THE UNIVERSITY OF CHICAGO

SINGLE CELL ANALYSIS STRATEGIES TO UNDERSTAND NOISY AND DYNAMIC NF-κB NETWORK BEHAVIOR

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE PRITZKER SCHOOL OF MOLECULAR ENGINEERING

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

PARTHIV PATEL

CHICAGO, ILLINOIS

DECEMBER 2021

Table of Contents

Table of Figures

Computer-vision reveals hidden variables underlying NF-κB activation in single cells

Environmental noise enables sensitive detection and transcriptional decoding of cytokine inputs

Abstract

One of the major open questions in biology today is how cells operate robustly despite the presence of biological noise in their environment. Signaling pathways are often depicted as having defined responses to environmental signals, and their protein cascades often are shown to have specific responses to doses, timings, and signaling behavior(*1*, *2*). One example of this robust signaling is the NF-κB immune pathway which controls responses to many pathogens and cytokines. In healthy organisms, NF-κB provides a graduated response based on dose that enables its' subsequent clearance efficiently. In tissue, signaling molecules often have fluctuating concentrations(*1*, *3*, *4*), yet are able to still create these robust gene expression profiles(*5*–*7*). Inputs received by cells are unavoidably noisy due to variable secretion and translation of signaling factors, propagation through tissues, and fluctuations in molecular concentrations(*4*, *8*, *9*). The molecular mechanisms behind this robust signaling despite presence of input noise are not fully understood.

Understanding the effect of network noise on signaling is an active area of research, however little attention has been given to the single cell specificity and sensitivity under noise. It is clear that noise is abundant in the input to NF-κB – after detecting pathogens immune cells secrete a wide range of cytokines that serve as input to nearby tissue cells. The input received by any single cell thus depends on the dynamics of cytokine secretion and is subject to the noise in the extracellular environment. Here, I take a microfluidic approach to understand how and why there is heterogeneity in activation of NF-κB, how noise in the input affects NF-κB signaling and transcription, and what mechanism the cells use to process and filter this noise.

Introduction

Individual cells often show highly variable and heterogeneous responses in a wide range of contexts, from immune signaling to transcriptional activation and to drug response(*10*, *11*). Following stimulation with signaling molecules, in pathways such as NF-κB, a portion of cells in a population will strongly activate, while others will completely ignore the stimulus and will not activate(*12*). Understanding the source of variable specificity in NF-κB activation is of immense importance to fundamentally understanding gene regulation, signaling, immunity, and development, as well as in predicting variable drug response and tolerance. Understanding how this important signaling network behaves can inform behavior in many adjacent biological pathways.

Understanding the effect of network noise on signaling is an active area of research and while some of signaling variability can be accounted for by internal fluctuations in protein concentrations, little attention has been given to noise in the input itself. This in turn sets up two modes of noise that can affect the dynamics and activation of a pathway. The first is the variation in cell state that gives rise to intrinsic noise. The second is the external noise in an input that can engage the biological pathway. Here we explore the impact of NF-κB sensitivity through variation in intrinsic noise, and the impact of NF-κB specificity through perturbation by external noise.

Despite previous demonstration of heterogeneous signaling responses in a wide range of contexts like development, immune response, infection and cancer, it remains difficult to explore the molecular and cellular mechanisms that drive variable behavior in single cells and the underlying source of individual cell specificity are unknown in many contexts. This is because

once a cellular state is perturbed by a signal, that state ceases to exist. To also understand the sensitivity of cells to environmental noise, it becomes imperative to probe response to fluctuating environments as well. It is clear that noise is abundant in the input to NF-κB – after sensing pathogens, immune cells can secrete a wide range of cytokines that serve as input to nearby tissue cells(*13–15*). The input received by any single cell thus depends on the dynamics of cytokine secretion and is subject to the noise in the extracellular environment. How cells deal with noisy external fluctuations and create appropriate signaling responses is not well understood, and it is central to many signaling problems in health and disease.

The well curated responses generated by cell populations despite noise in the environment and single cell variability(*16*, *17*) may be the result of noise dampening features of gene regulatory network motifs(*18–20*). These network motifs can potentially reduce the detrimental effects of noise, however, these features inevitably reduce the responsiveness (sensitivity), speed and temporal resolution of signaling systems. Many cellular signaling pathways manage to combine high sensitivity and fast speed with noise tolerance, which are seemingly conflicting properties to exist in the same signaling system. To understand these properties, we first look at the general structure and behavior of the NF-κB pathway

NF-κB Canonical Pathway Dynamics

Transcription factor dynamics in gene network regulation is an active area of research and NF-κB is a historically well studied example of a dynamically regulated gene. Previous studies within the Tay Group and elsewhere have shown that oscillations in the localization of NF-κB transcription factors like p65 encode the type and timing of the downstream response.

Hundreds of genes use NF-κB through a DNA sequence with a consensus of 5'-GGGRNWYYCC-3' (R=purine, Y=pyrimidine, W= A or T, N=any) as an enhancer and promotor activation target. The site binds dimers of proteins that contain Rel A (p65), Rel B, p50, p52 and c-rel. Of the 15 combinations, 13 have been demonstrated and the remaining two have yet to be described(*21*, *22*). Rel-A, Rel-B and c-Rel are transcriptional activators, while p50 and p52 homodimers are repressors unless bound with secondary proteins. As a result of this variability many regulatory patterns can be established with these handful of proteins.

The full network and subnetworks of proteins in the NF-κB cascade is complex and has many potential activators and regulators. We look specifically at the canonical pathway of NF-κB. Cells in the resting steady state contain cytoplasmic NF-κB proteins. The dedicated inhibitor and transporter of these proteins is the IκB-family molecules which shuttle the NF-κB proteins from nucleus to cytoplasm(*23*). The canonical p65/p50 heterodimers is regulated primarily by IκBα. As a result of this bottlenecked regulation, IκB becomes a control point for the expression of downstream genes and acts as a loaded spring for the downstream translocation and activation of inflammatory genes. The canonical NF-κB activation is thus reliant on the modification and inhibition of the IκB inhibitors. While this can happen in many ways, in the canonical pathway a complex called IκB Kinase (IKK) phosphorylates IκB proteins leading to their degradation(*24*, *25*). This releases NF-κB dimers from IκB and allow them to actively shuttle and remain in the nucleus.

IKK is canonically composed of three subunits: two kinases, IKKα and iKKβ and an NF-κB-Essential-Modifier (NEMO) which acts as a regulatory subunit that tethers the two other IKK proteins into a complex. The activation of this complex is tied to TNF stimulation and is activated

in a well-understood process(*26*, *27*). TNF stimulation recruits TRADD, RIP-1, cIAP, TRAF2, TAB and TAK1 into a megacomplex with the linear ubiquitin assembly complex stabilized by linear and k63-linked polyubiquitin chains. When IKK is incorporated into the megacomplex, NEMO is ubiquitinated and phosphorylation of IKK resulting in kinase activity induction(*28*, *29*). The composition of these megacomplexes depend on ubiquitin modifications and in turn result in different paths and timings of IKK and NF-κB activation.

The enormous amount of genes activated by NF-κB proteins and their dynamic activation modalities give this network a diverse and complex system of activation resulting in potentially many phenotypes and downstream cell behaviors(*30*).

Improper signal processing has been linked to many conditions including chronic inflammation, infection and cancer. Loss of function studies alongside biochemical and systems biology based studies have elucidated the NF-κB network components and the helped the field develop mathematical models to understand and predict the behavior of the pathway(*12*, *31*). Despite this, the significant heterogeneity and noise in the cellular environment make it difficult to reach an accurate and precise answer to many questions that rely on quantitative signaling behavior.

Understanding Dynamics Inputs in NF-κB

When cells are stimulated with signaling molecules (i.e. TNF or LPS), some of them show complete cytoplasm-to-nucleus translocation of the p65 transcription factor (the hallmark of pathway activation), while others ignore the stimulus and show no translocation and no NF-κB target gene expression(*12*). The fraction of cells that respond to a stimulus (signaling input)

increases according to dose and despite this seemingly heterogeneous activation, NF-κB nevertheless manages to mount specific responses to different signaling molecules, taking into account their concentrations and temporal ordering(*6*, *32*).

Through prior studies of NF-κB under time-varying inputs it has been shown that periodic stimulation entrains the NF-κB oscillations in single cells and in the population(*33*). This creates synchronous oscillations with dramatically reduced variability and significantly increased target mRNA expression. By contrast, with a constant stimulation, the behavior is very varied and as a result preserves some heterogeneity in the population. This underscores the heterogeneous input dynamics in the signaling pathway; timing, amplitude and duration of stimulation all have important roles in encoding NF-κB oscillations. Short-pulsed stimulation revealed that NF-κB integrates TNF area-under-curve (AUC). Cellular microenvironments are highly dynamic and present time-varying signals for time varying inputs.

There has been much research into the underlying mechanisms driving NF-κB response under dynamic signaling inputs under different ligands and with different time-varying stimuli. Recently there has been work into the general response patterns of NF-κB. It has been shown that there NF-κB responds to absolute differences in cytokine concentrations. This may be partially responsible for the differences in expression levels in time-varying signals. Recent work in NF-κB also showed comprehensive gene expression changes that are responsible for changing responsiveness to signals after the initial signaling cascade. It is likely that epigenetic changes and transcription factor mobilization after the initial stimulation is responsible for these developed and evolved adaptative responses. These interesting insights pave the way to help us understand how noise changes the NF-κB network and leads to downstream responsiveness and adaptability.

It was also recently found that intrinsic noise in the TNF receptor and IκB transcription broadens entrainment regions in response to periodic stimulation. In this case the intrinsic noise improves information processing creating more robust entrainment and oscillation quality. This observation led to the discovery that NF-κB can jump between two stable oscillation modes, termed "mode-hopping"

Identification of NF-κB network components that deal with noise

NF-κB is a thoroughly explored pathways and many of the key network components have been identified; this has allowed the community to make a comprehensive in-silico model. This has facilitated the identification of important regulatory molecules and network motifs that deal with noise. These fundamental findings pave the way for the hypotheses presented in the following sections.

One such noise regulator is the IKK amplification cascade between receptor and NF-κB; this process leads to the digital activation of the pathways under the canonical stimulation. There are many extensions to the canonical modality that could increase the efficiency of the cascade such as positive feedback from TRIF-mediated response(*14*). Digital activation is an efficient strategy to deal with noise in transmission; an all or none response makes it easy to determine activated cells and can give a graduated response based on input dosage.

The roles of repression complexes such as the TNFAIP3 complex is thought to act through negative feedback on the IKK cascade and thus is crucial to the long-term responsiveness to TNF(*34*). IκB is the well-known mechanism that creates the oscillations in the response and thus

may have a key characteristic as a noise regulator. Negative feedback here eliminates long-residence times of NF-κB in the nucleus and limits exaggerated responses in the population.

While NF-κB is well prepared with noise reducing mechanism, through simulations, we find that IκB and IKK driven oscillations allow NF-κB to take advantage of cellular intrinsic and extrinsic noise and encode and decode periodic signals through entrainment and noisy inputs through noise adaptive behaviors. This postulate will be experimentally investigated in this thesis. These results provide several directions for further study and will help us thoroughly examine the molecular mechanisms behind noise tolerance and usage by NF-κB. The experimentally examine and perturb live cells to understand the system, we use a microfluidic setup to precisely engage cells and track them over the course of a treatment.

Microfluidic Methods

The use of lab-on-chip tools has been adopted in a wide variety of scientific fields and hundreds of applications that increase the speed of, miniaturize, or enable previously unfeasible assays have emerged in the last decade. The microfluidic toolbox offers many advantages that make it a very attractive resource for biological study: reduced sample volume, streamlined assays, precise and predictable fluid flow, control of spatiotemporal chemical gradients, and integration with sensors, controllers, and automation systems. The main avenues of exploration in this field has thus far honed in on the development of tools that replace conventional methods with proof-of-principle applications.

Microfluidics takes advantage of fluid physics at small scales. One of the most important phenomena that dominates at the micro-scale is laminar flow and as a result the surface effects it

facilitates. Laminar flow results in predictable fluid behavior and minimal mixing between fluid streams. Manipulation of fluids at the micro-scale are precise and reproducible. These characteristics are useful to not only control spatiotemporal chemical gradients but also to manipulate samples via the fluid forces. Operating in a laminar flow regime generally implies that mixing of fluids requires special consideration. Some assays require the incorporation of channel geometries that enable efficient fluid mixing within a desired residence time. Surface effects arise when the high surface area to volume ratio of fluids constrained at the micro scale exist.

Microfluidics has enabled us to cover large distances and fill many gaps in knowledge by combining the mathematical modelling that helped bring this network into the light with the use of microfluidics to enable unique dynamic signal propagation and single cell analysis. Prior work has been done to develop and utilize high-throughput experimental procedures based on microfluidic cell culture devices. These devices are tailored to culture and control cells and follow NF-κB dynamics to allow for subsequent gene expression analysis.

The microfluidic chip that we plan to use in this experiment enables dynamic cell stimulation across the time domain at a range of concentrations. Stimulating cells with a function generator can be a powerful technique for probing gene networks and for understanding signaling pathways. The chip that we will use generates a wide-range of precisely formulated dynamic chemical signals at a high sampling rate for use in stimulating live cultured cells and to measure their dynamic responses. The chip uses a combinatorial selection of discrete input concentrations to create a digital to analog conversion of a time domain function. The chip has been validated and can create different waveforms and noise features to help in the study of stochasticity in cellular processing.

Cells were imaged in a microfluidic chip mounted on a Nikon Eclipse Ti Microscope to record fluorescence time lapse movies of p65-dsRed and H2B-GFP via a Retiga-SRV CCD camera (QImaging) in 5 min time intervals. After illumination correction the raw data was segmented and tracked through the time lapse movie by a, imaging pipeline in MATLAB.

Objectives

We will use microfluidic design principles and bioengineering approaches to interrogate the implications of noise in NF-κB; the plan of approach is outlined in the following aims:

Aim 1: Understand the effect of cellular heterogeneity on information processing in NF-κB specificity.

The NF-κB system can identify distinct inputs and mount an input-specific response. This information is encoded by the dynamics of the signaling cascade and decoded by the cell's transcriptional machinery. We will test how fluctuations in the internal cellular environment affect the specificity of NF-κB response. This will be done using a different computer vision classifiers that can predict activated vs non-activating cells. To answer questions about why some cells activate, we use a machine learning algorithm to train a classifier, then apply the classifier to unstimulated cells. We can then collect classified unstimulated cells to perform downstream analysis through methods such as immunofluorescent staining.

Aim 2: Understand the effect of signal fluctuations on NF-κB sensitivity.

We will look into comparisons of modes of stimulation (constant signal or periodic signal) with one degree of additive noise (white, colored with given amplitude). By testing the effects of noise on different parameters of the system such as activation probability, time to nuclear translocation

and expression level of target genes. We began our investigation of noise sensitivity by using a mathematical model of NF-κB response in fibroblasts. The model has been extensively validated and recapitulates the major dynamics of the system. We then use the predictions from this model to establish experimental tests to verify if we see the expected behavior in living cells.

To do this we rely on a fluorescently tagged p65 mouse fibroblast line, which allows us to temporally monitor NF-κB translocation. The cells will grow in a microfluidic cell culture device that is able to generate temporal signals at .5 sec intervals using a microfluidic mixing chip. This allows us to create custom-inputs for different chambers on the same device.

References

1. B. Snijder, R. Sacher, P. Rämö, E. M. Damm, P. Liberali, L. Pelkmans, Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*. **461**, 520–523 (2009).

2. S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggan, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, A. Raj, Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. **546**, 431–435 (2017).

3. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. *Science (80-.).* **297**, 1183–1186 (2002).

4. J. M. Raser, E. K. O'Shea, Noise in gene expression: origins, consequences, and control. *Science*. **309**, 2010–3 (2005).

5. J. Das, M. Ho, J. Zikherman, C. Govern, M. Yang, A. Weiss, A. K. Chakraborty, J. P. Roose, Digital signaling and hysteresis characterize ras activation in lymphoid cells. *Cell*. **136**, 337–51 (2009).

6. Q. Zhang, S. Gupta, D. L. Schipper, G. J. Kowalczyk, A. E. Mancini, J. R. Faeder, R. E. C. Lee, NF-κB Dynamics Discriminate between TNF Doses in Single Cells. *Cell Syst.* **5**, 638-645.e5 (2017).

7. R. Losick, C. Desplan, Stochasticity and cell fate. *Science (80-.).* **320** (2008), pp. 65–68.

8. M. Voliotis, R. M. Perrett, C. McWilliams, C. A. McArdle, C. G. Bowsher, Information

transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E326–E333 (2014).

9. P. S. Swain, M. B. Elowitz, E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12795–12800 (2002).

10. R. A. Kellogg, C. Tian, T. Lipniacki, S. R. Quake, S. Tay, Digital signaling decouples activation probability and population heterogeneity. *Elife*. **4**, e08931 (2015).

11. N. Drayman, P. Patel, L. Vistain, S. Tay, HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *Elife*. **8** (2019), doi:10.7554/eLife.46339.

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

13. Q. Han, N. Bagheri, E. M. Bradshaw, D. A. Hafler, D. A. Lauffenburger, J. C. Love, Polyfunctional responses by human T cells result from sequential release of cytokines. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1607–1612 (2012).

14. W. Hu, A. Jain, Y. Gao, I. M. Dozmorov, R. Mandraju, E. K. Wakeland, C. Pasare, Differential outcome of TRIF-mediated signaling in TLR4 and TLR3 induced DC maturation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13994–9 (2015).

15. K. Thurley, L. F. Wu, S. J. Altschuler, Modeling Cell-to-Cell Communication Networks Using Response-Time Distributions. *Cell Syst.* **6**, 355-367.e5 (2018).

16. T. Suda, J. Suda, M. Ogawa, Disparate differentiation in mouse hemopoietic colonies derived from paired progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2520–4 (1984).

17. M. Adler, Y. Korem Kohanim, A. Tendler, A. Mayo, U. Alon, Continuum of Gene-Expression Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. *Cell Syst.* **8**, 43-52.e5 (2019).

18. B. Ghosh, R. Karmakar, I. Bose, Noise characteristics of feed forward loops. *Phys. Biol.* **2**, 36–45 (2005).

19. R. E. C. Lee, S. R. Walker, K. Savery, D. A. Frank, S. Gaudet, Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Mol. Cell*. **53**, 867–879 (2014).

20. S. Mangan, U. Alon, Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11980–5 (2003).

21. S. T. Smale, Dimer-specific regulatory mechanisms within the NF-κB family of transcription factors. *Immunol. Rev.* **246**, 193–204 (2012).

22. Q. Zhang, M. J. Lenardo, D. Baltimore, 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell*. **168**, 37–57 (2017).

23. A. Hoffmann, A. Levchenko, M. L. Scott, The I B – NF- B Signaling Module : Temporal Control and Selective Gene Activation. **298**, 1241–1245 (2002).

24. M. Karin, How NF-kB is activated: the role of the IkB kinase (IKK) complex (available at https://www.nature.com/articles/1203219.pdf?origin=ppub).

25. M. S. Hayden, S. Ghosh, Signaling to NF-κB. *Genes Dev.* **18** (2004), pp. 2195–2224.

26. M. Vincendeau, K. Hadian, A. C. Messias, J. K. Brenke, J. Halander, R. Griesbach, U. Greczmiel, A. Bertossi, R. Stehle, D. Nagel, K. Demski, H. Velvarska, D. Niessing, A. Geerlof, M. Sattler, D. Krappmann, Inhibition of Canonical NF-κB Signaling by a Small Molecule Targeting NEMO-Ubiquitin Interaction. *Nat. Publ. Gr.* (2015), doi:10.1038/srep18934.

27. J.-M. Feng, Z. Jiang, P. Gao, X. Yu, P. Mu, R. Zhang, S. Bi, R. Fang, C. Wang, Q. Jiang, M. Lv, Pathway STING − B Activation in the cGAS κ NF- Are Essential for IRF3 and β IKK − NEMO NEMO–IKKb Are Essential for IRF3 and NF-kB Activation in the cGAS–STING Pathway. *J. Immunol. by guest January J. Immunol.* **23** (2018) (available at http://www.jimmunol.org/content/early/2017/09/22/jimmun).

28. Y. Dondelinger, M. Darding, M. J. M. Bertrand, H. Walczak, Poly-ubiquitination in TNFR1-mediated necroptosis. *Cell. Mol. Life Sci.* **73**, 2165–2176 (2016).

29. M. S. Hayden, S. Ghosh, Shared Principles in NF-κB Signaling. *Cell*. **132**, 344–362 (2008).

30. T. Lawrence, The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb. Perspect. Biol.* **1**, a001651 (2009).

31. S. Krishna, M. H. Jensen, K. Sneppen, L. P. Kadanoff, Minimal model of spiky oscillations in NF- B signaling (available at http://www.pnas.org/content/pnas/103/29/10840.full.pdf).

32. M. Son, A. G. Wang, H.-L. Tu, M. O. Metzig, P. Patel, K. Husain, J. Lin, A. Murugan, A. Hoffmann, S. Tay, NF-κB responds to absolute differences in cytokine concentrations. *Sci. Signal.* **14** (2021), doi:10.1126/SCISIGNAL.AAZ4382.

33. M. Heltberg, R. A. Kellogg, S. Krishna, S. Tay, M. H. Jensen, Noise Induces Hopping between NF-κB Entrainment Modes. *Cell Syst.* **3**, 532-539.e3 (2016).

34. B. Mengel, S. Krishna, M. H. Jensen, A. Trusina, Theoretical analyses predict A20 regulates period of NF-kB oscillation (2009) (available at http://arxiv.org/abs/0911.0529).

Computer-vision reveals hidden variables underlying NF-κB activation in single cells

Parthiv Patel[1,2], Nir Drayman[1,2], Ping Liu[3], Mustafa Bilgic[3], Savaş Tay [1,2]*

[1]The University of Chicago, Pritzker School of Molecular Engineering, Chicago, IL;

[2]The University of Chicago, Institute for Genomics & Systems Biology, Chicago, IL;

[3]Illinois Institute of Technology, Department of Computer science, Chicago, IL

* Correspondence: tays@uchicago.edu

One Sentence Summary: An image-based machine learning approach discovers mechanisms underlying variable activation of NF-κB pathway in single cells

Summary

Individual cells are heterogeneous when responding to environmental cues. Under an external signal certain cells activate gene regulatory pathways, while others completely ignore that signal. Mechanisms underlying cellular heterogeneity are often inaccessible because experiments needed to study molecular states destroy the very states we need to examine. Here, we developed an image-based support vector machine-learning model to uncover molecular variables controlling activation of the immune pathway NF-κB. Computer-vision analysis predicts the identity of cells

that will respond to cytokine stimulation and shows that pathway activation is pre-determined by minute amounts of "leaky" NF-κB (p65:p50) localization to the nucleus. Mechanistic modeling revealed that ratio of NF-κB to its inhibitor IκB pre-determines leakiness and activation probability of a cell. While cells transition between molecular states, they maintain their overall probabilities for NF-κB activation. Our results demonstrate how computer-vision can study living-cells and discovers mechanisms behind heterogenous single-cell activation under pro-inflammatory stimuli.

Introduction

Individual cells show unpredictable and highly variable responses in a wide range of contexts, from immune signaling to transcriptional activation and to drug response(*1–7*). For example, following stimulation with signaling molecules, a portion of cells in a population will strongly activate inflammatory response pathways like NF-κB and downstream transcription factors, while others will completely ignore the stimulus and will not activate(*8*). Determining the sources of cellular variability is of immense importance for fundamentally understanding gene regulation, signaling, immunity, and development, as well as in predicting variable drug response and tolerance(*9–12*). Despite previous demonstration of heterogeneous signaling responses in a wide range of contexts(*8*, *13*, *14*) like development, immune response, infection and cancer, it remains difficult to explore the molecular and cellular mechanisms that drive variable behavior in single cells and the underlying causes of cellular variability are unknown in many contexts.

Here, we study the NF-κB system to investigate the underlying sources of cellular heterogeneity using a computer vision approach applied to microfluidic live-cell imaging experiments. NF-κB is a key transcriptional pathway that is activated by many signaling molecules

involved in immunity(*15*, *16*) and controls the expression of hundreds of pro-inflammatory and cell fate genes(*17*). Dysregulation of NF-κB is implicated in many physiological conditions including infection, autoimmunity, and cancer(*15*, *18*). Live-cell imaging and single-cell analysis have shown that NF-κB activation is highly variable in single mammalian cells. When cells are stimulated with signaling molecules (i.e. TNF or LPS), some of them show complete cytoplasm-to-nucleus translocation of the p65 transcription factor (the hallmark of pathway activation), while others ignore the stimulus and show no translocation and no NF-κB target gene expression (**Figure 1.1a**). The fraction of cells that respond to a stimulus (signaling input) increases in a dose-dependent manner(*19*) and despite this seemingly noisy single-cell behavior, NF-κB nevertheless manages to mount specific responses to different signaling molecules, taking into account their concentrations and temporal ordering, and regulates gene expression in an input specific manner(*3*, *13*). Despite the plethora of research uncovering response relationships to input conditions, it is still unclear as to what underlying molecular components or mechanisms control variability in NF-κB pathway activation alongside the input. The accurate functioning of signaling pathways like NF-κB is crucial to maintaining healthy immunity and immune development and integration of internal cell state and external stimuli is important to understand the functional consequences of noise in the system (**Figure 1.1b)**. Despite several studies that significantly contributed to our understanding of cellular variability and noise(*1*, *20–24*), the ability of NF-κB and other signaling pathways to accurately interpret complex environmental signals and create specific gene expression responses is mostly unexplained.

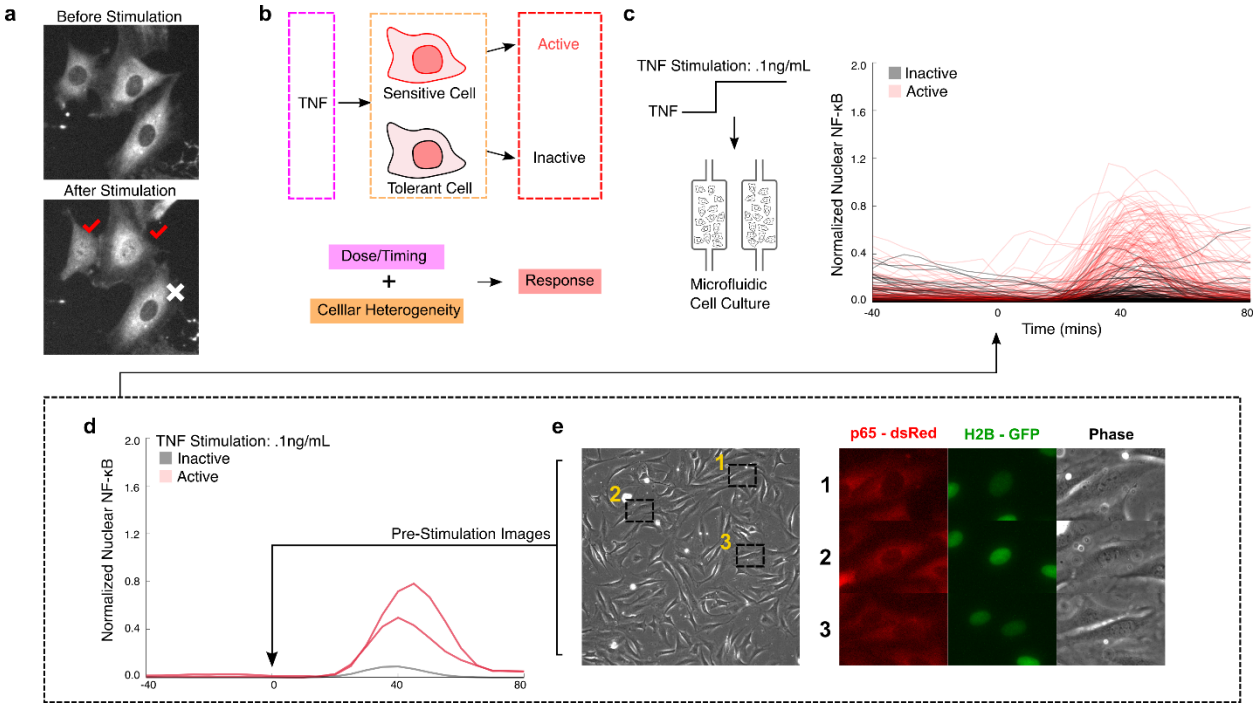


Figure 1.1: Pre-stimulation cell images can be used to generate a predictive model for NF-κB activation. (a) Under TNF stimulation, a fraction of individual cells persists in an un-activated state and do not show nuclear NF-κB localization, while others activate and NF-κB transcription factor p65 accumulates in the nucleus. Example images show activated cells, indicated with red check marks. **(b)** Analysis of single cells under constant dose of TNF shows variable single cell activation in the population: a given cell may or may not respond to the TNF stimulus. It is unclear whether single cell variability is due to purely stochastic processes (i.e. if a given cell can randomly achieve activated state), or if it is deterministic where only sensitive cells activate under the TNF input. **(c)** We use microfluidic cell culture to stimulate cells with TNF signals and image the nuclear localization of NF-κB over time in single cells. Analysis of individual cells reveals NF-κB localization traces (0.1ng/mL TNF stimulation shown on the right, stimulation starts at t=0). **(d)** Single cell traces reveal heterogenous activation and subpopulations of active and inactive cells. **(e)** We record pre-stimulation images of mouse 3T3 fibroblast cells that express p65-dsRed and H2B-GFP reporters and feed them into our machine learning pipeline.

The challenge in understanding the cause of heterogenous responses in signaling and gene regulation is a classic observer effect problem - we can only identify the responding (i.e. activated) cells after we stimulate the cell population with signaling molecules, which will inevitably perturb the cellular states we wish to examine. Many signaling pathways, including NF-κB, contain

multiple feedbacks that upon exposure to an external signal rapidly change the molecular composition of the pathway(*25*) which fundamentally limits the accuracy and power of repeat stimulation experiments(*8*, *26*). This presents a fundamental difficulty in determining how pre-existing cell-to-cell differences impact the probability of any given cell to respond to a stimulus, especially in short term cell states that arise from systemic changes in negative and positive feedback regulation. One way to circumvent this problem is by use of mathematical modeling(*27*, *28*). Modeling of pathway dynamics can reveal important insights and general patterns of noisy events, but, are bound by the underlying assumptions of the models and can offer many plausible explanations for cellular heterogeneity(*29*) or require monitoring multiple components of the system for validation(*30*). Because of these experimental and theoretical limitations, molecular mechanisms underlying important cellular behaviors like variable drug response, digital pathway activation, and signal tolerance are limited and often unclear(*8*, *31–33*).

To study the molecular mechanisms behind variable NF-κB activation in single cells we adopted an image-based machine learning approach to predict which individual cells will activate the NF-κB pathway in response to an inflammatory stimulus. On many occasions, machine learning algorithms have been shown to exceed human decision making on complex game problems(*34*, *35*). Machine learning has also previously been used to extract pathophysiological outcome predictions from images of tissue and inform classification of differentiation, disease state and infection(*36*, *37*). By imaging living cell morphologies before, during and after chemical stimulation, we were able to use the cell image before stimulation to predict whether that cell will activate the NF-κB pathway or not. We developed an image-based support vector machine (SVM) to predict outcomes of chemical stimulation in individual cells (see Supplemental Information). Our computer-vision based method classifies cells into responding and non-responding groups

based solely on the unperturbed cell's image and is able to predict which cells will respond to or ignore TNF stimulation with 79.4% accuracy. As this prediction is done on cells that are not chemically stimulated, our approach allows studying how the cell's unperturbed state differs between responding vs. nonresponding cells.

Results

Machine learning predicts single-cell NF-κB activation from pre-stimulation images

To develop a predictive machine learning model for NF-κB activation in individual cells, we first performed live-cell stimulation experiments to generate a reference data set. We used a high throughput microfluidic cell culture platform(*38*) to chemically stimulate and quantitatively measure NF-κB response in cultured mouse fibroblast cells (**Figure 1.1c-e**). These cells express p65-dsRed and H2B-GFP fluorescent reporters(*8*) to track NF-κB nuclear translocation in real-time. In activated cells, cytoplasmic NF-κB (p65) rapidly moves into the nucleus in a digital fashion(*8*). Cells were first imaged unperturbed for 1.5 hours, stimulated using the automated microfluidic system with TNF (0.1 ng/mL) and monitored for 6 hours. Custom image analysis software was used to track the nuclear localization of p65 and assign a label to each cell (activated vs. not-activated)(*38*).

To determine the characteristics of single cells that are responsible for activation we then assigned representative features for single cell images and analyzed predictor correlations (**See Supplementary Information**) on our defined cell features (**Figure 1.2a**) in a single dosage dataset (**Supplementary figure 1.1**). We also expanded the experimental dataset from a single TNF dose to a range of doses and reported their individual accuracies (0.005 to 5 ng/mL, n=3456, **Supplementary Figure 1.2a**). Analyzed together using the SVM algorithm, the total accuracy is

79.4%. We also tested an alternate cell activation criterion using different single cell p65 trace patterns as categories (**Supplementary Figure 1.3a**) on a single dose (0.05 ng/mL) and found that while some trace patterns are able to be classified by the model, this does not generally hold and has lower accuracy than binary classification (**Supplementary Figure 1.3b**). For our binary classifier labelling we use a mean nuclear p65 peak height threshold of 500 (nuclear $p65_{peak}$ - nuclear $p65_{t=0}$) (**Supplementary Figure 1.4a**). We also tested with different machine learning models and using different thresholding values for activation based on peak height and chose our threshold to balance the sample number and prediction accuracy of activated cells. While both CNN and tree classifiers are able to classify activation more accurately than the SVM, feature analysis for both CNN and tree classifiers is typically inaccessible or difficult. Overall, we can accurately predict individual cells' TNF response from their pre-stimulation images, which strongly suggests the existence of deterministic causes underlying NF-κB response in single cells.

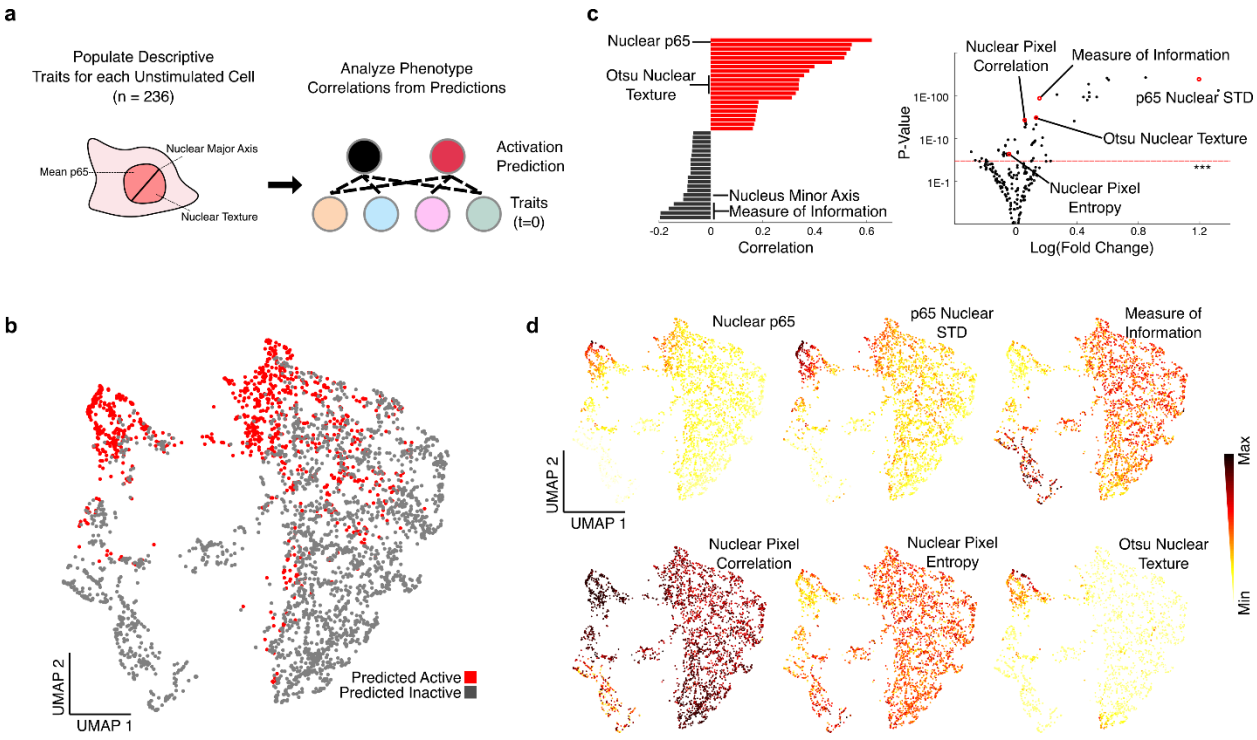


Figure 1.2: Machine-learning predicts NF-κB activation in single cells from pre-stimulation images. (a) We generate descriptive traits for each single cell image and analyze correlations between pre-stimulation traits and NF-κB activation probability. **(b)** Using a subset of highly predictive features (n=8), UMAP corroborates the clustering of a highly predictive fraction of cells that are TNF sensitive and TNF resistant. The cells indicated with red dots have high probability of activating NF-κB under TNF stimulation. **(c)** We determined the top feature of activation probability to be the mean nuclear fluorescence of the NF-κB (p65-dsRed) signal in the nucleus before any exposure to TNF. The plot shows the correlation of traits to single cell activation probability, with log(fold change) of traits from predicted active cells to predicted inactive cells on the x-axis and significance of difference on the y axis determined by t-test. Other highly predictive features include the standard deviation of nuclear p65, mean nuclear phase intensity, major axis length and a texture feature describing information measure of correlation in the nucleus, as well as aggregative image features like Otsu dimension, and SFTA (*** = 0.001). **(d)** UMAP plots of several highly predictive features determining NF-κB activation in single cells.

Leaky NF-κB localization is the primary predictor of single-cell activation

To understand the characteristics that underlie activation, we then identified a subset of highly predictive features that correlate with the cell's likelihood to become activated, including the basal (pre-stimulus) level and standard deviation of nuclear NF-κB and several texture features (**Figure 1.2b)**. Using these features, cells can be visualized by Uniform Manifold Approximation and Projection (UMAP), which present "TNF-sensitive" and "TNF-resistant" cell regions **(Figure 1.2c)**. Each point on the UMAP projection represents a cell which occupies a particular state with different amplitudes for the image features we previously determined. The UMAP plot shows cells segregate through variables unrelated to activation; we found that features relating to size and sum of protein are responsible for creating this divide (**Supplemental Figure 1.4b)**. Analysis of the contribution of individual cell morphology features to the prediction revealed that most of the variation in single cell predictions is explained by a single feature – the nuclear p65-dsRed levels before stimulation ($r = 0.62$), which showed a significant difference between TNF-sensitive and TNF-resistant cells.

While nuclear p65 was the most significant, we also explore the relationship between different features to see the effect of coupling on our variables: despite the very significant p-values for many features, often times many features are coupled in non-linear ways. While a large amount of variance can be explained using nuclear NF-κB, additional features increase the prediction for different cell states (**Supplementary Figure 1.5a,b**). Interestingly, the combination of p65, H2b and phase features can all used in one decision tree to isolate a subpopulation of inactive cells (**Supplementary Figure 1.5d**). To understand some of the underlying processing occurring within the algorithm, we looked into correlations in this subpopulation. Dissecting this

highly coupled subpopulation from Supplementary Figure 5d in turn showed remarkable correlation with many features that gave high p-values but did not have high fold changes.

Although nuclear p65 prior to stimulation was the most predictive feature for cell activation, nuclear texture features and H2b channel features are also highly predictive components of the classifier. To investigate whether other variables were acting as proxies for nuclear p65, we look to see if we could substitute phase image alone for biological markers in the other channels (accuracy = 73%, **Supplementary Figure 1.5c**). While some of the features did show high correlation (otsu: r=0.600), many other features had much lower correlations (**Supplementary Figure 1.6**). The clustering of correlations shows that there may be multiple underlying biological features that are being represented. For example, H2b feature sensitivity may be indicative of exogenous factors like nuclear import/export(*39*) and cell cycle(*40*) playing a role in NF-κB activation probability. Using gray level co-occurrence matrix (GLCM) features from the phase channel, we used metrics like entropy and measure of information that reflect different features of adjacent image pixels to determine if nuclear texture was predictive of NF-κB activation. The mapping of these various texture features onto our UMAP space (**Figure 1.2d**) show that highly activated cells have more heterogenous nuclear morphology with low pixel correlation, and that resistant cells have more homogenous nuclei with high pixel correlation. These features are suggestive of chromatin openness(*41–44*) and epigenetic factors having a potential relationship to NF-κB activation.

<u>Cells transition between multiple states but maintain their overall propensity for pathway activation</u>

To further characterize the unperturbed cell states, we look at the time window leading up to activation. By using multiple pre-stimulation images, we have the unique opportunity to look at the trajectories of individual cells in the UMAP space over time and see the short-term state transitions of a given cell before exposure to TNF (**Figure 1.3a**). Surprisingly, we find that the resting (unstimulated) cells do not occupy a fixed parameter state and instead show dynamic transitions between different states. In other words, important state features such as nuclear p65 level and nuclear texture (indicative of chromatin availability) actively fluctuate in unstimulated cells. Plotting several features over time for single cells reveals coordinated changes in nuclear texture and nuclear NF-κB (p65) before stimulation (**Figure 1.3b**) and is further observed by plotting the covariance between variables (**Supplemental Figure 1.6**). As cells change over time, the image features that are changing are often correlated. This could suggest that there is compensation from some features to others as cells dynamically change state. Importantly, while there are often changes in state features during these transitions, the overall prediction score for NF-κB activation for these single cells does not change significantly (**Figure 1.3c**). The cells maintained their overall propensity for NF-κB activation under TNF stimulation, while actively transitioning between multiple cellular states. This result indicates that cellular sensitivity towards cytokine stimulation may be stable despite the fluctuations present in our features and within the cells before stimulation. To look at how stable the prediction score is at earlier timepoints, we look at single cell autocorrelation of the prediction score. The correlation remains high even 35 mins prior stimulation ($r = 0.78$, $n = 1109$ cells, pairwise Pearson's Correlation) (**Figure 1.3d**). While 35 minutes may not be enough to capture the regime of long term state (ex. Cell cycle), it captures the relevant timescales of the NF-κB system state right before stimulation, as cells experience changes in pathway protein composition even at steady state(*25*). This indicates that while the cells

do make transitions on the UMAP space to different states, they tend to stay within the same grouping with respect to activation prediction.

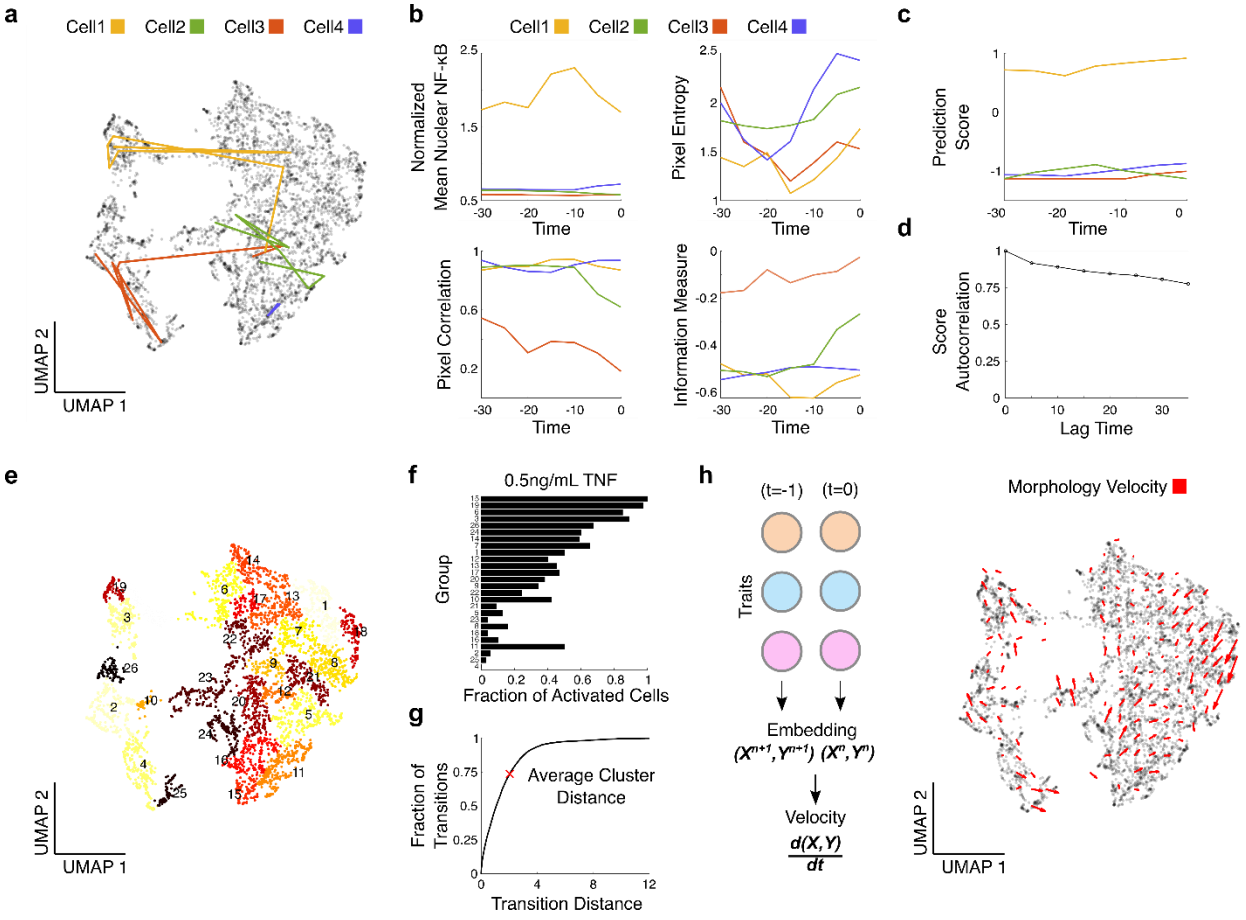


Figure 1.3: Resting cells dynamically transition between different states but maintain their overall NF-κB activation probability. Imaging of single cells before, during and after stimulation enables analysis of single cell state stability. **(a)** We use selected cellular features across multiple time points to analyze the trajectory of single cells across the latent UMAP space before stimulation. Colored lines show the temporal trajectories of 4 example cells on the UMAP plot before stimulation with TNF. Cells transition between various points before TNF stimulation. **(b)** Individual cells show coordinated changes in state features while transitioning. We show the dynamic changes in the level of various predictive features for these cells. Time is given in minutes before stimulation (stimulation starts at t=0). **(c)** While cells transition between different points with different transition distances, their probability to activate NF-κB (given by the prediction score of our algorithm) remains the same. **(d)** Autocorrelation of the prediction score among single cells across pre-stimulation time points remains stable 35 mins before stimulation (r = 0.78, n = 1109 cells, pairwise Pearson's Correlation) **(e)** Single cells are clustered in the UMAP space, by

Figure 1.3 (continued) local adjacency into communities to look for activation differences by state (color indicates group). **(f)** Analyzing endpoint data after TNF stimulation by individual state clusters shows different fractions of activated cells for different states. There is activation similarity in adjacent clusters, and **(g)** 73.73% of all cell transitions are within or to an adjacent cluster. Shown is cumulative distribution function of state transitions by transition distance (Red X indicates average distance between clusters). **(h)** Aggregated trajectories reveal a state velocity depicting activation score movement across the population.

To confirm that the cells maintain their activation probability through stimulation, we clustered all single cells to look for differences between states after TNF is introduced (**Figure 1.3e**) (see **Supplementary Information**). Analysis of the activation fraction (i.e. the population fraction of cells activating NF-κB) after TNF stimulation of the different state clusters reveals that the different states indeed have different activation fractions (**Figure 1.3f**). Adjacent cell clusters have similar activation probabilities and 73.71% of all transitions happened within or to an adjacent state cluster (**Figure 1.3g**). In other words, while cells transition between various parameter states, they maintain their overall sensitivity towards TNF stimulation. We also looked at how dosage plays a part in the magnitude and fraction of activation. Interestingly, while the magnitude of activation followed similar behavior across states (**Supplementary Figure 1.7b**), the fraction of activated cells varies across dose for different states (**Supplementary Figure 1.7c**). While some cells remain persistently sensitive or resistant, there are other cells that have their activation profiles dictated by TNF dose. This supports the hypothesis that pre-existing cell features govern activation or resistance to TNF input.

To see if there was regularity in transitions, we looked at average cellular transitions across the UMAP space. Aggregating single cell trajectories across the UMAP space can be used to understand the state velocity across the population (**Figure 1.3h**). We find that the average

transition of single cells follows generalized patterns and indicates that while the transitions might appear non-uniform at the single cell level, there are average population level movements.

Mechanistic modelling shows that IκB:p65 ratio drives p65 leakiness and activation probability of a given cell

Our machine learning analysis revealed that leaky (pre-stimulation) p65 localization to be the main predictor of single cell activation of NF-κB. There are many different NF-κB regulators that can influence p65 nuclear localization in resting unstimulated cells. This "leakiness" has been attributed to mechanisms of nucleo-cytoplasmic shuttling of IκB and NF-κB (p65) in resting cells(*39*, *45*, *46*) and observations that basal nuclear NF-κB abundances have important biological consequences on pathway activation has been elucidated(*47*, *48*). However, the NF-κB pathway is robust to environmental fluctuations and noise, and many built-in negative feedback mechanisms **(Figure 1.4a)** prevent spontaneous activation (nuclear import of p65) in the absence of stimulation. Nevertheless, our imaging data reveals that many cells show minute amounts of "leaky" nuclear p65 (localization without stimulus) (**Supplementary Figure 1.8**), which has been observed in cells(*39*) but not thoroughly investigated regarding activation. We find that this small but significant pre-stimulus p65 localization pre-determines the sensitivity of the mouse fibroblast cells in responding to upcoming TNF challenges.

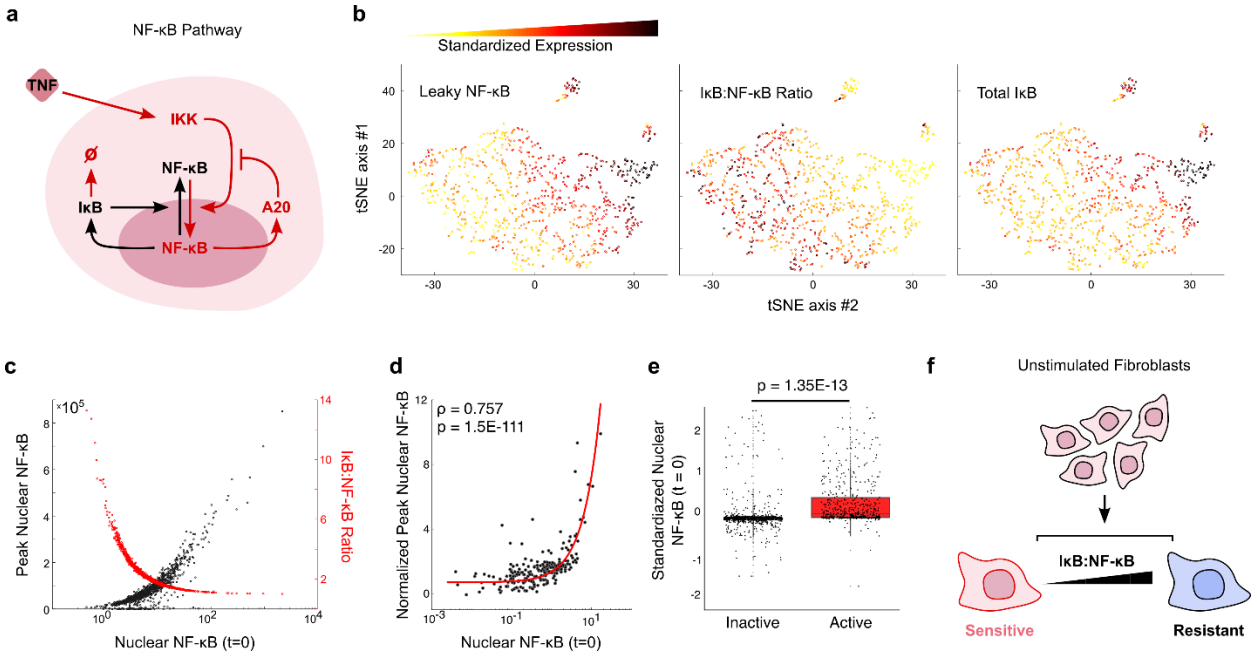


Figure 1.4: Simulations suggest that leaky NF-κB localization and overall NF-κB response is predetermined by the ratio of IκB to NF-κB proteins in single cells. (a) Simplified schematic of the NF-κB pathway used in simulations. IκB provides negative feedback to the pathway, preventing NF-κB nuclear localization in unstimulated cells. **(b)** tSNE mapping of simulated single cells on NF-κB pathway protein levels shows anti-correlated Nuclear NF-κB and IκB/NF-κB ratio in single cells. **(c)** Simulations predict a correlation between nuclear NF-κB at t=0 and NF-κB peak height after stimulation and consequently an inverse correlation between NF-κB at t=0 and the IκB/NF-κB ratio. **(d)** Activated live single cells confirm the prediction and show significant correlation between nuclear NF-κB at t=0 and NF-κB peak height after stimulation. **(e)** Pre-stimulation NF-κB nuclear fluorescence accounts for a high degree of variance and activated live single cells have a significantly higher level than inactive cells. **(f)** Simulations suggest that increasing IκB/NF-κB ratio make cells more resistant to activation under TNF stimulation.

To understand the molecular mechanism behind p65 nuclear leakiness, and specifically how it could lead to increased TNF sensitivity and NF-κB activation probability, we explored a mathematical model of NF-κB pathway in single cells(*8*, *14*). This model accurately reproduces the experimentally measured NF-κB dynamics in a wide range of conditions in the mouse fibroblast cells used in our study.[1,6,11,12] We performed simulations and generated thousands of

dynamic traces that describe nuclear NF-κB (p65) localization in single cells stimulated with TNF. We then analyzed the molecular composition of each simulated cell and its NF-κB activation profile. We found that variability in levels of IκB, the main inhibitor of NF-κB (p65) keeping it in the cytoplasm, explains the observed NF-κB leakiness. In addition to binding to p65 and keeping it in the cytoplasm in unstimulated cells, IκB acts as a dynamic negative feedback regulator of the pathway, since IκB is a direct target gene of NF-κB. IκB is produced when NF-κB pathway is activated and p65 enters the nucleus. It is of note that the leaky nuclear p65 we observe is correlated to total p65 present in the cell. Because many of these variables are entangled in the pathway, changes in one protein typically reflect changes across many proteins in the pathway.

Using dynamic simulations, we perturbed the IκB/NF-κB ratio in single cells prior to TNF stimulation and determined the resulting likelihood of pathway activation for single cells (n=1000) **(Supplementary Figure 1.9b)**. We find a major difference in cellular activation probability and peak NF-κB amplitude for different IκB levels (**Supplementary Figure 1.9c**). Counterintuitively, cells with high initial IκB levels require a smaller TNF dose to achieve NF-κB activation, whereas cells with low initial IκB levels are very resistant to any TNF dose. This surprising finding is explained by the facts that the probability of activation depends on the IκB/NF-κB ratio, and not the total level of IκB, and that the IκB/NF-κB ratio is anti-correlated to the total IκB level **(Figure 1.4b, Supplementary Figure 1.9a)**. Expression differences between activators and repressors have the capability of producing large variance in phenotype differences like we see with NF-κB activation(*49–51*). Our simulations suggest that a pre-existing imbalance in the NF-κB negative feedback is responsible for increased TNF sensitivity in single cells, and that the activation probability of individual cells is pre-determined by the molecular ratio of IκB to NF-κB in the cell **(Figure 1.4c)**.

Experiments show that IκB:NF-κB ratio is the main determinant of cellular activation probability

To validate our modeling predictions experimentally, we analyzed live single cells by time-lapse microscopy and immunofluorescence. First, we found that those cells that responded to TNF stimulus had total nuclear NF-κB (p65) at t=0 significantly correlated with the NF-κB peak height after stimulation (**Figure 1.4d**, p=1.5e-111), and on average, show a five-fold higher level of "leaky" nuclear p65 before stimulation than those that did not respond (**Figure 1.4e**, p = 1.35E-13). It is important to note that the low level of nuclear leakiness we observe is only about 12% of the total fluorescence in a given cell and is far below what is seen during activation. This small but significant difference shows that the regulation of the steady state (i.e. pre-stimulus) NF-κB localization is an important determinant of NF-κB activation and demonstrates the power of utilizing computer vision to analyze single cell responses.

To experimentally validate our hypothesis that the IκB/NF-κB ratio is driving activation probability from the modeling, we stained unstimulated cells for IκB-α protein expression and analyzed the relationship between IκB and NF-κB in individual cells using immunofluorescence **(Figure 1.5a)**. We segmented the nuclear and cytoplasmic compartments and found that, as predicted by our simulations, there is a significant inverse correlation between the IκB/NF-κB ratio and leaky nuclear p65 localization (**Figure 1.5c**, ρ=-0.26, p=2.7E-11), which is the main feature that determines cell activation upon TNF stimulation.

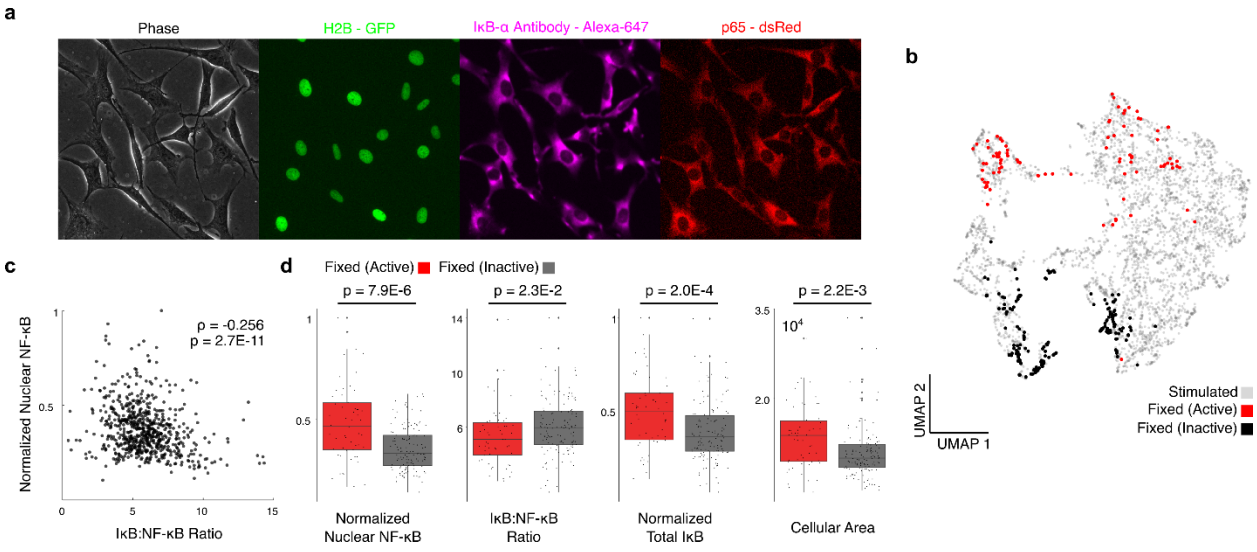


Figure 1.5: Single-cell activation is largely pre-determined by the NF-κB/IκB ratio. We validated our machine learning and modeling predictions in immunofluorescence staining experiments. **(a)** Image of unperturbed 3T3 cells stained with IκB-α with fluorescent p65 and H2B reporters. **(b)** High scoring fixed cells are mapped onto UMAP. **(c)** There is a significant inverse correlation between leaky nuclear NF-κB localization and IκB/NF-κB Ratio. **(d)** IκB/NF-κB ratio is significantly correlated with nuclear NF-κB state at t=0 in predicted fixed cells, and there is a significant difference between IκB/NF-κB ratio, leaky nuclear NF-κB localization and total IκB as well as cellular area.

Next, we mapped the fixed-and-stained cells onto our UMAP visualization of the previous live imaging data (**Supplementary Figure 1.2b**) allowing us to infer IκB localization on our previous experiments for high scoring cells (**Figure 1.5b**). We find that there is indeed a significant difference in nuclear NF-κB, IκB/NF-κB Ratio, and total IκB as well as cellular area, validating our prediction that IκB/NF-κB ratio is driving activation probability in NF-κB (**Figure 1.5d**).

Discussion

The underlying molecular sources of noisy and variable cellular responses can be challenging to study due to a fundamental difficulty of analyzing cellular states without disturbing them. Here, we adopted an image-based machine learning approach to predict the identity of cells that will respond to a biochemical signal and experimentally investigated them. Overall, our demonstration of machine learning in the identification and exploration of noisy NF-κB activation enables cellular analysis through a prospective lens. We supplemented our computer-vision approach with additional single cell experiments and mechanistic modeling to study how cellular states affect the probability of heterogenous signaling in immune regulation.

Our study revealed several molecular determinants of cellular states that lead to variable signaling responses in the NF-κB pathway. Previously, there has been conjecture as to how NF-κB signaling is regulated, and metrics such as cell cycle(*40*) stage and basal NF-κB levels(*39*) have previously been implicated. In addition to this, we find that the cell-to-cell variability in NF-κB activation can largely be explained by a pre-existing difference in the ratio of the NF-κB and its inhibitor, IκB. In cells that are not stimulated by a ligand, the p65 (NF-κB) transcription factor is bound by its inhibitor IκB and is kept in the cytoplasm. When the cells are exposed to a ligand, IκB degrades and NF-κB transcription factors like p65 rapidly translocate to the nucleus and regulate inflammatory genes. These genes are expressed in a ligand specific manner and have important functions in immunity, pathogen-host interactions, as well as development of adaptive immune cells. Therefore, studying which cells would show NF-κB translocation (the hallmark of NF-κB pathway activation) is important to understanding of cellular information processing and cellular variability during immunity.

We find that the cells with a lower IκB: NF-κB protein ratio show small amounts of translocation to the nucleus even without ligand stimulation, which we term “leakiness”. We also found that leaky cells that have a low IκB: NF-κB ratio are more sensitive to TNF stimulation, and readily respond when exposed to this signal. It remains an open question as to what leads to this variation and leakiness in the IκB/NF-κB ratio: one intriguing possibility is that epigenetic variance in genes encoding for NF-κB network components enforce the variable activation chance we observe akin to certain modes of drug tolerance(*32, 52*). Such epigenetic changes could perhaps explain the observed predicative power of nuclear texture features we measured (**Figure 1.2d**), and we plan to address this possibility in follow-up studies.

Another possible explanation for our observations is that cells that are less sensitive to TNF (i.e. those that are not leaky) do not have a functioning NF-κB network. However, in our experiments the cells that show leakiness as well as those that are not leaky all activate and show complete NF-κB translocation at sufficiently high TNF doses, which shows that they have intact IKK/ NF-κB network and transport mechanisms. On the other hand, our pathway simulations showed that simple changes in the protein abundance of p65 and IκB are sufficient to observe leakiness and increased sensitivity to TNF. Such protein abundance changes can happen due to a variety of factors, including the possible epigenetic factors that we discussed above, or cells being exposed to previous signals that change their transcriptional states.

Through dynamical imaging of cells before and after stimulation, we found that resting (unstimulated) cells mostly maintain their sensitivity state at various time points, indicating the stable nature of the underlying NF-κB architecture in response to TNF. However, while cells maintain their overall sensitivity towards future TNF stimulation, these cells actively sample various molecular states. We find that most of the NF-κB response variability is explainable, but

the cells do undergo spontaneous state transitions, which is plausible explanation for the NF-κB activation variability that cannot be explained by our algorithm alone. While the NF-κB system has many components to help maintain its sensitivity towards chemical signals, there is still an element of stochastic state fluctuations that result in creating diversity in single cell and ultimately population behavior. Such phenotypic diversity was shown to be advantageous for a cell population when responding to rapidly changing environmental conditions. Cytokines like TNF often activate multiple signaling pathways, and stratified state-based activation could provide a way to achieve specific responses for specific cell states at different dose ranges.

Our findings suggest an alternate characterization for signaling variability in mammalian signal transduction and offers the interesting possibility that activation heterogeneity is a proxy for cell state that a cell can interpret and act on(*53*). The prevalence of these state niches is further evidence that cell populations are well equipped at differentiating diverse and dynamic signals through cellular specialization even in an equilibrium population. In fact, the matching of activation probability, heterogeneity in the IκB/NF-κB ratio and nuclear morphology is indicative of regulation mechanisms in place for decoding different signaling patterns for a variety of different cellular states.

Materials and Methods

Experimental Model and Subject Details

Knockout p65-/- mouse 3T3 fibroblasts were engineered with p65-DsRed under the native p65 promotor (Tay et al., 2010) and a minimum fluorescence clone was selected to represent endogenous expression of

NF-κB and the pathway dynamics. A ubiquitin-promotor driven H2B-GFP cassette provides a nuclear marker for image processing.

Microfluidic Cell Culture

Cell culture chambers are made of PDMS and coated with fibronectin (FC010-10MG) overnight. Cells are seeded at a constant density of ~20,000 cells/cm^2. Cells are taken at 100% confluence by incubating with .25% trypsin-EDTA (25200-056) for 5mins prior to loading and are cultured for 5 hours before stimulation. Cells were cultured using standard conditions for cell culture (5% CO_2 and 37°C) and maintained using an incubation chamber during imaging. TNF--α (PMC3014_3671982503) was diluted in Fluorobrite DMEM media (A1896701) with 2x glutamax (35050061), pen/strep (15140-122) and FBS (16140071) for stimulation of NIH 3T3 cells. Vials of stimulation media was pressurized at 5psi, kept on ice, and connected to the chip via microbore tubing (PEEK, Vici). The microfluidic device is mounted on the microscope.

We image using Nikon eclipse ti2 microscope to capture both phase and fluorescence images of cultured cells at a 20x magnification. We use a Hamamatsu ORCA-Flash4.0 V3 Digital CMOS Camera (C13440) to capture an image every 5mins for the duration of the experiment. Custom Matlab scripts were used for image processing. NF-κB activation was quantified as mean nuclear fluorescence intensity after background correction. For peak analysis data was smoothed and normalized (Matlab functions smooth and zscore) followed by peak detection. Activation label is crafted from binary threshold for nuclear p65.

Fixed-Cell immunofluorescence

Fibroblasts were seeded in microfluidic chamber and allowed to attach. Cells were rinsed with a 1 min wash of PBS and fixed using a 4% paraformaldehyde solution at 5 hr post attachment. Cells were fixed for 10 minutes at room temperature and blocked and permeabilized with a 10% BSA, 0.5% Triton-X solution in PBS for one hour. Cells were then incubated with primary antibodies in a staining solution (2% BSA, 0.1% Triton-X in PBS) for another one hour at 37°C. Cells were then washed again with PBS and incubated with

secondary antibodies in staining solution for 1 hr at 37°C. Primary antibody used was a Rabbit polyclonal to IKB alpha (ab7217). Secondary antibodies was Goat polyclonal Secondary Antibody to Rabbit IgG - H&L (Alexa Fluor® 647) (G378361).

For the image analysis of fixed cells we used cell images taken at the time point immediately preceding fixation (5 minutes). We then analyzed fixed images and used cell coordinates to match up the fixation analysis with the image analysis that goes into the ML pipeline. The images taken immediately prior to fixation are cells in an unperturbed state similar to prestimulation cells in the dosage experiments. We did not have to do anything in the way of specifically transforming/aligning either to map with each other in UMAP.

Data processing

Predicting the activation of cells using microscopy images can be formulated as an object classification problem. Data processing details are as follows: we first use min-max normalization to scale the pixel intensity into the range of [0, 1]. Then, we crop each cell as a 64 x 64 image patch centered around the cell nucleus. We split the dataset into 10 folds and set up cross-validation experiment to evaluate our model. In each run, eight folds are used for training, one-fold is used for validation, and the last fold is used for evaluation.

Extraction of Texture Features and SVM Model

Hand crafted feature extraction was done during nuclear segmentation and for each tracked single cell. The number of all extracted parameters is 236 and are shown in **Supplementary Table 2**. Texture extraction was done using several custom Matlab scripts. SVM Model was run using cubic kernel with an overall accuracy of 73%. Fitting was done with 10-fold cross validation with 10% dropout on a dataset with [0.005, 0.05, 0.5, 5 ng/mL] TNF input images (n=3456) as well as an independent dataset with 0.1ng/mL TNF

input images (n=2113) prior to stimulation (**Supplementary Figure 1.4**). Dose feature is represented as a categorical variable in the analysis. Feature creation code is provided at:

https://github.com/parthivapatel/ML_cellim.git

Validation with Additional Models

We validate and report the accuracies using other classifiers as well (**Supplementary Table 1**). All models use the same 5-fold cross-validation split. The training fold is split into train-validation to grid search the best hyper-parameters. Finally, we report the average performance over the test folds. The models have the following hyper-parameter settings:

Logistic Regression: we use Logistic Regression with liblinear solver and L2 regularization, and optimize the best C values (for regularization) on the validation folds.

Support Vector Machine with rbf kernel (SVM-rbf**)**: we tune this SVM with `rbf' kernel to optimize different C values and gamma values (for kernel).

Support Vector Machine with polynomial kernel (SVM-poly): this SVM uses a polynomial kernel. The hyper-parameters are the C value and the degree of the polynomial kernel.

Random Forest: we choose to use 100 decision trees as the base estimators for the random forest classifier.

Adaboost: this model has the same number of estimators with the random forest. The base estimator is a decision tree classifier with a maximum depth of 1.

Convolutional Neural Network: there are two convolutional layers, each followed by max-pooling layers. After flattening the output of the second max-pooling layer, two fully connected layers are applied with a tanh activation function. Dropout layers follow both dense layers with a dropout ratio of 0.5 during training. The last layer is a sigmoid function to calculate the probability of two classes. We used the Keras library (https://keras.io/) for the Convolutional Neural Network implementation. We used tensorflow

(https://www.tensorflow.org) as the backend for Keras. We used scikit-learn (https://scikit-learn.org/) for the other classifiers.

State Velocity

Individual cells at time 0 were mapped onto Uniform Manifold Approximation and Projection using parameters (nearest neighbors=12, Chebyshev distance metric, minimum distance=.1) across multiple doses. Multiple additional pre-stimulation time points were then embedded onto the UMAP projection using nearest neighbor embedding. UMAP embedding was turned into a grid of (i,j) components and morphological velocity was approximated using $\frac{1}{n}\sum_{i=0}^{t}\vec{x}_{i,j}^{t}$ where $\vec{x}_{i,j}^{t}$ is the vector of displacement of a pre-stimulation single cell mapped onto UMAP between t and t-1 in the grid (i,j) for all cells. The average displacement vector across all cells is shown in the state velocity.

Clustering

Clustering in **Figure 3** was done using Louvain community detection derived from created adjacency matrix from the nearest neighbor descent algorithm on cell database with hand-crafted features. Single cells move across Louvain partitions between time points at a frequency of less than 73.71%. Clustering for cell traces used in **Supplementary Figure 4** was done through k-means clustering using a k=6.

Statistical Tests

Figure 2c uses Two-sample t-test to evaluate p-values with n = 1028 for positive predicted and n = 2428 for negative predicted. Red line shows significance level of $p = 0.001$.

Figure 4e uses Two-sample t-test to evaluate p-value with n = 1028 for positive predicted and n = 2428 for negative predicted.

Figure 5d uses Two-sample t-test to evaluate p-value on high scoring prediction fixed cells with n = 45 for positive predicted and n = 74 for negative predicted. Misclassification error was minimized by using high scoring cells only.

Mathematical modeling of NF-κB pathway

Using a hybrid stochastic deterministic model of the NF-κB pathway, published previously (*8*, *14*), we simulated 1000 single cells exposed to 20 TNF concentrations. The hybrid model based on Gillespie algorithm uses verified intrinsic noise(*8*, *14*) in TNF receptor-ligand binding and in transcription of IκBα and A20 which form the main negative-feedback loops leading to oscillations

Simulation was done at 100s time steps with 10hour equilibrium period from initial conditions, 6m of a single TNF pulse, then 2h of evolution. Analysis was done on first peak of resulting cell traces of Nuclear NF-κB. Simulated single cells at t=0 can be found in **Supplementary Table 3** for reference TNF concentrations.

See **Supplementary Table 4** for abbreviations, and **Supplementary Table 5** for parameters. NF-κB and TNFR are distributed in a lognormal distribution with means of 10^5 and 10^4 molecules and parameters μ and σ sigma (-1/4, √1/2) and (-1,√2). ODEs for the model are listed below.

Fast Reactions

$$\frac{d[IKKKa]}{dt} = k_a \times B(t) \times (K_N - [IKKKa]) \times \frac{k_{A20}}{k_{A20} + [A20]} - ki \times [IKKKa]$$

$$\frac{d[IKKn]}{dt} = -[IKKKa]^2 \times k_1 \times [IKKn] + k_4 \times (K_{NN} - [IKK_N - IKK_A - [IKKi])$$

$$\frac{d[IKKa]}{dt} = [IKKKa]^2 \times k_1 \times [IKKn] - k_3 \times [Ikka] \times (k_2 + [A20])$$

$$\frac{d[IKKi]}{dt} = k_3 \times [IKKa] \times \frac{k_2 + [A20]}{k_2} - k_4 \times [IKKi]$$

$$\frac{d[I\kappa B_p]}{dt} = a_2 \times [IKKa] \times [I\kappa B] - t_p \times [I\kappa B_p]$$

$$\frac{d[NF\kappa B|I\kappa B_p]}{dt} = a_3 \times [IKKa] \times [NF\kappa B|I\kappa B] - t_p \times [NF\kappa B|I\kappa B_p]$$

$$\frac{d[NF\kappa B]}{dt} = c_{6a} \times [NF\kappa B|I\kappa B] - a_1 \times [NF\kappa B] \times [I\kappa B] + t_p \times [NF\kappa B|I\kappa B] - i_1 \times [NF\kappa B]$$

$$\frac{d[NF\kappa B_n]}{dt} = i_1 \times [NF\kappa B] - a_1 \times k_v \times [I\kappa B_n] \times [NF\kappa B_n]$$

$$\frac{d[A20]}{dt} = c_4 \times [A20_t] - c_5 \times [A20]$$

$$\frac{d[A20_t]}{dt} = c_1 \times [G_{A20}] - c_3 \times [A20_t]$$

$$\frac{d[I\kappa B]}{dt} = -a_2 \times [IKKa] \times [I\kappa B] - a_1 \times [I\kappa B] \times [NF\kappa B] + c_4 \times [I\kappa B_t] - c_{5a} \times [I\kappa B] - i_{1a} \times [I\kappa B] + e_{1a} \times [I\kappa B_n]$$

$$\frac{d[I\kappa B_n]}{dt} = -a_1 \times k_v \times [I\kappa B_n] \times [NF\kappa B] + i_{1a} \times [I\kappa B] - e_{1a} \times [I\kappa B_n]$$

$$\frac{d[I\kappa B_t]}{dt} = c_{1a} \times [G_{I\kappa B}] - c_3 \times [I\kappa B_t]$$

$$\frac{d[NF\kappa B|I\kappa B]}{dt} = a_1 \times [I\kappa B] \times [NF\kappa B] - c_{6a} \times [NF\kappa B|I\kappa B] - a_3 \times [IKKa] \times [NF\kappa B|I\kappa B] + e_{2a} \times [NF\kappa B|I\kappa B_n]$$

$$\frac{d[NF\kappa B|I\kappa B_n]}{dt} = a_1 \times k_v \times [I\kappa Ba_n] \times [NF\kappa B_n] - e_{2a} \times [NF\kappa B|I\kappa B_n]$$

Reporter is Transcribed Cooperatively

$$\frac{d[R_t]}{dt} = c_1 \times \frac{[G_R]^n}{k_r^n - [G_R]^n} - c_3 \times [R_t]$$

Slow Reactions

$$\frac{d[B]}{dt} = k_b \times [TNF] \times (M - [B]) - k_f \times [B]$$

$$\frac{d[G_{A20}]}{dt} = q_1 \times [NF\kappa B_n] \times (A_N - [G_{A20}]) - q_2 \times [I\kappa B_n] \times [G_{A20}]$$

$$\frac{d[G_{I\kappa B}]}{dt} = q_1 \times [NF\kappa B_n] \times (A_{Na} - [G_{I\kappa B}]) - q_2 \times [I\kappa B_n] \times [G_{I\kappa B}]$$

$$\frac{d[G_R]}{dt} = q_1 \times [NF\kappa B_n] \times (A_{Nr} - [G_R]) - q_2 \times [I\kappa B_n] \times [G_R]$$

Stochastic Functions for Receptors, A20, IκB, and reporter genes

$$[R_r^b] = kb \times [TNF_{ext}]$$

$$[R_r^d] = kd$$

$$[R^b] = q_1 \times [NF\kappa B_n]$$

$$[R^d] = q_2 \times [I\kappa B_n]$$

References

1. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. *Science (80-.).* **297**, 1183–1186 (2002).

2. U. Ben-David, B. Siranosian, G. Ha, H. Tang, Y. Oren, K. Hinohara, C. A. Strathdee, J. Dempster, N. J. Lyons, R. Burns, A. Nag, G. Kugener, B. Cimini, P. Tsvetkov, Y. E. Maruvka, R. O'Rourke, A. Garrity, A. A. Tubelli, P. Bandopadhayay, A. Tsherniak, F. Vazquez, B. Wong, C. Birger, M. Ghandi, A. R. Thorner, J. A. Bittker, M. Meyerson, G. Getz, R. Beroukhim, T. R. Golub, Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*. **560**, 325–330 (2018).

3. R. A. Kellogg, C. Tian, M. Etzrodt, S. Tay, Cellular Decision Making by Non-Integrative Processing of TLR Inputs. *Cell Rep.* **19**, 125–135 (2017).

4. A. K. Chakraborty, J. Das, J. Zikherman, M. Yang, C. C. Govern, M. Ho, A. Weiss, J. Roose, Molecular origin and functional consequences of digital signaling and hysteresis during Ras activation in lymphocytes. *Sci. Signal.* **2**, pt2 (2009).

5. J. Das, M. Ho, J. Zikherman, C. Govern, M. Yang, A. Weiss, A. K. Chakraborty, J. P. Roose, Digital signaling and hysteresis characterize ras activation in lymphoid cells. *Cell*. **136**, 337–51 (2009).

6. N. Drayman, P. Patel, L. Vistain, S. Tay, HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *Elife*. **8** (2019), doi:10.7554/eLife.46339.

7. A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science*. **298**, 1241–5 (2002).

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

9. R. Avraham, N. Haseley, D. Brown, C. Penaranda, H. B. Jijon, J. J. Trombetta, R. Satija, A. K. Shalek, R. J. Xavier, A. Regev, D. T. Hung, Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell*. **162**, 1309–21 (2015).

10. M.-C. W. Lee, F. J. Lopez-Diaz, S. Y. Khan, M. A. Tariq, Y. Dayn, C. J. Vaske, A. J. Radenbaugh, H. J. Kim, B. M. Emerson, N. Pourmand, Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4726-35 (2014).

11. L. Otsuki, A. H. Brand, Cell cycle heterogeneity directs the timing of neural stem cell activation from quiescence. *Science*. **360**, 99–102 (2018).

12. S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M. A. Fischbach, K.-K. Wong, K. Brandstetter, B. Wittner, S. Ramaswamy, M. Classon, J. Settleman, A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell*. **141**, 69–80 (2010).

13. R. A. Kellogg, C. Tian, T. Lipniacki, S. R. Quake, S. Tay, Digital signaling decouples activation probability and population heterogeneity. *Elife*. **4**, e08931 (2015).

14. R. A. Kellogg, S. Tay, Noise Facilitates Transcriptional Control under Dynamic Inputs. *Cell*. **160**, 381–392 (2015).

15. T. Lawrence, The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb. Perspect. Biol.* **1**, a001651 (2009).

16. A. Hoffmann, D. Baltimore, Circuitry of nuclear factor κB signaling. *Immunol. Rev.* **210** (2006), pp. 171–186.

17. Q. Zhang, M. J. Lenardo, D. Baltimore, 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell*. **168**, 37–57 (2017).

18. N. D. Perkins, The diverse and complex roles of NF-κB subunits in cancer. *Nat. Rev. Cancer*. **12**, 121–132 (2012).

19. Q. Zhang, S. Gupta, D. L. Schipper, G. J. Kowalczyk, A. E. Mancini, J. R. Faeder, R. E. C. Lee, NF-κB Dynamics Discriminate between TNF Doses in Single Cells. *Cell Syst.* (2017), doi:10.1016/j.cels.2017.10.011.

20. A. Keshelava, G. P. Solis, M. Hersch, A. Koval, M. Kryuchkov, S. Bergmann, V. L. Katanaev, High capacity in G protein-coupled receptor signaling /631/80/86/1999 /631/553/1044 /13/62 /13/95 /13/106 /14/34 /129 /139 article. *Nat. Commun.* **9**, 1–8 (2018).

21. B. Snijder, R. Sacher, P. Rämö, E. M. Damm, P. Liberali, L. Pelkmans, Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*. **461**, 520–523 (2009).

22. J. M. Raser, E. K. O'Shea, Noise in gene expression: origins, consequences, and control. *Science*. **309**, 2010–3 (2005).

23. J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, J. S. Weissman, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*. **441**, 840–846 (2006).

24. R. Losick, C. Desplan, Stochasticity and cell fate. *Science (80-.).* **320** (2008), pp. 65–68.

25. M. L. Scott, T. Fujita, H. C. Liou, C. P. Nolan, D. Baltimore, The p65 subunit of NF-κB regulates IκB by two distinct mechanisms. *Genes Dev.* **7**, 1266–1276 (1993).

26. M. Sumit, A. Jovic, R. R. Neubig, S. Takayama, J. J. Linderman, A Two-Pulse Cellular Stimulation Test Elucidates Variability and Mechanisms in Signaling Pathways. *Biophys. J.* **116**, 962–973 (2019).

27. H. Liu, F. Zhang, S. K. Mishra, S. Zhou, J. Zheng, Knowledge-guided fuzzy logic modeling to infer cellular signaling networks from proteomic data. *Sci. Rep.* **6**, 35652 (2016).

28. R. Cheong, A. Hoffmann, A. Levchenko, Understanding NF-kappaB signaling via mathematical modeling. *Mol. Syst. Biol.* **4**, 192 (2008).

29. Z. E. Perlman, M. D. Slack, Y. Feng, T. J. Mitchison, L. F. Wu, S. J. Altschuler, Multidimensional drug profiling by automated microscopy. *Science (80-.).* **306**, 1194–1198 (2004).

30. K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, G. P. Nolan, Causal protein-signaling networks derived from multiparameter single-cell data. *Science (80-.).* **308**, 523–529 (2005).

31. R. Satija, A. K. Shalek, Heterogeneity in immune responses: from populations to single cells. *Trends Immunol.* **35**, 219–229 (2014).

32. S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggan, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, A. Raj, Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. **546**, 431–435 (2017).

33. A. Raj, A. van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*. **135**, 216–26 (2008).

34. D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis, Mastering the game of Go without human knowledge. *Nature*. **550**, 354–359 (2017).

35. M. Campbell, A. J. Hoane, F. Hsu, Deep Blue. *Artif. Intell.* **134**, 57–83 (2002).

36. F. Buggenthin, F. Buettner, P. S. Hoppe, M. Endele, M. Kroiss, M. Strasser, M. Schwarzfischer, D. Loeffler, K. D. Kokkaliaris, O. Hilsenbeck, T. Schroeder, F. J. Theis, C. Marr, Prospective identification of hematopoietic lineage choice by deep learning. *Nat. Methods*. **14**, 403–406 (2017).

37. J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, G. van den Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, C. O. Hughes, R. Raine, J. Hughes, D. A. Sim, C. Egan, A. Tufail, H. Montgomery, D. Hassabis, G. Rees, T. Back, P. T. Khaw, M. Suleyman, J. Cornebise, P. A.

Keane, O. Ronneberger, Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).

38. R. A. Kellogg, R. Gómez-Sjöberg, A. A. Leyrat, S. Tay, High-throughput microfluidic single-cell analysis pipeline for studies of signaling dynamics. *Nat. Protoc.* **9**, 1713–1726 (2014).

39. C. K. Shrum, D. Defrancisco, M. K. Meffert, Stimulated nuclear translocation of NF-κB and shuttling differentially depend on dynein and the dynactin complex. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2647–2652 (2009).

40. J. M. Ankers, R. Awais, N. A. Jones, J. Boyd, S. Ryan, A. D. Adamson, C. V. Harper, L. Bridge, D. G. Spiller, D. A. Jackson, P. Paszek, V. Sée, M. R. H. White, Dynamic NF-κb and E2F interactions control the priority and timing of inflammatory signalling and cell proliferation. *Elife*. **5** (2016), doi:10.7554/eLife.10473.

41. S. I. Murata, K. Mochizuki, T. Nakazawa, T. Kondo, N. Nakamura, H. Yamashita, Y. Urata, T. Ashihara, R. Katoh, Detection of underlying characteristics of nuclear chromatin patterns of thyroid tumor cells using texture and factor analyses. *Cytometry*. **49**, 91–95 (2002).

42. B. Palcic, Nuclear texture: can it be used as a surrogate endpoint biomarker? *J. Cell. Biochem. Suppl.* **19**, 40–6 (1994).

43. A. Doudkine, C. Macaulay, N. Poulin, B. Palcic, Nuclear texture measurements in image cytometry. *Pathologica*. **87**, 286–99 (1995).

44. H. E. Danielsen, W. Kildal, J. Sudbø, [Digital image analysis in pathology--exemplified in prostatic cancer]. *Tidsskr. Nor. Laegeforen.* **120**, 479–88 (2000).

45. I. Mikenberg, D. Widera, A. Kaus, B. Kaltschmidt, C. Kaltschmidt, Transcription Factor NF-κB Is Transported to the Nucleus via Cytoplasmic Dynein/Dynactin Motor Complex in Hippocampal Neurons. *PLoS One*. **2**, e589 (2007).

46. N. C. Bauer, P. W. Doetsch, A. H. Corbett, Mechanisms Regulating Protein Localization. *Traffic*. **16**, 1039–1061 (2015).

47. R. E. C. Lee, S. R. Walker, K. Savery, D. A. Frank, S. Gaudet, Fold Change of Nuclear NF-κB Determines TNF-Induced Transcription in Single Cells. *Mol. Cell*. **53**, 867–879 (2014).

48. A. V. Bagaev, A. Y. Garaeva, E. S. Lebedeva, A. V. Pichugin, R. I. Ataullakhanov, F. I. Ataullakhanov, Elevated pre-activation basal level of nuclear NF-κB in native macrophages accelerates LPS-induced translocation of cytosolic NF-κB into the cell nucleus. *Sci. Rep.* **9**, 1–16 (2019).

49. E. Rotem, A. Loinger, I. Ronin, I. Levin-Reisman, C. Gabay, N. Shoresh, O. Biham, N. Q. Balaban, Regulation of phenotypic variability by a threshold-based mechanism underlies

bacterial persistence. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12541–6 (2010).

50. N. E. Buchler, F. R. Cross, Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **5**, 272 (2009).

51. M. Son, J. Kaspar, S. J. Ahn, R. A. Burne, S. J. Hagen, Threshold regulation and stochasticity from the MecA/ClpCP proteolytic system in *Streptococcus mutans* competence. *Mol. Microbiol.* **110**, 914–930 (2018).

52. A. Brock, H. Chang, S. Huang, Non-genetic heterogeneity a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* **10** (2009), pp. 336–342.

53. E. A. Torre, E. Arai, S. Bayatpour, C. L. Jiang, L. E. Beck, B. L. Emert, S. M. Shaffer, I. A. Mellis, M. E. Fane, G. M. Alicea, K. A. Budinich, A. T. Weeraratna, J. Shi, A. Raj, Genetic screening for single-cell variability modulators driving therapy resistance. *Nat. Genet.*, 1–10 (2021).

Acknowledgements:

Author Contributions: P.P. did microfluidic experiments; P.L. and P.P. ran machine learning pipeline; P.P. analyzed the experimental data with assistance from N.D.; M.B and S.T supervised the work. All authors wrote the manuscript.

Correspondence: Savaş Tay, Institute for Molecular Engineering, The University of Chicago, tays@uchicago.edu

Funding: This work was supported by NIH grants R01GM128042 and R01GM127527 to S.T. Ping Liu was supported by NSF CAREER Award #1350337.

Competing Interests: The authors declare that they have no competing interests.

Data and Materials Availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The source code and image dataset can be accessed at: https://github.com/parthivapatel/ML_cellim.git **(DOI: 10.5281/zenodo.5361956)**

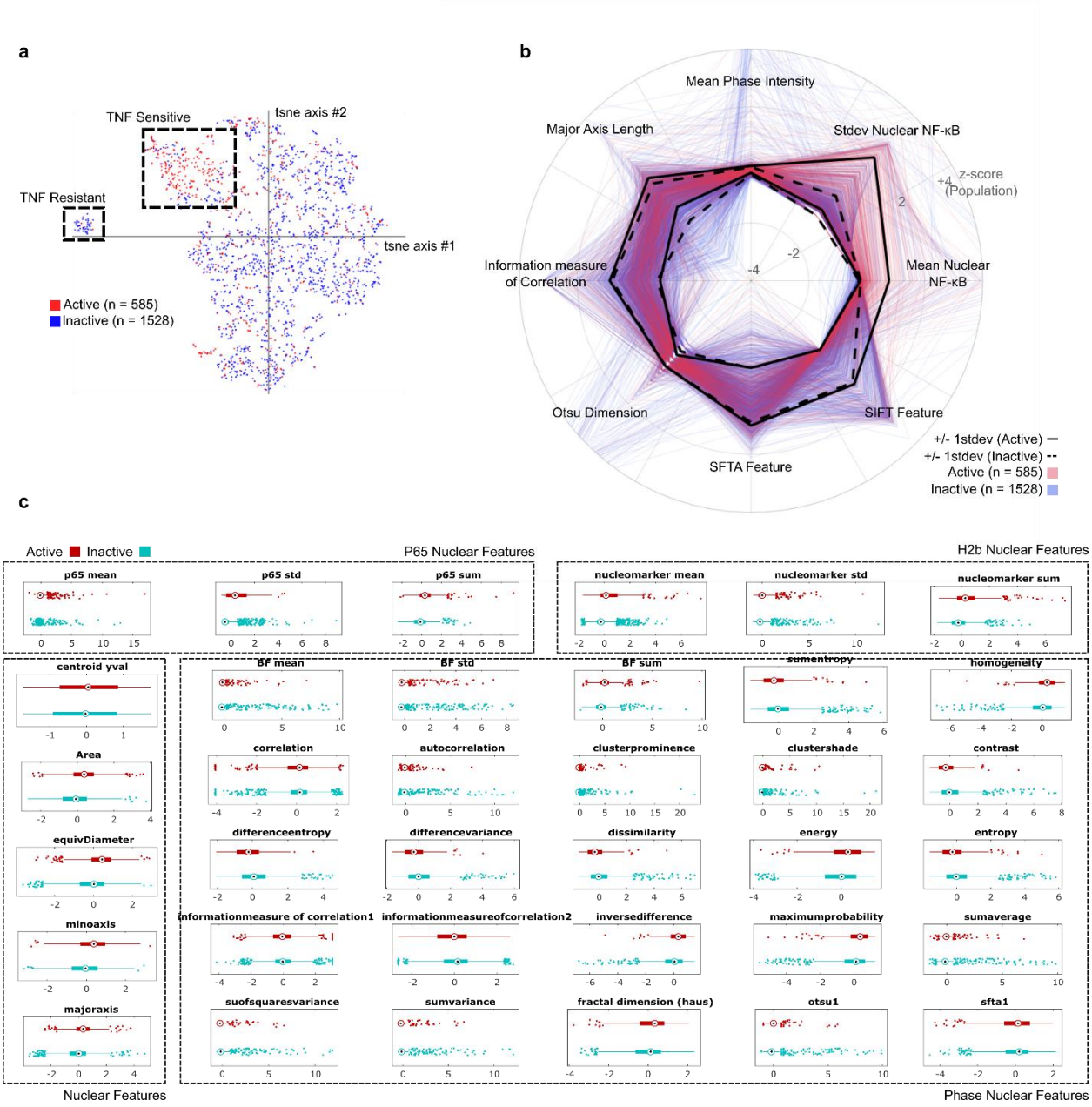
Supplementary Information

Computer-vision reveals hidden variables underlying NF-κB activation in single cells

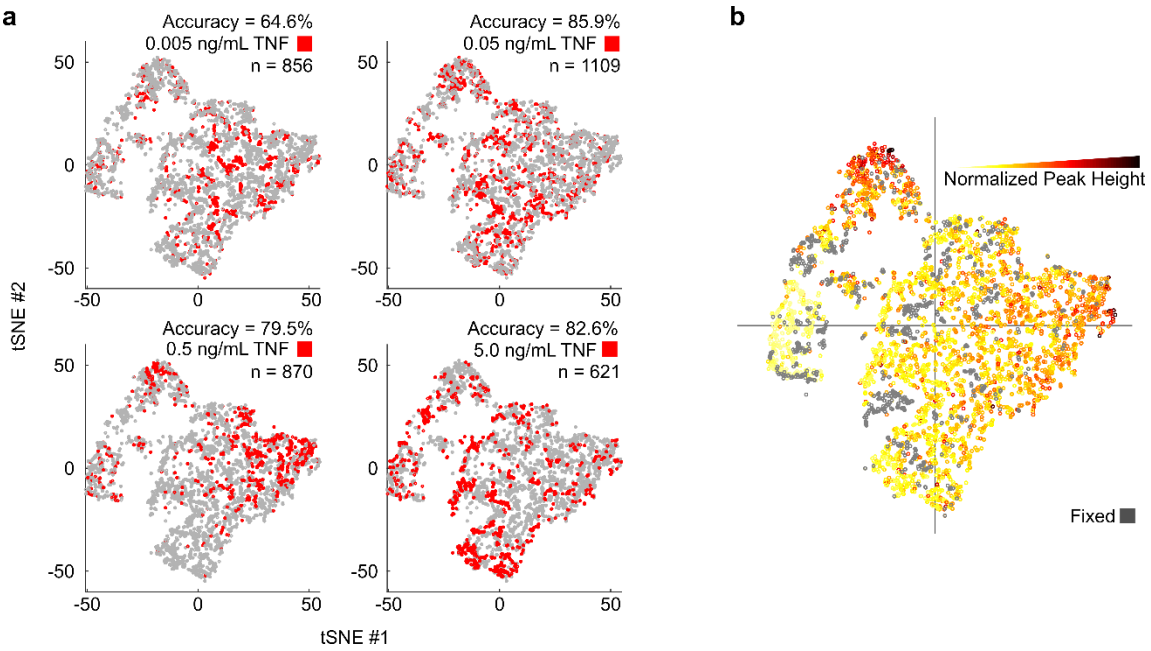
Parthiv Patel[1], Nir Drayman[1,2], Ping Liu[3], Mustafa Bilgic[3], Savaş Tay [1,2]*

Supplementary Materials for this manuscript include the following:

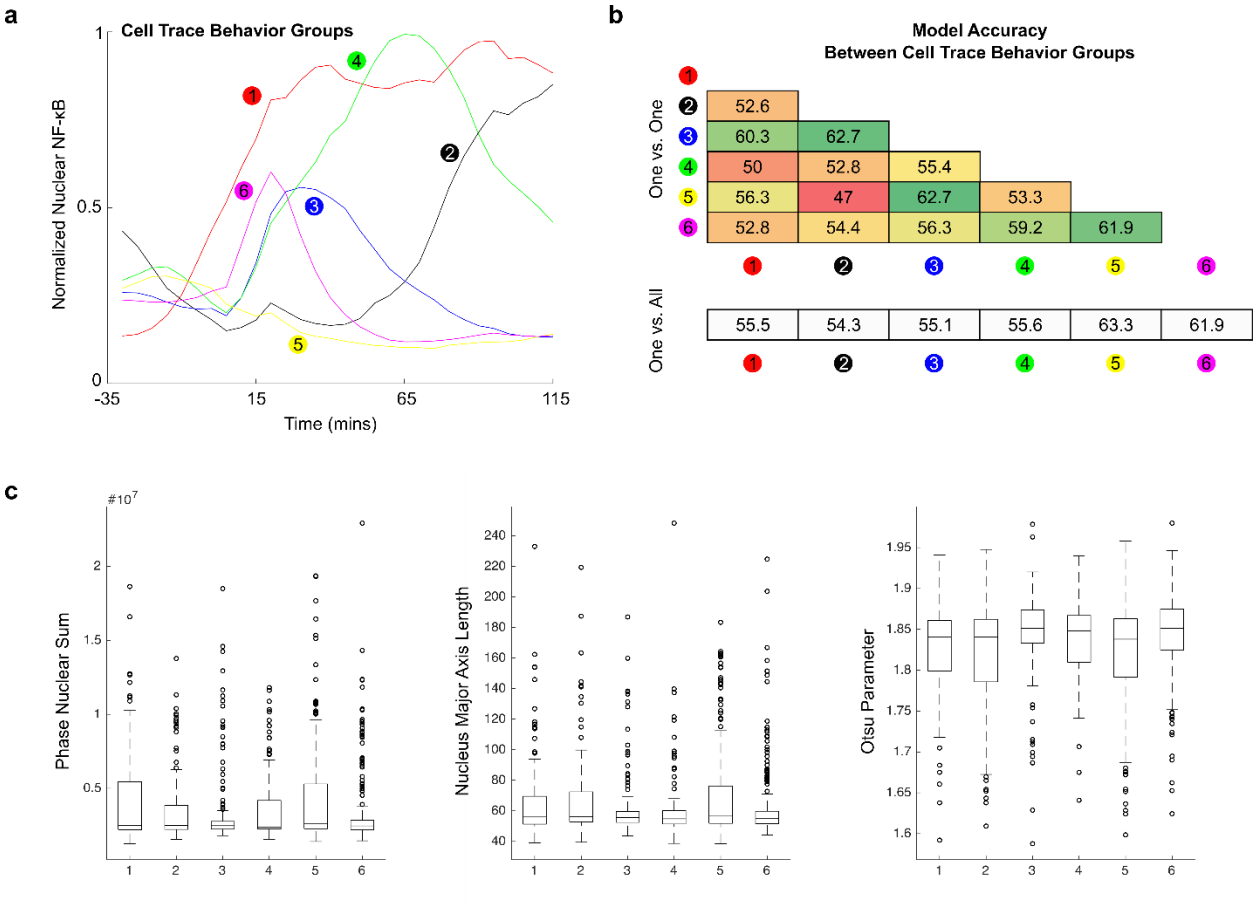
Supplementary Figure 1-9



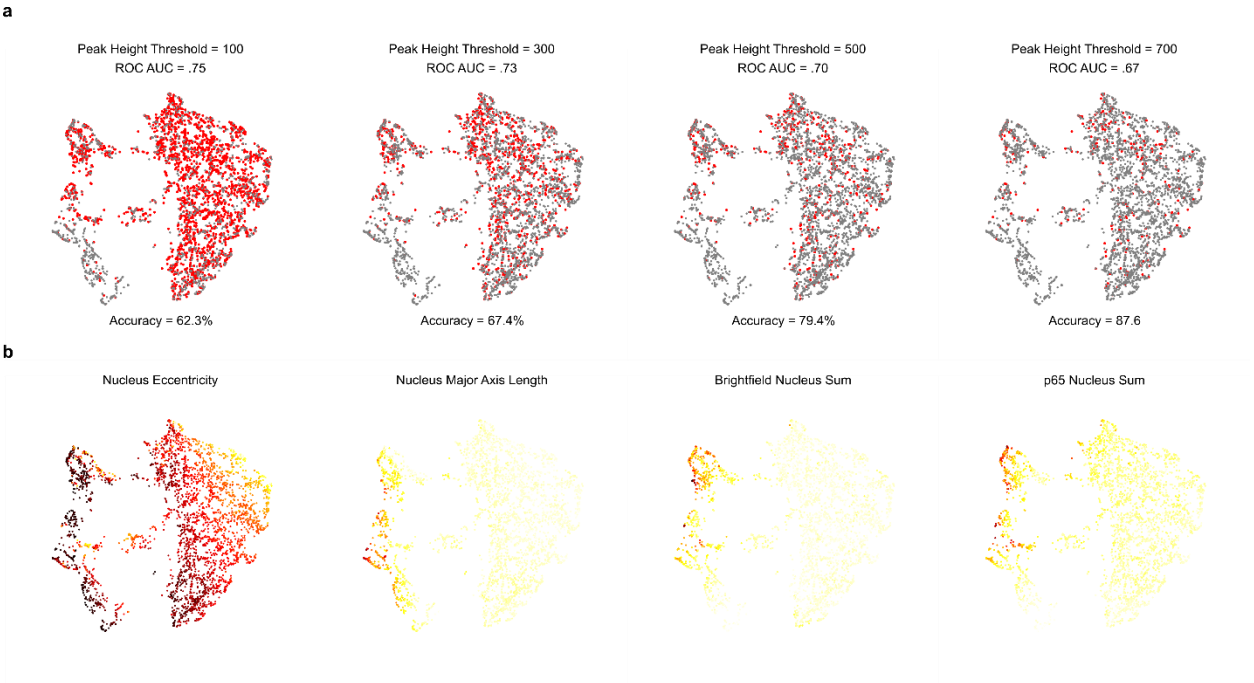
Supplementary Figure 1.1. Independent Repeat of single dose prediction for TNF activation **(a)** tSNE of unstimulated cells at 0.1ng/mL TNF. **(b)** Highly predictive features for each cell **(c)** Standardized descriptors for unstimulated fibroblasts are shown in boxplots with active cells (red) and inactive cells (blue)



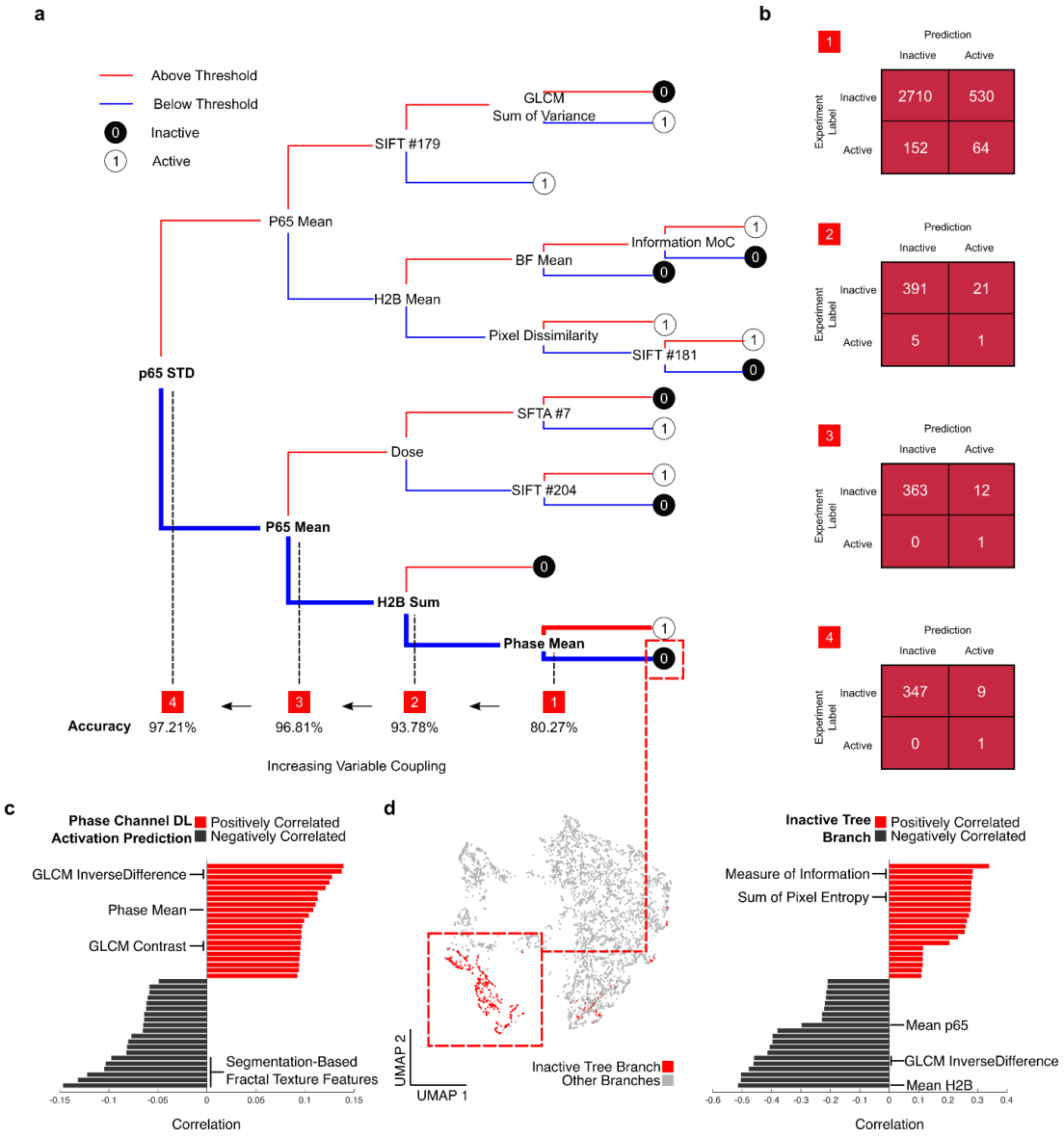
Supplementary Figure 1.2. Single cell population distribution in tSNE space (a) tSNE of different TNF dose stimulations (0.005, 0.05, 0.5, 5.0 ng/mL TNF) from t=0 descriptive variables. Cells from a dose are indicated in red and individual SVM model accuracies are listed. **(b)** tSNE visualization for NF-κB peak heights with overlay of fixed cells.



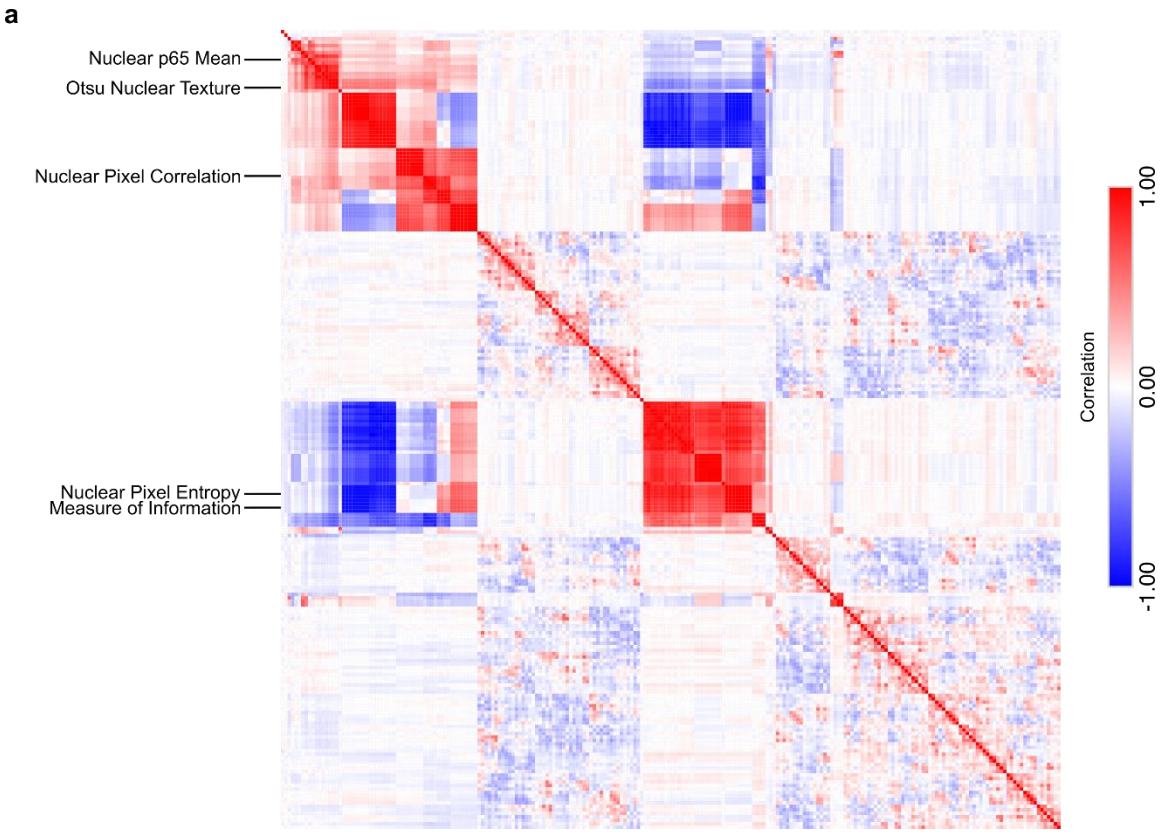
Supplementary Figure 1.3. Cell trace dynamics activation criteria for prediction shows modest prediction results (a) Single cell traces were clustered using k-means clustering (k=6). Median traces for each group is shown. **(b)** Classifier accuracies show how well individual groupings compare to other individual groupings and to all other groupings. **(c)** Some features, including phase nuclear sum, nucleus major axis length, and otsu parameter discriminate between trace behavior types.



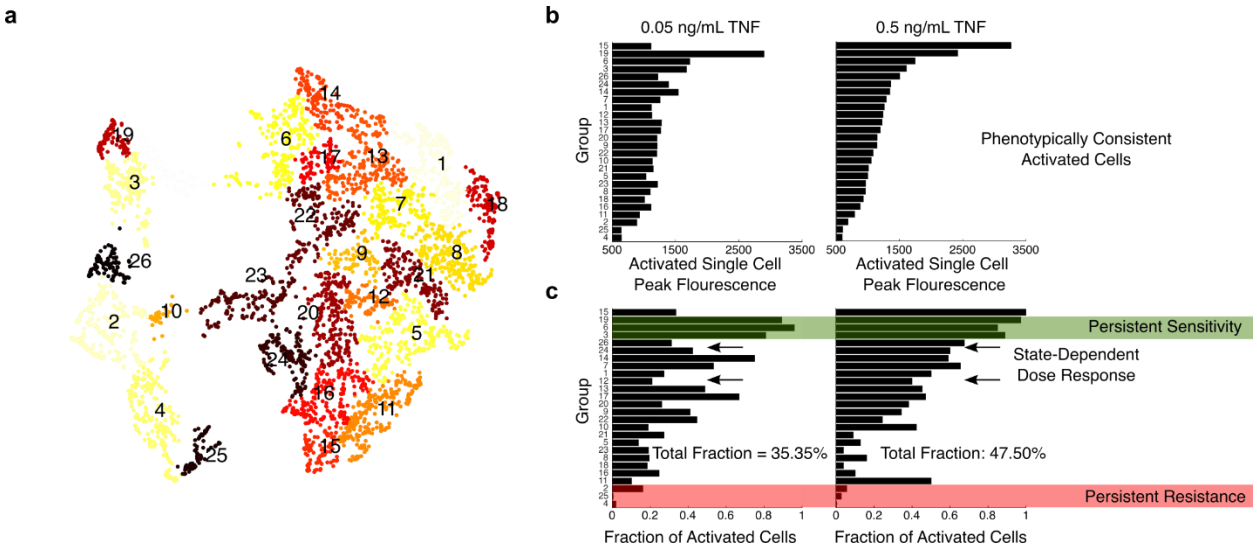
Supplementary Figure 1.4. Activation level thresholding shows minimal differences in ROC AUC (a) Cell label assignment using different mean nuclear peak height threshold values. Analysis throughout the manuscript uses peak height threshold of 500. **(b**) Cell features that track with the divide between clusters include nuclear size metrics like nuclear eccentricity and major axis length. Similarly, protein levels of p65 and total brightfield sum also track with this divide.



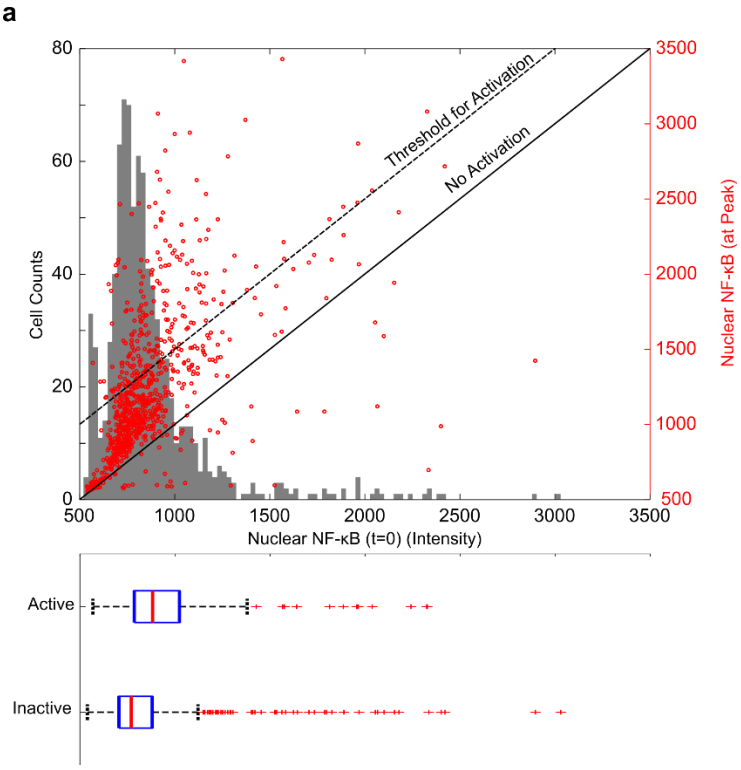
Supplementary Figure 1.5. Multiple variable coupling leads to improved prediction (a) A single tree model in an ensemble tree classifier is shown with thresholded decision boundaries for different variables (blue, under threshold; red, above threshold). An individual tree branch is analyzed for the power of variable coupling and accuracy is shown with the addition of addition features **(b)** Individual classification accuracy is shown in each stage in the analyzed tree branch. **(c)** Correlation between the phase classifier and feature variables is shown with no correlation between prediction labels and p65 or H2b features. **(d)** Cell mapping in the UMAP projection is shown for the group of analyzed cells in the individual tree branch. Correlation found for features in this grouping of cells is shown. Individual features have a higher correlation when predicting this individual grouping of cells.



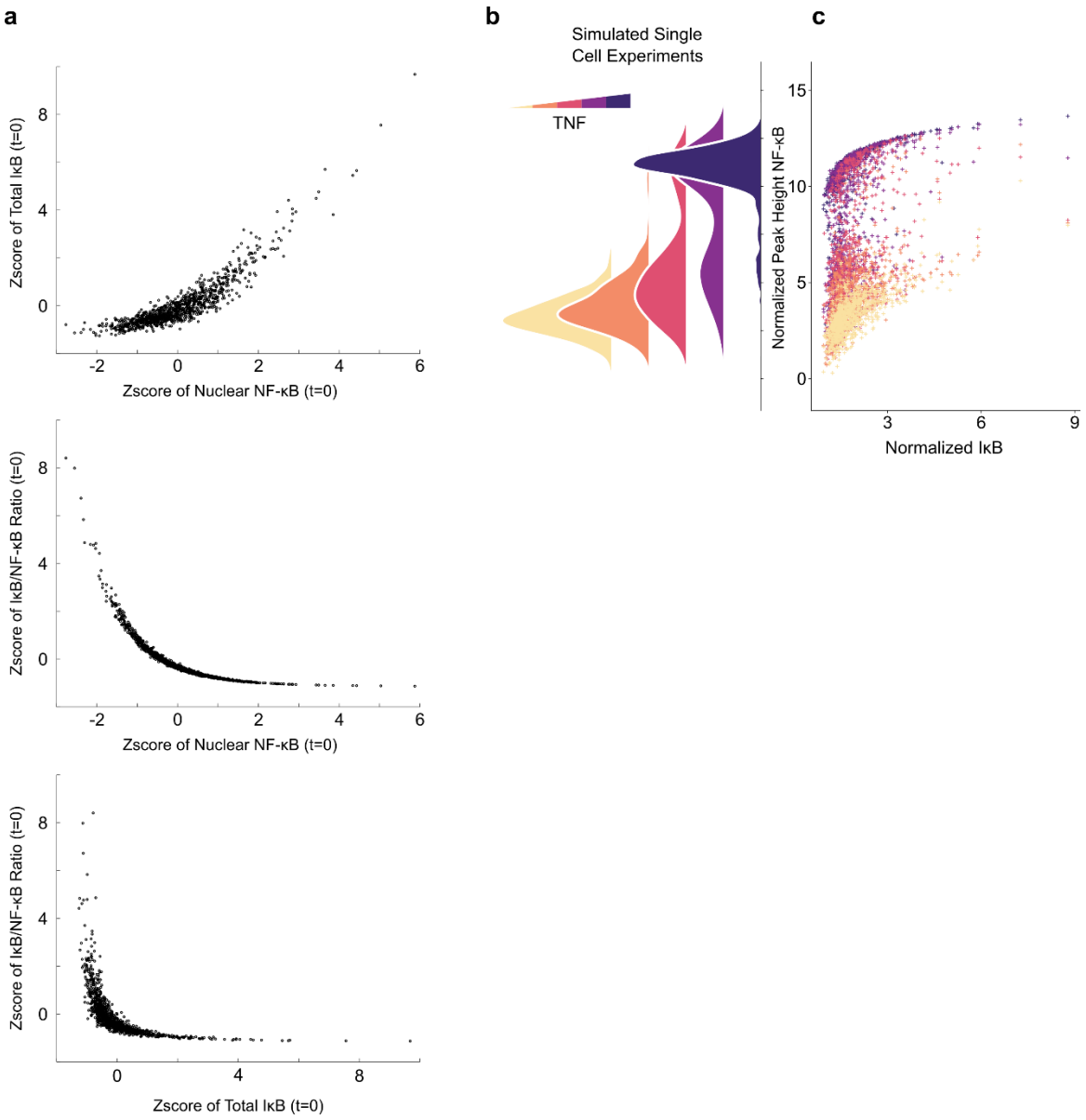
Supplementary Figure 1.6. Image feature covariance matrix (a) Covariance matrix for features colored by Correlation (red =1, blue =-1).



Supplementary Figure 1.7. Phenotypic analysis of cell state reveal both persistent and dose regulated states. (a) We use single cells across the UMAP space, clustered by local adjacency into communities to look for phenotypic differences in state by dosage **(b)** Activated single cells show consistent phenotype across different doses of TNF (0.05 ng/mL and 0.5ng/mL, left and right) **(c)** but show differences in the fraction of activated cells based on state. There are consistently sensitive and resistant groups of cell phenotypes, but also cellular states that have dose dependent differences in activation fraction



Supplementary Figure 1.8. p65 levels in unstimulated cells (a) Baseline levels of nuclear p65 is shown in the histogram. Cells that are classified as activated must reach a threshold of >500 peak-height. Boxplot is shown below comparing active and inactive classifications.



Supplementary Figure 1.9. Simulation comparisons between total IκB, IκB:NFκB, and Nuclear NFκB (a) Simulation comparisons of key features: total IκB, IκB:NFκB, and Nuclear NFκB. **(b)** Mathematical modeling simulations show single cell nuclear NF-κB peak heights increase with increasing TNF dose. **(c)** High IκB levels in cells require a smaller TNF dose to achieve NF-κB activation.

Environmental noise enables sensitive detection and transcriptional decoding of cytokine inputs

Parthiv Patel[1†], Weerapat Pittayakanchit [2†], Minjun Son[1], Arvind Murugan[2,3 *],

and Savaş Tay[1*]

[1]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637

[2]Department of Physics, The University of Chicago, Chicago, Illinois 60637

[3]The James Franck Institute, The University of Chicago, Chicago, Illinois 60637

* Correspondence: tays@uchicago.edu, amurugan@uchicago.edu

† Contributed equally

Summary:

Cells communicate effectively in noisy environments despite the detrimental effects of noise on information transfer. Understanding how cells deal with environmental perturbations when detecting and decoding biochemical inputs has immense importance for studies of healthy and dysregulated signaling. Here, we investigate how NF-κB pathway detects, discriminates and transcriptionally interprets noisy immune inputs, by using live-cell analysis and mathematical

modeling. Surprisingly, we find that single-cell NF-κB activation is dramatically enhanced through addition of noisy fluctuations to an otherwise constant cytokine signal. Cells perceive a noisy cytokine stimulus to be 10-times more potent than a constant one at the same average dose. Noisy stimulation expands cytokine specific transcriptional response and switches its profile to an anti-cancer program with acute inflammatory, anti-proliferation and anti-differentiation signatures. We find that this noise-enhanced signal sensitivity is explained by nonlinear adaptive dynamics – rectified adaptation – in the NF-κB network. Our results indicate a beneficial role for external fluctuations for the detection of otherwise undetectable weak signals, and show how cells exploit non-trivial molecular mechanisms to thrive in noisy environments.

Introduction:

Cells are exposed to randomly fluctuating levels of signaling molecules in their natural environment [1–5]. Burst-like protein secretion, thermal fluctuations, disordered tissue structure and interference from neighboring cells all contribute to the noisiness of the extracellular signaling environment[6–10]. Noise is considered detrimental to information transfer and signaling in general, both in physical and biological systems[11–13]. Despite constant influence of noise on cellular inputs, biological signaling networks manage to create robust and finely tuned gene expression profiles[14–16]. How cells deal with noisy external fluctuations and create appropriate signaling responses is not well understood, and it is central to many signaling problems in health and disease.

The well curated responses generated by cells may be the result of noise dampening features of gene regulatory network motifs[17]. These network motifs can potentially reduce the detrimental effects of noise and limit the response to input fluctuations through mechanisms like time-

averaging[17,18]. However, such features inevitably reduce the responsiveness (sensitivity), speed and temporal resolution of signaling systems. On the other hand, many cellular signaling pathways manage to combine high sensitivity and fast speed with noise tolerance, which are seemingly conflicting properties to exist in the same signaling system. Here, we carefully examine such a sensitive and dynamic signaling system central to immunity, the NF-κB pathway, for its ability in responding and potentially eliminating the effects of noisy fluctuations in a cytokine signal.

NF-κB is an important example of a dynamically regulated transcriptional network that responds to a range of signaling inputs relevant to immunity[19–22]. NF-κB controls the expression of hundreds of genes in response to a wide range of immune stimuli including signaling molecules secreted from host cells (i.e. cytokines and chemokines)[23–25]. NF-κB coordinates basic functions in innate immunity, immune development, and is involved in pathophysiological outcomes including autoimmune disorders and cancer[26–28]. Cells in the resting (unstimulated) state contain cytoplasmic NF-κB transcription factors, including p65. Upon cellular stimulation, NF-κB transcription factors rapidly shuttle to the nucleus and activate downstream gene expression. One of the transcriptional targets of NF-κB is its own inhibitor IκB. Newly synthesized IκB causes NF-κB to shuttle back to the cytoplasm, thereby creating oscillations in nuclear NF-κB and in target gene expression. These oscillations enable NF-κB to constantly monitor the signal types and amplitude in the extracellular environment and persist while external cytokine signals are present.

Increasing signal amplitude (cytokine dose) causes more cells in the population to activate, and do so in a faster way compared to lower doses[14,29]. NF-κB response time (activation speed) ranges from 15 minutes for the highest doses of cytokine inputs, and can be as slow as 90 minutes for the lowest doses. NF-κB has been a model system for understanding the role of protein

dynamics in signal specific gene regulation[19,30,31], and the influence of cellular variability and internal (i.e. transcriptional) fluctuations on dynamic NF-κB signaling has been studied[32,33]. However, the effect of noisy external inputs to an immune signaling pathway like NF-κB has not been studied to date, and how cells robustly communicate in noisy signaling environments remains a fundamentally important open question **(Figure 2.1a)**.

The addition of noise to a signal typically degrades the transfer of information to the output, both for physical and biological signaling systems[7,11,12]. Here, we ask how biological systems deal with input noise and address if signaling pathways like NF-κB can actually utilize noise in the input to improve signaling fidelity. We also investigate how input noise influences the transcriptional decoding of cytokine inputs received by cells. Theoretically, the sensitivity of a system to small signals can be improved by addition of white noise through a range of effects collectively called stochastic resonance[34]. In stochastic resonance, frequencies in the noise spectrum can amplify similar frequencies in the responding system and create greater output[11]. The precise mechanism of stochastic resonance varies, but in general, noise in the input induces a state change or excitation[11] that would not happen in the absence of noise. We reasoned that biological signaling systems like NF-κB could exhibit stochastic resonance since they are non-linear and exhibit dynamics of different time scales. NF-κB has multiple feedback loops leading to oscillations, and exhibits digital activation[14,35]: an obvious example of nonlinearity that could benefit from noise. Despite the favorable theoretical arguments, stochastic resonance has not been experimentally seen in living cells, and whether it plays a role in immune signaling is not known.

Results

Dynamic modeling predicts that input noise enhances NF-κB activation in single cells

We hypothesized that the multiple feedback time-scales in the NF-κB network could resonate with various frequencies in noisy biochemical inputs and create a more robust activation effect in live cells **(Figure 2.1b).** To explore this possibility, we first simulated the mathematical model of the NF-κB network in single cells under both constant and fluctuating (noisy) concentration of the pro-inflammatory cytokine TNF **(Figure 2.1c)**. The simulated input signal, when supplied at a constant (noise-free) concentration results in minimal NF-κB translocation (**Figure 2.1d)**. Surprisingly, a noisy signal at the same mean dose causes cells to robustly activate, with 100% of cells in the population activating the NF-κB pathway. To confirm that cells were not responding to the maximum TNF values in the fluctuating noisy signal, we modeled a noisy TNF signal at an even lower level than the constant signal, and observed that the cells still activated at a rate higher than the constant signal **(Figure 2.1e)**. Our simulations predicted that a noisy input will activate a much larger fraction of cells in the population than a constant input dose, even at the maximum value of the noisy input. Thus, our simulations suggest that a low dose input can activate cells with the addition of chemical noise, and that this enhanced response cannot be explained by the maximum dose alone. Such simulations show potential for subthreshold signal detection with the addition of noise and imply that noise can help increase the sensitivity of NF-κB response.

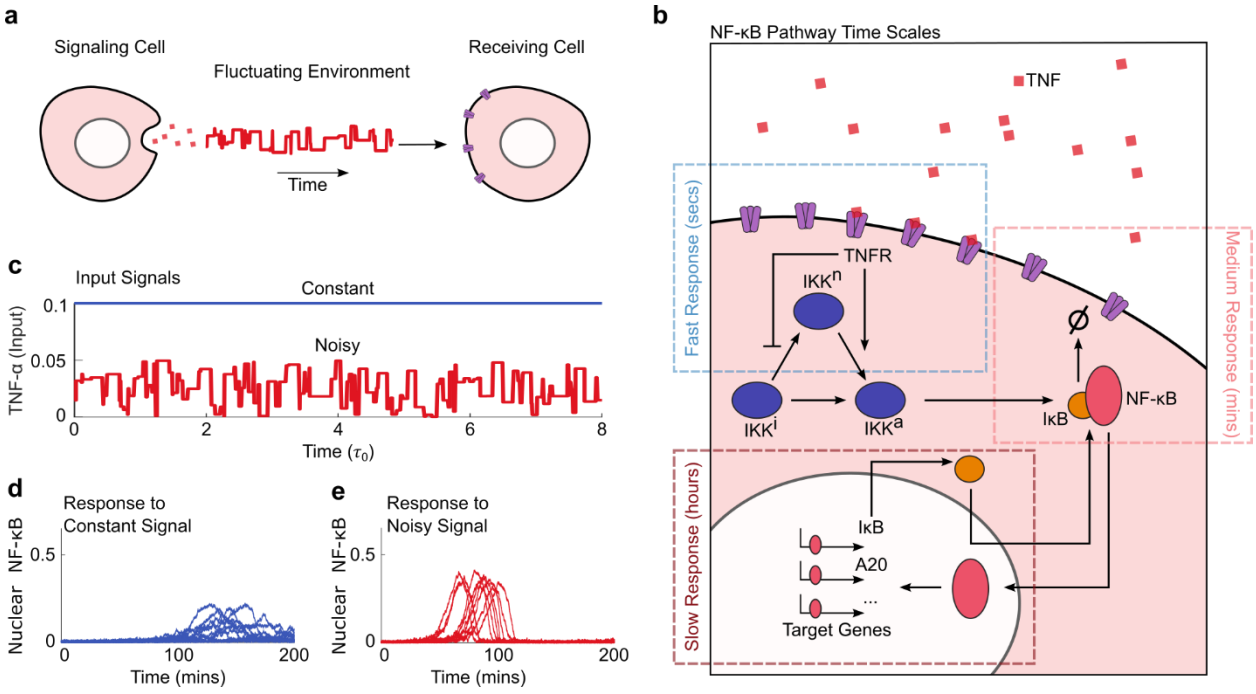


Figure 2.1: Simulations predict that weak (sub-threshold) cytokine inputs may evoke a strong NF-κB response upon addition of chemical noise to those inputs. **(a)** Extra-cellular environments are subject to chemical fluctuations, which make signals between cells to become noisy. Signal receiving cells process such noisy inputs. **(b)** NF-κB pathway has complex non-linear dynamics on a wide range of timescales, from fast receptor dynamics to the slower pulse-like nuclear translocation of NF-κB. **(c)** Dynamic input profiles used for simulations with the mathematical model of NF-κB pathway, under both constant (blue) and fluctuating (red) input levels from the cytokine TNF. **(d, e)** In simulations, fluctuating TNF levels cause strong activation of NF-κB nuclear translocation in all cells. However, constant TNF inputs induce lower NF-κB activity and in fewer cells, despite the fact that TNF dose they see exceed maximum of the fluctuating input.

Live-cell experiments reveal increased NF-κB activation under noisy cytokine signals

To experimentally test whether addition of chemical noise to a cytokine input would indeed enhance the activation of NF-κB in live cells, we studied nuclear localization dynamics of NF-κB in response to real-time fluctuations of TNF. We cultured 3T3 mouse fibroblast cells with p65-

dsRed reporter in an automated microfluidic cell culture platform, and quantitatively measured single cell NF-κB nuclear translocation in live-cell imaging experiments under a range of time-varying TNF concentrations **(Figure 2.2a)**. Our automated microfluidic device delivers time-varying dynamic chemical signals to cell culture chambers, including signals that contain predetermined levels and frequencies of dose fluctuations. The device creates 154 pre-determined concentration levels between 0.01 ng/mL and 0.11 ng/mL TNF with 0.5 second timesteps[36] **(Figure 2.2b)**, which is much faster than the typical timescales of NF-κB activation dynamics measured in live cells (tens of minutes). Cells were incubated in this device and exposed to constant TNF signals whose concentrations range from 0.001 ng/ml to 1ng/mL **(Supplementary Figure 2.1; Supplementary Figure 2.2)** and a noisy signal (Gaussian white noise) with mean concentration of 0.05ng/mL and Signal to Noise Ratio (SNR) of 5/1 **(Figure 2.2c; Supplementary Figure 2.3)**. During stimulation experiments, cells were imaged with time-lapse fluorescence microscopy, and individual cells were tracked using custom image processing software[37].

We found that cells responded much more sensitively to the noisy input: The fraction of responding cells in the population as well as the nuclear NF-κB amplitude in individual cells showed significant enhancement under the noisy signal when compared with an equivalent dose constant signal (constant dose at 0.05ng/mL) **(Figure 2.2d; Supplementary Figure 2.3)**. A randomly selected set of activated single-cell traces demonstrate this dramatic increase in NF-κB activity **(Figure 2.2e and 2.2f)**. In experiments with constant TNF at 0.05 ng/mL, only 33% of cells are activated, but under Gaussian white noise stimulation at the same time-averaged TNF dose, 85% of the single cells in the population are activated **(Figure 2.2i)**.

The amplitude of single cell NF-κB traces measured under the low-dose noisy signal are similar to those stimulated with much higher constant doses of TNF (mean integrated TNF dose for noisy signal is 0.05ng/mL). Furthermore, the fraction of responding cells is also much higher under the noisy input. To achieve the 80% activated fraction under constant input, we needed to use 1 ng/mL constant TNF, which is a 20-fold dose increase over the mean profile and 10-fold increase over the max value of the noisy signal **(Figure 2.2g)**. These results clearly show that individual cells perceive a noisy signal to be more than 10-fold more potent and respond accordingly. Such noise enhanced signal response is highly surprising and was not experimentally seen in living cells before.

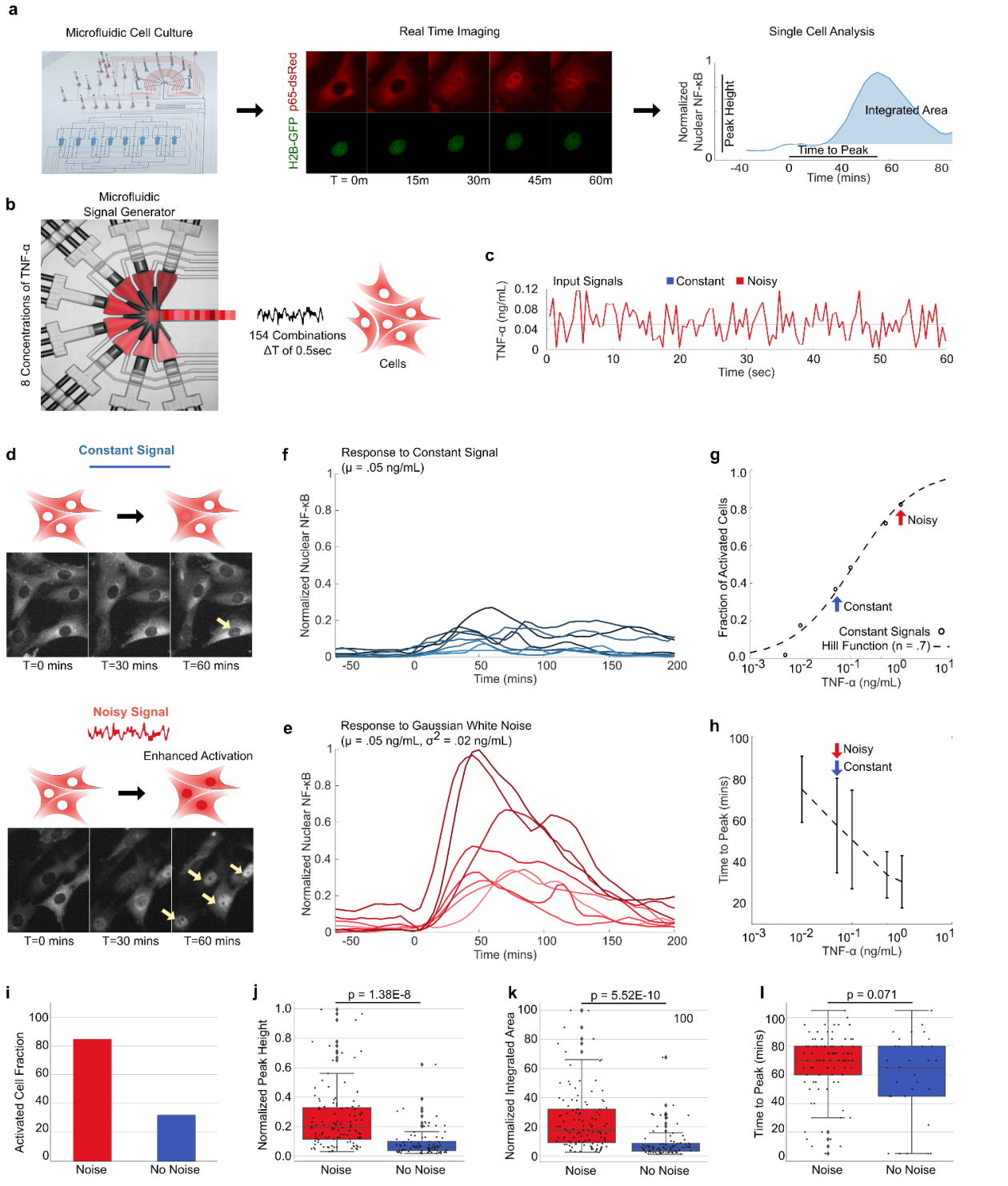


Figure 2.2 : Live-cell stimulation experiments show that addition of white noise to a weak TNF input causes NF-κB to respond very sensitively to that input. **(a)** Microfluidic live cell stimulation experiments allow analysis of single cells in dynamic environments. Cells cultured under dynamic signals are tracked via live cell microscopy, and NF-κB nuclear localization is analyzed in

Figure 2.2 (continued) individual cells. **(b)** Cells are exposed to signals generated by dynamic combination of several TNF concentrations, producing 154 variable exposures at 0.5 second timesteps. Actual picture of the microfluidic device is shown on the left. Colors indicate controllable valves that regulate input media. **(c)** Gaussian white noise with 5:1 signal to noise ratio (red) is added to the control TNF signal with a mean at 0.05 ng/mL (blue), and this chemical signal is delivered to hundreds of live cells. **(d)** Fluorescent images of cells expressing dsRed-p65 reporter that are exposed to noisy (red) or noise-free (blue) signals. Activated cells, showing nuclear NF-κB localization, are indicated with arrows. **(e)** Single cell traces under constant TNF exposure show significantly less NF-κB translocation to the nucleus. **(f)** Activated single cell traces measured under noisy input. **(g)** Population fraction of activated cells under different constant doses of TNF, measured in the same device. Blue and red arrows indicate the dose perceived by cells in stimulation experiments with constant and noisy inputs in (e) and (f). The noisy input is perceived by cells to be at a much high dose than the constant input. **(h)** NF-κB response time at different constant doses of TNF. Both noisy and noise-free inputs lead to the same response time. **(i)** Comparison of activation fraction for noisy vs. constant signals. Noisy input induces more cells to activate in the population, despite that fact that both inputs have the same mean TNF dose. **(j)** Comparison of peak height normalized to maximum peak value for noisy vs. constant exposure. **(k)** Comparison of integrated area of fluorescent readout for nuclear translocation normalized to maximum integrated area for noisy vs. constant exposure. **(l)** Peak timings for activated cells remain the same despite variation in peak height and integrated area.

There are several components of the single cell NF-κB localization profiles that are relevant when comparing NF-κB activation. First peak height and integrated area of nuclear fluorescence are linked to target gene expression; different levels of NF-κB in the nucleus are correlated with different phenotypic gene response[14,25]. For a noisy input, while the normalized NF-κB peak height and integrated area is consistent with a dramatically higher TNF dose **(Figure 2.2j and 2.2k)**, however we found that the NF-κB response time (time to first peak) is similar to the noise-free stimulation. Both noisy and noiseless signals show the same response time, with the same single-cell distribution and mean around 65 mins, which is consistent with activation at a constant dose of 0.05ng/mL **(Figure 2.2h and 2.2l)**. This result shows that while noisy fluctuation of TNF signal increases the perceived dose for NF-κB activation and amplitude, the speed of NF-κB activation is not increased due to the addition of noise.

Overall, our experiments confirmed the modeling prediction of dramatically increased noise-induced activation of NF-κB in living cells; further, the activation arises as a cellular response to input fluctuations and cannot be explained as perception of the maximum dose. This remarkable demonstration of noise enhanced signaling activation in living cells highlights how biological systems can exploit non-trivial physical effects to take advantage of noisy environments.

Noisy cytokine input leads to switching of transcriptional programs to anti-cancer responses

To evaluate the functional effects of cytokine input noise, we probed gene expression response in cells under both noisy and constant TNF stimuli. We stimulated cells again through our automated microfluidic cell culture platform, then extracted cells at 90min, 180min and 300min timepoints. Approximately 150 cells were lysed and frozen for pooling in RNA sequencing (SMART-seq) for each timepoint in triplicate (**Figure 2.3a**). Analyzing the whole transcriptome gene expression differences between the two conditions reveals many differences across a range of transcriptional programs. We looked at all genes that were significantly upregulated or downregulated (pval < 0.01, and foldchange > 0.5) through noisy stimulation compared to noise-free stimulation, and found 2540 genes that had significant variation of which a majority are upregulated in response to noisy stimulation. This indicates that noisy input largely enhances or broadens gene expression response compared to a constant cytokine signal, consistent with increased NF-κB activity in these cells. We narrowed down our analysis to the two broad categories of noise sensitive gene expression response using hierarchical clustering at two different time scales after exposure to TNF (early: 90 min, and late: 300 min).

We applied gene ontology analysis (GO) to these significantly changed genes using GSEA[38]. Gene ontology aggregates annotations and known information about genes to simplify and standardize their function[39]. We can then analyze GO overlap between our significantly changed genes to interpret how expression changes can lead to functional changes in the cell. Gene ontology showed upregulation in genes related to cell cycle (*e2f, Hdac3*), ribonucleotide binding (*e2f, zbtb12*), anti-cancer programs (*p53, Jdp2, Sall4, Ubn1, Tead2*) and chromosome modifying activity (*Mta1, Ubn1, Barx2)* at the early time point under noisy stimulation. (**Figure 2.3b**). There is also upregulation of NF-κB coregulators: transcription factors with >15% of total upregulated target genes include *Jdp2, Mta1, Mzf1, Sall4, Hdac3, e2f* and *Klf13*[40–44]. These genes bind to and have known activity with NF-κB. Such modifications of NF-κB and its co-regulators during the early response hint at differences in gene expression that will appear at the later time point. Further, this transcription factor response alongside the broad enhancement of the DREAM complex and p53 program indicate a cellular effort to limit differentiation and proliferation while preparing to address DNA damage within the cell[45]. This coordinated response at the early 90min timepoint suggest that a noisy environment could induce a programmed response to prevent cancer progression and proliferation of damaged cells[46].

While some of the early transcription factors continue to be upregulated highly at the 300m timepoint (*Hdac3, Klf13, Sal4, JDP2*), several new transcription factors emerge at this time point including *Lef1* from the WNT pathway and PRDM6 which inhibits proliferation (**Figure 2.3c**). The chromatin regulating protein *Morc2* is upregulated in both early and late noise upregulated genes. *Morc2* is responsible for many gene silencing effects via modification of chromatin through interaction with the HUSH complex[47]. It is of note that at the late timepoint, there is also an

increase in transcriptional regulation and protein phosphorylation and modification activity. While many of the upregulated gene expression programs point towards increased cell cycle regulation, there is also upregulation in anti-cancer markers to limit proliferation as well[48–50] (**Figure 2.3d and 2.3e**). Exposure to noisy stimuli thus evokes enhanced expression of anti-proliferation and anti-differentiation gene expression programs at late time points as well.

These upregulated gene expression programs combined suggest that noise in the signaling environment may create a more robust anti-cancer response in the NF-κB transcriptional response (**Figure 2.3f**). The significant difference in transcriptional phenotype of cells under noisy and noise free cytokine levels in the environment is consistent with the pleiotropic response of NF-κB under different physiological conditions[51]. It has been shown that different NF-κB temporal dynamics create different gene expression responses[52], thus the noise-responsive phenomenon we observed appear in agreement with NF-κB studies that show the adaptability of innate immune response to different environmental signals.

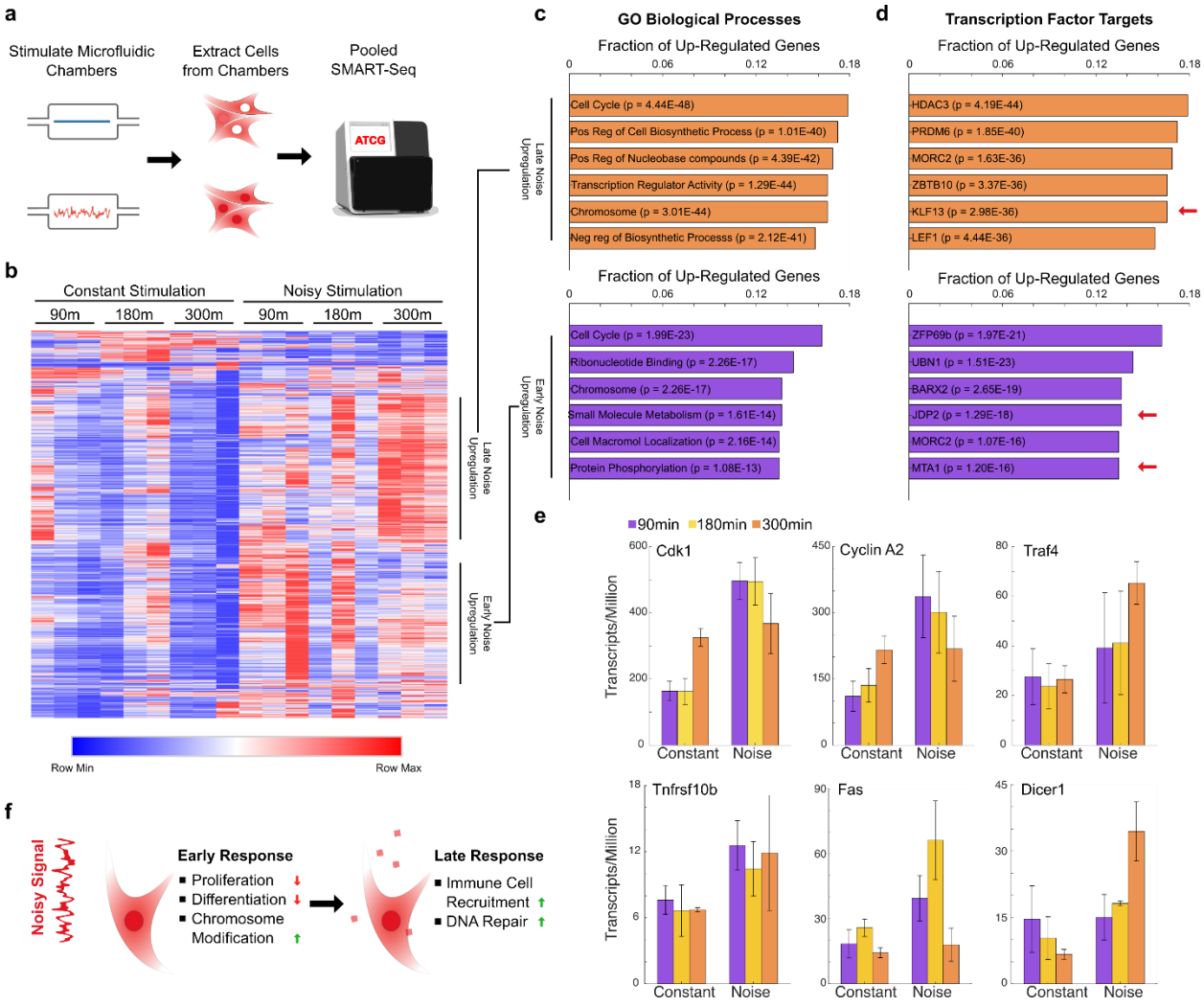


Figure 2.3. RNA sequencing analysis shows gene expression programs that significantly change under noisy TNF stimulation. **(a)** Cells were loaded and stimulated in microfluidic chambers, then extracted and pooled for SMART-sequencing. **(b)** Heatmap of genes that are expressed at significantly different rates under noisy and constant input, for each timepoint ($p < .01$, fold-change >.5) **(c)** Gene Ontology for early and late genes show upregulation of cell cycle and chromosome modification related gene expression under noisy TNF stimulation. Late term genes show increased transcription regulator activity under noisy input. **(d)** Transcription factors that are upregulated under noisy input include NF-κB coregulators (shown with red arrow) as well as transcription factors lef1 and Hdac3. **(e)** Expression of cell growth, anti-cancer, and anti-proliferation gene markers are also increased under noisy stimulation. (**f)** Functional summary of the gene expression changes under noisy TNF stimulation. These changes point to increased immune cell recruitment, reduced differentiation, reduced proliferation, and activation of DNA damage repair programs.

Early modifications caused by noisy input change the expression of TNF specific genes at later timepoints

Whole transcriptome analysis under noisy stimulation indicated upregulated chromatin modifications at early time points and upregulation of NF-κB co-regulators. These early modifications may change NF-κB target gene expression at the later time points. To explore this possibility we analyzed expression differences between noisy and constant TNF stimulus in genes with NF-κB specific binding sites[26,53]. We use hierarchical clustering to cluster genes that are amplified by or dampened by noisy stimulation (**Figure 2.4a**).

Interestingly, genes that typically respond within the first 90 minutes after TNF stimulation show similar expression profiles between noisy and constant stimulation despite the increased overall gene expression response under noisy input. These include the transcription factors *Junb* and *Stat5a* as well as regulatory proteins *Tnfaip3* and *Nfkbia*. At the later time points, however, the NF-κB transcriptional response diverges between the noisy and constant stimulation groups at 180 minutes and 300 minutes. This behavior suggests that the whole genome transcriptional activity that we observe in the early 90 min timepoint dictates and changes the specific NF-κB response for later time points. The early NF-κB specific response remains largely unchanged despite interaction with NF-κB cofactors like *Hdac3* and *Klf13*. However, many NF-κB target genes show upregulation upon noisy stimulation only at later time points[54,55]. (**Supplementary Figure 2.4**). This suggests that the chromatin modifying and NF-κB coregulator transcription

factor activity we observe at 90 minutes is changing the NF-κB-specific response at later time points[56].

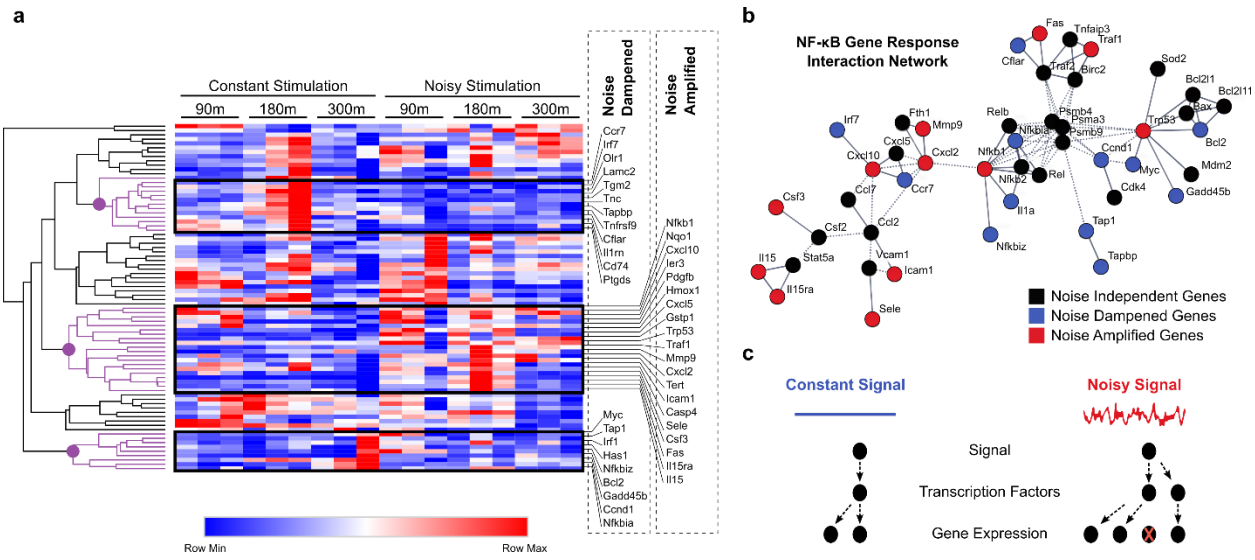


Figure 2.4. Noisy stimulation achieves early NF-κB pathway modifications that modulate TNF specific gene expression at later time points. **(a)** NF-κB associated genes were isolated and compared across stimulation patterns and time. Most genes increase in amplitude under noisy stimulation, but some are dampened. Genes that respond within the first 90 minutes show similar expression profiles between noisy and constant stimulation. However, gene expression responses at later time points (180 and 300 minutes) are significantly different under noisy and constant stimulation, both for noise amplified and noise dampened genes. **(b)** Mapping the highest confidence interactions between NF-κB responsive genes shows that the noise amplified genes stem from Cxcl2 interacting partners, whereas noise dampened genes are affiliated with p53 inhibition and canonical NF-κB response. (**c**) Overall, noisy stimulation changes NF-κB-specific gene expression and also increases the bandwidth of total gene expression through multiple coregulated transcription factors. Under constant TNF signal fewer transcription factors and target genes are activated. Under the noisy signal at the same dose, a higher number of transcription factors and a larger gene expression amplitude is achieved.

To understand where NF-κB specific response functionally changes with the addition of noise, we mapped the interactions of the response genes to each other using STRING[57]. STRING

is a platform that maps interactions between different proteins using historical knowledge, and enables understanding and visualizing interactions in a pathway. We use STRING to create and analyze a network of functional interactions between NF-κB specific proteins. We then specifically map the noise dampened and noise induced genes onto the STRING network. High confidence protein interactions are shown in a map of NF-κB response proteins and we then use the clusters generated through hierarchical clustering to assign each protein in the interaction network a grouping based on expression behavior given noisy stimulation (**Figure 2.4b**).

Plotting the network of these response genes shows that genes amplified under noisy input stem from *Nfkb1*, including the *CXCL*, *CCL*, and *IL* proteins, whereas noise dampened genes are affiliated with canonical NF-κB response given TNF and p53 inhibition (*Myc*, *Bcl2*). Our analysis of the general whole transcriptome programs given noisy stimulation add insight on this result as well. Cellular release of these Cxcl cytokines is used in neutrophil recruitment[58,59] and lef1 is also known for mediating cell-cell communication. IL and CCL proteins are also known cell-cell communication cytokines involved in immune cell recruitment. The upregulation of these cytokines as well as our prior observation of upregulation of anti-differentiation and anti-proliferation programs show remarkably coordinated action by cells from noisy cytokine signaling. These programmed behaviors suggest that a noisy TNF signal can prepare a microenvironment for an inflammatory response through immune cell engagement while limiting the growth potential of the effected cell. Changing the dynamics of stimulation has many different effects on gene expression. Here we observe a noisy stimulation inducing NF-κB-specific gene expression changes and also an overall increase to the bandwidth of total gene expression through activation of multiple coregulated transcription factors (**Figure 2.4c**).

Rectified adaptation as a mechanism for noise enhanced sensitivity in NF-κB

To study the mechanisms behind input noise enhanced NF-κB activity, we theoretically evaluated how fundamental components of the NF-κB signaling network respond to fluctuating inputs. In particular, adaptive dynamics[17,18,60,61], known to be present in the NF-κB pathway[62,63] and also widely across biology, show a large response to a change in input but the output soon returns to its prior resting value. Hence, the output of adaptive dynamics is naturally more sensitive to input fluctuations than to the steady state level of that input. Adaptive dynamics are commonly implemented by either Incoherent FeedForward Loops (iFFLs) or negative feedback loops.

It is known that iFFLs are present in the NF-κB pathway[62–64], and given our knowledge of gene expression changes with the addition of noise, we simulated a general iFFL in a simple input-output model upstream of IKK **(Figure 2.5a)**. It is also known that the IKK family proteins are mediators for p53 and e2f as well as activators for *Nfkb1* signaling[65–68]. The gene expression changes we observe suggest that network features surrounding IKK could be responsible for some of the gene expression variability we observe. Under a general iFFL, upward and downward fluctuations in the input signal cause equal and opposite changes in the response, thus cancelling each other. Consequently, noise in the input has no net impact on the output for a regular iFFL network (**Figure 2.5c**). However, we found that a simple modification of the iFFL led to a dramatically different conclusion (**Figure 2.5b**); if protein concentration (Y) in the network has a natural floor, the response to positive and negative input changes is asymmetric. Such a node in the network acts much like a diode or a `rectifier' in an electrical circuit, allowing responses to only upward but not downward fluctuations in the input. Such rectification in the presence of noise

leads to counterintuitive behaviors in physical systems, from molecular motors to Feynman's ratchet[69]. In the NF-κB signaling context, we find that such a floor-rectified iFFL responds predominantly to upward fluctuations in the input, and thus can lead to a higher sensitivity to noise as seen in experiments (**Figure 2.5d and 2.5e**).

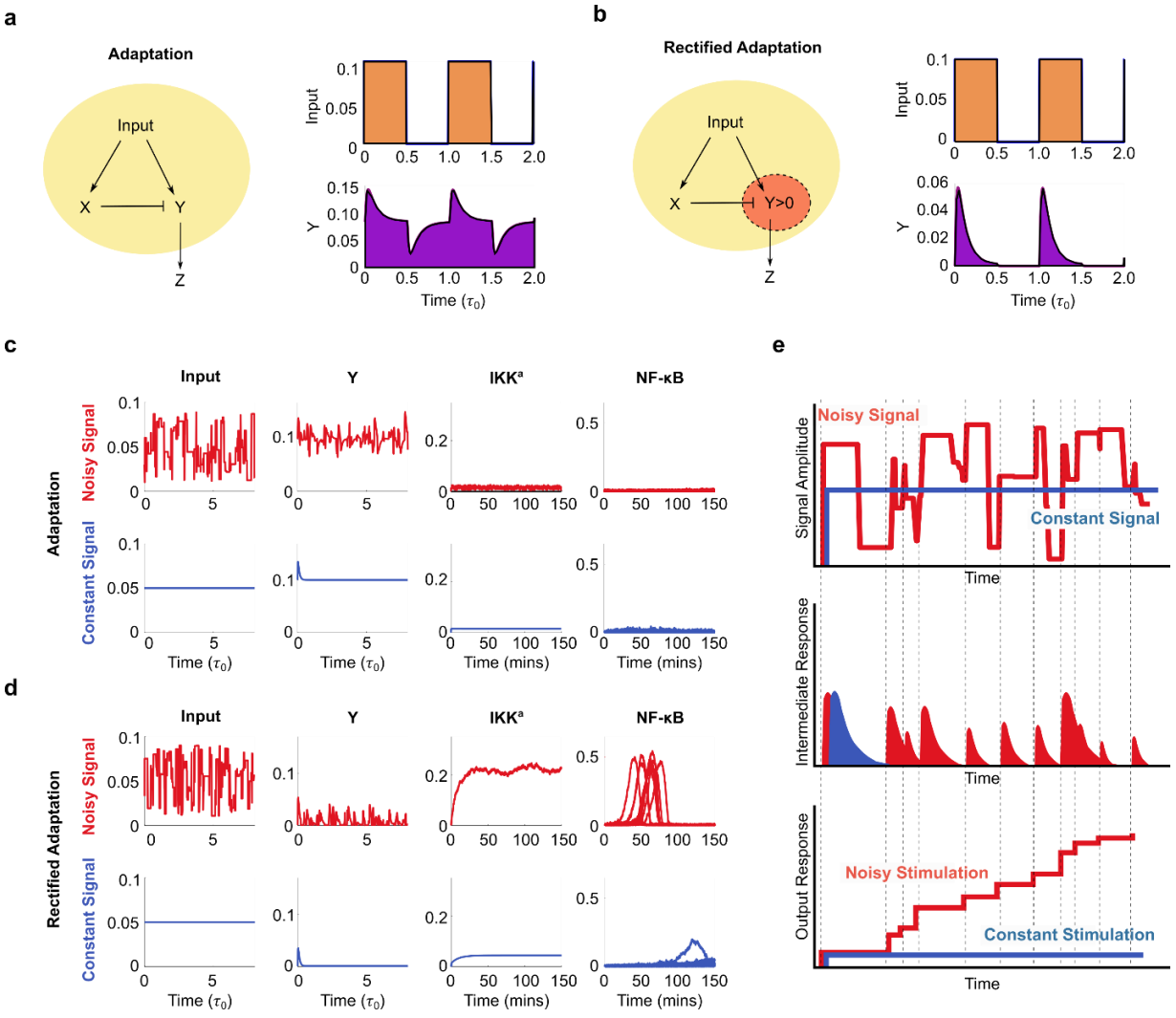


Figure 2.5: Rectified adaptation leads to enhanced NF-κB activation in simulations. **(a)** The Incoherent Feed Forward Loop, shown in the schematic on the left, demonstrates adaptation. A step-like change in the input produces a transient response in Y that dies out and the system returns to the resting level. Such motifs are found in the NF-κB pathway and molecules like IKK are

Figure 2.5 (continued) known to show an adaptive response to step changes in TNF. **(b)** Rectified adaptation, where step ups produce a response, but step downs produce no response because the output molecule Y is already at near-zero levels in steady state. **(c, d)** When subject to fluctuating input signals, the response of the conventional adaptive circuit to step ups and downs cancel each other out. However, the output of the rectified adaptation circuit builds up over time, since it only responds to step ups of TNF input. Consequently, active IKK levels rise with time, causing nuclear translocation of NF-κB in simulations. (**e**) Summary of the theoretical mechanism behind NF-κB enhancement by the interaction of noisy input and rectified adaptation. Through rectified adaptation, upswings in the noisy input produce pulses in intermediate protein activation, while downswing in the input do not. This leads to an increasing NF-κB output response under noisy input that accumulates over time. This mechanism also reproduces the slow NF-κB response time we observed under noisy stimulation in Fig 2.

Sinusoidal stimulation reveals resonant timescales in NF-κB activation

To evaluate the timescales involved in noise sensitivity of NF-κB, we derived a prediction for NF-κB response to periodic single frequency stimulation. In stochastic resonance, the response is highest when the input timescale (i.e. various frequencies in the noisy input) matches the corresponding timescale in the responding system. Our simulation suggested that NF-κB activation should be the highest for inputs whose period match the timescale of the adaptive response. For longer input periods, cells are expected to behave like they are simply experiencing periodic stimulation at the maximum seen dose (**Figure 2.6a**). The model also predicts that cell activation will saturate at frequencies higher than the adaptation timescale (**Figure 2.6b**).

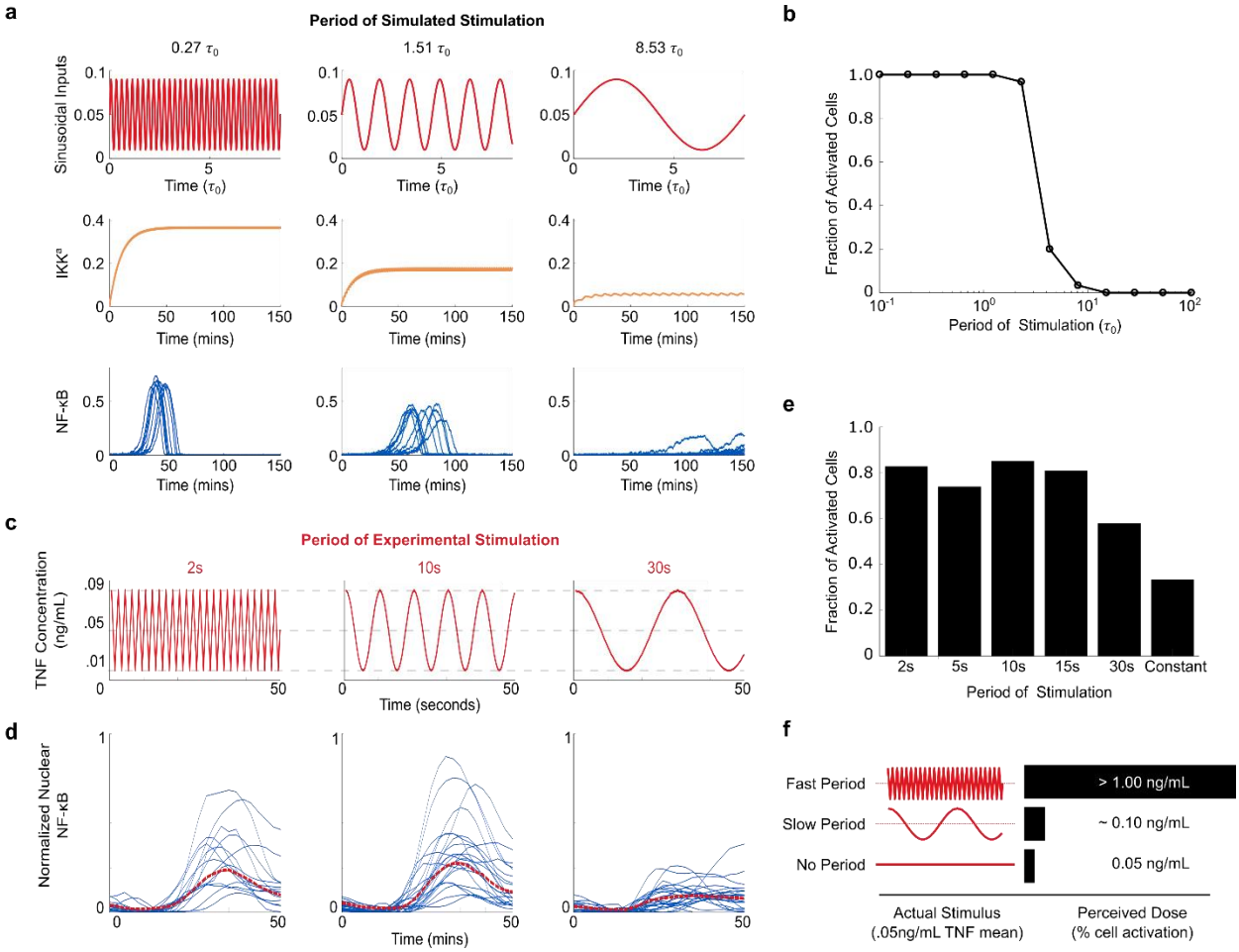


Figure 2.6: Periodic stimulation uncovers time-scales behind input-noise enhancement of NF-κB response. NF-κB shows resonance under fast oscillating TNF inputs. **(a)** Model simulations allow deriving predictions for response to different periodic inputs. Simulations show NF-κB activation only for periods comparable of faster than the adaptation timescale τ_0. For longer periods, the IKK response to different cycles of TNF input do not build on each other and there is no NF-κB activation. **(b)** The model also predicts that the fraction of cells activated saturates at a high fraction for frequencies higher than the adaptation timescale. **(c)** Experimental stimulation profiles for cosine exposure at different periods ranging from 2s to 30s. **(d)** Activated single cell traces in live cell imaging experiments show similar profiles for exposures under 30second period, but dramatically reduced peak height for 30second period (mean of populations in red). **(e)** Low period cosine stimulation increases fraction of activated cells in the population but increasing period of cosine stimulation after cutoff reduces activated fraction. **(f)** Cells exposed to slow periods behave as an intermediate between cells exposed to fast periods and constant signals. There is a 10-fold increase in the perceived dose with the addition of fast fluctuations to the signal.

To experimentally test our predictions, we measured single cell activation using a periodic cosine TNF stimulus at different periods, ranging from 2 s to 30 s (**Figure 2.6c**). These chemical stimuli are generated using our microfluidic device, and cells were imaged by time-lapse microscopy during stimulation. Activated single-cell traces show higher NF-κB peak heights for stimulation periods under 30 s (**Figure 2.6d; Supplementary Figure 2.5**) and faster response times with shorter periods (**Supplementary Figure 2.6**). For 30s and higher**,** cells fail to show the significant activation seen under stimuli with shorter periods. When comparing activation profiles under different time-scales, we find that cells perceive fast time scales (<15 sec period) at an effective constant dose greater than 1 ng/mL TNF, which is much higher than the actual used mean dose of 0.05 ng/mL in these experiments. These results show that NF-κB resonated at the input periods shorter than 15 seconds. As we increase the timescale to above 15 sec, cells perceive the stimulus at an effective dose of approximately 0.1 ng/mL TNF **(Figure 2.6e and 2.6f)**. Similar misperception of oscillatory signals on a minutes timescale has been previously observed in yeast[70,71]. Our experimental results here confirm the predictions from the model and suggest that noise perception at fast time scales (~ 5 s to 15 s) via a rectified Incoherent FeedForward Loop is contributing to the increased NF-κB activation observed in experiments with gaussian white noise.

Discussion

In summary, our results show that input noise can enhance the strength of NF-κB response to weak environmental (cytokine) signals via a mechanism that resembles stochastic resonance, and offer

a novel mode of cellular regulation for this important transcription factor. We find that NF-κB is sensitive to the dynamics of a pleiotropic factor like TNF, and temporally random modulation of the TNF signal can create significantly different phenotypic responses. This raises the intriguing possibility that biological noise can help discriminate regulatory signals. For example, the noise level in TNF could serve as a proxy for the number of signaling cells in the environment, where higher noise corresponds to higher number or density of surrounding cells. In such a scenario, the NF-κB signaling properties uncovered here would serve as an quorum sensing-like mechanism[72], with distinct activation programs only at high cell density for weak signals. Or more generally, a noisy environmental signal could mimic an intrinsic biological phenomenon. For example, immune regulatory cells such as natural killers have cell membrane associated TNF[73], and it is possible that attracted natural killers could reproduce such noisy signaling.

Input noise perception as an unexplored regime in cellular regulation may play an important physiological role in communication and signaling in fluctuating environments for many additional regulatory systems. The gene expression differences from NF-κB specific transcription of noise-stimulated cells show how adaptable the NF-κB pathway is to dynamic stimuli. Not only does the signal enhance NF-κB activation, but also leads to alternative response upon a secondary stimulation. Changes due to transcription of proteins involved in regulation of the pathway will lead to alternative outcomes given a secondary stimulation.

More generally, we found that cytokine noise induces a broad anti-cancer transcriptional response by limiting the growth potential of cells and activating anti-cancer transcriptional programs. It also prepares cells for the further recruitment of neutrophils. Interestingly, the behavior and expression patterns that we find is also a signature in tumorigenesis[74,75]. It is possible

that key oncogenic mutations could lead to the dysregulation of this program and as a result, induce cancer instead protecting against it.

Transcription factor dynamics in gene network regulation and signaling is an important element in mammalian cell response to many agonists - many of these transcription factors do behave non-linearly and rely on surpassing concentration thresholds of ligand molecules to produce all-or-none responses[76]. We found that noisy fluctuations of an inflammatory signal causes cells to perceive this signal to be much more potent. Our discovery of noise enhanced signal sensitivity in living cells, which was theoretically predicted in many settings but was not experimentally observed. This behavior emerges from the interaction of noisy cytokine inputs, NF-κB pathway resonance at fast time-scales, and rectified adaptation in the NF-κB pathway. These results thus demonstrate how biological systems can exploit highly non-trivial and interesting physics to turn a limitation such as environmental fluctuations into a strength.

References

1. Ventura, A. C. *et al.* Utilization of extracellular information before ligand-receptor binding reaches equilibrium expands and shifts the input dynamic range. doi:10.1073/pnas.1322761111

2. Suderman, R., Bachman, J. A., Smith, A., Sorger, P. K. & Deeds, E. J. Fundamental trade-offs between information flow in single cells and cellular populations. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5755–5760 (2017).

3. Hasty, J., Pradines, J., Dolnik, M. & Collins, J. J. Noise-based switches and amplifiers forgene expression. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2075–2080 (2000).

4. Ma, C. *et al.* A clinical microchip for evaluation of single immune cells reveals high functional heterogeneity in phenotypically similar T cells. *Nat. Med.* **17**, 738–743 (2011).

5. Han, Q. *et al.* Polyfunctional responses by human T cells result from sequential release of

cytokines. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1607–1612 (2012).

6. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (80-.).* **297**, 1183–1186 (2002).

7. Johnson, H. A. Thermal noise and biological information. *Q. Rev. Biol.* **62**, 141–52 (1987).

8. Tabbaa, O. P., Nudelman, G., Sealfon, S. C., Hayot, F. & Jayaprakash, C. Noise propagation through extracellular signaling leads to fluctuations in gene expression. *BMC Syst. Biol.* **7**, 94 (2013).

9. Grün, D., Kester, L. & Van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640 (2014).

10. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12795–12800 (2002).

11. McDonnell, M. D. & Ward, L. M. The benefits of noise in neural systems: bridging theory and experiment. *Nat. Rev. Neurosci.* **12**, 415–426 (2011).

12. McDonnell, M. D. & Abbott, D. What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. *PLoS Comput. Biol.* **5**, e1000348 (2009).

13. Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. A. & Bowsher, C. G. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E326–E333 (2014).

14. Tay, S. *et al.* Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–71 (2010).

15. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J. & Doyle, J. Robustness of cellular functions. *Cell* **118**, 675–685 (2004).

16. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).

17. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11980–5 (2003).

18. Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–73 (2009).

19. Ashall, L. *et al.* Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. *Science* **324**, 242–6 (2009).

20. Covert, M. W., Leung, T. H., Gaston, J. E. & Baltimore, D. Achieving stability of

lipopolysaccharide-induced NF-κB activation. *Science (80-.).* **309**, 1854–1857 (2005).

21. Hoffmann, A. & Baltimore, D. Circuitry of nuclear factor κB signaling. *Immunological Reviews* **210**, 171–186 (2006).

22. Lee, T. K. *et al.* A noisy paracrine signal determines the cellular NF-kappaB response to lipopolysaccharide. *Sci. Signal.* **2**, ra65 (2009).

23. Hao, S. & Baltimore, D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.* **10**, 281–288 (2009).

24. Lane, K. *et al.* Measuring Signaling and RNA-Seq in the Same Cell Links Gene Expression to Dynamic Patterns of NF-κB Activation. *Cell Syst.* **4**, 458-469.e5 (2017).

25. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241–5 (2002).

26. Zhang, Q., Lenardo, M. J. & Baltimore, D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell* **168**, 37–57 (2017).

27. Hayden, M. S., West, A. P. & Ghosh, S. NF-κB and the immune response. *Oncogene* **25**, 6758–6780 (2006).

28. Hayden, M. S. & Ghosh, S. NF-κB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* **26**, 203–34 (2012).

29. Zhang, Q. *et al.* NF-κB Dynamics Discriminate between TNF Doses in Single Cells. *Cell Syst.* (2017). doi:10.1016/j.cels.2017.10.011

30. Cheong, R., Hoffmann, A. & Levchenko, A. Understanding NF-kappaB signaling via mathematical modeling. *Mol. Syst. Biol.* **4**, 192 (2008).

31. Krishna, S., Jensen, M. H., Sneppen, K. & Kadanoff, L. P. Minimal model of spiky oscillations in NF-□B signaling.

32. Kellogg, R. A. & Tay, S. Noise Facilitates Transcriptional Control under Dynamic Inputs. *Cell* **160**, 381–392 (2015).

33. Heltberg, M., Kellogg, R. A., Krishna, S., Tay, S. & Jensen, M. H. Noise Induces Hopping between NF-κB Entrainment Modes. *Cell Syst.* **3**, 532-539.e3 (2016).

34. Collins, J. J., Chow, C. C. & Imhoff, T. T. Stochastic resonance without tuning. *Nature* **376**, 236–238 (1995).

35. Kellogg, R. A., Tian, C., Lipniacki, T., Quake, S. R. & Tay, S. Digital signaling decouples activation probability and population heterogeneity. *Elife* **4**, e08931 (2015).

36. Piehler, A., Ghorashian, N., Zhang, C. & Tay, S. Universal signal generator for dynamic cell stimulation. *Lab Chip* **17**, 2218–2224 (2017).

37. Kellogg, R. A., Gómez-Sjöberg, R., Leyrat, A. A. & Tay, S. High-throughput microfluidic single-cell analysis pipeline for studies of signaling dynamics. *Nat. Protoc.* **9**, 1713–1726 (2014).

38. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

39. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).

40. Bui-Nguyen, T. M. *et al.* NF-κB signaling mediates the induction of MTA1 by hepatitis B virus transactivator protein HBx. *Oncogene* **29**, 1179–1189 (2010).

41. Zhao, B. *et al.* Inflammatory Micro-environment Contributes to Stemness Properties and Metastatic Potential of HCC via the NF-κB/miR-497/SALL4 Axis. *Mol. Ther. - Oncolytics* **15**, 79–90 (2019).

42. Ankers, J. M. *et al.* Dynamic NF-κb and E2F interactions control the priority and timing of inflammatory signalling and cell proliferation. *Elife* **5**, (2016).

43. Araki, K., Kawauchi, K. & Tanaka, N. IKK/NF-κB signaling pathway inhibits cell-cycle progression by a novel Rb-independent suppression system for E2F transcription factors. *Oncogene* **27**, 5696–5705 (2008).

44. Lim, C. A. *et al.* Genome-wide Mapping of RELA(p65) Binding Identifies E2F1 as a Transcriptional Activator Recruited by NF-κB upon TLR4 Activation. *Mol. Cell* **27**, 622–635 (2007).

45. Sammons, M. A., Nguyen, T. A. T., McDade, S. S. & Fischer, M. Tumor suppressor p53: From engaging DNA to target gene regulation. *Nucleic Acids Research* **48**, 8848–8869 (2020).

46. Quaas, M., Müller, G. A. & Engeland, K. p53 can repress transcription of cell cycle genes through a p21 WAF1/CIP1-dependent switch from MMB to DREAM protein complex binding at CHR promoter elements. *Cell Cycle* **11**, 4661–4672 (2012).

47. Tchasovnikarova, I. A. *et al.* Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat. Genet.* **49**, 1035–1044 (2017).

48. Kawauchi, K., Araki, K., Tobiume, K. & Tanaka, N. p53 regulates glucose metabolism through an IKK-NF-κB pathway and inhibits cell transformation. *Nat. Cell Biol.* **10**, 611–618 (2008).

49. Lowe, J. M. *et al.* P53 and NF-κB coregulate proinflammatory gene responses in human macrophages. *Cancer Res.* **74**, 2182–2192 (2014).

50. Ishak Gabra, M. B. *et al.* IKKβ activates p53 to promote cancer cell adaptation to glutamine deprivation. *Oncogenesis* **7**, 1–12 (2018).

51. Hayden, M. S. & Ghosh, S. Signaling to NF-κB. *Genes and Development* **18**, 2195–2224 (2004).

52. Son, M. *et al.* NF-κB responds to absolute differences in cytokine concentrations. *Sci. Signal.* **14**, (2021).

53. Pahl, H. L. Activators and target genes of Rel/NF-κB transcription factors. *Oncogene* **18**, 6853–6866 (1999).

54. Wan, F. & Lenardo, M. J. Specification of DNA binding activity of NF-kappaB proteins. *Cold Spring Harbor perspectives in biology* **1**, (2009).

55. Sen, R. & Smale, S. T. Selectivity of the NF-{kappa}B response. *Cold Spring Harbor perspectives in biology* **2**, (2010).

56. Natoli, G. Control of NF-kappaB-dependent transcriptional responses by chromatin organization. *Cold Spring Harbor perspectives in biology* **1**, (2009).

57. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

58. Richmond, A. NF-κB, chemokine gene transcription and tumour growth. *Nature Reviews Immunology* **2**, 664–674 (2002).

59. Sawant, K. V. *et al.* Neutrophil recruitment by chemokines Cxcl1/KC and Cxcl2/MIP2: Role of Cxcr2 activation and glycosaminoglycan interactions. *J. Leukoc. Biol.* JLB.3A0820-207R (2020). doi:10.1002/JLB.3A0820-207R

60. Tu, Y. & Rappel, W.-J. Adaptation in Living Systems. *Annu. Rev. Condens. Matter Phys.* **9**, 183–205 (2018).

61. Briat, C., Gupta, A. & Khammash, M. Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. *Cell Syst.* **2**, 15–26 (2016).

62. Inoue, K. *et al.* Oscillation dynamics underlie functional switching of NF-κB for B-cell activation. *npj Syst. Biol. Appl.* **2**, 16024 (2016).

63. Workman, L. M. & Habelhah, H. TNFR1 signaling kinetics: Spatiotemporal control of three phases of IKK activation by posttranslational modification. *Cell. Signal.* **25**, 1654–1664

(2013).

64. Jensen, M. H. & Krishna, S. Inducing phase-locking and chaos in cellular oscillators by modulating the driving stimuli. *FEBS Lett.* **586**, 1664–1668 (2012).

65. Xia, Y. *et al.* Phosphorylation of p53 by IκB kinase 2 promotes its degradation by β-TrCP. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2629–2634 (2009).

66. Karin, M. How NF-kB is activated: the role of the IkB kinase (IKK) complex.

67. Heger, K. & Dixit, V. M. TBK1 and IKKε restrain cell death. *Nat. Cell Biol.* **20**, 1330–1331 (2018).

68. NF-κB activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase | The EMBO Journal. Available at: https://www.embopress.org/doi/full/10.1093/emboj/18.23.6694. (Accessed: 13th April 2020)

69. Bier, M. Brownian ratchets in physics and biology. *Contemp. Phys.* **38**, 371–379 (1997).

70. Mitchell, A. & Lim, W. Cellular perception and misperception: Internal models for decision-making shaped by evolutionary experience. *BioEssays* **38**, 845–849 (2016).

71. Mitchell, A., Wei, P. & Lim, W. A. Oscillatory stress stimulation uncovers an Achilles' heel of the yeast MAPK signaling network. *Science* **350**, 1379–83 (2015).

72. Antonioli, L., Blandizzi, C., Pacher, P., Guilliams, M. & Haskó, G. Quorum sensing in the immune system. *Nat. Rev. Immunol.* **18**, 537–538 (2018).

73. Caron, G. *et al.* Human NK cells constitutively express membrane TNF-α (mTNFα) and present mTNFα-dependent cytotoxic activity. *Eur. J. Immunol.* **29**, 3588–3595 (1999).

74. Xia, Y., Shen, S. & Verma, I. M. NF-κB, an active player in human cancers. *Cancer immunology research* **2**, 823–830 (2014).

75. Mussbacher, M. *et al.* Cell type specific roles of nf-kb linking inflamation and thrombosis. *Frontiers in Immunology* **10**, 85 (2019).

76. Chakraborty, A. K. *et al.* Molecular origin and functional consequences of digital signaling and hysteresis during Ras activation in lymphocytes. *Sci. Signal.* **2**, pt2 (2009).

Acknowledgements: We acknowledge Alan Selewa for earlier contributions to this work.

Author Contributions: P.P did microfluidic experiments with assistance from M.S. and P.P. analyzed the experimental data; W.P. ran simulation pipeline. A.M and S.T supervised the work.

Correspondence: Savaş Tay, Institute for Molecular Engineering, The University of Chicago, tays@uchicago.edu, Arvind Murugan, Department of Physics, The University of Chicago, Chicago, Illinois 60637, amurugan@uchicago.edu

Data and Materials Availability: All data is available in the main text or supplemental materials

Funding: This work was supported by NIH grant R01GM128042 and R01GM127527 to S.T.

Supplementary Information

Environmental noise enables sensitive detection and transcriptional decoding of cytokine inputs

Parthiv Patel[1†], Weerapat Pittayakanchit [3†], Minjun Son[1,2], Arvind Murugan[3,4 *], Savaş Tay[1,2, *]

Materials and Methods

TNF-α Stimulation Using Microfluidic Cell Culture

We use the signal generator chip described previously (18). PDMS chambers coated with fibronectin (FC010-10MG) and cells were seeded at constant density ~20,000 cells/cm2. Cells were cultured for 5 hours before stimulation; they were taken at 100% confluence and incubated with .25% Trypsin-EDTA for 5mins (25200-056) prior to loading. Standard culture conditions of 5% CO2 and 37°C were maintained using an incubation chamber for imaging and cell culture. Mouse TNF-α (PMC3014_3671982503) was diluted in Fluorobrite DMEM media (A1896701) with 2x glutamax (35050061), pen/strep (15140-122) and FBS (16140071) in vials pressured with 5% CO2 at 5psi and kept on ice. Microbore tubing (PEEK, Vici) was used to connect the TNF-α supply to the chip. For continuous pumping input, the on-chip peristaltic pump was operated at a sampling rate of .5 seconds of a combination (3,4, or 5 concurrently open) of the following inputs: [.2 .1 .05 .025 .0125 .00625 .003125 .001625] ng/mL TNF-α.

Image Acquisition and Data Processing

We use an automated Nikon eclipse ti2 microscope and capture fluorescence images (red and green channels for p65 and H2B reporters) at 20× magnification via a Hamamatsu ORCA-Flash4.0 V3 Camera (C13440) every 3-5 min for 1-10 hr. Microfluidic device is mounted on the microscope. Custom Matlab scripts were used for image processing (available on request). NF-κB activation was quantified as sum of nuclear fluorescence intensity. For peak analysis, data were smoothed followed by peak detection using a combination of integrated area, first derivative and peak height.

NF-κB Reporter Cell Line

Creation of mouse (3T3) fibroblasts displaying near-endogenous p65 levels was previously described (Tay et al., 2010). Knockout p65−/− mouse 3T3 fibroblasts were engineered to display p65-DsRed under the 1.5 kb p65 promoter (4). A minimum fluorescence clone was selected to achieve near-endogenous expression level to represent NF-κB wild-type dynamics. Addition of a ubiquitin-promoter driven H2B-GFP cassette provides a nuclear marker to facilitate automated image processing.

RNA Sequencing

Cells were removed from microfluidic chambers using .25% Trypsin-EDTA for 5mins (25200-056) to free the cells from microfluidic surface, then flushed with water to remove from device. Cells were placed in lysis buffer containing 0.2% (vol/vol) Triton X-100, RNase inhibitor, oligoDT primer and dntps. Protocol for SMART seq was followed thereafter. Samples were pooled then sequenced on NovaSeq platform. Expression levels are expressed as transcripts per million and hierarchical clustering was done using the metric: (1-pearson correlation) on gene expression

levels. Upregulated and downregulated genes were found using ($p < .01$, fold-change >.5). Gene ontology and transcription factor analysis was done using GSEA analysis on differentially expressed genes. STRING network was defined using curated knowledge databases of protein-protein associations and BIND database using high confidence interactions. Analysis was limited to proteins that contain the NF-κB binding motif and are within the interacting genes in the NF-κB pathway.

Mathematical modeling of NF-κB pathway

A comprehensive model for the NF-κB system with 26 variables was first elaborated by Hoffman in 2002[20]. Krishna et al. reduced the model to only five non-linear coupling equations[14,25] while still retaining the essential dynamics of the pathway:

$$\frac{dN_n}{dt} = \frac{K_I}{K_I + I} k_{Nin}(N_{tot} - N_n) - k_{lin} I \frac{N_n}{K_N + N_n}$$

$$\frac{dI_m}{dt} = k_t N_n^2 - \gamma_m I_m$$

$$\frac{dI}{dt} = k_{tl} I_m - \alpha [IKK]_a (N_{tot} - N_n) \frac{I}{K_I + I}$$

$$\frac{d[IKK]_a}{dt} = k_a [TNF]([IKK]_{tot} - [IKK]_a - [IKK]_i) - k_i [IKK]_a$$

$$\frac{d[IKK]_i}{dt} = k_i [IKK]_a - k_p [IKK]_i \frac{k_{A20}}{k_{A20} + [A20][TNF]}$$

where N_n is the nuclear NF-κB concentration, I_m is the IκB mRNA level, I is the concentration of cytoplasmic IκB protein, $[IKK]_a$ is the level of active IκB kinase, and $[IKK]_i$ is the level of inactive IκB kinase. This simple model reproduces spiky NF-κB oscillations seen in experiments and entrainment of such oscillations by time-varying TNF signals. The last two equations above are a simplified representation of signal transduction from external TNF levels to IKK levels[14,25]. Note the existence of both positive and negative regulation of IKK by TNF, characteristic of an incoherent feedforward loop[9]: TNF increases the amount of active IKK but also decreases active IKK indirectly through inactive IKK.

In reality, signal transduction between external TNF and IKK is known to be more complex, involving receptor dynamics and numerous intermediate complexes. These details include incoherent feedforward loops of different timescales in addition to the one modelled here[23,24]. Such features may be revealed by the response to high frequency TNF signals used in this study but cannot be predicted by the last two equations in this model since they are based on experiments involving slowly changing TNF (timescale of hours).

In this paper we modify the last two equations to capture faster signaling processes that might be visible to fast TNF signals (timescale of seconds). We continue to model the pathway between TNF and IKK as having general positive and negative regulation but with an unspecified timescale τ_0. This approach gives a simple coarse-grained model of signaling without making assumptions about the exact molecular identity of the molecules involved in these fast processes:

$$\frac{dN_n}{dt} = \frac{K_I}{K_I + I} k_{Nin}(N_{tot} - N_n) - k_{lin} I \frac{N_n}{K_N + N_n}$$

$$\frac{dI_m}{dt} = k_t N_n^2 - \gamma_m I_m$$

$$\frac{dI}{dt} = k_{tl} I_m - \alpha [IKK]_a (N_{tot} - N_n) \frac{I}{K_I + I}$$

$$\tau_0 \frac{dX}{dt} = -\alpha_x (X - TNF)$$

$$\tau_0 \frac{dY}{dt} = -\beta_y (Y - c_0) - \beta_x X \left(\frac{Y^n}{Y_0^n + Y^n} \right) + \beta_{TNF} TNF$$

$$\frac{d[IKK]_a}{dt} = -\gamma_{IKK} [IKK]_a + \gamma_y (Y - c_0)$$

Here TNF is the input signal of interest, X is promoted by TNF, and Y is promoted by TNF but suppressed by X. The dynamics downstream of IKK (i.e., NF-κB and Iκ B) are exactly as in the model of Krishna et al.[14,25]. The value of the parameters we use in these equations are summarized in Table 1.

Parameters for simplified NF-κB model[14,25]

Parameter	Value	Parameter	Value
k_{Nin}	5.4 min^{-1}	α_x	10
k_{lin}	0.018 min^{-1}	β_y	60
k_l	1.03 $\mu M^{-1} min^{-1}$	β_x	60
k_{tl}	0.24 min^{-1}	β_{TNF}	60.6

K_I	0.035 μM	n	10
K_N	0.029 μM	y_0	0.001
γ_m	0.017 min^{-1}	γ_{IKK}	1
α	1.05 $\mu M^{-1}min^{-1}$	γ_y	30
N_{tot}	1 μM	τ_0	1 min
		τ_{noise}	0.0667 τ_0

Table 1: The first two columns give parameters of the simplified NF-κB system (k_{lin} to N_{tot}), taken directly from [25]. The last columns give parameters for the rectified adaptation model, following design principles laid out in [9] for Incoherent Feed-Forward Loop-based adaptive circuits.

We have two sources of noise in our simulation, one from TNF, and the other from the stochasticity inside the NF-κB network. Additive noise in the TNF signal is simulated from two random distributions. The amplitude of the noise is drawn from a uniform distribution $[-\sigma, \sigma]$, and the time difference between the change of amplitudes of noise is drawn from an exponential distribution with mean equals to τ_{noise}. Meanwhile, the stochasticity inside the NF-κB networks (only the first three equations) is simulated from the Langevin noise $\eta(0, \epsilon)$ where $\epsilon = 0.01\ \tau_0^{-1/2}$.

Rectified adaptation

The existence of both positive and negative interactions (i.e., an Incoherent Feed Forward Loop), leads to adaptation[9]. Adaptation in a chemical circuit refers to a transient response to a step change

in the input after which the circuit output eventually settles down to its initial resting value. In our model, the resting value of Y is c_0. For non-zero c_0 = 0.1, we find that Y responds adaptively to both step ups and step downs in the input (TNF).

However, if $c_0 = 0$, Y only responds to step ups in TNF and not step downs, a phenomenon we term `rectified adaptation'. Such a circuit will perceive noise as an ever-increasing signal, and trigger an activation threshold much larger than the standard deviation of the noise. Thus, we find that a simple limit of the commonly found adaptation motif in biology explains stochastic resonance seen in NF-κB activation.

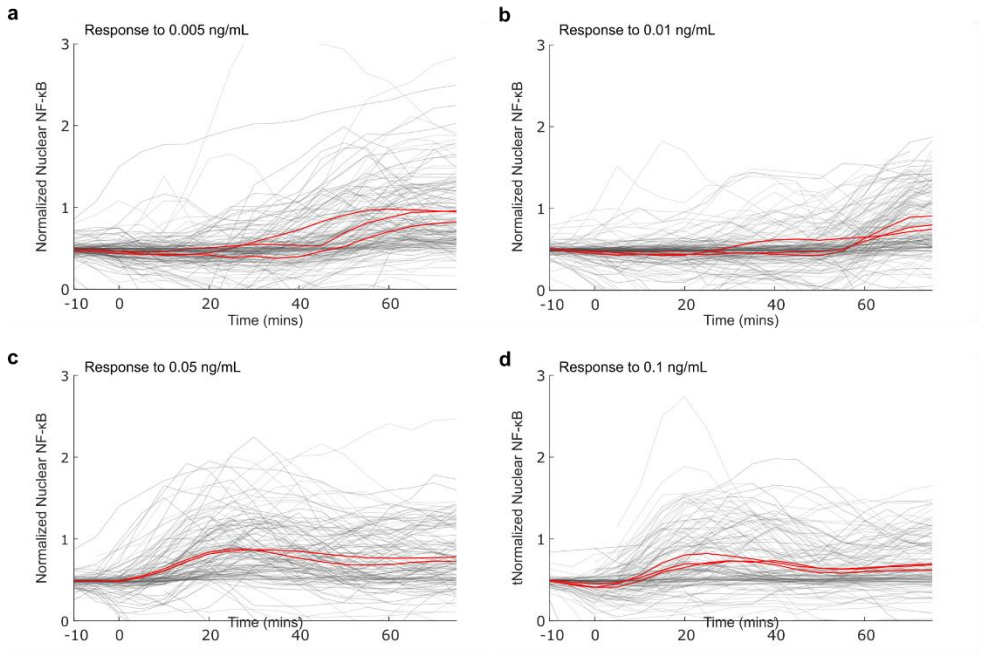
Details of simulations in each figure

In Figure 1d-f, NF-κB simulated using the rectified adaptative circuit (i.e., $c_0 = 0$) activates from noisy signal of low moving average. For the constant signal, we set the TNF level to be 0.1. For the noisy signal, we set the average value of TNF to be 0.05 and σ to be 0.04.

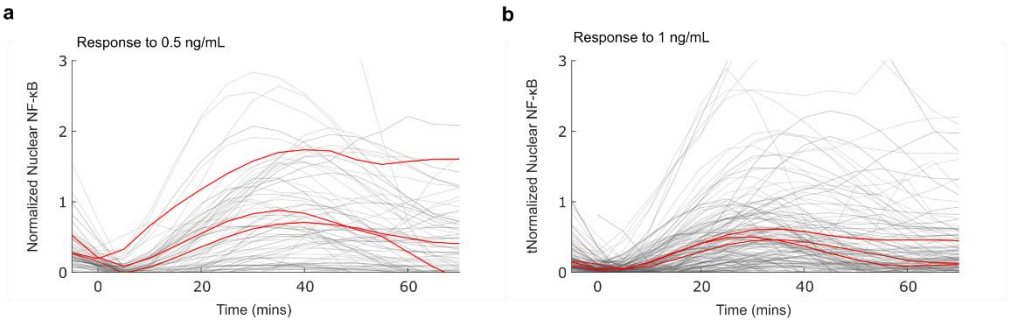
In Figure 3c-d, NF-κB only responds to noisy signal if the adaptation is rectified (i.e., if $c_0 = 0$) but not with conventional adaptation ($c_0 = 0.1$). Similar to Figure 1d-f, for constant signal, we set the TNF level to be 0.05. For the noisy signal, we set the average value of TNF to be 0.05 and σ to be 0.04.

In Figure 3e, NF-κB in the rectified adaptive circuit ($c_0 = 0$) shows decreasing responses as the period of the input signal increases. In creating Figure 3e, we use noiseless sinusoidal signals of periods $0.2667\tau_0$, $1.5085\tau_0$, $8.5333\ \tau_0$. For all periods, we set the average value of TNF to be 0.05 and σ to be 0.04.

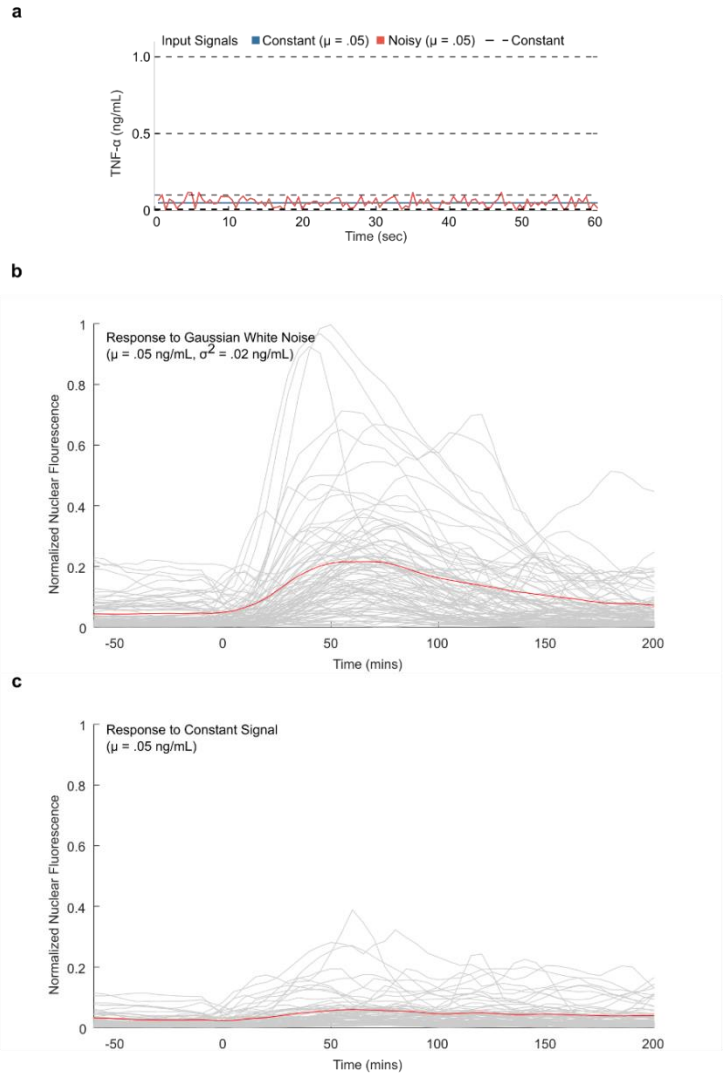
In Figure 3f, NF-κB in the rectified adaptive circuit shows decreasing fraction of cells activated as the period of the input signal increases. The purpose of Figure 3e-f is to understand how the time scale of the adaptation impacts the activation of NF-κB. Here, similar to Figure 3e, we remain the same average TNF level and the amplitude of the sine wave at 0.05 and 0.04, respectively. We vary the range of period to be from $0.1\tau_0$ to 100 τ_0.



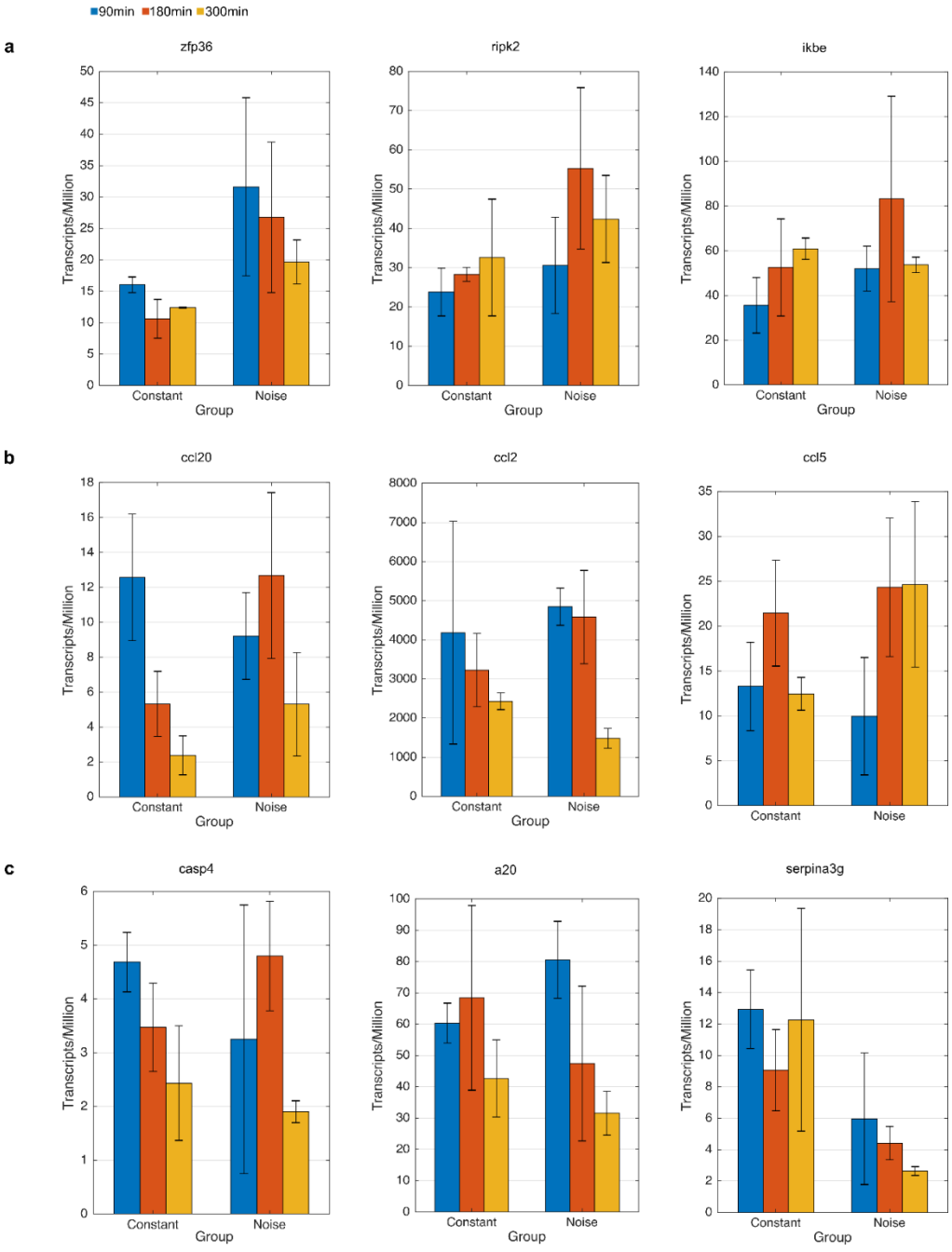
Supplementary Figure 2.1: Comparison of NF-κB single cell traces at constant pulse feeding for low doses. (a-d) Constant doses of .1ng/mL and less with constant stimulation show characteristic response of NF-κB activation.



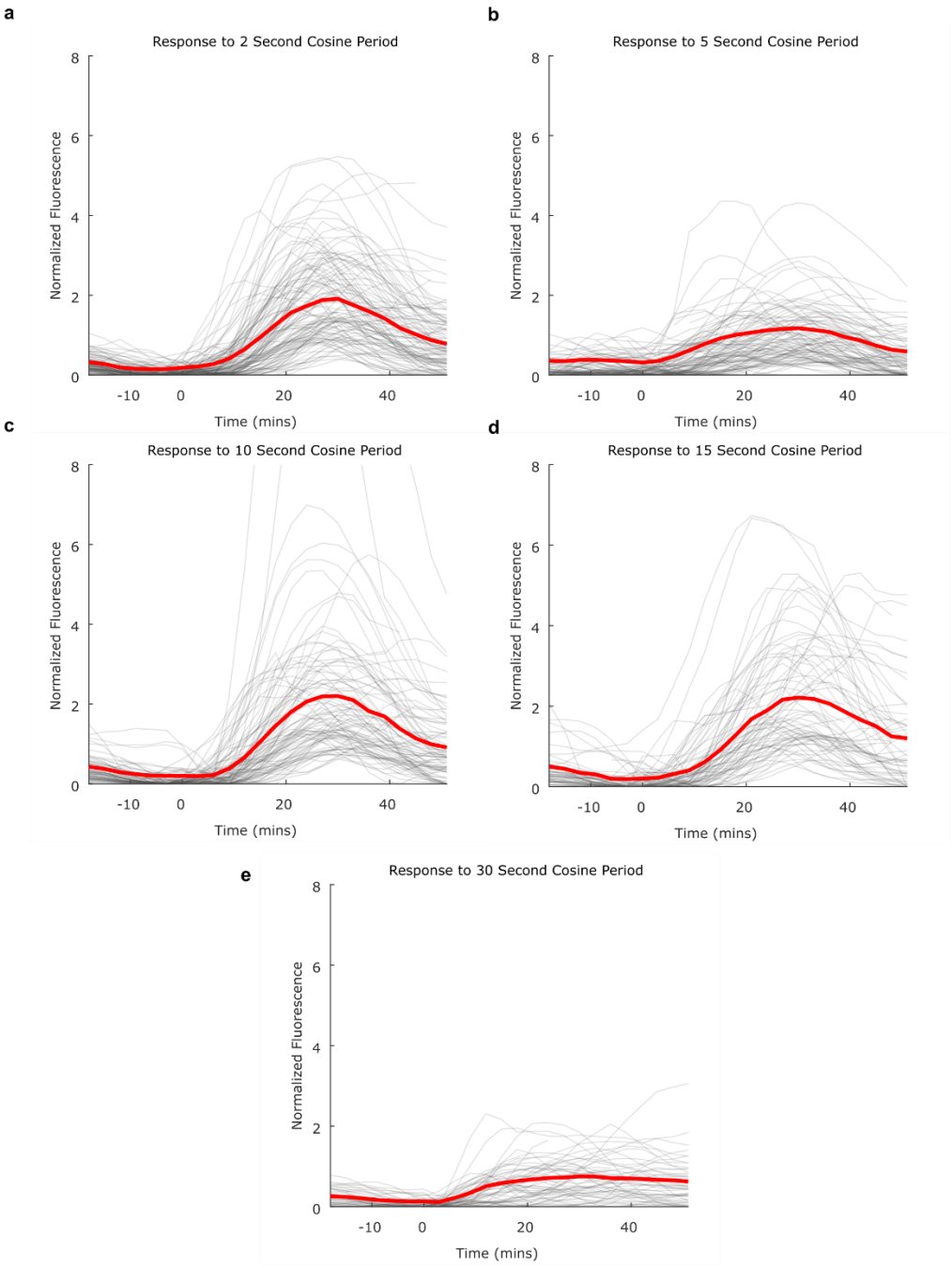
Supplementary Figure 2.2: Comparison of NF-κB single cell traces at constant pulse feeding for high doses. (a-d) Constant doses of 1 ng/mL and .5 ng/mL with constant stimulation show characteristic response of NF-κB activation with high activated fraction and fast response times.



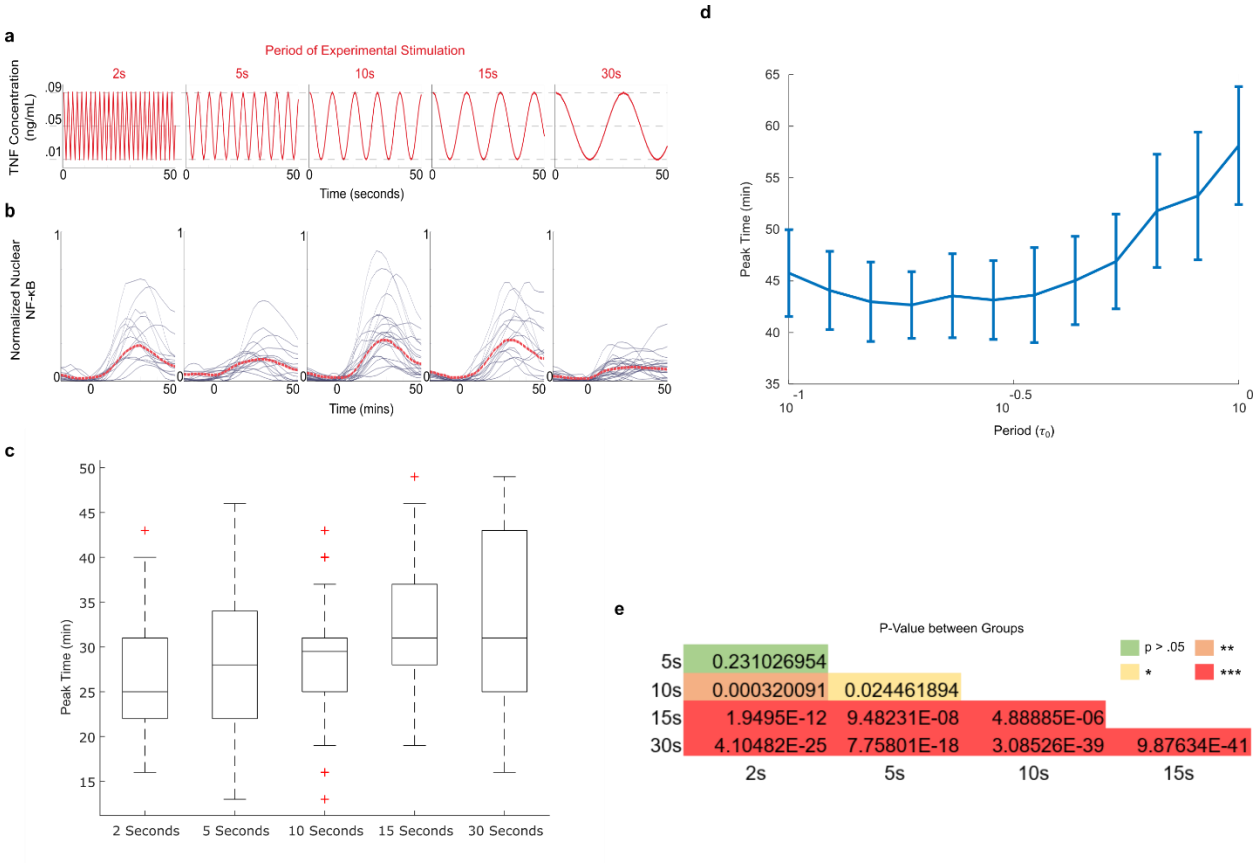
Supplementary Figure 2.3: Cells are sensitive to the addition of noise in TNF signal. (a) Stimulation comparison between different tested conditions **(b)** single cell traces measured under noisy stimulus exposure. **(c)** Single cell traces under constant TNF exposure with equal integrated mean as noisy stimulus show significantly less NF-κB translocation to the nucleus.



Supplementary Figure 2.4: NF-κB target genes are show different behaviors when exposed to a noisy signal (a) There are three different behaviors that present amongst classic NF-κB target genes. The first behavior is an increase expression across multiple time points. **(a)** The second behavior is an increased duration of high expression. This is present amongst CCL class genes. **(c)** The last grouping of gene expression behavior is a downregulation of expression.



Supplementary Figure 2.5: Single cell traces of response to cosine stimulus show increasing with slower periods (a-e) Activated single cell traces show similar profiles for exposures under 30second period, but dramatically reduced peak height for 30second period (mean of populations in red).



Supplementary Figure 2.6: Experiments confirm simulation prediction of increasing response time with slower period. (a) Experimental stimulation profiles for cosine exposure at different periods ranging from 2s to 30s. **(b)** Activated single cell traces in live cell imaging experiments (mean of populations in red) **(c)** Stochastic simulation of 100 cells using same methodology as F6b (simulation of cosine stimulation) comparing peak timing and period of stimulation **(d)** Experiments show increasing mean and tail for peak time in cosine stimulation experiments. **(e)** One tailed heteroskedastic t-test between groups reveal significant differences in peak timings between groups

Raw Data and Image Analysis Software: Single cell NF-κB fluorescence microscopy images, analysis software and analyzed single cell traces for all experiments can be found in the following server:

https://github.com/parthivapatel/NoisyInputSignaling

Conclusions

We attempt to solve a key question in cellular biology: how do cells successfully function despite a noisy environment. Using a medically-relevant signaling pathway that is well understood as a model system can help yield insight into the mechanisms that use or curtail noise and help us fundamentally understand sensitivity and specificity in this medically relevant pathway. Through this, we impact a wide range of fields including immunity, pharmacology, and signaling as well as systems biology and biophysics. Our research fundamentally increases understanding of how cells respond accurately to noisy signals and how signal transduction can be robust in a constantly changing environment. Mechanistic investigation into this problem is possible because of the unique application of microfluidics and single cell analysis.

Studying the detrimental or possibly beneficial role of input noise in the NF-κB pathway and its signal transduction is highly innovative. Our studies using quantitative single-cell measurements show how internal variability in network components play a role in controlling gene expression and response to cytokine inputs. Further, studying baseline noisy signal has been investigated for the first time in immune signaling networks. The role of input noise has been studied in neuroscience previously and theoretical studies exist for other networks, but can only be experimentally studied with a system capable of systematically supplying endogenous noisy signals and quantitatively measuring response. Further characterization of cells exposed to noisy signals showed significant gene expression changes though limitation of growth potential of cells by activating anti-cancer transcriptional programs.

Similarly, by using a machine learning method to investigate activation a priori we addressed questions surrounding the fundamental heterogeneity in NF-κB. This method, however, is useful for not only NF-κB, but for many other binary events as well such as drug survival, response in many gene circuits that exist outside immune signaling, and infection. This methodology and approach illuminates and helps decouple the underlying features of cell signaling without perturbing the initial state of those cells.

Mathematical analysis combined with quantitative analysis of NF-κB dynamics in response to noisy and non-uniform signaling continues to produce new insights into signaling, cellular decision making, and the study of biological variability. These approaches coupled with advances through microfluidic and biological engineering helps us pioneer design principles for single cell research for the community.

Appendix

Through the exploration of noise in transcription factor response, we circled around the fundamental question of how to measure transcription factor engagement in a high-throughput manner while still decoupling the effect of noise in transcription factor binding. Due to the effect of cofactors, enhancers, repressors and many other DNA modulating proteins, transcription factor response and gene expression does not always adhere to the input.

In the following section, I explore preliminary efforts to enable the usage of single cell RNA sequencing data to gather unique insights into transcription factor engagement and to understand some of the fundamental questions underlying noise in transcription factor response. Through hidden information within RNA sequencing data relating to the reference genome we are able to gather information about underlying motif usage in single cells. Using this we can infer behaviors for the transcription factors that bind these motifs.

motiFATE Uses Hidden RNA-seq Underlying Information to Inform Single Cell Transcriptional Factor Engagement

Parthiv Patel[1], Nir Drayman[1], and Savaș Tay[1*]

[1]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637

* Correspondence: tays@uchicago.edu

Abstract:

Single-cell RNA sequencing has emerged as a powerful tool often used to understand heterogeneity and transitions in cellular state. Single cell sequencing outputs however are either cost prohibitive or have standard levels of gene expression that often limit the scope of the analysis to specific target genes. Often, many additional studies and analysis are required to understand how noisy single cell data can relate to dynamics of a pathway and the kinetics of transcription factors. Here we demonstrate a simple, yet robust, application of sequencing data to derive transcription factor regulation from single cell RNA sequencing data. We used our platform, motiFATE, on hematapoetic stem cell differentiation, PBMCs and EBV infected cells to understand how downstream expression is driven by transcription factor adaptability. MotiFATE showcases the adaptability of motif elements and enabled a reconstruction of the cell states, differentiation trajectories, transcription factor binding dynamics, and new transcription factor identification. This study establishes how RNA-derived transcription binding analysis can be used to understand cellular hierarchy and fate decisions in development.

Introduction

Throughout their study, multicellular organisms have been classified and categorized into fundamental components, such as organs and tissues. More recently, study has gone into organizing cellular hierarchy through cell types and differentiation landscapes(*1–4*). This narrowing of scope in turn comes with understanding of differences in form and function of different components. Characterizing differences in cell type and further, cell states has become extremely accessible through data-rich experimental techniques such as cytof and rna seq(*5–7*), and in turn has led to large scale analysis of features that correlate with transitions through lineage.

There has been thorough analysis and methodology used in construction and analyzing these lineages(*8–10*) and these tools have emerged as powerful ways to organize and understand how cells transition from one state to another. However, the analysis of what drives changes in state is more often focused on expression of genes. Specifically, limitations in multiplexity of different sequencing strategies have limited the exploration of transcription factor engagement through the course of these cellular lineages. Often, specific prior knowledge of transcription factors is required to attribute meaning from data from transcription factor through to gene expression and has limited broad unbiased studies of transcription factor engagement. Despite the limited options to identify transcription factor states in single cells, tools such as Gene Ontology(*11*) and STRING(*12*) help to identify known sources of transcription factor utilization.

Similarly, while exploration of the details of gene expression differences has yielded many insights into how groups of cells function, the methods that are used for these experiments often fails to encapsulate how adaptable the cellular machine truly is. Despite the same pathway being used, a single transcription factor can represent vastly different programs(*13*). Conventional analysis often

serves to undermine the adaptability of these extremely adaptable transcription factors and the broad role of cofactors in regulating the binding activity of these transcription factors(*14–16*).

Here, using motiFATE, we systematically evaluated RNA-derived features from the host genome to map nearly 33,000 motif sequences to their corresponding single cells and quantify how the adaptability of known transcription factors plays a role in transcription factor programming across cell type differentiation. We then leveraged our findings to construct an unsupervised framework for predicting and understanding unknown transcription factor dynamics across cellular hierarchies.

RNA sequencing data contains underlying hidden genomic information

Our initial goal was to identify and map motif sequences to their corresponding single cells. Underlying gene expression data contains a wealth of underutilized knowledge and hidden data. A single gene transcript can represent data from the promotor and within the gene itself given prior knowledge of the genome. To utilize this information in single cell analysis we used known information about a population's genetic information and genome to populate possible motifs for each gene within a window of 5000bp from the transcriptional start site. We then used a moving window of 8bp to populate the frequency of motifs within each gene and used this information to transform single cell gene expression to motif utilization (**Figure 3.1a)**.

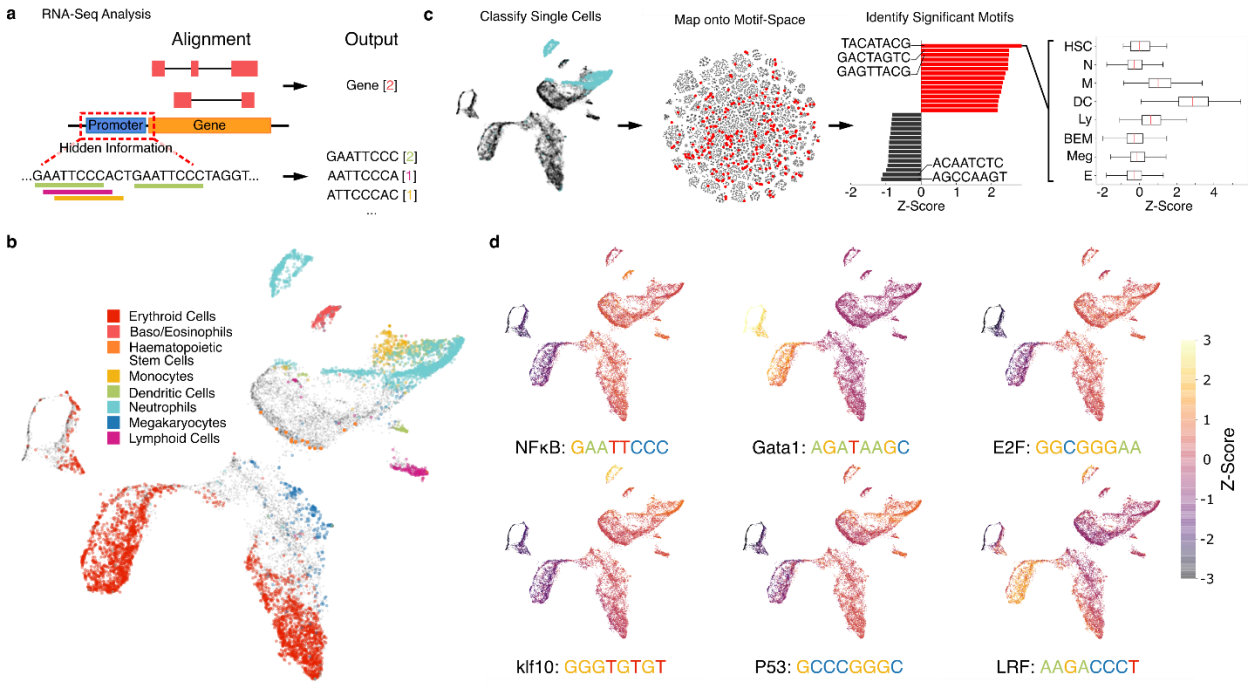


Figure 3.1: RNA sequencing data hides underlying information about the genome. (a) We use underlying genetic information to populate motif utilization by applying motifs used in each gene to existing RNA sequencing data **(b)** Single cell mapping of hematopoietic stem cell differentiation onto UMAP space with cell states identified. **(c)** Comparison of motif state across cell types can identify significantly overutilized and underutilized motifs in single cell. **(d)** Using known transcription factor binding motifs, we can visualize and confirm the distribution of transcription factor engagement across the population in different cell types.

To explore the value of this additional data we use hematopoietic cell differentiation to benchmark known transcription factor dynamics. We compare motif utilization in well understood cell types to address the validity of comparing transcription factor engagement in single cells (**Figure 3.1b**). We used an existing hematopoietic stem cell dataset(5) to mapped single cells onto a dimensionally reduced space using UMAP. We can compare the results from using motif-based features alone and gene-expression based features alone and find that we are able to recapture the neighborhood approximated of different clusters (using k-means; **Supplementary Figure 3.1**) Here we looked

specifically at dendritic cells and populated the single cell distribution of motifs onto a dimensionally reduced motif space and compared motif utilization across different cell types. We found many significant motifs that have a high prevalence in dendritic cells. Of these, some represent known transcription factors (**Figure 3.1c**).

While we can use unsupervised classification and comparison of motifs to identify new and particularly variable motifs, we can also use known transcription factor binding to inform and understand single cell distribution of transcription factor engagement (**Figure 3.1d**). For example, it is known that there is upregulated Gata1 and LRF expression in the erythroid cells, and upregulated Klf10 expression in Neutrophils and we can see that the transcription factor engagement follows our expectation. Using known motifs allows us to validate transcription factor engagement patterns that are known in these different cell types as well as identify new modes and timings of engagement in a population.

Motif utilization variability in single cells recapitulates known variability in transcription factor binding variability

To explore motif binding variability in single cells we looked specifically at the known binding motif for NF-κB and look into how NF-κB use similar motifs in PBMC single cell gene-expression data (10x 33k single cell dataset). We first mapped motifs into a dimensionally reduced space using tSNE (**Figure 3.2a**) by defining the hamming distance of every motif to 6 benchmark motifs including ‘AAAAAAAA’ and ‘GGGGGGGG’ to help visualize hamming distance differences among motifs. Hamming distance is a metric often used to measure similarity of two different sequences accounting for frameshifts, insertions and deletions. We then identified motifs that fall within 2 hamming distance of a general consensus sequence of NF-κB: ‘GAATTCCC’ to look at

all motifs that fall under utilization by NF-κB (**Figure 3.2b**). We term this sum as transcription factor engagement. Interestingly in PBMCs, NF-κB engagement appears to fall within two populations: one with high engagement and one with low engagement.

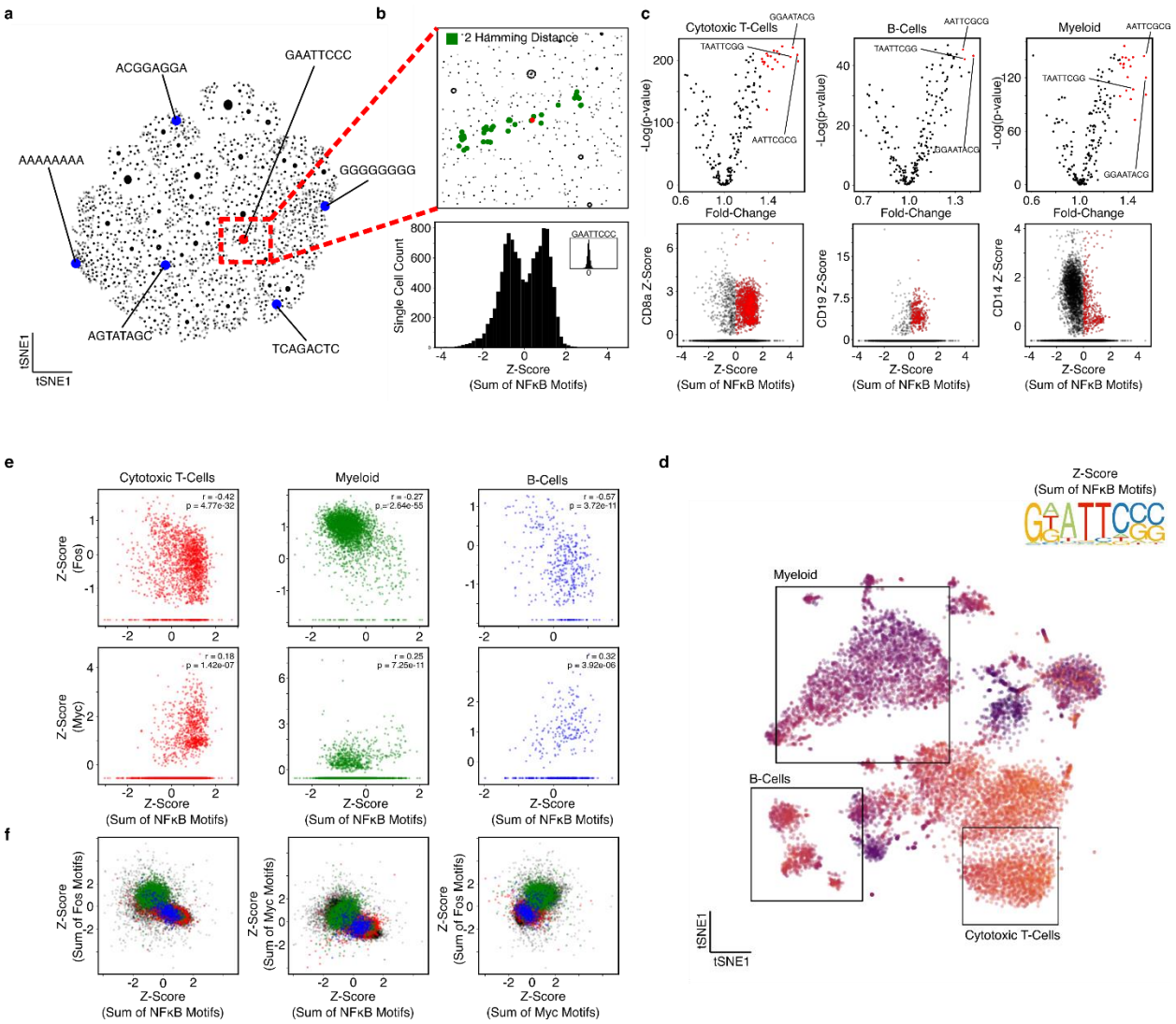


Figure 3.2: Variability in motif utilization matches known transcription factor consensus. (a) Motifs are mapped to dimensionally reduced tSNE space based on hamming distance to 6 benchmark motifs **(b)** NF-κB motifs within 2 hamming distance of "GAATTCCC" are used sum total NF-κB engagement **(c)** Comparison of NF-κB motifs across different cell types show upregulation of similar motifs for highly engaged cells within the population of each cellular sub-type **(d)** Aggregated weighted frequency within the population recapitulates the known consensus binding sequence of NF-κB **(e)** Comparisons across cell type for different transcription factors show differences and general trends between transcription factors for different cell types. **(f)** While cell-type specific behavior follows general trends, when all cell types are shown, there are clear differences in aggregated motif behavior.

We then use these two populations of NF-κB engagement to look at NF-κB-specific binding differences among single cells in different cell types. We split the categories by using the sum of NF-κB motif utilization and setting the threshold at 0 z-score. What we find is that across multiple cell types (cytotoxic T-cells, B-cells, and myeloid cells), there is persistent upregulation of specific binding motifs in highly-engaged cells (**Figure 3.2c**). However, interestingly, there are not significant cell-type specific differences across motif utilization and matches up consistently with the weighted average across all single cells. While this may be an NF-κB specific feature, interestingly, we find that the weighted average of motif utilization matches up with the known consensus sequence for NF-κB (**Figure 3.2d**).

We next looked at multiple transcription factors to try to understand the landscape for transcription factor engagement. We hypothesized that we could tease out effects of crosstalk and inhibition across different transcription factors by looking at transcription factor engagement across the population for different transcription factor engagement levels (**Figure 3.2e**). We first use different cell types to determine if there are cell type specific differences in crosstalk across transcription factors. We compare Myc, NF-κB, and Fos and find that while there are differences in the distribution of cells, the overall trend between transcription factors remains the same. While Myc is correlated with NF-κB engagement, Fos is the opposite and is anticorrelated. Interestingly when we plot al cell types together, the engagement for Myc and NF-κB switches to an overall downward trend (**Figure 3.2f**). This suggests that while the general utilization of transcription factors remains consistent within the population, individual cell-types may diverge in cross-talk of transcription factors.

Motif utilization can be used for transcription factor discovery

While there is an immense value to understanding existing and known transcription factor dynamics and engagement, there is also an equal importance to discovery of new mechanisms and transcription factors. We can use motiFATE to probe unknown cellular transitions to identify transcription factor dynamics along the transition trajectory.

We use EBV infected B-cells and look at re-activation in these cells. Re-activation from latency has immediate importance in EBV infections and leads to viral relapse in patients affected with EBV. We use this model system to investigate why reactivation happens and why some cells are able to escape reactivation. To do this, we first track the viral progression of cells after re-activation to get an approximate trajectory of cells along the transition landscape (**Figure 3.3a**). By breaking up cells across this viral fraction trajectory we can group cells into multiple groups (**Figure 3.3b**). Across all groups there are motifs that are upregulated and downregulated (**Figure 3.3c**), however in group 'D,' due to the high fraction of viral RNA, we find that there are not as many significant motifs upregulated. This could be due to the decreased host cell RNA content present within this highly infected group.

We are also able to find some motifs that are highly upregulated in group 'A.' Among these motifs are candidates for new transcription factor binding sites. We find palindromic motifs among those upregulated, which highly suggests the presence of a transcription factor binding site.

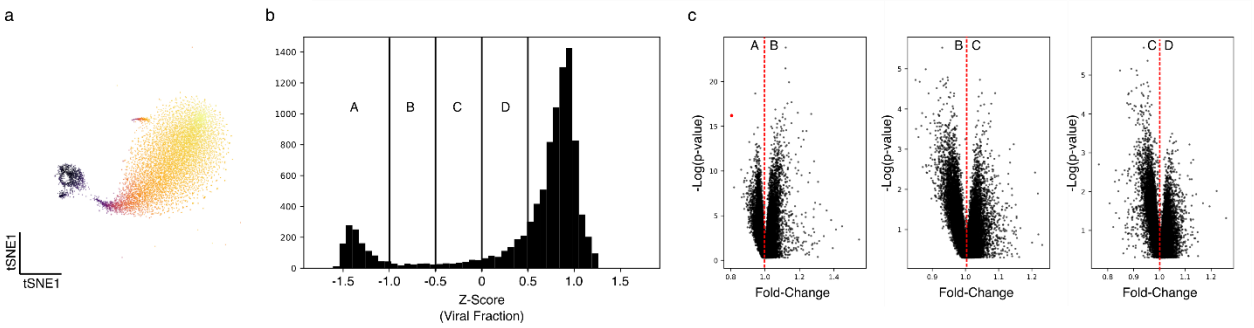


Figure 3.3: motiFATE can be used for transcription factor discovery. (a) single cell EBV reactivation in B cells is shown in dimensionally reduced tSNE. Cells are colored by viral fraction **(b)** Viral fraction separates cells that have undergone reactivation and those that have resisted along with cells in transition states **(c)** High significance motifs can be used to identify transcription factor binding targets for further downstream study.

Discussion

Single-cell RNA sequencing has emerged as a powerful tool often used to understand heterogeneity in cellular state for diverse populations. However, the wealth of data that RNA sequencing represents is underutilized and hides a wealth of hidden information about the genome. Here we demonstrate a simple yet robust application of sequencing data to derive transcription factor regulation from single cell RNA sequencing data. We use our platform, motiFATE to understand transcription factor engagement across different cell types, differentiation trajectories and ultimately use it to identify candidates for novel transcription factors. This study has establishes how RNA-derived transcription binding analysis can be used to understand cellular hierarchy and fate decisions in development.

References

1. G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, R. Jaenisch, A. Regev, E. S. Lander, A. Affiliations, Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming (2017), doi:10.1101/191056.

2. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

3. T. Suda, J. Suda, M. Ogawa, Disparate differentiation in mouse hemopoietic colonies derived from paired progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2520–4 (1984).

4. S. Jang, S. Choubey, L. Furchtgott, L. N. Zou, A. Doyle, V. Menon, E. B. Loew, A. R. Krostag, R. A. Martinez, L. Madisen, B. P. Levi, S. Ramanathan, Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. *Elife*. **6** (2017), doi:10.7554/ELIFE.20487.

5. I. C. Macaulay, V. Svensson, C. Labalette, L. Ferreira, F. Hamey, T. Voet, S. A. Teichmann, A. Cvejic, Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep.* **14**, 966–977 (2016).

6. C. E. Teh, J.-N. Gong, D. Segal, T. Tan, C. J. Vandenberg, P. L. Fedele, M. S. Y. Low, G. Grigoriadis, S. J. Harrison, A. Strasser, A. W. Roberts, D. C. S. Huang, G. P. Nolan, D. H. D. Gray, M. E. Ko, Deep profiling of apoptotic pathways with mass cytometry identifies a synergistic drug combination for killing myeloma cells. *Cell Death Differ. 2020 277*. **27**, 2217–2233 (2020).

7. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, S. Dudoit, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. **19**, 477 (2018).

8. R. Losick, C. Desplan, Stochasticity and cell fate. *Science (80-.).* **320** (2008), pp. 65–68.

9. A. McKenna, G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier, J. Shendure, Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*. **353**, aaf7907 (2016).

10. A. K. Casasent, A. Schalck, R. Gao, E. Sei, A. Long, W. Pangburn, T. Casasent, F. Meric-Bernstam, M. E. Edgerton, N. E. Navin, Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell*. **172**, 205-217.e12 (2018).

11. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25** (2000), pp.

25–29.

12. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. Von Mering, STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

13. T. H. Leung, A. Hoffmann, D. Baltimore, One Nucleotide in a κB Site Can Determine Cofactor Specificity for NF-κB Dimers. *Cell*. **118**, 453–464 (2004).

14. J. Berg, S. Willmann, M. Lässig, Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol. 2004 41*. **4**, 1–12 (2004).

15. Z. Wang, P. Wang, Y. Li, H. Peng, Y. Zhu, N. Mohandas, J. Liu, Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies. *Signal Transduct. Target. Ther. 2021 61*. **6**, 1–16 (2021).

16. I. L. Ibarra, N. M. Hollmann, B. Klaus, S. Augsten, B. Velten, J. Hennig, J. B. Zaugg, Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nat. Commun. 2020 111*. **11**, 1–16 (2020).

Acknowledgements:

Author Contributions: P.P. developed and ran computational pipeline; N.D. did RNA-seq pipeline for EBV infected cells; S.T supervised the work.

Correspondence: Savaş Tay, Institute for Molecular Engineering, The University of Chicago, tays@uchicago.edu

Funding:

Competing Interests:

Data and Materials Availability:

Supplementary Information

motiFATE Uses Hidden RNA-seq Underlying Information to Inform Single Cell Transcriptional Factor Activity

Parthiv Patel[1], Nir Drayman[1], Savaş Tay[1]

Materials and Methods

Motif Utilization Calculation

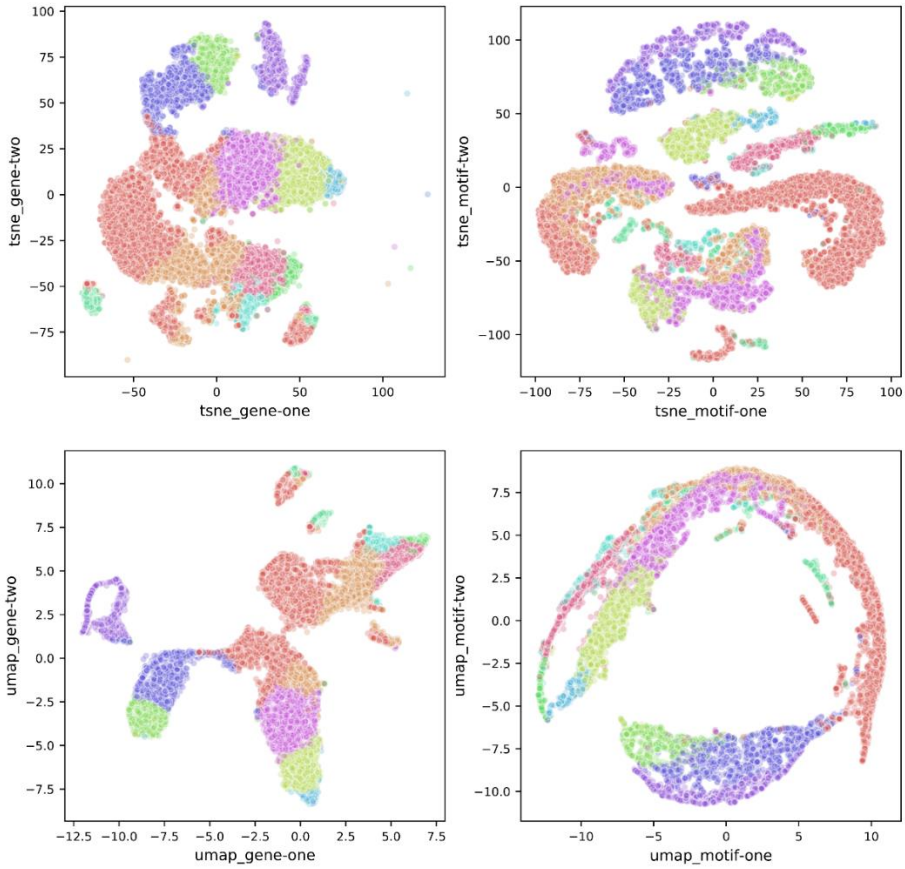
We use reference genome "GRCh38.p12.genome.fa" to infer the genetic information of single cells in a mouse host. Using this reference genome, we populate all possible motifs that occur within 5000bp of the transcriptional start site by using an 8bp sliding window for this study (these attributes are tunable based on the use case). By identifying the motifs that are present in all genes we construct a matrix that represent the motif counts for each gene. We then transform a given single cell RNA-seq dataset by matrix multiplying the motif matrix with single cell RNA-seq matrix to get motif utilization for each single cell.

For visualization of motif-space, we benchmark all motifs to 6 standard sequences including 'AAAAAAAA' and 'GGGGGGGG' to find hamming distance between them. We then dimensionally reduce this down to two variables and plot an approximation to sequence similarity within a 2-d plot.

Transcription Factor Engagement Calculation

To find transcription factor engagement, we aggregate motifs that have similar sequences and sum the total utilization for each motif. To find similar motifs, we use hamming distance to measure

the sequence similarity between a base motif and all other motifs. We then apply a cutoff of 2 hamming distance to find only very similar motifs to the base. Summing up motifs that are within 2 hamming distance results in a comprehensive view of transcription factor adaptability and generalizability. We then standardize these values for comparison and visualization on t-SNE or UMAP.



Supplementary Figure 3.1: Motifs can be used to recapture dimensionally reduced visualization. (a) Using both tSNE and UMAP shows that we can recapture cell transition states from gene-expression alone using just motifs utilization.