

# Wisdom of crowds for robust gene network inference

Daniel Marbach<sup>1,2,11</sup>, James C Costello<sup>3-5,11</sup>, Robert Küffner<sup>6,11</sup>, Nicole M Vega<sup>3-5</sup>, Robert J Prill<sup>7</sup>, Diogo M Camacho<sup>3-5,10</sup>, Kyle R Allison<sup>3-5</sup>, The DREAM5 Consortium<sup>8</sup>, Manolis Kellis<sup>1,2</sup>, James J Collins<sup>3-5,9</sup> & Gustavo Stolovitzky<sup>7</sup>

Reconstructing gene regulatory networks from high-throughput data is a long-standing challenge. Through the Dialogue on Reverse Engineering Assessment and Methods (DREAM) project, we performed a comprehensive blind assessment of over 30 network inference methods on *Escherichia coli*, *Staphylococcus aureus*, *Saccharomyces cerevisiae* and *in silico* microarray data. We characterize the performance, data requirements and inherent biases of different inference approaches, and we provide guidelines for algorithm application and development. We observed that no single inference method performs optimally across all data sets. In contrast, integration of predictions from multiple inference methods shows robust and high performance across diverse data sets. We thereby constructed high-confidence networks for *E. coli* and *S. aureus*, each comprising ~1,700 transcriptional interactions at a precision of ~50%. We experimentally tested 53 previously unobserved regulatory interactions in *E. coli*, of which 23 (43%) were supported. Our results establish community-based methods as a powerful and robust tool for the inference of transcriptional gene regulatory networks.

successfully used to address many biological problems<sup>8-11</sup>, yet when applied to the same data, they can generate disparate sets of predicted interactions<sup>2,3</sup>.

Understanding the advantages and limitations of different network inference methods is critical for their effective application in a given biological context. The DREAM project is a framework to enable such an assessment through standardized performance metrics and common benchmarks<sup>12</sup> (<http://www.the-dream-project.org/>). DREAM is organized around annual challenges, whereby the community of network inference experts is solicited to run their algorithms on benchmark data sets, participating teams submit their solutions to the challenge and the submissions are evaluated<sup>12-14</sup>.

Here we present the results for the transcriptional network inference challenge from DREAM5, the fifth annual set of DREAM systems biology challenges. The community of network inference experts was invited to infer genome-scale transcriptional regulatory networks from gene-expression microarray data sets for a prokaryotic model organism (*E. coli*), a eukaryotic model organism (*S. cerevisiae*), a human pathogen (*S. aureus*) and an *in silico* benchmark (Fig. 1).