

acquired later in evolution, this would be an intriguing example of how two regulatory circuits may have become connected in order to optimize properties of *S. aureus* as a pathogen. Note, however, that the direct, AgrA-mediated activation of *psm* genes is much stronger than RNAIII-dependent repression, and that the mechanism of the latter is as yet unknown.

In summary, Queck et al. present for the first time RNAIII-independent target gene regulation of the *agr* locus in *S. aureus*. Their results could allow us a glimpse at how the stepwise occurrence of two different regulatory circuits may have helped *S. aureus* on its way to becoming a leading human pathogen. Similarly, other bacteria may have integrated regulatory RNAs into pre-existing protein-based circuits in or-

der to develop from noninfectious bugs into species that cause much harm. Nonetheless, RNAIII remains an intriguing molecule, since we have only begun to understand its manifold functions in *S. aureus* virulence and the underlying mechanisms of riboregulation.

REFERENCES

- Arvidson, S., and Tegmark, K. (2001). *Int. J. Med. Microbiol.* 291, 159–170.
- Bassler, B.L., and Losick, R. (2006). *Cell* 125, 237–246.
- Benito, Y., Kolb, F.A., Romby, P., Lina, G., Etienne, J., and Vandenesch, F. (2000). *RNA* 6, 668–679.
- Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y., Jacquier, A., et al. (2007). *Genes Dev.* 21, 1353–1366.
- Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C., et al. (2005). *EMBO J.* 24, 824–835.
- Novick, R.P. (2003). *Mol. Microbiol.* 48, 1429–1449.
- Novick, R.P., Ross, H.F., Projan, S.J., Komblum, J., Kreiswirth, B., and Moghazeh, S. (1993). *EMBO J.* 12, 3967–3975.
- Queck, S.Y., Jameson-Lee, M., Villaruz, A.E., Bach, T.-H.L., Khan, B.A., Sturdevant, D.E., Ricklefs, S.M., Li, M., and Otto, M. (2008). *Mol. Cell* 32, 150–158.
- Recsei, P., Kreiswirth, B., O'Reilly, M., Schlievert, P., Gruss, A., and Novick, R.P. (1986). *Mol. Gen. Genet.* 202, 58–61.
- Wang, R., Braughton, K.R., Kretschmer, D., Bach, T.H., Queck, S.Y., Li, M., Kennedy, A.D., Dorward, D.W., Klebanoff, S.J., Peschel, A., et al. (2007). *Nat. Med.* 13, 1510–1514.

Taming Data

Alan Aderem^{1,*} and Ilya Shmulevich¹

¹The Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103, USA

*Correspondence: aderem@systemsbiology.org

DOI 10.1016/j.chom.2008.09.011

A challenge in systems-level investigations of the immune response is the principled integration of disparate data sets for constructing predictive models. InnateDB (Lynn et al., 2008; <http://www.innatedb.ca>), a publicly available, manually curated database of experimentally verified molecular interactions and pathways involved in innate immunity, is a powerful new resource that facilitates such integrative systems-level analyses.

The innate immune system is governed by complex networks of molecular interactions, which when perturbed can result in dysfunction such as inflammatory disease. Systems biology approaches, which incorporate powerful high-throughput measurement technologies with computational analysis and mathematical modeling, are allowing us to study such molecular systems on multiple levels of regulation (Gilchrist et al., 2006; Oda and Kitano, 2006). These include protein-DNA interactions in transcriptional regulatory networks, microRNA regulatory networks, epigenetic processes involving DNA methylation and chromatin remodeling, protein-protein interactions in signaling networks, as well as autocrine and

paracrine communication networks between cells. The construction of mathematical models, inferred from such global data sets measured dynamically over time and integrated with information available in public databases, will make it possible to predict cellular behavior under new conditions or environmental perturbations and, ultimately, will enable the development of systematic intervention strategies for reversing pathophysiological effects.

One of the most significant challenges in systems-level investigations of the immune response is the principled integration of disparate types of data for constructing such predictive models. The availability of information about molecular interactions and pathways is critical for

addressing this challenge as it can (1) significantly reduce the search spaces for data-driven network reconstruction since the inferred dynamical rules of interaction can be constrained or biased toward the known network structures; (2) mitigate the sample size requirements for network inference since prior information about the connectivity is available; and (3) facilitate the interpretation of complex data sets, which can be visualized in the context of such pathways. In a recent issue of *Molecular Systems Biology*, Lynn et al. (2008) describe a powerful new resource that facilitates such integrative systems-level analyses: InnateDB (<http://www.innatedb.ca>) is a publicly available, manually curated database of

experimentally verified human and mouse molecular interactions and pathways involved in innate immunity.

InnateDB contains more than 3500 curated innate immunity-relevant interactions involving around 1000 genes. In addition, to improve coverage of the innate immunity interactome, over 100,000 known interactions and approximately 2500 pathways from major public databases are also included and intelligently grouped to avoid redundancy between databases. The manually curated data are richly annotated with a variety of useful details, including supporting publications, interaction detection methods, interaction types, information about cells or tissues, and other contextual annotations. Where available, such interaction annotations from other databases were also imported into InnateDB. Importantly, all interaction data in InnateDB are available for download in a standard format, allowing computational biologists to analyze the interaction data using the tools of their choice and integrate such analyses with other measurement data. Additionally, detailed orthology predictions have been integrated into InnateDB in order to facilitate comparative network analysis of human, mouse, and cow.

InnateDB interacts with the user using a flexible web-based interface that offers searching by molecules, interactions, or pathways and provides lists of synonyms and external identifiers for genes and proteins. More complex queries are possible with Boolean operators, and molecular interactions can be searched by interaction type, cell or tissue type, and other annotations. The user can also search for genes that belong to a particular pathway along with all the genes that interact with them. Such searches can then be visualized using the Cerebral plug-in (Barsky et al., 2007) for Cytoscape (Shannon et al., 2003). This useful tool, also created by the InnateDB team, can visualize a network (e.g., the result of a pathway search), using a localization-based layout whereby the nodes of the network are arranged according to their subcellular localization, such as the nucleus, cytoplasm, cell surface, and so on. Furthermore, if expression data are also provided, they can be superimposed on the network by coloring the nodes in accordance with the expression levels of the corresponding genes or proteins.

The creators of InnateDB have clearly recognized the need for a flexible integrated analysis framework in order to achieve the promise of systems approaches to understanding immunity. Indeed, in addition to the highly useful Cerebral plug-in, which contains several standard analysis tools, such as *k*-means clustering, InnateDB also offers basic pathway or gene ontology enrichment analysis tools for identifying particular pathways or ontology terms that are enriched in differentially expressed gene sets. Although these analysis capabilities are useful and can generate novel insights, one can readily envision the need for more sophisticated analytical approaches that may include other data types or modeling formalisms. These could include integration of interaction and pathway information with transcription factor binding site prediction algorithms (Lähdesmäki et al., 2008) or transcription factor localization measurements using high-throughput technologies such as ChIPseq, proteomic measurements, microRNA regulatory networks and their expression measurements, as well as statistical and mathematical approaches for reconstructing and simulating biomolecular networks (Bonneau et al., 2007; Ramsey et al., 2008).

These approaches undoubtedly require flexible and adaptable software architectures that can support the rapid development of integrated tools for analyzing heterogeneous data typical in systems biology (Boyle et al., 2008). Such software architecture is being used and further developed as part of a large NIAID-funded international consortium focusing on systems approaches to immunology (<http://www.systemsimmunology.org>). The data generated by the project, which include genomic (SNPs and forward genetic methods using ENU mutagenesis), transcriptomic (genome-wide exon-level expression), proteomic (membrane-bound or secreted proteins using mass spectrometry or phospho-signaling using multiparameter flow cytometry), and interaction data (protein-DNA interactions measured using ChIP-chip or ChIPseq), are made available through a web-based data portal that provides access to the raw and processed data as well as to higher-level analysis results (Ramsey et al., 2008; Korb et al., 2008). Ultimately, seamless integration of the data in this portal with InnateDB's interaction and

pathway information would enable more powerful analyses to be carried out, thus yielding new insights into the nature of the molecular mechanisms by which the immune system responds to infectious disease by inciting innate inflammatory reactions and instructing adaptive immune responses. A number of large-scale programs directed at a systems-wide understanding of immunity are currently being funded by the NIH. These include programs on innate immunity (<http://www.systemsimmunology.org>), signal transduction (<http://www.signaling-gateway.org>), lipidomics (<http://www.lipidmaps.org>), glycomics (<http://www.functionalglycomics.org>), inflammation (<http://www.gluegrant.org>), and the Immunology Database and Analysis Portal (ImmPort) (<https://www.immport.org>). In addition, the Bill and Melinda Gates Foundation have funded a number of large-scale vaccine initiatives (<http://www.cavd.org>). All of these programs are generating a mountain of data that need to be stored and analyzed. InnateDB promises to be an important brick in the wall.

REFERENCES

- Barsky, A., Gardy, J.L., Hancock, R.E., and Munzner, T. (2007). *Bioinformatics* 23, 1040–1042.
- Bonneau, R., Facciotti, M.T., Reiss, D.J., Schmid, A.K., Pan, M., Kaur, A., Thorsson, V., Shannon, P., Johnson, M.H., Bare, J.C., et al. (2007). *Cell* 131, 1354–1365.
- Boyle, J., Cavnor, C., Killcoyne, S., and Shmulevich, I. (2008). *BMC Bioinformatics* 9, 295.
- Gilchrist, M., Thorsson, V., Li, B., Rust, A.G., Korb, M., Roach, J.C., Kennedy, K., Hai, T., Bolouri, H., and Aderem, A. (2006). *Nature* 441, 173–178.
- Korb, M., Rust, A.G., Thorsson, V., Battail, C., Li, B., Hwang, D., Kennedy, K.A., Roach, J.C., Rosenberger, C.M., Gilchrist, M., et al. (2008). *BMC Immunol.* 9, 7.
- Lähdesmäki, H., Rust, A.G., and Shmulevich, I. (2008). *PLoS ONE* 3, e1820. 10.1371/journal.pone.0001820.
- Lynn, D.J., Winsor, G.L., Chan, C., Richard, N., Laird, M.R., Barsky, A., Gardy, J.L., Roche, F.M., Chan, T.H., Shah, N., et al. (2008). *Mol. Syst. Biol.* 4, 218.
- Oda, K., and Kitano, H. (2006). *Mol. Syst. Biol.* 2, 2006.0015.
- Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., Li, B., Gilchrist, M., Gold, E.S., Johnson, C.D., Litvak, V., et al. (2008). *PLoS Comput. Biol.* 4, e1000021. 10.1371/journal.pcbi.1000021.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). *Genome Res.* 13, 2498–2504.