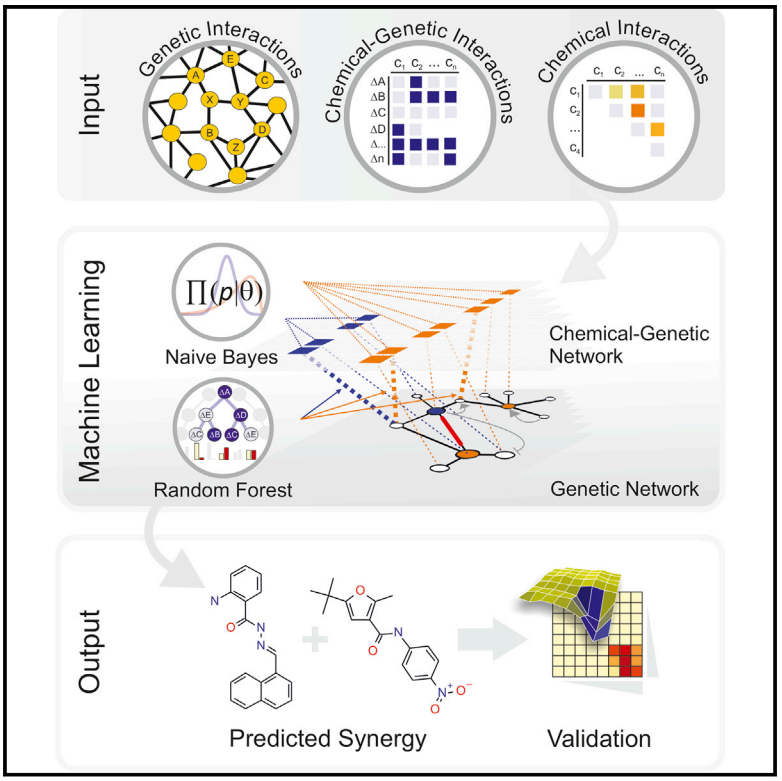
Article

Cell Systems

Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning

Graphical Abstract



Authors

Jan Wildenhain, Michaela Spitzer, Sonam Dolma, ..., David S. Bellows, Gerard D. Wright, Mike Tyers

Correspondence

md.tyers@umontreal.ca

In Brief

A general experimental and machine learning strategy is developed to predict synergistic compound combinations from chemical-genetic interaction data and chemical structural features.

Highlights

- A chemical-genetic interaction matrix (CGM) of compounds against genotypes
- A systematic chemical interaction matrix between genotype-specific inhibitors
- Machine learning models of structural features and CGM interactions that predict synergism
- Synergistic combinations that exhibit species-selective effects against pathogenic fungi

Wildenhain et al., 2015, Cell Systems *1*, 383–395
December 23, 2015
http://dx.doi.org/10.1016/j.cels.2015.12.003

CellPress

Cell Systems
Article

Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning

Jan Wildenhain,[1,5] Michaela Spitzer,[1,2,5] Sonam Dolma,[3,6,7] Nick Jarvik,[3,8] Rachel White,[1,9] Marcia Roy,[1,10] Emma Griffiths,[2,11] David S. Bellows,[3,12] Gerard D. Wright,[2] and Mike Tyers[1,4,*]

[1]Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK
[2]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON L8N 3Z5, Canada
[3]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada
[4]Institute for Research in Immunology and Cancer, Department of Medicine, Université de Montréal, Montréal, QC H3C 3J7, Canada
[5]Co-first author
[6]Present address: Arthur and Sonia Labatt Brain Tumor Research Centre, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada
[7]Present address: Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 1A1, Canada
[8]Present address: Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada
[9]Present address: Edinburgh Cancer Research Centre, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XR, UK
[10]Present address: Centre for Clinical Brain Sciences and Centre for Neuroregeneration, The University of Edinburgh, Edinburgh EH16 4SB, UK
[11]Present address: Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
[12]Present address: School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand
*Correspondence: md.tyers@umontreal.ca
http://dx.doi.org/10.1016/j.cels.2015.12.003

SUMMARY

The structure of genetic interaction networks predicts that, analogous to synthetic lethal interactions between non-essential genes, combinations of compounds with latent activities may exhibit potent synergism. To test this hypothesis, we generated a chemical-genetic matrix of 195 diverse yeast deletion strains treated with 4,915 compounds. This approach uncovered 1,221 genotype-specific inhibitors, which we termed cryptagens. Synergism between 8,128 structurally disparate cryptagen pairs was assessed experimentally and used to benchmark predictive algorithms. A model based on the chemical-genetic matrix and the genetic interaction network failed to accurately predict synergism. However, a combined random forest and Naive Bayesian learner that associated chemical structural features with genotype-specific growth inhibition had strong predictive power. This approach identified previously unknown compound combinations that exhibited species-selective toxicity toward human fungal pathogens. This work demonstrates that machine learning methods trained on unbiased chemical-genetic interaction data may be widely applicable for the discovery of synergistic combinations in different species.

INTRODUCTION

The modern era of drug discovery has been dominated by the "magic bullet" concept developed by Ehrlich more than 100 years ago (Strebhardt and Ullrich, 2008). This concept is predicated on the notion that a pathogen, genetic mutation, or physiological defect can be remedied by a single chemical agent. This approach has proven successful in the development of anti-infective agents and specific enzyme inhibitors to treat particular conditions. However, despite investments in genome-based approaches for target identification, validation, and screening, the current repertoire of approved drugs targets stands at only ~500 proteins, with at most only a further 500 targets under active exploration (Overington et al., 2006; Rask-Andersen et al., 2014). The druggable genome has been focused on a select number of enzyme and receptor target classes, suggesting that only a minor fraction of the potential target space has been tapped to date. Cell-based phenotypic screens explore all possible targets but with the caveat that the mechanism of action can be difficult if not impossible to discern (Nijman, 2015).

It has long been evident that the development, physiology, and phenotype of an organism are controlled by complex genetics (Waddington, 1957). Recent systematic genetic screens have uncovered the depth of this complexity, which manifests as a vast network of genetic interactions (Costanzo et al., 2010). A genetic interaction between two genes is observed when a phenotype caused by a mutation in the first gene is exacerbated by a mutation in the second gene, such that the combined effect exceeds the sum of each individual effect (Mani et al., 2008). In the extreme case, the synthetic combination of two or more non-lethal mutations may result in a lethal genetic interaction (Dobzhansky, 1946). Systematic genetic screens in budding yeast have revealed that while only ~1,000 of the ~6,000 genes are essential at least 200,000 documented genetic interactions occur between non-essential genes or non-lethal alleles of essential genes (Costanzo et al., 2010; Chatr-Aryamontri et al., 2015). This dense genetic architecture reflects the global hierarchy of cellular subsystems and allows the cell to coordinate