SYSTEMS BIOLOGY

Accurate information transmission through dynamic biochemical signaling networks

Jangir Selimkhanov,[1*] Brooks Taylor,[1*] Jason Yao,[2] Anna Pilko,[2] John Albeck,[3] Alexander Hoffmann,[4,5] Lev Tsimring,[4,6] Roy Wollman[2,4,7†]

Stochasticity inherent to biochemical reactions (intrinsic noise) and variability in cellular states (extrinsic noise) degrade information transmitted through signaling networks. We analyzed the ability of temporal signal modulation—that is, dynamics—to reduce noise-induced information loss. In the extracellular signal–regulated kinase (ERK), calcium (Ca^{2+}), and nuclear factor kappa-B (NF-κB) pathways, response dynamics resulted in significantly greater information transmission capacities compared to nondynamic responses. Theoretical analysis demonstrated that signaling dynamics has a key role in overcoming extrinsic noise. Experimental measurements of information transmission in the ERK network under varying signal-to-noise levels confirmed our predictions and showed that signaling dynamics mitigate, and can potentially eliminate, extrinsic noise–induced information loss. By curbing the information-degrading effects of cell-to-cell variability, dynamic responses substantially increase the accuracy of biochemical signaling networks.

The role of biological signaling networks is to reliably transmit specific information about the extracellular environment to downstream effectors, allowing the cell to adjust its physiological state to changing conditions. The stochasticity of molecular interactions that underlies various forms of "noise" in biological systems can interfere with signal transduction and degrade the transmitted information. How signaling networks perform their core functions in the presence of noise is a fundamental question. Information-theoretic approaches allow estimation of the maximal possible information transmission capacity of noisy biochemical networks (*1*–*11*). Previous applications of such methods to the analysis of signaling networks suggested that as a result of noise, cells lose most of the information about the concentration of ligands (*12*–*14*). Thus far, the information-theoretic analyses of signaling networks have been based on scalar measurements performed at a single time point. However, the information on activating ligands is often encoded using a dynamic signal represented by a multivariate vector that contains a single cell's response at multiple time points (*15*–*18*).

[1]Department of Bioengineering, University of California–San Diego, La Jolla, CA 92093, USA. [2]Department of Chemistry and Biochemistry, University of California–San Diego, La Jolla, CA 92093, USA. [3]Department of Molecular and Cellular Biology, University of California–Davis, Davis 95616, USA. [4]San Diego Center for Systems Biology, La Jolla, CA 92093, USA. [5]Institute for Quantitative and Computational Biosciences and Department of Microbiology, Immunology, and Molecular Genetics, University of California–Los Angeles, Los Angeles, CA 90025, USA. [6]BioCircuits Institute, University of California–San Diego, La Jolla, CA 92093, USA. [7]Cell and Developmental Biology Section, Division of Biological Sciences, University of California–San Diego, La Jolla, CA 92093, USA.
*These authors contributed equally to this work. †Corresponding author. E-mail: rwollman@ucsd.edu

To test the hypothesis that dynamic responses contain more information than static responses, we performed single live-cell measurements of three key signaling pathways (Fig. 1): extracellular signal–regulated kinase (ERK), calcium (Ca^{2+}), and nuclear factor kappa-B (NF-κB) [supplementary materials (SM) section 1.1]. Fully automated computational image analysis (SM section 1.2) allowed us to measure the response of 910,121 individual live cells (figs. S1 to S6 and tables S1 to S3). The large sample size was instrumental for analyzing high-dimensional multivariate responses. In all three pathways, there was substantial variability within the dynamic (Fig. 1, C to E) and nondynamic (Fig. 1F) single-cell responses across multiple concentrations of activating ligands.

To analyze the implications of noise on information loss, we used an information-theoretic approach to calculate the information transmission capacity of a dynamic signaling network. The information transmission capacity [also referred to as channel capacity (*19*)] is measured as the maximal possible mutual information between the measured response and the activating ligand concentration. To calculate the mutual information between a dynamic response (a vector) and the ligand concentration (a scalar), we expanded on a previously described algorithm (*14*). The algorithm uses continuous multidimensional response data and a k–nearest-neighbor approach to estimate the conditional probability density for each cell's response (SM section 2). We thus estimated the information transmission capacity of the dynamic response and of several types of static responses. For all single–time point static scalar responses, we found transmission capacity (<1 bit) (*12*, *13*) (Fig. 2A). However, across all three signaling pathways, the dynamic response had significantly higher information transmission capacity than several scalar responses previously described (*20*, *21*) (Fig. 2, B and C, Student's t test, $P < 0.05$ for all comparisons, table S6). These estimates should be considered as lower bounds because they do not exclude variability resulting from experimental imperfections.

To elucidate the origins of the enhanced information transmission capacity of dynamic signaling responses, we developed a mathematical theory using information-theoretic formalism (SM section 3). The theory explicitly accounts for the information-degrading effects of intrinsic and extrinsic noise sources in the context of multivariate responses. Intrinsic noise adds to uncertainty in all dimensions (i.e., time points) independently from one another. In contrast, the extrinsic variability in cellular states produces fluctuations that are constrained by the signaling network that generates the dynamics. Therefore, the fluctuating components generated by extrinsic noise at different time points are deterministically dependent on one another. As a result, intrinsic and extrinsic noise sources have different effects on the information transmission capacity of multivariate responses. In the case of purely intrinsic noise, additional measurements increase the information logarithmically because of simple ensemble averaging (*12*). In the case of purely extrinsic noise, a sufficient number of dynamical measurements can provide complete information about the a priori uncertain internal state of the cell and therefore lead to a substantial gain in the information about the activating ligand (Fig. 3A).

To test our analytical prediction that the multivariate dynamic response can completely eliminate the information loss that results from introduction of extrinsic noise (SM section 3), we used computer simulations of ERK responses based on a published kinetic model (SM section 4.1) (*22*) (fig. S15). We generated sets of simulated ERK activity trajectories in response to an increasing number of ligand concentrations. We varied model values for ERK and mitogen-activated protein kinase kinase (MEK) according to a uniform distribution (±20% mean value) to mimic extrinsic noise and measured the information transmission capacity. Our analysis supported the analytical prediction and showed that whereas the univariate response, based on maximal ERK dynamics, had limited information transmission capacity, the dynamic multivariate response can transmit complete information about ligand concentration (Fig. 3B). An intuitive demonstration for the limitation of univariate response and the ability of multivariate response to overcome extrinsic noise is shown in Fig. 3, C and D. Superficially, the trajectories of two populations of simulated responses of ERK activity to two input concentrations of epidermal growth factor (EGF) appear overlapping (Fig. 3C), but in fact, they are completely separable when considering joint distributions (*23*). Plotting the distribution of ERK activity at t = 9 and t = 24 min on a two-dimensional (2D) plane (Fig. 3D) shows that the responses to a single varied parameter input lies on a one-dimensional (1D)

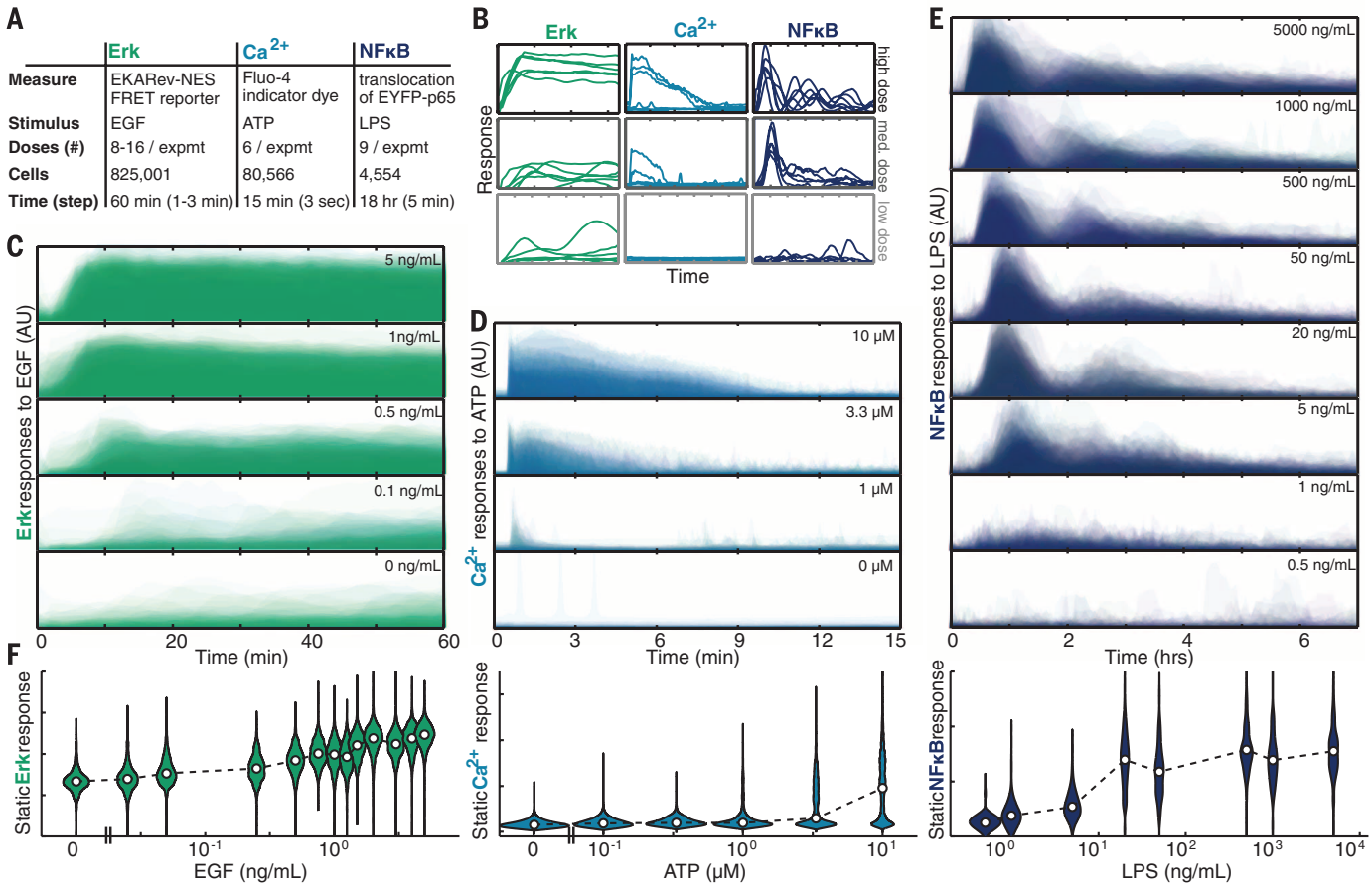

Fig. 1. Single-cell measurement of the dynamic response of ERK, Ca^{2+}, and NF-κB. (**A**) Overview of single-cell data analyzed in this work. (**B**) Examples of single-cell response dynamic trajectories. (**C** to **E**) Temporal histograms of several representative dosages for ERK (C), Ca^{2+} (D), and NF-κB (E). Color intensity reflects the probability density of a cellular response magnitude at each time point. *Y* axis in (B) to (E) is the same for each pathway and is in arbitrary units (AU), representing the Förster resonance energy transfer (FRET) to cyan fluorescent protein (CFP) ratio reported by the EKARev ERK biosensor (C), intensity of Ca^{2+} indicator dye Fluo-4 (D), and ratio of nuclear to cytoplasmic localization of an enhanced yellow fluorescent protein (EYFP)–p65 reporter (E). (**F**) Violin plot of the maximally separable static response in the three signaling pathways. Shape width shows response distribution (areas are equal), and point is the median response in each condition. EGF, epidermal growth factor; ATP, adenosine triphosphate; LPS, lipopolysaccharide.

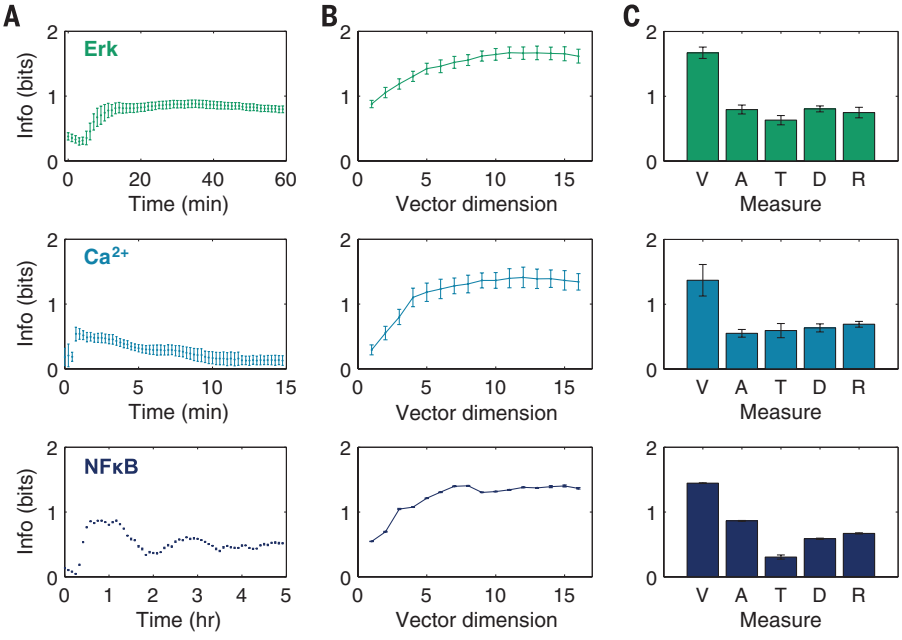

Fig. 2. Information transmission capacity of static and dynamic ERK, Ca^{2+}, and NF-κB responses. (**A**) Information transmission capacity calculated from static scalar response distribution based on single–time point measurements. (**B**) Information transmission capacity calculated from multivariate dynamic responses as a function of the dimension of the multivariate vector. The multivariate vector was subsampled using a uniform grid centered on the middle time point (fig. S19). (**C**) Comparison of the multivariate vector (V) measurement to the following scalar responses: maximum response amplitude (A), maximum response time (T), maximal rate of response (D), ratio of maximum response amplitude to initial response amplitude (R). Error bars are SEMs from six biological replicates for ERK and four for Ca^{2+}, and SDs from five jackknife iterations for NF-κB (tables S1 to S3). The multivariate vector information transfer was significantly greater than all scalar measures ($P < 0.05$, Student's *t* test, table S6).

curve within a 2D space. The two 1D manifolds for different inputs are completely separated from each other (inset), but overlap considerably in any 1D projection. This simple example demonstrates how the extrinsic variability of a single parameter can in principle be completely eliminated with measurements from only two time points.

The accuracy of a response can be characterized by its signal-to-noise ratio (SNR). The mutual information and the system's SNR are related; however, this relationship is strongly affected by the noise properties (intrinsic versus extrinsic) and the type of the response. Our analytical theory predicts a different relationship between mutual information and SNR for three different types of responses: (i) scalar responses that do not distinguish between intrinsic and extrinsic noise; (ii) multivariate responses without any dynamic component (redundant measurements) that can only reduce intrinsic noise; and (iii) dynamic responses that combine the benefits of redundant measurements with efficient mitigation of extrinsic variability. We varied the SNR in the ERK network by partial inhibition of the ERK kinase MEK with six different dosages of the inhibitor U0126. At each MEK inhibition level, we measured ERK response to eight EGF levels. A total of 48 conditions were measured in four biological replicates (fig. S23). At each MEK inhibitor level, we calculated the mutual information and the SNR from single-cell responses (SM section 4.3.1). In total, Fig. 4 contains 535,107 cell responses (tables S4 and S5). As expected, for a scalar response, the formula relating the mutual information and the overall SNR is in very good agreement with our experimental measurements. The theoretical prediction of mutual information for the redundant measurement case requires knowledge of intrinsic–to–extrinsic noise ratio (IER). IER was estimated in two ways: by (i) quantifying the fluctuations in the later (quasistationary) portion of the response time series of our ERK data (fig. S21A) (ii) using data for repeated measurements of single-cell responses (*24*) (fig. S21B). The predicted mutual information based on redundant responses required IER values that are two to four orders of magnitude higher than experimentally estimated IER values (SM section 4.3.3) (fig S17). In contrast, the measured mutual information values were in good agreement with the theoretical prediction for a dynamic response based on a computational ERK model (SM section 4.1) (fig. S18). Overall, this analysis demonstrates that the substantial information gain from multivariate measurements is indeed the direct result of the dynamic nature of ERK response.

The robustness of biological systems is epitomized by their ability to function in the presence of a large variability in cellular states (*25*, *26*). Signaling dynamics allow biochemical networks to mitigate variability in the cellular state and thereby maximize the information transmission capacity of signaling networks. Although the theory and observations presented here focus on the information transmission capacity of the dynamics of a single signaling molecule, the extension of our analysis to the case of multiple signaling molecules responding to one ligand is straightforward. Not all of the information contained in the dynamical responses may actually

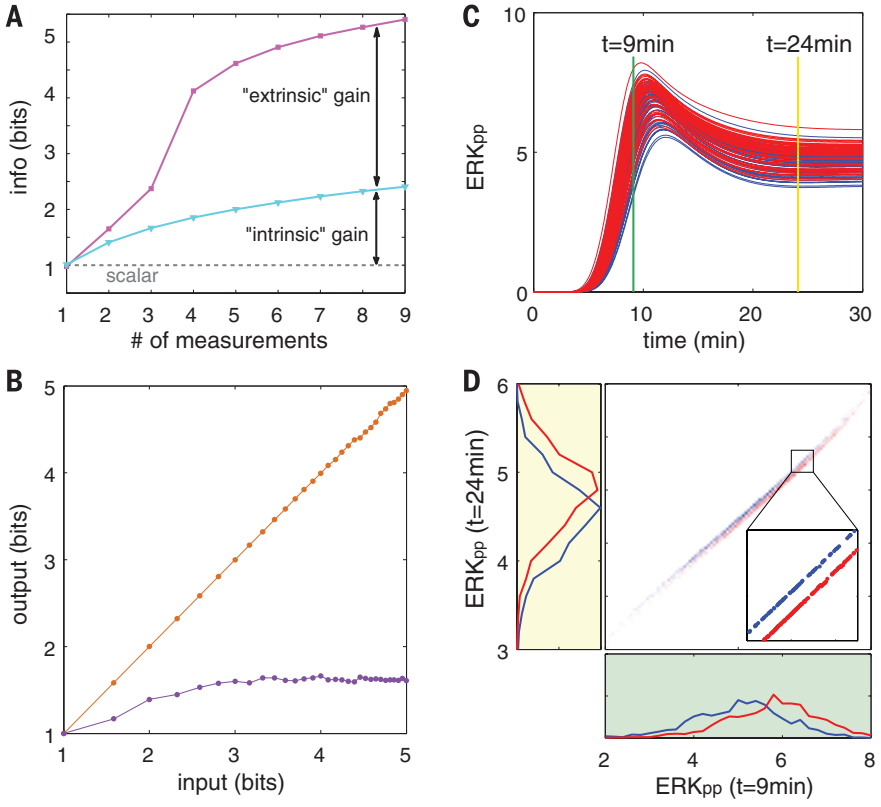

Fig. 3. Theoretical decomposition of information loss caused by intrinsic and extrinsic noise. (**A**) Graphical representation of the analytical expression for the gain in mutual information from overcoming intrinsic (cyan) and extrinsic (magenta) noise sources obtained from random linear Gaussian inputs and outputs with three parameters (*19*). (**B**) Information transmission capacity of dynamic (orange) and static (maximal response, purple) responses calculated using simulated trajectories from the computational model of ERK (*22*) with only the extrinsic noise contributing to cell response variability. (**C**) Example of ERK trajectory variability for two different inputs levels (red and blue). Variability was generated using a uniform distribution of a single parameter, MEK values, that was varied by ±20%. (**D**). Two-dimensional histogram (center) and marginal distributions (left and bottom) for the two input levels (shown in red and blue) at two time points (t = 9 and 24 min) from the trajectories in (C). Because only a single parameter was varried, the responses vary on a 1D curve. As a result, although the univariate marginal distributions show substantial response overlap, the 2D distribution shows completely separated response levels (inset).

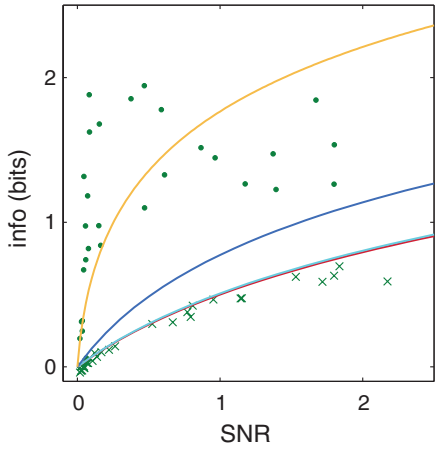

Fig. 4. Measured information gain is a result of ERK dynamics' ability to mitigate extrinsic noise. Experimental measurement of the mutual information between ERK response and EGF measured as a function of the response signal-to-noise ratio (SNR). Each marker represents calculations of SNR and mutual information from the dynamic (dot) and maximal scalar (cross) responses of cells from an eight-well dose-response experiment. Data shown are calculated based on 535,107 single-cell responses from 29 experiments with six doses of MEK inhibitor U0126 (tables S4 and S5). Lines represent theoretical predictions of the mutual information as a function of SNR for three types of responses: static scalar (red line), redundant measurements where the multivariate response has no dynamics (dark and light blue lines) calculated based on two independent estimates of IER (*19*) (fig. S21), and dynamic response (orange) that can mitigate both intrinsic and extrinsic noise.

be used by cells. Yet, because reliable information transmission is a fundamental function of cellular signaling networks, it is plausible that evolutionary pressures shaped the cellular machinery to maximize the reliable decoding of multivariate dynamic signals.

REFERENCES AND NOTES

1. M. D. Brennan, R. Cheong, A. Levchenko, *Science* **338**, 334–335 (2012).
2. A. Rhee, R. Cheong, A. Levchenko, *Phys. Biol.* **9**, 045011 (2012).
3. C. Waltermann, E. Klipp, *Biochim. Biophys. Acta* **1810**, 924–932 (2011).
4. F. Tostevin, P. R. ten Wolde, *Phys. Rev. Lett.* **102**, 218101 (2009).
5. W. H. de Ronde, F. Tostevin, P. R. ten Wolde, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **82**, 031914 (2010).
6. W. H. de Ronde, F. Tostevin, P. R. Ten Wolde, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **86**, 021913 (2012).
7. G. Tkacik, C. G. Callan Jr., W. Bialek, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12265–12270 (2008).
8. A. M. Walczak, G. Tkačik, W. Bialek, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **81**, 041905 (2010).
9. P. Mehta, S. Goyal, T. Long, B. L. Bassler, N. S. Wingreen, *Mol. Syst. Biol.* **5**, 325 (2009).
10. R. C. Yu *et al.*, *Nature* **456**, 755–761 (2008).
11. M. Behar, A. Hoffmann, *Biophys. J.* **105**, 231–241 (2013).
12. R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, A. Levchenko, *Science* **334**, 354–358 (2011).
13. S. Uda *et al.*, *Science* **341**, 558–561 (2013).
14. M. Voliotis, R. M. Perrett, C. McWilliams, C. A. McArdle, C. G. Bowsher, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E326–E333 (2014).
15. N. Hao, B. A. Budnik, J. Gunawardena, E. K. O'Shea, *Science* **339**, 460–464 (2013).
16. A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, *Science* **298**, 1241–1245 (2002).
17. S. D. M. Santos, P. J. Verveer, P. I. H. Bastiaens, *Nat. Cell Biol.* **9**, 324–330 (2007).
18. J. E. Purvis, G. Lahav, *Cell* **152**, 945–956 (2013).
19. T. M. Cover, J. A. Thomas, *Elements of Information Theory 2nd* (Wiley-Interscience, ed. 2, Hoboken, NJ 2006).
20. R. E. C. Lee, S. R. Walker, K. Savery, D. A. Frank, S. Gaudet, *Mol. Cell* **53**, 867–879 (2014).
21. C. Cohen-Saidon, A. A. Cohen, A. Sigal, Y. Liron, U. Alon, *Mol. Cell* **36**, 885–893 (2009).
22. O. E. Sturm *et al.*, *Sci. Signal.* **3**, ra90 (2010).
23. T. Schreiber, *Phys. Rep.* **308**, 1–64 (1999).
24. J. E. Toettcher, O. D. Weiner, W. A. Lim, *Cell* **155**, 1422–1434 (2013).
25. G. von Dassow, E. Meir, E. M. Munro, G. M. Odell, *Nature* **406**, 188–192 (2000).
26. N. Barkai, S. Leibler, *Nature* **387**, 913–917 (1997).

ACKNOWLEDGMENTS

We are indebted to T. Kemp (Univ. of California–San Diego) for discussion of the random Gram matrix theory. Supported by the NIH grants P50-GM085764 (R.W., A.H., L.T.), R01-GM089976 (L.T), and R01-GM071573 (A.H).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1370/suppl/DC1
Supplementary Text
Figs. S1 to S25
Tables S1 to S6
References (*27*–*38*)

17 April 2014; accepted 12 November 2014
10.1126/science.1254933

ONCOGENE REGULATION

An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element

Marc R. Mansour,[1,2] Brian J. Abraham,[3]* Lars Anders,[3]* Alla Berezovskaya,[1] Alejandro Gutierrez,[1,4] Adam D. Durbin,[1] Julia Etchin,[1] Lee Lawton,[3] Stephen E. Sallan,[1,4] Lewis B. Silverman,[1,4] Mignon L. Loh,[5] Stephen P. Hunger,[6] Takaomi Sanda,[7] Richard A. Young,[3,8]† A. Thomas Look[1,4]†

In certain human cancers, the expression of critical oncogenes is driven from large regulatory elements, called super-enhancers, that recruit much of the cell's transcriptional apparatus and are defined by extensive acetylation of histone H3 lysine 27 (H3K27ac). In a subset of T-cell acute lymphoblastic leukemia (T-ALL) cases, we found that heterozygous somatic mutations are acquired that introduce binding motifs for the MYB transcription factor in a precise noncoding site, which creates a super-enhancer upstream of the *TAL1* oncogene. MYB binds to this new site and recruits its H3K27 acetylase–binding partner CBP, as well as core components of a major leukemogenic transcriptional complex that contains RUNX1, GATA-3, and TAL1 itself. Additionally, most endogenous super-enhancers found in T-ALL cells are occupied by MYB and CBP, which suggests a general role for MYB in super-enhancer initiation. Thus, this study identifies a genetic mechanism responsible for the generation of oncogenic super-enhancers in malignant cells.

In cancer cells, monoallelic expression of oncogenes can occur through a variety of mechanisms, including chromosomal translocation, alterations in promoter methylation, parental imprinting, and intrachromosomal deletion (*1*–*3*). A quintessential example is TAL1^{d}, an ~80–kilobase (kb) deletion on chromosome 1p33 that is found in 25% of cases of human T cell acute lymphoblastic leukemia (T-ALL). The deletion results in overexpression of *TAL1*, an oncogene coding for a basic helix-loop-helix transcription factor, by mediating fusion of *TAL1* coding sequences to the regulatory elements of the ubiquitously expressed gene "*SCL*-interrupting locus" (*STIL*) (*4*–*6*). However, we previously reported that a substantial proportion of T-ALLs, including the Jurkat T-ALL cell line, have monoallelic overexpression of *TAL1* but lack either the TAL1^{d} abnormality or a chromosomal translocation of the *TAL1* locus (*7*, *8*).

We hypothesized that cis-acting genomic lesions affecting *TAL1* regulatory sequences might account for monoallelic *TAL1* activation. Chromatin immunoprecipitation (ChIP)–sequencing (ChIP-seq) analysis of Jurkat cells revealed aberrant histone H3 lysine 27 acetylation (H3K27ac), a mark of active transcription, starting upstream of the *TAL1* transcriptional start site and extending across the first exons (Fig. 1A) (*9*, *10*). Regions with such rich and broad H3K27ac marks have been termed super-enhancers (also stretch enhancers or locus control regions) and are commonly found at genes that determine cell identity in embryonic stem (ES) cells and in tumor cells at oncogenes critical for the malignant cell state (*11*–*17*). The super-enhancer encompassing *TAL1* in Jurkat cells was aberrant, in that it was not present in fetal thymocytes, normal CD34+ hematopoietic stem and progenitor cells (HSPCs), or in other T-ALL cell lines, such as TAL1^{d}-positive RPMI-8402 cells and DND-41 T-ALL cells that lack *TAL1* expression (Fig. 1A) (*9*). Of note, chromatin conformation capture experiments recently performed in Jurkat cells identified a looping interaction involving an enhancer site 8 kb upstream of the transcription start site (TSS), which coincides with the locations of both the aberrant super-enhancer and the positive autoregulatory binding sites for members of the TAL1 complex in this cell line (Fig. 1A, red arrow) (*9*, *18*).

Sequencing of the genomic DNA region encompassing this site identified a heterozygous 12–base pair (bp) insertion (GTTAGGAAACGG) that aligned precisely with the TAL1, GATA3, RUNX1, and HEB ChIP-seq peaks (Fig. 1B). Among eight additional TAL1-positive T-ALL cell lines, MOLT-3 cells also harbored an abnormal heterozygous 2-bp insertion (GT) at the same site (Fig. 1B), whereas none of 10

[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA. [2]Department of Haematology, UCL Cancer Institute, University College London, London WC1E 6BT, UK. [3]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [4]Division of Pediatric Hematology-Oncology, Boston Children's Hospital, MA 02115, USA. [5]Department of Pediatrics, Benioff Children's Hospital, University of California San Francisco, CA 94143, USA. [6]Pediatric Hematology/Oncology/BMT, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO 80045, USA. [7]Cancer Science Institute of Singapore, National University of Singapore, and Department of Medicine, Yong Loo Lin School of Medicine, 117599, Singapore. [8]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.
*These authors contributed equally and are listed in alphabetical order. **†Corresponding author. E-mail: thomas_look@dfci.harvard.edu (A.T.L.); young@wi.mit.edu (R.A.Y.)**

Accurate information transmission through dynamic biochemical signaling networks

Jangir Selimkhanov *et al.*

Science **346**, 1370 (2014);
DOI: 10.1126/science.1254933