Noise Propagation and Information Transmission in the Tumor Necrosis Factor Signaling

Pathway

by

Alexander Jaewoong Rhee

A dissertation submitted to Johns Hopkins University in conformity with the

requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

May, 2014

Abstract

Biological noise is generally defined as the non-genetic variability that arises in populations. For instance, identical twins, although very similar in appearance, will commonly display slightly different phenotypes. Likewise, daughter cells sharing the same genetic material may differentiate along divergent paths. In the past decade, there have been considerable advances in understanding the genetic mechanisms underpinning this variability; however, there still remain unanswered questions surrounding how signaling networks contribute to biological noise and how this noise sets limitations on intracellular information transmission. In the first half of this thesis, we demonstrate that a linear relationship between signal transduction responses allows one to quantify and map the propagation of noise along different parts of a signaling network, even if the network is complex and partially defined. We discover that the JNK pathway generates higher noise than the NF-κB pathway while the activation of c-Jun adds a greater amount of noise than the activation of ATF-2. In addition, by analyzing the negative feedback mechanisms mediated by the protein A20, we find that A20 can suppress noise in the activation of ATF-2 by separately inhibiting the tumor necrosis factor (TNF) receptor complex and JNK pathway. In the second half of this thesis, we will describe an integrative theoretical and experimental framework, based on the formalism of information theory, to quantitatively predict and measure the amount of information transduced by molecular and cellular networks. Analyzing TNF signaling, we find that individual TNF signaling pathways transduce information only sufficient for accurate binary decisions, and an upstream bottleneck limits the

information gained via multiple integrated pathways. In this dissertation, we demonstrate that the application of engineering concepts proves to be of great utility in uncovering novel characteristics of biological noise. We anticipate that these contributions will help move biology closer towards a more predictable and rule-based engineering discipline allowing us to design de novo biological solutions to pressing issues.

Advisor: Andre Levchenko, PhD.

Readers: Joel Bader, PhD; Andre Levchenko, PhD.

Thesis Committee: Joel Bader, PhD; Andre Levchenko, PhD; Jin Zhang, PhD.

Acknowledgements

First of all, I would like to thank my advisor Dr. Levchenko for giving me the chance to work in his lab. Having the opportunity to work alongside some of the world's greatest scientists does not come by very often, and I feel very fortunate to have had the experience. I am also very grateful for the flexibility given to me to pursue interests outside of science, encouragement to pursue novel ideas, and endless patience during my development as a scientist. I would also like to thank Dr. Raymond Cheong whom I worked with on a day-to-day basis for many years. He was instrumental in sharpening my thought process, and I am indebted to him for his endless generosity and patience. I would like to thank my committee members Drs. Joel Bader and Jin Zhang for taking their time to give helpful suggestions for improving this body of work. Lastly, I would like to thank all the friends I have made in Baltimore in the past years for supporting me and making my time in Baltimore immensely more enjoyable.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction and background

1.1 Biological noise

What makes a biological organism unique? Typically, genetic material will account for the majority of differences present in two unrelated individuals. However, organisms that have identical origins, such as clones or identical twins, will necessarily share the same genetic material but may still exhibit significant differences in phenotype[1]. Therefore, there exist non-genetic factors that contribute significantly to biological diversity, commonly known as biological noise. To understand the origins of this variability, it is helpful to re-assess some of our intuitions of the world.

We often think of the physical laws of our day-to-day lives as being foreseeable and deterministic. A clock's pendulum will swing with a predictable cadence, or gravity's pull on an apple will allow us to anticipate its trajectory. However, nature at the microscopic level is subject to an entirely different set of non-deterministic phenomena. At this scale, Brownian motion and molecular vibrations instead of Newtonian physics are the predominant forces governing the behavior of small molecules. Typically, most intracellular molecules exist in sufficient abundance such that these stochastic properties are masked; the behavior of molecules in large quantities can be predicted by known chemical kinetics. However, since cells often have very limited quantities of certain critical molecules, stochastic behavior will predominate under these circumstances.

For example, genes that are present as one or few copies per cell will often lead to

stochastic transcription factor binding and unbinding events that lead to a series of

subsequent repercussions that may influence cellular decision making.   In the following

sections, we will detail how cellular noise has been characterized in the past decade and

the phenotypic consequences

1.2 Sources of Biological Noise

Evidence of cell-to-cell variability is not new. Over half a century ago,

researchers reported that the production of beta-galactosidase in individual *Escherichia*

coli cells exhibited an "all-or-none" phenomenon, either fully induced or not expressed at

all[2]. In the following decades, there were additional breakthroughs in the mathematical

modeling of stochastic chemical kinetics[3-8]. However, further investigations into

biological noise were halted by a lack of available technology with single cell resolution.

More recently, an explosion of interest in the field was ignited after Elowitz et al.

pioneered the equivalent dual-reporter method of gene expression in Escherichia coli[9,10].

This method involves the simultaneous expression of two distinguishable fluorescent

reporter proteins in individual cells under statistically equivalent conditions (e.g. identical

promoters, equivalent integration sites, etc.)[9-11]. The observed difference between

reporter expression within a cell is thought to result from stochastic chemical kinetics that

randomly and independently affect both reporters, and is referred to as intrinsic noise.

The remaining reporter variability originates from the factors that simultaneously affect

both reporters equally within an individual cell but vary from cell to cell, and is referred

to as extrinsic noise. The extrinsic factors can include the expression levels of RNA

polymerase, ribosome number, cell size, or cell cycle stage, all thought to affect the

reporters in a similar manner within a cell.
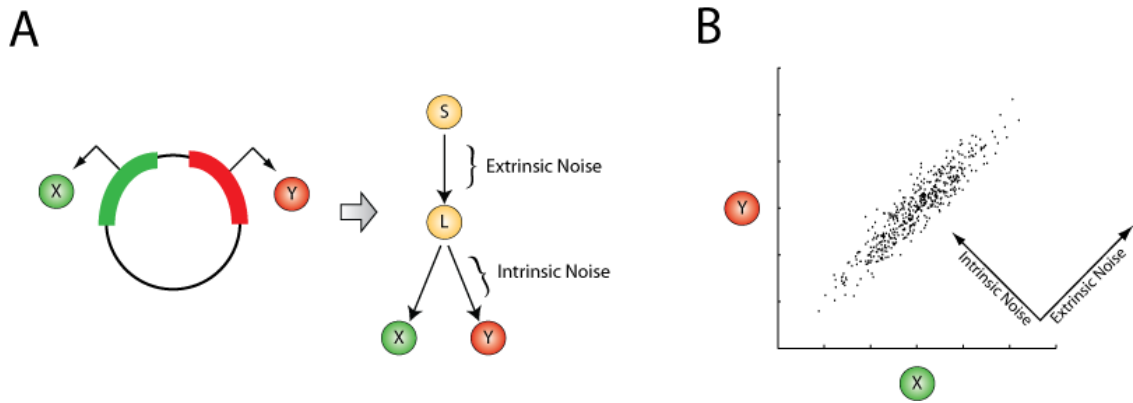
Figure 1.1: Dual reporter method for noise decomposition. (A) Schematic of the equivalent dual reporter method. Two genes that encode for two distinguishable and statistically equivalent fluorescent reporters (**X** and **Y**) can conceptually be reformulated as a 4-node branch motif. **S** can represent the cellular genetic background, **L** can represent the overall activity of the gene expression machinery in a given cell, and **X** and **Y** can represent the expression levels of the reporters. Thus, extrinsic noise is introduced in the segment **S** → **L**, and intrinsic noise is introduced in the segments downstream of **L**, intrinsic noise is acquired. (B) Simulated results for the reporters given in *A*. Each point corresponds to the expression level of the preorters in a single cell. Extrinsic noise causes points to spread out along the diagonal **Y** = **X** while intrinsic noise causes the points to spread out in the direction orthogonal to this line.

Because the two reporters are equivalent, the dissimilarity between the expression

of the reporters in a given cell can be ascribed to stochasticity or intrinsic noise. As a

result, the variance of this difference can further be shown to be proportional to the

intrinsic noise (see section 2.2.1). For these reasons, we can understand why extrinsic

noise is typically depicted to be in the direction of the line $Y = X$, and the intrinsic noise

is orthogonal to this line (Fig. 1.1*B*). In the following sections, we will detail recent

studies characterizing these two classes of noise.

1.2.1 Intrinsic Noise

An early theoretical model of the intrinsic noise of a single gene suggested that intrinsic noise originated from low transcription and high translation rates[12]. Few mRNA molecules would be created in a single cell, but each transcript would be directly responsible for large bursts of protein. Over long periods of time, this behavior would be predicted to cause significant cell-to-cell variability. To investigate this hypothesis, Ozbudak et al. sought to characterize the dependence of noise strength, as represented by the fano factor $\frac{\sigma^2}{\mu}$, on transcriptional and translational activity by creating a prokaryotic model with a chromosomally incorporated GFP reporter gene[13]. The transcriptional and translational rates were independently varied by increasing the activity of an inducible promoter upstream of the reporter gene and introducing point mutations in the mRNA ribosomal binding sites respectively. They found that the noise strength was largely independent of the rate of transcription but strongly dependent on the rate of translation, which provided confirmation that translational efficiency was the primary driving mechanism behind intrinsic noise.

Subsequent studies found additional evidence to validate this hypothesis. Normally, there is considerable difficulty in imaging protein production at the single molecule level. However, Yu et al. engineered a YFP molecule to localize at the membrane of E. coli cells where it could be detected with single molecule resolution[14]. They found that protein production occurs in bursts, each burst originating from a stochastically produced mRNA molecule. Similar results were found in a separate set of

experiments by Cai et al. who found single cell bursts of beta-galactosidase by E. coli cells[15].

In contrast with prokaryotic cells, intrinsic noise in eukaryotic cells was found to have a stronger dependence on stochastic transcriptional events. Blake et al., in a similar experiment to that undertaken by Ozbudak et al., constructed a genetic network within isogenic yeast cells that was under the control of both native and artificial transcriptional regulation[16]. By manipulating the induction of the promoters and swapping codon variants, they were able to determine that transcriptional activity contributed significantly more than translational activity to cell-to-cell variability in yeast cells. Direct evidence of this transcriptional dependence in eukaryotes was finally found after Chubb et al. were able to visualize transcriptional bursting in Dictyostelium discoideum with the use of the ms2 mRNA reporter system[17]. One hypothesis for the eukaryotic noise dependence on transcriptional activity was chromatin remodeling. Normally inaccessible to transcription factors, DNA condensed around histones would become available to transcriptional machinery only after nucleosome architecture was altered to expose regions of DNA. Indirect evidence for this hypothesis was provided by Raser et al[11]. They identified that gene expression noise is promoter specific in yeast cells and that some promoters are well described by a stable promoter state with infrequent transitions. This behavior matches well with a model where the promoter is activated and deactivated by slow chromatin remodeling kinetics.

In higher eukaryotic systems, evidence was found of similar behavior. Raj et al. demonstrated that mammalian cells have considerable cell-to-cell variability in both reporter and native gene transcripts, well described by a model with random gene

activation/inactivation events. They also found that genes that were proximally co-located displayed correlated bursting behavior, whereas distal genes did not[18]. Although there is still some controversy over this phenomenon[19], such behavior would be anticipated if cellular transcription machinery had only localized DNA access, fitting well with a chromatin remodeling theory. Notably, they found that randomness generated by biomolecular fluctuations played an insignificant role in transcriptional noise.

Overall, the above experimental findings have shed considerable insight into the nature of biological noise that originates from short-term stochastic fluctuations of critical molecules. In the following section, we will outline current knowledge of the sources of extrinsic noise.

1.2.2 Extrinsic Noise

Although much progress has been made in understanding the nature of intrinsic noise, the origins and impact of extrinsic noise are less clearly understood. One problem characterizing extrinsic noise is that by definition, any factor that causes two molecules to covary can be ascribed to extrinsic noise. Depending on how the system is defined, extrinsic noise can be defined to encompass pathway specific noise or cell-wide factors such as translation efficiency. Thus, the scope of the interpretation is broad and ill-framed, unlike intrinsic noise. Yet despite this limitation, several groups have characterized the contributions of cell size[11,19,20], cell cycle stage[21], chromosomal location[22], and environmental microfluctuations[23,24] to cell-cell variability.

Another major hypothesized source of extrinsic noise is the variable cell-to-cell concentration of gene expression machinery (metabolites, ribosomes, polymerases)[9,10]. The stochastic production of such essential proteins would likely contribute to the observed heterogeneous population-level distribution, but others have suggested that another factor, the random asymmetric partitioning by means of cell division, could play a significant role. Researchers have hypothesized that during cell division, the mother cell will partition its cell materials according to a random binomial distribution resulting in an uneven allocation of cell materials within the daughter cells[25-27]. Thus, a single founding cell, over many cell divisions, would generate a population of cells with an uneven distribution of gene expression machinery. A later experimental study by Rosenfeld et al. sought to experimentally assess this theory by examining how a series of cellular divisions affected the dilution of fluorescent proteins in E. coli[21]. They confirmed that partitioning of the fluorescent proteins followed a binomial distribution and found that although autocorrelations for intrinsic noise decayed rapidly, autocorrelations for total noise lasted approximately one cell cycle. In higher eukaryotes, Sigal et al. reported a longer cellular partition memory[28]. By tagging many proteins with YFP, they were able to demonstrate that human cells had a persistent memory of several cell cycles of any initial unequal distribution of proteins. Consequently, it would require several generations for a high expressing cell to become a low expressing cell

1.3 Consequences of biological noise

At first glance, biological noise appears to be a maladaptive trait. The biochemical noise underpinning intracellular signaling invariably subjects cells to the forces of randomness; thus, handicapping their ability to make accurate decisions from environmental information[21,29,30]. However, investigations in the past decade have shown that although noise could be considered deleterious at the single cell level, the heterogeneity may be beneficial at the population level. For example, evidence has shown that biological noise is under positive selection pressure[22], facilitates adaptive evolution[31], and provides a mechanism for cellular diversity[32]. Here in the following sections we describe specific consequences of biological noise in a variety of model organisms.

1.3.1 Stochastic state switching

One population-level benefit of stochastic gene expression noise is that it can serve as an engine to increase phenotypic heterogeneity within a group of cells. This diversity can increase the odds that at a fraction of the cells will persist in the face of negative selection pressure and replenish the population in the future. An illustrative example of this behavior can be found in the transient bacterial state of competence. In the competent state, from its immediate surroundings, an individual bacterial cell takes in DNA through its membrane and incorporates it into its genome through recombination.

However, only a random portion of the entire bacterial population will be in the competent state at any given time. Clearly, this 'bet-hedging' strategy ensures that some fraction of bacteria will acquire beneficial or harmful DNA present in the local microenvironment. If the DNA is deleterious, the damage is limited; only a fraction of the population will suffer. Otherwise, if the DNA is beneficial, a fraction of the cells will benefit from this conferred fitness advantage increasing the odds that the population will continue to persist in the face of future selection pressures. Therefore, although individual cells may occasionally fare poorly under this mechanism, the population will prosper as a whole. Guel et al. studied this phenomenon in B. Subtilis[33]. The genetic network underlying differentiation into the competent state has been well characterized to rely on the protein ComK, a key regulator that affects hundreds of genes[33-36]. By monitoring genes that are critically involved in determining competence, Guel et al. found that small stochastic perturbations in ComK could quickly escalate allowing the cell to temporarily exit the stable vegetative state and transiently enter into competence.

Similar network mechanisms were found in mouse embryonic stem (ES) cells. The transcription factors Nanog and Oct4 are two of several transcription factors in ES cells that determine pluripotency. Typically, ES cells exist in a stable non-pluripotent state; however, Kalmar et al. showed that stochastic fluctuations in the transcripts of both Nanog and Oct4 allow individual cells to excitably exit the stable state and transiently acquire pluripotency[37]. This mechanism of transient state switching would permit a fraction of the cellular population to be continuously primed for differentiation, a tactic beneficial to the population in environmental situations that require rapid decision-making and adaptation.

1.3.2 Heterogeneous cellular responses

As previously demonstrated, although biological noise can have a substantial impact on cellular decision making at short time scales, cellular decision making is also dependent on the state of thousands of proteins at any given time[19,38-40]. These protein levels can vary significantly from cell-to-cell causing individual cells to exhibit heterogeneous behavior in response to uniform physiological stimuli[30,41-44]. Spencer et al. examined this phenomenon by investigating how TNF-related apoptosis-inducing ligand (TRAIL), a potent initiator of apoptosis in human cells, initiated fractional cell death in a population of cells[45]. By quantifying caspase activation, an early predictor of cell death, after exposure to TRAIL, they found that the time between TRAIL exposure and caspase activation was highly variable within a clonal population and could largely be ascribed to existing apoptotic protein concentrations prior to TRAIL exposure. These protein states could be inherited but rapid protein synthesis would inevitably cause a rapid divergence in sister cells such that they would be no more alike than a pair of randomly selected cells.

A similar study was conducted by Cohen et al[46]. In a tour-de-force undertaking, they created over 1200 human lung carcinoma cell lines, each with a unique protein tagged with YFP. They then subjected these cells to a drug that caused DNA strand breaks, transcription inhibition, and ultimately cell death to determine how cell survival was dependent on protein concentrations. They found that 20 to 30 hours after the addition of the drug, the protein dynamics of individual cells began to diverge dramatically. Although most proteins displayed little to no change in abundance, the

cellular concentration of 24 proteins progressed to a bimodal distribution in a fashion that occasionally corresponded to cellular outcomes. The concentrations of two proteins in particular, a RNA helicase and a DNA replication factor, were found to be highly correlated with cell fate: an upregulation correlated with cell survival while a downregulation was correlated with cell death. Knocking down the RNA helicase with RNA interference accelerated cell death, suggesting a causal mechanism between protein concentration and cell fate. Thus, they concluded that the survival of individual cells was dependent on the concentrations of specific proteins.

We have seen that biological noise can be directly responsible for cellular phenotypic heterogeneity through a variety of mechanisms: stochastic fluctuations in the transcriptional and translational machinery or the variable abundance of cellular proteins. In the following section, we will briefly describe the model pathway chosen for investigation.

1.4 The tumor necrosis factor signaling pathway
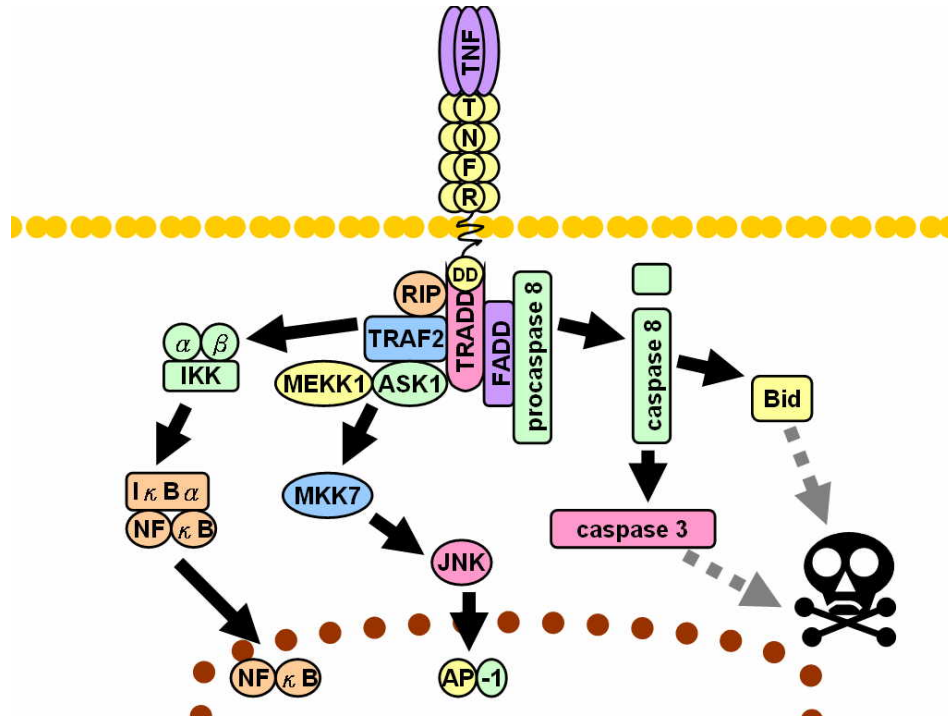
Figure 1.2: The TNF signaling pathway. A schematic of the TNF signaling pathway. Briefly, TNF activates the TNF receptor which then activates both the NF-κB pathway and the JNK mediated pathway causing the nuclear translocation of NF-κB and the AP-1 family of transcription factors including c-Jun, and ATF-2. The single cell nuclear concentrations of the transcription factors can then be quantified via immunofluorescence. Taken from[47]

The model intracellular signal transduction system in which we chose to investigate biological noise is the tumor necrosis factor (TNF) signaling pathway. TNF was first noted for initiating apoptosis in tumor cells, and since then, it has been identified to play a critical role in a wide range of pathologies including sepsis, apoptosis, diabetes, cancer, and autoimmune diseases. This pathway is well suited for investigations into cell-to-cell variability as it has been well-studied and is a model system for understanding heterogeneity in mammalian cells[48-53]. Herein, we will briefly

describe the general mechanisms underpinning the activation of this pathway (Fig. 1.2). We refer the reader to published reviews for further detail[54-56].

Soluble TNF primarily signals through the TNFR1 receptor which then, upon activation, recruits several adapter proteins. In the c-Jun N-terminal kinase (JNK) pathway, these activated adapter proteins will recruit and activate apoptosis-stimulated kinase 1 (ASK1), a mitogen-activated kinase kinase kinase (MAPKKK). This enzyme will then initiate a cascade of kinases including JNK that will ultimately result in the activation of several downstream transcription factors including c-Jun and ATF-2.

Highly regulated, NF-κB has been shown to play a pivotal role in many pathologies including inflammation[57] and cancer[58,59]. Thus, under normal conditions, NFκB is bound and sequestered in the cytoplasm by the three IκB isoforms (α,β,γ) preventing association with DNA in the nucleus. In a pathway parallel to JNK, as part of the TNF receptor complex, active adapter proteins will recruit and activate IκB kinase (IKK) which will which then mark the isoforms of IκB for degradation allowing free NF-κB to enter the nucleus and initiate transcription. In the nucleus, NF-κB immediately upregulates the expression of a number of genes including two proteins that mediate negative feedback: IκBα and A20. After IκBα is rapidly synthesized, it binds to and escorts NF-κB out of the nucleus completing a negative feedback loop while A20 destabilizes the TNF receptor complex, hindering prolonged TNF signaling.

1.5 Aims and significance of this research

The sections above have described how, in the past decade, innovation in experimental and mathematical methodologies have considerably advanced our knowledge of the fundamental characteristics of biological noise. We have shown that there is a deep understanding of the stochastic mechanisms that cells employ at the molecular level to generate diversity and substantial knowledge on how differential cell-to-cell protein abundance affects cellular behavior. Yet despite these advances, there still remain outstanding questions on how biological noise propagates through cellular signal transduction networks and how this noise impedes cellular processing of environmental information. To answer these two questions, we will organize this thesis along two major lines.

Specific Aim 1 – Develop mathematical and experimental methodologies to decompose noise within intracellular signaling networks (Chapter 2).

We will first describe a novel mathematical and experimental framework developed as a natural extension of the equivalent dual reporter methodology pioneered by Elowitz et al.[9]. We then use this framework to decompose the noise propagation in the TNF signaling pathway and establish that this method can easily scale to aid in the deconvolution of larger more complicated signaling networks. Finally, we will demonstrate that this framework can be robust and yield useful and even predictive information in the presence of negative feedback loops.

14

Specific Aim 2 – Quantify the amount of information that can be maximally transmitted through the TNF signaling pathway and assess cellular mechanisms to increase information transmission (Chapter 3).

Utilizing principles grounded in information theory, we will develop a novel mathematical framework to quantify the total amount of information that can be passed through the TNF signaling pathway. We will investigate how utilizing multiple branches can facilitate the transfer of additional information, and address, through the inclusion of multiple reporter branches, the effectiveness of time-averaging as a mechanism to average stochastic NF-κB fluctuations and increase information throughput.

Chapter 2: Decomposing noise in the TNF Pathway

2.1 Introduction

The great advantage of the equivalent dual reporter method is the separation of intrinsic noise from extrinsic noise in an experimentally measurable way. This method has been extended to analyze signaling networks; however, it requires simultaneous measurement of two reporters per signaling node of interest[60] which can quickly become experimentally intractable as the size of the system increases.

Expanding on the success of the equivalent dual reporter method, non-equivalent dual reporters have been utilized to great effect in characterizing sources of cell-to-cell variability. For instance, by comparing a reporter for a signaling pathway of interest to a reporter of a constitutively expressed gene, one can separate pathway-specific from general gene expression noise[61,62]. Alternatively, multiple reporters placed within a serial gene expression network can facilitate a comprehensive decomposition of the noise propagation[23]. However, these methods utilize designed networks whose structure is known a priori, facilitating the construction of a specific mathematical framework that then enables such a thorough decomposition. In addition, although equivalent and non-equivalent reporter methods have proven to yield important scientific insights, both methods require reporter genes to be inserted into cells which can hamper efforts to rapidly assess biological noise in a variety of signaling networks. Furthermore, both

reporter methods are limited to the analysis of biological noise at the gene expression level. Thus, despite substantial advances in the characterization of genetic noise, we lack tools needed to understand noise in intracellular signaling.

Here, we present a mathematical generalization of the equivalent dual reporter method that enables meaningful decomposition of signaling network noise using non-equivalent dual reporters. These reporters do not need to be genetically encoded, thus dramatically increasing the scope of systems that can be analyzed. Using this framework, we were able to quantify the relative noisiness of both the downstream mitogen activated protein kinase (MAPK) and NF-κB signaling pathways. We also show that this methodology can be used to identify previously unappreciated feedback mechanisms affecting both MAPK and NF-κB pathways. Overall, this new methodology is revealing and experimentally facile to implement in a system where detailed knowledge of the relevant biochemical mechanisms is unavailable.

2.2 Results

2.2.1 Derivation of noise decomposition framework from equivalent reporter framework

The method for noise decomposition proposed here can be understood as a generalization of the well-known extrinsic/intrinsic noise decomposition pioneered by Elowitz et al.[9]. To demonstrate the relationship between the methods, we note that by conceptualizing the propagation of additive noise through the equivalent dual reporter

17

system, a 4-node branch motif naturally emerges (Fig. 1.1A). The input node S

represents an external factor that modulates the activity of the entire motif. The

intermediate node L represents the noisy intracellular representation of S. In turn, L

modulates the activity of X and Y. For an equivalent reporter system, S can represent the

cellular genetic background, L can represent the overall activity of the gene expression

machinery in a given cell, and X and Y are the expression levels of the reporter genes. In

this case, extrinsic noise is introduced between S and L, and intrinsic noise is introduced

downstream of L. The mathematical expressions defining total, extrinsic, and intrinsic

noise given by Elowitz et al. are shown in non-normalized form in Eq. **2.2.1-3**.

$$\sigma_{tot}^2 = \tfrac{1}{2}(\langle X^2 + Y^2 \rangle - 2\langle X \rangle \langle Y \rangle) = \tfrac{1}{2}(var(X) + var(Y)) \qquad (2.2.1)$$

$$\sigma_{ext}^2 = \langle XY \rangle - \langle X \rangle \langle Y \rangle = cov(X,Y) \qquad (2.2.2)$$

$$\sigma_{int}^2 = \tfrac{1}{2}\langle (X - Y)^2 \rangle = \sigma_{tot}^2 - \sigma_{ext}^2 = \tfrac{1}{2}var(X - Y). \qquad (2.2.3)$$

As illustrated in Section 2.5.2, the total noise is identical to the average variance of the

reporters (Eq. **2.2.1**), a sensible result when the reporters are equivalent. The collection

of factors within a single cell that causes the two reporters to change in synchrony is

defined as the extrinsic noise and is mathematically defined as the covariance between

the reporters (Eq. **2.2.2**), also a sensible result. Thus, the remaining noise is the

difference between the total and extrinsic noise and is defined as the intrinsic noise (Eq.

2.2.3). Because the two reporters are equivalent, the dissimilarity between the expression

of the reporters in a given cell can be ascribed to stochasticity or intrinsic noise. As a

result, the variance of this difference can further be shown to be proportional to the

intrinsic noise (Eq. **2.5.9**). For these reasons, we can understand why extrinsic noise is typically depicted to be in the direction of the line $Y = X$, and the intrinsic noise is orthogonal to this line (Fig. 1.1*B*).

In the more general case in which X and Y are non-equivalent reporters, the assumptions supporting Eqs **2.2.1-3** are no longer valid, and the framework must be reformulated for the more general non-equivalent case. We will demonstrate later on that the equations describing the equivalent reporter method are a special subset of our more general non-equivalent reporter framework.
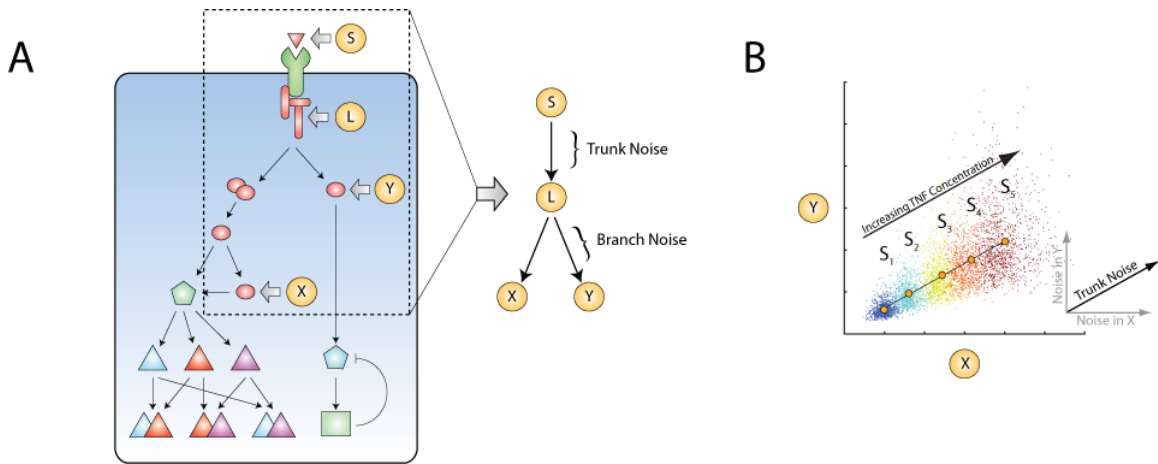
Figure 2.1: Non-equivalent reporters for noise decomposition. (A) A region of interest (ROI) for decomposition is selected from a larger complex intracellular signaling system. The components within the ROI can then be further simplified to a 4-node motif comprised of a ligand, *S*, that binds to its native receptor which sends a signal to a signaling intermediary, the receptor complex *L*. The signal from *S* then propagates down two parallel branches to the readouts *X* and *Y*. We denote the variability that causes coordinated fluctuations in the reporters *X* and *Y* as the *trunk noise* while the noise uniquely contributed by each branch is termed the *branch noise*. (B) Simulated results for individual cells expressing the readouts (*X* and *Y*) given in panel *A* under 5 input levels as denoted by the distinct colors. The means of the readout for each input level are indicated by the circles and fitted by regression to form a basis for decomposition. The observed variability in the $X - Y$ plane is a function of both the trunk and branch noises. The trunk noise adds noise along the basis, and hence each branch noise will add noise parallel to its corresponding axis while orthogonal to the other branch noise axis.

For non-equivalent reporters, S can represent an external stimulus (e.g. ligand concentration), L can represent a signaling intermediate, and X and Y can represent non-equivalent downstream signaling outputs (Fig. 2.1A). In this case, we refer to *trunk noise* as the noise introduced upstream of L and *branch noise* as the noise introduced by a specific branch downstream of L. The biochemical properties (e.g. molecule number variation, stochastic chemical kinetics, etc.) underlying the trunk and branch noise contributions will depend on the specific network being analyzed; thus, the magnitude of noise within the two branches may be unequal. In the instance in which the trunk and branch noise are independent, additive, have zero mean, and X and Y are linearly related to one another (but not necessarily along the line $Y = X$), we can show that the noise values are given by the following:

$$\sigma_{\eta_L}^2 = cov(X, Y) \qquad\qquad (2.2.4)$$

$$\sigma_{\eta_X}^2 = var(X) - \frac{\sigma_{\eta_L}^2}{r} \qquad\qquad (2.2.5)$$

$$\sigma_{\eta_Y}^2 = var(Y) - r \cdot \sigma_{\eta_L}^2 \qquad\qquad (2.2.6)$$

where r represents the slope of the average relationship of Y versus X (see Section 2.5). Eq. **2.2.4** reveals that the trunk noise $\sigma_{\eta_L}^2$ is proportional to the covariance term, thus it is mathematically analogous to extrinsic noise. Similar to the definition of intrinsic noise, the branch noise can then be calculated as the difference between the total noise specific to the branch and the trunk noise (Eq. **2.2.5-6**). Additionally, these equations show that

the three noise components can be extracted from joint measurements of X and Y without knowledge of L, which may be experimentally inaccessible.

Graphically, this system can be depicted in X-Y space as follows (Fig. 2.1B). In the total absence of noise, for a given input S, the activity of X and Y in all cells would be identical and map to a single point as shown (Fig. 2.1B). Experimentally, we can estimate this point by exposing many cells to the same stimulus S and computing the average value of X and Y. If S were allowed to vary, then the locus of points would trace out a line defined by the changing input signal. We refer to this line parameterized by S as the geometrical *basis* for the noise decomposition and experimentally estimate it via reduced major axis regression.

For a given S, by introducing only the trunk noise, the spread of (X, Y) activity of individual cells will lie along the basis line, as trunk noise is equivalent to noisy cellular interpretation of the value of the stimulus S. Given that the noise in the two branches are mutually independent, each branch will contribute noise parallel to its respective axis and orthogonal to the noise associated with the other branch. Experimentally, we observe this effect as a two-dimensional distribution for X and Y whose orientation depends on the direction of the basis and the relative magnitude of the trunk and branch noise terms. Finally, we observe that for a given level of S, the graphical depiction in Fig. 2.1B simplifies to the case in Fig. 1.1B if the reporters are equivalent. In particular, $r = 1$ for equivalent reporters so that the basis becomes $Y = X$ and the trunk noise along this line becomes the extrinsic noise. Furthermore, we note that if we average the branch noise values, we will arrive at the definition of intrinsic noise as expected:

21

$$\tfrac{1}{2}(\sigma_X^2 + \sigma_Y^2) = \tfrac{1}{2}\big(var(x) + var(y)\big) - \sigma_{trunk}^2 = \sigma_{tot}^2 - \sigma_{ext}^2 = \sigma_{int}^2.$$

Thus, the equivalent reporter framework is a special case of the more general non-equivalent reporter framework. Next, we will show that these equations can be applied to a biological system which can provide insights into the nature of biological noise in signaling networks.

2.2.2 Pathway-specific noise in the TNF signaling network

Using the above generalized noise decomposition framework, we sought to quantify the noise contributed by the c-Jun and NF-κB pathways when activated by tumor necrosis factor (TNF), a model system for understanding signaling heterogeneity in mammalian cells[48-53,63] to create a detailed understanding of noise propagation through this signaling network. We exposed mouse embryonic fibroblast cells to a wide range of TNF concentrations to elicit the full dynamic response of the transcription factors. For each TNF concentration, we measured the nuclear concentrations of the transcription factors in hundreds of individual cells using quantitative immunocytochemistry (Fig 2.2*B*). We examined the responses at the 30 min. time point, because the translocation of both transcription factors reach their maximum values at this time indicating similar operational timescales[64].
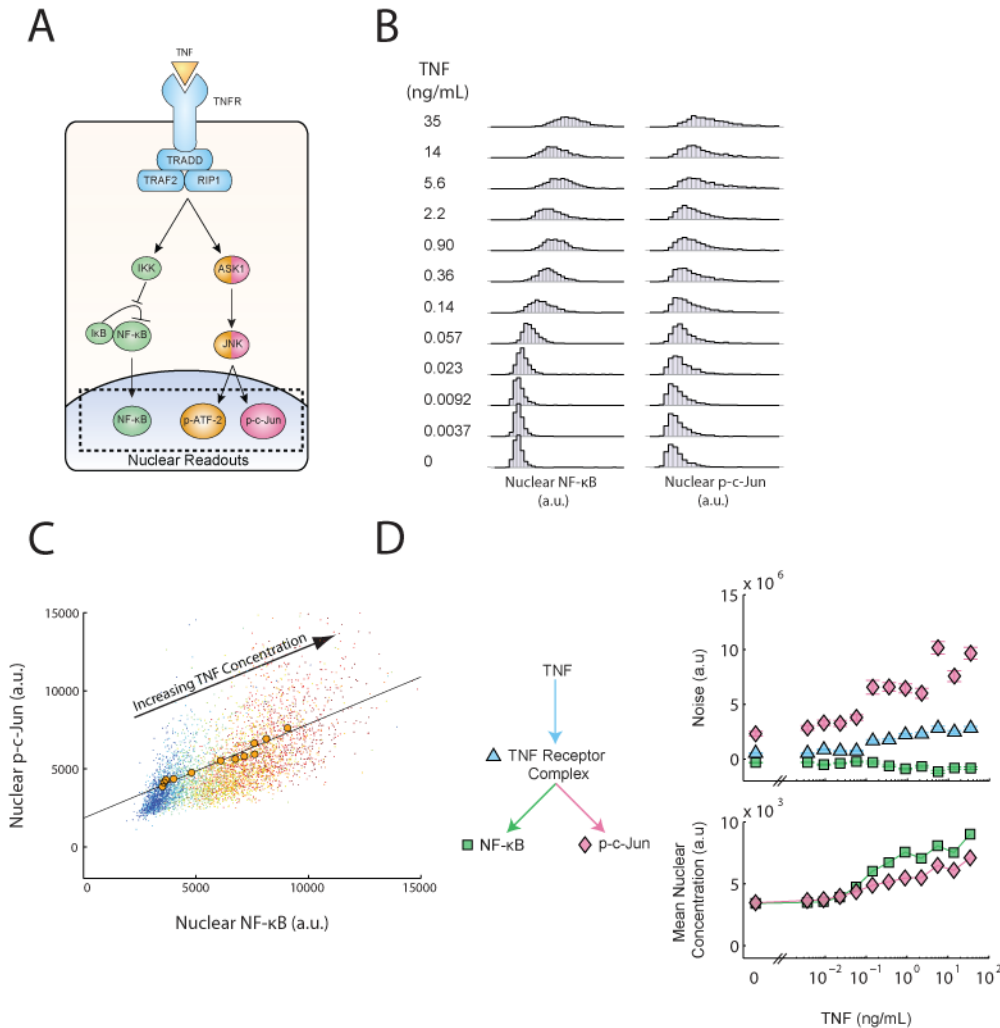
Figure 2.2: Decomposition of the TNF–NF-κB–p-c-Jun signaling pathway. (A) A schematic of the TNF–NF-κB–JNK signaling pathway. Briefly, TNF activates the TNF receptor which then activates both the NF-κB pathway and the JNK mediated pathway causing the nuclear translocation of the transcription factors NF-κB, p-c-Jun, and p-ATF-2. The single cell nuclear concentrations of the transcription factors can then be quantified via immunofluorescence. **(B)** Distributions of NF-κB and p-c-Jun nuclear concentrations in response to TNF. The coordinated single cell nuclear localization of NF-κB and p-c-Jun were measured for their response to a 30-min exposure of TNF and used in calculations to decompose pathway noise. **(C)** Scatter plot of the data given in panel *B*. Individual points are representative of single cells and each color represents a unique TNF concentration as listed in panel *B*. Means at each TNF concentration are denoted by the circles and fit with linear regression to form a basis for noise decomposition. **(D)** The noise decomposition of the TNF–NF-κB–JNK pathway of the data given in panel *B* (top) and the corresponding mean nuclear concentration of both transcription factors (bottom). This figure is shown again as Figure 2.3*C*.

In response to a stimulus, parallel signaling branches can have different dose dependencies leading to complex overall response characteristics, including biphasic ones resulting in complex and highly non-linear behavior[65-67]. However, surprisingly, we find that NF-κB and p-c-Jun levels are proportional to each other over 4 orders of magnitude (Fig. 2.7*A*). Thus, even though the average NF-κB and p-c-Jun levels are nonlinear functions of TNF (Fig. 2.2*D*), they are linearly related (Fig. 2.2*C*).

To better understand the factors contributing to the overall observed variability, we applied Eq. **2.4-6** to decompose the observed noise into a common trunk noise and branch noises specific the NF-κB and JNK pathways (Fig. 2.2*D*). We observed that for the NF-κB pathway, the trunk noise was slightly greater than NF-κB branch noise. Whereas for the JNK pathway, the c-Jun branch noise was greater than the trunk noise. Therefore, although both responses are subject to the noise resulting from common upstream signaling components, the NF-κB pathway introduces less noise to the signaling output in comparison to the JNK pathway. We find that the inflection point in the dose response of the trunk and c-Jun branch noise roughly mirrors the inflection point found in the dose response of the p-c-Jun and NF-κB mean nuclear concentration. The notable similarity in the dose response is likely due to a general correlation found between the scaling of noise and mean protein abundance[68].

2.2.3 Noise decomposition of the TNF network
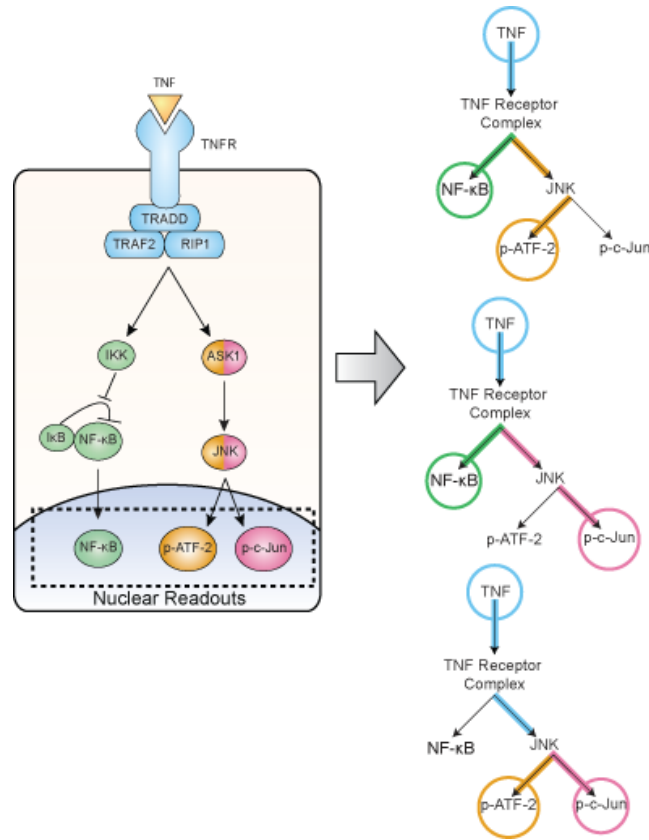
Figure 2.3: Disaggregation of the TNF pathway into 3 4-node motifs. A schematic illustrating the reduction of the TNF–NF-κB–JNK signaling pathway into a 6-node network which is then partitioned into three experimentally tractable 4-node motifs covering all possible transcription factor pairings. Each 4-node motif consists of a TNF input, a signaling intermediary (either the TNF receptor complex or JNK) and two readouts of transcription factor activity.

Next, we sought to demonstrate how our method can be extended to analyze larger, more complex signaling networks. We observed that many signaling networks, including that of TNF, consist of multiple levels of branching raising the question of how much noise each part of the network contributes to the downstream responses. For instance, the TNF network branches into the NF-κB and JNK pathways, and the JNK pathway subsequently branches to activate two transcription factors: ATF-2 and c-Jun.

25

To decompose the noise in this 6-node system, we considered multiple 4-node branch

motifs embedded within the network (Fig. 2.3). We can decompose the noise of each

motif in isolation, but since the three motifs have overlapping portions, we can assemble

a more detailed noise decomposition map of the original network. To perform this

decomposition, we measured, in parallel experiments, the joint pair-wise TNF responses

of NF-κB and p-ATF-2, NF-κB and p-c-Jun, and p-ATF-2 and p-c-Jun.
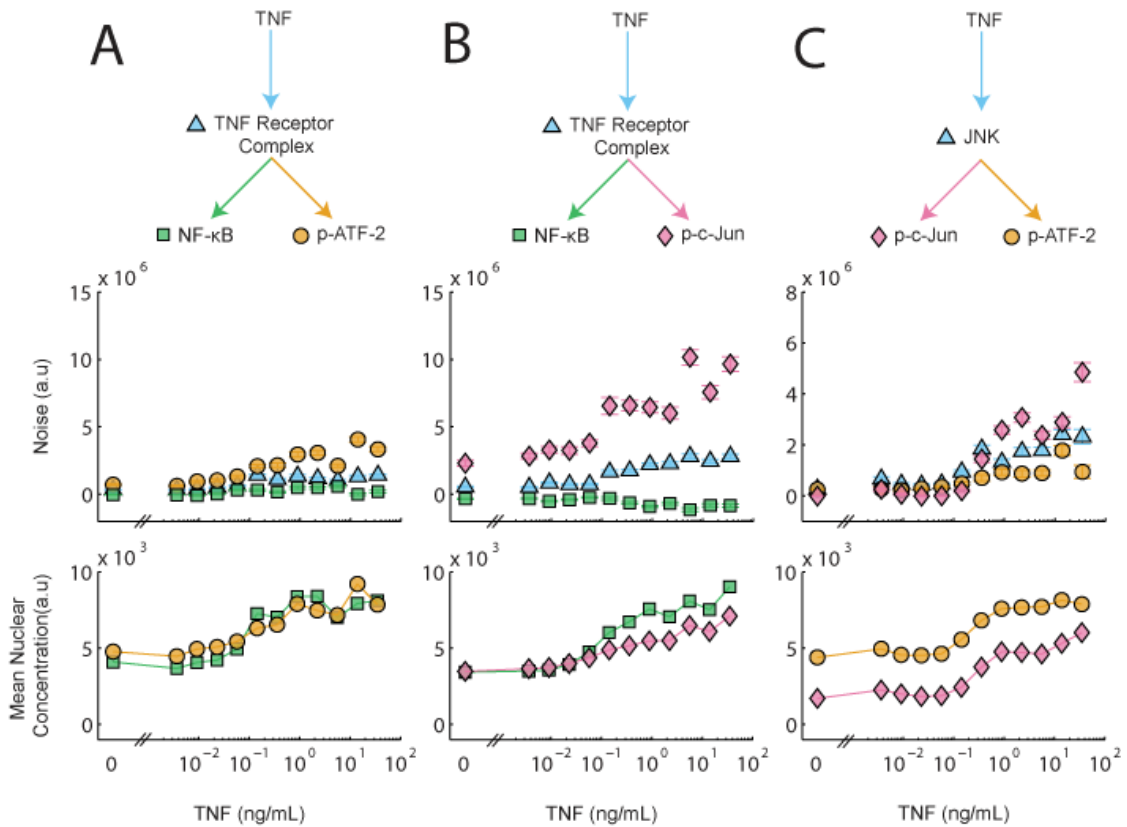
Figure 2.4: Noise Decomposition of the TNF–NF-κB–JNK motifs. (A-C) The noise decomposition of
the 4-node motifs given in Fig. 2.2 (top) and the corresponding mean nuclear concentration of both
transcription factors (bottom). The JNK branch specific noise is higher than both the NF-κB branch
specific noise and the TNF–TNFR trunk noise. Within the JNK pathway the c-Jun branch noise is greater
than the ATF-2 branch noise at higher TNF concentrations.

26

First, we found that the results for the NF-κB/p-ATF-2 pair (Fig. 2.4*A*) were similar to that of the NF-κB/p-c-Jun pair analyzed earlier (Fig. 2.4*B*). Further quantitative analysis revealed that of the noise in the fully activated TNF-NF-κB pathway, ~90% can be ascribed to the trunk portion shared with the TNF–JNK pathway, and the remaining ~10% can be ascribed to the NF-κB specific branch. In comparison, in the TNF-ATF-2 pathway, only approximately 30% of the noise in the ATF-2 pathway originates from the trunk, and the remaining ~70% of the noise arises from the remaining JNK pathway. Next, examining the results for the NF-κB/p-c-Jun pair, (Fig. 2.4*B*) we observe that ~80% of the p-c-Jun noise originates from the c-Jun-specific branch, suggesting that there may be slightly greater noise in the TNFR-c-Jun pathway than in the TNFR-ATF-2 pathway. Indeed, when we directly decomposed the p-ATF-2/p-c-Jun pair, we observed greater noise specific to the c-Jun pathway than compared to the ATF-2 pathway at the higher concentrations of TNF (Fig. 2.4*B*).

The pair-wise analysis can be used to assign relative noise contributions to each part of the TNF signaling network (Fig. 2.5*A*). For instance, if as a reference we assign a noise value of 1 to the initial TNF-TNFR segment, then the noise value in the TNFR-NF-κB segment is ~0.1 (See Section 2.5.5), in order to be consistent with our observation above that the signaling segment upstream of the NF-κB and JNK branch point contributes 90% of the total noise in the TNF- NF-κB pathway and the downstream NF-κB noise contributes the remaining ~10%. Similar calculations can be used to compute the relative noise contributions from the remaining segments. Due to the overlapping portions of the multiple 4-node motifs, certain segments, such as the TNFR-NF-κB

segment, will have multiple noise decomposition estimates. In this situation, the

estimates were averaged (See Section 2.5.5). This analysis yields the relative noise

values shown in Fig. 2.5*A*. Interestingly, the total noise of the TNF-NF-κB pathway is

approximately 30% of the noise present in TNF-ATF-2 pathway and 20% of the noise in

TNF-c-Jun pathway, indicating an asymmetry of pathway specific noise between the JNK

and NF-κB branches in TNF signal processing. This result provides an explanation for

our prior results which demonstrated that the information carrying capacity of the NF-κB

pathway is greater than that of the JNK pathway, with the capacity of both pathways

influenced by a common TNF receptor-level bottleneck[48].
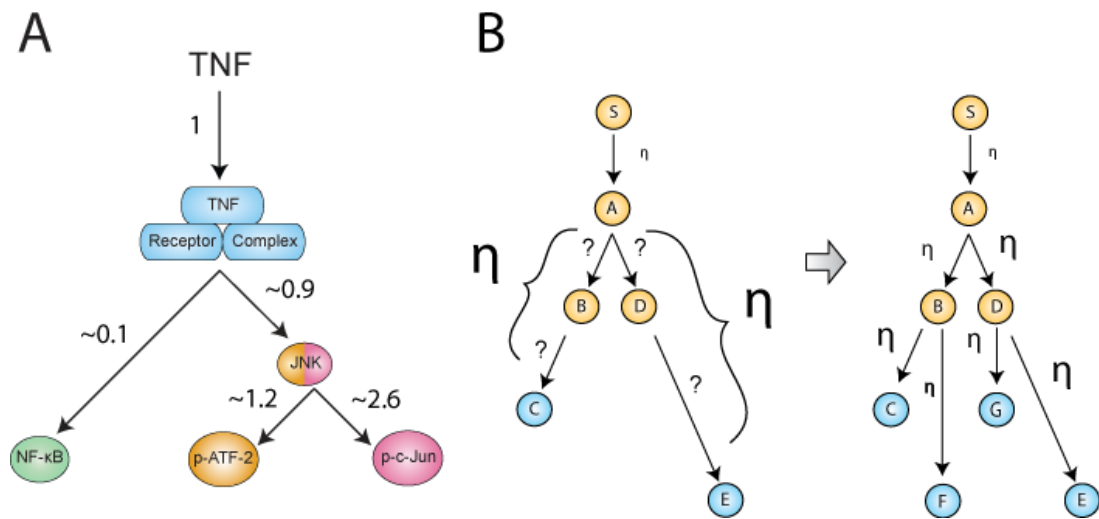
Figure 2.5: Noise decomposition of larger networks. **(A)** Noise decomposition map normalized to the noise that is contributed by the common TNF-TNFR segment, based upon the data in panels B-D. The map demonstrates asymmetry in the amount of noise contributed by the NF-κB and JNK pathways, and shows that the majority of noise in the JNK pathway is contributed downstream of the TNF receptor complex. **(B)** Illustration of noise decomposition of a larger network. Given a hypothetical signaling network of 6 nodes and two readouts (C and E), only 3 noise values can be ascertained. With the addition of two new readouts (F and G), a more comprehensive noise decomposition map can be constructed that can guide further investigations.

Because of the inherent scalability, this noise decomposition methodology can be easily expanded to analyze the noise propagation through larger signaling networks. For example, given a hypothetical signaling network (Fig. 2.5B) and the two downstream readouts, C and E, we can provide only limited noise mapping: the noise contribution of the $A \rightarrow C$ and $A \rightarrow E$ pathway segments in addition to the $S \rightarrow A$ trunk noise. However, with the addition of two more readouts (F and G), we can in principle resolve seven noise values and reconstruct a detailed noise decomposition of the entire network. Such a network noise map would allow one to prioritize further investigations into the physical basis of the noise and identify the portions of the network in which one may expect to find molecular mechanisms that regulate variability.

2.2.4 The impact of feedback on transcription factor variability
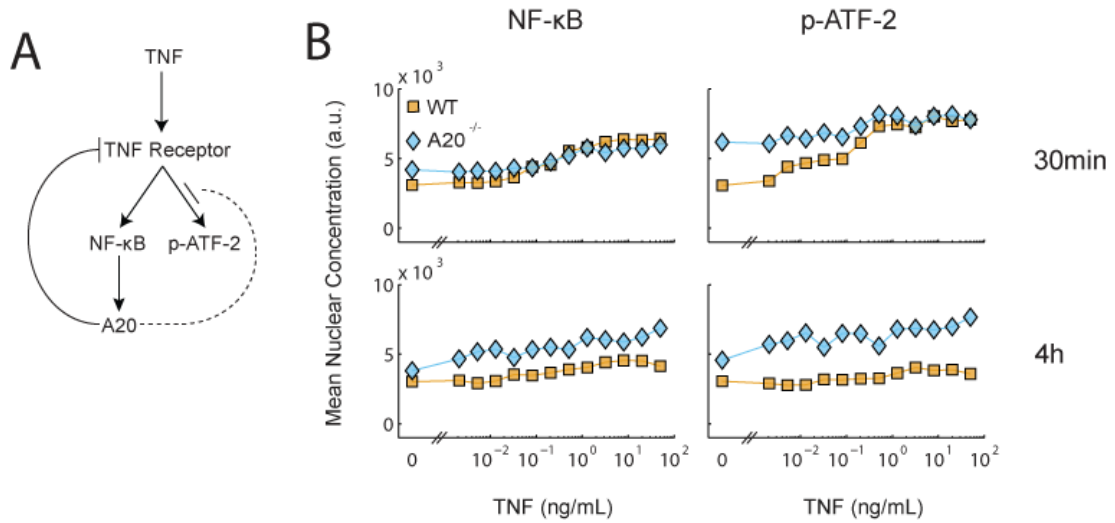
Figure 2.6: A20 functions as a late-acting negative feedback loop. (A) Schematic of the A20 feedback loop. At 4 hours, after upregulation, A20 interferes with the functionality of the TNF receptor complex (solid) and inhibits the JNK pathway (dotted). **(B)** Dose response curves for the mean nuclear concentration of NF-κB and p-ATF-2 in response to TNF in both WT and A20$^{-/-}$ cells at 30 min. and 4 hours.

Negative feedback is a well-known mechanism that cells can use to modulate biochemical noise. To quantitatively demonstrate the effect of negative feedback on noise in TNF signaling, we performed a noise decomposition in wildtype cells and cells lacking A20, an enzyme well-known to inhibit TNF-induced NF-κB activity by destabilizing the TNF receptor complex[69-72] (Fig. 2.6*A*). Destabilization of the receptor complex has been reported to further mitigate downstream JNK activation[73]; however, this mechanism is still controversial[74,75].

We also compared the noise decompositions at the 30 min. and 4 hr. timepoint in these cells (Fig. 2.6*B*), as induced expression of A20 is negligible at the earlier timepoint but maximal by the latter timepoint[63,76]. Importantly, we note that there exists a

consistent linear relationship between NF-κB and p-ATF-2 across two timepoints and

across both wildtype and A20$^{-/-}$ cells, enabling direct comparison of the noise

decomposition results among all four conditions (Fig. 2.7*B*).
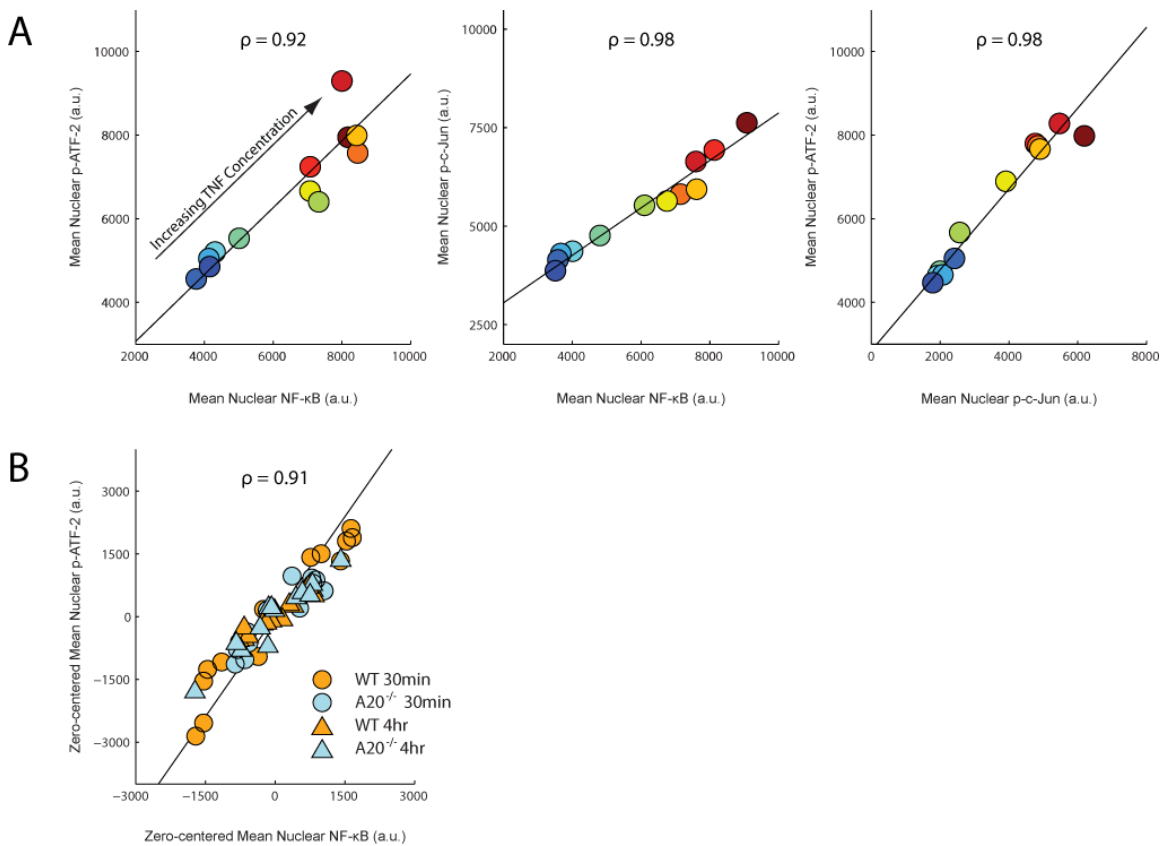
Figure 2.7: Response pairs are linearly related. **(A)** Means of the data given in Fig. 2.4*A-C* are shown. Each circle represents the mean response to a distinct concentration of TNF. The best fit regression lines are shown and used as the basis for noise decomposition. **(B)** Data given in Fig. 2.7*B* were combined, centered about the origin, and plotted. The slope of the linear relationship between NF-κB and p-ATF-2 is the same at both the 30 min. and 4 hr. time point for both WT and A20$^{-/-}$ cells.

At the 30 min. timepoint, we observed that the trunk noise was, on average, slightly

greater in A20$^{-/-}$ cells than wildtype cells, corroborating the ability of A20 to regulate

both NF-κB and JNK pathways at the receptor complex level, whereas there was no

difference in NF-κB-specific noise (Fig. 2.8). The difference in trunk noise was greater

at the 4 hr. timepoint, likely reflecting the difference between the effects of lower basally

expressed A20 versus that of highly induced A20. Unexpectedly, we also observed

markedly larger ATF-2 branch-specific noise in A20$^{-/-}$ cells compared to wildtype cells

with the difference between the cell types being greater at the 4 hr. timepoint than at the

30 min. timepoint. This result indicated that A20 can repress the JNK pathway in a

manner independent from its effects on the TNF receptor complex. At the time that this

prediction was made there was no known direct inhibition of the JNK pathway by A20,

but a later study by Won et al. verified that A20 directly binds to and represses ASK1, a

kinase in the JNK pathway that has no known direct effects on the NF-κB pathway[74].

We note that although negative feedback could potentially violate our assumption that the

trunk and branch noise levels are independent, this experiment demonstrates that on a

practical basis, our noise decomposition framework can yield sensible and even

predictive results regarding the effects of negative feedback on signaling noise.
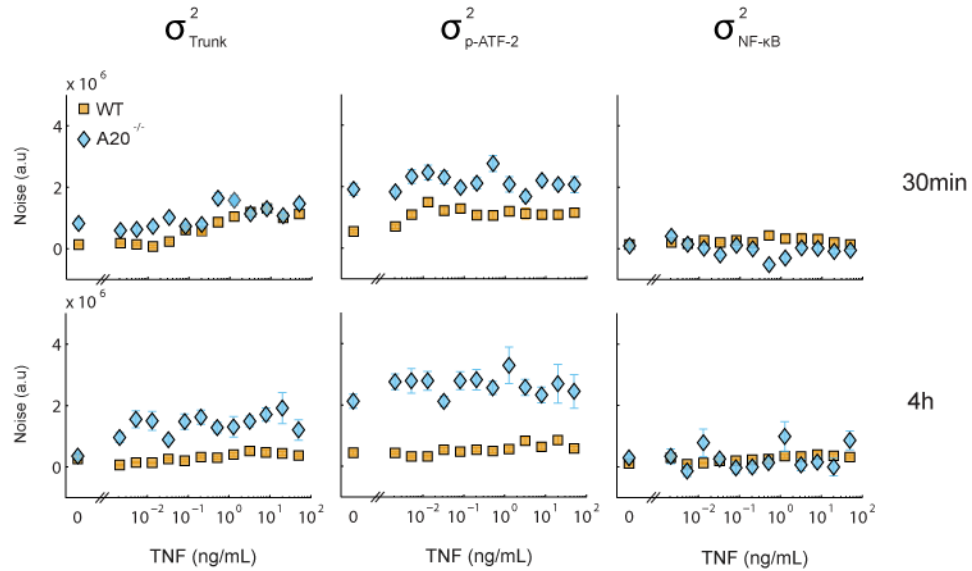
Figure 2.8: Noise decomposition of the WT and A20⁻ᐟ⁻ data. Noise decomposition of the TNF–NF-κB–ATF-2 signaling pathway in WT and A20⁻ᐟ⁻ cells at 30 min. and 4 hours for the dataset shown in panel *B*. Absence of the A20 protein does not affect the noise in the NF-κB branch but causes an increase in the amount of noise in the trunk and ATF-2 branch at both 30 min. and 4 hours. This observation corroborates known information about the mechanisms of A20 regulation.

2.3 Discussion

By utilizing the linear relationships between downstream effectors of the TNF pathway, we developed a mathematical and experimental framework that enables noise decomposition in intracellular signal transduction. This method distinguishes trunk from branch noise and can be derived as a natural extension of extrinsic/intrinsic noise analysis. Corroborating previous results, we showed that there is a greater amount of noise present in the JNK branch than the NF-κB branch and that both branches are subject to a sizable contribution of noise from the TNF receptor complex. More detailed noise mapping of the JNK pathway revealed that within the JNK sub-network, p-c-Jun is subject to greater noise than p-ATF-2.

Examining the impact of negative feedback on noise expression, we found further

evidence that A20 is able to suppress variability at the level of the TNF receptor. We

also unexpectedly discovered an additional mechanism of JNK noise suppression

consistent with a recent observation of the direct inhibition of ASK1 by A20. Although

negative feedback can theoretically complicate the mathematical decomposition by

allowing interactions between noise parameters, we nonetheless observed that the

nonequivalent dual reporter method can be robust to its presence and can provide a useful

first approximation. Furthermore, at a minimum, the noise analysis presented here can be

used to characterize the noise and provide a basis for quantitative comparison against

predictions generated by computational models incorporating details of biochemical

feedbacks.

Although this method requires a linear relationship between the reporters, we

believe it does not tightly constrain the general applicability. In most biological signaling

systems, nonlinear signal-dose responses align to allow for optimal information transfer

which results in responses that are approximately linearly related[77]. Furthermore, we

expect that in the case of nonlinear relationships, this method can be easily extended by

replacing the slope parameter r with the local slope dY/dX. Indeed, the basis used for

decomposition, which is presented here as a line, could be a curve, a 2-dimensional

surface, or higher dimensional manifold depending on the number of responses of interest

and their interrelationships.

We envision that this method and such further generalizations could enable better

measurements of noise which will open avenues into understanding its molecular

underpinnings and usher in a deep understanding of the nature of variability in the TNF

pathway and even more complex signaling systems.

2.4 Materials and methods

2.4.1 Cell culture

Wildtype and A20$^{-/-}$ 3T3-immortalized mouse embryonic fibroblasts (kind gift

from A. Hoffmann, Univ. of California, San Diego) were maintained in low glucose

Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% calf bovine

serum (American Type Culture Collection) and 10 U/mL each of penicillin and

streptomycin (Invitrogen). P65–GFP cells (kind gift from M. Covert, Stanford) were

maintained in high glucose Dulbecco's modified Eagle's medium (Invitrogen)

supplemented with 10% fetal bovine serum (American Type Culture Collection) and 10

U/mL each of penicillin and streptomycin (Invitrogen). Cells were seeded at a density of

approximately 150 cells/mm^2 onto 15mm diameter circular coverslips (Fisher Scientific)

coated with 0.1% gelatin (Sigma), placed in 6 well plates, and then serum starved in

medium with reduced serum concentration (0.1%) overnight before experimentation.

2.4.2 Immunocytochemistry

After exposure to murine TNF (Roche) at the specified concentrations and

duration, the cells were washed 3 times with ice-cold phosphate buffered saline (PBS,

Invitrogen) and fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 20 minutes. The cells were then permeabilized in 0.1% triton X-100 (Sigma) for 5 minutes and blocked in 10% goat serum (Invitrogen) for 60 minutes. Next, the cells were incubated in primary antibody solution. Primary antibody concentrations used were 1:100 rabbit anti-p65 antibody (Santa Cruz), 1:100 mouse anti-phospho-ATF-2 antibody (Santa Cruz), 1:100 mouse anti-phospho-c-Jun (Santa Cruz), 1:100 rabbit anti-phospho-c-Jun (Cell Signaling).

Finally, the cells were incubated in a secondary antibody solution consisting of 1:200 Alexa Fluor 488-conjugated goat anti-rabbit and 1:200 Alexa Fluor 594-conjugated goat anti-mouse antibodies (Invitrogen) for 60 minutes, and 2 μg/mL Hoechst-33258 (Sigma) for 60 minutes. All solutions were made in 10% goat serum (Invitrogen) in PBS, and cells were washed with PBS in between each step. To minimize experimentally-induced variability and to enable quantitative comparisons across conditions, all concentrations of TNF and all cell lines were assayed at the same time using common reagents. Finally, the stained coverslips were mounted on glass microscope slides and imaged on an Axiovert 200M inverted epifluorescence microscope (Zeiss) equipped with Slidebook 4.2 (Intelligent Imaging Innovations). On average, over 350 cells were imaged per experimental condition.

2.4.3 Image and data analysis

Image processing and data analysis were performed using Matlab R2009a (MathWorks). Background correction, nucleus segmentation, and quantification of

nuclear concentrations of NF-κB, phospho-ATF-2, and phospho-c-Jun were performed as described previously[48]. Programs are available upon request. Top and bottom 2nd percentiles of data were discarded to reduce the influence of outliers on the estimates of variance.

2.5 Linear noise decomposition

2.5.1 Derivation of the trunk and branch noise values
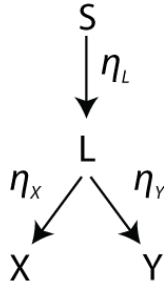
Figure 2.9: Model of 4-node network.

Here, we derive the noise decomposition equations shown in Eqs. **2.2.4-6** in the main text. We begin by examining the top portion of the four node motif shown in Fig. 2.9 where a discrete signal S is transmitted to an intermediary node L. The intermediary L can then be described as a function of the signal S and a stochastic noise term η_L, defined as the trunk noise:

$$L = f(S) + \eta_L. \tag{2.5.1}$$

Downstream, the signal bifurcates along two separate pathways to the readouts X and Y. Since X and Y are both affected by L, they can be represented as distinct linear

37

functions of L (or of some transformation of L) plus a corresponding stochastic noise term η_X or η_Y, defined as the branch noise. This is represented as follows, where m and b are the linear coefficients,

$$X = m_X \cdot L + b_X + \eta_X \qquad (2.5.2)$$

$$Y = m_Y \cdot L + b_Y + \eta_Y.$$

We assume the above noise terms are independent, additive, and with zero mean. By taking the variance of Eq. **2.5.1**, we find that the magnitude of the trunk noise is equal to $var(L)$. Since the noise terms are independent, they have zero pairwise covariance; thus, by taking the covariance of X and Y and rearranging we obtain

$$\frac{cov(X,Y)}{m_X m_Y} = var(L) = \sigma_{\eta_L}^2.$$

This choice defines the trunk noise in the units of L, whereas dividing or multiplying by r, as defined below, can convert the trunk noise into units of X or Y, respectively. Furthermore, by taking the variance of Eq. **2.5.2** we obtain

$$var(X) = m_X^2 \cdot var(L) + \sigma_{\eta_X}^2 \qquad (2.5.3)$$

$$var(Y) = m_Y^2 \cdot var(L) + \sigma_{\eta_Y}^2,$$

where the branch noise terms $\sigma_{\eta_X}^2$ and $\sigma_{\eta_Y}^2$ denote the variance of η_X and η_Y. We can therefore see that

$$\sigma_{\eta_L}^2 = cov(X, Y) \tag{2.5.4}$$

$$\sigma_{\eta_X}^2 = var(X) - \frac{\sigma_{\eta_L}^2}{r} \tag{2.5.5}$$

$$\sigma_{\eta_Y}^2 = var(Y) - r \cdot \sigma_{\eta_L}^2, \tag{2.5.6}$$

where $r = \frac{m_Y}{m_X}$. Importantly, r is the slope of the line of Y versus X in the absence of noise. This line is parameterized directly by L and indirectly by S. Thus, the line can be obtained by calculating the regression of the average of Y versus the average of X at various levels of S, allowing r to be experimentally estimated (Fig. 2.7). Since the variances of X and Y and their covariance are experimentally measurable, Eqs. **2.5.4-6** allow for direct estimation of the branch and trunk noises.

2.5.2 Relation to the method of Elowitz et al.

In this section, we will demonstrate that the trunk/branch decomposition is a more generalized formulation of the methods pioneered by Elowitz et al.[9,10]. To begin, we note the non-normalized definitions of the intrinsic and extrinsic noise:

$$\eta_{int}^2 = \frac{1}{2}\langle(X - Y)^2\rangle; \quad \eta_{ext}^2 = \langle XY\rangle - \langle X\rangle\langle Y\rangle. \tag{2.5.7}$$

From this definition, we can immediately see that the extrinsic noise is equivalent to the previously defined trunk noise (Eq. **2.5.4**). In the case of equivalent dual reporters, X and Y are statistically equivalent, hence $\langle X \rangle = \langle Y \rangle$ and $r = 1$. By taking advantage of these properties, we can then enumerate several parallels between the trunk/branch and the intrinsic/extrinsic methodologies.

To begin, by taking the average of the branch noises from Eqs. **2.5.5-6**, we obtain the expression for intrinsic noise:

$$\frac{1}{2}\left(\sigma_{\eta_X}^2 + \sigma_{\eta_Y}^2\right) = \frac{1}{2}\left(var(X) + var(Y) - 2cov(X,Y)\right)$$

$$= \frac{1}{2}\left(\langle X^2 \rangle - \langle X \rangle^2 + \langle Y^2 \rangle - \langle Y \rangle^2 - 2\langle XY \rangle + 2\langle X \rangle\langle Y \rangle\right)$$

$$= \frac{1}{2}\left(\langle X^2 \rangle + \langle Y^2 \rangle - 2\langle XY \rangle\right)$$

$$= \frac{1}{2}\langle (X - Y)^2 \rangle.$$

$$(2.5.8)$$

Using the above relationships, we can also easily prove statements made in the main text about extrinsic and intrinsic noise. First, using Eq. **2.5.8**, we can show that intrinsic noise is proportional to the variance of the difference in reporter expression:

$$\frac{1}{2}\langle (X - Y)^2 \rangle = \frac{1}{2}\left(var(X) + var(Y) - 2cov(X,Y)\right)$$

$$= \frac{1}{2}var(X - Y).$$

$$(2.5.9)$$

40

Lastly, since total noise is defined as the sum of the intrinsic and extrinsic noise values, we sum the contributions and find that the total noise can be rewritten as the average variance of the reporters:

$$\sigma_{tot}^2 = \sigma_{int}^2 + \sigma_{ext}^2 = \frac{1}{2}\left(\sigma_{\eta_X}^2 + \sigma_{\eta_Y}^2\right) + cov(X,Y)$$

$$= \frac{1}{2}\left(var(X) + var(Y)\right).$$

(2.5.10)

2.5.5 Noise decomposition of larger networks

To decompose a larger system such as the 6-node TNF–NF-κB–JNK network, we first note that it has three embedded 4-node motifs (Fig. 2.3). By decomposing each motif, we find that although we can obtain a single noise estimate for each segment of the larger network, for one portion of the network, we obtain two redundant estimates. For this particular segment, we average these two estimates to obtain a final estimate. For example, by decomposing the noise in the NF-κB/p-c-Jun pairing, we find that 76% of the noise in p-c-Jun can be ascribed to the TNFR to p-c-Jun segment, while the remaining 24% is due to noise at the TNF-TNFR level. In a similar fashion, from the p-ATF-2/p-c-Jun pair we find that 63% of the noise in p-c-Jun can be ascribed to the JNK to p-c-Jun segment. Thus, the signaling segment connecting TNFR to JNK must contribute 76% – 63% = 13% to the variance in p-c-Jun. To assign relative noise contributions to each part of the TNF signaling network, as described in the main text, we

41

normalize all values to the TNF to TNFR segment. Thus, the TNF to TNFR segment

becomes 1, the JNK to p-c-Jun segment becomes $\frac{63\%}{24\%} \sim 2.6$, and we arrive with a

normalized estimate of $\frac{13\%}{24\%} \sim 0.5$ for the TNFR to JNK segment. We conduct the same

analysis by utilizing the noise decomposition from the NF-κB/p-ATF-2 and p-ATF-2/p-c-

Jun pairings and arrive at an estimate of 1.3 for the TNFR to JNK segment. We then

average the two figures to arrive at a final noise estimate of 0.9 for the TNFR to JNK

segment (Fig. 2.5*A*).

2.6 Experimental considerations

2.6.1 Experimental noise compensation

In order to properly measure true biological noise in cellular signaling systems,

the experimental error needs to be quantified and removed from the total measured

variability. In our previous work, using the correlation between direct GFP fluorescence

and the indirect anti-GFP immunofluorescent signal, we estimated that immunostaining

accounts for less than ~12% of the measured variance in the anti-GFP signal[48].

To further validate this estimate, we obtained p65-knockout MEF cells that were

reconstituted with a p65 – GFP fusion protein[50]. The cells were stimulated with a range

of TNF concentrations, fixed, and then immunostained. We observed a strong linear

correlation between the direct and stained p65 measurements, on average $\rho \sim 0.94$ (Fig.

2.10).

As previously shown[48], if we assume that the immunostained p65 measurement is proportional to the p65 concentration and all distributions are Gaussian, we can estimate that $1 - 0.94^2 \sim 12\%$ of the observed variance is contributed by experimental noise which is supported by previous reports[48,78]. Therefore, to correct for the experimental noise, we reduced all variance quantities by 12%. We find that this does not significantly alter any conclusions.
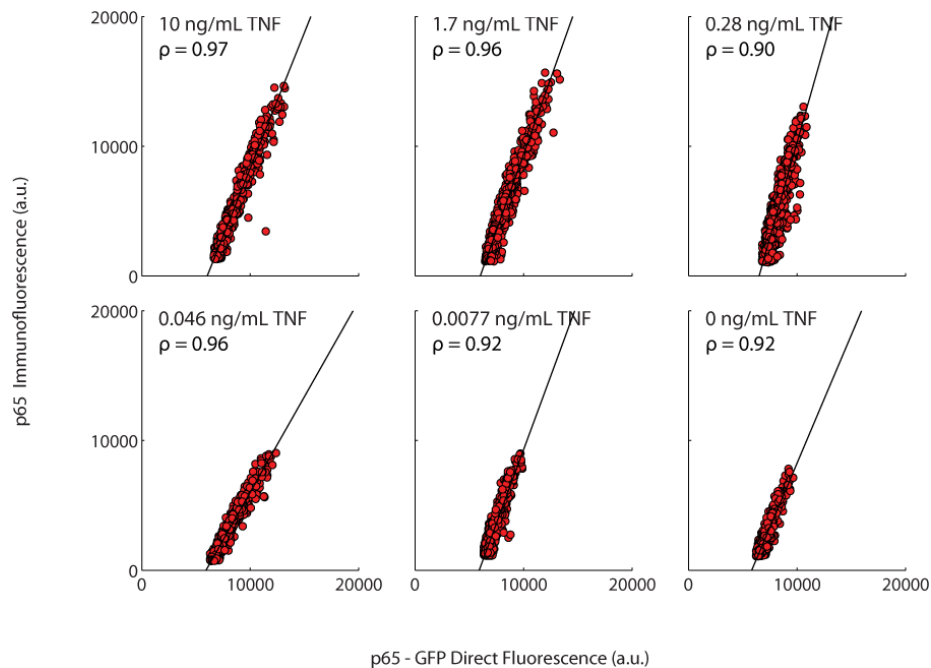
Figure 2.10: Experimental noise associated with NF-κB immunofluorescence. p65-knockout cells stably expressing a p65–GFP fusion protein were exposed to a range of TNF concentrations, fixed, and immunostained for p65. The correlation between the direct and immunostained p65 measurements was used to estimate the amount of error that arises from immunostaining.

2.6.2 Numerical estimation of branch and trunk noise

We observe that at low expression levels, our noise decomposition methodology will often give unreliable estimates. This is likely due to experimental noise overwhelming the true signal when the target protein expression level is low or absent. Therefore, to estimate the percent branch or trunk noise for a given pathway, we calculate the percent branch or trunk noise for all TNF concentrations at which the protein of interest is fully expressed (> 0.1ng/mL) and then average these calculations to arrive at a final estimate for the pathway.

We also note that due to the sensitive nature of covariances and variances to experimental error, the noise decomposition will occasionally yield slightly negative noise for branches that contribute relatively little noise to the total variability. In such cases, we interpret the results to indicate negligible noise, rather than the reduction of the total amount of noise.

Chapter 3. Quantifying information in the TNF pathway

3.1 Introduction[79]
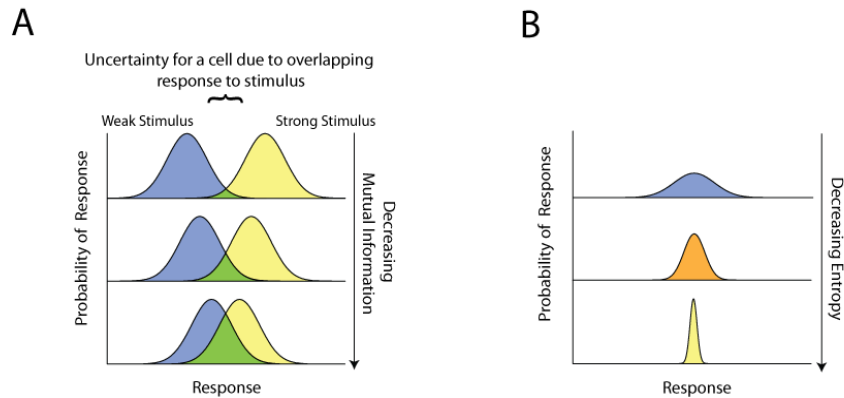
Figure 3.1: (A) Noise can limit the amount of information a cell can obtain about a stimulus. The magnitude of noise is evidenced in the breadth of the probability distribution of the response to a given stimulus. For sufficiently large noise, a cell which can encounter strong or weak stimuli cannot use its response to discern which stimulus was encountered with absolute precision. Consequently, from the cell's perspective, noise leads to a loss of information about the input. The amount of mutual information between the stimulus and cellular response also suffers such that the greater the overlap between distributions, the less mutual information is communicated. **(B) Entropy can be understood as a measure of dispersion.** A wider probability distribution corresponds to an increase in the uncertainty of the cellular response and consequently, entropy.

In their *in vivo* environment, cells are constantly awash in a sea of hormones, cytokines, morphogens, and other receptor ligands released by other cells. Each of these molecular signals can be thought of as being sent with the intent of communicating a specific message or action for the receiving cell to perform. Within the recipient cell, the information contained within the chemical messages must be captured and processed by the cell's biochemical circuitry, which typically involves feedback loops, crosstalk, and delays. These control functions are commonly executed by dedicated sets of kinases and transcription factors to ensure that the appropriate cellular response is activated. Since

the mechanisms behind this complex function are biochemical in nature, molecular noise can greatly hamper the propagation of signals[21,29,30]. As a result, the message can get distorted and cells may not be able to acquire a precise perception of their surroundings. Biological noise can perhaps more adequately described as stochastic cell-cell variability and can be experimentally observed by sampling the distribution of responses by a group of genetically identical cells exposed to the same stimulus. If, for example, the distribution of responses elicited by a weak stimulus overlaps with the distribution elicited by a strong stimulus, a cell whose response value falls within the overlap will not be able to discern with absolute certainty which stimulus was present (Fig. 3.1*A*). This inability to resolve distinct stimuli represents a loss of information about the input. Traditional metrics for noise related to the standard deviation or variance primarily quantify the magnitude of noise and do not directly indicate the degree to which noise hampers the discrimination of different inputs. Likewise, both deterministic and stochastic mathematical models, although able to capture dynamic trends, require *a priori* knowledge or assumptions of the underlying molecular mechanisms and ultimately fail to describe how signaling fidelity is affected by variability. In order to quantify the degree to which noise affects the fidelity of the message, or specifically to determine what a biological signaling system can or cannot communicate accurately, it is useful to turn to information theory.

Originally developed by Claude Shannon for the purpose of data compression and the analysis of man-made communication systems, information theory provides a mathematical framework to quantify the amount of information that can be transmitted through a noisy communication channel. A differentiating strength of this type of

analysis, especially pertinent in cell signaling, is that only input and output measurements are required, thereby obviating the need for a detailed understanding of the signaling system. With information theory, any complex system can be reduced to a black box communications channel and analyzed. When details of the underlying system are available, they can be included as part of the analysis, leading to an even deeper understanding. Some examples of biological systems that have benefitted from such an analysis include neural networks[80] and, more recently, gene regulation networks[81,82], particularly in developmental biology[83], and signal transduction networks[48].

3.2 Applications of information theory in biology

A major advantage of the information theoretic framework described above is that it can be easily implemented in a wide range of scenarios absent of any knowledge of the internal mechanisms or complexity of the system. The key to conducting such an analysis is to identify the boundaries of the communication channel and thus specify its input and output. As such, many applications of information theory to biology have been to characterize the information transmission capacity of specific signaling systems or network structures. One early example can be found in the application of information theory to the neural coding problem in neuroscience[84-86]. To acquire information about the outside world, a sensation is processed by a sensory organ into a stream of electrical impulses, called action potentials, which travel along a highway of neurons to the brain. The brain receives the neural signals and then proceeds to decode the information to recreate the original sensation. However, what is not immediately evident is how

47

neurons encode such vivid depictions of the environment into a simple series of electrical pulses that can be decoded with remarkable fidelity.

One simple way to represent sensory information is to encode it into the rate of neuronal firing, which can be easily measured by counting the number of spikes within a timeframe and then averaging over time. Alternatively, information might be represented in the relative position of the spikes, referred to as a temporal coding scheme. Experimentally measuring such a temporal code involves discretizing a time interval into bins and then assessing if a spike is present in each bin, designating a 1 to represent a full bin or 0 to indicate an empty one to generate a fixed length series of binary digits (Fig. 3.2*A*). This block of binary digits would then represent a code that a neuron would send. The capacity provides a way to evaluate rate coding, temporal coding, or any other hypothesized information coding mechanism based upon the ability to carry information. For example, information theory was used in the early analysis of neural codes to determine that temporal codes offer a greater potential to transmit information than simple rate codes[87].

Similar analyses can be used to evaluate how neural information transfer evolves over time. Since neural networks have the ability to learn, it may be possible for them to adapt to different sources of information. For example, the infomax principle[88], when applied to neurological sensing, posits that the brain can dynamically adjust to different inputs in order to maximize the amount of information provided by a sensory organ. Indeed, it has been shown that neural spikes display neural codes that adapt as the stimulus to the sensory organ changes to ensure that the amount of information transmitted is maximized[89,90].
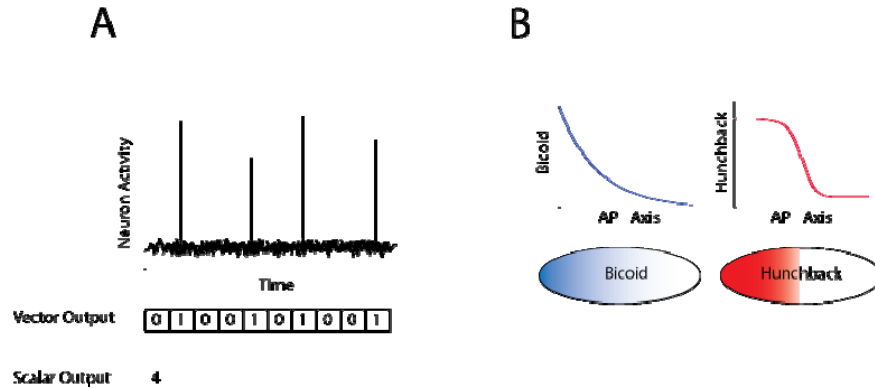
Figure 3.2: Information theory in biological contexts. (A) Quantifying a neural spike train as a scalar or vector. Neural activity consists of intermittent spikes known as action potentials. A series of spikes is known as a neural spike train. data spike train can be quantified as the total number of spikes over a given time period giving a scalar output. Alternatively, time can be divided into small time intervals such that the number of spikes occurring in each time interval is 1 or 0, enabling the spike train to be quantified as a binary vector output. As the total time frame is made longer, the vector becomes longer, and it becomes increasingly harder to adequately sample all possibilities in the entire vector space. **(B) Bicoid and hunchback gradient in the *Drosophila melanogaster* embryo.** In the developing embryo of *Drosophila melanogaster*, pre-deposited bicoid maternal mRNA is translated into a bicoid protein gradient along the anterior-posterior axis. Because bicoid is a cooperative transcriptional activator of hunchback, the smooth bicoid gradient leads to expression of hunchback in a much sharper concentration gradient which delineates the anterior and posterior halves of the embryo.

Another example of the importance of the fidelity of information transfer is found in the development of the embryo of the fruit fly, *Drosphila melanogaster*. In early developmental stages, the embryo consists of an undifferentiated collection of nuclei embedded in the common cytoplasm forming a so-called syncytium. Each nucleus must accurately determine its physical position within the embryo in order to adopt the appropriate developmental fate. To communicate information about position, a morphogen, a biochemical signal with a spatially graded distribution, typically encodes positional information via concentration. Any error in this process can lead to a fruit fly with body parts in the wrong locations or of the wrong sizes, which is often lethal.

Accuracy in the systems that communicate between the morphogen and fate decision processes is of paramount importance, thus we would expect to see sufficient information communicated from the morphogen to the molecular mechanisms involved in cell-decision making. For instance, a morphogen essential in patterning the anterior-posterior (A-P) embryonic axis is the bicoid transcription factor. Bicoid, in turn, induces expression of hunchback protein in a concentration dependent manner. Interestingly, bicoid concentration decreases steadily from the anterior to the posterior end of the embryo, whereas hunchback concentration falls off sharply in the middle of the embryo in a "switch-like" fashion (Fig. 3.2B). This observation has led to the hypothesis that bicoid concentration encodes positional information that is transmitted to hunchback, thereby enabling a cell to determine whether it is located in the anterior or posterior half of the embryo. Until recently however, it was unclear whether this long-standing hypothesis could withstand quantitative scrutiny, as gene expression in individual cells is an inherently noisy process[9], which along with other sources of cell-to-cell or embryo-to-embryo variability could interfere with transmission of the positional information. To examine the capacity of the bicoid-hunchback communication channel in the presence of such noise, Tkacik et al.[83] used data collected by Gregor et al.[91] that simultaneously quantified bicoid and hunchback concentrations throughout many embryos, yielding a sample of their joint distribution. From this data, Tkacik et al. estimated that the mutual information between bicoid and hunchback and found experimentally that there was 1.5 ± 0.2 bits of positional information transmitted. Because 1 bit is the minimum needed to perfectly specify the A-P boundary (a binary outcome), it was concluded that the capacity of the bicoid-hunchback channel was sufficient for each cell to accurately determine

whether they are located in the front or back half of the embryo. These measurements have been recently extended to show that multiple morphogens in combination are sufficient for each cell to uniquely determine its location along the A-P axis[92].

A common theme throughout the prior examples is that information can be thought of as a distinct quantity that cellular systems necessarily require for survival. The examined studies provide support for the premise that cells seek to acquire information sufficient only to ensure continued existence and that any additional capacity can be acquired but presumably at a higher energetic cost to the cell. By placing physical upper limits to the transfer of information in biological systems, information theory can direct a novel line of inquiry in well-established systems. For example, in the fruit fly *D. melanogaster*, we can quantify to what degree each molecular mechanism contributes to create such complex patterns of morphogen gradients that ultimately lead to the differentiation of the adult fruit fly.

Here, in the following sections, we extend information theory concepts to analyzing biochemical signaling networks, whose information transfer capacities were previously generally unknown. We develop a general integrative theoretical and experimental framework to predict and measure the mutual information transduced by one or more signaling pathways. Applying this framework to analyze a 4-dimensional compendium of single cell responses to tumor necrosis factor (Fig. 3.3*A*), an inflammatory cytokine that initiates stochastic signaling at physiologic concentrations spanning ~4 orders of magnitude[51,52,93-97], shows that signaling via a network rather than a single pathway can abate the information lost to noise. Furthermore, we find that an information bottleneck can restrict the maximum information a network can capture.

3.3 Results
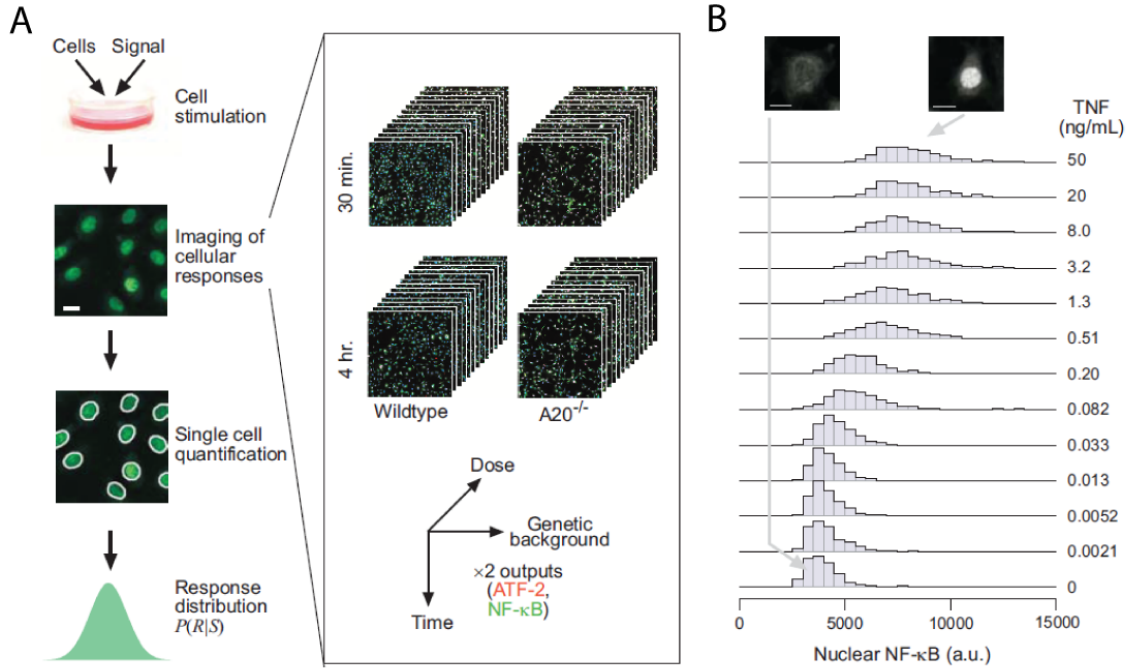
Figure 3.3: Experimental portion of information theoretic analysis of cell signaling fidelity. (**A**) Experimental flowchart for sampling the conditional response distribution at single cell resolution using immunocytochemistry, and resulting 4-dimensional compendium of multiple responses in cells of multiple genetic backgrounds to multiple TNF concentrations, at multiple time points. The data was collected in a single experiment, allowing controlled, quantitative comparisons along each dimension. (**B**) Distributions of noisy NF-κB nuclear translocation responses to 30 min. TNF (examples shown at top) used to compute the channel capacity of the TNF-NF-κB pathway. (Scale bars, 20 μm)

The mutual information, $I(R;S)$, measured in bits, is the binary logarithm of the maximum number of input signal values (S), such as ligand concentrations, a signaling system can perfectly resolve on the basis of its noisy output responses (R)[98]. One bit of information can resolve two different signal values, two bits resolves four values, etc. More generally,

$$I(R;S) = \int_S \int_R P(R,S) \log_2\left(\frac{P(R,S)}{P(R)P(S)}\right) dRdS .$$ (3.3.1)

52

The joint distribution $P(R,S)$ determines the marginal distributions $P(R)$ and $P(S)$ and hence also the mutual information, and can be decomposed as $P(R,S) = P(S) P(R|S)$. The response distribution, $P(R|S)$, is experimentally accessible by sampling responses of individual isogenic cells to various signal levels (Fig. 3.3*B*) and its spread reflects the noise magnitude given any specific input. The signal distribution, $P(S)$, reflects potentially context-specific frequencies at which a cell experiences different signal values. Although the amount of information might thus vary from case to case, one can also determine the maximal amount of transducible information, given the observed noise (see Section 3.6). This quantity, known as the channel capacity[98], is a general characteristic of the signaling system and the signal-response pair of interest, and can thereby be experimentally measured without making assumptions about the (possibly nonlinear) relationship between R and S, signal power, or noise properties.

Using immunocytochemistry, we assayed nuclear concentrations of the transcription factor NF-κB in thousands of individual mouse fibroblasts 30 min. after exposure to various TNF concentrations (Fig. 3.3*A*), choosing this time point because NF-κB translocation peaks at 30 min. regardless of the concentration used, initiating expression of early response inflammatory genes[51,52,97,99]. The NF-κB response value in a single cell could yield at most 0.92 ± 0.01 bits of information which is equivalent to resolving $2^{0.92} = 1.9$, or about 2, concentrations of the TNF signal, thus essentially only reliably indicating whether TNF is present or not (See Sections 3.6.2 and 3.7, regarding the low experimental uncertainty.) A bimodal input signal distribution, $P(S)$, with peaks at low and high TNF concentrations maximizes the information (Fig. 3.4), supporting the

notion of essentially binary (digital) sensing capabilities of this pathway[96], although we did not observe bimodal output responses, $P(R|S)$.
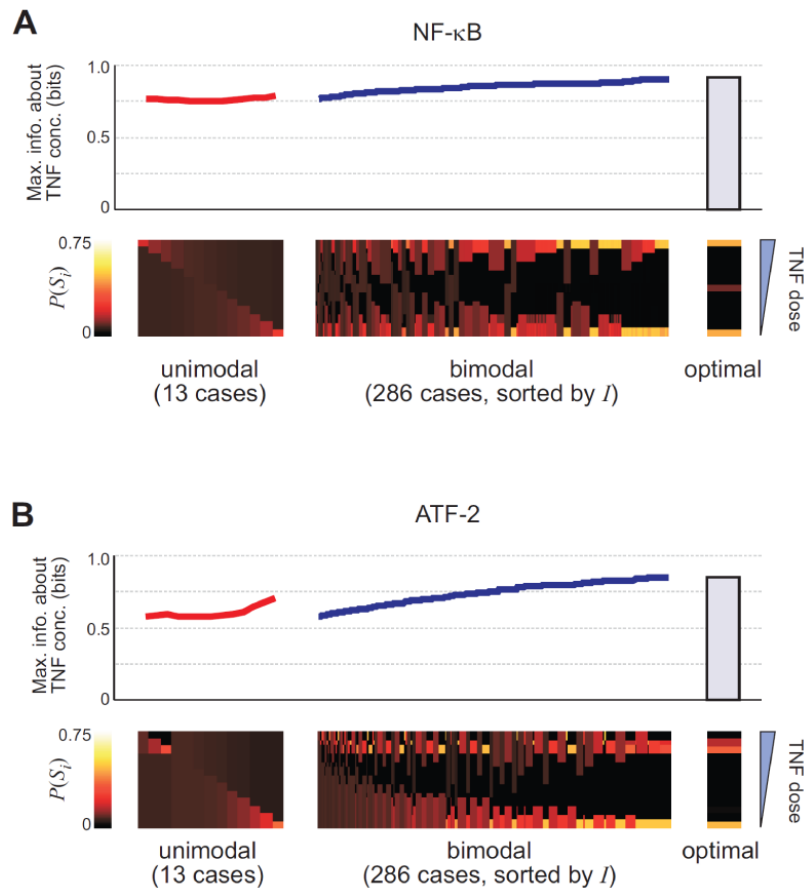
Figure 3.4: Maximum mutual information about TNF concentration. (**A**) The top graph shows the maximum mutual information between TNF concentration and nuclear NF-κB concentration at 30 min. under a unimodal constraint (sorted in order of the 13 possible locations of the mode), bimodal constraint (testing all 286 possible locations of the two modes and the intervening minimum, sorted in increasing order of mutual information), and no constraint (optimal). The bottom heat maps show the signal distributions that yield the maximum mutual information under the various constraints. Each column in the heat map represents a signal distribution (a set of probabilities that sum to 1), each row corresponds to a specific signal value (TNF concentration), and the color indicates the probability associated with that signal value. The optimal value is approached by multiple bimodal distributions in which only very high and very low TNF concentrations are represented. (**B**) Same as panel A except the response analyzed is nuclear phospho-ATF-2.

Noise also limits other canonical pathways, including signaling by platelet derived growth factor (PDGF), epidermal growth factor[100], and G-protein coupled receptors[101] to ~1 bit (Fig. 3.5*A-C*, Table 3.1). Even the most reliable system we examined, morphogen gradient signaling through the receptor Torso in *Drosophila* embryos[102], was limited to 1.61 bits (Fig. 3.5*D*, Table 3.1), corresponding to just ~3 distinguishable signal levels.

Signal	Response	Maximum mutual information (bits)
TNF	NF-κB	0.92 ± 0.01
TNF	ATF-2	0.85 ± 0.02
TNF	NF-κB and ATF-2	1.05 ± 0.02
PDGF	NF-κB	0.67 ± 0.01
PDGF	ATF-2	0.74 ± 0.01
PDGF	NF-κB and ATF-2	0.81 ± 0.02
EGF	Erk (fold-change)	0.60 ± 0.03
UDP	Peak Ca^{2+}	1.22 ± 0.03
UDP	Integrated Ca^{2+}	1.07 ± 0.02
Position	Doubly phosphorylated Erk	1.61 ± 0.05

Table 3.1: Experimentally measured channel capacity of various signaling pathways
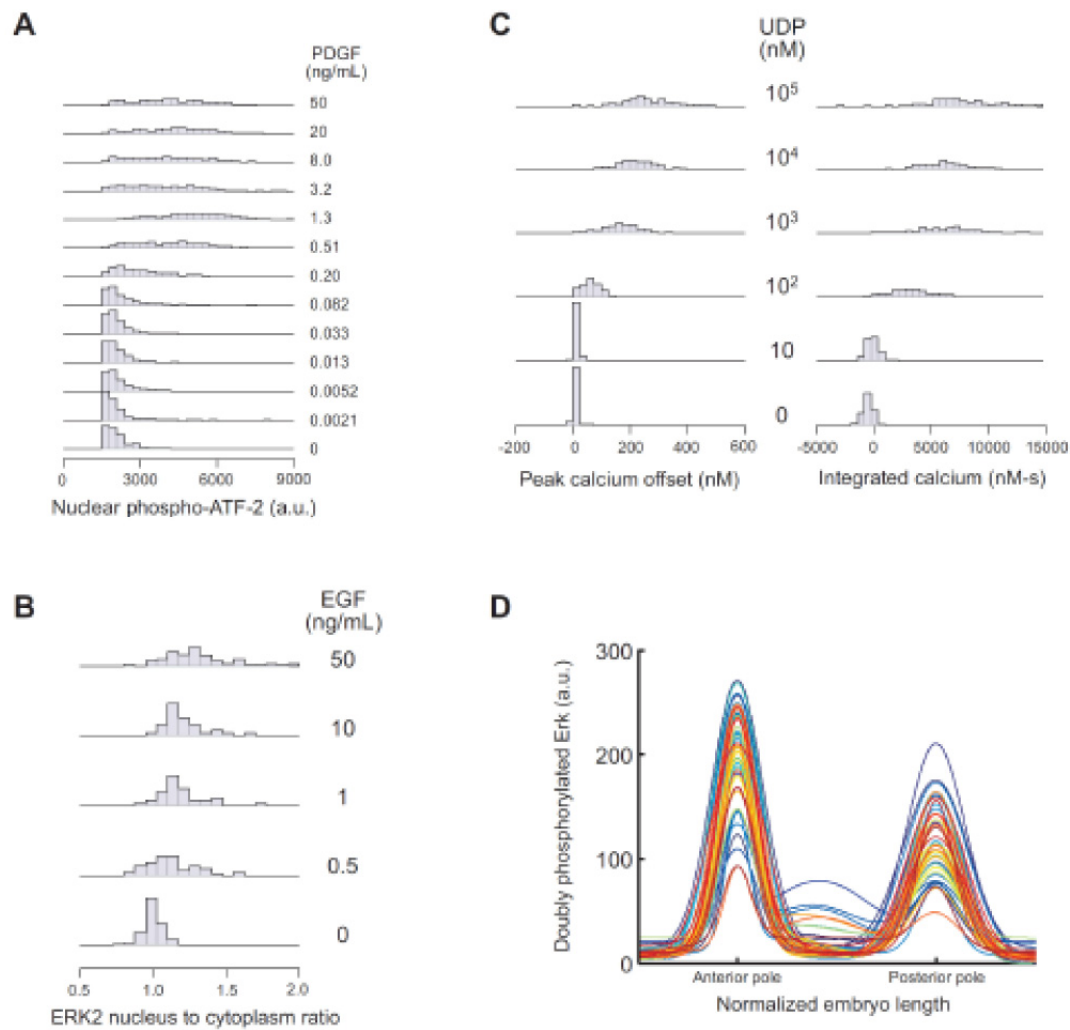
Figure 3.5: Response distributions for various signaling systems. The data shown here were used to compute some of the channel capacity values reported in Table S1. (**A**) nuclear phospho-ATF-2 concentrations in mouse fibroblasts following 30 min. exposure to TNF at the indicated concentrations, as measured by immunofluorescence. (**B**) Fold-change in extracellular signal regulated kinase 2 (ERK2) nucleus to cytoplasm ratio in human lung cancer cells in response to 10 min. epidermal growth factor (EGF) exposure, as measured in single live cells (see Section 3.5.5). (**C**) Peak calcium concentration (left) and time-integrated calcium dynamics (right, integrated over 120 sec) in RAW264.7 macrophages following exposure to uridine diphosphate (UDP), a stimulus for the P2Y family of G protein-coupled receptors. Data was obtained courtesy of M. Simon (California Institute of Technology), see[101]. (**D**) Concentrations of doubly phosphorylated Erk along the perimeter of wildtype *Drosophila melanogaster* embryos between nuclear cycles 10 and 14, as determined by immunofluorescence. Each curve is fitted to an individual embryo and normalized so that peak Erk activities occur at the anterior and posterior poles. Data was obtained courtesy of S. Shvartsman

The pathways examined above are examples of individual biochemical communication channels that capture relatively low amounts of information about signal intensity, which would allow only limited reliable decision making by a cell. However, information in biological systems is typically processed by networks comprising multiple communication channels, each transducing information about the signal. For instance, a transcription factor often regulates many genes, a receptor many transcription factors, and a diffusible ligand many cells. The outputs of such multiple channels together can provide more information about the signal than the output of any one channel (see Section 3.8). Subsequently, downstream signaling processes that converge to co-regulate common effectors, biological processes, or physiologic functions can provide the point needed to integrate the multiple outputs to realize the benefit of increased aggregate information (Fig. 3.6). To provide a unified framework for analyzing such various networks, we first theoretically investigated information gained by network signaling in general, then experimentally tested the predictions made by the theory when applied to a specific system.
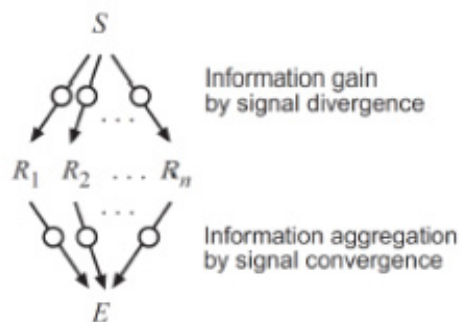
Figure 3.6: Information flow through multiple communication channels that diverge then converge. Signaling through multiple communication channels to the responses R_1, R_2, ..., R_n can increase the amount of information transduced about the input signal, S, as compared to the information transferred by an individual channel. This information can be aggregated through downstream convergence at a common effector, E.

We considered two information theoretic models, similar to models of population coding in neural systems[103-105], for transmitting a signal S through multiple channels to the responses R_1, R_2, …, R_n, under the assumption of Gaussian variables (see Section 3.9). The bush model utilizes independent channels (topologically resembling an upside down shrub) (Fig. 3.7A), whereas the tree model signals through a common channel ("trunk") to the intermediate, C, before diverging into independent branches (Fig. 3.7B). The information resulting from the bush model is

$$I_{\text{bush}}(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1 + n\frac{\sigma_S^2}{\sigma_{S\to R}^2}\right),$$ (3.3.2)

where σ_S^2 is the variance of the signal distribution, and $\sigma_{S\to R}^2$ is the noise (variance) introduced in each branch. Thus, the information can grow logarithmically with the number of branches without an upper bound. In contrast, the information resulting from the tree model is

$$I_{\text{tree}}(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1 + \frac{n\sigma_S^2/\sigma_{C\to R}^2}{1 + n\sigma_{S\to C}^2/\sigma_{C\to R}^2}\right),$$ (3.3.3)

where $\sigma_{S\to C}^2$ and $\sigma_{C\to R}^2$ are the trunk and branch noises, respectively (see Section 3.9). As the number of branches increases, the information asymptotically approaches an upper limit equal to the mutual information between the input signal and the common intermediate, thus the information lost to noise in the trunk determines the maximum throughput of a tree network.

The key difference between bush and tree networks is the absence or presence of this trunk-based information bottleneck. The biochemical structure of a network can resemble a tree, but if the trunk presents little information limitation, the bush model

lacking a bottleneck might best estimate the capacity of the network. Additionally, the

bush and tree models make various semi-quantitative predictions (see Section 3.10), such

as the information captured by a network based on the capacities of its component

pathways. For example, for a bush network comprising two pathways each with 1 bit

responses, Eq. **3.3.2** implies $\sigma_S^2 / \sigma_{S \to R}^2 = 3$ and that together they should yield

$\frac{1}{2} \log_2 (1 + 2(3)) = 1.4$ bits.
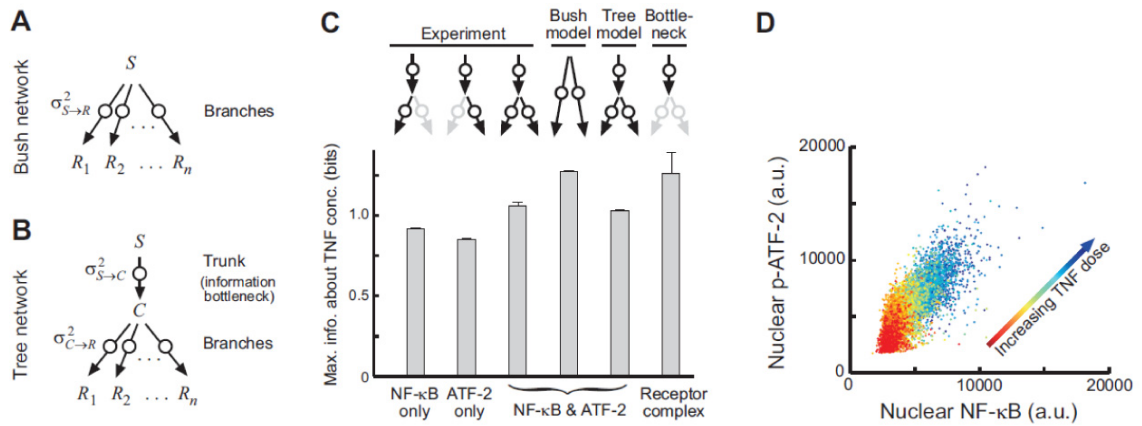
Figure 3.7: Information gained by signaling through a network comprising multiple communication channels. (**A**) Schematic of a bush network with independent channels lacking an information bottleneck. (**B**) Schematic of a tree network with channels sharing a common trunk that forms an information bottleneck. (**C**) Comparison of bush and tree model predictions for the capacity of the TNF network to experimental values. At 30 min., the NF-κB and ATF-2 pathways together capture more information about TNF concentration than either pathway alone (bars 1-3), and the tree rather than bush model accurately predicts this increase (bars 3-5). The tree model further predicts a receptor level bottleneck of 1.26 ± 0.13 bits (bar 6). In all panels, circles represent noise introduced in the indicated portions of the signaling network; see text for definition of symbols. (**D**) Joint distribution of NF-κB and ATF-2 responses to 30 min. stimulation of TNF. Each datapoint represents a single cell, and each concentration of TNF examined is shown using a distinct color.

TNF activates the NF-κB and c-Jun N-terminal kinase (JNK) pathways, stimulating

nuclear localization of NF-κB and phosphorylated activating transcription factor-2 (ATF-

2) (Fig. 3.8*A*), respectively[106]. To determine if the TNF signaling network contains a

significant upstream information bottleneck limiting the information captured by these

pathways, we examined whether the bush (bottleneck absent) or tree (bottleneck present)

network model better approximates the network (Fig. 3.6). The models are applicable

because the NF-κB (Fig. 3.3*B*) and ATF-2 (Fig. 3.8*B*) response distributions are

approximately Gaussian at all TNF concentrations.
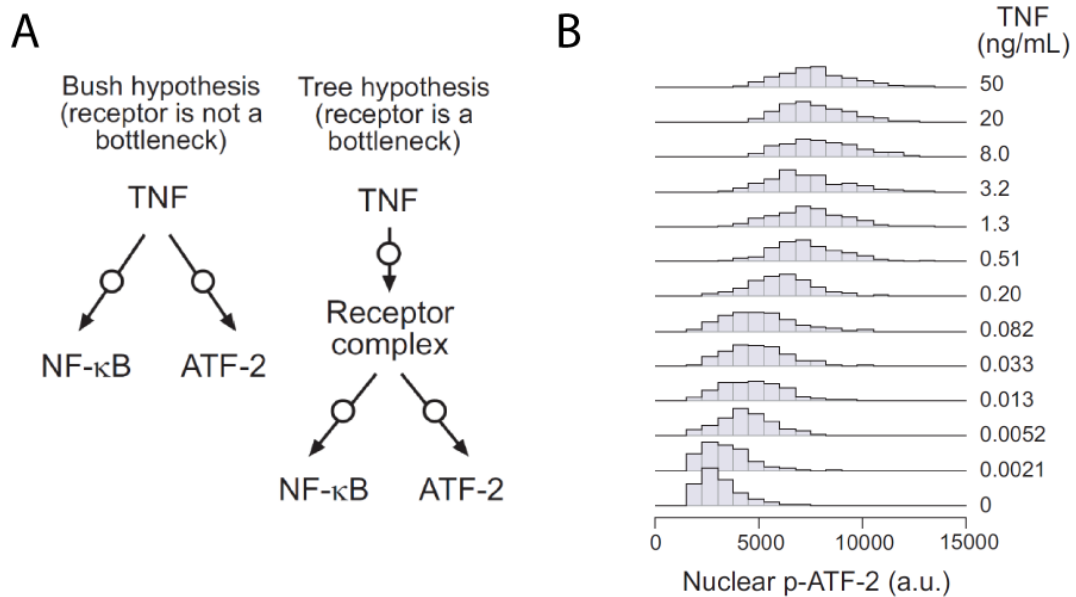
Figure 3.8: Bush and tree representations of the TNF signaling network. (A) Schematics of information flow through the TNF signaling network highlighting the experimentally testable hypotheses of whether the network lacks (bush model, left) or contains (tree model, right) an information bottleneck due to the steps of receptor complex activation common to multiple TNF signaling pathways. **(B)** Distribution of ATF-2 activity in response to TNF. Histograms showing the distribution of nuclear phospho-ATF-2 concentrations in mouse fibroblasts in response to 30 min. TNF exposure at the indicated concentrations.

We found that NF-κB alone yielded at most 0.92 bits of information about TNF

concentration, and ATF-2 alone yielded at most 0.85 ± 0.02 bits (Fig. 3.4*B*, Table 3.1).

Together, the bush model predicts that these pathways jointly yield 1.27 ± 0.01 bits (Fig.

3.7*C*) and a similar model assuming independent pathway responses that are not

necessarily Gaussian likewise predicts an increase to 1.13 ± 0.01 bits. The actual

information determined by dual staining immunocytochemistry (Fig. 3.7*D*) was 1.05 ±

0.02 bits, much lower than both predictions (Fig. 3.7*C*), demonstrating that the bush

model does not approximate the TNF network well. In contrast, the tree model predicts

1.03 ± 0.01 bits, matching the experimental value within error (Fig. 3.7*C*), and also

correctly predicts the statistical dependency between the responses given the signal (Fig.

3.9).
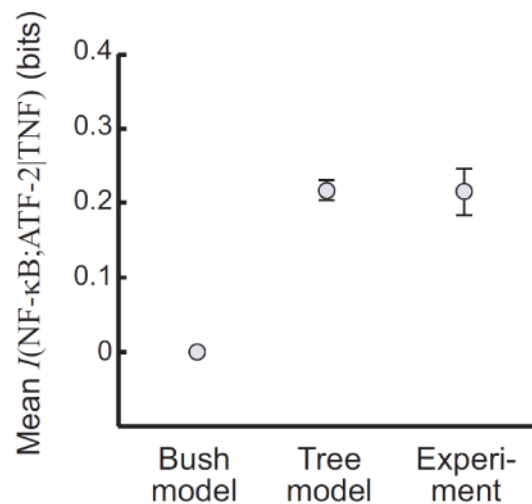
Figure 3.9: Statistical dependence between NF-κB and ATF-2 responses to TNF. Plot shows the
experimentally measured statistical dependence between the NF-κB and ATF-2 responses, as quantified by
the mean value of I(NF-κB; ATF-2 | TNF) (see Sections 3.6.3 and 3.10.3), compared to values predicted by
the bush and tree network models. The bush model predicts conditional independence between the
responses and hence zero mutual information, but the tree model predicts conditional dependence resulting
from the common trunk with mutual information of 0.22 ± 0.01 bits, which corresponds exactly with the
experimentally observed value of 0.22 ± 0.03 bits. Conditional dependence between the responses may also
arise from crosstalk between the pathways, but there is likely insufficient time for substantial crosstalk to
occur following 30 min. TNF exposure.

The correspondence between the tree model predictions and experimental measurements strongly indicates that the network contains an information bottleneck. The tree model predicts the maximum information that can pass through the bottleneck is 1.26 ± 0.13 bits (Fig. 3.7C), corresponding to just $2^{1.26} = 2.3$ distinguishable TNF concentrations. The known biochemistry of TNF signaling implies the bottleneck (trunk) comprises the steps of TNF receptor complex activation common to both pathways, including ligand binding, receptor trimerization, and complex formation and activation. Since all TNF signaling passes through the receptor complex, multiple pathways in the TNF signaling network, activated at the 30 min. time point, only modestly increase the information about TNF concentration regardless of the number of pathways or their fidelity.

We next considered whether networks comprising multiple target genes can capture substantial amounts of information through time integration. If the target gene product lifetime is long compared to its transcription and translation time scales, the accumulated protein concentration is approximately proportional to the time integral of signaling activity, thereby averaging out temporal fluctuations[107,108]. However, the biochemical readout of protein synthesis can introduce extra noise confounding determination of the information contained in the time integral. Fortunately, the maximum information captured by a tree network, in which the time integral of transcription factor activity is the intermediate signal activating multiple independent target genes (Fig. 3.10A, inset), is determined by the trunk (time integration) rather than branch noise (readout mechanism). We measured the information captured by such tree networks in cells stably transfected with different copy numbers of a gene for a stable

green fluorescent protein (GFP)[109] reporting on NF-κB activity (Fig. 3.10*B*). Using the tree model to extrapolate the extent of the bottleneck, under the assumption that ~10 hrs TNF exposure induces similar expression level and noise for each gene, indicates that 1.64 ± 0.36 bits is the maximum information that integrating NF-κB activity over the experimental time period can yield about TNF concentration (Fig. 3.10*A*), regardless of the readout mechanism.

To understand why information was only moderately higher compared to a single time point (1.64 versus 0.92 bits), we monitored GFP expression in individual cells, finding that, for any given cell, GFP accumulated linearly in time in a nearly deterministic fashion, although its onset and accumulation rate varied from cell to cell (Fig. 3.10*C*). This is consistent with observations made using live cell probes[51,52,96] showing NF-κB dynamics to be essentially deterministic over the experimental time scale within each cell, but distinct across cells. We thus conclude that the ability of time integration to increase the information about TNF concentration is limited by the lack of rapid temporal fluctuations that would otherwise be suppressed by integration over the 10 hour response.
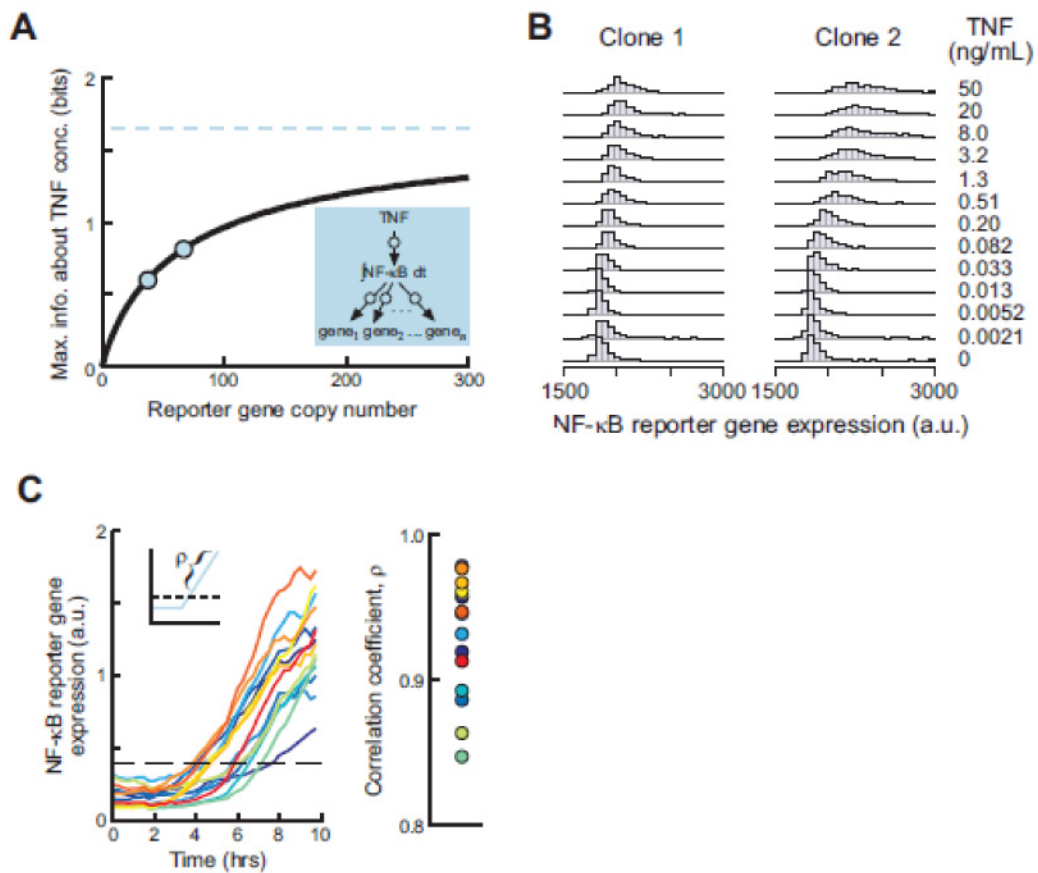
Figure 3.10: Information gained by signaling through a network of multiple genes. Information gained by signaling through networks of multiple genes. (**A**) Plot shows the unique curve (solid black) determined by the tree model (inset), passing through the experimentally determined values (circles), for information as a function of the number of copies of a NF-κB reporter gene. The upper limit, corresponding to the maximum information captured by integrating NF-κB activity over time, is 1.64 ± 0.36 bits (blue dashed line). (**B**) Expression level distributions of clonal cell lines containing different numbers of copies of an NF-κB reporter gene in response to ~10 hrs of TNF. (**C**) Time courses corresponding to individual cells showing cell-to-cell differences in the onset and rate of NF-κB reporter gene expression (left). In each cell, expression is nearly linear and deterministic in time, as quantified by the correlation coefficient (right) of the time course following onset of expression (shown schematically in inset on left).

3.4 Discussion

By treating biochemical signaling systems as information theoretic communication channels, we have rigorously and quantitatively shown that in a single cell noise can substantially restrict the amount of information transduced about input intensity, particularly within individual signaling pathways. The bush and tree network models, which provide a unified theoretical framework for analyzing branched motifs widespread in natural and synthetic signaling networks, further demonstrated that signaling networks can be more effective in information transfer, although bottlenecks can also severely limit the information gained. Receptor level bottlenecks restrict the TNF and also PDGF signaling networks and may be prevalent in other signaling systems[48].

We explored several strategies that a cell might employ to overcome restrictions due to noise. We found that negative feedback can suppress bottleneck noise, which can be offset by concomitantly reduced dynamic range of the response. Time integration can increase the information transferred, to the extent that the response undergoes substantial dynamic fluctuations in a single cell over the physiologically relevant time course. The advantage of collective cell responses can also be substantial, but limited by the number of cells exposed to the same signal or by the information present in the initiating signal itself.

Responses incorporating the signaling history of the cell might also increase the information[110,111]. For instance, responses relative to the basal state (fold-change response) might be less susceptible to noise arising from diverse initial states[100], although

this does not necessarily translate into large amounts of transferred information (Table 3.1). Similarly, for the reporter gene system described here, ~0.5 bits of additional information can be obtained if a cell can determine expression levels at both early and late time points. However, noise in the biochemical networks a cell uses to record earlier output levels and to later compute the final response may nullify the information gain potentially provided by this strategy. Overall, we anticipate that the information theory paradigm can extend to the analysis of noise mitigation strategies and information transfer mechanisms beyond those explored here, in order to determine what specific signaling systems can do reliably despite noise.

3.5 Materials and methods

3.5.1 Cell culture

Wildtype and A20$^{-/-}$ 3T3-immortalized mouse embryonic fibroblasts (kind gift from A. Hoffmann, Univ. of California, San Diego) were maintained in low glucose Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% calf bovine serum (American Type Culture Collection) and 10 U/mL each of penicillin and streptomycin (Invitrogen). Cells were seeded at a density of approximately 150 cells/mm^2 onto 15mm diameter circular coverslips (Fisher Scientific) coated with 0.1% gelatin (Sigma) placed in a 35mm diameter dish, then serum starved in medium with reduced serum concentration (0.1%) overnight before experimentation.

3.5.2 Immunocytochemistry

After exposure to murine TNF (Roche) or murine PDGF-BB (Sigma) at the specified dose and duration, the cells were washed 3 times with ice-cold phosphate buffered saline (PBS, Invitrogen), and fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 20 minutes. The cells were then permeabilized in 0.1% triton X-100 (Sigma) for 5 minutes, and blocked in 10% goat serum (Invitrogen) for 60 minutes. Next, the cells were incubated in primary antibody solution consisting of 1:100 rabbit anti-p65 antibody (Santa Cruz), 1:100 mouse anti-phospho-ATF-2 antibody (Santa Cruz), and 2 μg/mL Hoechst-33258 (Sigma) for 60 minutes. Finally, the cells were incubated in secondary antibody solution consisting of 1:200 Alexa Fluor 488-conjugated goat anti-rabbit and 1:200 Alexa Fluor 594-conjugated goat anti-mouse antibodies (Invitrogen) for 60 minutes. All solutions were made in PBS, and cells were washed with PBS in between each step. To minimize experimentally-induced variability and to enable quantitative comparisons across conditions, all doses of TNF and all cell lines were assayed at the same time using common reagents. Finally, the stained coverslips were mounted on glass microscope slides and imaged on an Axiovert 200M inverted epifluorescence microscope (Zeiss) equipped with Slidebook 4.2 (Intelligent Imaging Innovations). On average, 350 cells were imaged per experimental condition.

In Fig. 3.13, immortalized human umbilical vein endothelial cells (kind gift from the late J. Folkman, Harvard) expressing GFP[112], were stained with 1:100 mouse anti-GFP antibody (Roche) paired with 1:200 Alexa Fluor 594-conjugated goat anti-mouse antibodies.

3.5.3 NF-κB reporter gene

Wildtype 3T3 mouse embryonic fibroblasts were infected with lentiviruses containing a gene for Turbo GFP whose promoter was under the control of NF-κB (Cignal lenti NF-κB reporter, from SA Biosciences). Lentiviral transfection was performed according to the manufacturer's recommendation using a multiplicity of infection of ~200 in the presence of 1 μg/mL polybrene (Sigma), followed by selection in 6 μg/mL puromycin (Sigma). After two rounds of infection, cells were clonally seeded in a 48-well plate and tested for response to TNF. Cells that displayed high levels of GFP fluorescence were individually selected and cultured to create clonal lines of cells. GFP expression was monitored in live cells on a Zeiss Axiovert 200M microscope, or measured in cells that were fixed by exposure to 4% paraformaldehyde for 20 minutes.

3.5.4 Image and data analysis

Image processing, data analysis, and information theoretic calculations were performed using Matlab R2006a (MathWorks). Background correction, nucleus segmentation, and quantification of nuclear concentrations of NF-κB and phospho-ATF-2 were performed as described previously[64].

3.5.5 ERK2 translocation

Nuclear translocation of ERK2 was measured using ERK2-YFP clone C7, which is a H1299 human non-small cell lung cancer cell line clone expressing YFP-tagged ERK2 and mCherry-tagged CBX5 (chromobox 5), a protein with persistent nuclear localization and unconnected to ERK2 signaling[100] (generous gift from Drs. C. Cohen-Saidon and U. Alon, Weizmann Institute). The cell line was maintained as described in[100]. Prior to experimentation, the cells were seeded into a 4-well LabTek optical chamber coated with fibronectin (Sigma) and allowed to attach in serum starved conditions for 5 hours. Within the LabTek well, the cells were maintained in transparent medium consisting of a riboflavin- and phenol red-free formulation of the RPMI medium (Athena Enzyme Systems custom medium) supplemented with 10 U/mL each of penicillin and streptomycin. ERK2 and CBX5 expression was monitored in live cells on a Zeiss Axiovert 200M microscope. Measurements were made for 5 minutes to establish a baseline (zero dose) and for 40 minutes following the addition of EGF (Peprotech) in transparent medium to the well via syringe pump. Information theoretic calculations were performed for individual cell responses at 10 minutes EGF exposure, the time at which ERK2 nuclear translocation peaked.

3.6 Numerical computations of mutual information

3.6.1 Bias correction and error estimate

Mutual information between two variables can be computed from discretized data using the standard formula[98]:

$$
\begin{aligned}
I(R; S) &= H(R) - H(R \mid S) \\
&= \sum_j -P(R = r_j) \log_2 P(R = r_j) \\
&\quad - \sum_i P(S = s_i) \left(\sum_j -P(R = r_j \mid S = s_i) \log_2 P(R = r_j \mid S = s_i) \right)
\end{aligned}
\tag{3.6.1}
$$

where H is the entropy functional, the marginal distribution of the response is given by

$$
P(R = r_j) = \sum_i P(S = s_i) P(R = r_j \mid S = s_i)
\tag{3.6.2}
$$

where the values of R (i.e., r_j) are discretized into N_R bins and the values of S (i.e. S_i) are discretized into N_S bins. In the case that the response R is, for example, a two-dimensional vector then each element of R is discretized into N_R bins for N_R^2 bins in total. The formula for mutual information, written in the form shown in Eq. **3.6.1**, highlights the dependence on $P(R|S)$ which is given by the single cell response data, and $P(S)$ which is chosen or assumed.

In the limit of infinitely small bins but infinitely many datapoints per bin, the discrete mutual information computed using Eq. **3.6.1** converges to the true continuous

value. However, given finite (limited) data, direct estimates of mutual information using Eq. **3.6.1** are biased[113]. Bias likewise contaminates estimates of the maximum mutual information, also known as the channel capacity[98]. Since we are able to obtain large samples, typically consisting of ~300 single cell responses per signal value, we are far away from the severely undersampled regime[110,114], and the bias resulting from finite sample size can be corrected by adapting universal estimators described in[111,115].

In particular, we consider the series expansion of the mutual information in terms of inverse powers of sample size:

$$I_{\text{biased}} = I_{\infty} + \frac{a_1}{N} + \frac{a_2}{N^2} + \cdots \tag{3.6.3}$$

where I_{biased} is the biased estimate of the mutual information, I_{∞} is the unbiased estimate of the mutual information, N is the total number of samples, and the a_i are coefficients that depend on underlying distribution of the signal and the response. The quantity I_{∞}, which we wish to estimate, may be the value of the mutual information under a specific distribution of the signal, or the maximum value under all possible distributions of the signal. When N is sufficiently large, as in our case, terms of second order or larger are negligible in comparison to the first order term $\sim 1/N$, and the estimated mutual information is a linear function of inverse sample size.

We used jackknife sampling to estimate this linear function. In particular, we sampled fractions of the data, ranging from ~60% to 100%, without replacement and

computed the discretized mutual information, I_{biased}. Notably, when computing the discretized mutual information, the boundaries of the bins were chosen so that each bin of the marginal distribution $P(R)$ has approximately equal density (under the assumption that $P(S)$ is uniformly distributed), as in[116]. Then, we plotted the mutual information with respect to inverse sample size, and extrapolated to infinite sample size, i.e. $1/N \to 0$, to obtain I_∞ (Fig. 3.11).
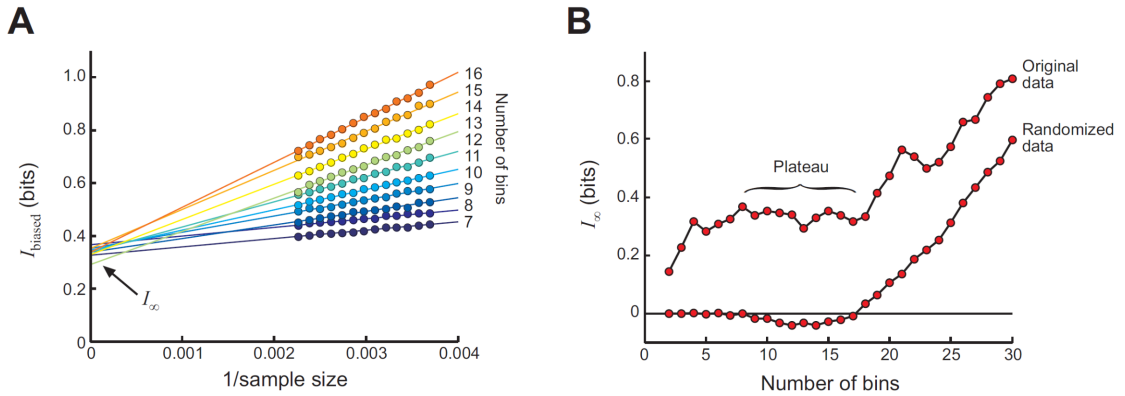
Figure 3.11: Determination of unbiased mutual information. *A*, Linear extrapolation to infinite sample size to determine I_∞ (see Eq. **2.1.3**). *B*, I_∞ plateaus for those numbers of bins for which I_∞ computed for randomized data is slightly negative. The estimate and error for the unbiased mutual information are taken as the mean and standard deviation, respectively, of the I_∞ values within the plateau. The data shown in this figure illustrate the computation $I(\text{NF-}\kappa\text{B};\text{ATF-2}|\text{TNF}=50\text{ng/mL})$.

The extrapolation procedure was performed for different numbers of response bins. When the number of bins is small, I_∞ is an underestimate because differential responses are not distinguished by the coarse discretization. For a moderate number of bins, I_∞ is constant, indicating that the unbiased mutual information is captured. The range of bin numbers for which this occurs is also known as the "plateau" region[115]. For a large number of bins, I_∞ increases because the sample size is not large enough to support very fine discretization, and the linear approximation breaks down. Other

popular approaches for selection of the appropriate coarseness of the data[116,117] are conceptually very similar.

When computing the channel capacity (see Section 3.6.2) for a single response (scalar), e.g. the maximum value of I(NF-κB;TNF), we observed that the plateau region extended to at least 50 bins, a result of the large sample size (~350 cells per TNF dose). The mutual information and its error was estimated as the average and standard deviation, respectively, of the values of I_∞ obtained from 10 to 50 bins, inclusive. When computing the maximum channel capacity for two responses (vector), e.g. the maximum value of I(NF-κB,ATF-2;TNF), the plateau region was typically between 4 and 15 bins (Fig. 3.11). The plateau region was smaller due to the larger ratio between response space and the number of datapoints for two responses which scales as the square of the number of bins, compared to that for a single response which scales linearly in the number of bins. Furthermore, for the channel capacity of either scalar or vector responses, for some bin numbers the value of I_∞ computed on data randomized by shuffling pairings of signals and responses can be negative, though not statistically significantly different than zero[118]. Empirically, we found that these bin numbers reliably indicated the plateau region. The value and error of the mutual information was likewise taken as the average and standard deviation, respectively, of the values of I_∞ computed on the non-randomized data in the plateau (Fig. 3.11).

3.6.2 Computing the channel capacity given *P(R|S)*

In this section, we describe the methods used to determine the channel capacity of the signaling unit, *C(R;S)*, that is, the maximum value of *I(R;S)* under all possible input distributions *P(S)*, given the experimental conditional response data *P(R|S)*. Formally, this can be stated as an optimization problem:

$$C(R;S) = \max_{P(S)} I(R;S) \text{ such that } \sum_i P(S_i) = 1 \qquad (3.6.4)$$
$$P(S_i) \geq 0$$

The constraints ensure that the probability associated with each signal bin is between 0 and 1 inclusive, and the total probability sums to 1. Importantly, since *I(R;S)* is a concave function of *P(S)*, and the constraints are also concave (linear) functions of *P(S)*, there is a single global maximum for *I(R;S)*[98].

The concavity of *I(R;S)* enables easy identification of its maximum value and the corresponding *P(S)* by a variety of algorithms. One fast and simple method to maximize the mutual information is the well-known Blahut-Arimoto algorithm[98], which by iteratively optimizing the mutual information over the marginal and conditional distributions of the input, converges on the input distribution that yields the maximum mutual information. The solution identified by the algorithm was checked using the Karush-Kuhn-Tucker conditions, which for this problem were both necessary and sufficient conditions satisfied by the optimal solution[119]. The Blahut-Arimoto algorithm

can further be run on jackknife samples as described above in Sec. 3.6.1, in order to

obtain unbiased estimates of the maximum mutual information.
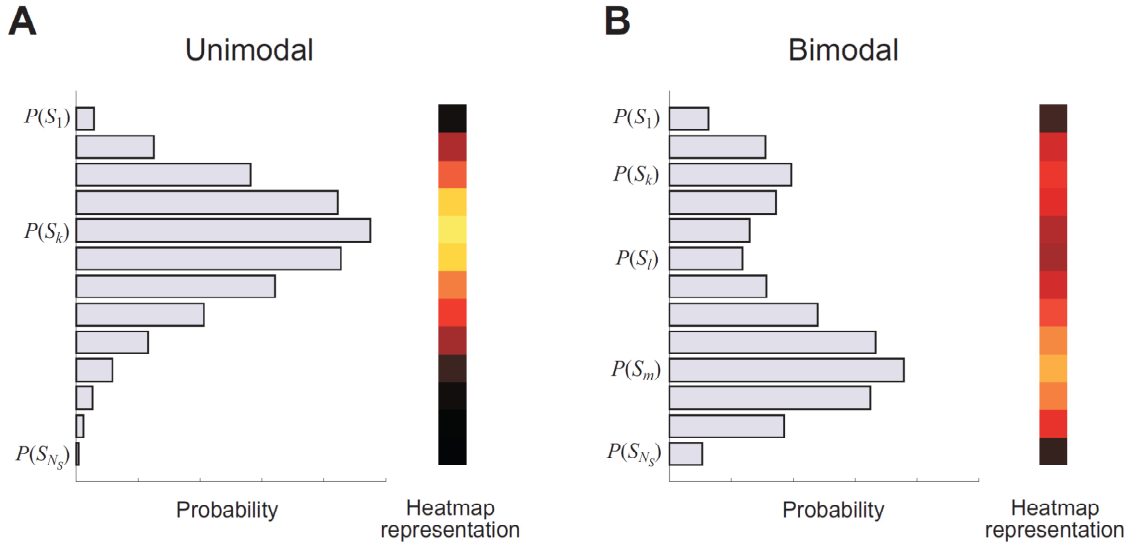
Figure 3.12: Schematic representation of unimodal and bimodal constraints. *A*, Unimodal probability distribution for the signal where the peak occurs at signal value S_k. *B*, Bimodal probability distribution for the signal where the peaks occur at S_k and S_m, with a local minimum at S_l. The corresponding heatmap representations are shown for comparison to Fig. 3.15.

It is well-known that the $P(S)$ that yields the global maximum may be highly

spiky or discontinuous, which may not represent a physically reasonable distribution.

Hence, it is prudent to consider the maximum information that can be achieved when

$P(S)$ is constrained to be "smooth" in some sense. Smoothness constraints are

cumbersome to implement and not guaranteed to yield optimal solutions using a modified

Blahut-Arimoto algorithm[120], but these difficulties can be surmounted using a gradient

ascent method. In particular, in order to enforce additional constraints on $P(S)$, we

utilized a gradient ascent method, specifically Matlab's fmincon function. (Technically,

fmincon minimizes a function, but by using $-I(R;S)$ as the objective function, the

maximum value of $I(R;S)$ is identified.) In the absence of additional constraints, fmincon and the Blahut-Arimoto algorithm yielded identical results.

We point out that the signals that are produced from multiple sources, as in the case of inflammatory signaling, should exhibit a unimodal (normal-like) shape, or they can be bimodal (e.g. inflammation that is either absent or present), with each of the modes having a similar shape for the same reason. This suggests using a definition of "smoothness" that is somewhat different from traditional constraints on derivatives of the distribution (see, e.g.,[121]). Namely, we insist that the distribution $P(S)$ that attains the channel capacity is either unimodal or bimodal.

First, we explored the information capacity that could be obtained if $P(S)$ was constrained to be a unimodal distribution. The corresponding optimization problem was written as:

$$\max_{P(S)} I(R;S) \text{ such that } \sum_i P(S_i) = 1 \tag{3.6.5}$$
$$P(S_i) \geq 0$$
$$P(S_1) \leq P(S_2) \leq \cdots \leq P(S_k)$$
$$P(S_k) \geq P(S_{k+1}) \geq \cdots \geq P(S_{N_S})$$

for $1 \leq k \leq N_S$. The additional constraints ensured that the single peak of the input distribution is at $P(S_k)$ (Fig. 3.12). The maximization was then performed for each of the

76

N_S possible positions of the peak. For the TNF dose response experiments, the value of N_S was 13.

Second, we explored the mutual information that could be obtained if $P(S)$ was constrained to be a bimodal distribution. The corresponding optimization problem was:

$$\max_{P(S)} I(R;S) \text{ such that } \sum_i P(S_i) = 1 \tag{3.6.6}$$

$$P(S_i) \geq 0$$
$$P(S_1) \leq P(S_2) \leq \cdots \leq P(S_k)$$
$$P(S_k) \geq P(S_{k+1}) \geq \cdots \geq P(S_l)$$
$$P(S_l) \leq P(S_{l+1}) \leq \cdots \leq P(S_m)$$
$$P(S_m) \geq P(S_{m+1}) \geq \cdots \geq P(S_{N_S})$$

for $1 \leq k < l < m \leq N_S$. These constraints ensured that the two peaks of the input distributions occur at $P(S_k)$ and $P(S_m)$ and the intervening local minimum occurred at $P(S_l)$ (Fig. 3.12). The maximization was then performed for the $\binom{N_S}{3}$ possibilities for the locations of the two peaks and the local minimum. For the TNF dose response experiments, all $\binom{N_S}{3} = \binom{13}{3} = 286$ possibilities were tested.

For both the unimodal and bimodal constrained optimizations, we note that the added constraints are concave (linear) functions of $P(S_i)$. As a result, the Karush-Kuhn-Tucker conditions again guarantee existence of a unique global optimum and enable it to be verified[119].

To enable fair comparison of the maximum mutual information under no, unimodal, or bimodal constraints (as shown in Fig. 3.5), we performed all calculations using $N_R = 10$ response bins without performing bias corrections. Due to the large sample size, we estimate that the bias is less than 0.017 bits (using the formulas of[122]), and thus does not affect the conclusions drawn. In all other figures and text, the maximum mutual information is reported without unimodal or bimodal constraints and is corrected for bias using the method described above in Sec. 3.6.1.

3.6.3 Computing $I(R_1;R_2|S)$ given $P(R_1,R_2|S)$

In this section, we describe the method used to compute directly from the data the mutual information between two responses resulting from a specific signal value. The corresponding formula is:

$$I(R_1;R_2 \mid S) = \sum_{R_1,R_2} P(R_1,R_2 \mid S) \log_2 \frac{P(R_1,R_2 \mid S)}{P(R_1 \mid S)P(R_2 \mid S)} \tag{3.6.7}$$

Notably, in comparison to the procedures used to maximize mutual information (Sec. 3.6.2), computing $I(R_1;R_2|S)$ can be performed solely with the conditional response data $P(R_1,R_2|S)$ and does not require any assumptions about other distributions. In particular, one does not need to assume the distribution $P(S)$. Nonetheless, bias correction must still be performed to yield reliable estimates of the mutual information.

The bias correction is performed similarly to the method described above (Sec. 3.6.1). The data is binned into N_R bins along the first response R_1 and N_R bins along the second response R_2, with the bin boundaries chosen so that the marginal distributions are equally partitioned into the bins. Jackknife samples are used to extrapolate to the mutual information I_∞ that would be obtained with infinite sample size, as $1/N \to 0$. Then I_∞ is plotted versus the number of bins, N_R, and the plateau region is identified as the bin numbers for which I_∞ computed on randomized data is slightly negative. The unbiased estimate of the mutual information and its error are taken as the average and standard deviation of I_∞ values within the plateau (as in Fig. 3.11).

3.6.4 Computing $I(R_1,R_2;S)$ assuming conditionally independent responses given the signal

The key assumption of the bush network model (see Sec. 3.8.2) is that the responses are conditionally independent given the signal. For the case of two responses, R_1 and R_2, this implies that

$$P_{\text{bush}}(R_1, R_2 \mid S) = P(R_1 \mid S)P(R_2 \mid S) \tag{3.6.8}$$

The joint conditional distribution, constructed in this way from the marginals, can be used to estimate the channel capacity that could be obtained if the responses were the result of signaling via a bush network. The computation is performed by maximizing the

mutual information yielded by $P_{bush}(R_1,R_2|S)$ over all possible $P(S)$ using the algorithms described in Sec. 3.6.2 to yield unbiased estimates of the maximum mutual information.

3.7 Effect of experimental noise on mutual information

In this section, we determine the amount of observed cell-to-cell variability that can be ascribed to true biological variability versus experimental noise, in order to evaluate the degree to which estimates of mutual information are affected by experimental noise. With respect to the experimental noise, we are primarily concerned with the accuracy with which concentrations of cellular species, particularly nuclear NF-κB, can be determined by immunofluorescence. Analogous to the method used to separate total noise into extrinsic and intrinsic noise[123], the total observed variability can be partitioned into true biological variability and immunochemical noise by simultaneous co-measurement of the species of interest.
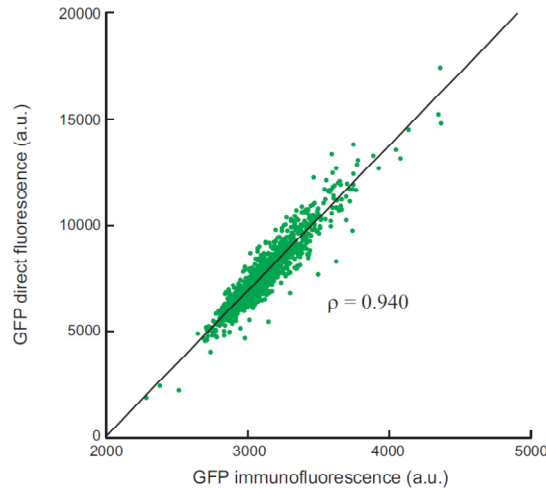
Figure 3.13: Experimental variability associated with immunofluorescence. Cells stably expressing GFP in the nucleus were fixed and immunostained for GFP. In each cell, nuclear GFP concentration was determined by measuring direct fluorescence from GFP and by GFP immunofluorescence. The graph shows the GFP measurements obtained for 1,096 cells. There is a tight linear relationship between direct fluorescence (proportional to GFP concentration) and immunofluorescence, with a correlation coefficient of 0.940.

First, we determined the level of experimental noise that can be generally ascribed

to immunofluorescence. Using cells stably expressing GFP, we measured nuclear

concentrations of GFP by direct measurement of GFP fluorescence and by

immunofluorescence using GFP-specific antibodies. We observed an excellent linear

correspondence between the direct and stained GFP measurements, with a correlation

coefficient of $\rho = 0.940$ (Fig. 3.13). Now, if we take the direct GFP measurement to be

(proportional to) the true GFP concentration, then it is reasonable to define the

experimental noise as the variance of the stained GFP measurement given the true value

determined by direct fluorescence. Likewise, the total variability is given by the variance

of the stained GFP measurement. Then, under Gaussian assumptions (cf. Eq. **3.9.4**), the

portion of the total variability resulting from experimental noise is

$$\frac{\text{var(stained GFP | true GFP)}}{\text{var(stained GFP)}} = \frac{(1-\rho^2)\,\text{var(stained GFP)}}{\text{var(stained GFP)}} = 1-\rho^2 \qquad (3.7.1)$$

Thus, about 12% ($1 - 0.940^2 = 0.116$) of the total observed variance resulted from immunofluorescence noise. In reality, the direct GFP fluorescence is a slightly noisy (due to shot noise, etc.) measurement of the true GFP concentration. This extra source of noise implies that 12% is a slight over-estimate of the actual portion of the total variance that results from immunofluorescence.
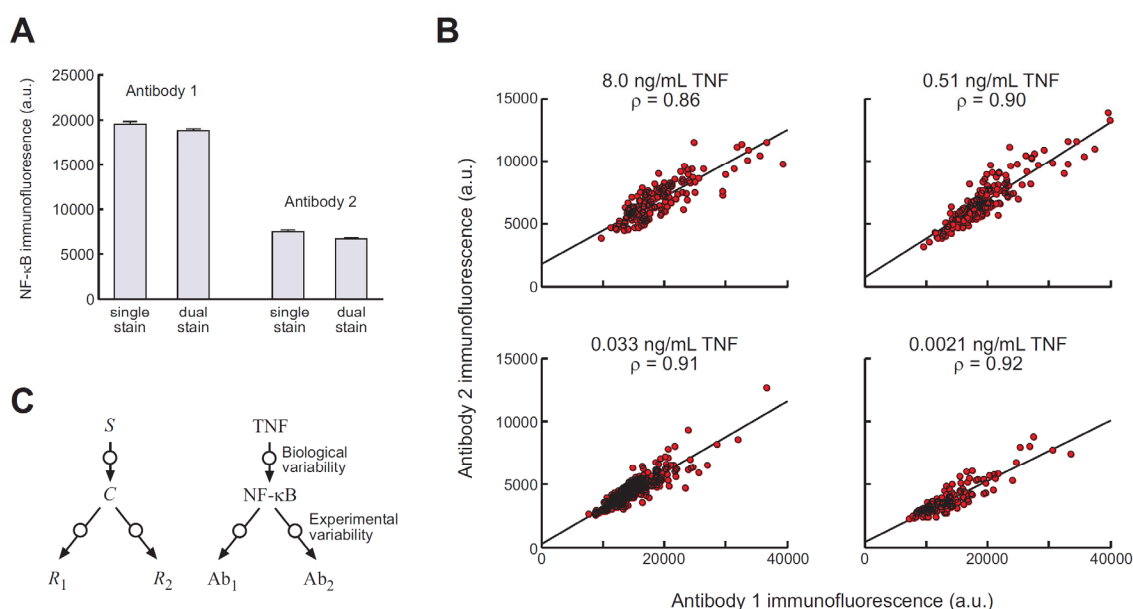
Figure 3.14: Experimental variability associated with NF-κB immunofluorescence. *A*, Wildtype fibroblasts exposed to 8.0 ng/mL TNF for 30 minutes were stained with two different antibodies specific to NF-κB applied individually (single stain) or simultaneously (dual stain). The average NF-κB immunofluorescence was similar for single and dual staining, indicating minimal interference between the two antibodies. *B*, Wildtype fibroblasts were exposed to the indicated doses of TNF for 30 minutes, then dual stained for NF-κB. At all doses, there was a tight linear relationship between the immunofluorescence of the two antibodies with correlation coefficient of ~0.90. *C*, Variability in the dual staining experiment can be analyzed as a tree network. The trunk of the network transduces TNF dose into the true NF-κB concentration, and the branches transduce the true NF-κB concentration into the concentration measured by the antibodies (Ab$_1$, Ab$_2$) by immunofluorescence. The variability associated with the trunk represents the true biological variability, and the variability associated with the branches represents experimental noise.

To confirm this result specifically for immunofluorescence measurements of NF-κB, we performed another experiment in which the p65 subunit of NF-κB was simultaneously stained by two distinct antibodies. The antibodies were chosen to be specific to different termini of p65 to minimize interference with one another. We confirmed that dual staining did not substantially affect the measurements yielded by the individual antibodies (Fig. 3.14). We found that, across a wide range of TNF doses, there was a linear correspondence between the two stained NF-κB measurements with a correlation coefficient of $\rho \approx 0.90$ (Fig. 3.15). Since both stained measurements are affected by experimental noise, neither measurement should be taken to represent the true NF-κB concentration, and Eq. **3.7.1** does not apply. Instead, we note that, under Gaussian assumptions, the correlation between the joint measurements is the product of the correlations between each measurement and the true value:

$$\rho^2_{R_1,R_2|S} = \rho^2_{C,R_1|S}\rho^2_{C,R_2|S} \tag{3.7.2}$$

where R_1 and R_2 are the measured levels of NF-κB and C is the actual level of NF-κB. (This expression can be obtained, for example, by considering a Gaussian tree network in which the trunk represents biological variability and the branches represent experimental noise (Fig. 3.14).) Since, in this experiment, the measurement noises both result from immunofluorescence, we expect that their contributions to the total variability are similar, i.e. $\rho^2_{C,R_1|S} \approx \rho^2_{C,R_2|S}$. Under this assumption, then, the proportion of the observed variance that can be ascribed to measurement noise is

$$\frac{\sigma^2_{C \to R_1}}{\text{var}(R_1 \mid S)} \approx 1 - \rho_{R_1, R_2 \mid S} \qquad\qquad (3.7.3)$$

In our experiment, this shows that ~10% (1 − 0.90) of the total observed variance is due to experimental noise and the rest is true biological variability. This result is consistent with the conservative estimate of 12% obtained from the GFP experiment above.

Next, we estimate the effect of this level of experimental noise on the measured amount of mutual information. We note that the mutual information is determined by the signal-to-noise ratio of the pathway, ϕ, as in Eq. **3.10.1**:

$$I(R; S) = \frac{1}{2} \log_2 \left(1 + m^2 \frac{\sigma^2_S}{\sigma^2_{S \to R}} \right) \quad \Rightarrow \quad \phi \equiv m^2 \frac{\sigma^2_S}{\sigma^2_{S \to R}} = 2^{2I} - 1 \qquad (3.7.4)$$

For the TNF-NF-κB pathway, whose maximum mutual information is I(NF-κB;TNF) = 0.916 bits, the corresponding signal-to-noise ratio is $\phi = 2.56$. The above experiments show that approximately 10% of the denominator of ϕ is due to experimental noise. Thus, continuing the Gaussian assumption, the true value could be as high as 2.56/0.90 = 2.84. Plugging into Eq. **3.7.4**, this implies that the true maximum mutual information may be 0.971 bits. Stated another way, the mutual information between the true p65 concentration and the antibody measurement is not smaller than $\sim \frac{1}{2} \log_2 (1 - 0.9) = 1.66$ bits, which is substantially larger than the measured channel capacity of about 0.92 bits

between the TNF signal and the antibody measurement. Hence the measurement itself is not a bottleneck that substantially decreases the apparent value of the mutual information, whether the TNF-NF-κB relation is Gaussian or not.

Finally, we note that in the TNF-NF-κB pathway, accounting for experimental noise as an additive Gaussian process led to correcting the channel capacity by about 0.055 bits. For other signal-response pairs (e.g. Table 3.1) in which the initial estimate for mutual information is lower than that of the TNF-NF-κB pathway, accounting for experimental noise will lead to a smaller increase due to the monotonic relationship between mutual information and ϕ. Thus, in this study, 0.055 bits is the largest and most conservative value for the extent to which mutual information is underestimated due to experimental noise.

3.8 Information captured by multiple versus individual responses

In this section, we show that the responses of multiple communication channels can obtain more information about a signal than the response of the individual channels. In particular, we explore the values for the mutual information resulting from two responses, $I(R_1,R_2;S)$, can attain relative to the mutual information resulting from the individual responses, $I(R_1;S)$ and $I(R_2;S)$. First, we prove that $I(R_1,R_2;S)$ is at least as large as the greater of $I(R_1;S)$ and $I(R_2;S)$. Then, we prove that if the responses result

from independent signaling processes, then $I(R_1,R_2;S)$ is necessarily larger than $I(R_1;S)$

and $I(R_2;S)$. Finally, we show that $I(R_1,R_2;S)$ has no upper bound and can take on large

values, for example, if the noise in the two responses is negatively correlated. The reader

should consider exploring[124] for discussion of relations among mutual informations in

more general multivariate dependencies models.

3.8.1 The lower bound of $I(R_1,R_2;S)$ is the greater of $I(R_1;S)$ and $I(R_2;S)$

The chain rule for mutual information gives the following relation:

$$I(R_1,R_2;S) = I(R_1;S) + I(R_2;S \mid R_1) \tag{3.8.1}$$

Since mutual information is always non-negative, $I(R_2;S \mid R_1) \geq 0$. Thus

$I(R_1,R_2;S) \geq I(R_1;S)$. By instead applying the chain rule conditioned on R_2, we can

likewise show that $I(R_1,R_2;S) \geq I(R_2;S)$. The combination of these inequalities

demonstrates that a lower bound for the information that two responses provide about a

signal is

$$I(R_1,R_2;S) \geq \max(I(R_1;S), I(R_2;S)) \tag{3.8.2}$$

This lower bound is achieved when either $I(R_2;S|R_1)$ or $I(R_1;S|R_2)$ equals zero, that is

when one response is conditionally independent of the signal given the other response,

implying no improvement in information using the two responses together. In other

86

words, the information provided by the two responses together is not smaller than the information provided by the more informative individual response.

Notably, the proof of this lower bound does not depend on whether R_1 or R_2 are scalars or vectors, a fact that will be utilized in Section 3.8.2.

3.8.2 $I(R_1,R_2;S)$ is strictly greater than the lower bound for conditionally independent responses

In this section, we consider the case in which responses R_1 and R_2 are conditionally independent given the signal, corresponding to the scenario in which the responses are generated by signaling processes that do not interact, other than sharing a common signal. Below, we prove that conditional independence necessarily implies that the mutual information of the responses together is strictly greater than the lower bound, implying a gain of information compared to either response alone. The proof of this statement does not depend on whether R_1 and R_2 are scalars or vectors. Applying the proof to the case in which R_1 and R_2 are scalars implies that the responses of two signaling pathways considered together, one which yields output R_1 and the other which yields output R_2, is more informative than either pathway alone. If instead R_1 is a vector representing a set of outputs from some (arbitrarily complicated) signaling system then the proof implies that adding the conditionally independent response R_2, representing either a scalar output of a separate pathway or a vector output of a separate signaling system, also increases the information about the signal.

87

<u>Theorem:</u> If $I(R_1;S) > 0$, $I(R_2;S) > 0$, and R_1 and R_2 are conditionally independent given S, then $I(R_1,R_2;S) > \max(I(R_1;S), I(R_2;S))$ (strictly greater than the lower bound).

<u>Proof:</u> Suppose without loss of generality that R_1 is the most informative response, i.e. $I(R_1;S) \geq I(R_2;S) > 0$. The chain rule for mutual information allows us to write

$$I(R_1,R_2;S) = I(R_1;S) + I(R_2;S \mid R_1) \tag{3.8.3}$$

To prove that $I(R_1,R_2;S)$ is strictly greater than the lower bound, $I(R_1;S)$, we must prove that $I(R_2;S|R_1) > 0$. This can be proven by contradiction.

Mutual information cannot be negative, so assume that $I(R_2;S|R_1) = 0$. This implies that R_2 and S are conditionally independent given R_1, which implies that for any given values of R_1, R_2, and S, the following holds:

$$
\begin{aligned}
P(R_2,S \mid R_1) &= P(R_2 \mid R_1)P(S \mid R_1) \\
\frac{P(R_1,R_2,S)}{P(R_1)} &= P(R_2 \mid R_1)\frac{P(R_1,S)}{P(R_1)} \\
P(R_1,R_2 \mid S)P(S) &= P(R_2 \mid R_1)P(R_1 \mid S)P(S) \\
P(R_2 \mid S) &= P(R_2 \mid R_1)
\end{aligned}
\tag{3.8.4}
$$

where in the last line we used the conditional independence of R_1 and R_2 given S. Since this holds for all values, we can sum the equation over all possible values of R_1, yielding

$$\sum_{R_1} P(R_1)P(R_2|S) = \sum_{R_1} P(R_1)P(R_2|R_1)$$
$$P(R_2|S)\sum_{R_1} P(R_1) = P(R_2)$$
$$P(R_2|S) = P(R_2)$$

(3.8.5)

Finally, this implies that

$$P(R_2,S) = P(R_2|S)P(S) = P(R_2)P(S)$$

(3.8.6)

This shows that a necessary condition for $I(R_1,R_2;S)$ to equal the lower bound is that R_2 and S are unconditionally independent. However, this would imply that R_2 is not informative about S, contradicting the assumption that $I(R_2;S) > 0$. Therefore, the conditions of the claim imply that $I(R_1,R_2;S)$ is strictly greater than $I(R_1;S)$ and also strictly greater than $I(R_2;S)$.

3.8.3 The upper bound of $I(R_1,R_2;S)$ is infinity

In this section we show that $I(R_1,R_2;S)$ has an infinite upper bound, by considering a simple example. Consider the case in which the responses are scalars given by the equations

$$R_1 = S + \gamma_1$$
$$R_2 = S + \gamma_2$$

(3.8.7)

where γ_1 and γ_2 are noise terms independent of S. If the noise terms have non-zero variance, then the information provided by each individual response, $I(R_1;S)$ and $I(R_2;S)$, is finite.

Now, suppose further that the noise terms are correlated. In the extreme, suppose that they are exactly negatively correlated such that $\gamma_1 = -\gamma_2$. Biologically, this situation might be approached if the there is strong mutually repressive crosstalk between the two pathway branches, or when both branches are competing for the same signaling molecule to activate them. Then, given knowledge of R_1 and R_2, their average yields:

$$\tfrac{1}{2}(R_1 + R_2) = S + \tfrac{1}{2}(\gamma_1 + \gamma_2)$$
$$= S$$

(3.8.8)

Hence, knowledge of R_1 and R_2 allows the noiseless recovery of the exact value of S. If S is a continuous variable, which requires an infinite number of bits to specify exactly, then $I(R_1, R_2; S) = \infty$.

More rigorously, using the methods of Sec. 3.9, one can show that the mutual information of the system described by Eq. **3.8.7** is

$$I(R_1, R_2; S) = \frac{1}{2} \log_2 \left(1 + 2 \frac{\sigma_S^2}{\sigma_{S \to R}^2} \frac{1}{1 + \rho} \right)$$ if S is a normally distributed stochastic variable

with variance σ_S^2, and γ_1 and γ_2 are normally distributed each with variance σ_S^2 and correlation ρ. (See also[125].) As $\rho \to -1$, it is easy to see that $I(R_1, R_2; S) \to \infty$. From this example, we conclude that $I(R_1, R_2; S)$ is unbounded from above.

3.9 Information theoretic analysis of bush and tree networks

3.9.1 Preliminaries

In this section, we consider signaling networks that take a single signal S and broadcast the signal out to n communication channels yielding the responses R_1, R_2, …, R_n. We are interested in the amount of information that the responses jointly yield about the signal, i.e. $I(R_1,…,R_n;S)$. To gain semiquantitative insight into such pathways, we assume that $(R_1, R_2, …, R_n, S)$ is a multivariate normal distribution of dimension $n + 1$, as detailed in the sections below. The Gaussian assumption enables the mutual information

to be solved analytically. The resulting formulas allow us to understand the relative influences of the various sources of noise on the information gathering ability of the signaling network and to predict the value of the mutual information. In order to provide a self-contained description of the theoretical framework that is accessible to both specialists and non-specialists alike, we here provide a complete and detailed derivation of the formulas. However, we caution the reader that the formulas will not hold, in general, for non-Gaussian distributions of the variables.

First, we establish some mathematical formulas which will be used in the derivation of the mutual information for specific network structures. First, a well-known result in information theory is that a multivariate normal distribution of dimension n has an entropy of

$$H = \frac{1}{2} \log_2 \left(\left(2\pi e \right)^n |\Sigma| \right) \tag{3.9.1}$$

where $|\Sigma|$ is the determinant of the covariance matrix of the distribution[98]. Since the marginal and conditional distributions of a multivariate normal distribution are themselves normal, it is easy to see that

$$I(R_1,...,R_n;S) = H(R_1,...,R_n) - H(R_1,...,R_n \mid S)$$

$$= \frac{1}{2}\log_2\left((2\pi e)^n \left|\Sigma_{\vec{R}}\right|\right) - \frac{1}{2}\log_2\left((2\pi e)^n \left|\Sigma_{\vec{R}\mid S}\right|\right)$$

$$= \frac{1}{2}\log_2\left(\frac{\left|\Sigma_{\vec{R}}\right|}{\left|\Sigma_{\vec{R}\mid S}\right|}\right) \tag{3.9.2}$$

where $\left|\Sigma_{\vec{R}}\right|$ and $\left|\Sigma_{\vec{R}\mid S}\right|$ are the determinants of the covariance matrix of the responses and

the responses given the signal, respectively.

If we consider just one response, R, then the determinants are given by

$$\left|\Sigma_R\right| = \mathrm{var}(R), \tag{3.9.3}$$

$$\left|\Sigma_{R\mid S}\right| = \mathrm{var}(R \mid S) = (1 - \rho^2)\,\mathrm{var}(R), \tag{3.9.4}$$

yielding

$$I(R;S) = -\frac{1}{2}\log_2(1 - \rho^2) \tag{3.9.5}$$

As expected intuitively, when there is zero correlation between R and S, their mutual

information is zero. In comparison, the information increases as the correlation

approaches +1 or −1. If the correlation is perfect (exactly +1 or −1), the information is

infinite. Note that this deterministic relation between the information and the correlation

is a direct consequence of Gaussian assumption about the involved variables. In general,

mutual information among two variables is not smaller than the value calculated using the Gaussian assumption.

Finally, we establish the following lemma, which enables us to compute the determinants for multiple responses resulting from either bush or tree signaling networks:

Lemma: The determinant of the $n \times n$ matrix $\mathbf{Q}$ whose entries are given by $q_{ij} = m_i m_j a + \delta_{ij} b_i$ is $\left(\prod_i b_i \right) \left(1 + \sum_i m_i^2 \frac{a}{b_i} \right)$. Here, the Kronecker delta notation ($\delta_{ij} = 1$ if $i = j$, and is zero otherwise) indicates that b_i terms only appear in the diagonal elements of $\mathbf{Q}$.

Proof: A basic property of the matrix determinant is that it is invariant to elementary row addition and subtraction (also known as Gaussian elimination). That is, adding or subtracting a multiple of one row to/from another row does not change the determinant. Therefore, the determinant does not change if we subtract $\frac{m_i}{m_{i-1}}$ times row $i - 1$ from row i, for each of the rows $i = n, n - 1, \ldots, 2$. These operations yield:

$$|\mathbf{Q}| = \begin{vmatrix} m_1^2 a + b_1 & m_1 m_2 a & m_1 m_3 a & m_1 m_4 a & \cdots \\ m_2 m_1 a & m_2^2 a + b_2 & m_2 m_3 a & m_2 m_4 a & \cdots \\ m_3 m_1 a & m_3 m_2 a & m_3^2 a + b_3 & m_3 m_4 a & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}$$

$$= \begin{vmatrix} m_1^2 a + b_1 & m_1 m_2 a & m_1 m_3 a & m_1 m_4 a & \cdots \\ -\frac{m_2}{m_1} b_1 & b_2 & 0 & 0 & \cdots \\ 0 & -\frac{m_3}{m_2} b_2 & b_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}$$

(3.9.6)

Next, we reduce row 2 by adding $\dfrac{\frac{m_2}{m_1} b_1}{m_1^2 a + b_1}$ times row 1 to row 2, yielding:

$$= \begin{vmatrix} d(1) & m_1 m_2 a & m_1 m_3 a & m_1 m_4 a & \cdots \\ 0 & \frac{d(2)}{d(1)} & \frac{m_2 m_3 a b_1}{d(1)} & \frac{m_2 m_4 a b_1}{d(1)} & \cdots \\ 0 & -\frac{m_3}{m_2} b_2 & b_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}$$

(3.9.7)

where we have introduced the notation $d(n) = \left(\prod_{i=1}^{n} b_i \right) \left(1 + \sum_{i=1}^{n} m_i^2 \frac{a}{b_i} \right)$.

Now, we claim that after reduction of subsequent rows, that the diagonal

element of row $k \geq 2$ equals $\dfrac{d(k)}{d(k-1)}$, and the elements of row k to the

right in columns $h = k + 1, \ldots, n$ equal $\dfrac{m_k m_h a \prod_{i=1}^{k-1} b_i}{d(k-1)}$.

The claim can be proven by induction. Clearly the claim holds for row 2.

Assume the claim holds for row k. Then, the reduction of row $k + 1$ is

performed by multiplying row k by $\dfrac{\frac{m_{k+1}}{m_k}b_k}{d(k)/d(k-1)}$ and adding it to row k

+ 1. For the diagonal element of row $k + 1$, this yields

$$\underbrace{\frac{m_k m_{k+1}a\prod\limits_{i=1}^{k-1}b_i}{d(k-1)}}_{\substack{\text{the element above the}\\\text{diagonal of row }k+1}}\frac{\frac{m_{k+1}}{m_k}b_k}{d(k)/d(k-1)}+b_{k+1}=\frac{m_{k+1}^2 a\prod\limits_{i=1}^{k}b_i+b_{k+1}d(k)}{d(k)}$$

$$=\frac{\left(\prod\limits_{i=1}^{k+1}b_i\right)\left(m_{k+1}^2\dfrac{a}{b_{k+1}}+1+\sum\limits_{i=1}^{k}m_i^2\dfrac{a}{b_i}\right)}{d(k)}$$

$$=\frac{d(k+1)}{d(k)} \tag{3.9.8}$$

and for the element in column $h > k + 1$ to the right of the diagonal, the

row reduction yields

$$\frac{m_k m_h a\prod\limits_{i=1}^{k-1}b_i}{d(k-1)}\frac{\frac{m_{k+1}}{m_k}b_k}{d(k)/d(k-1)}=\frac{m_{k+1}m_h a\prod\limits_{i=1}^{k}b_i}{d(k)} \tag{3.9.9}$$

This proves the claim also holds for row $k + 1$, completing the induction.

Since the determinant of the fully row reduced (upper triangular) matrix is the product of its diagonal elements, the desired determinant telescopes to

$$|\mathbf{Q}| = d(1)\frac{d(2)}{d(1)}\frac{d(3)}{d(2)}\cdots\frac{d(n)}{d(n-1)} = d(n) = \left(\prod_{i=1}^{n} b_i\right)\left(1 + \sum_{i=1}^{n} m_i^2 \frac{a}{b_i}\right) \qquad (3.9.10)$$

as claimed.

3.9.2 Information captured by a Gaussian bush network

In this section, we derive formulas for the mutual information, under Gaussian conditions, between a signal and multiple linear responses activated by a "bush" network. The key feature of a bush network is that the network branches into multiple signaling pathways at the level of the signal, so that each response is conditionally independent given the signal.
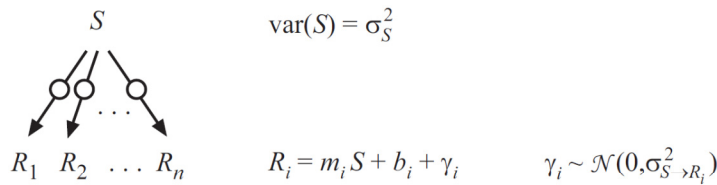
Figure 3.15: Model of a bush signaling network. Each pathway branch transduces the signal into a linear response R_i, with gain m_i, bias b_i, and noise γ_i.

The formal formulation of this model is as follows (Fig. 3.15). The signal S is a normally distributed stochastic variable with variance σ_S^2. Each pathway $i = 1, 2, \ldots, n$

97

yields a linear response $R_i = m_i S + b_i + \gamma_i$ where m_i and b_i are the slope (gain) and

intercept (bias) respectively between R_i and S in the absence of cellular variability, and γ_i

is a stochastic variable representing cellular variability in the response R_i. We assume

that γ_i is normally distributed from cell-to-cell with variance $\sigma^2_{S \to R_i}$ and that the γ_i terms

are independent of each other. As a result, each of the R_i is normally distributed because

each is the sum of two normally distributed variables, and the R_i are conditionally

independent given the signal S. Note that in this general formulation that each pathway

can have different values for the slope, intercept, and magnitude of noise.

Since the variance of independent variables add, the variance of each response is

$\text{var}(R_i) = m_i^2 \sigma_S^2 + \sigma^2_{S \to R_i}$. Similarly, the covariance between any two responses is

$\text{cov}(R_i, R_j) = m_i m_j \sigma_S^2$ (for all $i \neq j$). Thus, using the lemma in Sec. 3.9, the determinant

of the response covariance matrix is:

$$\left(\Sigma_{\vec{R}} \right)_{ij} = m_i m_j \sigma_S^2 + \sigma^2_{S \to R_i} \delta_{ij} \Rightarrow \left| \Sigma_{\vec{R}} \right| = \left(\prod_{i=1}^{n} \sigma^2_{S \to R_i} \right) \left(1 + \sum_{i=1}^{n} m_i^2 \frac{\sigma_S^2}{\sigma^2_{S \to R_i}} \right) \qquad (3.9.11)$$

When S is given, the variance and covariance terms reduce to $\text{var}(R_i \mid S) = \sigma^2_{S \to R_i}$ and

$\text{cov}(R_i, R_j \mid S) = 0$. Then, again using the lemma, the corresponding determinant

evaluates to

$$\left(\Sigma_{\bar{R}|S}\right)_{ij} = \sigma^2_{S \to R_i} \delta_{ij} \Rightarrow \left|\Sigma_{\bar{R}|S}\right| = \left(\prod_{i=1}^{n} \sigma^2_{S \to R_i}\right)$$

(3.9.12)

Finally, using Eq. **3.9.2**, the mutual information between the responses together and the signal is

$$I(R_1, ..., R_n; S) = \frac{1}{2}\log_2\left(1 + \sum_{i=1}^{n} m_i^2 \frac{\sigma_S^2}{\sigma^2_{S \to R_i}}\right)$$

(3.9.13)

The ratio $\sigma_S^2 / \sigma^2_{S \to R_i}$ can be considered to be a signal-to-noise ratio [98], where σ_S^2 represents the signal power and $\sigma^2_{S \to R_i}$ is the noise (variability) in transmitting from S to R_i. The slope m_i can be considered to be a factor that normalizes $\sigma^2_{S \to R_i}$, or more specifically, allows the individual $\sigma^2_{S \to R_i}$ to be compared in similar units. Thus, the mutual information of the n responses together can be obtained by summing the signal-to-noise ratios of the n pathways, when those ratios are given in comparable units. The formula also enables determination of which pathways dominate the mutual information obtained by integrating multiple responses together.

When the n pathways are equivalent the formula simplifies to Eq. **3.3.2**. In particular, if all the $m_i = 1$ and the magnitude of the pathway variability is the same $\sigma^2_{S \to R} = \sigma^2_{S \to R_i}$ for each pathway $i = 1, ..., n$, then the mutual information is:

$$I(R_1,...,R_n; S) = \frac{1}{2}\log_2\left(1 + n\frac{\sigma_S^2}{\sigma_{S\to R}^2}\right) \qquad\qquad (3.9.14)$$

As expected intuitively, the formula reveals that the information increases as the noise

introduced by each branch ($\sigma_{S\to R}^2$) decreases with respect to the spread in the input (σ_S^2).

Furthermore, the information grows logarithmically with the number of responses

measuring the signal, in an unbounded fashion.

3.9.3 Information captured by a Gaussian tree network

In this section, we derive formulas for the mutual information, under Gaussian

conditions, between a signal and multiple linear responses activated by a "tree" network.

The key feature of a tree network is that the signal activates a common "trunk" before

branching into the individual pathways. The trunk terminates at the point of branching

denoted as C, i.e. the last common intermediate shared by the pathways. Thus, the

responses are conditionally independent given C, but not conditionally independent given

the signal. In comparison, responses of bush network are conditionally independent

given the signal.

$$S \qquad\qquad \text{var}(S) = \sigma_S^2$$

$$C \qquad\qquad C = m_C S + b_C + \gamma_C \qquad \gamma_C \sim \mathcal{N}(0, \sigma_{S \to C}^2)$$

$$R_1 \; R_2 \; \ldots \; R_n \qquad R_i = m_i C + b_i + \gamma_i \qquad \gamma_i \sim \mathcal{N}(0, \sigma_{S \to R_i}^2)$$
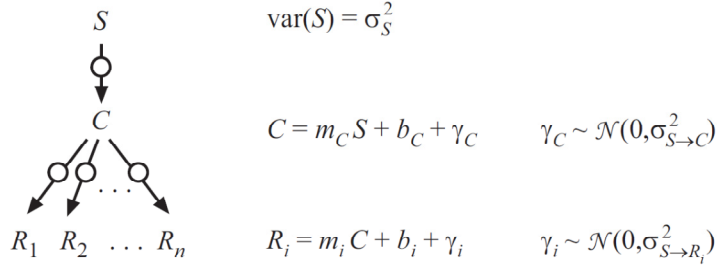
Figure 3.16: Model of a tree signaling network. The common trunk of the network transduces the signal into the intermediate linear response C with gain m_C, bias b_C, and noise γ_C. Each downstream pathway branch then transduces C into the linear response R_i with gain m_i, bias b_i, and noise γ_i.

The formal formulation of the tree network model (Fig. 3.16) is similar to that of bush network model. The signal S is a normally distributed stochastic variable with variance σ_S^2. C is the last common intermediate in the pathways measuring the signal, with $C = m_C S + b_C + \gamma_C$, where m_C and b_C are the slope (gain) and intercept (bias) respectively between S and C in the absence of cellular variability, and γ_C is a stochastic variable representing cellular variability in the common trunk. In particular, we assume that γ_C is normally distributed from cell-to-cell with variance $\sigma_{S \to C}^2$.

Each downstream pathway yields a response $R_i = m_i C + b_i + \gamma_i$ where, similarly, m_i and b_i are the slope (gain) and intercept (bias) respectively between R_i and C in the absence of cellular variability, and γ_i is a stochastic variable representing cellular variability in the branch from C to R_i. We assume that γ_i is normally distributed from cell-to-cell with variance $\sigma_{S \to R}^2$. All of the noise terms γ_i and γ_C are independent of each other and independent of S.

Substituting the definition for R_i into the definition of C reveals that R_i is normally distributed, and on average a linear function of S with slope (gain) $m_C m_i$ and intercept (bias) $m_i b_C + b_i$:

$$R_i = m_C m_i S + (m_i b_C + b_i) + m_i \gamma_C + \gamma_i \tag{3.9.15}$$

From this formula, it is easy to see that the variance of each response is

$\mathrm{var}(R_i) = m_C^2 m_i^2 \sigma_S^2 + m_i^2 \sigma_{S \to C}^2 + \sigma_{C \to R_i}^2$ and that the covariance between any two responses

is $\mathrm{cov}(R_i, R_j) = m_C^2 m_i m_j \sigma_S^2 + m_i m_j \sigma_{S \to C}^2$ (for all $i \neq j$). Thus, using the lemma in Sec.

3.9.1, the determinant of the response covariance matrix is:

$$\left(\Sigma_{\bar{R}} \right)_{ij} = m_i m_j (m_C^2 \sigma_S^2 + \sigma_{S \to C}^2) + \sigma_{C \to R_i}^2 \delta_{ij} \Rightarrow \left| \Sigma_{\bar{R}} \right| = \left(\prod_{i=1}^{n} \sigma_{C \to R_i}^2 \right) \left(1 + \sum_{i=1}^{n} m_i^2 \frac{m_C^2 \sigma_S^2 + \sigma_{S \to C}^2}{\sigma_{C \to R_i}^2} \right)$$

$$(3.9.16)$$

When S is given, the variance and covariance terms reduce to

$\mathrm{var}(R_i \mid S) = m_i^2 \sigma_{S \to C}^2 + \sigma_{C \to R_i}^2$ and $\mathrm{cov}(R_i, R_j \mid S) = m_i m_j \sigma_{S \to C}^2$ (for all $i \neq j$). Then, again

using the lemma, the determinant is:

$$\left(\Sigma_{\bar{R}|S} \right)_{ij} = m_i m_j \sigma_{S \to C}^2 + \sigma_{C \to R_i}^2 \delta_{ij} \Rightarrow \left| \Sigma_{\bar{R}|S} \right| = \left(\prod_{i=1}^{n} \sigma_{C \to R_i}^2 \right) \left(1 + \sum_{i=1}^{n} m_i^2 \frac{\sigma_{S \to C}^2}{\sigma_{C \to R_i}^2} \right) \tag{3.9.17}$$

102

Finally, using Eq. **3.9.2**, the mutual information between the responses together and the signal is

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2 \left(\frac{1 + \sum_{i=1}^{n} m_i^2 \dfrac{m_C^2 \sigma_S^2 + \sigma_{S \to C}^2}{\sigma_{C \to R_i}^2}}{1 + \sum_{i=1}^{n} m_i^2 \dfrac{\sigma_{S \to C}^2}{\sigma_{C \to R_i}^2}} \right)$$

$$= \frac{1}{2}\log_2 \left(1 + \frac{\sum_{i=1}^{n} m_i^2 \dfrac{m_C^2 \sigma_S^2}{\sigma_{C \to R_i}^2}}{1 + \sum_{i=1}^{n} m_i^2 \dfrac{\sigma_{S \to C}^2}{\sigma_{C \to R_i}^2}} \right) \tag{3.9.18}$$

Similar to the bush network, the information obtained from a tree network depends on signal-to-noise ratios. The information depends on two key ratios: (1) $\sigma_S^2 / \sigma_{C \to R_i}^2$, the signal power versus the noise in the downstream branches, and (2) $\sigma_{S \to C}^2 / \sigma_{C \to R_i}^2$, the noise in the trunk versus the noise in the downstream branches. The slope m_i can again be considered to be a factor that normalizes the noise in the downstream branch, $\sigma_{C \to R_i}^2$, enabling the noises to be compared in equivalent units. Likewise, the slope m_C normalizes the signal power σ_S^2. Again, the formula enables determination of which sources of variability dominate the mutual information obtained by integrating multiple responses together.

Notably, the tree network contains a bush network embedded within, i.e. the network consisting of C and the downstream branches. The results for bush-type

networks show that as the number of branches in the network grows, the information that the responses together yield about C grows without bound. However, the information that those responses yield about the signal S approaches a limit:

$$n \to \infty \quad \Rightarrow \quad \sum_{i=1}^{n} \frac{m_i^2}{\sigma_{C \to R_i}^2} \to \infty \quad \Rightarrow \quad I(R_1,...,R_n;S) \to \frac{1}{2}\log_2\left(1 + m_C^2 \frac{\sigma_S^2}{\sigma_{S \to C}^2}\right) = I(C;S) \quad (3.9.19)$$

The equivalence to $I(C;S)$ can be seen by considering a bush network (Eq. **3.9.13**) with a single branch from S to C with slope (gain) m_C and cellular variability magnitude $\sigma_{S \to C}^2$. (The data processing inequality[98] yields the same upper limit, i.e. if $S \to C \to (R_1, ..., R_n)$ form a Markov chain, then $I(R_1,...,R_n;S) \le I(C;S)$, but Eq. **3.9.19** shows that the limit is actually approached through the use of many pathway branches.) Thus, many downstream branches allow a very accurate and informative estimate of C, but the information that these branches can obtain about S is limited by the bottleneck resulting from noise in the trunk portion of the pathway from S to C.

Finally, when the n downstream branches are equivalent the formula simplifies to Eq. **3.3.3** in the main text. In particular, if all the $m_i = 1$ and the magnitude of the variability in the branches is the same $\sigma_{C \to R}^2 = \sigma_{C \to R_i}^2$ for $i = 1, ..., n$, and we further assume for simplicity that $m_C = 1$, then the mutual information becomes:

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1 + \frac{n\sigma_S^2 / \sigma_{C \to R}^2}{1 + n\sigma_{S \to C}^2 / \sigma_{C \to R}^2}\right) \qquad (3.9.20)$$

Again, the simplified formula highlights the dependence of the information on the two key signal-to-noise ratios and the number of downstream branches.

3.10 Predictions made by the bush and tree network models

The Gaussian, linear response models for tree and bush networks described in Sec. 3.8 and 3.9 make specific quantitative predictions for mutual information. Both models make predictions for the information that multiple responses yield about the signal, based on the amount of information that the individual responses yield about the signal. The models also predict the mutual information between the responses. For the tree model, one can further predict the information capacity of the trunk. In this section, we derive formulas that enable such predictions. We illustrate the methods given experimental data for $n = 2$ responses, although they generalize to larger n.

3.10.1 Predicting $I(R_1,R_2;S)$ for the Gaussian bush network

Eq. **3.9.13** shows that the information captured by multiple responses emanating from a bush network depends on the sum of the signal-to-noise ratios for the individual branches. Reversing the relations, these ratios can be obtained from the information captured by the individual responses. In particular, we may compute ϕ_1, the signal-to-noise ratio for branch #1, as follows:

$$I_1 \equiv I(R_1; S) = \frac{1}{2}\log_2\left(1 + m_1^2 \frac{\sigma_S^2}{\sigma_{S \to R_1}^2}\right) \quad \Rightarrow \quad \phi_1 \equiv m_1^2 \frac{\sigma_S^2}{\sigma_{S \to R_1}^2} = 2^{2I_1} - 1 \qquad (3.10.1)$$

Likewise, for branch #2, $\phi_2 = 2^{2I_2} - 1$. Then, Eq. **3.9.13** predicts that the mutual information captured by the two responses together is simply:

$$I_{12} \equiv I(R_1, R_2; S) = \frac{1}{2}\log_2\left(1 + \phi_1 + \phi_2\right) \qquad (3.10.2)$$

3.10.2 Predicting $I(R_1, R_2; S)$ for the Gaussian tree network

Eq. **3.9.18** shows that the information captured by multiple responses emanating from a tree network depends on the sums of two signal-to-noise ratios, namely $\sigma_S^2 / \sigma_{C \to R_i}^2$ and $\sigma_{S \to C}^2 / \sigma_{C \to R_i}^2$, whose values are normalized by the slopes (gains) m_i and m_C. For each branch, the latter ratio $\phi_{C,i} \equiv m_i^2 \sigma_{S \to C}^2 / \sigma_{C \to R_i}^2$ can be obtained by rearranging expressions given in Sec. 3.9 for the overall conditional variance and covariance of the responses:

$$
\begin{aligned}
\phi_{C,1} &= m_1^2 \frac{\sigma_{S \to C}^2}{\sigma_{C \to R_1}^2} = \frac{\frac{m_1}{m_2}\operatorname{cov}(R_1, R_2 | S)}{\operatorname{var}(R_1 | S) - \frac{m_1}{m_2}\operatorname{cov}(R_1, R_2 | S)} \\[2mm]
\phi_{C,2} &= m_2^2 \frac{\sigma_{S \to C}^2}{\sigma_{C \to R_2}^2} = \frac{\frac{m_2}{m_1}\operatorname{cov}(R_1, R_2 | S)}{\operatorname{var}(R_2 | S) - \frac{m_2}{m_1}\operatorname{cov}(R_1, R_2 | S)}
\end{aligned}
\qquad (3.10.3)
$$

The variance and covariance terms can be measured directly from the experimental data. The ratio of the slopes (gains) m_2/m_1 (or its inverse) can also be determined experimentally as the slope of the best fit line through the average values of R_2 plotted against the average values of R_1 that are induced by various levels of the signal S.

The other key ratio, $\phi_{S,i} \equiv m_1^2 m_C^2 \sigma_S^2 / \sigma_{C \to R_i}^2$, can be obtained from Eq. **3.9.18** using $\phi_{C,i}$. For branch #1, this is done as follows

$$I_1 \equiv I(R_1; S) = \frac{1}{2} \log_2 \left(1 + \frac{m_1^2 m_C^2 \sigma_S^2 / \sigma_{C \to R_1}^2}{1 + m_1^2 \sigma_{S \to C}^2 / \sigma_{C \to R_1}^2} \right) \quad \Rightarrow \quad \phi_{S,1} \equiv m_1^2 m_C^2 \sigma_S^2 / \sigma_{C \to R_i}^2 = (2^{2I_1} - 1)(1 + \phi_{C,1})$$

(3.10.4)

Likewise, for branch #2, we have $\phi_{S,2} = (2^{2I_2} - 1)(1 + \phi_{C,2})$. Together, Eq. **3.9.18** then predicts that the mutual information captured by the two responses together is simply:

$$I_{12} \equiv I(R_1, R_2; S) = \frac{1}{2} \log_2 \left(1 + \frac{\phi_{S,1} + \phi_{S,2}}{1 + \phi_{C,1} + \phi_{C,2}} \right) \tag{3.10.5}$$

3.10.3 Predicting $I(R_1;R_2|S)$ for the Gaussian bush and tree networks

The quantity $I(R_1;R_2|S)$ measures the amount of information one can obtain about a response R_1 with knowledge of the other response R_2, or vice versa, given the signal. It can be measured experimentally, e.g. by performing the computations of Sec. 3.6.1 on data obtained from single cells co-stained for multiple responses. These experimental measurements can then be compared to the values predicted from the bush and tree models.

The key assumption in the bush model is that the responses are conditionally independent given the signal. Therefore, the bush model predicts $I_{bush}(R_1;R_2|S) = 0$.

On the other hand, the tree model assumes that the responses are not conditionally independent, and hence $I(R_1;R_2|S)$ is greater than zero. Since R_1 and R_2 are assumed to be jointly normally distributed, the mutual information can be predicted by considering the correlation between the responses (Eq. **3.9.5**). In this case, the correlation is:

$$\rho^2 = \frac{\text{cov}^2(R_1, R_2 \mid S)}{\text{var}(R_1 \mid S)\,\text{var}(R_2 \mid S)}$$

$$= \frac{(m_1 m_2 \sigma_{S \to C}^2)^2}{(m_1^2 \sigma_{S \to C}^2 + \sigma_{C \to R_1}^2)(m_2^2 \sigma_{S \to C}^2 + \sigma_{C \to R_2}^2)}$$

$$= \frac{(m_1^2 \frac{\sigma_S^2}{\sigma_{C \to R_1}^2})(m_2^2 \frac{\sigma_S^2}{\sigma_{C \to R_2}^2})}{(m_1^2 \frac{\sigma_{S \to C}^2}{\sigma_{C \to R_1}^2} + 1)(m_2^2 \frac{\sigma_{S \to C}^2}{\sigma_{C \to R_2}^2} + 1)}$$

$$= \frac{\phi_{C,1} \phi_{C,2}}{(\phi_{C,1} + 1)(\phi_{C,2} + 1)} \tag{3.10.6}$$

where we used the ϕ notation of Sec. 3.10.2. The values of $\phi_{C,1}$ and $\phi_{C,2}$ can be obtained experimentally using the methods also described in Sec. 3.10.2. Then, plugging into Eq. **3.9.5** yields the predicted information:

$$I_{\text{tree}}(R_1; R_2 \mid S) = -\frac{1}{2}\log_2\left(1 - \rho^2\right)$$

$$= \frac{1}{2}\log_2\left(1 + \frac{\phi_{C,1}\phi_{C,2}}{1 + \phi_{C,1} + \phi_{C,2}}\right) \tag{3.10.7}$$

3.10.4 Predicting *I(C;S)* for the Gaussian tree network

In a tree network, the common trunk from S to C sets a limit on the information about the signal that can be transmitted to the downstream branches, and this limit is given by $I(C;S)$. Eq. **3.9.19** shows that for a Gaussian tree network, $I(C;S)$ depends solely on the ratio $m_C^2 \sigma_S^2 / \sigma_{S \to C}^2$. By examining the definitions of $\phi_{S,1}$ and $\phi_{C,1}$ from Sec. 3.10.2 it can be easily seen that $m_C^2 \sigma_S^2 / \sigma_{S \to C}^2 = \phi_{S,1} / \phi_{C,1}$. Thus, the predicted value of $I(C;S)$ is

109

$$I(C;S) = \frac{1}{2}\log_2\left(1 + \frac{\phi_{S,1}}{\phi_{C,1}}\right)$$

(3.10.8)

$I(C;S)$ can also be predicted from $\phi_{S,2}/\phi_{C,2}$ if a second response was measured (and so on for three or more responses), and the predicted values can be averaged together to yield a final prediction.

3.10.5 Predicting $I(R_1,…,R_n;S)$ for the Gaussian tree network for an arbitrary number of identical branches

The mutual information for a tree network whose branches have identical levels of noise is given by Eq. **3.9.20**

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1 + \frac{n\sigma_S^2/\sigma_{C\to R}^2}{n\sigma_{S\to C}^2/\sigma_{C\to R}^2 + 1}\right)$$

(3.10.9)

where the values of m_C and m_i have been subsumed into σ_S^2 and $\sigma_{S\to C}^2$, respectively. This formula shows that the information essentially depends on just three parameters: n, $\sigma_S^2/\sigma_{C\to R}^2$, and $\sigma_{S\to C}^2/\sigma_{C\to R}^2$. Here, we show how to fit this equation to experimental data. To simplify the algebra, we will denote the noise ratios as $\phi_S \equiv \sigma_S^2/\sigma_{C\to R}^2$ and

110

$\phi_C \equiv \sigma^2_{S \to C} / \sigma^2_{C \to R}$. Furthermore, we define θ_n to be a function of the mutual information resulting from n responses as:

$$\frac{n\phi_S}{n\phi_C + 1} = 2^{2I(R_1,\dots,R_n;S)} - 1 \equiv \theta_n \qquad (3.10.10)$$

Suppose that the mutual information has been experimentally measured for two different values of n (i.e., n_1 and n_2) and the ratio n_1/n_2 is also known. First, we will show how to extrapolate to $n \to \infty$. To do this we solve Eq. **3.10.10** for n, yielding

$$n = \frac{\theta_n}{\phi_S - \theta_n \phi_C} \qquad (3.10.11)$$

Then, by writing Eq. **3.10.11** for n_1 and n_2, dividing, and rearranging we obtain

$$\frac{n_1}{n_2} = \frac{\dfrac{\theta_{n_1}}{\phi_S - \theta_{n_1}\phi_C}}{\dfrac{\theta_{n_2}}{\phi_S - \theta_{n_2}\phi_C}} \Leftrightarrow \frac{\phi_S}{\phi_C} = \frac{\theta_{n_1}\theta_{n_2}(1 - \frac{n_1}{n_2})}{\theta_{n_1} - \theta_{n_2}\frac{n_1}{n_2}} \qquad (3.10.12)$$

Thus, the ratio ϕ_S/ϕ_C depends only on experimentally accessible quantities. Examination of Eqs. **3.9.19** and **3.10.8** shows that this ratio allows us to directly compute the mutual information resulting from an infinite number of branches as:

$$n \to \infty \Rightarrow I(R_1, \ldots, R_n; S) \to \frac{1}{2} \log_2 \left(1 + \frac{\phi_S}{\phi_C} \right) \tag{3.10.13}$$

Next, suppose that we wish to compute the mutual information for some other

value of n (or, at least for some other value of n/n_2 if the exact value of n_2 is not known).

Then, replacing n_1 with an arbitrary value $n > 0$ in Eq. **3.10.12** and solving for θ_n gives

$$\frac{\phi_S}{\phi_C} = \frac{\theta_n \theta_{n_2} (1 - \frac{n}{n_2})}{\theta_n - \theta_{n_2} \frac{n}{n_2}} \Rightarrow \theta_n = \frac{\theta_{n_2} \frac{n}{n_2} \frac{\phi_S}{\phi_C}}{\theta_{n_2} (\frac{n}{n_2} - 1) + \frac{\phi_S}{\phi_C}} \tag{3.10.14}$$

which is a quantity consisting of all known values except n (or n/n_2). Thus, inverting the

definition of θ_n gives the desired mutual information as a function of n (or n/n_2):

$$I(R_1, \ldots, R_n; S) = \frac{1}{2} \log_2 (1 + \theta_n) \tag{3.10.15}$$

Chapter 4. Conclusions

In this dissertation we have developed new mathematical and experimental methodologies to characterize how both noise and information propagate through intracellular signaling in further detail.

4.1 Summary of results

In chapter 2, we find that the linear correlation of the average dose responses for reporter pairs allows us to develop a mathematical and experimental framework to decompose noise in intracellular signal transduction networks. By applying this natural extension of the dual reporter method to the TNF signaling network, we found that the JNK branch contributes more noise than the NF-κB branch. Further detailed noise mapping revealed that within the JNK branch, that the c-Jun branch contributes more noise than the ATF-2 branch. We then considered the effects of negative feedback on noise propagation by examining the effects of negative regulation by A20. By applying this framework to wildtype and A20 knockout cell lines, we determined that A20 can suppress noise both at the TNF receptor and at the ATF-2 branch level. Although negative feedback could possibly violate the assumptions of independent noise terms, we find that this method provides, at minimum, a useful and even predictive first approximation of noise propagation within signaling networks. In addition, because of the inherent scalability, this methodology can readily decompose larger networks.

In chapter 3, we used principles of information theory to develop a mathematical model to quantify the amount of information transduced by biochemical signaling pathways. We found that the TNF pathway was, at best, able to transfer 1 bit of information, which is only sufficient for accurate binary decisions. We then evaluated the incorporation of additional pathways to allow for increased information transfer. We found that the addition of an extra pathway, at minimum, adds no additional information but at best, can provide a limitless bound of information that can be transferred. With this understanding, we then developed the bush and tree network models. We found that the TNF signaling pathway is best modeled as a tree network; thus, the receptor level creates an information bottleneck. We then investigated the use of multiple pathways to aid in information transfer in the context of time averaging. We found that time averaging does not significantly add additional information because NF-κB activity behaves deterministically.

4.2 Future outlook and directions

In the past, gene expression systems in isolation, as reviewed in this dissertation, have undergone a very thorough characterization to reveal how biological noise is shaped by genetic networks. However, more recently, we have seen a transition towards examining components upstream of gene expression to understand how noise can propagate through these signal transduction networks. In the future, we envision that with the advent of novel technologies to aid in the rapid characterization of signaling pathways we will be able to achieve a deeper understanding of how signaling pathways

can shape the propagation of noise at the same level of detail as our current understanding of genetic networks. This capability will allow us to create a more comprehensive picture of the limitations of biochemical signaling and aid us in understanding the population level benefits of signaling motifs. In parallel, we find that information theory has steadily advanced into the biological lexicon. Initially used for the spike decoding of individual neurons, it is starting to find application in the modeling of chemotaxis, embryonic drosophila patterning, and signal transduction. We imagine that in the future, this mathematical framework can help trace the information flow through signaling networks which will help us understand what biochemical mechanisms are directly responsible for the impedance and propagation of information.

Holistically, these incremental gains in knowledge will advance us towards developing more general mathematical principles and engineering heuristics to draw biology closer to a more rule-based predictive science. Similar to the historical progression of electronics, we believe that once these biological principles are discovered and established, the engineering ethos will allow us to create whole cell models which will allow us to develop novel de novo solutions to pressing biological problems and uncover critical insights into existing cellular organisms.

Bibliography

1   Raser, J. M. & O'Shea, E. K. Noise in Gene Expression: Origins, Consequences, and Control. *Science* **309**, 2010-2013, doi:10.1126/science.1105891 (2005).

2   Novick, A. & Weiner, M. ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. *Proceedings of the National Academy of Sciences* **43**, 553-566 (1957).

3   Arkin, A., Ross, J. & McAdams, H. H. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells. *Genetics* **149**, 1633-1648 (1998).

4   Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340-2361, doi:10.1021/j100540a008 (1977).

5   Gillespie, D. T. Chemical Langevin equation. *Journal of Chemical Physics* **113**, 297-306 (2000).

6   Gillespie, D. T. The Chemical Langevin and Fokker−Planck Equations for the Reversible Isomerization Reaction†. *The Journal of Physical Chemistry A* **106**, 5063-5071, doi:10.1021/jp0128832 (2002).

7   Gillespie, D. T. The multivariate Langevin and Fokker--Planck equations. *American Journal of Physics* **64**, 1246-1257 (1996).

8   Arkin, A. & Ross, J. Statistical Construction of Chemical Reaction Mechanisms from Measured Time-Series. *The Journal of Physical Chemistry* **99**, 970-979, doi:10.1021/j100003a020 (1995).

9   Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183-1186, doi:10.1126/science.1070919 (2002).

10   Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* **99**, 12795-12800, doi:10.1073/pnas.162041399 (2002).

11   Raser, J. M. & O'Shea, E. K. Control of Stochasticity in Eukaryotic Gene Expression. *Science* **304**, 1811-1814, doi:10.1126/science.1098641 (2004).

12   Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* **98**, 8614-8619, doi:10.1073/pnas.151588598 (2001).

13   Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat Genet* **31**, 69-73 (2002).

14   Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* **311**, 1600-1603, doi:10.1126/science.1119623 (2006).

15   Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485-490 (2008).

16   Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633-637 (2003).

17   Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional Pulsing of a Developmental Gene. *Current Biology* **16**, 1018-1025 (2006).

18   Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Meth* **5**, 877-879 (2008).

19   Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).

20   Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861-864 (2006).

21   Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene Regulation at the Single-Cell Level. *Science* **307**, 1962-1965, doi:10.1126/science.1106914 (2005).

22   Becskei, A., Kaufmann, B. B. & van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat Genet* **37**, 937-944 (2005).

23   Pedraza, J. M. & van Oudenaarden, A. Noise Propagation in Gene Networks. *Science* **307**, 1965-1969, doi:10.1126/science.1109090 (2005).

24   Stockholm, D. *et al.* The Origin of Phenotypic Heterogeneity in a Clonal Cell Population In Vitro. *PLoS One* **2**, e394, doi:10.1371/journal.pone.0000394 (2007).

25   Berg, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of Theoretical Biology* **71**, 587-603, doi:http://dx.doi.org/10.1016/0022-5193(78)90326-0 (1978).

26   Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* **43**, 95-100 (2011).

27   Rigney, D. R. Stochastic model of constitutive protein levels in growing and dividing bacterial cells. *Journal of Theoretical Biology* **76**, 453-480, doi:http://dx.doi.org/10.1016/0022-5193(79)90013-4 (1979).

28   Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643-646 (2006).

29   Perkins, T. J. & Swain, P. S. Strategies for cellular decision-making. *Mol Syst Biol* **5** (2009).

30   Albeck, J. G., Burke, J. M., Spencer, S. L., Lauffenburger, D. A. & Sorger, P. K. Modeling a Snap-Action, Variable-Delay Switch Controlling Extrinsic Cell Death. *PLoS Biol* **6**, e299, doi:10.1371/journal.pbio.0060299 (2008).

31   Espinosa-Soto, C., Martin, O. & Wagner, A. Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. *BMC Evolutionary Biology* **11**, 5 (2011).

32   Balázsi, G., van Oudenaarden, A. & Collins, James J. Cellular Decision Making and Biological Noise: From Microbes to Mammals. *Cell* **144**, 910-925, doi:10.1016/j.cell.2011.01.030 (2011).

33   Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545-550 (2006).

34   Iber, D., Clarkson, J., Yudkin, M. D. & Campbell, I. D. The mechanism of cell differentiation in Bacillus subtilis. *Nature* **441**, 371-374 (2006).

35   Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. Tunability and Noise Dependence in Differentiation Dynamics. *Science* **315**, 1716-1719, doi:10.1126/science.1137455 (2007).

36   Çağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J. & Süel, G. M. Architecture-Dependent Noise Discriminates Functionally Analogous Differentiation Circuits. *Cell* **139**, 512-522, doi:10.1016/j.cell.2009.07.046.

37   Kalmar, T. *et al.* Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. *PLoS Biol* **7**, e1000149, doi:10.1371/journal.pbio.1000149 (2009).

38   Janes, K. A. *et al.* A Systems Model of Signaling Identifies a Molecular Basis Set for Cytokine-Induced Apoptosis. *Science* **310**, 1646-1653, doi:10.1126/science.1116598 (2005).

39   Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences* **103**, 13004-13009, doi:10.1073/pnas.0605420103 (2006).

40   Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691 (2003).

41   Albeck, J. G. *et al.* Quantitative Analysis of Pathways Controlling Extrinsic Apoptosis in Single Cells. *Molecular Cell* **30**, 11-25, doi:10.1016/j.molcel.2008.02.012 (2008).

42   Geva-Zatorsky, N. *et al. Oscillations and variability in the p53 system*. Vol. 2 (2006).

43   Goldstein, J. C., Kluck, R. M. & Green, D. R. A Single Cell Analysis of Apoptosis: Ordering the Apoptotic Phenotype. *Annals of the New York Academy of Sciences* **926**, 132-141, doi:10.1111/j.1749-6632.2000.tb05607.x (2000).

44   Lahav, G. *et al.* Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet* **36**, 147-150 (2004).

45   Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428-432 (2009).

46   Cohen, A. A. *et al.* Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science* **322**, 1511-1516, doi:10.1126/science.1160165 (2008).

47   Subclavian. Simple representation of TNFR signaling pathways. Dashed lines represent multiple steps. (2007). <http://en.wikipedia.org/wiki/File:TNF_signaling.jpg>.

48   Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science* **334**, 354-358, doi:10.1126/science.1204553 (2011).

49   Tay, S. *et al.* Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267-271 (2010).

50   Lee, T. K. *et al.* A Noisy Paracrine Signal Determines the Cellular NF-{kappa}B Response to Lipopolysaccharide. *Sci. Signal.* **2**, ra65 (2009).

51   Ashall, L. *et al.* Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. *Science* **324**, 242-246 (2009).

52   Nelson, D. E. *et al.* Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science* **306**, 704-708 (2004).

53   Turner, D. A. *et al.* Physiological levels of TNFα stimulation induce stochastic dynamics of NF-κB responses in single living cells. *Journal of Cell Science* **123**, 2834-2843, doi:10.1242/jcs.069641 (2010).

54   Liu, Z.-g. Molecular mechanism of TNF signaling and beyond. *Cell Res* **15**, 24-27 (2005).

55   Chen, G. & Goeddel, D. V. TNF-R1 Signaling: A Beautiful Pathway. *Science* **296**, 1634-1635, doi:10.1126/science.1071924 (2002).

56   Tracey, K. J. & Cerami, A. Tumor Necrosis Factor, Other Cytokines and Disease. *Annual Review of Cell Biology* **9**, 317-343, doi:doi:10.1146/annurev.cb.09.110193.001533 (1993).

57   Barnes, P. J. & Karin, M. Nuclear Factor-κB — A Pivotal Transcription Factor in Chronic Inflammatory Diseases. *New England Journal of Medicine* **336**, 1066-1071, doi:doi:10.1056/NEJM199704103361506 (1997).

58   Karin, M., Cao, Y., Greten, F. R. & Li, Z.-W. NF-[kappa]B in cancer: from innocent bystander to major culprit. *Nat Rev Cancer* **2**, 301-310 (2002).

59   Karin, M. & Greten, F. R. NF-[kappa]B: linking inflammation and immunity to cancer development and progression. *Nat Rev Immunol* **5**, 749-759 (2005).

60   Bowsher, C. G. & Swain, P. S. Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences* **109**, E1320-E1328, doi:10.1073/pnas.1119407109 (2012).

61   Rinott, R., Jaimovich, A. & Friedman, N. Exploring transcription regulation through cell-to-cell variability. *Proceedings of the National Academy of Sciences* **108**, 6329-6334, doi:10.1073/pnas.1013148108 (2011).

62   Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699-706 (2005).

63   Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IκB-NF-κB Signaling Module: Temporal Control and Selective Gene Activation. *Science* **298**, 1241-1245, doi:10.1126/science.1071914 (2002).

64   Cheong, R., Wang, C. J. & Levchenko, A. High Content Cell Screening in a Microfluidic Device. *Molecular & Cellular Proteomics* **8**, 433-442, doi:10.1074/mcp.M800291-MCP200 (2009).

65   Dyson, S. & Gurdon, J. B. The Interpretation of Position in a Morphogen Gradient as Revealed by Occupancy of Activin Receptors. *Cell* **93**, 557-568 (1998).

66   Chen, S.-h., Masuno, K., Cooper, S. B. & Yamamoto, K. R. Incoherent feed-forward regulatory logic underpinning glucocorticoid receptor action. *Proceedings of the National Academy of Sciences* **110**, 1964-1969, doi:10.1073/pnas.1216108110 (2013).

67   Bronnikov, G. E., Zhang, S.-J., Cannon, B. & Nedergaard, J. A Dual Component Analysis Explains the Distinctive Kinetics of cAMP Accumulation in Brown

Adipocytes. *Journal of Biological Chemistry* **274**, 37770-37780, doi:10.1074/jbc.274.53.37770 (1999).

68   Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636-643 (2006).

69   Heyninck, K. & Beyaert, R. A20 inhibits NF-κB activation by dual ubiquitin-editing functions. *Trends in Biochemical Sciences* **30**, 1-4 (2005).

70   Coornaert, B., Carpentier, I. & Beyaert, R. A20: Central Gatekeeper in Inflammation and Immunity. *Journal of Biological Chemistry* **284**, 8217-8221, doi:10.1074/jbc.R800032200 (2009).

71   Vereecke, L., Beyaert, R. & van Loo, G. The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. *Trends in Immunology* **30**, 383-391 (2009).

72   Baer, M. *et al.* Tumor Necrosis Factor Alpha Transcription in Macrophages Is Attenuated by an Autocrine Factor That Preferentially Induces NF-kappa B p50. *Mol. Cell. Biol.* **18**, 5678-5689 (1998).

73   Devin, A., Lin, Y. & Liu, Z. g. The role of the death-domain kinase RIP in tumour-necrosis-factor-induced activation of mitogen-activated protein kinases. *EMBO reports* **4**, 623-627, doi:10.1038/sj.embor.embor854 (2003).

74   Won, M. *et al.* Novel anti-apoptotic mechanism of A20 through targeting ASK1 to suppress TNF-induced JNK activation. *Cell Death Differ* **17**, 1830-1841 (2010).

75   Kelliher, M. A. *et al.* The Death Domain Kinase RIP Mediates the TNF-Induced NF-κB Signal. *Immunity* **8**, 297-303 (1998).

76   Werner, S. L. *et al.* Encoding NF-κB temporal control in response to TNF: distinct roles for the negative regulators IκBα and A20. *Genes & Development* **22**, 2093-2101, doi:10.1101/gad.1680708 (2008).

77   Yu, R. C. *et al.* Negative feedback that improves information transmission in yeast signalling. *Nature* **456**, 755-761 (2008).

78   Dubuis, J. O., Samanta, R. & Gregor, T. Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol Syst Biol* **9** (2013).

79   Rhee, A., Cheong, R. & Levchenko, A. The application of information theory to biochemical signaling systems. *Physical Biology* **9**, 045011 (2012). http://stacks.iop.org/1478-3975/9/i=4/a=045011

80   Borst, A. & Theunissen, F. E. Information theory and neural coding. *Nat Neurosci* **2**, 947-957 (1999).

81   Tkacik, G., Callan, C. G. & Bialek, W. Information capacity of genetic regulatory elements. *Physical Review E* **78**, 011910 (2008).

82   Tkacik, G., Walczak, A. M. & Bialek, W. Optimizing information flow in small genetic networks. *Physical Review E* **80**, 031920 (2009).

83   Tkačik, G., Callan, C. G. & Bialek, W. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences* **105**, 12265-12270, doi:10.1073/pnas.0806077105 (2008).

84   MacKay, D. M. & McCulloch, W. S. The Limiting Information Capacity of a Neuronal Link. *Bulletin of Mathematical Biophysics* **14**, 127-135 (1952).

85   Rapoport, A. & Horvath, W. J. The theoretical channel capacity of a single neuron as determined by various coding systems. *Information and Control* **3**, 335-350, doi:10.1016/s0019-9958(60)90917-7 (1960).

86   Stein, R. B. The Information Capacity of Nerve Cells Using a Frequency Code. *Biophysical Journal* **7**, 797-826 (1967).

87   Berry, M. J., Warland, D. K. & Meister, M. The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences* **94**, 5411-5416 (1997).

88   Linsker, R. Perceptual Neural Organization: Some Approaches Based on Network Models and Information Theory. *Annual Review of Neuroscience* **13**, 257-281, doi:doi:10.1146/annurev.ne.13.030190.001353 (1990).

89   Fairhall, A. L., Lewen, G. D., Bialek, W. & de Ruyter van Steveninck, R. R. Efficiency and ambiguity in an adaptive neural code. *Nature* **412**, 787-792 (2001).

90   Sharpee, T. O. *et al.* Adaptive filtering enhances information transmission in visual cortex. *Nature* **439**, 936-942 (2006).

91   Gregor, T., Tank, D. W., Wieschaus, E. F. & Bialek, W. Probing the Limits to Positional Information. *Cell* **130**, 153-164 (2007).

92   Dubuis, J. O., Tkacik, G., Wieschaus, E. F., Gregor, T. & Bialek, W. Positional information, in bits. *arXiv.org:1201.0198 [q-bio.MN]* (2011).

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

94   Cheong, R., Wang, C. J. & Levchenko, A. High content cell screening in a microfluidic device. *Mol Cell Proteomics* **8**, 433-442 (2009).

95   Werner, S. L. *et al.* Encoding NF-kappaB temporal control in response to TNF: distinct roles for the negative regulators IkappaBalpha and A20. *Genes Dev* **22**, 2093-2101 (2008).

96   Tay, S. *et al.* Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267-271 (2010).

97   Cheong, R. *et al.* Transient IkappaB kinase activity mediates temporal NF-kappaB dynamics in response to a wide range of tumor necrosis factor-alpha doses. *J Biol Chem* **281**, 2945-2950 (2006).

98   Cover, T. M. & Thomas, J. A. *Elements of information theory.* (Wiley, 1991).

99   Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241-1245 (2002).

100   Cohen-Saidon, C., Cohen, A. A., Sigal, A., Liron, Y. & Alon, U. Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol Cell* **36**, 885-893 (2009).

101   Bao, X. R., Fraser, I. D., Wall, E. A., Quake, S. R. & Simon, M. I. Variability in G-protein-coupled signaling studied with microfluidic devices. *Biophys J* **99**, 2414-2422 (2010).

102   Coppey, M., Boettiger, A. N., Berezhkovskii, A. M. & Shvartsman, S. Y. Nuclear trapping shapes the terminal gradient in the Drosophila embryo. *Curr Biol* **18**, 915-919 (2008).

103   Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat Rev Neurosci* **7**, 358-366 (2006).

104   Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995-999 (2008).

105   Schneidman, E., Bialek, W. & Berry II, M. J. Synergy, redundancy, and independence in population codes. *J Neurosci* **23**, 11539-11553 (2003).

106   Wajant, H., Pfizenmaier, K. & Scheurich, P. Tumor necrosis factor signaling. *Cell Death Differ* **10**, 45-65 (2003).

107   Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci U S A* **105**, 17256-17261 (2008).

108   Krishna, S., Jensen, M. H. & Sneppen, K. Minimal model of spiky oscillations in NF-kappaB signaling. *Proc Natl Acad Sci U S A* **103**, 10840-10845 (2006).

109   Thierfelder, S., Ostermann, K., Gobel, A. & Rodel, G. Vectors for glucose-dependent protein expression in Saccharomyces cerevisiae. *Appl Biochem Biotechnol* **163**, 954-964 (2011).

110   Nemenman, I., Lewen, G. D., Bialek, W. & de Ruyter van Steveninck, R. R. Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Comput Biol* **4**, e1000025 (2008).

111   Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R. & Bialek, W. Entropy and information in neural spike trains. *Phys Rev Lett* **80**, 197-200 (1998).

112   Freedman, D. A. & Folkman, J. Maintenance of G1 checkpoint controls in telomerase-immortalized endothelial cells. *Cell Cycle* **3**, 811-816 (2004).

113   Paninski, L. Estimation of entropy and mutual information. *Neural Comput* **15**, 1191-1253 (2003).

114   Nemenman, I., Shafee, F. & Bialek, W. in *Advances in Neural Information Processing Systems* Vol. 14 (eds T. G. Dietterich, S. Becker, & Z. Ghahramani) 95-100 (MIT Press, 2002).

115   Slonim, N., Atwal, G. S., Tkacik, G. & Bialek, W. Information-based clustering. *Proc Natl Acad Sci U S A* **102**, 18297-18302 (2005).

116   Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7 (2006).

117   Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys Rev E* **69**, 066138 (2004).

118   Slonim, N., Atwal, G. S., Tkacik, G. & Bialek, W. Estimating mutual information and multi-information in large networks. *arXiv cs.IT/0502017* (2005).

119   Chong, E. K. P. & Zak, S. H. *An introduction to optimization, 3rd ed.*, (Wiley, 2008).

120   Tkacik, G., Callan, C. G., Jr. & Bialek, W. Information capacity of genetic regulatory elements. *Phys Rev E* **78**, 011910 (2008).

121   Nemenman, I. & Bialek, W. Occam factors and model independent Bayesian learning of continuous distributions. *Phys Rev E* **65**, 026137 (2002).

122   Panzeri, S., Senatore, R., Montemurro, M. A. & Petersen, R. S. Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* **98**, 1064-1072 (2007).

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

124   Margolin, A. A., Wang, K., Califano, A. & Nemenman, I. Multivariate dependence and genetic networks inference. *IET Syst Biol* **4**, 428 (2010).

125   Magesacher, T., Odling, P., Sayir, J. & Nordstrom, T. in *IEEE International Symposium on Information Theory.*

Curriculum Vitae

Education

> **Ph.D.** **Johns Hopkins University,** Baltimore, MD   2014
>
> Department of Biomedical Engineering
>
> Advisor: Dr. Andre Levchenko
>
> **B.S.** **University of Toronto,** Toronto, Canada   2005
>
> Department of Chemical Engineering

Other Professional Experience

> *Research Associate*   2006 – 2007
>
> **Fio Corporation,** Toronto, Canada
>
> *Junior Process Engineer*   2005—2006
>
> **Hatch Ltd.,** Toronto, Canada

Awards & Honors

NSERC Postgraduate Scholarship M (Declined. Current   2007

school outside Canada)

Ontario Graduate Scholarship (Declined. Current school   2007

outside Ontario)

NSERC Undergraduate Student Research Award   2003 – 2004

Ransom Scholarship in Chemical Engineering   2002

Peer Reviewed Publications

<u>**A Rhee**</u>*****, R Cheong*****, A Levchenko. "Noise decomposition of

intracellular biochemical signaling networks." (In preparation).

*Denotes equal contribution.

Annaluru, N. *et al.* Total Synthesis of a Functional Designer

Eukaryotic Chromosome. *Science* **344**, 55-58,

doi:10.1126/science.1249252 (2014). **(Featured in Science's News**

Focus.)

R Cheong, <u>**A Rhee**</u>, CJ Wang, I Nemenman, A Levchenko.

"Information transduction capacity of noisy biochemical signaling

networks." *Science*. 334(6054):354-8 (2011). **(Featured in Science's Perspectives and Biopolymers' Research highlights.)**

S Fournier-Bidoz, TL Jennings, JM Klostranec, W Fung, **A Rhee,** et al. "Facile and Rapid One-Step Mass Preparation of Quantum-Dot Barcodes". *Angewandte Chemie International*. 120(30): 5557 (2008).

JM Klostranec, Q Xiang, GA Farcas, JA Lee, **A Rhee,** et al. "Convergence of Quantum Dot Barcodes with Microfluidics and Signal Processing for Multiplexed High-Throughput Infectious Disease Diagnostics". *Nano Letters*. 7(9): 2812-2818 (2007).

JA Lee, A Hung, S Mardyani, <u>A Rhee</u>, et al. "Toward the Accurate Read-out of Quantum Dot Barcodes: Design of Deconvolution Algorithms and Assessment of Fluorescence Signals in Buffer. *Advanced Materials* 19, 3113-3118, (2007).

T Kim, **A Rhee**, CM Yip. "Force-induced insulin dimer dissociation: a molecular dynamics study". *J. Am. Chem. Soc.* 128(16): 5330-5331 (2006).

Review Articles

A Rhee, R Cheong, A Levchenko. "The application of information theory to biochemical signaling systems." *Phys. Bio*. 9(4):045011 (2012).

Conferences and Presentations

R Cheong, **A Rhee**, CJ Wang, I Nemenman, A Levchenko. "Information transduction capacity of noisy biochemical signaling networks." Mathematical Biosciences Institute Workshop 2: Stochastic Processes in Cell and Population Biology. Columbus, OH, USA (Oct 24 2011 – Oct 28 2011).

R Cheong, **A Rhee**, CJ Wang, I Nemenman, A Levchenko. "The amount of information transduced by noisy signaling pathways: bottlenecks and strategies to overcome them." Gordon Research Conference: Stochastic Physics in Biology. Ventura, CA, USA (Jan 23 2011 – Jan 28 2011).

R Cheong, **A Rhee**, CJ Wang, I Nemenman, A Levchenko. "Advantages and limitations of network-based information processing

in biological signaling systems." 5th q-bio Conference on Cellular Information Processing. **Selected for oral presentation.** Santa Fe, NM, USA (Aug 10 2011 – Aug 13 2011).

R Cheong, **A Rhee**, I Nemenman, A Levchenko. "Information processing in the TNF-stimulated NF-κB signaling pathway in fibroblasts: what thousands of little cells can tell you?" 3rd q-bio Conference on Cellular Information Processing. Santa Fe, NM, USA (Aug 5 2009 – Aug 9 2009).

T.Kim, **A Rhee,** CM Yip. "Molecular Dynamics Studies of Insulin-Insulin Interactions". *Biophysical Society 48th Annual Meeting.* Baltimore, MD, USA (Feb 14 2004 – Feb 18 2004).

Service and Leadership

Head of Household   2007 – 2013

Incentive Mentoring Program (IMP)

Band Founder and Lead Guitarist   2011 – 2013

Fishcloset

BME PhD Council President   2009 – 2010

Johns Hopkins University