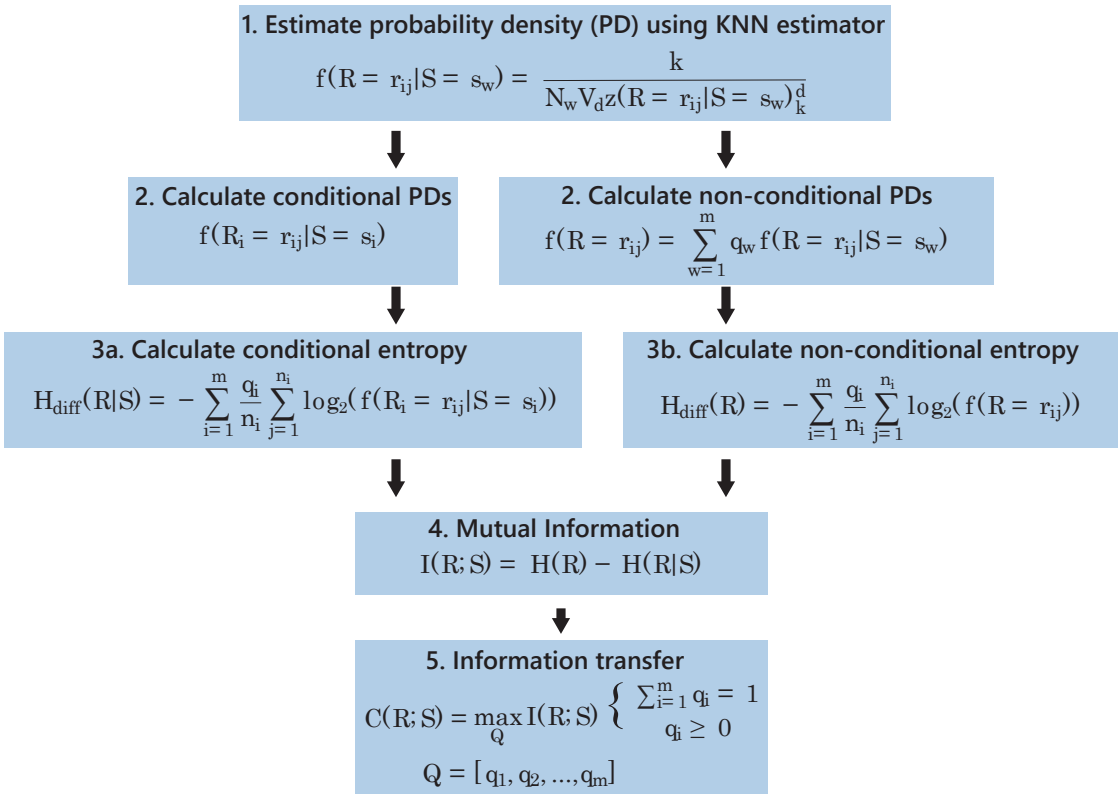

Figure 3.2: General schema for estimation of information transmission.

In order to estimate the information transfer (I) between an input (S) and an output (R) using well known formula

$$I(R;S) = H(R) - H(R|S). \tag{3.1}$$

we need to calculate Shannon entropies $H(R)$ and $H(R|S)$. The general scheme of our approach is shown in Fig. 3.2. First, given that our data is continuous, we need to define

how we will estimate these entropies. For a continuous probability density $f(x)$ of some observable X, the Shannon entropy is defined as differential entropy

$$H_{\text{diff}}(X) = -\int_{-\infty}^{\infty} f(x) log_2(f(x))dx. \tag{3.2}$$

Following change of variable of integration, Equation 3.2 becomes

$$H_{\text{diff}}(X) = -\int_{0}^{1} log_2(f(x))dy. \tag{3.3}$$

where $y = \int_{-\infty}^{x} f(t))dt$ is the cumulative probability density. We can estimate y by the cumulative probability distribution of N_x observations using

$$H_{\text{diff}}(X) = -\sum_{j=1}^{N_x} \delta_j log_2(f(x_j)), \tag{3.4}$$

where δ_j is the probability of observing x_j, $P(X = x_j)$.

Using Equation 3.4 as our basis, we will now illustrate how to obtain $H_{\text{diff}}(R|S)$ and $H_{\text{diff}}(R)$, given that our experimental data only contains conditional probabilities of a responses.

For the conditional case, $H_{\text{diff}}(R|S)$, since all n_i responses in R_i are equally likely, $\delta_j = \frac{1}{n_i}$, we can estimate probability density of a single response r_{ij} directly from all the other responses to $S = s_i$,

$$H_{\text{diff}}(R_i|S = s_i) = -\sum_{j=1}^{n_i} \frac{1}{n_i} log_2(f(R_i = r_{ij}|S = s_i)). \tag{3.5}$$

Here $f(R_i = r_{ij}|S = s_i)$ represents the probability density of response r_{ij} in R_i given all the other responses r_i (in R_i) to the signal $S = s_i$. We will explain how to estimate $f(R_i = r_{ij}|S = s_i)$ later in the derivation. With the probability of a given signal, $q_i = P(S = s_i)$, we can then sum the conditional entropies of each signal to get overall conditional entropy,

$$H_{\text{diff}}(R|S) = \sum_{i=1}^{m} q_i H_{\text{diff}}(R_i|S = s_i) = -\sum_{i=1}^{m} q_i \sum_{j=1}^{n_i} \frac{1}{n_i} log_2(f(R_i = r_{ij}|S = s_i)). \tag{3.6}$$

The case of estimating $H_{\text{diff}}(R)$ requires special attention, since we do not have access to non-conditional probabilities of responses. The difficulty arises from the fact that our estimate of non-conditional density of a single response, $f(R = r)$ is dependent on the probability of the input signals that generated all other responses. One approach is to estimate probability density that a given response r occurred in response to a given input signal (s_w), for each of the signals in S, by effectively placing that response into R_w and estimating the probability density for r is if it were also a response to s_w. Using total probability, for every response r in R, we can estimate the probability density within each set of responses R_w and sum over m such densities multiplied by the probability q_w of the signal that generated responses in R_w, as follows

$$f(R = r) = \sum_{w=1}^{m} q_w f(R = r|S = s_w). \tag{3.7}$$

Plugging 3.7 into 3.4, we get

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} \delta_{ij} log_2(f(R = r_{ij})). \tag{3.8}$$

The key difference between $f(R_i = r_{ij}|S = s_i)$ and $f(R = r_{ij})$, is that the former calculates the conditional probability density of a response among all other responses to the same signal, while the latter estimates non-conditional probability density of the response, combining law of total probability and conditional probability density of the response belonging to each of the subsets of responses R_i. To get δ_k we must account for the different probabilities associated with observing responses from different input signals and the number of responses n_i obtained for each input signal:

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(f(R = r_{ij})). \tag{3.9}$$

Now that we have formulas for $H_{\text{diff}}(R)$ and $H_{\text{diff}}(R|S)$, we need to estimate the probability densities $f(R = r|S = s_i)$ in 3.7 and $f(R_i = r_{ij}|S = s_i)$ in 3.5. This can be accomplished with the k-nearest neighbor (KNN) estimator,

$$f(x_j|X) = \frac{k}{N_x V_d z(x_j|X)_k^d} \tag{3.10}$$

where

$$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} \tag{3.11}$$

is the the volume of a unit sphere of dimension d (also dimension of x_j), N_x is the number of x_j in X, and $z(x_j)_k$ is the Euclidean distance to the kth nearest neighbor in X from x_j[56].

Applying this estimator to Equations 3.5 and 3.7, we get

$$H_{\text{diff}}(R|S) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(\frac{k}{n_i V_d z(r_{ij}|R_i)_k^d}). \tag{3.12}$$

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(\sum_{w=1}^{m} q_w \frac{k}{n_w V_d z(r_{ij}|R_w)_k^d}) \tag{3.13}$$

where $z(r_{ij}|R_i)_k$ is the distance from response r_j in R_i to the kth nearest neighbor in R_i, while $z(r_{ij}|R_w)_k$ is the distance from response r_{ij} in R_i to the kth nearest neighbor in R_w (Fig. **??**).

Without the knowledge of q_i, we are unable to estimate the information transfer $I(R;S)$ using Equations 3.13 and 3.12. However, the maximum information transfer (C) can be calculated with

$$C(R;S) = \max_{Q}\{I(R;S)\}, \tag{3.14}$$

where $Q = [q_1, q_2, ..., q_m]$, such that $\sum_{i=1}^{m} q_i = 1$ and $q_i \geq 0$[21]. This corresponds to the maximum possible informatin transfer between input S and output R.

Using our new algorithm we estimated the information transmission capacity of the dynamic response and of several types of static responses. For all single time point static scalar responses we found transmission capacity (<1 bit) similar to what was previously reported[19,106] (Fig. 3.4A). However, across all three signaling pathways, the dynamic response had significantly higher information transmission capacity than several possible scalar responses (Fig. 3.4B-C), p-value <0.05 for all comparisons).

The key advancement of our theory with respect to earlier work[102] is that it explicitly accounts for the differences between the information-degrading effects of in-

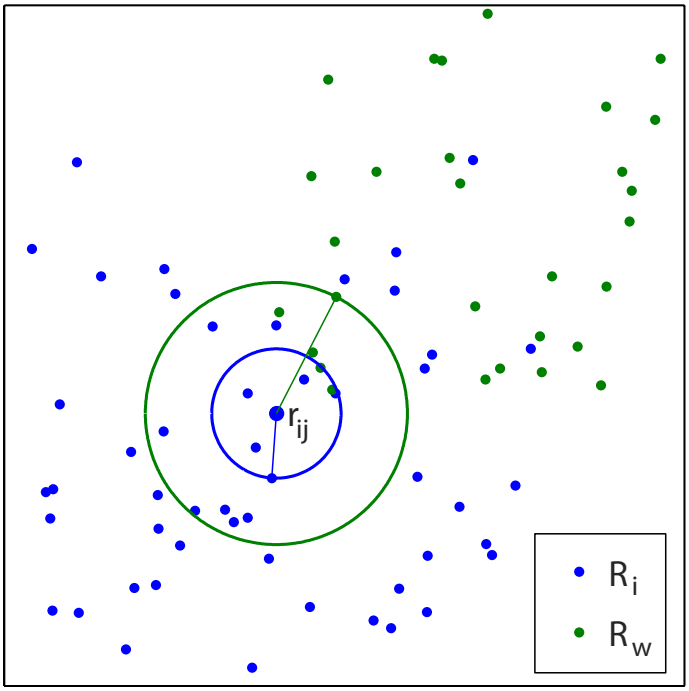

Figure 3.3: Using K-nearest neighbor radius to define probability distribution. Representation of k-nearest neighbor calculation for $k = 5$. The blue circle radius is the distance to the fifth closest neighbor within the same input response represented by blue points. The green circle radius is the distance to the fifth closest neighbor to a different input response (green points). Visualization created by Roy Wollman.

trinsic and extrinsic noise sources in the context of multivariate responses. Intrinsic noise adds to uncertainty in all dimensions (i.e. time points) independently from one another. In contrast, the extrinsic variability in cellular state produces fluctuations that are constrained by the signaling network that generates the dynamics. Therefore, the fluctuating components generated by extrinsic noise at different time points are deterministically dependent on one another. As a result, intrinsic and extrinsic noise sources have different effects on the information transmission capacity of multivariate responses. In the case of purely intrinsic noise, additional measurements increase the information logarithmically due to simple ensemble averaging (in accord with earlier findings in the Levchenko laboratory[19]). In the case of purely extrinsic noise, a., sufficient number of

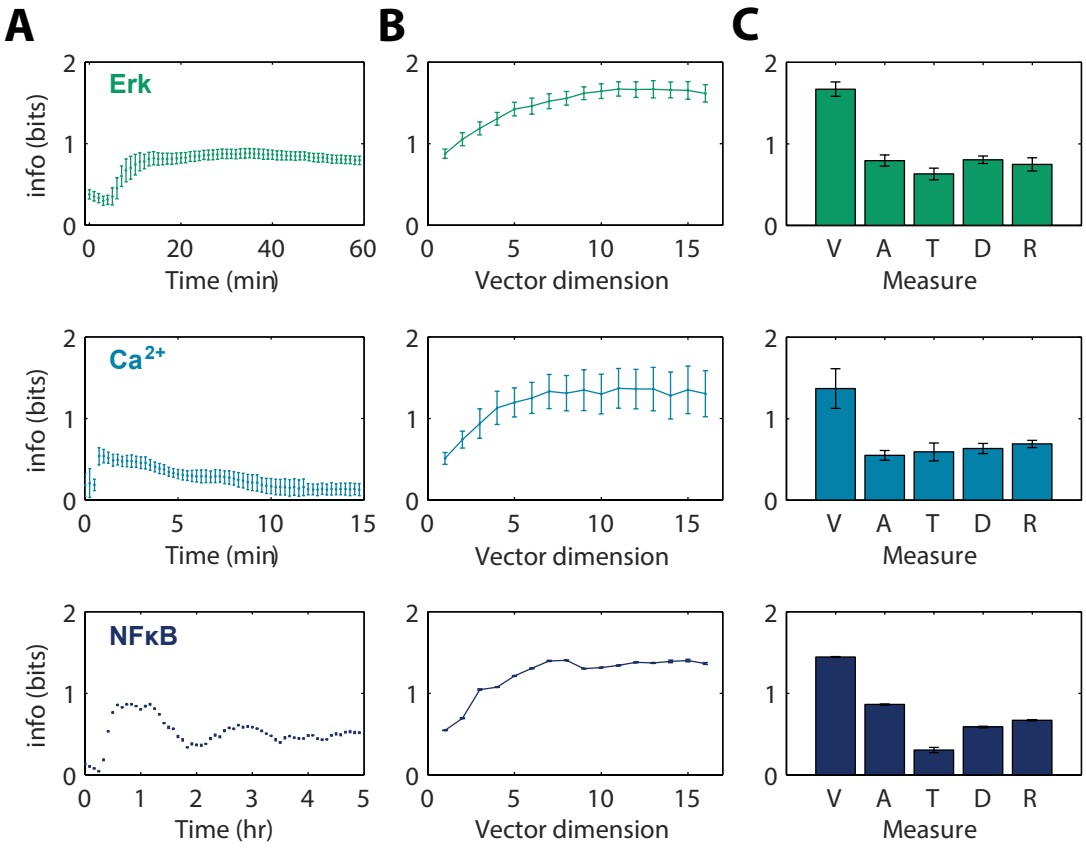

Figure 3.4: Information transmission capacity of static and dynamic responses. **(A)**Information transmission capacity calculated from static scalar response distribution based on single time point measurements.**(B)** Information transmission capacity calculated from multivariate dynamic responses as a function of the dimension of the multivariate vector. The multivariate vector was subsampled using a uniform grid centered on the middle time point.**(C)** Comparison of the multivariate vector (V) measurement to the following scalar responses: maximum response amplitude (A), maximum response time (T), maximal rate of response (D), ratio of maximum response amplitude to initial response amplitude (R). Error bars are SEMs from six biological replicates for ERK and four for Ca^{2+}, and SDs from five jackknife iterations for NF-κB . The multivariate vector information transfer was significantly greater than all scalar measures ($p < 0.05$, Student's t test). All analysis performed by Jangir Selimkhanov.

dynamical measurements can provide complete information about the a priori uncertain internal state of the cell and therefore lead to a substantial gain in the information about the activating ligand. Consequently, while an infinite number of measurements are required to completely eliminate information loss due to intrinsic noise, complete elimination of the information loss due to extrinsic noise only requires that the effective dimension of the dynamic response (or number of linearly independent response measurements) is higher than the number of independent uncertain parameters that determine the dynamics of the signaling network. In case when a signaling network is affected by both intrinsic and extrinsic noise, the multivariate dynamical response provides significant gain in the information transmission capacity by both ensemble averaging the intrinsic noise and effective elimination of the extrinsic noise. Figure 3A shows how an increase in response dimension allows cells to overcome the degrading effects of intrinsic and extrinsic noise sources.

Predicting channel capacity as a function of the signal-to-noise ratio

To test our analytical theory and demonstrate the ability of signaling dynamics to overcome extrinsic noise we used computer simulations of ERK responses based on a published kinetic model[95]. The mathematical model recapitulated the experimental data of ERK dynamics allowing us to simulate ERK responses resulting solely from extrinsic noise (i.e. ”cell-to-cell” variability of model parameters). Our theoretical analysis predicts that the multivariate dynamic response can completely eliminate the information loss that results from introduction of extrinsic noise. To test this prediction, we generated sets of simulated ERK trajectories in response to an increasing number of input levels, varying the model parameters for ERK and Mek according to a uniform distribution ($\pm$20% mean value). These simulation results were used to calculate the mutual information between ERK response and the level of the activating ligand. Our analysis

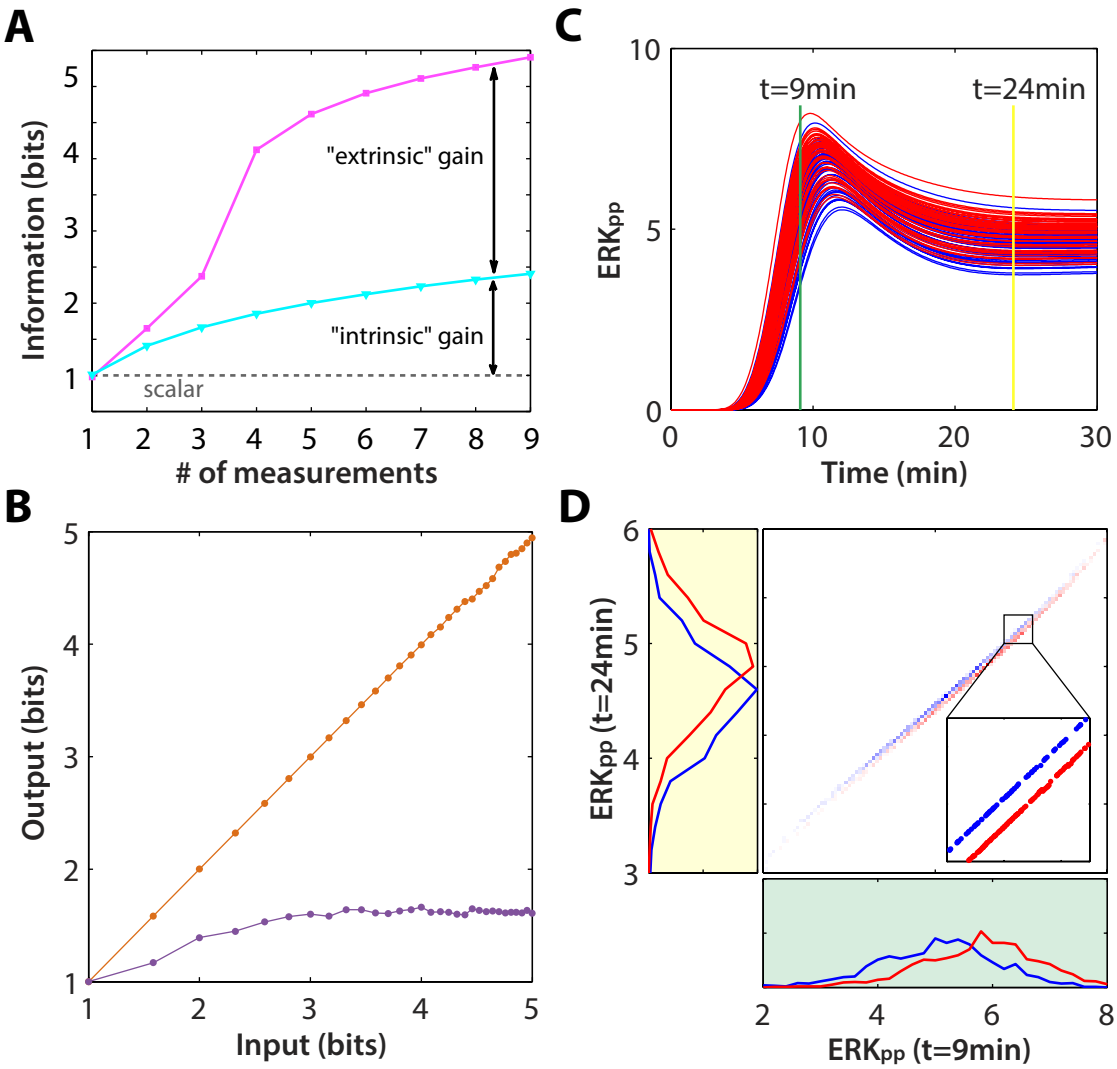

Figure 3.5: Theoretical decomposition of information loss. **(A)** Graphical representation of the analytical expression for the gain in mutual information from overcoming intrinsic (cyan) and extrinsic (magenta) noise sources obtained from random linear Gaussian inputs and outputs with three parameters[75]. **(B)** Information transmission capacity of dynamic (orange) and static (maximal response, purple) responses calculated using simulated trajectories from the computational model of ERK (22) with only the extrinsic noise contributing to cell response variability. **(C** Example of ERK trajectory variability for two different inputs levels (red and blue). Variability was generated using a uniform distribution of a single parameter, MEK values, that was varied by $\pm 20\%$. **(D)** . Two-dimensional histogram (center) and marginal distributions (left and bottom) for the two input levels (shown in red and blue) at two time points (t = 9 and 24 min) from the trajectories in **(C)**. Because only a single parameter was varied, the responses vary on a 1D curve. As a result, although the univariate marginal distributions show substantial response overlap, the 2-dimensional distribution shows completely separable response levels (inset). Model analysis and noise decomposition performed by Roy Wollman and Lev Tsimring.

(Fig. 3.5B) supports the analytical prediction and show that the dynamic multivariate response can indeed transmit complete information about the activating ligand. In contrast, the information transmission capacity of a univariate response, based on maximal ERK levels, was limited and therefore was unable to faithfully transmit complete information about the activating ligand. An intuitive demonstration for the limitation of univariate response and the ability of multivariate response to overcome extrinsic noise is shown in Fig. 3.5C-D. Superficially, the trajectories of two population of simulated ERK pp responses to two input levels of EGF appear overlapping (Fig. 3.5C), but in fact, they are completely separable when considering joint distributions in a procedure known as embedding[81]. Plotting the distribution of ERK-pp at t=9 and t=24 minutes on a two-dimensional plane (Fig. 3.5D) shows that the responses to a single varied parameter input lies on a one-dimensional manifold within a two-dimensional embedding space. The two one-dimensional manifolds for different inputs are completely separated from each other (inset), but overlap considerably in any one- dimensional projection. This simple example demonstrates how the extrinsic variability of a single parameter can in principle be completely eliminated using measurements from only two time points.

The accuracy of a response can be characterized by its signal-to-noise ratio (SNR). The mutual information and the system's SNR are related, however this relationship is strongly affected by the noise properties (intrinsic vs. extrinsic) and the type of the response. Our analytical theory describes mutual information as a function of SNR for three different types of responses: (i) scalar responses that do not distinguish between intrinsic and extrinsic noise, (ii) redundant vector responses (repeated measurements) that can only reduce intrinsic noise, and (iii) dynamic responses that combine the benefits of redundant measurements with very efficient mitigation of extrinsic variability. Among these three, the dynamic response yields the greatest mutual information for the same SNR.

To test our theoretical predictions of the dynamic response gain with respect to static and redundant, we collected dynamic ERK response data at different SNR levels. We varied the SNR in the ERK network by partial inhibition of the ERK kinase MEK-1

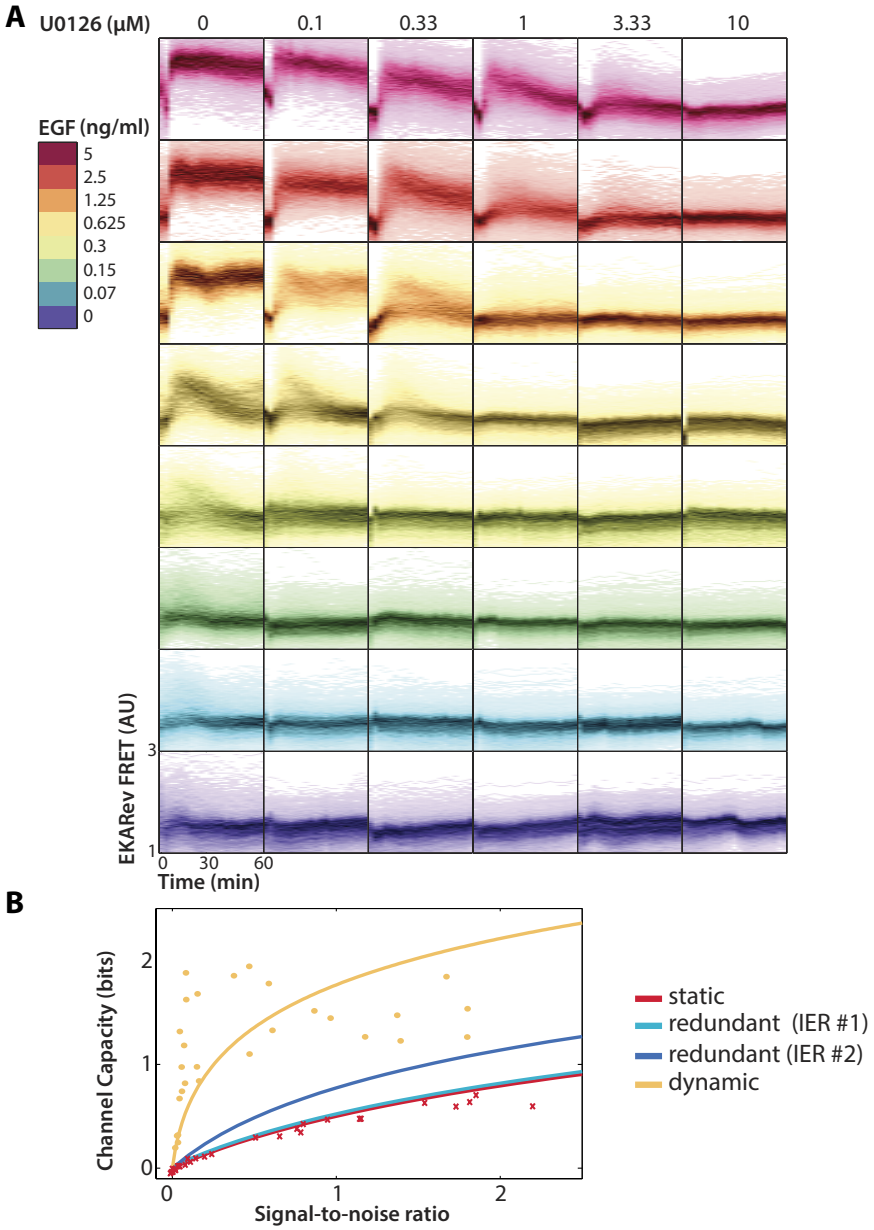

Figure 3.6: Measured information gain is a result of ERK dynamics' ability to mitigate extrinsic noise. **(A)** Single cell eight-well dose-response experiments. A total of 535,107 single cell responses are shown, from 48 experiments with six doses of MEK-1 inhibitor U0126 and 8 doses of the input ligand EGF. **(B)** Experimental measurement of the mutual information between ERK response and EGF measured as a function of the response signal-to-noise ratio (SNR). Each marker represents calculations of SNR and mutual information from the dynamic (dot) and maximal scalar (cross) responses of cells from **(A)**. Lines represent theoretical predictions of the mutual information as a function of SNR for three types of responses: static scalar (red line), redundant measurements where the multivariate response has no dynamics (dark and light blue lines) calculated based on two independent estimates of IER[75], and dynamic response (orange) that can mitigate both intrinsic and extrinsic noise. Data collected by Anna Pilko, and analysis performed by Roy Wollman.

using six different dosages of the inhibitor U0126 (Fig. 3.6A). At each MEK inhibition level we measured ERK response to eight EGF levels. A total of 48 conditions were measured in four biological replicates: at each MEK inhibitor level we calculated the mutual information and the SNR using single cell responses. In total, Fig. 3.6 contains 535,107 cell responses. As expected, for a scalar response, the formula relating the mutual information and the overall SNR, is in very good agreement with our experimental measurements. The theoretical prediction of mutual information for the redundant measurement case requires knowledge of the ratio of intrinsic to extrinsic noise (IER). IER was estimated in two ways by: (i) quantifying the fluctuations in the later (quasi-stationary) portion of the response time series of our ERK data (ii) using data and a fitted Hill function as used previously[101], where repeated measurements were performed on single cell responses. The predicted mutual information based on redundant responses using either IER estimates is significantly below the experimentally measured values. In contrast, the measured mutual information values are in good agreement with the theoretical prediction for a dynamic response based on a computational ERK model. Overall this analysis demonstrates that the substantial information gain from multivariate measurements is indeed the direct result of the dynamic nature of ERK response.

The robustness of biological systems is epitomized by their ability to function in the presence of large uncertainties[7]. A major source of uncertainty is the variability in cellular states, e.g., protein concentration within individual cells. We showed that signaling dynamics allow biochemical networks to mitigate this major source of extrinsic noise and thereby maximize the information transmission capacity of signaling networks. Our estimates of the information transmission capacities should be considered as lower bounds because part of what we refer to as intrinsic noise may in fact be the result of experimental imperfections. Therefore, the intrinsic-to- extrinsic ratio may be even lower than our estimates, which would further support our claim that dynamic response can substantially increase the information transmission capacity. Furthermore, information about the input ligand may be encoded in the dynamics of multiple signaling molecules. While the theory and observations presented here focus on the information

transmission capacity of the dynamics of a single signaling molecule, the extension of our analysis to the case of multiple signaling molecules responding to one ligand is straightforward. It is important to note that not all of the information contained in the multivariate dynamical responses may actually be used by cells. Yet, because reliable information transmission is a fundamental function of cellular signaling networks, it is plausible that evolutionary pressures shaped the cellular machinery to maximize the reliable decoding of multivariate dynamic signals.

Methods

Cell Culture

MCF10a cells (ATCC CRL-10317) were maintained in F12/DMEM media (Life Technologies) supplemented with serum, insulin and hydrocortisol according to [3]. RAW 267.4 cells (ATCC TIB-71) were maintained in maintained in a DMEM (CellGro 10-013) medium supplemented with 10% fetal bovine serum, 20mM HEPES, and 1x penicillin/streptomycin. For the analysis of the ERK pathway, MCF10a constitutively expressed the FRET sensor EKARev-NES [1]. This sensor has been optimized for dynamic range [6] and indeed we say up to 200% increase on the FRET/CFP ratio. Analysis of Ca2+ pathway was done by loading cells with 1 μM Flou-4 using PowerLoad (Life Technologies) loading solution according to manufacuer instructions.

Analysis of NF-κB was performed in RAW264.7 macrophages which were sequentially create a stable line with lentivirus vectors containing EYFP-RelA (driven by the endogenous 1.4 kilobase sequence directly upstream of RelA) and H2B-mCherry. Double-stable lines were made by successive selection, then further FACS-enriched. Cells were used in imaging from passage 16 to passage 20, then discarded. 20 hours prior to experiment start, cells were replated in Ibidi 8-well slide at a density of 50,000/sq. cm. 2 hours before the experiment, 1/3 of the total media volume was drawn off and mixed with stimulus, which was then injected into the chamber precisely at experiment

start.

Image aquisition/processing

Analysis of information transmission of dynamic vector responses can only be done from single live cell measurements. Furthermore, the analysis algorithm developed is this paper is based on the estimation of the probability density for each cellular response based on its degree of similarity to the response of cells with very similar responses. While the algorithm corrects for sample size using a jackknife procedure to eliminate possible bias, we have empirically found that robust estimation substantially benefits from large sample sizes. Therefore, a key technical aspect of this work is the ability to collect very large numbers of single cell responses. To that goal we have optimized both the image acquisition and analysis to increase the sample size without compromising on measurement accuracy. Below we outline the key steps that were taken to allow the collection of the signaling responses of 910,121 cells overall.

MCF10a image acquisition was done in a 96-well plate format (Coaster 3094) on a Nikon Ti microscope using a 10x 0.45NA Plan Apo objective. To increase the number of cell images the camera that was used (Zyla 5.5, Andor) has a large sensor size (21.8 mm). Furthermore, an additional de-magnifying 0.7x optovar was light path. The microscope was controlled through custom software written in MATLAB that uses the open-source micro-manager [4] as the driver layer, which controls microscope hardware. Image acquisition was completely automated and relayed on the Nikon Perfect Focus (PFS) to perform multi-well imaging while maintaining focus. The custom software performed periodic checks to verify that focus is maintained. To speed acquisition and minimize photo-damage and photo-bleaching, the acquisition rate of each channel was determined based on need: where the fluorescence channels used for measurement were acquired in every time point, the channels that were used for image segmentation were acquired at 8x-slower rate. To minimize background fluorescence, and since MCF10a cells were only imaged for a short period of time (<1 hour), cells were imaged in PBS

supplemented with HEPES buffer.

RAW 264.7 cells were automatically imaged every 5 minutes, using Zeiss's Zen interface, on a Zeiss AxioVert fitted with an incubation box/stage (which was both temperature- and CO_2-controlled), a 40x oil-immersion objective, LED fluorescence excitation, and a CoolSnap HQ2 camera. We collected DIC, mCherry, and YFP images over 18.5 hours at 12 stage positions per experimental condition. Cells were initially incubated for >2 hrs to allow temperature stabilization, then stage positions (X-, Y-, and Z-position) were set immediately before the initiation of aquisition. Automated analysis was performed as in chapter 2 in RAW 264.7 cells, and using similar techniques in MCF10a cells.

ERK Model Simulations

For numerical simulations, we adopted ODE model of ERK signaling network from Sturm *et al.*[95]. The model incorporates dynamics from RasGTP through Raf and Mek down to ERK phosphorylation. Following the previously-described model[95], we used the input concentration of RasGTP as a proxy for extracellular EGF, varying its value over several orders of magnitude. We found that the ERKpp dynamics generated by the model closely matched the dynamics of FRET signal recorded in experiments (Figure 3.7). Given these realistic trajectories, we used the model to test some of the results predicted by the earlier discussed theory.

Simulations with extrinsic noise

To illustrate the effect of using multi-dimensional measurements to eliminate the contribution of extrinsic noise to the information transfer of the system, we calculated MI using model simulation trajectories of ERKpp as the response and the input RasGTP. The range of input RasGTP (2500 to 22500) was chosen to minimize saturation at both ends of response. Except for ERK and MEK, model parameter values were kept consistent with Strum *et al.* for all simulations. To generate extrinsic noise we

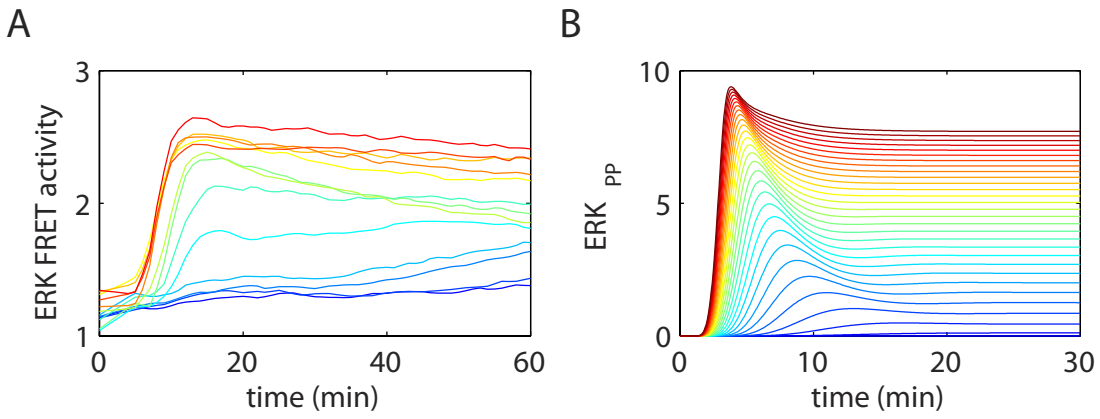

Figure 3.7: ERK model trajectories. Model simulation comparison to experimental ERK FRET trajectories. (A) Mean response of ERK FRET sensor to persistent EGF input. (B) ERKpp response trajectories from simulations of the ERK model[95] for increasing amounts of RasGTP. Modeling performed by Roy Wollman and Jangir Selimkhanov.

randomly sampled ERK and MEK values from uniform distributions ($\pm$20%) centered around the nominal values presented in Strum *et al.*. The model was allowed to reach a steady state with the chosen ERK and MEK parameters at the lowest value of Ras-GTP (2500), before applying inducing amount of RasGTP. The model was simulated for 30min. As the number of input levels of RasGTP within the input range increased, we found that the multi-dimensional measurement of ERKpp ($0:3:30min$) resulted in MI equivalent to the number of input levels (orange curve in Fig 3B). In stark contrast, the scalar measurement MI shows a saturation around 2 bits (purple curve in Fig 3B), which is consistent with theoretical prediction that at least 3 measurements are required to completely remove extrinsic noise.

To further demonstrate the underlying principles that allow for elimination of extrinsic noise, we plot 50 ERKpp trajectories generated from two nearby input values of RasGTP (Fig 3C). The overlap between these trajectories might suggest that it would be practically impossible to distinguish between the two input signal values. Furthermore, considering two different time points (vertical lines), this is exactly the case given the overlap between ERKpp response values at those time points (left and bottom 1-D

histograms in Fig 3D). If we consider those two points together, however, we can clearly see a separation between the two input levels as the 2-D histogram in Fig 3D indicates.

Fitting vector MI vs SNR data

To fit dynamic ERK response experimental inhibition data in Figure 3B (green filled circles), we applied theoretical description of mutual information and SNR (see Section IIC: Extrinsic and intrinsic noise) to our ERK model. We calculated sensitivities of ERK response at 10 equally spaced time points on the trajectory with respect to the signal (Ras_GTP) and 7 model protein parameters (Mek, ERK, Raf, Phase1, Phase2, Phase3, Phase4) near the middle induction level and nominal parameter values. Constructing Gram matrices with these sensitivities, we were then able to calculate MI as a function of SNR, derived earlier[83]:

$$I = \frac{1}{2}\log_2 \frac{|\hat{\mathbf{G}}_+ + \hat{\mathbf{I}}\sigma_{\tilde{\xi}}^2|}{|\hat{\mathbf{G}} + \hat{\mathbf{I}}\sigma_{\tilde{\xi}}^2|} \tag{3.15}$$

We assumed equal coefficient of variation (CV) for all parameters. Intrinsic noise was calculated based on experimental IER ratio. Varying the number of parameters contributing to the extrinsic noise in the model, we were able to fit the mean MI vs SNR curve to the experimental data (Fig. 3.8. For a given number of parameters, we generated the mean MI vs SNR curve for all combinations of parameters. The best fit was obtained with two parameters accounting for the extrinsic noise in the model. This could be thought of as the number of effective system parameters contributing to extrinsic noise that our dynamic measurements can overcome.

Sampling dimension for vector response

To select the time points that should be part of the vector of dynamic responses for increasing dimension of the vector in Figure 4, we used a simple strategy. For a given time frame of the trajectory (60min for ERK, 15min for Ca^{2+}, 5hrs for NF-κB) and chosen dimension $\mathbf{N}$, we selected $\mathbf{N}+\mathbf{2}$ equally spaced response values throughout

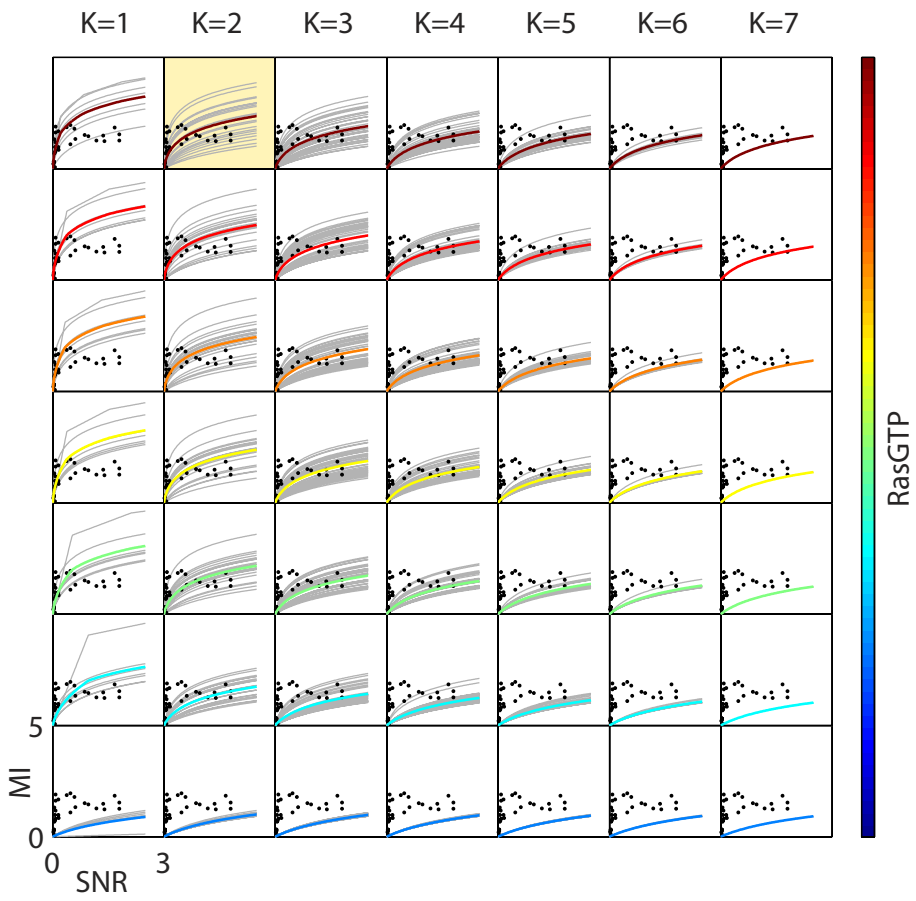

Figure 3.8: Fitting ERK model to dynamic information/SNR. Fitting ERK model to dynamic MI vs SNR data using analytic theory approach. We used K number of parameters and RasGTP input level as the two fitting parameters for the model. Grey MI vs SNR curves were obtained for are all combinations of K parameters (out of 7). Colored curves (input RasGTP level) correspond to the mean of all of those combinations of K parameters. The black dots represent experimentally obtained values for vector MI shown in Figure 4, that the model (colored curves) were fit to. Yellow highlighted box corresponds to the best fit for K and RasGTP level. Fitting performed by Roy Wollman and Jangir Selimkhanov.

the given time frame, removing the first and the last values. For example, for vector of dimension 1, we chose response value at the center of the given time frame, while for

dimension 2, we chose values located at the one third and two thirds points of the time frame (Fig. 3.9).

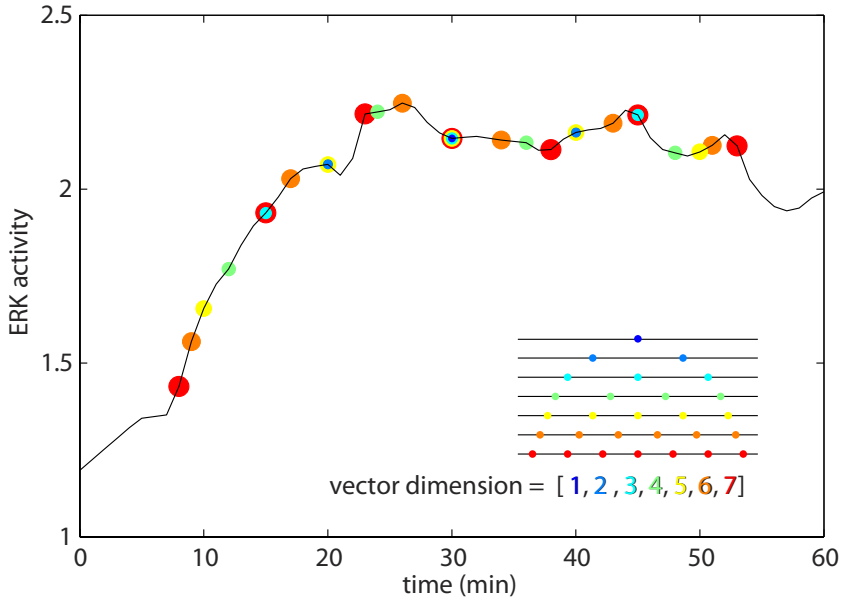

Figure 3.9: Dimension sampling in dynamics. Dimension sampling approach for vector dimension in Fig. 3.4B.

Experimental noise analysis

Signal-to-Noise Ratio (SNR)

To calculate ERK signal-to-noise ratio (SNR), we defined the signal magnitude σ_r^2 as the variance of average responses over all m input levels of EGF:

$$\sigma_r^2 = \frac{1}{m}\sum_{i=1}^{m}((\frac{1}{m}\sum_{w=1}^{m}\frac{1}{n_w}\sum_{j=1}^{n_w}r_{wj}) - \frac{1}{n_i}\sum_{j=1}^{n_i}r_{ij}))^2 \tag{3.16}$$

Noise magnitude was defined as the average of the variances of n_i responses to a single input level of EGF:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m}(\frac{1}{n_i}\sum_{j=1}^{n_i}(\frac{1}{n_i}\sum_{w=1}^{n_i} r_{iw} - r_{ij})^2 \tag{3.17}$$

SNR is then σ_r^2/σ_n^2.

Autocorrelation of ERK response

We performed autocorrelation analysis on ERK trajectories to gain insight into our sampling time sampling of the data. According to the analysis, decay of autocorrelation function shows that on average self-correlation is lost after 11min (Fig. 3.10). This suggests that ERK measurements taken every 10 min are, on average, independent from one another.

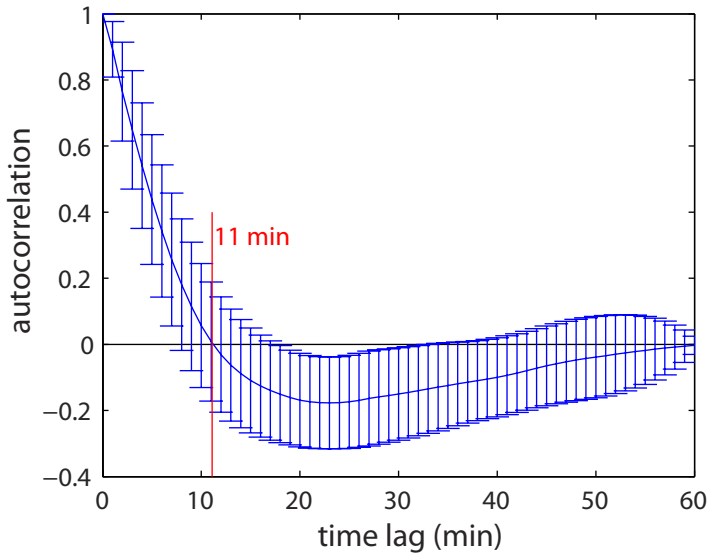

Figure 3.10: Autocorrelation in the ERK response. Mean autocorrelation of ERK response trajectories (bars represent standard deviation). Decay of autocorrelation function shows that on average self-correlation is lost after 11min. Analysis performed by Roy Wollman and Jangir Selimkhanov.

Intrinsic-to-Extrinsic Ratio (IER)

To calculate ERK intrinsic-to-extrinsic noise ratio (IER) from our data, we defined intrinsic noise as combination of stochasticity inherent to biochemical reactions in signal transduction and measurement noise, while extrinsic noise was defined by the variability in individual cell states. To estimate the upper bound on the experimental IER, we used the fact that our sampling of ERK response was faster than ERK dynamics (based on the autocorrelation of ERK response), to calculate intrinsic noise. Using second portion of ERK trajectory, where ERK levels did not change significantly between successive time point measurements, we estimated the intrinsic noise (σ^2_ξ) as the variance of the differences in ERK expression between successive time points (Fig. 3.11 red). To get the extrinsic noise (σ^2_e), we estimated the total noise according to equation 3.17 (Fig. 3.11 cyan) for the second portion of ERK trajectory, and simply subtracted the intrinsic noise from the total noise. The mean IER ($\sigma^2_\xi / \sigma^2_e$) for all experimental conditions was estimated to be **0.024**.

While IER estimate from our data best matches our theoretical analysis, it is limitted to our assumption about ERK dynamics. For completeness, we decided to use previously obtained data[101] to estimated an upper bound on IER. We chose ERK data from Toettcher *et al.*[101], who were able to measure multiple ERK responses within a single cell. To estimate IER from their data, we used a slightly different approach. Using Hill function fit (exponent of 2, previously shown[101]) for each cell, we estimated intrinsic noise as the variance of the differences between experimental values and the model fit (Fig. 3.11 red). Similarly, to get the total noise, we fit the same function to all of the experimental points and calculated the variance of the differences between experimental values and the model fit (Fig. 3.11 cyan). To get extrinsic noise, we simply subtracted the intrinsic noise from the total noise. The mean ration of intrinsic to extrinsic noise was estimated to be **1.14**.

Chapter 3 contains material originally published as "Accurate information transmission through dynamic biochemical signaling networks". Selimkhanov, J.*, Taylor,

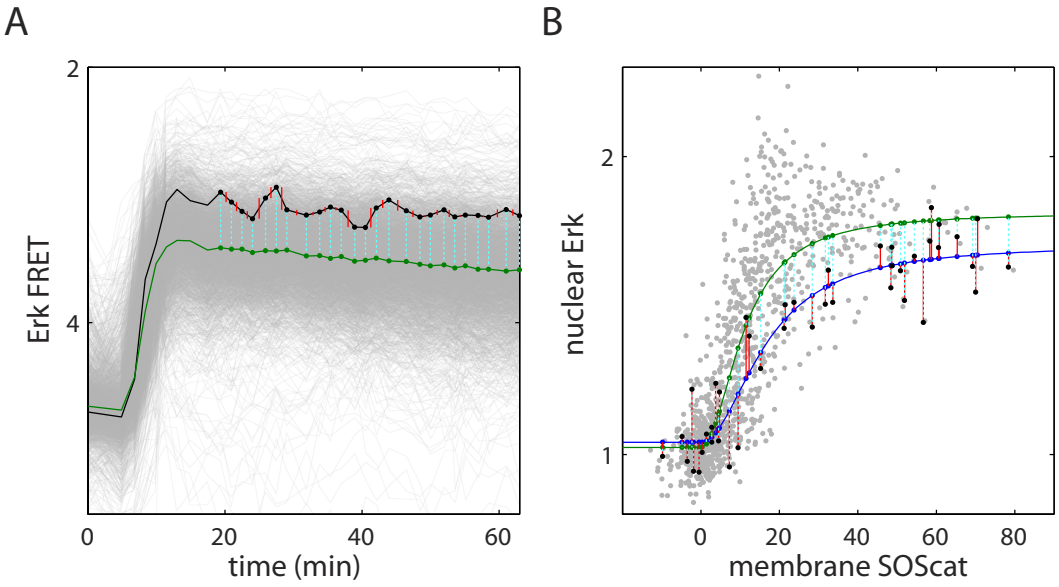

Figure 3.11: Estimating extrinsic and intrinsic noise in ERK data. **(A)** Using our single cell ERK data, intrinsic noise was estimated by the mean of the mean of squared errors between successive ERK trajectory points (red). Total noise was estimated by the mean of squared errors (cyan) between single ERK trajectory and average of all trajectories (green). Extrinsic ratio was obtained from the difference between total noise and intrinsic noise. The mean ratio of intrinsic to extrinsic noise was estimated to be **0.024**. **(B)** Using data derived from literatucaptionre[101], we fit a Hill function to the data and calculated the mean squared error between the fit for each cell (intrinsic noise) and between the fit for all points and each cell (total noise). The IER was estimated to be **1.14**. Analysis performed by Roy Wollman.

B.*, Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. ***Science*** (2014). Copyright permission to republish here was granted by AAAS.

Chapter 4

Distinct single cell signaling characteristics conferred by the MyD88 and TRIF Pathways in TLR4 activation

Introduction

Toll-like receptor signaling involves a complex network of at least 12 different TLRs engaging in physical and functional interactions with a wide variety of signal transduction proteins[68]. However, all information is transduced via two signaling pathways: the myeloid differentiation primary response gene 88 (MyD88)-dependent, and/or the TIR domain-containing adaptor protein-inducing interferon-Îš (TRIF)-dependent pathways, forming a bow-tie network structure[68]. Among all the TLR family members, TLR4 is the first described in mammals[13] and the only one that employs both pathways to trigger inflammation and innate immune responses[47].

TLR4 recognizes bacterial lipopolysaccharide (LPS), a component of the outer membrane of Gram-negative bacteria, and activates the pleiotropic transcription factors

NF-κB and interferon response factor 3 (IRF3). Upon binding to LPS and dimerizing, TLR4 recruits MyD88 at the plasma membrane, triggering the initial activation of the NF-κB -controlling kinase IKK[47]. TLR4 also undergoes dynamin-dependent endocytosis and is trafficked to the early endosome[42], where it interacts with TRIF and its adapter molecule (TRAM), initiating the TRIF-dependent pathway that leads to IRF3 activation and a second wave of IKK activity[47].

NF-κB dynamics are stimulus-specific[22] and are thought to represent a signaling code that determines downstream cellular responses[10]. Interestingly, while NF-κB activity is oscillatory in response to tumor necrosis factor (TNF), it was reported to be steady in fibroblasts responding to LPS due to autocrine TNF feedback[22;111;53]. However, in macrophages TNF does not contribute substantially to LPS signaling[16], and studies of the population average[46;48] suggested that MyD88 may be responsible for an early NF-κB peak[46] and TRIF for a later phase[117]. But how NF-κB dynamics at the single cell level are encoded in macrophages responding to TLR ligands remains an open and important question.

There remains a fundamental disconnect between robust signaling and gene expression behaviors observed at a population level, and the variability that characterizes individual cell responses[4;92]. This variability limits the capacity for reliable biochemical information transduction[19], a characteristic required for mounting appropriate physiological responses to diverse external signals. We require, then, a better understanding of the origins, control, and consequences of molecular network noise that drives response variability. Preliminary modeling efforts account for experimentally measured distribution of binary cell fates[58], or have identified potential sources of noise relevant to NF-κB signaling particularly in the expression of NF-κB feedback genes[73 44]. However, this variability has not been contextualized with noise sources in receptor-proximal signaling modules, and signaling model simulations do not generally account or match the measured variability in single cell signaling activities[9].

Molecular mechanisms not only determine the dynamics of signaling, but also harbor potential noise sources that determine cell-cell variability. In the case of the

TLR4 network, TRIF signaling occurs from the endosome, dependent to unknown degrees on endosomal trafficking[39], while MyD88 signaling is initiated from the plasma membrane: these branches are both temporally and spatially separated[42]. MyD88 additionally contains a death domain (DD) which mediates homotypic interactions to form a macromolecular complex (the "Myddosome)[55]. These mechanisms, though identified, have not yet been integrated into a full reaction network to understand their role in TLR4 signaling.

Here we report the iterative development of a mathematical model of the TLR4-to-NF-κB signaling network in macrophages in the context of quantitative biochemical and live cell imaging experimental studies. We reveal how specific molecular mechanisms within the MyD88 and TRIF pathways control dynamical features of the NF-κB response, as well as associated cell-to-cell variability, which together determine the information transduction capacity of pathogen-responsive signaling.

Roles of MyD88 and TRIF in a model of TLR4 signaling

Within the TLR4 signaling network four modules may be distinguished (Fig. 4.1A): a TLR4 module, which transduces the presence of LPS into downstream kinases activities (IKKK and TBK1) via MyD88 and TRIF pathways; an IRF module, which transduces TBK1 activity to output levels of phosphorylated nuclear IRF3; an IKK module, which generates IKK activity profiles from activated TRIF and MyD88 inputs; and an NF-κB module, which determines nuclear NF-κB activity as a function of input IKK and IκB. We have previously established mathematical models for the IKK and NF-κB modules[49;111], and focused here on establishing the topology, parameters, and behavior of the TLR4-proximal signaling module in macrophages, the primary pathogen-sensors and effectors of the innate immune response.

We first measured IKK activation dynamics in response to 1ng/mL and 100ng/mL LPS in populations of bone-marrow derived macrophages (BMDMs) derived from wild-type (wt), *Myd88*$^{-/-}$, and *Trif*$^{-/-}$ mouse strains. In *Trif*$^{-/-}$ BMDMs, IKK activ-

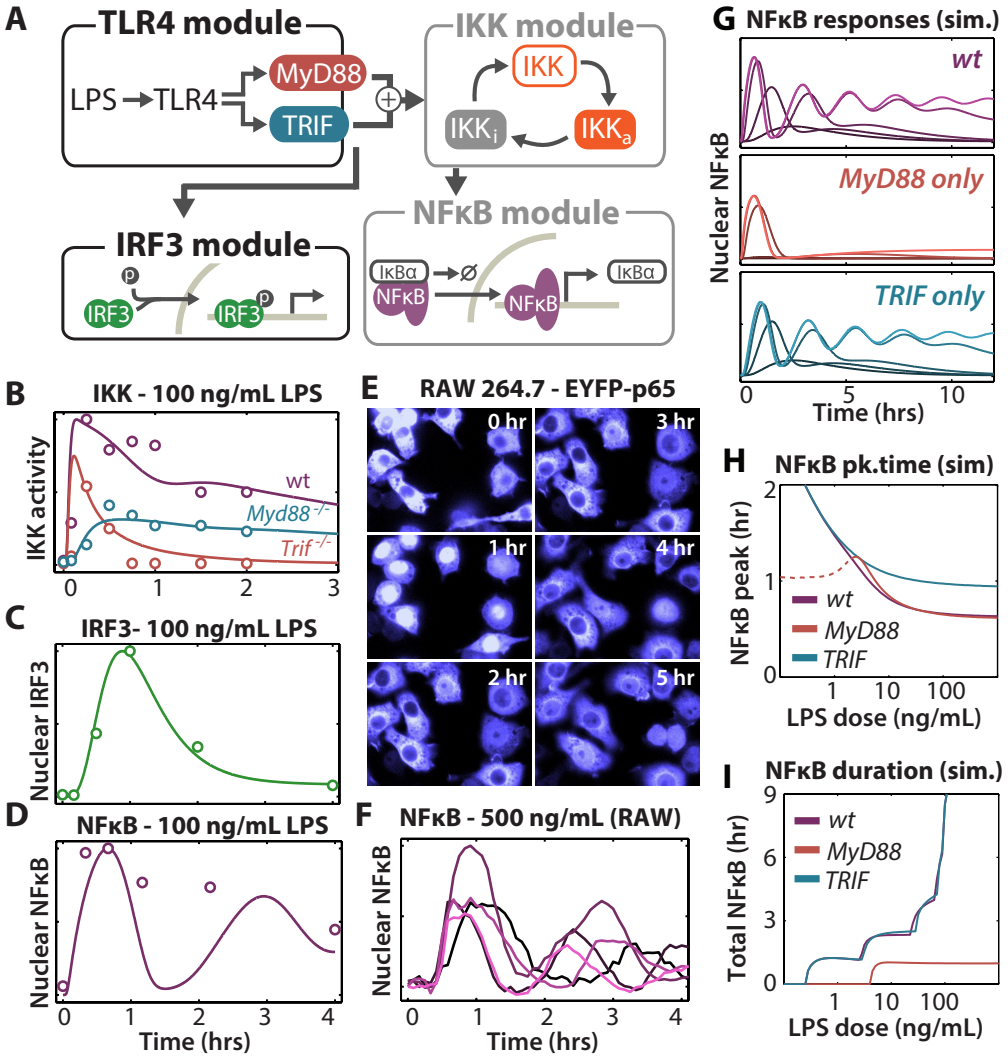

Figure 4.1: Computationally modeling MyD88- and TRIF-dependent activation of IKK/NF-κB in TLR4 signaling. **(A)** The TLR4 signaling network depicted as four regulatory modules. **(B)** Time course of active IKK in response to LPS (100ng/mL) in wt (purple), $Trif^{-/-}$ (red), and $Myd88^{-/-}$ (blue) BMDMs. Line: model simulation. Circles: IKK kinase assay produced with co-immunoprecipitates of NEMO. **(C)** Phosphorylated nuclear IRF3 time course in response to LPS (100 ng/mL) in wt BMDMs. Line: model simulation. Circles: nuclear immunoblots for phosphorylated IRF3. **(D)** Time course of active nuclear NF-κB in response to LPS (100 ng/mL) in BMDMs. Line: model simulation. Circles: EMSA using NF-κB -specific double-stranded oligonucleotide probe. **(E)** Time-lapse images of live RAW 264.1 macrophages (expressing EYFP-p65) responding to LPS (500ng/ml). **(F)** NF-κB responses to 500ng/mL LPS in five randomly selected single cells. **(G)** Simulated NF-κB dynamics from 0.1ng/mL to 1μg/ml (line color from dark to bright), in wt or single-pathway (MyD88 only, TRIF only) conditions. **(H)** Model-predicted nuclear NF-κB 1st peak times, by LPS dose, in wild-type or single-pathway conditions. Dashed lines indicate weak peaks (<1% of maximum peak observed). **(I)** Model-predicted durations of total nuclear NF-κB activity ([NF-κB]$_{nuc}$>0.03 μM), by LPS dose. Initial model asssmbly and parameterization by Zhang Cheng.

ity was induced early and transiently, reaching a peak at least 15 minutes earlier than in *Myd88*$^{-/-}$ cells (open circles in Fig. 4.1B, Fig. 4.8). In contrast, later-activating *Myd88*$^{-/-}$ BMDMs show more persistent signaling, well past the hour-long window of activation in *Trif*$^{-/-}$ cells. Wild-type IKK dynamics show both early activation and long duration characteristic of summing MyD88 and TRIF activity. We similarly quantified the dynamics of downstream transcription factors NF-κB (measured by EMSA) and IRF3 (measured by phosphorylated western blot) (open circles in Fig. 4.1C-D, Fig. 4.8). Using these data to parameterize simple models of the TLR4 and IRF3 modules (Fig. 4.12, tables 4-4, "Methods" section), we were unable to obtain a good fits to EMSA-measured NF-κB activity profiles, although IKK dynamics and IRF3 dynamics were recapitulated for all genotypes and LPS concentrations (Fig. 4.1 B-D, Fig. 4.8). (This model did not include autocrine cytokine mechanisms as these play little role in macrophages responding to LPS[16]). The key discrepancy lay in the model's propensity to produce oscillations as a result of the delayed negative feedback loop mediated by IκBα (Fig. 4.1D). The fact that the population measurements did not reveal these suggested the possibility that single cell dynamics were obscured by a high degree of cell-to-cell variability. To resolve this discrepancy, we performed live-cell imaging (Fig. 4.1E) in a RAW264.7 macrophage cell line stably transduced with a lentivirus containing EYFP-p65 (driven by the *Rela* promoter). Following stimulation with LPS concentrations ranging from 500 pg/mL to 5 μg/ml, we tracked single cell NF-κB translocation. Responses are first evident at 1ng/mL, and increasing doses revealed oscillatory translocation patterns (Fig. 4.1F), in agreement with model simulations. These results suggest that that oscillatory dynamics in NF-κB signaling are not stimulus-specific, but are instead intrinsic in the NF-κB -IκB signaling module.

Using the model, we explored the pathway-specific roles in encoding NF-κB dynamics to a range of LPS concentrations (Fig. 4.1G, dark to bright lines). As expected, NF-κB activated solely by MyD88 showed no second-phase activity. While late dynamics remained intact in the TRIF-only condition, they showed a slowed, reduced first phase. First-peak time in wt or MyD88-only cells, then, was 20 min. faster than

in TRIF-only cells (Fig. 4.1H). In contrast, the control of NF-κB response duration was entirely TRIF-dependent, with MyD88 producing only transient responses (Fig. 4.1I). In sum, our analysis illustrated that the two TLR4-responsive pathways encode specific aspects of NF-κB dynamics: early, transient activation of IKK by MyD88 dictates the initial timing of the response, while slower, persistent activation of IKK by TRIF encodes a longer duration of NF-κB activity. This analysis provides a framework for studies of how underlying molecular mechanisms in MyD88 and TRIF activation determine NF-κB dose-responsiveness, dynamics, and cell-to-cell variability.

Signalosome formation determines MyD88 signaling

MyD88 signaling is thought to involve the formation of a signaling complex, the Myddosome, which is formed by 6 subunits of MyD88, 4 of IRAK4 and 4 of IRAK1[55 27], prompting us to examine its implicit signaling characteristics. We modeled Myddosome formation in accordance to a sequential binding proposal[55] in which the LPS-TLR4 complex (C) functions as a seed that may attract two MyD88 monomers (M), forming CM_2, which in turn functions as a building block to form C_2M_4 and C_3M_6. The C_3M_6 complex constitutes a molecular platform that forms the macromolecule Myddosome by incorporating IRAK4 and IRAK1 (Fig. 4.2A and supplemental text). As previously hypothesized[27], we found that the process was inherently cooperative, leading us to approximate the reaction scheme by Hill kinetics (Fig. 4.10). Assuming MyD88 subunit interaction ratios ($k_f/_b$ from Fig. 4.2A) in a range of 0.1-10 resulted in a range of fitted Hill coefficients between 1.8-3.1 (Fig. 4.2B). We then employed our TLR4 model to study the ramifications of the range of predicted Hill coefficients (Fig. 4.2C and D). While the *Myd88*$^{-/-}$ model predicts a slowly saturating peak response to increasing LPS, peak NF-κB in the *Trif*$^{-/-}$ condition shows a faster switch from off-to-on, depending on the strength of the MyD88 subunit interaction ratio. We measured the first-peak amplitude in the single cell RAW 264.7 response (Fig. 4.2E and F) and observed an increase from minimal to nearly-saturated peak NF-κB activation

over a range of 1-10 ng/mL LPS. These values were used to tune the effective EC_{50} and Hill coefficient of the model (supplemental text), providing a best-fit Hill coefficient of 3 (corresponding to $k_f/_b = 6.1$, red line in Fig. 4.2B-D).

A second prediction of cooperative signalosome formation regards signal duration. Given a set of dose-dependent transient inputs (i.e. upstream TLR activation at the plasma membrane, quickly turned off by receptor endocytosis), higher levels of cooperativity cause robust switch-off behavior, saturating not only the amplitude, but also the total amount of the signaling output (Fig. 4.2G). In a highly cooperative system (Hill coefficient n = 3 in a *Trif*$^{-/-}$ model, Fig. 4.2G, bottom left), outputs above the activation threshold are nearly indistinguishable from one another, resulting a saturated dose-response curve for integrated IKK activity (Fig. 4.2G, bottom right). Integrating measured IKK activity in *Myd88*$^{-/-}$ and *Trif*$^{-/-}$ BMDMs (Fig. 4.2H, Fig. 4.8) shows that total MyD88-driven IKK activity saturates more quickly than TRIF-mediated IKK activity, supporting this prediction.

Outside of a narrow dose range, MyD88 induces robust on-or-off responses. This behavior was predicted in a cooperative model, and measured by peak NF-κB activity. Once the on-state is reached, MyD88-mediated signaling is quickly saturated. This dynamic behavior is partially explained by cooperative Myddosome formation, but additionally requires a robust means of downregulation. In the case of TLR4 signaling, rapid TLR4 endosomal internalization, which limits the duration of MyD88 signaling, might function as such a downregulation mechanism. We thus turned to the processes of endosome trafficking, to study how these determine the balance between the MyD88 and TRIF pathways.

Endosome translocation and maturation in signaling

Ligand-bound TLR4 first induces MyD88 signaling from the plasma membrane, and then is endocytosed to induce TRIF signaling from the early endosome[42]. However, unbound TLR4 also traffics along with recycled plasma membrane to endosomes as part

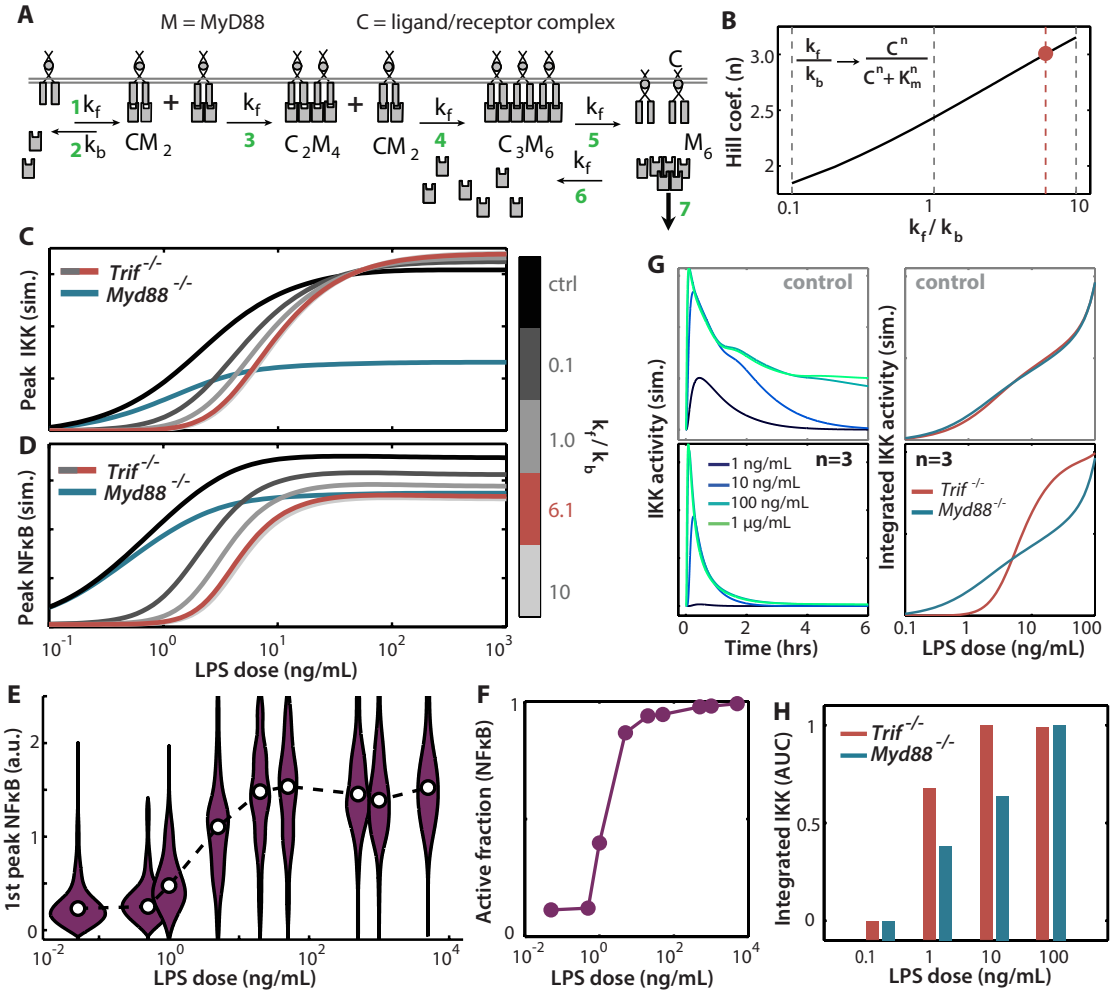

Figure 4.2: Signalosome formation shapes the IKK and NF-κB dose response. (A) The 7-step Myddosome assembly model **(B)** Time course of active IKK in response to LPS (100ng/mL) in wt (purple), *Trif*$^{-/-}$ (red), and *Myd88*$^{-/-}$ (blue) BMDMs. Line: model simulation. Circles: IKK kinase assay produced with co-immunoprecipitates of NEMO. **(C-D)** Dose response of peak IKK **(C)** and nuclear NF-κB **(D)** activity predicted by the full TLR4 model for indicated Hill coefficients (from Myddosome assembly model, and corresponding to different $k_f/_b$ ratios) in *Trif*$^{-/-}$ cells (grey gradient lines) and *Myd88*$^{-/-}$ (purple line). Control (ctrl) indicates the case with Hill coefficient =1. **(E)** Violin plot of first-peak NF-κB amplitudes vs. LPS dose in single cells. Shapes show total response distribution (normalized to the same area), and points are the median response. **(F)** Fraction of cells showing NF-κB activation above a given peak amplitude threshold (>0.55a.u.) as a function of LPS dose. **(G)** Simulated total IKK response to different doses of LPS with Hill coefficient (n) = 1 (no cooperativity, top two panels) or 3 (bottom two panels). Left: time courses in *Trif*$^{-/-}$ model. Right: integrated IKK responses vs. LPS dose in *Myd88*$^{-/-}$ and *Trif*$^{-/-}$ models. **(H)** Integrated IKK response in *Myd88*$^{-/-}$ and *Trif*$^{-/-}$ BMDMs from quantified kinase assay. Responses in both G and H were normalized to total observed range. Hill fitting and modeling performed by Zhang Cheng.

of normal constitutive processes[39], where it could encounter free LPS or endocytosed CD14-LPS complexes in the endosome. We sought to measure the relative contributions of these constitutive and ligand-induced TLR4 endocytosis to NF-κB signaling. We decomposed endosomal signaling via flux analysis, calculating the receptor-ligand complex transfer from the plasma membrane to the early endosome (activation via ligand-induced shuttling), as well as the amount of endosomal LPS-TLR4 binding (activation via constitutive shuttling). Overall, ligand-induced shuttling proved to be more important at all doses of LPS (Fig. 4.3A and B). Indeed, outside of very high doses and early timepoints (Fig. 4.3B), constitutive shuttling generally provides negative flux, partially balancing ligand-induced traffic. These conclusions appear to be robust to a high degree of parameter perturbation (Fig. 4.13,"Methods").

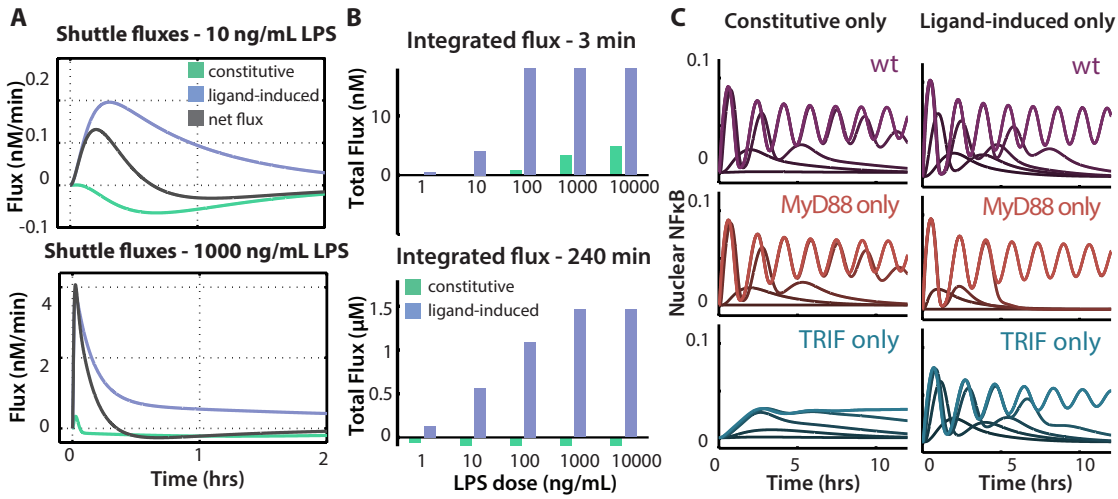

Figure 4.3: TLR4 endosomal transport drives transition from MyD88 to TRIF. (A) Predicted mechanism-specific fluxes of the endosomal ligand-TLR4 complex with 10ng/ml and 1μg/ml LPS stimulation. "Constitutive" flux is the sum of LPS-TLR4 binding and unbinding in the early endosome. "Ligand-induced" flux is the sum of ligand-TLR4 complex endocytosis and exocytosis. The net flux is the sum of constitutive and ligand-induced flux, as well as the degradation of ligand-TLR4 complex in the early endosome. **(B)** Model-predicted integrated flux over short (0-3 min.) and long (0-4hr.) term for LPS doses from 1ng/mL to 10μg/ml. **(C)** Simulated nuclear NF-κB responses to LPS in in silico knockout conditions, which are either able to internalize TLR4 via constitutive means only (left) or via ligand-induced shuttling only (right). LPS stimuli range from 1ng/ml to 1μg/mL (dark to bright lines). Flux analysis performed by Zhang Cheng.

We additionally examined in silico whether these flux-based conclusions would be reflected in the relative requirement of each trafficking mechanism, an isolation that would be difficult to attain experimentally. However, when either mechanism was removed from the model, total TLR-mediated NF-κB activation was actually enhanced (Fig. 4.3C, cf. Fig. 4.1F). By decreasing transport away from the plasma membrane, TRIF signaling is indeed lowered (and nearly eliminated in the constitutive shuttle-only model), but MyD88 signaling is greatly enhanced in both cases, as more TLR4 persists on the plasma membrane, well above the Myddosome's threshold of activation. Thus while ligand-induced shuttling proves essential in TRIF activation, both shuttling modalities are critical for enforcing the transience of MyD88 activation. While endosomal shuttling controls simultaneous downregulation of MyD88 activity and upregulation of TRIF, the maturation of early endosomes into late endosomes is associated with the degradation of TLR4 and shutoff of the TRIF response[40;60]. We measured this process in single cells by observing single-phagosome maturation in RAW 264.7 cells exposed to E. coli conjugated to the pH-responsive dye pHrodo red (Fig. 4.4A). Fluorescent intensity of endosome-localized pHrodo increases over pH ranges associated with the progress of endosome through early to late stages before eventually plateauing, on average, after 10-12 hrs (Fig. 4.14,"Methods").

We associated segmented pHrodo/endosome spots to single cells, and measured the fluorescent increase of these single cell spots over time, normalized to their final state (measured at 17.5 hrs, Fig. 4.4B). We then computed the time taken for each cell to cross given specific maturation thresholds: depending on the threshold applied, a range of maturation times is generated, resembling a normal distribution with a mean ranging from 4 to 10 hrs and a standard deviation of 1-2 hrs (Fig. 4.4C). These times suggested a distribution of delays, after which endosomal ligand-TLR4 complexes would undergo a rapid degradation.

To determine the maturation threshold that functionally downregulates TRIF signaling (Fig. 4.4D), we measured the timing of final NF-κB downregulation in single cells (Fig. 4.4E) revealing a distribution centered at roughly 6.5-7 hrs (Fig. 4.4F). In our

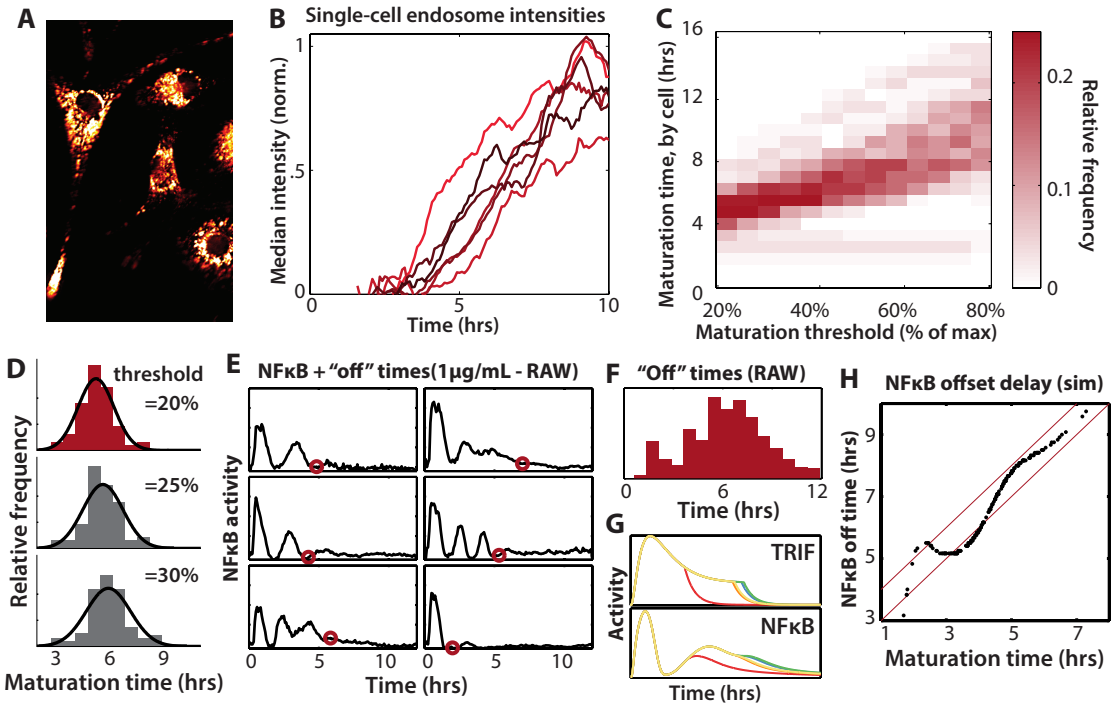

Figure 4.4: Variable endosome maturation shapes termination of TRIF signaling. **(A)** Visualized maturation of endosomes by following the phagocytosis of pHrodo. **(B)** single cell endosome development, as measured by median spot intensity increase over time (normalized to the maximum intensities reached in that cell) **(C)** single cell maturation histograms, showing time taken to cross a threshold (as a function of threshold), applied to the endosome development curves. **(D)** Histograms for low threshold values, with corresponding Gaussian fits. Selected distribution is shown in red. **(E)** The downregulation of NF-κB signaling ("off" time) of 6 randomly-selected cells (highlighted by a red circle in each trajectory). **(F)** Full histogram of computed off times for all measured cells. **(G)** Examples of model's predicted active TRIF and nuclear NF-κB trajectories in response to 500ng/ml LPS after introducing a cell-to-cell variable endosome maturation delay. **(H)** Predicted NF-κB shutoff time vs. corresponding endosome maturation time from 200 single cell simulations described in **(G)**. 2 (lower line) and 3 (upper) hour delays between NF-κB -off and endosome maturation time are indicated by red lines.

model, distributed TRIF and NF-κB shutoff times (Fig. 4.4G) were closely correlated, and the effective delay between them was about 2-3 hrs (Fig. 4.4H, 2 red lines). The best-fit maturation time, then, was estimated to match a normal distribution of 4.4 ± 1.2 hrs corresponding to a 20% maturation threshold (Fig. 4.4D in red). These results reveal the intricate relationship between constitutive and ligand induced trafficking of TLR4 to the endosome, as well as endosomal maturation in determining transience or duration of NF-κB , and suggest that the measured cell-to-cell variability in these processes may directly contribute to the cell-to-cell variability observed in single cell NF-κB signaling studies.

Explaining NF-κB variation with extrinsic noise

Prior work has sought to explain the variability of NF-κB dynamics via intrinsic noise arising from the small number of molecules involved in IκBÎś transcription initiation[113;93]; however, in several other systems, pre-existing cell-to-cell differences (extrinsic noise) was found to be the dominant driver of variability in cellular responses and decisions (as calculated in Chapter 3 and reported elsewhere)[112]. Measured NF-κB dynamics are highly variable and desynchronized across individuals: we observed wide distributions of peak timing, amplitude and duration at all doses (Fig. 4.5A). We extended our modeling work to match these behaviors by focusing on extrinsic noise that affect rates of reactions[112]. In addition to the measured distribution of endosome maturation delays, τ_{mature}, we identified three key receptor-proximal processes in the signal transduction cascade that may be subject to cell-to-cell differences: TLR4 synthesis, TRIF activation, and MyD88 activation (Fig. 4.5B). We introduced extrinsic variability into these processes by applying a zero-mean log-normally distributed multiplier into corresponding reaction rates (Fig. 4.5B bottom). These processes were considered representative, with log-normal distributions being the result of multiplicative combinations of multiple distributed protein species. We also considered downstream processes such as cellular RelA concentration, but found that variability in RelA concentration was

only weakly correlated with overall NF-κB dynamics[52] (Fig. 4.16,"Methods"). A model variant that lacks endosomal maturation distribution but incorporates the three protein distributions predicted variance in total NF-κB activity by amplitude scaling only âĂŞ knowing the amplitude of the 2nd peak, for instance, allows nearly perfect prediction of total NF-κB activity (r=0.98, Fig. 4.15A,"Methods"). By contrast, a model incorporating only a variable τ_{mature} predicts modulation via duration only: for a majority of cells, there is no correlation between 2nd-phase amplitude and total activity (Fig. 4.15B, "Methods"). The measured correlation in single cell data between these two quantities is positive, but weak, indicating a combination of duration and amplitude modulation (r = 0.65, Fig. 4.15C, "Methods"D), and the full model with varied protein and degradation delays provides a good match for this observed relationship (Fig. 4.15D,"Methods"). The full-variation model was fit to the full range of measured single cell responses, from 0.5ng/mL to 5μg/ml LPS (Fig. 4.5A). Our earlier parameterization provided us with the mean of each of our distributed protein levels, leaving us to fit noise levels (i.e. σ of each log-normal distribution). To do so, we used metrics (1st/2nd peak amplitude and total activity) to compress a large set of dynamic trajectories to univariate distributions. Model results were then optimized to fit to measured trajectories by comparing distributions of both simulated and measured metrics, minimizing the Kolmogorov-Smirnov distance between them at all doses (Fig. 4.17, "Methods").

Resultant single cell simulations qualitatively matched experimentally observed responses (Fig. 4.5C and D). At the 5ng/ml dose, nearly all cells show first-peak activity, with a smaller subset of cells exhibiting a second-peak (Fig. 4.5C). At 5μg/ml, nearly all of the cells show a secondary peak, with some cells continuing to oscillate for up to 8-9 hours (Fig. 4.5D). In both cases, the values sampled from each distribution for all of the four varied parameters are shown next to the resultant simulation trajectory. Histograms of overlaid experimental and simulation distributions of dynamical metrics show good agreement across the full range of tested doses (Fig. 4.5E). Finally, Fourier analysis of single cell dynamics revealed the emergence of a distinct harmonic signature, corresponding to 2-hour periodic activity (Fig. 4.5F). Identical analyses performed on

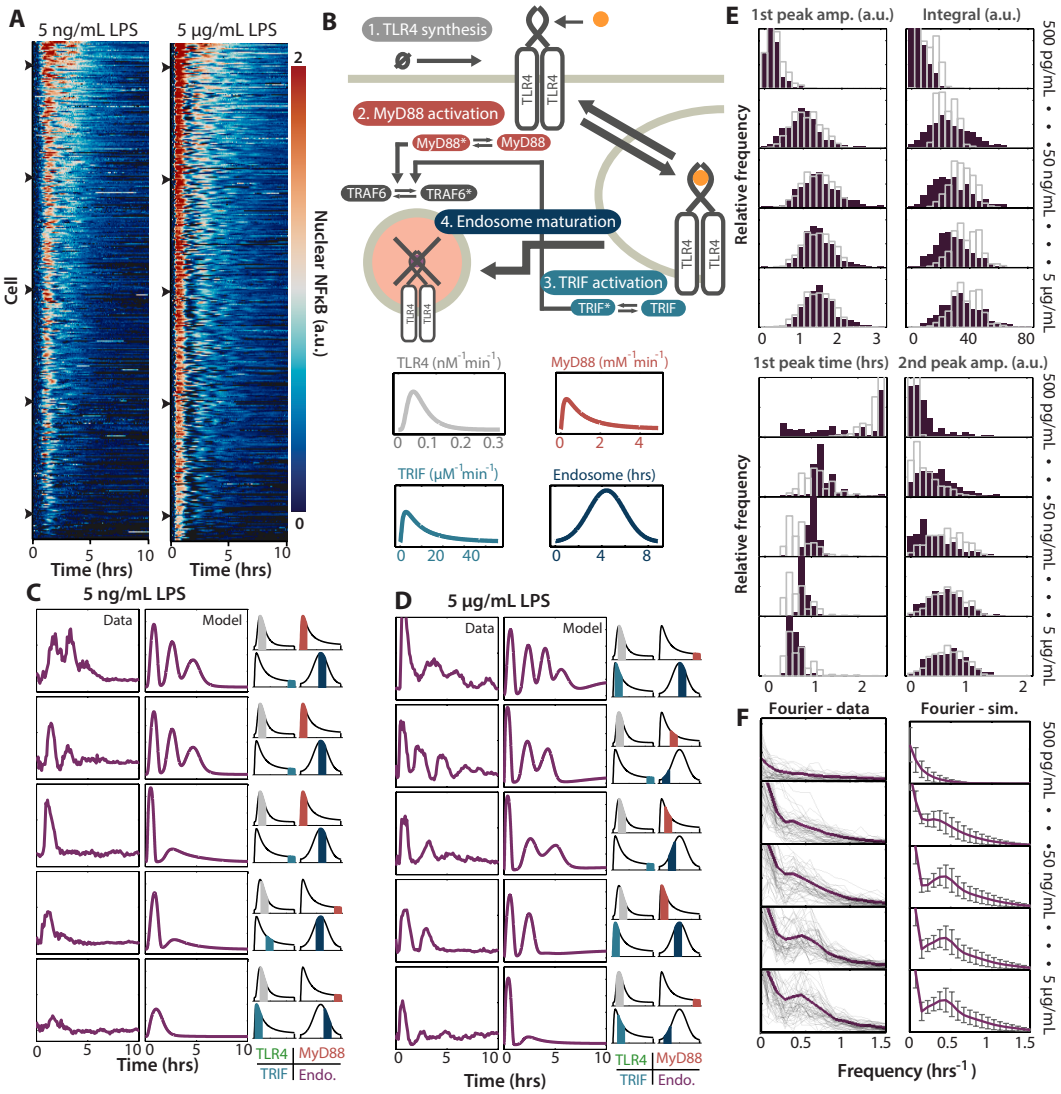

Figure 4.5: Distributed, oscillatory NF-κB nuclear translocations in response to LPS can be fitted using an extrinsic noise model. **(A)** Heatmaps of single cell NF-κB responses to both 5ng/mL and 5μg/ml LPS. **(B)** Location of four varied reactions (indicated by number and colors in top panel) in model topology and the corresponding rate distributions (bottom) which represent cell-to-cell variability/extrinsic noise: synthesis rates of TLR4, TRIF, MyD88 and the endosome maturation time were varied. **(C-D)** Five typical NF-κB trajectories in experiments, the best-fit single cell simulation (minimum Euclidean distance) and the corresponding parameter values in each case for 5ng/ml LPS **(C)** and 5μg/ml LPS **(D)**. **(E)** Metrics (1st peak time and amplitude, 2nd-peak time, and integrated activity) were used to characterize distributions of NF-κB responses in RAW264.7 cells (solid color) and best-fit simulations (gray outlines) at each of 5 doses of LPS. (X- and Y-limits are held constant for each metric at all doses). **(F)** Fourier transformations of dynamics, by dose: (left) RAW 264.7 trajectories are shown as thin black lines, and the root-mean-square (RMS) of all measured responses is shown in purple. (right) Simulations: RMS magnitude is shown in purple, with standard deviations from 500 simulations. Extrinsic model fitting aided by Zhang Cheng.

500 iterations of the varied single cell model shows a similar frequency distribution: trajectories are dominated by low-frequency information, but higher doses give rise to a consistent, distinct harmonic peak at roughly 0.5 h-1. We note that the period encoded in these responses appears to be characteristic of the NF-κB delayed negative feedback system: similar periods have been reported in response to other stimuli[57;38].

Dynamic specifity encoded by variation in MyD88/TRIF

The extrinsic noise single cell model recapitulated many aspects of observed behavior, and could now be used to examine the individual contributions of introduced noise sources. Simulated and measured NF-κB dynamics show high amounts of overall activation variability, particularly at late timepoints (Fig. 4.6A and B). We ranked both quantified and simulated dynamics by total response level, then applied the same ordering to the generated parameter values for each simulation (Fig. 4.6C). We show that parameters determining TRIF activation and inactivation make the largest contributions to the response strength (correlation, r, between $\log_{kaTRIF}$ and ranking = 0.61, and correlation between ÏĎ$_{mature}$ and ranking = 0.64), while ligand-TLR4 and MyD88 activation play weaker roles (r=0.30 and 0.18, respectively). To examine how each noise source affects specific dynamical features of the NF-κB response, we repeated simulations using univariate distributions (Fig. 4.6D, example trajectories shown on top). Variation in MyD88 causes modulation of 1st-peak timing and amplitude, but fails to modulate total activity as strongly as variation in either TRIF activation and inactivation. Additionally, we note that MyD88 variation leads to a roughly 2-fold difference in 1st-peak amplitude, much smaller than the 8-fold difference in 2nd-peak amplitude caused by (equivalent) TRIF noise, or the 5-fold difference in signaling duration (nuclear NF-κB > 0.015 μM) induced by the range of maturation delays.

Our analysis suggests that the variability of NF-κB dynamics in macrophages is largely driven by variability in TRIF signaling. To better isolate the effects of TRIF signaling variability, we examined IRF3 dynamics, which, unlike NF-κB , is activated

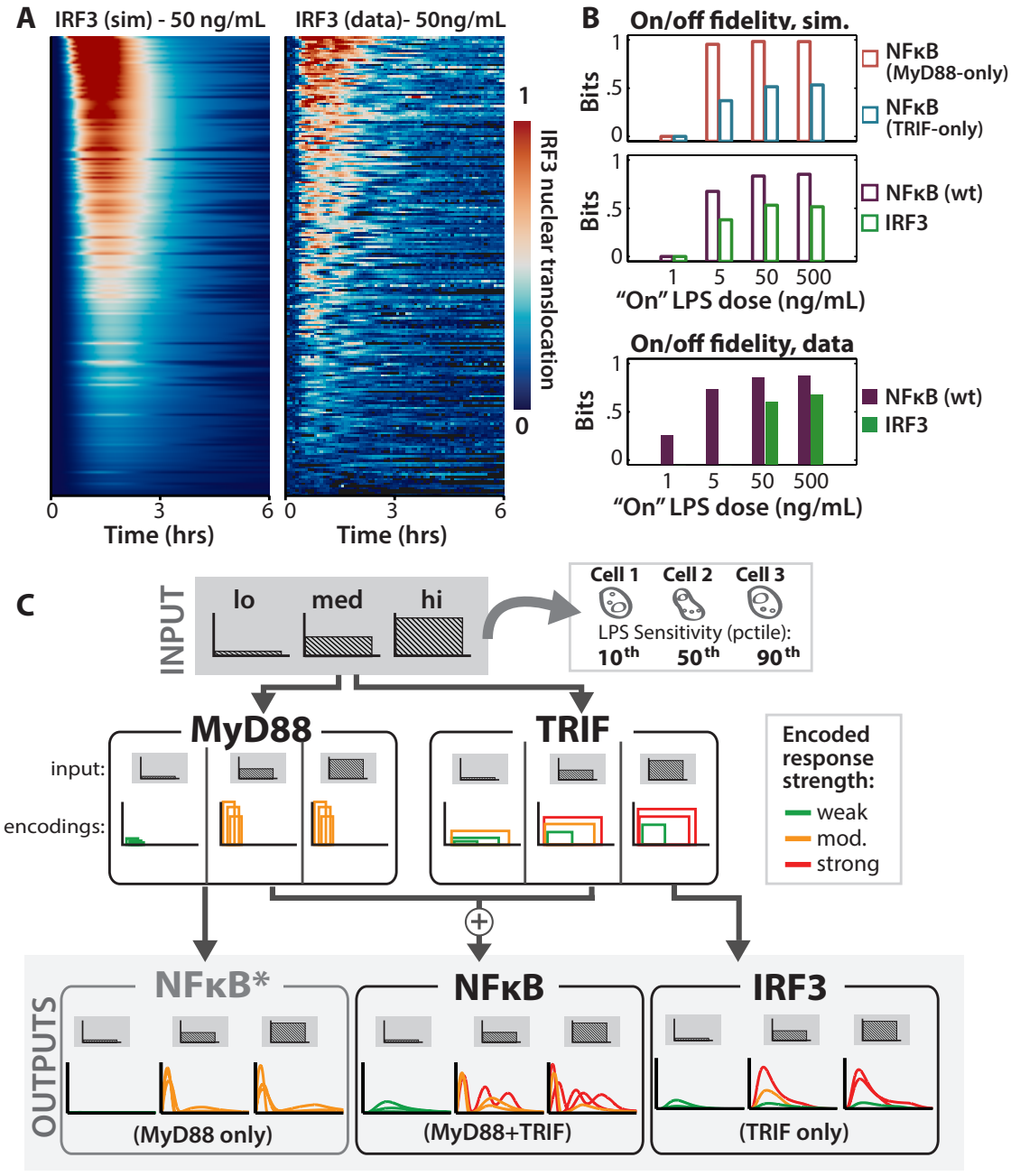

Figure 4.6: Isolating the effects of variability in specific signaling processes on the NF-κB and IRF3 dynamics. (A) Spectrum of single cell NF-κB responses to 500 ng/mL LPS, ordered by overall response strength. **(B)** Identically-ordered model simulation, using full varied parameter set. **(C)** Relationship between parameters and response strength: points are aligned with rows, showing the specific parameter value for all varied quantities, along with total correlation between sort order and parameter magnitude. **(D)** (Top) Randomly-generated NF-κB trajectories resulting from single-parameter variation models in response to 500ng/mL LPS. Metrics for integrated activity, 1st-peak amplitude and time, 2nd-peak amplitude, and total duration (>15nM) are plotted as a function of individually varied parameters.

solely by the TRIF pathway, and is therefore a more direct readout for the variability of TRIF signaling. Indeed, simulating IRF3 activity in our full-variation model predicts a broadly distributed range of IRF3 responses (Fig. 4.7A), with a high proportion of cells showing little-to-no-response (relative to NF-κB activity at the same dose, 30% vs 5%). To test this prediction, we engineered a RAW264.7 cell line expressing mVenus-IRF3 in which IRF3 nuclear translocation was signal-dependent. single cell experimental data also revealed a range of responses that were similarly distributed. (We also note oscillatory unexpected characteristics that warrant future study.) The lack of a robust "on" state means that transduction via the TRIF pathway inherently involves information loss: no matter how high of an input is applied, some cells will act indistinguishably from non-activated ones.

To quantify this phenomenon, we applied the information theory formalism outlined in Chapter 3 to measure the differences in the ability of MyD88 and TRIF branches to robustly encode stimulus inputs into transcription factor outputs. We calculated channel capacity between response distributions to an "off" dose (0.5 ng/mL LPS) and an "on" dose of 1, 5, 50, or 500 ng/mL LPS (Fig. 4.7B). In each case, a channel capacity of 1 bit would indicate perfect "off" vs. "on" encoding, allowing perfect distinction of the two states based on measured maximal outputs.

As expected, the ability to distinguish between off and on rises with dose, but our model predicts that the MyD88 pathway is less noisy than its counterpart. Both *Trif*$^{-/-}$ and wt NF-κB responses show a sharp increase in transmission capability between 1ng/mL and 5ng/mL, or at our hypothesized "switch point" for MyD88 activation (Fig. 4.7B top). Better information transmission is achieved for all doses than either *Myd88*$^{-/-}$ NF-κB or IRF3 responses, where activity is solely controlled by TRIF (Fig. 4.7B top/middle). Our measured datasets (1, 5, 50, and 500ng/mL LPS for NF-κB , and 50 and 500ng/mL for IRF3) confirm this prediction: even at very high input doses of LPS, cell-to-cell variability leads to non-robust responses in the absence of structurally-enforced control (Fig. 4.7B, bottom).

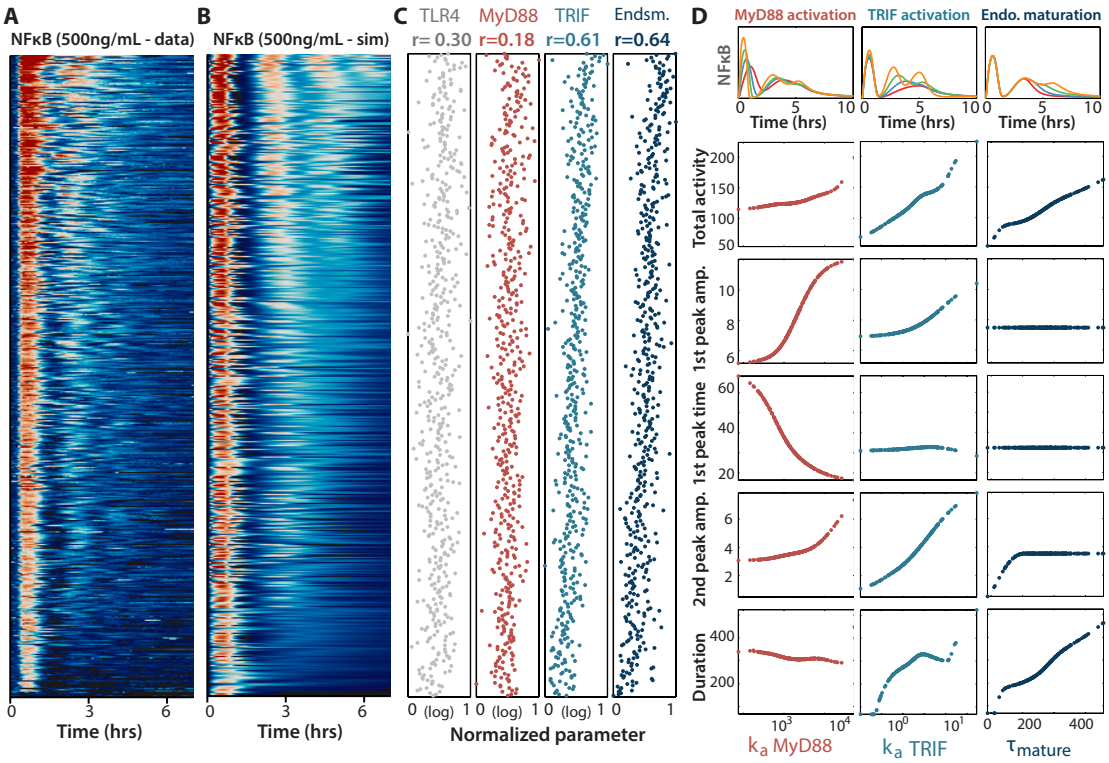

Figure 4.7: MyD88 and TRIF pathways encode distinct dynamics and cell-to-cell variability of NF-κB responses. **(A)** Heatmap of modeled (left) and measured (right) IRF3 trajectories in response to 50 ng/mL LPS. **(B)** Calculated channel capacity between outputs to a 0.5 ng/mL input (i.e. no visible activation, or "off") and a varied "on" LPS dose. One bit indicates 100% distinction between "on" and "off" in all measured individuals. (top) Simulated *Myd88*$^{-/-}$ and *Trif*$^{-/-}$ NF-κB dynamics. (middle) simulated wt NF-κB and IRF3 dynamics. (bottom) NF-κB and IRF3 responses in RAW 264.7 cells at 1, 5, 50, and 500ng/mL LPS (NF-κB), or 50 and 500ng/mL LPS (IRF3). **(C)** Summarizing effects of TLR4 response encoding by MyD88/TRIF on IRF3, NF-κB , and NF-κB * (hypothetical; activated only by MyD88) outputs.

Discussion

Here we have pursued an iterative approach of computational modeling and experimentation to develop a predictive understanding of how TLR4-responsive NF-κB dynamics in single cells are encoded by MyD88 and TRIF pathways. We distinguished between several dynamic features and mapped these to underlying regulatory network topologies and molecular biochemical characteristics. In turn, these mechanisms determine the variability of single cell responses in the population. Having achieved a base parameterization of the model, we added extrinsic noise to match variability observed in single cell data. Specifically, we first experimentally measured one potential source of noise (single cell endosome maturation). We then fitted three others, related to receptor-proximal protein species, by comparing distributions of relevant single-dimension metrics. While the metric-match methodology could be applied to other systems, we were aided in this case by the temporal separation of our two pathways: as a result, aspects of wild-type responses map uniquely to the TRIF or the MyD88 pathway.

We found that endotoxin-responsive NF-κB dynamics in macrophages are oscillatory, contrary to previous population-level observations in MEFs[22;110]. This finding was first predicted by a model of the TLR4 signaling module, which was parameterized based on published quantitative data and IKK measurements. It was then confirmed in single cell experiments, which also revealed substantial cell-to-cell variability that explained why such oscillatory dynamics were not evident in prior population level measurements. Thus oscillatory dynamics may be a more profound and conserved feature of NF-κB signaling than previously thought, being independent of the stimulus, and solely a property of IκB feedback (previously derived theoretically[57]). Structural characterization of MyD88[27] revealed that it oligomerizes into a large signaling molecular complex, or signalosome. This oligomerization has been hypothesized to generate positive cooperativity (38, 39), and similar signalsome-based cooperativity has been suggested to generate bistability in the apoptosis decision process (40). However, though signalsome-forming adaptor proteins have been shown to provide high specificity by selectively

recruiting substrates (41), their effects on the dynamics of signal transduction remain unclear[27;82]. Our study suggests that MyD88-signalsome-mediated cooperativity, modeled by and fitted to a Hill equation, is sufficient to explain a dose threshold in both IKK activity and NF-κB translocation. It ensures not only the reliability of MyD88 signaling but also its transience (Fig. 4.2G, together with the translocation module, Fig.3), while reducing its scalability in response to different doses (Fig. 4.2H). Several other signal transmission systems upstream of NF-κB have been shown to exhibit similar thresholding behavior (mediated by either cooperativity or positive feedback), including in the TNF receptor-R1-IKK signal transduction axis[105], as well as B-cell receptor (BCR) signaling[90], where CARMA1 mediates IKK-dependent positive feedback. It is worth noting that the positive feedback motif in BCR signaling does not ensure the transience but rather the prolongation of NF-κB dynamics. Receptor endocytosis provides temporal separation between NF-κB activation from the MyD88 and TRIF sub-pathways: though this delay is small relative to the oscillatory period caused by IκB feedback, it is sufficiently long to ensure that first-peak timing is largely driven by MyD88 signaling, which thereby determines the oscillatory phase of the cell. As LPS, CD14, MD-2, and TLR4 form stable complexes[72], and LPS itself is stable, receptor trafficking serves to determine the duration of MyD88 signaling, and endosomal maturation serves to determine the duration of TRIF signaling. Our work suggests that the process of deactivation via endosome maturation plays a role in determining total signal duration. single cell measurements of this process showed high variability, adding another source of noise in encoding late NF-κB activity.

Inflammatory gene expression programs induced by TLR4 signaling reach their maximal induction at different times[34], and may differentially depend on NF-κB dynamics. Our study suggests that genes dependent on the first peak of NF-κB activity may show less cell-to-cell variability than those dependent on late NF-κB activity. Consistent with this suggestion, a recent single cell gene expression study noted subsets of genes that were unimodally, universally expressed early in dendritic cells exposed to LPS, followed by a multimodal, highly-variable distribution of inflammatory gene

expression at late times[86].

Innate immune responses must ensure both high sensitivity and tight control that provides for appropriate responses that minimize the potentially destructive consequences to surrounding tissue[67]. The integration of two pathways with distinct signaling topologies and systems characteristics may enable the TLR4 network accomplish these competing tasks. Our results suggest that in the face of high extrinsic noise associated with macrophage and micro-environment heterogeneity the MyD88 pathway is still able to provide for reliable first signaling response due to a signalosome topology (Fig. 4.7C left); however, this is limited in duration and not scalable with dose, minimizing the risk of inflammatory damage. In contrast, the TRIF pathway exploits cell-to-cell variability to provide a prolonged response only in a fraction of cells that is scalable with dose (Fig. 4.7C, right). By limiting the number of cells that can produce cytokine at levels at which they potentially act systemically we speculate the network behavior also reduces the risk of inflammatory shock while providing for scalability in systemic immune activation. Recent work delineating expression control and functions of TNF is consistent with this view: whereas local autocrine functions of TNF only require MyD88-mediated signaling events, TRIF-mediated signaling mechanisms are required for full expression of TNF and its (non-local) paracrine functions[16].

By combining the two pathway responses (Fig. 4.7C, bottom-middle), every cell in a population is able to mount a minimum response to the presence of LPS to provide a digital and reliable local response, and this response is able to scale (in subsets of cells) with the extent of the threat, tuning systemic immune functions appropriately. Our study thus reveals how distinct topologies of innate immune signaling pathways either mitigate or exploit molecular network noise to provide appropriate responses at the cell, tissue or organism level appropriate to the pathogen threat while carefully controlling the risk of endotoxic shock.

Methods

Biochemical Assays

Wild-type and gene-deficient C57BL/6 mice were housed at UCSD in accordance with protocols authorized by the Institutional Animal Care and Use Committee. Bone Marrow-Derived Macrophages (BMDMs) were generated by culturing 6 x 10^6 bone marrow cells from mouse femurs in suspension in L929-conditioned DMEM on 15cm plates for 7 days, before replating them for experiments on day 8. LPS was sourced from Sigma (B5:055). Antibodies against RelA/p65 (sc-372) were from Santa Cruz Biotechnology. EMSA and kinase assays were performed as previously described[88].

Computational Modeling

Model Derivation

Our modeling efforts in this paper focused largely on the TLR4 module. Model structure and parameters were derived as follows:

Ligand capture and endocytosis

Our current understanding of the mechanism by which TLR4 senses LPS is: LPS first binds to LBP with high affinity (3.5nM), and then CD14 chaperones LPS from LBP to TLR4-MD-2[100]. CD14 is responsible for the uptake of LPS and also controls endocytosis of the receptor-ligand complex in a Syk-dependent manner[119]. Thus, CD14 is explicitly modeled here. Early studies found that there is no enzymatic digestion of the cell surface-bound LPS[54]. Thus, LPS is only degraded in the endosome compartment. We simplified the receptor and co-receptor binding into the following processes

1. LPS concentration transformation factor. The molecular weight of LPS is estimated to be 10kDa (see http://textbookofbacteriology.net/endotoxin.html). The

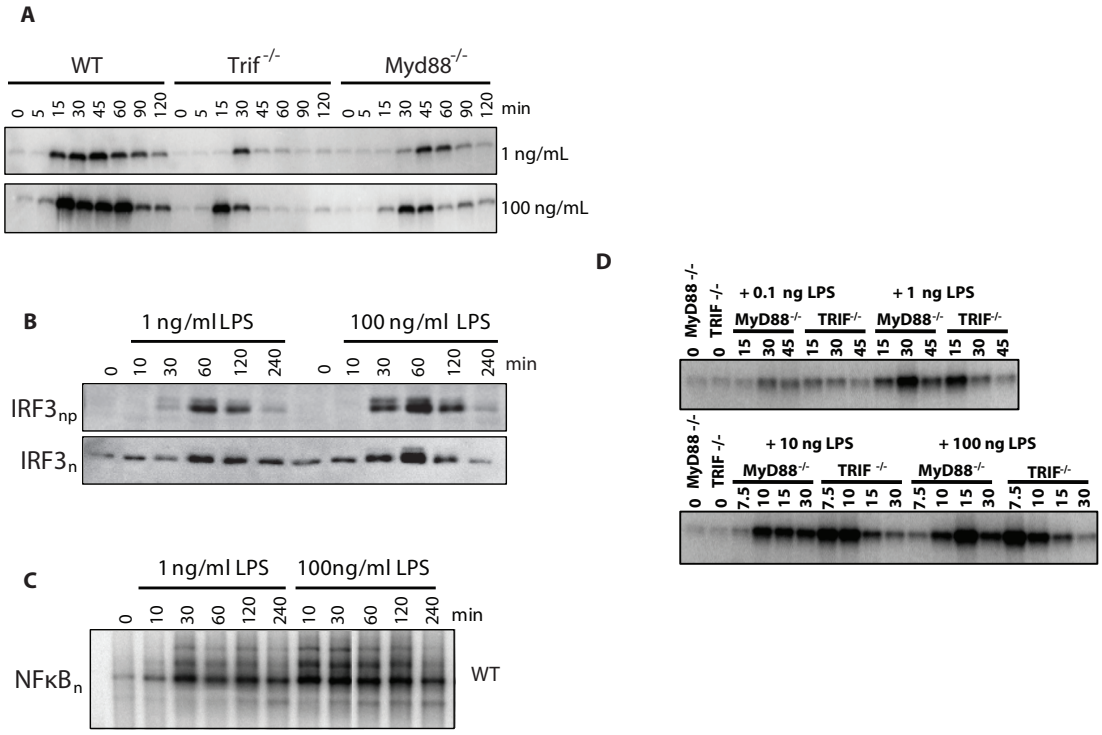

Figure 4.8: Biochemical assay blots. **(A-C)** Blots quantified in fig. 4.1. **(A)** IKK kinase assay blot (produced with co-immunoprecipitates with NEMO), performed on wt and knockout cell lines for 1ng/ml and 100ng/ml LPS stimulation. **(B)** IRF3 nuclear phosphorylation western blot. **(C)** Nuclear NF-κB electromobility shift assay (EMSA). **(D)** IKK kinase assay quantified in fig. 4.2for indicated doses in BMDMs derived from indicated knockout strains. IKK was co-immunoprecipitated with NEMO. Vertical numbers above each the gel is the time in minutes. Biochemical experiments performed by Diana Ourthiague.

maximum dose modeled was 10 μg/ml, which equates to 1 μM. (10μg/ml = 1μM).

2. TLR4 concentration and distribution in macrophages without stimuli. It is estimated that the number of TLR4 receptors per cells is close to 50,000. Total cell volume is around 4000 μm^3 (Bionumbers database ID: 103566). Thus, the default concentration of TLR4 is expected to be near 20 nM. As we know that the majority of TLR4 resident in the cell surface for macrophage before treatment[108], we assume a ratio of roughly 1/20 between surface and endosome distribution of TLR4.

3. LPS binds CD14 near the cell surface (k_1, k_2, k_3, k_4). Effective concentrations of CD14 (including its membrane and soluble forms) are set by its synthesis (k_3) and degradation rates (k_4), and the dissociation constant between CD14 and LPS (k_2/k_1) is estimated to be on the order of 8-10 nMciteAkashi2003. Using this dissociation constant as a starting point, we fit data measuring the (CD14-dependent) LPS uptake rate of cells[28] to get the values for k_1, k_2, k_3, and k_4. The resultant model fit is shown fig. 4.9, left.

4. Captured LPS binds to TLR4/MD2 complex (k_5, k_6) and ligand endocytosis (k_8, k_{12}). TLR4 is generally bound with MD2, and the binding affinity (k_6/k_5) between LPS and TLR4/MD2 is about 3-5 nM[1]. Again, using the dissociation constant to constrain parameters, we fit remaining degrees of freedom to a TLR4 endocytosis assay (56): in this case, both the binding and the ligand-induced endocytosis rate constants (k_5, and k_{12},respectively) were fit (see best fit result in fig. 4.9, right). Finally, CD14-LPS endocytosis is assumed to match the fitted value for CD14-dependent TLR4-LPS endocytosis, or $k_8 = k_{12} = 0.066$ $\mu M^{-1} min^{-1}$. For this and the above fit, we employed the "patternsearch" algorithm in the MATLAB R2014a Optimization Toolbox.

5. TLR4 synthesis, degradation, and shuttling (k_7, k_{10}, k_{11}, k_{15}) The estimated time

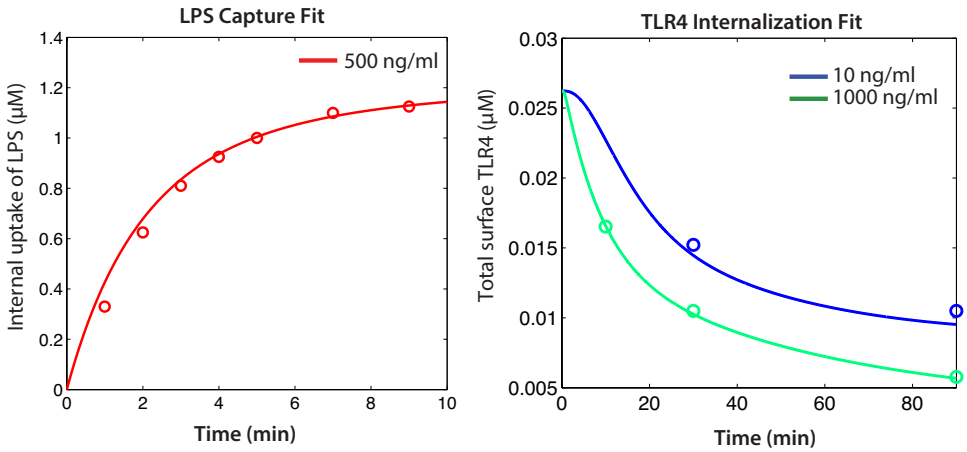

Figure 4.9: Model fitting to literature data (1). Preliminary fitting to (left) LPS uptake[100] and (right) TLR4 endocytosis staining[119] assays was used in conjunction with dissociation constants gathered from literature. (Open circles = assay data. Solid lines = model prediction).

scale for total receptor turnover is between 1-5 times per hour[94]. We therefore set the constitutive endocytosis rate of TLR4 (k_{10}) at 0.028 min-1. We additionally know that at steady state, given the assumption of the membrane:endosome distribution of TLR4, (k_{11}+k_{15})/k_{10} = $[TLR4]_{pm}$/$[TLR4]_{en}$ = 20, and that $[TLR]_{pm}$ = 20*k_7/k_{15} ,or roughly 0.02 μM. This allows for the direct estimation of rate constants related to TLR4 synthesis (k_7), degradation (k_{15}), and shuttling (k_{11}).

6. Ligand recycling from early endosome to plasma membrane (k_{13}, k_9). It has been established that endosomal recycling in eukaryotes occurs via one of two processes: a fast process via early endosome (half-life around 5 mins), and a slower one by Rab11-containing recycling endosomes with half-life around 15 20 mins[87 98]. Recently, it has been found that upon stimulation, TLR4 can be recycled via Rab11 containing recycling endosome[119]. We therefore assume that receptor-ligand complexes recycle via the slow routine with a rate constant about 0.04 min-1 (k_{13}) and free ligand recycle with the same rate (k_9 = k_{13}).

Kinetic threshold generated by signalosome formation

Based on the structure and behavioral differences following site-directed mutagenesis, Lin et al. suggested a sequential assembly model for Myddosome[55]. In essence, LPS binding induces both dimerization and higher-order oligomerization of TLR4[27]. MyD88 is then recruited to the receptors through homotypic TIR-TIR interactions and facilitates Myddosome assembly with IRAK4 and IRAK1 or IRAK2 through homotypic DD-DD interactions[27]. According to this hypothesis, we proposed a six-step model for Myddosome formation shown in Fig. 4.2A.

Table 4.1: MyD88 reaction schema

Step	Reaction	k	Description
1	C_0 + 2M => CM_2	k_f	The TLR4-LPS dimer (C) recruits two MyD88 (M) molecules forming a complex (CM_2).
2	CM_2 => C + 2M	k_b	Disassociation
3	2 CM_2 => C2M4	k_f	Two CM_2 complexes bind and form C_2M_4
4	2 CM_2 + C_2M_4 => C_3M_6	k_f	C_2M_4 binds to CM_2 forming C_3M_6.
5	C_3M_6 => 3 C + M_6	k_f	In the C_3M_6 complex, interactions of nearby MyD88 molecules form hexamer (M_6), releasing TLR4 dimers.
6	M_6 => 6 M	k_f	Disassociation to monomers
7	M_6 =>Myddosome	k_{cat}	M_6 continues to form the Myddosome, activating TRAF6

We use singular rate constants k_f to represent the rate of allforward reactions, including binding and M_6 formation, and k_b as the rate constants of all backward reactions

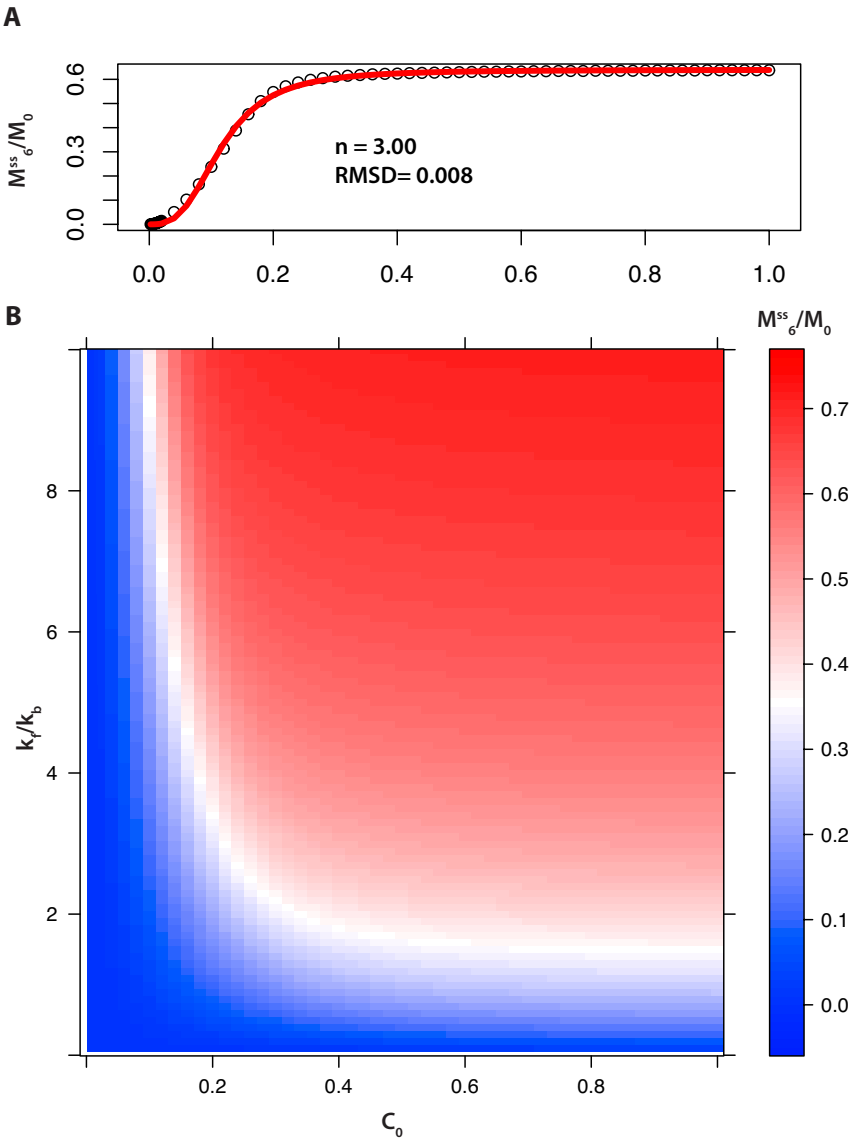

Figure 4.10: Approximating sequential binding using Hill kinetics. **(A)** The relative concentrations of M_{6ss}/M_0 versus the relative input concentration C_0/M_0 (dots). Parameters are k_f=0.61, k_b=0.1 and M_0=10. The data were fitted by a Hill equation (Hill constant n = 3) by using the *nlminb* function in R programing language (solid line). The root-mean-square deviation (RMSD) for the fit is 0.008. **(B)** The dose-response of M_{6ss}/M_0 to C_0/M_0, by varying the fraction k_f/k_b (changing k_f only). Sequential MyD88 modeling and Hill fitting performed by Zhang Cheng.

(disassociation). We suppose that the steps 1 to 6 are fast relative to reaction 7. So based on the quasi-steady-state hypothesis, these reactions quickly reach equilibrium given an initial concentration of the TLR4-LPS dimer (C_0). The Myddosome formation rate in 7 then is in proportion to the corresponding steady-state level of M_6 (v_7=k_{cat}*M_{6ss}).

Relative concentrations of M_{6ss}/M_0 versus the relative input concentration C_0/M_0 are plotted in upper-left panel of Fig. 4.10 (circles). Parameters are k_f =1, k_b = 0.1 and M_0 = 1. These dots can be fitted by a Hill equation with Hill constant n = 3.2 (line). Thus, the Myddosome formation or MyD88 activation rate rate can be written as v_7=k_{cat}*M_{6ss} = k_{cat}*$C_0{}^n$/ ($C_0{}^n$ + $k_m{}^n$)*M_0. In Fig. 4.10, we varied the fraction k_f/k_b by changing k_f only and present the dose-response between M_{6ss} and C_0 in a heat map. Fig. 4.2 shows best-fit Hill constants between 1.8-3.2 correspond to ratios of kf/kb ratios from 0.1 to 10. We fitted MyD88 EC_{50} and Hill coefficient (k_{20} and k_{21}), within this allowable range, using error terms generated from our NF-κB and IKK data.

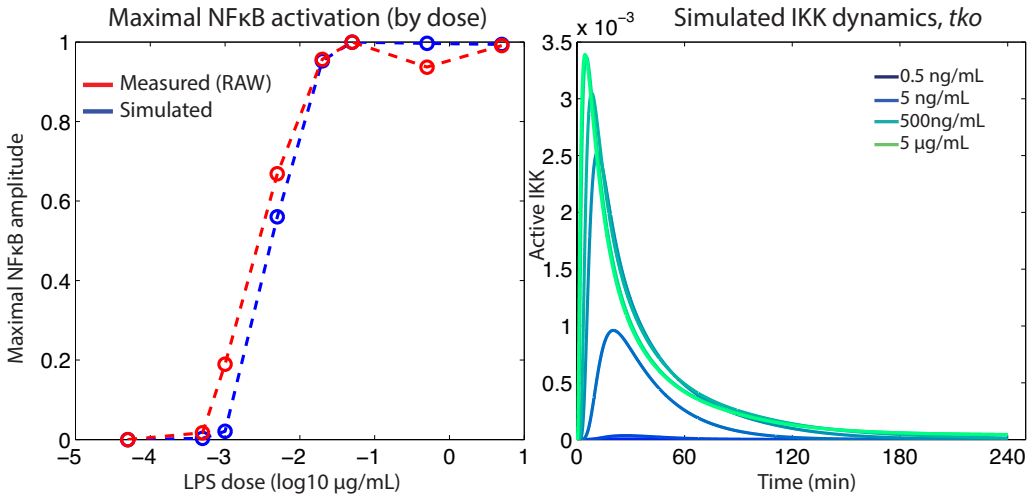

Figure 4.11: Model fitting to literature data (2). Fitting modeled EC_{50} and Hill coefficient to aspects of the MyD88 response (1st-peak amplitude in RAW264.7 cells NF-κB responses and IKK activity in *Trif*$^{-/-}$ BMDMs).

Model Structure

As outlined in Fig. 4.1, our reaction schema is modular: subsets, or pathways, could be distiguished, and idividual modules interact unidirectionally with other modules at a single discrete point. Thus, the TLR4 module activates IKK and TBK kinases via MyD88 and TRIF. The IKK module[11], in turn, transduces dynamic signals, degrading IκB to activate the NF-κB module[111]. This module produces nuclear NF-κB activity as a function of IÎžB degradation and synthesis control, while a separate IRF3 module produces nuclear IRF activity as a function of import/export mechanisms controlled by TBK1 (which is, in turn, interfaces with the TLR4 module via TRIF). Following parameterization outlined previously, we are left with the full schema shown in Fig. 4.12, and the full set of parameter values for our newly-parameterized models are shown in Tables 4-4 (parameters in the NF-κB module are as described previously[111].) In all tables, "pm" refers to species on the plasma membrane, "en" refers to species in the early endosome, and "n" refers to species in the nucleus.

Table 4.2: TLR4 module parameters

No.	Reaction	Value	Unit	Description
1	LPS => LPSpm	4.2	min^{-1}	Binding LPS to CD14
2	LPSpm => LPSen	0.17	min^{-1}	LPS translocation
3	LPSen => LPSpm	0.15	min^{-1}	LPS export
4	LPSen =>	0.07	min^{-1}	Endosomal degradation of LPS
5	LPSpm + TLR4pm => TLP4LPSpm	200	$\mu\text{M}^{-1}\text{min}^{-1}$	Association of LPS and TLR4
6	TLP4LPSpm => LPSpm + TLR4pm	0.6	min^{-1}	Disassociation of TLR4LPS
7	LPSen + TLR4en => TLR4LPSen	200	$\mu\text{M}^{-1}\text{min}^{-1}$	Association of LPS and TLR4
8	TLR4LPSen => LPSen + TLR4en	0.6	min^{-1}	Disassociation of complexed TLR4-LPS
9	=> TLR4pm	2.31e-5	$\mu\text{M}^{-1}\text{min}^{-1}$	TLR4 generation
10	TLR4en =>	0.07	min^{-1}	TLR4 depletion
11	TLR4pm => TLR4en	0.021	min^{-1}	TLR4 internalization
12	TLR4en => TLR4pm	0.14	min^{-1}	TLR4 recycling

Table 4.2: TLR4 module parameters (cont.)

No.	Reaction	Value	Unit	Description
13	TLR4LPSpm => TLR4LPSen	0.087	min^{-1}	TLR4-LPS cendocytosis
14	TLR4LPSen => TLR4LPSpm	0.04	min^{-1}	TLR4-LPS complex recycling
15	TLR4LPSen =>	0.07	min^{-1}	Receptor complex degradation
16	MYD88 => MYD88* ; TLR4LPSpm	21.5	$\mu\text{M}^{-1}\text{min}^{-1}$	MyD88 activation (by receptor complex)
17		3	Hill coef.	
18		0.022	EC50: μM	
19	MYD88* => MYD88	0.987	min^{-1}	MyD88 inactivation
20	TRIF => TRIF* ; TLR4LPSen	10767	$\mu\text{M}^{-1}\text{min}^{-1}$	TRIF activation by receptor complex
21	TRIF* => TRIF	0.027	min^{-1}	TRIF inactivation

Table 4.3: IKK module parameters

No.	Reaction	Value	Unit	Description
22	TRAF6 => TRAF6* ; MyD88*	115	$\mu\text{M}^{-1}\text{min}^{-1}$	TRAF6 activation by MyD88*
23	TRAF6 => TRAF6* ; TRIF*	0.75	$\mu\text{M}^{-1}\text{min}^{-1}$	TRAF6 activation by TRIF*
24	TRAF6* => TRAF6	0.47	min^{-1}	TRAF6 inactivation
25	IKKK => IKKK* ; TRAF6*	0.00098	$\mu\text{M}^{-1}\text{min}^{-1}$	IKKK activation by TRAF6*
26	IKKK => IKKK*	5e-7	min^{-1}	Basal activation
27	IKKK* => IKKK	0.25	min^{-1}	Inactivation
28	IKK => IKK* ; IKKK*	520	$\mu\text{M}^{-1}\text{min}^{-1}$	Activation
29	IKK => IKK*	0.00005	min^{-1}	Basal activation
30	IKK* => IKK	0.02	min^{-1}	Inactivation
31	IKK* => IKKi	0.15	min^{-1}	Inactivation
32	IKKi => IKK	0.02	min^{-1}	Inactivation

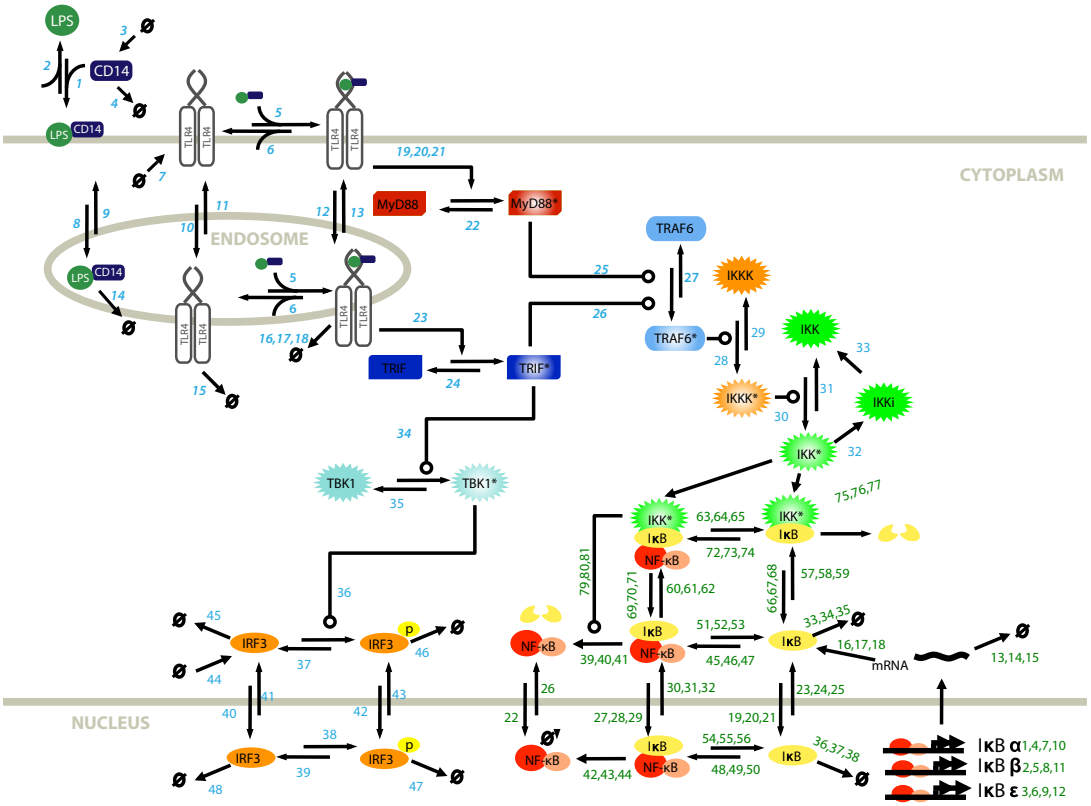

Figure 4.12: TLR4-NF-κB model schema. Full schematic diagram of the model reaction network. The arrows represent the reaction and the numbers close to the arrows indicate the parameter indexes, listed below. Green parameter indices specify values specific to the NF-κB module, and blue indices indicate IKK, TLR4, or IRF3 module parameters. Initial model structure by Zhang Cheng.

Table 4.4: IRF3 module parameters

No.	Reaction	Value	Unit	Description
33	TBK1 => TBK1* ; TRIF*	0.0057	$\mu M^{-1}min^{-1}$	Activation
34	TBK1* => TBK1	0.179	min^{-1}	Inactivation
35	IRF3 => IRF3* ; TBK1	12.8415	$\mu M^{-1}min^{-1}$	Activation
36	IRF3* => IRF3	0.0000010142	min^{-1}	Inactivation
37	IRF3n => IRF3n*	0.0143	min^{-1}	Activation
38	IRF3n* => IRF3*	0.0194	min^{-1}	Inactivation
39	IRF3 => IRF3n	0.00091935	min^{-1}	Translocation
40	IRF3n => IRF3	3.6871	min^{-1}	Translocation
41	IRF3* => IRF3n*	0.0275	min^{-1}	Translocation
42	IRF3n* => IRF3*	0.4145	min^{-1}	Translocation
43	=> IRF3	1.0546	$\mu M^{-1}min^{-1}$	Synthesis
44	IRF3* =>	0.015	min^{-1}	Degradation
45	IRF3n =>	0.0091	min^{-1}	Degradation
46	IRF3 =>	0.0013	min^{-1}	Degradation
47	IRF3n* =>	0.0584	min^{-1}	Degradation

Endosome flux and maturation

In order to test the parameter dependence/robustness of shuttle flux contribution, we measured the ”relative sum of flux difference” (RSFD). We then analyzed RSFD as a function of altered parameters involved in endocyotis/recycling (Fig. 4.13). More positive values indicate a higher contribution of ligand-induced shuttling (LI), and more negative values indicate a higher contribution of constitutive shuttling (CS). The formula for RSFD is given by:

$$RSFD(t) = \frac{\int_0^t flux_{LI}dt - \int_0^t flux_{CS}dt}{max(\int_0^t flux_{LI}dt, \int_0^t flux_{CS}dt)} \tag{4.1}$$

Endosome maturation was performed as shown in Fig. 4.14. For each image taken of the pHrodo-*E. coli*, we first used a high-pass filter (designed to target spacial

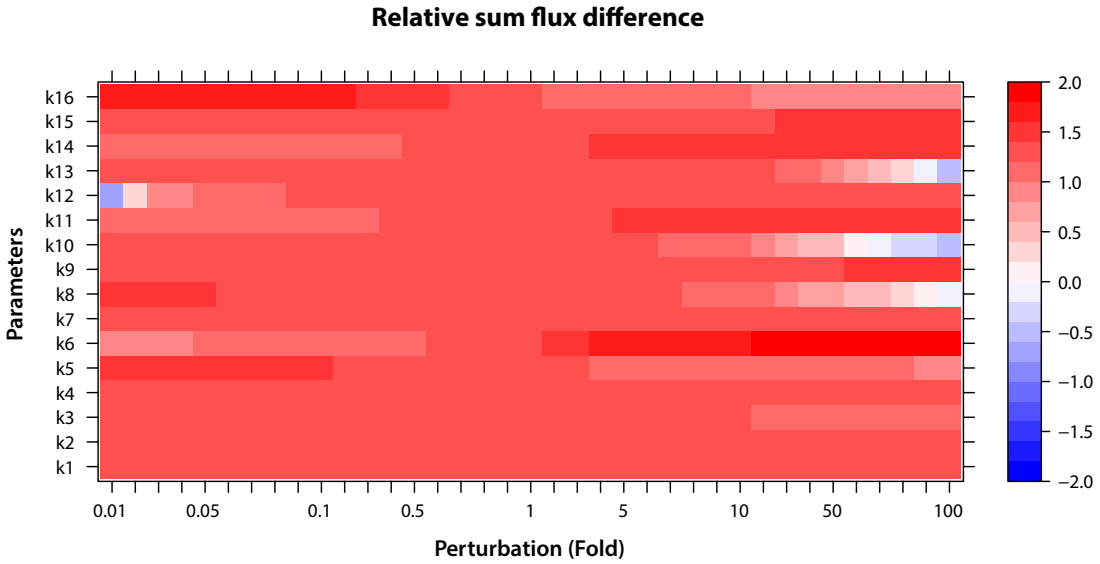

Figure 4.13: The parameter dependence of the contribution of shuttling modes. RSFD as a function of individually increased/decreased parameters. Only decreasing ligand-induced endocytosis rate (k_{12}) to 1/100 of its original value, or increasing k_8, k_{10} and k_{13} to >50-fold of fitted values cause the relative sum of flux difference become negative (i.e. such that the constitutive shuttling would contribute more flux than the ligand-induced one). Flux analysis performed by Zhang Cheng; biochemical experiments performed by Diana Ourthiague.

features <10 pixels in diameter) to emphasize individual spots and better resolve phagosomes that were slightly above or below the focal plane. On the first iteration of the measurement, the algorithm would select an appropriately late frame (after >10 hrs of stimulation) and determine an appropriate threshold to identify bright spots, using the speckle-noise thresholding algorithm outlined in Chapter 2. All subsequent frames, after high-pass filtering, were then thresholded using this value, and individual dots were counted, their intensities measured, and assigned to cells according to the tracking algorithm's estimate of cytoplasmic boundaries. Each cell, then, posessed a discrete set of "dots" that increased in brightness over time before eventual saturation at roughly 10-12 hrs.

In modeling this process, we noted that much of the cell-to-cell variation was not in the rate of brightness increase (or maturation rate), but rather the delay time before initial maturation was noted. For this reason, we chose to model the process as a variable discrete delay, after which degradation of bound TLR4 would rapidly increase. In our model, this allowed partial decoupling between amplitude and duration, which was a good match to the data that we measured (Fig. 4.15).

Modeling dynamic variation using extrinsic noise

After decoupling the morphological change in the LPS response from the signaling response (described in Chapter 2), we observed little predictive effect of RelA abundance on downstream dynamics (Fig. 4.16). Indeed, contrary to previous hypotheses[96;109], increased RelA was actually associated, however weakly, with *decreased* late NF-κB activity. We chose, instead, to focus on modeling upstream sources of noise: recent information-theoretic work has suggested that the IκB-NF-κB network contributes only 10% of the total noise (or information loss) to the system response[79]. We focused, as noted previously, on TLR4 synthesis variation (which may also be associated with MD2), and MyD88/TRIF variation, which were held equal. These distributions were assumed to have log-normal distributions, with previously-parameterized mean values.

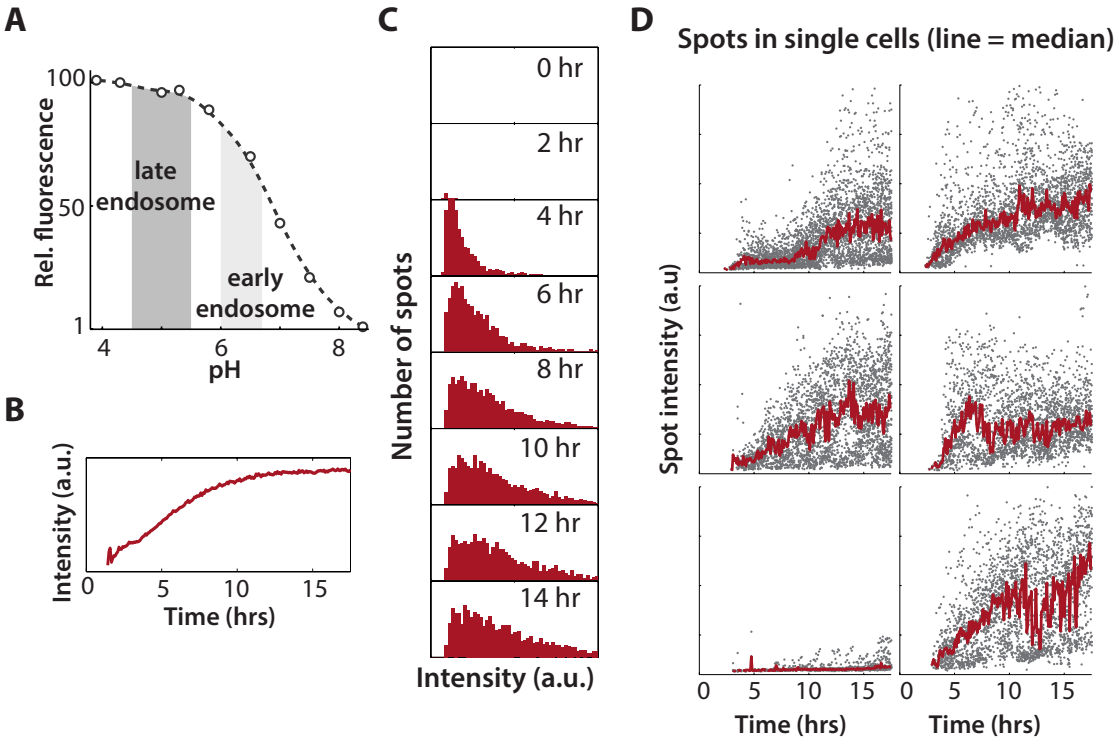

Figure 4.14: Measuring endosomal maturation. Fitting and modeling endosome maturation-induced degradation of TLR4 complexes. **(A)** Fluorescence intensity versus pH for the pHrodo dye. **(B)** Average spot intensity over time following phagocytosis of pHrodo-conjuaged *E.coli* in RAW 264.7 cells. **(C)** Histogram of all spot intensity distributions over time. Fluorescence is a function of local concentration and pH, eventually stabilizing to a log-normal distribution. **(D)** Spots are assigned to automatically-segmented cells, giving a single cell set of phosomes that evolve over time (red line: median. Gray dots: individual spots).

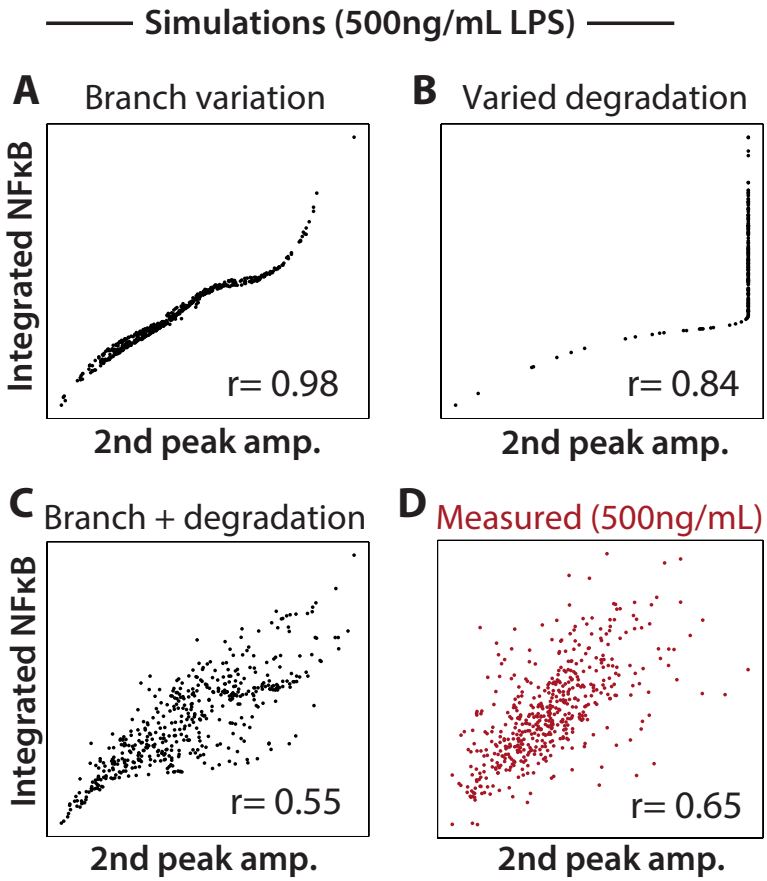

Figure 4.15: Modeling endosomal maturation in NF-κB dynamics. Modulation of total NF-κB activity in models incorporating protein variation and/or endosomal maturation variation. "Branch variation" refers to a model variant incorporating variable TRIF/MyD88 signaling levels only, and "Varied degradation" refers to a model variant incorporating only the endosomal maturation distribution fitted earlier.

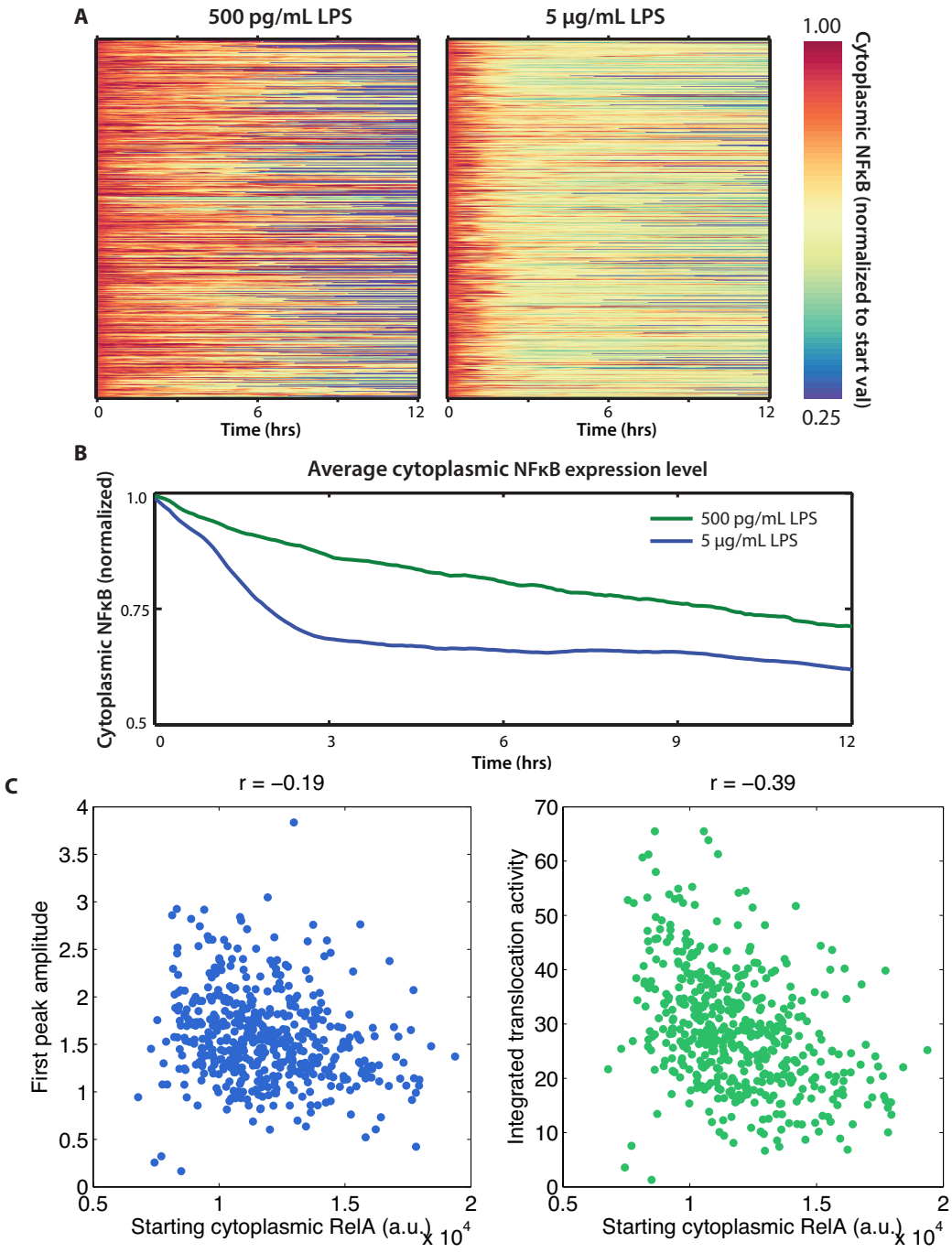

Figure 4.16: RelA abundance and dynamics. RelA expression is not correlated with NF-κB dynamics. **(A)** Heatmaps of single cell cytoplasmic RelA show a dose dependent (morphologically-related) decrease in apparent expression over time, contrary to previous reports (66). **(B)** Average decrease for all cells. **(C)** RelA expression is generally uncorrelated with NF-κB dynamics - r^2 values are generally between 0.05-0.16 for expression and dynamics, depending on the metric used.

Our scan, then, was restricted to two values.

In order to match branch- and receptor-associated variation to the variation observed in the measured single cell dynamics, we minimized the residuals left from a Kolmogorov-Smirnov test performed upon a distribution of NF-κB signal responses, which were compressed to single metrics (1st peak amplitude, 2nd peak amplitude, and total atcivity). The average residual, of distributions at 5 different doses of LPS spanning 0.5ng/mL to 5μg/mL LPS, was scaled from 0-1 (using the maximal and minimal residual observed for the search range of log-normal variabilities) to create a "fit score". Individual-metric fit scores, and averaged fit scores over all metrics, are shown in Fig. 4.17.

Chapter 4 contains material originally published as "Distinct Dynamical Characteristics Conferred by the MyD88 and TRIF Pathways in TLR4 Signaling". Cheng, Z.*, Taylor, B.*, Ourthiague, D., and Hoffmann, A. ***Science Signaling*** (2015 - in review). Copyright permission to republish here was granted by AAAS.

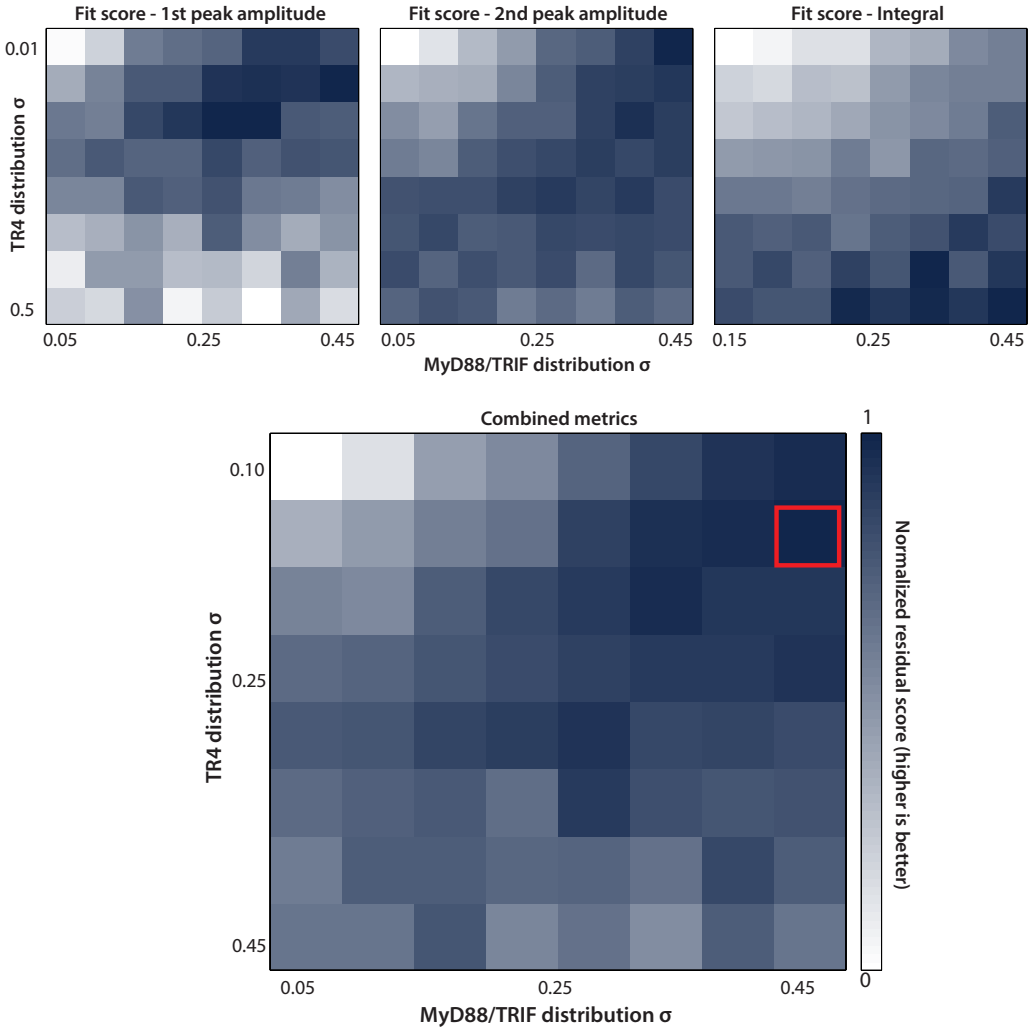

Figure 4.17: K-S scoring of modeled variabilities. Single cell simulations reflect quantitative metrics characterizing measured single cell responses. Scores between measured dynamics (quantified using several metrics) and simulations were calculated using residuals of a two-sample Kolmogorov-Smirnov test: each sample was the distribution, at a given dose, of simulation and measured behaviors. Test residuals were summed, inverted and normalized (such that a lower residual sum gives a score closer to 1) for all 5 doses (0.5ng/mL - 5 μg/mL) to give a total score for each set of input standard deviations. These scores were then combined and scaled to choose the best-fit standard deviation pair. (σ_{TLR4} = 0.15 and $\sigma_{MyD88/TRIF}$ = 0.45).

Chapter 5

Discussion and future directions

Predictive and mechanistic understanding of how individual NF-κB dynamics are assembled into a greater whole offers exciting potential. The potential in medicine, of moving beyond a crude, binary understanding of NF-κB activation in various forms of disease, is tantalizing: we could diagnose conditions, for instance, based on the dynamic responses of cells isolated from a patient. In pharmacology, we might move past simplistic targeting of network elements, and instead modulate dynamic responses, minimizing deleterious side effects[11]. However, much work remains to be done in understanding not only how NF-κB dynamics function in specific contexts, but also the biological basis for signal-specific responses. Below, we outline three future directions, which build on this work and move towards these goals.

1. **Application in primary cells**. Though the RAW264.7 cells used in these studies better approximate true innate immune cells, they are still strongly compromised. Simply put, they don't mount strong responses to many of the things that we know macrophages can respond to (e.g. TLR9 stimuli or TNF). Given this, we risk strongly underestimating the nature of "stimulus specificity". The lack of publications using the existing EGFP-p65 mouse[23] suggests there are technical limitations to measurement, but an improved mouse line would allow significant progress, both by using crosses with the extensive array of knockout

lines currently available to do sufficiency/requirement testing on network components, and the study of dynamics in better-approximated disease models, such as tissue-associated macrophages[74], or immunomodulated macrophages[62].

2. **Study of co-signaling dynamics**. While NF-κB has been studied in great detail, it is almost never active in isolation. Multiple families of transcription factors, such as AP-1, cJun, and STAT are often activated as part of the same cascade of events that triggers NF-κB . Any study seeking to extract information from NF-κB alone will necessarily be incomplete, as it may only be one piece in a larger, coherent whole. Fortunately, the analytical frameworks presented here, both for image analysis, and interpretation of data in an information context, would readily scale to the study of cells with multiple fluorescently-tagged proteins.

3. **Understanding decoding mechanisms**. We recognize that all our efforts so far only constitute one half of the information transmission puzzle. Measured single cell dynamics can teach about the network mechanisms by which information is encoded, but stop there: without coupling dynamics to downstream processes (be they behavioral decisions or gene expression), we cannot link the two. Several hypotheses have been put forward about how transcription factor dynamics might potentially be decoded[32;70], and technologies have recently become available to allow high-throughput transcriptome analysis[114;86] in single cells, but the next barrier to overcome will be gathering dynamics and results in the same mammalian cell.

Bibliography

[1] Akashi, S., Saitoh, S.-i., Wakabayashi, Y., Kikuchi, T., Takamura, N., Nagai, Y., Kusumoto, Y., Fukase, K., Kusumoto, S., Adachi, Y., Kosugi, A., and Miyake, K., 2003: Lipopolysaccharide interaction with cell surface Toll-like receptor 4-MD-2: higher affinity than that with MD-2 or CD14. *The Journal of experimental medicine*, **198**(7), 1035–1042.

[2] Albeck, J. G., Mills, G. B., and Brugge, J. S., 2013: Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals. *Molecular Cell*, **49**(2), 249–261.

[3] Almaden, J. V., Tsui, R., Liu, Y. C., Birnbaum, H., Shokhirev, M. N., Ngo, K. A., Davis-Turak, J. C., Otero, D., Basak, S., Rickert, R. C., and Hoffmann, A., 2014: A Pathway Switch Directs BAFF Signaling to Distinct NFκB Transcription Factors in Maturing and Proliferating B Cells. *Cell Reports*, **9**(6), 2098–2111.

[4] Altschuler, S. J., and Wu, L. F., 2010: Cellular Heterogeneity: Do Differences Make a Difference? *Cell*, **141**(4), 559–563.

[5] Ashall, L., Horton, C. A., Nelson, D. E., Paszek, P., Harper, C. V., Sillitoe, K., Ryan, S., Spiller, D. G., Unitt, J. F., Broomhead, D. S., Kell, D. B., Rand, D. A., Sée, V., and White, M. R. H., 2009: Timing and Specificity of NF-kB âĂŞ Dependent Transcription. *Science*, **324**(April), 242–246.

[6] Baker, R. G., Hayden, M. S., and Ghosh, S., 2011: NF-κB, inflammation, and metabolic disease. *Cell Metabolism*, **13**, 11–22.

[7] Barkai, N., and Leibler, S., 1997: Robustness in simple biochemical networks. *Nature*, **387**, 913–917.

[8] Barken, D., Wang, C. J., Kearns, J., Cheong, R., Hoffmann, A., and Levchenko, A., 2005: Comment on "Oscillations in NF-kappaB signaling control the dynamics of gene expression". *Science (New York, N.Y.)*, **308**(April), 52; author reply 52.

[9] Basak, S., Behar, M., and Hoffmann, A., 2012: Lessons from mathematically modeling the NF-κB pathway. *Immunological Reviews*, **246**, 221–238.

[10] Behar, M., and Hoffmann, A., 2010: Understanding the temporal codes of intracellular signals. *Current opinion in genetics & development*, **20**(6), 684–93.

[11] Behar, M., and Hoffmann, A., 2013: Tunable signal processing through a kinase control cycle: The IKK signaling node. *Biophysical Journal*, **105**(1), 231–241.

[12] Ben-Neriah, Y., and Karin, M., 2011: Inflammation meets cancer, with NF-κB as the matchmaker. *Nature immunology*, **12**(8), 715–723.

[13] Blander, J. M., and Medzhitov, R., 2004: Regulation of phagosome maturation by signals from toll-like receptors. *Science (New York, N.Y.)*, **304**(5673), 1014–8.

[14] Bonizzi, G., and Karin, M., 2004: The two NF-κB activation pathways and their role in innate and adaptive immunity. *Trends in Immunology*, **25**(6), 280–288.

[15] Bradhurst, C. J., Boles, W., and Xiao, Y., 2008: Segmentation of bone marrow stromal cells in phase contrast microscopy images. *2008 23rd International Conference Image and Vision Computing New Zealand*, 1–6.

[16] Caldwell, A. B., Cheng, Z., Vargas, J. D., Birnbaum, H. a., and Hoffmann, A., 2014: Network dynamics determine the autocrine and paracrine signaling functions of TNF. *Genes & Development*, 2120–2133.

[17] Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. a., Chang, J. H., Lindquist, R. a., Moffat, J., Golland, P., and Sabatini, D. M., 2006: CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, **7**(10), R100.

[18] Chan, Y. K., and Gack, M. U., 2015: RIG-I-like receptor regulation in virus infection and immunity. *Current Opinion in Virology*, **12**(Figure 1), 7–14.

[19] Cheong, R., Rhee, A., Wang, C. J., Nemenman, I., and Levchenko, A., 2011: Information transduction capacity of noisy biochemical signaling networks. *Science (New York, N.Y.)*, **334**(6054), 354–8.

[20] Collins, F., 2010: Has the revolution arrived? *Nature*, **464**(April 2010), 674–675.

[21] Cover, T., and Thomas, J., 2006: *Elements of Information Theory, 2nd Edition*. Wiley-Interscience.

[22] Covert, M. W., Leung, T. H., Gaston, J. E., and Baltimore, D., 2005: Achieving stability of lipopolysaccharide-induced NF-kappaB activation. *Science (New York, N.Y.)*, **309**(5742), 1854–7.

[23] De Lorenzi, R., Gareus, R., Fengler, S., and Pasparakis, M., 2009: GFP-p65 knock-in mice as a tool to study NF-kappaB dynamics in vivo. *Genesis (New York, N.Y. : 2000)*, **47**(5), 323–9.

[24] De Ronde, W. H., Tostevin, F., and Ten Wolde, P. R., 2010: Effect of feedback on the fidelity of information transmission of time-varying signals. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **82**(September).

[25] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S., 2002: Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, **297**(5584), 1183–6.

[26] Escoubet-Lozach, L., Benner, C., Kaikkonen, M. U., Lozach, J., Heinz, S., Spann, N. J., Crotti, A., Stender, J., Ghisletti, S., Reichart, D., Cheng, C. S., Luna, R., Ludka, C., Sasik, R., Garcia-Bassets, I., Hoffmann, A., Subramaniam, S., Hardiman, G., Rosenfeld, M. G., and Glass, C. K., 2011: Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. *PLoS genetics*, **7**(12), e1002401.

[27] Ferrao, R., Li, J., Bergamin, E., and Wu, H., 2012: Structural insights into the assembly of large oligomeric signalosomes in the Toll-like receptor-interleukin-1 receptor superfamily. *Science signaling*, **5**(226), re3.

[28] Gagner, J. A., Ulevitch, R. J., and Tobias, P. S., 1995: Lipopolysaccharide(LPS) Signal Transduction and Clearance. *Journal of Biological Chemistry*, **270**(10), 5320–5325.

[29] Gilmore, T. D., 2006: Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene*, **25**, 6680–6684.

[30] Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., and Shet, V., 2013: Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. *CoRR*, **abs/1312.6**, 1–13.

[31] Gutschow, M. V., Hughey, J. J., Ruggero, N. a., Bajar, B. T., Valle, S. D., and Covert, M. W., 2013: Single-cell and population NF-κB dynamic responses depend on lipopolysaccharide preparation. *PloS one*, **8**(1), e53222.

[32] Hansen, A. S., and O'Shea, E. K., 2013: Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Molecular systems biology*, **9**(704), 704.

[33] Hao, N., Budnik, B. a., Gunawardena, J., and O'Shea, E. K., 2013: Tunable signal processing through modular control of transcription factor translocation. *Science (New York, N.Y.)*, **339**(6118), 460–4.

[34] Hao, S., and Baltimore, D., 2009: The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nature immunology*, **10**(3), 281–8.

[35] Hayden, M. S., and Ghosh, S., 2008: Shared principles in NF-kappaB signaling. *Cell*, **132**(3), 344–62.

[36] Hodneland, E., Kögel, T., Frei, D. M., Gerdes, H.-H., and Lundervold, A., 2013: CellSegm - a MATLAB toolbox for high-throughput 3D cell segmentation. *Source code for biology and medicine*, **8**(1), 16.

[37] Hoffmann, A., Levchenko, A., Scott, M. L., and Baltimore, D., 2002: The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science (New York, N.Y.)*, **298**(December), 1241–1245.

[38] Hughey, J. J., Gutschow, M. V., Bajar, B. T., and Covert, M. W., 2014: Single-cell variation leads to population invariance in NF-κB signaling dynamics. *Molecular Biology of the Cell*, **26**, 583–590.

[39] Husebye, H., Aune, M. H., Stenvik, J. r., Samstad, E., Skjeldal, F., Halaas, O., Nilsen, N. J., Stenmark, H., Latz, E., Lien, E., Mollnes, T. E., Bakke, O., and Espevik, T., 2010: The Rab11a GTPase controls Toll-like receptor 4-induced activation of interferon regulatory factor-3 on phagosomes. *Immunity*, **33**(4), 583–96.

[40] Husebye, H., Halaas, O. y., Stenmark, H., Tunheim, G., Sandanger, O. y., Bogen, B., Brech, A., Latz, E., and Espevik, T., 2006: Endocytic pathways regulate Toll-like receptor 4 signaling and link innate and adaptive immunity. *The EMBO journal*, **25**(4), 683–92.

[41] Jones, T. R., Carpenter, A., and Golland, P., 2005: Voronoi-Based Segmentation of Cells on Image Manifolds. *Computer Vision for Biomedical Image Application*.

[42] Kagan, J. C., Su, T., Horng, T., Chow, A., Akira, S., and Medzhitov, R., 2008: TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nature immunology*, **9**(4), 361–8.

[43] Kaikkonen, M. U., Spann, N. J., Heinz, S., Romanoski, C. E., Allison, K. a., Stender, J. D., Chun, H. B., Tough, D. F., Prinjha, R. K., Benner, C., and Glass, C. K., 2013: Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell*, **51**(3), 310–25.

[44] Kalita, M. K., Sargsyan, K., Tian, B., Paulucci-Holthauzen, A., Najm, H. N., Debusschere, B. J., and Brasier, A. R., 2011: Sources of cell-to-cell variability in canonical nuclear factor-κB (NF-κB) signaling pathway inferred from single cell dynamic images. *The Journal of biological chemistry*, **286**(43), 37741–57.

[45] Karin, M., and Greten, F. R., 2005: NF-kappaB: linking inflammation and immunity to cancer development and progression. *Nature reviews. Immunology*, **5**(10), 749–59.

[46] Kawai, T., Adachi, O., Ogawa, T., Takeda, K., and Akira, S., 1999: Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity*, **11**(1), 115–22.

[47] Kawai, T., and Akira, S., 2010: The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature immunology*, **11**(5), 373–84.

[48] Kawai, T., Takeuchi, O., Fujita, T., Inoue, J., Mühlradt, P. F., Sato, S., Hoshino, K., and Akira, S., 2001: Lipopolysaccharide stimulates the MyD88-independent pathway and results in activation of IFN-regulatory factor 3 and the expression of a subset of lipopolysaccharide-inducible genes. *Journal of immunology (Baltimore, Md. : 1950)*, **167**(10), 5887–94.

[49] Kearns, J. D., Basak, S., Werner, S. L., Huang, C. S., and Hoffmann, A., 2006: IκBα provides negative feedback to control NF-κB oscillations, signaling dynamics, and inflammatory gene expression. *Journal of Cell Biology*, **173**(5), 659–664.

[50] Kellogg, R. a., Gómez-Sjöberg, R., Leyrat, A. a., and Tay, S., 2014: High-throughput microfluidic single-cell analysis pipeline for studies of signaling dynamics. *Nature protocols*, **9**(7), 1713–26.

[51] Lavin, Y., Winter, D., Blecher-gonen, R., David, E., Keren-shaul, H., Merad, M., Jung, S., and Amit, I., 2014: Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment. *Cell*, **159**(6), 1312–1326.

[52] Lee, R. E. C., Walker, S. R., Savery, K., Frank, D. a., and Gaudet, S., 2014: Fold Change of Nuclear NF-κB Determines TNF-Induced Transcription in Single Cells. *Molecular cell*, 1–13.

[53] Lee, T. K., Denny, E. M., Sanghvi, J. C., Gaston, J. E., Maynard, N. D., Hughey, J. J., and Covert, M. W., 2009: A noisy paracrine signal determines the cellular NF-kappaB response to lipopolysaccharide. *Science signaling*, **2**(93), ra65.

[54] Lentschat, a., El-Samalouti, V. T., Schletter, J., Kusumoto, S., Brade, L., Rietschel, E. T., Gerdes, J., Ernst, M., Flad, H., and Ulmer, a. J., 1999: The internalization time course of a given lipopolysaccharide chemotype does not correspond to its activation kinetics in monocytes. *Infection and immunity*, **67**(5), 2515–21.

[55] Lin, S.-C., Lo, Y.-C., and Wu, H., 2010: Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. *Nature*, **465**(7300), 885–90.

[56] Loftsgaarden, D. O., and Quesenberry, C. P., 1965: A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, **36**(3), 1049–1051.

[57] Longo, D. M., Selimkhanov, J., Kearns, J. D., Hasty, J., Hoffmann, A., and Tsimring, L. S., 2013: Dual delayed feedback provides sensitivity and robustness to the NF-κB signaling module. *PLoS computational biology*, **9**(6), e1003112.

[58] Loriaux, P. M., Tesler, G., and Hoffmann, A., 2013: Characterizing the relationship between steady state and response using analytical expressions for the steady states of mass action models. *PLoS computational biology*, **9**(2), e1002901.

[59] Mantovani, A., Sica, A., Sozzani, S., Allavena, P., Vecchi, A., and Locati, M., 2004: The chemokine system in diverse forms of macrophage activation and polarization. *Trends in immunology*, **25**(12), 677–86.

[60] McGettrick, A. F., and O'Neill, L. A., 2010: Localisation and trafficking of Toll-like receptors: an important mode of regulation. *Current Opinion in Immunology*, **22**(1), 20–27.

[61] Medzhitov, R., 2008: Origin and physiological roles of inflammation. *Nature*, **454**(July), 428–435.

[62] Mills, C. D., and Ley, K., 2014: M1 and M2 Macrophages: The Chicken and the Egg of Immunity. *Journal of innate immunity*, 716–726.

[63] Mualla, F., Scholl, S., Sommerfeldt, B., Maier, A., and Hornegger, J., 2013: Automatic cell detection in bright-field microscope images using sift, random forests, and hierarchical clustering. *IEEE Transactions on Medical Imaging*, **32**(12), 2274–2286.

[64] Murphy, K., Travers, P., Walport, M., and Janeway, C., 2012): *Janeway's immunobiology, 8th edn*. Garland Science.

[65] Natoli, G., 2012: NF-κB and chromatin: Ten years on the path from basic mechanisms to candidate drugs. *Immunological Reviews*, **246**, 183–192.

[66] Nelson, D. E., Ihekwaba, a. E. C., Elliott, M., Johnson, J. R., Gibney, C. a., Foreman, B. E., Nelson, G., See, V., Horton, C. a., Spiller, D. G., Edwards, S. W., McDowell, H. P., Unitt, J. F., Sullivan, E., Grimley, R., Benson, N., Broomhead, D., Kell, D. B., and White, M. R. H., 2004: Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science (New York, N.Y.)*, **306**(5696), 704–8.

[67] Nilsson, R., Bajic, V. B., Suzuki, H., di Bernardo, D., Björkegren, J., Katayama, S., Reid, J. F., Sweet, M. J., Gariboldi, M., Carninci, P., Hayashizaki, Y., Hume, D. a., Tegner, J., and Ravasi, T., 2006: Transcriptional network dynamics in macrophage activation. *Genomics*, **88**(2), 133–42.

[68] Oda, K., and Kitano, H., 2006: A comprehensive map of the toll-like receptor signaling network. *Molecular systems biology*, **2**, 2006.0015.

[69] O'Dea, E., and Hoffmann, A., 2010: The regulatory logic of the NF-kappaB signaling system. *Cold Spring Harbor perspectives in biology*, **2**(1), a000216.

[70] Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G., 2013: Latent enhancers activated by stimulation in differentiated cells. *Cell*, **152**(1-2), 157–71.

[71] Otsu, N., 1979: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-9**(1), 62–66.

[72] Park, B. S., Song, D. H., Kim, H. M., Choi, B.-S., Lee, H., and Lee, J.-O., 2009: The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature*, **458**(7242), 1191–1195.

[73] Paszek, P., Ryan, S., Ashall, L., Sillitoe, K., Harper, C. V. C., Spiller, D. G. D., Rand, D. D. a., and White, M. R. H. M., 2010: Population robustness arising from cellular heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(25), 11644–9.

[74] Pollard, J. W., 2009: Trophic macrophages in development and disease. *Nature reviews. Immunology*, **9**(4), 259–70.

[75] Purvis, J. E., Karhohs, K. W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G., 2012: p53 Dynamics Control Cell Fate. *Science*, **336**(June), 1440–1444.

[76] Purvis, J. E., and Lahav, G., 2013: Encoding and decoding cellular information through signaling dynamics. *Cell*, **152**(5), 945–56.

[77] Rand, U., Rinas, M., Schwerk, J., Nöhren, G., Linnes, M., Kröger, A., Flossdorf, M., Kály-Kullai, K., Hauser, H., Höfer, T., and Köster, M., 2012: Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular systems biology*, **8**(584), 584.

[78] Rhee, A., Cheong, R., and Levchenko, A., 2012: The application of information theory to biochemical signaling systems. *Physical Biology*, **9**, 045011.

[79] Rhee, A., Cheong, R., and Levchenko, A., 2014: Noise decomposition of intracellular biochemical signaling networks using nonequivalent reporters. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(48).

[80] Schmitt, O., and Reetz, S., 2008: On the Decomposition of Cell Clusters. *Journal of Mathematical Imaging and Vision*, **33**(1), 85–103.

[81] Schreiber, T., 1999: Interdisciplinary application of nonlinear time series methods. *Physics Reports*, **308**(May 1998), 1–64.

[82] Scott, J. D., and Pawson, T., 2009: Cell signaling in space and time: where proteins come together and when they're apart. *Science (New York, N.Y.)*, **326**(November), 1220–1224.

[83] Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R., 2014: Accurate information transmission through dynamic biochemical signaling networks. *Science*, **346**(6215), 1370–1373.

[84] Selinummi, J., Ruusuvuori, P., Podolsky, I., Ozinsky, A., Gold, E., Yli-Harja, O., Aderem, A., and Shmulevich, I., 2009: Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images. *PloS one*, **4**(10), e7497.

[85] Sezgin, M., and Sankur, B., 2004: Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, **13**(1), 220.

[86] Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P., Gertner, R. S., Gaublomme, J. T., Yosef, N., Schwartz, S., Fowler, B., Weaver, S., Wang, J., Wang, X., Ding, R., Raychowdhury, R., Friedman, N., Hacohen, N., Park, H., May, A. P., and Regev, A., 2014: Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, **509**(7505), 363–9.

[87] Sheff, D. R., Daro, E. A., Hull, M., and Mellman, I., 1999: The receptor recycling pathway contains two distinct populations of early endosomes with different sorting functions. *The Journal of cell biology*, **145**(1), 123–39.

[88] Shih, V. F.-S., Kearns, J. D., Basak, S., Savinova, O. V., Ghosh, G., and Hoffmann, A., 2009: Kinetic control of negative feedback regulators of NF-kappaB/RelA determines their pathogen- and cytokine-receptor signaling specificity. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(24), 9619–24.

[89] Shih, V. F.-S., Tsui, R., Caldwell, A., and Hoffmann, A., 2010: A single NFκB system for both canonical and non-canonical signaling. *Cell research*, **21**(1), 86–102.

[90] Shinohara, H., Behar, M., Inoue, K., Hiroshima, M., Yasuda, T., Nagashima, T., Kimura, S., Sanjo, H., Maeda, S., Yumoto, N., Ki, S., Akira, S., Sako, Y., Hoffmann, A., Kurosaki, T., and Okada-Hatakeyama, M., 2014: Positive feedback within a kinase signaling complex functions as a switch mechanism for NF-κB activation. *Science (New York, N.Y.)*, **344**(6185), 760–4.

[91] Smale, S. T., 2012: Dimer-specific regulatory mechanisms within the NF-kappaB family of transcription factors. *Immunological reviews*, **246**, 193–204.

[92] Snijder, B., and Pelkmans, L., 2011: Origins of regulated cell-to-cell variability. *Nature reviews. Molecular cell biology*, **12**(2), 119–125.

[93] Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., and Sorger, P. K., 2009: Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, **459**(7245), 428–32.

[94] Steinman, R., and Mellman, I., 1983: Endocytosis and the recycling of plasma membrane. *The Journal of cell . . .* , 1–27.

[95] Sturm, O. E., Orton, R., Grindlay, J., Birtwistle, M., Vyshemirsky, V., Gilbert, D., Calder, M., Pitt, A., Kholodenko, B., and Kolch, W., 2010: The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Science signaling*, **3**(153), ra90.

[96] Sung, M.-H., Li, N., Lao, Q., Gottschalk, R. a., Hager, G. L., and Fraser, I. D. C., 2014: Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF-κB signaling. *Science signaling*, **7**(308), ra6.

[97] Sung, M.-H., Salvatore, L., De Lorenzi, R., Indrawan, A., Pasparakis, M., Hager, G. L., Bianchi, M. E., and Agresti, A., 2009: Sustained oscillations of NF-kappaB produce distinct genome scanning and gene expression profiles. *PloS one*, **4**(9), e7163.

[98] Taguchi, T., 2013: Emerging roles of recycling endosomes. *Journal of biochemistry*, **153**(6), 505–10.

[99] Tay, S., Hughey, J. J., Lee, T. K., Lipniacki, T., Quake, S. R., and Covert, M. W., 2010: Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature*, **466**(7303), 267–71.

[100] Tobias, P. S., Soldau, K., Gegner, J. a., Mintz, D., and Ulevitch, R. J., 1995: Lipopolysaccharide binding protein-mediated complexation of lipopolysaccharide with soluble CD14.

[101] Toettcher, J. E., Weiner, O. D., and Lim, W. a., 2013: Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell*, **155**(6), 1422–1434.

[102] Tostevin, F., and Ten Wolde, P. R., 2009: Mutual information between input and output trajectories of biochemical networks. *Physical Review Letters*, **102**(May), 1–4.

[103] Tsai, D.-m., 1995: A Fast Thresholding Selection Procedure for Multimodal and Unimodal Histograms. *Pattern Recognition Letters*, **16**(June), 653–666.

[104] Tsimring, L. S., 2014: Noise in biology. *Reports on Progress in Physics*, **77**, 026601.

[105] Turner, D. a., Paszek, P., Woodcock, D. J., Nelson, D. E., Horton, C. a., Wang, Y., Spiller, D. G., Rand, D. a., White, M. R. H., and Harper, C. V., 2010: Physiological levels of TNFalpha stimulation induce stochastic dynamics of NF-kappaB responses in single living cells. *Journal of cell science*, **123**(Pt 16), 2834–43.

[106] Uda, S., Saito, T. H., Kudo, T., Kokaji, T., Tsuchiya, T., Kubota, H., Komori, Y., Ozaki, Y.-i., and Kuroda, S., 2013: Robustness and compensation of information transmission of signaling pathways. *Science (New York, N.Y.)*, **341**(August), 558–61.

[107] Wang, W., 2007: Study on Cell Cluster Splitting. *Fourth International Conference on Image and Graphics (ICIG 2007)*, 427–431.

[108] Wang, Y., Chen, T., Han, C., He, D., Liu, H., An, H., Cai, Z., and Cao, X., 2007: Lysosome-associated small Rab GTPase Rab7b negatively regulates TLR4 signaling in macrophages by promoting lysosomal degradation of TLR4. *Blood*, **110**(3), 962–71.

[109] Wang, Y., Paszek, P., Horton, C. a., Yue, H., White, M. R. H., Kell, D. B., Muldoon, M. R., and Broomhead, D. S., 2012: A systematic survey of the response of a model NF-κB signalling pathway to TNFα stimulation. *Journal of theoretical biology*, **297**, 137–47.

[110] Werner, S. L., Barken, D., and Hoffmann, A., 2005: Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science (New York, N.Y.)*, **309**(5742), 1857–61.

[111] Werner, S. L., Kearns, J. D., Zadorozhnaya, V., Lynch, C., O'Dea, E., Boldin, M. P., Ma, A., Baltimore, D., and Hoffmann, A., 2008: Encoding NF-kappaB temporal control in response to TNF: distinct roles for the negative regulators IkappaBalpha and A20. *Genes & development*, **22**(15), 2093–101.

[112] Wilkinson, D. J., 2009: Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature reviews. Genetics*, **10**(fEbruAry), 122–133.

[113] Williams, R., Timmis, J., and Qwarnstrom, E., 2014: Computational Models of the NF-KB Signalling Pathway. *Computation*, **2**(4), 131–158.

[114] Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F., and Quake, S. R., 2014: Quantitative assessment of single-cell RNA-sequencing methods. *Nature methods*, **11**(1), 41–6.

[115] Wu, D., and Zhang, Q., 2007: A Novel Approach for Cell Segmentation Based on Directional Information. *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, 920–923.

[116] Wynn, T. a., Chawla, A., and Pollard, J. W., 2013: Macrophage biology in development, homeostasis and disease. *Nature*, **496**(7446), 445–455.

[117] Yamamoto, M., Sato, S., and Hemmi, H., 2003: Role of Adaptor TRIF in the MyD88-Independent Toll-like Receptor Signaling Pathway. *Science*, **301**(August), 640–643.

[118] Yu, R. C., Pesce, C. G., Colman-Lerner, A., Lok, L., Pincus, D., Serra, E., Holl, M., Benjamin, K., Gordon, A., and Brent, R., 2008: Negative feedback that improves information transmission in yeast signalling. *Nature*, **456**(December), 755–761.

[119] Zanoni, I., Ostuni, R., Marek, L. R., Barresi, S., Barbalat, R., Barton, G. M., Granucci, F., and Kagan, J. C., 2011: CD14 controls the LPS-induced endocytosis of Toll-like receptor 4. *Cell*, **147**(4), 868–80.

[120] Zaritsky, A., Natan, S., Horev, J., Hecht, I., Wolf, L., Ben-Jacob, E., and Tsarfaty, I., 2011: Cell motility dynamics: a novel segmentation algorithm to quantify multi-cellular bright field microscopy images. *PloS one*, **6**(11), e27593.